Cotranscriptional folding of the lncRNA Xist A-repeats indicates a modular structure

Alisha N. Jones[1,2*], Frank Gabel[3], Stefan Bohn[1], Gregory Wolfe[1], Michael Sattler[1,2,*]

[1] Institute of Structural Biology, Helmholtz Zentrum München, Neuherberg, Germany

[2] Bavarian NMR Center, Department of Chemistry, Technical University of Munich, Garching, Germany

[3] Univ. Grenoble Alpes, CEA, CNRS, IBS, Grenoble 38042, France

* To whom correspondence should be addressed:
sattler@helmholtz-muenchen.de, alisha.jones@helmholtz-muenchen.de

Keywords: Xist, lncRNA, structure, SHAPE, NMR, SAXS, cryoEM

Abstract

The A-repeat region of the lncRNA Xist is required for X-chromosome inactivation in placental mammals and comprises 8.5 copies of a conserved 24-nucleotide motif separated by uracil-rich linkers. Several experimentally-supported secondary structures have been proposed for the A-repeats. Here, we consider the effects of cotranscriptional folding and present a comprehensive biochemical and structural analysis of the Xist A-repeats. We show that under native-like conditions the A-repeat region folds into two independent subdomains. The repeats are modularly arranged, forming stable AUCG tetraloop hairpins and inter-repeat dimers. The similarity of chemical probing profiles in cells and *in vitro* argues that this structure represents the predominant native conformation of the Xist A-repeat *in vivo*. However, the A-repeat conformation is dynamic and affected by buffer conditions, rationalizing the disparity between previously proposed structures. Our results provide a starting point for studying the interaction with RNA binding proteins and molecular mechanisms linked to X-inactivation. Our integrative approach establishes a pipeline for investigating cotranscriptional pathways of large RNAs and their structures.

Introduction

Long noncoding RNAs (lncRNAs) are non-protein-encoding transcripts with lengths of at least 200 nt, with some spanning up to several thousands of nucleotides (Kung, Colognori, and Lee 2013). lncRNAs are emerging as important regulators of gene expression and cellular functions, as transcriptional regulators at the level of chromatin, as scaffolding platforms, and as guides for RNP localization (Arun, Diermeier, and Spector 2018; Chaumeil, Baccon, and and E. Heard 2006; Diermeier et al. 2016; Lewandowski et al. 2019; Ponting, Oliver, and Reik 2009; Wang and Chang 2011; Wutz, Rasmussen, and Jaenisch 2002). The molecular functions of lncRNAs are expected to depend on local and global structures and dynamic conformations, which are modulated by RNA binding proteins (Chu et al. 2015; Jones et al. 2020; Kim et al. 2016, 2020; Somarowthu et al. 2015; Uroda et al. 2019). Of the known ~50,000 lncRNAs in humans, to date, fewer than 100 have been studied on a structural level. One of those studied is the lncRNA Xist (X-inactive specific transcript), a 17 kb long transcript that is required for dosage compensation in placental mammals (Loda and Heard 2019; Penny et al. 1996; Da Rocha and Heard 2017; Wutz et al. 2002). The primary sequence of the Xist lncRNA consists of six interspersed, repetitive domains: A, F, B, C, D, and E. Together, the A and F repeats comprise a transcript known as Rep A (Allaman-Pillet et al. 2000; Liu, Somarowthu, and Pyle 2017; Pintacuda, Young, and Cerase 2017; Wutz et al. 2002). Each of these six repeat regions contributes to the transcriptional silencing of the inactive X chromosome; a process known as XCI. The A-repeats, in a domain located at the 5' end of the transcript that spans approximately 450 nucleotides, perform one of the more critical roles in XCI: in their absence, Xist can coat, but not transcriptionally silence, the inactive X chromosome (Allaman-Pillet et al. 2000; Wutz et al. 2002).

Analysis of the A-repeat primary sequence revealed a 24-nucleotide conserved element that is repeated 7.5 or 8.5 times in mouse or human, respectively (Wutz et al. 2002). The conserved elements, designated 'A-repeats', are separated by uracil-rich sequences. An A-repeat was initially predicted to form two hairpins. However, structural analysis by NMR revealed that only the major hairpin is thermodynamically stable and forms a novel AUCG tetraloop structure, while the second hairpin dimerizes and can mediate duplex formation with other regions in the A-repeat RNA (Duszczyk et al. 2011; Duszczyk, Zanier, and Sattler 2008; Wutz et al. 2002). The secondary structure of the Xist A-repeats has been extensively investigated both in cells and *in vitro* and more than fifteen secondary structures have been proposed, based on various chemical probing and enzymatic cleavage assays (Duszczyk et al. 2011; Fang et al. 2015; Liu et al. 2017; Lu et al. 2016; Maenner et al. 2010; Ponti et al. 2018; Smola et al. 2016). These structural models suggest that the RNA adopts a secondary structure that is either modular (where the A-repeats form repetitive duplex-containing structures) or non-modular (where each repeat adopts a different fold) (Jones and Sattler 2019).

The variety of proposed secondary structures suggests that the Xist A-repeat region is dynamic and may sample multiple, dynamically interconverting conformations (Aguilar et al. 2022; Kim et al. 2020) that may be influenced by buffer conditions (Jones and Sattler 2019). This is reasonable, considering that RNA structures tend to be more dynamic and often utilize conformational dynamics to regulate function (Barnwal et al. 2016; Dallaire et al. 2016; Dethoff et al. 2012; Wang et al. 2021). In addition, the environmental/buffer conditions under which the RNA is studied, whether *in vitro* or in cells, may influence the structure and dynamics of RNAs. For RNAs studied

in vitro, purification of the transcript under denaturing conditions requires refolding of the RNA, which can lead to mis-folding, especially for longer RNAs (Chillón et al. 2015). Even when purifying the RNA under native conditions, buffer composition (i.e., salt concentration, pH) and temperature can impact the structure adopted by the RNA. While in-cell chemical probing of the RNA secondary structure occurs under native conditions, RNAs can adopt multiple folded states at any given time in the cell, leading to structural information that reflects an average of structural states (Morandi et al. 2021; Tomezsko et al. 2020). The secondary structural data obtained for the Xist A-repeats, having been studied both *in vitro* and in cells, are subject to the limitations of each of these approaches.

Selective 2' hydroxyl acylation analyzed by primer extension (SHAPE) experiments carried out on the Xist lncRNA in cells and *in vitro* reveals very similar reactivity profiles, suggesting that the structure analyzed *in vitro* also reflects the cellular context, with differences being attributed to protein binding (Smola et al. 2016; Smola, Calabrese, and Weeks 2015). Here, we report a comprehensive study of the secondary and tertiary structure of the human Xist A-repeats using an integrative approach combining SHAPE, NMR (nuclear magnetic resonance), SAXS (small angle X-ray scattering), and cryo-electron microscopy (cryo-EM). To assess the effects of cotranscriptional folding, we analyze various fragments of the A-repeat, mimicking different steps during the transcriptional synthesis of the RNA by the polymerase in cells. Our data show that the Xist A-repeats fold locally to form stable, structured elements comprising the previously reported AUCG tetraloops, and that the individual repeats are involved in dimerization with consecutive repeats by inter-repeat duplex formation via the 3' region. Our findings support a modular structure of the A-repeats that can adopt an overall tertiary fold by interactions involving the different repeats and the connecting linker regions.

Results

Xist A-repeat RNA structure is influenced by buffer composition in vitro

Considering that a range of divergent structural models have been proposed for the A-repeat RNA, we first analyzed the effects of experimental conditions. The secondary structure of the Xist A-repeats has been investigated *in vitro* by both chemical probing and enzymatic cleavage under four different buffer conditions, each resulting in a different secondary structure prediction (Duszczyk et al. 2011; Liu et al. 2017; Maenner et al. 2010; Smola et al. 2016) (**Table 1**). The most distinct differences between these buffer conditions concern the pH (ranging from 6.0 – 8.0) and salt content (100 to 150 mM KCl or NaCl, while $MgCl_2$ is absent or ranges from 3.25 mM to 15 mM), two factors which are known to influence how an RNA folds (Chen and Pollack 2016; Fürtig et al. 2003; Liu et al. 2017). We therefore prepared the Xist A-repeat RNA under the conditions described in each of the four studies, and used SAXS, NMR, and cryo-EM to study the effect of buffer conditions on the overall conformation of the RNA. Hereafter, these four conditions are referred to by the last name of the first author of each study: Maenner et al., Liu et al., Duszczyk et al., and Smola et al.

We first used SAXS to investigate the global conformation, intrinsic flexibility and degree of folding of the Xist A-repeats under each of the four conditions at 298 K. Scattering profiles were recorded at three RNA concentrations between 0.20 and 2.1 mg/mL (**Supplementary Table 1**). The

concentration series showed no inter-particle effects, and thus the highest concentration data were used for analysis (**Supplementary Figure 1**). Guinier analysis reveals that the radius of gyration (R_g), which is a measure of a molecule's compactness, varies for the Xist A-repeats depending on buffer composition (**Figure 1a, Table 2**) . The Xist A-repeat RNA is most compact under the buffer conditions used by Duszczyk et al., with an R_g value of 7.56 ± 0.02 nm (**Table 2**). The R_g of the Xist A-repeats increases for the Maenner et al., Smola et al., and Liu et al. buffer conditions with R_g values of 8.89 ± 0.18 nm, 9.83 ± 0.09 nm, and 12.10 ± 0.21 nm respectively, indicating a reduced compactness of the overall conformational ensemble in solution (**Table 2**). It was previously shown that the compactness of the RepA element of Xist (which is comprised of both the A and F repeats), as measured through the hydrodynamic radius (R_H), decreases with an increase in $MgCl_2$ concentration (Liu et al. 2017). By these standards, one might expect a similar trend to be observed for the four buffer conditions assayed in our SAXS studies: the Duszczyk buffer possesses no $MgCl_2$, and the concentration of $MgCl_2$ in the Maenner, Smola, and Liu buffers is 3.25 mM, 10 mM, and 15 mM, respectively. The R_g values we obtained do not, in fact, reflect a corollary relationship; however, because numerous components of the buffers vary for each of the buffers, it is not unexpected that we find no clear correlations between the buffer components and compactness of the Xist A-repeat RNA.

To assess the intrinsic flexibility of the Xist A-repeat RNA as a function of buffer condition, we generated dimensionless Kratky plots (**Figure 1b**) (Durand et al. 2010). For each of the buffer conditions, the dimensionless Kratky plot shows a profile characteristic of a partially compact particle with extended flexible elements. This is indicated by the steady increase of the curve as q^*R_g increases through $q^*R_g = 10$, followed by a plateau after a partial decrease at intermediate values. This result is not surprising, as lncRNAs are expected to be conformationally flexible (Kim et al. 2020). Interestingly, the dimensionless Kratky plots for each of the buffer conditions are not congruent, indicating structural differences depending on buffer conditions.

We investigated the degree of folding of the Xist A-repeats in each buffer by assessing the maximum pairwise distance (D_{max}) from the *P(r)* pair distance distribution function (**Figure 1c**). The D_{max} is the maximum extended length within the RNA. The Xist A-repeat transcript used in our experiments is 467 nucleotides in length. Considering that the distance between each base in an extended RNA is approximately 0.27 nm, a completely unfolded Xist A-repeat RNA transcript would correspond to a D_{max} of 126 nm. Regardless of buffer composition, each of the proposed Xist A-repeat RNA structural models is predicted to fold into an intricately structured RNA (Jones and Sattler 2019). As such, we anticipated a much shorter D_{max} for the folded Xist A-repeat RNA. Indeed, shorter and comparable D_{max} values are observed for the conditions used by Maenner et al., Smola et al., and Duszczyk et al., reflecting maximum extended lengths of 35, 33, and 30 nm, respectively. The D_{max} length for the Liu et al. buffer condition is 43 nm, suggesting the Xist A-repeat RNA is more loosely structured. Altogether, these results support the notion that the Xist A-repeats, regardless of buffer composition, possess a high fraction of structured domains.

We next used single particle cryo-EM to obtain a complementary analysis of the RNA dimensions in the various buffer conditions. Observation and analysis of individual Xist A-repeat molecules in vitreous ice reveals that the RNA adopts a partially elongated structure with several branched domains originating from the main helical branch (**Figure 2a**). Analysis of the lengths of the helical stems and branches from ~6,500 particles reveals a length distribution of 5 to 40 nm for all buffered conditions (**Figure 2b**). The particle length distribution is in excellent agreement with the

P(r) distribution curves obtained from SAXS analysis. A sub-analysis of the longest chain for a subset of approximately 100 particles reveals that the Liu et al. buffered condition results in the most extended RNA particles, consistent with D_{max} values obtained from SAXS (**Figure 2b**, inset). Due to the intrinsic flexibility of the Xist A-repeats, we were unable to obtain high-resolution 3D classifications.

As analysis of the Xist A-repeat RNA SAXS data suggested the overall RNA conformational ensemble may vary depending on the buffer condition, we used ^{1}H NMR to assess the presence of local RNA structure through observation of the imino proton resonances (**Figure 2c**). Imino protons of guanosines (H1) and uracils (H3) that are involved in base pair interactions are observable with chemical shifts between 10-15 ppm. The imino resonance profiles of the Xist A-repeats are not starkly different for the buffer conditions utilized by Smola et al., Liu et al., and Duszczyk et al. (note that, due to the glycerol present in the buffer used by Maenner et al., we were unable to obtain high quality NMR spectra). Two clusters of imino signals between 12.0-13.5 ppm correspond to G-C base pairs and are observed for Smola et al., Duszczyk et al., and Liu et al. conditions. Only minor differences are seen for the imino signals between each buffer, and the overall similarities are not surprising, as the chemical probing and enzymatic cleavage reactivity profiles reported by Smola et al., Liu et al., and Maenner et al. are also relatively similar (**Figure 2d, Supplementary File 1**). Taken together, the overall similarity of the imino NMR signal footprints indicates the presence of similar base pairing and folded RNA regions. However, the SAXS analyses reveal that the Xist A-repeats exhibit distinct compactness under varied buffer conditions, suggesting different conformational ensembles and arrangements of individual local structural elements.

Co-transcriptional folding reveals a modular A-repeat arrangement

Large RNAs can fold while being transcribed by RNA polymerase, a process known as cotranscriptional folding (Lai, Proctor, and Meyer 2013; Watters et al. 2016). Relative to *in vitro* studies, where the focus is generally on initiating refolding of fully transcribed RNA transcripts in a near physiological buffer, cotranscriptional RNA folding studies better replicate how an RNA folds under cellular conditions. Our SAXS data indicate that buffer composition strongly impacts the fold adopted by the Xist A-repeat RNA. Therefore, we investigated the cotranscriptional pathway of the Xist A-repeats by monitoring the folding of secondary structure elements using SHAPE chemical probing and NMR; and inspected the overall conformational ensemble using SAXS. For this, we arrested RNA transcripts at various points in their transcription in order to gain insight into the hierarchical folding patterns adopted by the Xist A-repeat RNA.

We first used SHAPE to probe the structure of the Xist A-repeats while they are being transcribed, to determine the folding adopted by the RNA. We generated DNA templates to yield eight Xist A-repeat constructs that extend from the 5' end of the RNA through the end of each A-repeat–Xist A-rep1 to Xist A-rep8–and probed with 1M7 immediately post-transcription at 310 K. Using this approach, the RNA is expected to be probed in its native folding state. Inspection of the SHAPE reactivity profiles for each arrested A-rep RNA shows that nucleotides of the uracil-rich region exhibit increased SHAPE reactivities relative to nucleotides that comprise the A-repeat elements. This indicates the uracil-rich linkers are single stranded. Specific to the A-repeats, we find increased SHAPE reactivity for nucleotides that comprise the AUCG tetraloop of each major hairpin (**Figure 3a**, aligned with blue boxes). Residues corresponding to the apical loop of the

second, so-called 'minor hairpin' (**Figure 3a**, aligned with green boxes) show little to no SHAPE reactivity relative to the AUCG nucleotides. These data suggest that the A-repeats adopt a stable secondary structure, consistent with the stable AUCG stem loop structure formed by the major hairpin and the minor hairpin being involved in secondary structure. Strikingly, a comparison of the SHAPE reactivity profiles of consecutive extended A-repeat RNA fragments reveals that the structured elements adopted by the Xist A-repeats are locally folded, as the SHAPE reactivity profile at the 5' end of the RNA does not change when additional nucleotides are appended to the 3' end of the transcript. The 3' ends of each RNA show increased SHAPE reactivity due to intrinsic flexibility, thereby indicating poorly folded nucleotides. Upon extension of the transcript, new structured domains are formed among the appended nucleotides (**Figure 3a**).

It is noteworthy that the SHAPE reactivity profiles for the eight 24-nucleotide A-repeats are highly similar, with the linkers between them showing increased SHAPE reactivity. We previously reported that the 24-nucleotide A-repeats dimerize with subsequent A-repeats via the 3' region of each repeat (Duszczyk et al. 2008), confirmed by imino NMR spectra that show base pairing for the 3' region. We thus collected 1D imino ^{1}H NMR experiments for each of the extended A-rep transcripts to analyze the base pairing in the context of these larger regions (**Figure 3b**). Indeed, the pattern of imino peaks observed for the duplex formed by a single A-repeat dimer is consistently maintained as the length of the RNA increases. For example, imino signals of an RNA comprising two repeats are highly similar and show the presence of U-A base pairs in the duplex regions formed by dimerization of the 3' region of the repeats (**Figure 3b**, bottom). This is also seen for larger RNA constructs, comprising 3-8 repeats, although spectral quality is challenged by line-broadening due to the molecular weight of the larger constructs. These data further support the analysis of the SHAPE data and confirm that the A-repeats fold to form repetitive structures of AUCG tetraloop hairpins that involve dimerization with subsequent A-repeat elements.

We used SAXS to investigate the conformational properties of each Xist A-repeat RNA cotranscriptional fragment (**Figure 3c, Supplementary Table 2**)**.** As a control, we carried out SAXS on the 14-mer AUCG tetraloop hairpin previously determined by NMR (Duszczyk et al. 2011). CRYSOL analysis (Svergun, Barberato, and Koch 1995) reveals a χ^2 value of 1.24 for the NMR structure (PDB ID 2y95, (Duszczyk et al. 2011)), indicating excellent agreement between the structure and the SAXS data (**Supplementary Figure 2**). We then analyzed SAXS data for the arrested RNA transcripts. Superposition of the SAXS curves of each stalled RNA reveals an expected increase in I(q) with increasing length of the RNA (**Figure 3c**). Along the same lines, a steady increase in the R_g is found as the length of the RNA increases (**Figure 3c,** inset, **Table 3**). Comparison of the dimensionless Kratky plot suggests a compact, globular conformation for the 14-mer major hairpin, the 24-mer duplex-forming A-repeat, and A-reps 1 and 2 (**Figure 3d**). However, as the length of the RNA increases through A-rep3, the Kratky plot shows characteristics of an extended and flexible molecule. This is indicated by the steady increase of the curve as q^*R_g increases. This trend is also observed for A-rep5 and A-rep6. Notably, these arrested RNA constructs also show high SHAPE reactivity at the 3' end of the RNA (**Figure 3a**), indicating the presence of regions with high flexibility. On the other hand, A-rep4 and A-rep8 have similar dimensionless Kratky profiles, where the curves come down at the highest R_g values, indicating the RNAs are rather compact. Consistent with this, the SHAPE reactivity at the 3' end of these RNA is relatively lower than what is observed for A-rep5 and A-rep6. These data suggest

that the first 4 and the full-length A-repeats may fold to form a compact structure, whereas intermediates are more flexible, likely due to poor folding of the 3' ends of the RNAs. This may indicate the presence of two more globular domains formed by 4 repeats that assemble in the full A-repeat RNA structure to form an overall globular fold comprised of repetitive inter-repeat dimers.

A structural model for cotranscriptionally folded Xist A-repeat RNA structure

Our SHAPE and NMR data suggest that the Xist A-repeats form a duplex structure, where the major hairpin of the first A-repeat is stably formed and coaxially stacked (through the duplex formed by the 'minor hairpins') over the major hairpin from the sequential A-repeat (Duszczyk et al. 2011). We used RNAComposer (Popenda et al. 2012) to generate a tertiary structure of two A-repeats forming a coaxially stacked duplex. Using CRYSOL, we evaluated the agreement of the RNAComposer-generated structural model with the experimental scattering curve obtained for the 24-mer from SAXS. The duplexed 24-mer shows a $\chi^2 = 9.3$ (**Figure 4a**), while a monomeric 24-mer (structure generated in RNAComposer) shows a $\chi^2 = 41.8$ (**Figure 4b**). The relative comparison of the agreement with the experimental SAXS data supports the duplex being a folding unit. The rather larger χ^2 value (9.26) likely reflects a poor prediction of the structure of the duplex by RNAComposer. DAMMIF *ab initio* bead modeling revealed a shape that is consistent with the coaxially stacked duplex (**Figure 4c**). Based on these data, alongside our NMR and SHAPE analysis, we propose the Xist A-repeat RNA folds by coaxial stacking of dimeric repeats, each comprising two AUCG tetraloop hairpins, and that the Xist A-repeat RNA overall adopts a modular structural arrangement, where two larger domains made of four repeats each form a compact higher order structure (**Figure 4d**).

The cotranscriptional folding of the A-repeat can be derived considering that nucleotides located at the 5' end of the RNA remain folded despite the addition of nucleotides to the 3' end of the transcript. To this end, we developed a pipeline to generate folding restraint files that build upon the structure adopted by the Xist A-repeats upon the addition of 3' nucleotides. These profiles were then used to fold each of the Xist A-rep arrested RNAs (**Supplementary Figure 3**). Briefly, the Xist A-rep2 RNA secondary structure was predicted using RNAstructure based on SHAPE and NMR data (D. H. Mathews n.d.); SHAPE reactivity data was included as soft folding restraints and unambiguous base pairs identified by NMR were set as hard folding restraints, in particular for the AUCG tetraloop hairpin and dimerization region. Then, to determine the structure of the Xist A-rep3 RNA, the SHAPE reactivity profiles of Xist A-rep3 and A-rep2 RNA were compared using a sliding window approach; for stretches of at least four nucleotides that show comparable SHAPE reactivities between the two RNAs, their corresponding structured elements (i.e., base pairs that are maintained) were added as hard restraints in the folding restraint file for A-rep3. For nucleotides of Xist A-rep3 that showed variable SHAPE reactivities relative to Xist A-rep2, SHAPE data (as soft restraints) was used to guide folding of that region of the RNA. The Xist A-rep3 RNA was then folded using the curated folding restraint file, and this process was iterated through the 8 arrested RNA Xist A-rep constructs. The resulting cotranscriptional folding pathway of the Xist A-repeats is illustrated in **Figure 5**. Here, we find that A-repeats 1 and 2, 3 and 4, 5 and 6, and 7 and 8 form duplex structures, as illustrated in **Figure 4d**.

Discussion

The lncRNA Xist is one of the most well-studied long non-coding RNAs to date. Considering the critical role that the A-repeats have in transcriptional silencing, it is not surprising that determining the secondary structure to further understand molecular mechanisms of XCI has been of great interest. Here, we consider the role of cotranscriptional folding and employ an integrative approach to investigate the structure of the Xist A-repeats. We show that the A-repeat RNA folds into two larger subdomains, each comprised of 4-5 repeats, with a modular arrangement of individual repeats by inter-repeat dimerization. Notably, the conformation and overall fold of the A-repeat is dynamic, consistent with significant effects of experimental conditions and buffer composition. Nevertheless, the basic topology and modular fold is expected to represent the native structure *in vivo*, as evidenced by the similarity of chemical probing profiles *in vitro* and in cells.

Of note, our data show that the conformational landscape of the Xist A-repeats varies under differential buffer conditions. This may also rationalize why different binding affinities have been reported for the Xist A-repeats with the PRC2 complex. In assays carried out by Davidovich et al., the Xist A-repeats are shown to bind PRC2 at 9 nM, whereas a study by Cifuentes-Rojas et al. reports a binding affinity of 80 nM (Cifuentes-Rojas et al. 2014; Davidovich et al. 2013, 2015). The differences between these two studies lie in buffer composition, where the $MgCl_2$ concentration and pH differ, and either KCl or NaCl is present in the buffer used to initiate refolding of the RNA (**Table 4**). It is likely that the buffer composition altered the structural landscape of the Xist A-repeats, thus modulating the PRC2 binding site(s). Further analysis that maps the protein binding sites to the structure would confirm this hypothesis.

As RNAs are known to fold cotranscriptionally, it is important to consider how an RNA folds in a cellular environment as it is being transcribed. Here, we have investigated the cotranscriptional pathway of the Xist A-repeats *in vitro.* The structure derived from our integrative approach confirms modular folding based on dimerization of individual repeats. Our data indicate that A-reps 1 and 2, 3 and 4, 5 and 6, and 7 and 8 form duplexes, **Figure 4**). The modular structure and topology of the A-repeat fold presented here is distinct from other proposed models, where in some cases the thermodynamically stable AUCG tetraloop is not formed. This, however, is not expected, considering cotranscriptional folding and the stability of the AUCG tetraloop. The integrative approach that we have employed here, combining low-resolution structure-probing with high-resolution structural biology data clearly demonstrates a modular structure of the A-repeat domain with the formation of AUCG tetraloop hairpin structures for each repeat, as well as inter-repeat dimerization.

Tertiary contacts of local structural elements provide an overall globular fold that nevertheless exhibits significant conformational flexibility. The conformational ensemble of the Xist A-repeat RNA is expected to be modulated by m^6A modification and the binding of RBPs, and thereby modulate the function of the lncRNA. The large differences we observe in the overall structure of the Xist A-repeats under varied buffer conditions further support the notion that the Xist A-repeats are highly subject to structural rearrangements. Our work provides an alternative approach that can be used to study the conformational dynamics of lncRNAs in solution (Morandi et al. 2021; Tomezsko et al. 2020), and sheds light on how cotranscriptional folding can be used to identify structured domains within lncRNA transcripts.

It has recently been shown that the Xist A-repeats are subject to m^6A modifications and that this has impacts on the structure adopted by the RNA (Jones et al. 2022; Patil et al. 2016), resulting in partial opening of the 5' hairpins formed by the A-repeats. Moreover, the Xist A-repeats are recognized and bound by numerous RNA binding proteins (RBPs), which in turn modulate the structure of the RNA (Da Rocha and Heard 2017). These factors may influence the cotranscriptional folding pathway in cells, and further investigation into how the Xist A-repeats fold as they are being transcribed in cells is of considerable interest.

Methods

DNA template preparation

DNA encoding the full-length Xist transcript (human) was kindly provided by Dr Carolyn Brown. DNA templates corresponding to the full-length Xist A-repeats (nucleotides 330-796, accession number gi|340393|gb|M97168.1) and each of the A-repeX RNAs were subcloned using primers purchased from Eurofins Genomics. The 5' primer for generating full-length and A-repeX RNAs was designed to contain the 17-nuclotide T7 promoter to facilitate *in vitro* transcription by the T7 polymerase. The 3' primer for the transcripts analyzed by SHAPE encoded for a 3' cassette (for primer annealing) following the final A-repeat. For A-repeX transcripts analyzed by NMR and SAXS, this cassette was not present. High-fidelity Phusion polymerase (New England Biolabs) was used to amplify the DNA constructs according to manufacturer's instructions. DNA was assessed and purified on a 1% TAE agarose gel, followed by extraction using the Wizard PCR DNA extraction and purification kit, according to manufacturer instructions.

RNA preparation

RNAs were transcribed *in vitro* using in-house prepared T7 polymerase. Briefly, 800 ng of template DNA was supplemented with 20 mM $MgCl_2$, 20X transcription buffer (100 mM Tris pH 8, 100 mM spermidine, 200 mM DTT), 5% PEG 8000, 10 mM of each rATP, rUTP, rGTP, and rCTP, and 0.03 mg of T7 polymerase. Transcriptions were incubated at 37°C for two hours followed by purification, NMR measurements, or the immediate addition of 1M7 as described below.

Full-length Xist RNAs analyzed using the buffer conditions listed in Table 1 were prepared as previously described and equilibrated in their respective buffers. Full-length Xist A-repeat RNA corresponding to the *ex vivo* Xist transcript analyzed by Smola et al., was purified under non-denaturing conditions using size exclusion chromatography, with buffer conditions matching the transcription reaction buffer listed above. RNAs were concentrated to 75 µM.

A-repeX RNAs analyzed by SHAPE were treated immediately post-transcription (after two hours) with an optimized concentration of 1M7 (25 mM) for 5 minutes at 37°C. Reactions were quenched with 6X denaturing buffer (8 M urea, 10 mM Tris pH 8.1, 40 mM EDTA), followed by purification using urea denaturing polyacrylamide gels (6.5 or 8%). Bands corresponding to the 1M7-modified RNA were excised from the gel and the RNA was extracted using the crush and soak method. Following extraction, RNAs were equilibrated against water and diluted to 1 µM.

Selective 2'OH acylation analyzed by primer extension

1M7-treated A-repX RNAs were reverse transcribed with Superscript III reverse transcriptase (purchased from ThermoFisher) according to manufacturer's instructions, using a 5' FAM labeled primer (purchased from Eurofins genomics). The resulting cDNA was precipitated with ethanol and redissolved in 100% HiDi formamide for analysis by capillary electrophoresis. Capillary electrophoresis was carried out on an ABI 3130XL genetic analyzer. QuSHAPE (Karabiber, McGinnis, and and K. M. Weeks 2013) was used to obtain SHAPE reactivities**.** Scripts to generate folding restraint files used to fold each A-repX arrested RNA, as well as the SHAPE data, are available at github.com/gpwolfe/ctf.

Nuclear magnetic resonance

All 1D NMR experiments were carried out on an 800 MHz spectrometer equipped with a cryogenic probe (Bruker) at 298 K and 280 K. 1D ^{1}H spectra were recorded using a WATERGATE flip back sequence (Piotto, Saudek, and Sklenář 1992) on RNAs concentrated to 75 µM. Spectra were processed and analyzed using TopSpin (v3.7).

Small angle X-ray scattering

Natively purified RNAs were concentrated to 0.5, 1.0, and 1.5 mg/mL in their respective buffers. Prior to measurement (carried out 298 K), RNAs were incubated at 310 K for 30 minutes. SAXS measurements were carried out at beamline P12 in Hamburg at the EMBL at the PETRA III storage ring. Buffer subtraction, scaling, and merging of data was performed in Primus.

Cryo-Electron Microscopy

Natively purified RNAs (10 µM) were incubated in their respective buffers at 310 K for 30 min prior to grid preparation. For cryoEM sample preparation, 4.5µl of the buffered RNA solution was applied to glow discharged Quantifoil 2/1 grids, blotted for 3.5s with force 4 in a Vitrobot Mark III (Thermo Fisher) at 100% humidity and 4°C, and plunge frozen in liquid ethane, cooled by liquid nitrogen. Cryo-EM data was acquired with an FEI Titan Krios transmission electron microscope using SerialEM software (Mastronarde 2005). Movie frames were recorded at a nominal magnification of 22,500X using a K3 direct electro detector (Gatan). The total electron dose of ~55 electrons per Å^2 was distributed over 30 frames at a calibrated physical pixel size of 1.0 Å. Micrographs were recorded in a defocus range of -0.5 to -3.0 µm. Particles were picked using Cryolo (Wagner et al. 2019) and micrographs analyzed in Relion (Scheres 2012).

Data availability

SHAPE reactivities and NMR restraints files used to fold the Xist A-repeats are available at github.com/gpwolfe/ctf. Experimental SAXS data have been deposited at SASDB under the accession code XXX.

References

Aguilar, Rodrigo, Kerrie B. Spencer, Barry Kesner, Noreen F. Rizvi, Maulik D. Badmalia, Tyler Mrozowich, Jonathan D. Mortison, Carlos Rivera, Graham F. Smith, Julja Burchard, Peter

J. Dandliker, Trushar R. Patel, Elliott B. Nickbarg, and Jeannie T. Lee. 2022. “Targeting Xist with Compounds That Disrupt RNA Structure and X Inactivation.” *Nature* 604(7904):160–66.

Allaman-Pillet, N., A. Djemai, C. Bonny, and D. F. Schorderet. 2000. “The 5′ Repeat Elements of the Mouse Xist Gene Inhibit the Transcription of X-Linked Genes.” *Gene Expr* 9:93–101.

Arun, Gayatri, Sarah D. Diermeier, and David L. Spector. 2018. “Therapeutic Targeting of Long Non-Coding RNAs in Cancer.” *Trends in Molecular Medicine* 24(3):257–77.

Barnwal, Ravi Pratap, Edmund Loh, Katherine S. Godin, Jordan Yip, Hayley Lavender, Christoph M. Tang, and Gabriele Varani. 2016. “Structure and Mechanism of a Molecular Rheostat, an RNA Thermometer That Modulates Immune Evasion by Neisseria Meningitidis.” *Nucleic Acids Res* 44(19):9426–37.

Chaumeil, J., P. L. Baccon, and A. Wut. and E. Heard. 2006. “A Novel Role for Xist RNA in the Formation of a Repressive Nuclear Compartment into Which Genes Are Recruited When Silenced.” *Genes Dev* 20(16):2223–2237.

Chen, Yujie and Lois Pollack. 2016. “SAXS Studies of RNA: Structures, Dynamics, and Interactions with Partners.” 7(4):512–26.

Chillón, Isabel, Marco Marcia, Michal Legiewicz, Fei Liu, Srinivas Somarowthu, and Anna Marie Pyle. 2015. “Native Purification and Analysis of Long RNAs.” *Methods Enzymol* 558(1):3–37.

Chu, C., QC Zhang, ST da Rocha, RA Flynn, M. Bharadwaj, JM Calabrese, T. Magnuson, E. Heard, and HY Chang. 2015. “Systematic Discovery of Xist RNA Binding Proteins.” *Cell* 161(2):404–16.

Cifuentes-Rojas, Catherine, Alfredo J. Hernandez, Kavitha Sarma, and Jeannie T. Lee. 2014. “Regulatory Interactions between RNA and Polycomb Repressive Complex 2.” *Mol Cell* 55(2):171–85.

D. H. Mathews, J. S. Reute. n.d. “RNAstructure: Software for RNA Secondary Structure Prediction and Analysis.” *BMC Bioinformatics* 11(129).

Dallaire, Paul, Huiping Tan, Keith Szulwach, Christopher Ma, Peng Jin, and François Major. 2016. “Structural Dynamics Control the MicroRNA Maturation Pathway.” *Nucleic Acids Res* 44(20):9956–64.

Davidovich, Chen, Xueyin Wang, Catherine Cifuentes-rojas, Karen J. Goodrich, R. Gooding, Jeannie T. Lee, and Thomas R. Cech. 2015. “Towards a Consensus on the Binding Specificity and Promiscuity of PRC2 for RNA.” *Mol Cell* 57(3):552–58.

Davidovich, Chen, Leon Zheng, Karen J. Goodrich, and Thomas R. Cech. 2013. “Promiscuous RNA Binding by Polycomb Repressive Complex 2.” *Nat Struct Mol Biol*.

Dethoff, Elizabeth A., Jeetender Chugh, Anthony M. Mustoe, and H. M. Al-Hashimi. 2012. “Functional Complexity and Regulation through RNA Dynamics.” *Nature* 482(7385):322–30.

Diermeier, Sarah D., Kung-Chi Chang, Susan M. Freier, Junyan Song, Osama El Demerdash, Alexander Krasnitz, Frank Rigo, C. Frank Bennett, and David L. Spector. 2016. “Mammary Tumor-Associated RNAs Impact Tumor Cell Proliferation, Invasion, and Migration.” *Cell Rep* 17(1):261–74.

Durand, Dominique, Corinne Vivès, Dominique Cannella, Javier Pérez, Eva Pebay-Peyroula, Patrice Vachette, and Franck Fieschi. 2010. “NADPH Oxidase Activator P67phox Behaves in Solution as a Multidomain Protein with Semi-Flexible Linkers.” *Journal of Structural Biology* 169(1):45–53.

Duszczyk, Malgorzata M., Anton Wutz, Vladimir Rybin, and Michael Sattler. 2011. “The Xist

RNA A-Repeat Comprises a Novel AUCG Tetraloop Fold and a Platform for Multimerization." *RNA* 17(11):1973–82.

Duszczyk, Malgorzata M., Katia Zanier, and Michael Sattler. 2008. "A NMR Strategy to Unambiguously Distinguish Nucleic Acid Hairpin and Duplex Conformations Applied to a Xist RNA A-Repeat." *Nucleic Acids Res* 36(22):7068–77.

Fang, Rui, Walter N. Moss, Michael Rutenberg-Schoenberg, and Matthew D. Simon. 2015. "Probing Xist RNA Structure in Cells Using Targeted Structure-Seq." *PLoS Genet* 11(12):e1005668.

Fürtig, Boris, Christian Richter, Jens Wöhnert, and Harald Schwalbe. 2003. "NMR Spectroscopy of RNA." *ChemBioChem* 4(10):936–62.

Jones, Alisha N., Giuseppina Pisignano, Thomas Pavelitz, Jessica White, Martin Kinisu, Nicholas Forino, Dreycey Albin, and Gabriele Varani. 2020. "An Evolutionarily-Conserved RNA Structure in the Functional Core of the LincRNA Cyrano." *Rna* 26(9):1234–46.

Jones, Alisha N. and Michael Sattler. 2019. "Challenges and Perspectives for Structural Biology of LncRNAs – the Example of the Xist LncRNA A-Repeats." *Journal of Molecular Cell Biology* 11(10):845–59.

Jones, Alisha N., Ekaterina Tikhaia, André Mourão, and Michael Sattler. 2022. "Structural Effects of M6A Modification of the Xist A-Repeat AUCG Tetraloop and Its Recognition by YTHDC1." *Nucleic Acids Research* 1–13.

Karabiber, F., J. L. McGinnis, and O. V. Favoro. and K. M. Weeks. 2013. "QuShape: Rapid, Accurate, and Best-Practices Quantification of Nucleic Acid Probing Information, Resolved by Capillary Electrophoresis." *RNA* 19(1):63–73.

Kim, Doo Nam, Bernhard C. Thiel, Tyler Mrozowich, Scott P. Hennelly, Ivo L. Hofacker, Trushar R. Patel, and Karissa Y. Sanbonmatsu. 2020. "Zinc-Finger Protein CNBP Alters the 3-D Structure of LncRNA Braveheart in Solution." *Nature Communications* 11(1):1–13.

Kim, Jiyoung, Kotb Abdelmohsen, Xiaoling Yang, Supriyo De, Ioannis Grammatikakis, Ji Heon Noh, and Myriam Gorospe. 2016. "LncRNA OIP5-AS1/Cyrano Sponges RNA-Binding Protein HuR." *Nucleic Acids Res* 44(5):2378–92.

Kung, Johnny T. Y., David Colognori, and Jeannie T. Lee. 2013. "Long Noncoding RNAs: Past, Present, and Future." *Genetics* 193(3):651–69.

Lai, Daniel, Jeff R. Proctor, and Irmtraud M. Meyer. 2013. "On the Importance of Cotranscriptional RNA Structure Formation." *RNA* 19:1461–73.

Lewandowski, Jordan P., James C. Lee, Taeyoung Hwang, Hongjae Sunwoo, Jill M. Goldstein, Abigail F. Groff, Nydia P. Chang, William Mallard, Adam Williams, Jorge Henao-Meija, Richard A. Flavell, Jeannie T. Lee, Chiara Gerhardinger, Amy J. Wagers, and John L. Rinn. 2019. "The Firre Locus Produces a Trans-Acting RNA Molecule That Functions in Hematopoiesis." *Nature Communications* 10(5137):1–13.

Liu, Fei, Srinivas Somarowthu, and Anna Marie Pyle. 2017. "Visualizing the Secondary and Tertiary Architectural Domains of LncRNA RepA." *Nat Chem Biol* 13(3):282–89.

Loda, Agnese and Edith Heard. 2019. "Xist RNA in Action: Past, Present, and Future." *PLoS Genetics* 15(9):1–17.

Lu, Zhipeng, Qiangfeng Cliff Zhang, Byron Lee, Ryan A. Flynn, Martin A. Smith, James T. Robinson, Chen Davidovich, Anne R. Gooding, Karen J. Goodrich, John S. Mattick, Jill P. Mesirov, Thomas R. Cech, and Howard Y. Chang. 2016. "RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure." *Cell* 165(5):1267–79.

Maenner, Sylvain, Magali Blaud, Laetitia Fouillen, Anne Savoye, Virginie Marchand, Agnès Dubois, Sarah Sanglier-Cianférani, Alain Van Dorsselaer, Philippe Clerc, Philip Avner,

Athanase Visvikis, and Christiane Branlant. 2010. “2-D Structure of the a Region of Xist RNA and Its Implication for PRC2 Association.” *PLoS Biol* 8(1):1–16.

Mastronarde, David N. 2005. “Automated Electron Microscope Tomography Using Robust Prediction of Specimen Movements.” *Journal of Structural Biology* 152(1):36–51.

Morandi, Edoardo, Ilaria Manfredonia, Lisa M. Simon, Francesca Anselmi, Martijn J. van Hemert, Salvatore Oliviero, and Danny Incarnato. 2021. “Genome-Scale Deconvolution of RNA Structure Ensembles.” *Nature Methods* 18(3):249–52.

Patil, Deepak P., Chun Kan Chen, Brian F. Pickering, Amy Chow, Constanza Jackson, Mitchell Guttman, and Samie R. Jaffrey. 2016. “M6A RNA Methylation Promotes XIST-Mediated Transcriptional Repression.” *Nature* 537(7620):369–73.

Penny, G. D., G. F. Kay, S. A. Sheardown, S. Rastan, and N. Brockdorff. 1996. “Requirement for Xist in X Chromosome Inactivation.” *Nature* 379:131–37.

Pintacuda, Greta, Alexander N. Young, and Andrea Cerase. 2017. “Function by Structure: Spotlights on Xist Long Non-Coding RNA.” *Front Mol Biosci* 4(90):1–11.

Piotto, Martial, Vladimir Saudek, and Vladimir Sklenář. 1992. “Gradient-Tailored Excitation for Single-Quantum NMR Spectroscopy of Aqueous Solutions.” *Journal of Biomolecular NMR* 2(6):661–65.

Ponti, Riccardo Delli, Alexandros Armaos, Stefanie Marti, and Gian Gaetano Tartaglia. 2018. “A Method for RNA Structure Prediction Shows Evidence for Structure in LncRNAs.” *Frontiers in Molecular Biosciences* 5(111):1–14.

Ponting, Chris P., Peter L. Oliver, and Wolf Reik. 2009. “Evolution and Functions of Long Noncoding RNAs.” *Cell* 136(4):629–41.

Popenda, Mariusz, Marta Szachniuk, Maclej Antczak, Katarzyna J. Purzycka, Piotr Lukasiak, Natalia Bartol, Jacek Blazewicz, and Ryszard W. Adamiak. 2012. “Automated 3D Structure Composition for Large RNAs.” *Nucleic Acids Research* 40(14):e112.

Da Rocha, Simão T. and Edith Heard. 2017. “Novel Players in X Inactivation: Insights into Xist-Mediated Gene Silencing and Chromosome Conformation.” *Nat Struct Mol Biol* 24(3):197–204.

Scheres, Sjors H. W. 2012. “RELION: Implementation of a Bayesian Approach to Cryo-EM Structure Determination.” *Journal of Structural Biology* 180(3):519–30.

Smola, M. J., T. W. Cristy, K. Inoue, C. O. Nicholson, M. Friedersdorf, J. D. Keene, D. M. Lee, J. M. Calabrese, and K. M. Weeks. 2016. “SHAPE Reveals Transcript-Wide Interactions, Complex Structural Domains, and Protein Interactions across the Xist LncRNA in Living Cells.” *Proc Natl Acad Sci U S A* 113(37):10322–27.

Smola, Matthew J., J. Mauro Calabrese, and Kevin M. Weeks. 2015. “Detection of RNA-Protein Interactions in Living Cells with SHAPE.” *Biochemistry* 54(46):6867–75.

Somarowthu, S., M. Legiewicz, I. Chillon, M. Marcia, and F. Li. and A. M. Pyle. 2015. “HOTAIR Forms an Intricate and Modular Secondary Structure.” *Molecular Cell* 58(2):353–361.

Svergun, D., C. Barberato, and M. H. J. Koch. 1995. “CRYSOL– a Program to Evaluate X-Ray Solution Scattering of Biological Macromolecules FromAtomic Coordinates.” *Journal of Applied Crystallography* 28(1):768–73.

Tomezsko, Phillip J., Vincent D. A. Corbin, Paromita Gupta, Harish Swaminathan, Margalit Glasgow, Sitara Persad, Matthew D. Edwards, Lachlan Mcintosh, Anthony T. Papenfuss, and Paromita ✉ Gupta. 2020. “Determination of RNA Structural Diversity and Its Role in HIV-1 RNA Splicing.” *Nature* 582:438.

Uroda, Tina, Eleni Anastasakou, Annalisa Rossi, Jean-Marie Teulon, Jean-Luc Pellequer, Paolo Annibale, Ombeline Pessey, Alberto Inga, Isabel Chillón, and Marco Marcia. 2019.

"Conserved Pseudoknots in LncRNA MEG3 Are Essential for Stimulation of the P53 Pathway." *Molecular Cell* 75(5):982–95.

Wagner, Thorsten, Felipe Merino, Markus Stabrin, Toshio Moriya, Claudia Antoni, Amir Apelbaum, Philine Hagel, Oleg Sitsel, Tobias Raisch, Daniel Prumbaum, Dennis Quentin, Daniel Roderer, Sebastian Tacke, Birte Siebolds, Evelyn Schubert, Tanvir R. Shaikh, Pascal Lill, Christos Gatsogiannis, and Stefan Raunser. 2019. "SPHIRE-CrYOLO Is a Fast and Accurate Fully Automated Particle Picker for Cryo-EM." *Communications Biology* 2(1):1–13.

Wang, Jiaxu, Tong Zhang, Zhang Yu, Wen Ting Tan, Ming Wen, Yang Shen, Finnlay R. P. Lambert, Roland G. Huber, and Yue Wan. 2021. "Genome-Wide RNA Structure Changes during Human Neurogenesis Modulate Gene Regulatory Networks." *Molecular Cell* 81(23):4942-4953.e8.

Wang, Kevin C. and Howard Y. Chang. 2011. "Molecular Mechanisms of Long Noncoding RNAs." *Molecular Cell* 43:904–14.

Watters, Kyle E., Eric J. Strobel, Angela M. Yu, John T. Lis, and Julius B. Lucks. 2016. "Cotranscriptional Folding of a Riboswitch at Nucleotide Resolution." *Nat Struct Mol Biol* 23(12):1124–31.

Wutz, Anton, Theodore P. Rasmussen, and Rudolf Jaenisch. 2002. "Chromosomal Silencing and Localization Are Mediated by Different Domains of Xist RNA." *Nat Genet* 30(2):167–74.

Acknowledgements

We thank Dr Carolyn Brown for providing the Xist DNA template, Dr Dominik Lenhart for the synthesis of 1M7 and Dr Arie Geerlof for preparation of T7 polymerase. We thank Johanna Voigt, Manuel Rohl and Markus Mueller for assistance with sample preparation. We thank Dr Gerd Gemmecker and Dr Sam Asami for help with NMR experiments, Dr Ralf Stehle for assistance with SAXS analysis, and members of the Sattler group for helpful discussions. The synchrotron SAXS data were collected at beamline P12 operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany). We would like to thank Stefano Da Vela for the assistance in using the beamline. NMR measurements at the Bavarian NMR center, and access to cryo-EM instrumentation at the MPI for biochemistry, Martinsried.

Author contributions

ANJ carried out SHAPE and NMR experiments, with contributions from JV and analyzed SHAPE and NMR data. SB and ANJ carried out cryo-EM experiments and analyzed the data. ANJ and FG analyzed SAXS data. GW wrote scripts for generating SHAPE restraints, with contributions from ANJ. ANJ and MS designed the study and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Funding

We acknowledge funding by the Deutsche Forschungsgemeinschaft (German Research Foundation) SFB1309 project 325871075 (M.S.) and TRR267 project 403584255 (M.S.).

Tables

Table 1. Buffer conditions used to probe the secondary structure of the Xist A-repeats.

Reference	Buffer condition
Duszczyk, et al.	298 K; 10 mM NaH_2PO_4/Na_2HPO_4 pH 6.0, 100 mM NaCl, 0.02 mM EDTA, 0.02% azide
Maenner, et al.	298 K; 20 mM HEPES-KOH pH 7.9, 100 mM KCl, 0.2 mM EDTA pH 8.0, 0.5 mM DTT, 0.5 mM PMSF, 20% glycerol, 3.25 mM $MgCl_2$
Liu, et al.	310 K/298 K; 25 mM K-HEPES pH 7.0, 0.1 mM Na-EDTA, 150 mM KCl, 15 mM $MgCl_2$
Smola, et al.	310 K; 100 mM HEPES pH 8.0, 100 mM NaCl, 10 mM $MgCl_2$

Table 2. SAXS parameters obtained for each of the RNAs studied.

Buffer condition (298 K)	R_g (nm)	D_{max} (nm)
Liu et al.,	12.10 ± 0.21	42.98
Duszczyk et al.,	7.56 ± 0.02	30.33
Smola et al.,	9.83 ± 0.09	33.03
Maenner et al.,	8.89 ± 0.18	34.69

Table 3. SAXS parameters obtained for each of the arrested A-rep RNAs studied.

RNA construct (298 K)	R_g (nm)
14-mer	2.12 ± 0.05
24-mer	3.08 ± 0.02
A-rep1	4.10 ± 0.07
A-rep2	5.20 ± 0.09
A-rep3	6.11 ± 0.19
A-rep4	7.20 ± 0.10
A-rep5	7.86 ± 0.20
A-rep6	12.86 ± 0.43
A-rep8	11.39 ± 0.13

Table 4. Buffer conditions and binding affinities for the Xist A-repeats with PRC2.

Reference	Binding affinity	Buffer condition
Davidovich, et al.	9.1 nM	50 mM Tris HCl, pH 7.5 (@ RT), 100 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM $ZnCl_2$, 2 mM 2- mercaptoethanol, 0.1 mg/ml BSA, 0.1 mg/ml yeast tRNA, 0.05 % v/v NP40 and 5 % v/v glycerol
Cifuentes-Rojas, et al.	80.9 nM	50 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM $MgCl_2$, 10 µg/ml BSA, 0.05% NP40, 1 mM DTT, 20U RNaseOUT and 5% Glycerol in 20 µl

Figures

Figure 1

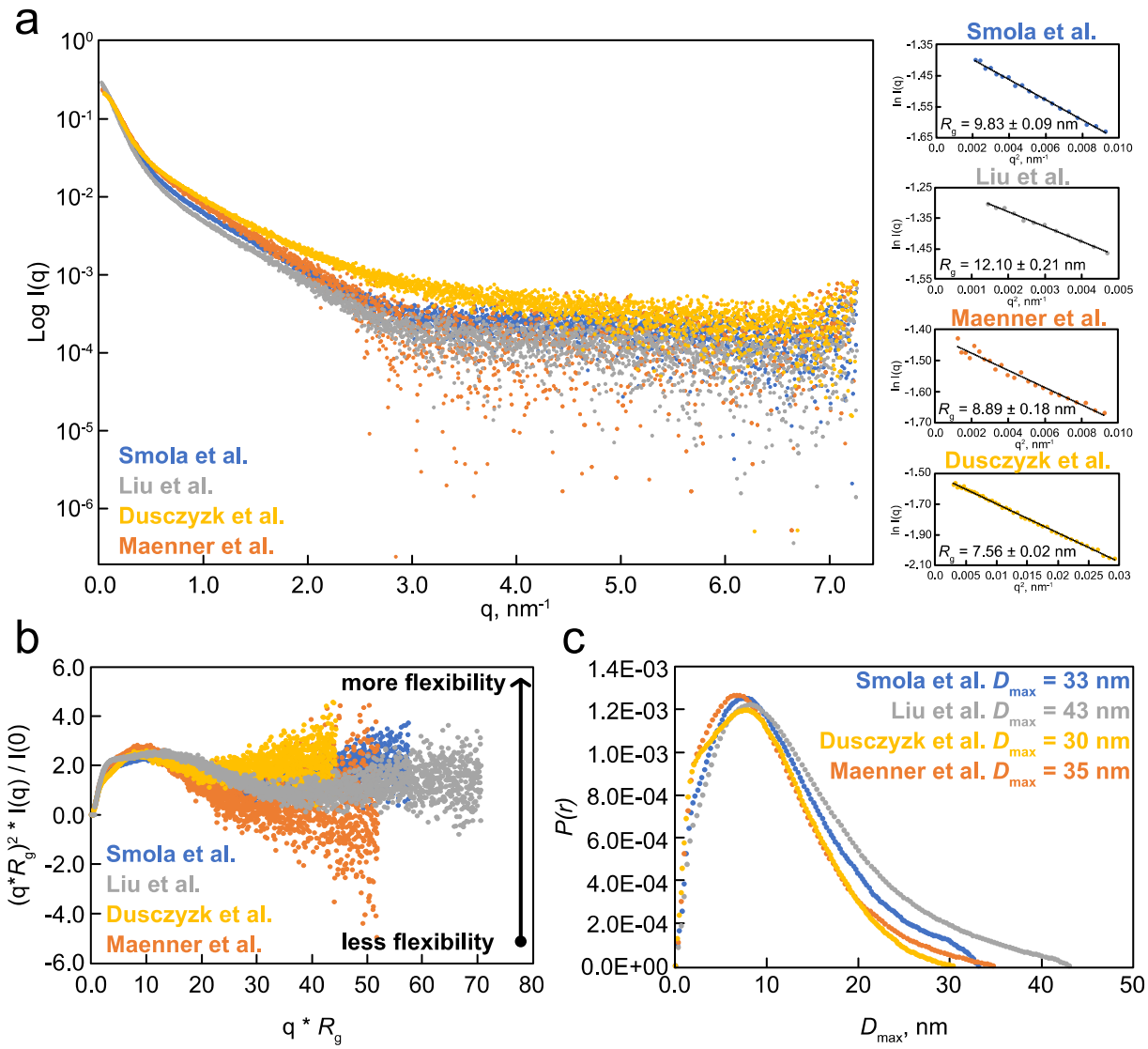

Figure 1. SAXS analysis of the Xist A-repeats under various buffered conditions at 298 K. a) SAXS curves for buffers corresponding to Dusczyzk et al., Maenner et al., Smola et al., and Liu et al. The radius of gyration, as determined from Guinier analysis, is shown on the right. The straight line is the linear regression fit in the range of q < 1.3/R_g. b) Dimensionless Kratky plot for each of the four buffered conditions. c) *P(r)* curve for each of the four buffered conditions, with D_{max} values listed.

Figure 2

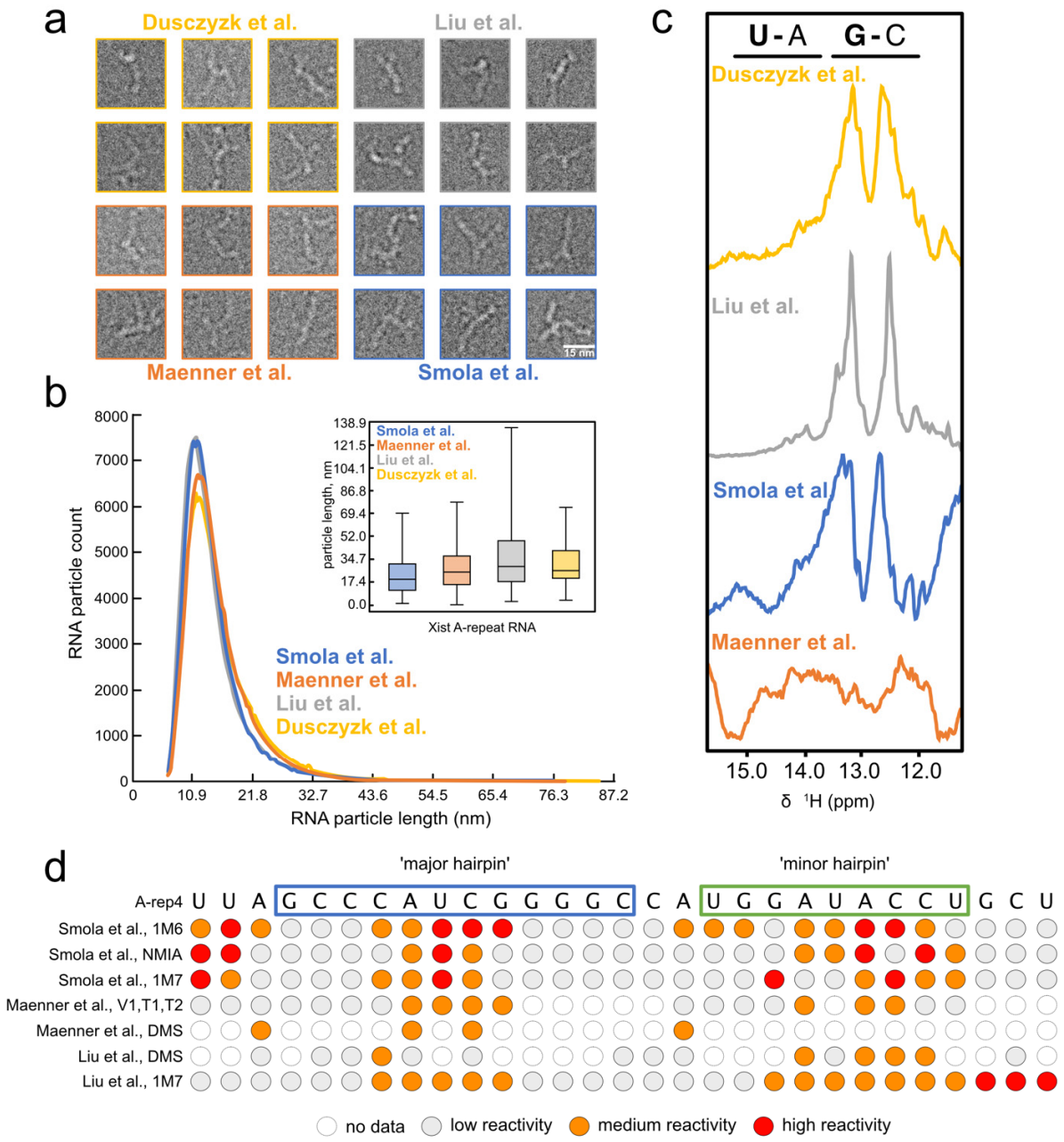

Figure 2. Effect of buffer on the structure of the Xist A-repeat RNAs. a) Individual particles of the Xist A-repeats in vitreous ice, in buffers used by Maenner (orange), Smola (blue), Liu (grey), and Dusczyzk (yellow) et al. b) Distribution of maximum distances found in particles analyzed by cryoEM. The inset shows the lengths of only the longest particles for approximately 100 particles for each buffered condition. c) ^{1}H 1D imino fingerprint spectra of the Xist A-repeats under varying buffer conditions at 298K. d) Representative chemical probing (1M6, NMIA, 1M7, and DMS) and enzymatic cleavage (V1, T1, and T2) data for Xist A-repeat 4 as determined by Smola, et al., Maenner et al., and Liu et al. See also Supplementary File 1.

Figure 3

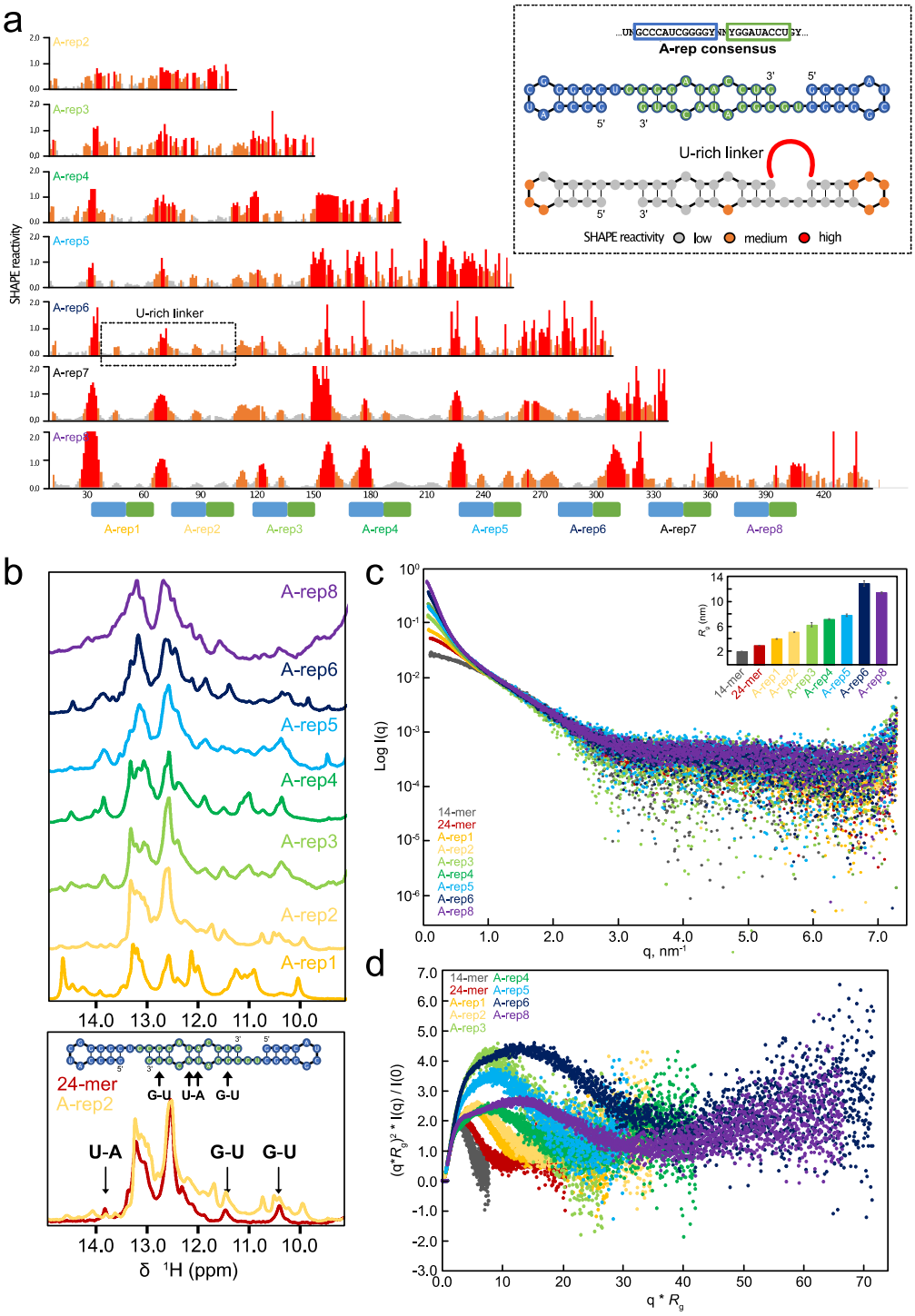

Figure 3. SHAPE and SAXS analysis of the Xist A-repeats throughout the course of cotranscriptional folding. a) The SHAPE reactivity profile for extended Xist A-repeat RNA fragments. The inset box shows the SHAPE reactivity profile mapped to a duplex formed by two consecutive A-repeats. b) ^{1}H 1D imino fingerprint spectra for each of the Xist A-repeat cotranscriptionally folded RNAs. c) Dimensionless Kratky plot for extended Xist A-repeat RNA fragments.

Figure 4

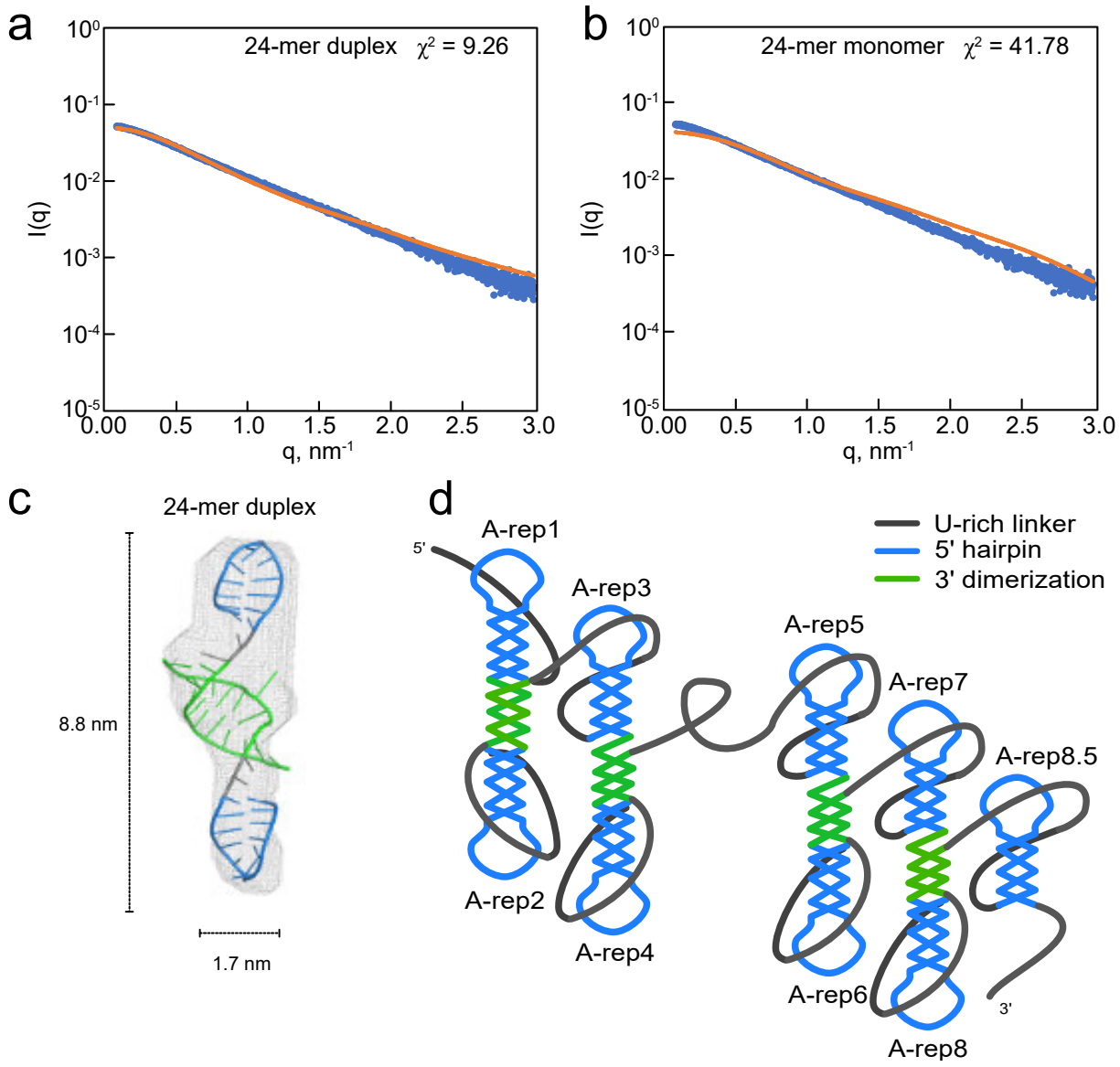

Figure 4. Structural model of the Xist A-repeats. c) CRYSOL goodness of fit curve for the 24-mer SAXS curve (highest concentration curve at 1.98 mg / mL) with a duplexed 24-mer RNA structure and a b) monomeric 24-mer structure. c) The RNAComposer model for the duplexed 24-nucleotide Xist A-repeat RNA element fits well inside the 3D DAMMIF reconstruction of the structure based on SAXS data. d) Cartoon of the proposed tertiary structure of the full-length Xist A-repeats.

Figure 5

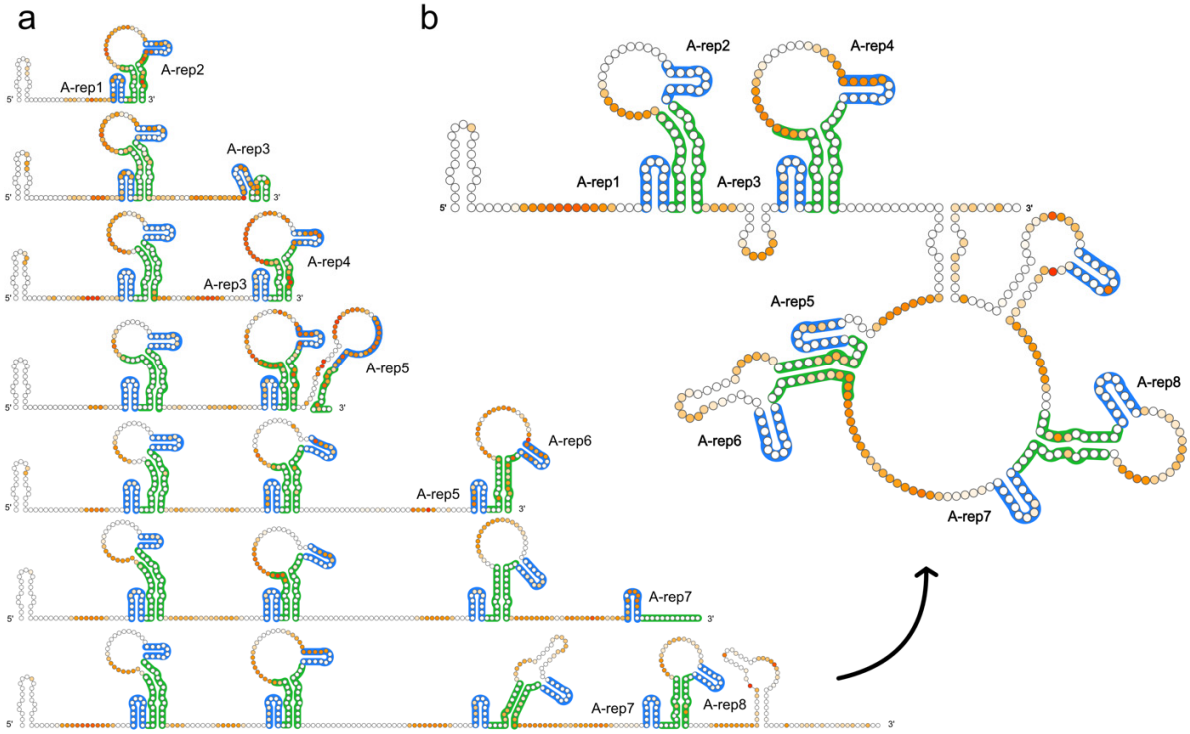

Figure 5. Cotranscriptional folding pathway of the Xist A-repeats. a) The secondary structures of the Xist A-rep1 – A-rep8 RNA arrested RNAs; for simplicity, nucleotides between each A-rep duplex are represented as single stranded nucleotides. b) Secondary structure of the full-length Xist A-repeats based on SHAPE restraints.