



RNA

A PUBLICATION OF THE RNA SOCIETY

The Xist RNA A-repeat comprises a novel AUCG tetraloop fold and a platform for multimerization

Malgorzata M. Duszczek, Anton Wutz, Vladimir Rybin, et al.

RNA 2011 17: 1973-1982 originally published online September 26, 2011

Access the most recent version at doi:[10.1261/rna.2747411](https://doi.org/10.1261/rna.2747411)

Supplemental Material

<http://rnajournal.cshlp.org/content/suppl/2011/08/31/rna.2747411.DC1.html>

References

This article cites 36 articles, 9 of which can be accessed free at:

<http://rnajournal.cshlp.org/content/17/11/1973.full.html#ref-list-1>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)

To subscribe to *RNA* go to:
<http://rnajournal.cshlp.org/subscriptions>

The Xist RNA A-repeat comprises a novel AUCG tetraloop fold and a platform for multimerization

MALGORZATA M. DUSZCZYK,^{1,2,5} ANTON WUTZ,³ VLADIMIR RYBIN,⁴ and MICHAEL SATTLER^{1,2,6}

¹Institute of Structural Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany

²Biomolecular NMR and Center for Integrated Protein Science Munich at Department Chemie, Technische Universität München, 85747 Garching, Germany

³Wellcome Trust Centre for Stem Cell Research, Cambridge CB2 1QN, United Kingdom

⁴Structural and Computational Biology Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany

ABSTRACT

X-chromosome inactivation (XCI) in female mammals depends on the noncoding RNA X inactivation specific transcript (Xist). The mechanism of chromosome-wide silencing by *Xist* is poorly understood. While it is established that the 5' region of Xist RNA, comprising the A-repeats and holding 7.5–8.5 copies of a conserved 26-mer sequence, is essential for Xist-mediated silencing, high-resolution structural information for the A-repeats is not available. Here, we report the three-dimensional solution structure of a 14-mer hairpin in the 5' region of a human A-repeat. This hairpin is remarkably stable and adopts a novel AUCG tetraloop fold, the integrity of which is required for silencing. We show that, contrary to previous predictions, the 3' region of single or tandem A-repeats mediates duplex formation in vitro. Significantly, mutations in this region disrupt the inter-repeat duplex formation in vitro and abrogate the silencing function of Xist A-repeats in vivo. Our data suggest that the complete A-repeat region may be stabilized by inter-repeat duplex formation and, as such, may provide a platform for multimerization and specific recognition of the AUCG tetraloops by *trans*-acting factors.

Keywords: X-inactivation; Xist RNA; A-repeats; nuclear magnetic resonance (NMR); structural biology; RNA tetraloop

INTRODUCTION

Dosage compensation is a mechanism that evolved to compensate for the difference in X-linked gene expression in species with different numbers of X chromosomes between the sexes (Lucchesi et al. 2005). The mammalian solution to dosage compensation is X-chromosome inactivation (XCI), the inactivation of one of the two X chromosomes in females. Initiation of XCI takes place early in development and depends on the large noncoding RNA X-inactivation specific transcript (Xist), which is unique to placental mammals (Penny et al. 1996). XCI is initiated by Xist transcription and coating of the future inactive X chromosome (Xi) in *cis*, which coincides with transcriptional shutdown through an unknown mechanism (Sheardown et al. 1997).

Differential treatment of two X chromosomes in a single cell is facilitated by regulatory mechanisms of so-called

“counting” of the number of X chromosomes and “choice” to inactivate all but one of them (Avner and Heard 2001). These mechanisms are controlled from the X chromosome itself. The X-inactivation center (Xic) contains several regulatory elements including the *Xist* gene and an overlapping gene for *Tsix*, another noncoding RNA that is transcribed in antisense orientation and suppresses *Xist* function (Lee and Lu 1999). Based on the observation of Xist-Tsix sense-antisense duplexes in vivo, it was speculated that these could function as regulatory RNAs and contribute to the differential treatment of the X chromosomes in an RNAi-like fashion (Ogawa et al. 2008), although details of such a role remain to be clarified.

Xist RNA accumulation on the Xi and its ability to trigger silencing are functionally separable (Wutz et al. 2002). The so-called “A-repeats” located at the 5' end of Xist are conserved among all placental mammals and are essential for the initiation of silencing. Several other regions are redundantly responsible for chromosome association (Fig. 1A). In humans, the A-repeats are constituted of 8.5 copies of a highly conserved 26-nucleotide motif, connected by long U-rich linkers (Fig. 1B; Supplemental Fig. S1). Xist RNA lacking the A-repeats accumulates on the Xi

⁵Present address: Institute of Molecular Biology and Biophysics, Eidgenössische Technische Hochschule Zürich, 8093 Zürich, Switzerland.

⁶Corresponding author.

E-mail sattler@helmholtz-muenchen.de.

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.2747411>.

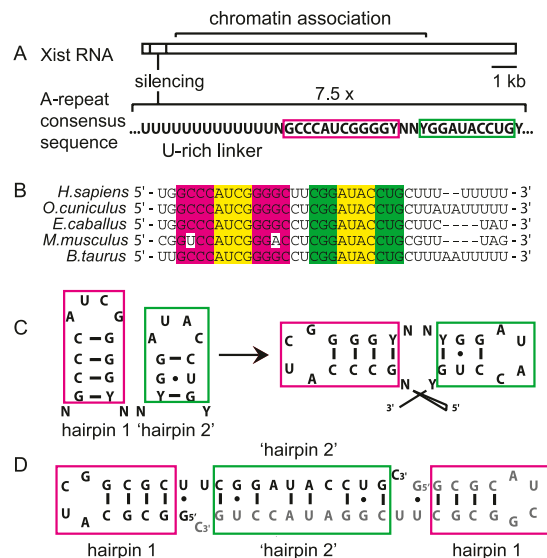


FIGURE 1. Schematic structure and sequence of Xist RNA and its A-repeats. (A) Xist RNA is a long (15 kb in mouse, 17 kb in human) noncoding RNA. The A-repeats located at the 5' end are essential for silencing, while other regions are redundantly responsible for chromosome association. (B) Sequence alignment of the sixth human A-repeat shows that it is highly conserved among placental mammals. (C) The A-repeats consist of 7.5–8.5 copies of a conserved sequence predicted to fold into two hairpins, connected by long U-rich linkers. (N) any nucleotide, (Y) C/U. (D) In vitro, “hairpin 2” is not formed but mediates intermolecular duplex formation with a second A-repeat through intermolecular A-U base pairs. Duplex formation in the NMR-modified NMR-XCR construct is represented.

but is not able to induce silencing. It has been proposed that the A-repeats are a recruitment site for factors regulating gene repression. Following Xist-coating, Polycomb complexes are recruited to the Xi (Silva et al. 2003; de Napoles et al. 2004; Plath et al. 2004; Schoeftner et al. 2006). The Polycomb repressive complex 2 (PRC2) is responsible for the establishment of histone H3-K27 methylation marks along the Xi and is thus required for long-term X inactivation (Plath et al. 2003). The A-repeats have been proposed to recruit PRC2 to the Xi, consistent with binding of A-repeats to PRC2 components, including Ezh2 and Suz12, in vivo and in vitro (Zhao et al. 2008; Kanhere et al. 2010; Maenner et al. 2010). More recently, it was reported that the A-repeats are required for Xist RNA processing through binding to the alternative splicing factor ASF/SF2, although this interaction is not considered to influence Xist's silencing function (Royce-Tolland et al. 2010).

The molecular mechanisms of XCI are still poorly understood. A single A-repeat has been predicted to fold into a double hairpin structure where the two hairpins possibly stack on top of each other (Fig. 1C; Wutz et al. 2002). We previously reported that, *in vitro*, only hairpin 1 in the 5' part of the A-repeat is formed, while the predicted "hairpin 2" in the 3' region does not fold but instead mediates dimerization of the 26-mer A-repeat, as shown in Figure 1D (Duszczek et al. 2008). Recently, a two-dimensional

structural model has been proposed for the A-repeat region based on foot-printing, chemical modification, and FRET data (Maenner et al. 2010). In this model, the complete A-repeat region is predicted to fold into two extended stem-loop structures that involve inter-repeat pairing of A-repeats, but without the formation of the earlier predicted hairpin structures within individual A-repeats.

Here, we show that—in contrast to this model (Maenner et al. 2010)—the 5′ region of single A-repeats comprises a remarkably stable conserved RNA hairpin. NMR-spectroscopy reveals that the solution structure of this hairpin adopts a novel AUCG tetraloop fold. The 3′ region of the A-repeat does not adopt the predicted hairpin structure but is involved in inter-repeat duplex formation. By mutational analysis *in vitro* and *in vivo* using a mouse embryonic stem (ES) cell system with an inducible *Xist* transgene, we show that sequence variations in the 3′ region that destabilize intermolecular dimerization of a single A-repeat *in vitro* are unable to induce silencing *in vivo*. Based on our structural and functional analysis *in vitro* and *in vivo*, we propose a novel model for the architecture of the *Xist* RNA A-repeats and discuss the implications for their molecular function.

RESULTS

Structure of the AUCG RNA tetraloop in the Xist A-repeat

We determined the solution structure of the 14-mer hairpin located in the 5' part of an Xist RNA A-repeat using heteronuclear multidimensional NMR (Supplemental Methods; Supplemental Figs. S2–S6). The sequence of this 14-mer hairpin (Fig. 2D) is derived from the sixth human A-repeat (Fig. 1B; Supplemental Fig. S1). The third G-C base-pair is replaced by a C-G base pair to facilitate chemical shift assignments, and the closing G-U base pair is replaced by G-C due to superior spectral quality. Previous studies have shown that sequence variations in the stem of hairpin 1 are tolerated and do not influence Xist activity if base-pairing is retained (Wutz et al. 2002). Moreover, an *Xist* transgene with this exact A-repeat sequence for hairpin 1 is functional and mediates silencing in vivo (see NMR-XCR in Fig. 4D, see below). The formation of a monomeric hairpin in the NMR sample was confirmed as described (Duszczuk et al. 2008). Comparison of 2D (^1H , ^1H) TOCSY and NOESY spectra demonstrates that the conformation of the AUCG tetraloop in the 14-mer is the same as in the context of the full 26-mer single A-repeat (Supplemental Fig. S4).

The structure of the 14-mer A-repeat hairpin 1 was calculated with ARIA/CNS (Brunger et al. 1998; Linge et al. 2003) and is based on 292 NOE-derived distance restraints (Supplemental Fig. S3), as well as 25 restraints derived from residual dipolar couplings (RDCs). After subsequent re-

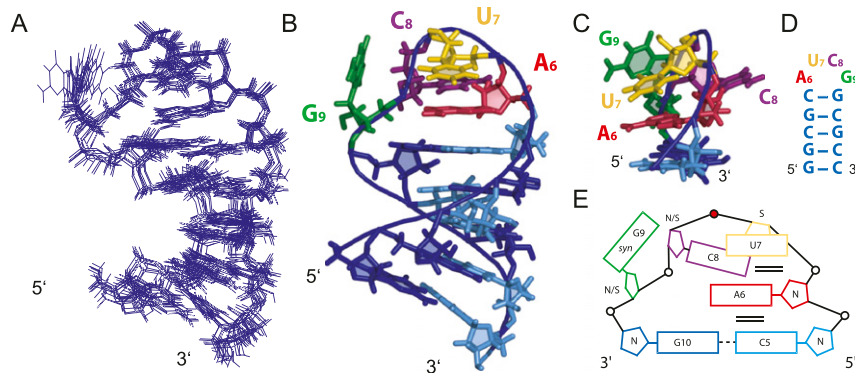


FIGURE 2. Solution structure of the Xist RNA A-repeat AUCG tetraloop. (A) View of the NMR ensemble of 10 final AMBER-refined structures, superimposed over all heavy atoms. (B) View of the lowest energy structure. The loop nucleotides are colored red (A), yellow (U), purple (C), and green (G), as shown in the secondary structure in D. (C) View of the AUCG tetraloop from the side. (E) Schematic view of the AUCG tetraloop. Nucleotides are colored according to B, C, and D. Bases are labeled with the base name and “syn” if the base is in the “syn”-conformation and not in the more common “anti”-conformation. Sugars are labeled with N (North, C3'-endo), S (South, C2'-endo), or N/S (mixed), according to their conformation. The phosphate colored red is the turning phosphate in the tetraloop. Stacking between bases is represented by a double horizontal line and hydrogen bonding by a dotted line.

finement using AMBER, a final ensemble of 10 structures (Fig. 2A) was selected based on restraint violation analysis and back-calculation of chemical shifts (Supplemental Fig. S6). A summary of structural statistics is given in Table 1.

The 14-mer hairpin 1 adopts a well-defined stem-loop structure, with an A-form helical stem that is capped by a structured AUCG tetraloop (Fig. 2). The fold of the A⁶-U⁷-C⁸-G⁹ tetraloop is stabilized by 5' extended stacking of the A6 and U7 base onto the A-form helical stem. The phosphodiester backbone is reversed between U7-C8, allowing the C8 base to fold back into the minor groove with the H5/H6 side pointing inward. Following C8, the backbone is kinked toward the major groove. The G9 base is bulged out in a *syn* conformation. One of the G9 phosphate oxygens points toward A6 and forms a potential C—H...O=P hydrogen bond (Wahl and Sundaralingam 1997; Henn et al. 2004) with A6 H2, inferred from the donor-acceptor distance (2.2 Å) and geometry. Although the chemical shift of the A6 H2 proton is not unusual, deshielding effects from the H-bond could be compensated for by the ring currents of the bases stacked above and below the A6. Unique features of the tetraloop structure are the kink in the backbone between C8 and G9 and the exposure of the G9 purine, which is rarely observed in RNA hairpin loops. Continuation of 5' stacking of A6 and U7 and deviation from A-helical stacking for C8 and G9, as well as the bulged-out conformation of G9, are consistent with the NOE data, experimental dihedral angles, and unusual chemical shifts (Supplemental Material). The solvent-exposed conformation of G9 is further supported by NMR paramagnetic relaxation enhancement data (M Duszczek, T Madl, and M Sattler, in prep.). Experimental J-couplings for residues C8 to G10 indicate conformational dynamics and

averaging of sugar puckers and the ϵ/γ angles (Supplemental Table S1). The exposure of G9 and the indication of conformational dynamics of residues C8 to G10 suggest that potential molecular recognition of the AUCG tetraloop, for example, by a cognate protein, may involve an induced fit.

In summary, the Xist RNA AUCG tetraloop structure is stabilized by a 5' extended base stacking and cross-loop hydrogen bonding. The 3' side of the loop is less well-defined due to conformational dynamics of residues C8 and G9. The G9 base is solvent-exposed and easily accessible for potential intermolecular interactions.

The Xist A-repeat AUCG loop is a novel tetraloop motif

Tetraloops are some of the most abundant RNA structural elements, and early sequence comparisons on ribosomal RNAs revealed three predominant and thermodynamically stable families (Woese et al. 1990), namely the UNCG-, CUUG-, and GNRA-type tetraloops. Supplemental Figure S7 shows a structural comparison of the Xist AUCG tetraloop with three representative members of the three major families: UUCG (Allain and Varani 1995), CUUG (Jucker and Pardi 1995), and GAGA (Jucker et al. 1996). The AUCG loop is distinct from these known tetraloops in many aspects: Comparison of the base-pairing pattern reveals that most tetraloop structures are, in fact, diloops (a base pair is formed between the first and fourth base within the tetraloop), while the AUCG loop is a *bona fide* tetraloop. In the AUCG tetraloop, two bases stack at the 5' side, while in other tetraloops two or more bases are found stacking at the 3' side. The position of the turning phosphate between the second and third residue, as seen in the AUCG tetraloop, is distinct from other tetraloops, where the turning phosphate is found between the first and second nucleotide. The A6 H2 to G9 phosphate contact across the AUCG tetraloop is another distinct feature. Finally, the solvent exposure of the G9 purine is unusual and rarely observed. Therefore, we conclude that the Xist A-repeat hairpin 1 represents a novel RNA tetraloop motif.

The Xist RNA AUCG tetraloop hairpin is remarkably stable

To assess the stability of the AUCG tetraloop hairpin, we recorded UV melting curves. As most AUCG tetraloop hairpins in the A-repeats contain four GC base pairs (Supplemental Fig. S1), we compared the melting points of a 12-mer RNA, containing the most abundant A-repeat

TABLE 1. NMR restraints and structural statistics for the A-repeat AUCG tetraloop 14-mer

NMR restraints ^a	No RDCs	RDCs
Total restraints	413 (30 per residue)	438 (31 per residue)
NOE-derived distance restraints		292
Intra-residue		176
Inter-residue		116
H-bond distance restraints		25
Torsion angles ^b		96
RDCs	0	25
Structural statistics ^c		
NOE violations		
Number (>0.2 Å)	0.4	0.6
Maximum violations (Å)	0.33	0.35
Torsion violations		
Number (>5°)	0.2	0.1
Maximum violations (°)	20.7	5.9
RDC violations		
Number (>2 Hz)	12.5	1.6
Q-factor ^d	0.61 ± 0.34	0.23 ± 0.01
Maximum violations (Hz)	20	3.8
Global coordinate precision (Å) ^e		
All heavy atoms	0.82 ± 0.24	0.76 ± 0.19
5–10 (loop)	0.65 ± 0.2	0.65 ± 0.23
1–5 and 10–14 (stem)	0.66 ± 0.19	0.62 ± 0.18
RMSD from ideal geometry		
Bonds lengths (Å)	0.01 ± 0.00	0.01 ± 0.00
Bonds angles (°)	2.03 ± 0.03	2.07 ± 0.03
Average AMBER energy (kcal/mol)	−2422 ± 3	−2410 ± 11

^aFinal force constants used for square-well penalty functions were 32 kcal·mol^{−1}·Å^{−2} for NOEs and Watson–Crick hydrogen bonds, 200 kcal·mol^{−1}·rad^{−2} for torsion angles, and 0.1 kcal·mol^{−1}·Hz^{−2} for residual dipolar couplings (RDCs). The force constants were raised during the simulated annealing protocol as described in the Supplemental Material. Error bounds of ± 2 Hz were used for all residual dipolar coupling restraints.

^bA-form duplex backbone torsion angles derived from high-resolution crystal structures were used for the double-helical part as described in the Supplemental Material: α (290°–310°), β (170°–190°), γ (40°–60°), δ (70°–90°), ε (200°–220°), and ζ (280°–300°).

^cStructural statistics are averages per structure calculated for the bundle for the 10 selected AMBER-refined lowest-energy structures.

^dThe RDC Q-factor is calculated as $\text{rms}(D^{\text{calc}} - D^{\text{obs}})/\text{rms}(D^{\text{obs}})$, where D^{calc} and D^{obs} are calculated and observed RDC values, respectively (Cornilescu et al. 1998).

^eThe global coordinate precision is given as the average pairwise Cartesian coordinate RMSD of heavy atoms in the 10 selected AMBER-refined lowest-energy structures in the NMR ensemble.

sequence (gcccAUCGgggc), to gcccGAAAgggc and gcccUUUgggc 12-mers, containing an exceptionally stable apical tetraloop and an arbitrary UUUU tetraloop, respectively. Note, that the melting temperature of a GAAA tetraloop is only 5.5°C higher than the UUUU tetraloop, which, in turn, is more stable than AAAA or CCCC (Antao et al. 1991). Therefore, any tetraloop with a significantly higher melting point than UUUU may be considered unusually stable. Remarkably, in both NMR buffer (10 mM sodium phosphate pH 6.5, 100 mM NaCl) and high salt conditions (1 M NaCl), all three RNAs melted above 85°C, rendering a fit of the melting point difficult, as the maximum in the first derivative of the absorbance is not observed (data not shown). The melting curves and their first derivative are shown in Supplemental Figure S8. To be able to compare the stability of the three tetraloops, UV

melting curves were recorded under low salt conditions (0.5 mM sodium phosphate pH 6.5, 5 mM NaCl; condition A) and even more destabilizing conditions (0.5 mM sodium phosphate pH 6.5, 5 mM NaCl, 25% formamide; condition B). Table 2 shows the melting points for the three 12-mer RNAs. The AUCG tetraloop hairpin is marginally more stable than the UUUU tetraloop hairpin, while, as expected, the GAAA tetraloop is much more stable. We conclude that the remarkable stability of the A-repeat AUCG hairpin can mostly be attributed to the G-C-rich stem. It has been previously shown that the interruption of a four G-C base pair stem by one A-U base pair strongly reduces the thermal stability (Antao et al. 1991). Therefore, both the conserved AUCG loop and the complete conservation of G-C rich stems in the A-repeat region (Supplemental Fig. S1) support a functional role for these hairpins as stable, independent folding units.

A-repeat variants that disrupt dimerization in vitro are inactive in vivo

Previously, we have reported that, within a single 26-mer A-repeat, the predicted “hairpin 2” mediates dimerization with a second A-repeat in vitro (Fig. 1D; Supplemental Fig. S9; Duszczek et al. 2008). This observation suggests that dimerization or multimerization involving the predicted “hairpin 2” sequence could play a role for Xist function

in vivo, consistent with the fact that at least five A-repeats are required for Xist activity. Moreover, there is a high local concentration of the A-repeats in vivo, as they are connected by relatively short linkers and are localized at the X chromosome (Wutz et al. 2002). In order to test whether duplex formation via the predicted “hairpin 2” sequence can mediate multimerization in the context of multiple A-repeats, we analyzed a tandem A-repeat RNA, where two A-repeats are connected by a short (UAUACUU) linker, using NMR and analytical ultracentrifugation. The overall similarity and the presence of a characteristic uridine imino signal in the 1D imino-proton NMR spectrum of this double A-repeat RNA suggests that it folds in a similar topology as two single A-repeats (Figs. 3, 4A). The downfield-shifted imino signal has the same chemical shift as the uridine imino proton in the intermolecular A-U base pairs that

TABLE 2. RNA melting temperatures of a 12-mer A-repeat AUCG tetraloop hairpin compared to hairpins with a random UUUU and an unusually stable GAAA tetraloop

RNA hairpin sequence	T _m (°C) low salt; condition A	T _m (°C) low salt; 25% formamide; condition B
gccAUCGgggc	82.1 ± 0.4	69.0 ± 0.2
gccUUUUgggc	81.1 ± 0.1	67.8 ± 0.3
gccGAAAgggc	>85	73.7 ± 0.4

Melting temperatures were measured by UV melting (Supplemental Fig. S8) in triplicate.

mediate dimerization of a single A-repeat (see Fig. 1D). Sedimentation velocity analytical ultracentrifugation (AUC) confirms that the single 26-mer A-repeat (NMR-XCR) is a dimer at 0.1 mM concentration, consistent with our previous analysis by NMR (green in Fig. 4B; Duszczek et al. 2008). In contrast, the tandem A-repeat RNA, where the two A-repeats are connected by a short (UAUACUU) linker, is monomeric in solution (magenta in Fig. 4B). The monomeric state of the tandem A-repeat and the overall similarity of the NMR spectra of the single A-repeat 26-mer and the tandem A-repeats argues that the uridine imino NMR signal discussed above represents intramolecular A-U base pairs involving the two “hairpin 2” sequences in the tandem repeat (Fig. 4A). Thus the tandem A-repeat forms an intramolecular duplex via the “hairpin 2” sequences, as opposed to the intermolecular duplex formation that we have observed for the single 26-mer A-repeat. This suggests that multiple A-repeats may form higher-order structures that are stabilized by inter-repeat duplex formation involving the 3′ regions of the individual A-repeat sequences.

We then designed single A-repeat “hairpin 2” variants that should disrupt A-repeat dimerization. A “U20C” variant (blue in Fig. 4) disrupts the intermolecular A-U base pairs and, thus, should strongly weaken the dimerization. In a second variant, “hairpin 2” is capped by a GUAA tetraloop (black in Fig. 4), a member of the stable GNRA tetraloop family, which is expected to force the 3′ part of the A-repeat into a hairpin conformation. Analytical ultracentrifugation experiments of these two variants show that both A-repeat variants are monomeric at 0.1 mM concentration, in contrast to the original single 26-mer A-repeat (NMR-XCR).

The functional activity of the A-repeat dimerization variants was tested *in vivo* using an inducible *Xist* expression system in mouse embryonic stem cells. This assay has previously been used to show that chromosomal association and spreading of *Xist* RNA can be functionally separated from silencing (Wutz et al. 2002). In male mouse ES cells, single-copy variant *Xist* RNA transgenes are introduced into a locus on the X chromosome under control of an inducible promoter. When *Xist* expression is induced by addition of doxycycline, the activity of the *Xist* transgene

can be monitored by measuring cell survival in cell culture, as successful *Xist*-mediated repression of the single male X causes cell death. Southern blot analysis of the endogenous and the transgenic *Xist* DNA is used to confirm that the transgenes are successfully introduced into the ES cell system. Similar expression levels of the constructs are confirmed based on comparably sized *Xist* clusters by RNA FISH. Cell survival is then analyzed after five days of differentiation after *Xist* induction by doxycycline and expressed relative to cell survival of parallel cultures in which *Xist* was not induced.

Figure 4C shows Southern blot analysis of the endogenous *Xist* DNA and the *Xist* transgenic DNA, confirming that the transgenes were successfully introduced into the ES cell system (lanes 1 and 2 show two independent clones). NMR-XCR contains 12 copies of the NMR-modified A-repeat connected by short linkers of eight uridines, as does U20C, causing the bands to run at approximately equal height. GUAA appears as a shorter and weaker band as it has only seven copies of the NMR-modified A-repeat. Note that both variants (U20C and GUAA) contain the same NMR-modified “hairpin 1” sequence as NMR-XCR. Similar expression levels of all three constructs are demonstrated by comparably sized *Xist* clusters by RNA FISH (data not shown). Figure 4D shows cell survival of differentiating ES cells after five days in cell culture. The two controls shown are wild-type *Xist* (with 7.5 copies of the A-repeat) and Δ XS, in which the A-repeats are deleted, a construct that is unable to induce silencing (Wutz et al.

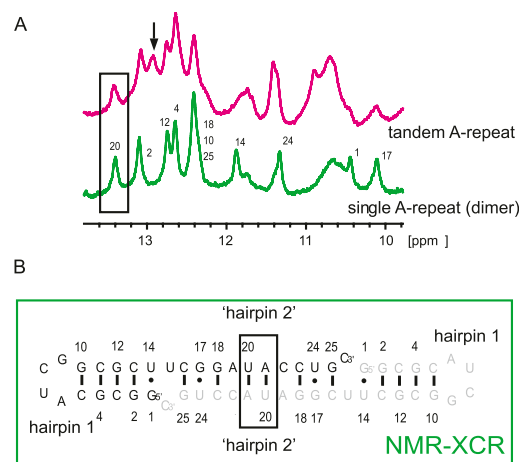


FIGURE 3. The tandem A-repeat folds in a similar way as a duplex of two single A-repeats. (A) 1D imino proton NMR spectra of the single A-repeat dimer (in green) and of a tandem A-repeat (in magenta; color coding as in Fig. 4) are shown, with residue numbers labeled. The uridine imino signal of A-U base pairs is seen in both RNAs (black box) and is consistent with intermolecular duplex formation of the single A-repeat and intramolecular duplex formation in the tandem repeat (cf. Fig. 4) via the “hairpin 2” region. The additional imino peak assigned (black arrow) in the tandem repeat spectrum presumably corresponds to the G-C base pair closest to the linker region between the two repeats. (B) Schematic drawing of the single A-repeat dimer. A-U base pairs, boxed in black, are involved in the dimerization platform that is *intermolecular* in the single A-repeat and *intramolecular* in the tandem A-repeat.

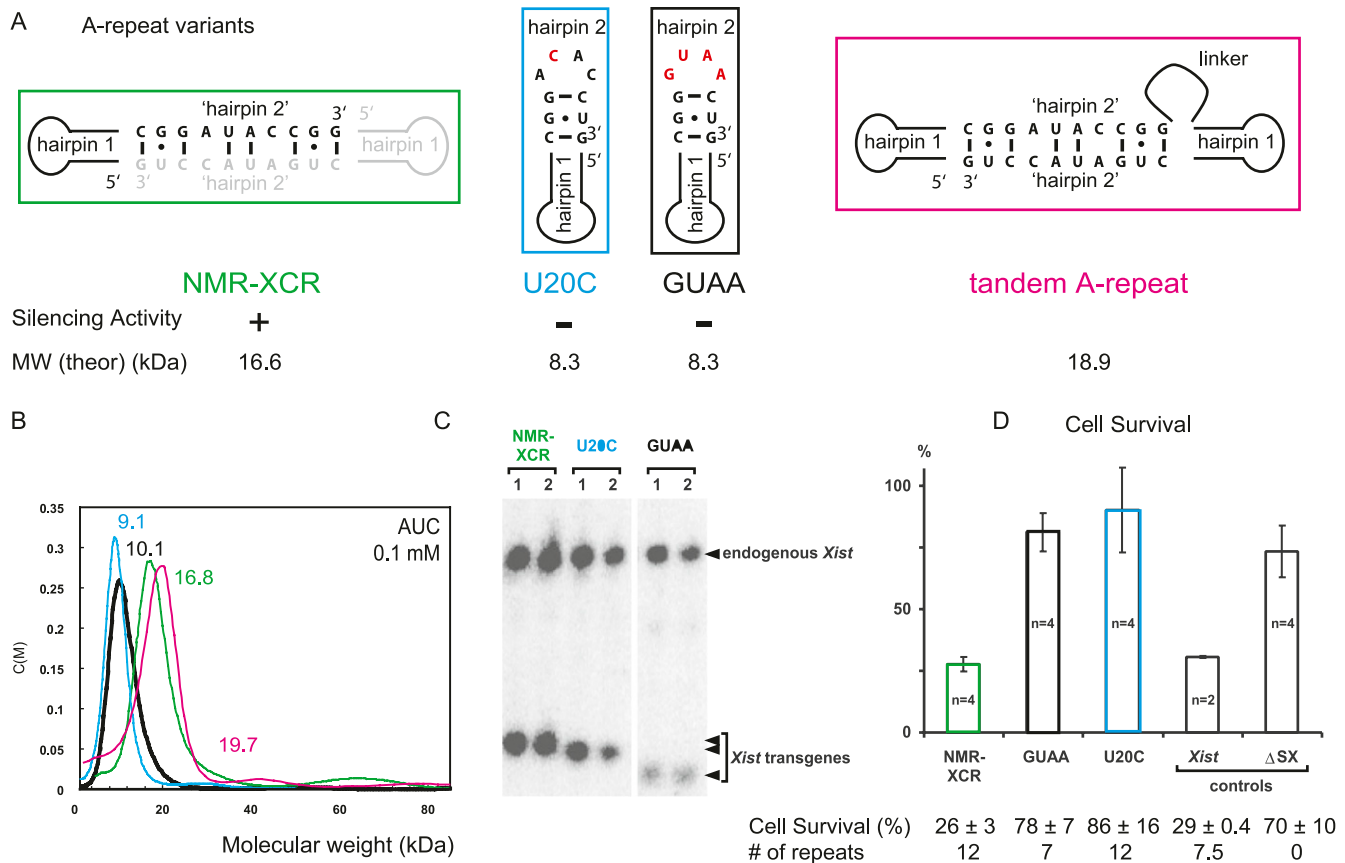


FIGURE 4. A-repeat “hairpin 2” variants that disrupt dimerization do not initiate X-inactivation in vivo. (A) The NMR-XCR single 26-mer A-repeat dimerizes in vitro via the “hairpin 2” sequence (in green). Variants of “hairpin 2,” U20C (in blue), and GUAA (in black) are expected to disrupt this dimerization. The expected fold of the tandem A-repeat is boxed in pink. The color coding of these four RNAs is followed throughout the figure. (B) The oligomerization of these four RNAs was characterized by analytical ultracentrifugation. All RNAs, except NMR-XCR, are monomeric. (C,D) Functional activity of the A-repeat variants in vivo. (C) Southern blot analysis shows successful introduction of the variant Xist RNA transgenes into a mouse ES cell system. (D) The percentage of cells (*n* is the number of measurements in triplicates) surviving in differentiating cultures in the presence of doxycyclin is given as a measure of silencing. The number of A-repeats in a given transgene construct and the silencing activity are indicated below each variant. Wild-type Xist and a variant where the A-repeats are deleted (Δ SX) are used as controls. Cell survival of the U20C and GUAA variants in cell culture is comparable with the variant where the A-repeats are deleted. An Xist transgene with modified A-repeats (NMR-XCR) apparently has slightly stronger activity than the wild-type Xist. The latter has less copies of the A-repeat; in fact, the NMR-modified A-repeat may have slightly less activity per A-repeat unit than the wild type.

2002). NMR-XCR silencing is comparable to wild-type Xist. Its stronger activity is likely due to the higher A-repeat copy number. Strikingly, the U20C A-repeat construct shows no activity compared to NMR-XCR. Also the GUAA variant is severely impaired in its silencing activity. Thus, both of these mutations that disrupt duplex formation via the “hairpin 2” region are inactive, consistent with the hypothesis that duplex formation via the 3' region of A repeats is required for silencing in vivo.

DISCUSSION

Correlation of the AUCG tetraloop hairpin structure with Xist silencing

Previous mutational studies on the Xist RNA A-repeats using the in vivo assay described above showed that the

length and sequence of the spacer by which the A-repeats are joined do not influence Xist activity and that at least five A-repeat copies are required for effective silencing (Wutz et al. 2002). Inversion of the hairpin 1 stem sequence does not influence activity, while disruption of stem base-pairing inhibits silencing. This is fully consistent with our three-dimensional structure and melting point measurements and supports the formation of a highly stable AUCG tetraloop hairpin with a G-C rich stem in the context of the complete A-repeat RNA. Interestingly, a variant with antisense A-repeats abolished silencing. As the sequence of stem 1 can be inverted without affecting activity, the loss of function in the antisense variant must be linked to the antisense loop. This further underpins the importance of not only a stable intact stem, but also of the AUCG tetraloop sequence for function. A-repeats with a scrambled hairpin 1 loop sequence (UAGC) show reduced (40% cell

survival after Xist induction compared to 25% for wild-type A-repeats) silencing activity. We speculate that the residual function could reflect some unique structural feature of the wild-type loop in the scrambled loop variant, in spite of the different loop sequence. The inactivity of the antisense variant and the diminished activity of the scrambled tetraloop variant argue that not the size but rather specific structural features of the tetraloop conformation are required for function. This is further underlined by the full conservation of the AUCG loop in mouse Xist A-repeats and an almost full conservation in placental mammals (Supplemental Fig. S1). Sequence variations are only found for the second and the fourth position of the tetraloop in human Xist RNA A-repeats. The uridine in the second position of the AUCG tetraloop nucleotide is replaced by adenosine in some repeats. This variation is consistent with the observed stacking of the corresponding nucleotide and most likely structurally silent. The fourth nucleotide is conserved as purine (Supplemental Fig. S1). As this base is exposed, the tetraloop structure is expected to tolerate the observed sequence variation. Taken together, our data suggest that the stable tetraloop fold is a conserved feature of the A-repeat RNA.

Functional implications for “hairpin 2” dimerization: A model of the Xist RNA A-repeats

Previous mutational analysis of the A-repeats did not reveal any sequence-function relationships for the predicted “hairpin 2,” except that its complete removal abolished Xist function. The results presented here suggest that duplex formation involving this sequence is required for silencing *in vivo*. This is supported by our mutational analysis which provides a clear correlation between the potential of the “hairpin 2” sequence for duplex formation and functional activity of A-repeats *in vivo*. A functional role for multimerization of A-repeats is a tempting hypothesis, (1) as a minimum of five repeats is required for Xist function, and (2) as the A-repeats, which are connected by short linkers, have a high local concentration on the Xi chromosome. Multimerization of several A-repeats either within a single Xist RNA molecule (as shown in Figs. 3, 4) or between different ones could thus help to bring the AUCG tetraloops in close spatial proximity as illustrated in Figure 5.

Recently, a 2D model of the complete Xist A-repeat RNA was proposed based on foot-printing, chemical modification, and FRET data (Maenner et al. 2010). In this model, the A-repeat region is predicted to fold into two long stem-loops involving inter-repeat duplex formation, albeit without the formation of the stable AUCG tetraloop for the hairpin 1 region. Our finding that the hairpin 1 adopts a highly stable and unique structure strongly argues that it represents an independent basic folding unit and that the tertiary fold of the A-repeat region is constructed by sequential formation of the individual tetraloop structures.

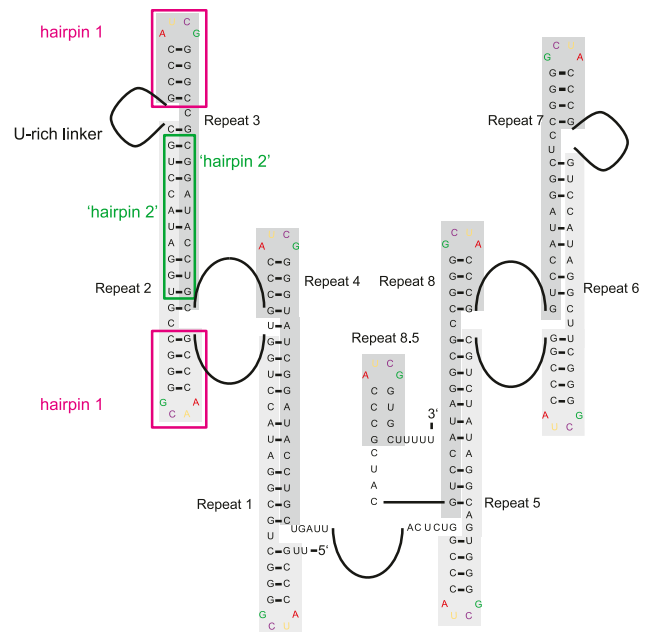


FIGURE 5. A model for the Xist RNA A-repeats. Inter-repeat dimerization of the “hairpin 2” sequence brings together the AUCG tetraloops of different A-repeats. Consecutive repeats are highlighted by different background color. The free energy (ΔG) of this fold at 37°C predicted by the Mfold 3.2 software (Zuker 2003) is -69.8 (kcal/mol) and is consistent with the structural and biochemical data reported here and in Maenner et al. (2010).

Significantly, the model for the structure of the full A-repeat region shown in Figure 5 reconciles all our structural and functional data as well as the foot-printing, chemical modification and FRET data that were recently used to propose a different model (Supplemental Fig. S10; Maenner et al. 2010). This previously proposed model is not compatible with our *in vivo* data as it cannot explain the loss of activity of the U20C A-repeat variant (Fig. 4). Mutation of this uridine should not influence the two-dimensional structure, as proposed by Maenner et al., but will strongly destabilize the inter-repeat duplex formation in the model, shown in Figure 5.

There are several implications from our structural and biochemical analysis for molecular functions of the A-repeats. (1) Mutational analysis of the hairpin 1 and the AUCG tetraloop sequence demonstrates that the integrity of the loop is required for Xist-mediated silencing. This suggests that the AUCG loop could be specifically recognized by a *trans*-acting factor. For example, subunits of the PRC2 complex that have been shown to play a role in Xist-mediated silencing are potential binding partners (Zhao et al. 2008; Kanhere et al. 2010). (2) The presentation of multiple AUCG motifs by inter-repeat duplex formation (Fig. 5) could support cooperative binding of the Xist A-repeats to multiple RNA binding domains, which, individually, might have low AUCG-binding affinity. This hypothesis is in agreement with the observation that several PRC2 core components bind more efficiently to multiple

A-repeat RNA *in vitro* (Maenner et al. 2010), most notably the Suz12 component that needs the complete A-repeat region for efficient binding. (3) The inter-repeat duplex formation could also contribute to the differential treatment of the two X chromosomes by prevention of Xist-Tsix duplex formation as suggested by Ogawa et al. (2008). The proposed RNAi-like Xist repression on the Xi by short RNAs originating from these Xist-Tsix duplexes could then be prevented, allowing Xist RNA to accumulate and initiate silencing. (4) Finally, by intermolecular multimerization of several Xist RNA molecules, the A-repeats could serve as a nucleation center by which PRC2 recruitment and thus silencing is spread over the Xi. Although we cannot exclude the possibility that RNA binding proteins alter the fold of the Xist RNA *in vivo*, our observation that mutations that abolish duplex formation also inhibit silencing suggests that the Xist A-repeats may act as a multimerization platform that is required for the functional activity of Xist RNA.

MATERIALS AND METHODS

Sample preparation

^{13}C , ^{15}N uniformly labeled and unlabeled Xist RNA A-repeat constructs were prepared as described previously (Duszczek et al. 2008) and are listed in Supplemental Table S2. NMR samples were prepared in 10 mM $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$, pH 6.0, 100 mM NaCl, 0.02 mM EDTA, 0.02% azide in 95% H_2O , 5% D_2O , or 100% D_2O . RNA concentrations ranged between 0.8 and 1.2 mM. Samples for analytical ultracentrifugation were prepared in NMR buffer at 0.1 mM concentration. Prior to all measurements, samples were heated to 95°C for 5 min, followed by snap-cooling on ice with the rationale to trap the kinetically favored intramolecular monomeric hairpin conformation over a possible intermolecular dimer. The homogenous formation of a monomeric 14-mer hairpin was unambiguously confirmed by measuring the relative intensity of cross and diagonal peaks in a HNN-COSY spectrum (Dingley and Grzesiek 1998) recorded on a 50% ^{13}C , ^{15}N -labeled, 50% unlabeled sample in H_2O (Supplemental Fig. S9; Duszczek et al. 2008). Partial alignment of RNA for residual dipolar coupling measurements was achieved by adding 18 mg mL^{-1} filamentous bacteriophage Pf1 (ASLA Biotech) to an unlabeled sample.

NMR spectroscopy

NMR experiments were recorded at 5°C (assignment and NOESY spectra involving exchangeable protons) or 25°C (assignment, NOESY spectra involving nonexchangeable protons, measurements of J-couplings, and RDCs) on Bruker DRX600, 800, or 900 spectrometers equipped with cryoprobes. Spectra were processed with NMRPipe (Delaglio et al. 1995) and analyzed using NMRVIEW (Johnson and Blevins 1994). Chemical shifts in the Xist RNA 14-mer AUCG hairpin were assigned as described previously (Duszczek and Sattler 2011).

Structure calculation and refinement

Generation of hydrogen bond and inter-proton distance restraints, torsion angle restraints, and residual dipolar coupling

restraints as input for the structure calculations is described in detail in the Supplementary Information. Initial structures were calculated with ARIA/CNS 1.2 (Brunger et al. 1998; Linge et al. 2003) with a mixed Cartesian and torsion angle dynamics simulated annealing protocol from an extended starting structure using NOE, torsion angle, hydrogen bonding, and planarity restraints. Based on consistency between back-calculated proton-proton distances and experimental NOEs and on low energy criteria, 20 best ARIA-calculated structures were selected for refinement with the SANDER module of AMBER 9 (University of California, San Francisco, <http://ambermd.org>). In the refinement protocol, 25 RDCs were included with a single floating alignment tensor. The restrained MD refinement was performed with the Cornell et al. (1995) force field with the generalized-Born solvation model (Bashford and Case 2000). The final ensemble of 10 structures was selected from the 20 AMBER-refined structures based on the agreement between experimental and back-calculated proton chemical shifts (using the program NUCHEMICS [Wijmenga et al. 1997]) for the loop nucleotides while maintaining low restraint violations and force field energies. RMSDs and angles in the final structures were calculated using MOLMOL (Koradi et al. 1996). Molecular graphics were generated using Pymol (DeLano 2002).

In vivo studies

Cloning of variant Xist constructs, generation of transgenic cell lines, and growing of ES cell cultures were performed as described previously (Wutz et al. 2002).

Analytical ultracentrifugation

The oligomeric state of the A-repeats variants was investigated by monitoring sedimentation properties in centrifugation experiments at 0.1 mM concentration in NMR buffer. The sedimentation velocity profiles were collected by monitoring the absorbance signal at 260 nm as the samples were centrifuged in a Beckman Optima XL-A centrifuge fitted with a four-hole AN-60 rotor and double-sector aluminium centerpieces (45,000 rpm, 20°C). Molecular weight distributions were determined by the C(s) method implemented in the Sedfit software (Schuck 2000). Buffer density and viscosity corrections were made according to data published by Laue et al. (1992). The partial specific volume of the A-repeat RNA variants was taken as 0.53 mL/g.

RNA UV melting studies

RNA oligonucleotides were purchased from Fisher Scientific. They were dissolved to OD 80–85 in NMR buffer without further purification and diluted 20-fold into either H_2O (condition A) or H_2O , 25% formamide (condition B). Absorbance versus temperature was recorded at 254 nm (condition A) or 270 nm (condition B) on a Varian-Cary UV-visible spectrophotometer equipped with a Peltier temperature control device. Standard 1-cm-path-length quartz cuvettes were used for the measurements. RNA melting curves were recorded at a heating rate of 0.2°C/min from 10°C to 95°C in triplicates for each condition. Before each measurement the diluted samples were heated to 95°C for 5 min followed by snap-cooling on ice for 5 min. Melting temperatures were derived from the maxima of first derivatives of the melting curves

calculated using Savitzky-Golay smoothing (polynomial order 2, 20 window points) after reducing and evenly spreading the data points (every 5 points) using the Origin 8.1G software.

Deposition of coordinates, chemical shifts, and restraints

Coordinates for the ensemble of 10 AMBER-refined structures of the 14-mer AUCG tetraloop hairpin have been deposited into the Protein Data Bank (accession code 2Y95). NMR chemical shift assignments and other restraints used to calculate the structural ensemble have been deposited into the BMRB (accession code 16714).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank Bernd Simon for help with NMR experiments and structure calculations, Richard Stefl for sharing the AMBER-refinement protocol, Frank Nelissen, Katia Zanier, and Alexander Gasch for help with RNA sample preparation, Arthur van Aerschot and Guy Schepers (K.U.Leuven) for initial melting temperature studies, and Brighton Samatanga and Christophe Maris (ETH Zurich) for assistance in recording the UV melting data. We acknowledge the Bayerisches NMR Zentrum (BNMRZ) in Garching, the Biomolecular NMR Centre in Frankfurt, and Wolfgang Bermel (Bruker, Karlsruhe) for spectrometer time. We are grateful to Eric Westhof, Sybren Wijmenga, Elena Conti, Asifa Akhtar, Elisa Izaurrealde, and Alexander Beribisky for discussions and critical reading of the manuscript. M.D. acknowledges support by an “E-STAR” Marie Curie Host fellowship for Early Stage Research Training funded by the EC’s FP6, contract MEST-CT-2004-504640. This work was supported by the EU STREP FSG-V-RNA, contract LSHG-CT-2004-503455, and the DFG funded Cluster of Excellence, Center for integrated Protein Science, Munich (CiPS-M). Initial work was supported by the EMBL Heidelberg. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions: M.M.D. transcribed and purified RNA for structural and biochemical analysis, performed the structure determination, and recorded and analyzed the RNA UV melting curves. A.W. tested functional activity of Xist RNA variants in mouse ES cells. V.R. performed analytical ultracentrifugation experiments and analyzed the data. M.M.D. and M.S. designed the study and wrote the paper. All authors discussed the results and commented on the manuscript.

Received March 23, 2011; accepted August 10, 2011.

REFERENCES

- Allain FHT, Varani G. 1995. Structure of the P1 helix from group-I self-splicing introns. *J Mol Biol* **250**: 333–353.
- Antao VP, Lai SY, Tinoco I. 1991. A thermodynamic study of unusually stable RNA and DNA hairpins. *Nucleic Acids Res* **19**: 5901–5905.
- Avner P, Heard E. 2001. X-chromosome inactivation: counting, choice and initiation. *Nat Rev Genet* **2**: 59–67.
- Bashford D, Case DA. 2000. Generalized Born models of macromolecular solvation effects. *Annu Rev Phys Chem* **51**: 129–152.
- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, et al. 1998. Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**: 905–921.
- Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* **117**: 5179–5197.
- Cornilescu G, Marquardt JL, Ottiger M, Bax A. 1998. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J Am Chem Soc* **120**: 6836–6837.
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. 1995. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biol NMR* **6**: 277–293.
- DeLano WL. 2002. *The PyMOL molecular graphics system*. DeLano Scientific, San Carlos, CA.
- de Napoles M, Mermoud JE, Wakao R, Tang YA, Endoh M, Appanah R, Nesterova TB, Silva J, Otte AP, Vidal M, et al. 2004. Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev Cell* **7**: 663–676.
- Dingley AJ, Grzesiek S. 1998. Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide $^2J_{NN}$ couplings. *J Am Chem Soc* **120**: 8293–8297.
- Duszczak MM, Sattler M. 2011. 1H , ^{13}C , ^{15}N and ^{31}P chemical shift assignments of a human Xist RNA A-repeat tetraloop hairpin essential for X-chromosome inactivation. *Biomol NMR Assign* doi: 10.1007/s12104-011-9328-z.
- Duszczak MM, Zanier K, Sattler M. 2008. A NMR strategy to unambiguously distinguish nucleic acid hairpin and duplex conformations applied to a Xist RNA A-repeat. *Nucleic Acids Res* **36**: 7068–7077.
- Henn M, Jurkschat K, Mansfeld D, Mehring M, Schürmann M. 2004. Synthesis and structure of and DFT-studies on 1,3,5-[P(O)(i-PrO) $_2$] $_3C_6H_3$ and its $CHCl_3$ adduct: Analysis of the $Cl_3C-H \cdots OP$ hydrogen bond. *J Mol Struct* **697**: 213–220.
- Johnson BA, Blevins RA. 1994. NMR View: A computer program for the visualization and analysis of NMR data. *J Biol NMR* **4**: 603–614.
- Jucker FM, Pardi A. 1995. Solution structure of the CUUG hairpin loop: A novel RNA tetraloop motif. *Biochemistry* **34**: 14416–14427.
- Jucker FM, Heus HA, Yip PF, Moors EHM, Pardi A. 1996. A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J Mol Biol* **264**: 968–980.
- Kanhere A, Viiri K, Araújo CC, Rasaiyaah J, Bouwman RD, Whyte WA, Pereira CF, Brookes E, Walker K, Bell GW, et al. 2010. Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Mol Cell* **38**: 675–688.
- Koradi R, Billeter M, Wüthrich K. 1996. MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* **14**: 51–55.
- Laue T, Shah B, Ridgeway T, Pelletier S. 1992. *Analytical ultracentrifugation in biochemistry and polymer science*. Royal Society of Chemistry, Cambridge, UK.
- Lee JT, Lu N. 1999. Targeted mutagenesis of *Tsix* leads to nonrandom X inactivation. *Cell* **99**: 47–57.
- Linge JP, Habeck M, Rieping W, Nilges M. 2003. ARIA: Automated NOE assignment and NMR structure calculation. *Bioinformatics* **19**: 315–316.
- Lucchini JC, Kelly WG, Panning B. 2005. Chromatin remodeling in dosage compensation. *Annu Rev Genet* **39**: 615–651.
- Maenner S, Bland M, Fouillen L, Savoye A, Marchand V, Dubois AS, Sanglier-Cianféran S, Van Dorsselaer A, Clerc P, Avner P, et al.

2010. 2-D structure of the A region of Xist RNA and its implication for PRC2 association. *PLoS Biol* **8**: e1000276. doi: 10.1371/journal.pbio.1000276.
- Ogawa Y, Sun BK, Lee JT. 2008. Intersection of the RNA interference and X-inactivation pathways. *Science* **320**: 1336–1341.
- Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N. 1996. Requirement for *Xist* in X-chromosome inactivation. *Nature* **379**: 131–137.
- Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, de la Cruz CC, Otte AP, Panning B, Zhang Y. 2003. Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**: 131–135.
- Plath K, Talbot D, Hamer KM, Otte AP, Yang TP, Jaenisch R, Panning B. 2004. Developmentally regulated alterations in Polycomb repressive complex 1 proteins on the inactive X chromosome. *J Cell Biol* **167**: 1025–1035.
- Royce-Tolland ME, Andersen AA, Koefman HR, Talbot DJ, Wutz A, Tonks ID, Kay GF, Panning B. 2010. The A-repeat links ASF/SF2-dependent *Xist* RNA processing with random choice during X inactivation. *Nat Struct Mol Biol* **17**: 948–954.
- Schoeftner S, Sengupta AK, Kubicek S, Mechtler K, Spahn L, Koseki H, Jenuwein T, Wutz A. 2006. Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J* **25**: 3110–3122.
- Schuck P. 2000. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys J* **78**: 1606–1619.
- Sheardown SA, Duthie SM, Johnston CM, Newall AE, Formstone EJ, Arkell RM, Nesterova TB, Alghisi GC, Rastan S, Brockdorff N. 1997. Stabilization of *Xist* RNA mediates initiation of X chromosome inactivation. *Cell* **91**: 99–107.
- Silva J, Mak W, Zvetkova I, Appanah R, Nesterova TB, Webster Z, Peters AHFM, Jenuwein T, Otte AP, Brockdorff N. 2003. Establishment of histone H3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 Polycomb group complexes. *Dev Cell* **4**: 481–495.
- Wahl MC, Sundaralingam M. 1997. C—H...O hydrogen bonding in biology. *Trends Biochem Sci* **22**: 97–102.
- Wijmenga SS, Kruithof M, Hilbers CW. 1997. Analysis of H-1 chemical shifts in DNA: Assessment of the reliability of H-1 chemical shift calculations for use in structure refinement. *J Biol NMR* **10**: 337–350.
- Woese CR, Winker S, Gutell RR. 1990. Architecture of ribosomal RNA: Constraints on the sequence of “tetra-loops”. *Proc Natl Acad Sci* **87**: 8467–8471.
- Wutz A, Rasmussen TP, Jaenisch R. 2002. Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nat Genet* **30**: 167–174.
- Zhao J, Sun BK, Erwin JA, Song J-J, Lee JT. 2008. Polycomb proteins targeted by a short repeat RNA to the mouse X-chromosome. *Science* **322**: 750–756.
- Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415.