BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Nucleotide Contributions to the Structural Integrity and DNA Replication Initiation Activity of Noncoding Y RNA

Iren Wang,[†] Madzia P. Kowalski,[‡] Alexander R. Langley,[‡,@] Raphaël Rodriguez,[§,▽] Shankar Balasubramanian,[§,∥] Shang-Te Danny Hsu,*[,†,§,⊥,#] and Torsten Krude*[,‡]

[†]Institute of Biological Chemistry, Academia Sinica, 128, Section 2, Academia Road, Taipei 11529, Taiwan

[‡]Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, United Kingdom

[§]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

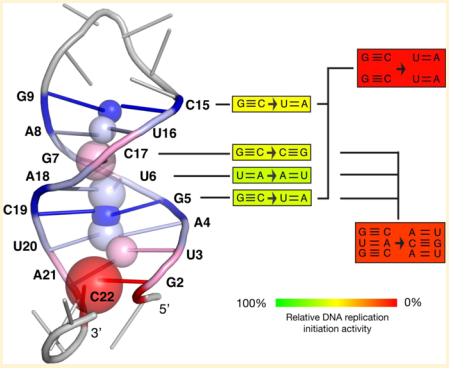
[∥]Cancer Research UK Cambridge Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, United Kingdom

[⊥]Institute of Biochemical Sciences, National Taiwan University, 1, Section 4, Roosevelt Road, Taipei 10617, Taiwan

[#]Institute of Bioinformatics and Structural Biology, National Tsing Hua University, 101, Section 2, Kuang-Fu Road, Hsinchu 30013, Taiwan

Supporting Information

ABSTRACT: Noncoding Y RNAs are small stem−loop RNAs that are involved in different cellular processes, including the regulation of DNA replication. An evolutionarily conserved small domain in the upper stem of vertebrate Y RNAs has an essential function for the initiation of chromosomal DNA replication. Here we provide a structure−function analysis of this essential RNA domain under physiological conditions. Solution state nuclear magnetic resonance and far-ultraviolet circular dichroism spectroscopy show that the upper stem domain of human Y1 RNA adopts a locally destabilized A-form helical structure involving eight Watson−Crick base pairs. Within this helix, two G:C base pairs are highly stable even at elevated temperatures and therefore may serve as clamps to maintain the local structure of the helix. These two stable G:C base pairs frame three unstable base pairs, which are located centrally between them. Systematic substitution mutagenesis results in a disruption of the ordered A-form helical structure and in the loss of DNA replication initiation activity, establishing a positive correlation between folding stability and function. Our data thus provide a structural basis for the evolutionary conservation of key nucleotides in this RNA domain that are essential for the functionality of noncoding Y RNAs during the initiation of DNA replication.



Y RNAs are abundant small noncoding RNAs that are present in all vertebrates.[1,2] In humans, there are four distinct Y RNAs, termed hY1, hY3, hY4, and hY5, which range in size from 83 to 112 nucleotides. Despite their relatively small size, Y RNAs are involved in several independent cellular pathways.[3−6] They were described originally as the RNA component of Ro ribonucleoprotein particles (Ro RNPs) that also contain proteins Ro60 and La.[7,8] Ro RNPs are currently implicated in noncoding RNA quality control and nucleolytic RNA degradation.[3,9]

Independent of Ro RNP assembly, the function of Y RNAs is essential in the replication of chromosomal DNA in vertebrates.[10−14] In an unbiased screen for DNA replication factors, Y RNAs were biochemically purified from human cell extracts as essential components for the reconstitution of chromosomal DNA replication in a cell free system.[10] In this system, chromosomal DNA replication is initiated in late G1 phase template nuclei when they are incubated in a cytosolic cell extract from proliferating cells.[15] Depletion or degradation of endogenous Y RNAs from this extract abrogates initiation, while addition of exogenous Y RNAs to an extract depleted of its endogenous Y RNAs restores initiation.[10,12,13] Furthermore, degradation of Y RNAs in proliferating somatic cells inhibits DNA replication, and their functional depletion in fish and amphibian embryos leads to arrested development and early embryonic death after the mid-blastula transition.[10,11,13,16]

Structural analyses have shown that the different cellular roles of Y RNAs correlate with distinct and evolutionarily conserved structural elements. All Y RNAs form characteristic stem−loop structures.[17,18] The 5′ and 3′ termini hybridize to form the lower and upper stem domains with a large internal loop (Figure 1). The lower stem domain contains a bulged cytosine and has a short single-stranded 3′-polyuridine tail at one end and a small single-stranded loop at the other end. This lower stem domain is essential for Ro RNP formation and contains the binding sites for Ro60 and La proteins.[19]

Received: April 17, 2014
Revised: August 21, 2014
Published: August 25, 2014

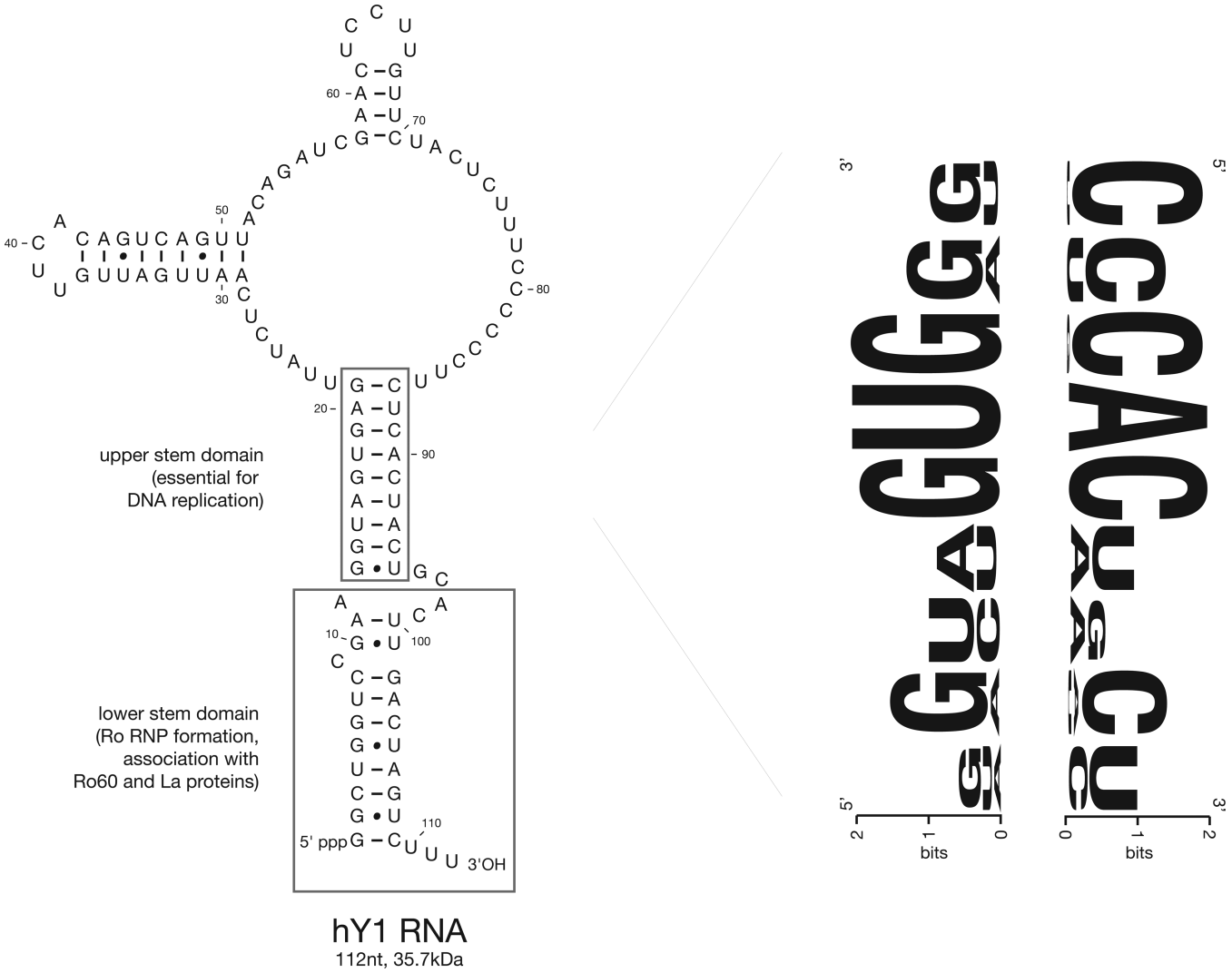


Figure 1. Human Y1 RNA and evolutionary conservation of the upper stem domain. The nucleotide sequence and secondary structure of human Y1 RNA, as derived from sequence alignment and enzymatic and chemical probing,[17] are shown on the left. The conserved lower and upper stem domains are boxed. Evolutionary conservation of nucleotides in the upper stem is shown by WebLogos on the right.

Enzymatic and chemical probing experiments suggest that the highly conserved upper stem domain is double-stranded, as predicted from computational free energy minimization and nucleotide sequence alignment.[17,18,20] Systematic mutagenesis and functional testing of Y RNAs have established that the upper stem domain is not only essential but also fully sufficient for Y RNA function in chromosomal DNA replication.[13] This domain has a highly conserved GUG-CAC nucleotide sequence element (Figure 1), and mutations disrupting this sequence element in hY1 RNA inactivate its DNA replication initiation function.[13] Interestingly, the adenosine residue embedded in this motif (position A90 in hY1 RNA) appears to have a weak reactivity to chemical probing by dimethyl sulfate,[17,18] suggesting that base pairing within this essential domain may exhibit significant conformational fluctuations. Collectively, these observations suggest that structural elements in the upper stem domain may be linked to the DNA replication initiation function of Y RNAs. However, there is a lack of structural information about this domain under physiological conditions.

In this study, we examine the structure−function relationship of the conserved upper stem of hY1 RNA using biophysical and functional assays. We combine solution state nuclear magnetic resonance (NMR) and far-UV circular dichroism (CD) spectroscopy with systematic mutagenesis of the upper stem, and functional testing of DNA replication initiation activity of the resulting RNAs. Imino proton NMR signals are ideal reporters of nucleic acid structures and dynamics, because they can be observed only when they are involved in stable hydrogen bonds.[21,22] The chemical shifts and transverse relaxation times (T_2) of the imino protons reflect the strength of the hydrogen bonding and thus provide information about local secondary structures. Furthermore, we employ ^{15}N spin relaxation anaylsis to monitor the local dynamics of the imino nitrogen atoms as well as CLEANEX-PM measurements to determine the solvent exchange rates of individual imino protons.[23,24] These quantitative NMR analyses are complemented by global structural analysis using far-UV CD spectroscopy. While solution state NMR spectroscopy provides insights into the thermal stability and dynamics of individual imino groups that are hydrogen bonded, thermal melting experiments by far-UV CD spectroscopy provide quantitative information about the relative proportions of different functional conformations in the population of a given RNA in solution. Collectively, our

dx.doi.org/10.1021/bi500470b | *Biochemistry* 2014, 53, 5848−5863

Table 1

wt fw	5′-GATTTAGGTGACACTATAGGTAGTGAGTTCTTCTCACTACTTT-3′
wt rev	5′-AAAGTAGTGAGAAGAACTCACTACCTATAGTGTCACCTAAATC-3′
mut1 fw	5′-GATTTAGGTGACACTATAGGTAACAAGTTCTTCTTGTTACTTT-3′
mut1 rev	5′-AAAGTAACAAGAAGAACTTGTTACCTATAGTGTCACCTAAATC-3′
mut2 fw	5′-GATTTAGGTGACACTATAGGTAGAGAGTTCTTCTCTCTACTTT-3′
mut2 rev	5′-AAAGTAGAGAGAAGAACTCTCTACCTATAGTGTCACCTAAATC-3′
mut3 fw	5′-GATTTAGGTGACACTATAGGTAGTCAGTTCTTCTGACTACTTT-3′
mut3 rev	5′-AAAGTAGTCAGAAGAACTGACTACCTATAGTGTCACCTAAATC-3′
mut4 fw	5′-GATTTAGGTGACACTATAGGTATTGAGTTCTTCTCAATACTTT-3′
mut4 rev	5′-AAAGTATTGAGAAGAACTCAATACCTATAGTGTCACCTAAATC-3′
mut5 fw	5′-GATTTAGGTGACACTATAGGTAGTGATTTCTTATCACTACTTT-3′
mut5 rev	5′-AAAGTAGTGATAAGAAATCACTACCTATAGTGTCACCTAAATC-3′
mut6 fw	5′-GATTTAGGTGACACTATAGGTATTGATTTCTTATCAATACTTT-3′
mut6 rev	5′-AAAGTATTGATAAGAAATCAATACCTATAGTGTCACCTAAATC-3′

biophysical data provide evidence of a partially destabilized A-form helical structure of the upper stem domain of hY1 RNA and suggest a cooperative contribution of individual conserved base pairs to the structural stability and DNA replication function of this domain.

MATERIALS AND METHODS

Synthesis and Purification of Wild-Type and Mutant Y RNAs. Templates for the *in vitro* transcription of wild-type and mutant RNAs under the control of an SP6 promoter site were generated by annealing synthetic DNA oligonucleotides (Sigma-Genosys). Sequences of the oligonucleotide pairs are listed in Table 1.

RNAs were synthesized by *in vitro* transcription using SP6 RNA polymerase, as described previously.[10,13] Isotope-labeled RNAs were synthesized by replacing all four rNTPs by [^{13}C,^{15}N]rNTPs (Silantes). Nucleotide sequences of the 25 nucleotide transcripts are listed in Table 2.

Table 2

wt	5′-GGUAGUGAGUUCUUCUCACUACUUU-3′
mut1	5′-GGUAACAAGUUCUUCUUGUUACUUU-3′
mut2	5′-GGUAGAGAGUUCUUCUCUCUACUUU-3′
mut3	5′-GGUAGUCAGUUCUUCUGACUACUUU-3′
mut4	5′-GGUAUUGAGUUCUUCUCAAUACUUU-3′
mut5	5′-GGUAGUGAUUUCUUAUCACUACUUU-3′
mut6	5′-GGUAUUGAUUUCUUAUCAAUACUUU-3′

RNAs were purified by anion exchange chromatography on a MonoQ column (Fisher Scientific). RNA was loaded in 50 mM sodium acetate (pH 5.2), washed in 0.1 M NaCl and 50 mM sodium acetate (pH 5.2), and eluted with a linear 0.1 to 0.6 M NaCl gradient in 50 mM sodium acetate (pH 5.2). The wt and mutant RNAs eluted between 0.35 and 0.42 M NaCl. The size and purity of all *in vitro* synthesized RNA were confirmed using 8 M urea denaturing polyacrylamide gel electrophoresis and staining with SYBR Gold (Invitrogen). Single-stranded DNA oligonucleotides 43, 37, 20, and 16 nucleotides in length (Sigma Genosys) were used as molecular weight markers.

Computational Analyses. Secondary RNA structures were calculated using the Mfold version 3.2 RNA folding algorithm (web server at http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/rna-form1.cgi) under default conditions.[25,26]

WebLogo analysis was performed by using sequence data from a dialign sequence alignment of all complete and annotated vertebrate Y RNAs[1] [available as a CLUSTAL X (1.81) multiple sequence alignment at http://www.bioinf.uni-leipzig.de/Publications/SUPPLEMENTS/07-004/Y_dialign.aln]. Nucleotide sequences corresponding to the nine consecutive bases on the 5′ side of the upper stem and the complementary bases located on the 3′ side were identified (between positions 13 and 22 and positions 114 and 127 of the clustal X alignment, respectively). WebLogos were generated from these aligned sequence fragments by the online tools available at http://weblogo.berkeley.edu.[27]

Cell Culture, Extracts, and Fractionation. Human EJ30 bladder carcinoma cells were cultured and synchronized in late G1 phase by mimosine, and template nuclei were isolated from these cells as described previously.[15,28] The cytosolic extract of asynchronously proliferating human HeLa cells was obtained from Cilbiotech (Mons, Belgium) and fractionated by anion exchange chromatography on Q-sepharose and arginine sepharose into protein fractions QA and ArFT, respectively, as described previously.[29,30] Fractions were concentrated to 10 mg/mL protein using Vivaspin 6 spin columns (Sartorius).

DNA Replication *in Vitro* Assays. Standard *in vitro* DNA replication reactions were performed as described previously,[10,13,31] containing template nuclei from late G1 phase human EJ30 cells, a buffered mix or ribonucleoside and deoxyribonucleoside triphosphates and digoxigenin-11-dUTP (Roche) as a tracer. For functional testing, the purified RNAs were added at a concentration of 100 nM together with 7 μg of each protein fraction (QA and ArFT) to a reaction volume of 50 μL. To analyze DNA replication reactions, nuclei were fixed after a 2.5 h reaction time and spun onto polylysine-coated glass coverslips. Digoxigenin-labeled replicated DNA was detected by anti-digoxigenin fluorescein-conjugated F_{ab} fragments (Roche), and total DNA was counterstained with propidium iodide as described previously.[15,29,30] Confocal fluorescence microscopy was performed on a Leica SP1 microscope using 40× lens magnification; individual channels were recorded simultaneously, and the percentages of replicating nuclei were determined from randomly chosen microscopic fields. At least 300 nuclei were scored per reaction.

NMR Spectroscopy. ^{13}C- and ^{15}N-labeled wild-type (wt) hY1 stem RNA and unlabeled RNAs were synthesized for NMR analysis on a milligram scale and purified. The RNA samples were precipitated with ethanol and lyophilized, and their quantities were determined by weighing. All NMR samples were dissolved in 100 mM potassium phosphate buffer (pH 7.5) with 2 mM $MgCl_2$ in 10% D_2O (v/v) or in 100%

dx.doi.org/10.1021/bi500470b | *Biochemistry* 2014, 53, 5848−5863

D_2O, with sample concentrations between 0.2 and 0.8 mM. All NMR experiments were conducted using a Bruker AVANCE 500, AVANCE 600, AVANCE III 600, or AVANCE 800 spectrometer equipped with a cryogenic probe. The hydrogen bond pairings were confirmed through the cross-hydrogen bond scalar coupling using the quantitative $^2J_{NN}$ HNN-COSY experiment[32] at 278 K. The quantitative values of $^{2Hb}J_{NN}$ coupling constants are calculated by the ratio of cross-peak to diagonal peak intensities $[I_c/I_d = \tan^2(2\pi J_{NN}T)]$, where the transfer time $2T$ used for HNN-COSY is 30 ms. The sequential assignments of imino and aromatic protons were obtained by a series of homonuclear two-dimensional (2D) nuclear Overhauser effect spectroscopy (NOESY) experiments[21] at various sample temperatures ranging between 278 and 310 K, with various mixing times (100, 200, and 300 ms) to resolve spectral overlapping problems. The 2D homonuclear TOCSY experiments[33] were conducted to gain information about each pyrimidine spin system (H5 and H6 protons).[34,35] The NMR experiments were processed and analyzed with Bruker TopSpin version 3.0 or with NMRPipe[36] and Sparky packages (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco, CA). The imino proton T_2 values were determined by recording a series of one-dimensional (1D) NMR spectra with the 1-1 solvent suppression pulse sequence with different relaxation delays (τ = 1.5, 2.5, 4, 6.5, 9, 11.5, 16.5, 21.5, 36.5, and 60 ms) at 278, 288, 298, and 308 K. The transverse relaxation times were obtained by fitting the observed peak intensities as a function of relaxation delay to a single-exponential decay.

^{15}N spin relaxation measurements of the imino nitrogens were recorded at 278, 288, and 298 K at 14.1 T. For the longitudinal relaxation measurements, the recovery delays were set to 100, 250, 500, 1000, and 1500 ms; for the transverse relaxation measurements, the relaxation delays were set to 17, 34, 51, 68, 85, and 119 ms. The relaxation time constants were derived by extracting peak heights using the rate analysis module within the Sparky package and subsequently fit to a single-exponential decay function using the software package GraphPad Prism (version 6.00 for Mac, GraphPad Software, La Jolla, CA) to extract the longitudinal (T_1) and transverse (T_2) relaxation time constants.

The hydrogen exchange rates of imino protons were determined using the CLEANEX-PM experiment[23,24] with spin-lock mixing times of 5, 10, 15, 25, 50, 75, 100, 125, 150, and 200 ms, and the water suppression was achieved by using a 3-9-19 Watergate scheme. To ensure that the system reaches equilibrium between pulses, a long relaxation delay of 5 s was used for the measurements.

UV Melting Measurement. For UV melting measurements, the RNA samples were diluted with 100 mM potassium phosphate buffer and 2 mM $MgCl_2$ (pH 7.5) to a concentration of 2−3 μM in a 400 μL volume. The UV absorbance was monitored with a JASCO V630 UV−Vis spectrophotometer with a bandwidth of 1.5 nm using a quartz cuvette with a path length of 1.0 cm. Absorbance changes at 260, 280, and 295 nm were collected over the temperature range of 20−100 °C, and RNA samples were overlaid with 20 μL of paraffin oil to prevent evaporation artifacts during measurements. The ramping rate for the temperature was 1 °C/min, and data points were collected every 1 °C when the target temperature reached and was kept within 0.1 °C for 5 s.

Far-UV CD Spectroscopy. The RNA samples used for NMR were reused for far-UV CD spectroscopy after their NMR spectra had been recorded. The samples were diluted with the same buffer to approximately 5−15 μM in a 300 μL volume for CD measurements using a quartz cuvette with a path length of 0.1 cm. After the CD measurements, the nucleotide concentrations were confirmed by measuring the absorbance at 260 nm with an extinction coefficient of 285,062 M^{-1} cm^{-1} determined by OligoCalc.[37] The CD spectra were recorded using an Aviv model 202 spectrometer over 200−300 nm at 25 °C with a bandwidth of 1 nm, a data interval of 0.5 nm, and an averaging time of 1 s. The molar nucleotide ellipticity (degrees square centimeters per decimole) was calculated by using the equation $[\theta] = \theta/(10 \times C \times l \times n)$, where θ is the CD signal in millidegrees, C is the concentration in molar, l is the cell path length in centimeters, and n is the number of nucleotides. For thermal denaturation, the CD signal at 260 nm was collected with an averaging time of 5 s over a temperature range of 5−99 °C with an equilibrium period of 1.1 min between each measurement.

Analysis of CD Thermal Unfolding Data with a Three-State Model. The CD isotherms and UV melting curves of individual hY1 RNA variants were subject to unfolding equilibrium analysis by evoking a three-state model[38,39] defined as

$$\mathrm{N} \overset{K_{NI}}{\longleftrightarrow} \mathrm{I} \overset{K_{IU}}{\longleftrightarrow} \mathrm{U} \tag{1}$$

where N, I, and U correspond to the native, intermediate, and unfolded states, respectively. The two equilibrium constants of unfolding, K_{NI} and K_{IU}, can then be given as

$$K_{NI} = \frac{[\mathrm{I}]}{[\mathrm{N}]} \text{ and } K_{IU} = \frac{[\mathrm{U}]}{[\mathrm{I}]} \tag{2}$$

where [N], [I], and [U] are the concentrations of the N, I, and U states, respectively.

The observed CD or UV signal (S_{obs}) obtained from the three states is given as

$$S_{obs} = S_N f_N + S_I f_I + S_U f_U \tag{3}$$

where f_N, f_I, and f_U are the fractional populations of the N, I, and U states, respectively, and S_N, S_I, and S_U are the signals of the N, I, and U states, respectively. Under equilibrium conditions, the fractional populations can be presented as

$$f_N = \frac{1}{1 + K_{NI} + K_{NI}K_{IU}} \tag{4}$$

$$f_I = \frac{K_{NI}}{1 + K_{NI} + K_{NI}K_{IU}} \tag{5}$$

$$f_U = \frac{K_{NI}K_{IU}}{1 + K_{NI} + K_{NI}K_{IU}} \tag{6}$$

The equilibrium constants can be transformed into a free energy difference between states i and j as

$$\Delta G_{ij} = -RT \ln K_{ij} \tag{7}$$

Therefore, eq 3 can be expressed as a function of free energies between different states and sample temperature, T:

$$S_{obs} = \frac{S_N + S_I e^{-\Delta G_{NI}/RT} + S_U e^{-(\Delta G_{NI} + \Delta G_{IU})/RT}}{1 + e^{-\Delta G_{NI}/RT} + e^{-(\Delta G_{NI} + \Delta G_{IU})/RT}} \tag{8}$$

The temperature dependence of Gibbs free energy ΔG_{ij} is defined as

dx.doi.org/10.1021/bi500470b | Biochemistry 2014, 53, 5848−5863

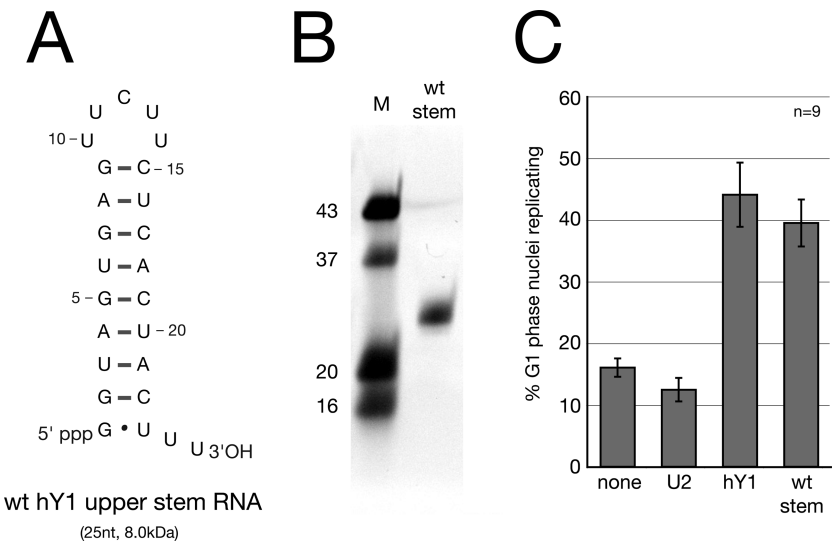


Figure 2. Synthesis and functional analysis of the upper stem domain of hY1 RNA. (A) Nucleotide sequences and predicted folding of the upper stem domain of wild-type (wt) hY1 stem RNA. RNA folding was calculated using the Mfold version 3.2 RNA folding algorithm under default conditions.[25,26] The wt stem RNA contains the predicted 9 bp double-stranded upper stem domain of hY1 RNA (see Figure 1), linked by a polypyrimidine loop and extended by a 3′ polyuridine tail. This RNA corresponds to ΔLS/ΔLP RNA described previously.[31] (B) RNA synthesis and purification. The wt stem RNA was synthesized by *in vitro* transcription, purified by anion exchange chromatography, separated on a denaturing 12% polyacrylamide gel, and stained with SYBR Gold. A series of single-stranded DNA oligonucleotides were used as a running marker (M), and their lengths in nucleotides are indicated. (C) The wt hY1 stem RNA has initiation activity for chromosomal DNA replication *in vitro*. Template nuclei from late G1 phase human cells were incubated with a cytosolic extract depleted of endogenous Y RNAs (i.e., protein fractions QA and ArFT[10]), supplemented with the indicated purified exogenous RNAs (100 nM). U2 snRNA was used as a negative control and full-length hY1 RNA as positive control.[10] Proportions of replicating nuclei were determined by immunofluorescence microscopy. Mean values ± the standard error of the mean (SEM) are shown for nine independent experiments (n = 9).

$$\Delta G_{ij} = \frac{T_{m,ij} - T}{T_{m,ij}} \Delta H(T_{m,ij}) + \Delta C_{p,ij}(T - T_{m,ij}) + T \times \Delta C_{p,ij} \times \ln\left(\frac{T_{m,ij}}{T}\right) \quad (9)$$

where T_m is the melting temperature and ΔC_p is the change in heat capacity. We made the assumption that $\Delta C_p = 0$ for the system. Hence, eq 9 is reduced to

$$\Delta G_{ij} = \frac{T_{m,ij} - T}{T_{m,ij}} \Delta H(T_{m,ij}) \quad (10)$$

The nonlinear regression of the data fitting was conducted using GraphPad Prism.

RESULTS

Functional Activity of the Upper Stem of Wild-Type hY1 RNA. For comparative structure−function analysis, we used a domain of hY1 RNA that is both functionally active and small enough to allow NMR and far-UV CD analyses (Figure 2). A synthetic RNA consisting of the upper stem domain of wild-type hY1 RNA, a five-nucleotide polypyrimidine linker loop, and a 3′ poly(U) tail (Figure 2A) is sufficient to provide the initiation function of wt hY1 RNA.[13] We synthesized this short wt hY1 stem RNA by *in vitro* transcription and purified it to homogeneity by anion exchange chromatography (Figure 2B). We confirmed the ability of this RNA to initiate chromosomal DNA replication in a human cell free system (Figure 2C). In this system, template nuclei from late G1 phase human cells initiate semiconservative chromosomal DNA replication upon being incubated in a cytosolic extract from proliferating human cells, which contains all essential soluble DNA replication proteins and endogenous Y RNAs.[10,15] This system can be modified to allow the functional testing of exogenous RNAs.[10,13] In this modification, the endogenous Y RNAs are first depleted from the cytosolic extract by biochemical fractionation and the exogenous RNAs to be tested are then added to the depleted extract. Using this experimental setup, we found that a background level of approximately 15% of template nuclei replicated in the absence of any exogenously added RNAs (Figure 2C). Addition of human U2 snRNA as a negative control did not increase the proportion of replicating nuclei, whereas addition of either full-length hY1 RNA or the shorter hY1 stem RNA increased their proportion to approximately ≥40% (Figure 2C), all in agreement with earlier reports.[10,13] Next, we examined the structure of this small active RNA in aqueous solution by NMR and far-UV CD spectroscopy.

The Upper Stem Domain of hY1 RNA Forms a Stable Hairpin Structure in Solution with Eight Watson−Crick Hydrogen Bond Base Pairings. To obtain detailed structural information about the wt hY1 stem RNA by high-resolution NMR spectroscopy, we synthesized a uniformly ^{13}C- and ^{15}N-labeled version of this RNA and purified it to homogeneity. Control experiments confirmed that it initiates DNA replication *in vitro* as effectively as unlabeled hY1 stem RNA (data not shown). To directly identify the hydrogen bond pairing patterns, we recorded a 2D HNN-COSY NMR spectrum[32] of this ^{13}C- and ^{15}N-labeled RNA (Figure 3A). The sequential NMR assignments of the imino protons of G2−G9 on the 5′ side and of C15−C22 on the 3′ side of the RNA were made unambiguously (Figure 3B), indicating conventional Watson−Crick base pairing of these two complementary sides of the stem. These results are consistent with the prediction made by the Mfold RNA folding algorithm[25] (see Figure 2A). The NMR

dx.doi.org/10.1021/bi500470b | *Biochemistry* 2014, 53, 5848−5863

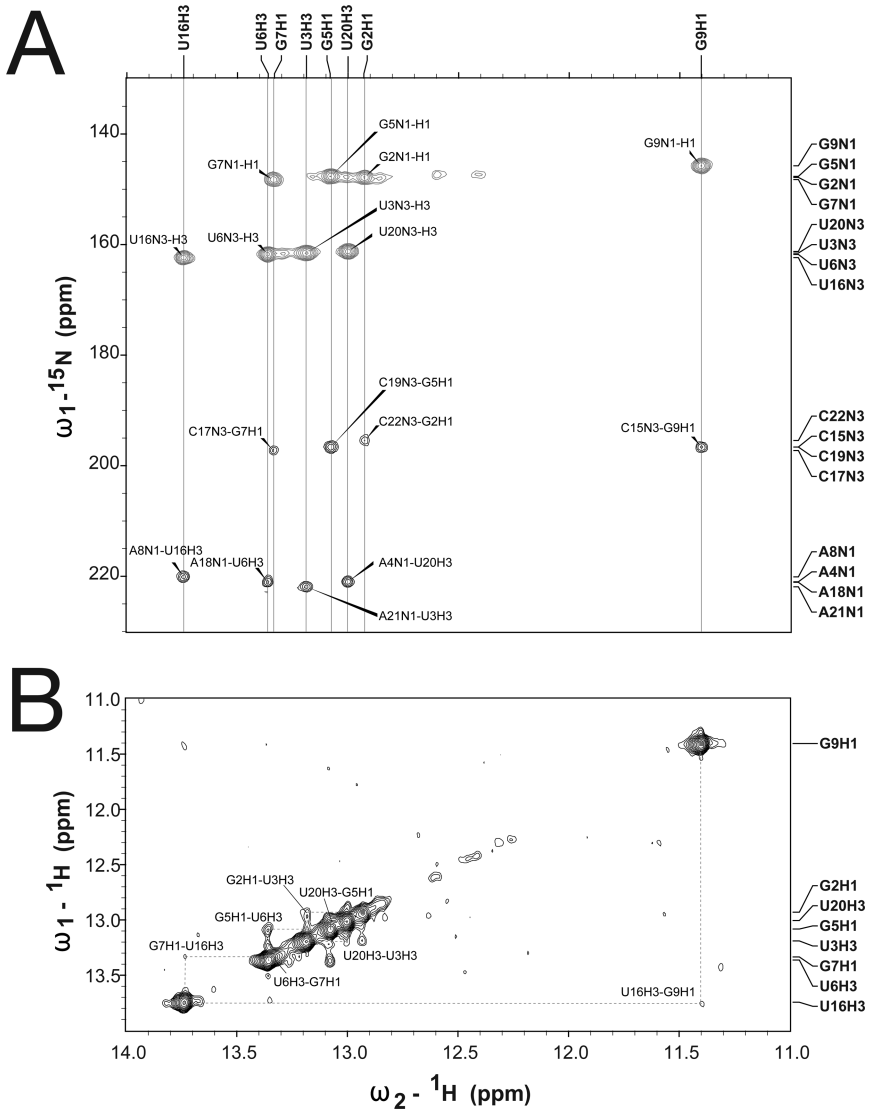


Figure 3. NMR structure analysis of wt hY1 stem RNA. (A) Two-dimensional HNN-COSY spectrum of the ^{13}C- and ^{15}N-labeled wt hY1 stem RNA recorded at 11.7 T (corresponding to a proton Larmor frequency of 500 MHz) and 278 K. Eight imino hydrogen-bonded base pairs between N1 of adenines and H3 of uridines (A:U) or N3 of cytosines and H1 of guanines (C:G) are observed and labeled accordingly. (B) Sequential assignments of the imino protons made using the 2D NOESY spectrum recorded at 11.7 T and 280 K with a mixing time of 200 ms.

assignments of the aromatic protons are in good agreement with the assignments for the eight imino protons (Figure S1 of the Supporting Information). However, NMR assignments of the nucleotides at 3′ and 5′ ends, as well as those located in the short internal linker loop, cannot be made unambiguously because of the lack of corresponding NOE cross-peaks, which could result from the lack of well-defined three-dimensional (3D) structures in these flanking regions.

Individual Nucleotides Provide Helical Stability to the Upper Stem of hY1 RNA. To obtain independent secondary structure information for the upper stem of hY1 RNA, we recorded the far-UV CD spectrum of the wt hY1 stem RNA (Figure 4A). It exhibits a prominent negative signal at 210 nm, which corresponds to a canonical A-form RNA duplex,[40] thus confirming the overall helical structure. We next recorded a series of 1D ^{1}H NMR spectra at different temperatures to investigate the contributions of individual hydrogen bonds to the thermal stability of the upper stem of hY1 RNA (Figure 4B) and determined the temperature coefficients ($\Delta\delta/\Delta T$) of the imino protons that are hydrogen bonded to form the stem (Figure 4C). A low temperature coefficient (its absolute value) indicates that the chemical shift is more inert to temperature changes and hence structurally more stable in response to thermal expansion and/or unfolding. The results show low temperature coefficients for the imino protons of G5, G9, and U16 (Figure 4C), indicating that they might be involved in stable hydrogen bonds in the upper stem of hY1 RNA. In contrast, the imino protons of G2, U3, U6, G7, and U20 exhibit larger temperature coefficients, suggesting that their respective base pairings are considerably unstable at physiological temperature.

We next determined the imino proton transverse relaxation times (T_2) for the hydrogen-bonded imino protons as a function of temperature (Table S1 of the Supporting Information). The imino proton T_2 encompasses two contributions: (1) the molecular tumbling rate that increases with temperature and therefore leads to a longer T_2 value and (2) the solvent exchange rate that also increases with temperature but leads to shorter T_2 values because of the chemical exchange process. These opposing effects, namely an

dx.doi.org/10.1021/bi500470b | Biochemistry 2014, 53, 5848−5863

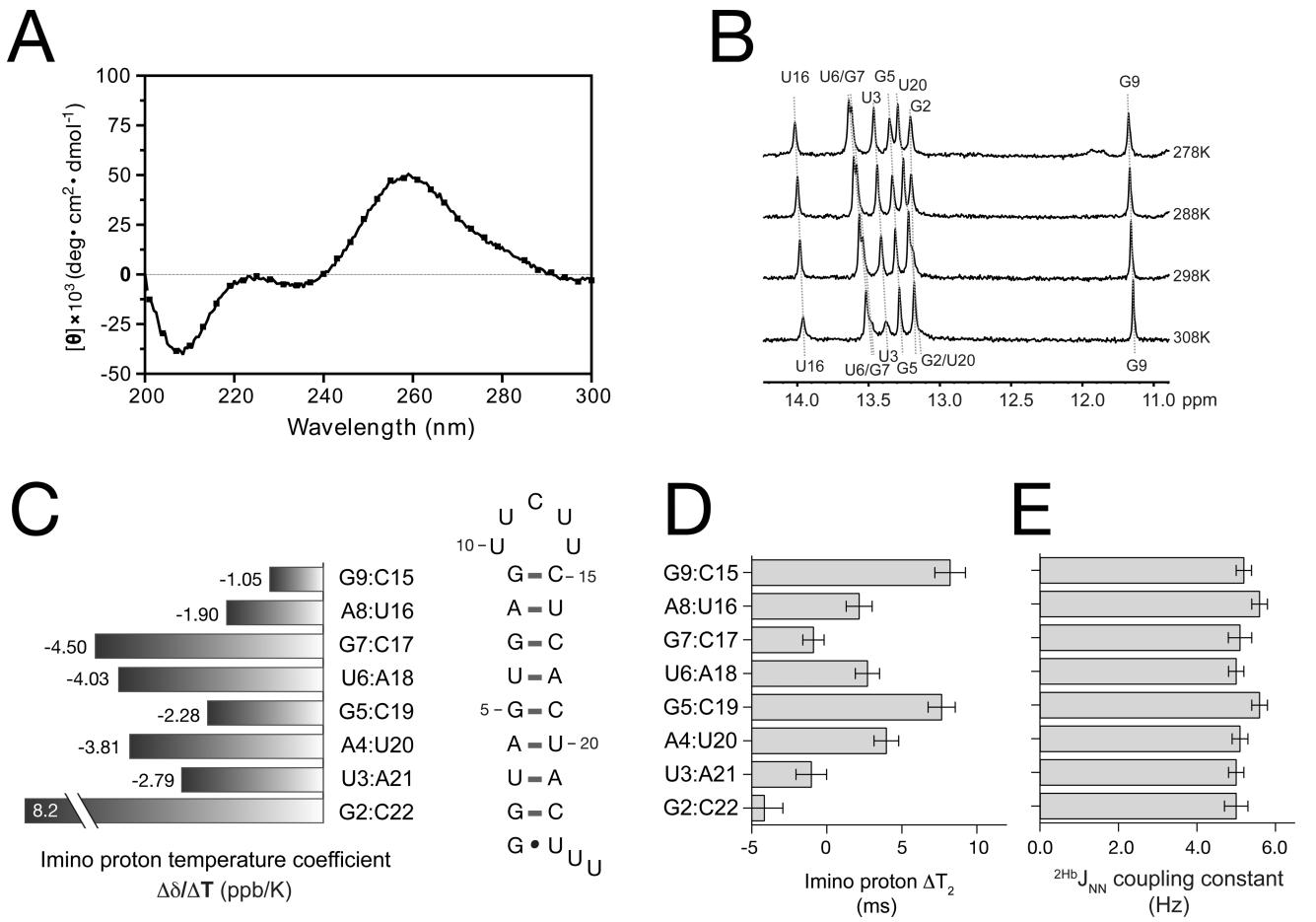


Figure 4. Biophysical analyses of base pair stability parameters of the wt hY1 stem RNA. (A) Far-UV CD spectrum of wt hY1 stem RNA. (B) Temperature series of 1D imino proton NMR spectra of wt hY1 upper stem RNA. Assignments of the imino protons are shown above the peaks in the top spectrum taken at 278 K, and shifts of these peaks with increasing temperatures are indicated by dotted lines across the spectra shown below. (C) Imino proton temperature coefficients of the eight base pairings. Temperature coefficients ($-\Delta\delta/\Delta T$ in parts per billion per kelvin) for each base pair are shown as a bar chart and placed next to their position in the corresponding schematic of wt hY1 stem RNA. (D) Imino proton ΔT_2 values for each base pair of the hY1 upper stem RNA. ΔT_2 values were calculated by subtracting T_2(298 K) from T_2(278 K) and plotted for each of the base pairs. (E) $^{2Hb}J_{NN}$ coupling constants of Watson−Crick base pairs. The coupling constants were derived from the HNN-COSY spectrum of the ^{13}C- and ^{15}N-labeled wt hY1 stem RNA recorded at 11.7 T and 278 K.

increased T_2 caused by a faster molecular tumbling rate and a decreased T_2 caused by an enhanced solvent exchange rate, coexist, and the apparent temperature dependence of the observed T_2 depends on which of these two parameters dominates at a specific temperature. For comparison, we calculated the difference in the imino proton T_2 values at 298 and 278 K [$\Delta T_2 = T_2(298\ \text{K}) - T_2(278\ \text{K})$] for each base pair of the hY1 upper stem RNA (Figure 4D). A high ΔT_2 value indicates a high stability of the corresponding base pair due to marginal solvent exchange and accelerated molecular tumbling rates at higher sample temperatures. The G5:C19 and G9:C15 base pairs are therefore the most stable base pairs in this hY1 stem RNA, whereas the base pairing stability decreases gradually toward the end of the double-stranded domain at the G2:C22 base pair and, interestingly, toward the central G7:C17 base pair (Figure 4D).

To further evaluate the relative strengths of these hydrogen bonds, we determined the cross-hydrogen bond J coupling constants between hydrogen bond donors and acceptors of the imino protons ($^{2Hb}J_{NN}$) using the HNN-COSY experiment[32] conducted at 278 K. The G5:C19 and A8:U16 base pairs exhibit $^{2Hb}J_{NN}$ values (5.6 ± 0.2 Hz) slightly larger by ∼0.5 Hz than those of the central U6:A18 and G7:C17 base pairs and the base pairs toward the end of the helix (Figure 4E and Table S1 of the Supporting Information). Nevertheless, the differences in the $^{2Hb}J_{NN}$ coupling constants are marginal, suggesting that the base pairings within the hY1 upper stem RNA are equally stable at 278 K. The differences in the $^{2Hb}J_{NN}$ coupling constants may become more pronounced when the sample temperature is higher. However, the HNN-COSY experiment is relatively insensitive, and the quantitative measurements at higher temperatures were hindered by severe line broadening of the imino protons (see Figure 5A).

To examine conformational dynamics within the hY1 upper stem RNA, we conducted ^{15}N spin relaxation analysis for the imino nitrogen atoms to investigate changes in local dynamics as a function of temperature (Figure 5). We recorded ^{15}N−^{1}H HSQC-type spectra at different temperatures (Figure 5A) and determined the spin relaxation times (Figure 5B,C). On one hand, the longitudinal relaxation times (T_1) of all the imino nitrogens decreased monotonously with an increase in temperature, which is expected from the increase in tumbling rates at elevated temperatures. On the other hand, the transverse relaxation times (T_2) significantly increased monot-

dx.doi.org/10.1021/bi500470b | Biochemistry 2014, 53, 5848−5863

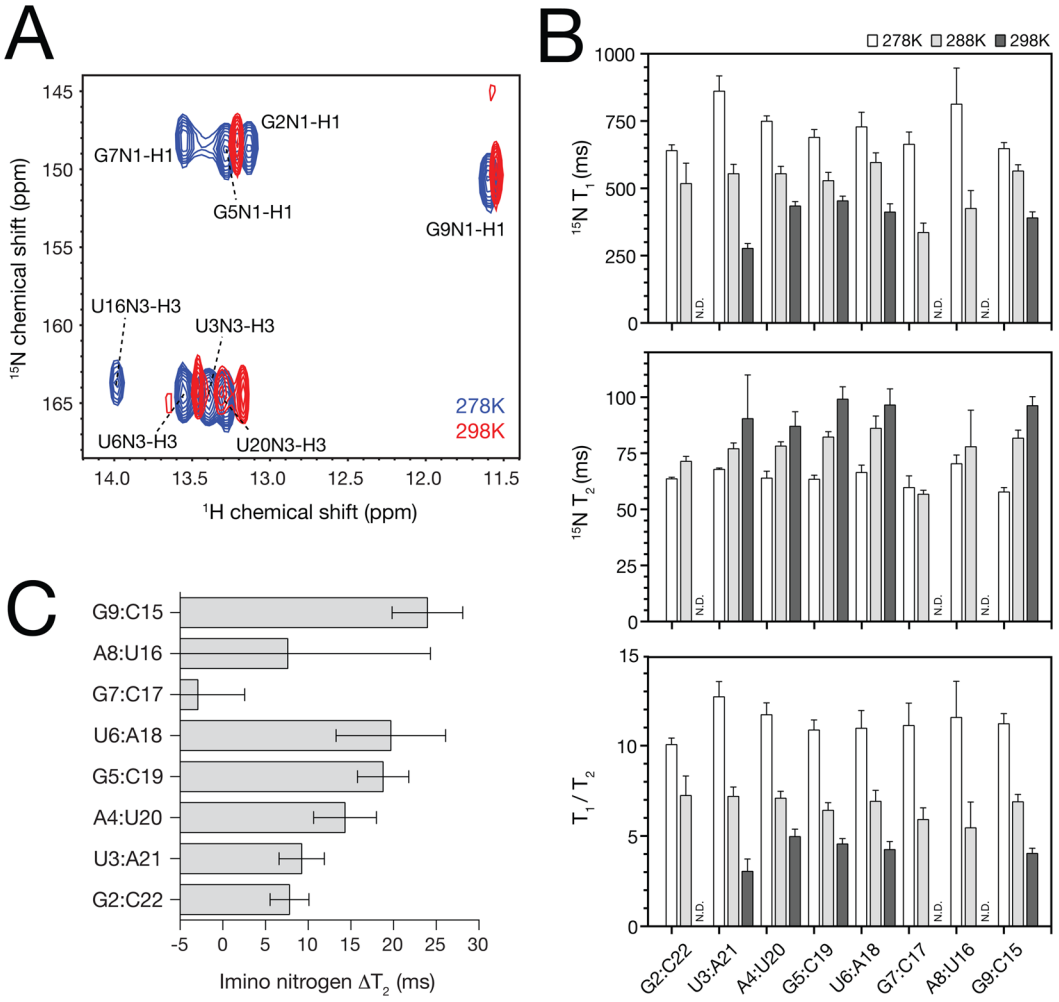


Figure 5. ^{15}N relaxation analysis of wt hY1 stem RNA as a function of temperature. (A) ^{15}N−^{1}H HSQC spectra of wt hY1 stem RNA with a T_2 delay time of 17 ms at 278 K (blue) and 298 K (red). (B) Longitudinal and transverse relaxation time constants, T_1 and T_2, respectively, of the imino nitrogen and their ratio (T_1/T_2) are shown in the top, middle, and bottom panels, respectively. (C) Imino nitrogen ΔT_2 values for each base pair of the hY1 upper stem RNA. ΔT_2 values were calculated by subtracting T_2(288 K) from T_2(278 K) and plotted for each of the base pairs.

onously between 278 and 298 K. This is consistent with the ^{1}H T_2 analysis (Figure S2 of the Supporting Information). In particular, the ^{15}N T_2 is significantly shorter for the G7:C17 base pair at 288 K, suggesting the presence of heterogeneous conformations that lead to additional relaxation processes due to chemical exchange. The chemical exchange contribution to the transverse relaxation is compensated by a much faster molecular tumbling rate at higher temperatures, which is reflected in the overall decrease in the T_1/T_2 ratios (Figure 5B). For G2:C22, G7:C17, and A8:U16 base pairs, their respective imino ^{15}N−^{1}H cross-peaks were broadened beyond detection at 298 K (Figure 5A), suggesting that the solvent exchange process is particularly pronounced for these base pairs. These base pairs also showed the lowest ΔT_2 values (Figure 5C), indicating that they are less stable than the other base pairs in this stem at higher temperatures.

To quantify the hydrogen exchange rates of individual imino protons (k_{HX}), we conducted 1D CLEANEX-PM experiments at 288, 293, and 298 K (Figure 6).[23,24] Below 288 K, the hydrogen exchange process was too slow to be quantified by the CLEANEX-PM experiment (data not shown). While most of the base pairs exhibited appreciable hydrogen exchange rates at 288 K, ranging between 1.4 and 2.5 s^{-1}, the G5:C19 and G9:C15 base pairs did not show appreciable imino proton signals to allow an accurate determination of their respective hydrogen exchange rates (Figure 6A−C). Only at 293 and 298 K could the hydrogen exchange rate of G5:C19 be estimated (1.4 ± 0.2 s^{-1}), which is substantially smaller than those of the other base pairs, ranging between 4 and 7 s^{-1} (Table S2 of the Supporting Information). In the case of the G9:C15 base pair, the hydrogen exchange was too slow to be characterized by CLEANEX-PM even at 298 K. These data therefore support the conclusion that these two base pairs are significantly more stable at higher temperatures than the other base pairs in the helix. (As an aside, the overall hydrogen exchange rates at 293 K are surprisingly higher than those at 298 K, and the reason for that remains to be established.)

Collectively, our NMR results suggest that the stronger and more stable hydrogen bonds of the G5:C19 and G9:C15 base pairs may serve as clamps to maintain the stability of the upper stem domain of hY1 RNA, whereas the base pairs present between these stable base pairs at U6:A18, G7:C17, and A8:U16 (especially G7:C17) and toward the fraying end of the double-stranded RNA helix are intrinsically unstable.

Functional Analysis of Mutant hY1 Stem RNAs. A mutation of the essential and conserved GUG-CAC sequence

dx.doi.org/10.1021/bi500470b | *Biochemistry* 2014, 53, 5848−5863

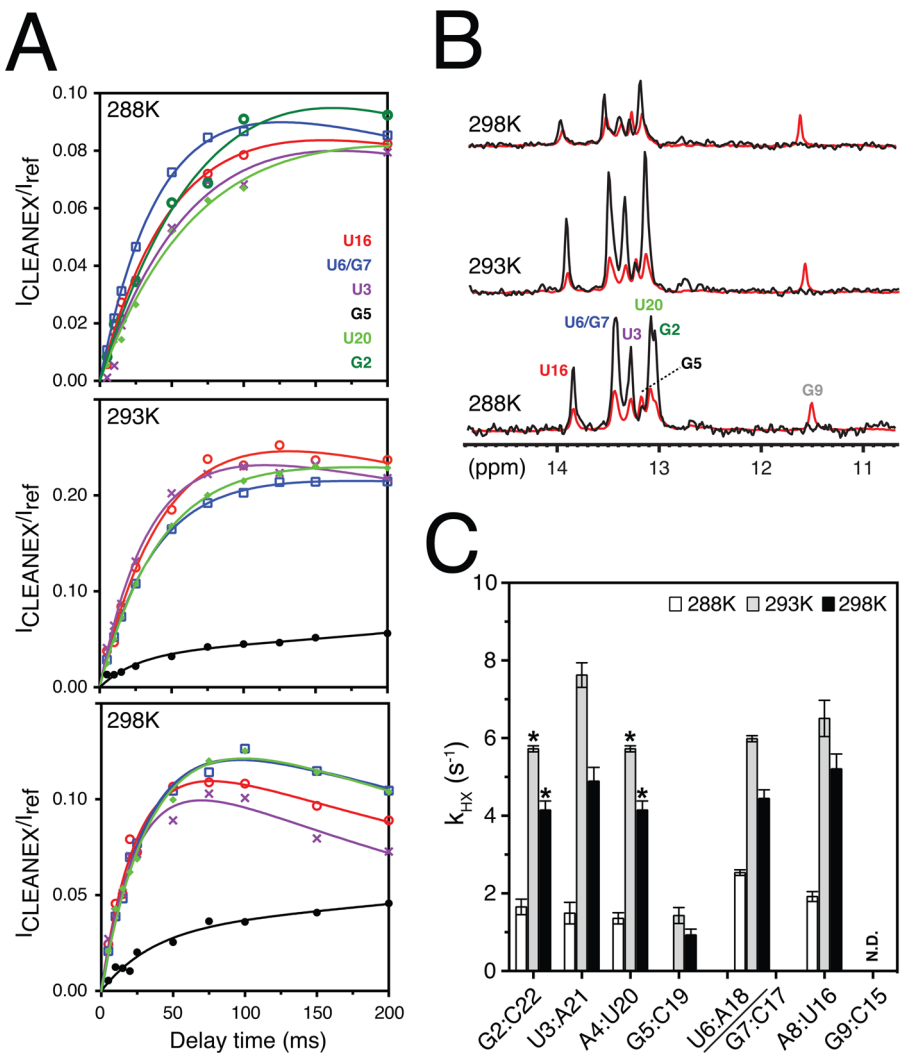


Figure 6. Imino proton hydrogen exchange rates (k_{HX}) of hY1 stem RNA determined by CLEANEX-PM. (A) Intensity buildup curve as a function of spin-lock delay time. The CLEANEX-PM data recorded at 288, 293, and 298 K are shown in the top, middle, and bottom panels, respectively. (B) Comparison of the 1D CLEANEX-PM spectra with a spin-lock delay of 200 ms (black) with a reference spectrum (red). It shows that the hydrogen exchange rate of the G9:C15 base pair is too slow to show a CLEANEX-PM signal even at 298 K. (C) Bar charts of imino proton k_{HX} values at 288, 293, and 298 K. The U6/G7 imino proton resonances were overlapped in the 1D CLEANEX-PM spectra at 288, 293, and 298 K, so that their k_{HX} values were presented as a single group. The k_{HX} values of G2 and U20 at 293 and 298 K are indicated by asterisks as their imino proton resonances are overlapped.

element within the upper stem domain into ACA-UGU abolishes the initiation function of wt full-length hY1 RNA.[13] This 3 bp motif comprises the highly stable G5:C19 and the two unstable U6:A18 and G7:C17 base pairs in the hY1 stem RNA (see Figures 4–6). To see if a loss of function correlates with a loss of structural integrity of this RNA, we first designed, synthesized, and purified a short mutant hY1 stem RNA (mt1) that carries this 3 bp substitution (Figure 7A,B). In functional DNA replication assays, mutant mt1 RNA did not increase the proportion of replicating nuclei above background levels (Figure 7C). We therefore conclude that the 3 bp substitution mutation inactivates the initiation function of the small upper stem RNA for DNA replication *in vitro*, as it does in the full-length background of hY1 RNA.[13]

For a more detailed analysis, we next designed, synthesized, and purified a series of four individual base pair substitution mutants that span the essential GUG-CAC sequence motif (mt2–mt4) and the base pair at the loop end of the upper stem (mt5) at nucleotide position G9:C15 that shows strong helical stability (Figure 7A,B). We also made a double mutant (mt6) combining the base pair substitutions of the stable G:C base pairs at nucleotide positions G5:C19 and G9:C15 (Figure 7A,B). In functional assays, the single-substitution mutants mt2–mt5 each increased the proportions of replicating nuclei to intermediate levels between the background level observed without RNA and the high level seen in the presence of wt hY1 stem RNA (Figure 7C). In contrast, the double-substitution mutant mt6 did not increase the proportion of nuclei replicating above background levels, like the triple mutant mt1.

Taken together, these functional assays indicate that single substitutions of those base pairs that show high helical stability in the upper stem domain of wt hY1 RNA, or of those intrinsically unstable central base pairs, compromise the ability of the small stem hY1 to initiate DNA replication, but only to a limited extent. Importantly, these individual contributions to function are additive because substitutions of two or three of these base pairs together abrogate the RNA function completely. We therefore investigated next how these

dx.doi.org/10.1021/bi500470b | *Biochemistry* 2014, 53, 5848–5863

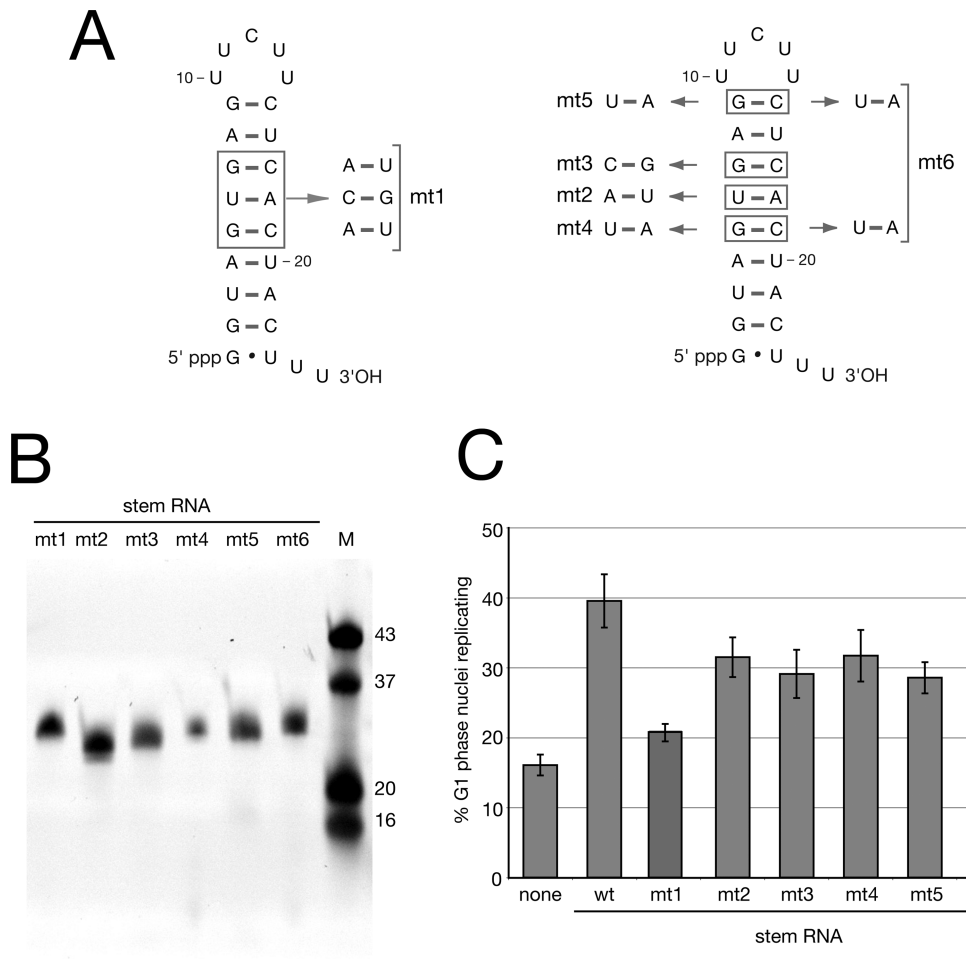


Figure 7. Functional analysis of mutant hY1 stem RNAs. (A) Nucleotide sequence and predicted folding of base substitution mutant stem RNAs. The RNA sequences to be mutated are boxed on the wild-type reference sequences, and substituted base pairs are indicated by arrows for the GUG-box mutant (mt1) on the left and for the four single-base pair substitutions (mt2–mt5) and the double mutant (mt6) on the right. (B) RNA synthesis and purification. RNAs were synthesized, purified, and visualized as detailed for Figure 2B. (C) Functional analysis of mutant hY1 stem RNAs. The ability of the indicated RNAs to initiate chromosomal DNA replication *in vitro* was assayed as detailed for Figure 2C. Mean values ± the standard error of the mean (SEM) of six to nine independent experiments (n = 6–9) are shown.

mutations change the structure of the stem RNAs and if altered structures correlate with impaired functionality.

The GUG-CAC Motif in the Upper Stem of hY1 RNA Is Essential for Maintaining Structural Stability. For the structural analysis of the mutant stem RNAs, we conducted 1D ^{1}H NMR and far-UV CD spectroscopy (Figure 8). In contrast to the wt hY1 stem RNA, the imino proton NMR spectrum of mt1 is less resolved and more heterogeneous and the peak intensities are very low, compared with the signals from the aromatic protons (Figure 8A). The aromatic proton signals can serve as a reference because they do not exchange with bulk solvent, so that the peak intensities are proportional to the sample concentration. This result indicates that mt1 RNA has a stem–loop structure much less stable than that of wt hY1 RNA. The far-UV CD spectrum of mt1 lacks the negative signal at 210 nm that corresponds to the canonical A-form helix but instead exhibits a positive signal at 218 nm (Figure 8B, top panel), indicating that its structure is significantly different from that of wt hY1 stem RNA. For RNA, the CD signal at 265 nm is sensitive to base stacking, and its decrease in intensity generally occurs because of thermal melting and/or solvent perturbation.[41−43] We observed an increased signal intensity at 265 nm for mt1 RNA (Figure 8B), which might result from anomalous base stacking that is structurally heterogeneous. We conclude that the conserved GUG-CAC motif in the upper stem of hY1 RNA not only is essential for its function during DNA replication but also plays an essential role in maintaining the stable A-form RNA helical structure of this domain under physiological conditions in solution.

The imino proton NMR spectra of mt2 and mt5 are more similar to that of wt, whereas those of mt3, mt4, and mt6 are less resolved and more heterogeneous (Figure 8A). The far-UV CD spectra reveal that wt and mt2 RNAs possess a similar helical A-form structure, and the negative signal at 210 nm is even stronger in mt2 than in wt (Figure 8B, bottom panel). These data indicate that the U6:A18 to A6:U18 mutation of mt2, a swap in base pairing, has no major effect on the overall structure of the upper stem RNA. In contrast, an increasing level of conformational divergence from the A-form helical structure, as deduced from the CD signals at 210 nm, can be observed for the mutants targeting the conserved G residues at position 5, 7, or 9, namely, mt4, mt3, mt5, and mt6 (Figure 8B). However, these single and double mutants are not as disruptive to the overall structure as triple mutant mt1. These

dx.doi.org/10.1021/bi500470b | *Biochemistry* 2014, 53, 5848–5863

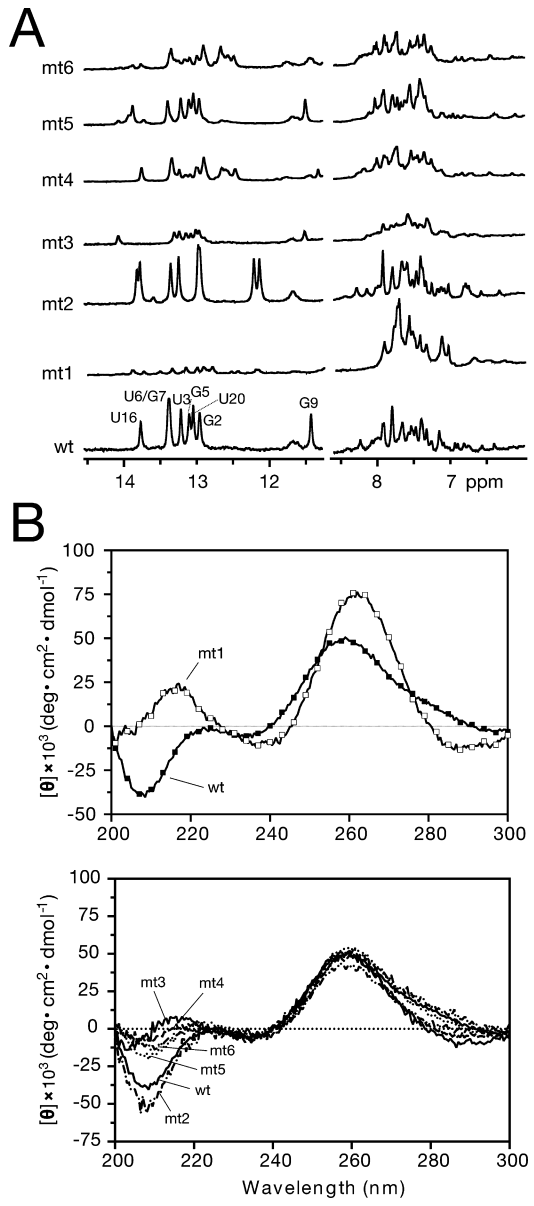


Figure 8. NMR and far-UV CD spectroscopy of mutant hY1 stem RNAs. (A) Comparison of the 1D NMR spectra of imino (at 11.2−14.5 ppm) and aromatic (at 6−8.5 ppm) protons of wt and mutant hY1 stem RNAs at 278 K. Assignments of imino protons for wt RNA are shown above the peaks in the bottom spectrum. The imino proton spectra of wt and mt5 are very similar, while those mt3, mt4, mt6, and mt1 are less resolved and more heterogeneous, indicating that these mutants are more conformationally heterogeneous than wt. (B) Far-UV CD spectra. A comparison of wt with mt1 is shown in the top panel and with mt2−mt6 in the bottom panel. The CD signals at 210 nm for wt and mt2 are similar and suggest A-form helical structure, while mt5, mt6, mt4, mt3, and mt1 are increasingly less similar to wt, suggesting increasingly noncanonical A-form helical structures.

CD results are consistent with the more heterogeneous imino proton resonances in these mutants (Figure 8A).

We looked in more detail at the structure of mt5, which contains the single-base pair substitution outside of the conserved GUG-CAC motif at the top end of the stem domain (Figure S2 of the Supporting Information). As the imino proton spectrum of mt5 is similar to that of the wt, we completed the assignments of imino protons for comparative analysis. The chemical shift of the imino proton of U9 in mt5 is upfield-shifted to ∼11.4 ppm, similar to that of G9 in wt (Figure S2A of the Supporting Information). Additionally, the imino proton T_2 values of mt5 are also similar to those of the wt, particularly so for the G5 and U9 nucleotides (Figure S2B,C and Table S1 of the Supporting Information). These observations indicate that the G:C-to-U:A replacement at position 9 alone does not significantly perturb the chemical environment or molecular dynamics of the helix around the loop; i.e., mt5 and wt hY1 stem RNA are structurally similar at physiological temperature, despite the difference in base composition. These minor changes to the structure of the upper stem in mt5 correlate with only a marginal reduction in the DNA replication function of this mutant (Figure 7C). Importantly, when this substitution of G9:C15 to U9:A15 is combined with the substitution of G5:C19 to U5:A19 in double mutant mt6, stronger structural changes are seen (Figure 8), and the DNA replication initiation activity becomes inhibited (Figure 7C).

In summary, our NMR and CD data show that individual nucleotides in the conserved GUG-CAC motif and the terminal G:C base pair contribute additively to the maintenance of the structural integrity of hY1 upper stem RNA. The contribution of individual base pairs to structural integrity correlates with the contribution of these base pairs to the DNA replication function of this domain.

The Thermal Stability of hY1 Stem RNA Mutants Directly Correlates with Biological Activity. In the final set of experiments, we employed thermal melting experiments to extract additional thermodynamic parameters from the wt and mutant hY1 stem RNAs. We monitored the changes in the UV absorption at 260 nm (Figure S3 of the Supporting Information) and in the molar ellipticity measured by CD spectroscopy and plotted these parameters as a function of temperature to allow comparison of relative thermal stabilities of the hY1 stem RNAs (Figure 9A). The melting curves of all these RNAs exhibit more than one transition for both UV absorption and CD isotherms. We therefore applied a three-state thermal unfolding model to determine their associated thermodynamic parameters (see Materials and Methods). In the three-state model, three populations that correspond to the native (N), intermediate (I), and unfolded (U) states were deconvoluted as a function of temperature (Figure 9A, insets). The higher I−U transition temperatures are highly correlated between the two different spectroscopic techniques for each RNA, while the lower N−I transition temperatures are less correlated (Figures S3 and S4 and Table S3 of the Supporting Information). As the amplitude of the first N−I transition of UV absorbance is very small and thus difficult to evaluate for confident data fitting results, far-UV CD appears to be more sensitive to conformational changes in the hY1 RNA variants.

To correlate the thermal stabilities of the hY1 RNA variants with their respective DNA replication activities, we calculated their native populations (N states) at 37 °C based on their far-UV CD isotherms (Figure 9B), to match the temperature at which all the functional DNA replication experiments were conducted. This analysis shows that the N state populations of wt, mt2, mt4, and mt5 are similarly high at 37 °C (93 ± 2, 93 ± 2, 96 ± 1, and 90 ± 4%, respectively), that of mt3 is intermediate (78 ± 10%), and those of mt6 and mt1 are significantly lower (66 ± 7 and 63 ± 6%, respectively) (Figure 9B). We then plotted these N state proportions against the percentages of nuclei replicating in the presence of these RNAs (Figure 10) and obtained a correlation (R^2) of 0.76 between

dx.doi.org/10.1021/bi500470b | Biochemistry 2014, 53, 5848−5863

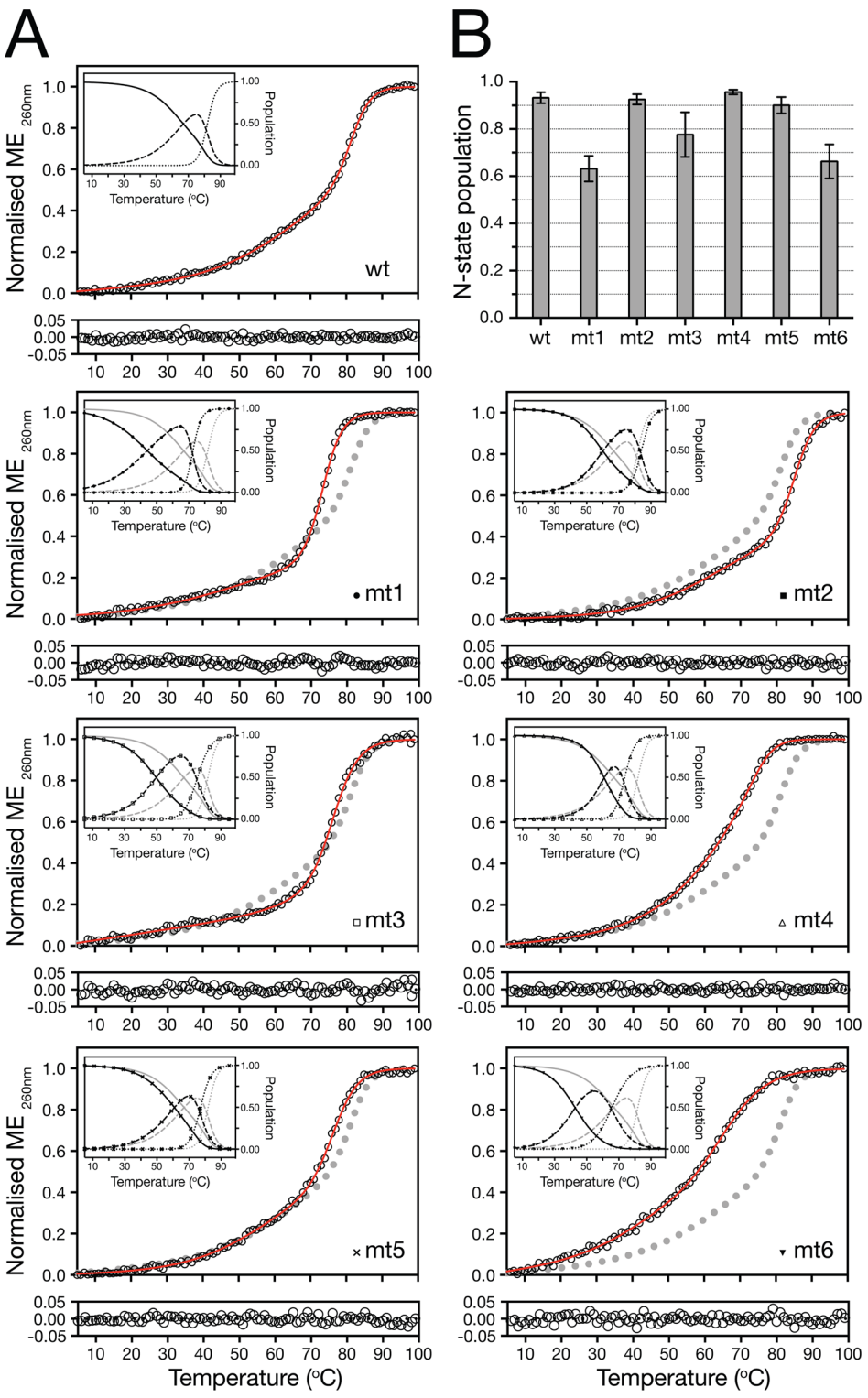


Figure 9. Thermal denaturation of wt and mutant hY1 stem RNAs by CD spectroscopy. (A) Representative far-UV CD melting isotherms of the hY1 stem RNA variants. The CD signals at 260 nm were normalized prior to nonlinear regression using a three-state folding model shown as red lines. The residuals of data fitting are shown below the raw data. Insets show the relative populations of the N (native), I (intermediate), and U (unfolded) states as solid, dashed, and dotted lines, respectively, as a function of temperature. The fitting results of the wild type are shown as gray lines for comparison with those of the variants. (B) Bar chart of the N state populations of individual hY1 RNA variants at 37 °C. The N state fractions of wt, mt2, mt4, and mt5 are >90%, while that of mt1 is only 63 ± 6% with a significant amount of the I state (37 ± 6%). Additionally, mt3 and mt6 also exhibit appreciable amounts of I states (22 ± 10 and 33 ± 7%, respectively). The error bars correspond to the standard error of the mean (SEM) of three independent experiments (n = 3).

the thermal stability and the functionality, indicating that thermal stability positively correlates with DNA replication initiation activity of the upper stem domain of Y RNAs. Although how the native and intermediate states differ in their

dx.doi.org/10.1021/bi500470b | *Biochemistry* 2014, 53, 5848−5863

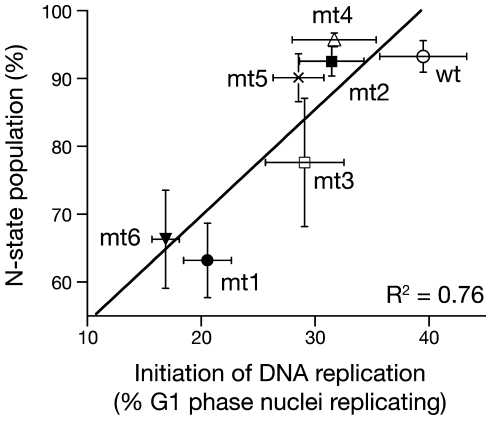


Figure 10. DNA replication initiation activity correlates with the N state population of hY1 stem RNAs. Correlation plot between DNA replication initiation efficiencies and the N state populations of wt and mutant hY1 stem RNAs. Data were taken from Figures 7C and 9B, respectively.

tertiary structures based on our NMR and far-UV CD data remains unclear, it is clear that the reductions in imino proton NMR intensities in mt1, mt3, and mt6 reflect a higher degree of structural heterogeneity under native conditions, which are in agreement with their lower native state populations under the same condition (Figures 8 and 9). It should be noted that the three-state folding analysis shows that the first N−I transition takes place when the CD_{260} signal loss is <20% for all variants (Table S3 of the Supporting Information). In contrast to conventional melting temperature estimation, i.e., temperature at which 50% CD signal loss observed, our analysis shows that even subtle structural changes in the tertiary structures of the A-form RNA helical stem, which cause only marginal CD signal changes, can lead to significant functional loss.

Collectively, the results show that the conserved GUG-CAC motif and the two stable G5:C19 and G9:C15 base pairs are critical for the structure and function of the upper stem domain of hY1 RNA. We found that mutations in the nucleotide sequence in this critical region that lead to structural destabilization also result in significant loss of its biological activity to initiate DNA replication.

DISCUSSION

In this study, we have provided a nucleotide-specific structure−function analysis of the conserved upper stem domain of human Y1 RNA in solution under physiological conditions. The key findings are summarized schematically in Figure 11. On the left-hand side of this graphic, we have indicated the stability of individual base pairs along the axis of an A-form RNA helix as spheres. We have used sphere size and color to represent the two biophysical parameters indicating base pair stability (i.e., imino proton temperature coefficients and ΔT_2, respectively). On the right-hand side, we have represented the biological activity of mutants in key base pairs of this RNA by color-coded boxes. The four partially active single-base pair substitution mutants are shown next to their position in the helix model and the two inactive double or triple mutants further to the right. Our NMR and far-UV CD data confirm that the conserved Y RNA upper stem domain adopts an A-form helical structure involving eight Watson−Crick base pairings, which is consistent with computational predictions. Within this helical domain, the G5:C19 and G9:C15 base pairs [represented by small dark blue spheres (Figure 11)] are the most stable, which clamp an inherently unstable central section of the helix at positions U6:A18, G7:C17, and A8:U16 [represented by large pink and light blue spheres (Figure 11)]. To link these structural data with a function of this Y RNA domain in the replication of chromosomal DNA, we performed systematic substitution mutagenesis of key bases and functional testing of the mutant RNAs. Our data showed that mutating individual base pairs had only mild to moderate effects, while mutating both of the two most stable base pairs or mutating the central GUG motif resulted in a disruption of the ordered A-form helical structure and a loss of DNA replication initiation activity of the mutant RNAs [represented by colored square boxes (Figure 11)].

These RNA variants were subjected to spectroscopic analysis to characterize their structures and folding stabilities. The imino proton NMR and far-UV CD spectra of the hY1 stem RNAs demonstrated that a small variation in the sequence composition that is considered identical for most free energy secondary structure predictors (e.g., swapping a G:C base pair to a C:G base pair as in mt3) can in fact lead to profound structural perturbation and destabilization. Quantitative analysis of the thermal unfolding of these RNAs by far-UV CD spectroscopy allowed us to estimate the populations of natively folded species of the hY1 stem RNAs, which presumably correspond to the functional states. On the basis of the population distribution profiles, a positive correlation was observed at 37 °C between the populations of the native states and their ability to initiate DNA replication in functional assays *in vitro*.

The structure and function of full-length noncoding Y RNAs have been conserved during vertebrate evolution. Earlier chemical and enzymatic probing experiments have shown that Y RNAs form stem−loop structures in different vertebrate organisms.[17,18,20] Nucleotide sequences on both strands of the upper stem are highly conserved in vertebrate Y RNAs,[1,2] and in particular, the GUG-CAC motif is an island of near-complete nucleotide conservation within this domain (Figure 1). Nucleotide conservation at a given position of a small noncoding RNA strongly suggests a conservation of function for this domain because selective pressure during evolution acts on function. It has been demonstrated experimentally that the upper stem of Y RNAs is essential for the initiation of chromosomal DNA replication in different vertebrate organisms.[13,16]

Within the upper stem domain, we have identified by NMR spectroscopy two G:C base pairs that confer stability to the A-form helix of this domain [at positions G5:C19 and G9:C15 of the small stem RNA, corresponding to positions G17:C91 and G21:C87, respectively, of the full-length hY1 RNA (Figure 1)]. The sequence and position of these base pairs are evolutionarily conserved in vertebrate Y RNAs (Figure 1). However, there is minor sequence variation between different Y RNAs, suggesting that some redundancy is built into this system. For instance, vertebrate Y1 RNAs have the paired AG dinucleotide at positions 8 and 9 like hY1; however, Y3 RNAs lack the A so that the adjacent paired G is moved down to position 8, whereas Y4 and Y5 RNAs instead tend to have paired GG dinucleotide pairs at positions 8 and 9.[1,2] Consistent with this naturally occurring local sequence variation, we found that a mutation at this position 9 (mt5) had only mild effects on the structure and function of the small hY1 stem RNA. Importantly, the highly stable G:C base pair at the lower

dx.doi.org/10.1021/bi500470b | *Biochemistry* 2014, 53, 5848−5863

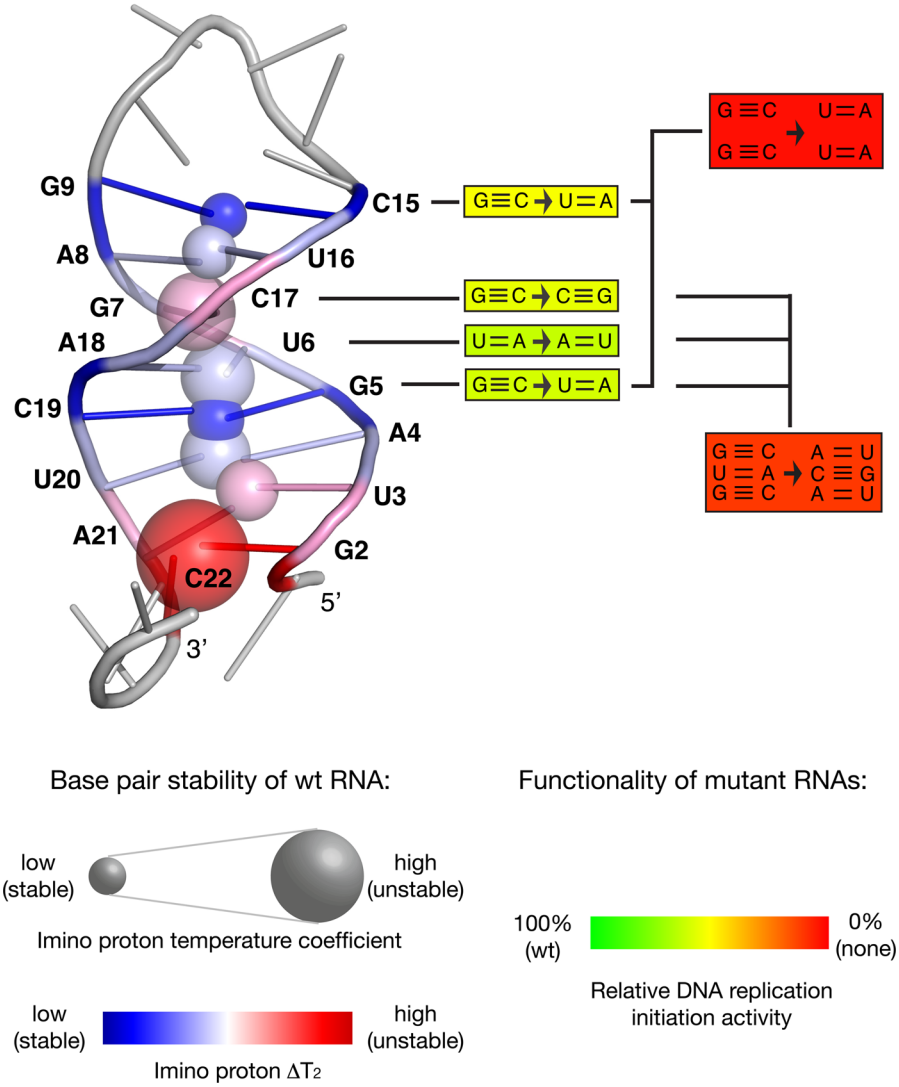


Figure 11. Structure–function determinants for the hY1 stem RNA. On the left-hand side, we plotted a structure model of the hY1 upper stem RNA domain as an A-form RNA helix with eight canonical Watson–Crick base pairs (using PyMOL in cartoon nucleic acid mode). The imino proton of each base pair is symbolized by a colored sphere in the center of the helix. The size of each sphere is proportional to the relative instability of the respective imino proton, as deduced from their temperature coefficients (data taken from Figure 4C). The color of each sphere (and the associated base pair) represents the hydrogen bond stability, from blue (most stable) to red (least stable). The represented values are the imino proton T_2 differences between 278 and 298 K (data taken from Figure 4D and Table S1 of the Supporting Information). On the right-hand side, we summarized the functionalities of mutant hY1 upper stem RNAs containing the indicated single- or multiple-base pair substitutions. A color spectrum from green to red represents the relative DNA replication initiation activity of these mutants specified in the indicated boxes. The top value of 100% (green) corresponds to the percentages of late G1 phase nuclei replicating in the presence of the wt hY1 stem RNA, while the bottom value of 0% (red) corresponds to the percentages of late G1 phase nuclei replicating in the absence of Y RNAs (data taken from Figure 7C).

G5:C19 position is highly conserved and present in almost all vertebrate Y RNAs (see Figure 1).

Our NMR data also identified three inherently unstable base pairs in the upper stem domain at positions U6:A18, G7:C17, and A8:U16 of the stem RNA [corresponding to positions U18:A90, G19:C89, and A20:U88, respectively, of the full-length hY1 RNA (see Figure 1)], of which the central G7:C17 base pair is the most unstable. These base pairs exhibit severe line broadening of the respective imino protons at elevated temperatures, large temperature coefficients, and enhanced transverse relaxation rates (Figures 4 and 5). These unstable base pairs are positioned between the stable clamp. In vertebrate Y RNAs, an adenosine residue in this motif at position U6:A18 (corresponding to position U18:A90 of the full-length hY1 RNA) also shows a weak reactivity to chemical probing by dimethyl sulfate, which could indicate that it may not be base-paired all the time.[17,18] These data therefore suggest that the overall A-form helix of the upper domain is locally stabilized by the two G:C clamps and that the unstable nucleotides between them are able to unfold locally or may even flip out of the overall stable helix. Importantly, two of these unstable base pairs in the upper stem are very highly conserved in vertebrate Y RNAs (Figure 1).

Taken together, the strong conservation of both stable and unstable base pairs predicts that their contribution to the structural stability of this domain is relevant for the biological function of Y RNAs. In direct experimental support of this prediction, we found that mutating the two stable G:C clamps of the upper stem domain of hY1 RNA together (mutant mt6) leads to both structure disruption and functional impairment of

dx.doi.org/10.1021/bi500470b | *Biochemistry* 2014, 53, 5848–5863

the stem RNA in DNA replication. Furthermore, combinations of mutations of both stable and unstable base pairs in the conserved GUG-CAC trinucleotide motif (mutant mt1) also abolish the initiation function of the upper stem (this study) and of full-length hY1 RNA.[13] It is therefore highly likely that this evolutionarily conserved motif of inherently unstable base pairs clamped either side by relatively stable base pairs is essential for the conserved function of Y RNAs in DNA replication.

The molecular mechanism by which Y RNAs exert their function in the initiation of chromosomal DNA replication in vertebrates is not yet known. The biophysical and functional analysis of the essential upper stem reported here now allows us to make several predictions with regard to a possible molecular mechanism that can be tested in future experiments. Published data support physical interactions between Y RNAs and proteins that are essential for the initiation step of DNA replication in vertebrates, including the origin recognition complex (ORC), associated preinitiation complex proteins Cdc6, Cdt1, and others.[16,31] Furthermore, Y RNAs associate with chromatin in a "catch-and-release" mechanism by binding to unreplicated chromatin before DNA replication is initiated and local displacement from sites where DNA replication has been initiated.[31] Apart from Ro60 and La proteins, Y RNAs interact with many other cellular proteins,[6,31,44,45] and it is likely that Y RNAs exert their function through molecular interactions with specific proteins that are also essential for the initiation of DNA replication. These proteins have not been identified unanimously yet, although ORC, Cdc6, and Cdt1 are emerging as strong candidates.[16,31] It is tempting to speculate, therefore, that the interaction of a key initiation protein with the upper stem of Y RNAs may exploit the complex local helical instability of this domain, so that unstable base pairs may denature or flip out as part of, for example, an allosteric activation (or antirepression) process. Base flipping is an established mechanism by which DNA- and RNA-modifying proteins access and chemically react with individual nucleotides that are normally embedded in double-stranded RNA and DNA helices, including dsRNA adenosine deaminases (ADARs) involved in RNA editing.[46−48] Clearly, those DNA replication initiation proteins that specifically interact with the upper stem domain of vertebrate Y RNAs need to be identified and purified. Future experiments will then allow the structural and functional characterization of molecular interactions between these proteins and the upper stem domain of Y RNAs.

■ ASSOCIATED CONTENT

Supporting Information

Four supplemental data figures (Figures S1−S4) and three supplementary tables (Tables S1−S3). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: sthsu@gate.sinica.edu.tw.
*E-mail: tk218@cam.ac.uk.

Present Addresses

[@]A.R.L.: MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, United Kingdom.
[▽]R.R.: Centre de Recherche de Gif, Institut de Chimie des Substances Naturelles, CNRS, 91198 Gif-sur-Yvette Cedex, France.

Author Contributions

I.W. and M.P.K. contributed equally to this work.

Funding

This work was supported by the National Science Council (100-2113-M-001-031-MY2 and 101-2627-M-001-004), Academia Sinica, Taiwan, and the Association for International Cancer Research (AICR Project Grant 10-0570 to T.K.). The Balasubramanian laboratory is supported by program funding from Cancer Research UK and by a Senior Investigator Award from the Wellcome Trust. S.-T.D.H. was supported by a Career Development Award (CDA-00025/2010-C) from the International Human Frontier Science Program. M.P.K. was supported by a doctoral training grant from the Biotechnology and Biological Sciences Research Council (BBSRC), and A.R.L. was supported by a research studentship from the Medical Research Council (MRC).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The NMR spectra were obtained at the Biomolecular NMR facility, Department of Chemistry, University of Cambridge, and the Core Facility for Protein Structural Analysis, supported by the National Core Facility Program for Biotechnology, Taiwan. We thank Dr. Yuan-Chao Lou of the Biophysics Core Facility, Scientific Instrument Center at Academia Sinica, Taiwan, for providing technical assistance and support with data of the far-UV CD experiments.

■ REFERENCES

(1) Mosig, A., Guofeng, M., Stadler, B. M. R., and Stadler, P. F. (2007) Evolution of the vertebrate Y RNA cluster. *Theory Biosci. 129*, 9−14.
(2) Perreault, J., Perreault, J. P., and Boire, G. (2007) The Ro Associated Y RNAs in Metazoans: Evolution and Diversification. *Mol. Biol. Evol. 24*, 1678−1689.
(3) Sim, S., and Wolin, S. L. (2011) Emerging roles for the Ro 60-kDa autoantigen in noncoding RNA metabolism. *Wiley Interdiscip. Rev.: RNA 2*, 686−699.
(4) Verhagen, A. P., and Pruijn, G. J. (2011) Are the Ro RNP-associated Y RNAs concealing microRNAs? Y RNA-derived miRNAs may be involved in autoimmunity. *BioEssays 33*, 674−682.
(5) Krude, T. (2010) Non-coding RNAs: New players in the field of eukaryotic DNA replication. *Subcell. Biochem. 50*, 105−118.
(6) Hall, A. E., Turnbull, C., and Dalmay, T. (2013) Y RNAs: Recent developments. *Biomol. Concepts 4*, 103−110.
(7) Lerner, M. R., Boyle, J. A., Hardin, J. A., and Steitz, J. A. (1981) Two novel classes of small ribonucleoproteins detected by antibodies associated with lupus erythematosus. *Science 211*, 400−402.
(8) Hendrick, J. P., Wolin, S. L., Rinke, J., Lerner, M. R., and Steitz, J. A. (1981) Ro small cytoplasmic ribonucleoproteins are a subclass of La ribonucleoproteins: Further characterization of the Ro and La small ribonucleoproteins from uninfected mammalian cells. *Mol. Cell. Biol. 1*, 1138−1149.
(9) Chen, X., Taylor, D. W., Fowler, C. C., Galan, J. E., Wang, H. W., and Wolin, S. L. (2013) An RNA degradation machine sculpted by Ro autoantigen and noncoding RNA. *Cell 153*, 166−177.
(10) Christov, C. P., Gardiner, T. J., Szüts, D., and Krude, T. (2006) Functional requirement of noncoding Y RNAs for human chromosomal DNA replication. *Mol. Cell. Biol. 26*, 6993−7004.
(11) Christov, C. P., Trivier, E., and Krude, T. (2008) Noncoding human Y RNAs are overexpressed in tumours and required for cell proliferation. *Br. J. Cancer 98*, 981−988.

(12) Krude, T., Christov, C. P., Hyrien, O., and Marheineke, K. (2009) Y RNA functions at the initiation step of mammalian chromosomal DNA replication. *J. Cell Sci. 122*, 2836−2845.
(13) Gardiner, T. J., Christov, C. P., Langley, A. R., and Krude, T. (2009) A conserved motif of vertebrate Y RNAs essential for chromosomal DNA replication. *RNA 15*, 1375−1385.
(14) Langley, A. R., Chambers, H., Christov, C. P., and Krude, T. (2010) Ribonucleoprotein particles containing non-coding Y RNAs, Ro60, La and nucleolin are not required for Y RNA function in DNA replication. *PLoS One 5*, e13673.
(15) Krude, T. (2000) Initiation of human DNA replication in vitro using nuclei from cells arrested at an initiation-competent state. *J. Biol. Chem. 275*, 13699−13707.
(16) Collart, C., Christov, C. P., Smith, J. C., and Krude, T. (2011) The midblastula transition defines the onset of Y RNA-dependent DNA replication in *Xenopus laevis*. *Mol. Cell. Biol. 31*, 3857−3870.
(17) van Gelder, C. W., Thijssen, J. P., Klaassen, E. C., Sturchler, C., Krol, A., van Venrooij, W. J., and Pruijn, G. J. (1994) Common structural features of the Ro RNP associated hY1 and hY5 RNAs. *Nucleic Acids Res. 22*, 2498−2506.
(18) Teunissen, S. W., Kruithof, M. J., Farris, A. D., Harley, J. B., Venrooij, W. J., and Pruijn, G. J. (2000) Conserved features of Y RNAs: A comparison of experimentally derived secondary structures. *Nucleic Acids Res. 28*, 610−619.
(19) Wolin, S. L., and Steitz, J. A. (1984) The Ro small cytoplasmic ribonucleoproteins: Identification of the antigenic protein and its binding site on the Ro RNAs. *Proc. Natl. Acad. Sci. U.S.A. 81*, 1996−2000.
(20) Green, C. D., Long, K. S., Shi, H., and Wolin, S. L. (1998) Binding of the 60-kDa Ro autoantigen to Y RNAs: Evidence for recognition in the major groove of a conserved helix. *RNA 4*, 750−765.
(21) Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, p 320, Wiley, New York.
(22) Furtig, B., Richter, C., Wohnert, J., and Schwalbe, H. (2003) NMR spectroscopy of RNA. *ChemBioChem 4*, 936−962.
(23) Hwang, T. L., van Zijl, P. C., and Mori, S. (1998) Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *J. Biomol. NMR 11*, 221−226.
(24) Hwang, T. L., Mori, S., Shaka, A. J., and van Zijl, P. C. (1997) Application of phase-modulated CLEAN chemical EXchange spectroscopy (CLEANEX-PM) to detect water-protein proton exchange and intermolecular NOEs. *J. Am. Chem. Soc. 119*, 6203−6204.
(25) Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res. 31*, 3406−3415.
(26) Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol. 288*, 911−940.
(27) Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: A sequence logo generator. *Genome Res. 14*, 1188−1190.
(28) Krude, T. (1999) Mimosine arrests proliferating human cells before onset of DNA replication in a dose-dependent manner. *Exp. Cell Res. 247*, 148−159.
(29) Szüts, D., Christov, C., Kitching, L., and Krude, T. (2005) Distinct populations of human PCNA are required for initiation of chromosomal DNA replication and concurrent DNA repair. *Exp. Cell Res. 311*, 240−250.
(30) Szüts, D., Kitching, L., Christov, C., Budd, A., Peak-Chew, S., and Krude, T. (2003) RPA is an initiation factor for human chromosomal DNA replication. *Nucleic Acids Res. 31*, 1725−1734.
(31) Zhang, A. T., Langley, A. R., Christov, C. P., Kheir, E., Shafee, T., Gardiner, T. J., and Krude, T. (2011) Dynamic interaction of Y RNAs with chromatin and initiation proteins during human DNA replication. *J. Cell Sci. 124*, 2058−2069.
(32) Dingley, A. J., and Grzesiek, S. (1998) Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide $^{2}J_{NN}$ couplings. *J. Am. Chem. Soc. 120*, 8293−8297.
(33) Bax, A., and Davis, D. G. (1985) Mlev-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy. *J. Magn. Reson. 65*, 355−360.
(34) Hsu, S. T., Chou, M. T., and Cheng, J. W. (2000) The solution structure of $[d(CGC)r(aaa)d(TTTGCG)]_2$: Hybrid junctions flanked by DNA duplexes. *Nucleic Acids Res. 28*, 1322−1331.
(35) Hsu, S. T., Chou, M. T., Chou, S. H., Huang, W. C., and Cheng, J. W. (2000) Hydration of $[d(CGC)r(aaa)d(TTTGCG)]_2$. *J. Mol. Biol. 295*, 1129−1137.
(36) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277−293.
(37) Kibbe, W. A. (2007) OligoCalc: An online oligonucleotide properties calculator. *Nucleic Acids Res. 35* (Web Server Issue), W43−W46.
(38) Hecky, J., and Muller, K. M. (2005) Structural perturbation and compensation by directed evolution at physiological temperature leads to thermostabilization of β-lactamase. *Biochemistry 44*, 12640−12654.
(39) Sancho, J. (2013) The stability of 2-state, 3-state and more-state proteins from simple spectroscopic techniques ... plus the structure of the equilibrium intermediates at the same time. *Arch. Biochem. Biophys. 531*, 4−13.
(40) Kypr, J., Kejnovska, I., Renciuk, D., and Vorlickova, M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res. 37*, 1713−1725.
(41) Mikulecky, P. J., and Feig, A. L. (2004) Heat capacity changes in RNA folding: Application of perturbation theory to hammerhead ribozyme cold denaturation. *Nucleic Acids Res. 32*, 3967−3976.
(42) Baumann, C., Xirasagar, S., and Gollnick, P. (1997) The trp RNA-binding attenuation protein (TRAP) from *Bacillus subtilis* binds to unstacked Trp leader RNA. *J. Biol. Chem. 272*, 19863−19869.
(43) Loret, E. P., Georgel, P., Johnson, W. C., Jr., and Ho, P. S. (1992) Circular dichroism and molecular modeling yield a structure for the complex of human immunodeficiency virus type 1 trans-activation response RNA and the binding region of Tat, the trans-acting transcriptional activator. *Proc. Natl. Acad. Sci. U.S.A. 89*, 9734−9738.
(44) Fabini, G., Raijmakers, R., Hayer, S., Fouraux, M. A., Pruijn, G. J., and Steiner, G. (2001) The heterogeneous nuclear ribonucleoproteins I and K interact with a subset of the ro ribonucleoprotein-associated Y RNAs in vitro and in vivo. *J. Biol. Chem. 276*, 20711−20718.
(45) Fouraux, M. A., Bouvet, P., Verkaart, S., van Venrooij, W. J., and Pruijn, G. J. (2002) Nucleolin associates with a subset of the human Ro ribonucleoprotein complexes. *J. Mol. Biol. 320*, 475−488.
(46) Roberts, R. J., and Cheng, X. (1998) Base flipping. *Annu. Rev. Biochem. 67*, 181−198.
(47) Cheng, X., and Roberts, R. J. (2010) *Base flipping*, pp 1−8, John Wiley & Sons, Chichester, U.K.
(48) Kuttan, A., and Bass, B. L. (2012) Mechanistic insights into editing-site specificity of ADARs. *Proc. Natl. Acad. Sci. U.S.A. 109*, E3295−E3304.