1770–1782 *Nucleic Acids Research, 2008, Vol. 36, No. 6* *Published online 5 February 2008*
doi:10.1093/nar/gkm1085

Isoenergetic penta- and hexanucleotide microarray probing and chemical mapping provide a secondary structure model for an RNA element orchestrating R2 retrotransposon protein function

Elzbieta Kierzek[1,2], Ryszard Kierzek[2], Walter N. Moss[1], Shawn M. Christensen[3,4], Thomas H. Eickbush[3] and Douglas H. Turner[1,5,*]

[1]Department of Chemistry, University of Rochester, RC Box 270216, Rochester, NY 14627-0216, USA, [2]Institute of Bioorganic Chemistry, Polish Academy of Sciences, 60-714 Poznan, Noskowskiego 12/14, Poland, [3]Department of Biology, University of Rochester, RC Box 270211, Rochester, NY 14627, USA, [4]Department of Biology, University of Texas at Arlington, Arlington, TX 76019 and [5]Department of Pediatrics, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642, USA

Received September 10, 2007; Revised November 16, 2007; Accepted November 19, 2007

ABSTRACT

LNA (locked nucleic acids, i.e. oligonucleotides with a methyl bridge between the 2′ oxygen and 4′ carbon of ribose) and 2,6-diaminopurine were incorporated into 2′-*O*-methyl RNA pentamer and hexamer probes to make a microarray that binds unpaired RNA approximately isoenergetically. That is, binding is roughly independent of target sequence if target is unfolded. The isoenergetic binding and short probe length simplify interpretation of binding to a structured RNA to provide insight into target RNA secondary structure. Microarray binding and chemical mapping were used to probe the secondary structure of a 323 nt segment of the 5′ coding region of the R2 retrotransposon from *Bombyx mori* (R2Bm 5′ RNA). This R2Bm 5′ RNA orchestrates functioning of the R2 protein responsible for cleaving the second strand of DNA during insertion of the R2 sequence into the genome. The experimental results were used as constraints in a free energy minimization algorithm to provide an initial model for the secondary structure of the R2Bm 5′ RNA.

INTRODUCTION

Knowledge of the secondary structure of an RNA is the first step in studying structure–function relationships. Rapid production of reliable secondary structures would allow generation of databases of RNA secondary structures that could be searched in the same way that genome databases are searched. Most definitive secondary structures have been determined by sequence comparison (1,2), but often there are not enough homologous sequences for this method. Free energy minimization can provide a secondary structure model for a single sequence (3,4) and can also facilitate structure comparison (5,6). When tested against known secondary structures of roughly 150 000 nt of RNA domains with each having less than 700 nt, free energy minimization with the program RNAstructure predicts an average of only 73% of known canonical base pairs (4). Within a few kcal/mol, however, a structure with 87% of known canonical base pairs is generated on average. This suggests that structure predictions can be improved by constraining free energy minimization with experimental data and this has been demonstrated for chemical mapping data (4).

Binding of oligonucleotides can provide additional insight into RNA secondary structure (7–9) and generate data that can be used to constrain the free energy minimization. Development of oligonucleotide microarrays allows rapid determination of sequences that bind to a folded RNA (10,11). For example, Southern and coworkers have made arrays containing all DNA oligonucleotides from monomers to as long as 21-mers that are complementary to an RNA target and used them to probe large RNAs (12,13). However, binding of unmodified DNA oligonucleotides is not optimal for providing constraints for structure prediction algorithms. Since binding of unmodified DNA to RNA is relatively weak (14), most binding is to probes longer than decamers. Moreover, the free energy of binding is very sequence

*To whom correspondence should be addressed. Tel: +1 585 275 3207; Fax: +1 585 276 0205; Email: turner@chem.rochester.edu

dependent for a given probe length. Both characteristics complicate interpretation of binding data.

A short probe on a microarray will only capture a folded RNA target when the complementary region of the target RNA is largely single stranded or only weakly paired. One way to interpret microarray data is to assume that a probe will only capture an RNA if the middle nucleotide of the probe is complementary to a target nucleotide that is not in a Watson–Crick base pair flanked by Watson–Crick base pairs in the target secondary structure. This can provide a constraint for prediction of secondary structure by computational methods (15,16). This assumption was sufficient to explain binding of 7-mers to *Escherichia coli* 5S rRNA (15), but not of 9-mers to the 3′ untranslated region of *Bombyx mori* R2 RNA (16). This suggests that interpretation of microarray data can be simplified by using shorter oligonucleotides. Most nucleotides not in Watson–Crick pairs in structured RNAs are in loops with fewer than seven such contiguous nucleotides. Thus, oligonucleotides shorter than 7-mers should be useful probes of structure.

Here, a new method using microarrays of short isoenergetic oligonucleotides is coupled with chemical modification and free energy minimization calculations to model the secondary structure of a recently discovered RNA. The isoenergetic microarray is composed of pentamer and hexamer 2′-*O*-methyl RNA probes modified by inclusion of LNA (locked nucleic acids, i.e. oligonucleotides with a methyl bridge between the 2′ oxygen and 4′ carbon of ribose) and 2,6-diaminopurine. The modified nucleotides provide similar predicted free energies of binding to a completely unfolded complementary RNA, independent of base content and sequence. Thus, all probes have similar stringency for binding and relative binding to target depends primarily on the free energy required to unfold target. This simplifies interpretation of binding in terms of target structure when compared to that required when the binding also depends on base content and sequence of each probe. The modifications also provide sufficient free energy to allow the capture of target RNA by short probes, which simplifies interpretation because self-folding of probes is eliminated. RNA, 2′-*O*-methyl RNA, and LNA all favor A-form helixes (17–19), so hybridization with 2′-*O*-methyl and LNA backbones minimizes helix distortions. The 2′-*O*-methyl and LNA backbones are also relatively stable chemically and resistant to nuclease digestion (20–25).

The 323 nt from near the 5′ end of the coding region of the R2 retrotransposable element from *B. mori* (R2Bm 5′ RNA) were probed by an isoenergetic oligonucleotide microarray and by small chemicals. This region is important during integration of the sequence into a specific site in the *B. mori* genome (26). In particular, it signals one protein of a probable R2 protein dimer to bind DNA downstream of the insertion site thus positioning the R2 protein for second strand cleavage of the DNA. The results demonstrate the power of combining the isoenergetic microarray approach with chemical mapping and free energy minimization to provide an initial model for the secondary structure of a novel RNA.

MATERIALS AND METHODS

Materials

A, C, G and U 2′-*O*-methyl RNA phosphoramidites and C6-aminolinker for oligonucleotide synthesis were purchased from Glen Research and Proligo. LNA phosphoramidites were synthesized as described previously (27). Phosphoramidites of LNA 2,6-diaminopurine riboside and 2′-*O*-methyl-2,6-diaminopurine riboside were synthesized according to newly developed procedures (28).

Restriction enzymes: XcmI and SacI were from New England BioLabs. Taq polymerase was a product of Promega. AmpliScribe transcription kit was from Epicentre. DNA oligonucleotides for PCR and *in vitro* reverse transcription were purchased from Integrated DNA Technologies.

The γ ^{32}P-ATP was purchased from Perkin Elmer. Reverse transcriptase SuperScript III, T4 polynucleotide kinase and agarose were from Invitrogen; dNTPs and ddNTPs were from Amersham Biosciences. Dimethyl sulfate (DMS) and 1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimide metho-*p*-toluenesulfonate (CMCT) were from Aldrich and kethoxal was from ICN Biomedicals. *N*-methylisatoic anhydride (NMIA) was from Molecular Probes. Silanized slides and probe-clip press seal incubation chambers for hybridization experiments were purchased from Sigma.

Chemical synthesis of modified oligonucleotide probes

Oligonucleotides were synthesized by the phosphoramidite approach on an ABI 392 synthesizer. The modified oligonucleotides used as probes for microarrays were synthesized with a C6-aminolinker on the 5′-end. Oligonucleotides were deprotected and purified according to published procedures (29) with several changes. Oligonucleotides were cleaved from support and base labile protecting groups were removed by incubation for 16 h with 40% methylamine in water at room temperature. The MMTr protecting group on C6-aminolinker was removed by incubation for 3 h at room temperature in 80% acetic acid. Deprotected oligonucleotides were purified on F_{254}, TLC plates in: nPrOH/NH_4OH/H_2O 55:45:10 v/v/v.

Molecular weights were confirmed by mass spectrometry (LC MS Hewlett Packard series 1100 MSD with API-ES). Concentrations of all oligonucleotides were determined from predicted extinction coefficients for RNA and measured absorbance at 260 nm at 80°C (30,31). It was assumed that 2′-*O*-methyl RNA–LNA chimeras and RNA strands with identical sequences have identical extinction coefficients.

Preparation of isoenergetic microarrays

Microarrays were prepared on agarose-coated slides according to the method described by Afanassiev *et al.* (32) with changes. Silanized slides were coated with 2% agarose activated by $NaIO_4$. On dried slides, 0.4 µl of 200 µM of each probe were spotted and incubated for 4 h at 37°C in a 100% humidity chamber. The remaining aldehyde groups on microarrays were reduced with 35 mM $NaBH_4$ solution in PBS buffer (137 mM NaCl, 2.7 mM

KCl, 4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4, pH 7.5) and ethanol (3:1 v/v). Then slides were washed in water at room temperature (3 washes for 30 min each), and in 1% SDS solution at 55°C for 1 h, and finally in water at room temperature (3 washes for 30 min each) and dried at room temperature overnight.

Synthesis and purification of 5′ end open reading frame of R2 RNA from *B. mori*

The DNA sequence of the R2Bm 5′ region was PCR amplified with primers: 5′CGCAGAACTGGCAGGTC CAACCAG3′ and 5′GCGTAATACGACTCACTATAG GGCCGGTGTAACCCGGATGGCTG3′, which contains the T7 promoter. Eight PCR reactions with each containing 70 ng of DNA template, 10 pmol of each primer and 2.5 U of Taq DNA polymerase in 50 µl were run according to the protocol from Promega. R2Bm 5′ RNA was made from 1 µg of PCR template by *in vitro* transcription with an AmpliScribe transcription kit, and purified on an 8% polyacrylamide denaturing gel.

Buffers, folding method and native gel electrophoresis

Native gel electrophoresis indicated that R2Bm 5′ RNA forms a single common structure under various folding conditions. R2Bm 5′ RNA radioactively labeled at the 5′-end with γ ^{32}P-ATP (15 000 c.p.m. per lane) was folded in buffers: (i) 200 mM NaCl, 5 mM $MgCl_2$, 10 mM Tris–HCl, pH 8.0 (0.2 Na^+/5 Mg^{2+}/10T) and (ii) 1 M NaCl, 5 mM $MgCl_2$, 10 mM Tris–HCl, pH 8.0 (1 Na^+/5 Mg^{2+}/10T) using several conditions: (a) incubation for 5 min at 65°C and slowly cooling to room temperature, (b) incubation for 15 min at 65°C and slowly cooling to room temperature, (c) incubation for 2 min at 90°C and slowly cooling to room temperature, (d) incubation for 45 min at room temperature. After folding, samples were analyzed by electrophoresis on a 15 cm, 4% polyacrylamide nondenaturing gel containing Tris–borate–EDTA, pH 8.0 with a running buffer of Tris–borate–EDTA (89 mM Tris, 89 mM boric acid, 2 mM sodium EDTA, pH 8.0). The gel was pre-electrophoresed at 350 V for 0.5 h. Electrophoresis was at 350 V for 1 h at 4°C. Dried gels were analyzed by exposing to X-ray film and also to a phosphorimager screen. In all conditions, the same single band was obtained. For further experiments, R2Bm 5′ RNA was refolded in buffer (1) or (2) with incubation for 5 min at 65°C and slowly cooling to room temperature.

Hybridization conditions

R2Bm 5′ RNA was radioactively labeled with γ ^{32}P-ATP on the 5′-end according to standard procedure and purified on an 8% polyacrylamide denaturing gel. For hybridization, labeled R2Bm 5′ RNA was used at an approximate concentration of 10 nM. The hybridizations were performed as described previously (15) and in the same buffers used for folding. R2Bm 5′ RNA was incubated with the microarray for 18 h at room temperature or 4°C using probe-clip press seal incubation chambers with 200 µl of hybridization buffer. After hybridization, buffers with R2Bm 5′ RNA were poured out and slides were washed in buffers with the same salt concentrations for 1 min at 0°C. (One minute is the estimated half-life predicted at 0°C for binding of the least stable probe (#9) that binds strongly to its exact complement.) Then, slides were dried by slow centrifugation in a clinical centrifuge and covered with saran wrap. Hybridization was visualized by exposure to a phosphorimager screen, which was then scanned on a Molecular Dynamics 840 Storm Phosphorimager. Quantitative analysis was done with ImageQuant 5.2 software. Binding was considered strong when the integrated intensity was $\geq 1/3$ of the strongest integrated intensity for a given condition. Experiments were repeated at least three times and the average of the data is presented.

Chemical mapping

Chemical mapping was performed according to procedures described earlier (15,33). Dimethyl sulfate (DMS) was used to modify adenosine and cytidine, kethoxal to modify guanosine and CMCT to modify uridine. For chemical mapping, 1 pmol of R2Bm 5′ RNA was taken for each reaction and folded in 0.2 Na^+/5 Mg^{2+}/10T, as described above. Then tRNA carrier was added to give a total RNA concentration of 8 µM and the solution was incubated for 10 min at room temperature. To a 9 µl sample, 1 µl of DMS or kethoxal solution was added. DMS was diluted in ethanol and used at a final concentration of 60 mM. Kethoxal was diluted in ethanol/water (1:3 v/v) to give a final concentration of 160 mM. After modification with kethoxal, 3 µl of 35 mM potassium borate solution was added to stabilize products of modification. For modification with CMCT, 9 µl of CMCT solution was added to the 9 µl of RNA sample. CMCT was diluted in an appropriate buffer to give a final concentration of 625 mM in the reaction mixture. Chemical modification reactions were performed for 20 min at room temperature. Reactions were stopped by ethanol precipitation on dry ice.

Chemical mapping with *N*-methylisatoic anhydride (NMIA) was done as described (34) with some changes. For each reaction, 1 pmol of R2Bm 5′ RNA was taken and refolded as described above. To a 9 µl sample, 1 µl of NMIA solution (1 mg NMIA/42 µl DMSO) was added. Samples were incubated for 3 h at room temperature. Reactions were stopped by ethanol precipitation on dry ice.

Primer extension reactions

DNA primers for primer extension reactions had sequences: 5′GACGGTCCTCGCGGTCCG3′ (61–78), 5′CGCAGGTATCGCAAGGCC3′ (104–121), 5′GGCTT CAGGTCTATTTTCTTTATTTGAC3′ (184–211), 5′CC CCGCACAGTCGGGTTATC3′ (245–264) and 5′CGCA GAACTGGCAGGTCC3′ (306–323), where the numbers in parentheses denote the complementary region of R2 5′ RNA. Primers were labeled on the 5′-end with γ ^{32}P-ATP according to standard procedure. For each reaction, 1 pmol of primer was used. Primer extension was performed at 55°C with reverse transcriptase SuperScript III and Invitrogen's buffers and protocol. Reactions were stopped by adding loading buffer containing formamide

and EDTA and chilling to 0°C. Products were heated and then separated on a 12% or 16% polyacrylamide denaturing gel. The gels were analyzed with the ImageQuant 5.2 program and products were identified by comparing to sequencing lanes for A, C, T and G and to control lanes. Modifications were initially identified by visual inspection of autoradiograms and were considered strong or medium when the band corresponding to chemical modification had at least 6 times, or 2–6 times, respectively, the integrated intensity of the equivalent band in the control lane, as quantified with ImageQuant 5.2. DNA sequences complementary to the R2 sequences C117–G134, G122–C150, G123–G140 and G135–C150 were not able to prime reverse transcription.

RESULTS

Isoenergetic microarrays

Nearest neighbor parameters are available for predicting the stabilities of RNA/2′-*O*-methyl RNA duplexes in 100 mM NaCl (28,35). Stabilities of RNA/2′-*O*-methyl RNA duplexes can be increased by substituting LNA nucleotides for some 2′-*O*-methyl nucleotides (27) and by substituting 2,6-diaminopurine (D) for adenosine (28). The enhancement in stability can be approximated from:

$$\begin{aligned}\Delta G^\circ_{37}\,&(\text{modified probe/RNA, 100 mM NaCl}) = \\ &\Delta G^\circ_{37}(2'\text{–}O\text{–MeRNA/RNA, 100 mM NaCl}) \\ &- 0.53\ n_{5'tL} - 1.28\ n_{iAL/UL} - 1.58\ n_{iGL/CL} \\ &- 1.34\ n_{iDL} - 0.14\ n_{3'tUL} - 1.23\ n_{3'tAL/CL/GL/DL} \\ &- 1.50\ n_{add.3'GLmismatch}\end{aligned} \quad \mathbf{1}$$

Here ΔG°_{37} (2′-*O*-MeRNA/RNA, 100 mM NaCl) is the free energy change at 37°C for duplex formation with Watson–Crick pairs in the absence of LNA nucleotides as calculated with nearest neighbor parameters (28) with the m(5′-DD)/r3′-UU) nearest neighbor ΔG°_{37} approximated as −1.5 kcal/mol. The symbol $n_{5'tL}$ is the number (0 or 1) of 5′ terminal LNAs; $n_{iAL/UL}$, $n_{iGL/CL}$ and n_{iDL} are the number of internal LNAs in AU, GC and DU pairs, respectively; $n_{3'tU}$ and $n_{3'tAL/CL/GL/DL}$ are the number (0 or 1) of LNAs that are U or not U, respectively, and are in a Watson–Crick base pair at the 3′ end of a Watson–Crick paired helix; $n_{add3'GL\ mismatch}$ is the number (0 or 1) of LNA 3′ G in mismatches, GA, GG or GU, for a hexamer. Equation (1) was derived from experimental results at 100 mM NaCl (27,28,35), but different salt conditions are expected to change stability in a sequence-independent manner. Parameters in Equation (1) are those reported by Pasternak *et al.* (28,36). Equation (1) allowed design of 2′-*O*-methyl RNA–LNA chimeric oligonucleotides with roughly isoenergetic binding to unstructured RNA.

On the basis of microarray results for 7-mers binding to *E. coli* 5S rRNA (15), oligonucleotides with predicted ΔG°_{37} values of −10 kcal/mol provide excellent probes of secondary structure. The 319 pentamer-binding sites on the R2Bm 5′ RNA have 256 unique sequences. It was possible to design 71 pentamers with predicted ΔG°_{37} more favorable than −8.5 kcal/mol and another 5 pentamers with ΔG°_{37} between −7.7 and −8.3 kcal/mol were also used. To expand the number of sites probed, a 3′ LNA G was added to 156 probe sequences to increase binding. If the 3′ LNA G is complementary to a C in the RNA target, then binding is predicted to be enhanced by an average of 3.5 kcal/mol at 37°C (36). If the 3′ LNA G is opposite to an A, G or U in the RNA target, then the stability enhancement is roughly sequence independent, averaging 1.5 kcal/mol (36). Of the 232 probes specific for R2Bm 5′ RNA, 45 and 7 probes have two and three complementary binding sites, respectively, on the target RNA. Altogether this covers 291 sites, which is 91% of all possible pentamer-binding sites on R2Bm 5′ RNA. From all possible sites, 28 were omitted because calculated free energies of even highly modified probes could not provide reasonable binding. Figure 1 shows the distribution of number of probes versus predicted ΔG°_{37} of binding to an unstructured target. Table S1 lists all the probes used. This library provides roughly 25% of the coverage required for a microarray able to interrogate all 1024 possible pentamer sequences in RNA.

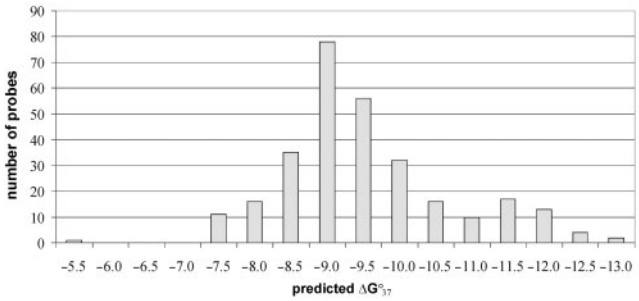

Figure 1. Number of probes with predicted ΔG°_{37} falling within the free energy windows shown on the *x*-axis. The ΔG°_{37} is predicted for binding to an unfolded RNA with the sequence from R2Bm 5′ RNA that is Watson–Crick complementary to the first 5 nt of the probe. If the probe is a hexamer, then the pairing of the 3′ LNA G with the R2Bm 5′ RNA sequence is also included.

The average predicted free energy of binding of the modified library to complementary sites on the R2Bm 5′ RNA is -9.8 ± 1.2 kcal/mol which is 4.8 kcal/mol more stable than the average for unmodified pentamers. The stability enhancement is large, and just as important, the difference between free energies is relatively small. For an equivalent library of unmodified 2′-*O*-methyl RNA pentamers, 70% had predicted free energies between −3.6 and −6.6 kcal/mol for binding to complementary unstructured pentamer RNA. For the modified pentamers and hexamers, 79% had predicted free energies between −8.7 and −11.7 kcal/mol for binding to their R2Bm 5′ RNA sites if unstructured (Figure 1). This corresponds to about a 100-fold range in equilibrium constants at 37°C.

Structure probing with isoenergetic microarrays at 200 mM NaCl, 5 mM $MgCl_2$ at room temperature

Bombyx mori live at room temperature, ~23°C, and have an average body temperature of ~24°C (37). The structure of R2Bm 5′ RNA was probed at room temperature with isoenergetic microarrays (Figure 2) having the sequences listed in Table S1. Hybridization results are collected in Tables 1 and 2 and illustrated in Figures 2 and 3.

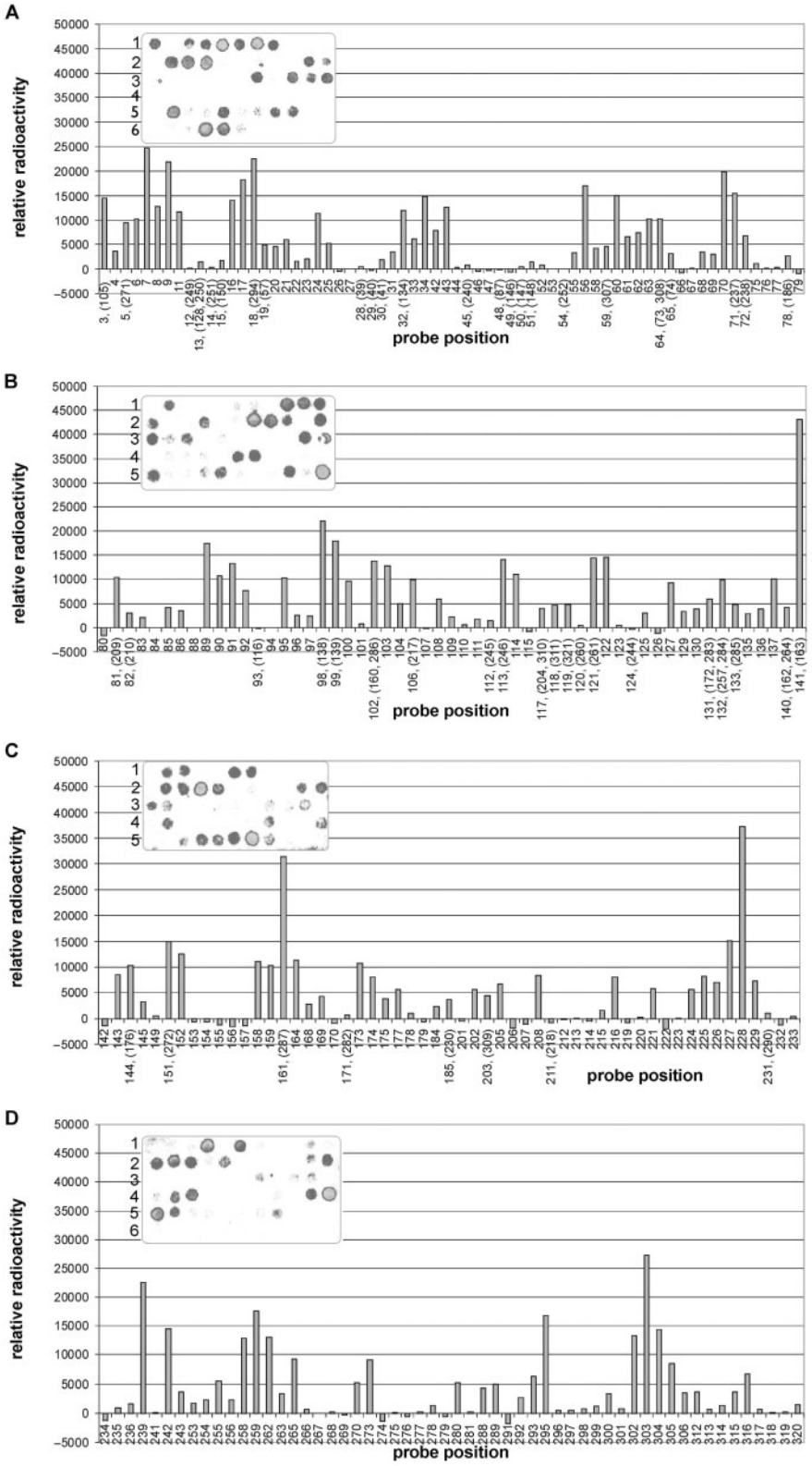

Figure 2. Hybridization results on isoenergetic library for R2 5′ RNA from *B. mori* in buffer 200 mM NaCl, 5 mM $MgCl_2$, 10 mM Tris–HCl, pH 8.0 (0.2 Na^+/5 Mg^{2+}/10T) at room temperature. Probes on microarrays are labeled by the number of the R2 5′RNA complementary to the third nucleotide of the probe and are spotted in the orders listed, starting at the left side of each row. **(Panel A)** Row 1: 3–9, 11–14; row 2: 15–25; row 3: 26–34, 42, 43; row 4: 44–54; row 5: 55, 56, 58–66; row 6: 67–72, 75–79. (**Panel B**) Row 1: 80–86, 88–91; row 2: 92–102; row 3: 103, 104, 106–114; row 4: 115, 117–126; row 5: 127, 129–133, 135–137, 140, 141. (**Panel C**) Row 1: 142–145, 149, 151–156; row 2: 157–159, 161, 164, 168–171, 173, 174; row 3: 175, 177–179, 184, 185, 201–203, 205, 206; row 4: 207, 208, 211–216, 219–221; row 5: 222–229, 231–233. (**Panel D**) Row 1: 234–236, 239, 241–243, 253–256; row 2: 258, 259, 262, 263, 265–270, 273; row 3: 274–281, 288, 289, 291; row 4: 292, 293, 295–303; row 5: 304–306, 312–319; row 6: 320.

Table 1. Probes binding R2Bm 5′ RNA strongly and used as constraints[a]

Probe sequence 5′ to 3′	Center of binding site	ΔG°_{37} (base pairing) kcal/mol[b]
$C^{L}DC^{L}CG^{L}$	7	−9.02
$UD^{L}CD^{L}Cg^{L}$	9	−7.82
$CD^{L}UC^{L}GG^{L}$	56	−11.63
$CUC^{L}CD^{L}G^{L}$	227	−11.97
$DC^{L}UC^{L}Cg^{L}$	228	−10.04
$CG^{L}CD^{L}Cg^{L}$	259	−10.27
$D^{L}DC^{L}CD^{L}G^{L}$	303	−11.80

[a]Hybridization was in 200 mM NaCl, 5 mM $MgCl_2$, 10 mM Tris–HCl, pH 8.0 (0.2 Na^+/5 Mg^{2+}/10T) at room temperature. Nucleotides with and without a superscript L are LNA and 2′-*O*-methyl, respectively. D represents 2,6-diaminopurine. G^L and g^L represent a 3′ terminal LNA G that forms a G^LC or mismatch pair, respectively, with R2Bm 5′ RNA. Probes are used as constraints if they do not have potentially strong binding alternative sites, or if there is no strong binding by the probe completely complementary to an alternative binding site.
[b]Calculated from Equation (1), which assumes 100 mM Na^+.

Each probe is identified by a number corresponding to the number of the R2Bm 5′ RNA nucleotide in the middle of the sequence completely complementary to the first 5 nt of the probe. Hexamer probes all have a 3′ terminal LNA G which is denoted in Tables 1 and 2 by a g^L or G^L in order to indicate a terminal g^LA, g^LG or g^LU mismatch or a G^LC pair, respectively, with the R2Bm 5′ RNA-binding site. A second binding site with equal complementarity to a probe is listed in parentheses. Only strong binding was considered. In general, few probes strongly bind R2Bm 5′ RNA at room temperature. In 0.2 Na^+/5 Mg^{2+}/10T, only 25 of 232 probes strongly bind R2Bm 5′ RNA.

Structure probing with isoenergetic microarrays under different conditions

Hybridization experiments were also done for 0.2 Na^+/5 Mg^{2+}/10T at 4°C and for 1 Na^+/5 Mg^{2+}/10T at room temperature and at 4°C. There are probes that bind under all conditions, but generally binding depends on conditions applied for hybridization (see Table S1). For 0.2 Na^+/5 Mg^{2+}/10T, more probes bound strongly at 4°C than at room temperature. At both temperatures, more probes bind in 1 Na^+/5 Mg^{2+}/10T than in 0.2 Na^+/5 Mg^{2+}/10T. The differences in binding due to conditions can come from both small structural changes and different free energies of base pairing and folding. When specificity is less, however, interpretation is more complicated. Thus the R2Bm 5′ RNA secondary structure was modeled from room temperature results in 0.2 Na^+/5 Mg^{2+}/10T.

Secondary structure of 5′ R2 RNA from *B. mori* deduced from microarray results

Microarray results in 0.2 Na^+/5 Mg^{2+}/10T at room temperature were used as constraints in predicting the secondary structure of R2Bm 5′ RNA. This buffer corresponds to that used for biochemical experiments (26) and provides the greatest stringency for binding to probe. Interpretation of microarray results is complicated by multiple potential binding sites, including both fully complementary and mismatched sites. For each strongly binding probe, predicted free energies of binding to fully complementary and mismatched sites were calculated for unmodified RNA probes binding to unstructured RNA target with the program RNA structure in (38) bimolecular binding mode (Table S1) using nearest neighbor parameters for RNA/RNA duplexes (29,39). The difference between free energies of binding to complementary and alternative binding sites is likely similar for RNA/RNA and for the modified probes. This facilitates identification of probes that might bind with mismatches and the sites where they might bind. Strongly binding probes with alternative binding sites having free energy predicted by the bimolecular mode of RNAstructure to be more favorable than −6.0 kcal/mol at 37°C and that strongly bind their exact complement were not used as constraints for prediction of secondary structure by RNAstructure. A probe was also not used as a constraint if the bimolecular binding mode of RNAstructure predicted an alternative site less favorable than −6.0 kcal/mol but within 2 kcal/mol of the complementary site and the alternate site bound strongly to its exact complement. The secondary structure of the R2Bm 5′ RNA was predicted with RNAstructure 4.4 in folding mode with the middle nucleotide of strong binding sites for probes constrained not to be in an AU or GC pair flanked on each side by an AU or GC pair. The hybridization constraints used for prediction are listed in Table 1 and shown in Figure 3 along with the predicted structure.

Once a secondary structure is generated, it is possible to compare predicted free energies of binding with observed binding because the free energy required to unfold local structure can be estimated. This prediction eliminated some of the possible cross-hybridization sites identified with RNAstructure, which does not consider folding of the RNA target. Inspection of the proposed secondary structure also revealed that the hairpin centered at G287 could bind several of the probes. To test this hypothesis, the hairpin, 5′GCCUGUGGGUCAGGC3′, was synthesized and binding to the microarray was measured. It bound strongly to probes 161(287), 99(139) and 242. Table 2 lists the probable binding sites of strongly binding probes that were not used as constraints. Figure 3 shows those that are relatively certain.

Chemical mapping of R2 5′ RNA from *B. mori*

To test the structure deduced from microarray data, R2Bm 5′ RNA was chemically mapped in 0.2 Na^+/5 Mg^{2+}/10T at room temperature (Figures 3 and 4). Not many regions are accessible for chemical reactivity, thus confirming a very condensed structure. Loop regions are suggested by strong reactivity at G33, U34, A35, A36 and A37, and medium reactivity at U285, G286, G287, G288 and U289. There are also many medium modifications from A182 to A200. In this region, almost every base is modified during mapping, but only A200 is strongly modified. Mapping with *N*-methylisatoic anhydride (NMIA), which identifies flexible sugars (34), is largely in agreement with results from DMS, kethoxal and CMCT (Figure 3).

When strong modifications by DMS, kethoxal and CMCT and constraints from microarray data are

Table 2. Probes binding R2Bm 5′ RNA strongly but not used as constraints[a]

Probe sequence 5′ to 3′	Center of binding site	ΔG°_{37} (base pairing) kcal/mol[b]	Possible cross hybridization site(s)[c]	Probable binding site(s) (ΔG°_{37})[d]
GGCC^{L}C	3(105)	−9.72	122/123	3 or 105
AU^{L}CC^{L}GG^{L}	17	−11.23	99	17 (−6.8)
CD^{L}UC^{L}CG^{L}	18(294)	−11.59	99/100	18 (−6.0)
				99/100 (−6.7)
U^{L}UD^{L}CC^{L}g^{L}	34	−8.16	7	34 (−8.2)
CG^{L}UC^{L}Cg^{L}	60	−10.57	Many	122/123 (−9.7)
				238/239 (−5.7)
CC^{L}UC^{L}Gg^{L}	70	−10.55	56	70 (−7.3)
U^{L}CC^{L}UC^{L}G^{L}	71(237)	−12.42		71 (−9.2)
				237 (−7.9)
D^{L}CC^{L}UG^{L}g^{L}	89	−10.28	98/99	98/99 (−11.2)
			241/242	
CCC^{L}GA	98(138)	−8.05	121/122	98 (−8.5)
			261/262	
ACC^{L}CG^{L}	99(139)	−8.79	121/122	99 (−7.8)
			261/262	287 (−8.5)
GGCC^{L}C	105(3)	−9.72	122/123	3 or 105
CCC^{L}GC	121(261)	−9.22	98/99	98/99 (−6.7)
			138/139	
GCC^{L}CG	122	−9.22	Many	99 (−6.7)
CCC^{L}GA	138(98)	−8.05	121/122	98 (−8.5)
			261/262	
ACC^{L}CG^{L}	139(99)	−8.79	121/122	99 (−7.8)
			261/262	287 (−8.5)
D^{L}DD^{L}CC^{L}g^{L}	141(163)	−8.94	Many	Many
GCC^{L}GG	151(272)	−9.26		303 (−4.6)
AC^{L}CC^{L}Ag^{L}	161(287)	−9.89	99/100	287 (−9.9)
			139/140	
D^{L}DD^{L}CC^{L}g^{L}	163(141)	−8.94	Many	Many
U^{L}CC^{L}UC^{L}g^{L}	237(71)	−9.99		71 (−9.2)
				237 (−7.9)
U^{L}GU^{L}CC^{L}g^{L}	239	−9.77	122/123	122/123 (−6.5)
			60	239 (−5.6)
GCC^{L}UG^{L}	242	−9.39	89	122 (−5.3)
			122	287 (−6.7)
CCC^{L}GC	261(121)	−9.22	98/99	98/99 (−5.1)
			138/139	
GCC^{L}GG	272(151)	−9.26		303 (−4.6)
AC^{L}CC^{L}Ag^{L}	287(161)	−9.89	99/100	287 (−9.9)
			139/140	
CD^{L}UC^{L}Cg^{L}	294(18)	−9.51	99/100	18 (−6.0)
				99/100 (−6.7)
GCD^{L}UC^{L}g^{L}	295	−9.33	7/8	295 (−4.1)

[a]Hybridization was in 200 mM NaCl, 5 mM $MgCl_2$, 10 mM Tris–HCl, pH 8.0 (0.2 Na^+/5 Mg^{2+}/10T) at room temperature. Nucleotides with and without a superscript L are LNA and 2′-*O*-methyl, respectively. D represents 2,6-diaminopurine. G^{L} and g^{L} represent a 3′ terminal LNA G that forms a G^{L}C or mismatch pair, respectively, with R2Bm 5′ RNA. Probes are used as constraints if they do not have potentially strong binding alternative sites, or if there is no strong binding by the probe completely complementary to an alternative binding site.

[b]Calculated from Equation (1), which assumes 100 mM Na^+.

[c]A possible cross-hybridization site binds its Watson–Crick complementary probe or a directly adjacent probe strongly and when not folded is predicted by the bimolecular binding mode of RNAstructure 4.4 to have a free energy more favorable than −6.0 kcal/mol for binding the indicated probe, which binds with at least one non-Watson–Crick pair (see Table S1).

[d]Probable binding sites are those expected to have ΔG°_{37} at least as favorable as −4.0 kcal/mol for binding to the proposed secondary structure and that bind tightly their exactly complementary probe. Estimates of ΔG°_{37} (kcal/mol) for binding used nearest neighbor parameters for 2′-*O*-methyl RNA/RNA at 100 mM Na^+ (28) with the assumption that a GU pair is equivalent to an AU pair along with nearest neighbor parameters for target RNA folding at 1 M Na^+ (29,38). Correction for the difference in salt concentrations and temperature would lead to predicted binding being more favorable. No prediction is made for probe 3 (105) because thermodynamics have not been measured for model systems containing multibranch loops with more than four helixes or for complicated pseudoknots. Free energy values in parentheses are for the probable alternative site listed adjacent (see text for description of calculation).

used together as chemical mapping constraints in RNAstructure 4.4, the predicted structure is identical to that predicted with only constraints from microarray data (Figure 3). When only strong modifications by DMS, kethoxal and CMCT are used as constraints, however, the structure shown in Figure 4 is predicted. This is also the structure predicted in the absence of constraints. Its predicted ΔG°_{37} of folding is −154.6 kcal/mol, which is 1.5 kcal/mol more favorable than the predicted ΔG°_{37} for the structure in Figure 3.

NMIA mapping data were not used in generating the structures shown in Figures 3 and 4 because the exact

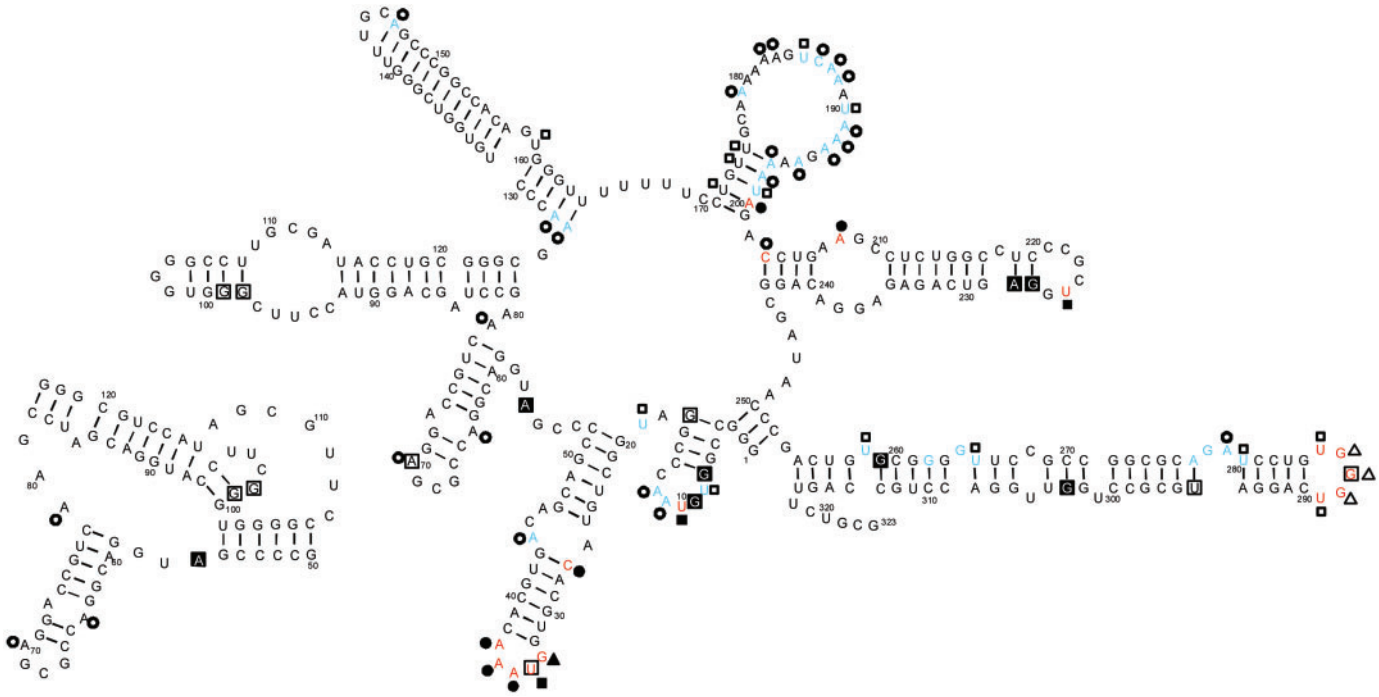

Figure 3. Prediction of secondary structure of R2Bm 5′ RNA from RNAstructure 4.4 (4) using chemical mapping (excluding NMIA) and microarray hybridization constraints or hybridization constraints alone at room temperature in buffer 200 mM NaCl, 5 mM $MgCl_2$, 10 mM Tris–HCl, pH 8.0 (0.2 Na^+/5 Mg^{2+}/10T). Hybridization constraints required that the nucleotide complementary to the third nucleotide of a strongly binding probe with a unique binding site could not be in a Watson–Crick pair flanked on both sides by Watson–Crick pairs. Chemical mapping constraints required that nucleotides strongly modified by DMS, kethoxal or CMCT could also not be in a Watson–Crick pair flanked on both sides by Watson–Crick pairs. The pseudoknot shown on the left is an alternate folding of nucleotides 50–123. RNAstructure does not allow pseudoknots. Probes providing hybridization constraints: 7, $9g^L$, $56G^L$, $227G^L$, $228g^L$, $259g^L$, $303G^L$, where g^L and G^L indicate a hexamer with a 3′ terminal LNA G that forms a mismatch or a G–C pair, respectively. Pentamer or hexamer probes that bind tightly but have more than one potentially strong binding site: 3(105), $17G^L$, $18G^L$ ($294g^L$), $34g^L$, $60g^L$, $70g^L$, $71G^L$($237g^L$), $89g^L$, 98(138), 99(139) 121(261), 122, $141g^L$($163g^L$), 151(272), $161g^L$($287g^L$), $239g^L$, 242, $295g^L$. Probes not used: 10, 36–38, 86, 162, 165–167, 180–183, 187–200, 247, 248, 264, 290. Strong binding sites of heptamer probes: 8, 9, 10, 239. Chemical mapping constraints: 10, 27, 33, 34, 35, 36, 37, 200, 208, 225. Symbols: filled circle—strong DMS; open circle—medium DMS; filled square—strong CMCT; open square—medium CMCT; filled triangle—strong kethoxal; open triangle—medium kethoxal; A, C, G, or U within filled square—middle nucleotide of site of strong binding used as hybridization constraint; A, C, G, or U within open square—probable middle site of strongly binding probe not used as hybridization constraint; red A, C, G, or U—strong NMIA, blue A, C, G, or U—medium NMIA. None of the nucleotides used as hybridization constraints react with NMIA.

rules for interpreting NMIA data are not known (34). NMIA modifies riboses that are flexible and is therefore not expected to modify nucleotides in Watson–Crick pairs flanked by Watson–Crick pairs. Strong or moderate NMIA reactivity is seen for 1 and 6 nt in Watson–Crick pairs flanked by Watson–Crick pairs in the structures in Figures 3 and 4, respectively. The same difference between structures is seen for moderate DMS and CMCT reactivity. Thus the NMIA and moderate DMS and CMCT data are more consistent with the structure in Figure 3.

Assigning probabilities to base pairs in the proposed structure

The RNAstructure 4.4 algorithm generates additional structures that are consistent with the experimental constraints but with less favorable predicted free energies. The free energies of predicted structures can be used in a partition function method (40) to assign a probability to each base pair (Figure 5). This calculation considers all structures allowed by the constraints and weights each base pair by the sum of the Boltzmann weights for each structure in which it appears. This provides a quantitative measure of the certainty of prediction for each base pair. Red indicates >99% probability and unshaded <50% probability with other colors indicating intermediate probabilities.

DISCUSSION

Secondary structures are definitively known for only a few classes of RNA and those are mostly ribozymes (2,41–43). Here, it is shown that constraints from isoenergetic microarray and chemical mapping data can be coupled with free energy minimization to allow rapid modeling of the secondary structure of an RNA. The test RNA, whose structure has not been previously studied, is a part of the transcript coding for the R2 retrotransposon in *B. mori*. This segment of the RNA transcript orchestrates a change in protein function during retrotransposition (26).

Advantages of isoenergetic microarrays of short oligonucleotides

Interpretation of microarray data can be complicated because binding depends on the differences in free energies

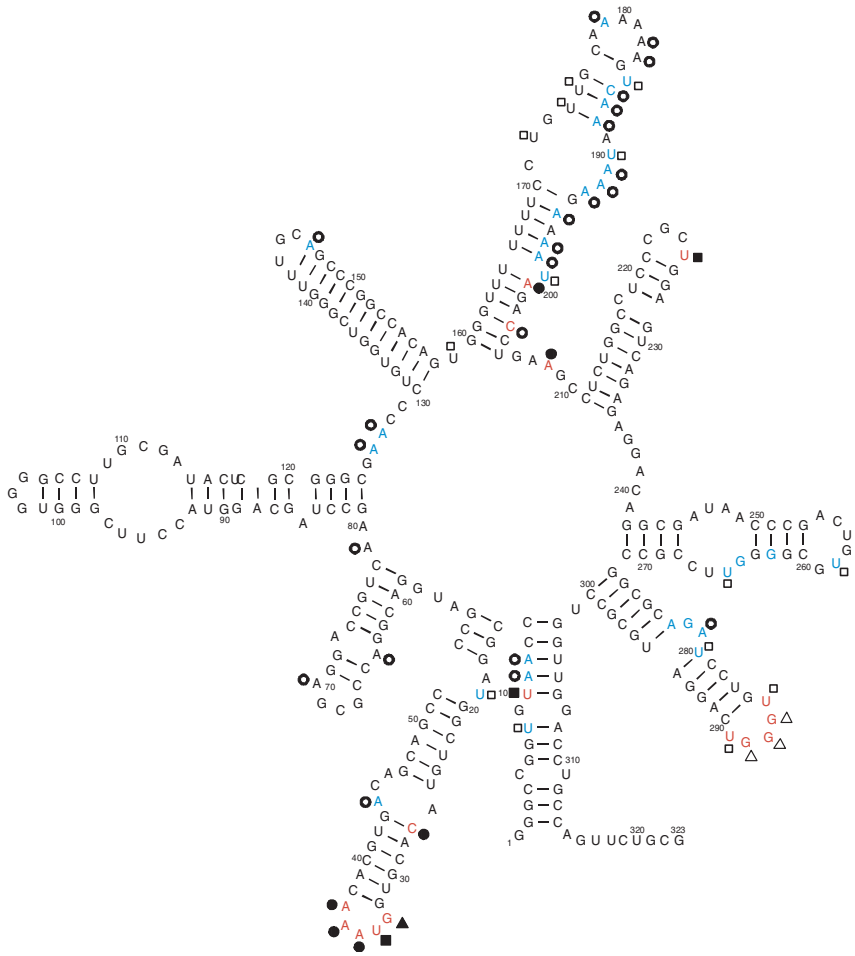

Figure 4. Prediction of secondary structure of R2Bm 5′ RNA from RNAstructure 4.4 with no constraints or with chemical mapping constraints only (excluding NMIA) at room temperature in buffer 200 mM NaCl, 5 mM $MgCl_2$, 10 mM Tris–HCl, pH 8.0 (0.2 Na^+/5 Mg^{2+}/10T). Prediction with hybridization constraints alone and combined with chemical mapping constraints is shown in Figure 3. See caption to Figure 3 for symbols and for description of chemical mapping constaints.

for breaking self-structure in the probe and target and for binding of probe to target. The LNA and 2,6-diaminopurine modifications employed here allow the use of pentamer and hexamer probes, which eliminates self-folding of probe. Short probes also reduce problems associated with prediction of target unfolding because fewer nucleotides need to be unfolded. They are also less likely to allow coaxial stacking with flanking target helixes at both ends of the probe. Coaxial stacking enhances binding (11,44), but is not included in algorithms for prediction of oligonucleotide binding (45). Nucleotide modifications allow design of probes with a limited range of free energies predicted for binding to unpaired RNA, thus further simplifying interpretation. There are only 1024 and 4096 different pentamer and hexamer sequences, respectively, so that 'universal' microarrays applicable to any RNA could be manufactured.

One disadvantage of pentamers and hexamers is that some are likely to have more than one perfectly matched site on a long RNA. Such ambiguous probes cannot provide initial constraints for algorithms predicting secondary structure. They do, however, provide an additional check on a predicted structure because there must be at least one accessible site available for any strongly binding probe (Figure 3, Table 2). Figure 6 presents a flowchart summarizing the steps in modeling and testing a secondary structure as facilitated by microarray and chemical mapping data.

Proposed secondary structure of part of the 5′ half of the open reading frame for R2 RNA from *B. mori*

Figure 3 shows the secondary structure predicted for R2Bm 5′ RNA with constraints from microarray hybridizations in 0.2 Na^+/5 Mg^{2+}/10T at room temperature. After omitting probes with ambiguous binding sites from the microarray data, only nucleotides 7, 9, 56, 227, 228, 259 and 303 were constrained (Table 1). The same structure is predicted when microarray data by themselves are used to constrain folding by RNAstructure 4.4 and when microarray data in conjunction with chemical mapping by DMS, kethoxal and CMCT are used as constraints in RNAstructure 4.4. Only strong chemical modifications were used for constraints, but with the minor exception of nucleotide 127, medium modifications are also consistent with the predicted structure.

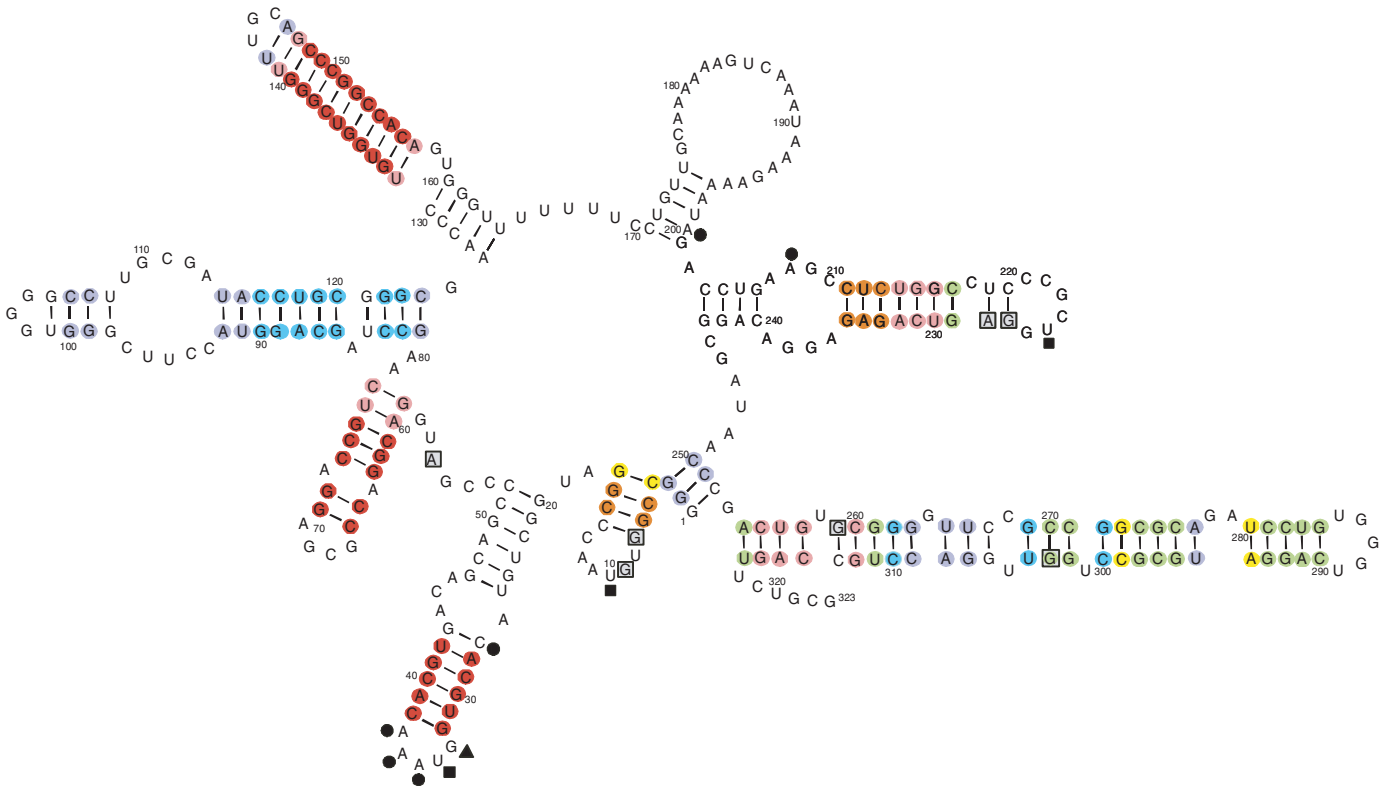

Figure 5. Probability of base pairs in predicted structure of R2Bm 5′ RNA: symbols: red shaded (i.e. A, C, G, or U)—BP probability ≥99%, pink shaded 'N'—99% > BP probability ≥95%, Orange shaded 'N'—95% > BP probability ≥90%, Green shaded 'N'—90% > BP probability ≥80%, Yellow shaded 'N'—80% > BP probability ≥70%, blue shaded 'N'—70% > BP probability ≥60%, purple shaded 'N'—60% > BP probability >50%, unshaded 'N'—50% ≥ BP probability. Chemical mapping, used as constraints: filled circle—strong DMS; filled square—strong CMCT; filled triangle—strong kethoxal. Probes binding: Letter 'N' within gray shaded box—middle nucleotide of site of strong binding used as constraints. Probability is from a partition function calculation in which base pairs are consistent with constraints from microarray binding and strong chemical reactivity and are weighted by the sum of Boltzmann weights for the structures containing them.

NMIA data were not used as constraints because all the rules for reactive nucleotides are not yet known (34). With the minor exception of nucleotide 127, however, the NMIA modifications are also consistent with the proposed structure. Nucleotide 127 is the third purine in the sequence 5′GAA/3′UUU near the end of a helix. This may be a dynamic region of the structure, as discussed below. Two base pairs not predicted, C221–G226 and G253–U318 could also form dynamically.

Interestingly, little reactivity or binding to microarray is observed for the region from G50–G123 even though it contains a 5X5 nucleotide internal loop. Moreover, nucleotides 94–99 could form a CUUCGG tetraloop (46–48). An alternate fold for this region is the pseudoknot containing a CUUCGG tetraloop as shown in Figure 3. RNAstructure does not allow pseudoknots.

One test of the structure is whether it has reasonable binding sites for strongly binding probes not used as constraints (Figure 3). Table 2 lists all the probes not used as constraints that nevertheless bind strongly and also lists potential and probable cross-hybridization sites for these probes. With the exception of probe 3(105), all the probes with multiple potential binding sites have at least one site in the proposed secondary structure that is expected to be a good binding site. Probe 3(105) may bind to site 3 and/or 105, but this cannot be predicted well because thermodynamic parameters have not been measured for complicated multibranch loops and pseudoknots.

The predicted secondary structure is rich in double-stranded regions, with only 35% of nucleotides not in canonical base pairs when the pseudoknot is assumed. There is one clearly accessible region, hairpin loop G33–A37. All 5 nt in the loop react strongly with chemicals. Only oligonucleotide $34g^{L}$ probes this region and it binds strongly.

No other regions are as strongly reactive to chemicals, but all nucleotides in hairpin loop U285–U289 react strongly with NMIA and moderately with kethoxal or CMCT. In hairpin loop U9–C13, U10 is strongly and U8, A11 and A12 are moderately modified by both standard mapping reagents and NMIA. Strong binding of probes 7 and $9g^{L}$ is also consistent with this hairpin.

The partition function calculation (Figure 5) identifies five hairpins as highly probable, those containing loops G33–A37, G67–A70, U143–C145, C221–G226 (or perhaps C222–U225) and U285–U289. Surprisingly, three of the five loops are not very reactive to chemicals. The long hairpin closing loop U143–C145, however, is consistent with the observation that region G122–C150 is inaccessible for long DNA primers which were tried for reverse transcription in the chemical mapping experiments.

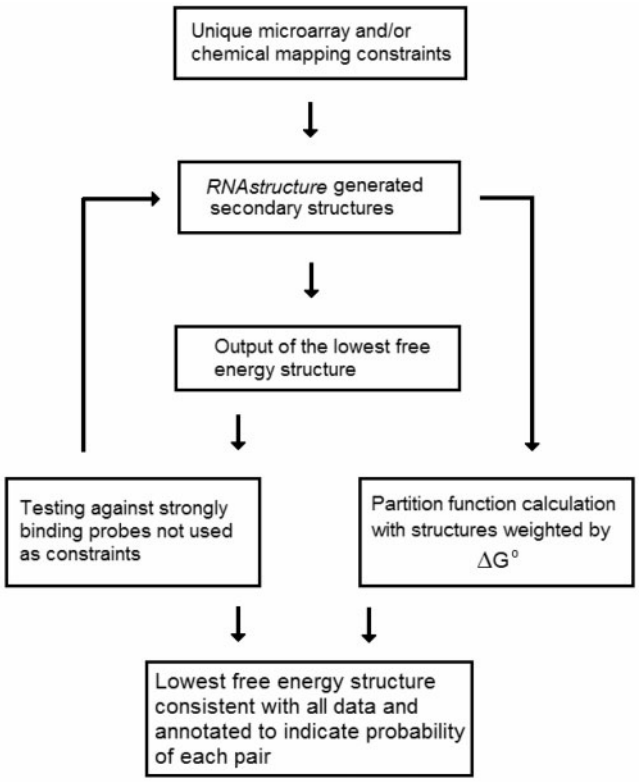

Figure 6. Flowchart for modeling RNA secondary structure with constraints from microarray binding and chemical modification. A unique microarray constraint is defined as a strongly binding probe without an alternate target site that strongly binds its exactly complementary probe and is predicted to have a ΔG°_{37} more favorable than −6.0 kcal/mol for binding of the mismatched probe. If the strongly binding probe has an alternative site that strongly binds its exactly complementary probe and has a predicted ΔG°_{37} for mismatched probe less favorable than −6.0 kcal/mol but within 2 kcal/mol of the predicted ΔG°_{37} for its completely complementary probe, then the strongly binding probe is also not used as a constraint. Predicted ΔG°_{37} values are for RNA/RNA duplexes formed between initially unstructured strands.

Primers complementary to regions C117–G134 or G122–C150 or G123–G140 or G135–C150 do not anneal to target RNA even though the predicted duplex melting temperature is reasonable and they cannot fold into stable self-structures. However, the primer complementary to region G104–G121 primes well.

There is a large hairpin loop in the predicted structure of R2Bm 5′ RNA: U176–A196. Unfortunately, probes 180–183 and 186–200, complementary to that region, could not be used because of low predicted stability of binding. Probes designed for this region: 177g, 178g, 179g, 184g and 185g (230g) do not bind in buffer 0.2 Na^+/5 Mg^{2+}/10T at room temperature, and at 4°C only probe 177g binds. In 1 Na^+/5 Mg^{2+}/10T at room temperature, however, probes 184g and 185g (230g) bind. Perhaps this region is not truly open but rather has dynamic interactions. For example, there is the possibility of pseudoknot formation involving nucleotides 163–169 with any of the A_n sequences between A178 and A198. Another possibility is tertiary interactions between the A_n sequences and the CCAC sequence of nucleotides 153–156 as has been seen for an equivalent sequence in a pseudoknot (49). Most of the A_n nucleotides were moderately modified by DMS and NMIA except for A200, which was strongly modified. Nucleotides U172, U174 and U175 are also moderately reactive to chemicals. This is the only region with so many moderate chemical modifications, consistent with it being dynamic. It would not be surprising, however, if other base pairs unshaded in Figure 5, to represent <50% probability, are also dynamic.

The R2 protein binds to both the 3′ end and this 5′ fragment of R2 RNA (26). There are no identical secondary structure elements, however, to suggest a common binding site for the protein. Several identical or similar sequences of 4–6 nt are found in single-stranded regions, but the motifs are different at the 3′ end (50). The two RNAs confer different properties to the R2 protein, so multiple binding sites are not surprising.

Constraints from oligonucleotide binding do not overlap constraints from chemical modification

Binding of oligonucleotides identifies regions of weak or no intramolecular base pairing in an RNA. Interestingly, of 14 central nucleotides for probes with unambiguous probable binding sites, only one is strongly modified by a chemical (Figure 3). Evidently, oligonucleotide binding can provide information complementary to that available from chemical modification. More overlap can be expected for an RNA with a smaller percentage of nucleotides in canonical base pairs.

NMIA reactivity largely overlaps with reactivity of DMS, kethoxal and CMCT

When strong and moderate hits are considered, 41 of 50 nt modified strongly or moderately by DMS, kethoxal or CMCT are also modified by NMIA. Conversely, 41 of 45 nt modified by NMIA are also modified by DMS, kethoxal or CMCT (Figure 3). Evidently, the small molecules give similar information.

Comparison to previous studies of structured RNA binding to microarrays

To our knowledge, this is the first application of oligonucleotide arrays to facilitate modeling of a new RNA secondary structure. Southern and coworkers used DNA microarrays to reveal changes in RNA secondary structure (12,13) and to explore factors important for oligonucleotide binding to tRNA (11). In general, the factors identified by Mir and Southern (11) are consistent with the results shown here in Figure 3. A possible exception is that they found: 'There is no high yield from heteroduplexes that would fail to incorporate the whole arm of a stem. Although heteroduplexes that extend beyond the end of a stem into a loop are formed, no significant yield is seen from those that would require the oligonucleotide to partially penetrate a second stem.' The strongly binding pentamers and hexamers found here are usually too short to incorporate the whole arm of a stem. The apparent difference in results may only reflect the definition of 'high yield'. Mir and Southern studied binding as a function of oligonucleotide length and

found that the yield decreased when length was shorter or longer than required to extend to the end of a stem. Alternatively, the apparent difference could reflect a difference between DNA and 2′-*O*-methyl probes. The DNA and 2′-*O*-methyl backbones favor B- and A-form structures, respectively, so 2′-*O*-methyl probes may provide a more regular interface when partially invading a helix.

The pentamers and hexamers used here are shorter and more isoenergetic than the heptamers (15) and nonamers (16) used previously to test this microarray method. Interpretation of results for nonamers was complex because tight binding was observed in cases where the middle nucleotide of the probe was complementary to a target nucleotide already in a Watson–Crick pair flanked by Watson–Crick pairs. Thus some nonamers had sufficient binding strength to compensate for the free energy required to open three consecutive Watson–Crick pairs in the target. Interpretation of heptamer binding to the 120 nt *E. coli* 5S rRNA was relatively straightforward, but required more approximations than for the pentamers and hexamers used here. The available results suggest that a universal microarray containing isoenergetic 5- to 7-mers of all possible sequences would facilitate rapid modeling of RNA secondary structures. This would require a total of 21 504 probes.

CONCLUSIONS

This article proposes a secondary structure for the novel R2Bm 5′ RNA on the basis of free energy minimization and results from hybridizations on isoenergetic RNA microarrays and from chemical modification. The isoenergetic library was designed on the basis of thermodynamic data detailing the effects of LNA and 2,6-diaminopurine substitutions in 2′-*O*-methyl RNA/RNA hybrids. The results clearly identify five hairpin loops and also identify regions most important for additional experiments such as site-directed mutagenesis. The results suggest that a combination of microarray binding and chemical modification experiments with free energy minimization provide a rapid way to model RNA secondary structure. Using oligonucleotides to probe RNA structure is an emerging technology and while useful, it is prudent to keep in mind that little is known about some of the factors controlling binding of oligonucleotides to folded RNA. More information will no doubt be extracted from microarray data as knowledge of these factors increases. Even with the given limitations, it is clear that an oligonucleotide microarray-based approach can facilitate extending searchable databases of genome sequences to searchable databases of RNA secondary structures.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was supported by NIH grants GM22939 (D.H.T.), 1R03 TW1068 (R.K. and D.H.T.), GM42790 (T.H.E.) and by the Polish State Committee for Scientific Research grants 2 P04A 03729 (R.K.), N N301 3383 33 (E.K.). Funding to pay the Open Access publication charges for this article was provided by NIH.

Conflict of interest statement. None declared.

REFERENCES

1. Pace,N.R., Thomas,B.C. and Woese,C.R. (1999)Probing RNA structure, function, and history by comparative analysis. In Gesteland,R.F., Cech,T.R. and Atkins,J.F. (eds), *The RNA World,* 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 113–141.
2. Cannone,J.J., Subramanian,S., Schnare,M.N., Collett,J.R., D'Souza,L.M., Du,Y., Feng,B., Lin,N., Madabusi,L.V. *et al*. (2002) The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, **3**, 2 [http://www.biomedcentral.com/1471-2105/3/2].
3. Zuker,M. (1989) On finding all suboptimal foldings of an RNA molecule. *Science*, **244**, 48–52.
4. Mathews,D.H., Disney,M.D., Childs,J.L., Schroeder,S.J., Zuker,M. and Turner,D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl Acad. Sci. USA*, **101**, 7287–7292.
5. Mathews,D.H. and Turner,D.H. (2002) Dynalign: an algorithm for finding the secondary structure common to two RNA sequences. *J. Mol. Biol.*, **317**, 191–203.
6. Chen,J.H., Le,S.Y. and Maizel,J.V. (2000) Prediction of common secondary structures of RNAs: a genetic algorithm approach. *Nucleic Acids Res.*, **28**, 991–999.
7. Lewis,J.B. and Doty,P. (1970) Derivation of the secondary structure of 5S RNA from its binding of complementary oligonucleotides. *Nature*, **225**, 510–512.
8. Uhlenbeck,O.C., Baller,J. and Doty,P. (1970) Complementary oligonucleotide binding to the anticodon loop of fMet transfer RNA. *Nature*, **225**, 508–510.
9. Uhlenbeck,O.C. (1972) Complementary oligonucleotide binding to transfer RNA. *J. Mol. Biol.*, **65**, 25–41.
10. Milner,N., Mir,K.U. and Southern,E.M. (1997) Selecting effective antisense reagents on combinatorial oligonucleotide arrays. *Nat. Biotechnol.*, **15**, 537–541.
11. Mir,K.U. and Southern,E.M. (1999) Determining the influence of structure on hybridization using oligonucleotide arrays. *Nat. Biotechnol.*, **17**, 788–792.
12. Sohail,M., Akhtar,S. and Southern,E.M. (1999) The folding of large RNAs studied by hybridization to arrays of complementary oligonucleotides. *RNA*, **5**, 646–655.
13. Ooms,M., Verhoef,K., Southern,E.M., Huthoff,H. and Berkhout,B. (2004) Probing alternative foldings of the HIV-1 leader RNA by antisense oligonucleotide scanning arrays. *Nucleic Acids Res.*, **32**, 819–827.
14. Sugimoto,N., Nakano,S., Katoh,M., Matsumura,A., Nakamuta,H., Ohmichi,T., Yoneyama,M. and Sasaki,M. (1995) Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry*, **34**, 11211–11216.
15. Kierzek,E., Kierzek,R., Turner,D.H. and Catrina,I.E. (2006) Facilitating RNA structure prediction with microarrays. *Biochemistry*, **45**, 581–593.
16. Duan,S.H., Mathews,D.H. and Turner,D.H. (2006) Interpreting oligonucleotide microarray data to determine RNA secondary structure: Application to the 3′ end of Bombyx mori R2 RNA. *Biochemistry*, **45**, 9819–9832.
17. Cummins,L.L., Owens,S.R., Risen,L.M., Lesnik,E.A., Freier,S.M., McGee,D., Guinosso,C.J. and Cook,P.D. (1995) Characterization of fully 2′-modified oligoribonucleotide hetero- and homoduplex

hybridization and nuclease sensitivity. *Nucleic Acids Res.*, **23**, 2019–2024.
18. Adamiak,D.A., Rypniewski,W.R., Milecki,J. and Adamiak,R.W. (2001) The 1.19 angstrom X-ray structure of 2′-O-Me(CGCGCG)$_2$ duplex shows dehydrated RNA with 2-methyl-2,4-pentanediol in the minor groove. *Nucleic Acids Res.*, **29**, 4144–4153.
19. Nielsen,K.E., Rasmussen,J., Kumar,R., Wengel,J., Jacobsen,J.P. and Petersen,M. (2004) NMR studies of fully modified locked nucleic acid (LNA) hybrids: solution structure of an LNA:RNA hybrid and characterization of an LNA:DNA hybrid. *Bioconjugate Chem.*, **15**, 449–457.
20. Cramer,H. and Pfleiderer,W. (2000) Nucleotides LXIV[1]: synthesis, hybridization and enzymatic degradation studies of 2′-O-methyl-oligoribonucleotides and 2′-O-methyl/deoxy gapmers. *Nucleosides Nucleotides Nucleic Acids*, **19**, 1765–1777.
21. Majlessi,M., Nelson,N.C. and Becker,M.M. (1998) Advantages of 2′-O-methyl oligoribonucleotide probes for detecting RNA targets. *Nucleic Acids Res.*, **26**, 2224–2229.
22. Sproat,B.S., Lamond,A.I., Beijer,B., Neuner,P. and Ryder,U. (1989) Highly efficient chemical synthesis of 2′-O-methyloligoribonucleotides and tetrabiotinylated derivatives – novel probes that are resistant to degradation by RNA or DNA specific nucleases. *Nucleic Acids Res.*, **17**, 3373–3386.
23. Jepsen,J.S., Sorensen,M.D. and Wengel,J. (2004) Locked nucleic acid: a potent nucleic acid analog in therapeutics and biotechnology. *Oligonucleotides*, **14**, 130–146.
24. Frieden,M., Hansen,H.F. and Koch,T. (2003) Nuclease stability of LNA oligonucleotides and LNA-DNA chimeras. *Nucleosides Nucleotides Nucleic Acids*, **22**, 1041–1043.
25. Morita,K., Hasegawa,C., Kaneko,M., Tsutsumi,S., Sone,J., Ishikawa,T., Imanishi,T. and Koizumi,M. (2002) 2′-O,4′-C-ethylene-bridged nucleic acids (ENA): highly nuclease-resistant and thermodynamically stable oligonucleotides for antisense drug. *Bioorg. Med. Chem. Lett.*, **12**, 73–76.
26. Christensen,S.M., Ye,J.Q. and Eickbush,T.H. (2006) RNA from the 5′ end of the R2 retrotransposon controls R2 protein binding to and cleavage of its DNA target site. *Proc. Natl Acad. Sci. USA*, **103**, 17602–17607.
27. Kierzek,E., Ciesielska,A., Pasternak,K., Mathews,D.H., Turner,D.H. and Kierzek,R. (2005) The influence of locked nucleic acid residues on the thermodynamic properties of 2′-O-methyl RNA/RNA heteroduplexes. *Nucleic Acids Res.*, **33**, 5082–5093.
28. Pasternak,A., Kierzek,E., Pasternak,K., Turner,D.H. and Kierzek,R. (2007) The chemical synthesis of 2′-O-methyl-2,6-diaminopurine riboside and LNA-2,6-diaminopurine riboside and their influence on the thermodynamic properties of 2′-O-methyl RNA/RNA heteroduplexes. *Nucleic Acids Res.*, **35**, 4055–4063.
29. Xia,T.B., SantaLucia,J., Burkard,M.E., Kierzek,R., Schroeder,S.J., Jiao,X.Q., Cox,C. and Turner,D.H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, **37**, 14719–14735.
30. Borer,P.N. (1975) Optical properties of nucleic acids, absorption, and circular dichroism spectra. In Fasman,G.D. (ed.), *CRC Handbook of Biochemistry and Molecular Biology: Nucleic Acids,*, Vol. 1, 3rd edn. CRC Press, Cleveland, OH, pp. 589–595.
31. Richards,E.G. (1975) Use of tables in calculation of absorption, optical rotatory dispersion, and circular dichroism of polyribonucleotides. In Fasman,G. D. (ed.), *CRC Handbook of Biochemistry and Molecular Biology: Nucleic Acids*, Vol. 1, 3rd edn. CRC Press, Cleveland, OH, pp. 596–603.
32. Afanassiev,V., Hanemann,V. and Wölfl,S. (2000) Preparation of DNA and protein micro arrays on glass slides coated with an agarose film. *Nucleic Acids Res.*, **28**, e66.
33. Ziehler,W.A. and Engelke,D.R. (2000) Probing RNA structure with chemical reagents and enzymes. *Curr. Protocol. Nucleic Acid Chem.*, **2**, 6.1.1–6.1.21.
34. Merino,E.J., Wilkinson,K.A., Coughlan,J.L. and Weeks,K.M. (2005) RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.*, **127**, 4223–4231.
35. Kierzek,E., Mathews,D.H., Ciesielska,A., Turner,D.H. and Kierzek,R. (2006) Nearest neighbor parameters for Watson-Crick complementary heteroduplexes formed between 2′-O-methyl RNA and RNA oligonucleotides. *Nucleic Acids Res.*, **34**, 3609–3614.
36. Pasternak,A., Kierzek,E., Pasternak,K., Fratczak,A., Turner,D.H. and Kierzek,R. (2007) The thermodynamics of 3′-terminal pyrene and guanosine for the design of isoenergetic 2′-O-methyl-RNA-LNA chimeric oligonucleotide probes of RNA structure. *Biochemistry,* in press (DOI: 10.1021/bi701758z).
37. Ploye,H. (1979) Endothermy and partial thermoregulation in the silkworm moth, *Bombyx mori*. *J. Comp. Physiol.*, **129**, 315–318.
38. 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.
39. Turner,D.H. (2000) Conformational changes. In Bloomfield,V.A., Crothers,D.M. and Tinoco,I. (eds), *Nucleic Acids: Structures, Properties and Functions*. University Science Books Sausalito, California, pp. 259–334.
40. Mathews,D.H. (2004) Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA*, **10**, 1178–1190.
41. Szymanski,M., Specht,T., Barciszewska,M.Z., Barciszewski,J. and Erdmann,V.A. (1998) 5S rRNA Data Bank. *Nucleic Acids Res.*, **26**, 156–159.
42. Brown,J.W. (1998) The Ribonuclease P Database. *Nucleic Acids Res.*, **26**, 351–352.
43. Sprinzl,M., Horn,C., Brown,M., Ioudovitch,A. and Steinberg,S. (1998) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.*, **26**, 148–153.
44. Walter,A.E., Turner,D.H., Kim,J., Lyttle,M.H., Muller,P., Mathews,D.H. and Zuker,M. (1994) Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl Acad. Sci. USA*, **91**, 9218–9222.
45. Mathews,D.H., Burkard,M.E., Freier,S.M., Wyatt,J.R. and Turner,D.H. (1999) Predicting oligonucleotide affinity to nucleic acid targets. *RNA*, **5**, 1458–1469.
46. Tuerk,C., Gauss,P., Thermes,C., Groebe,D.R., Gayle,M., Guild,N., Stormo,G., Daubentoncarafa,Y., Uhlenbeck,O.C. *et al.* (1988) CUUCGG hairpins – extraordinarily stable RNA secondary structures associated with various biochemical properties. *Proc. Natl Acad. Sci. USA*, **85**, 1364–1368.
47. Antao,V.P. and Tinoco,I. Jr (1992) Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucleic Acids Res.*, **20**, 819–824.
48. Allain,F.H.T. and Varani,G. (1995) Structure of the P1 helix from group I self-splicing introns. *J. Mol. Biol.*, **250**, 333–353.
49. Cornish,P.V., Stammler,S.N. and Giedroc,D.P. (2006) The global structures of a wild-type and poorly functional plant luteoviral mRNA pseudoknot are essentially identical. *RNA*, **12**, 1959–1969.
50. Ruschak,A.M., Mathews,D.H., Bibillo,A., Spinelli,S.L., Childs,J.L., Eickbush,T.H. and Turner,D.H. (2004) Secondary structure models of the 3′ untranslated regions of diverse R2 RNAs. *RNA*, **10**, 978–987.