Published online 3 April 2008

Nucleic Acids Research, 2008, Vol. 36, No. 9 ***2981–2989***
doi:10.1093/nar/gkn134

Formation of the double helix: a mutational study

Mehrdad Majlessi and Michael M. Becker*

Gen-Probe Incorporated, 10210 Genetic Center Drive, San Diego, CA 92121-4362, USA

Received January 19, 2008; Revised March 5, 2008; Accepted March 10, 2008

ABSTRACT

To investigate the mechanisms by which oligonucleotides hybridize to target molecules, the binding of two oligodeoxynucleotide probes to RNA targets was measured over a broad range of temperatures. Mutations were then scanned across each DNA/RNA hybrid to map, at single base resolution, sequences important for hybridization. Despite being unrelated in sequence, each hybrid formed by a similar mechanism. In the absence of secondary structure, two stretches of bases, termed nucleation regions, cooperated with one another by a looping mechanism to nucleate hybridization. Mutations inside each nucleation region strongly decreased hybridization rates, even at temperatures well below the melting temperature (T_m) of the hybridized duplex. Surprisingly, nucleation regions were detected in a RNA target but not a corresponding DNA target. When either nucleation region was sequestered in secondary structure, the hybridization rate fell and the mechanism of hybridization changed. Single-stranded bases within the nucleation region of the probe and target first collided to form a double helix. If sufficiently G + C rich, the double helix then propagated throughout the oligonucleotide by a strand invasion process. On the basis of these results, general mechanisms for the hybridization of oligonucleotides to complementary and mutant targets are proposed.

INTRODUCTION

Early studies of hybridization examined double helix formation of long nucleic acids over a broad range of temperatures (1). At temperatures well below the melting temperature (T_m) of the double helix, hybridization occurred by a collision mechanism characterized by a positive energy of activation. The rate-limiting step was the collision of two strands to form a small complementary double helix. Once formed, the small double helix was postulated to rapidly propagate by a zippering mechanism to complete double helix formation. In contrast, at temperatures near T_m, hybridization was shown to take place by a nucleation mechanism characterized by a negative activation energy in which two single strands were in rapid equilibrium with a small double helix. If sufficiently stable, the small duplex was postulated to rapidly propagate by a zippering mechanism to complete double helix formation. Although these studies suggested a basic mechanism for the formation of the double helix, they left many critical questions unanswered. At which point along a single strand does the initial collision or nucleation event take place? Do collision and nucleation mechanisms initiate at the same sequences? What are the base composition and length requirements for these sequences? What effects do mutations inside or outside these sequences have on association rates? How does secondary structure affect these sequences? Are there other hybridization mechanisms? Here we seek answers to these questions by examining the hybridization of two oligodeoxynucleotide probes of unrelated sequence to complementary as well as mutated targets. On the basis of our measurements, we present a general mechanism for the hybridization of oligonucleotides to complementary and mutant targets that also encompasses specific hybridization mechanisms proposed by others (2–6).

MATERIALS AND METHODS

Oligonucleotides

Oligoribonucleotide targets and deoxyoligonucleotide acridinium ester-labeled probes (7) were prepared as previously described (8). Probes P1, P3-5 and P7-11 were labeled between nucleotides 16 and 17, P2 was labeled between nucleotides 12 and 13 and P6 was labeled between nucleotides 15 and 16.

Hybridization measurements

Kinetic and thermodynamic measurements of hybridization reactions conducted for 60 min or longer were carried out as previously described using 0.5 fmol of acridinium ester-labeled probe (9). The hybridization buffer consisted

*To whom correspondence should be addressed. Tel: +1 858 731 5966; Fax: +1 858 731 5900; Email: mickb@gen-probe.com

of 0.1 M lithium succinate, 8.5% wt/vol lithium lauryl sulfate, 125 mM lithium hydroxide, 95 mM succinic acid, 1.5 mM EDTA and 1.5 mM EGTA (pH 5.1). For shorter hybridization times, a modified protocol was used. An acridinium ester-labeled probe (0.5 fmol) was dissolved in 40 µl of 2× hybridization buffer, equilibrated to the desired temperature and 40 µl of target dissolved in water was added to initiate hybridization. After the desired hybridization time, 200 µl of sulfite solution [0.014 $NaSO_3$, 0.042 borate (pH.8)] was automatically injected into the sample followed 12 s later by automated injection of 200 µl of peroxide solution (0.12% H_2O_2, 1.5 N NaOH) and the resultant chemiluminescence was measured for 2 s in a Leader I Luminometer (Gen-Probe Incorporated, San Diego, CA, USA).

T_m measurements

T_m measurements were conducted as previously described (8) in hybridization buffer using an oligonucleotide concentration of 15 µg/ml. Unless otherwise noted, all melting curves consisted of a single well-defined melting transition.

Hybridization rate constant calculations

The formation and dissociation of the double helix is described by Equation 1 where A and C are single-stranded nucleic acids and B is a double-stranded hybrid.

$$A + C \underset{k_2}{\overset{k_1}{\rightleftarrows}} B \quad \mathbf{1}$$

The rate of formation of B is given by:

$$\frac{d[B]}{dt} = k_1[A][C] - k_2[B] \quad \mathbf{2}$$

When the concentration of C does not change, hybridization is pseudo-first order and integration of Equation 2 yields:

$$1 - \frac{[A]}{[A_0]} = \left(1 - K'\right)\left(1 - e^{-(k_1 C_0 + k_2)t}\right) \quad \mathbf{3}$$

where

$$K' = \frac{k_2}{C_0 k_1 + k_2} \quad \mathbf{4}$$

and A is the free concentration of A, A_0 is the total concentration of A, k_1 is the association rate constant, k_2 is the dissociation rate constant, C_0 is the concentration of target, t is the hybridization time and $(1 - [A]/[A_0])$ is the fraction of hybrid at time t. When k_2 is negligible, Equation 3 reduces to the well-known C_0t Equation 5, which assumes that hybridization is an irreversible process:

$$1 - \frac{[A]}{[A_0]} = \left(1 - e^{-(k_1 C_0)t}\right) \quad \mathbf{5}$$

When k_2t is large Equation 3 reduces to the Langmuir expression for equilibrium binding:

$$1 - \frac{[A]}{[A_0]} = \left(1 - K'\right) = \frac{K_{eq} C_0}{K_{eq} C_0 + 1} \quad \mathbf{6}$$

where K_{eq} equals k_1/k_2. When deletions, mutations or increases in temperature significantly enhance dissociation rates, k_2t can become large and hybridization rate constants cannot be calculated using Equation 3 because it reduces to Equation 6. Attempts to calculate hybridization rates from Equation 5 under these conditions yields rate constants that increase with decreased hybridization time. Since the value of k_2 is, in general, not known, we carried out hybridization analysis at short hybridization times (3 and 8 min) to minimize the contribution of k_2 to hybridization kinetics. The dependence of $(1 - [A]/[A_0])$ on C_0 was fit by Equation 3 for both hybridization times to yield one set of values for k_1 and k_2 that satisfied both times. Although the values of k_1 at these short hybridization times are accurate, the values of k_2 are less accurate and are not reported here. As a check of the accuracy of this approach, we measured the hybridization kinetics of HivScr at 58°C over a broad range of target concentrations and times. At this temperature, HivScr exhibited one of the largest k_2 of any oligonucleotide examined here. The values of k_1 and k_2 calculated from Equation 3 at 3 and 8 min were identical to the values of k_1 and k_2 calculated at any hybridization time leading up to the attainment of equilibrium (Supplementary Data and results not shown), validating the approach.

Nuclease digestion studies

Nucleases S1 or RNAse A and V1 were used to probe for single- and double-stranded regions of RNA, respectively. Digestions were carried out at 40°C (Eco 1082) or 37°C (HivScr) in a buffer consisting of 0.1 M lithium succinate (pH 5.0) and LiCl (0.3 M); V1 digestions also contained 4 mM $MgCl_2$ to support enzymatic activity. Eco 1082 was labeled at either its 3′ or 5′ end with fluorescein. V1 digestion patterns of flourescein-labeled HivScr depended upon whether fluorescein was at the 3′ or 5′ end. Furthermore, V1 digestion patterns of 5′ labeled-HivScr changed when flourescein was replaced by TAMARA suggesting that attachment of a dye to the 5′ end of HivScr was disturbing V1 digestion patterns. To eliminate this potential artifact, two triethylene glycol linkers (Glen Research, Sterling, VA) were inserted between the 5′ end of HivScr and flourescein. The resultant V1 digestion patterns matched the V1 digestion patterns of HivScr labeled at the 3′ end with flourescein. Digestion products were resolved by HPLC using a Dionex PA-100 column and buffers A [25 mM Tris (pH 8.0) and B (25 mM Tris (pH 8.0); 0.5 M sodium perchlorate] where buffer B was varied from 5 to 50% over 50 min.

RESULTS

To probe the mechanism whereby oligonucleotides hybridize to one another, the hybridization properties of

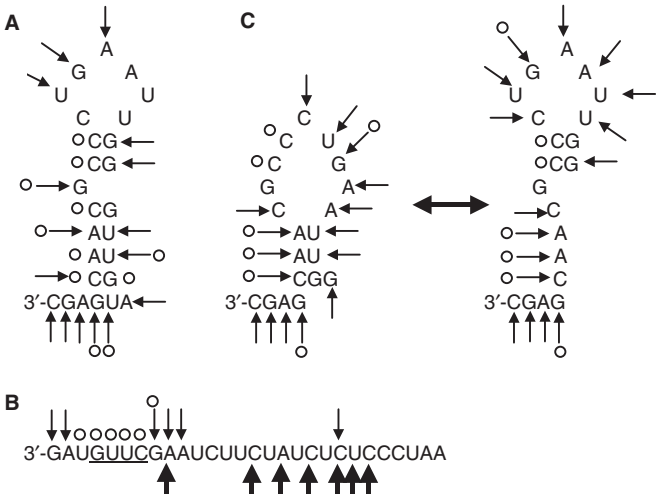

Figure 1. Structure of (**A**) Eco 1082 (**B**) HivScr and (**C**) Δ6Eco1082 as determined by nuclease digestion studies. Arrows denote S1 cleavage sites, circles denote V1 cleavage sites and bold arrows denote RNAse A cleavage sites.

a 26-base long DNA/RNA hybrid, termed Eco1082 (Figure 1A; P1 + T1; Table 1), and a 28-base long DNA/RNA hybrid, termed HivScr (P6 + T10; Table 1), were examined. To measure hybridization kinetics, oligodeoxynucleotide probes P1 or P6 were labeled with a chemiluminescent acridinium ester reporter molecule (2). Previously, we showed that the chemiluminescent properties of acridinium ester could be used to accurately measure hybridization of a probe to a target molecule without significantly affecting the hybridization process (4).

The temperature dependence of the hybridization kinetics of Eco 1082 ($T_m = 77°C$) is summarized in Figure 2. At temperatures above 67°C ($1/T = 0.00294$), hybridization rates decreased with temperature, consistent with a nucleation process. At temperatures below 67°C, hybridization rates increased with temperature rather than remained constant with temperature, as expected for a collision process. Thus, the Eco 1082 target appears to be folded at temperatures below 67°C, whereas at higher temperatures, unfolding increases its hybridization rate.

To identify sites within Eco1082 that initiated hybridization, mismatches or deletions were introduced into a complementary oligoribonucleotide target (T1) and hybridization to the acridinium ester-labeled probe (P1) was measured at 40°C. Hybridization rate constants (k_1) were determined by measuring the fraction of hybrid formed in the presence of increasing concentrations of target over a broad range of target concentration (1 to 10^5 fmol).

As summarized in Table 1, at 40°C, mutation (T2) or deletion (T3) of the 5′end but not the 3′end (T4) of the Eco1082 target decreased hybridization rates 24- to 275-fold, respectively. Sequences near the 5′end of Eco1082 may have initiated hybridization because they were located at the end of the oligonucleotide. To test this hypothesis, the first 4 bases from the 5′ end were translocated to the 3′end. At 40°C, the resultant target

Table 1. Hybridization rates [k_1(strands^{-1} s^{-1})] of Eco1082 (P1 + T1) and HIVScr (P6 + T10) and their derivatives[a]

Sequence name	Sequence	Rate constant (k_1)
P1	GCTCGTTGCGGGACTTAACCCAACAT	
T1	CGAGCAACGCCCUGAAUUGGGUUGUA	5.12×10^5
T2	CGUGCAACGCCCUGAAUUGGGUUGUA	2.13×10^4
T3	CAACGCCCUGAAUUGGGUUGUA	1.87×10^3
T4	CGAGCAACGCCCUGAAUUGGGU	5.63×10^5
P2	GTTGCGGGACTTAACCCAACATGCTC	
T5	CAACGCCCUGAAUUGGGUUGUACGAG	4.26×10^5
T6	CAACGCCCUGAAUUGGGUUGUACCAG	2.05×10^4
P3	GGTCGTTGCGGGACTTAACCCAACAT	
T7	CCAGCAACGCCCUGAAUUGGGUUGUA	3.02×10^5
P4	AATCGTTGCGGGACTTAACCCAACAT	
T8	UUAGCAACGCCCUGAAUUGGGUUGUA	5.28×10^4
P5	IITCGTTGCGGGACTTAACCCAACAT	
T9	GGAGCAACGCCCUGAAUUGGGUUGUA	6.5×10^4
P6	CTACAAGCTTAGAAGATAGAGAGGGATT	
T10	GAUGUUCGAAUCUUCUAUCUCUCCCUAA	4.63×10^6
T11	UUCGAAUCUUCUAUCUCUCCCUAA	4.76×10^6
P7	GCTCGTTGCGGGACTTAACC	
T12	CGACGAACGCCCUGAAUUGG	8×10^5
P8	GCTCGTTGCGGGACTTAAUU	
T13	CGACGAACGCCCUGAAUUGG	6×10^5
P9	GCGCGTTGCGGGACTTAAUU	
T14	CGGGCAACGCCCUGAAUUGG	1.6×10^4
P10	GCTCGTTGCGGGACTTAA	
T15	CGAGCAACGCCCUGAAUU	2.43×10^5
P11	GCGCGTTGCGGGACTTAA	
T16	CGGGCAACGCCCUGAAUU	1.08×10^4

[a]Experiments employing Eco 1082 probes (P1-P5,P7-P11) were carried out at 40° while experiments employing HIVScr probe P6 were carried out at 42°C. Probe sequences (P1-11) are highlighted in bold and all read from left to right in the 5′ to 3′ direction. Hybridization rates were calculated from Equation (3) for three and eight hybridization times and exhibited an accuracy of 15% (1 SD).

(T5) hybridized to a complementary probe (P2) only 1.2-fold slower than to the wild-type target, but 21-fold faster than a mutant target containing a mismatch in the translocated sequence (T6). Thus the 5′end of Eco1082, when translocated to the 3′end, could still initiate hybridization.

The ability of sequences at the 5′end of Eco1082, but not the 3′end, to initiate hybridization may be related to the G + C richness of the 5′end. To test this hypothesis, the first 2 bases at the 5′end of Eco1082, GC, were replaced by GG (P3 + T7), AA (P4 + T8) or II (P5 + T9; I is inosine) and hybridization rates measured to complementary targets at 40°C. Substitution of GC with GG only decreased hybridization rates 1.7-fold whereas substitution of GC with AA or II decreased hybridization rates 8- to10-fold. We conclude that the ability of the 5′end of Eco1082 to function as a nucleation region was enhanced by the presence of G + C rich bases.

To map the boundaries of the nucleation region located at the 5′end of Eco1082 at single nucleotide resolution, a mismatch scanning experiment was performed (Figure 3A). Individual base pairs along Eco1082 were replaced by GG mismatches (solid diamonds) and hybridization measured for 60 min at 40°C. Since 40°C is well below the T_m of Eco1082, hybridization rates were

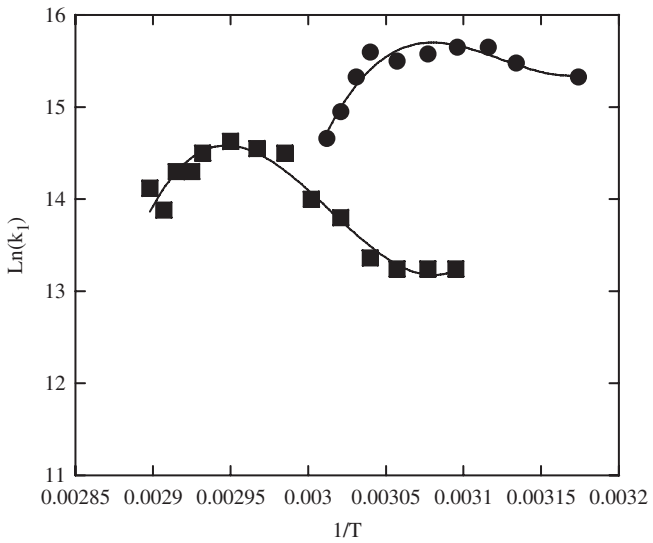

Figure 2. Dependence of the natural logarithm of the hybridization rate constant (k_1 molar concentration of strands^{-1} sec^{-1}) with the inverse of temperature (°K) for HIVScr (circles) and Eco1082 (squares). Hybridization rates were calculated by Equation (3) for hybridization times of 3 and 8 min.

first calculated from Equation 5, which assumes that during its formation the double helix does not dissociate (i.e. k_2 is negligible). At 40°C, a single nucleation region spanning nucleotides 1 to 9 was observed (Figure 3A). Within this region, GG mismatches at positions 3 and 6 weakly decreased hybridization rates. Thus, either several small nucleation regions exist between nucleotides 1through 9 or GG mismatches at positions 3 and 6 were weak mutations that did not strongly decrease hybridization rates. To distinguish between these two possibilities, AA mismatches were inserted at position 3 or 6. As summarized in Figure 3A, the AA mismatches decreased hybridization rates as strongly as did GG mismatches at adjacent positions. AA mismatches at positions 13 and 22, positions that were located outside the nucleation region, had no affect on hybridization. Thus, at 40°C, Eco1082 contained only one nucleation region spanning nucleotides 1 to 9.

To confirm that k_2 was negligible at 40°C, the hybridization kinetics of 9 GG or AA mutants within the nucleation region were reanalyzed at 3 or 8 min and hybridization rates were calculated using Equation 5. The results (Figure 3A) were indistinguishable from the results measured at 60 min, confirming that k_2 was negligible.

Figure 3B summarizes the results of a mismatch scanning experiment of Eco1082 at 61°C. GG mismatches were placed at all positions except position 13, which contained an AA mismatch as in Figure 3A, and at positions 3 and 6, which contained a GT and AA mismatch, respectively. The latter two mismatches were chosen since they decreased hybridization rates more strongly than GG mismatches (Figure 3A). Because of enhanced dissociation rates at 61°C, hybridization rates were calculated using Equation 3 for the 3-min and 8-min hybridization times. The resultant mutational profile (Figure 3B) was very similar to that detected at 40°C (Figure 3A).

The temperature dependence of the hybridization kinetics of the HivScr hybrid is summarized in Figure 2. In contrast to Eco1082, HivScr (T_m = 65°C) does not contain G + C rich sequences at either of its ends. At temperatures below 55°C ($1/T$ = 0.0305), hybridization rates of HivScr did not vary significantly with temperature, consistent with a collision mechanism. At higher temperatures, hybridization rates fell with increased temperature, consistent with a nucleation mechanism (10).

Figure 4A summarizes the results of a GG mismatch scanning experiment of HIVScr at 42°C. Hybridization rates were first calculated from Equation 5 for a hybridization time of 60 min. Mutation (Figure 4A; GG mismatches) or deletion (T11; Table 1) of the first 4 bases of HIVScr did not decrease hybridization rates. Thus, in contrast to Eco1082, HivScr did not harbor a strong nucleation region at its end. Instead, two nucleation domains spanning nucleotides 5 to12 and 19 to 27 were detected. Within the first nucleation region, a GG mismatch at position 8 only weakly decreased hybridization rates. As discussed above, this result could signify the presence of two small nucleation regions on either side of nucleotide 8 or, alternatively, that a GG mismatch at position 8 was incapable of destabilizing the nucleation region spanning nucleotides 5–12. To distinguish between these possibilities, six different mismatches were inserted at position 8 as well as at various positions inside and outside the nucleation regions of HivScr or Eco1082. When introduced at positions 6, 7 or 9 of the nucleation region of HivScr (Figure 4A), or at position 3 of the nucleation region of Eco1082 (Figure 3A), at least one of these mismatches strongly decreased hybridization rates. In contrast, when inserted at position 8 or at position 17 of HivScr, which lies outside of the nucleation region of HivScr, none of these mismatches strongly decreased hybridization rates (Figure 4A). Thus, the failure of any mismatch at position 8 to alter hybridization rates of HivScr suggests that this nucleotide does not lie within a nucleation region. However, if two independent nucleation regions lie on either side of nucleotide 8 then complete inactivation of either region by a mismatch would only reduce hybridization rates 2-fold. In contrast, mutation of either region reduced hybridization rates 4.5- to 7-fold. Thus, these two nucleation regions cooperate with one another to initiate hybridization. We conclude that a single nucleation region spans nucleotides 5 to12 in HivScr and that a number of different mismatches at position 8 can be accommodated within the nucleation region. We further note that because mutations at position 8 decreased the T_m of HivScr as strongly as mutations at position 9 (Supplementary Data and results not shown), mutational effects on hybridization rates did not correlate well with mutational effects on T_m.

To determine whether k_2 was negligible at 42°C, the hybridization kinetics of wild-type HivScr and eight different mutants scanning the two nucleation regions were re-examined at 3 and 8 min. Hybridization rates were calculated using Equation 5. For six of the mutants analyzed, the hybridization rates measured at 3, 8 or

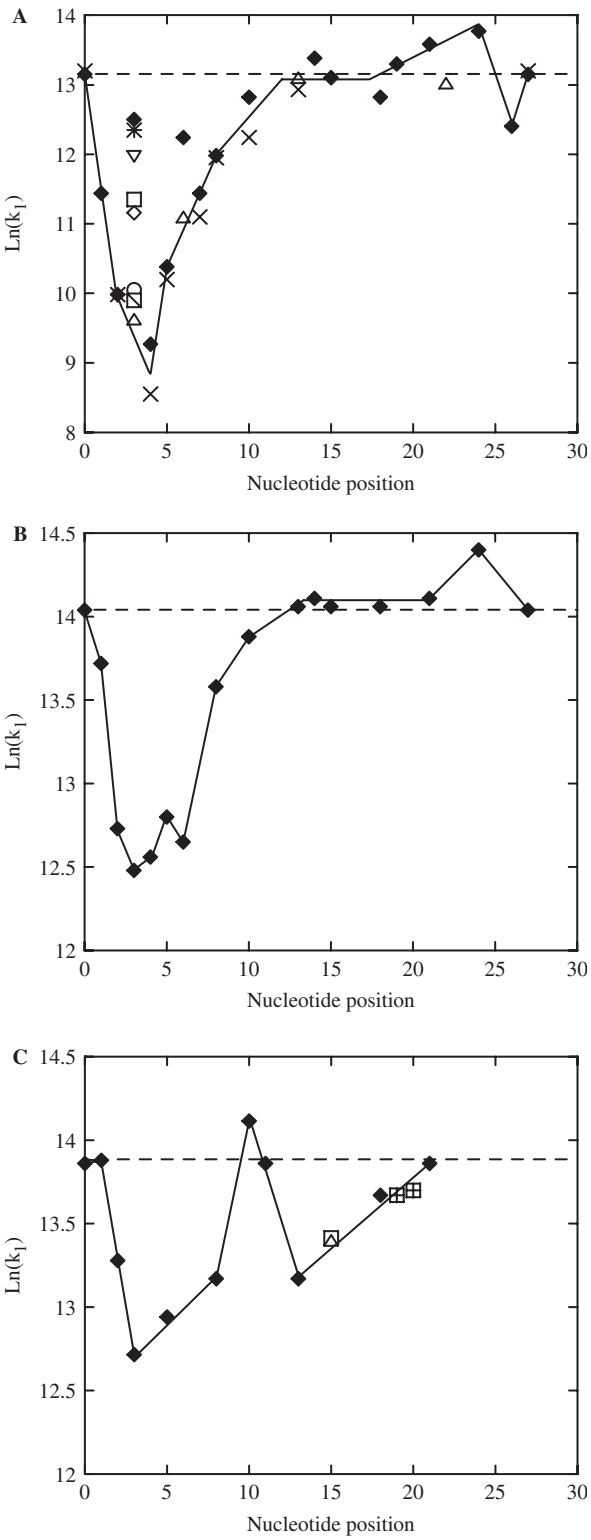

Figure 3. Mismatch scanning experiments of Eco 1082 at 40°C (**A**) or 61°C (**B**) and Δ6 Eco1082 at 40°C (**C**). The natural logarithm of the hybridization rate constant (k_1) is plotted against the nucleotide position of the mutant. Position 0 and 27 (Eco1082) or 0 and 21 (Δ6Eco1082) denote the hybridization results for the wild-type oligonucleotide. The solid line outlines the effects of the strongest mutation at each nucleotide position examined. (A) Hybridization rates calculated by Equation (5) for a 60-min hybridization time are shown for GG (filled diamond), AA (triangle), TT (circle), TG (upside down triangle), AG (square), GA (open diamond) or GT (square with diagonal) mismatches where the first nucleotide denotes the probe strand and the second nucleotide denotes the target strand. Results calculated from Equation (5) for GG mismatches (positions 2 and 3) at 3 min (cross) or for GG (positions 2–5, 7, 8, 10) and AA mismatches (positions 3and 13) at 8 min hybridization times (x) are also shown. (B) Hybridization rates were calculated by Equation (3) for 3- and 8-min hybridization times for the strongest mutants in (A). (C) Hybridization rates were calculated by Equation (3) for 3- and 8-min times for GG (filled diamond), AA (triangle), AG (square) and CC (cross hatched square) mismatches.

60 min were the same. For two mutants (positions 21 and 23), the rate constant calculated using Equation 5 increased with shorter times, indicating that k_2 was not negligible for these mutants. The hybridization rates for these mutants were therefore calculated using Equation 3 for the 3-min and 8-min hybridization times. The mutational profile summarizing these results (Figure 4B) closely resembled the mutational profile using Equation 5 (Figure 4A).

To determine the hybridization kinetics of HivScr at a higher temperature where hybridization occurred by a nucleation mechanism, wild-type and mutant HivScr targets were hybridized for 3 and 8 min at 58°C. For many of the mutants examined, the percentage of hybridized molecules did not increase with time, indicating that equilibrium had been reached. Because the fast hybridization rates of these mutants were difficult to measure, we determined their equilibrium affinities using Equation 6. The affinities of the mutants, as well as that of wild-type HivScr, were measured at longer times (60 min) to allow equilibrium to be reached. The resultant equilibrium mismatch scanning profile is shown in Figure 5A. A similar equilibrium analysis of Eco1082 at 61°C was also conducted by carrying out hybridization overnight where equilibrium was reached (Figure 5B). For both DNA/RNA hybrids, the equilibrium mismatch-scanning profiles differed markedly from the kinetic mismatch-scanning profiles (compare Figure 5A to Figure 4B and Figure 5B to Figure 3B). Furthermore, the equilibrium mismatch-scanning profiles of the two hybrids were more similar to each other than the kinetic mismatch-scanning profiles were to each other. In general, mismatches reduced equilibrium affinities more than hybridization rates (compare Figure 3B and 5A).

The mismatch scanning experiments of Eco1082 and HivScr examined the hybridization of DNA probes to RNA targets. To examine a DNA target, hybridization of the HivScr DNA probe to a DNA target was examined at 42°C by mismatch scanning analysis. In contrast to an RNA target, GG mismatches in the DNA/DNA hybrid only weakly decreased hybridization rates (1.5-fold) and no distinct nucleation regions were detected (Figure 4C).

To examine the potential role of secondary structure in the hybridization kinetics of Eco1082 and HivScr, Mfold (11) was used to identify possible regions of secondary structure within these oligonucleotides. For Eco1082, the RNA target exhibited a potential stem–loop structure, 3′-cgaGCAACgCCcugaauuGGGUUGUa-5′ (capital letters denote base-paired structures). Optical melting studies revealed a well-defined melting transition in the DNA

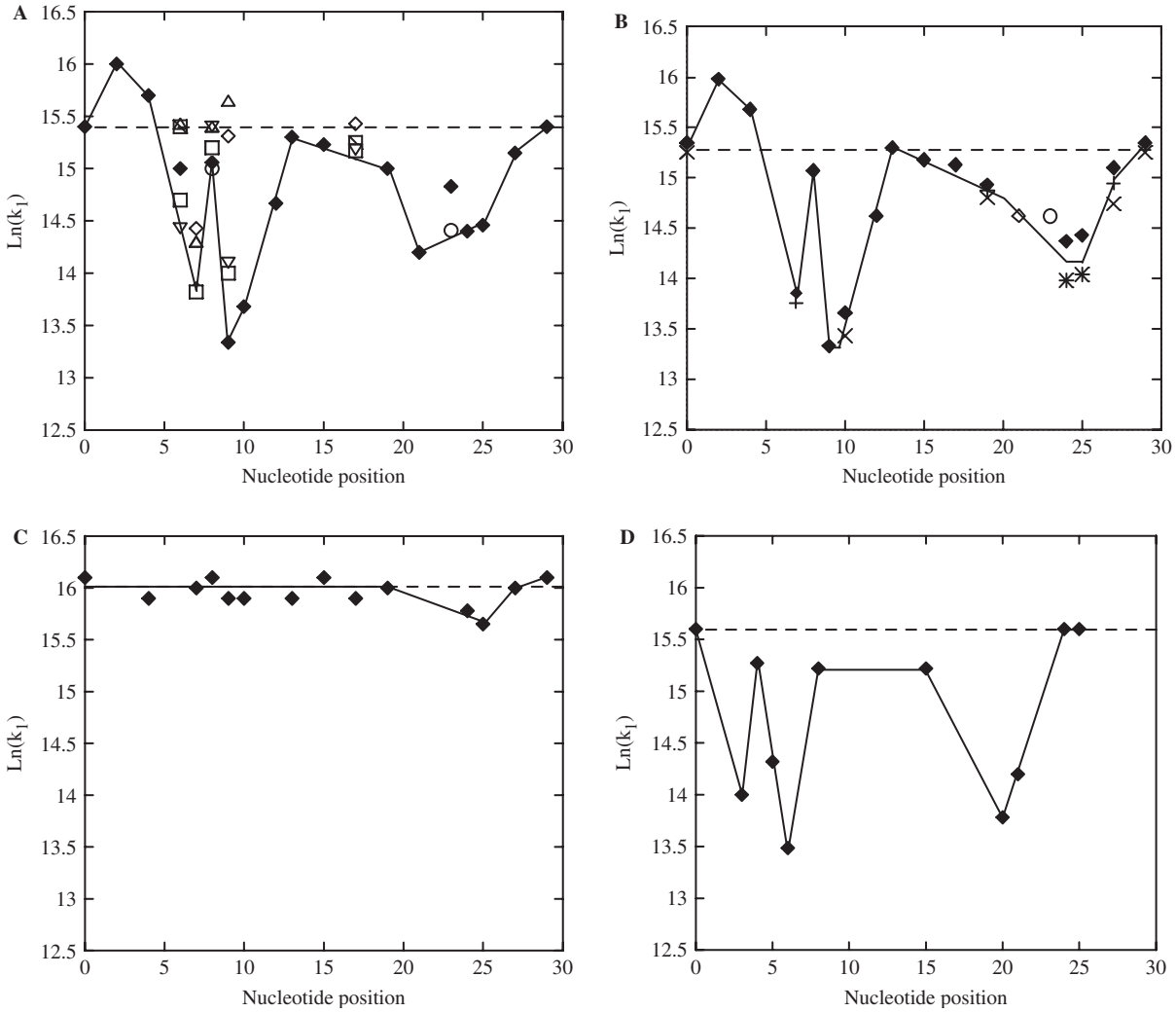

Figure 4. Mismatch scanning experiments at 42°C of HIVScr (**A–C**) and Δ4HIVScr (**D**). (A) Hybridization rates were calculated by Equation (5) for a 60-min hybridization time for GG (filled diamond), AA (triangle), TT (circle), TG (inverted triangle), AG (square), GA (open diamond) or GT (square with diagonal) mismatches. (B) Hybridization rates from (A) for GG mismatches at positions 2, 4, 7–10, 12, 13, 15, 17, 19, 24, 25, 27 are denoted as (filled diamond). Hybridization rates calculated from Equation (5) for GG mismatches at positions 7, 10, 19, 24, 25, 27 at 3 or 8 min are denoted as (cross) and (x), respectively. Results calculated by Equation (3) for 3- and 8-min hybridization times are shown for GG (open diamond) and TT (circle) mutants. (C) Mismatch scanning experiments of HIVScr where the target strand was DNA. Hybridization rates calculated by Equation (3) for 3- and 8-min hybridization times are shown for GG (filled diamond) mismatches. (D) Hybridization rates were calculated by Equation (3) for 3- and 8-min hybridization times for GG (filled diamond) mismatches.

strand of Eco1082 (T_m = 51°C) and an increase in hyperchromicity over a broad range of temperatures (40–85°C) for the RNA strand (Supplementary Data). For HivScr, a much smaller stem–loop structure, 3′-GAUGuucgaaucuucUAUCucucccuaa-5′, was identified in the RNA target. Optical melting studies of HivScr revealed well-defined melting transitions [temperatures below our hybridization experiments (results not shown)] in the DNA strand (T_m = 26°C) and the RNA strand (T_m = 33°C).

To confirm the potential secondary structure in the Eco 1082 and HivScr RNA targets, nuclease digestion studies using enzymes specific for single-stranded RNA (S1 and RNAse A) or double-stranded RNA (V1) were carried out. Digestions were done in the same buffer used to carry out hybridization rate measurements except that LiCl was substituted for lithium lauryl sulfate to prevent inhibition of nuclease activity.

As summarized in Figure 1A, at 40°C, Eco 1082 adopted a stem–loop structure that closely matched that predicted by computer folding algorithms. The single-stranded regions were only digested by S1 and not V1, and all V1 cuts only occurred in the stem region. The appearance of V1 as well as S1 cuts within the stem suggests that it is capable of breathing.

At 37°C, HivScr adopted a single-stranded structure with a small hairpin loop at one end (Figure 1B). The hairpin loop (3′-UGUUCG-5′) contains a stable tetraloop, 3′-GUUC-5′, previously identified by others (12) flanked by a GU base pair. The tetraloop was strongly digested by V1 while the GU flanking base pair was weakly digested by V1.

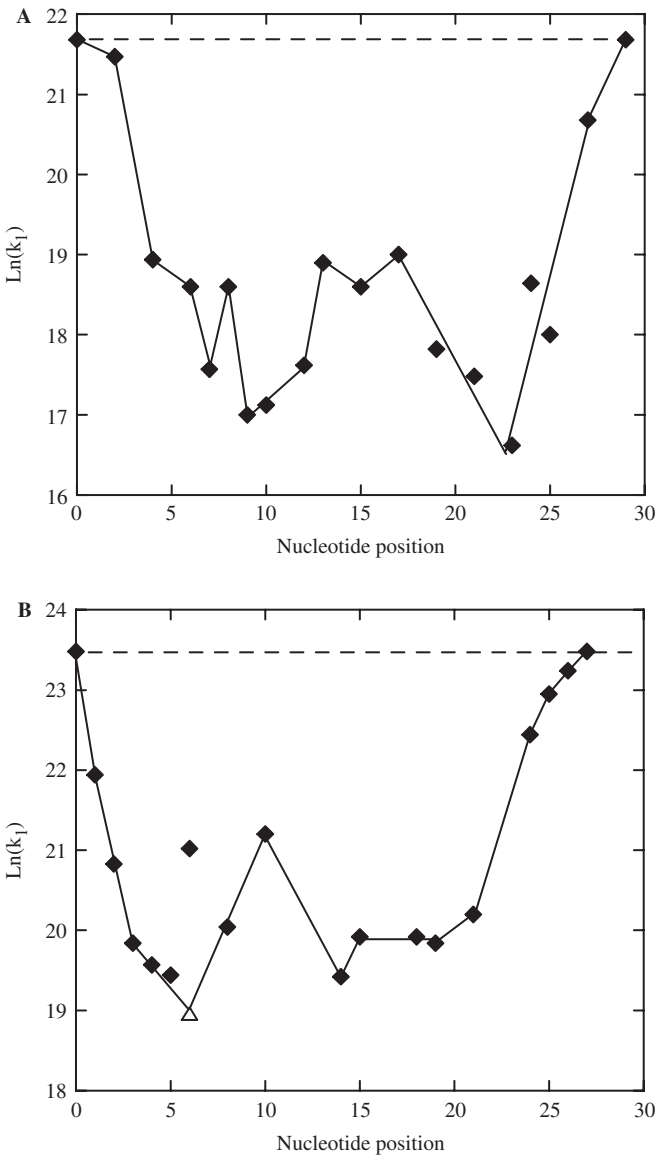

Figure 5. Mismatch scanning experiments of HIVScr (**A**) and Eco1082 (**B**). Equilibrium affinities (molar concentration of strands^{-1}) for GG mismatches (filled diamond) or an AA mismatch (triangle) were calculated at 58°C after 60 min hybridization (A) or after overnight hybridization (B).

To further address the potential role of stem–loop structure in the hybridization kinetics of HivScr, the first four bases of both the DNA and RNA strands of HivScr were deleted to eliminate the small hairpin loop structure. The resultant mutant, Δ4 HivScr (P6 + T11), hybridized at 42°C at the same rate as HivScr (Figure 4D). As shown in Figure 4D, a mismatch scanning profile of Δ4HivScr was nearly identical to HivScr (Figure 4B) except that it was shifted 4 bases by the deletion. We conclude, therefore, that secondary structure is not responsible for the presence of the two nucleation regions within HivScr.

To address the potential role of secondary structure in the hybridization kinetics of Eco1082, the last six bases of the DNA and RNA strand of Eco1082 were deleted to yield a DNA/RNA hybrid with little potential secondary structure (RNA target: cgagcaacgCCcugaauuGG). Nuclease digestion studies of the resultant mutant, Δ6Eco 1082 (P7 + T12), argue that 6Eco1082 exists as a mixture of two hairpin structures each containing a very small stem (Figure 1C).

At 40°C, the Δ6Eco1082 probe hybridized to the deleted RNA target 1.6-fold faster than full-length Eco1082 probe hybridized to its full-length target (Table 1). In contrast, when the Δ6Eco1082 probe was hybridized to full-length Eco1082 RNA target (P7 + T1), the hybridization rate fell 4-fold (data not shown). Thus formation of a stem–loop structure in the full-length Eco1082 target decreased hybridization rates of the Δ6Eco1082 probe. As shown in Figure 3C, the mismatch scanning profile of Δ6Eco1082 differed from wild-type Eco1082 (Figure 3A). In contrast to the one strong nucleation region in Eco1082, Δ6 Eco1082 exhibited two weaker nucleation regions like HivScr and Δ4HivScr (Figure 4B and D).

The finding of two nucleation regions within 082 exhibited two weaker Δ6Eco1082 suggests that, as observed in HivScr, both regions cooperate with one another to nucleate hybridization. To test this hypothesis, the RNA target was not altered, but the last 2 CC bases in the Δ6 Eco1082 probe were changed to UU such that the last 2 bp of the resultant DNA/RNA hybrid were weak GU rather than strong GC base pairs. The resultant DNA/RNA hybrid, UUΔ6Eco1082 (P8 + T13), hybridized 1.3-fold more slowly than the wild-type hybrid (P7 + T12) at 40°C (Table 1). More importantly, a GG mismatch at position 3 in UUΔ6Eco1082 (P9 + T14) decreased hybridization rates 38-fold (Table 1) whereas the same mutation in Δ6Eco1082 only decreased hybridization rates 3.3-fold (Figure 3C). We conclude that weakening the last 2 GC bases of Δ6Eco1082 greatly impairs the ability of the first nucleation region to initiate hybridization and thus both nucleation regions cooperate to nucleate hybridization. In further support of this hypothesis, the last 2 bases of Δ6Eco1082 were deleted. The resultant mutant, Δ8Eco1082 (P10 + T15), hybridized 2.5-fold slower than UUΔ6Eco1082 and a GG mismatch at position 3 (P11 + T16) decreased hybridization rates of Δ8Eco1082 22-fold (Table 1).

DISCUSSION

The mechanisms used to form a double helix in Eco1082 and HivScr depended upon whether the RNA target contained secondary structure or not. Nuclease digestion studies clearly revealed that the Eco 1082 RNA target adopts a stem-loop structure throughout most of its length. This secondary structure is present at 40°C and melts out at higher temperatures, explaining why Eco1082's hybridization rate increased with increasing temperature (Figure 2). At 40°C, the Eco1082 target exhibited one nucleation region. The majority of bases within the single nucleation are base-paired, leaving only 3 bases as a single-stranded tail in the DNA probe and/or RNA target. Consistent with this expectation, deletion of

the first 4 bases strongly decreased hybridization rates at 40°C. Transposition of these bases to the other end of Eco 1082, to recreate a new single-stranded tail, restored hybridization. The finding that decreases in the G + C content of these bases strongly decreased hybridization rates is also consistent with the proposed secondary structure of Eco1082 since A + T bases will greatly decrease the stability of a small duplex formed at the single-stranded tail (13). When hybridization was carried out at a higher temperature (61°C), the single nucleation region persisted but mutations within it were much less deleterious than they were at lower temperatures. This change in the importance of the nucleation region can be explained by weakening of the stem–loop structure at the higher temperature, which otherwise competes with base pairing within the nucleation region and decreases the probability of productive nucleation.

Once Eco1082 nucleates at the dangling single-stranded tail how does hybridization proceed? Since a small duplex formed at the tail will be unstable, and because the remaining probe or target is folded, the probe must strand-invade the double-stranded region adjacent to the tail and continue invasion until the final duplex is formed. The alternative mechanism in which Eco1082 exists in a dynamic equilibrium between unfolded and folded conformations with hybridization taking place from the unfolded conformer is unlikely since lowering the temperature from 61 to 40°C should strongly disfavor the unfolded conformer, greatly decreasing hybridization rates in contrast to our experimental results (compare Figures 3A and 3B). This two-step hybridization mechanism proposed for Eco 1082 has also been proposed for the binding of antisense RNAs to their RNA targets (2,3). Both cases examined helix formation initiated at accessible single-stranded bases that propagated by a strand-invasion process to complete duplex formation.

In contrast to Eco1082, the RNA target of HivScr was largely single-stranded. Although a small hairpin was detected at the 3′ end of HivScr, its removal had little impact on hybridization rates and no effect on the location of nucleation regions. Whereas Eco1082 exhibited one nucleation region, HivScr exhibited two. Both nucleation regions were required for hybridization since mutation of either nucleation region alone strongly decreased hybridization rates. If either nucleation region alone could nucleate hybridization, inactivation of either region would only decrease hybridization rates 2-fold in contrast to our results.

How do two nucleation regions initiate hybridization to the HivScr target? We propose that the HivScr probe collides randomly over the length of the unstructured target. When collision takes place at one of the two nucleation regions, the resultant small duplex is sufficiently stable to allow the probe to loop over and collide with the second nucleation region. The two regions together are then sufficiently stable to allow propagation of the rest of the duplex by a zippering mechanism. Both nucleation regions must cooperate with one another and, thus, inactivation of either one reduces hybridization >2-fold. The failure of Eco1082 to exhibit a second nucleation region can be attributed to the secondary structure adjacent to the first nucleation region which will inhibit looping. Weakening of this secondary structure should allow looping to take place in Eco1082. Consistent with this model, removal of 6 bases from the 5′ end of the Eco1082 RNA, which eliminated most of the secondary structure target, resulted in the appearance of two nucleation regions in Eco1082 that cooperated with one another to nucleate hybridization. The weakness of the second nucleation region in Δ6Eco1082, coupled with the finding that its hybridization rate was only 1.6-fold higher than Eco1082, suggests that the secondary structure overlapping the second nucleation region (Figure 1C) competes with looping. A looping mechanism also appears to be involved in the interconversion of hairpin structures (6) as well as in the formation of RNA duplexes that proceed by way of a kissing complex formed between two RNA loops (4,5). Formation of a small duplex between two RNA loops was proposed to facilitate base pairing at a second site by a proximity effect, which then propagates by a strand invasion process to complete duplex formation.

The four nucleation regions identified in Δ6Eco1082 and HivScr bear no obvious sequence relationship to one another and their location cannot be predicted by nearest-neighbor thermodynamic parameters (14). The sequence of these regions is clearly important since mutations within these sequences can strongly decrease hybridization rates. We hypothesize that sequences that nucleate hybridization do so because they are sufficiently stable to form a small nucleating duplex and because they are flexible enough to rapidly undergo the conformational changes required to form this helix. Because different sequences will undergo these conformational changes more readily than others, different sequences will initiate hybridization at different rates. For two identical sequences, the probability of adopting a particular conformation will depend upon flanking sequences (15). Thus, the ability of a sequence to adopt a conformation necessary to initiate hybridization will depend upon its position within an oligonucleotide in accord with our experimental observations. For the DNA/RNA hybrids examined here, single base mismatches within their nucleation regions strongly decreased hybridization rates even though the hybridization temperature was well below the T_m of the hybridized double helix. The effect of a mismatch on hybridization rates, however, varied greatly between different mismatches.

The presence of two nucleation regions in an oligonucleotide that cooperate with one another by looping gives rise to unusual effects of mutations that extend over long distances. For example, weakening or deletion of the last 2 bases of one nucleation region in the ΔEco1082 probe greatly accentuated the effect of a single mismatch located 16 bases away in the second nucleation region.

In contrast to a RNA target, when a corresponding DNA target was hybridized to the DNA strand of HivScr, no distinct nucleation regions were detected. This observation reveals that many sequences in the DNA target are capable of nucleating hybridization and hence mutation of any given base has little effect on hybridization rates. We speculate that the stability of short nucleation regions in

RNA/DNA hybrids may be more sequence dependent than corresponding nucleation regions in DNA/DNA hybrids because of enhanced interstrand stacking in A versus B conformations (16).

In summary, for oligonucleotides, the following picture of hybridization emerges from our work. Hybridization is initiated by random collisions of a probe to single-stranded regions of a target. In most cases, the resultant complex is unstable and dissociates. However, for certain sequences, the resultant complex is sufficiently stable to nucleate the hybridization process. The small duplex formed at this stage is unlikely to be a classic Watson–Crick double helix because its sensitivity to mismatches differs markedly from a Watson–Crick helix; indeed, nucleation regions may not be helical. For example, many different mismatches within one of the nucleation regions of HivScr (position 8, Fig. 4A) only weakly decreased hybridization rates. Although previous measurements hypothesized that nucleation was rapidly followed by propagation of the duplex by a zippering mechanism, our measurements reveal a more complex process. For unstructured targets, once the first nucleation region is formed, the remaining single-stranded probe loops over and collides with a second region of the target to form a second nucleation region. The two regions together are sufficiently stable to allow propagation of the rest of the duplex by a zippering mechanism. Both nucleation regions cooperate with one another and thus inactivation of either one strongly reduces hybridization. The failure of a single nucleation region to complete duplex formation by itself may be due to the fact that base pairs in the nucleation regions have a higher probability of dissociating than forming an additional base pair. As the length of the probe and target get longer, the number of favorable nucleation regions within each will increase and thus the effect of mutations on any given nucleation region will fall.

For structured probes or targets, hybridization initiates at single-stranded bases within the first nucleation region. If looping of the probe to a second nucleation region is blocked by secondary structure duplex formation at the first nucleation region continues by a strand-invasion mechanism to complete hybrid formation. Stacking interactions between the duplex of the nucleation region and adjacent base pairs in the folded probe or target strand may facilitate the strand invasion process (17).

The finding that mutations within nucleation regions can strongly decrease hybridization rates even at temperatures well below the T_m of a duplex has important consequences for optimizing hybridization specificity. To discriminate against a mutation, a probe can be positioned along a target such that the mutation overlaps an initiation region in the probe, thus, inhibiting hybridization. In contrast, positioning a probe along a target such that the mutation does not overlap a nucleation region allows the mutation to be tolerated. In contrast to their effect on association rates, all bases except those at the very end of Eco1082 or HIVScr, when mutated, strongly decreased hybridization rates when hybridization was allowed to reach equilibrium. Thus, the only strong mutations that could be tolerated at equilibrium were those close to the end of the oligonucleotide.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We gratefully acknowledge the technical assistance of Abhijit Mazumder, Cathy Nguyen and James Rader. Funding to pay the Open Access publication charges for this article was provided by Gen-Probe Incorporated.

Conflict of interest statement. None declared.

REFERENCES

1. Cantor,C.R. and Schimmel,P.R. (1980) *Biophysical Chemistry Part III: The Behavior of Biological Molecules*. W. H. Freeman and Company, NY.
2. Kittle,J.D., Simons,R.W., Lee,J. and Kleckner,N. (1989) Insertion Sequence IS10 anti-sense pairing initiates by an interaction between the 5' end of the target RNA and a loop in the antisense RNA. *J. Mol. Biol.*, **210**, 561–572.
3. Thisted,T., Sorensen,N.S., Wagner,E.G.H. and Gerdes,K. (1994) Mechanism of post-segregational killing: Sok antisense RNA interacts with Hok mRNA via its 5′-end single-stranded leader and competes with the 3′-end of Hok mRNA for binding to the mok translational initiation region. *EMBO J.*, **13**, 1960–1968.
4. Tomizawa,J. (1984) Control of ColE1 Plasmid replication: the process of binding of RNA I to the primer transcript. *Cell*, **38**, 861–870.
5. Persson,C., Wagner,E.G.H. and Nordstrom,K. (1990) Control of replication of plasmid R1: structures and sequences of the antisense RNA, CopA, required for its binding to the target RNA, Copt. *EMBO J.*, **9**, 3767–3775.
6. Wenter,P., Furtig,B., Hainard,A., Schwalbe,H. and Pitsch,S. (2005) Kinetics of photoinduced RNA folding by real-time NMR spectroscopy. *Angew. Chem. Int. Ed.*, **44**, 2600–2603.
7. Becker,M., Lerum,V., Dickson,S., Nelson,N.C. and Matsuda,E. (1999) The double helix is dehydrated: evidence from the hydrolysis of acridinium ester-labeled probes. *Biochemistry*, **38**, 5603–5611.
8. 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.
9. Mazumder,A., Majlessi,M. and Becker,M.M. (1998) A high throughput method to investigate oligodeoxynucleotide hybridization kinetics and thermodynamics. *Nucleic Acids Res.*, **26**, 1996–2000.
10. Wetmur,J.G. and Davidson,N. (1968) Kinetics of renaturation of DNA. *J. Mol. Biol.*, **31**, 349–370.
11. Zucker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
12. Jucker,F.M. and Pardi,A. (1995) Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif. *Biochemistry*, **44**, 14416–14427.
13. Nelson,J.W. and Tinoco,I. (1982) Comparison of the kinetics of ribooligonucleotide, deoxyoligonucleotide and hybrid oligonucleotide double-strand formation by temperature-jump kinetics. *Biochemistry*, **21**, 5289–5295.
14. Sugimoto,N., Nakano,S., Katoh,M., Matsumura,A., Nakamuta,H., Ohmichi,T., Yoneyama,M. and Saski,M. (1995) Thermodynamic parameters to predict stability of RNA/DNA duplexes. *Biochemistry*, **34**, 11211–11216.
15. Chuprina,V.P., Lipanov,A.A., Fedoroff,O.Y., Kim,S., Kintanar,A. and Reid,B. (1991) Sequence effects on local DNA topology. *Proc. Natl Acad. Sci. USA*, **88**, 9087–9091.
16. Westhof,E. and Fritsch,V. (2000) RNA folding: beyond Watson-Crick pairs. *Structure*, **8**, R55–R65.
17. Mir,K.U. and Southern,E.M. (1999) Determining the influence of structure on hybridization using oligonucleotides arrays. *Nat. Biotechnol.*, **8**, 788–792.