R-loop stability as a function of RNA structure and size

Ralf Landgraf, Chi-hong B. Chen and David S. Sigman*

Department of Biological Chemistry, School of Medicine and Department of Chemistry and Biochemistry, Molecular Biology Institute, University of California–Los Angeles, Los Angeles, CA 90024-1570, USA

Received May 8, 1995; Revised and Accepted July 25, 1995

ABSTRACT

The sequence-specific formation of R-loops can be assayed using RNAs which overlap a *HindIII* cleavage site in a 3.5 kb plasmid. Chemical modification of the displaced DNA strand has permitted stabilization of these R-loops and allowed a systematic investigation of the dependence of these triple-stranded structures on the chain length and structure of the input RNA. RNAs as short as 50 nt form stable R-loops if 5-allylamine uridines (U^{aa}-RNA) are used in place of normal uridines; normal RNAs must be 100 nt long to form R-loops quantitatively. Since acetic anhydride decreases the hybridization efficiency of U^{aa}-RNAs, the positive charge of the RNAs must diminish the electrostatic repulsion of the three negatively charged phosphodiester backbones. The dependence of R-loop stability on the length of RNA can be simulated with a random walk model, which also applies to strand migration within Holiday junctions. R-loop hybridization provides a versatile method to generate single-stranded DNA in a sequence-selective manner.

INTRODUCTION

R-loops are triple-stranded structures formed between RNA and duplex DNA in which the RNA strand displaces the DNA strand of identical sequence. These structures rely exclusively on Watson–Crick base-pairing for sequence recognition. *In vitro* they form in 70% formamide at neutral pH at a temperature of ~55°C, although the precise conditions depend on the G–C content (1,2). Upon dilution of the formamide the RNA–DNA heteroduplex is not stable with respect to the DNA–DNA duplex, but the rate of displacement of RNA by DNA is sufficiently slow to allow isolation of the R-loop by chromatography or electrophoresis.

Initial studies on R-loops focused on visualization of the genetic organization of the 18S and 25S ribosomal RNA genes of yeast via electron microscopy (3). The study of R-loop hybridization via electron microscopy was in fact central to the discovery of splicing in the 5'-terminus of adenovirus 2 late mRNA (4,5). A subsequent application of R-loop hybridization was in the isolation of recombinant DNA fragments which gave rise to moderately abundant mRNAs (6).

In the course of developing these applications several important features of R-loop formation were determined. One of the parameters which influences R-loop formation is the DNA–DNA strand separation temperature (T_{ss}). T_{ss} is the temperature at which 50% of the strands separate. This temperature is different from the T_m , which is the temperature at which 50% of bases are no longer base paired in a particular sequence yet remain in a double-stranded structure (2). A sharp temperature dependence has been observed for the formation of R-loops consisting of 2900 nt yeast 26S rRNA and its complementary DNA (T_{ss} 49°C). The $t_{1/2}$ for R-loop formation at 43°C was 4 h. The $t_{1/2}$ decreased dramatically to about 30 min as the incubation temperature was increased to 48°C, only 1°C less than the T_{ss} . Although the rate of formation of R-loops is linearly dependent on RNA concentration near the T_{ss} , the rate does not directly depend on RNA concentration at 43°C, suggesting that the nucleation of heteroduplexes is not rate limiting at temperatures 6°C lower than the T_{ss} .

Two procedures have been used to increase the yield of R-loop formation *in vitro* (1). The first is chemical modification of the displaced strand with glyoxal. This blocks displacement of the RNA by the single-stranded DNA formed in the R-loop. The second approach is to randomly cross-link the target DNA (every ~3 kb) with trioxsalen, 4,5',9-trimethylpsoralen. This proves to be especially useful if nicked high molecular weight DNAs are used, because the high formamide concentration during R-loop formation will otherwise facilitate the undesired complete strand separation of the DNA substrate.

R-loop formation provides an intriguing method for sequence-specific formation of single-stranded DNA within an extensive region of Watson–Crick B-DNA. Potential applications would include the synthesis of substrates to explore mechanisms of transcription elongation (7) and DNA repair (8), as well as the development of strategies for sequence-specific DNA scission which rely on the hydrolysis of DNA by single strand-specific nucleases. Earlier work in our laboratory has demonstrated that RNAs of ~100 nt can direct the sequence-specific scission of duplex DNA via an R-loop intermediate if the RNA is derivatized with 1,10-phenanthroline-copper (9). The yield of this scission reaction ranged from 5 to 10%, but it was not clear whether this was limited by the stability of the R-loops or the chemical nuclease activity. In addition, since R-loops are transient steady-state intermediates in the enzymatic synthesis of RNA by RNA polymerase, understanding the factors which contribute to R-loop

*To whom correspondence should be addressed

stability would provide insight into this fundamental biochemical mechanism (10,11).

While previous studies successfully outlined some of the general features of R-loops, no data is available with respect to the minimum size of RNA probes that can be used to generate R-loops at high yield. This minimum size and methods which allow a further decrease of this RNA length would be of importance to applications that rely on this targeting scheme for directed scission or DNA modification. In this communication we examine R-loops prepared with RNAs from ~30–300 nt. We have found that normal RNAs must be ~100 nt long to form stable R-loops quantitatively. R-loops prepared with RNAs containing 5-allylamine uridine (U^{aa}) instead of uridine form more stable R-loops than normal RNAs of identical sequence and length. Using U^{aa} -containing RNAs (U^{aa} -RNA), in conjunction with chemical modification of the displaced strand with glyoxal, quantitative R-loop recovery is possible even if only 45 nt RNAs are used. The length dependence for R-loop formation can be understood in terms of a random walk model which has been previously applied to the spontaneous strand migration of Holiday junctions (12).

MATERIALS AND METHODS

In vitro transcription and RNA purification

RNA for the plasmid-based test system was generated as described earlier (13) using the Promega Riboprobe System II. pGEM-820 (pGEM-2; Promega) containing the 5'-region of the *Escherichia coli lacZ* gene from -19 to +63 downstream of the T7 promoter was linearized with *EcoRI* and RNA was generated from the T7 promoter. The complementary strand can be generated from the SP6 promoter after linearization with an appropriate restriction enzyme. 5-Allylamine-UTP (U^{aa} ATP) was substituted for UTP in the reaction mixture where applicable. The RNA was purified by DNase I treatment and ethanol precipitation. RNA for the scission of pETenHB (a pET-3A-based vector; 14) was also generated from the T7 promoter after linearization with *Bgl*II and was processed in the same way.

R-loop formation

The final volume for R-loop formation was 50 μ l and the reaction contained 1 pmol linearized plasmid, RNA (1–10 pmol), 70% deionized formamide, 83 mM PIPES, pH 7.8, 33 mM NaCl and 10 mM EDTA. The reaction was heated to 55°C in a thermocycler and the temperature was lowered to 45°C over a period of 4 h. The R-loops formed in this way were stable in 70% formamide at room temperature or 4°C for several days. Formamide was redistilled nucleic acid grade from BRL. Aliquots were deionized over BioRad mixed bed resin AG 501-X8(D) and stored at -20°C.

*Hind*III protection assay

For protection assays of non-glyoxal-treated R-loops, 5 μ l of the 50 μ l R-loop reaction was removed and diluted with *Hind*III reaction buffer (50 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$) to a final volume of 45 μ l. *Hind*III (5 μ l, 50 U) was added and restriction digestion carried out at room temperature for 1 h. Glyoxal-treated R-loops were directly dissolved in *Hind*III buffer after ethanol precipitation and were treated with 50 U *Hind*III in 50 μ l for 1 h.

Figure 1. Schematic of the arrangement of restriction sites on the two plasmid test system used to analyze R-loop formation and R-loop-directed scission. The 2.9 kb pGEM-820 plasmid is linearized with *Bgl*II for use as a target for R-loop formation and with *EcoRI* for use as a template for T7-based transcription. R-loop formation destroys the *Hind*III site located inside the R-loop, 16 bp from the T7 starting point of transcription, which is the basis of the R-loop assay. In the same way pGEM-4 is used to generate different length RNAs to determine the length dependency of R-loop formation.

Glyoxal fixation

The R-loop reaction was cooled to 4°C and a 1/7 vol. 40% deionized glyoxal (final concentration 1 M) was added. The reaction was incubated for 2 h at 12°C and formamide and glyoxal were removed by ethanol precipitation. Special care was taken to assure that the reaction mixture did not warm up in the presence of glyoxal. Aliquots of glyoxal were deionized over BioRad mixed bed resin AG 501-X8(D) and stored at -20°C.

Random walk simulation

Random walks were simulated on a PC using a program written in Quick Basic (Microsoft), which is available upon request. One hundred R-loops were represented through an appropriately dimensioned array. Any element in this array has two position markers assigned, representing the position of the DNA-RNA junction moving in from either side of the R-loop. Random movement was simulated by independently increasing both position values either one integer up or down. The direction of movement depends on whether a random number between 0 and 100 fell above or below a pre-set threshold. This threshold is referred to as exchange bias and represents the percent probability for the exchange of RNA by DNA. For R-loops at the starting position (no RNA exchanged) only displacement of RNA is scored. Elements for which the sum of both position markers reached the length of the RNA were removed from the simulation and scored as complete exchange. The distribution of all position markers throughout the population was monitored after every step to obtain a representation of the equilibrium established at the flanks of the R-loop.

Figure 2. Titration curve of the RNA:DNA ratio for the formation of R-loops. (A) *Hind*III protection assay after R-loop formation with *Bgl*II-linearized pGEM-820 and 111 nt U^{aa}-RNA derived from the T7 promoter and *Eco*RI-digested template. R-loops were formed at increasing ratios of RNA to DNA. Lane 1, no RNA; lane 2, 1:10; lane 3, 1:4; lane 4, 1:1; lane 5, 2.5:1; lane 6, 10:1; lane 7, 1 kb marker. (B) Graphic representation of the results of the protection assays shown in (A). Values were obtained by scanning the negative of the above image and intensities were corrected for differences in the size of the fragments.

RESULTS

Assay systems for R-loop formation

Two plasmid test systems were used to analyze the requirements for R-loop formation (Fig. 1). The first system, pGEM-820, is a derivative of pGEM-2 in which the sequence from -19 to +63 of the *lac* operon was cloned downstream of the T7 promoter. The second system is based on insert-free pGEM-4 and contains a similar promoter arrangement but allows the generation of RNAs of different length using the polylinker region of this vector. The RNA transcribed from the T7 promoter of either system overlaps a *Hind*III restriction site by 16 nt. Since *Hind*III digestion is blocked when an R-loop is formed at its cleavage site, the percentage of target DNA containing stable R-loops can be quantified by the extent of enzyme inhibition.

Stoichiometry of R-loop formation

The RNA concentration dependence of R-loop formation at a fixed percentage of formamide (70%) and constant buffer conditions has

Figure 3. Relative stability of R-loops with regular RNA and U^{aa}-RNA. *Hind*III protection assay on 0.3 μ g pGEM-820 after R-loop formation with a 10-fold excess of a 111 nt RNA, derived from T7 and terminated at the *Eco*RI site. Lane 1, no RNA; lane 2, regular RNA; lane 3, U^{aa}-RNA.

been measured at different U^{aa}-RNA:DNA ratios using the *Hind*III protection assay (Fig. 2A). U^{aa}-RNA was used in our initial experiments because its competence to form R-loops was established by its ability to direct the sequence-specific scission of DNA when coupled to the chemical nuclease 1,10-phenanthroline-copper (9). The linear dependence of R-loop formation on the concentration of input RNA in the presence of a fixed concentration of 2×10^{-8} M DNA (Fig. 2B) indicates that the dissociation constant must be at least one order of magnitude lower. The abrupt end-point of maximal R-loop formation is consistent with this high affinity binding and demonstrates that R-loop formation is a stringent method of hybridization.

Stability of R-loops

A salient feature of the titration curve of Figure 2B is that the maximum R-loop yield does not exceed 73%, even at a 10-fold excess of RNA over DNA. Since the R-loop reactions are subjected to a 1:10 dilution in order to carry out the *Hind*III digestion, the less than quantitative protection from enzymatic digestion probably reflects dissociation of the R-loops during enzymatic digestion. Although large R-loops have sufficient kinetic stability to be isolated by electrophoretic or chromatographic techniques (6,13), the yield obtained with R-loops generated with short RNAs is likely to be reduced relative to their initial concentrations in 70% formamide.

A comparison of the stability of R-loops formed with U^{aa}-RNA and regular RNA shows a significantly higher yield of R-loop recovery for the modified RNA in *Hind*III protection assays (Fig. 3).

Enhanced stability of R-loops formed with U^{aa}-RNA

Since R-loop recovery is a function of the relative stability of the DNA-DNA and RNA-DNA duplex structures at low concentrations of formamide, any modification of a purine or pyrimidine leading to a more stable heteroduplex should result in an increased lifetime of the R-loops. The *Hind*III digest confirms that U^{aa}-RNA forms more stable R-loops than regular RNA. There are two possible reasons for this increased stability. The primary amine introduced through U^{aa} could stabilize the RNA-DNA duplex by placing a positive charge into the major groove. A second possible reason could be that formation of a more extended conjugated π electron system between the uridine and the allyl spacer of U^{aa} could improve base stacking interaction. Similar effects have been

reported for 5-propyne-UTP and 5-propyne-CTP (15). In the case of U^{aa}-RNA the charge effect seems to be dominant. Exhaustive modification with acetic anhydride causes a loss of the increased stability. The stability of R-loops formed with normal RNA treated in this manner with acetic anhydride is unchanged.

Length dependency

The applicability of R-loop formation as a method to target DNA modification and scission reagents to specific sequences raises the question of the minimum length of RNA required for efficient R-loop formation and subsequent scission. To address this question R-loop formation using different length RNAs was monitored through a *Hind*III protection assay. RNAs were used in a 10-fold molar excess over the linearized test plasmid and consisted either of regular nucleotides or had uridine substituted by U^{aa}. All RNAs (except for the 111 nt probe generated and tested using pGEM-820) were derived from and targeted to pGEM-4, which contains a polylinker region that facilitates the generation of RNAs of different length. Figure 4A shows the result of a *Hind*III protection assay after 1:10 dilution of the R-loop reaction. A strong length dependency is observed for both types of RNA. The increased stability resulting from the use of U^{aa}-RNA is reflected in the ability to achieve efficient R-loop formation with shorter RNAs. While ~10% of the R-loops formed with a 45 nt RNA are stable if U^{aa}-RNA is used, no stable R-loops were detected after a 1 h incubation with *Hind*III using regular RNA. Maximum protection is achieved with RNAs of ~100 nt. Since the titration (Fig. 2) already indicated that R-loop formation is probably quantitative, methods to stabilize the R-loops prior to removal or dilution of the formamide were sought.

Glyoxal modification

This instability of R-loops had previously been observed by electron microscopy for R-loops formed with ribosomal RNA (1). R-loop stabilization was achieved by modification of the displaced DNA strand with glyoxal, a single strand-specific reagent which modifies single-stranded guanines in such a way that no Watson-Crick DNA-DNA base pairing can occur after removal of formamide (16). This effectively eliminates competition by the displaced strand and thereby prevents strand migration. Glyoxal is added to the R-loop reaction mixture in the cold after R-loops are formed and is removed after 2 h by either dialysis or ethanol precipitation. The effect of glyoxal fixation on R-loop stability is shown in Figure 4B. Apart from the fixation step, the underlying experiment is the same as for Figure 4A.

Since re-hybridization of the DNA strands is blocked by glyoxal modification of the displaced strand, the yields of R-loops reflect the factors which govern the formation of R-loops and not those which contribute to their dissociation at reduced concentrations of formamide. The increase in R-loop yield from 73 to 100% after glyoxal stabilization of hybridization mixtures containing greater than a 1:1 stoichiometry of RNA to DNA indicates that dissociation upon dilution was responsible for the incomplete yield. The experiments further demonstrate that the increased yield of R-loop formation observed with U^{aa}-RNA reflects the greater kinetic stability of these heteroduplexes relative to those formed with normal RNAs. For U^{aa}-RNA quantitative R-loop formation can be achieved with RNAs between 45 and 66 nt, while 66–111 nt are required for normal RNAs. However, between 27

Figure 4. Dependence of the R-loop yield (represented as percent protection against *Hind*III digestion) on RNA length and glyoxal modification. (A) *Hind*III protection assay of R-loops formed using different length RNAs and linearized pGEM-4 plasmid without glyoxal fixation. (B) *Hind*III protection assay after R-loop formation and glyoxal fixation. RNAs are identical to (A) and are derived from and targeted to pGEM-4.

and 45 nt a very strong length dependency for the formation of R-loops with U^{aa}-RNA is observed. Minor variations in these values will depend on the specific base composition of the RNAs, since the shorter RNAs will have a non-random nucleotide composition.

Model for R-loop formation

Based on the above results and previously published information on larger R-loops (1,2,6) a model for R-loop formation with small RNA probes has been developed which shows that the rates and yield of R-loops can be described as a function of two thermal parameters, the strand separation temperature for the full-length DNA target (T_{ss}) and the local strand separation temperature (T_{ss}^{loc}) (2). These parameters are distinct from the melting temperature, T_m , which is the temperature at which 50% of the base pairs are no longer Watson-Crick base paired. T_{ss}^{loc} is the temperature at which the target sequence of DNA melts and becomes accessible to hybridization by RNA. It can never exceed

Figure 5. Model of the different stages in R-loop formation. Stage II represents the starting point of the temperature gradient. This temperature permits local but not global strand separation. The gradient used in our experiments was linear over 4 h from 55 to 45°C. T , actual temperature; T_{ss}^{loc} , local strand separation temperature at which strand separation of the target region occurs.

T_{ss} , the dissociation temperature of the target (essentially irreversible for large DNAs) or the strand separation temperature of the RNA–DNA heteroduplex (T_{ss} -RNA). Of the three temperatures described above, only T_{ss} is readily experimentally accessible, because it can be assayed by electrophoresis. Therefore, the standard temperature profile for R-loop formation uses a temperature gradient that starts just below T_{ss} (for the pGEM plasmid 55°C) and slowly drops to 45°C over a period of 4 h (6). Since 45°C is at the low end of the scale for T_{ss} values, it is likely to cover most naturally occurring values for T_{ss}^{loc} . After execution of the annealing profile R-loops are stored at 4°C.

The model presented in Figure 5 represents different stages in the continuous temperature gradient between 55 and 45°C characterized by the value of the experimental temperature (T) relative to the key parameters T_{ss}^{loc} , T_{ss} and T_{ss} -RNA. The critical stages in this model are stages III–V.

At stage III T is above T_{ss}^{loc} , but below T_{ss} -RNA. The window of opportunity for R-loop formation between those two temperatures reflects the higher thermodynamic stability of heteroduplexes

Figure 6. Illustration of the random walk model for RNA displacement showing a simplified model of the structure of a three strand junction during branch migration. D, DNA; R, RNA (DNA in a Holiday junction); T, template strand.

to regular DNA duplexes in 70% formamide. The best condition for R-loop formation would be a high T_{ss} -RNA relative to a low T_{ss}^{loc} . A higher G+C content may raise T_{ss} -RNA, but it would also increase T_{ss}^{loc} relative to T_{ss} and enhance the danger of complete DNA strand separation. Base-modified RNAs (e.g. U^{aa} -RNAs) could give higher values for T_{ss} -RNA without changing T_{ss}^{loc} and provide a wider temperature range for successful RNA hybridization.

At stage IV the RNA is properly annealed, but the surrounding DNA is still partially separated (T is just above T_{ss}^{loc}). Under these conditions the heteroduplex is stable.

At stage V the temperature has dropped just below T_{ss}^{loc} . DNA and RNA exchange is possible in stages V and VI, but their rates are higher at the elevated temperature. The thermodynamic stability of the heteroduplexes does not exclude an exchange of RNA by DNA under these conditions, but does define the conditions for the exchange at the borders of the heteroduplex. Although accurate knowledge of T_{ss}^{loc} would theoretically allow the development of an optimized individualized annealing profile, experimental determination of T_{ss}^{loc} is not feasible. Once the temperature has fallen below T_{ss}^{loc} RNA that has been displaced completely by DNA is no longer in dynamic exchange and will be unable to re-associate to its hybridization site because of the kinetic stability of the duplex DNA. The model therefore can explain the experimental observation that the dependence of R-loop formation on the length of RNA will not be a function of RNA concentration.

Model for random walk simulation

In order to verify the above model we compared the empirical data presented in Figure 4B with a random walk simulation for the exchange of RNA by DNA which has previously been successfully applied to the exchange of DNA by DNA in a Holiday junction (12). The model is based on the migrating junction between two competing strands of DNA or RNA which are complementary to the same template (Fig. 6). Assuming the central unpaired region remains constant during migration, base pairing of the DNA at position 1 has to be accompanied by a dissociation of the heteroduplex at position 6. Conversely, extension of the heteroduplex to position 5 requires dissociation of the DNA at position -1. Because of the uncertainty of the

structure of the branch at the terminus of the RNA, the initiation of strand migration is not considered in this computer model.

Results of the random walk simulation

To simulate the net movement of the branch point of a heteroduplex in which there is biased displacement a program was written which generates a population of 100 R-loops. The branch point in each element of this population can either move forwards (i.e. toward RNA displacement) or backwards (resulting in DNA displacement) at any given step in the simulation. Since RNA-DNA heteroduplexes are substantially more stable than DNA-DNA duplexes, at each step there will be bias for the displacement of DNA by RNA. By definition, biases $<50\%$ favor the displacement of DNA by RNA. A bias of $>50\%$ would lead to the displacement of RNA by DNA. A bias of 50% indicates that the two types of base pairs are equivalently stable (as in a Holiday junction). For R-loops at the starting point only forward movement is allowed and R-loops that have completed the displacement (i.e. elements of the population where the branch point has covered the entire length of the RNA) are considered dissociated, removed from the population and scored as complete. Due to the symmetric nature of the R-loop, exchange occurs from both sides of the loop simultaneously. Our test system uses relatively long sequences with the simplifying assumption that sequence-dependent differences in the composition of both flanks can be ignored. For short RNAs or sequences with pronounced A-T- or G-C-rich sequences this simplification will obviously result in deviations from the model.

Figure 7A shows the percentage of R-loops in a given population in which a 100 nt RNA has been completely displaced as a function of the number of steps carried out in the random walk. The examples show the effect of variations of the bias for forward over backward movement. The exchange of the 100 nt RNA at a 49% bias indicates that displacement of the 100 nt RNA by DNA is still possible under these conditions, but requires a significantly larger number of steps. A better understanding of the relationship between the bias of the exchange reaction and the stability of RNAs of different length can be gained from Figure 7B. Rather than scoring the number of R-loops in the population which completed exchange, this graph shows the distribution of branch points within an entire population of 200 nt R-loops. After a given number of steps the position of the branch point of every individual element of the population was scored and plotted against the length of the RNA. This procedure was carried out at different numbers of steps. The distribution soon reaches an equilibrium that is defined by the exchange bias. For a 50% bias an even distribution over the entire length of the RNA is soon found. However, for biases $<50\%$ (i.e. DNA displacement is favored over RNA displacement) an area develops in the center of the RNA which is practically free of branch points. The consequence for a 200 nt RNA at a 47% exchange bias is a core region of 120 nt that is practically stable, while ~40 nt at each flank are in an exchange equilibrium. Simulations for larger RNAs give essentially the same distribution at the flanks, but an increased stable core region where the probability of exchange approaches zero.

Correlation of the random walk simulation with experimental results

An RNA would become unstable if this core region drops below the minimum length required for stable hybridization. Figure 7C gives another representation of the occurrence of the stable core of

Figure 7. Results of the random walk simulation of branch migration. (A) Percentage of 100 nt RNAs that have been displaced from the R-loop (out of a population of 100) as a function of the number of steps in a random walk simulation. The simulation assumes exchange from both flanks of the R-loop and was carried out at three different biases for the forward exchange (values $>50\%$ indicate that the displacement of RNA by DNA is favored). (B) Distribution of branch points in a population of R-loops along a 200 nt RNA at various values for the exchange bias. The population was scored five times after 100 000 steps and average percentages are presented. (C) Comparison of the empirical R-loop recovery (exp.) after glyoxal treatment as presented in Figure 4B and stability estimates based on (B) as a function of the RNA length and combined nucleotide distances from the termini respectively. Experimental data are presented as percent R-loops determined by restriction digestion protection. For simulation data the area under the curve in (B) is presented normalized for maximum 100%. The data are shown uncorrected for 47% exchange bias and corrected (displaced by 20 nt) for 39 and 47% bias.

the heteroduplex as a function of RNA length and exchange bias. The area under the curve in Figure 7B can be taken as an indicator of the extent to which an RNA undergoes exchange up to a specific point from the outer flanks inward. This area is plotted as a function of the combined nucleotide distances from both termini. In the case of the above example for 47% exchange bias nucleotides are being displaced up to ~40 nt on both sides. As a result, only 120 nt of a 200 nt RNA can be considered to be in a stable core heteroduplex 100% of the time. Additional nucleotides will contribute to the formation of a stable core region. The resulting plot for a 47% exchange bias is presented, together with the empirical length dependency shown in Figure 4B.

The shapes of these plots are identical, but the simulated curve reaches its maximum at the nucleotide length at which a stable core region starts to develop. The minimum length of this central core region, necessary for stable RNA hybridization, cannot be determined from the simulation data. However, it can be measured as the intercept on the x -axis in the plot of experimental R-loop stability versus RNA length (Fig. 4) and is ~20 nt. By displacing the computed curve by this amount to higher values for the RNA length (Fig. 7C) the simulation can be brought into good agreement with the experimental data. The behavior of a 60 nt RNA as predicted by the random walk model will in fact match that of an RNA of 80 nt. An unexpected result of the simulation is that despite the greater stability of the heteroduplexes formed with U^{aa}-RNA relative to normal RNA, reflected by the 47% versus 39% exchange bias, a similar minimum length for the core region is estimated. The shape of the plot is affected by the percent bias, but not the minimal size needed for R-loop formation. The explanation for this observation is that the greater stability has a cumulative effect in a random walk mechanism due to the large number of steps required (see Fig. 7A), but only a minor effect on thermodynamic properties. An 8% difference in the off rate would only translate to the difference in stability of a 20mer versus an 18mer, which would not be differentiable in this analysis.

For the model to describe the length dependency for the formation of R-loops (Fig. 7B) the exchange equilibrium has to be established rapidly relative to the temperature profile used for formation. In the simulation the equilibrium was reached after ~10 000 steps. Under aqueous conditions the step time for Holiday junctions has been estimated to be of the order of milliseconds in the absence of bivalent cations (12). Our own data confirm that similar values are found for the displacement of RNA by DNA in aqueous buffer. Formamide, the absence of magnesium ions and elevated temperature are likely to ensure a high rate of exchange, so that an equilibrium will be established within several seconds during stage V of R-loop formation (Fig. 5). This demonstrates two points: (i) the length dependency observed with both types of RNA can be explained by RNA displacement through branch migration; (ii) Figure 5 outlines a reasonable model for the reactions that occur during the temperature gradient used to form the R-loops.

DISCUSSION

The formation of an R-loop at the recognition site of a restriction enzyme blocks hydrolysis. The extent of inhibition upon R-loop formation therefore provides a rapid and convenient method for determining the fraction of a target DNA converted into an R-loop. In this study this assay was carried out in two ways. In one set of experiments R-loops were formed between the plasmid DNA and RNA in 70% formamide and then diluted 1:10 in order

to provide solvent conditions appropriate for *Hind*III catalysis. The linearity of inhibition by RNA of enzymatic hydrolysis of linearized plasmid DNA (Fig. 2B) indicates that the stoichiometry of R-loop formation is 1:1 mol RNA:DNA and that the dissociation constant of the R-loops must be less than 1 nM in 70% formamide. However, despite this high affinity binding complete protection of the target DNA from scission was not observed; ~25% of the DNA remained susceptible to *Hind*III digestion within 1 h of digestion at room temperature. The probable reason for this observation is that upon dilution of formamide and over the time course of the assay the RNA is displaced from the R-loop in the area covering the *Hind*III site. This instability, which reflects the lability of RNA-DNA heteroduplexes relative to DNA-DNA duplexes at low formamide concentrations, had previously been noted as a shortening of R-loops in electron microscopy in the 2–3 kb size range and is responsible for the marginal stability of R-loops formed with non-base-modified 50 nt RNAs.

The strand exchange reaction which causes displacement of RNA from R-loops at low formamide concentrations can be completely eliminated if the displaced DNA strand is derivatized with glyoxal prior to formamide removal. A dramatic increase in the extent of inhibition of *Hind*III digestion is apparent. A 50 nt RNA, which gave only marginal protection against *Hind*III digestion without glyoxal modification, causes 80% inhibition with glyoxal modification. Moreover, complete inhibition of *Hind*III digestion is observed with RNAs as long as 100 nt if glyoxal modification is carried out; there is no dissociation of R-loops upon dilution of the formamide. Using glyoxal-modified DNA it is possible to focus exclusively on the relative stability of R-loops formed with RNAs of different length and composition in 70% formamide.

In addition to providing a method for the sequence-selective preparation of single-stranded DNA within normal B-DNA, glyoxal modification has revealed interesting structural insights associated with the formation of R-loops. The first is that abiological modification of bases can enhance the stability of heteroduplexes. For example, RNAs prepared with U^{aa}ATP are more stable than normal RNAs of identical sequence. This is reflected in the lower rate of dissociation of R-loops in low formamide. It is also evident in comparing the lengths of RNA essential for formation of stable R-loops in 70% formamide. Quantitative formation of R-loops is achieved with 50 nt U^{aa}-RNAs, but 100 nt are essential for normal RNAs. The source of the higher stability found for U^{aa}-substituted RNA appears to be predominantly the positive charge provided by the protonated primary amine, because modification of this abiological RNA with acetic anhydride abolishes its enhanced binding affinity. The protonated allylamine would be located in the major groove and is likely to diminish electrostatic repulsion between the negatively charged phosphodiester backbones. Based on canonical B and A form DNA this effect would be of special importance in a heteroduplex, where the phosphate backbone is brought into even closer proximity compared with the B-helix. The advantage of using substituted nucleotides to increase the yields of R-loops with shorter RNAs is that R-loops are stabilized without raising the temperature required to initially melt the target region. The use of RNAs with a higher G+C ratio does not share this advantage, because the thermal stability of the R-loops and the target strand will increase in parallel. The use of 5-propyne-cytosine and -uridine might allow the use of RNAs short enough to be efficiently and economically generated by solid phase synthesis (15).

By stabilizing R-loops with glyoxal experimental results and theoretical simulations of R-loop formation in 70% formamide can be compared. These studies have demonstrated that the stable central core of the R-loop can only be formed with RNAs >20 nt. The length of the heteroduplex with an R-loop is a function of the relative bias for RNA versus DNA exchange in a random walk mechanism. This bias is in turn a reflection of the relative stability of the two duplexes and is therefore influenced by any modification that exclusively increases the heteroduplex stability. Our results provide an explanation for the observation that above a 1:1 ratio of RNA to DNA the yield of R-loops depends on RNA length but not RNA concentration. The underlying idea is that an equilibrium of strand migration reactions is established rapidly at the end of the heteroduplex. The position of this equilibrium determines the minimum length for an RNA that can maintain an R-loop despite exchange reactions on both flanks, but this equilibrium will not depend on the concentration of the RNA in solution.

In summary, quantitative formation of R-loops can be achieved with 50 nt U^{aa}-RNAs or 100 nt normal RNAs. These R-loops can be stabilized by blocking the re-annealing of the displaced DNA strand through chemical modification of the bases. Sequence-selective formation of single-stranded DNAs is therefore possible. One application of these results, discussed in the accompanying communication, is that it has permitted the development of a method for scission of the DNA sequence matching the RNA initially used to form the R-loop.

ACKNOWLEDGEMENTS

This research was supported by the Human Genome Project grant

HG00255-01 and the award of the Merck Academic Development Program Predoctoral Fellowship CO92591 to RL.

REFERENCES

- 1 Kaback,D.B., Angerer,L.M. and Davidson,N. (1979) *Nucleic Acids Res.*, **6**, 2499–2517.
- 2 Thomas,M., White,R.L. and Davis,R.W. (1976) *Proc. Natl. Acad. Sci. USA*, **73**, 2294–2298.
- 3 Woolford,J.L., Jr and Rosbash,M. (1979) *Nucleic Acids Res.*, **6**, 2483–2497.
- 4 Chow,L.T., Gelinas,R.E., Broker,T.R. and Roberts,R.J. (1977) *Cell*, **12**, 1–8.
- 5 Berget,S.M., Moore,C. and Sharp,P.A. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 3171–3175.
- 6 Rosbash,M., Blank,D., Fahmer,K., Hereford,L., Ricciardi,R., Roberts,B., Ruby,S. and Woolford,J. (1979) *Methods Enzymol.*, **68**, 454–469.
- 7 Nudler,E., Goldfarb,A. and Kashlev,M. (1994) *Science*, **265**, 793–796.
- 8 Selby,C.P. and Sancar,A. (1993) *Science*, **260**, 53–58.
- 9 Sigman,D.S., Chen,C.-h.B. and Gorin,M.B. (1993) *Nature*, **363**, 474–475.
- 10 Tantin,D. and Carey,M. (1994) *J. Biol. Chem.*, **269**, 17397–17400.
- 11 Daube,S.S. and von Hippel,P.H. (1994) *Biochemistry*, **33**, 340–347.
- 12 Panyutin,I.G. and Hsieh,P. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 2021–2025.
- 13 Chen,C.-h.B., Gorin,M.B. and Sigman,D.S. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 4206–4210.
- 14 Poole,S.J., Kauvar,L.M., Drees,B. and Kornberg,T. (1985) *Cell*, **40**, 37–43.
- 15 Wagner,R.W., Mateucci,M.D., Lewis,J.G., Gutierrez,A.J., Moulds,C. and Froehler,B.C. (1993) *Science*, **260**, 1510–1513.
- 16 Broude,N.E. and Budowsky,E.I. (1971) *Biochim. Biophys. Acta*, **254**, 380–388.