Double stranded scission of DNA directed through sequence-specific R-loop formation

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

R-loop formation with short (100 nt) RNAs provides a highly flexible and stringent method to achieve sequence-specific separation of target DNA at any given sequence. After stabilization of R-loops with glyoxal and removal of the RNA through RNase treatment the remaining single-stranded DNA bubble provides a highly favorable substrate for attenuated micrococcal nuclease. We investigated this method for sequence-specific scission of double-stranded DNA and achieved quantitative scission of 3–5 kb plasmids. The applicability to larger size DNA is demonstrated through specific excision of the intervening segment between two R-loops from a P1 plasmid of ~120 kb.

INTRODUCTION

R-loops are triple-stranded structures containing 1 mol RNA and 1 mol duplex DNA in which the RNA strand forms a heteroduplex with its complement and displaces the DNA strand of identical sequence. These structures are stable in 70% formamide and rely on Watson–Crick base pairing for sequence recognition (1,2). Upon removal of the formamide the displaced DNA strand re-anneals to its DNA complement and promotes dissociation of the RNA from the heteroduplex.

Earlier studies have focused on R-loops prepared with RNAs of ~2–3 kb (3,4). In the accompanying communication we have analyzed the formation of R-loops prepared with RNAs <300 nt long. These studies have led to the conclusion that RNAs as short as 50 nt can form R-loops quantitatively in 70% formamide. Although short RNAs are displaced from these R-loops upon dilution of formamide, chemical modification of the displaced DNA single strand with a base-reactive reagent such as glyoxal blocks re-annealing of the duplex and ensures that the DNA remains separated (1). This stabilized single-stranded DNA, which would have the same sequence as the RNA used to form the R-loop, can now serve as a substrate for further biochemical manipulations.

The formation of R-loops with short RNA probes has several potential applications. In this communication we develop a method for directing the double-stranded scission of DNA at any chosen sequence within plasmids. The protocol, summarized in Figure 1, consists of the following steps: (i) sequence-specific

R-loop formation; (ii) chemical modification of the displaced single strand of DNA with base-specific modification reagents to stabilize the R-loop and block renaturation of the DNA; (iii) hydrolysis of the RNA used for R-loop formation to render both DNA strands sensitive to hydrolysis by single strand-specific nucleases; (iv) activation by calcium ions of a high K_m mutant of micrococcal nuclease [MN(a)] to cleave the single-stranded bubble formed by R-loop formation; (v) analysis of the reaction products. The underlying chemistry of this new approach has been developed using 3–5 kb plasmids. It has been implemented in the analysis of the exon/intron structure of human L-aspartate (D-aspartate) O-methyltransferase (10,11) cloned into a 120 kb P1 plasmid.

A flexible method of DNA scission should find use in the determination of distances between two sequences in a genome, analysis of the exon/intron structure of a transcription unit and in the molecular definition of chromosomal translocations. The use of R-loops to direct the sequence-specific scission of DNA by RNAs derivatized with the chemical nuclease phenanthroline-copper has previously been explored (8,9). Although the specificity of scission was high, the yield of directed cleavage was too low to be practically useful. Single strand scission approached 70%, but double strand cutting did not exceed 10%.

MATERIALS AND METHODS

Construction of the attenuated micrococcal nuclease

The attenuated mutant (L37A, Y113A) of micrococcal nuclease [MN(a)] was generated in two steps by PCR mutagenesis from pONF-1 (12), a generous gift of Prof. John Gerlt (University of Maryland), as described earlier (7). Following expression the protein was purified on a phosphocellulose column and its attenuated phenotype was established by comparison with the wild-type protein. A fusion protein between GST and MN(a) was prepared by PCR amplification and in-frame ligation into the *EcoRI* site of pGEX-3x, as described earlier (7).

MN(a) mutants were expressed in *Escherichia coli* DH5 α cells by induction with 0.5 mM isopropylthio- β -D-galactoside (IPTG) at 0.6 OD₆₀₀ for 1.5 h. The cell pellet was resuspended in 10 ml 50 mM HEPES pH 6.8, and lysed by sonification. The low salt supernatant was applied to a MonoS cation exchange column and was eluted with a linear 100 ml gradient from 100 to 500 mM KCl,

*To whom correspondence should be addressed

50 mM HEPES pH 6.8, at 0.5 ml/min. Both proteins were >95% homogeneous as judged by SDS-PAGE (yield 40–50 mg/l cells). Batch purification of the GST-MN(a) fusion protein can also be accomplished using glutathione-Sepharose 4B (Pharmacia) (13).

In vitro transcription and RNA purification

RNA was synthesized as previously described from the plasmid-based test system pGEM-820, which contains the 5' region of the *E. coli lacZ* gene from –19 to +63 downstream of the T7 promoter (8). 5-Allylamine UTP (U^{aa} TP) 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 (8). The reaction mixture was heated to 55°C in a thermocycler and the temperature lowered to 45°C over a period of 4 h. The R-loops formed this way were stable in 70% formamide at room temperature or 4°C for several days.

*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 water to a final volume of 40 μ l. Five microliters of 10 $\times$ *Hind*III buffer (500 mM Tris-HCl pH 8.0, 100 mM MgCl₂) and 5 μ l (50 U) *Hind*III were added and the restriction digestion was carried out at room temperature for 1 h. Glyoxal-treated R-loops were directly dissolved in *Hind*III buffer after ethanol precipitation and were digested with 50 U *Hind*III in 50 μ l for 1 h.

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 removed by ethanol precipitation. Special care was taken to assure that the reaction mixture did not warm up in the presence of glyoxal.

Scission of R-loops

For non-glyoxal-derivatized R-loops 5 μ l of the R-loop reaction (containing ~0.2 μ g plasmid) were removed and diluted to a final volume of 50 μ l containing 6 mM CaCl₂ and 0.5 μ g (0.12 μ M) GST-MN(a) or the same concentration (0.1 μ g) of MN(a). After 1 h at room temperature the reaction was stopped by addition of EGTA to a final concentration of 10 mM. The products were separated on an 0.8% agarose gel and visualized by ethidium bromide staining. Cutting yields were calculated based on scanning results of the negatives obtained from the above gels. Absorbance values were corrected for the difference in length of the observed fragments.

Glyoxal-treated R-loops were resuspended after ethanol precipitation in 100 μ l 50 mM Tris-HCl pH 7.8, 6 mM CaCl₂. Aliquots of 20 μ l were treated for 30 min with 0.05 mg/ml RNase A at

Figure 1. Schematic of the plasmid-based test system used to analyze R-loop formation and R-loop-directed scission. The 2.9 kb plasmid is linearized with *Bgl*II for use as a target for R-loop formation and with *Eco*RI for use as a template for T7-based transcription. R-loop formation destroys the *Hind*III site located inside the R-loop, 10 bp from the T7 starting point of transcription, which is the basis of the R-loop assay. *Eco*RI/*Hind*III-digested plasmids are used as approximate markers for scission at either end of the R-loop. Sequence dictated scission involves steps (i)–(iv) as outlined in the Abstract and Discussion. Protection against restriction digestion with *Hind*III after step (ii) allows quantification of R-loop formation.

37°C. Scission of the RNA-free plasmid and analysis of the products was carried out as described above.

Radiolabeled R-loops were obtained by PCR amplification of a 311 bp region of pGEM-820 containing the central 111 bp target region located between the T7 promoter and the *Eco*RI site (see Fig. 3B). Both strands were labeled separately by using appropriately 5'-labeled primers. The R-loops were formed and stabilized with glyoxal as described above. Scission (without RNase treatment) was carried out in the presence of non-radiolabeled plasmid at a concentration that was identical to a normal plasmid scission experiment.

RESULTS

R-loop formation

The map of plasmid pGEM-820 (15), used in the majority of the experiments described here, is presented in Figure 1. Since a *Hind*III restriction site is located within the transcribed region, which can be used as the target for R-loop formation, the formation of these loops can be studied independently, because they block *Hind*III scission (5). Products of *Hind*III digestion also calibrate the size of the products of R-loop-directed scission.

R-loop formation, hybridization of RNAs to DNA in 70% formamide, is a powerful and flexible technique to target any sequence of DNA with a specificity determined by the length of the RNA used in the hybridization reaction. In the accompanying publication (5) we have demonstrated that normal RNAs as short as 100 nt and 5-allylamine U-RNA (U^{aa}-RNA) as short as 50 nt can be used to generate single-stranded regions of DNA of identical sequence to the hybridizing RNA. With either RNA quantitative R-loop formation could be achieved after modification of the displaced strand with glyoxal, a single strand-specific reagent that modifies guanines preferentially (16,17). Displacement of the RNA from the R-loop is blocked because this chemical modification prevents re-annealing of the duplex DNA.

Attenuated micrococcal nuclease

A high K_m mutant (L37A, Y113A) of micrococcal nuclease [MN(a)] has been prepared as both a glutathione-S-transferase fusion protein [GST-MN(a)] and as the free enzyme to cleave the single-stranded regions of DNA formed by R-loops. These mutants retain the wild-type enzyme's catalytic preference for the 3'-end of single-stranded Ts and cleave double-stranded DNA sluggishly (6,7,18–21). In addition to requiring calcium ions for activity, which permits regulation of its activity, this mutant enzyme shows reduced hydrolysis of nicked DNA. To establish this point, identical concentrations of M-13 single-stranded DNA, nicked double-stranded pGEM-820 (DSN) and ligase-repaired nicked double-stranded pGEM-820 (DSL) were incubated with either attenuated MN or wild-type MN.

The progress of the reaction was analyzed by labeling the 5' hydroxyl groups of the DNA substrate with T4 kinase at $t = 0$ and at various time points after quenching the reaction mixture with EGTA. DNA is present at a concentration of 50 μ M, expressed as pentanucleotides, the smallest substrate for endonucleolytic cleavage (20). For the three substrates the DNA concentration exceeds the enzyme concentration and the K_m of either enzyme and ensures that the concentration of nicked substrate is not rate limiting. Since ligase-treated pGEM-820 shows a 50% reduction in incorporation prior to MN hydrolysis, ~50% of the single-stranded nicks of DSN were repaired. Under conditions where the rate of scission of single-stranded M13 by both the wild-type and attenuated MN is approximately the same the attenuated enzyme shows a substantially decreased activity on nicked double-stranded DNA relative to the wild-type nuclease (Fig. 2B). These preferences are maintained in the GST fusion construct of MN(a).

Enzymatic scission of R-loops

R-loop-directed scission by GST-MN(a) was tested by preparing a short R-loop substrate composed of a 311 bp PCR product derived from pGEM-820 and a 111 nt RNA complementary to the

Figure 2. Analysis of substrate preference of wild-type MN (10 nM) (A) versus attenuated MN (40 nM) (B). After various time points scission was stopped and quantitated by labeling generated 5' hydroxyl groups with [γ - 32 P]ATP. The generation of 5'-ends is shown for three different substrates: single-stranded M13 (SS), nicked pGEM-820 plasmid (DSN) and ligase-treated pGEM-820 (DSL).

center of this 311 bp fragment. After fixation with glyoxal R-loops were treated with 0.12 μ M GST-MN(a) for 1 h and products separated on an denaturing acrylamide gel (Fig. 3A) next to sequencing lanes (not shown). The displaced non-template strand (lane 3) is quantitatively cut and the products are clustered around the end of the R-loop, indicating a highly preferred scission of the displaced single-stranded region. However, specific double-stranded scission of the R-loop also requires scission of the hybridized (template) strand at a rate that is substantially above background scission. Although scission of the template strand does occur at the nucleotides immediately adjacent to the heteroduplex, cleavage of this strand is rate limiting, because the heteroduplex is a very poor substrate (Fig. 3A, lane 5). The scission of the template strand that is observed is probably the result of incomplete or weak hydrogen bonding in the adjacent 5–10 nt, which generates a partially single-stranded stretch of DNA. The result of these preferences is a mode of double-stranded scission which generates two sets of products, corresponding to cleavage at either side of the

Figure 3. (A) Scission of a labeled pGEM-820-derived 311 bp fragment containing a central 111 bp R-loop. Partial restriction digests with R-loop-free fragment labeled on the non-template (displaced) strand identify the location of the R-loop (lanes 1 and 2). Lanes 3 and 5, treatment of R-loop-containing fragment with 0.12 μ M GST-MN(a) (lane 3, displaced, non-template strand labeled; lane 5, template strand labeled). Lane 4, R-loop-free fragment (template strand labeled) after treatment with 0.12 μ M GST-MN(a). The schematic of the R-loop indicates the position of both restriction sites as they would appear with the template strand labeled (lanes 4 and 5). (B) Summary of scission sites on the template strand from (A) (open arrows) and comparison with the sequence and overall double-stranded scission (solid arrows) using either T7- or SP6 (opposite strand)-derived RNA to form the R-loop. The arrows represent relative intensities within each scission experiment.

R-loop but not its complete degradation. Figure 3B summarizes the results of this scission experiment at the nucleotide level and compares the overall rate of scission for either side of this R-loop with that obtained with RNA derived from the opposite strand, using the SP6 promoter. The results are consistent with the conclusion that efficient double-stranded scission requires the presence of thymidine in the sequence surrounding the end of the R-loop. In the case of the RNAs used in this experiment these are

Figure 4. R-loop-directed scission of pGEM-820 (0.01 μ M) with GST-MN(a) (0.1 μ M). Lane 1, linearized control plasmid; lane 2, scission after R-loop formation and glyoxal fixation; lane 3, scission after R-loop formation, fixation and RNase H treatment; lane 4, as lane 3 but with a high (0.5 mg/ml) concentration of RNase A instead of RNase H; lane 5, HindIII digest of R-loop-free plasmid; lane 6, EcoRI digest of R-loop-free plasmid; lane 7, R-loop-free plasmid without glyoxal treated with GST-MN(a); lane 8, R-loop reaction with a 100 nt sequence of unrelated RNA and treatment with GST-MN(a); lane 9, as lane 7 but treated with glyoxal and RNase A; lane 10, as lane 8 but treated with glyoxal and RNase A. Recovery is reduced in lanes 2-4, 9 and 10 compared with the overall DNA yield for reactions that were merely diluted for scission (lane 5-8), because precipitation and several washing steps are essential after glyoxal treatment.

mainly contributed by the TATA elements of the corresponding promoters.

Optimizing the yield of R-loop-directed scission

Two steps were added to the protocol to enhance the yield of double-stranded scission. The first was to stabilize the R-loop by chemical modification of the displaced strand with glyoxal to prevent displacement of the RNA strand (1,5). The second step was hydrolysis of the RNA used for R-loop formation after glyoxal fixation with either RNase H or RNase A. Both modifications were incorporated into the protocol in order to make both DNA strands fully accessible for hydrolysis by single strand-specific nucleases.

In Figure 4 the result of the scission of pGEM-820 (2.9 kb) by GST-MN(a) targeted by the T7 derived 111 nt RNA (lanes 2-4) and a 100 nt pGEM-820-unrelated RNA (lanes 8 and 10) is presented. The advantage of glyoxal modification and RNase treatment is apparent from these results. Scission with GST-MN(a) after fixation of the R-loop but without removal of the RNA (lane 2) gives specific scission with a pattern that corresponds to the EcoRI and HindIII digests, but the overall yield of scission is low (in this case 15-20%) and differs from RNA to RNA according to the base composition in the positions on the flanks of the R-loop. This experiment is analogous to scission of the 311 bp fragment and the pattern is indicative of double-stranded scission at the flanks of the R-loop. The stronger bands in this case closely match the HindIII digest, which is consistent with the results shown in Figure 3A and B.

However, removal of the RNA by digestion with RNase H after fixation with glyoxal gives a scission yield >90% (lane 3). Furthermore, only two fragments are generated by this digest, indicating a complete removal of the R-loop. Both results are consistent with the conclusion that combined glyoxal and RNase treatment yields a single-stranded DNA bubble which presents a highly favorable substrate for the enzyme. The same scission

Figure 5. (A) Outline of the pETenHB (4.9 kb) test system. AflIII-linearized plasmid served as the substrate for the scission experiment presented in (B). (B) R-loop-directed scission of pETenHB (4.9 kb) with GST-MN(a) or MN(a). Lane 1, linearized control plasmid; lane 2, glyoxal-fixed R-loop + GST-MN(a); lane 3, glyoxal-fixed R-loop + MN(a); lane 4, glyoxal-fixed and RNase A-treated R-loop + GST-MN(a); lane 5, *Xba*I digest of R-loop-free plasmid; lane 6, *Bgl*II digest of R-loop-free plasmid (lower band not indicative of R-loop border due to a second *Bgl*II site); lane 7, R-loop-free plasmid + GST-MN(a); lane 8, R-loop + GST-MN(a); lane 9, R-loop + MN(a); outside lane to the right, 1 kb marker.

products can be obtained by substituting RNase H with high concentrations (0.5 mg/ml) of RNase A (lane 4). This result can be explained by dissociation of the RNA from the ends of the heteroduplex. These short stretches of single-stranded RNA are hydrolyzed by RNase A. The disadvantage of using high concentrations of RNase A (a substantial amount of the DNA was trapped in the origin of the gel) was overcome in subsequent experiments (e.g. Fig. 5) by using less RNase A (0.05 mg/ml) at higher temperatures (37°C instead of room temperature) to favor RNA dissociation. The specificity of R-loop formation and scission is demonstrated in lanes 7–10. Reactions without RNA (lanes 7 and 9) or with sequence-unrelated RNA (lanes 8 and 10) show no scission by the nuclease either with (lanes 7 and 8) or without (lanes 9 and 10) fixation and RNase treatment. The high K_m mutant of MN can be replaced by 10 U mung bean nuclease used in an overnight digestion at 4°C.

To confirm these findings in a different test system we generated a 145 nt RNA of the T7 promoter in the pET-based vector pETenHB (4.9 kb), containing the coding region of the DNA binding portion of the engrailed homeodomain protein (13) (Fig. 5). Treatment with glyoxal and RNase A (0.05 mg/ml for 30 min at 37°C) and subsequent scission with GST-MN(a) yields quantitative scission of the 4.9 kb plasmid without detectable background. The fragments generated by this scission match the smaller *Xba*I fragment (lane 5) and the larger *Bgl*II fragment (lane 6), which is indicative of a complete removal of the R-loop. Minor differences in the reactivity of MN(a) and GST-MN(a) with R-loops formed in the absence of glyoxal stabilization (lanes 8 and 9) vanish upon stabilization (lanes 2 and 3). The multiple bands reflective of partial cleavage of the DNA strand involved in heteroduplex formation disappear upon RNase hydrolysis. This set of experiments demonstrates the advantages gained by glyoxal fixation and RNase hydrolysis. Experiments with both plasmids demonstrate that R-loop-directed scission by attenuated mutants of micrococcal nuclease is an efficient and reproducible method to achieve sequence-specific cleavage of a double-stranded DNA sequence that depends solely on the choice of the RNA used for R-loop formation.

Scission of a 120 kb P1 plasmid

Although 3–5 kb plasmids (Figs 4 and 5) provide an opportunity to test the feasibility of R-loop-directed scission, the utility of the

methodology lies in the analysis of larger (~100 kb) P1 plasmids which are not likely to have restriction sites conveniently located at desirable positions. However, these larger plasmids will also pose methodological problems not present in 3–5 kb plasmids. For example, the nuclease must exhibit higher stringency for single-stranded DNA relative to double-stranded DNA. Moreover, the conditions for R-loop formation might lead to the preferential melting of A-T-rich regions, which in turn could result in considerable background scission. To evaluate these features further R-loop-directed scission by GST-MN(a) was used in the analysis of the exon/intron structure of P1 plasmids containing a genomic copy of the gene encoding human L-isoaspartate O-methyltransferase, a protein which has been strongly conserved throughout evolution (11).

The design of this experiment is outlined in Figure 6. The templates for *in vitro* transcription of suitable RNA probes for this scission experiment were derived from the target DNA by PCR as outlined in Figure 6A. The results of scission of the target by GST-MN(a) at different enzyme concentrations were analyzed by Southern blotting using probes which specifically target the insert or the flanking regions, respectively (Fig. 6B). Even though separation on this gel system does not allow accurate sizing of the intervening sequence, a product band between 12 and 20 kb is apparent, which is generated in an RNA-dependent manner (compare Fig. 6B, lane 3 and 6). This band is only visible using a probe directed towards the region between both R-loop targets and not when the same blot is probed for the outside vector sequence. Scission in lane 6 is apparently quantitative with respect to the disappearing parent band. The difference in band intensity can be attributed to the fact that the insert-specific probe shows partial overlap with the area targeted for scission. Equal intensities are obtained when the same lane is analyzed with an outside probe (Fig. 6C, lane 3 and 6). This experiment demonstrates the feasibility of direct distance measurements between two sequences of choice on a substrate that exceeds 100 kb.

However, the results in Figure 6B also highlight some of the problems which become apparent when the current protocol was applied to large sized DNA. The concentration of R-loop targets is reduced 30-fold compared with the plasmid scission experiment. At enzyme concentrations similar to the plasmid scission experiment (100 nM, Fig. 6B, lane 5) weak scission is observed. A strong scission product is observed at a 3-fold higher enzyme concentra-

A

Figure 6. Scission of a 120 kb P1 plasmid at two positions targeted through R-loop formation. (A) Scheme of the test system and location of the sequences used to generate the RNAs for R-loop formation and the radiolabeled probes for Southern analysis. The generation of templates by PCR with T7 promoter containing primers is outlined for exon 7 (E 7). (B) Southern analysis of P1 scission products after R-loop formation and scission for 1 h at room temperature with different concentrations of GST-MN(a). The products were analyzed by standard gel electrophoresis on a 0.5% agarose gel and probed with an exon 7a-derived probe. Lanes 1-3, (no RNAs during R-loop formation) GST-MN(a) at 30 (1), 100 (2) and 300 nM (3); lanes 4-6, (both RNAs present during R-loop formation) GST-MN(a) at 30 (4), 100 (5) and 300 nM (6); lane 7, untreated P1. (C) The same gel as in (B) probed with a probe derived from exon 3.

tion, at the expense of considerable background scission. Comparison of the signal obtained with the outside probe also reveals that a large amount of the scission that occurs at this high enzyme concentration is RNA independent. Substantially less background scission is observed at 100 nM nuclease. This enzyme concentration range corresponds approximately to the threshold (>300 nM) where non-specific interaction can be observed between DNA and GST-MN(a) in gel shift experiments (7).

Background scission is attributable to the likely complete separation of highly nicked DNA when R-loops are formed in 70% formamide. Incomplete re-association will therefore provide favorable substrates for MN(a). Comparison of several preparations (data not shown) shows that the extent to which the plasmids denature as a result of the R-loop formation procedure is a reflection of the individual plasmid preparation. It therefore appears that at this stage the methodology described is useful for direct measurement of distances between probe sequences, but 120 kb may represent the upper limit of the applicability of the technique.

DISCUSSION

Restriction endonucleases and PCR have played a central role in the analysis of genome structure. Despite the limited repertoire and length of endonuclease recognition sites and the possibility that they may not be available in a region of interest, creative multi-step 'Achilles heel' protection strategies have enhanced their flexibility and applicability (22,23). The utility of PCR in genomic applications has been greatly enhanced by the recent development of protocols which allow reliable amplification of DNAs as long as 40 kb (23).

The R-loop-based methodology described here should complement both of these important methods for the analysis of high molecular weight DNA. Its advantage relative to any methodology which relies on restriction enzymes is its greater inherent flexibility. While the size of the largest fragment that can be recovered after R-loop-directed scission is a reflection of the single strand specificity of the nuclease used for scission and the quality of the DNA starting material, there is no intrinsic limit to the distance between two R-loops that can be targeted for scission through R-loop formation. Although both the R-loop method and PCR will have problems with high molecular weight nicked DNA, PCR may be more robust, because only a fraction of the DNA target molecules must remain intact for measurable amounts of product to be generated.

Two avenues show promise for avoiding the extensive formation of single-stranded DNA, which may arise from nicked DNA, during R-loop formation. The first is psoralen cross-linking (25), which has previously been used in electron microscopy studies (1). In this case the undesired complete strand separation of the DNA is prevented through limited cross-linking of the DNA. Secondly, repairing the DNA substrate with polynucleotide kinase and ligase dramatically increases the yield of sequence-specific scission. Ultimately, the best method to provide high molecular weight DNA of high quality would be the use of agarose embedded DNA. Preliminary experiments have identified commercially available agarose matrices which are stable under the conditions of R-loop formation and allow pulsed field electrophoresis of large sized DNA in gel slice R-loop formation and nuclease scission. Current work focuses on the necessary parameter adjustments for R-loop-directed scission under these modified conditions.

The single strand specificity of the nuclease used in the scission of the single-stranded bubble is another factor that must be addressed. MN(a) or its fused counterpart GST-MN(a) has several advantages as the nuclease for use in the R-loop-based method. In the first place, it retains the dependence on calcium ions for catalytic activity characteristic of wild-type MN. As a result, the enzyme can be readily activated by calcium addition and quenched by the calcium-specific ligand EGTA. Secondly, its sluggish catalytic activity permits incubations long enough to hydrolyze the

single-stranded bubble and yet avoid overdigestion of double-stranded DNA. This pronounced preference for single-stranded DNA allows efficient degradation of the bubble produced as a consequence of R-loop formation. Finally, compared with wild-type MN, MN(a) cleaves the DNA opposite single strand nicks with highly reduced affinity. This is of special importance in the treatment of large sized DNA, which is bound to encounter nicking as a result of the isolation procedure. Mung bean nuclease, which has a more stringent single strand specificity than MN(a) (26,27), can be used as the nuclease and has the advantage of being commercially available. However, neither MN(a) nor mung bean nuclease are absolutely specific for single-stranded DNA. After prolonged incubation duplex DNA will be degraded. Methods for improving the cleavage chemistry might include the linking of MN(a) to a hybridizing DNA or RNA via a cibacron blue tether. We have previously demonstrated that this dye can direct MN(a) prepared as a fusion protein with glutathione S-transferase because it has a high affinity for the catalytic site of the transferase (7). Concentrating the nuclease at the site of the bubble will minimize the potential for non-specific background scission.

The R-loop-dependent methodology for double strand-specific DNA scission achieves virtually quantitative scission of plasmid sized DNA. Initial studies have indicated that the method is also applicable to a P1 plasmid DNA which has a molecular weight of ~120 kb. The problems which arise are those associated with manipulation of this much larger DNA and are not necessarily inherent in the method for targeting a specific DNA sequence for scission. Current research is focused on minimizing these problems so that this flexible method of DNA scission can have general application to genomic analysis.

ACKNOWLEDGEMENTS

This research was supported by the Human Genome Project grant HG00255 and the award of 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 Rosbash,M., Blank,D., Fahrner,K., Hereford,L., Ricciardi,R., Roberts,B., Ruby,S. and Woolford,J. (1979) *Methods Enzymol.*, **68**, 454–469.
- 5 Landgraf,R., Chen,C.-h.B. and Sigman,D.S. (1995) *Nucleic Acids Res.*, **23**, 3516–3523.
- 6 Corey,D.R., Pei,D. and Schultz,P.G. (1989) *J. Am. Chem. Soc.*, **111**, 8523–8525.
- 7 Landgraf,R., Chen,C.-h.B. and Sigman,D.S. (1994) *Biochemistry*, **33**, 10607–10615.
- 8 Chen,C.-h.B., Gorin,M.B. and Sigman,D.S. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 4206–4210.
- 9 Sigman,D.S., Chen,C.-h.B. and Gorin,M.B. (1993) *Nature*, **363**, 474–475.
- 10 Tsai,W. and Clarke,S. (1994) *Biochem. Biophys. Res. Commun.*, **203**, 491–497.
- 11 MacLaren,D.C., O'Connor,C.M., Xia,Y.R., Mehrabian,M., Klisak,I., Sparkes,R.S., Clarke,S. and Lusk,A.J. (1992) *Genomics*, **14**, 852–856.
- 12 Takahara,M., Hibler,D.W., Barr,P.J., Gerlt,J.A. and Inouye,M. (1985) *J. Biol. Chem.*, **260**, 2670–2674.
- 13 Smith,D.B. and Johnson,K.S. (1988) *Gene*, **67**, 31–40.
- 14 Poole,S.J., Kauvar,L.M., Drees,B. and Kornberg,T. (1985) *Cell*, **40**, 37–43.
- 15 Murakawa,G.J., Chen,C.-h.B., Kuwabara,M.D., Nierlich,D. and Sigman,D.S. (1989) *Nucleic Acids Res.*, **17**, 5361–5369.
- 16 Broude,N.E. and Budowsky,E.I. (1971) *Biochim. Biophys. Acta*, **254**, 380–388.
- 17 Nakaya,K., Takenaka,O., Horinishi,H. and Shibata,K. (1968) *Biochim. Biophys. Acta*, **161**, 23–31.
- 18 Cuatrecasas,P., Wilchek,M. and Anfinsen,C.B. (1969) *Biochemistry*, **8**, 2277–2284.
- 19 Cuatrecasas,P., Fuchs,S. and Anfinsen,C.B. (1967) *J. Biol. Chem.*, **242**, 1541–1547.
- 20 Anfinsen,C.B., Cuatrecasas,P. and Taniuchi,H. (1971) *Enzymes*, **4**, 177–204.
- 21 Gerlt,J.A. (1992) *Enzymes*, **20**, 95–136.
- 22 Ferrin,L.J. and Camerini-Otero,R.D. (1991) *Science*, **254**, 1494–1497.
- 23 Strobel,S.A. and Dervan,P.B. (1991) *Nature*, **350**, 172–174.
- 24 Cheng,S., Fockler,C., Barnes,W.M. and Higuchi,R. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 5695–5699.
- 25 Johnston,B.H., Johnson,M.A., Moore,C.B. and Hearst,J.E. (1977) *Science*, **197**, 906–908.
- 26 Johnson,P.H. and Laskowski,M.S. (1970) *J. Biol. Chem.*, **245**, 891–898.
- 27 Kedzierski,W. and Laskowski,M.S. (1973) *J. Biol. Chem.*, **248**, 1277–1280.