Biochemistry

 Volume 32, Number 24 June 22, 1993

Accelerated Publications

A Distinct Intron-DNA Structure in Simian Virus 40 T-Antigen and Adenovirus 2 E1A Genes†

Inho Lee and Jacqueline K. Barton*

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125

Received March 5, 1993; Revised Manuscript Received April 15, 1993

ABSTRACT: Distinct structures delineating the introns of simian virus 40 T-antigen and adenovirus 2 E1A genes have been discovered. The structures, which are centered around the branch points of the genes inserted in supercoiled double-stranded plasmids, are specifically targeted through photoactivated strand cleavage by the metal complex tris(4,7-diphenyl-1,10-phenanthroline)rhodium(III). The DNA sites that are recognized lack sequence homology but are similar in demarcating functionally important sites on the RNA level. The single-stranded DNA fragment corresponding to the coding strand of the E1A gene was also found to fold into a structure apparently identical to that in the supercoiled E1A gene based on the recognition by the metal complex. Further investigation of different single-stranded DNA fragments showed that the structure requires the sequences at both ends of the intron plus the flanking sequences but not the middle of the intron. These results provide direct evidence that the positions of these introns are structurally encoded on the DNA level.

Since the discovery of introns, much attention has been focused on the mechanism of the processing of the nascent transcript to the messenger RNA, a process which involves a myriad of factors and precise splicing of the folded RNA molecule (Padgett et al., 1986; Green, 1986; Maniatis & Reed, 1987; Cech, 1986). Since most eukaryotic genes have been found to contain introns, their persistence in a higher order structure of genes suggests that introns confer an evolutionary advantage to the organism, an advantage great enough to justify maintaining the elaborate splicing machinery. But what function(s) does (do) introns serve? Here we report finding an unusual structure in the intron DNA of two mammalian viral genes. The results we describe establish that in these genes the position of introns are encoded structurally in the DNA.

Structural probing of the intron DNA was carried out with tris(4,7-diphenyl-1,10-phenanthroline)rhodium(III) [$Rh(DIP)_3^{3+}$] (Figure 1), a metal complex designed in our laboratory which lacks hydrogen-bonding functionalities and instead targets distinct nucleic acid sites through *shape selection* (Pyle & Barton, 1990; Chow & Barton, 1992a). This site-specific targeting may be monitored sensitively since, upon photoactivation with ultraviolet light (313–332 nm),

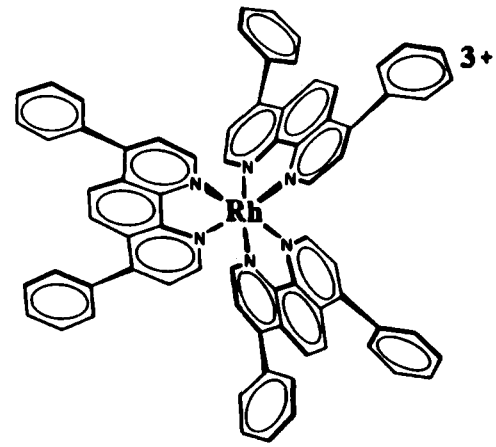

FIGURE 1: Tris(4,7-diphenyl-1,10-phenanthroline)rhodium(III) [$Rh(DIP)_3^{3+}$].

† We are grateful to the NIH for their support of this work.
* To whom correspondence should be addressed.

0006-2960/93/0432-6121$04.00/0

the metal complex promotes strand cleavage directly at its binding site (Pyle & Barton, 1990; Kirschenbaum, 1989). A secondary reaction which is not structure-specific is also apparent at guanines. Detailed product analysis of cleavage by $Rh(DIP)_3^{3+}$ at sites of altered structure has not been carried out. However, mechanistic studies (Sitlani et al., 1992) on DNA photocleavage by $Rh(phen)_2phi^{3+}$, an analogue of $Rh(DIP)_3^{3+}$, on B-form DNA oligonucleotides indicate direct C3′-H atom abstraction by a photoexcited ligand radical, and this mechanism is thus far consistent with the observed cleavage by $Rh(DIP)_3^{3+}$. Using this cleavage methodology, $Rh(DIP)_3^{3+}$ has been shown to recognize unusual non-B-DNA tertiary structures such as a cruciform (Kirschenbaum et al., 1988) and a Holliday junction (Waldron et al., unpublished results). $Rh(DIP)_3^{3+}$ and its structural analogue $Co(DIP)_3^{3+}$ also promote cleavage on Z-DNA, but no cleavage is apparent on A-form nucleic acids or on uncoiled single-stranded sites (Kirschenbaum, 1989; Barton & Raphael, 1985). With RNA as a substrate $Rh(DIP)_3^{3+}$ also specifically targets G-U mismatches within helical regions of RNA, while the A-RNA double helices or uncoiled regions are not cleaved (Chow & Barton, 1992b).

$Rh(DIP)_3^{3+}$ and $Co(DIP)_3^{3+}$ have been shown earlier (Müller et al., 1987) to target functionally significant sites in the genome of simian virus 40 (SV40), a circular double-stranded DNA. In particular, a strong site specifically recognized by the rhodium complex was identified in the intron contained in the large T-antigen and the small t-antigen genes. Low-resolution studies (Keene & Elgin, 1984; McCutchan et al., 1984) with enzymatic probes have also pointed to structural polymorphism in genomic DNA at positions before and after genes. These intriguing observations prompted the present study of intron DNA recognition on both the SV40 T-antigen gene and the adenovirus 2 (Ad2) E1A gene.

MATERIALS AND METHODS

Plasmids. Plasmids containing the genes of interest as well as a control cruciform target site were provided by Prof. James L. Manley of Columbia University. The plasmids were constructed by cloning *Hin*dIII fragments of the two genes into the expression vector pSP64, a pBR322 derivative, from Promega. The pSP64-SVT plasmid contains SV40 sequences from 4002 to 5171, and the pSP64-E1A plasmid contains Ad2 sequences from 500 to 1569. The pre-mRNA of the SV40 early gene is spliced alternatively to generate the mRNAs for the large T-antigen and small t-antigen. [See Berk and Sharp (1978a).] The E1A pre-mRNA is also alternatively spliced to give three major products, the 9S, the 12S, and the 13S mRNAs. [See Berk and Sharp (1978b).]

Single-Stranded DNA Fragments. Four single-stranded DNA (ssDNA) fragments for the E1A intron were synthesized: a 174-mer containing the full intron plus 33 nucleotides of exon sequences flanking the 5′ end and 27 nucleotides flanking the 3′ end; a 114-mer representing the complete intron; a 70-mer representing the 3′ two-thirds of the intron; and an 85-mer corresponding to the two ends of the intron including the branch site and the $Rh(DIP)_3^{3+}$ cleavage site on the supercoiled DNA plus 15 nucleotides of flanking exon sequences at both ends but with the middle of the intron deleted. The 174-mer was prepared by first generating a double-stranded DNA fragment using PCR with two primers and digesting that dsDNA fragment with *Ava*I which gives a 174-bp fragment with a four-base 5′ overhang and then separating the two strands of the 174-bp fragment on a denaturing polyacrylamide gel. The 114-mer, 70-mer, and 85-mer were all prepared by solid-phase synthesis using phosphoramidite chemistry. The 114-mer was purified on the NENSorb column (New England Nuclear). The 70-mer and the 85-mer were purified by HPLC (Waters).

Low-Resolution Mapping of the Plasmids. Supercoiled plasmid DNA (100 μM nucleotides) in 20 μL of 20 mM Tris-HCl, pH 7.4, and 25 mM NaCl is photolyzed at 313 nm for 2 min with a Hg/Xe lamp (Oriel) in the presence of 10 μM $Rh(DIP)_3^{3+}$ and ethanol precipitated. The DNA is resuspended and digested with a restriction enzyme and then with nuclease S1 (pH 4.5). The reaction mixture is loaded directly on a 1% agarose gel and electrophoresed. The gel is stained in 0.5 μg/mL ethidium bromide, destained in 1 mM $MgSO_4$, and photographed irradiated from below with UV light. Relatively long times of irradiation were used in these experiments, both to explore the possibility of cleavage on linear fragments as well as to visualize more easily the specific fragments observed after photolysis of the supercoiled plasmids.

High-Resolution Mapping of the Plasmids. The plasmid (100 μM nucleotides) is photolyzed at 332 nm for 1–4 min in 20 mM Tris-HCl, pH 7.4, and 25 mM NaCl in the presence of 5–10 μM $Rh(DIP)_3^{3+}$. The DNA is then washed with 1% sodium dodecyl sulfate to remove the metal complex and ethanol precipitated three to four times. Thereafter, the DNA is digested with a restriction enzyme, *Acc*I for E1A and *Bsm*I for T-antigen, that cuts the plasmids relatively close to the site of photocleavage. The resulting linear plasmid is end-labeled with ^{32}P using polynucleotide kinase for E1A and terminal deoxynucleotidyl transferase for T-antigen (Sambrook et al., 1989) and digested with another restriction enzyme, *Xba*I for E1A and *Bst*XI for T-antigen, that gives rise to a labeled fragment (234 bp in length for E1A and 238 bp for SV40) containing the site of photocleavage. The fragment is then isolated on a nondenaturing polyacrylamide gel, denatured, and electrophoresed on a denaturing polyacrylamide gel together with samples sequenced by the Maxam–Gilbert method (Maxam & Gilbert, 1980).

High-Resolution Mapping of the ssDNA Fragments. The 174-mer and the 85-mer are labeled at the 5′ end with ^{32}P using polynucleotide kinase, and the 114-mer and the 70-mer are labeled at the 3′ end with ^{32}P using terminal deoxynucleotidyl transferase. All the labeled fragments are purified again on denaturing polyacrylamide gels before irradiation for 2–8 min in the presence of 1–10 μM $Rh(DIP)_3^{3+}$. Then the DNA is ethanol precipitated twice and run on analytical denaturing polyacrylamide gels together with samples sequenced by the Maxam–Gilbert method.

RESULTS AND DISCUSSION

Low-Resolution Mapping of the Rhodium Complex Cleavage Sites in the Plasmids. Low-resolution mapping experiments, schematically illustrated in Figure 2A, were carried out on the SV40 (Figure 2B) and Ad2 (Figure 2C) plasmids. Digestion with two different restriction enzymes and S1 nuclease after photolysis in the presence of $Rh(DIP)_3^{3+}$ reveals the specificity of single-strand cleavage by the rhodium complex. On each plasmid a site within the intron insert is specifically targeted. Two other sites of specific cleavage are evident, one of which corresponds to a cruciform and another of which is located near the origin of replication for the plasmid; both sites had been identified earlier in mapping studies of pBR322 with $Co(DIP)_3^{3+}$ (Barton & Raphael, 1985). At lower levels of irradiation of supercoiled plasmids followed by linearization but without S1 nuclease treatment, fragments

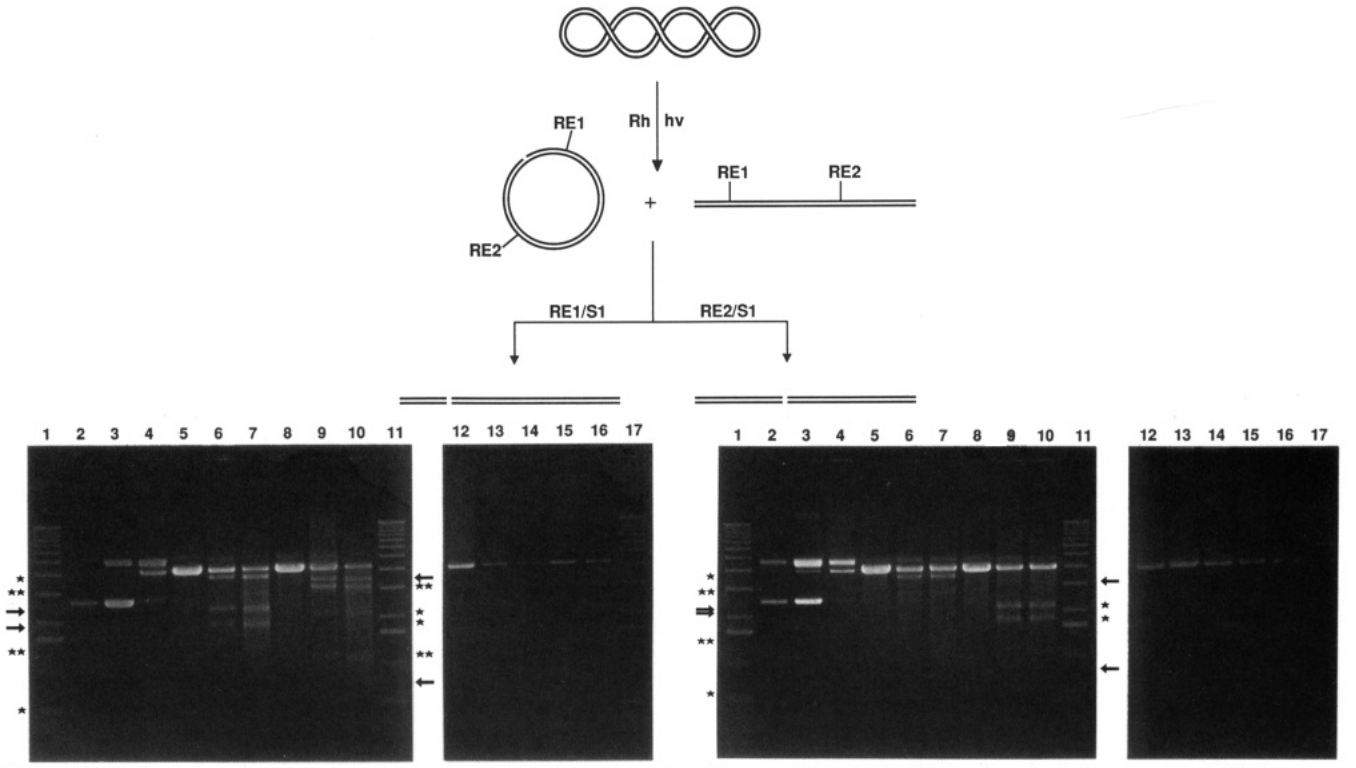

FIGURE 2: Targeting of distinct tertiary structures by $Rh(DIP)_3^{3+}$ on plasmids containing the SV40 T-antigen and the adenovirus 2 E1A genes. (A, top) Schematic illustration of the procedure used to identify regions specifically targeted by rhodium photocleavage on the supercoiled plasmids. Rhodium photocleavage leads to the conversion of form I DNA to a mixture of forms II and III. After linearization of the plasmid with a single-site restriction enzyme (RE1) and digestion with S1 nuclease (which cleaves opposite the rhodium-induced nick), discrete fragments should be evident if the rhodium photocleavage is centered at specific sites on the plasmid. The same protocol but using a different restriction enzyme (RE2) permits the unique localization of the specific site cleaved, based upon the length of the fragments obtained. To examine whether supercoiling is required for specific targeting, an equivalent protocol was employed but the plasmids were digested first with the restriction enzyme, followed by photolysis in the presence of the metal complex and digestion with S1 nuclease. (B, bottom left) Fragmentation map of sites cleaved specifically by $Rh(DIP)_3^{3+}$ on supercoiled pSP64-SVT, containing the SV40 T-antigen intron: lane 1, 1-kb marker (Bethesda Research Laboratories); lane 2, DNA incubated with $Rh(DIP)_3^{3+}$ but without irradiation at 313 nm; lane 3, DNA irradiated at 313 nm but without $Rh(DIP)_3^{3+}$; lane 4, DNA irradiated at 313 nm in the presence of $Rh(DIP)_3^{3+}$; lane 5, DNA irradiated at 313 nm without $Rh(DIP)_3^{3+}$ and digested with *Pvu*I and S1; lane 6, DNA irradiated at 313 nm with $Rh(DIP)_3^{3+}$ and then digested with *Pvu*I; lane 7, DNA irradiated at 313 nm with $Rh(DIP)_3^{3+}$ and then digested with *Pvu*I and S1; lane 8, DNA irradiated at 313 nm without $Rh(DIP)_3^{3+}$ and digested with *Eco*RI and S1; lane 9, DNA irradiated at 313 nm with $Rh(DIP)_3^{3+}$ and then digested with *Eco*RI; lane 10, DNA irradiated at 313 nm with $Rh(DIP)_3^{3+}$ and then digested with *Eco*RI and S1; lane 11, 1-kb marker; lane 12, plasmid first linearized with *Pvu*I and incubated with $Rh(DIP)_3^{3+}$; lane 13, plasmid first linearized with *Pvu*I and irradiated at 313 nm without $Rh(DIP)_3^{3+}$; lane 14, plasmid first linearized with *Pvu*I and irradiated at 313 nm without $Rh(DIP)_3^{3+}$ and digested with S1; lane 15, plasmid first linearized with *Pvu*I and irradiated at 313 nm with $Rh(DIP)_3^{3+}$; lane 16, plasmid first linearized with *Pvu*I and irradiated at 313 nm with $Rh(DIP)_3^{3+}$ and digested with S1; lane 17, 1-kb marker. Arrows indicate the fragments formed (2350 and 1900 base pairs in length with *Pvu*I digestion, and 3510 and 740 base pairs in length with *Eco*RI digestion, with a margin of error of 50 base pairs) as a result of specific cleavage within the intron (lanes 7 and 10). Supercoiling is also required for the targeting of these structures (lanes 15 and 16). (C, bottom right) Fragmentation map of sites cleaved specifically by $Rh(DIP)_3^{3+}$ on supercoiled pSP64-E1A, containing the Ad2 E1A intron: lane 1, 1-kb marker; lane 2, DNA incubated with $Rh(DIP)_3^{3+}$ but without irradiation at 313 nm; lane 3, DNA irradiated at 313 nm but without $Rh(DIP)_3^{3+}$; lane 4, DNA irradiated at 313 nm in the presence of $Rh(DIP)_3^{3+}$; lane 5, DNA irradiated at 313 nm without $Rh(DIP)_3^{3+}$ and then digested with *Pvu*I and S1; lane 6, DNA irradiated at 313 nm with $Rh(DIP)_3^{3+}$ and then digested with *Pvu*I; lane 7, DNA irradiated at 313 nm with $Rh(DIP)_3^{3+}$ and then digested with *Pvu*I and S1; lane 8, DNA irradiated at 313 nm without $Rh(DIP)_3^{3+}$ and then digested with *Hin*dIII and S1; lane 9, DNA irradiated at 313 nm with $Rh(DIP)_3^{3+}$ and then digested with *Hin*dIII; lane 10, DNA irradiated at 313 nm with $Rh(DIP)_3^{3+}$ and then digested with *Hin*dIII and S1; lane 11, 1-kb marker; lane 12, plasmid first linearized with *Pvu*I and incubated with $Rh(DIP)_3^{3+}$; lane 13, plasmid first linearized with *Pvu*I and irradiated at 313 nm without $Rh(DIP)_3^{3+}$; lane 14, plasmid first linearized with *Pvu*I and irradiated at 313 nm without $Rh(DIP)_3^{3+}$ and digested with S1; lane 15, plasmid first linearized with *Pvu*I and irradiated at 313 nm with $Rh(DIP)_3^{3+}$; lane 16, plasmid first linearized with *Pvu*I and irradiated at 313 nm with $Rh(DIP)_3^{3+}$ and digested with S1; lane 17, 1-kb marker. Arrows indicate the fragments formed (2040 and 2010 base pairs in length with *Pvu*I digestion and 3360 and 690 base pairs in length with *Hin*dIII digestion with a margin of error of 50 base pairs) as a result of specific cleavage within the intron (lanes 7 and 10). The stars denote fragments resulting from cleavage at the cruciform site and double stars indicate cleavage at the origin. No specific sites are targeted if the plasmid is first linearized (lanes 15 and 16), showing the requirement of supercoiling in the structures targeted.

are observed primarily which correspond to specific cleavage at the cruciform site rather than the intron site, which indicates that the cleavage at the intron site is primarily single-stranded. It should be noted that the relative intensities of the cleaved fragments vary between plasmids. Also shown are the digestions after photolysis in the presence of rhodium of the linearized rather than supercoiled plasmid. In this case no specific fragments are apparent. Like the cruciform site, therefore, the structures targeted in the introns are supercoil-dependent.

High-Resolution Mapping of the Rhodium Complex Cleavage Sites in the Introns. To determine precisely the sequences which are involved in the recognition by the rhodium complex, high-resolution mapping of the intron cleavage sites, schematically illustrated in Figure 3A, was carried out on both SV40 T-antigen and Ad2 E1A introns. Panels B and C of Figure 3 display the results for the T-antigen and E1A introns, respectively. Within the coding strand of the T-antigen intron (Figure 3B), specific photoactivated cleavage by $Rh(DIP)_3^{3+}$ is observed within the sequence 5′-AAAC-

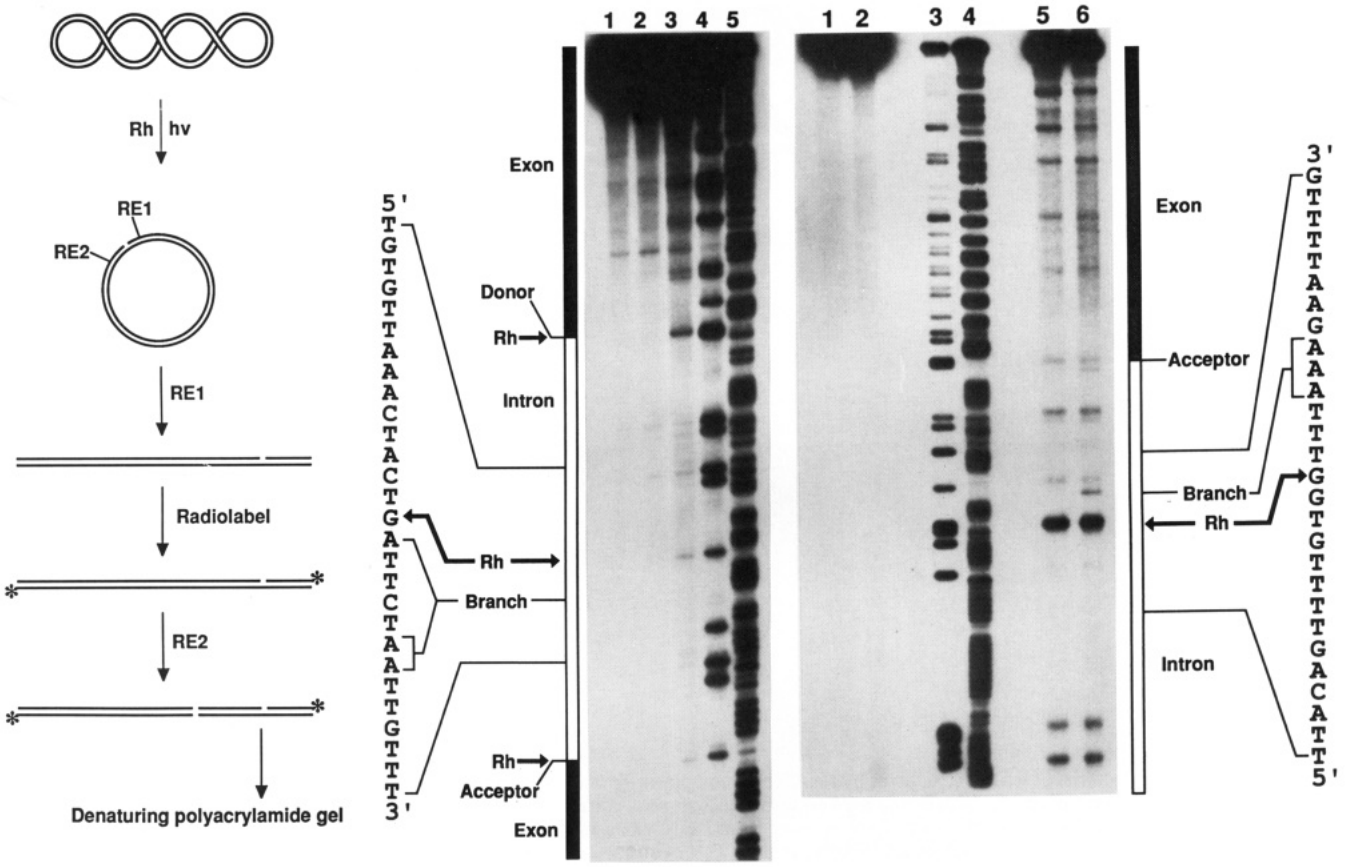

FIGURE 3: High-resolution mapping of tertiary structures targeted by $Rh(DIP)_3^{3+}$ within the introns of SV40 T-antigen and Ad2 E1A genes. (A, left) Schematic illustration of the protocol used to identify specific sites cleaved by $Rh(DIP)_3^{3+}$ on the supercoiled plasmids. The plasmid is photolyzed at 332 nm in the presence of 5–10 μM $Rh(DIP)_3^{3+}$. The DNA is then digested with a restriction enzyme that cuts the plasmid relatively close to the site of photocleavage. The resulting linear plasmid is end-labeled with ^{32}P and digested with another restriction enzyme that gives rise to a labeled fragment containing the site of photocleavage. The fragment is then analyzed on a denaturing polyacrylamide gel. (B, center) Site-specific cleavage of the SV40 T-antigen intron DNA coding strand: lane 1, the pSP64-SVT fragment in the absence of both irradiation and $Rh(DIP)_3^{3+}$; lane 2, the pSP64-SVT fragment after irradiation in the absence of $Rh(DIP)_3^{3+}$; lane 3, the pSP64-SVT fragment after irradiation for 4 min in the presence of 5 μM $Rh(DIP)_3^{3+}$; lane 4, Maxam–Gilbert G reaction; lane 5, Maxam–Gilbert T + C reaction. Cleavage (arrows) by $Rh(DIP)_3^{3+}$ is observed at the donor and the acceptor sites and at the G six bases to the 5′ side of the major branch point and adjacent to the minor branch point. (C, right) Site-specific cleavage of the Ad2 E1A intron DNA coding strand: lane 1, the pSP64-E1A fragment in the absence of both irradiation and $Rh(DIP)_3^{3+}$; lane 2, the pSP64-E1A fragment after irradiation in the absence of $Rh(DIP)_3^{3+}$; lane 3, Maxam–Gilbert G reaction; lane 4, Maxam–Gilbert T + C reaction; lane 5, the pSP64-E1A fragment after irradiation for 1 min in the presence of 10 μM $Rh(DIP)_3^{3+}$; lane 6, the pSP64-E1A fragment after irradiation for 2 min in the presence of 10 μM $Rh(DIP)_3^{3+}$. Intense cleavage (arrow) by $Rh(DIP)_3^{3+}$ is observed at the G four bases to the 5′ side of the branch point.

TACTG*A*TTCT*AA*T-3′, where the site cleaved is given in boldface; the major branch point (italicized) is six nucleotides away from the cleavage site, and the minor branch point (also italicized) is adjacent to the cleavage site. Also, a still stronger cleavage site is apparent on the sequence 5′-TGTCTACAG-TAAGTGAA-3′ which corresponds precisely to the 5′ end of the small t-antigen intron and on the sequence 5′-GTATTT-TAGATTCCAAC-3′ which corresponds precisely to the 3′ end of the large T-antigen and small t-antigen introns. The cleavage intensities near the branch point and at the acceptor site are low but are significant in that they are not observed in the control experiments in the absence of rhodium or light, and this metal-specific cleavage is consistently reproduced in multiple experiments. All cleavage experiments were conducted under single-hit conditions. Within the coding strand of the E1A intron (Figure 3C), strong cleavage is observed within a substantially different sequence, 5′-GTTTTGTG-GTTT*AAA*GA-3′, where the site cleaved is highlighted; the branch point (italicized) is four nucleotides to the 5′ side. Although weak cleavage is evident at the acceptor site, the experimental design in mapping of the Ad2 E1A intron did not facilitate examination at high resolution of rhodium cleavage at its 5′ end. In both cases (data not shown) no specific cleavage is evident on the non-coding strand. On both fragments only guanines are cleaved, but not all guanines;

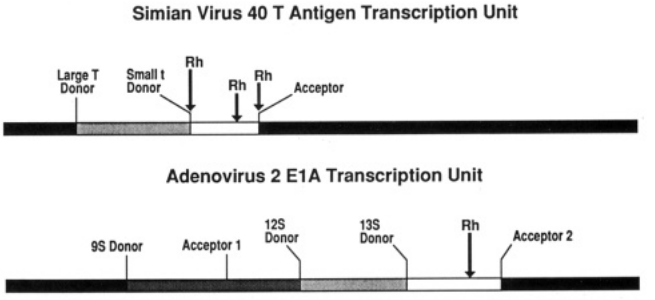

FIGURE 4: Schematic illustration of the results from high-resolution structural mapping of introns with $Rh(DIP)_3^{3+}$. The two genes are represented, roughly to scale, by the bars. Solid shaded elements represent exons; the unshaded elements, the introns of interest. The sites of cleavage by the rhodium complex are marked. On the T-antigen gene the cleavage is seen most strongly at the donor site and less strongly at the acceptor site and six bases upstream from the branch site. On the E1A gene a strong cleavage is apparent four bases from the branch site; the acceptor site is also cleaved here, though to a much lesser degree. The sequences within these regions differ substantially, however. The shape-selective metal complex appears, therefore, to mark distinct structures at functionally but not sequentially equivalent sites within the DNA introns.

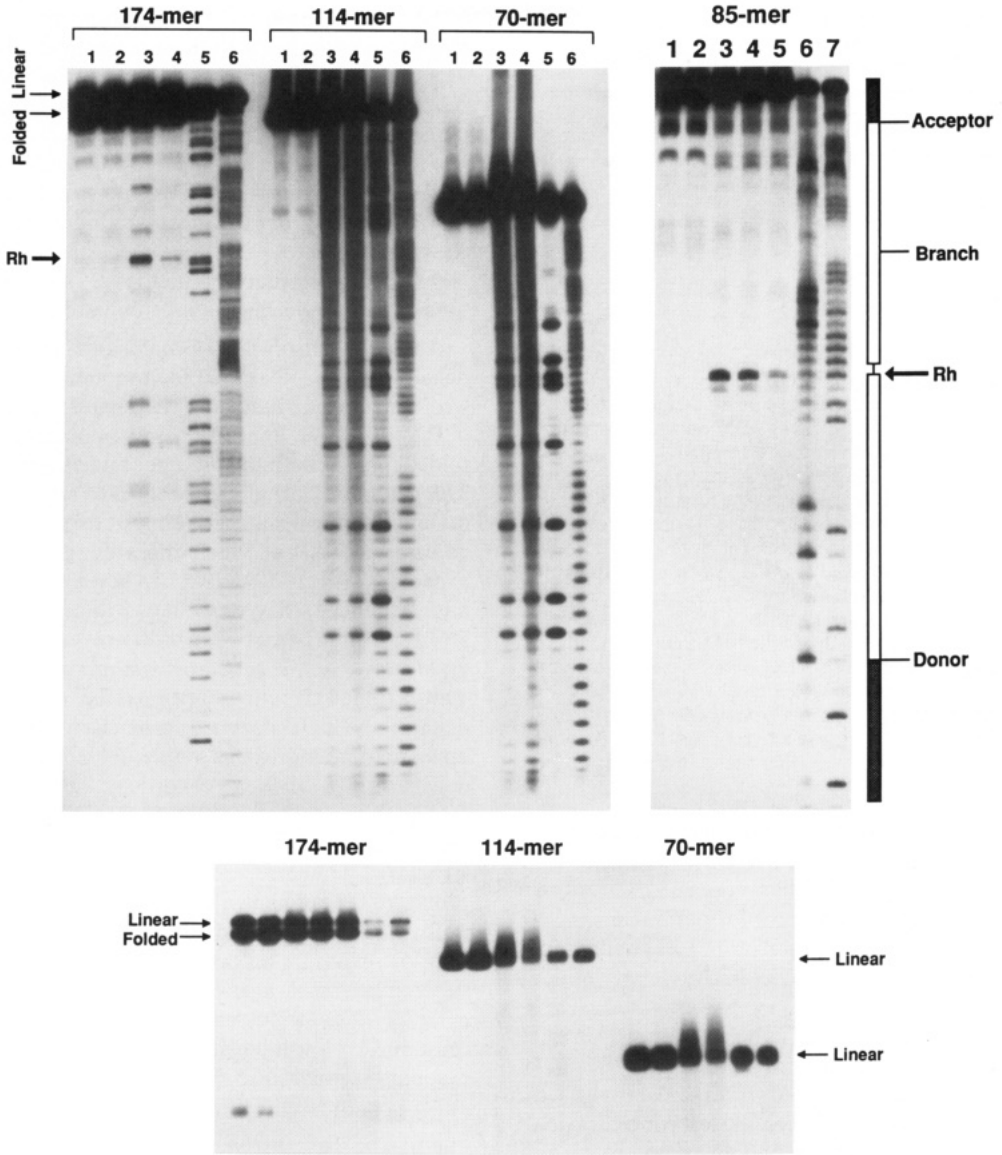

FIGURE 5: Structural probing of single-stranded DNA fragments of the E1A intron. Single-stranded DNA fragments containing elements of the intron and neighboring flanking sequences of the E1A gene were constructed to establish the requirements for structural targeting by $Rh(DIP)_3^{3+}$. (A, top left) The 174-mer, which consists of the entire intron plus exon flanking sequences at both ends, is cleaved specifically by $Rh(DIP)_3^{3+}$ (lanes 3 and 4) while the 114-mer (entire intron but with no flanking sequences) and the 70-mer (3′ two-thirds of the intron) are not cleaved specifically (lanes 3 and 4). Cleavage on the 174-mer occurs at the nucleotide four bases upstream from the branch point, as was found on the supercoiled double-stranded plasmid. Note also with the 174-mer but not the shorter fragments that a diffuse band is apparent which migrates with a faster mobility than a denatured 174-mer; this band likely corresponds to the folded form. Lanes 1, fragment incubated with $Rh(DIP)_3^{3+}$ in the absence of irradiation; lanes 2, fragment after irradiation in the absence of $Rh(DIP)_3^{3+}$; lane 3 for the 174-mer, fragment after irradiation in the presence of $Rh(DIP)_3^{3+}$; lane 4 for the 174-mer, fragment after irradiation in the presence of $Rh(DIP)_3^{3+}$ and 2 mM $MgCl_2$; lanes 3 for the 114-mer and the 70 mer, fragment after irradiation for 2 min in the presence of $Rh(DIP)_3^{3+}$; lanes 4 for the 114-mer and the 70 mer, fragment after irradiation for 4 min in the presence of $Rh(DIP)_3^{3+}$; lane 5, Maxam–Gilbert G reaction; lane 6, Maxam–Gilbert T + C reaction. (B, bottom) Low-exposure autoradiogram showing the migration patterns of the three ssDNA fragments in a denaturing polyacrylamide gel. The 174-mer runs as two separate bands. The two bands can be excised and the DNA isolated and sequenced to show identity. The DNA from the two bands, when reloaded on a gel separately, each gives rise again to two bands, indicating that under the gel running conditions the folded form of the fragment is in equilibrium with the unfolded form. (C, top right) $Rh(DIP)_3^{3+}$ photocleavage of the 85-mer. The 85-mer consists of the two ends and flanking sequences of the intron, including the branch point, covalently linked. This ssDNA is cleaved specifically seven nucleotides upstream from the cleavage site of the 174-mer, which is ten nucleotides upstream from the branch site. Lane 1, fragment incubated with $Rh(DIP)_3^{3+}$ in the absence of irradiation; lane 2, fragment after irradiation but in the absence of $Rh(DIP)_3^{3+}$; lane 3, fragment after irradiation in the presence of $Rh(DIP)_3^{3+}$; lane 4, fragment after irradiation in the presence of $Rh(DIP)_3^{3+}$ and 2 mM $MgCl_2$; lane 5, fragment after irradiation in the presence of $Rh(DIP)_3^{3+}$ and 4 mM $MgCl_2$; lane 6, Maxam–Gilbert G reaction; lane 7, Maxam–Gilbert T + C reaction.

the reaction is not simply base-specific but must reflect a structure being recognized.

Therefore, *within both intron DNAs functionally important sites on the coding strand are specifically targeted by the metal complex*. On the T-antigen intron DNA, the donor site is cleaved strongly, and to a lesser extent the acceptor site and a site adjacent to the branch point are also cleaved. On the E1A intron DNA, the strongest site cleaved is the site adjacent to the branch point, where functionally the lariat forms in the splicing of the mRNA. The fact that the relative intensities for cleavage near the branch point compared to the intron termini differ for the two DNA fragments may reflect subtle differences in the structures formed. These sites lack sequence homology; however, they are similar in that they demarcate functionally important elements on the RNA level. These observations are schematically presented in Figure 4.

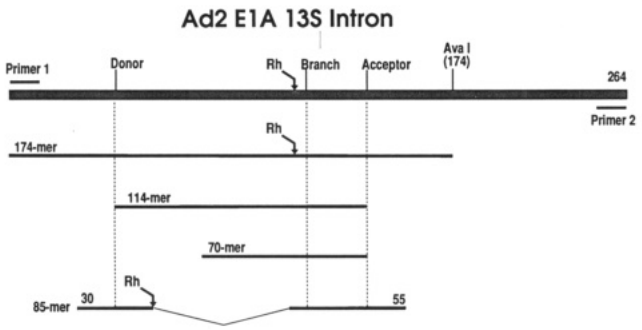

FIGURE 6: Schematic illustration of results from structural probing of the single-stranded DNA fragments of the E1A intron coding strand. The shaded bar at the top represents a portion of the E1A gene around the 13S intron. The cleavage site on the supercoiled plasmid, four nucleotides upstream from the branch point, is marked Rh. The 174-mer, shown next as a solid line, consisting of the entire intron plus flanking sequences, is cleaved at the identical site, while the 114 mer (the intron from end to end) and the 70-mer (two-thirds of the intron at the 3′ end) are not specifically cleaved. The 85-mer, which consists of the covalently linked ends of the intron plus flanking sequences, is cleaved at a site seven nucleotides upstream from that on the supercoiled plasmid and the 174-mer. The results indicate that the structure targeted on the supercoiled DNA corresponds to that of the folded single-stranded 174-mer. Specific cleavage by the 85-mer in the vicinity of the site targeted on the 174-mer suggests that it folds into a similar but perhaps more loosely packed structure. Targeting by $Rh(DIP)_3^{3+}$ of the intron site requires the borders of the intron and its flanking sequences, rather than the central portion of the intron.

That the sequences cleaved within these functional domains are so different points to the targeting of a structure(s) encoded by these sequences rather than of the sequence itself. The notion that this recognition would be conformation- or shape-selective is reasonable since the rhodium probe lacks hydrogen-bonding functionalities. The sequences in the vicinity of the sites targeted, as well as the DNA cleavage patterns themselves, eliminate the conformation targeted as being primarily in the A- or Z-form, uncoiled, or folded into a cruciform. Several experiments to characterize the folded structure were carried out using other chemical probes. When the salt concentration was varied, cleavage by the rhodium complex was found to reach a maximum between 25 and 50 mM; this, too, contrasts targeting of the cruciform site (Kirschenbaum et al., 1988). Diethyl pyrocarbonate, dimethyl sulfate, and OsO_4 were also used to examine the SV40 intron structure at high resolution, but no specific hypersensitivities were apparent. The regions adjacent to the $Rh(DIP)_3^{3+}$ cleavage sites were modified, though not strongly, by the chemical reagents. It is likely that the folded structure is too tight for the chemical reagents to react with the protected bases.

Probing the Structure of the ssDNA Fragments Corresponding to the Coding Strand of the E1A Intron. Given the sensitivity of rhodium cleavage for the coding strand, we next examined whether the DNA coding strand itself would be sufficient to generate the structure specifically recognized by the rhodium complex. Initially, three single-stranded DNA fragments (174-mer, 114-mer, and 70-mer) of the E1A 13S intron were prepared, since particularly strong cleavage had been observed on the E1A intron at high resolution. The 174-mer contained the entire intron plus 33 nucleotides of flanking sequences at the 5′ end and 27 nucleotides at the 3′ end. The 114-mer contained the intron from end to end, and the 70-mer contained the 3′ two-thirds of the intron including the branch point and the site of rhodium complex cleavage in double-stranded DNA. These single-stranded sequences were end-labeled with ^{32}P and then photolyzed in the presence of $Rh(DIP)_3^{3+}$.

The results from cleavage of these single-stranded DNA fragments with the rhodium complex are shown in Figure 5A. The two shorter fragments show a nonspecific G reaction, but the longer one containing the intron plus flanking exon sequences is cleaved specifically at exactly the same nucleotide position as that on the supercoiled double-stranded DNA. Several cleavage sites are apparent on this single-stranded DNA fragment, but the most striking is the site neighboring the branch point. This result suggests strongly that the 174-mer folds into a structure which is very similar, if not identical, to the structure within the supercoiled double-stranded DNA.

A folded form of the 174-mer is furthermore evident on the denaturing gel. No similar high-mobility bands were apparent with the shorter fragments. This band is more clearly evident in Figure 5B in a lightly exposed autoradiogram showing the folded form in both the presence and absence of $Rh(DIP)_3^{3+}$. This gel shows two bands, one of which migrates at the expected mobility for a denatured, single-stranded 174-mer but, in addition, another of which migrates at a faster mobility than expected. When these two bands were excised from the gel and sequenced, they were shown to be identical. When the DNAs from these two bands are electrophoresed again, separately, each regenerates the mixture of the two bands. Thus a folded structure appears to be in equilibrium with the denatured, single-stranded form. Remarkably, the 174-mer appears to maintain this stable, folded structure even under denaturing (8 M urea) conditions. The observation of the folded form in the absence of $Rh(DIP)_3^{3+}$ shows that the metal complex is not necessary to induce the folding of the structure. However, experiments using alternate chemical probes (data not shown) suggest that $Rh(DIP)_3^{3+}$ preferentially stabilizes the folded structure. It is noteworthy also that low-exposure autoradiograms (not shown) of the SV40 supercoiled DNA also show a high-mobility band that may correspond to a folded form; this may account for the reproducibly low intensity of cleavage seen in high-resolution mapping compared to that seen in low-resolution mapping.

Importantly, the 174-mer, but not the 114-mer corresponding to the full intron, folds into this discrete structure. Therefore, the intron DNA itself is not sufficient to form the structure; sequences flanking within the exon are also required. It is noteworthy that flanking exon sequences also appear necessary in splicing (Fu et al., 1991; Newman & Norman, 1992).

To examine the minimum sequences needed for folding of the structure, we then synthesized another single-stranded DNA fragment that consisted of sequences at the 5′ end and 3′ end of the intron including the branch point and 15 nucleotides of flanking sequences at either end but with the sequences in the middle of the intron deleted. Again, as evident in Figure 5C, this DNA was found to be cleaved specifically by the rhodium complex. In this case, however, the cleavage observed is seven bases to the 5′ side of the site cleaved on the supercoiled and 174-mer DNA fragments (or ten bases from the branch site). Cleavage now occurs at a T rather than a G and corresponds directly to the linkage point of the two partial fragments. The cleavage pattern here, furthermore, is not restricted to a single site but is dispersed over two bases. This difference of seven nucleotides is smaller than the size of the metal complex; if the positioning of the complex in the binding site were slightly shifted, such a change in cleavage position could arise. In addition, despite the strong rhodium cleavage, in this case no discrete band of higher mobility than the full fragment was evident on the denaturing gel. These observations, taken together, suggest that this fragment may

not fold as tightly into the intron structure as does the 174-mer, but it still folds so as to create a comparable, perhaps looser, recognition site for $Rh(DIP)_3^{3+}$. Figure 6 summarizes the results with the single-stranded fragments.

Functional Implications of the Structure. A distinct structure has therefore been identified and may be specifically targeted in the intron coding strand of SV40 T-antigen and Ad2 E1A genes. The structure, centered around the branch site of the introns, is supercoil-dependent, sensitive to variations in salt concentration, and weakly sensitive to modifications by chemical reagents. Experiments with the single-stranded DNA fragments show that the structure can be formed by the coding strand alone and is remarkably stable to denaturants and that sequences flanking the intron (i.e., sequences into the exons) are essential in forming the structure. The ends of the intron and the flanking sequences seem to play a key role in the proper folding of the structure. This structure may bear some relationship to the recently proposed Holliday-like structure for spliceosomes (Steitz, 1992), especially given the specific targeting of DNA Holliday junctions and cruciforms by $Rh(DIP)_3^{3+}$.

The results described here provide evidence that the position of an intron is encoded on DNA. These results may offer support for the exon shuffling theory, proposed by Gilbert (Gilbert, 1987; Hall et al., 1989), which postulates that genes evolved through the shuffling of exons which code for functional domains of proteins. Whether or not the DNA structure described here evolved from structures necessary for RNA splicing, such a DNA structure could now serve to demarcate the intron or exon boundary. A structural delineation of exon boundaries would facilitate the maintenance of the integrity of the exons during the shuffling process. The structure present in the two intron DNA regions described in this paper and specifically targeted by the shape-selective $Rh(DIP)_3^{3+}$ could be similarly recognized by enzymes involved in recombination. It is interesting to consider in this context that $Rh(DIP)_3^{3+}$ also targets cruciform structures which are important protein recognition elements in recombination. Certainly that intron positions are structurally delineated at the DNA level speaks to their functional importance.

ACKNOWLEDGMENT

We thank Prof. James L. Manley of Columbia University for the construction of the plasmids as well as for stimulating discussions at the outset of this project.

REFERENCES

Barton, J. K., & Raphael, A. L. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 6460–6464.
Berk, A. J., & Sharp, P. A. (1978a) *Proc. Natl. Acad. Sci. U.S.A. 75*, 1274–1278.
Berk, A. J., & Sharp, P. A. (1978b) *Cell 14*, 695–711.
Cech, T. R. (1990) *Annu. Rev. Biochem. 59*, 543–568.
Chow, C. S., & Barton, J. K. (1992a) *Methods Enzymol. 212*, 219–242.
Chow, C. S., & Barton, J. K. (1992b) *Biochemistry 31*, 5423–5429.
Fu, X.-D., Katz, R. A., Skalka, A. M., & Maniatis, T. (1991) *Genes Dev. 5*, 211–220.
Gilbert, W. (1987) *Cold Spring Harbor Symp. Quant. Biol. 52*, 901–905.
Green, M. R. (1986) *Annu. Rev. Genet. 20*, 671–708.
Hall, D. W., Liu, Y., & Shub, D. A. (1989) *Nature 340*, 574–576.
Keene, M. A., & Elgin, S. C. R. (1984) *Cell 36*, 121–129.
Kirshenbaum, M. R. (1989) Doctoral Dissertation, Columbia University.
Kirshenbaum, M. R., Tribolet, R., & Barton, J. K. (1988) *Nucleic Acids Res. 16*, 7948–7960.
Maniatis, T., & Reed, R. (1987) *Nature 325*, 673–678.
Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.
McCutchan, T. F., Hansen, J. L., Dame, J. B., & Mullins, J. A. (1984) *Science 225*, 625–628.
Müller, B. C., Raphael, A. L., & Barton, J. K. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 1764–1768.
Newman, A. J., & Norman, C. (1992) *Cell 68*, 743–754.
Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., & Sharp, P. A. (1986) *Annu. Rev. Biochem. 55*, 1119–1150.
Pyle, A. M., & Barton, J. K. (1990) *Prog. Inorg. Chem. 38*, 413–475.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sitlani, A., Long, E. C., Pyle, A. M., & Barton, J. K. (1992) *J. Am. Chem. Soc. 114*, 2303–2312.
Steitz, J. A. (1992) *Science 257*, 888–889.