DISTINCT INTRON DNA STRUCTURES IN SIMIAN VIRUS 40 T-ANTIGEN AND ADENOVIRUS 2 E1A GENES

Thesis by

Inho Lee

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1994

(Submitted February 19, 1994)

Acknowledgements

Of all the persons who have guided me into personal and professional maturity, my advisor, Prof. Jacqueline K. Barton, ranks second only to my mother in importance. Jackie has been patient throughout my long career as a graduate student. She has lifted me up from depths of depression and showed me that perseverance is perhaps the most important quality in a scientist. She has always encouraged me with her constant optimism and enthusiasm. I thank her for teaching me these important lessons and for the opportunities she has provided me.

I am grateful to Prof. James L. Manley of Columbia University for the gift of the plasmids and for interesting and helpful discussions at the outset of this project. I am also indebted to Dr. Suzy Chen of Columbia University for technical assistance in amplifying the plasmids.

I would like to thank past members of the Barton group who have helped me become a proficient technician; they include Mindy Kirshenbaum, Takashi Morii, Anna Marie Pyle, and Alan Friedman. I wish to thank the rest of past and present members of the Barton group for stimulating discussions, soul-searching conversations, hearty laughs, and good times together: Michelle Arkin, Donna Campisi, Christy Chow, Sheila David, Cindy Dupureur, Richard Hartshorn, Marilena Fitzsimons, Dan Hall, Brian Hudson, Yonchu Jenkins, Tim Johann, Scott Klakamp, Achim Krotz, Ai Ching Lim, Susanne Lin, Cathy Murphy, Michael Pustilnik, Niranjan Sardesai, Tom Shields, Ayesha Sitlani, Bob Terbrueggen, and Kaspar Zimmermann. I am also indebted to Mo Renta who has helped me in many different ways over the years.

My life as a graduate student would have been too trying at times without the support of my family and friends. I wish to thank my friends Bruce, Mark, and Elliot for their long and enduring friendship and for discussions of whys and wherefors of doctoral dissertations. Closer to lab, I would like to thank Ayesha, Niranjan, Ai Ching, and Michelle for their friendship. None of this would have been possible, of course, without my mother, for obvious reasons; but she has been much more. She has been constantly, unfailingly on my side, gently reminding me of the

importance of the process and not just the end. I thank her, among other things, for her wisdom. I would also like to thank my stepfather, my sister and brother-in-law, and my uncle and his family for their support. And finally there is one who deserves more than I could ever pay back; she has been patient and understanding and supportive throughout these years; she has endured more than anybody should; she has helped me persevere and grow in countless ways. I thank my wife Eve for all that she is and has done for me. In addition, I wish to acknowledge an emotion too deep and complicated to be labeled thanks, an emotion I feel towards my son Torin who contributed plenty to my long career as a graduate student, for which I feel no regrets, only goodness and rightness.

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 fragments corresponding to the coding strands of the genes were also found to fold into a structure apparently identical to that in the supercoiled genes based on the recognition by the metal complex. Further investigation of different single-stranded DNA fragments with other structural probes, such as another metal complex bis(1,10-phenanthroline)(phenanthrenequinone diimine)rhodium(III), AMT (4'-aminomethyl-4,5',8-trimethylpsoralen), restriction enzyme Mse I, and mung bean nuclease, showed that the structures require the sequences at both ends of the intron plus the flanking sequences but not the middle of the intron. The two ends form independent helices which interact with each other to form the global tertiary structures. Both of the intron structures share similarities to the structure of the Holliday junction, which is also known to be specifically targeted by the former metal complex. These structures may have arisen from early RNA intron structures and may have been used to facilitate the evolution of genes through exon shuffling by acting as target sites for recombinase enzymes.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1.

Chapter 2.

LIST OF TABLES

Chapter 1.

Structures in intron DNA as an indication of a possible functional significance of introns: Comparison to other known DNA structures

1.1 Introduction

The discovery of introns was a major event in the history of biology (1). It overturned the dogma of the gene as a discrete and contiguous unit of DNA and ushered in a new era of biology with much attention being focused on RNA processing. Introns are stretches of DNA that separate the coding sequences (exons) of a gene. The RNA after transcription contains both the intron and the exons. Introns are spliced out of these RNA molecules to give rise to mature mRNA molecules. The mechanism of RNA splicing (schematically illustrated in Figure 1.1), a process which involves a myriad of factors and precise splicing of the folded RNA molecule, has been very well studied to date (2). However, it still remains a mystery why introns exist at all. What possible function(s) could introns serve? A plausible answer to this question is the exon theory of genes proposed by Walter Gilbert (3). This theory proposes that exons code for discrete functional domains of proteins and that the introns serve as spacers separating these functional gene segments. By being thus separated the exons can be recombined ("shuffled") to produce new proteins with different functional domains. For such shuffling to occur, the exons would have to be somehow demarcated so that the enzymes responsible for recombination can recognize their boundaries. One mechanism of demarcation would be through a structure distinctly different from the regular double helix of DNA. Thus, finding a structure in the intron-exon boundaries would be a first step to showing a possible function of the introns. Evidence is presented in this thesis that shows the DNA in the intron-exon boundaries of two viral genes can form distinct structures *in vitro*.

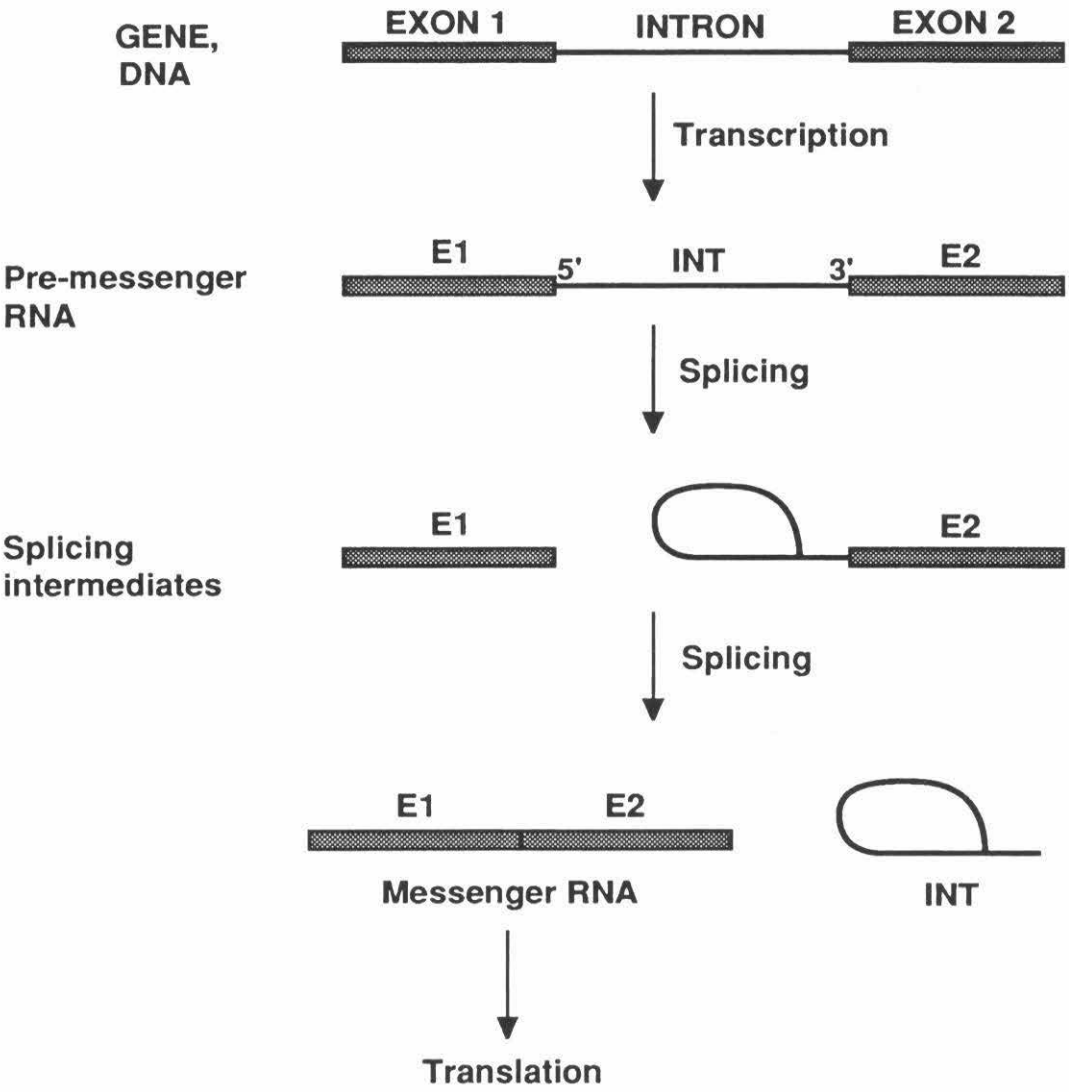


Figure 1.1 A schematic illustration of steps involved in nuclear pre-mRNA splicing. The gene consists of exons and introns. Transcription generates a pre-mRNA containing the exons and the introns. Splicing takes place within the spliceosome complex and produces the mature mRNA and the lariat intron.

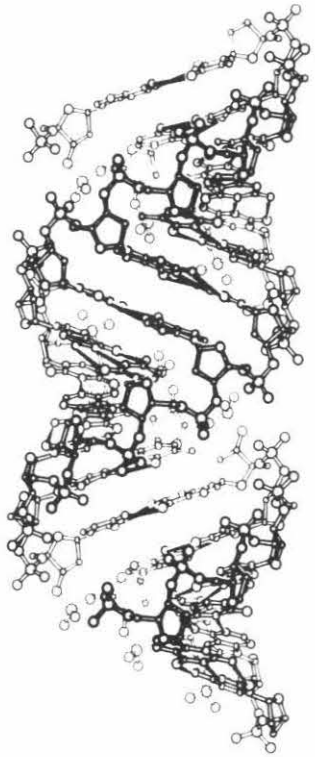
1.2 Known DNA Structures

Several structures in DNA have already been discovered. These are Z-DNA, cruciforms, Holliday junctions, H-DNA, bent DNA, and telomeric quadruplexes. Although the physiological importance of these structures have not been conclusively proven, their existence *in vitro*, and indirect evidence for their existence *in vivo*, suggest a possible function for them in processes such as replication or transcription.

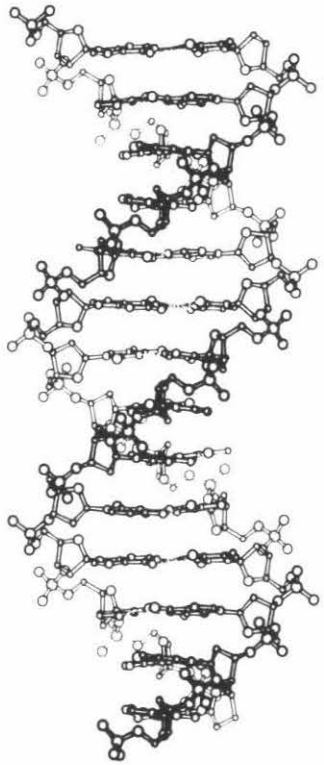
Z-DNA was first characterized by X-ray crystallography by Rich and coworkers (4). It is a left-handed double-helix rather than a right-handed one as in B-form DNA. Its sugar-phosphate backbone follows a zig-zag pattern along the helix. Its major groove is shallow and wide, so much so that it can hardly be called a groove, and its minor groove is deep and narrow. It is formed by alternating purine-pyrimidine sequences, primarily by alternating GC. Such sequences were found in the Simian Virus 40 transcriptional enhancer region (5), suggesting a role for them in regulation of transcription. More recently it has been found that transcription of the human c-myc gene is associated with formation of Z-DNA in three discrete regions of the gene as assayed by anti-Z antibodies (6). The implication of this finding is that the three segments capable of assuming the Z-conformation do so at different stages of transcription and thus regulate its process by interacting or interfering with protein factors. Such findings lend more support for the hypothesis that Z-DNA, dynamically formed under the stress of transcription, may play a role in regulating transcription.

Cruciforms can be formed from palindromic sequences of DNA and have been found to form *in vitro* in supercoiled plasmids such as pBR322 (7). Its structure in two dimensions resembles a cross, hence the name cruciform. It is formed by one continuous strand of DNA which first unwinds and then re-anneals so that each

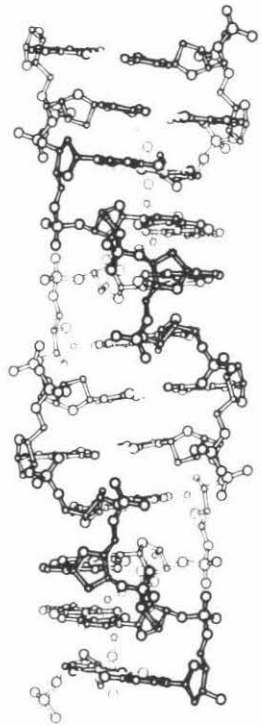
Figure 1.2 A-, B-, and Z-form DNA. Side views and top views of computer generated models of the three forms of DNA double helix are shown. Adapted from W. Saenger, *Principles of Nucleic Acid Structure*, Springer Verlag, New York, 1984.



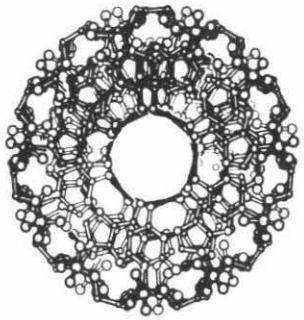
A-DNA

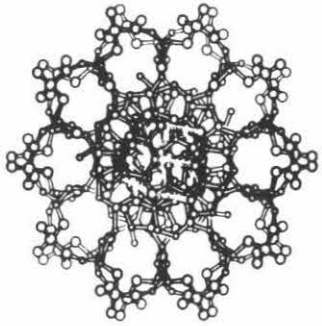


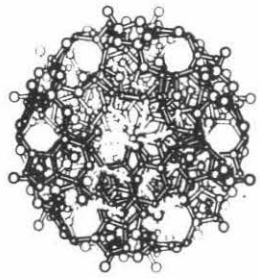
B-DNA



Z-DNA







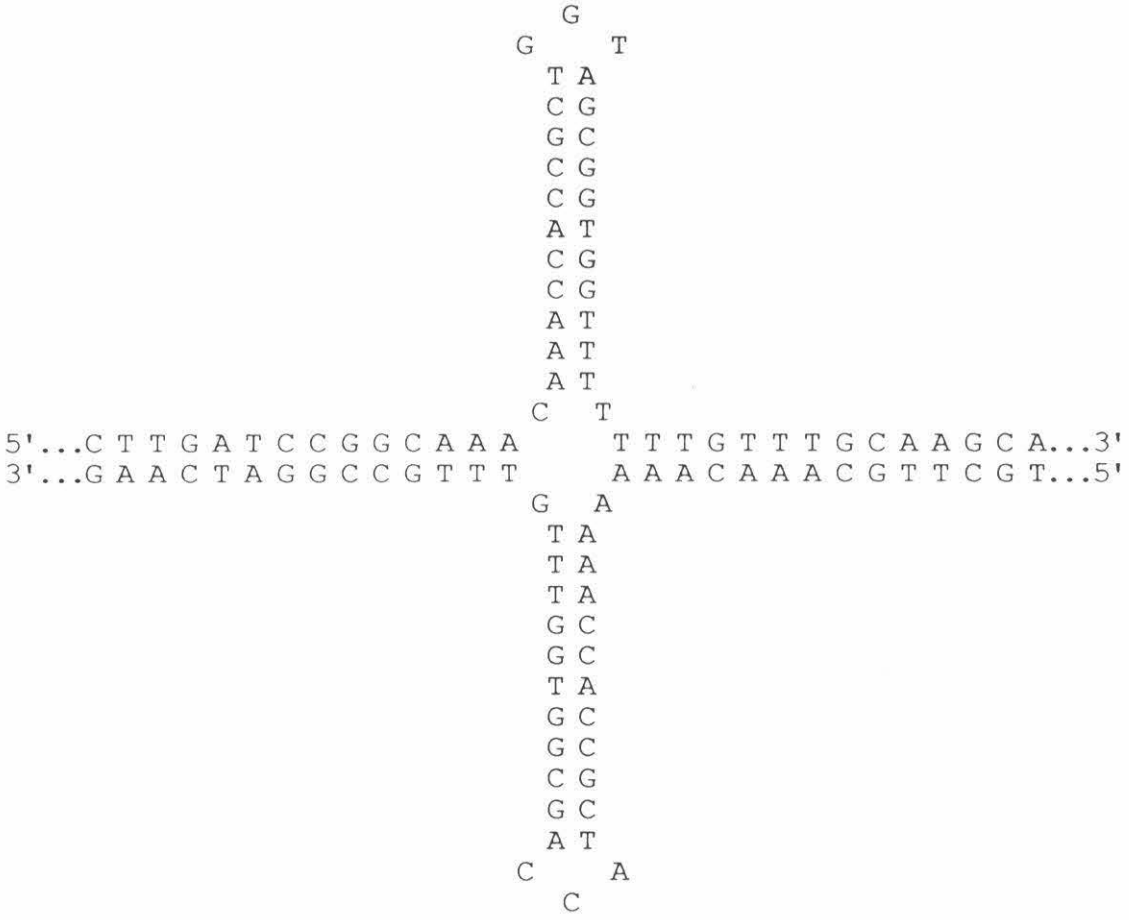


Figure 1.3 A two-dimensional representation of a cruciform structure from plasmid pBR322.

strand separately forms new duplexes that jut out from the main duplex. The tips of the protruding duplexes are single-stranded loops. Single-strand specific nucleases were used to probe the loops which arise as a result of the cruciform formation in supercoiled plasmids. Supercoiling was necessary for the site-specific cleavage of the plasmids, indicating that the stress of supercoiling was driving the formation of the cruciform. It has also been shown that transcription *in vitro* of yeast plasmids causes supercoiling (8), which leads to the hypothesis that cruciform may form during transcription *in vivo*. And this has recently been shown to be the case (9). So it appears that another structural perturbation of DNA may also be associated with transcription, possibly playing a role in its regulation.

Holliday junctions are similar in structure to cruciforms in that there are four helices branching out from the "junction," but, unlike cruciforms, which are formed by one continuous stretch of DNA, Holliday junctions are formed by the joining of two separate helices. The structure was proposed to be an intermediate in recombination events (10). There have not yet been any hard evidence to support the existence of Holliday junctions *in vivo*; however, synthetic Holliday junctions have been well characterized by Lilley and coworkers (11). The synthetic junctions prefer a certain way of stacking the helices (referred to as "stacked X") depending on the bases at the center of the junctions, so that they stay in one conformation or the other, which can be observed by their different migrations through non-denaturing polyacrylamide gels. More recently, they have found naturally occurring enzymes which cut the synthetic junctions at specific positions *in vitro*, greatly strengthening the case for natural existence and potential importance of Holliday junctions (12).

H-DNA, or hinged DNA, is formed by naturally occurring sequences of alternating TC steps with the corresponding AG on the other strand (13). Part of this segment of repeated sequences unwinds and forms a triple-stranded helix with the

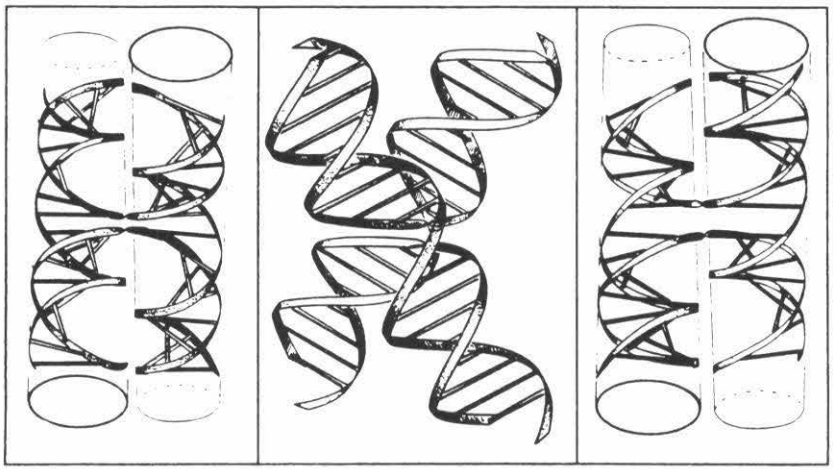


Figure 1.4 Three views of the right-handed, antiparallel stacked-X structure of a synthetic Holliday junction. In the face view the X-shape of the overall structure formed by the two helices is evident. The major and minor groove sides are so labeled based on the type of groove present on either side of the face at the center of the junction. Adapted from D. M. Lilley & R. M. Clegg (ref. 11).

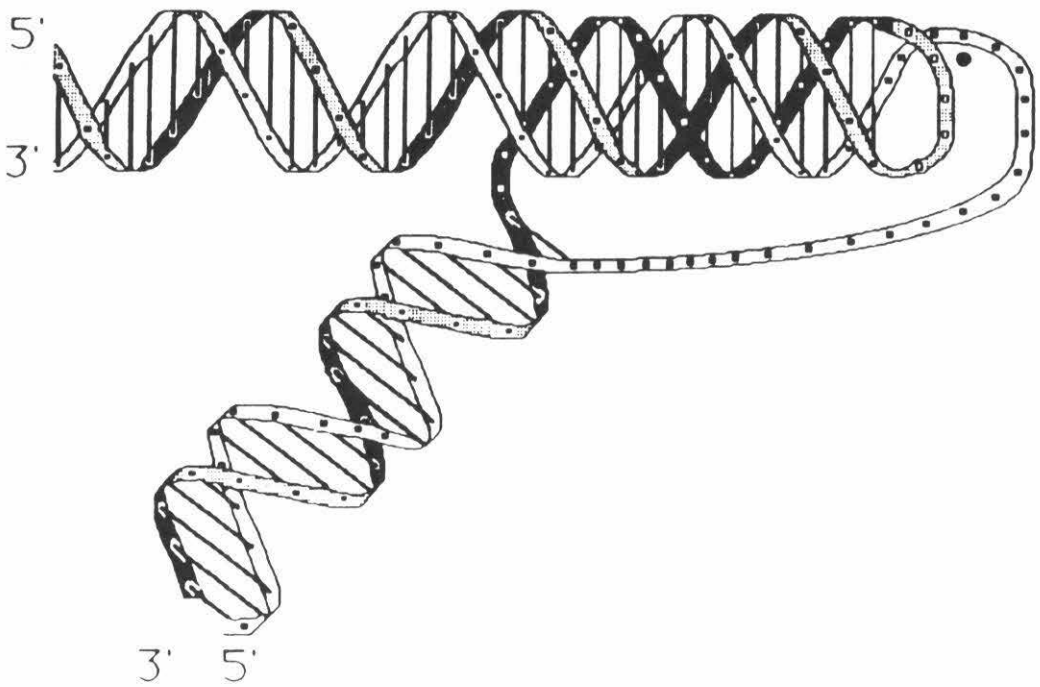


Figure 1.5 Three-dimensional model for H-DNA. In the upper-left portion of the model, the triple helix and the lone single-strand is evident. The structure causes a sharp kink in the direction of the overall DNA molecule. Adapted from H. Htun & J. E. Dahlberg (ref. 13a).

other half of the segment, leaving the complementary strand single-stranded. The resulting structure produces a sharp kink in the overall stretch of DNA in which it occurs, leading to its "hinged" appearance. The formation of this structure was observed to relax negative supercoils in the plasmid in which the sequence occurs, suggesting again that DNA structures form under the stress of supercoiling and thus may play a role in processes that generate supercoils. However, H-DNA has not been shown to exist *in vivo* and the significance of the structure still remains unknown.

Bent DNA is formed by poly-A tracts of more than 5 residues (14). The spacing of the poly-A tracts along the DNA helix appears to be important in producing the overall curvature of DNA (15). When the A tracts are separated by a full helical period, i.e., ten to eleven bases, or multiples thereof, from the beginning of one segment to the beginning of another segment, the stretch of DNA containing these segments achieves a global curvature. When the separation is out of phase with the helical repeat, no curvature is observed. This indicates the curvature produced by the A tract is toward one groove, the minor groove (14), and therefore proper spacing of the A tracts produces repeated curvatures toward one side of the helix, which ultimately leads to the global curvature. The most striking occurrence of bent DNA is in kinetoplast minicircles of trypanosomes (16). The minicircles contain a segment of about 200 base-pairs which are highly curved; these segments form almost a complete circle and can be seen as such under the electron microscope (17). Bent DNA has also been observed in the origins of replication of several different systems such as SV40 (18), yeast (19), and bacteriophage lambda (20). SV40 has also been shown to contain bent DNA at the termini of replication and transcription (21). Occurrence of bent DNA at such functional sites suggests a functional role for the structure, such as a recognition site for proteins factors and/or enzymes, which is

Figure 1.6 Schematic illustration of bent DNA. The overall curvature is generated by in-phase repeat of poly A tracts. Adapted from L. E. Ulanovsky & E. N. Trifonov (1987) *Nature* **326**: 720-722.

indeed postulated for these systems. For example the bent site in the origin of replication of the SV40 genome is postulated to be a binding site for a dimer of the large T-antigen, which is a regulator of the replication process (18).

Quadruplexes occur at the ends of chromosomes, which are termed telomeres (22). The telomeres contain an unusual repeat of sequences of the type $G_4T_{2\text{-}4}$ which form a structure known as the G-quartet (23). At the ends of the chromosomes there are single-stranded copies of the G repeat (22). These ends are postulated to fold back on themselves so that the guanines pair with one another to form a short four-stranded structure. Monovalent cation is a requirement for this structure. It is thought that the cation stabilizes the structure by coordinating to the O6 of the guanines in the center of the quadruplex (23). Crystal (24) and solution (25) structures of the G-quartet have been solved since the first proposal of the structure, and they have confirmed the earlier models based on biochemical studies. It is not known for sure whether the G-quartet is formed by one strand at the end of a chromosome or if the ends of two chromosomes, in a hair-pin structure of G-G pairs, come together in an antiparallel fashion to form the G-quartet. In either case, the structure is thought to be a component in maintaining the ends of chromosomes which may otherwise be exposed to damage or degradation (22).

Less defined structural polymorphisms have also been observed in interesting regions of genomes: mung bean nuclease was shown to cleave at sites before and after genes in *Plasmodium* (26); in *Drosophila* micrococcal nuclease cleaved in spaces surrounding the genes but not in the protein coding regions (27). These nucleases are single-strand specific, and their targeting of regions between genes suggests perturbations in the conformation of DNA that include single-stranded regions. Whether these are distinct and stable structures have not been investigated, and the low level of resolution in these studies preclude any specific discussion of these

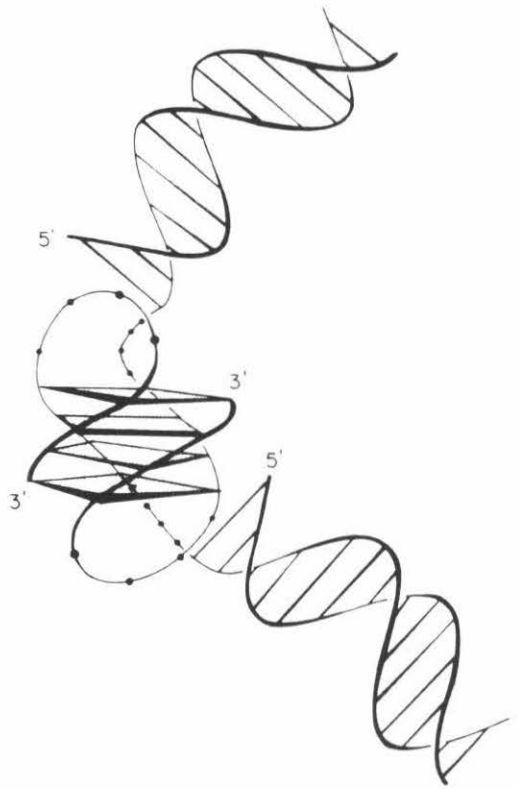


Figure 1.7 A schematic illustration of a quadruplex structure of telomeres. The diagram shows the manner in which two telomeres may be linked together by the quadruplex structure. The shaded planes represent the G-quartets, and the solid dots represent thymine residues. Adapted from C. H. Kang, et al. (ref 24).

structural perturbations. However, the obvious conclusion one can draw from these observations is that DNA structure is not uniform along the long stretches of a genome and that structural polymorphisms exist at functionally significant sites. Whether these structural perturbations serve a real function is a question that is still being addressed by the various studies on DNA structures.

All of the above examples make it amply clear that DNA is not a static molecule of stored information as initially thought but a dynamic molecule able to assume various unusual structures. These structures may be the points of interactions with many protein factors in diverse set of processes, most important and common of which are replication and transcription. It appears that the information content of DNA is not limited to its sequence of bases but includes the various perturbations of its conformation, which arise as a consequence of certain patterns in its primary sequence. It seems likely that evolution would have taken advantage of the varieties of DNA structures and developed mechanisms of utilizing them, so it is reasonable to hypothesize that these structures carry out some function in the cellular processes. Evidence for this case is getting stronger, but we still lack conclusive proof. Such proof will be hard to obtain; in the meantime, however, more and more interesting DNA structures are being discovered.

1.3 Probes of DNA structures

For all of the structures mentioned above only the structure of Z-DNA and the G-quartet have been elucidated by x-ray crystallography. Other structures do not lend themselves easily to crystallization, partly due to their complexity and partly due to their dynamic nature. Therefore, these structures are deciphered with enzymatic and chemical probes, of which there are a few. These probes take advantage of the differences in conformations displayed by the unusual structures.

They target single-stranded regions, parts of the bases that are normally inaccessible, or double-stranded regions that may have formed as a result of the unusual structure. They do not give very detailed information on the structures, and sometimes the results can be inconclusive; but used together they give enough useful data to arrive at plausible structures.

1.3.1 Enzymatic probes

These are mainly single-strand specific nucleases such as S1 nuclease or mung bean nuclease. Though simple in their action these enzymes can provide very valuable data on the nature of the DNA structure in question: used carefully they cleave only the single-stranded regions of the structure, thus allowing us to determine loops and bulges. S1 nuclease has been used to probe cruciforms and H-DNA (7, 13), and mung bean nuclease has been used in probing global structural patterns in the genome of *Plasmodium* (26).

Restriction endonucleases can be used to probe for double-stranded regions in unknown structures. A prediction of a double strand has to be made first, and the double strand must contain a suitable site for a restriction enzyme for this technique to be useful. The integrity of a tDNA structure has been demonstrated with this technique (28).

1.3.2 Chemical probes

Diethyl pyrocarbonate (DEPC): This reagent is specific for purines; it carbethoxylates the N7 position on the purines. The accessibility of the N7 of purines determines the degree of reactivity to DEPC. Thus, a perturbed structure, such as Z-DNA, in which the purines are in the syn conformation, will show increased

sensitivity to DEPC. Z-DNA within plasmids has indeed been shown to be hyper-reactive to DEPC, thus establishing the usefulness of this probe (29).

Osmium tetroxide: In the presence of pyridine OsO_4 covalently adds across the 5,6 double bond of pyrimidines (30). The accessibility of the double bond is again the determining factor of reactivity. Single-stranded regions or other highly perturbed structures would lead to solvent accessibility of the double bond and thus to hyper-reactivity toward OsO_4. Structural distortions at B-Z junctions have been probed successfully with OsO_4 by inserting a Z-forming sequence into an E. coli plasmid and treating the plasmid with OsO_4 (31).

Psoralens: This reagent acts by intercalating and then crosslinking thymines on opposite strands upon photoactivation at 340 nm (32). Though used more prevalently in determining RNA structures such as the ribosomal RNA (33), psoralen and its derivatives can be used to probe unknown DNA structures in which double-stranded regions are not clearly delineated. Shen and Hearst studied global structural characteristics of the linear single-stranded DNA genome of bacteriophage fd through psoralen crosslinking and analysis by electron microscopy and found that the crosslink was localized to one end of the genome believed to be the origin of replication (34). More recently Hopkins and coworkers examined the sequence preference of psoralen crosslinking in short double-stranded DNA fragments by cleaving crosslinked DNA with EDTA-Fe(II). The resulting fragmentation pattern was analyzed on denaturing polyacrylamide gels, and the site of crosslink was shown to have reduced intensity of cleavage (35).

EDTA-Fe(II) and MPE-Fe(II): Iron chelated to EDTA (ethylenediamine tetraacetate) is the functional center in these reagents. In MPE-Fe(II) the intercalator methidium is tethered to the EDTA moiety (36). Ferrous ion with peroxide undergoes Fenton chemistry, generating hydroxyl radicals which attack the sugar-

Table 1.1 Enzymatic and chemical probes of DNA structures

Probe	Target	Structures probed	References
S1 nuclease	Single-stranded DNA	Cruciforms, H-DNA	Panayotatos & Wells (7) Htun & Dahlberg (13)
Mung bean nuclease	Single-stranded DNA	Genome of *Plasmodium*	McCutchan, et al. (26)
Restriction endonucleases	Specific double-stranded sites	tDNA	Paquette, et al. (28)
Diethyl pyrocarbonate	N7 of purines	Z-DNA	Herr (29)
Osmium tetroxide	5,6 double bond of pyrimidines	B-Z junctions	Nejedly, et al. (31)
Psoralen	Thymines on opposite strands	Bacteriophage fd	Shen & Hearst (34)
EDTA-Fe(II)	Sugar-phosphate backbone	Bent DNA	Tullius, et al. (37)

phosphate backbone of DNA, leading to strand scission. As expected for such a reaction, there is no specific sequence preference for cleavage. This property makes them more useful as footprinting reagents rather than structural probes. However, they can also be used to obtain information on various structural perturbations of DNA. Tullius and coworkers have used EDTA-Fe(II) to probe structural features of bent DNA, among others (37). They observe that there are variations in the intensity of EDTA-Fe(II) cleavage when certain structures are present; the structure makes certain portions of it less accessible to the hydroxyl radicals and thus are cleaved to a lesser extent than other regions. EDTA-Fe(II) has also been used in the same manner to successfully sketch a structure for a large RNA molecule, the self-splicing Tetrahymena pre-rRNA (38).

1.3.3 Transition metal complexes as probes of DNA structures.

Transition metal complexes present an ideal system for probing differences in the shape of the target DNA (39). The metal center provides a variety of coordination geometries for ligands of various shapes and functionalities. The rigidity of the coordination complex locks the overall shape of the complex, in two enantiomeric forms, delta and lambda. The rigidity of the complex is a crucial factor in being able to predict and interpret the data in structural probings. The shape of the complex is a fixed entity, and this makes it possible for them to recognize specific complementary shapes on DNA. The shape of the complex also depends on the shape of the ligands, and variations in the type of the ligands used will give a variety of complexes with different overall shapes. Different types of ligands can be mixed within a single complex, giving rise to new shapes and properties. This type of shape recognition occurs even in nature; the zinc-finger and other zinc motifs in transcriptional regulatory proteins take advantage of the coordination chemistry of zinc in forming

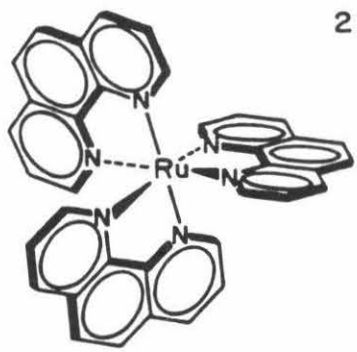
discrete structures around the metal center with cysteines and histidines, which are then able to recognize specific sequences of DNA (40).

Use of appropriate metal centers in the metal complexes enables the interpretation of the shape-selective binding to DNA. Ruthenium complexes have optical properties that depend very much on the extent of their binding to DNA and thus provide a spectroscopic handle on their interaction with DNA (41). Rhodium complexes, on the other hand, are photoreactive and cleave the DNA strand where they are bound upon activation with UV light (42). This property of the rhodium complexes makes them a very valuable tool for probing fine, and gross, structures of DNA at the nucleotide resolution, since the DNA can be radiolabeled before cleavage and analyzed on high-resolution polyacrylamide gels.

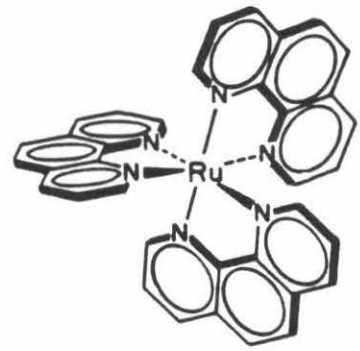
$[Ru(phen)_3]^{2+}$, tris(phenanthroline) ruthenium(II), one of the first metal complexes designed by Barton and coworkers, binds to DNA by intercalation or by surface-binding. Spectroscopic studies indicate that it exhibits enantioselectivity in intercalative binding to DNA, so that the delta isomer binds preferentially over the lambda isomer in the major groove of the right-handed double helix of DNA (43). This is so because the ancillary ligands which are not intercalated are necessarily oriented either along (in the case of the delta isomer) or perpendicular to (in the case of the lambda isomer) the direction of the groove. Here is an excellent illustration of the shape selectivity of these metal complexes: the shape of the complex complements the shape of the DNA, and the binding is thus facilitated, which can be detected using appropriate techniques.

$[Rh(phen)_2phi]^{3+}$, bis(phenanthroline)(phenanthrene quinone diimine) rhodium(III), is a highly sequence-selective photoreactive metal complex which has been used to probe fine structures in the major groove of DNA (44). The phi ligand is an avid intercalator (45) and provides a high DNA binding affinity. The ancillary

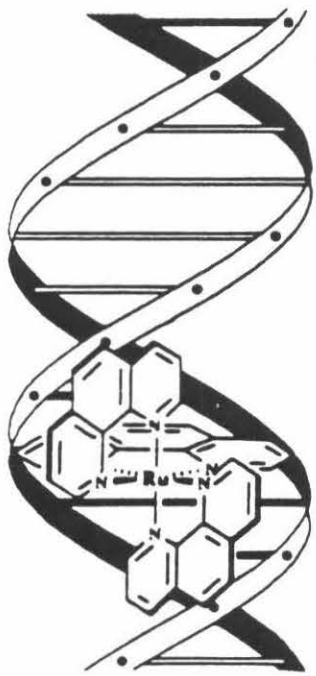
Figure 1.8 Intercalation of the Δ and Λ isomers of $[Ru(phen)_3]^{2+}$ to B-form DNA. Intercalation in the major groove is sterically favored for the D isomer due to the positioning of the non-intercalated ligands along the groove rather than perpendicular to it as in the case of the L isomer. Portions adapted from A. Sitlani (1993) Doctoral Dissertation, California Institute of Technology.



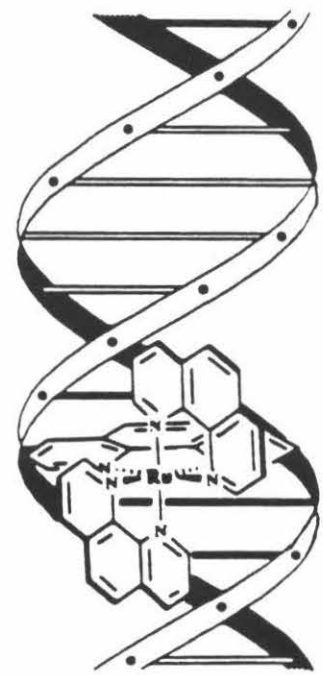
Λ



Δ

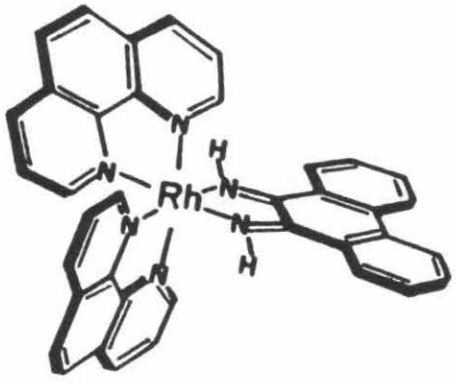


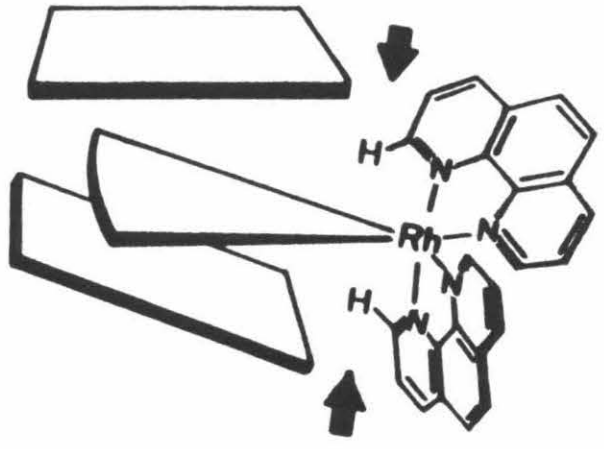
Λ



Δ

Figure 1.9 Shape selective recognition of DNA by $[Rh(phen)_2phi]^{3+}$through intercalation. The steric constraints imposed by the ancillary phenanthroline ligands promote intercalation of this complex between base pairs whose propeller twisting is such that the major groove is more open than the canonical B-form helix.


Rh
N
H

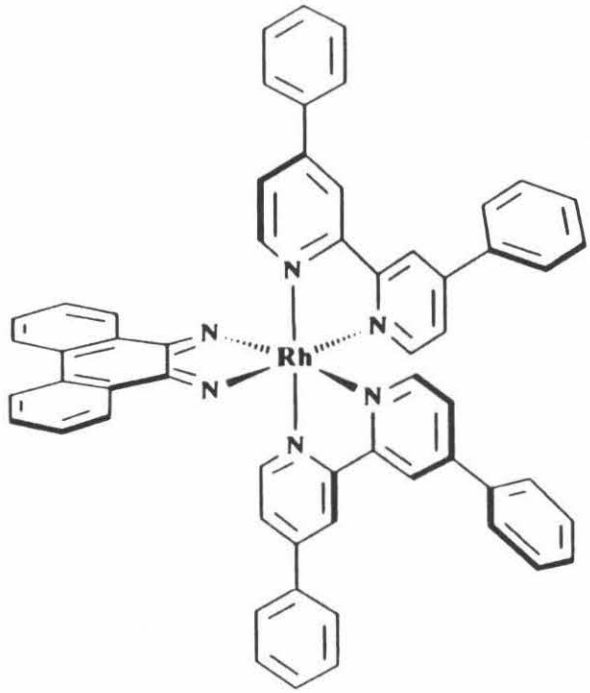

H
N
Rh
N
H
N

phen ligands, however, provide a steric bulk above and below the phi ligand and hinder the intercalation at sites where the distance between the base pairs is short due to propeller twisting. Thus intercalation is facilitated only at sites where propeller twisting creates an "open" base pair in the major groove (See Figure X). This is an example of more finely tuned shape selectivity of DNA by a metal complex. The mechanism of photocleavage of DNA by this complex has been extensively studied in this laboratory (46). The proposed mechanism involves an excited state phi cation radical, generated by photoactivation at 310 nm, which abstracts directly the C3' hydrogen atom from the deoxyribose, leading eventually to scission of the sugar-phosphate backbone.

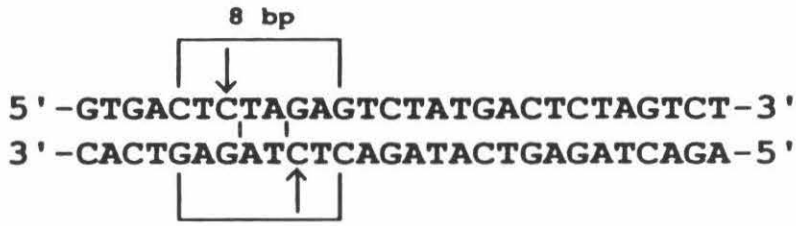
[RhDBP]$^{3+}$, bis(diphenylbipyridyl)(phi) rhodium(III), is a more recently designed metal complex which has the highly intercalative phi ligand and two bipyridyl ligands with phenyl groups which are likely to be less intercalative. This complex can recognize an eight base pair sequence within a 2000 base-pair plasmid (47). Interestingly, this high selectivity is achieved by a dimer of the metal complex, associated presumably through one of the ancillary diphenylbpy ligands, each of which occupies the overlapping half site in the eight base pair segment. This example illustrates another interesting feature of the transition metal complexes: they operate, at least on some level, on a common principle as most proteins that target specific DNA sequences of structures. First of all, they complement the shape of the target DNA, and in this example they are shown to be able to take a step further in DNA recognition by facilitating and increasing the affinity and specificity by dimerization, which proteins are known to do quite commonly.

[Rh(DIP)$_3$]$^{3+}$, tris(diphenylphenanthroline) rhodium(III), is the most structure-specific metal complex that recognizes very unusual structures of DNA. It has been shown to cleave Z-DNA sites in plasmids near the junctions between B-

Figure 1.10 [RhDBP]$^{3+}$ and the 8 base-pair site recognized by the complex. The cleavage of the 8 bp site is specific at nanomolar concentrations, suggesting that the high affinity may be achieved through dimerization of the complex. the arrows indicate the cleaved residues. Adapted from A. Sitlani (1993) Doctoral dissertation, California Institute of Technology.



$[Rh(4,4'\text{-diphenylbpy})_2(phi)]^{3+}$

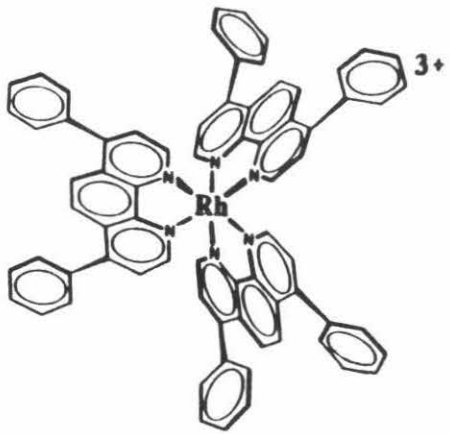


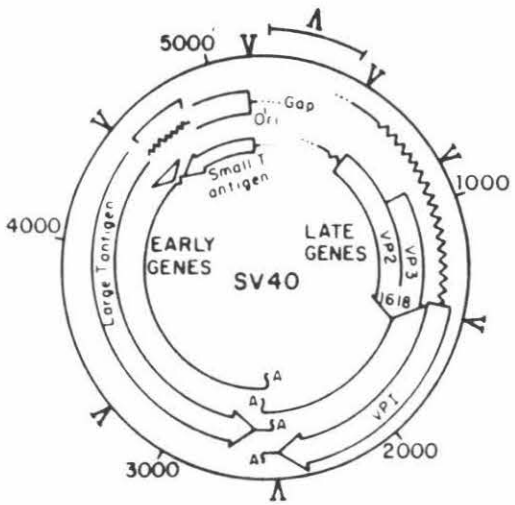
and Z-DNA (48). It was also observed to cleave in the middle of two adjacent helices of a cruciform in plasmid pBR322 (49, See Figure 1.11) and in the center of Holliday junctions (50). Its cleavage of the SV40 genome at functionally significant sites (51) has not been investigated at high resolution to determine the exact location of the cleavage sites. One of these sites, the intron cleavage site, is investigated in this thesis, which presents evidence for the recognition of distinct DNA structures in the SV40 intron and in the intron of another virus, Adenovirus 2.

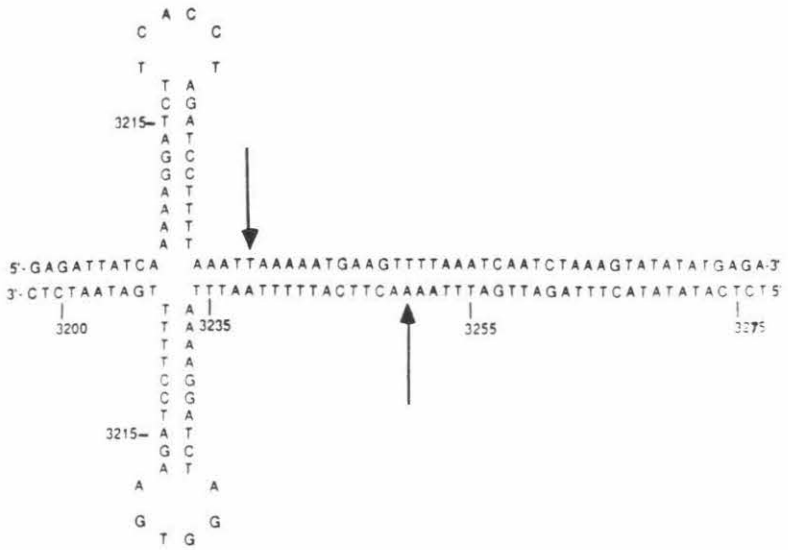
$Rh(DIP)_3^{3+}$ contains three bulky ligands, the diphenyl phenanthroline (DIP), and this makes it less favored for intercalation, though spectroscopy indicates that it is capable of intercalating and that the delta isomer is favored in intercalation as in $Ru(phen)_3^{3+}$ (52). Its cleavage properties are consistent with the $Rh(phen)_2phi^{3+}$, and its cleavage mechanism is likely to be similar to the C3'-H abstraction proposed for the phi complex. This indicates that the interaction of $Rh(DIP)_3^{3+}$ with its target sites is an intimate one and not a loose "surface-binding" interaction, which would not lead to the specific and strong cleavages observed. The varieties of the structures for which it is specific suggest, however, that there is more than one type of binding mode for this complex. Also consistent with its specificity, the complex does not target regular A-form or B-form helix or unstructured single-stranded DNA. The metal complex has a secondary, non-structure-specific, cleavage preference for G residues, which involves a different cleavage chemistry with attack at the base itself rather than the C3'-H of the sugar in the structure-specific interaction. This secondary cleavage is usually at lower level than the structure-specific cleavage.

$Rh(DIP)_3^{3+}$ has also been used to probe RNA structures; it specifically targets the TΨC loop and the G-U mismatch in the acceptor stem of $tRNA^{Phe}$ (53). The loop is single-stranded but structured in that the bases still stack on one another as in the usual A-conformation. The targeting of the G-U mismatch seems to depend on the

Figure 1.11 Targeting of unusual DNA structures by $Rh(DIP)_3^{3+}$. The plasmid map shows the sites of cleavage by the metal complex (marked with "Λ"). Note the cleavage occurs at functionally significant regions of the genome, including the intron of the Early genes (51). Cleavage of the cruciform occurs mainly on one helix of the structure away from the center.


3+
Rh
N


5000
1000
2000
3000
4000
Gap
Ori
Small T antigen
Large T antigen
EARLY GENES
LATE GENES
SV40
VP2
VP3
VP1
1618


5'-GAGATTATCA AAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGA-3'
3'-CTCTAATAGT TTTAATTTTTACTTCAAAATTTAGTTAGATTTCATATATACTCT-5'
3200
3215
3235
3255
3275

nature of the mismatch and not on other tertiary structural features of tRNA, and the cleavage always occurs on the 3' side of the U residue. These observations are consistent with the DNA recognition properties of the complex: it targets *unusually structured* nucleic acid molecules and not regular helices or random single strands.

Due to the complexities of the structures targeted by $Rh(DIP)_3^{3+}$ and lack of detailed information on them, such as crystallographic data, the interaction between the complex and its specific target sites are not well understood. The mode of binding for the above mentioned structure probably involves interactions with many parts of the structure that are adjacent to each other and create a kind of binding pocket for the complex. The high specificity of the complex for unusual sites makes it a very useful tool for detecting perturbations in structure along a long stretch of DNA. This property of the metal complex was utilized to detect and characterize the intron DNA structure described in this thesis.

1.4 Introns and RNA splicing

To fully understand the significance of introns, it is worthwhile to study the process by which introns are removed from the pre-mRNA transcripts of intron-containing genes (2). In the mammalian splicing system, with which we are mainly concerned, the splicing of the intron takes place within a complex of RNA and proteins called the spliceosome. The major building blocks of the spliceosome are, apart from the transcript itself, four snRNP's (small nuclear ribonucleoprotein particles) which are known by their RNA components U1, U2, U4-U6, and U5 snRNA's. In addition to the snRNP's there are many protein factors, whose precise number and nature are still to be determined.

The various components of the spliceosome recognize parts of the transcript and orient them into a structure capable of catalyzing the transesterification

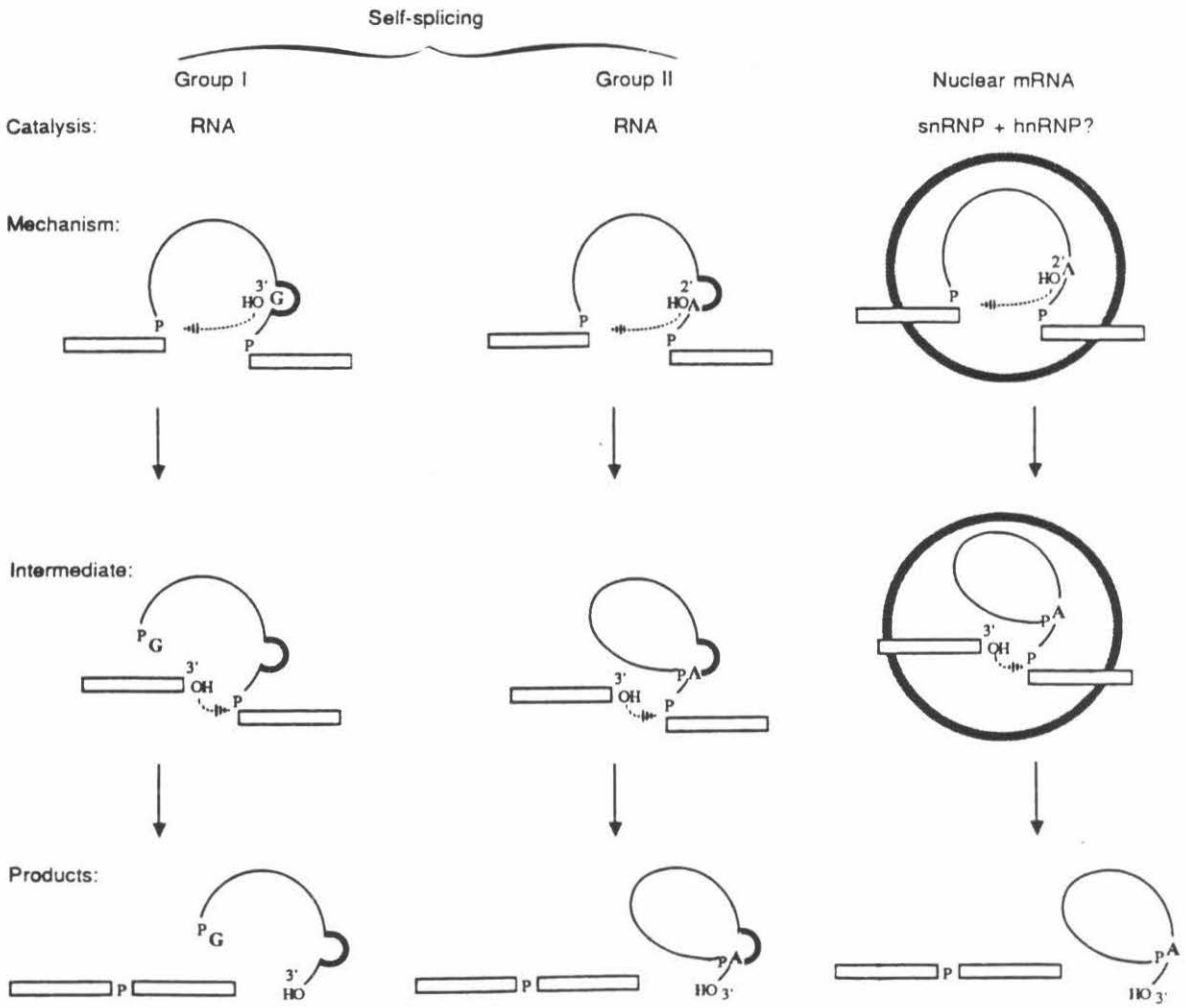


Figure 1.12 RNA splicing in group I, group II, and nuclear spliceosomal systems. All systems share a similar mechanism, and the products of the reactions are also similar in structure. Adapted from P. A. Sharp (1987) *Science* **235**: 766-771.

reactions required for cleavage and religation of the RNA strand. U1 recognizes the 5' splice site and U5 the 3' splice site. U2 anchors the branch point and interacts with U6 which join the complex at a stage to bring the three parts of the transcript together for the chemical steps of transesterification. In the actual reactions, which are thought to be catalyzed by the RNA, rather than the protein, components of the spliceosome, the 2' hydroxyl group of the branch point adenosine, located approximately 30 residues 5' to the 3' end of the intron, is involved in a nucleophilic attack on the 5' phosphate of the 5' end guanosine of the intron. This produces the 5' exon and the lariat intermediate containing the 3' exon at its end. The 3' hydroxyl group of the last nucleotide of the 5' exon then attacks the 5' carbon of the first nucleotide of the 3' exon, leading to the formation of the covalently joined exons and the lariat intron. The resulting mature mRNA is transported out of the nucleus into the cytoplasm for translation, and the intron is degraded.

In self-splicing systems the transesterification reactions are catalyzed by the transcripts themselves. In these cases the structure assumed by the transcript is of vital importance to catalysis (54). More and more evidence is accumulating that supports the hypothesis that the structure of the assembled RNA molecules in the spliceosome resembles that of the self-splicing RNA's, and that snRNA's catalyze the spliceosomal transesterification (55).

1.5 Implications for the intron DNA structure

The foregoing section makes it clear that introns at the RNA level must fold into a distinct structure for splicing to occur. Thus its function at that level can be described as presenting itself in a form that facilitates its own removal. In self-splicing systems this is taken a step further, and the intron actually catalyzes its own removal. At the DNA level, however, this function does not come into play simply

because DNA is not "spliced" as the transcripts are. Whether introns in DNA has any function is a question that has yet to be answered. This function, if it exists, must entail something more than the tautological "spacer-function," for such a function does not necessarily assume any active part in the intron. The discovery of a structure in the intron DNA, which will be described in detail in this thesis, opens up the possibility that introns may play more than a passive role in delineating boundaries of coding segments of genes, exons, as initially postulated in the exon theory of genes. Therefore, characterization of such an intron DNA structure and a search for similar structures in other systems may eventually lead to elucidation of the true significance of introns.

References:

1. (a) A. J. Berk & P. A. Sharp (1977) *Proc. Natl. Acad. Sci. USA* **74**: 3171-3175 (b) L. T. Chow, R. E. Gelinas, T. R. Broker, & R. T. Roberts (1977) *Cell* **12**: 1-8.
2. (a) R. A. Padgett, P. J. Grabowski, M. M. Konarska, S. Seiler, & P. A. Sharp (1986) *Ann. Rev. Biochem* **55**: 1119-1150. (b) M. R. Green (1986) *Ann. Rev. Genetics* **20**: 671-708. (c) E. Brody & J. Abelson (1985) *Science* **228**: 963-967. (d) J. A. Steitz, D. L. Black, V. Gerke, K. A. Parker, A. Kramer, D. Frendewey, & W. Keller (1988) *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*, pp. 115-154, Max L. Birnstiel, ed., Springer Verlag, Berlin. (e) C. Guthrie & B. Patterson (1988) *Ann. Rev. Genet.* **22**: 387-419.
3. (a) W. Gilbert (1978) *Nature* **271**: 501. (b) W. Gilbert (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52**: 901-905.
4. A. H.-J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. v. Boom, G. v. d. Marel, & A. Rich (1979) *Nature* **282**: 680-686.
5. A. Nordheim & A. Rich (1983) *Nature* **303**: 674-679.
6. B. Wittig, S. Wolfl, T. Dorbic, W. Vahrson, & A. Rich (1992) *EMBO J.* **11**: 4653-4663.
7. N. Panayotatos & R. D. Wells (1981) *Nature* **289**: 466-470.
8. G. N. Giaever & J. C. Wang (1988) *Cell* **55**: 849-856.
9. A. Dayn & S. Malkosyan, & S. M. Mirkin (1992) *Nuc. Acids Res.* **20**: 5991-5997.
10. R. Holliday (1964) *Genet. Res.* **5**: 282-304.
11. D. M. J. Lilley & R. M. Clegg (1993) *Ann. Rev. Biophys. Biomolec. Struct.* **22**: 299-328.
12. (a) A. Bhattacharyya, A. I. H. Murchie, E. v. Kitzing, S. Diekmann, B. Kemper, & D. M. J. Lilley (1991) *J. Mol. Biol.* **221**: 1191-1207. (b) C. A. Parsons, A. I. H. Murchie, D. M. J. Lilley, & S.C. West (1989) *EMBO J.* **8**: 239-246.

13. (a) H. Htun & J. E. Dahlberg (1988) *Science* **241**: 1791-1796 (b) H. Htun & J. E. Dahlberg (1989) *Science* **243**: 1571-1576.

14. H. C. M. Nelson, J. T. Finch, B. F. Luisi, & A. Klug (1987) *Nature* **330**: 221-226.

15. H.-S. Koo, H.-M. Wu, & D. M. Crothers (1986) *Nature* **320**: 501-506.

16. J. C. Marini, S. D. Levene, D. M. Crothers, & P. T. Englund (1982) *Proc. Natl. Acad. Sci. USA* **79**: 7664-7668.

17. J. Griffith, M. Bleyman, C. A. Rauch, P. A. Kitchin, & P. T. Englund (1986) *Cell* **46**: 717-724.

18. K. Ryder, S. Silver, A. L. DeLucia, E. Fanning, & P. Tegtmeyer (1986) *Cell* **44**: 719-725.

19. M. Snyder, A. R. Buchman, & R. W. Davis (1986) *Nature* **324**: 87-89.

20. K. Zahn & F. R. Blattner (1987) *Science* **236**: 416-422.

21. C. H. Hsieh & J. Griffith (1988) *Cell* **52**: 535-544.

22. E. H. Blackburn & J. W. Szostak (1984) *Ann. Rev. Biochem* **53**: 163-194.

23. (a) J. R. Williamson, M. K. Raghuraman, & T. R. Cech (1989) *Cell* **59**: 871-880. (b) W. I. Sundquist & A. Klug (1989) *Nature* **342**: 825-829.

24. C. H. Kang, X. Zhang, R. Ratliff, R. Moyzis, & A. Rich (1992) *Nature* **356**: 126-131.

25. F. W. Smith & J. Feigon (1992) *Nature* **356**: 164-168.

26. T. F. McCutchan, J. L. Hansen, J. B. Dame, & J. A. Mullins (1984) *Science* **225**: 625-628.

27. M. A. Keene & S. C. R. Elgin (1984) *Cell* **36**: 121-129.

28. J. Paquette, K. Nichoghosian, G. Qi, N. Beauchemin, & R. Cedergren (1990) *Eur. J. Biochem.* **189**: 259-265.

29. W. Herr (1985) *Proc. Natl. Acad. Sci. USA* **82**: 8009-8013.

30. G. C. Glikin, M. Vojtiskova, L. Rena-Descalzi, & E. Palecek (1984) *Nuc. Acids Res.* **12**: 1725-1735.

31. K. Nejedly, M. Kwinkowski, G. Galazka, J. Klysik, E. Palecek (1985) J. Biomolec. Struct. Dyn. **3**: 467-478.

32. G. D. Cimino, H. B. Gamper, S. T. Isaacs, & J. E. Hearst (1985) *Ann. Rev. Biochem* **54**: 1151-1193

33. (a) P. L. Wollenzien, J. E. Hearst, P. Thammana, & C. R. Cantor (1979) *J. Mol. Biol.* **135**: 255-269. (b) C. R. Cantor, P. L. Wollenzien, & J. E. Hearst (1980) *Nucl. Acids Res.* **8**: 1855-1872.

34. C.-K. J. Shen & J. E. Hearst (1976) *Proc. Natl. Acad. Sci. USA* **73**: 2649-2653.

35. J. T. Millard, M. F. Weidner, J. J. Kirchner, S. Ribeiro, & P. B. Hopkins (1991) *Nucl. Acids Res.* **19**: 1885-1891.

36. (a) R. P. Hertzberg & P. B. Dervan (1982) *J. Am. Chem. Soc.* **104**: 313-315. (b) R. P. Hertzberg & P. B. Dervan (1984) *Biochemistry* **23**: 3934-3945.

37. T. D. Tullius, B. A. Dombroski, M. E. A. Churchill & L. Kam (1987) *Methods Enz.* **155**: 537-559.

38. J. A. Latham & T. R. Cech (1989) *Science* **245**: 276-282.

39. A. M. Pyle & J. K. Barton (1990) *Prg. Inorg. Chem.* **38**: 413-475.

40. (a) N. P. Pavletich & C. O. Pabo (1993) *Science* 261: 1701-1707 (b) L. Fairall, J. W. R. Schwabe, L. Chapman, J. T. Finch, & D. Rhodes (1993) *Nature* **366**: 483-487.

41. C. J. Murphy & J. K. Barton. *Methods Enz.*, in press.

42. C. S. Chow & J. K. Barton (1992) *Methods Enz.* **212**: 219-242.

43. (a) J. K. Barton (1986) *Science* **233**: 727-734. (b) J. K. Barton, A. T. Danishevsky, & J. M. Goldberg (1984) *J. Am. Chem. Soc.* **106**: 2172-2176.

44. A. M. Pyle, E. C. Long, & J. K. Barton (1989) *J. Am. Chem. Soc.* **111**: 4520-4522.

45. A. M. Pyle, J. P. Rehmann, R. Meshoyrer, C. V. Kumar, N. J. Turro, & J. K. Barton (1989) *J. Am. Chem. Soc.* **111**: 3501-3058.

46. A. Sitlani, E. C. Long, A. M. Pyle, & J. K. Barton (1992) *J. Am. Chem. Soc.* **114**: 2303-2312.

47. A. Sitlani, C. Dupureur, & J. K. Barton (1994) *J. Am. Chem. Soc.* **115**: 12589-12590.

48. M. R. Kirshenbaum (1989) Doctoral Dissertation, Columbia University.

49. M. R. Kirshenbaum, R. Tribolet, & J. K. Barton (1988) *Nuc. Acids Res.* **16**: 7948-7960.

50. K. Waldron, J. Voulgaris, & J. K. Barton, unpublished results.

51. B. C. Müller, A. L. Raphael, & J. K. Barton (1987) *Proc. Natl. Acad. Sci. USA* **84**: 1764-1768.

52. (a) J. K. Barton, L. A. Basile, A. Danishevsky, & A. Alexandrescu (1984) *Proc. Natl. Acad. Sci. USA* **81**: 1961-1965. (b) C. V. Kumar, J. K. Barton, & N. J. Turro (1985) *J. Am. Chem. Soc.* 107: 5518-5523.

53. C. S. Chow & J. K. Barton (1992) *Biochemistry* **31**: 5423-5429.

54. T. R. Cech (1990) *Ann. Rev. Biochem* **59**: 543-568.

55. (a) H. D. Madhani & C. Guthrie (1992) *Cell* **71**: 803-817. (b) D. S. McPheeters & J. Abelson (1992) *Cell* **71**: 819-831. (c) C. F. Lesser & C. Guthrie (1994) *Science* **262**: 1982-1988. (d) E. J. Sontheimer & J. A. Steitz (1994) *Science* **262**: 1989-1996.

Chapter 2.

Low-resolution mapping of $Rh(DIP)_3^{3+}$ cleavage sites in plasmids containing the introns of Simian Virus 40 T-antigen and Adenovirus 2 E1A genes

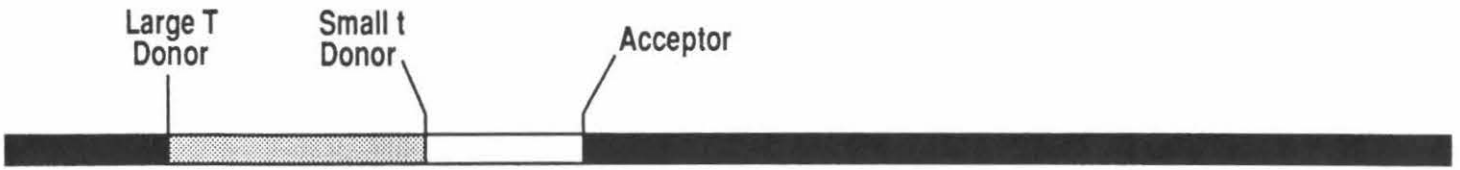
2.1 Introduction

Introns and RNA splicing were first discovered in Simian Virus 40 (SV40) and Adenovirus 2 (Ad2) transcription systems (1, 2). These viruses infect mammalian cells and often give rise to tumors in their hosts (3). Both have genes that are expressed early in the infection cycle to establish the viral takeover of the cell machinery. In SV40 this gene is the T-antigen gene, and in Ad2 it is the E1A gene. The chief functions of the gene products are believed to be in regulation of viral DNA replication. They also regulate the transcription of their own genes as well as others of the viral genome. Both of these genes are spliced in such a way (termed alternative splicing) that each leads to multiple protein products (2, 3). The T-antigen gene gives rise to, depending on the position of the 5' splice site, the large T-antigen or the small t-antigen. The E1A gene gives rise to, again with different 5' splice sites, several products which are known by the size of their mRNA's: 9S, 12S, and 13S. These transcription units are schematically illustrated in Figure 2.1.

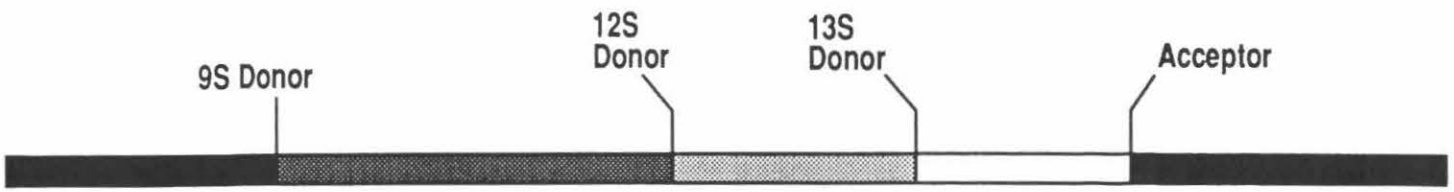
The genome of SV40, which is a circular plasmid of 5243 base pairs, was the first plasmid DNA used to test the specificity of the tris(diphenylphenanthroline) complexes (4). The cobalt complex, $Co(DIP)_3^{3+}$, was initially used to cleave the SV40 genome. Cleavage was indeed specific and not random (See Figure 1.11), and interestingly, the sites of cleavage corresponded with functionally significant regions of the genome, such as the origin of replication and beginnings and ends of genes, and most strikingly in the intron of the T-antigen gene. The surprise about the intron site was that at the DNA level the intron does not have any function that we know

Figure 2.1 Schematic illustration of the transcription units of the SV40 T-antigen and the Ad2 E1A genes. The clear area represents the intron sequence common to all of the alternative splicing, and the shaded areas represent the portions of introns for the particular alternative splicing reactions.

Simian Virus 40 T Antigen Transcription Unit



Adenovirus 2 E1A Transcription Unit



of, at least not one that would require an unusual structure as indicated by the specificity of $Co(DIP)_3^{3+}$ cleavage. The other sites at beginnings and ends of genes and at the origin of replication can be explained by the fact that these sites are indeed places where significant DNA biochemical events take place; many protein factors bind to these regions to initiate or to terminate DNA replication and transcription. It would be likely for these regions to contain unusual DNA structures.

Therefore, the specific and strong cleavage in the intron presented an intellectual challenge. What could be so unusual about a stretch of DNA that is apparently useless to the virus? This question is directly related to the fundamental difficulty we face in thinking about introns. The answer to this question would lead us a long way to understanding the evolutionary and biochemical significance of introns.

The early results on $Co(DIP)_3^{3+}$ cleavage of the SV40 genome led to the further examination of the intron site described in this thesis. The Adenovirus E1A intron was also taken up for study because of the similarity of its organization and functions to those of SV40 T-antigen gene, but there was no homology between the sequences of the two genes. This chapter describes the experiments leading to the establishment of the cleavage in these introns cloned in plasmids so that only the intron of interest and the surrounding exons are present among the vector sequences. In both cases specific cleavage by the rhodium complex, $Rh(DIP)_3^{3+}$, is observed in the introns of the cloned gene segments.

2.2 Experimental

Materials:

Plasmids containing the genes of interest as well as a control cruciform site

(pBR322 cruciform; see ref. 5) were provided by Prof. James L. Manley of Columbia University and amplified using standard cloning techniques (6). The plasmids were constructed by cloning Hind III fragments of the two genes into the expression vector pSP64, a pBR322 derivative, from Promega (Figure 2.2). The pSP64-SVT plasmid contains SV40 sequences from 4002 to 5171, and the pSP64-E1A plasmid contains Ad2 sequences from 500 to 1569. $Rh(DIP)_3^{3+}$, prepared as described previously (7), was available as a stock in the laboratory. S1 nuclease and the restriction enzymes were from Boehringer Mannheim Biochemicals. Molecular weight marker DNA was from Bethesda Research laboratories. Molecular biology grade reagents for buffers were from Sigma.

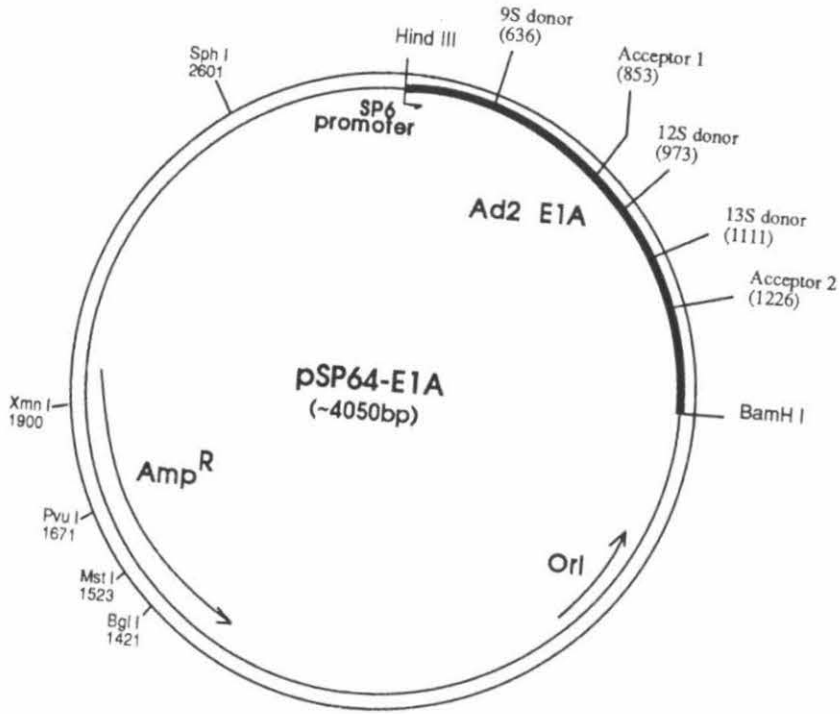
Instrumentation:

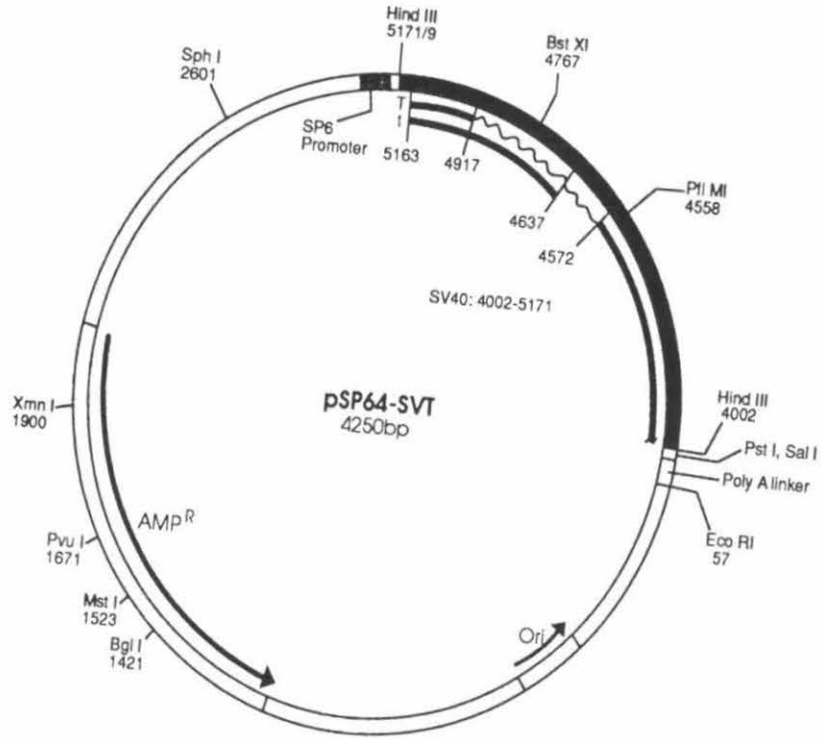
The light source used in photocleavage was a 1000 W Oriel Hg/Xe lamp (model 6140) fitted with a monochromator and a 300 nm cut-off filter. Quantitation of DNA was carried out by UV-VIS absorption spectrometry with Cary 219 Spectrophotometer.

Methods:

Supercoiled plasmid DNA (100 μM nucleotides) in 20 μl of 20 mM Tris-HCl, pH 7.4, 25 mM NaCl was photolyzed at 313 nm for 2 minutes with a Hg/Xe lamp 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.

Figure 2.2 Plasmid constructs containing the SV40 T-antigen gene and the Ad2 E1A gene. Portions of the genes are cloned into the expression vector pSP64 (Promega), which contains convenient restriction enzyme sites for mapping.


Sph I
2601
Hind III
9S donor
(636)
Acceptor 1
(853)
SP6
promoter
12S donor
(973)
Ad2 E1A
13S donor
(1111)
Acceptor 2
(1226)
pSP64-E1A
(~4050bp)
BamH I
Xmn I
1900
Amp^R
Pvu I
1671
Mst I
1523
Bgl I
1421
Ori


Hind III
5171/9
Bst XI
4767
Sph I
2601
T
1
SP6
Promoter
5163
4917
Pfl MI
4558
4637
4572
SV40: 4002-5171
pSP64-SVT
4250bp
Hind III
4002
Pst I, Sal I
Poly A linker
Xmn I
1900
AMP^R
Eco RI
57
Pvu I
1671
Mst I
1523
Bgl I
1421
Ori

For cleavage of the linearized plasmids, the plasmids were digested with the restriction enzyme and ethanol precipitated before photolysis with $Rh(DIP)_3^{3+}$.

For determining the salt dependence of photocleavage of the plasmids, the same protocol was used in buffers of NaCl concentrations varying from 10 to 100 mM.

2.3 Results and discussion

Low-resolution mapping of the $Rh(DIP)_3^{3+}$ cleavage sites in the introns:

The low-resolution mapping experiment is schematically illustrated in Figure 2.3. Two sets of experiments are carried out for each plasmid. In one the plasmid is photolyzed in the presence of $Rh(DIP)_3^{3+}$ and digested with a restriction enzyme which cuts the plasmid approximately diagonally across from the expected site of photocleavage. In the other experiment the plasmid is digested with a different restriction enzyme which cuts significantly far away from both the first enzyme site and the expected photocleavage site. In both cases the plasmid is finally treated with S1 nuclease to visualize single-strand nicks. This protocol locates unambiguously the site of single-strand, and double-strand, cleavage by the rhodium complex.

The mapping experiments were carried out on the SV40 T-antigen plasmid (pSP64-SVT; Figure 2.4) and the Ad2 E1A plasmid (pSP64-E1A; Figure 2.5). On each plasmid a site within the intron insert is specifically cleaved by the rhodium complex. Two other sites of specific cleavage are also evident, one of which corresponds to a cruciform, 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 $Co(DIP)_3^{3+}$ (5). At lower levels of irradiation of supercoiled plasmids followed by linearization but without S1 nuclease treatment, fragments are observed which correspond primarily to specific cleavage at the cruciform site rather than the intron

Figure 2.3 A 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.

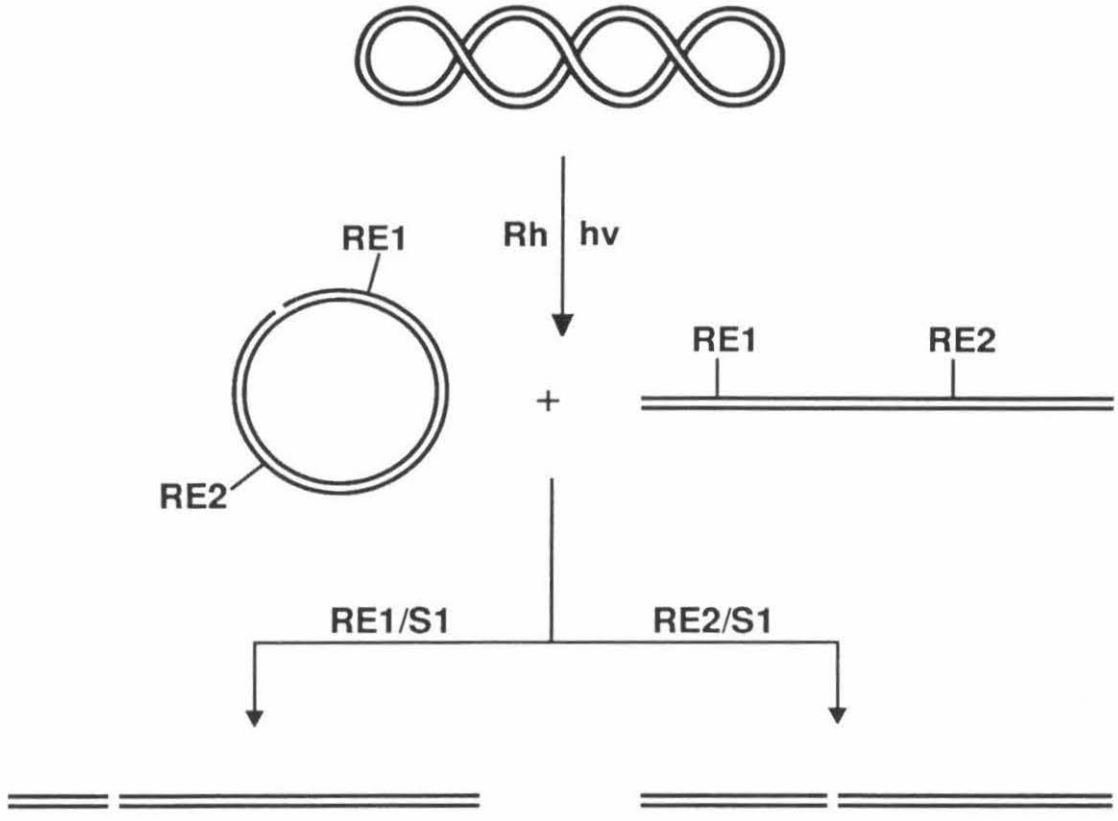

RE1
Rh
hv
RE1
RE2
+
RE2
RE1/S1
RE2/S1

Figure 2.4 Low-resolution 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; lane 2, DNA incubated with $Rh(DIP)_3^{3+}$ but without irradiation; 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 $Rh(DIP)_3^{3+}$, then digested with Pvu I; lane 7, DNA irradiated at 313 nm with $Rh(DIP)_3^{3+}$, 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 $Rh(DIP)_3^{3+}$, then digested with Eco RI; lane 10, DNA irradiated at 313 nm with $Rh(DIP)_3^{3+}$, 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, 16).

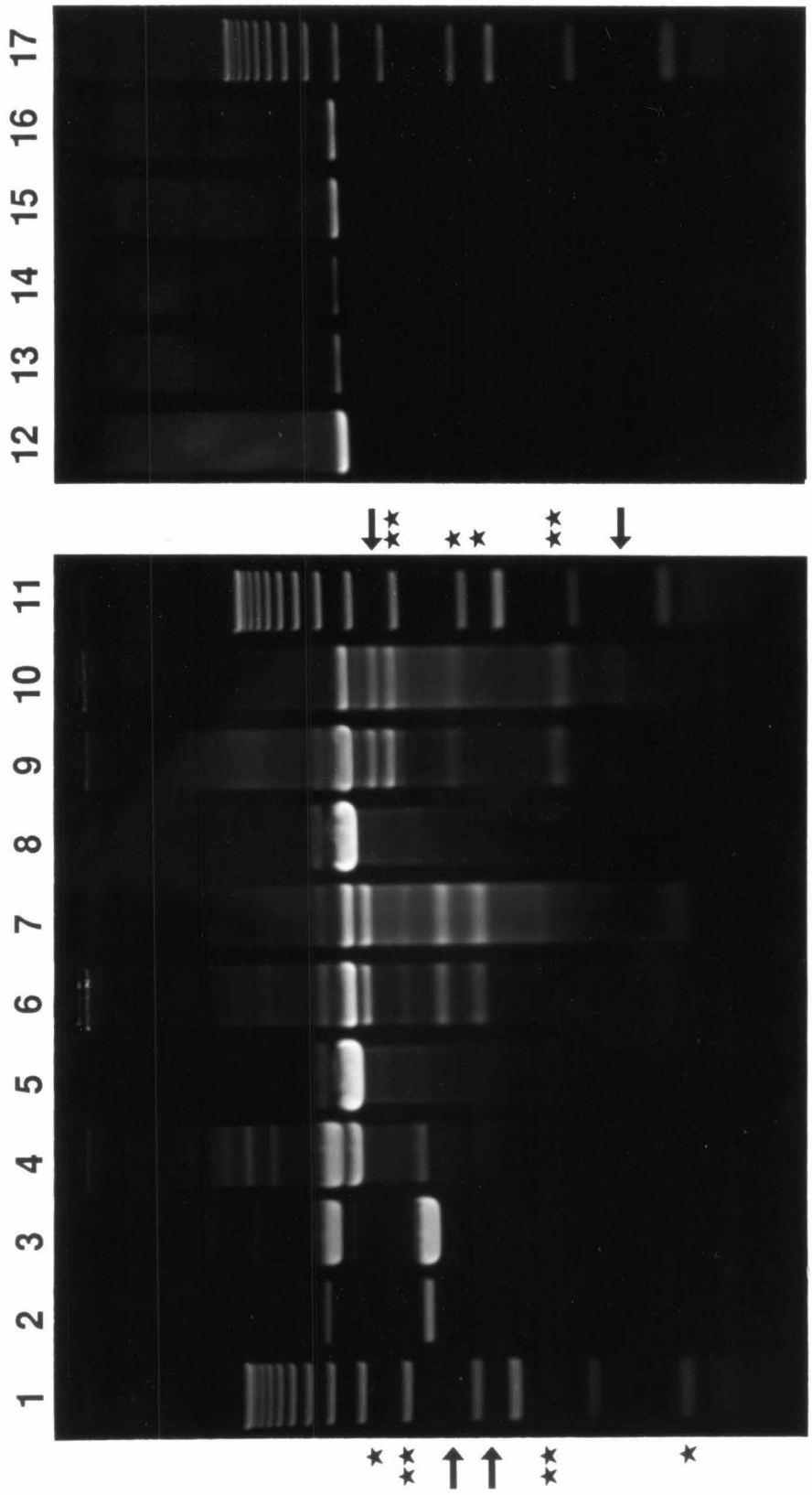
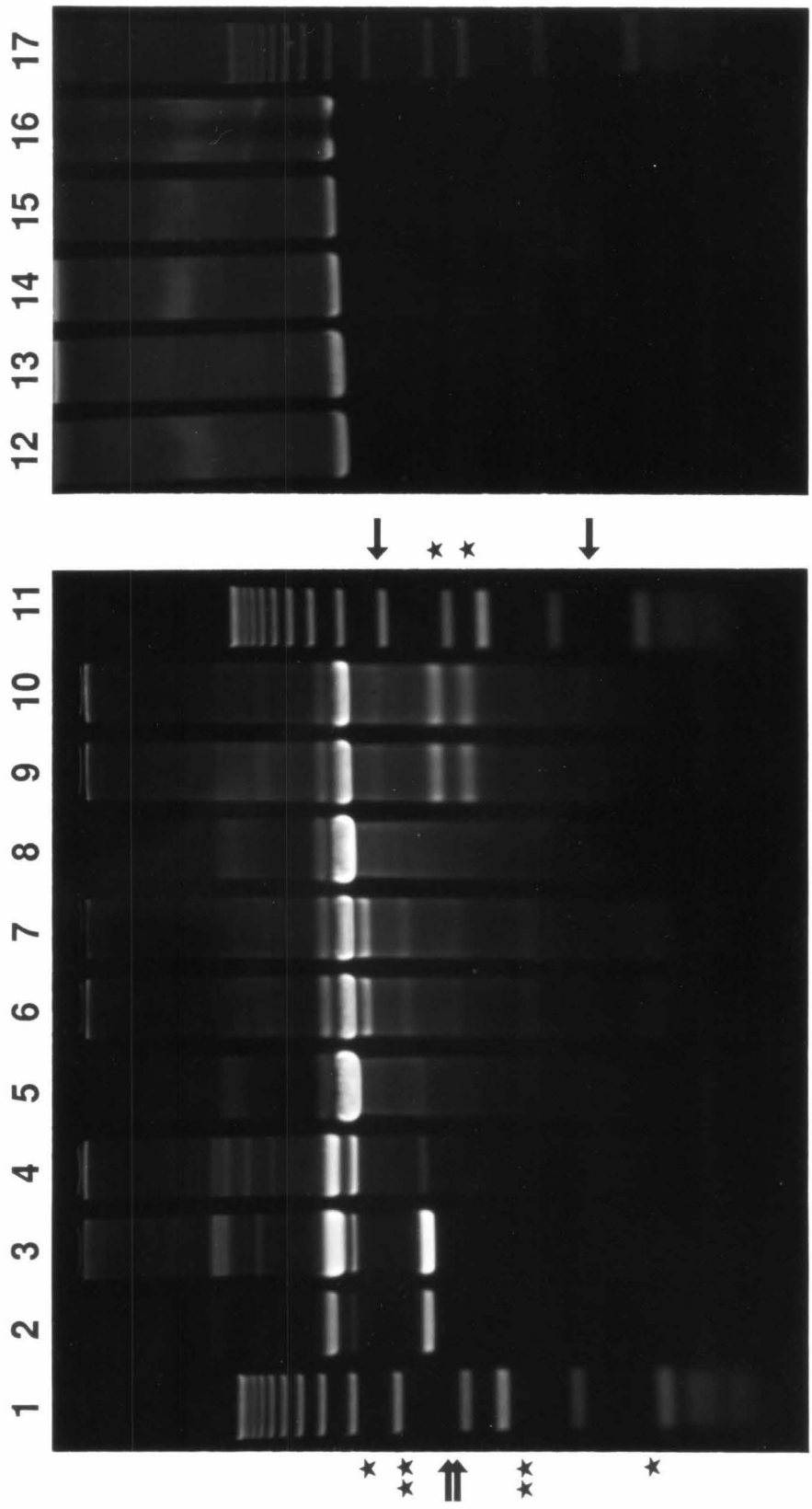

12
13
14
15
16
17
1
2
3
4
5
6
7
8
9
10
11

Figure 2.5 Low-resolution 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; 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 $Rh(DIP)_3^{3+}$, then digested with Pvu I and S1; lane 6, DNA irradiated at 313 nm with $Rh(DIP)_3^{3+}$, then digested with Pvu I; lane 7, DNA irradiated at 313 nm with $Rh(DIP)_3^{3+}$, then digested with Pvu I and S1; lane 8, DNA irradiated at 313 nm without $Rh(DIP)_3^{3+}$, then digested with Hind III and S1; lane 9, DNA irradiated at 313 nm with $Rh(DIP)_3^{3+}$, then digested with Hind III; lane 10, DNA irradiated at 313 nm with $Rh(DIP)_3^{3+}$, then digested with Hind III 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 Hind III 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, 16), showing the requirement of supercoiling in the structures targeted.


17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1

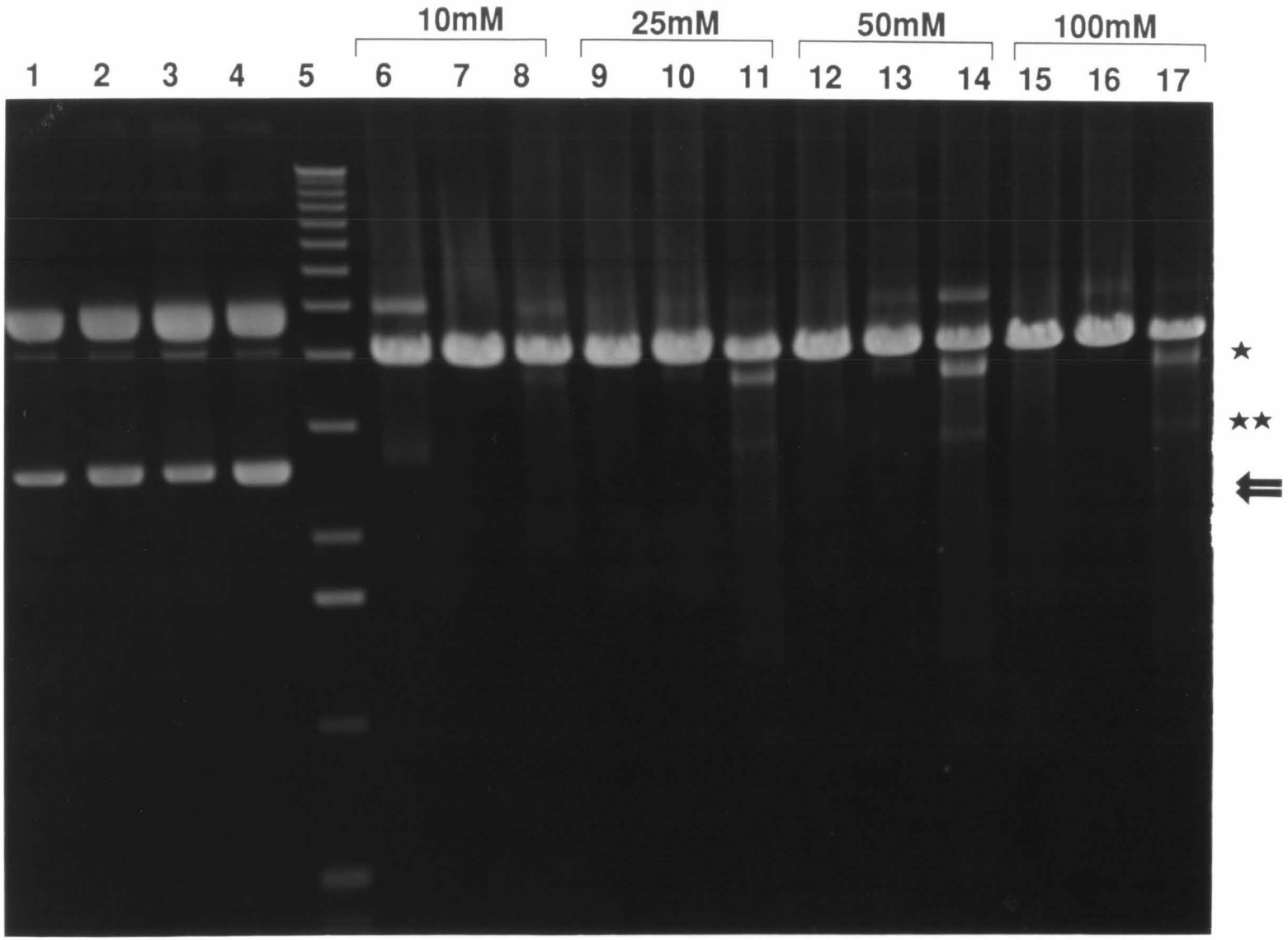
site. This indicates that the cleavage at the intron site is primarily single-stranded as opposed to the cleavage at the cruciform site which is primarily double-stranded. It should be noted that the relative intensities of the cleaved fragments vary between plasmids. Also of interest is the fact that the cloned gene fragment of SV40 behaved in an identical manner in cleavage with $Rh(DIP)_3^{3+}$ as the entire SV40 genome did (4). This shows that the entire genome is not necessary for the intron structure and suggests that supercoiling is sufficient to induce the structure.

When photolysis of the plasmids is carried out *after* first digesting the plasmids with one or the other of the chosen restriction enzymes, no apparent cleavage is observed at either the cruciform site or the intron site. This is consistent with the observation that cruciforms require supercoil stress to form (8) and indicates that the cleavage in the introns is also a result of a structure dependent on supercoiling.

Salt concentration dependence of the $Rh(DIP)_3^{3+}$ cleavage:

Ionic strength affects the structural conformations of many macromolecules such as DNA. To probe further into the characteristics of the structure recognized by $Rh(DIP)_3^{3+}$ in the introns of the two cloned genes, low-resolution mapping was carried out under a variety of NaCl concentrations, from 10 mM to 100 mM. Figure 2.6 shows the results from the cleavage of the Ad2 E1A plasmid. The variation in the NaCl concentration brought about a variation in the intensities of cleavage by the rhodium complex at the intron sites. Cleavage was weak at 10 mM and increased up to 50 mM and then decreased at 100 mM. Interestingly, cleavage at the cruciform site showed a different profile; it increased sharply at 25 mM and stayed intense up to 200 mM. The difference in the salt concentration dependence of the two sites suggests different types of structures at the two sites. The intron site appears to be more sensitive to variations in salt concentration, preferring a certain range instead

Figure 2.6 Low-resolution map of sites cleaved specifically by $Rh(DIP)_3^{3+}$ on supercoiled pSP64-E1A under increasing salt concentrations. Lane 1, DNA incubated with $Rh(DIP)_3^{3+}$ but without irradiation; lane 2, DNA irradiated at 313 nm but without $Rh(DIP)_3^{3+}$; lane 3. DNA irradiated at 313 nm in the presence of $Rh(DIP)_3^{3+}$ in 10 mM NaCl; lane 4, DNA irradiated at 313 nm in the presence of $Rh(DIP)_3^{3+}$ in 100 mM NaCl; lane 5, 1 Kb marker; lanes 6, 9, 12, & 15, DNA irradiated at 313 nm without $Rh(DIP)_3^{3+}$, in 10, 25, 50, or 100 mM NaCl, then digested with Pvu I and S1; lane 7, 10, 13, & 16, DNA irradiated at 313 nm with $Rh(DIP)_3^{3+}$, in 10, 25, 50, or 100 mM NaCl, then digested with Pvu I; lanes 8, 11, 14, & 17, DNA irradiated at 313 nm with $Rh(DIP)_3^{3+}$, in 10, 25, 50, or 100 mM NaCl, then digested with Pvu I and S1. Arrows indicate the fragments formed as a result of specific cleavage within the intron. Lone star indicates the cleavage at the cruciform site, and the double star indicates the cleavage at the origin or replication. The intensity of the intron cleavage, very faintly visible, increases up to 50 mM and then falls at 100 mM, while the intensity of cleavage at the cruciform site remains fairly constant from 25 mM to 100 mM.

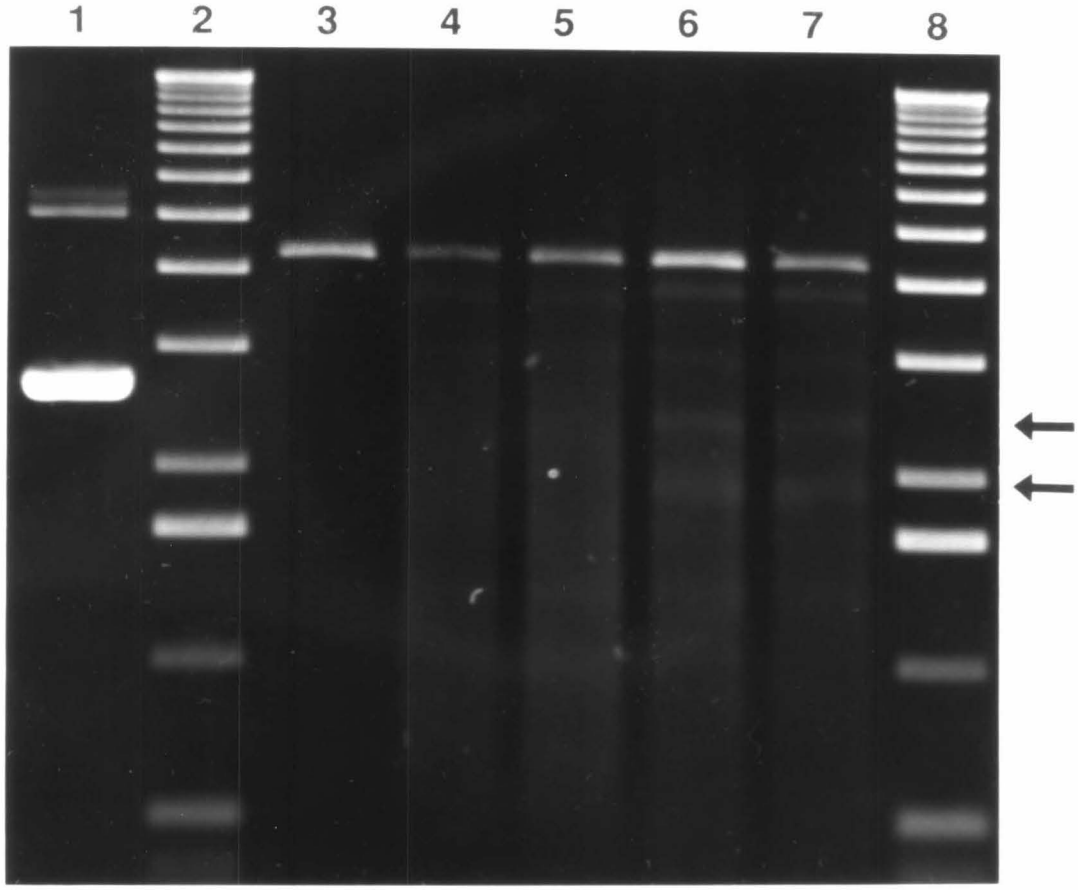

10mM
25mM
50mM
100mM
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
★
★★

of requiring a certain level of ionic strength and persisting thereafter like the cruciform.

Another simple way of testing for structural perturbations is to digest the supercoiled plasmid with single-strand specific nucleases such as S1 nuclease. This reveals sites on the plasmid that are single-stranded, which arise as a result of specific structures as well as simple melting of the strands. The plasmids containing the SV40 T-antigen gene was digested with S1 nuclease, in the presence and absence of $Rh(DIP)_3^{3+}$, without photolysis. Subsequent digestion with a restriction enzyme, as is done in the low-resolution photocleavage mapping experiments, showed that the intron site was sensitive to S1 in the presence of the metal complex but only very slightly so in the absence (Figure 2.7). This suggests that the structure is greatly stabilized by the metal complex but not necessarily induced by it.

In summary, the low-resolution mapping experiments confirm earlier findings and establish the presence of a conformationally altered DNA structure in the introns of SV40 T-antigen and Ad2 E1A genes. The structure is supercoil-dependent and sensitive to variations in salt concentration.

Figure 2.7 S1 nuclease mapping of pSP64-SVT, containing the SV40 T-antigen intron. The supercoiled plasmid (100 μM nucleotides) is digested with S1 nuclease in the presence and absence of 10 μM $Rh(DIP)_3^{3+}$ and then digested again with the restriction enzyme Pvu I to detect S1 sensitivity in the intron. lane 1, control plasmid; lanes 2 & 8, 1 Kb marker; lane 3, plasmid digested with Pvu I only; lanes 4 & 5, plasmid digested with S1 and then with Pvu I; lanes 6 & 7, plasmid digested with S1 in the presence of $Rh(DIP)_3^{3+}$ and then with Pvu I. S1 nuclease cleaves the plasmid at the site of rhodium complex photocleavage; the cleavage is greatly enhanced in the presence of the metal complex.


1
2
3
4
5
6
7
8

References

1. (a) A. J. Berk & P. A. Sharp (1977) *Proc. Natl. Acad. Sci. USA* **74**: 3171-3175 (b) L. T. Chow, R. E. Gelinas, T. R. Broker, & R. T. Roberts (1977) *Cell* **12**: 1-8.
2. (a) A. J. Berk & P. A. Sharp (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 1274-1278. (b) A. J. Berk & P. A. Sharp (1978) *Cell*, **14**, 695-711.
3. (a) M. Fried & C. Prives (1986) *Cancer Cells 4/DNA Tumor Viruses*, pp 1-16, Cold Spring Harbor Laboratory, Cold Spring Harbor. (b) U. Pettersson & R. J. Roberts (1986) *Cancer Cells 4/DNA Tumor Viruses*, pp. 37-57, Cold Spring Harbor Laboratory, Cold Spring Harbor.
4. B. C. Müller, A. L. Raphael, & J. K. Barton (1987) *Proc. Natl. Acad. Sci. USA* **84**: 1764-1768.
5. J. K. Barton & A. L. Raphael (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6460-6464.
6. J. Sambrook, E. F. Fritsch, & T. Maniatis (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed. (Cold Spring Harbor Laboratory Press)
7. M. R. Kirshenbaum, R. Tribolet, & J. K. Barton (1988) *Nuc. Acids Res.* **16**: 7948-7960.
8. N. Panayotatos & R. D. Wells (1981) *Nature* **289**: 466-470.

Chapter 3

High-resolution mapping of $Rh(DIP)_3^{3+}$ cleavage sites in the introns of Simian Virus 40 T-antigen and Adenovirus 2 E1A genes

3.1 Introduction

The low-resolution mapping of the plasmids gave us a definite indication of a structure (or structures) in the introns of the SV40 T-antigen and Ad2 E1A genes. However, a more detailed analysis of the intron sequences is necessary to determine precisely where the cleavage site is located and to draw conclusions on the nature of the structure being recognized by the metal complex. High-resolution mapping involves radiolabeling and isolating, after photocleavage, a relatively short fragment of the plasmid containing the cleavage site and analyzing it on a denaturing polyacrylamide gel. The cleavage site can be determined at the nucleotide resolution with this technique.

The results from these experiments reveal that the cleavage site is centered around the branch point of the introns. The functional significance at the RNA level of the sites of cleavage points to the possibility that the structure of the intron RNA is also present in the DNA. Though obviously the two structures cannot be the same, they arise from the same sequence which has a propensity to adopt a certain conformation. On the other hand, the DNA structure may be an independent entity on its own right which serve its own function. These ideas will be discussed more fully in Chapter 6.

Also described in this chapter is probing of the intron for unusual conformations with diethyl pyrocarbonate which modifies exposed N7 of purine residues. These results were not very conclusive but still indicated an altered conformation as suggested by the specific photocleavage by the rhodium complex.

3.2 Experimental

Materials:

The plasmids were originally a gift from Prof. James L. Manley of Columbia University. They were amplified using standard cloning techniques (1). Restriction enzymes and labeling enzymes were from Boehringer Mannheim, Bethesda Research Laboratories, or New England Biolabs. [α-^{32}P]dideoxy-ATP was from Amersham, and [γ-^{32}P]ATP was from New England Nuclear-Dupont. Molecular biology grade reagents for buffers were from Sigma or Boehringer Mannheim. Reagents for Maxam-Gilbert sequencing were from Aldrich. Diethyl pyrocarbonate was from Sigma.

Instrumentation:

The light source used in photocleavage was a 1000 W Oriel Hg/Xe lamp (model 6140) fitted with a monochromator and a 300 nm cut-off filter. Quantitation of DNA was carried out by UV-VIS absorption spectrometry with Cary 219 Spectrophotometer.

Methods:

The plasmid (100 μM nucleotides in 20 μl total volume) was photolyzed at 332 nm for 1-4 minutes in 20 mM Tris-HCl, pH 7.4, 25 mM NaCl in the presence of 5-10 μM $Rh(DIP)_3^{3+}$. The DNA was then washed with 1% sodium dodecyl sulfate to remove the metal complex and ethanol precipitated three to four times. Thereafter, the DNA was 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 was end-labeled with ^{32}P at the 5' end using polynucleotide kinase for E1A and at the 3' end using terminal deoxynucleotidyl transferase for T-

antigen (1) 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 was then isolated on a non-denaturing polyacrylamide gel, denatured, and electrophoresed on a denaturing polyacrylamide gel together with samples sequenced by the Maxam-Gilbert method (2).

3.3 Results and discussion

3.3.1 $Rh(DIP)_3^{3+}$ cleavage of the introns:

The technique of high resolution mapping of the intron cleavage sites is schematically illustrated in Figure 3.1 The supercoiled plasmid is photolyzed as in low-resolution mapping, but a different restriction enzyme is used to cut the plasmid relatively close to the cleavage site of interest. The resulting ends of the linearized plasmid are radiolabeled, and another enzyme is used to cut the plasmid again on the other side of the cleavage site to give a manageable piece to be analyzed on a denaturing polyacrylamide gel.

The experiment was carried out on both SV40 T-antigen and Ad2 E1A introns. Figures 3.2 and 3.3 display the results for the T-antigen and E1A introns, respectively. Within the coding strand of the T-antigen intron (Figure 3.2), specific photoactivated cleavage by $Rh(DIP)_3^{3+}$ is observed within the sequence 5'-AAACTACT**G**ATTCT*AA*T-3', where the site cleaved is highlighted; the major branch point (italicized) is 6 nucleotides away from the cleavage site, and the minor branch point (also italicized) is adjacent to the cleavage site. A stronger cleavage site is apparent on the sequence 5'-TGTCTACA**G**TAAGTGAA-3' which corresponds precisely to the 5' end of the small t-antigen intron and on the sequence 5' GTATTTTA**G**ATT-CCAAC-3' which corresponds precisely to the 3' end of the large T- and small t-

Figure 3.1 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.

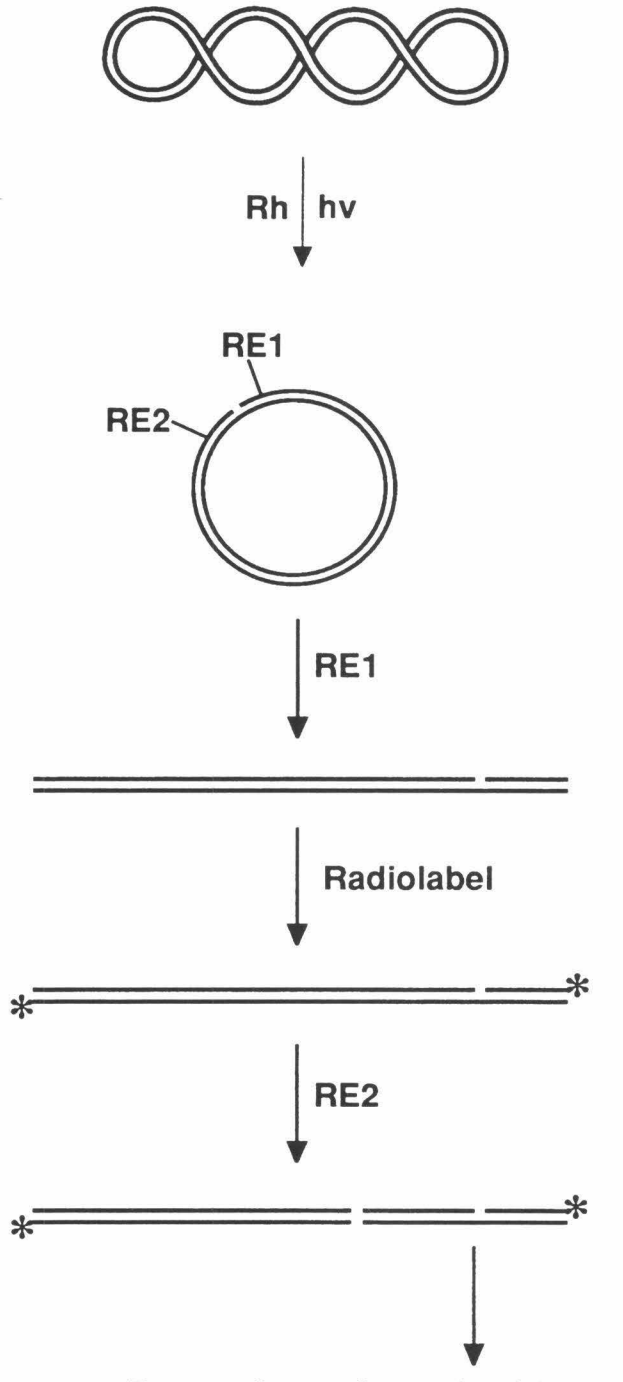

Rh
hv
RE1
RE2
RE1
Radiolabel
RE2
Denaturing polyacrylamide gel

Figure 3.2 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 minutes 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 6 bases to the 5' side of the major branch point and adjacent to the minor branch point.

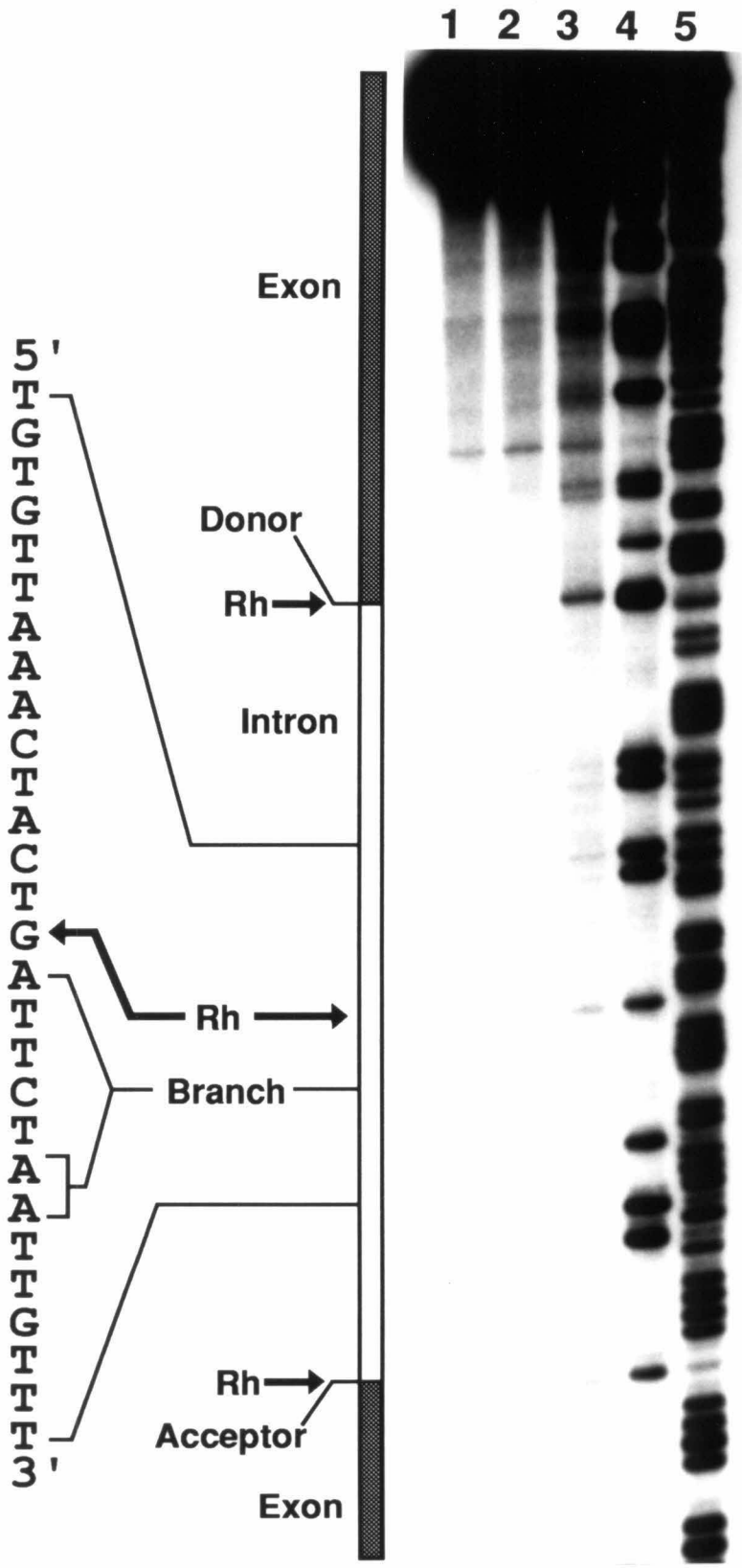
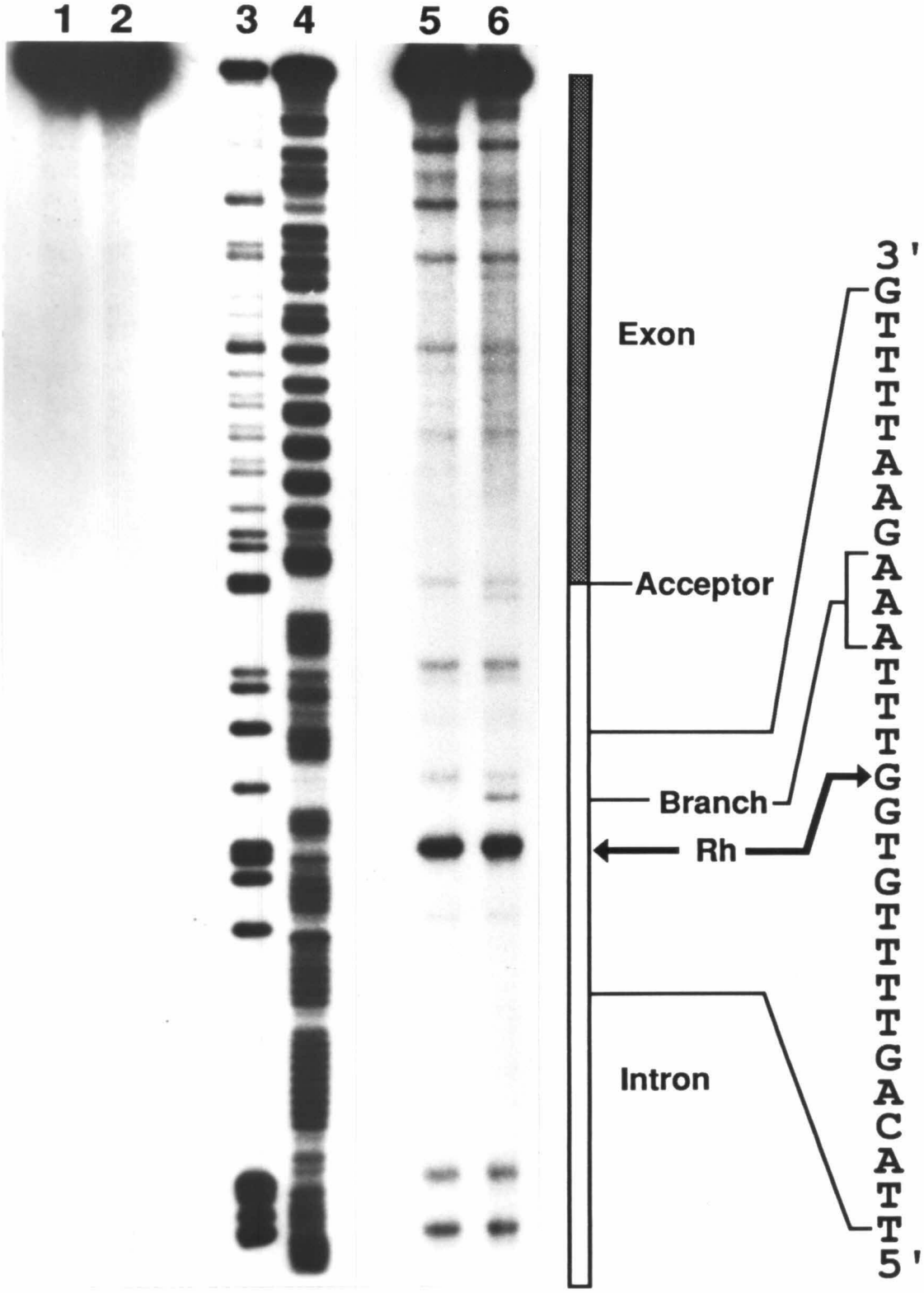

1 2 3 4 5
Exon
5'
TGTGTTAAACTACTTGATTCCTAATTGTTT
3'
Donor
Rh
Intron
Rh
Branch
Rh
Acceptor
Exon

Figure 3.3 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; lanes 5, the pSP64-E1A fragment after irradiation for 1 minute in the presence of 10 μM $Rh(DIP)_3^{3+}$; lane 6, the pSP64-E1A fragment after irradiation for 2 minutes in the presence of 10 μM $Rh(DIP)_3^{3+}$. Intense cleavage (arrow) by $Rh(DIP)_3^{3+}$ is observed at the G 4 bases to the 5' side of the branch point.


1 2
3 4
5 6
Exon
Acceptor
Branch
Rh
Intron
3'
GTTTTAAGAAATTTGGTGTTTTTGACATT
5'

antigen introns. Although the cleavage intensities near the branch point and at the acceptor site are low, the cleavage is consistently reproduced in multiple experiments.

Within the coding strand of the E1A intron (Figure 3.3), strong cleavage is observed within the sequence 5'-GTTTTGTG**G**TTT*AAAGA*-3' , which is substantially different from the SV40 intron. The site cleaved is again highlighted and the branch point (italicized) is 4 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.

When the opposite, the non-coding, strands are examined, no cleavage is apparent immediately opposite from the cleavage sites on the coding strands (Figure 3.4). This indicates that the coding strand alone forms the structure recognized by $Rh(DIP)_3^{3+}$. The non-coding strand does not apparently fold into any discrete structure that can be recognized by the metal complex. This is an interesting observation for which there is no ready explanation. The sequence of the coding strand obviously corresponds to the sequence of the RNA transcript that is to be processed. It is possible that the DNA may be assuming a structure similar to the structure of the RNA transcript. The implications of this possibility are discussed in chapter 6.

In summary, functionally important sites on the coding strand are specifically targeted by $Rh(DIP)_3^{3+}$ within both the SV40 and Ad2 introns. On the SV40 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 Ad2 intron DNA, the site adjacent to the branch point represents the strongest cleavage site. The fact that the intensities of cleavage in the two introns are different, though the sites of cleavage are similar, suggests that the two structures are not identical. They may share similar

Figure 3.4 High-resolution mapping of the non-coding strands for $Rh(DIP)_3^{3+}$ photocleavage. Protocol used was identical to the mapping of the coding strand, except for the radio-labeling of opposite strand. "Branch" and "Donor" indicate the nucleotide positions opposite the sites on the coding strand. The left panel, SV40 T-antigen intron non-coding strand: Lane 1, Maxam-Gilbert G reaction; lane 2, Maxam-Gilbert T + C reaction; lanes 3 to 6, fragment after photolysis of the plasmid at 332 nm with $Rh(DIP)_3^{3+}$ for 2, 4, 8, & 12 minutes, respectively; lane 7, fragment after irradiation of the plasmid at 332 nm without $Rh(DIP)_3^{3+}$; lane 8, fragment after incubation of plasmid with $Rh(DIP)_3^{3+}$ without irradiation. No specific cleavage is apparent, and only non-specific G reaction is observed. The right panel, Ad2 E1A intron: Lane 1, fragment after incubation of the plasmid with $Rh(DIP)_3^{3+}$ without irradiation; lane 2, fragment after irradiation of the plasmid at 332 nm; lanes 3 & 4; fragments after photolysis of the plasmid at 332 nm with $Rh(DIP)_3^{3+}$ for 1 and 2 minutes, respectively; lane 5, Maxam-Gilbert G reaction; lane 6, Maxam-Gilbert T + C reaction. The smearing near the region opposite the branch site does not appear to result from specific interactions and is not comparable to the intensity and the specificity of cleavage on the coding strand.

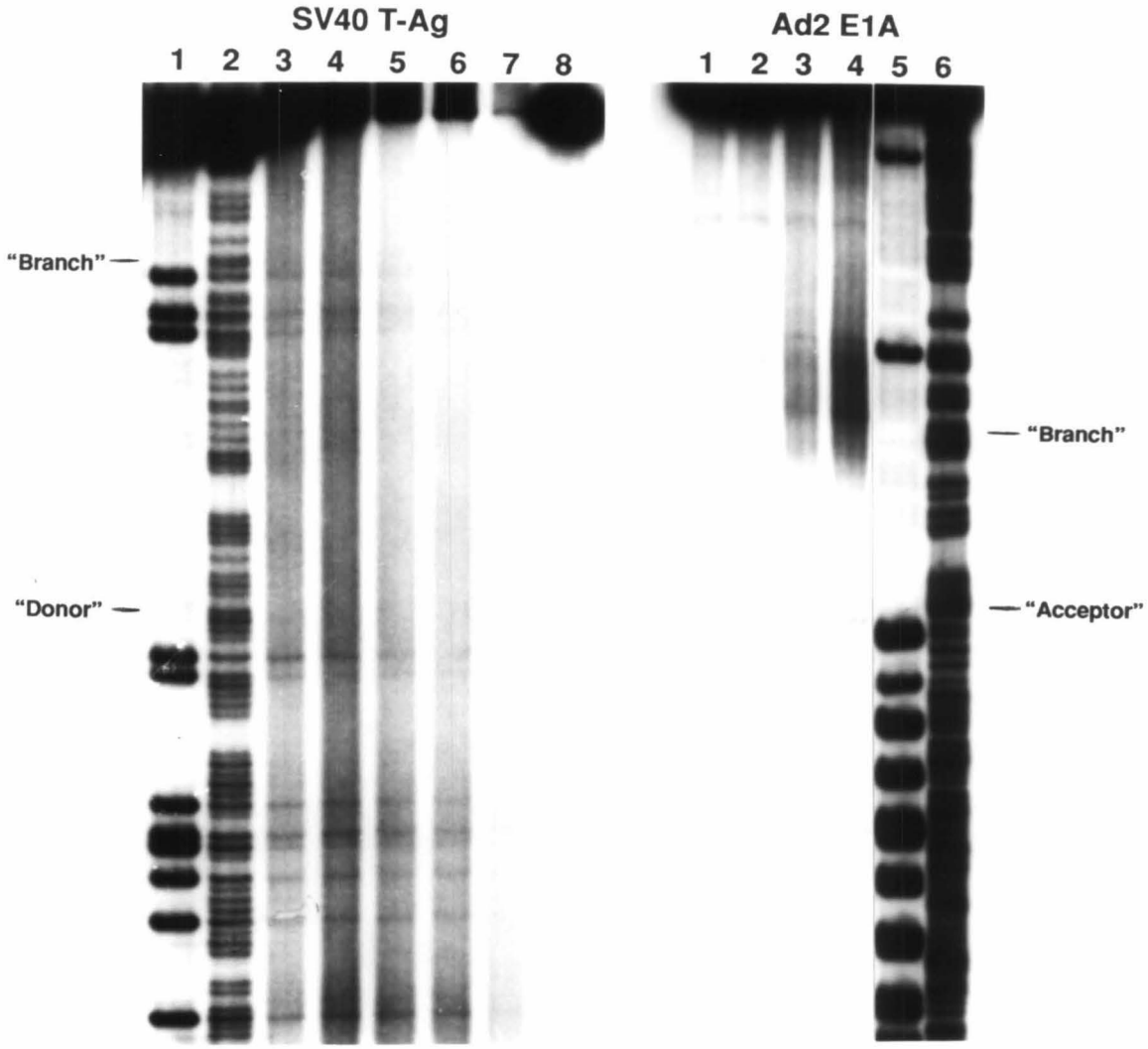
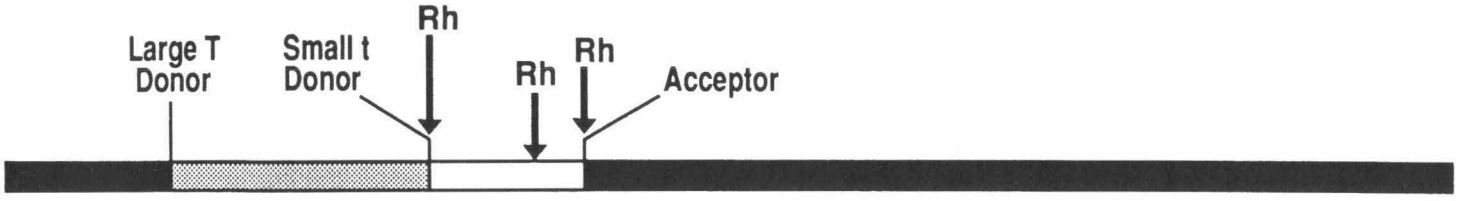
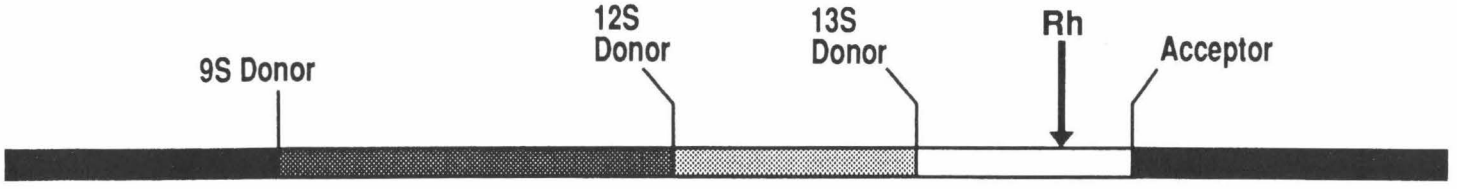

SV40 T-Ag
1 2 3 4 5 6 7 8
"Branch"
"Donor"
Ad2 E1A
1 2 3 4 5 6
"Branch"
"Acceptor"

Figure 3.5 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 and 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.

Simian Virus 40 T Antigen Transcription Unit



Adenovirus 2 E1A Transcription Unit



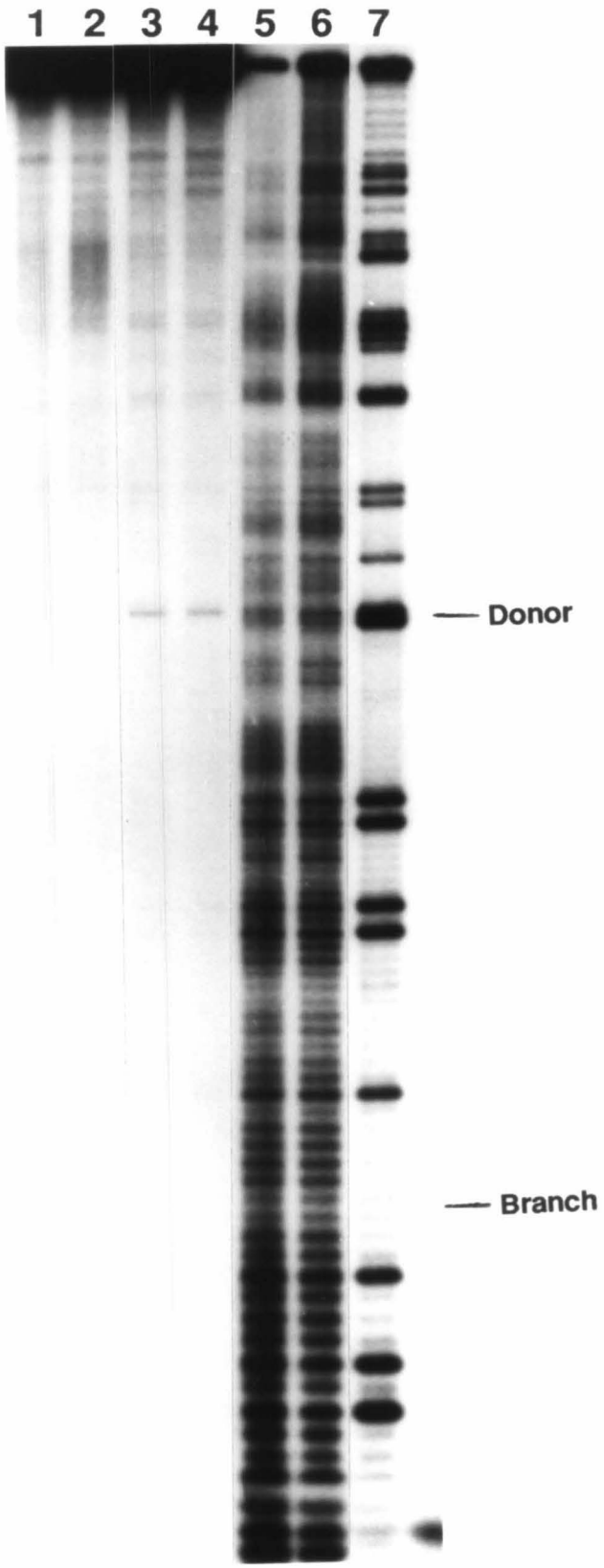
patterns, but the different recognition characteristics indicate that the overall structures are different from each other. The position of the structures relative to the ends of the introns may also be different in the two cases, leading to the observed preference for different positions of cleavage. An important observation is that these sites lack sequence homology but are similar in that they demarcate functionally important elements on the RNA level. These observations are schematically presented in Figure 3.5.

3.3.2 Diethyl pyrocarbonate modification of the introns

In an attempt to probe in more detail the characteristics of the intron structures, diethyl pyrocarbonate (DEPC) was used to detect conformational alterations in the introns in the supercoiled plasmids. DEPC, which modifies the exposed N7 of purines, has been used in the past to probe Z-DNA structures (3). In the current experiments the supercoiled plasmids were modified with DEPC and then digested with the same restriction enzymes used in photocleavage experiments for labeling and isolation of the fragment containing the intron site. The fragment is treated further with piperidine, as in Maxam-Gilbert sequencing, to effect strand scission at the modified nucleotides.

Figure 3.6 shows DEPC modification of the SV40 T-antigen intron. There were no major points of strong sensitivity to DEPC. However, the 3' end of the intron did show a pattern of modification that suggested a perturbed conformation; the sensitivity to DEPC was stronger on either side of the branch point than at the branch point itself or at regions farther away from the branch point. This still does not give us many clues to the specific conformation of the structure(s), but it is consistent with all of the previous observations, which indicate a supercoil-dependent structure in the intron DNA's of SV40 T-antigen and Ad2 E1A genes.

Figure 3.6 DEPC modification of the SV40 T-antigen intron. Lane 1, control fragment; lane 2, fragment after irradiation of plasmid at 332 nm; lanes 3 & 4, fragment after photolysis of the plasmid with $Rh(DIP)_3^{3+}$; lanes 5 & 6, fragment after modification of the plasmid with DEPC for 5 and 10 minutes respectively; lane 7, Maxam-Gilbert G reaction. The intensity of the modification is high on either side of the branch site, indicating a structural perturbation.


1 2 3 4 5 6 7
Donor
Branch

Modifications of the intron DNA with OsO_4 (data not shown) also displayed the same pattern of sensitivity on either side of the branch point.

The results of both the low-resolution and the high-resolution experiments support the initial hypothesis that the rhodium complex is targeting a distinct structure in the intron DNA. It is also consistent with the observations made earlier on the types of structures $Rh(DIP)_3^{3+}$ can recognize. They are dynamic and unusual structures formed under the stress of supercoiling. The intron target exhibits similar characteristics regarding its structural requirements. Additionally, it is more sensitive to changes in salt concentration than the other structures. It appears to be chiefly composed of the coding strand and not the non-coding strand; if the non-coding strand is involved in the structure, it is not being targeted by the rhodium complex. Chemical probes DEPC and OsO_4 did not detect any obvious hyper-sensitivity near the site of rhodium cleavage. This could be due to a tight folding of the structure so that the chemical probes do not have easy access to their points of attack. What emerges from these observations is a distinct and stable, though dynamic, structure formed by the coding strand in the intron DNA of SV40 and Ad2. In the next chapter the structure is characterized further by use of single-stranded DNA fragments corresponding to the coding strands of the introns.

References

1. J. Sambrook, E. F. Fritsch, & T. Maniatis (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed. (Cold Spring Harbor Laboratory Press)

2. A. M. Maxam & W. Gilbert (1980) *Methods Enzymol.*, **65**, 499-560.

3. (a) W. Herr (1985) *Proc. Natl. Acad. Sci. USA* **82**: 8009-8013. (b) M. R. Kirshenbaum (1989) Doctoral Dissertation, Columbia University.

Chapter 4.

Construction and characterization of single-stranded DNA fragments corresponding to the coding strands of the SV40 and Ad2 introns

4.1 Introduction

The sensitivity of rhodium cleavage for the coding strand but not for the non-coding strand indicated that the structure being recognized is formed by the coding strand alone. This would of course necessitate the melting of the two strands before formation of the structure. The dependence of cleavage on supercoiling, as observed in the low-resolution mapping experiments, is consistent with this model of the structure. The formation of the structure unwinds the DNA helix relieving the superhelical stress of the plasmid, and the final folding of the strands into the structure stabilizes the unwound state.

The question that arises from these observations is whether the coding strand alone, in the absence of the complementary non-coding strand, could fold into the structure recognized by the rhodium complex. A short fragment of single-stranded DNA corresponding to part of the coding strand of the genes in question that includes the necessary sequence for the structure may well have a tendency to fold into a stable structure that may resemble or may be identical to the structure formed in the double-stranded supercoiled plasmids. This possibility was investigated by synthesizing a series of single-stranded DNA fragments for both the SV40 and the E1A introns and testing them for sensitivity to rhodium complex cleavage. The results were very interesting: only the longest fragments containing the entire intron and the flanking exon sequences were cleaved specifically by the rhodium complex, indicating that exon sequences are involved in the structures. The identical position of cleavage in the ssDNA as compared to the cleavage in supercoiled plasmids

suggests the structures formed by the ssDNA's are very similar, if not identical, to the structures in the plasmids. This confirms the hypothesis that the coding strand alone is responsible for the specific recognition by the rhodium complex.

Further experiments with ssDNA fragments with deletions in the middle of the intron show that the essential sequences for a structure that can be specifically targeted by the metal complex are the ends of the intron and adjacent sequences of the flanking exons. Investigation of the structure of the ssDNA fragments were continued with enzymatic and chemical probes. The results indicate that the entire structure is composed of two stem and loop structures, each of which are formed by both exon and intron sequences; these two parts come together to build the final structure which provides a target site for the metal complex.

4.2 Experimental

Materials:

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 4-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 on an ABI DNA synthesizer. The 114-mer was purified on the NENSorb column (New England Nuclear). The 70-mer and the 85-mer were purified by HPLC (Waters).

Four ssDNA fragments were synthesized also for the T-antigen intron: a 136-mer containing the intron plus 35 nucleotides of flanking exon sequences at either end; a 116-mer containing the intron plus 25 nucleotides at either end; a 66-mer representing the complete intron; and a 40-mer representing the 3' two thirds of the intron, including the branch site and the $Rh(DIP)_3^{3+}$ cleavage site on the supercoiled DNA. All of these fragments were synthesized chemically on an ABI DNA synthesizer.

Restriction enzymes and labeling enzymes were from Boehringer Mannheim, Bethesda Research Laboratories, or New England Biolabs. [α-^{32}P]dideoxy-ATP was from Amersham, and [γ-^{32}P]ATP was from New England Nuclear-Dupont. Molecular biology grade reagents for buffers were from Sigma or Boehringer Mannheim. Reagents for Maxam-Gilbert sequencing were from Aldrich. AMT (4'-aminomethyl-4,5',8-trimethylpsoralen) was from HRI associates.

Instrumentation:

The light source used in photocleavage was a 1000 W Oriel Hg/Xe lamp (model 6140) fitted with a monochromator and a 300 nm cut-off filter. Quantitation of DNA was carried out by UV-VIS absorption spectrometry with Cary 219 Spectrophotometer.

Methods:

The 174-mer is labeled as follows: The 260 bp PCR fragment is labeled with

^{32}P at the 5' end using polynucleotide kinase. It is digested with Ava I to generate the 174 bp fragment, with 4 nucleotide 5' overhang, and the 86-mer, which are separated on a non-denaturing prep gel. The 174 bp fragment is then electro-eluted and run on a denaturing polyacrylamide gel to separate the two strands. All other ssDNA fragments are labeled with ^{32}P at the 5' end using polynucleotide kinase or at the 3' end using terminal deoxynucleotidyl transferase. All the labeled fragments are purified again on denaturing polyacrylamide gels.

The irradiation mixture contained, in a total volume of 20 μl of buffer (20 mM Tris-HCl, pH 7.4, 25 mM NaCl), labeled fragment and carrier DNA at 100 μM nucleotides and $Rh(DIP)_3^{3+}$ at 1-10 μM. The mixture was irradiated for 2 to 8 minutes at 332 nm with the Hg/Xe lamp. Then the DNA is ethanol precipitated twice and run on analytical denaturing polyacrylamide gels together with samples sequenced by the Maxam-Gilbert method.

Mung bean nuclease digestion is carried out in the same buffer used for photocleavage experiments. One to two units of the enzyme is added to the DNA, and the mixture is incubated at room temperature for 5 to 10 minutes. The samples are ethanol precipitated and analyzed on denaturing polyacrylamide gels.

Mse I digestion is carried out using the buffer required for the enzyme, which contains 50 mM NaCl. Incubation is at room temperature for 5 to 10 minutes, and the samples are analyzed on denaturing polyacrylamide gels.

Psoralen crosslinking is also carried out in the same buffer used for photo-cleavage experiments. Psoralen dissolved in water as a stock is added to 5 to 10 μg/ml to the reaction mixture in a total volume of 20 μl. Labeled DNA and carrier DNA are present at 25 to 50 μM nucleotides. The samples are irradiated with the Hg/Xe lamp at 340 nm for two to four minutes and ethanol precipitated. They are then treated with hydrazine and piperidine as in the usual Maxam-Gilbert

sequencing reactions and analyzed on denaturing polyacrylamide gels.

4.3 Results and discussion

4.3.1 $Rh(DIP)_3^{3+}$ cleavage of ssDNA fragments

$Rh(DIP)_3^{3+}$ cleavage of Ad2 E1A ssDNA fragments:

Initially three single-stranded DNA fragments (174-mer, 114-mer, and 70-mer) of the E1A 13S intron were prepared (Figure 4.1), 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 the flanking sequence 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 4.2. The two shorter fragments show the secondary, non-structure-specific, reaction of the metal complex with G residues, 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.

An interesting observation was made while purifying the 174-mer on denaturing polyacrylamide gels. The labeled 174-mer always ran as two separate bands, one at the expected place for a single-stranded fragment of that length and

Figure 4.1 Schematic illustration of the Ad2 E1A single-stranded DNA fragments. The top bar represents the intron and the flanking exons. Primers were used as indicated to synthesize the double stranded fragment through polymerase chain reaction. The 174-mer was generated by digesting this dsDNA with Ava I and then separating the two strands on a denaturing polyacrylamide gel. All of the other ssDNA fragments were synthesized chemically using an automated DNA synthesizer.

Ad2 E1A 13S Intron

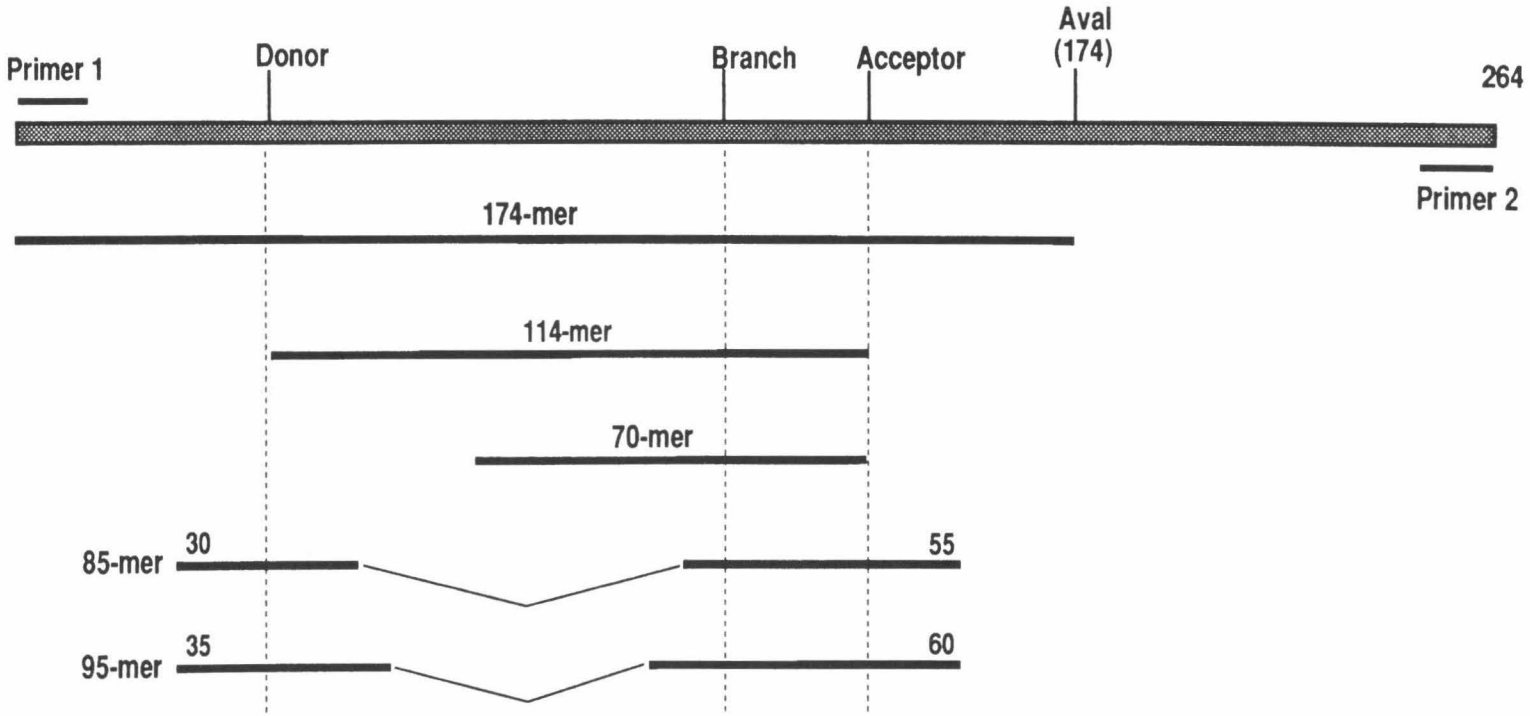
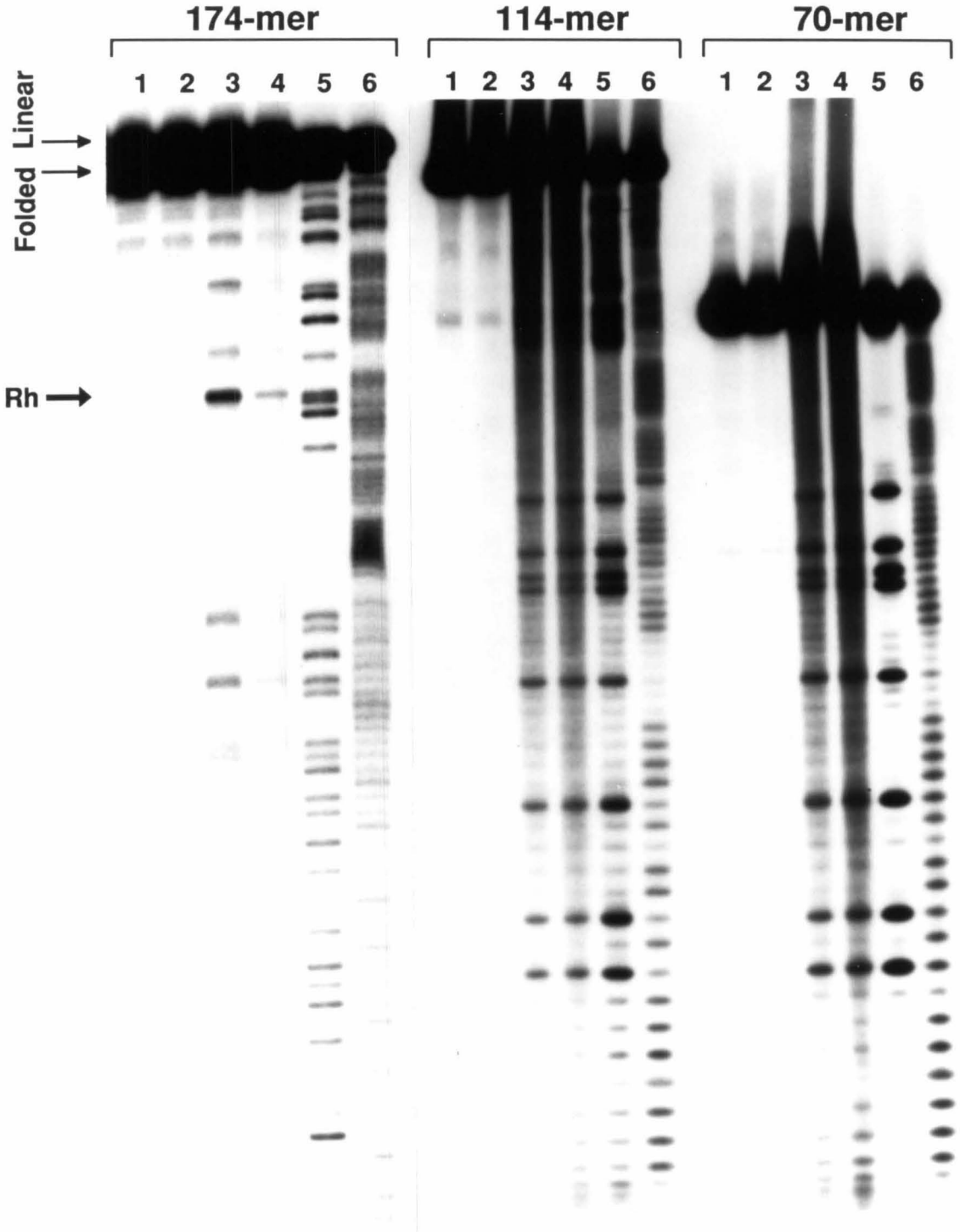


Figure 4.2 Structural probing of single-stranded DNA fragments of the E1A intron. The 174-mer, which consists of the entire intron plus exon flanking sequences at both ends, is cleaved specifically by $Rh(DIP)_3^{3+}$ (lanes 3 and 4) while the 114-mer (entire intron but with no flanking sequences) and 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 4 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 minutes in the presence of $Rh(DIP)_3^{3+}$; lanes 4 for the 114-mer and the 70 mer, fragment after irradiation for 4 minutes in the presence of $Rh(DIP)_3^{3+}$; lane 5, Maxam-Gilbert G reaction; lane 6, Maxam-Gilbert T + C reaction;

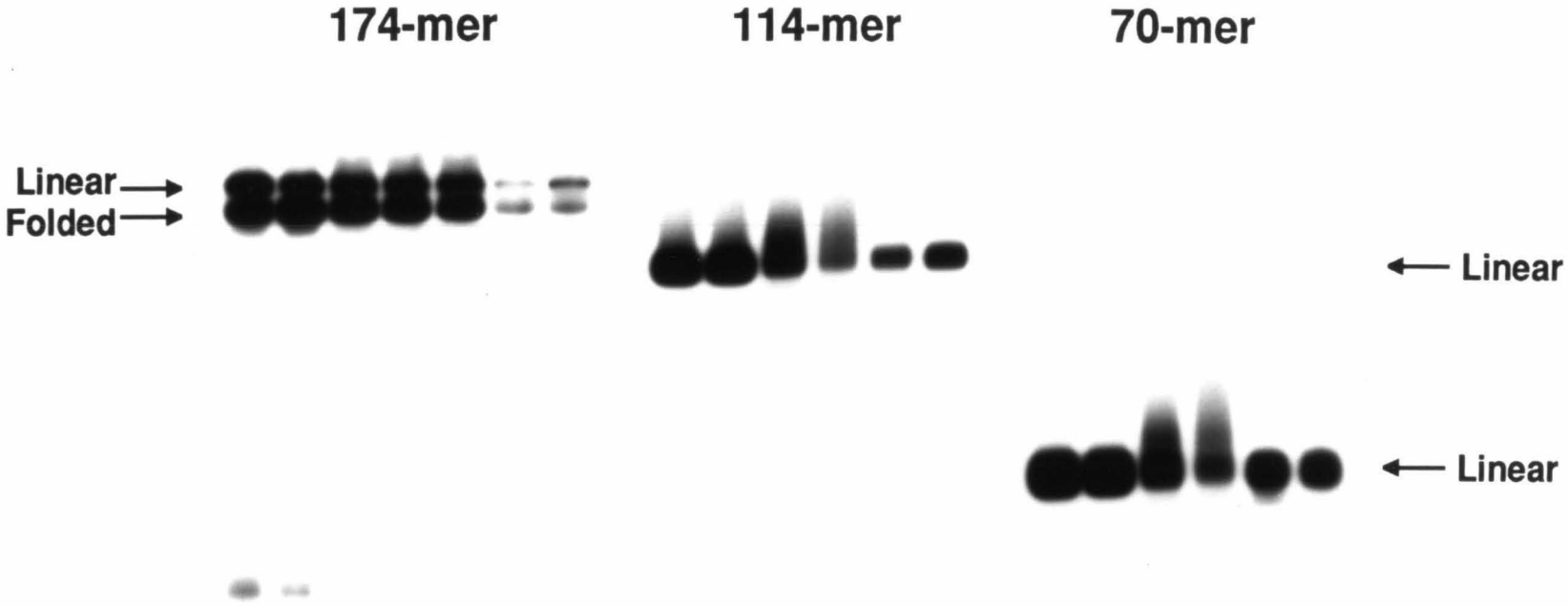

174-mer
114-mer
70-mer
1 2 3 4 5 6
1 2 3 4 5 6
1 2 3 4 5 6
Linear
Folded
Rh

another at a place expected for a much shorter fragment. When these two bands were excised from the gel and sequenced, they were shown to be identical. When the DNA 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. This property of the 174-mer is illustrated clearly in Figure 4.3, a lightly exposed autoradiogram of a denaturing polyacrylamide gel showing the folded form of the 174-mer both in the presence and absence of $Rh(DIP)_3^{3+}$. No similar high mobility bands were apparent with the shorter fragments.

Importantly the 174-mer, but not the 114-mer corresponding to the full intron, folds into this discrete structure. The intron DNA itself is not sufficient to form the structure; flanking sequences within the exon are also required. This indicates again the possibility that the DNA structure represents a similar structure in the pre-mRNA which has the intron and the exon sequences as in the DNA. It is known that exon sequences do participate in splicing by pairing with portions of snRNA's (1). The structure of pre-mRNA before the assembly of the spliceosome may be of relevance to the structure of ssDNA observed in the 174-mer. Though the dynamics of RNA structure and DNA structure are different, their secondary base-pairing may be similar or even identical, given the same primary sequence. This has been shown for at least one tDNA molecule which has the same secondary structure as its tRNA counterpart (2).

The observation of the folded form of the 174-mer in the absence of $Rh(DIP)_3^{3+}$ shows that the metal complex is not necessary to induce the folding of the structure. This is an important aspect of the structures that are targeted by metal complexes such as $Rh(DIP)_3^{3+}$. The interaction between the structure and the metal

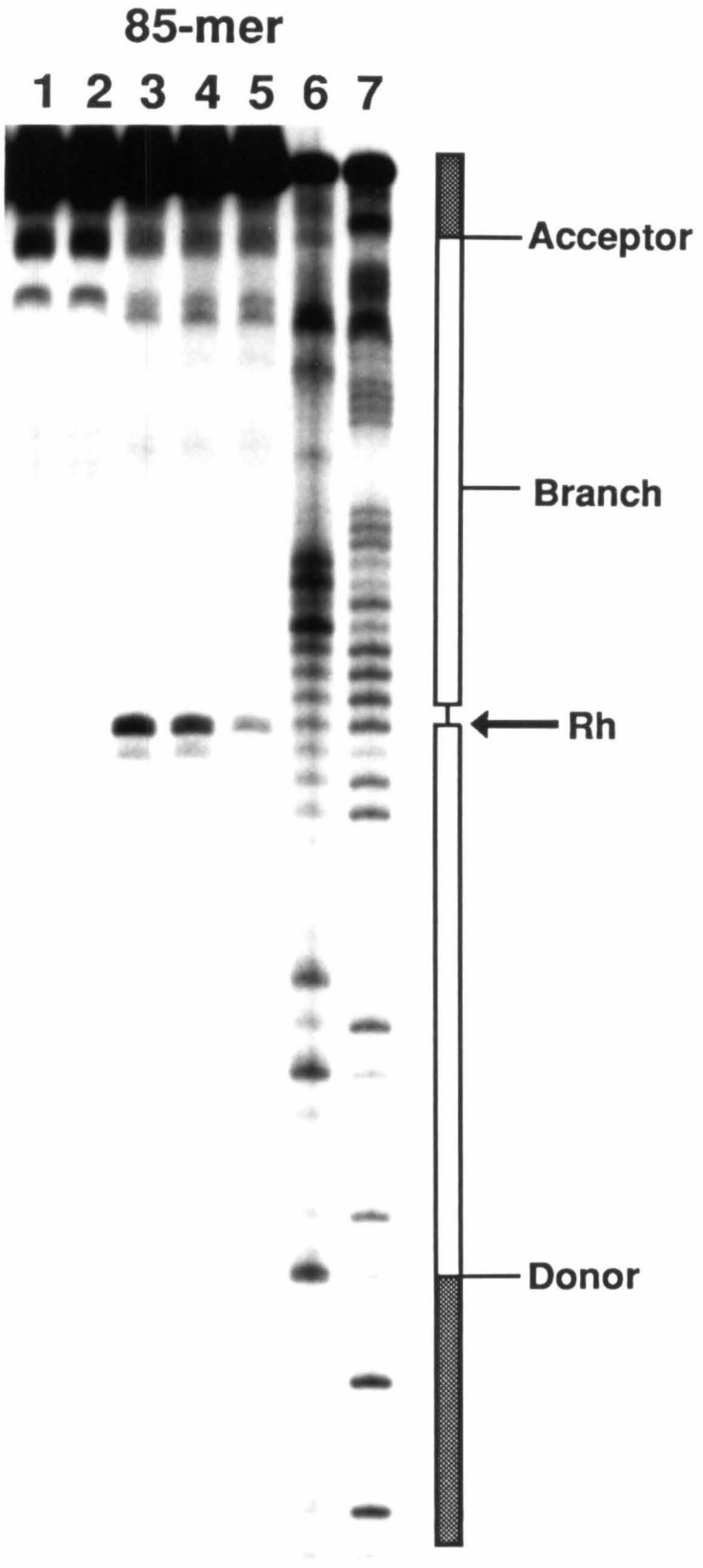
Figure 4.3 Low exposure autoradiogram showing the migration patterns of the three ssDNA fragments in a denaturing polyacrylamide gel. The 174-mer runs as two separate bands. The two bands can be excised and the DNA isolated and sequenced to show identity. The DNA from the two bands, when reloaded on 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.


174-mer
114-mer
70-mer
Linear
Folded
Linear
Linear

complex is based on the already existing structure, and the effects of the interaction on the DNA structure is independent of the initial recognition. The metal complex may, however, stabilize the structure by "locking" it with the molecular interactions and thus shift the equilibrium toward the folded form of the DNA.

To examine the minimum sequences needed for folding of the structure, another single-stranded DNA fragment was synthesized. It consisted of sequences at the 5' 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. The idea behind this construction was that since the rhodium cleavage occurred at the branch site which is near the 3' end of the intron, the sequences at the 3' end might be interacting with the sequences at the 5' end, again much as the potential pre-mRNA structure might be. This DNA was found to be cleaved specifically by the rhodium complex as shown in Figure 4.4. In this case, however, the cleavage observed is 7 bases to the 5' side of the site cleaved on the supercoiled and 174-mer DNA fragments (or 10 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 4.4 $Rh(DIP)_3^{3+}$ 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.


85-mer
1 2 3 4 5 6 7
Acceptor
Branch
Rh
Donor

In constructing the 85-mer we were investigating the possibility of the interaction of the two ends of the introns as a source of the structural perturbation observed in the intron DNA. The positive rhodium cleavage of the 85-mer suggested that this may be the case. To test this hypothesis further, the two portions of the 85-mer were separated into a 35-mer and a 55-mer which correspond to the 5' end and the 3' end of the intron respectively with sequences of the flanking exons. The 55-mer and the 35-mer were tested for cleavage by the rhodium complex, and no specific cleavage could be observed on these DNA's either singly or mixed together (data not shown).

The observation that the rhodium cleavage of the 85-mer occurred six bases 5' to the site on the 174-mer suggested an altered or incomplete structure was forming on the 85-mer as compared to the 174-mer. To test this theory another ssDNA fragment was constructed. It was 95 nucleotides in length and consisted of the entire 85-mer plus ten more nucleotides in the middle, five on each side, which are then covalently linked in the final ssDNA. Interestingly this fragment did not show any specific cleavage by $Rh(DIP)_3^{3+}$ (Figure 4.5), indicating that the structure was lacking the necessary conformation required for the recognition by the metal complex. At this point it is not clear why the 95-mer is not specifically targeted by the metal complex while the 85-mer is. The most likely explanation is that the structure forming in the middle of the ssDNA fragments requires a precise arrangement of the various components which make up the global structure. Thus the introduction of the ten extra bases in the middle hinders the correct folding of the entire structure. The components, however, may still be folding independently in the 95-mer. This possibility is investigated further in later sections.

The results so far, schematically illustrated in Figure 4.6, show that indeed the structure recognized by the metal complex requires the two ends of the introns.

Figure 4.5 Photocleavage of the E1A 95-mer with $Rh(DIP)_3^{3+}$ and $Rh(phen)_2phi^{3+}$ Lane 1, control fragment; lane 2, fragment irradiated for 2 minutes at 332 nm; lane 3 & 4, fragments photolyzed with $Rh(DIP)_3^{3+}$ at 332 nm for 1 minute and 2 minutes respectively; lanes 5 & 6; fragments photolyzed with $Rh(phen)_2phi^{3+}$ at 365 nm for 1 minute and 2 minutes respectively. $Rh(DIP)_3^{3+}$ does not cleave the 95-mer specifically, while $Rh(phen)_2phi^{3+}$ cleaves at several sites specifically.

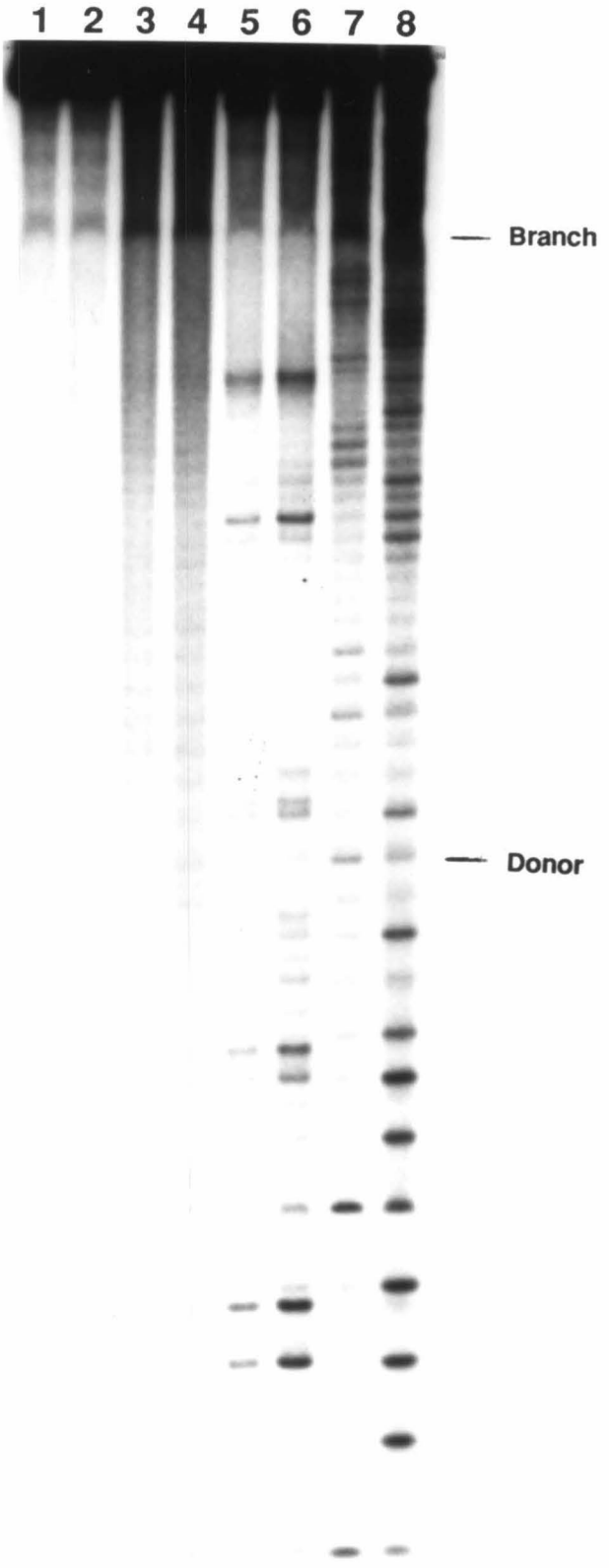
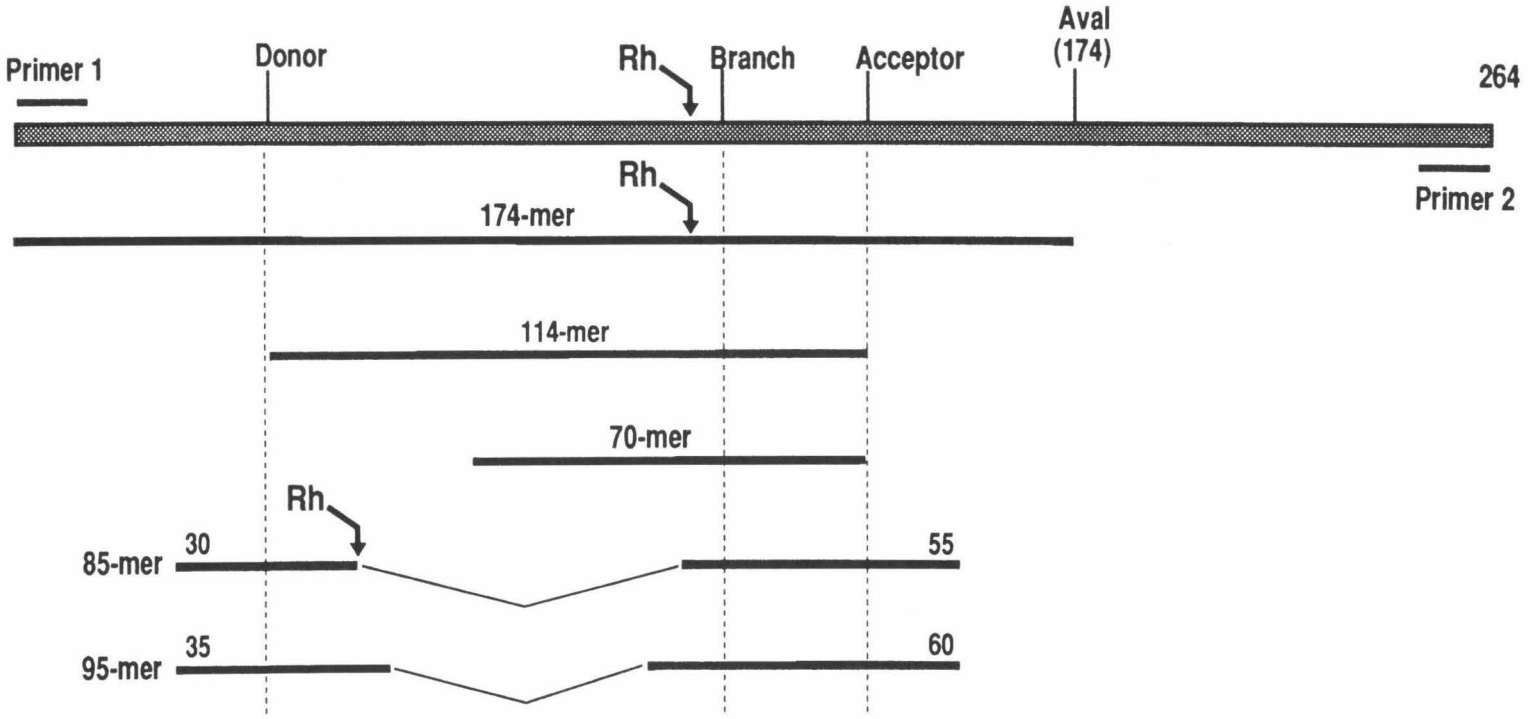

1 2 3 4 5 6 7 8
Branch
Donor

Figure 4.6 Schematic illustration of results from structural probing of the single-stranded DNA fragments of the E1A intron coding strand. The shaded bar at top represents a portion of the E1A gene around the 13S intron. The cleavage site on the supercoiled plasmid, 4 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 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 7 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.

Ad2 E1A 13S Intron



Whether the two ends fold into independent units which then interact to form the final structure, or the two ends are involved in more integral interactions is not answered by the results so far. Further probing of the ssDNA structure was conducted to answer this question and to arrive at a plausible model for the structure.

$Rh(DIP)_3^{3+}$ cleavage of SV40 T-antigen ssDNA fragments:

For the SV40 T-antigen intron three ssDNA fragments, comparable to the three made for the Ad2 intron, were initially made (Figure 4.7). They were a 116-mer containing the intron plus 25 nucleotides at either end; a 66-mer representing the complete intron; and a 40-mer representing the 3' two thirds of the intron, including the branch site and the $Rh(DIP)_3^{3+}$ cleavage site on the supercoiled DNA. These were tested for specific cleavage by $Rh(DIP)_3^{3+}$ under the same conditions as the Ad2 E1A ssDNA fragments. None of them showed any specific cleavage by the metal complex (Figure 4.8). This was a puzzling observation since in low-resolution experiments, the SV40 intron shows a stronger cleavage than the Ad2 intron, and the design of the ssDNA fragments for the two introns were identical in scheme. The 116-mer was expected to show specific cleavage as was the 174-mer for the Ad2 intron. One possible reason for the apparent lack of structure in the 116-mer could be that there wasn't enough of the exon sequences at the ends.

Therefore, a new fragment, a 136-mer, containing the intron plus 35 nucleotides of flanking exon sequences at either end was synthesized. This ssDNA was indeed cleaved specifically by $Rh(DIP)_3^{3+}$ (Figure 4.9). Moreover, it was cleaved at exactly the same nucleotide position as one of the sites the supercoiled plasmid was cleaved, the donor site. No cleavage was apparent, however, at the branch site, which is the weaker cleavage site in the supercoiled plasmid DNA. The ssDNA

Figure 4.7 Schematic illustration of the SV40 T-antigen single-stranded DNA fragments. The top bar represents the intron and the flanking exons. The solid lines represent the ssDNA fragments synthesized by automated solid-phase chemical synthesis.

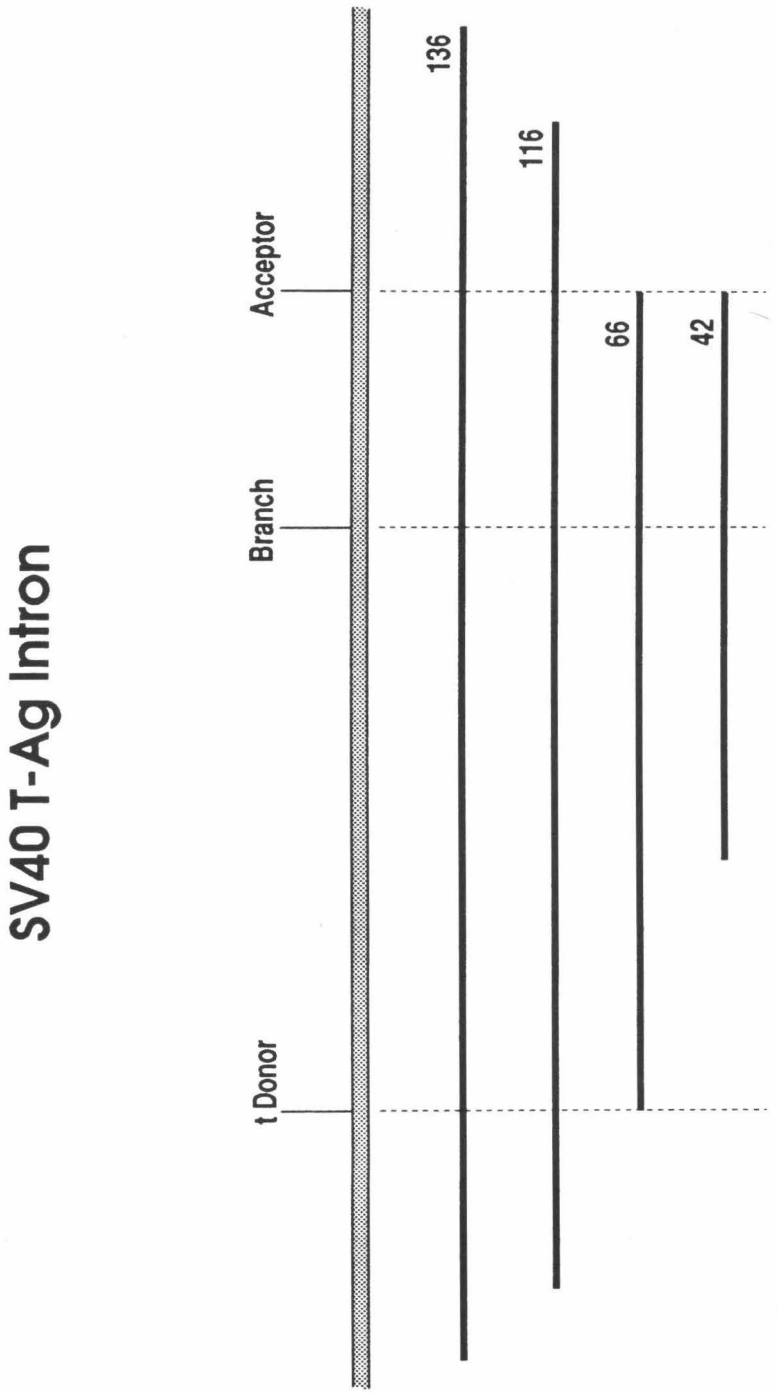

SV40 T-Ag Intron
t Donor
Branch
Acceptor
136
116
66
42

Figure 4.8 Structural probing of the ssDNA fragments of the SV40 T-antigen intron. The three initial ssDNA fragments for the intron were photolyzed with $Rh(DIP)_3^{3+}$ at 332 nm. None of the fragments showed any specific cleavage. All showed only the non-specific G reaction of the metal complex. For the 40-mer and the 66-mer: Lanes 1, fragments incubated with $Rh(DIP)_3^{3+}$ without photolysis; lanes 2, fragments irradiated at 332 nm for 4 minutes without $Rh(DIP)_3^{3+}$; lanes 3, 4, & 5, fragments photolyzed with $Rh(DIP)_3^{3+}$ at 332 nm for 1, 2, and 4 minutes respectively; lanes 6, Maxam-Gilbert G reaction; lanes 7, Maxam-Gilbert T + C reaction. For the 116-mer: lane 1, Maxam-Gilbert G reaction; lane 2, Maxam-Gilbert T + C reaction; lane 3, fragment incubated with $Rh(DIP)_3^{3+}$ without photolysis; lane 4, fragment irradiated at 332 nm for 4 minutes without $Rh(DIP)_3^{3+}$; lanes 5, 6, & 7, fragments photolyzed with $Rh(DIP)_3^{3+}$ at 332 nm for 1, 2, and 4 minutes respectively.

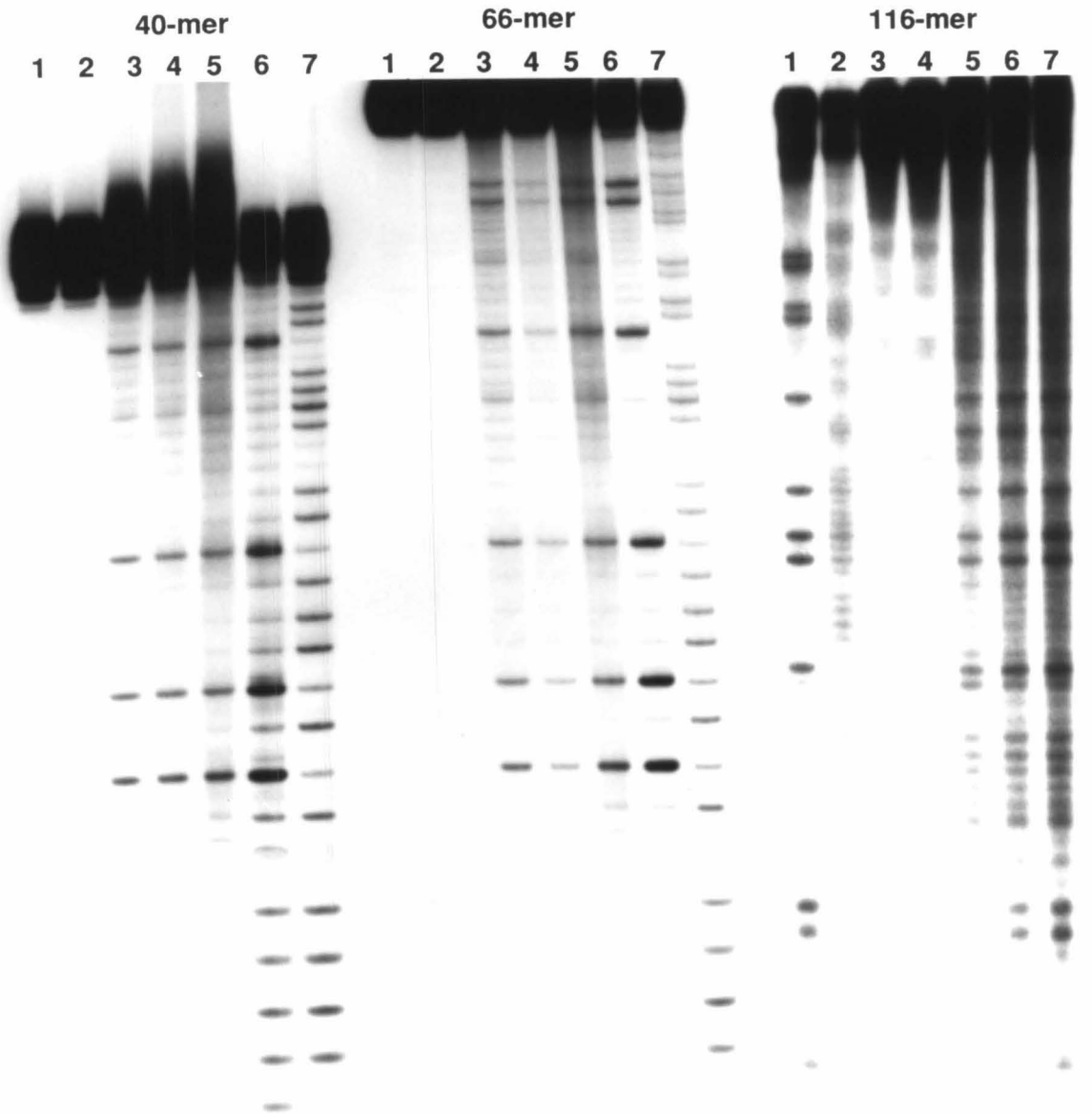

40-mer
1 2 3 4 5 6 7
66-mer
1 2 3 4 5 6 7
116-mer
1 2 3 4 5 6 7

Figure 4.9 Structural probing of the 136-mer for the SV40 T-antigen intron. A. 5'-end labeled 136-mer probed with $Rh(DIP)_3^{3+}$ and $Rh(phen)_2phi^{3+}$: Lane 1, control fragment; lane 2, fragment irradiated at 332 nm for 4 minutes without metal complex; lanes 3 & 4, fragment photolyzed with $Rh(DIP)_3^{3+}$ at 332 nm for 2 and 4 minutes respectively; lanes 5 & 6, fragment photolyzed with $Rh(phen)_2phi^{3+}$ at 365 nm for 2 and 4 minutes respectively; lane 7, Maxam-Gilbert G reaction; lane 8, Maxam-Gilbert T + C reaction. B. 3'-end labeled 136-mer: Lane 1, control fragment; lane 2, fragment irradiated at 332 nm for 4 minutes without metal complex; lane 3, fragment photolyzed with $Rh(DIP)_3^{3+}$ at 332 nm for 4 minutes; lane 4, fragment photolyzed with $Rh(phen)_2phi^{3+}$ at 365 nm for 4 minutes; lane 5, fragment digested with mung bean nuclease; lane 6, Maxam-Gilbert G reaction; lane 7, Maxam-Gilbert T + C reaction.

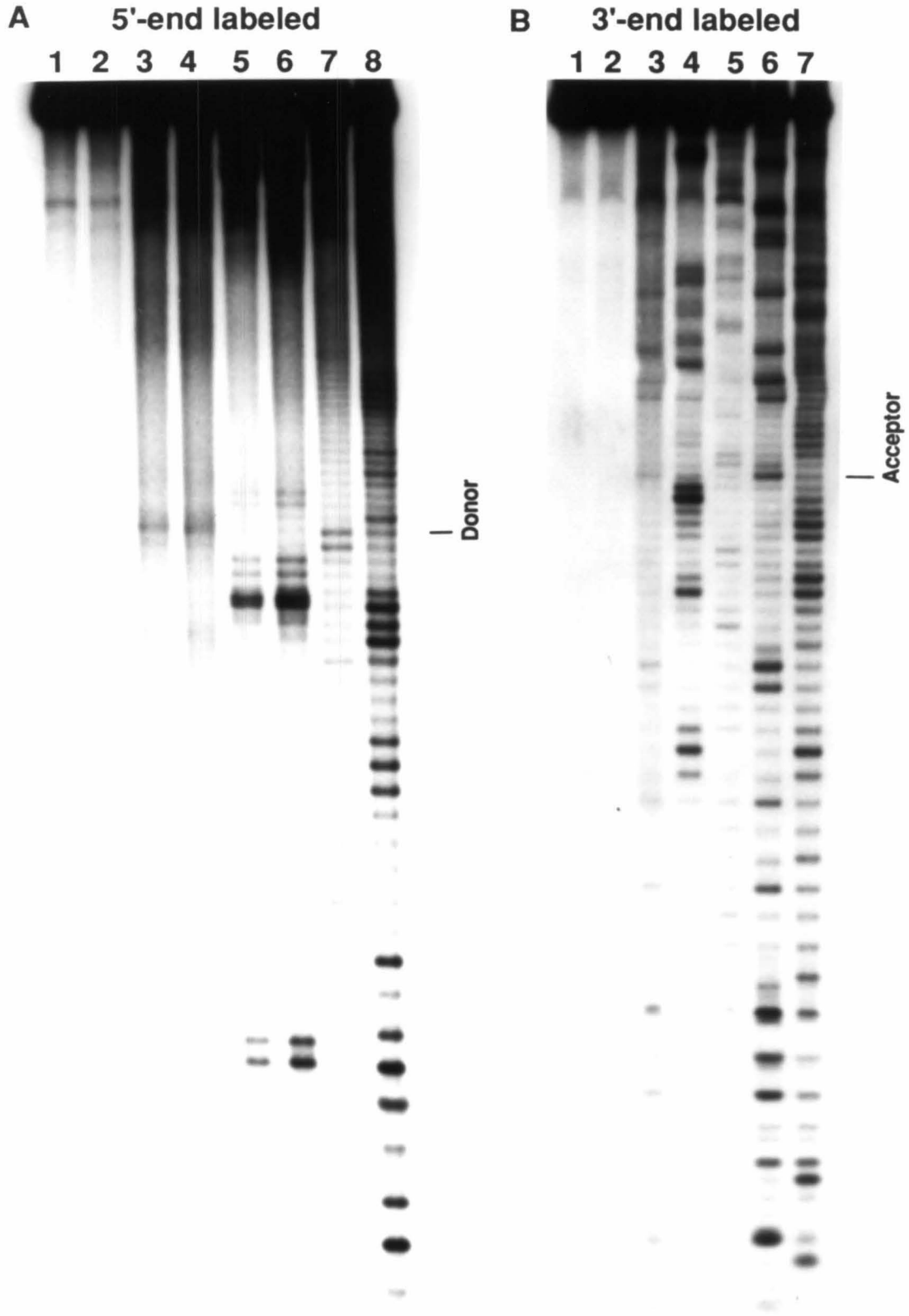
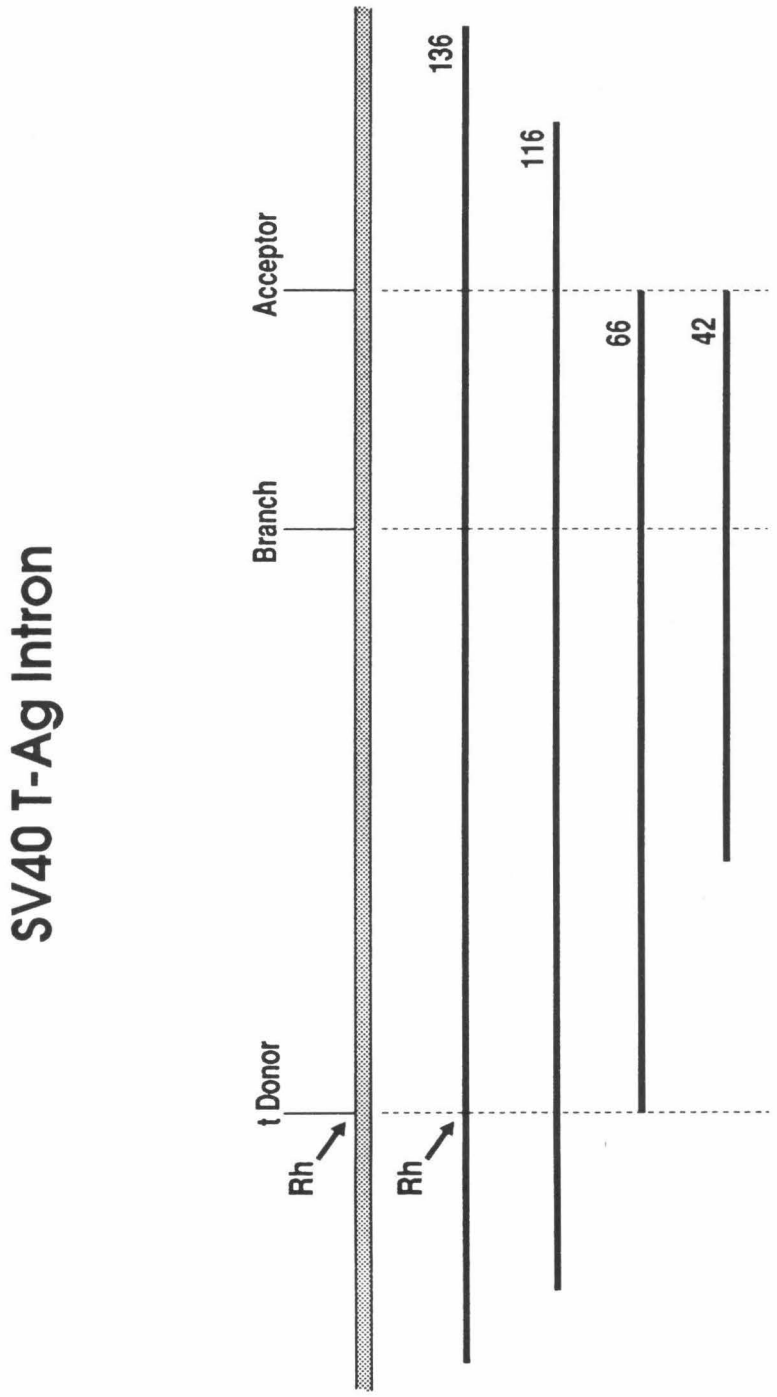

A
5'-end labeled
1 2 3 4 5 6 7 8
Donor
B
3'-end labeled
1 2 3 4 5 6 7
Acceptor

Figure 4.10 Schematic illustration of the structural probings of the SV40 T-antigen intron ssDNA fragments. The top bar represents the intron with its flanking exon sequences, and the solid lines represent the various ssDNA fragments. Cleavage by $Rh(DIP)_3^{3+}$ (denoted by "Rh") is seen in the dsDNA when supercoiled and on the longest ssDNA fragment, the 136-mer.


SV40 T-Ag Intron
t Donor
Branch
Acceptor
Rh
Rh
136
116
66
42

again appears to be folding into a structure similar, or identical, to the structure observed in the supercoiled dsDNA. The results for the SV40 intron are schematically illustrated in Figure 4.10.

That the two intron sequences behave in the same manner with respect to the rhodium cleavage suggests that they may be functionally equivalent. The sequences themselves, though not homologous, do code for functionally equivalent sites in RNA splicing. The structure assumed by the intron DNA's at these sites may also function in an equivalent manner. Whether or not this is at all true has not been established, and this thesis does not attempt to prove this function. Instead, the occurrence of the structures is considered significant, and their implications are explored.

4.3.2 $Rh(phen)_2phi^{3+}$ cleavage of the ssDNA fragments

$Rh(phen)_2phi^{3+}$ is a metal complex with a highly intercalative phi (phenanthrenequinone diimine) ligand and two phenanthroline ligands (3). It targets double stranded DNA sequences of the type 5'-pyr-pyr-pur-3'. The specificity for this sequence is based on the shape of the double helix at that site, which shows an opening toward the major groove due to the propeller twisting of the bases (4, 5). The intercalation of the phi ligand is hindered at sites where the major groove is narrow by the steric clash of hydrogens on the ancillary phenanthroline ligands with the bases above and below the intercalation site. An open major groove relieves this steric clash, and the intercalation is facilitated. This leads to the specificity seen in the cleavage of double-stranded DNA. $Rh(phen)_2phi^{3+}$ also cleaves tRNA specifically, but only at sites of triple base interactions and structured loops but not at double-stranded sites which are A-form (6). Thus $Rh(phen)_2phi^{3+}$ could be used to probe for B-form double-stranded sites in a DNA molecule of unknown structure, and it might

also detect tertiary structures in DNA, as it does in RNA, which are not simply double-stranded but may have tertiary interactions. $Rh(phen)_2phi^{3+}$ was used on the ssDNA fragments that showed positive cleavage and on some that did not compare with the results of the $Rh(DIP)_3^{3+}$ cleavage experiments.

For the E1A intron, the 174-mer (Figure 4.11) and the 85-mer (Figure 4.12 & 4.13) were initially tested for specific cleavage by $Rh(phen)_2phi^{3+}$. Both were cleaved specifically at several sites. There were distinct and reproducible cleavage sites in the exon sequences just beyond the 5' end or the 3' end of the intron. The sequences cleaved are 5'-T**CC**TG-3' and 5'-GT**T**TGT**C**TA-3' for the 85-mer and 5'-GT**C**TG-3' and 5'-GT**C**TA**C**AG-3', where the cleavage sites are highlighted. These cleavages are all at 5'-pyr-pyr-pur-3' sequences, indicating double stranded structures at either end of the fragments. However, cleavage at most of these sites were not limited to the one nucleotide but was spread over two or three nucleotides with the strongest cleavage occurring at the highlighted residue. This is not consistent with the intercalation model of the interaction of the complex with double-stranded DNA, which produces strong and unique cleavages. It is possible that the structures at these sites are involved in tertiary interactions.

$Rh(phen)_2phi^{3+}$ also cleaved, in both the 174-mer and the 85-mer, several bases 3' to the $Rh(DIP)_3^{3+}$ cleavage site. For the 85-mer the sequence cleaved was 5'-TT**G**TG-3', and for the 174-mer it was 5'-TA**TT**GT-3', where the cleavage sites are again highlighted. The sequence of the 85-mer does not conform to the usual double-stranded site for $Rh(phen)_2phi^{3+}$, and the cleavage in the 174-mer is spread over two residues. This suggests, along with the fact that these sites are adjacent to the $Rh(DIP)_3^{3+}$ cleavage sites, that the phi complex may be recognizing in this case a tertiary structure rather than regular double helices.

Figure 4.11 $Rh(phen)_2phi^{3+}$ and EDTA-Fe(II) cleavage of the 174-mer. Lane 1, Maxam-Gilbert G reaction; lane 2, Maxam-Gilbert T + C reaction; lane 3, control fragment; lane 4, fragment irradiated at 365 nm for 4 minutes; lane 5, fragment photolyzed with $Rh(phen)_2phi^{3+}$ at 365 nm for 4 minutes; lanes 6 & 7, fragment cleaved with EDTA-Fe(II) at room temperature for 10 and 20 minutes respectively.

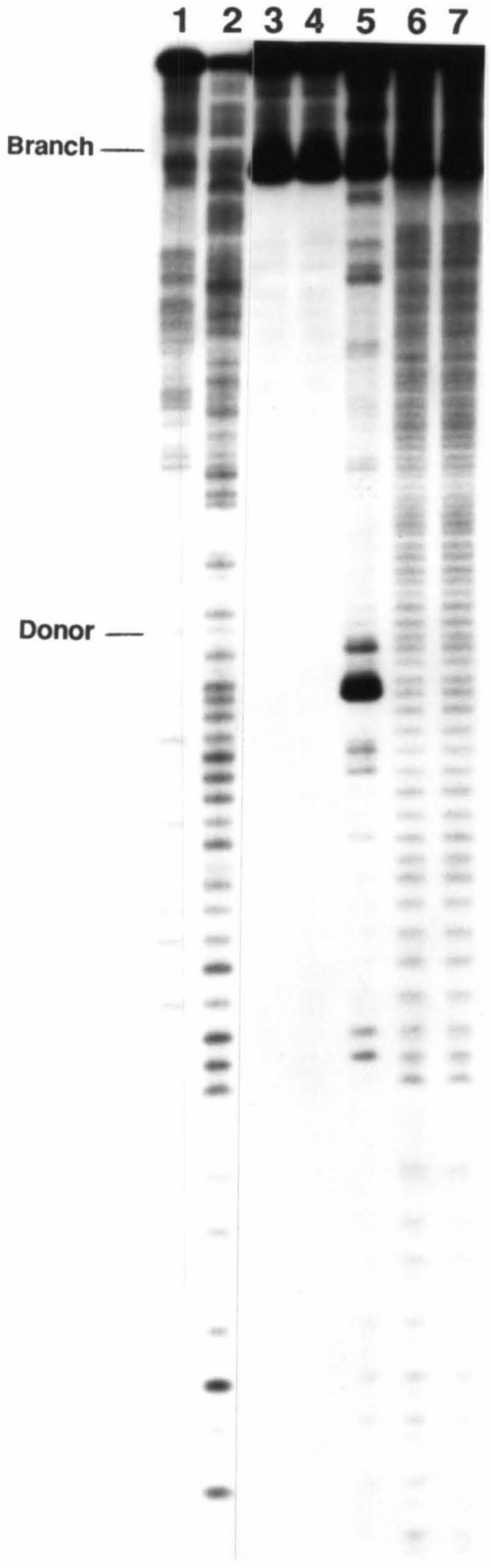

1 2 3 4 5 6 7
Branch
Donor

Figure 4.12 Structural probing of the 55-mer, the 85-mer, and the 95-mer with various probes. 5'-end labeled fragments. Lanes 1, control fragments; lanes 2, fragments irradiated at 332 nm without metal complex; lanes 3, fragments photolyzed with $Rh(DIP)_3^{3+}$ at 332 nm; lanes 4, fragments photolyzed with $Rh(phen)_2phi^{3+}$ at 365 nm; lanes 5, fragments digested with Mse I; lanes 6, fragments digested with mung bean nuclease; lanes 7, Maxam-Gilbert G reaction; lanes 8, Maxam-Gilbert T + C reaction. The various probes show structural similarities among the ssDNA fragments. Note that the 55-mer is not cleaved at all by $Rh(DIP)_3^{3+}$, and the 85-mer and the 95-mer are cleaved at identical sites at the 5' end by $Rh(phen)_2phi^{3+}$. Discussions of Mse I and mung bean nuclease digestions follow in later sections.

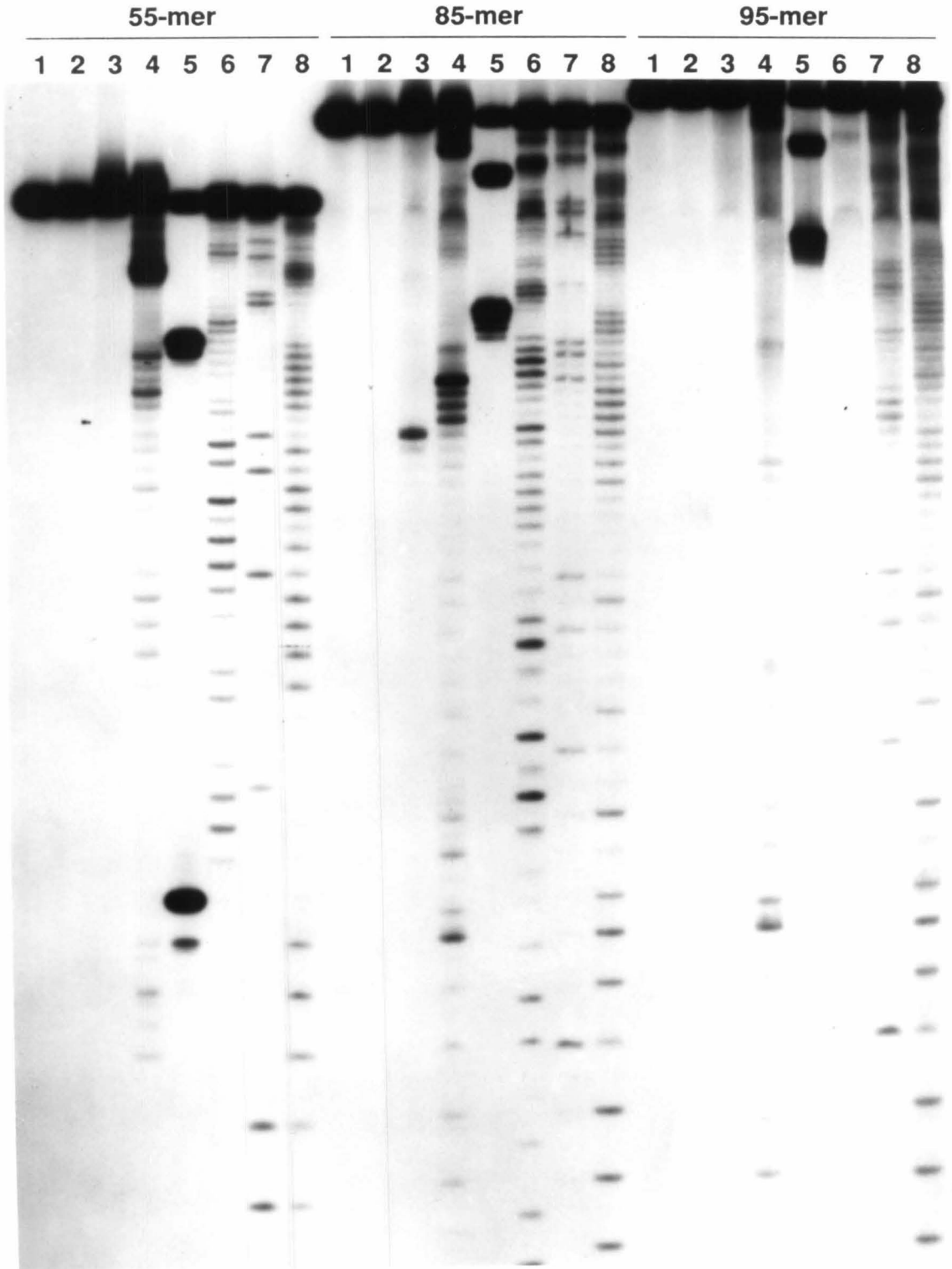
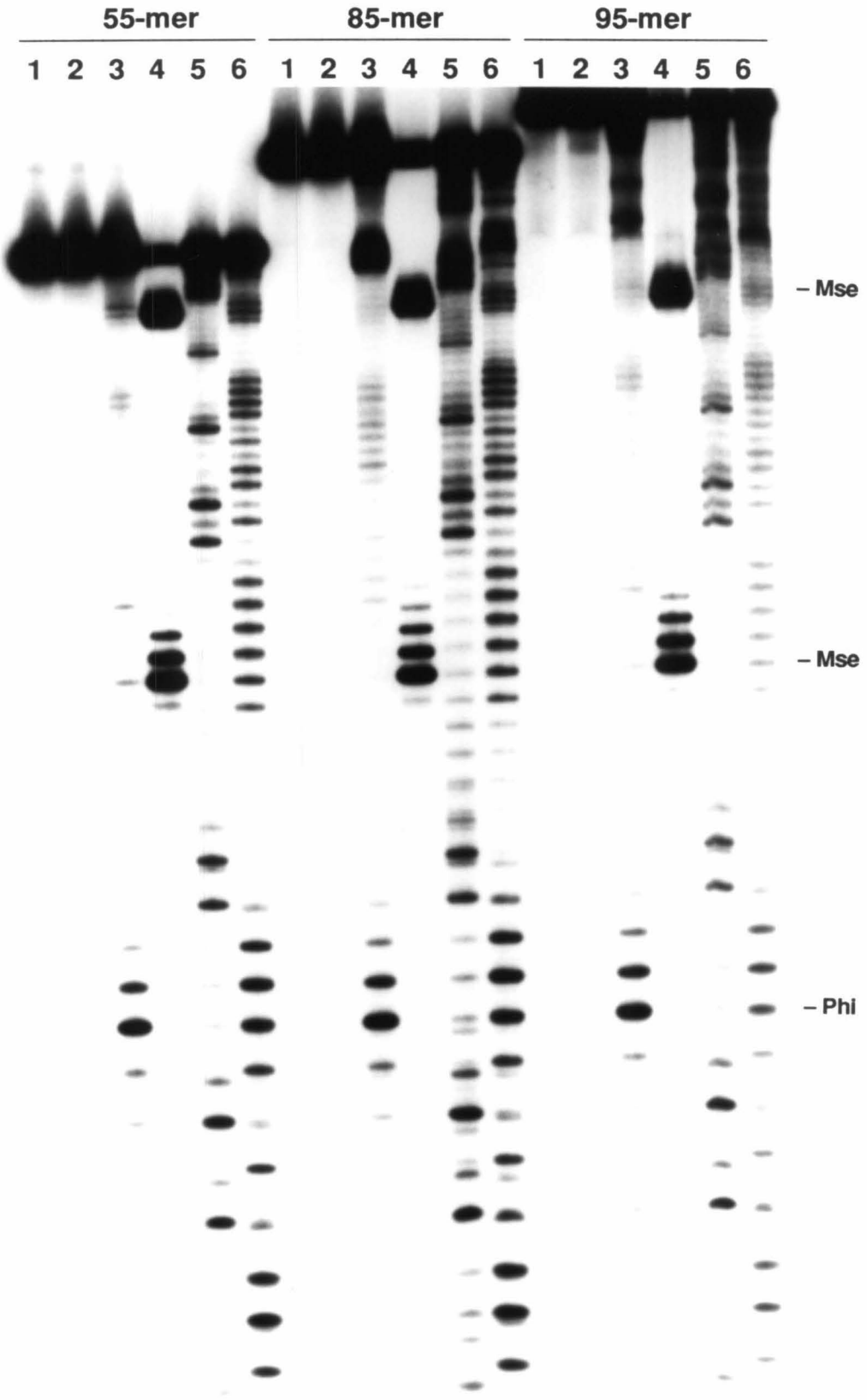

55-mer
85-mer
95-mer
1 2 3 4 5 6 7 8
1 2 3 4 5 6 7 8
1 2 3 4 5 6 7 8

Figure 4.13 Structural probing of the 55-mer, the 85-mer, and the 95-mer with various probes. 3'-end labeled fragments. Lanes 1, control fragments; lanes 2, fragments irradiated at 332 nm without metal complex; lanes 3, fragments photolyzed with $Rh(phen)_2phi^{3+}$ at 365 nm; lanes 4, fragments digested with Mse I; lanes 5, Maxam-Gilbert G reaction; lanes 6, Maxam-Gilbert T + C reaction. The three fragments contain identical structural elements at the 3' end: the positions of Mse I and $Rh(phen)_2phi^{3+}$ cleavage are identical.


55-mer
85-mer
95-mer
1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6
– Mse
– Mse
– Phi

The 95-mer (Figure 4.12 & 4.13) was also tested for cleavage by $Rh(phen)_2phi^{3+}$, and interestingly it was targeted at the same sites as the 85-mer in the exon sequences just beyond the 5' and 3' ends of the intron. However, in the middle of the fragment where the two parts are joined, it showed a different intensity and position of cleavage than the 85-mer. There are two sites in the middle of the 95-mer 5' to the 85-mer cleavage site, and the intensity of the cleavage is much weaker than the 85-mer. This is consistent with the lack of $Rh(DIP)_3^{3+}$ cleavage site, and indicates a different folding for the 95-mer in the crucial middle section where the two end parts seem to converge.

For the SV40 intron the 136-mer was tested for cleavage by $Rh(phen)_2phi^{3+}$ (Figure 4.9). As was the case for the 174-mer and the 85-mer, the 136-mer was specifically cleaved in the exon sequences just beyond the two ends of the intron. Additionally there was a cleavage site five bases 3' to the major branch point. This is near the $Rh(DIP)_3^{3+}$ cleavage site in the supercoiled DNA, and possibly $Rh(phen)_2phi^{3+}$ may be targeting the same structure from a different angle.

The $Rh(DIP)_3^{3+}$ and $Rh(phen)_2phi^{3+}$ cleavage of the ssDNA's seems to suggest so far that there are at least two double-stranded components to the DNA structures of the SV40 and Ad2 introns. The two helices appear to be formed at the two ends of the introns and involve the exon sequences as well as the intron sequences. In fact, the exon sequences are crucial to the structure as was demonstrated by the non-cleavage of the ssDNA consisting only of the intron sequences. However, since $Rh(phen)_2phi^{3+}$ can also potentially target tertiary DNA structures, we cannot conclude with certainty that there are indeed two independent helices in the structure of the ssDNA fragments. Experiments in the following sections address this problem as well as other aspects of the structure.

4.3.3 Salt concentration dependence of the cleavage of the Ad2 ssDNA fragments by $Rh(DIP)_3^{3+}$ and $Rh(phen)_2phi^{3+}$

Low-resolution mapping showed that the cleavage of the intron DNA by $Rh(DIP)_3^{3+}$ was sensitive to changes in salt concentration. If the structure formed by the coding strand ssDNA is the same structure observed in the supercoiled DNA, then we should expect the same sensitivity to the salt concentration for the ssDNA. Cleavage of the Ad2 intron 85-mer was carried out under a variety of salt concentrations, from 5 mM to 200 mM (5, 10, 25, 50, 100, & 200 mM). The $Rh(DIP)_3^{3+}$ cleavage pattern (Figure 4.14) was essentially the same as in the low-resolution mapping. The intensity was quite low at 5 mM and increased to a maximum at 25 mM and then gradually decreased until almost nonexistent at 200 mM. To compare this to the solubility of the metal complex, UV-Vis spectra of the metal complex was taken under the same salt concentrations (Figure 4.15). The solubility of $Rh(DIP)_3^{3+}$ decreased slightly with decreasing salt concentrations and did not follow the cleavage intensity profile for the 85-mer. For $Rh(phen)_2phi^{3+}$, however, the salt concentration profile was different (Figure 4.16). The intensity was low at 5 and 10 mM, rose sharply at 25 mM and increased still more up to 100 mM and then fell sharply at 200 mM. For the 95-mer the effect of the salt concentration variation was much less pronounced than for the 85-mer, again suggesting that its structure in the central region is different.

The two different profiles for the two rhodium complexes, and the fact that these are not determined by the solubility of the metal complexes, seem to suggest strongly that the variation in cleavage intensities is a result of a structural change rather than a property of the cleavage efficiency at different salt concentrations. Higher salt concentrations are likely to promote a closer packing of the structure overall and particularly in the middle section of the ssDNA's where two structural

Figure 4.14 Salt concentration dependence of the $Rh(DIP)_3^{3+}$ cleavage of the 85-mer. Lane 1, fragment incubated with $Rh(DIP)_3^{3+}$ without photolysis; lane 2, fragment irradiated at 332 nm without $Rh(DIP)_3^{3+}$; lanes 3 to 8, fragments photolyzed with $Rh(DIP)_3^{3+}$ at 332 nm in 5, 10, 25, 50, 100, and 200 mM NaCl respectively; lane 9, Maxam-Gilbert G reaction; lane 10, Maxam-Gilbert T + C reaction.


1 2 3 4 5 6 7 8 9 10

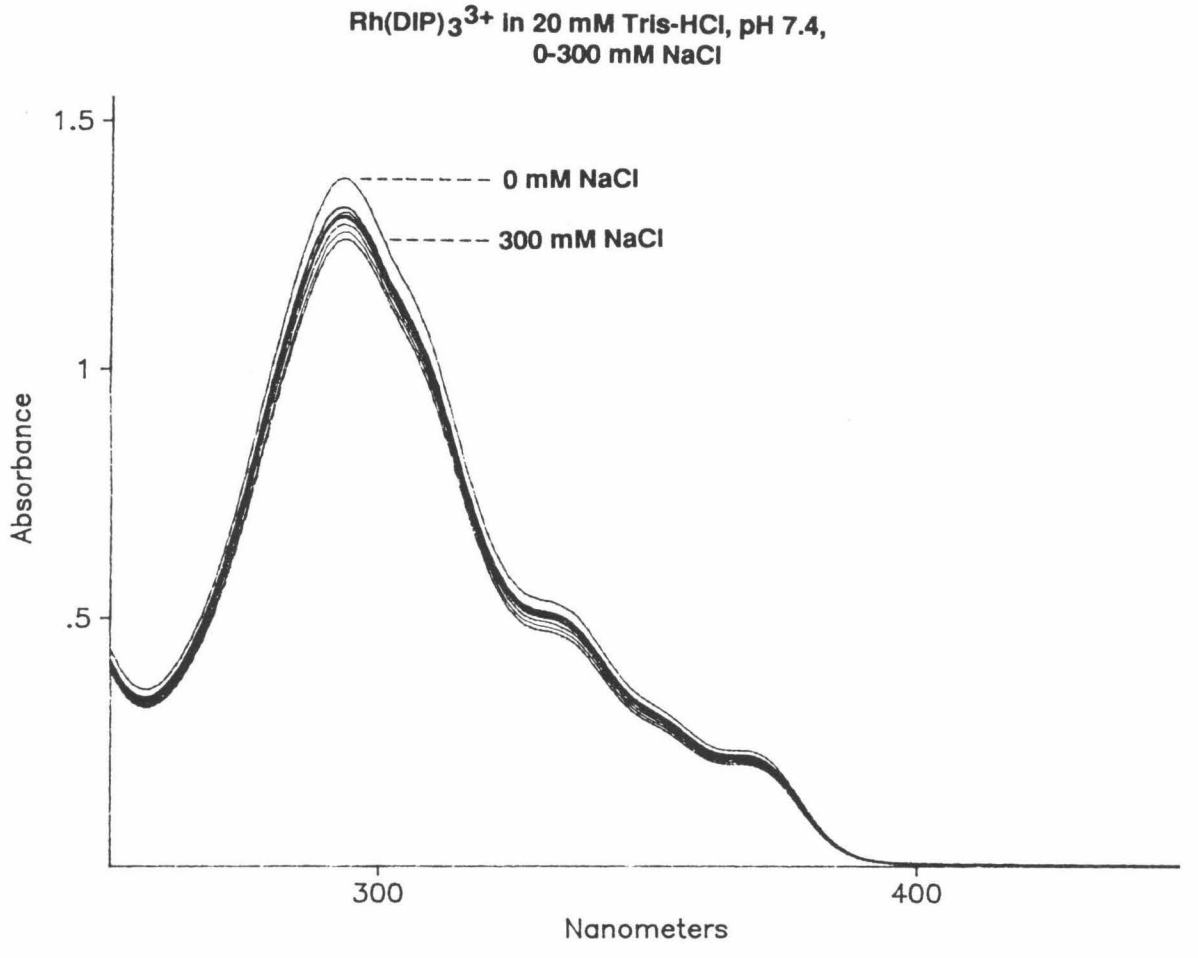
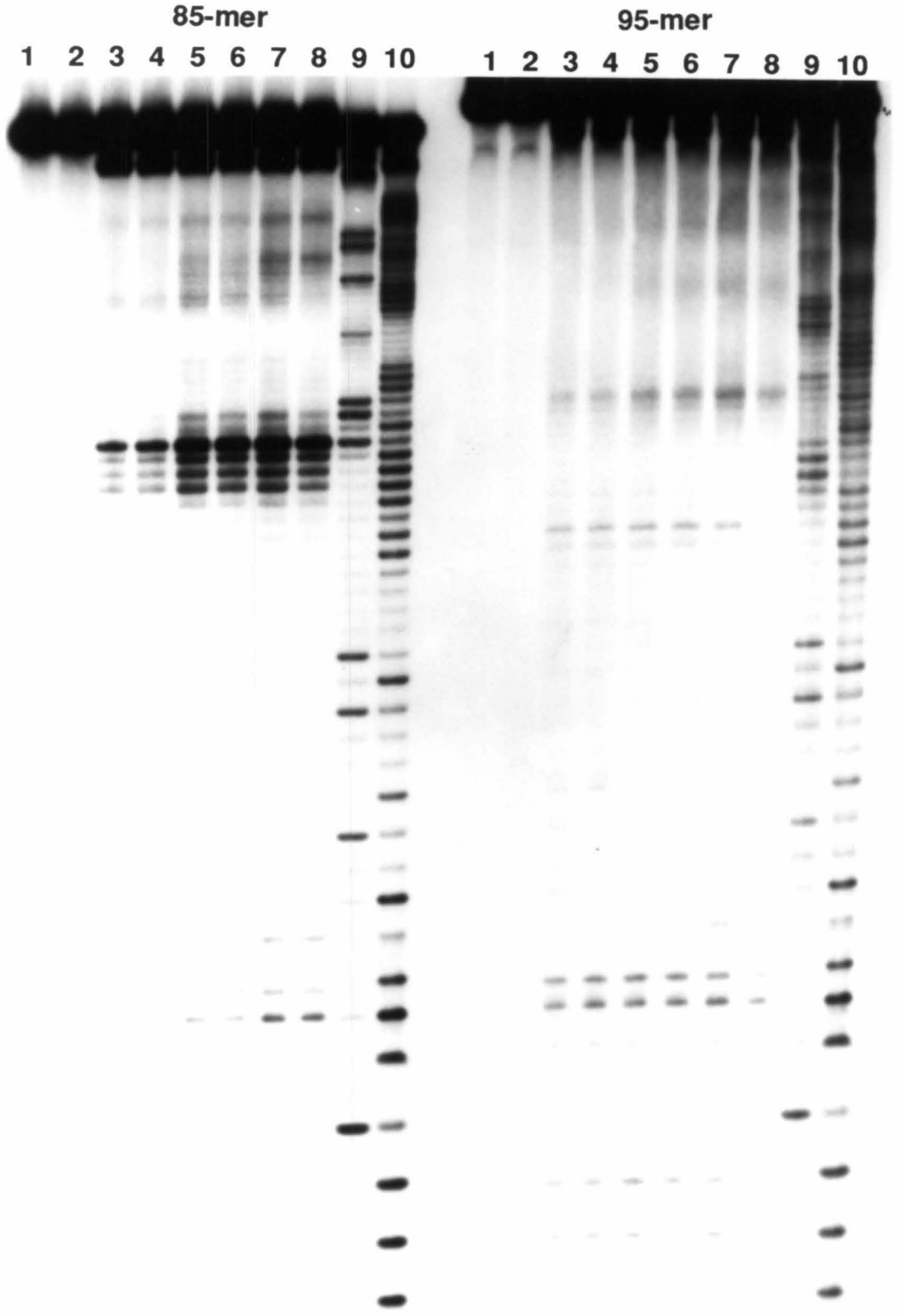


Figure 4.15 UV-Vis spectra of $Rh(DIP)_3^{3+}$ under NaCl concentrations from 0 to 300 mM. There is a slight drop in solubility of the complex under increasing salt concentrations.

Figure 4.16 Salt concentration dependence of the $Rh(phen)_2phi^{3+}$ cleavage of the 85-mer and the 95-mer. Lanes 1, fragment incubated with $Rh(phen)_2phi^{3+}$ without photolysis; lanes 2, fragment irradiated at 365 nm without $Rh(phen)_2phi^{3}$; lanes 3 to 8, fragments photolyzed with $Rh(phen)_2phi^{3+}$ at 365 nm in 5, 10, 25, 50, 100, and 200 mM NaCl respectively; lane 9, Maxam-Gilbert G reaction; lane 10, Maxam-Gilbert T + C reaction.


85-mer
1 2 3 4 5 6 7 8 9 10
95-mer
1 2 3 4 5 6 7 8 9 10

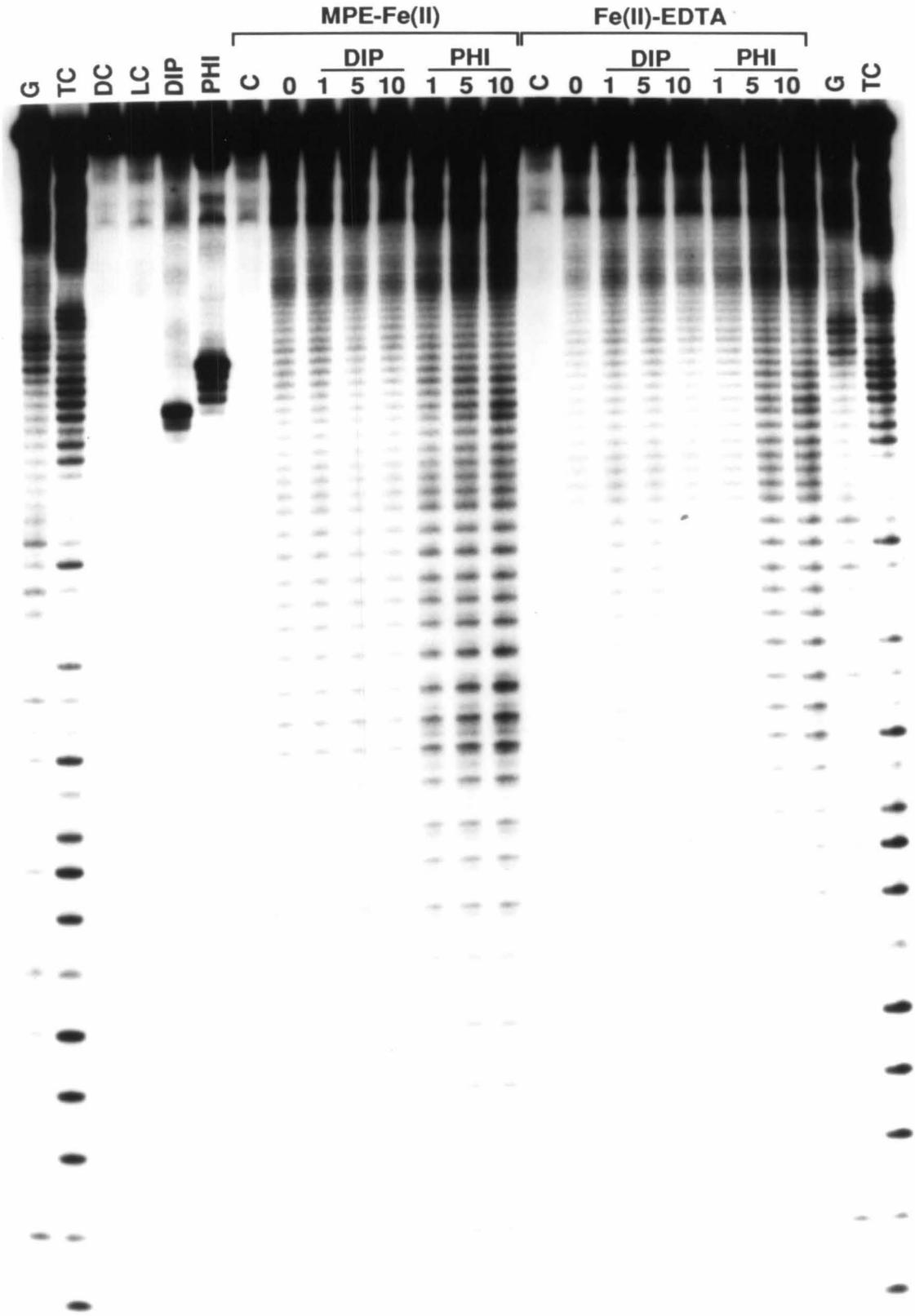
components seem to be converging. Thus a plausible explanation of the observed cleavage under the varying salt concentrations is that the binding site for $Rh(DIP)_3^{3+}$ is most likely getting smaller as the salt concentration increases, thus making the interaction more difficult, and the binding site for $Rh(phen)_2phi^{3+}$ is becoming more defined until it also gets too small at very high salt concentrations.

4.3.4 EDTA-Fe(II) and MPE-Fe(II) cleavage of the Ad2 ssDNA fragments

EDTA-Fe(II) and MPE-Fe(II) take advantage of the Fenton chemistry of the ferrous ion which generates hydroxyl radicals in aqueous solution. The hydroxyl radicals attack the sugar-phosphate backbone of DNA or RNA and cause strand scission at the point of attack (7). EDTA-Fe(II) was used by Tullius and coworkers to probe DNA structures such as bent DNA (8). MPE-Fe(II) is a derivative of EDTA-Fe(II), developed by Dervan and coworkers, which has an intercalator methidium attached to the EDTA moiety (9). The intercalator increases the efficiency of cleavage considerably since it brings the iron center close to the DNA backbone. Neither reagent, however, has affinity for specific sequences or shapes of DNA. The cleavage depends on the efficiency of the delivery of the hydroxyl radicals to the DNA backbone. Therefore, a folded structure of DNA or RNA with parts of it buried or shielded by other parts will be cleaved to different extents at different positions. A model of the tertiary folding can thus be made from the cleavage results given other constraints such as duplexes and loops.

EDTA-Fe cleavage was carried out primarily on the 85-mer which showed a very specific cleavage by $Rh(DIP)_3^{3+}$ at one site. The results (Figure 4.16), however, were not very striking. There was cleavage at all positions, and the variations in the intensity of cleavage was not great. The interesting regions that were cleaved more strongly than others were the branch site and three nucleotides following it to the 3'

Figure 4.17 Hydroxyl radical cleavage of the 85-mer. EDTA-Fe(II) and MPE-Fe(II) were used to probe the structure of the 85-mer. $Rh(DIP)_3^{3+}$ and $Rh(phen)_2phi^{3+}$ were also incubated with the DNA before hydroxyl radical cleavage in an attempt to footprint the complexes. A footprint was observed for $Rh(DIP)_3^{3+}$ but not for $Rh(phen)_2phi^{3+}$, which induced a hypersensitivity to hydroxyl radical cleavage at several sites. These results confirm a tertiary structure of the ssDNA fragment. G, Maxam-Gilber G reaction; TC, Maxam-Gilbert T + C reaction; DC, control fragment; LC, fragment irradiated at 332 nm without metal complex; DIP, fragment photolyzed with $Rh(DIP)_3^{3+}$; PHI, fragment photolyzed with $Rh(phen)_2phi^{3+}$; Lanes under MPE-Fe(II) or Fe(II)-EDTA, hydroxyl radical cleavage with the reagents: C, control; 0, cleavage with no metal complex; 1, 5, & 10, cleavage with 1, 5, and 10 μM metal complex in the reaction mixture.


MPE-Fe(II)
Fe(II)-EDTA
DIP
PHI
G
TC
DC
LC
DIP
PHI
C
0
1
5
10
1
5
10
C
0
1
5
10
1
5
10
G
TC

side. Several nucleotides on either side of this strongly cleaved stretch show a lower intensity of cleavage. The cleavage sites for $Rh(DIP)_3^{3+}$ and $Rh(phen)_2phi^{3+}$ are just 5' to this region of weak-strong-weak hydroxyl radical cleavage. Though this does indicate a structural polymorphism, it is not clear yet how the data could be interpreted. MPE-Fe(II) was also used to probe the structure, and the results obtained were identical to the EDTA-Fe(II) results.

Hydroxyl radical cleavage was also carried out in the presence of the two rhodium complexes in an attempt to footprint the complexes and to possibly gain more insight into the structure of the 85-mer. The concentration of the metal complexes were varied from 1 to 5 to 10 μM. A slight footprint was visible for $Rh(DIP)_3^{3+}$ but not for $Rh(phen)_2phi^{3+}$. However, $Rh(phen)_2phi^{3+}$ caused a hypersensitivity to hydroxyl radical cleavage just 5' to its cleavage site, and the hypersensitivity increased with increasing metal concentration. This shows nicely that the metal complexes actually bind to the site of their cleavage. The binding of $Rh(phen)_2phi^{3+}$ seems to cause a structural change in the 85-mer which makes parts of it more accessible to hydroxyl radicals. It is not clear at this point whether the change is toward a more compact folding or a shift of the orientation of the components, either could bring about the hypersensitivity observed next to the metal complex binding site. Two more regions, one about 16 residues 5' and the other about 7 residues 3' to the $Rh(phen)_2phi^{3+}$ cleavage site, also became hypersensitive to hydroxyl radical cleavage in the presence of the metal complex. This suggests that those regions may be interacting closely, perhaps directly, with the central portion where the binding sites of the rhodium complexes are located. This lends further support to all the previous observations which point to a distinctly folded structure of the ssDNA fragments in which the two halves of the molecules interact to form the overall structure.

4.3.5 Mung bean nuclease mapping of the ssDNA fragments

The single-strand specific mung bean nuclease was also used to test for loops or bulges that may be present in the intron structure. Cleavage of the 85-mer and the 95-mer yielded results that were consistent with the $Rh(phen)_2phi^{3+}$ cleavage results (Figure 4.17). We would not expect the nuclease to cleave at sites where the phi complex cleaved because the phi complex would presumably target double-stranded sites. Except for the site near the branch point where both $Rh(DIP)_3^{3+}$ and $Rh(phen)_2phi^{3+}$ cleave, there was no overlap of the mung bean nuclease cleavage and $Rh(phen)_2phi^{3+}$ cleavage. Cleavage of the 136-mer by mung bean nuclease was also mostly consistent with the $Rh(phen)_2phi^{3+}$ cleavage results. However, in all the ssDNA's tested, mung bean nuclease cleaved very close to the phi sites in some cases, which makes the interpretation of the results difficult. It is to be noted, however, that mung bean nuclease cleavage is not very consistent from experiment to experiment, and the dynamics of the structure and its equilibrium with the unfolded form of the ssDNA are factors that should be considered in interpreting the nuclease cleavage results. Thus, unfortunately, mung bean nuclease did not point us in any clear direction regarding the overall folding of the structure. That there are at least two parts of the structure seems clear so far, and these components seem to involve double-stranded units as indicated by $Rh(phen)_2phi^{3+}$ cleavage. It appears also that there are bulges or loops within the two parts.

4.3.6 Prediction of the secondary structure by computational folding

So far there have not been any definite clues to determine the base-pairing that may be involved in the two parts of the ssDNA structure. The use of a restriction enzyme to test for double-strandedness would give us definite proof of a base-pairing interaction. Before this can be done, however, we must be able to

Figure 4.18 Mung bean nuclease digestion of the 85-mer and the 95-mer. Lanes 1, control fragment; lanes 2, mung bean nuclease digestion for 5 minutes at room temperature; lanes 3, mung bean nuclease digestion for 10 minutes at room temperature; lanes 4, Maxam-Gilbert G reaction; lanes 5, Maxam-Gilbert T + C reaction. Cleavage patterns of the two fragments are identical at the 5' end of the intron.

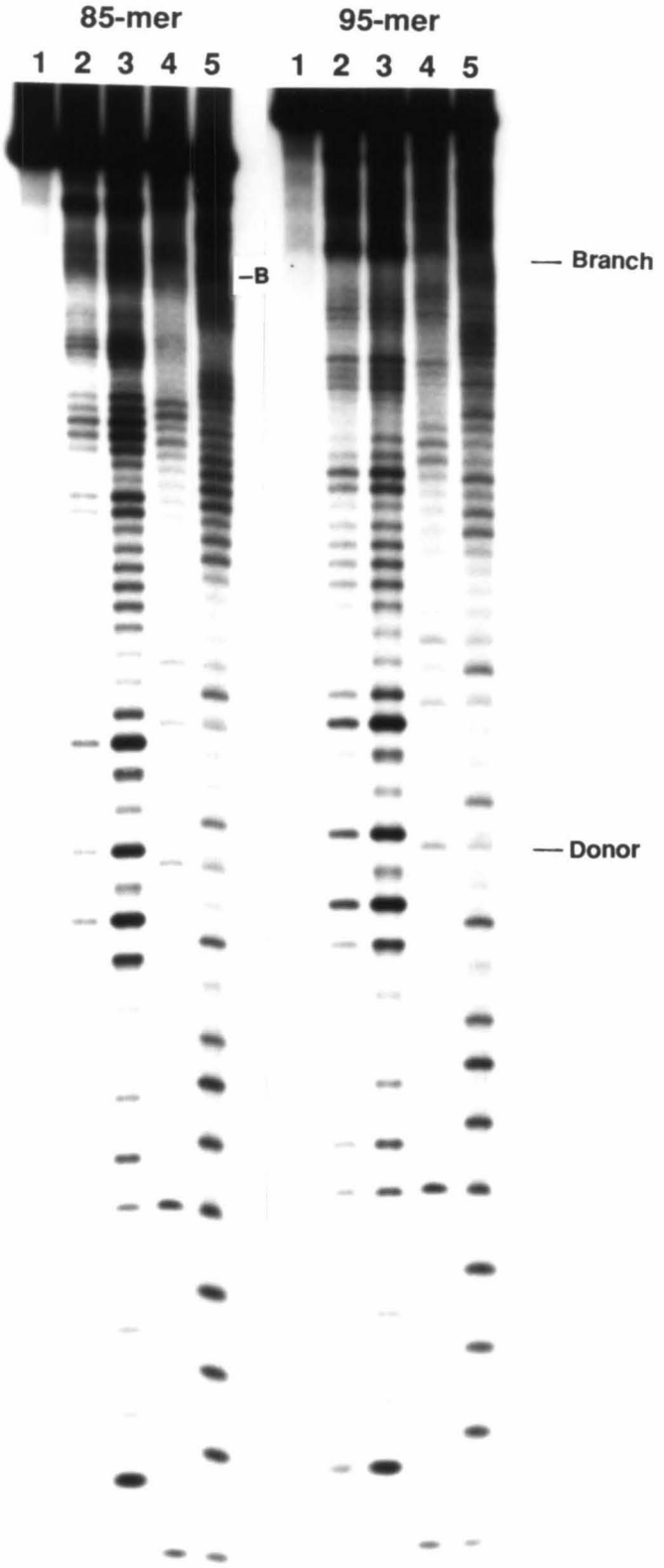
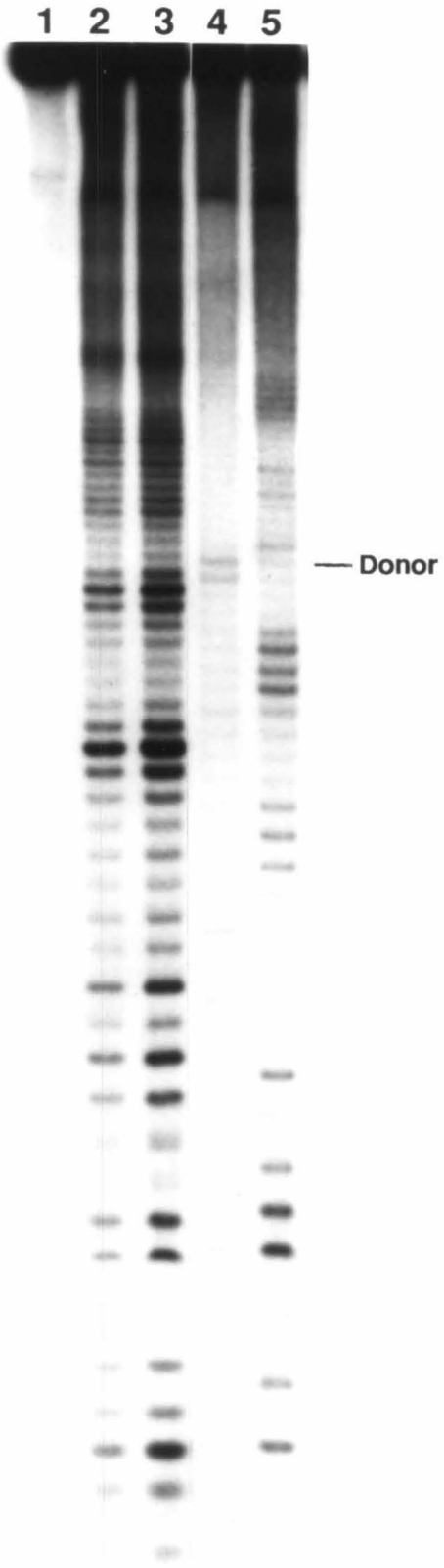

85-mer
95-mer
1 2 3 4 5
1 2 3 4 5
–B
— Branch
— Donor

Figure 4.19 Mung bean nuclease digestion of the 136-mer. Lane 1, control fragment; lane 2, mung bean nuclease digestion for 5 minutes at room temperature; lane 3, mung bean nuclease digestion for 10 minutes at room temperature; lane 4, Maxam-Gilbert G reaction; lane 5, Maxam-Gilbert T + C reaction.

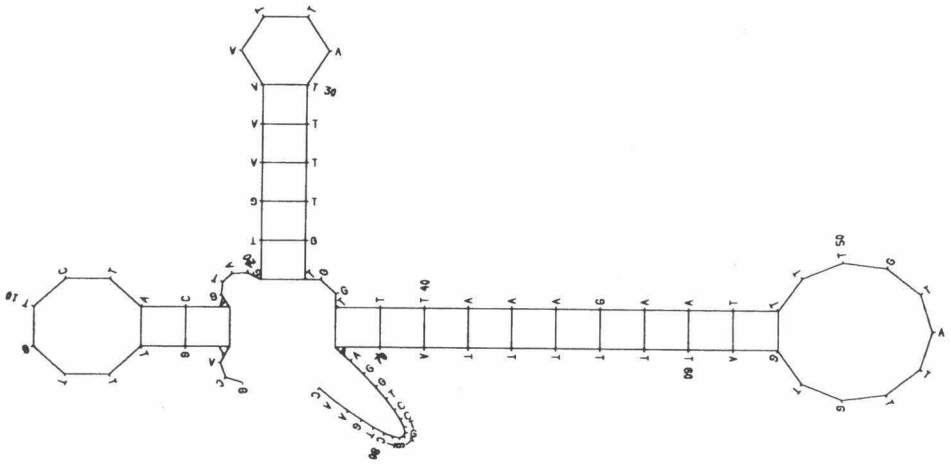

1 2 3 4 5
Donor

predict the base-pairing interaction. To this end the RNA secondary folding program of Zuker (10) was used to predict the secondary structure of the 174-mer, 85-mer, and also the 95-mer for the Ad2 intron and the 136-mer for the SV40 intron (Figures 4.20--4.23). Though these sequences are DNA, rather than RNA, their base-pairing interaction might be reasonably predicted by the RNA folding program. It is to be noted in support of this reasoning that the secondary structure of a tDNA was shown to be identical to its tRNA counterpart (2).

The computational folding of the Ad2 ssDNA fragments resulted in interesting possibilities for the secondary structures. For the 85-mer and the 95-mer the variation between different possibilities was less than that for the 174-mer, which could be folded into many different structures. However, one common feature among all the possibilities was a double helix at the 3' end of the intron with a stem of 11 base pairs and a loop of 9 bases. There was another stem-loop structure at the 5' end which was fairly predictable though it did not consistently appear in all structures. For the SV40 ssDNA there appeared a constant stem-loop structure at the 5' end which included a small bubble in the middle of the stem. A smaller stem-loop structure was also predicted at the 3' end with lower consistency. The middle portion had, however, a wide variation in possible base-pairing interactions and thus its secondary structure could not be predicted accurately.

The computational prediction of secondary structures cannot be taken too seriously without experimental verification of the predicted structures. Accurate secondary structural models can be arrived at through comparison of a large number of sequences belonging to various phylogenetic groups. Such studies have produced secondary structural models of ribosomal RNA's (11), which have since been largely confirmed by experimental data (12). Mammalian intron sequences do not show high conservation and thus such an analysis is not possible. The predicted

Figure 4.20 Computational prediction of the secondary structure of the 85-mer. Two representative predictions of the secondary structure are shown. Note the common helix in the 3' half of the molecule. The structures were predicted using Mfold of Zuker (10).



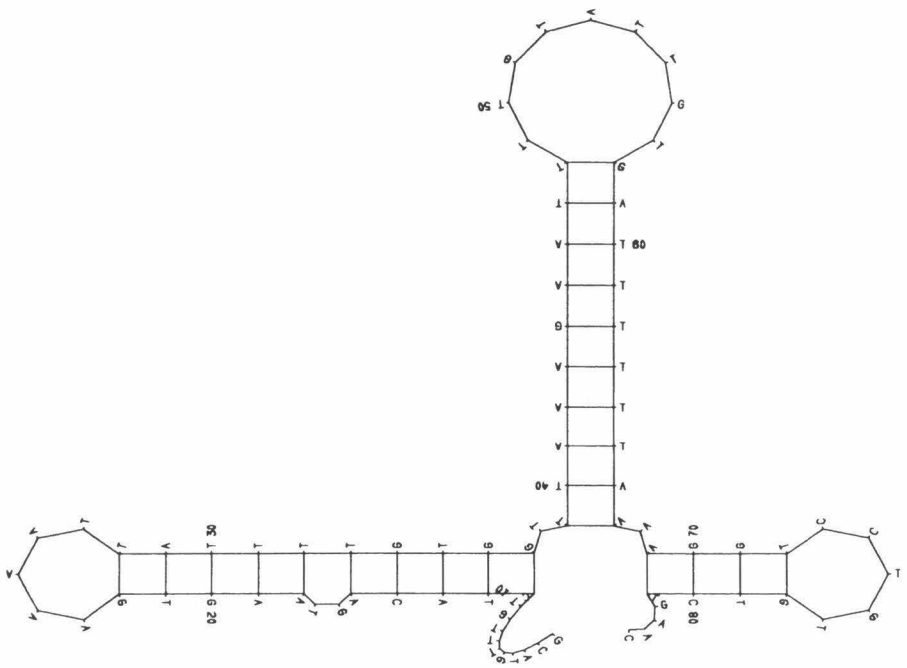
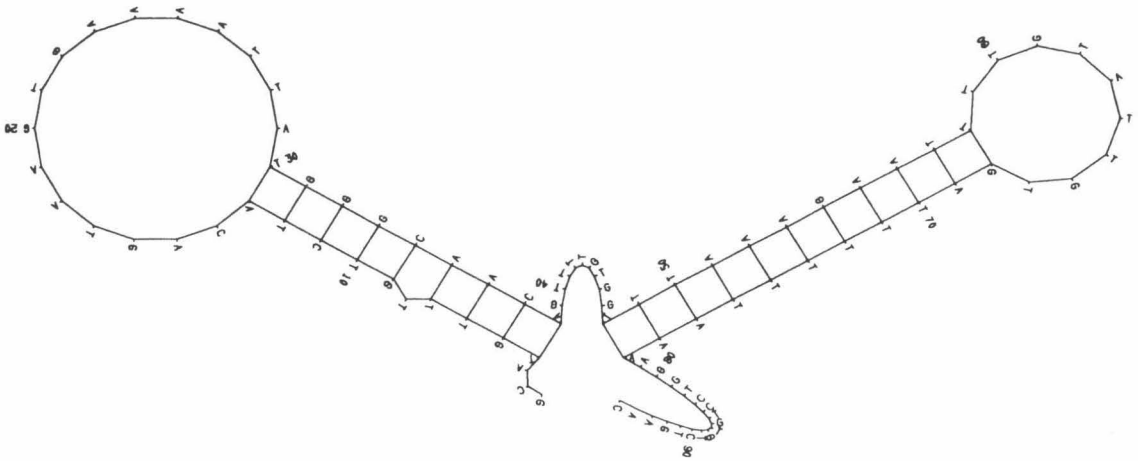


Figure 4.21 Computational prediction of the secondary structure of the 95-mer. Two representative predictions of the secondary structure are shown. Note the common helix in the 3' half of the molecule. The structures were predicted using Mfold of Zuker (10).



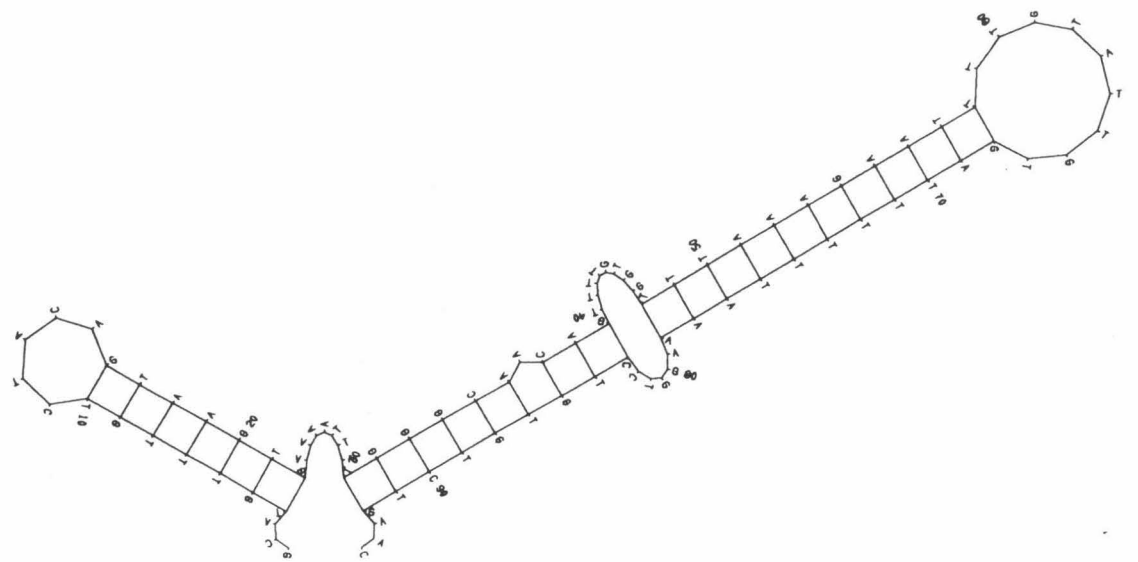


Figure 4.22 Computational prediction of the secondary structure of the 174-mer. One representative prediction of the secondary structure is shown. The helix found in the 3' half of the 85-mer and the 95-mer is still present though in a slightly different arrangement of base-pairs. The structure was predicted using Mfold of Zuker (10).

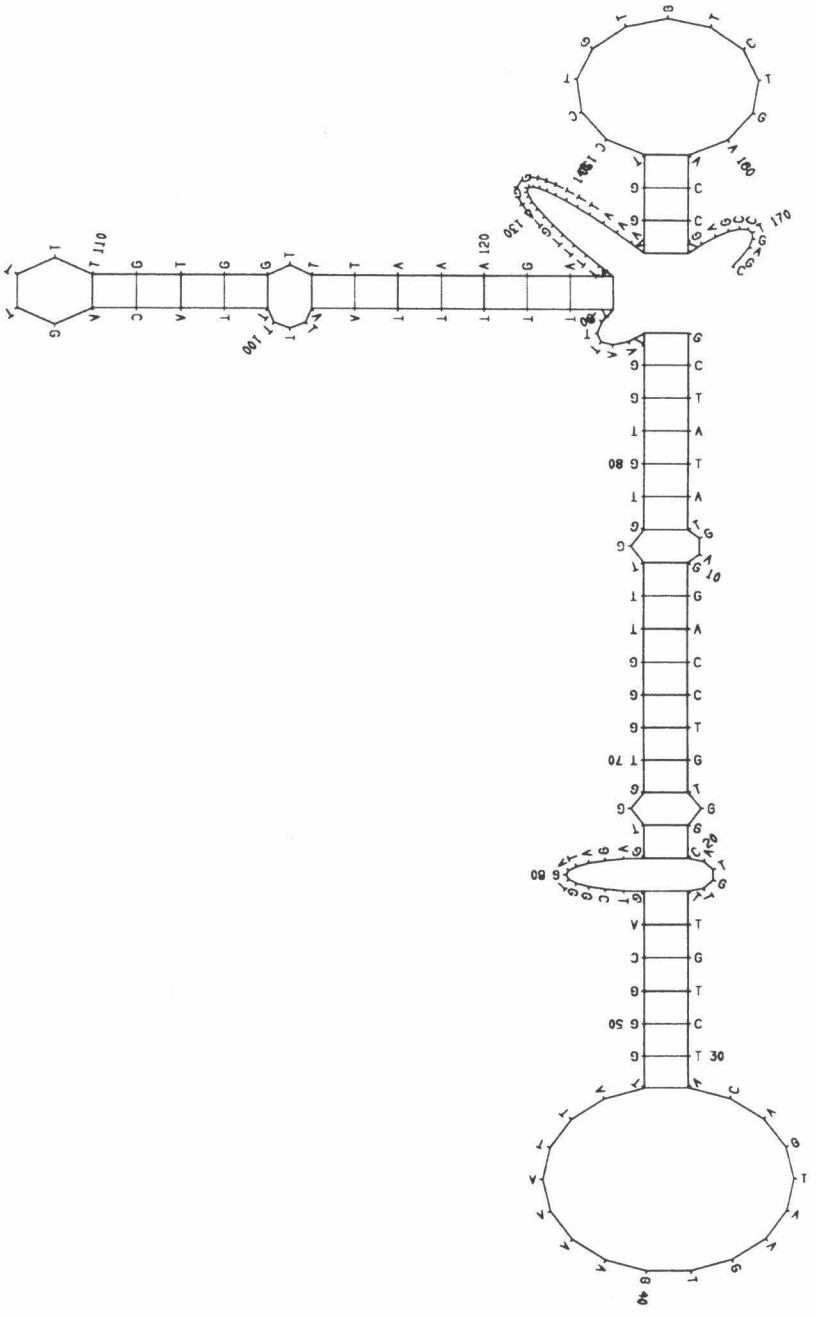
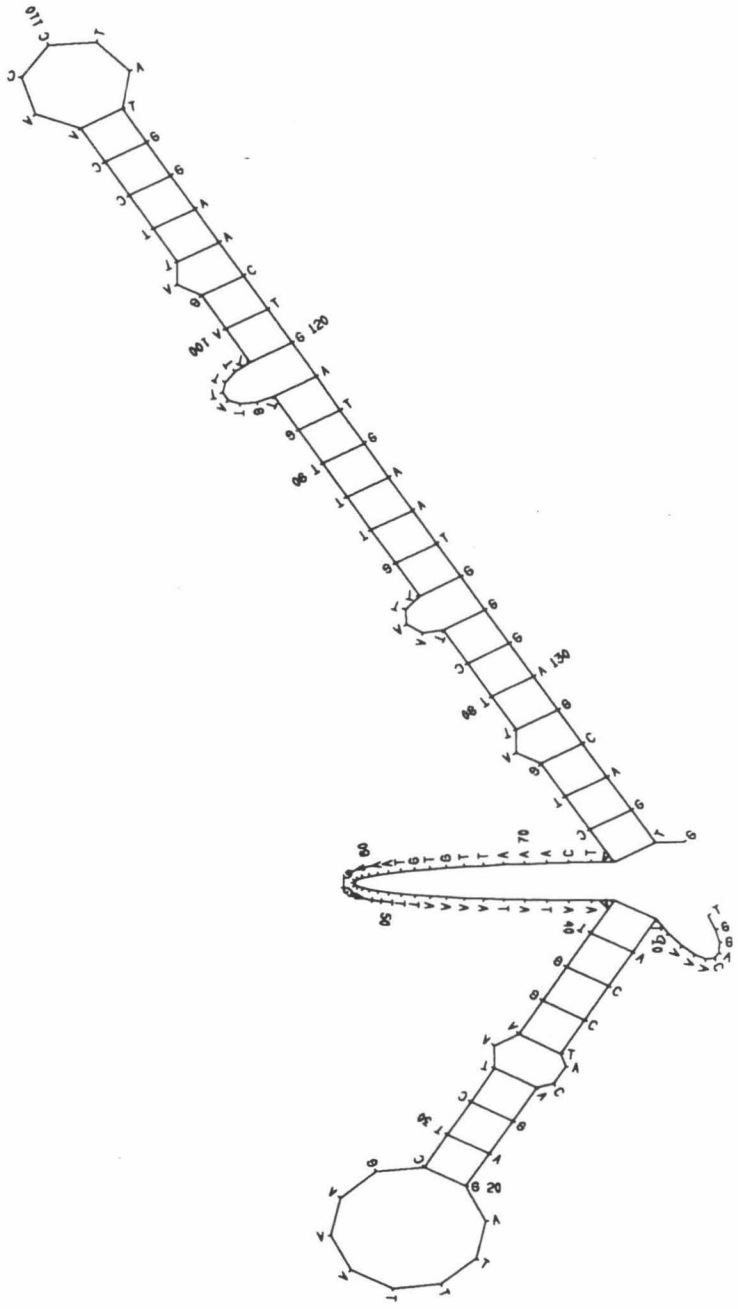


Figure 4.23 Computational prediction of the secondary structure of the 136-mer. One representative prediction of the secondary structure is shown. As found in the Ad2 E1A ssDNA structures there are two major helices in the two halves of the molecule. The structure was predicted using Mfold of Zuker (10).



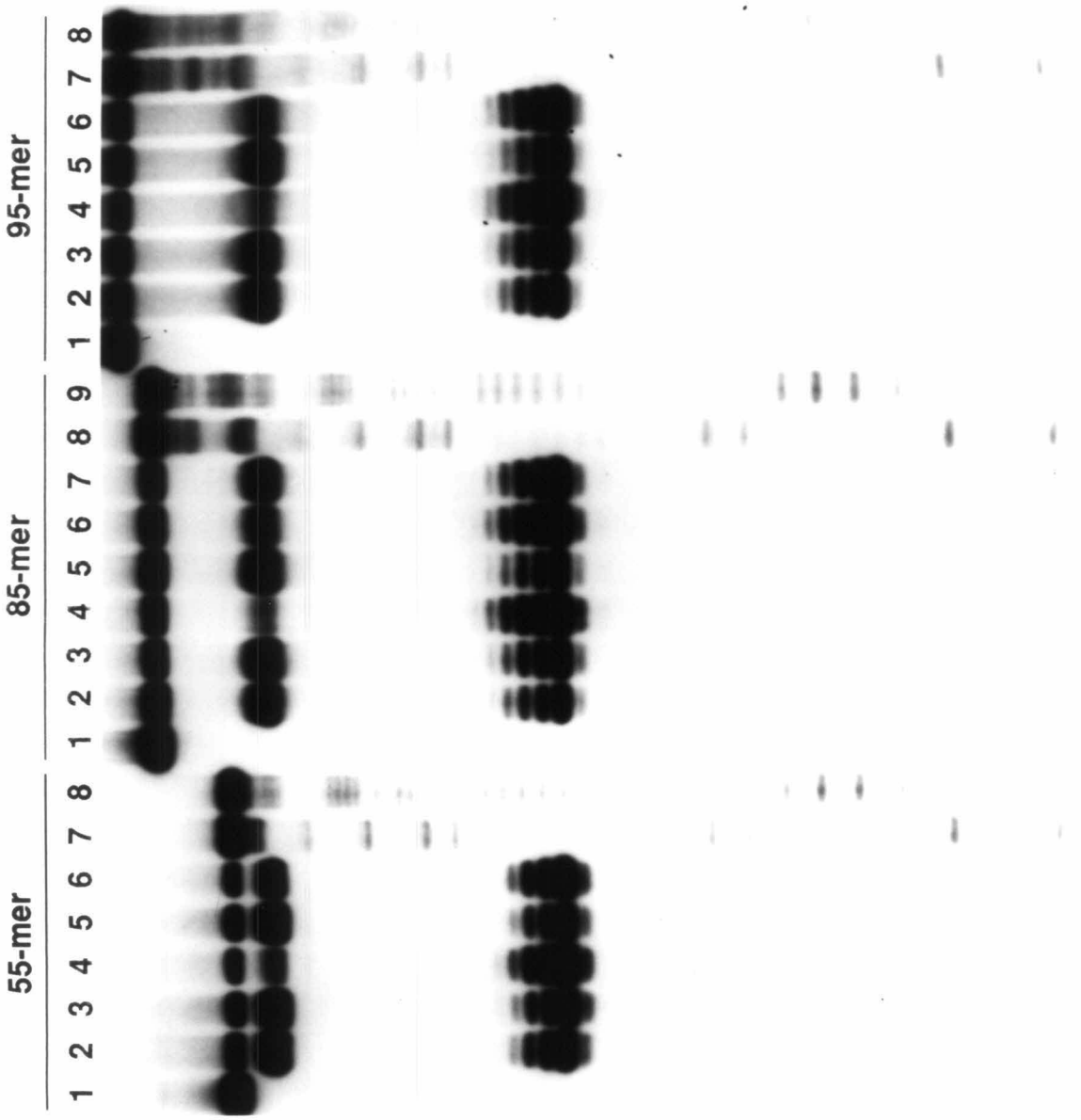
structural possibilities were analyzed for common features as mentioned above, and the models were tested experimentally as described in the following sections.

4.3.7 Restriction enzyme (Mse I) digestion of the Ad2 ssDNA fragments

The predicted double-stranded regions at the 3' end of the Ad2 ssDNA's was found to contain a site for Mse I which has a four base-pair recognition sequence of 5'-TTAA-3'. This sequence is found at the base of the stem at the 3' end of the ssDNA's. The less constant stem at the 5' end did not contain any suitable restriction enzyme site. And the predicted structures for the SV40 ssDNA also had no restriction enzyme site. The Ad2 ssDNA fragments were digested with Mse I to test for the presence of the predicted double helix. The 85-mer and the 95-mer were both cleaved at the expected site for the enzyme, providing proof of the existence of the double helix in those fragments (Figure 4.24). Interestingly, the 55-mer, which represents the 3' portion of the 85-mer, was also cleaved by Mse I at the expected site. This shows that the formation of this helix does not depend on the sequences in the 5' half of the fragment. However, it is clear from rhodium cleavage experiments that the two halves are necessary for the overall structure which contains the binding sites for the rhodium complexes. Thus the two structural components are first independently folded and then interact with each other to create the global structure.

The cleavage appears to be incomplete and thus two bands are visible on the gel, one corresponding to cleavage on one strand of the helix and the other on the opposite strand. There are also more than one band for each site indicating that the cleavage is "slippery." This could be caused by the dynamic nature of the structure, which may not be entirely stable and may exist in equilibrium with unfolded and partially folded states.

Figure 4.24 Mse I digestion of the 55-, 85-, and 95-mers. The fragments are labeled at the 3' end, which is common to all three fragments, with ^{32}P and then digested with the restriction enzyme Mse I for various lengths of time. The enzyme cleaves at identical positions in all three fragments showing that they posses a common structural element. Lanes 1, control fragments; lanes 2, fragment digested with Mse I in 50 mM NaCl at 37°C for 3 hours and then at 24°C for 9 hours; lanes 3, fragment digested with Mse I in 50 mM NaCl at 37°C for 3 hours; lanes 4, fragment digested with Mse I in 50 mM NaCl at 24°C for 12 hours; lanes 5, fragment digested with Mse I in 100 mM NaCl at 37°C for 3 hours; lanes 6, fragment digested with Mse I in 100 mM NaCl at 24°C for 12 hours; lane 7 for the 85-mer, fragment digested with Mse I in 100 mM NaCl at 37°C for 6 hours; lanes 7 for the 55-mer and the 95-mer, Maxam-Gilbert G reaction; lanes 8 for the 55-mer and the 95-mer, Maxam-Gilbert T + C reaction; lane 8 for the 85-mer, Maxam-Gilbert G reaction; lane 9 for the 85-mer, Maxam-Gilbert T + C reaction.

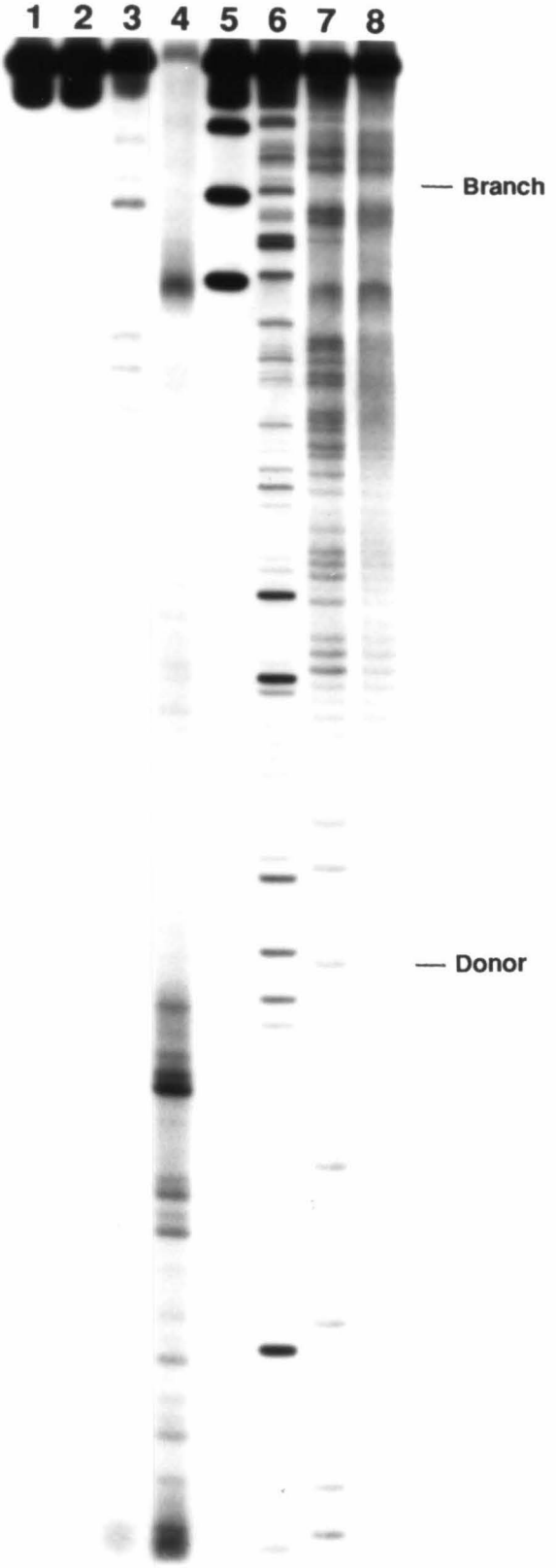

55-mer
1 2 3 4 5 6 7 8
85-mer
1 2 3 4 5 6 7 8 9
95-mer
1 2 3 4 5 6 7 8

Digestion of the 174-mer showed that the helix recognized by Mse I in the smaller fragments was present also in the 174-mer (Figure 4.25). Cleavage is more specific, and not as "slippery," in this case; however, multiple experiments showed that this property of the cleavage persisted in the 174-mer. Another site of Mse I cleavage was observed in the 174-mer which is within a helix predicted to be at the 5' half of the molecule. This site, however, does not appear as a complete four-base pair site for the enzyme in the predicted structure. It is also cleaved only on one strand and not the other, as is the site in the 3' helix. This may be due to the incompleteness of the site. It is also unclear how the enzyme recognizes the incomplete site. These questions remain unanswered, though the positive cleavage suggests a definite double helical structure.

4.3.8 Psoralen crosslinking of the ssDNA fragments

Since no suitable restriction enzyme site was found for the stem-loop structure in the 5' half of the ssDNA fragments, a different method of testing its existence was needed. Psoralen crosslinking is a method commonly used to probe nucleic acid structures (13). Psoralen and its derivatives are polycyclic intercalators which can add to the 5,6 double bond of thymines upon photoactivation. When there are two thymines on opposite strands in adjacent base pairs, the intercalated psoralen can crosslink the two thymines. Most studies of psoralen crosslinking have been on RNA structures (13, 14). In these studies the crosslink is reversed and the point of crosslink is determined by primer extension with reverse transcriptase which stops at the point of crosslink (15). For the DNA sequences studied here, however, such a method was not very suitable since the sequences were too short, and the primer needed for extension may overlap with the crosslink site. Hopkins and coworkers developed a method to determine the site of psoralen crosslinks in

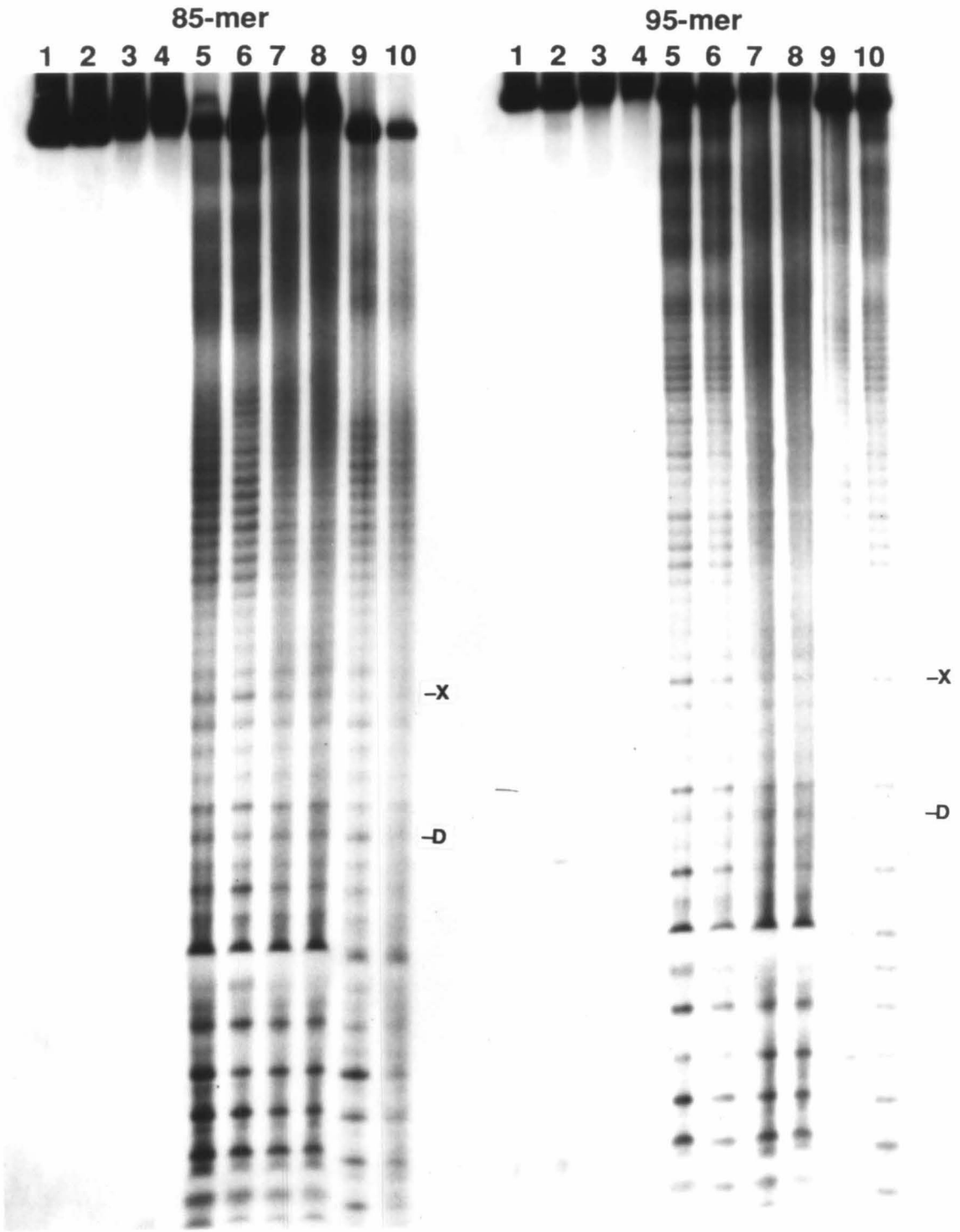
Figure 4.25 Structural probing of the 174-mer with various probes. Lane 1, control fragment; lane 2, fragment irradiated at 332 nm without metal complex; lane 3, fragment photolyzed with $Rh(DIP)_3^{3+}$; lane 4, fragment photolyzed with $Rh(phen)_2phi^{3+}$; lane 5, fragment digested with Mse I; lane 6, fragment digested with mung bean nuclease; lane 7, Maxam-Gilbert G reaction; lane 8, Maxam-Gilbert T + C reaction.


1 2 3 4 5 6 7 8
— Branch
— Donor

double-stranded DNA (16). Their method involved cleaving the DNA with EDTA-Fe(II) after psoralen crosslinking. The cleavage intensity analyzed on denaturing polyacrylamide gels decreased significantly at the site of the crosslink. This method produced positive results for double-stranded DNA, and the investigators were able to determine sequence preferences for psoralen crosslinking.

Crosslinking and mapping of the crosslinked site was carried out on the ssDNA fragments with AMT (4'-aminomethyl-4,5',8-trimethylpsoralen). The EDTA-Fe(II) cleavage method was tried in determining the site of crosslink, but it did not show any clear-cut delineation at any point in the sequence (data not shown). Thus a different approach was adopted. After crosslinking with psoralen the fragments were treated with hydrazine, a reagent for Maxam-Gilbert sequencing, which requires the 5,6 double bond of pyrimidines for its reaction with them. The samples are then treated with piperidine to effect strand scission. We should expect a lower level of hydrazine modification at the thymines to which psoralen has been crosslinked. This analysis was carried out on the 174-mer, the 85-mer, and the 95-mer for the Ad2 E1A intron and the 136-mer for the SV40 T-antigen intron. The results from the 174-mer and the 136-mer were unfortunately uninterpretable, and thus no conclusion could be drawn for the folding of those fragments. The 85-mer and the 95-mer, which so far show remarkable similarity to the 174-mer, did produce a noticeable site where hydrazine modification was lower than in control samples (Figure 4.26). It corresponded to the end of a predicted stem where there were indeed thymines on opposite strands. Thus that stem-loop structure, though less consistently predicted than the 3' stem-loop, seems to be a stable component of the structure of the 85-mer and the 95-mer. However, it is to be remembered that the 95-mer lacks any specific binding site for $Rh(DIP)_3^{3+}$, while the 85-mer has a distinct binding site for the rhodium complex. It appears that the two structural units are

Figure 4.26 Psoralen crosslinking of the 85-mer and the 95-mer. AMT (4'-amino-methyl-4,5',8-trimethylpsoralen) was used to probe for double-stranded regions in the 5' half of the molecule. Crosslink was detected by hydrazine modification of the crosslinked fragment. There was one definite position of a crosslink (marked with "X") several residues 3' to the donor site (marked with "D"). Lanes 1, control fragments; lanes 2, fragments irradiated at 340 nm for 3 minutes without AMT; lanes 3 & 4, fragments irradiated with 10 μg/ml AMT at 340 nm for 2 and 3 minutes respectively; lanes 5, control fragment treated with hydrazine; lanes 6, fragments irradiated without AMT and then treated with hydrazine; lanes 7 & 8, fragments irradiated for 2 and 3 minutes, respectively, with AMT and then treated with hydrazine; lanes 9, Maxam-Gilbert G reaction; lanes 10, Maxam-Gilbert T + C reaction.


85-mer
1 2 3 4 5 6 7 8 9 10
–X
–D
95-mer
1 2 3 4 5 6 7 8 9 10
–X
–D

present in both the 85-mer and the 95-mer, but the extra sequences in the middle of the 95-mer is hindering the formation of the global structure that is being recognized by the rhodium complex.

Though the crosslinking of the 174-mer could not be interpreted accurately, an extrapolation can be made from the crosslinking of the 85-mer and the 95-mer. Evidence so far indicates that the sequences present in the two shorter fragments fold independently and that the interaction of these structural components determine the overall shape of the structure. The fact that both the 174-mer and the 85-mer are cleaved specifically also indicates that the 85-mer folds into a structure that resembles the portion of the 174-mer that is responsible for metal complex recognition. $Rh(phen)_2phi^{3+}$ cleavage of the 174-mer and the 85-mer are also mostly coincident with one another. Mse I digest showed that the 3' stem-loop is present in both of them. Thus, it is reasonable to conclude that the 5' helix detected in the 85-mer and the 95-mer is also probably present in the 174-mer.

4.4 Summary

The results discussed in this chapter establish a structural model for the Ad2 E1A intron in some detail and for the SV40 T-antigen intron in somewhat less detail. The $Rh(DIP)_3^{3+}$ cleavage of the 174-mer and the 85-mer show that the structure required for metal complex recognition is assembled by sequences at the ends of the intron and parts of the flanking exon sequences. The specific cleavage of the 136-mer also shows the same requirement for the SV40 T-antigen intron. The fact that the full-length fragments for the two introns are targeted by $Rh(DIP)_3^{3+}$ at exactly the same site as the supercoiled version of the introns suggests that the ssDNA fragments assume a very similar, or perhaps identical, structure as the introns in the supercoiled plasmids. $Rh(phen)_2phi^{3+}$ cleavage suggests double-stranded portions

at either end of the ssDNA fragments and shows remarkable coincidence of cleavage sites in the 174-mer and the 85-mer. Mse I digest of the Ad2 E1A ssDNA fragments proved the existence of the stem-loop structure at the 3' end of the fragments as predicted by the computational base-pairing. And finally, psoralen crosslinking established the existence of another stem-loop at the 5' end of the 85-mer and the 95-mer, from which we can postulate that the 174-mer may also contain the same stem-loop. How the two structural units interact to form the global structure is still unknown. In the next chapter a structural model for the folding of the introns is proposed and discussed in terms of the data presented in this chapter.

References

1. (a) H. D. Madhani & C. Guthrie (1992) *Cell* **71**: 803-817. (b) D. S. McPheeters & J. Abelson (1992) *Cell* **71**: 819-831. (c) C. F. Lesser & C. Guthrie (1994) *Science* **262**: 1982-1988. (d) E. J. Sontheimer & J. A. Steitz (1994) *Science* **262**: 1989-1996.
2. J. Paquette, K. Nichoghosian, G. Qi, N. Beauchemin, & R. Cedergren (1990) *Eur. J. Biochem.* **189**: 259-265.
3. A. M. Pyle, E. C. Long, & J. K. Barton (1989) *J. Am. Chem. Soc.* **111**: 4520-4522.
4. A. M. Pyle, T. Morii, & J. K. Barton (1990) *J. Am. Chem. Soc.* **112**: 9432-9434.
5. A. M. Pyle & J. K. Barton (1990) *Prg. Inorg. Chem.* **38**: 413-475.
6. C. S. Chow, L. S. Behlen, O. C. Uhlenbeck, & J. K. Barton (1992) *Biochemistry* **31**: 972-982.
7. T. D. Tullius, B. A. Dombroski, M. E. A. Churchill & L. Kam (1987) *Methods Enz.* **155**: 537-559.
8. A. M. Burkhoff & T. D. Tullius (1987) *Cell* **48**: 935-943.
9. (a) R. P. Hertzberg & P. B. Dervan (1982) *J. Am. Chem. Soc.* **104**: 313-315. (b) R. P. Hertzberg & P. B. Dervan(1984) *Biochemistry* **23**: 3934-3945.
10. M. Zuker (1989) *Science* **244**: 48-52 (b) J. A. Jaeger, D. H. Turner, & M. Zuker (1989) *Proc. Natl. Acad. Sci. USA* **86**: 7706 -7710.
11. (a) C. R. Woese, L. J. Magrum, R. Gupta, R. B. Siegel, D. A. Stahl, J. Kop, N. Crawford, J. Brosius, R. Gutell, J. J. Hogan, & H. F. Noller (1980) *Nuc. Acids Res.* **8**, 2275-2293. (b) H. F. Noller, J. Kop, V. Wheaton, J. Brosius, R. Gutell, A. M. Kopylov, F. Dohme, W. Herr, D. A. Stahl, R. Gupta, & C. R. Woese (1981) *Nuc. Acids Res.* **9**, 6167-6189.
12. (a) S. Stern, B. Weiser, & H. F. Noller (1988) *J. Mol. Biol.* **204**, 447-481. (b) J. Egebjerg, N. Larsen, & R. A. Garrett (1990) *The Ribosome: Structure, Function, &*

Evolution, pp. 168-179, eds.W. A. Hill, et al., American Society for Microbiology, 1990, Washington.

13. G. D. Cimino, H. B. Gamper, S. T. Isaacs, & J. E. Hearst (1985) *Ann. Rev. Biochem* **54**: 1151-1193

14. (a) P. L. Wollenzien, J. E. Hearst, P. Thammana, & C. R. Cantor (1979) *J. Mol. Biol.* **135**: 255-269. (b) C. R. Cantor, P. L. Wollenzien, & J. E. Hearst (1980) *Nucl. Acids Res.* **8**: 1855-1872. (c) D. A. Wassarman & J. A. Steitz (1992) *Science* **257**: 1918-1925.

15. D. A. Wassarman (1993) *Mol. Biol. Reports* **17**: 143-151.

16. J. T. Millard, M. F. Weidner, J. J. Kirchner, S. Ribeiro, & P. B. Hopkins (1991) *Nucl. Acids Res.* **19**: 1885-1891.

Chapter 5.

Secondary structural models for the Ad2 E1A and SV40 T-antigen intron ssDNA fragments

5.1 Introduction

The preceding chapter described the results of various probings of the structure of the ssDNA fragments corresponding to the coding strands of the two introns and the flanking exons. The ssDNA fragments were synthesized because of the observation that the non-coding strand in the supercoiled state did not show any specific $Rh(DIP)_3^{3+}$ cleavage, and therefore the targeted structure was being formed by the coding strand alone. Specific cleavage of the ssDNA fragments containing the entire intron and a portion of the flanking exon sequences showed that this was indeed the case. A battery of chemical and enzymatic probes were used to characterize the structure of the ssDNA fragments, and computational modeling was successfully used to predict elements of the structures. The global structure of the intron DNA is composed of two major helices in the 5' and the 3' half of the introns. These two components are sufficient to create a binding site for $Rh(DIP)_3^{3+}$, as indicated by the specific cleavage of the 85-mer which lacks the middle of the intron. However, the middle portion is necessary for the complete folding of the intron to create the structure that is probably identical to the structure formed in the supercoiled plasmids. In this chapter, models for the secondary structures of the introns is presented and discussed in terms of the supporting experimental data.

5.2 Model for the structure of the Ad2 E1A intron DNA

The 174-mer and the 85-mer are considered in the analysis of the final structure since they both showed specific cleavage by $Rh(DIP)_3^{3+}$ and are thus

assumed to contain the required structural components for the metal complex binding. The starting point of the analysis is the constraint imposed on the structure by the Mse I digestion, which fixed a double-stranded region in the 3' half of the molecules. The second Mse I site in the 174-mer is also considered, though its validity is questionable due to its incomplete and one sided digestion. The psoralen crosslink site in the 5' half of the 85-mer is another constraint to be imposed on the structure. When these factors are taken into account in the computational folding (1) of the sequences, slightly different structures emerge as possibilities.

Figures 5.1 and 5.2 show the models for the 85-mer and the 174-mer respectively. The helix recognized by the restriction enzyme Mse I is prominent as the major structural component on the 3' side of the two sequences. The enzyme site is at the base of the stem from which other possible helices are also protruding. The cleavage site for $Rh(DIP)_3^{3+}$ is at the junction of the predicted helices in the 174-mer. This is consistent with the results of the studies of Holliday junctions in which synthetic junctions were cleaved with the rhodium complex (2). The cleavage of the structures occurred specifically at the junctions of the four helices in the major groove side, which complemented well the enzyme cleavages of the junctions which occurs on the minor groove side of Holliday junctions (3). The inference made from these observations is that the structure of the intron DNA may in some way resemble the structure of Holliday junctions. There are several helices protruding out from a common center, and the helices may stack to form a compact structure.

However, the cleavage site for $Rh(DIP)_3^{3+}$ in the 85-mer is away from the junction point of the helices. It is to be noted, however, that the original cleavage site for the 174-mer is still at the junction in the 85-mer; cleavage has shifted seven nucleotides to the 5' side. Such a shift in cleavage, which must result from a shift in the binding position or orientation, is understandable given the difference in the two

Figure 5.1 A structural model for the folding of the 85-mer of the Ad2 E1A intron. There are two major helices in the two halves and a minor one at the 3' end of the molecule. The two major helices are proposed to interact with each other to create the binding site for the metal complexes. Legend: Don., donor; Acc., acceptor; Br., branch; DIP, $Rh(DIP)_3^{3+}$ cleavage; Phi, $Rh(phen)_2phi^{3+}$ cleavage; Mse, Mse I cleavage; Psor., psoralen crosslink; •, mung bean nuclease cleavage.

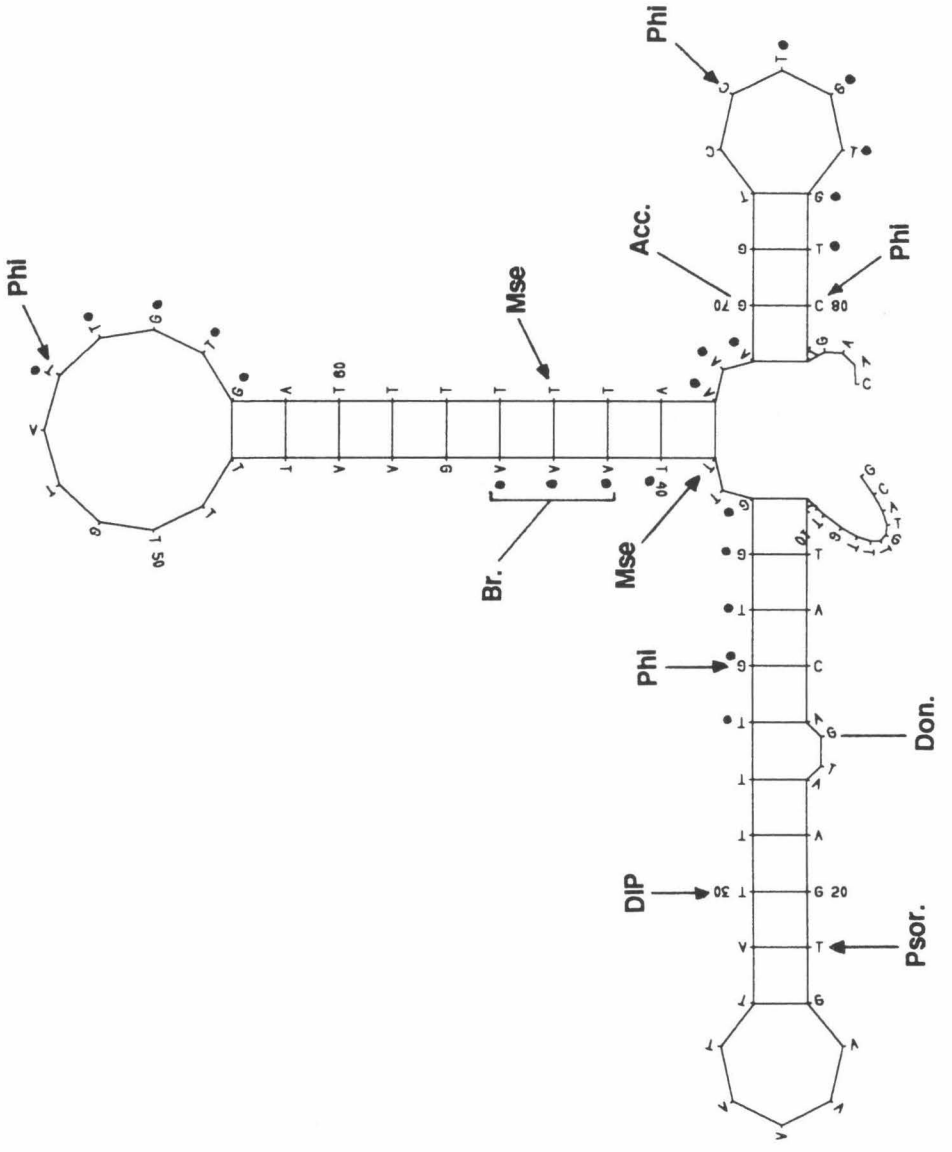
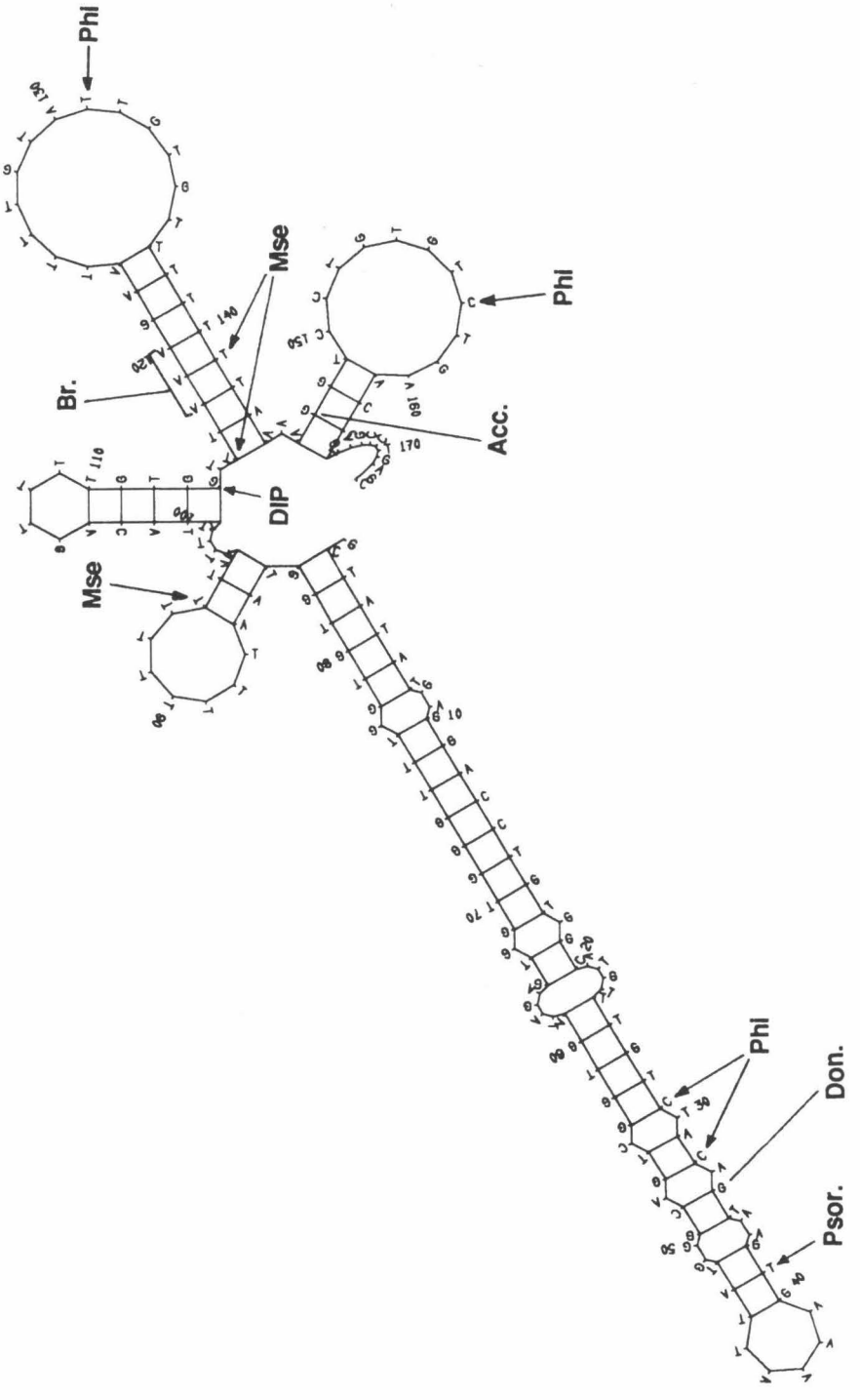

Phl
Acc.
Phl
Phl
Mse
Mse
Br.
Phl
Don.
DIP
Psor.

Figure 5.2 A structural model for the folding of the 174-mer of the Ad2 E1A intron. There are two major helices in the two halves of the molecule and three minor ones, two in the middle and one at the 3' end of the molecule. The two major helices are proposed to interact with each other, and likely also with the minor helices, to create the binding site for the metal complexes. Legend: Don., donor; Acc., acceptor; Br., branch; DIP, $Rh(DIP)_3^{3+}$ cleavage; Phi, $Rh(phen)_2phi^{3+}$ cleavage; Mse, Mse I cleavage; Psor., psoralen crosslink; •, mung bean nuclease cleavage.


Phl
Mse
Phl
Br.
Acc.
DIP
Mse
Phl
Don.
Psor.

sequences, and it is completely consistent with binding in the same general region given the size of the metal complex. This is an observation that strongly supports the idea that the two fragments share the essential components of the intron structure.

Psoralen crosslink was found at the tip of the stem that contains the $Rh(DIP)_3^{3+}$ cleavage site in the 85-mer. In the 174-mer it is also found at the tip of a less defined stem in the 5' half of the molecule. The lower (proximal to the center) half of this long helix is not verified by experimental data, and it is doubtful that it is essential to the overall folding, given the structural integrity of the 85-mer which lacks most of this part. It may be that these sequences are unstructured in solution and that the top half of the helix may associate with the helix or helices in the 3' half of the molecule, perhaps lying side by side with the helix that contains the Mse I site. This arrangement is also plausible in the 85-mer. In this regard it is useful to remember that the cleavage of a pBR322 cruciform occurred on the stems away from the center of the cruciform (4), which is an analogous situation to the cleavage in the 85-mer. If the cruciform structure resembles that of Holliday junctions, such a parallel arrangement of the helices is plausible, and this may also be the case for the intron structure, in which the two helices may be interacting in a similar way to the helices in a Holliday junction structure.

The role of the other smaller helices is not clear. Definite proof for these helices is lacking, except for the Mse I digestion of the small helix in the middle of the 174-mer. The small helix at the extreme 3' end of the sequences may be coaxial with the Mse I helix, forming a kind of platform with which the 5' helix may interact. In the 85-mer this seems a plausible model of the overall three-dimensional conformation of the molecule.

Some of the cleavage sites for $Rh(phen)_2phi^{3+}$ in both the 85-mer and the 174-mer are not easily interpretable. Presumably the complex has affinity for double-

stranded sites into which it can intercalate, but the cleavage sites are often predicted to be single-stranded loops. A possible explanation for this observation is that the single-stranded sites may be interacting with other parts of the molecule to form unusual structures, for which the complex has some affinity. $Rh(phen)_2phi^{3+}$ recognizes RNA triple base interactions and anticodon loops which are believed to be structured. Thus the recognition of loops in DNA is not out of the ordinary for the metal complex and may mean that the loops are not random but structured.

The mung bean nuclease cleavage data are the most difficult to interpret. Not only does the enzyme cleave the predicted single-stranded regions, but it also cleaves at regions that otherwise appear double-stranded, for example the Mse I cleavage site. The most plausible explanation for this observation is that enzymes such as the mung bean nuclease are more likely to give positive results than negative ones; a single-strand specific nuclease can recognize and digest parts of the DNA that may be undergoing changes or that may be fluctuating between two equilibrium states. Thus the Mse I site may be single-stranded in some states of the molecule and double-stranded in other. Both of these states will be recognized and acted upon by the appropriate enzyme and thus lead to the apparently conflicting results.

The model for the Ad2 E1A intron DNA structure is reasonably well predicted and verified by the various probes used. (For a schematic illustration of the model, see Figure 5.4.) The important features of the structure are the two main stem-loop components in the two sides of the intron which appear to be interacting to form the overall structure. Other helices may participate in making this structure which may have some resemblance to elements of the Holliday junction structure. A stretch of sequence in the middle of the intron appear to be unstructured and unnecessary for the overall structure.

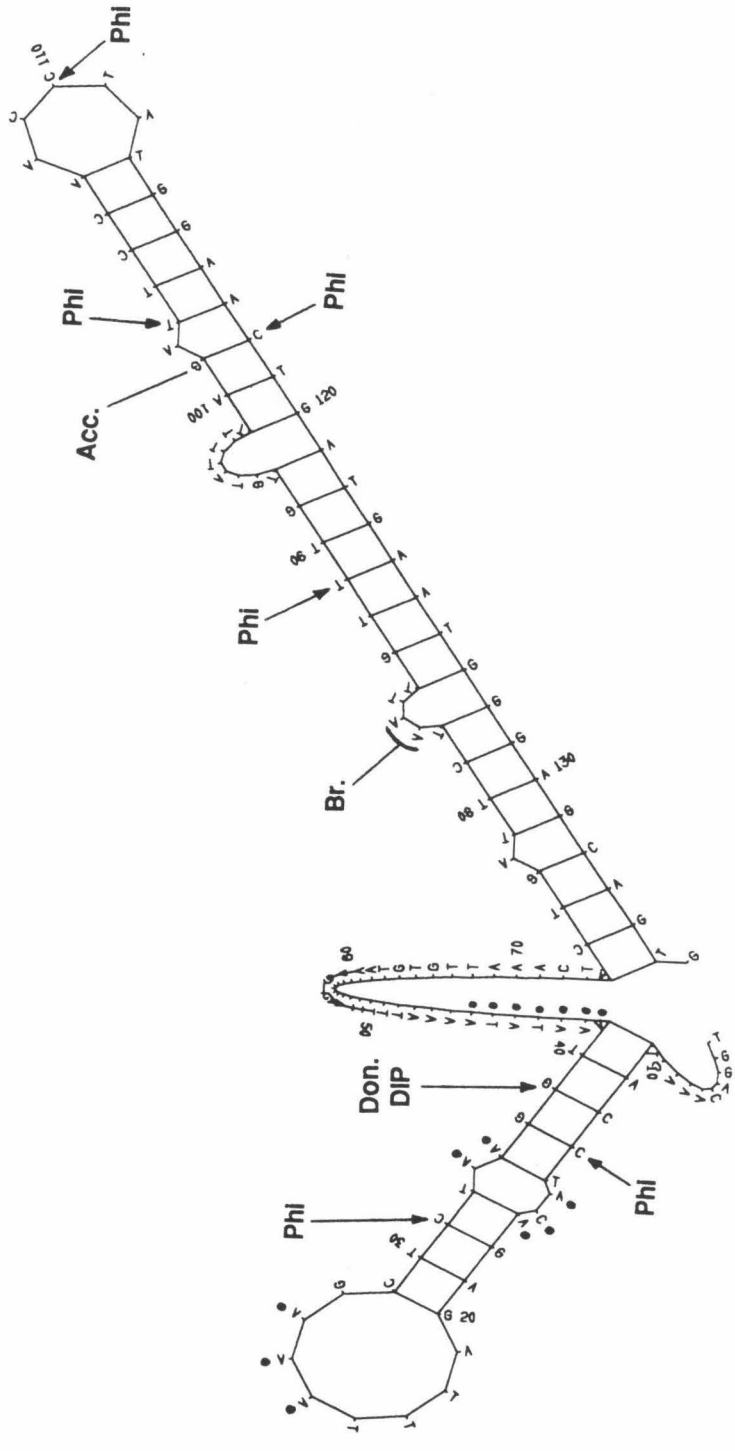
5.3 Model for the structure of the SV40 T-antigen intron DNA

For the SV40 ssDNA the available data do not allow a determination of a definite base-pairing interaction as was done for the Ad2 ssDNA. Nevertheless, the predicted structure agrees well with the available experimental data. Figure 5.3 shows the predicted structure for the SV40 T-antigen intron DNA. $Rh(DIP)_3^{3+}$ cleaves again at the bottom of the helix, close to the point where the two helices may meet in the three-dimensional structure. One of the $Rh(phen)_2phi^{3+}$ cleavage sites in the 3' stem structure conform to the usual double-stranded DNA site for the complex. Two of the sites are on opposite sides of a bulge of one nucleotide, which coincides with the 3' splice site of the intron. The last phi site is in the loop at the end of the helix, again suggesting that the complex might be targeting a structured loop. The $Rh(phen)_2phi^{3+}$ cleavage sites near the $Rh(DIP)_3^{3+}$ site are not the usual 5'-pyr-pyr-pur-3' sequences because of the bubble in the middle of the stem; it is possible that $Rh(phen)_2phi^{3+}$ may be interacting with some tertiary DNA structure created by the bulge.

Mung bean nuclease did not yield consistent results for the 3' helix, but the cleavage pattern for the 5' helix matches very well the predicted single-stranded regions. The middle region of no apparent structure consisting of about 35 nucleotides is not entirely cleaved by mung bean nuclease. Without further information on the structure in this region, a clear explanation of this observation is not possible; however, it is possible that parts of the middle region may actually be double-stranded or otherwise structured in solution, so that it is protected from attack by mung bean nuclease.

Thus the SV40 intron DNA also appears to involve two helical components at the ends of the intron that come together to form the structure that is targeted by the metal complexes. It is also likely in this case that the structure involves helix to helix

Figure 5.3 A structural model for the folding of the 136-mer of the SV40 T-antigen intron. There are two helices with an intervening region of no apparent structure. The two helices are proposed to interact to create a tertiary structural binding site for the metal complexes. Legend: Don., donor; Acc., acceptor; Br., branch; DIP, $Rh(DIP)_3^{3+}$ cleavage; Phi, $Rh(phen)_2phi^{3+}$ cleavage; Mse, Mse I cleavage; Psor., psoralen crosslink; •, mung bean nuclease cleavage.

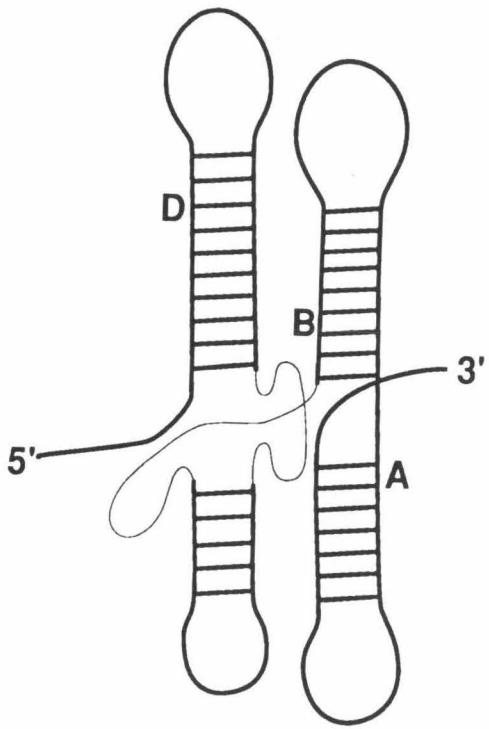

Phl
Phl
Phl
Acc.
Phl
Br.
Don.
DIP
Phl
Phl

interactions similar to that found in Holliday junctions; the two helices are likely to be parallel for this intron structure rather than coaxial. Part of the sequence in the middle may also participate in the formation of the overall structure, but there is no evidence to support such a role. A schematic illustration of the model is shown in Figure 5.4.

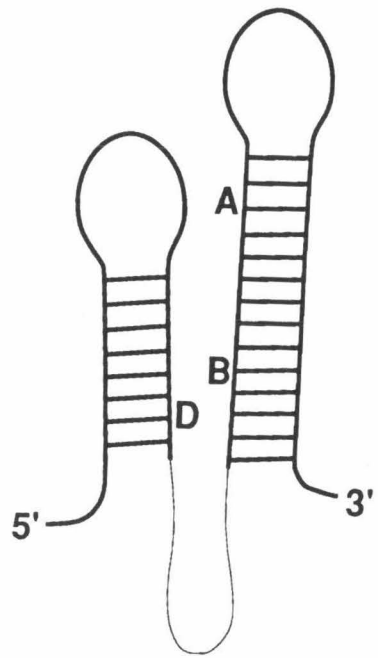
5.4 Summary

These two intron structures share many similarities In both structures exon sequences are involved, and there are two major components to the structure; one helix in the 5' half and the other in the 3' half. The middle region does not appear to be necessary for the integrity of the structure. $Rh(DIP)_3^{3+}$ cleavage sites are consistently near the "junction" of the helices, except in the 85-mer in which the site is shifted for plausible and explicable reasons. This points to the similarity of these structures to Holliday junction structures, where four helices lie stacked and side-by-side in a configuration known as "stacked-X" (3; See Figure 1.4). The two main helices in the intron structures most likely are arranged side-by-side with other possible helices coaxially stacked or otherwise placed to help stabilize the structure. This arrangement is also consistent with the cleavage results of the cruciforms (4) and synthetic Holliday junctions (2). The implications of these intron DNA structures are discussed in the next chapter.

Figure 5.4 Schematic representation of the intron DNA structures proposed for the SV40 T-antigen and the Ad2 E1A genes. The sequences around the splice sites form helices which interact with each other, and with other structural elements, to form a global structure that may resemble the structure of a Holliday junction. Legend: D, donor; A, acceptor; B, branch.



Ad2 E1A Intron DNA



SV40 T-antigen Intron DNA

References

1. M. Zuker (1989) *Science* **244**: 48-52 (b) J. A. Jaeger, D. H. Turner, & M. Zuker (1989) *Proc. Natl. Acad. Sci. USA* **86**: 7706 -7710.

2. K. Waldron, J. Voulgaris, & J. K. Barton. unpublished results.

3. D. M. J. Lilley & R. M. Clegg (1993) *Ann. Rev. Biophys. Biomolec. Struct.* **22**: 299-328.

4. M. R. Kirshenbaum, R. Tribolet, & J. K. Barton (1988) *Nuc. Acids Res.* **16**: 7948-7960.

5. C. S. Chow, L. S. Behlen, O. C. Uhlenbeck, & J. K. Barton (1992) *Biochemistry* **31**: 972-982.

Chapter 6

Functional and evolutionary implications of the intron DNA structures

6.1 Introduction

The discovery of the structures in the intron DNA of the two viruses, simian virus 40 and adenovirus 2, discussed in this thesis suggests a possible function of the introns in maintenance of the intron-exon structure of genes. What this function may be will be discussed in detail later in this chapter. However, in order to make sense of any argument for a functional importance of introns, we must first examine the evolution of the intron-exon arrangement of genes. The prevailing theory on the origin of introns postulates the existence of introns very early in the evolution of the gene structure of all organisms. According to this theory introns were subsequently lost in most prokaryotes but retained in eukaryotes in whom they facilitated further evolution into more complex forms of life, by enabling the recombination of exons into more complex gene arrangements. Another theory posits that introns were introduced late in the evolution of the gene structure, after the divergence of prokaryotes and eukaryotes. After their introduction the introns facilitated the evolution of eukaryotes through the same mechanism postulated by the former theory. Therefore, the end result of the intron-exon arrangement of genes is that it led to the evolution of new and more complex genes and thus to new and more complex life forms. The structures in the intron DNA, which may be a remnant of a past structure required for some function now lost, may have been utilized for a different purpose, such as shuffling of the exons, which then led to the current varieties of complex gene structures found in eukaryotes.

6.2 The origin and evolution of the intron-exon structure of genes

6.2.1 Exon theory of genes:

The exon theory of genes, proposed by Gilbert (1) soon after the discovery of introns (2) and expanded since then (3), postulates the existence of introns in the evolving gene structure of very early life forms before any of them diverged to form distinct groups of organisms. The intron's role was to assemble and reassort the exons as individual elements during the formation of these genes. After this initial evolution of genes, which was probably rapid on the evolutionary time scale, the introns made possible further assortment of the exons within genes over a longer time span by increasing the rate of functional exon recombination. If exons were contiguous, recombination between them would have to be carried out with amazing accuracy in order for the recombinant product to be functional. The introns provide a large space between exons over which recombination can occur without sacrificing the integrity of the exons. The presence and the length of the introns between exons also simply increases the frequency of recombination events between exons. Thus the overall effect of the presence of introns is to make recombination events between exons more frequent and less deleterious to functions of the gene products (3). This model of gene organization also suggests that the introns are likely to get longer as a result of frequent recombinations of exons. This is indeed observed to be the case: introns range from about 50 to 50,000 base pairs (3), while the range of exon lengths is much narrower and sharply centers around 40 amino acid residues, or 120 bp (4).

There are many examples of exon shuffling in genes that have arisen throughout evolution. The most striking of these is the use of eight exons encoding an epidermal-growth-factor-like domain present in three disparate genes; the low density lipoprotein (LDL) receptor, the epidermal growth factor precursor, and the

blood clotting factors IX and X (5). Another example is the nucleotide cofactor binding domains of glyceraldehyde-3-phosphate dehydrogenase (6), phospho-glycerate kinase (7), alcohol dehydrogenase (8), and pyruvate kinase (9), all of which are encoded by similar exons which are most likely to have arisen from the same original set of exons. Thus it is evident that exon shuffling has indeed occurred over the course of evolution as Gilbert postulated.

An example that points to the early origin of the intron is found in the gene for the ubiquitous enzyme triosephosphate isomerase, a gene that has apparently evolved to its final structure before the divergence of prokaryotes and eukaryotes (10). This enzyme carries out a basic step in glycolysis and gluconeogenesis, and its sequence is highly conserved across all species. The gene coding for triosephosphate isomerase has no introns in *E. coli* or in yeast; however, it has five introns in *Aspergillus nidulans* (11), six introns and chickens (12) and humans (13), and eight introns in maize (14). The positions of the introns in maize and the vertebrates are identical in five out of the eight introns of maize, suggesting that these introns were present before plants and animals diverged from each other. A plausible explanation for the three remaining introns is that the position of the one intron was shifted in animals while the other two introns were lost (10).

Thus the antiquity of introns and the shuffling of exons seems to be supported by available evidence. Much that is known about exons also support the notion that exons encode discrete domains of proteins. However, it has been argued that not all exons code for discrete structural domains and that what appears to be a discrete domain of a protein can be encoded by several exons (4). Examples in support of this argument are the gene structures of the serum albumin whose three specific binding functions are each encoded by four exons (15) and the serine proteinase binding function in ovomucoid which is encoded by two exons (16). This

observation can be easily explained and reconciled with the exon theory of genes by the phenomenon of retroposition, a process by which introns are removed over the course of evolution: a mature mRNA is copied back into DNA and inserted back into the genome. In some cases this would lead to no functional genes since the insertion point may not be transcribed at all. In other cases the insertion may be within an intron, and the restoration of proper splicing is all that is needed for the inserted DNA to act as a new complex exon encoding a single domain in the final product. Mutations over time within the exon itself can blur the previously distinct delineations between domains that might have existed before retroposition. Retroposition is seen in primates, for example, in the Alu element which is found in multiple copies throughout the genome and which is thought to originate from one or two source genes (17).

Yet another piece of evidence to support the exon theory of genes is the correlation of the chain length of a protein and the number of exons encoding the protein (4). When a plot of residue length versus number of exons is plotted for a number of proteins, a very nice positive correlation between the residue length and the number of exons is observed (Figure 6.1). This is consistent with the idea that exons code for discrete domains of protein structures and that larger and more complex proteins are put together by combining a number of exons. It is interesting to note in this regard that the peak residue length, in amino acids, of the exons is 40-45 for the same proteins plotted in Figure 6.1, and that the minimum size of peptides capable of assuming a stable folded structure has been estimated to be in the range of 20-40 residues (18). The average exon length corresponds well to the upper estimate for the chain length of a stable peptide.

The exon theory suggests the existence of intron-exon arrangement in the first genes that appeared in evolution. Introns facilitated, and perhaps catalyzed, the

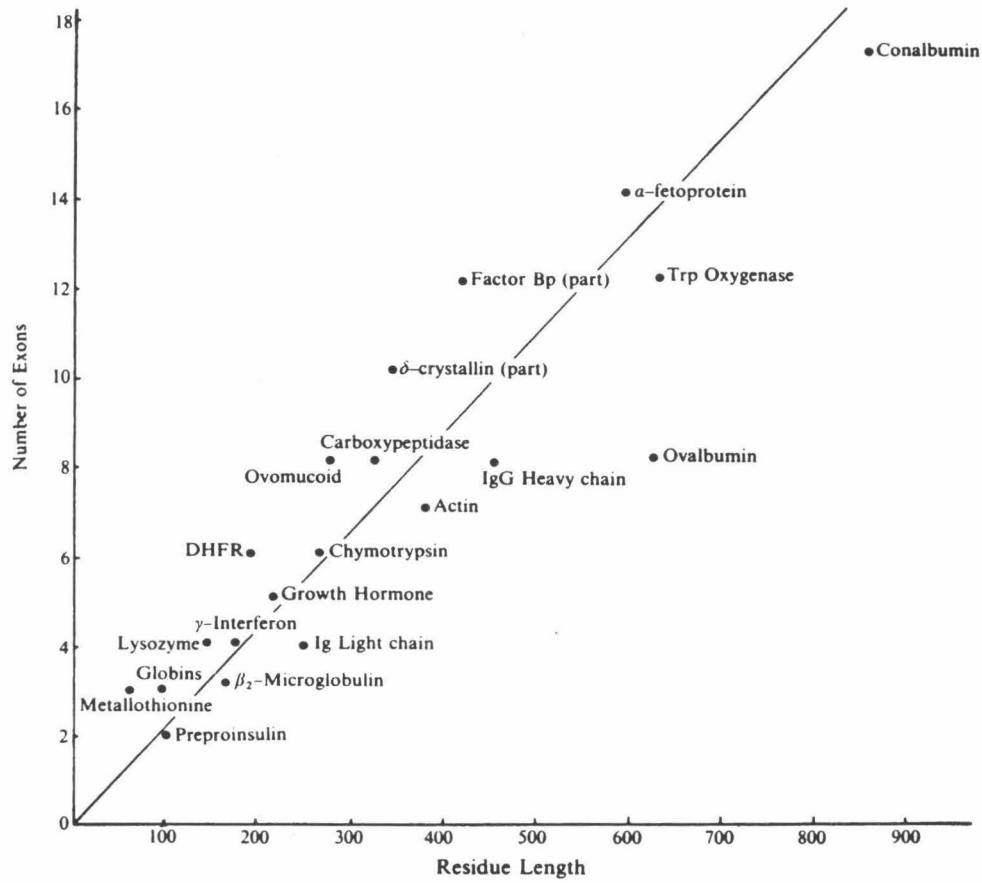


Figure 6.1 Plot of the number of exons against chain length for 20 different proteins. The full line represents one exon per 45 amino acid residues. Adapted from C. Blake (ref. 4).

assortment of message-coding exons into complex gene structures. Over time exons were fused to make longer and more complex ones. There was no specific evolutionary pressure on the length of the introns, and thus over time some of them gained a great deal in length while others stayed short. It still remains to be answered how the intron-exon structure originated in the first place. For this we must turn to the possibility of an RNA world.

6.2.2 The RNA world

The discovery of the catalytic RNA (19, 20) has brought about a breakthrough in our understanding of the origin of life. The puzzle of which came first, the nucleic acids or the proteins, seems to have been solved by the discovery that RNA can catalyze reactions which was previously thought to be the property only of protein enzymes. The reactions catalyzed by RNA are mainly limited to breaking and making of phosphodiester bonds; however, it has been shown recently that RNA can catalyze the hydrolysis of an aminoacyl ester bond (21). This raises the possibility that RNA might in theory have a variety of catalytic activities and that this might indeed have been the case in an RNA world which may have preceded proteins or DNA. RNA would have catalyzed all reactions in this primitive stage of life using their own functions arising from their structures and also utilizing cofactors such as metal ions or nucleotides (22). Uhlenbeck and coworkers have shown that RNA can be designed to use certain metals as cofactors (23).

In the RNA world, RNA would also serve as the store of genetic information as well as the gene products, the ribozymes. The gene structure would be organized in such a way that the introns excised themselves out to make the gene products which would be exons themselves. The exons, as well as the introns would fold into specific structures which gave them their functionalities. The intron's function

would simply be to excise itself out, while the functions of exons would be to serve the various processes required for primitive life. The genomes would have the capability to replicate themselves, or one of the gene products may catalyze the replication of the parent genome. Self-replication of RNA has been hinted at by experiments with sunY Tetrahymena self-splicing intron which was shown to catalyze the synthesis of complementary strand RNA by template-directed assembly of oligonucleotides (24).

First proteins would be simple homopolymers or oligopeptides of very few amino acids. These would first have minimal roles in the biochemistry of life, perhaps associating with and stabilizing RNA genomes or ribozymes. Thus the first protein "enzymes" would not add to the repertoire of activities already catalyzed by ribozymes but enhance the rate of reactions by stabilizing the structures of ribozymes and also by acting in the manner of cofactors. Gradually, the protein enzymes would start replacing the ribozymes, first through RNA-protein complexes and then by themselves. It is interesting to note that two of the most important and ancient processes, protein synthesis and nuclear pre-mRNA splicing, are catalyzed by RNA-protein complexes, by the ribosome and the spliceosome, respectively. Furthermore, evidence is accumulating that these processes are catalyzed by RNA rather than proteins (25, 26).

The common origin and the antiquity of introns becomes more evident when one examines the similarities between the three types of RNA splicing: group I, group II, and nuclear splicing (See Figure 6.2). Nuclear splicing and group I splicing were first discovered before group II splicing was recognized. Nuclear pre-mRNA splicing requires many trans-acting factors, and the intron is excised in a "lariat" form (27), while the self-splicing of Group I introns required only a guanosine cofactor and the intron was excised in a linear form (28). In Group II self-splicing

(29) no cofactor is required, and the intron is excised in a lariat form as in nuclear pre-mRNA splicing; the 2' hydroxyl group of the branch point nucleotide adenosine is directly involved in a nucleophilic attack on the 5' phosphate of the 5'-end nucleotide of the intron. This led to the proposal that nuclear splicing may have a mechanism similar to the self-splicing of Group II introns (30); it would be catalyzed, however, not by the pre-mRNA itself but by the snRNA's (small nuclear RNAs). The self-splicing introns have conserved sequence elements which fold into secondary and tertiary structures which provide the necessary framework for catalysis of the transesterification reactions. Pre-mRNAs, on the other hand, do not have such well conserved sequences and require participation of RNA and proteins in a complex called the spliceosome (31).

Recent experiments in structural probing and genetic analysis of snRNA sequences have suggested that the core structure of the spliceosome assembled by the pre-mRNA and the snRNA's may bear resemblance to the structures of group I and group II self-splicing introns (26). This raises the possibility then that all introns may have been self-splicing in the beginning of gene evolution and that some introns lost the ability to excise themselves out, perhaps as a result of recombination events within introns that destroyed the integrity of the intron sequences and structures. The splicing activity was provided by the remnants or fragments of spliced introns which assembled themselves into spliceosomes, which were then able to recreate the structure necessary for splicing of the now inert introns.

The RNA world scenario is consistent with the exon theory of genes in that the intron-exon arrangement was already built into the RNA genome as a result of the catalytic capabilities of the early RNA molecules. The discovery of the self-splicing intron and the increasing likelihood of RNA catalysis of nuclear splicing and the striking structural similarities among all types of splicing systems point to

the common origin of all introns in the RNA world. Evolution of proteins and DNA from RNA-based life forms is consistent with the biochemistry of today's life forms. DNA replication and transcription is entirely protein-based with ultra-fast reaction rates (3). RNA enzymes are known to be still functional in one system (20), other than self-splicing, and the evidence is growing for the functional roles of RNA in ribosomes (25) and spliceosomes (26). The intron-exon arrangement of genes was simply carried over when DNA replaced RNA as genetic material, which would have occurred through a process of reverse transcription of the RNA genome into DNA.

The switch to DNA as genetic material is likely to have occurred after protein enzymes were fully developed, for reasons cited above. The duration of the RNA world would have been relatively brief on the evolutionary time scale. RNA, being a much more labile molecule than DNA, would have undergone rapid changes leading to rapid evolution, and the closely following evolution of proteins would also have been rapid since it would have depended almost entirely on the properties of the RNA genome. Establishment of the DNA genome would most likely have preceded the divergence of the three main courses which evolution took; the eubacteria, the archaebacteria, and the eukaryotes (32). We know that eubacteria lack introns entirely, archaebacteria have some remnants of self-splicing (33), and that eukaryotes have introns in most of their genes.

The exon theory of Gilbert logically leads to the conclusion that introns were lost over the course of evolution of eubacteria and archaebacteria and that they were retained in eukaryotes. As the process of replication, transcription, and translation became more accurate and streamlined, the redundant non-informational DNA lost their original function in the RNA world and was under no selection pressure to persist. The loss of such useless sequences from bacteria is easily explained in

evolutionary terms, and it is also experimentally observed in modern day bacteria: E. coli loses unnecessary sequences promptly within a few generations (34). In contrast, mice retain unnecessary sequences for 15 generations or more (35). The bacteria gained in efficiency of genome organization and function while losing the evolutionary flexibility of the intron-exon arrangement of genes. In this sense, the genomes of prokaryotes can be considered very highly evolved and not inferior to the genomes of eukaryotes as is popularly assumed. However, the advantage gained by the eukaryotes was tremendous, as is clearly evident in the variety and complexity of the descendants of the early eukaryotes. Though it is difficult to explain—in strictly Darwinian terms and not imputing a purpose on the part of the early eukaryotes—why introns persisted in the eukaryotes, the fact remains that they did, and the effect of this persistence was an advantage in the long run. The persistence of introns in eukaryotes made possible further rearrangement of exons over time leading to ever more complex combinations of functions and phenotypes which would have been under the influence of natural selection at the organismal as well as the molecular level. The lost efficiency in genome organization and function was compensated for by the subsequently gained ability to adapt to a variety of conditions and led to the evolution of multicellular organisms inhabiting all habitats of the ecosystem.

6.2.3 Transposon theory of introns

The opposing view on the evolution of introns holds that introns were introduced late in the evolution of life forms on earth, certainly after the divergence of prokaryotes and eukaryotes (36-38). This makes at least some intuitive sense; prokaryotes do not have introns, and eukaryotes do. The assumption is that the introduction of introns in eukaryotes gave them an evolutionary advantage over the

prokaryotes, and the eukaryotes were then able to develop into complex multi-cellular organisms. The origin of introns in this theory is postulated to be the transposable elements found today in both prokaryotes and eukaryotes (39). The transposon would have had the ability to excise themselves out of genomes, and this need only have been carried over to the ability to excise RNA to give rise to RNA splicing. In support of this theory it is pointed out that the majority of protein domains are encoded by more than one exon, and thus it is unlikely that exons encoding discrete protein domains existed as such (38). It is also argued that the loss of every single intron by accident is far too improbable to have occurred (36, 38). Another reasonable argument is that exon length in mitochondria and chloroplasts is much less uniform than in nuclei where the arrangement of nucleosomes may have directed the uniform insertion of introns (38).

There are many problems with this theory. It does not provide a good explanation of the extremely wide range of intron length unless, as postulated by the exon theory of genes, numerous recombinations took place within introns over the course of evolution. The improbability of the loss of all introns in prokaryotes is a compelling argument for this theory, but it is to be noted that modern day bacteria lose useless sequences of DNA readily (34). The recent discovery of an intron in archaebacteria (33) also argues against the intron-late theory and supports the idea that introns existed from very early on. But the most compelling argument against the late appearance of introns is the plausibility of the RNA world. It is becoming increasingly likely that all splicing is catalyzed by RNA and that all of the splicing systems share the common structural elements (26). The discovery of self-splicing itself and the chemical reversibility of transesterification reactions catalyzed by RNA raise the possibility, as discussed in the previous section, that RNA was the sole catalytic molecule when life began.

Thus introns arose simply as one of the many properties of catalytic RNA, as the function of RNA that catalyzed breaking and joining of phosphodiester bonds, which would have been responsible also for replication of RNA genomes. Other RNA molecules with different catalytic activities would have existed as well, and these would have been replicated by the molecules carrying the replicase function, which would not have been very different from the degradative function and possibly could have been carried on the same ribozyme. The arrangement of RNA into longer pieces containing the various functionalities is not difficult to imagine in this scenario. The arrangement of exon-intron-exon is logical if we assume that certain ribozymes could not cleave or join phosphodiester bonds; only the "introns" would have been capable of this, and thus the arrangement would necessarily have to be alternating introns and exons. This theory is also consistently Darwinian in that it imputes no purpose on any part of the processes. Introns did not have any specific purpose to be inserted in between exons; they simply combined with the exons to create long RNA chains which then served as genomes. Once there, however, introns persisted lacking any strong selection pressure against them. The transition to DNA as the genetic material made the removal and loss of introns more difficult. One remaining puzzle in the exon theory of genes is why introns were lost in most prokaryotes while persisting in the eukaryotes. There is no readily apparent answer to this question so far. It is only observed that bacteria have extremely efficient genomes able to streamline itself back to its original efficiency when artificially altered (34).

6.3 Effect of introns and intron DNA structures on the course of evolution

The accidental organization of the intron-exon arrangement of genes had the effect that the protein coding exons would be assorted and recombined to form more

complex genes. This is the central tenet of the exon theory of genes, and it explains the evolutionary success, in terms of variety and complexity, of eukaryotes over prokaryotes. As mentioned in preceding sections, however, evolution does not have foresight, unless we are ready to invoke the force of a deity in shaping the course of evolution. Thus the intron-exon arrangement of genes cannot have been for a future evolutionary fitness. That it has been maintained so far in spite of its apparent biochemical inefficiency indicates a tremendous *advantage*, but not a *function*, in the strict evolutionary definition of the word, of the arrangement over that without introns. The accidental origin of introns does not make them an adaptation, especially when they may actually have been less advantageous in the early stage of evolution, when the DNA genome had been established and the lability of the RNA genome was no longer necessary for rapid changes and adaptations.

Introns fall into the category of characters called *exaptations* defined by Gould and Vrba (40). Exaptation is a character evolved for other uses, or for no function at all, but later coopted for its current role. Exaptations have *effects* rather than *functions* which are the properties of *adaptations*, features that promote fitness and are built by selection for their current role. Thus the original function of introns in assorting and assembling genes in the RNA world was no longer needed once the genetic material was transferred to DNA. But their presence was coopted for later usefulness in further recombination of the elements of genes into more complex ones over a longer evolutionary course. So the "function" ascribed to them in the exon theory of genes is strictly speaking an *effect* of their persistence.

The RNA origin of introns also raises the possibility that the structure of the RNA intron, which would have been necessary for its function, could have been carried over to DNA. Under certain circumstances, such as high superhelicity, the intron DNA could assume a structure similar to that of the intron RNA, This DNA

structure would have no function, as there would be nothing in the biochemistry of the cell to utilize it, but it is again reasonable to postulate that this structure could have been coopted for use by a newly evolved function such as a recombinase. The intron DNA structure would again fall into the category of exaptations. It would specifically aid in exon shuffling to create new genes. Such a structure would also increase the rate and accuracy of recombination events leading to more rapid evolution of genes, and thus adds a new dimension to the exon theory of genes by giving the introns themselves an active role in the evolution of genes.

The intron DNA structures found in the two viruses studied in this thesis show exactly the characteristics of such an RNA-derived DNA structure, and it is reasonable to postulate that they could be the kind of exaptation discussed above. It is interesting to note in this regard that only the coding strand is seen to fold into a stable structure targeted by the metal complex $Rh(DIP)_3^{3+}$, an observation consistent with the idea that the structures have their origins in RNA. However, the fact that exon sequences are also required for the structure is at odds with the idea that it originated from the intron RNA structure alone; it is possible that there were specific structures involving both the intron and the exon sequences in RNA which then was carried over to DNA. How much the intron DNA structures have to do with the current splicing of the transcripts is not clear; certainly, the splicing of these particular RNA transcripts requires the spliceosome and thus involves the snRNA sequences as well as the transcript itself. Whether the transcript itself assumes a structure similar to that found in DNA is being investigated in this laboratory (41).

There are no known enzymes that recognize specific intron sequences for recombination of the adjacent exon with other exons. So no experimental evidence can be given to support the notion that the intron structures are recombinase target sites. However, introns have been known to transpose, and the transposition of these

introns is dependent on the proteins encoded within the introns (42), suggesting that the proteins may actually be involved in cutting out and reinserting the introns. Admittedly this is a different process than exon shuffling, but it suggests a good probability for a similar mechanism in exon shuffling. Finally, the similarity of the intron DNA structures to Holliday junction structures, which are thought to be intermediates in recombination, also suggests the possibility of a similar mechanism in exon shuffling. There have already been discovered enzymes which cleave Holiday junction structures (43).

6.4 Future directions

The presence of the intron DNA structures in the two viruses, SV40 and Adenovirus 2, suggests a possible role for the introns in evolution. The first step to testing for the validity of this hypothesis would be to determine the generality of the intron DNA structures in various other systems. Among the many candidates for study may be the genome of Plasmodium studied by McCutchan, et al. (44), which has already been shown to have interesting structural polymorphisms, though at very low resolution of structural probing by mung bean nuclease. The human globin gene is already being studied in this laboratory (41). Another system worth studying may be the genes of LDL receptor, the epidermal growth factor precursor, and the blood clotting factors IX and X (5) which contain exons thought to have originated from one source through exon shuffling. The gene for triosephosphate isomerase (10), an ancient gene whose sequence is highly conserved across all species, is another good candidate for structural investigations.

The second approach to determining the evolutionary or physiological significance of the intron structures would be to isolate enzymes or factors that recognize the structures. The intron structures can be synthesized and attached to an

affinity column on which cell extracts of various systems can be tested for presence of specific binding proteins. If identified, the protein can be isolated and its binding properties can be studied.

Another line of inquiry would be to study the structure of the RNA counterpart of the introns in the pre-mRNA molecules. Structures that may be present in RNA may have similar characteristics to the DNA structures. The RNA structures may be involved in the initial assembly of spliceosomes, or they may also be involved in the actual processing of the pre-mRNA. Structures assumed by the complex of snRNA and pre-mRNA may also be of interest for study, as it was proposed that at one stage of splicing the complex may assume a Holliday-like structure (45). The DNA structure observed in the introns may be coincident with some of these RNA structures, or they may be entirely independent DNA structures on their own. Investigation into RNA structures would provide an answer to this question.

All of the proposed studies above would provide information on which a plausible argument for the evolutionary and physiological significance of the intron DNA structures and thus lead to a deeper understanding of the unexpected and puzzling phenomenon of introns.

References

1. W. Gilbert (1978) *Nature* **271**: 501.
2. (a) A. J. Berk & P. A. Sharp (1977) *Proc. Natl. Acad. Sci. USA* **74**: 3171-3175 (b) L. T. Chow, R. E. Gelinas, T. R. Broker, & R. T. Roberts (1977) *Cell* **12**: 1-8.
3. W. Gilbert (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52**: 901-905.
4. C. Blake (1983) *Nature* **306**: 535-537.
5. (a) T. C. Südhof, J. L. Goldstein, M. S. Brown, & D. W. Russell (1985) *Science* **228**: 815-822. (b) T. C. Südhof, D. W. Russell, J. L. Goldstein, M. S. Brown, R. Sanchez-Pescador, & G. I. Bell (1985) *Science* **228**: 893-895.
6. E. M. Stone, K. N. Rothblum, R. J. Schwartz (1985) *Nature* **313**: 498-500.
7. A. M. Michelson, C. C. F. Blake, S. T. Evans, & S. H. Orkin (1985) *Proc. Natl. Acad. Sci. USA* **82**: 6965-6969.
8. C.-I. Branden, H. Eklund, C. Cambillan, & A. J. Pryor (1984) *EMBO J.* **3**: 1307-1310.
9. N. Lonberg & W. Gilbert (1985) *Cell* **40**: 81-90.
10. W. Gilbert, M. Marchionni, & G. McKnight (1986) *Cell* **46**: 151-154.
11. G. L. McKnight, P. J. O'Hara, & M. L. Parker (1986) *Cell* **46**: 143-147.
12. D. Strauss & W. Gilbert (1985) *Mol. Cell. Biol.* **5**: 3497-3506.
13. J. R. Brown, I. O. Daar, & J. R. Krug (1985) *Mol. Cell. Biol.* **5**: 1694-1706.
14. M. Marchionni & W. Gilbert (1986) *Cell* **46**: 133-141.
15. F. A. Eiferman, P. R. Young, R. W. Scott, & S. M. Tilghman (1981) *Nature* **294**: 713-718
16. J. P. Stein, J. F. Catterall, P. Kristo, A. R. Means, & B. W. O'Malley (1980) *Cell* **21**: 681-687.

17. (a) G. B. Hutchinson, S. E. Andrew, H. McDonald, Y. P. Goldberg, R. Graham, J. M. Rommens, & M. R. Hayden (1993) *Nuc. Acids Res.* **21**: 3379-3383. (b) A. D. Bailey & C. K. J. Shen (1993) *Proc. Natl. Acad. Sci. USA* **90**: 7205-7209.

18. D. B. Wetlaufer (1981) *Adv. Prof. Chem.* **34**: 61

19. K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling & T. R. Cech (1982) *Cell* **31**: 147-157.

20. C. Guerriertakada, K. Gardiner, T. Marsh, N. Pace, & S. Altman (1983) *Cell* **35**: 849-857.

21. J. A. Piccirilli, T. S. McConnel, A. J. Zaug, H. F. Noller, T. R. Cech (1992) *Science* **256**, 1420-1424.

22. W. Gilbert (1986) *Nature* **319**: 618.

23. (a) T. Pan & O. C. Uhlenbeck (1992) *Biochemistry* **31**: 3887-3895. (b) T. Pan & O. C. Uhlenbeck (1992) *Nature* **358**: 560-563.

24. J. A. Doudna, S. Couture, & J. W. Szostak (1991) *Science* **251**: 1605-1608.

25. H. F. Noller, V. Hoffarth, & L. Zimniak (1992) *Science* **256**, 1416-1419.

26. (a) H. D. Madhani & C. Guthrie (1992) *Cell* **71**: 803-817. (b) D. S. McPheeters & J. Abelson (1992) *Cell* **71**: 819-831. (c) C. F. Lesser & C. Guthrie (1994) *Science* **262**: 1982-1988. (d) E. J. Sontheimer & J. A. Steitz (1994) *Science* **262**: 1989-1996.

27. (a) R. A. Padgett, P. J. Grabowski, M. M. Konarska, S. Seiler, & P. A. Sharp (1986) *Ann. Rev. Biochem* **55**: 1119-1150. (b) P. A. Sharp (1987) *Science* **235**: 766-771.

28. T. R. Cech (1990) *Ann. Rev. Biochem* **59**: 543-568.

29. (a) C. L. Peebels, P. S. Perlman, K. L. Mecklenburg, M. L. Petrillo, J. H. Tabor, K. A. Jarrel, & H.-L. Cheng (1986) *Cell* **44**: 213-223. (b) R. Van der Veen, A. C. Arnberg, G. van der Horst, L. Bonen, H. F. Tabak & L. A. Grivell (1986) *Cell* **44**: 225-234.

30. (a) P. A. Sharp (1985) *Cell* **42**: 397-400. (b) T. R. Cech (1986) *Cell* **44**: 207-210.

31. E. Brody & J. Abelson (1985) *Science* **228**: 963-967.

32. C. R. Woese (1981) *Sci. Am.* **244**: 98-125.

33. J.-L. Ferat & F. Michel (1993) *Nature* **364**: 358-361.

34. (a) W. R. Folk & P. Berg (1971) *J. Mol. Biol.* **58**: 595-610. (b) P. W. J. Rigby, B. D. Burleigh, & B. S. Hartley (1974) *Nature* **251**: 200-204.

35. (a) R. Jaennisch (1976) *Proc. Natl. Acad. Sci. USA* **73**: 1260-1264. (b) J. Johler, R. Timpi, & R. Jaennisch (1984) *Cell* **38**: 597-607.

36. F. H. C. Crick (1979) *Science* **204**: 264-271.

37. D. A. Hickey & B. Benkel (1986) *J. Theor. Biol.* **121**: 283-291.

38. T. Cavalier-Smith (1985) *Nature* **315**: 283-284.

39. (a) W. F. Doolittle & C. Sapienza (1980) *Nature* **284**: 601-603. (b) L. E. Orgel & F. H. C. Crick (1980) *Nature* **284**: 604-607. (c) D. A. Hickey (1982) *Genetics* **101**: 519-531.

40. S. J. Gould & E. S. Vrba (1982) *Paleobiology* **8**: 4-15.

41. T. Johann & J. K. Barton, unpublished results.

42. (a) I. G. Macreadie, R. M. Scott, A. R. Zinn, & R. A. Butow (1985) *Cell* **41**: 395-402. (b) F. Michel & B. F. Lang (1985) Nature 316: 641-643.

43. (a) A. Bhattacharyya, A. I. H. Murchie, E. v. Kitzing, S. Diekmann, B. Kemper, & D. M. J. Lilley (1991) *J. Mol. Biol.* **221**: 1191-1207. (b) C. A. Parsons, A. I. H. Murchie, D. M. J. Lilley, & S.C. West (1989) *EMBO J.* **8**: 239-246.

44. T. F. McCutchan, J. L. Hansen, J. B. Dame, & J. A. Mullins (1984) *Science* **225**: 625-628.

45. J. A. Steitz (1992) *Science* **257**: 888-889