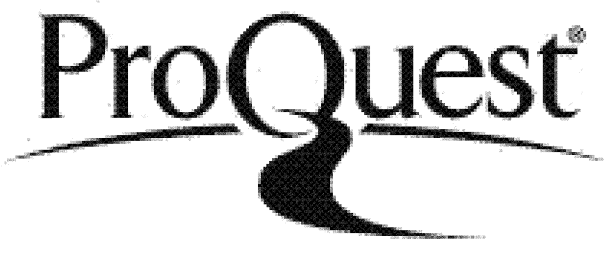
Investigation of the mechanism of pre-mRNA splicing using phosphorothioate analogues of RNA.

Andrew David Griffiths BSc. (Sp. Hons. Sheffield)

A thesis submitted in accordance with the requirements of the University of Leicester for the degree of Doctor of Philosophy.

Department of Biochemistry

December 1988



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Statement

This thesis is based on work conducted by the author in the Department of Biochemistry of the University of Leicester during the period between October 1985, and October 1988.

All the work recorded in this thesis is original unless otherwise acknowledged in the text or by references. None of the work has been submitted for any other degree in this or any other university.

Signed: ....A. Griffiths...

Date: ....19:12:88.....

I dedicate this work to my Grandmother (Lucy Hulley),
and to my Parents.

I cannot imagine how life would have been without their
unfailing love and support.

ACKNOWLEDGMENTS

I acknowledge the receipt of a studentship from the Medical Research Council which has enabled me to carry out the work described in this thesis.

I would also like to thank my supervisor, Dr. Ian Eperon, for his help and advice during the time I have spent in his laboratory, and for kindly reading this manuscript.

I also thank all the other members of the laboratory, past and present: Lucy, Judith, Peter, Whizz (Ian Graham), Tony, Caroline, Sonia, Ian (Reed), Ian (Wheatley), ATR (Andrew Turnbull-Ross), Andy (Lear), Julie, Jenny, David and Sophie, for their advice, practical assistance, and friendship.

Thanks also to all the other members of the Biochemistry department who have provided me with advice or who have been foolhardy enough to lend me equipment. Special thanks go to John Keyte and Jim Turner for the rapid and accurate synthesis of numerous oligonucleotides, without which much of this project would not have been possible.

Abstract

Investigation of the mechanism of pre-mRNA splicing using phosphorothioate analogues of RNA.

Andrew David Griffiths
1988

The Sp diastereoisomer of adenosine 5'-O-(1-thiotriphosphate) (ATPαS) was found to be incorporated into RNA transcribed by T7 RNA polymerase with an apparent K_M similar to that for ATP; the Rp diastereoisomer was neither a substrate nor a competitive inhibitor. The configuration of the thiophosphodiester link in the RNA produced was analysed with stereospecific nucleases. Surprisingly, the nucleases exhibited reduced discrimination compared to their activity on dinucleotides. The results show that phosphorothioate linkages in T7 RNA polymerase transcripts are in the Rp configuration. Thus, the transcription reaction proceeds with inversion of configuration at phosphorus.

In vitro transcription by T7 RNA polymerase in the presence of a nucleoside 5'-O-(1-thiotriphosphate) has been used to prepare pre-mRNA analogues of the small intron of a rabbit β-globin gene and flanking exon sequences. Incubation of transcripts prepared with ATPαS in a HeLa cell nuclear extract showed that the presence of the thionucleotide in a transcript inhibited splicing, but a novel product (termed E2*) was formed by cleavage three nucleotides upstream of the 3' splice site. This product was formed with the same kinetics as the intermediates of a normal splicing reaction, and its formation depended on the presence of ATP, Mg^{2+}, and intact small nuclear RNAs U1, U2, and U6. Hence, E2* formation would seem to be a splicing related phenomenon.

However, E2* formation does not require all the components necessary for *in vitro* splicing. E2* formation was supported by S100 extracts and mildly heat treated nuclear extracts, both of which were inactive in splicing. Retardation gel asssays and combined RNase T1 digestion and immunoprecipitation experiments indicated that large spliceosome-like complexes did not form on transcripts prepared with ATPαS. Furthermore, neither the absence of a functional 5' splice site nor polypyrimidine tract sequence prevented E2* formation, despite both these sequences being required for splicing. In addition, a nuclease activity has been identified in HeLa nuclear extract which is dependent on intact small nuclear RNAs U1 and U2. This activity may be involved in E2* formation.

Table of Contents

Index of Figures and Tables

(Figures and tables are found inserted into the text after the appropriate page)

Abbreviations

RNA	ribonucleic acid
[S]RNA	an RNA analogue with all phosphodiester links replaced by thiophosphodiester links
[sC]RNA	an RNA analogue with all phosphodiester links 5' to cytidines replaced by thiophosphodiester links
[sA]RNA	an RNA analogue with all phosphodiester links 5' to adenosines replaced by thiophosphodiester links
[sG]RNA	an RNA analogue with all phosphodiester links 5' to guanosines replaced by thiophosphodiester links
[sU]RNA	an RNA analogue with all phosphodiester links 5' to uridines replaced by thiophosphodiester links
[sC/sA]RNA	an RNA analogue with all phosphodiester links 5' to cytidines and adenosines replaced by thiophosphodiester links
DNA	deoxyribonucleic acid
cDNA	complementary DNA
RF	replicative form (of bacteriophage M13)
NTP	ribonucleoside 5'-triphosphate
ATP	adenosine 5'-triphosphate
NTPαS	nucleoside 5'-O-(1-thiotriphosphate)
dNTP	2'-deoxyribonucleoside 5'-triphosphate
ddNTP	2',3,-dideoxyribonucleoside 5'-triphosphate
C	cytosine
A	adenine
G	guanine
T	thymine
U	uracil
N	one of the five bases above
mRNA	messenger RNA
$tRNA^{phe}$	transfer RNA: phenylalanine specific
AMV	avian myeloblastosis virus
SDS	sodium dodecylsulphate
EDTA	diaminoethanetetra-acetic acid
EGTA	ethyleneglycol-*bis*-(β-aminoethyl ether) N,N,N',N'-tetraacetic acid
DTT	dithiothreitol
PMSF	phenylmethylsulphonylfluoride
Tris	Tris (hydroxymethyl)-aminomethane
MOPS	3-[N-morpholino]propanesulphonic acid
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid

IPTG	isopropyl-β-D-thiogalactopyranoside
BCIG	5-bromo, 4-chloro, 3-indolyl, β-galactoside.
RNase	ribonuclease
RNP	ribonucleoprotein
snRNP	small nuclear ribonucleoprotein
hnRNP	heterogeneous nuclear ribonucleoprotein
PAGE	polyacrylamide gel electrophoresis
HPLC	high performance liquid chromatography
TLC	thin layer chromatography
PEI	polyethyleneimine
PVA	polyvinyl alcohol
TEAB	triethylammonium bicarbonate

Chapter 1.

Introduction.

1. INTRODUCTION

Most metazoan RNA polymerase II transcripts must undergo a variety of post-transcriptional processing events before they become mature, translatable, messenger RNAs, and are exported from the nucleus. Usually the 5'-end of the RNA is modified shortly after initiation by the addition of 7-methyl guanosine triphosphate and methylation to form the cap structure (7mGpppmXpmY. . .) (Salditt-Georgieff *et al.*, 1980). Most transcripts are also modified by the addition of a poly(A) tail at the 3'-end (Darnell, 1982). Transcription usually continues beyond the 3'-end of the mature mRNA and an endonucleolytic cleavage generates the site of poly(A) addition. Approximately 0.1% of adenosine residues are methylated at position 6, usually at the sequence N_1-($^{G}/_{A}$)-m^6A-C-N_2 where N_1 is typically a purine and N_2 is rarely a guanosine (Schibler *et al.*, 1977; Chen-Kiang *et al.*, 1979).

Most RNA polymerase II transcripts must also have internal sequences (introns) precisely removed by cleavage-ligation reactions in the process of *splicing*.

The study described in this thesis is an attempt to use phosphorothioate analogues of RNA to investigate the mechanism of metazoan pre-mRNA splicing. This chapter gives a description of previous studies on the mechanism of pre-mRNA splicing, together with an overview of the use of nucleoside phosphorothioates in the investigation of other systems. Phosphorothioate analogues of RNA for use in this study were transcribed by bacteriophage T7 RNA polymerase. It was, therefore, important to know the stereochemistry of the reaction catalysed by this enzyme, since internucleotidic thiophosphodiester linkages are chiral groups. Chapter 3 describes the investigation of the stereochemistry of transcription by T7 RNA polymerase, together with an investigation of the stereospecificity of certain nucleases towards phosphorothioate-substituted RNA. Chapter 4 describes a study of the *in vitro* splicing of phosphorothioate-substituted RNAs and reveals a fragment, which is derived from the 3'- end of the pre-mRNA but which is neither a normal product nor intermediate of splicing, to be produced in a splicing-related reaction. A further investigation of the relationship of this novel reaction to splicing is described in Chapter 5, in particular the requirement (or otherwise) of a splicing complex is investigated.

1.1. MECHANISM OF SPLICING

The development of *in vitro* splicing systems has been central in the study of the biochemical mechanism of the splicing reaction. Cell-free splicing systems have been based on whole cell (Manley *et al.*, 1980) or nuclear extracts (Dignam *et al.*, 1983; Shapiro *et al.*, 1988) of cultured cells, usually HeLa cells, but plasmacytoma (Goldenberg and Hauser, 1983) and lymphoblastoid cells (Kedes and Steitz, 1987; Lowery and van Ness, 1988) have also been used. *Drosophila* nuclear extracts from Kc tissue culture cells or 0-12 hour embryos have also been found to support *in vitro* splicing (Rio, 1988). Although both whole-cell extracts (Kole and Weissman, 1982; Padgett *et al.*, 1983a) and nuclear extracts (Hernandez and Keller, 1983; Krainer *et al.*, 1984) splice pre-mRNA accurately, splicing is generally more efficient in nuclear extracts. Whole-cell extracts from yeast have also been developed which allow accurate and efficient splicing of yeast pre-mRNAs (Lin *et al.*, 1985). Although RNA polymerase II in extracts has been used to produce pre-mRNA substrates for *in vitro* splicing, both by transcription of appropriate DNA templates separately (Hardy *et al.*, 1984) and in a coupled transcription/splicing system (Padgett *et al.*, 1983a), substrates are generally synthesised by efficient *in vitro* transcription using purified bacteriophage RNA polymerases. Bacteriophage polymerases can be used to produce "run-off" transcripts of a eukaryotic structural gene fused to the appropriate bacteriophage promoter. SP6 RNA polymerase was the enzyme first used, and transcripts made using it were found to be spliced accurately when microinjected into *Xenopus* oocyte nuclei (Green *et al.*, 1983) and in HeLa cell nuclear extracts (Krainer *et al.*, 1984). Although SP6 transcripts can be capped enzymatically (Green *et al.*, 1983), SP6 polymerase will initiate transcription efficiently with a cap analogue such as G(5')ppp(5')G included in the transcription reaction(Konarska *et al.*, 1984). Other bacteriophage polymerases have also been used more recently. All the bacteriophage polymerases used (SP6, T7, and T3) seem to show very similar properties.

Both mammalian (Hardy *et al.*, 1984; Hernandez and Keller, 1983) and yeast (Lin *et al.*, 1985) splicing *in vitro* require Mg^{2+} and ATP as cofactors. Only low concentrations of Mg^{2+} were required (1 to 3mM optimum), higher concentrations being inhibitory. Although dATP will substitute for ATP (Krainer *et al.*, 1984; Ruskin and Green, 1985a), no other ribonucleoside triphosphates are effective cofactors (Hardy *et al.*, 1984; Lin *et al.*, 1985). Phosphonate analogues of ATP with non-hydrolysable α–β and β–γ linkages do not allow splicing, perhaps implying a role for different bond hydrolyses in different steps of splicing (Krainer *et al.*, 1984; Lin *et al.*, 1985).

The substrate itself requires neither a poly(A) tail nor complete exons to be spliced efficiently *in vitro*. The role of the 5' 7mGpppN cap structure is, however, less clear. Capped pre-mRNAs are spliced more efficiently, and

produce fewer aberrant splicing products than when capped RNA was used in nuclear extracts (Krainer *et al.*, 1984). In whole cell extracts the cap dependance was even more stringent (Konarska *et al.*, 1984) and addition of less than 10μM of cap analogues can strongly inhibit splicing of capped RNAs if added at the start of the reaction, but not at later times in the incubation, suggesting cap recognition may be an early step in splicing. Certain nuclear extract systems are inhibited strongly by cap analogues (A.L. Lear, personal communication), whereas others are only partly inhibited unless the extract was first pre-incubated with cap analogue and Mg^{2+} prior to addition of substrate RNA (Edery and Sonenberg, 1985).

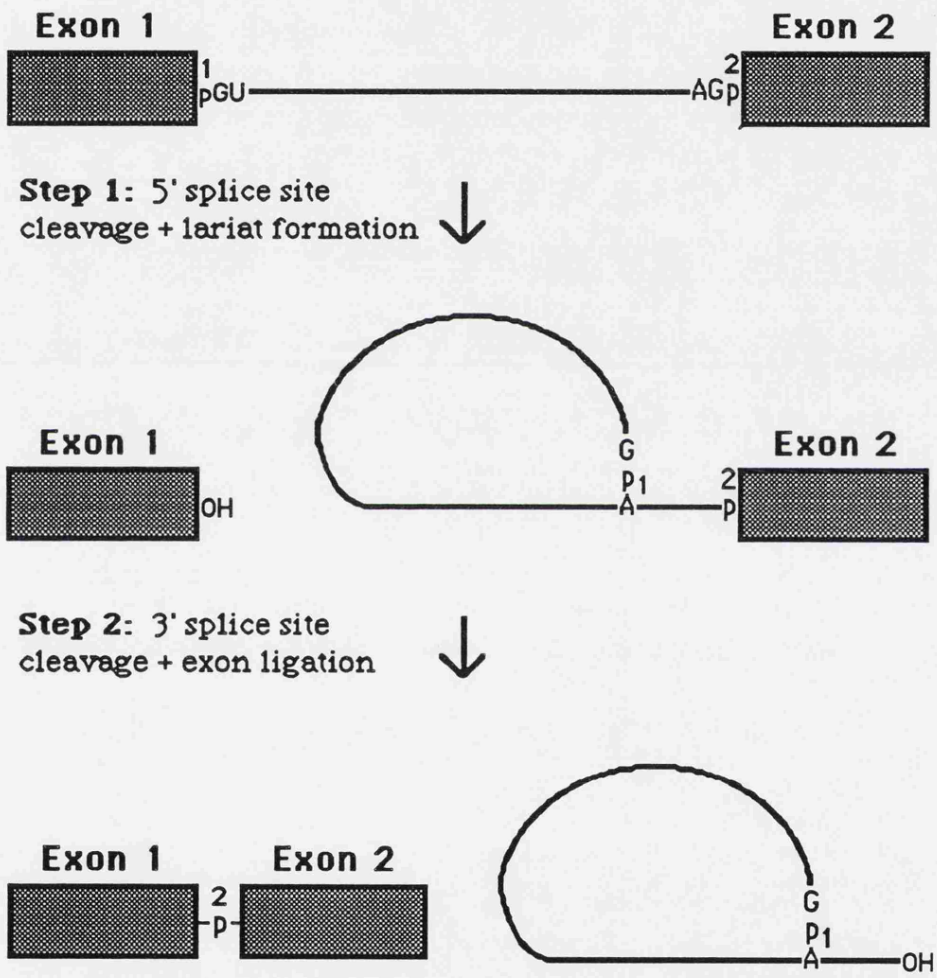
The pathway of pre-mRNA splicing has largely been determined by studies in which ^{32}P-labelled mammalian pre-mRNAs were spliced *in vitro*. Novel RNA structures in addition to precursor and ligated exons were found, and based on time courses and detailed characterisation of the novel structures a two-step splicing pathway was formulated (Fig. 1.1).

Krainer *et al.* (1984) carried out S1 nuclease mapping and primer-extension on unfractionated *in vitro* RNA processing products and identified a product cut at the 5' splice site and containing intron and 3'-exon, thereby providing the first evidence that 5' splice site cleavage precedes 3' splice site cleavage. No RNA generated by cleavage at the 3' splice site without exon ligation was ever found.

These findings were rapidly confirmed by *in vitro* splicing of ^{32}P-labelled RNA containing a single intron (Grabowski *et al.*, 1984; Padgett *et al.*, 1984; Ruskin *et al.*, 1984). The first detectable RNA processing event was cleavage at the 5' splice site, producing free 5'-exon and a molecule consisting of the intron still joined to the 3'-exon. Subsequently, accurately spliced RNA and the excised, intact intron were simultaneously observed. The two intron containing species were found not to be simple linear molecules as they had several unusual properties including: aberrant mobilities on denaturing polyacrylamide gels, an indication of an unusual topological structure; a block to reverse transcription near the 3' end of the intron mapping to the same site in both species; the presence of a nuclease resistant component. Krainer *et al.* (1984) had previously suggested that one or more products of *in vitro* processing contained a branch structure such as those detected in nuclear but not cytoplasmic RNA (Wallace and Edmonds, 1983) because a primer extension product was found whose 3' end mapped to a discrete site near the 3' end of the intron, yet no RNA cleavage was revealed at this site by S1 nuclease mapping. Oligonucleotide directed RNase H cleavage revealed the RNAs to be circular at their 5', but not 3' ends (Ruskin *et al.*, 1984). The nuclease resistant component was found to map to the same position as the primer extension stop (Ruskin *et al.*, 1984) and was identified as a branched trinucleotide as described by

Figure 1.1. **The pre-mRNA splicing pathway.**

A schematic representation of the pre-mRNA splicing pathway: boxes represent exons; the line represents an intron. The conserved GU at the 5' splice site, and AG at the 3' splice site, together with the A at the branch point are indicated. The phosphates at the 5' and 3' splice junctions are designated by a small 1 and 2, respectively.


Exon 1
Exon 2
1
pGU
AGp
2
Step 1: 5' splice site
cleavage + lariat formation
Exon 1
OH
G
p 1
A
2
p
Exon 2
Step 2: 3' splice site
cleavage + exon ligation
Exon 1
Exon 2
2
p
G
p 1
A
OH

Wallace and Edmonds (1983), containing 2'-5' and 3'-5' phosphodiester bonds joined to a single adenosine residue (Padgett *et al.*, 1984; Ruskin *et al.*, 1984; Konarska *et al.*, 1985). Hence, it was proposed that the intron-containing RNA products are in lariat form, with the 5' end of the intron joined by a 2'-5' phosphodiester bond to a specific adenosine near the 3' end of the intron (Padgett *et al.*, 1984; Ruskin *et al.*, 1984).

The kinetic and structural data led to the formulation of a two step splicing pathway (Padgett *et al.*, 1984; Ruskin *et al.*, 1984; Fig. 1.1). In the first step the substrate RNA is cleaved at the 5' splice site and the 5' end of the intron joined via a 2'-5' phosphodiester bond to an adenosine residue near the 3' splice site to produce a lariat molecule containing both the intron and 3' exon. The phosphorus in the 2'-5' phosphodiester link is from the 5' end of the intron and the 5' exon RNA species contains a 3' hydroxyl group as determined by nearest neighbour labelling and RNase T1 fingerprinting respectively (Konarska *et al.*, 1985). This reaction may be a direct transesterification, or the ligation reaction may constitute a kinetically rapid second reaction after 5' splice site cleavage. The second step of splicing involves cleavage at the 3' splice site. In a kinetically rapid second reaction or mechanistically coupled reaction the exons are ligated and the intron released in lariat form. RNase T1 fingerprinting of the spliced exons revealed that the phosphate in the 3'-5' phosphodiester bond joining the exons originated from the 5' terminal phosphate of the 3' exon (Padgett *et al.*, 1984) and the excised intron has a 3' hydroxyl group (Padgett *et al.*, 1984; Ruskin *et al.*, 1984).

The same reaction course has been demonstrated *in vitro* with *S. cerevisiae* splicing systems (Lin *et al.*, 1985) and also appears to be used *in vivo* since similar intermediates can be identified from mammalian (Zeitlin and Efstratiadis, 1984) and yeast pre-mRNAs (Rodriguez *et al.*, 1984; Domdey *et al.*, 1984) spliced *in vivo*.

1.2. PRE-MESSENGER RNA SEQUENCES INVOLVED IN SPLICING

Common sequence elements at exon/intron boundaries were identified not long after the discovery of split genes (Breathnach *et al.*, 1978; Breathnach and Chambon, 1981). The role of these conserved sequences and a third conserved region around the branch site (Langford and Gallwitz, 1983; Pikielny *et al.*, 1983; Keller and Noon, 1984) has been studied extensively both by comparing the sequences of a large number of different genes (Mount, 1982; Ohshima and Gotoh, 1987; Shapiro and Senapathy, 1987) and by studying the splicing of mutant genes.

1.2.1. The 5' splice site

In higher eukaryotes, computer sequence comparisons have revealed a consensus sequence around the 5' splice site (Mount, 1982; Padgett *et al.*, 1986; Ohshima and Gotoh, 1987; Shapiro and Senapathy, 1987). The consensus sequence for the seven GENBANK categories Primate, Rodent, Mammal, Vertebrate, Invertebrate, Plant and Viral is AG:GT($^{A}/_{G}$)AGT where the colon indicates the cleavage site (Shapiro and Senapathy, 1987). With the exception of position -3 relative to the site of 5' cleavage, which shows a bias towards A and C (Mount, 1982; Shapiro and Senapathy, 1987), sequences beyond this region appear random. The GT immediately 3' to the site of cleavage is almost completely conserved, however examples of GC at this position have also been reported (Dodgson and Engel, 1983; Erbil and Niessing, 1983; Wieringa *et al.*, 1983; King and Piatigorsky, 1983; Fischer *et al.*, 1984; Jongeneel *et al.*, 1986). With the exception of the virtually invariant GT the other consensus nucleotides are only moderately conserved.

The phylogenetic evidence for the importance of the 9 base consensus around the 5' splice site is supported by deletion analyses which showed that only the six 5' proximal intron nucleotides were required for normal *in vivo* splicing of rabbit β-globin pre-mRNA IVS-2 and human $^{G}\gamma$-globin pre-mRNA IVS-2 (Wieringa *et al.*, 1984; van Santen and Spritz, 1985). In addition, it was found that 5' and 3' splice sites from two different genes could be accurately spliced together (Chu and Sharp, 1981; Horowitz *et al.*, 1983) and synthetic 5' and 3' splice site consensus sequences are active *in vivo* (Rautmann *et al.*, 1984).

The study of defects in α and β-globin genes found in several forms of thalassaemia has revealed many examples of natural splicing mutants (for reviews see Treisman *et al.*, 1983a; Collins and Weissman, 1984). One class of mutations of normal 5' (or 3') splice sites reduces the yield of correctly spliced mRNA and frequently activates nearby cryptic (alternative) splice sites, causing production of aberrant mRNAs. Splice-site mutations have also been identified which cause exon-skipping, for example three spontaneous splice site mutations in dihydrofolate reductase (Mitchell *et al.*, 1986). New splice sites can also be created by a second class of mutations away from the wild type splice sites which improve the match of sequences to the splice site consensus, making them more competitive with the natural splice sites. The thalassaemia mutations, and other naturally occuring (Esumi *et al.*, 1983) and induced (Benyajati *et al.*, 1982) splice site mutants confirm the importance of the 5' splice site consensus sequence, the first class of mutations having been identified only in the 9 nucleotide consensus region.

However, the majority of splicing mutants analysed have been produced *in vitro*, and the effects of the mutations determined *in vivo*, *in vitro*, or both.

The most systematic study of the effect of point mutations on the 5' splice site consensus has been by Weissman and his group who have used as their model the rabbit β-globin IVS-2 5' splice site. All 16 possible combinations of dinucleotide at the site of the invariant GT of the IVS-2 5' splice site were made and their effect on splicing determined *in vitro* (Aebi *et al.*, 1987). None of the mutations abolished cleavage within the 5' splice region. Of the single point mutations of the GT sequence only the GT to GC transition yielded intermediates that could subsequently undergo step 2 of splicing, although the efficiency of step 2 was reduced even in this construct. Conversion to GA, GG or AT allowed accurate 5' cleavage with reduced efficiency, but the lariat intermediate was incapable of further processing. Hence, the second step of splicing appears to be dependent on the structure around the branchpoint, both the first and second nucleotides of the intron and the branch point nucleotide itself (Hornig *et al.*, 1986; see section 1.2.3) being important. Mutation of the GT to TT or CT allowed the formation of spliced products, but 5' cleavage was displaced upstream by one nucleotide. These two mutations generated a new GT or GC sequence one nucleotide upstream but also caused a one nucleotide shift in the optimal complementarity to the 5' end of U1 snRNA, which is thought to interact, at least in part, by base-pairing with the 5' splice site of pre-mRNA (Zhuang and Weiner, 1986; see section 1.3). If a second mutation is made in the GT to TT mutant two nucleotides upstream of the wild-type cleavage site, such as to restore better complementarity to U1 snRNA in the original base pairing frame, cleavage is shifted back to the wild-type site. Hence Aebi *et al.* (1987) proposed that the precise location of the 5' cleavage site is determined by most or all of the 5' consensus, rather than by the strictly conserved GPy. Conservation of the GPy sequence seems to be a result of the requirements of the second step of splicing. A more extensive series of 5' splice site mutants were construsted by Weber and Aebi (1988), including triple mutations. The location of the 5' cleavage site was found to be accurately predicted by calculation of the complementarity between the potential cleavage sites and U1 snRNA, adding further support to the hypothesis of Aebi *et al.* (1987).

In the above study the substrate for *in vitro* splicing did not contain IVS-1. If a complete rabbit β-globin gene transcript including exon 1 and IVS-1 was spliced *in vitro* somewhat different results were seen (Aebi *et al.*, 1986). Correct splicing was impaired by most mutations in the 5' consensus region of IVS-2 and the aberrant joining of exons 1 and 3 promoted. Mutating the conserved GT to AT or GA caused the strongest shift towards exon1 and 3 joining *in vitro*, followed by substitutions at +4, +5, +6, -3 and -1 relative to the 5' cleavage site. Mutations at positions -2 and +3 had no effect. In contrast to

the situation *in vitro*, mutations outside the conserved GT had no effect on the steady state levels of normally spliced mRNA *in vivo*. However, although GT to AT or GA mutations were cleaved at the normal 5' splice site, lariat intermediate accumulated *in vivo* indicating that branchpoint structure is important in the second step of splicing *in vivo* as well as *in vitro*.

The different results obtained when splice site mutants were processed *in vitro* and *in vivo* is of some interest. Many 5' splice site mutants showed no decrease in the level of spliced products *in vivo*, whereas effects on the efficiency of correct splicing were obvious *in vitro*. Pikielny and Rosbash (1985) had previously suggested a model in which although the rate of splicing is diminished, the level of spliced products *in vivo* may be altered very little, because pre-mRNA concentration builds up and drives the reaction. This situation would arise when splicing is not rate limiting in the production of mRNA. Splice site selection also differs: *in vivo*, the abnormal exon 1 to 3 "jump-splice" of the rabbit β-globin mutants was never seen with any 5' splice site mutants, despite this reaction being strongly promoted *in vitro* by the same mutations. Indeed, a low level of exon 1 to 3 "jump-splicing" was seen even with wild type rabbit β-globin pre-mRNA spliced *in vitro*. *In vivo*, in addition to causing accumulation of lariat intermediate, the GT to AT mutation resulted in the use of three cryptic 5' splice sites (Wieringa *et al.*, 1983; Aebi *et al.*, 1986). A similar effect was seen with the GT to GA mutation but only two of the above cryptic sites were used (Aebi *et al.*, 1986). To explain the difference between the *in vitro* and *in vivo* results Aebi *et al.* (1986) proposed the "first come first served" hypothesis: *in vivo* the nascent RNA strand can be rapidly bound by splicing factors and 5' and 3' splice sites are paired in a "committed complex" which undergoes splicing at some later time. Hence, a downstream 3' splice complex could be left to search for a cryptic 5' partner. *In vitro*, however, the IVS-2 3' splice site complex may be able to effectively compete for the IVS-1 5' splice site. Electron microscopic observations on the nascent transcripts of early *Drosophila* embryo genes suggest the presence of ribonucleoprotein particles located at splice junctions, and the coalescence of such particles to form what may represent committed complexes (Osheim *et al.*, 1985; Beyer and Osheim, 1988).

The functional importance of the conserved GT is supported by studies in other genes. A human β-globin gene was cloned from a foetus with β^{o} thalassaemia which had a GT to AT transition in the IVS-2 5' splice site (Treisman *et al.*, 1982). *In vivo* splicing was ineficient, the major processing event being use of a cryptic 5' splice site 47 nucleotides downstream of the wild type IVS-2 5' splice site. This result is generally consistent with the effect of the equivalent mutation on rabbit β-globin splicing *in vivo* (Wieringa *et al.*, 1983; Aebi *et al.*, 1986; see above). However, with the mutant human gene a low

level of "jump splicing" linking exons 1 and 3 was seen *in vivo* which, although seen *in vitro* with the mutant rabbit gene, was absent *in vivo*. In addition, a GT to GG transversion in the 5' splice site used for the production of the 12S RNA from the the adenovirus 2 EIA gene has been observed to cause complete switching of splicing from this site to the alternative 13S 5' splice site *in vivo* (Montell *et al.*, 1982). Mutation of the conserved GT of the human β-globin IVS-1 5' splice site to AT inactivated the splice site completely and led to the activation of three cryptic sites both *in vivo* (Treisman *et al.*, 1983b) and *in vitro* (Krainer *et al.*, 1984). The major splicing event observed *in vitro* with this mutant used a cryptic 5' splice site 16 nucleotides upstream of the normal IVS-1 splice junction. The major lariat intermediate (resulting from use of the cryptic 5' splice site above) was excised from a gel and the branch site found to be in the normal position (Ruskin *et al.*, 1984). However, RNA was not mapped from the region of the gel where lariat intermediate generated from use of the normal 5' splice site would be found, hence there was no indication as to whether there is a low level of "dead-end" lariat formation due to cleavage at the normal 5' splice site as found with rabbit β-globin IVS-2 *in vitro* and *in vivo* (Aebi *et al.*, 1987; see above). *In vitro* splicing of ^{32}P-labeled human β-globin transcripts, containing IVS-1 with the GT to AT mutation at the 5' splice site, did not, however, reveal any lariats produced from cleavage at the normal 5' splice site (Chabot and Steitz, 1987a).

Mutations at positions other than the conserved GT have variable effects, depending on the gene and splice site involved. This is, perhaps, only to be expected given the lack of sequence constraint at these positions. For example, although mutations outside the conserved GT of the rabbit β-globin 5' splice site have no effect on the steady state levels of normally spliced mRNA *in vivo* (Aebi *et al.*,1986; see above), mutating the the G 5 nucleotides downstream from the normal 5' cleavage site, or the T 6 nucleotides downstream from the normal 5' cleavage site to a C in human β-globin IVS1 reduces normal splicing and leads to activation of three cryptic sites *in vivo* (Treisman *et al.*, 1983b). *In vitro* the same three cryptic 5' splice sites were activated, but in contrast to *in vivo*, the total amount of spliced RNA observed was unaffected by the mutations (Krainer *et al.*, 1984). A double mutation of the bases 5 and 6 nucleotides downstream of the adenovirus 2 E1A 12S 5' splice site from GG to AT causes complete inactivation of splicing to this site (Solnick, 1981). Similarly, an A to T transversion 3 nucleotides downstream of the normal 13S 5' cleavage site of adenovirus 2 E1A gene will caused complete switching of splicing to the competing 12S site (Montell and Berk, 1984). The relatively severe effect of these latter two mutations may be atributable to the presence of a competing 5' splice site to which splicing may switch if a mutation has only a relatively minor negative effect on splicing efficiency at the other site.

5' splice junction sequences are more highly conserved in yeast than in metazoans, being in nearly all cases /GTA($^{C}/_{T}$)GT (Langford and Gallwitz, 1983; Teem *et al.*, 1984). In marked contrast to mammalian systems, mutations in the yeast 5' splice site consensus do not result in the use of cryptic splice sites (Gallwitz, 1982). Instead large amounts of unspliced RNA accumulate (Gallwitz, 1982; Pikielny *et al.*, 1983) or aberrant, non-productive 5' cleavage occurs (Jacquier *et al.*, 1985; Parker and Guthrie, 1985; Vijayraghavan *et al.*, 1986; Fouser and Friesen, 1986). As with metazoan systems, mutations at the first position of a yeast intron seem to be particularly deleterious for splicing. Substitution of an A for the first G of the yeast CYH2^{m} intron gives an RNA which both *in vivo* and *in vitro* blocks intron excision whilst still allowing 5' cleavage and lariat formation with reduced efficiency (Newman *et al.*, 1985). This lariat, which contains an A(2'-5')A phosphodiester bond, is a dead-end product, as is the case with similar metazoan structures. Fouser and Friesen (1986), using an actin-thymidine kinase fusion gene with a G to C mutation at position +1 or a U to G mutation at position +2 of the intron, found that not only was step 1 of splicing inhibited *in vivo*, but once again abortive lariat intermediates were formed. The mutation at position +1 had a greater effect than that at +2 further emphasising the particular importance of the first base of the intron. A mutation of G to A at position +1 of the intron of a yeast actin-HIS4 fusion gene has a similar *in vivo* effect to the same mutation in the CYH2^{m} intron whereas a G to C mutation at position +1 of the actin-HIS4 fusion gene results in failure to undergo any detectable 5' splice site cleavage and lariat formation *in vivo* and causes accumulation of precursor RNA (Vijayraghavan *et al.*, 1986). The situation *in vitro* is different; here, in contrast to the situation with the CYH2^{m} intron, mutations of positions +1 or +2 of the actin-HIS4 fusion gene intron to A or C respectively result in constructs that are incapable of performing step 1 of splicing *in vitro* (Vijayraghavan *et al.*, 1986).

Mutations at position +5 of yeast introns also seem to have a particularly dramatic effect on splicing. *In vivo*, a G to A transition at position +5 of the intron in an actin-HIS4 gene fusion still allows the 5' splice junction to be recognised, but with reduced efficiency (Parker and Guthrie, 1985; Vijayraghavan *et al.*, 1986). However, in addition to cleavage at the normal site, cleavage also occurs at an incorrect position 6 nucleotides upstream of the normal cleavage site, at the dinucleotide /AT (where the slash indicates the site of cleavage). The lariats formed are dead-end products. *In vitro*, this mutant is completely inactive and inhibits spliceosome assembly (Vijayraghavan *et al.*, 1986). Mutation of the G at intron position +5 of an actin-thymidine kinase fusion gene caused similar *in vivo* effects, the site being recognised with reduced efficiency and cleavage occuring at novel sites

to generate dead-end products, in addition to cleavage at the normal 5' junction (Fouser and Friesen,1986). The main aberrant cleavages were, however, at positions -5 and +5 relative to the wild type cleavage site in this construct. A G to A transition at intron position +5 of an rp51A fusion transcript also increases the steady state pre-mRNA level *in vivo* and causes a small accumulation of lariat intermediate, implying a reduction in the rate of both steps 1 and 2 of splicing (Jacquier *et al.*, 1985). A small amount of 5' cleavage occured 3 nucleotides upstream of the wild type 5' splice site at an /AT (where the slash indicates the site of cleavage). The lariats formed from cleavage at this site were inactive in step 2 of splicing, hence the effect is qualitatively similar to that seen with the yeast actin intron but the effect with the rp51A intron is not so strong.

Hence, mutations at position +5 of yeast introns seem to have pleiotropic effects, effecting an early an early step in splicing (as seen by the inhibition of spliceosome assembly (Vijayraghavan *et al.*, 1986)), the fidelity of 5' cleavage, and the efficiency of step 2 of splicing. It may be significant that, in mammalian 5' splice sites, position +5 is the most highly conserved base outside the GT at positions +1 and +2.

Mutations at other positions within the yeast 5' splice site consensus seem to have more modest consequences. For example a U to G mutation at position +6 relative to the wild type 5' cleavage site in an rp51A fusion gene transcript causes only a slightly reduced rate of step 1 of splicing *in vivo* (Jacquier *et al.*, 1985). No increase in lariat intermediate was observed.

1.2.2. The 3' splice site

Sequence comparisons of higher eukaryotic genes have also revealed a consensus sequence at the 3' splice site (Mount, 1982; Padgett *et al.*, 1986; Ohshima and Gotoh, 1987; Shapiro and Senapathy, 1987) which includes a pyrimidine rich region of variable length (the so called polypyrimidine tract), a nonconserved position, another pyrimidine, and an absolutely conserved AG dinucleotide. The consensus sequence established for the seven GENBANK categories Primate, Rodent, Mammal, Vertebrate, Invertebrate, Plant and Viral is $(^{T}/_{C})_2TT(^{T}/_{C})_6NCAG{:}G$ (Shapiro and Senapathy, 1987). The pyrimidine-rich region preceding the 3' splice site frequently extends further into the intron than the 14 bases indicated above. Both Mount (1982) and Shapiro and Senapathy (1987) noted the absence of other AG dinucleotides close to the 3' splice site. Shapiro and Senapathy (1987) found that less than 2% of sequences tested had an AG in the 10 nucleotides preceding the AG at the 3' splice site, and Mount (1982) suggested that this may imply a scanning

mechanism for recognition of the splice site by the splicing machinery.

Progressive deletions of the polypyrimidine tract of the rabbit β-globin gene IVS-2 (Weiringa *et al.*, 1984) or human $^{G}\gamma$-globin gene IVS-2 (van Santen and Spritz, 1985) caused progressive reduction of splicing efficiency *in vivo* and eventually a complete abolition of normal splicing. In the former case, splicing was shifted to a cryptic 3' site and the proportion of unspliced transcripts increased, whereas in the latter case cryptic 3' sites were not used to a great extent. These experiments show that the efficiency of splicing at the 3' splice site is determined by incremental contributions of the polypyrimidine tract. However, the precise length of the polypyrimidine tract required varies between genes, a fact reinforced by earlier deletions of the polypyrimidine tract of SV40 T-antigen RNA (Thimmapaya and Shenk, 1979; Volckaert *et al.*, 1979). Indeed, the role of the polypyrimidine tract upstream of the SV40 early region 3' splice site seems unusual compared to that in many other (non-alternatively spliced) RNAs studied. Its primary role seems not so much to enhance the efficiency of splicing, but rather to influence the selection of alternative 5' splice sites. Deletion of the polypyrimidine region, substitution of the pyrimidine-rich region with heterologous polypyrimidine tracts, and the insertion of purine rich fragments between the 3' AG and the polypyrimidine stretch had only minor effects on splicing efficiency, both *in vivo* and *in vitro* (Fu *et al.*, 1988). Splicing was significantly inhibited only when the polypyrimidine tract was replaced by a polypurine stretch, and even then inhibition was incomplete. However, mutations that increased the number of pyrimidine residues in the pyrimidine rich tract resulted in more efficient utilisation of the large T 5' splice site relative to the small t 5' splice site, while mutations that increased the purine content enhanced small t mRNA splicing. As with the 5' splice site, cryptic 3' splice sites can be activated *in vivo* by mutations within exons and introns that make the mutated sequence resemble a 3' splice site (Fukumaki *et al.*, 1982).

In vitro experiments revealed that, in human β-globin pre-mRNA, deletion of the UUAG sequence at the 3' end of IVS-1 (Ruskin and Green, 1985b) or an AG to GG or AC mutation in the same sequence (Reed and Maniatis, 1985; M. Green, cited in Reed and Maniatis, 1985) allowed 5' splice site cleavage and lariat formation to occur with reduced efficiency, but caused complete inhibition of exon ligation. Mutation the conserved AG at the 3' splice site of the rabbit β-globin IVS-2 to TG, CG, AT, or AA also abolished correct splicing both *in vivo* and *in vitro* (Aebi *et al.*, 1986), but in comparison with the same mutation in the human β-globin gene IVS-1, 5' splice site cleavage and lariat formation *in vitro* was more severely inhibited (Reed and Maniatis, 1985). *In vivo*, the mutations of the conserved AG of rabbit β-globin IVS-2 led to the use of a cryptic 3' splice site at the closest AG dinucleotide downstream of the

mutated 3' splice site. Aebi *et al.* (1986) also found the partially conserved C residue at the third position before the 3' splice site to be functionally significant, as mutation of this residue to an A (which occurs naturally with a frequency of only 4%) led to about 70% reduction in the rate of splicing as measured *in vitro*. Since lariat intermediate did not accumulate, the effect was probably at the level of step 1 of splicing. The same mutation did not diminish the level of correctly spliced RNA *in vivo*. In contrast, deletion of most of the polypyrimidine tract of human β-globin, while leaving the 3' UAG intact, drastically reduces the efficiency of 5' splice site cleavage (Ruskin and Green, 1985b). Therefore, although both elements of the 3' consensus are involved at an early stage in splicing, the polypyrimidine tract appears particularly important for 5' splice site cleavage and lariat formation.

In yeast, the 3' consensus is essentially similar to that in higher eukaryotes, but the polypyrimidine tract tends to be much less extensive (Langford *et al.*, 1984). The role of the sequences at the 3' splice site seems to differ somewhat in yeast and mammals. Truncated yeast rp51A transcripts lacking the 3' splice site sequences, but still containing the TACTAAC box at the branch point (see section 1.2.3) when processed *in vitro*, could still undergo both cleavage at the 5' splice site and lariat formation (Rymond and Rosbash, 1985). Furthermore, deletions of the yeast actin intron 3' splice site sequence were found to activate cryptic 3' splice sites, predominantly at the first AG downstream of the TACTAAC box *in vivo*, provided it is not closer than about 10 nucleotides (Langford and Gallwitz, 1983). In contrast to metazoan systems, where cryptic 3' splice sites consist of an AG preceded by a pyrimidine rich tract, and therefore closely resemble normal splice sites (Fukumaki *et al.*, 1982; Weiringa *et al.*, 1984), the nucleotides upstream of cryptic AG dinucleotides in yeast are more variable (Langford and Gallwitz, 1983). The polypyrimidine tract, therefore, appears to play a much less important role in yeast splicing than in metazoan systems. Mutations of the yeast actin 3' splice site AG:AG sequence (where the colon indicates the site of cleavage) to AC:AC appears, *in vivo*, to have no effect on the first step of splicing yet decreases the efficiency of the second step. What little 3' cleavage there is occurs at AC:AC. *In vitro* the result is as *in vivo* except in this case step two is completely inhibited (Vijayraghavan *et al.*, 1986). Splicing complex formation on this mutant substrate is apparently normal. The results with this mutant *in vitro* are, therefore, similar to those obtained with the equivalent mutation in human β-globin IVS-1 (Reed and Maniatis, 1985) except the yeast mutation has less effect on step one.

1.2.3. The Branchpoint

No highly conserved sequence elements can be identified within metazoan introns, despite the fact that specific interactions with intronic sequences must occur in the course of lariat formation. Several *in vivo* studies have established that the majority of sequences within the intron can be deleted or substituted without abolishing splicing (Khoury *et al.*, 1979; Thimmapaya and Shenk, 1979; Volckaert *et al.*, 1979; Wieringa *et al.*, 1984; van Santen and Spritz, 1985). *In vitro* studies soon revealed that splicing could proceed in human β-globin IVS-1 substrates in which the branch site region had been deleted, or the branch point A mutated to a G, by activation of cryptic branch points (Ruskin *et al.*, 1985). The sequences around mapped authentic and cryptic branch points (see Green, 1986 and references therein) were found not to be strongly conserved, indicating that branch formation does not have a strict sequence requirement. However, the distance upstream of the 3' splice site appears to be significant, nearly all mapped branch points being located 18 to 37 nucleotides upstream of the 3' splice site. One notable exception is the *in vitro* excision of a 216 nucleotide intron, delineated by the 9S 5' splice site and a 3' splice site 216 nucleotides downstream, from adenovirus E1A transcripts, which uses any one of 3 branch points located 51, 55 or 59 residues upstream of the 3' splice site (Gattoni *et al.*, 1988). Although there is no absolute sequence constraint, branch point formation does show a sequence preference. A consensus sequence $(^{C}/_{U})N(^{C}/_{U})U(^{A}/_{G})\underline{A}(^{C}/_{U})$ was derived (Ruskin *et al.*, 1984; Zeitlin and Efstratiadis, 1984). The nucleotide which forms the branch is underlined. Derivation was by a comparison of sequences around mapped branch points, and it is particularly noticeable that in all natural branch points mapped to date, except one (a C-branched lariat in IVS-1 of human growth hormone pre-mRNA; K. Hartmuth and A. Barta, cited in Hornig *et al.*, 1986), the nucleoside which forms the branch is adenosine. This correlates well with the data of Wallace and Edmonds (1983) who found 80% of the branched nucleosides in polyadenylated nuclear RNA to be adenosine. This work also indicated that the nucleotide immediately downstream of the branched residue was almost always a pyrimidine. Before branching was discovered Keller and Noon (1984) derived a consensus by computer analysis which was located near the 3' splice site of metazoan introns. This consensus, based upon its homology with the yeast TACTAAC box (see below), is $CU(^{A}/_{G})A(^{C}/_{U})$, and many mapped branch points have subsequently been found to conform well to this sequence.

The functional importance of the more upstream bases in the branch point consensus is reinforced by the observation that a 40 nucleotide deletion of human β-globin IVS-1, upstream of the branch point, which results in alteration of the bases 4 and 5 nucleotides upstream of the branch nucleotide,

causes both a decrease in splicing efficiency and use of a cryptic branch point as well as the normal branch point (Ruskin *et al.*, 1985). Switching lariat formation to different branch points can alter splicing efficiency both *in vitro* (Padgett *et al.*, 1985; Ruskin *et al.*, 1985) and *in vivo* (Rautmann and Breathnach, 1985) by a variable amount. In certain cases the cryptic branch points used do not conform to the consensus sequence (Padgett *et al.*, 1985; Ruskin *et al.*, 1985). Despite the fact that deletion or mutation of the mammalian conserved 3' AG sequence reduces the efficiency of 5' splice site cleavage and lariat formation (Ruskin and Green, 1985b; Reed and Maniatis, 1985; Aebi *et al.*, 1986; see section 1.2.2) the authentic branch point is still used when the UUAG sequence at the 3' end of human β-globin IVS-1 is deleted (Ruskin and Green, 1985b). However, on the basis of sequence comparison, Ohshima and Gotoh (1987) suggest that the presence of a sequence that could serve as a branch site, 18 to 37 nucleotides upstream of the 3' splice site, does not seem to be critical in distinguishing it from a 3' splice site like sequence. Hornig *et al.* (1986) converted the branch point A residue of rabbit β-globin IVS-2 to either G, C or T and spliced these constructs *in vitro*. All three mutations reduced the efficiency of IVS-2 splicing by approximately 50%. Branch formation occured at the normal branch point location with all four constructs, albeit with different efficiencies, namely A>C>G=U. In the case of G and U the majority of lariat formation was shifted to the A one nucleotide upstream of the normal branch point, presumably due to the natural preference of the system for an A residue. Only branches to A or C participated efficiently in the second step of splicing. These observations fit those of Wallace and Edmonds (1983), who found the predominant branched nucleotide in polyadenylated nuclear RNA to be an A residue. More detailed analysis (J.C. Wallace and M. Edmonds, cited in Hornig *et al.*, 1986) revealed C residues to be the next most frequently found. A naturally occuring C-branched lariat has been reported in IVS-1 of human growth hormone pre-mRNA (K. Hartmuth and A. Barta, cited in Hornig *et al.*, 1986).

Yeast branch point sequences differ from their mammalian counterparts in being much better conserved. Indeed, deletion mutagenesis of the yeast actin gene intron (Langford and Gallwitz, 1983) and rp51A gene intron (Pikielny *et al.*, 1983) revealed an internal intronic region essential for yeast splicing. Within this essential region is a conserved sequence element 5'-UACUAAC-3' (the "TACTAAC box" or "ICS") which occurs 20 to 55 nucleotides upstream of the 3' splice site in all sequenced yeast introns. The underlined A of the UACUA<u>A</u>C sequence was subsequently identified as the site of branching on lariat formation both *in vivo* (Domdey *et al.*, 1984; Rodriguez *et al.*, 1984) and *in vitro* (Lin *et al.*, 1985). In contrast to the situation in higher eukaryotes, cryptic branch points are not usually activated by mutations of the yeast

TACTAAC box. Deletion of the TACTAAC box abolishes *in vivo* splicing (Langford and Gallwitz, 1983; Pikielny *et al.*, 1983). Splicing of the yeast actin intron can also be abolished *in vivo* by point mutations converting UACUAAC to UCCUAAC or UACUACC (Langford *et al.*, 1984). The latter mutation, at the normal branch position, made in the yeast $CYH2^m$ gene intron also renders the pre-mRNA inert, both *in vivo* and *in vitro* (Newman *et al.*, 1985). In the yeast rp51A gene intron, this same mutation almost eliminates *in vivo* splicing, but a low level (about 0.1%) of splicing is detectable, and branching is to the normal location i.e. to a C. However, the former mutation, when made in the yeast β-galactosidase gene intron (along with another mutation, to UAUUAAC) results only in a modest decrease in the rate of step 1 *in vivo* (Jacquier *et al.*, 1985). Step 2 is unaffected. A mutation to UAUUAAC in the yeast actin gene intron also had no detectable influence on the efficiency or accuracy of splicing *in vivo* (Langford *et al.*, 1984). Fouser and Friesen (1986) mutated all seven positions of the TACTAAC box in the yeast actin gene intron and tested splicing *in vivo*. They concluded that all seven positions, but especially the last five, were important for the efficiency of step 1. In contrast to Newman *et al.* (1985), who mutated the $CYH2^m$ gene intron, the UACUACC mutation in the actin gene intron did not completely prevent splicing and this difference could be due to the different strains of yeast or different substrates used. Transversions to GACUAAC and UCCUAAC caused only minor accumulation of primary transcript. Lariat formation could still occur to C or G at the branch site, branching at G being the more efficient, however step 2 was almost completely abolished. Since some spliced transcript was found with C or G as the branched residue, an branched A is not absolutely essential for step 2.

Vijayraghavan *et al.* (1986) analysed point mutations in the yeast actin gene intron TACTAAC box and found that mutations to UAAUAAC, UACAAAC or UACUACC all caused a severe inhibition of step 1 both *in vivo* and *in vitro*. However, the branch site mutation still formed lariats to the normal branch site, i.e. to C, both *in vivo* and *in vitro*. Interestingly, mutating UACUAAC to UAUUAAC had little effect *in vivo*.

There is a single example of a cryptic branch point being used in yeast. This is at the last A of a UAGUAAG sequence upstream of the normal branch site (Cellini *et al.*, 1986). However, this is very unusual. Mutating the normal branch point consensus usually does not lead to cryptic branch points being used, even when nearby sequences match the UACUAAC consensus quite closely (Jacquier *et al.*, 1985; Fouser and Friesen, 1986; Jacquier and Rosbash, 1986).

These results would seem to imply that the sequences surrounding the branch point are responsible for determining the site of branch formation rather than the branch site A itself; that the role of identical nucleotides may

vary in different genes; and that nucleotides are often involved in more than one step of splicing, the specific requirements for the chemistry of these individual reactions sometimes varying considerably.

The role of the branch point consensus in yeast seems to differ from that of mammals. It has been shown (Langford and Gallwitz, 1983; Langford *et al.*, 1984) that inserting a TACTAAC box into a yeast construct can direct utilisation of a new 3' splice site at an AG directly downstream in yeast. The first AG downstream is preferred.

The difference between mammalian and yeast systems is clearly seen by the fact that metazoan pre-mRNAs lacking a perfect TACTAAC box are not spliced in yeast cells (Langford *et al.*, 1983) and by the fact that transcripts derived from the yeast rp51A gene can be accurately spliced in HeLa cell nuclear extract, but the branch is not in the TACTAAC box but at an A residue closer to the 3' splice site (Ruskin *et al.*, 1986). Deletion of the TACTAAC box did not prevent splicing in the mammalian system.

A degree of sequence complementarity has been identified between the branch point sequence and the 5' splice site sequence, and a base pairing interaction between these two sites has been suggested to be of functional importance in both yeast (Pikielny *et al.*, 1983) and mammalian systems (Konarska *et al.*, 1985). Attractive though this hypothesis is, for aligning the 5' splice site and branch point, there is no experimental support. It has not proved possible to rescue mutations in the branch sequence of yeast with compensating mutations in the 5' splice site sequence (Jacquier *et al.*, 1985; Fouser and Friesen, 1986). The fact that cryptic branch site sequences in metazoan introns frequently show no complementarity to the 5' splice site (Padgett *et al.*, 1985; Ruskin *et al.*, 1985) does not support the intramolecular base pairing hypothesis. A mammalian cis-competition system, consisting of rabbit β-globin gene construct with competing artificial 5' splice sites 25 nucleotides upstream of the wild type IVS-2 5' splice site (Eperon *et al.*, 1986) has been used to investigate this hypothesis in metazoans (S.A. Cunningham, personal communication). Complementarity between the branch point and several artificial (upstream) 5' splice sites was selectively increased by mutation of the branch point region but no alteration in the ratio of splicing to the two 5' splice sites was observed *in vivo* or *in vitro*. However, intramolecular base pairing between these two regions cannot be discounted on the basis of these results, since 5' splice site selection may occur at a step in splicing prior to base pairing between the 5' splice site and branch point.

1.2.4. Additional aspects of splice site selection

It is an interesting, and as yet unresolved problem as to why sequences within exons and introns that conform to the splice site consensus are normally inactive. It would seem that the consensus 5' and 3' splice site sequences, although essential, are not the sole determinants of splice site selection. Several ideas have been proposed to explain the selection of authentic splice sites in preferences to sequences that resemble splice junctions within exons and introns. First, several studies suggest that U1 snRNA is a major determinant of 5' splice site selection. There is direct genetic evidence that the 5' end of U1 snRNA interacts, at least in part, by base pairing with the 5' splice site (Zhuang and Weiner, 1986). These experiments, discussed in section 1.3, strongly support a role for intermolecular base pairing with U1 snRNA in the selection of the alternative 12S and 13S 5' splice sites in adenovirus E1A RNA. The computer study of Ohshima and Gotoh (1987) revealed that in 79% of cases in the 13 genes studied, the normal 5' splice site was more complementary to U1 than any other sequence in the adjacent intron or exon. Aebi *et al.* (1987) also found that a cryptic 5' splice site could be activated by decreasing the complementarity of the normal rabbit β-globin IVS-2 5' splice site to U1, however, this is not in itself very strong evidence for the base pairing model.

Secondly, so called "scanning models" have been proposed (Sharp, 1981; Lang and Spritz, 1983) in which a splicing factor (or factors) bind to a specific mRNA region and can scan the RNA unidirectionally (moving either upstream or downstream) until a splice site (5' or 3') is reached. There is, however, considerable evidence that a simple scanning mechanism does not operate. Cryptic sites are frequently activated both in intronic and exonic regions by mutation of the authentic splice site (see above). There are also many examples of pre-mRNA sequences which show stronger potential complementarity to U1 snRNA than does a nearby functional 5' splice, and yet these sequences are not used as splice sites (Ohshima and Gotoh, 1987; Shapiro and Senapathy, 1987). Hence, any scanning component does not simply seem to recognise the first potential splice site it reaches. Also, *in vitro trans* splicing between two transcripts with mutually complementary sequences inserted into their introns has been demonstrated (Solnick, 1985a; Konarska *et al.*, 1985b), where there is a discontinuity of phosphodiester bonds between the 5' and 3' splice sites. A low level of *trans* splicing occurs even between RNA fragments with no complementarity (Konarska *et al.*, 1985) in which case it cannot even be argued that a scanning component might be able to switch backbone in regions of secondary structure. There is, however, some debate as to whether *trans* splicing in such cases is actually due to transient base pairing (Solnick, 1986; Sharp and Konarska, 1986). Further evidence against a simple scanning model comes from the fact that distal splice sites were used *in*

vivo in some precursors containing tandem duplications of the 5' splice site (Kuhne *et al.*, 1983; Lang and Spritz, 1983). Distal 3' splice sites were also found to be used *in vivo* in some precursors containing tandem 3' splice site duplications (Kuhne *et al.*, 1983), whereas proximal 3' splice sites were used in others (Lang and Spritz, 1983). Similar discrepancies were found by Reed and Maniatis (1986). It has also been suggested that functional 5' and 3' splice sites closest to one another may be preferred *in vitro* (Reed and Maniatis, 1986), but even if this is true *in vivo*, it certainly does not account for many patterns of splice site selection observed (e.g. Kuhne *et al.*, 1983) unless other factors are taken into consideration (Reed and Maniatis, 1986).

The sequestration of splice sites in regions of secondary structure has also been proposed to influence splice site selection *in vivo* (Burnett, 1982; Breitbart and Nadal-Ginard, 1986; Ruiz-Opazo and Nadal-Ginard, 1987; Leff *et al.*, 1987). However, these proposed structures may have no significance *in vivo*, where experiments indicate that, in order to affect splice site selection, either extensive regions of sequence complementarity were needed, or loops on hairpins would have to be small (Solnick, 1985b; Solnick and Lee, 1987; Eperon *et al.*, 1988).

By using a sensitive cis-competition assay for splice site selection an important role for the sequences within which the splice site consensus itself resides has been found. This regional effect has been termed "splice site context" by Nelson and Green (1988) who constructed a series of splice site insertion mutants by inserting synthetic 5' and 3' splice sites, comprising only the conserved splice site bases, into various positions within the introns and exons of the rabbit and human β-globin genes. *In vitro* splicing of these constructs showed the synthetic 5' splice sites to be variably active in a position dependent manner. 5' splice site selection is, therefore, not solely based on match to the consensus (i.e complementarity to U1 snRNA). The authentic 5' splice site had an advantage in selection except when the synthetic site was nearby, presumably in the same favourable context. Even in the absence of the competing authentic site, the rabbit β-globin IVS-2 5' splice site was only used when near the position of the deleted authentic site. It may, therefore, be significant that mutations that inactivate 5' splice sites often lead to activation of cryptic sites near to the authentic site (Treisman *et al.*, 1983a; see section 1.2.1).

The sequence of the nine conserved bases around the 5' splice junction does, however, play an important role in splice site selection both *in vitro* (Nelson and Green, 1988) and *in vivo* (Eperon *et al.*, 1986).

The region around the authentic 3' splice site of human β-globin IVS-1 also appears to confer an advantage for 3' splice site selection. If the normal 3' splice site consensus sequence is deleted (both polypyrimidine tract and

conserved AG) and a synthetic 3' splice site inserted, splicing efficiency decreases rapidly when the site is moved a mere 12 nucleotides further downstream from one site to the next (Nelson and Green, 1988). Although Rautmann *et al.* (1984) reported a synthetic 3' splice site to be active *in vivo* in several positions, in this study a large drop in efficiency could have gone unnoticed.

There is a body of additional evidence supporting an important role for context in splice site selection. Of particular interest is the work of Reed and Maniatis (1986) who showed that large deletions or substitutions of exon sequences usually inactivate the adjacent 5' or 3' splice sites of human β-globin IVS-1 when precursors containing tandemly duplicated splice sites are spliced *in vitro*. Furthermore, even small differences in exon sequence can significantly shange splice site selection. A decrease in usage of the adjacent site was observed when normal exon is replaced by the highly homologous mouse or rabbit exon.

Mutations within exons can also affect alternative splicing of SV40 pre-mRNA *in vivo* (Somasekhar and Mertz, 1985). Intron sequences too, would seem to be significant: Kuhne *et al.* (1983) found that a duplicated rabbit β-globin gene IVS-2 5' splice junction missing sequences from six nucleotides downstream of the cleavage site is less efficiently used in competition with the downstream wild type site than a 5' junction containing 100 nucleotides of wild type sequence immediately downstream of the splice site. The nature of the sequences within 5' splice site flanking regions that influence splice site usage is, however, unknown. Any primary sequence element would have to be highly degenerate or dispersed since several computer searches have failed to identify them (Mount, 1982; Ohshima and Gotoh, 1987; Shapiro and Senapathy, 1987). Context effects could perhaps also result from local secondary or tertiary structures that facilitate efficient splice site utilisation.

Yeast mRNA precursors generally contain only one intron, and so selection of splice sites and branch points is a less complex task than in metazoans. Nevertheless, the conserved sequences at the splice junctions and branch point in yeast still do not seem adequate to direct accurate splicing in isolation. Consensus splice sequences occur in the coding regions of mRNAs such as the HIS4 gene (Donahue *et al.*, 1982) yet are not used for splicing, and duplications of the 5' splice introduced upstream and downstream of the authentic site in the yeast $CYH2^m$ gene are silent, even when the authentic site is deleted (Newman, 1987). Newman (1987), by making small deletions and rearrangements in the yeast $CYH2^m$ gene intron and characterising the effects on splicing, both *in vivo* and *in vitro*, has identified two discrete regions of the intron, distinct from the consensus sequences at the branch point and splice sites, which perform critical roles in splicing. These two

sequence elements appear to correspond to two elements of partial sequence complementarity lying downstream of the 5' splice site and upstream of the 3' splice site respectively. Potential helix forming sequences are usually found at these positions in yeast introns (Parker and Patterson, 1987). Deletion of the 3' complementary sequence abolishes splicing *in vivo* and *in vitro*, and prevents spliceosome assembly. However, splicing can still proceed if the 5' complementary sequence is also deleted. As deletion of the 5' sequence alone does not prevent splicing, a role for base pairing in the alignment or juxtaposition of spliceosome components interacting with the 5' splice site and branch point (Parker and Patterson, 1987) seems to be ruled out. The data are consistent with a model in which a feature in the vicinity of the 5' complementary element inhibits spliceosome assembly in an unknown way. The inhibition can be relieved by deleting the 5' element or providing a complementary 3' element with which it can base-pair. Mixing and matching the elements from the yeast ACT1 and $CYH2^{m}$ gene introns showed that these elements can cooperate in an intron-specific fashion and argue in favour of a sequence specific interastion between the two elements. The fact that these two regions of sequence complementarity are a conserved feature in yeast introns (Parker and Patterson, 1987) may suggest a common function in yeast splicing.

1.3. THE ROLE OF SMALL NUCLEAR RIBONUCLEOPROTEIN PARTICLES IN PRE-MESSENGER RNA SPLICING

Small nuclear ribonucleoproteins (snRNPs) are found in the nuclei of all eukaryotic cells. In mammalian nuclei there are six major snRNP species (U1-U6 snRNPs). U3 snRNP is nucleolus specific, whereas the others are nucleoplasmic and present at an abundance of 2×10^5 to 10^6 copies per cell. Additional, less abundant mammalian snRNPs have been discovered more recently, designated U7 (Galli *et al.*, 1983), U8-U10 (Reddy *et al.*, 1985) and U11 (Krämer, 1987a). Four novel U RNAs have also been found to be encoded by a Herpesvirus (Lee *et al.*, 1988), two of which are probably involved in polyadenylation as they have 5' end sequences perfectly complementary to the highly conserved AAUAAA polyadenylation signal. Additional snRNAs may yet be awaiting discovery. U6 snRNA possesses an unidentified 5' cap structure, and all the other identified snRNAs have an unusual 2-N,N-dimethyl-7-methylguanosine 5' terminal cap structure.

The snRNAs exist within the cell as complex ribonucleoprotein particles. In addition to the snRNA components, the major snRNPs contain at least 10 polypeptides. Biochemical fractionation and immunological experiments,

making use of the fact that some patients with autoimmune diseases, such as systemic lupus erythematosus (SLE), carry antibodies that react with the proteins of snRNPs (Lerner and Steitz, 1979), have enabled assignment of the snRNP specific polypeptides to the various specific polypeptides to the various snRNPs. The proteins B, B', D, D', E, F and G are present in all major snRNPs (Hinterberger *et al.*, 1983; Kinlaw *et al.*, 1983, Bringmann and Lührmann, 1986). Harris *et al.* (1988), after performing chemical cross-linking to study the organisation of proteins within U snRNPs, proposed that part of the U snRNP core of common proteins contained at least two asymmetrical copies of B:B':D:D':E:G with stoichiometries of 2:1:1:1:1:1 and 1:2:1:1:1:1. The U1 snRNP contains the unique determinants A, C and 68K (Fisher *et al.*, 1983; Fisher *et al.*, 1984; Pettersson *et al.*, 1984, Bringmann and Lührmann, 1986). The U2 snRNP uniquely contains A' and B" (Kinlaw *et al.*, 1983; Mimori *et al.*, 1984; Bringmann and Lührmann, 1986). U4 and U6 snRNAs are generally found associated in a complex in which U4 and U6 RNAs interact by intermolecular base-pairing (Bringmann *et al.*, 1984; Hashimoto and Steitz, 1984; Rinke *et al.*, 1985) Neither the U4/U6 snRNP nor the U5 snRNP has been found to contain any unique polypeptides (Bringmann and Lührmann, 1986).

Four classes of snRNP specific antibodies have been reported. The anti-(Sm) antibody reacts with U1, U2, U4, U5 and U6, along with many other snRNPs, including yeast (but not mammalian) U3: protein blotting indicated that the antibody recognises the B, B', D, E, F and G polypeptides (Mimori *et al.*, 1984) common to these snRNPs. Anti-(U1)RNP antibody (originally called anti-(RNP)) reacts only with the U1 snRNP and recognises the A, C and 68K polypeptides (Fisher *et al.*, 1983; Fisher *et al.*, 1984; Pettersson *et al.*, 1984). The anti-(U2)RNP reacts only with the U2 snRNP recognising the A' protein and possibly the B" protein (Mimori *et al.*, 1984). Antibodies have also been raised against the 2,2,7-trimethylguanosine (TMG) cap structure (Luhrmann *et al.*, 1982).

Until recently, functional homologues of mammalian snRNPs had not been identified in yeast. 24 *S. cerevisiae* snRNAs possessing the 5' TMG cap structure were, however, identified as a result of their immunoprecipitability by the anti-(TMG) antibody (Riedel *et al.*, 1986). Furthermore, a subset of yeast snRNAs contain potential Sm antigen-binding sites in their single stranded regions with a consensus sequence of $A(U_{3-6})G$ (Wise *et al.*, 1983; Ares, 1986; Guthrie *et al.*, 1986; Riedel *et al.*, 1986; Riedel *et al.*, 1987; Patterson and Guthrie, 1987, Siliciano *et al.*, 1987a; Siliciano *et al.*, 1987b; Kretzner *et al.*, 1987). This sequence motif is the binding site for the four proteins that comprise the "Sm" antigenic determinant (Mattaj and DeRobertis, 1985). Indeed, some of the human anti-(Sm) antisera can specifically immunoprecipitate snR7(L and S), snR14, snR19, and snR20 (the yeast homologues of mammalian U5, U4, U1 and

U2 RNAs respectively) directly from splicing extracts (Siliciano *et al.*, 1987b). These RNAs do, however, differ in several respects from their metazoan counterparts. With the exception of snR17 (yeast U3), where there are two copies, yeast snRNAs are encoded by single copy genes (Tollervey *et al.*, 1983; Tollervey and Guthrie, 1985). They are present at only 10-500 transcripts per cell (Wise *et al.*, 1983) and range in size from 120 up to 1000 nucleotides (Riedel *et al.*, 1986). Gene disruption techniques, however, have revealed that many of the yeast snRNA genes are not essential for growth (Tollervey *et al.*, 1983; Tollervey and Guthrie, 1985; Guthrie *et al.*, 1986), indicating them to be functionally redundant, or not required for splicing. Recent work has, however, identified the yeast homologues of U1, U2, U4, U5 and U6 snRNAs. These are summarised in Table 1.1. For most, the identity was established by clear homology with portions of the mammalian counterpart.

The first speculation that snRNPs might be involved in splicing was based on the observed complementarity between the 5' end of U1 snRNA and pre-mRNA splice junctions (Lerner *et al.*, 1980; Rogers and Wall, 1980). Anti-(snRNP) antibodies were subsequently shown to inhibit splicing, first in isolated nuclei (Yang *et al.*, 1981), and later in splicing extracts (Padgett *et al.*, 1983b; Krämer *et al.*, 1984). However, addition of partially purified snRNPs to antibody depleted extracts did not restore splicing activity (Krämer *et al.*, 1984) leaving these results open to question. Anti-(U2) antisera, in contrast to anti-(U1) antibody, failed to inhibit splicing (Padgett *et al.*, 1983b).

The involvement of an essential RNA component was, however, more strongly implicated by experiments in which extracts treated with micrococcal nuclease (which was subsequently inactivated by the addition of EGTA) were found to be completely inactive in RNA splicing, yet activity could be restored by the addition of partially purified snRNPs which were themselves devoid of splicing activity (Krainer and Maniatis, 1985). By hybridising DNA oligonucleotides to specific sequences in snRNAs and then using RNase H to degrade the RNA strand of the RNA-DNA duplex, specific snRNAs could be inactivated (Krämer *et al.*, 1984). This technique has shown that cleavage of the 5' end of U1 snRNA inhibits splicing (Krämer *et al.*, 1984; Krainer and Maniatis, 1985; Black *et al.*, 1985) and has also demonstrated a need for intact U2 (Krainer and Maniatis, 1985; Black *et al.*, 1985), U4 (Berget and Robberson, 1986; Black and Steitz, 1986) and U6 (Black and Steitz, 1986) snRNAs in the *in vitro* splicing reaction. Complementation experiments revealed that splicing activity could be recovered by mixing extracts in which different snRNAs were inactivated (Krainer and Maniatis, 1985; Black *et al.*, 1985; Berget and Robberson, 1986; Black and Steitz, 1986).

As discussed in section 1.4.5, purified U1 snRNP has been shown to bind specifically to 5' splice site sequences in RNA (Mount *et al.*, 1983; Tatei *et al.*,

Table 1.1. **Comparison of mammalian and yeast small nuclear RNAs involved in pre-mRNA splicing.**

Mammalian		Yeast		
designation	size	counterpart	size	References
U1	164	snR19	569	Kretzner *et al.* (1987) Siliciano *et al.* (1987b)
U2	188	snR20	1175	Ares (1986)
U4	145	snR14	160	Siliciano *et al.* (1987a)
U5	115	snR7	179	Patterson and Guthrie (1987)
U6	107	snR6	112	Brow and Guthrie (cited in Cheng and Abelson, 1987)

1987). RNase protection and RNase T1 protection/immunoprecipitation experiments have also shown that U1 snRNP binds to the 5' splice site and that U2 snRNP binds to sequences around the branch point in the course of splicing reactions (Black *et al.*, 1985; Chabot and Steitz, 1987; see section 1.4.5). An additional snRNP, possibly U5 interacts with the 3' splice site (Chabot *et al.*, 1985; see section 1.4.5). A variety of methods of fractionation have also been used to show that the mammalian spliceosome contains U2, U5, U4 and U6 snRNAs, and also possibly U1 snRNA. However, it has been proposed that the U1 snRNP may only associate transiently with the substrate RNA and dissociate before spliceosome formation (Konarska and Sharp, 1986). Similarly, yeast spliceosomes have been shown to contain snR6 (yeast U6), snR7 (yeast U5), snR14 (yeast U4) and snR20 (yeast U2) (Pikielny *et al.*, 1986).

That recognition of the 5' splice site of mammalian introns occurs, at least in part, by complementary base pairing with the 5' end of U1 snRNA has been demonstrated by a direct genetic test (Zhuang and Weiner, 1986). The experiments were designed to test whether point mutations in the alternative 12S and 13S 5' splice sites of the adenovirus E1A gene, could be suppressed by compensatory base changes in the 5' end of human U1 snRNA. Expression vectors containing either the wild-type or mutant E1A genes were used. Efficient suppression of a mutation at position +5 of the 12S splice site was observed. A cryptic 5' splice site was also found to be activated by a mutant U1 snRNA with one more base pair of potential complementarity to this site than has wild-type U1 snRNA. A mutant U1 snRNA with increased complementarity to the wild type 12S 5' splice site was observed to bias alternative splicing towards the 12S mRNA product. However, although base pairing between the the 5' splice site and U1 snRNA is necessary for splicing, this interaction alone is not sufficient since only exceedingly weak suppression of a mutation at position +3 of the 13S splice site was observed.

A base pairing interaction between yeast 5' splice sites and the 5' end of snR19 (yeast U1) RNA has also been established genetically (Séraphin *et al.*, 1988). A haploid yeast strain was constructed containing a chromosomal copy of the U1 RNA gene which had been disrupted by the scheme of Rothstein (1983) and which was complemented by a wild type copy of the gene, tagged with a URA3 marker on a centromeric plasmid. Mutant U1 RNA genes could then be introduced on a plasmid containing a TRP1 marker and the strains grown in the presence of 5-fluoro-orotic acid which selects against cells containing a functional URA3 gene. Hence, cells containing mutant U1 snRNAs that could provide a minimal U1 snRNP function could be obtained. All three possible changes at position 4 (excluding cap) of the U1 snRNA were tested. This position is expected to pair with position +5 of the intron (Kretzner *et al.*, 1987; Siliciano *et al.*, 1987b). No growth defects were observed when both

mutant and wild type U1 snRNAs were present, indicating that growth defects of the position 4 mutants are recessive.

When U1 RNA alone was present, only the mutant with a C to U change at position 4 could support growth, although at a reduced rate, consistent with the ability of this mutant to form a U-G base pair instead of the proposed wild type C-G base pair. The splicing of a yeast RP51A gene intron containing a G-A mutation at position +5 of the intron (the 5'II mutant) has previously been analysed *in vitro* (Jacquier *et al.*, 1985) and found to reduce splicing efficiency (see section 1.2.1). Pre-mRNA containing this mutation was more efficiently spliced in a strain containing U1 RNA with a C to U change at position 4 than in a strain with only wild type U1 snRNA. A mutant U1 with a C to G change at position 4 could also partially suppress the complementary 5' splice site mutation at position +5 of the intron, although a mutation of position 4 of U1 from C to A could not suppress a complementary mutation in the substrate. Mutations at positions 3 or 5 of U1 snRNA also did not affect the splicing of substrate RNA with a G to A mutation at position +5 of the intron.

With the 5'II mutant, a low level of aberrant cleavage was also detected at a site 3 nucleotides upstream of the normal 5' splice site (Jacquier *et al.*, 1985) and this cleavage led to a frozen lariat intermediate. In a yeast strain containing U1 RNA with a C to U mutation at position 4, there is an enhancement of both the aberrant cut, and the proper cleavage event relative to with wild type U1 RNA. Since the aberrant cleavage site shows no strong complementarity to the 5' end of U1 RNA, this strongly suggests that aberrant cutting is not due to a shift in the U1 RNA binding position at the 5' splice site. Although base pairing between U1 snRNA and position +5 of the conserved 5' consensus sequence (5'-GUAUGU-3') is important for efficient splicing, it apparently does not uniquely determine the specificity of cleavage site selection. In contrast, in a mammalian system, a cleavage one nucleotide upstream of the proper 5' cleavage site has been proposed to result from the shifting of U1 snRNA base pairing by one base compared to the wild type pairing, and the hypothesis tested by shifting cleavage back to the original site by the introduction of second-site mutations which restored complementarity in the original frame (Aebi *et al.*, 1987; Weber and Aebi, 1988, see section 1.2.1).

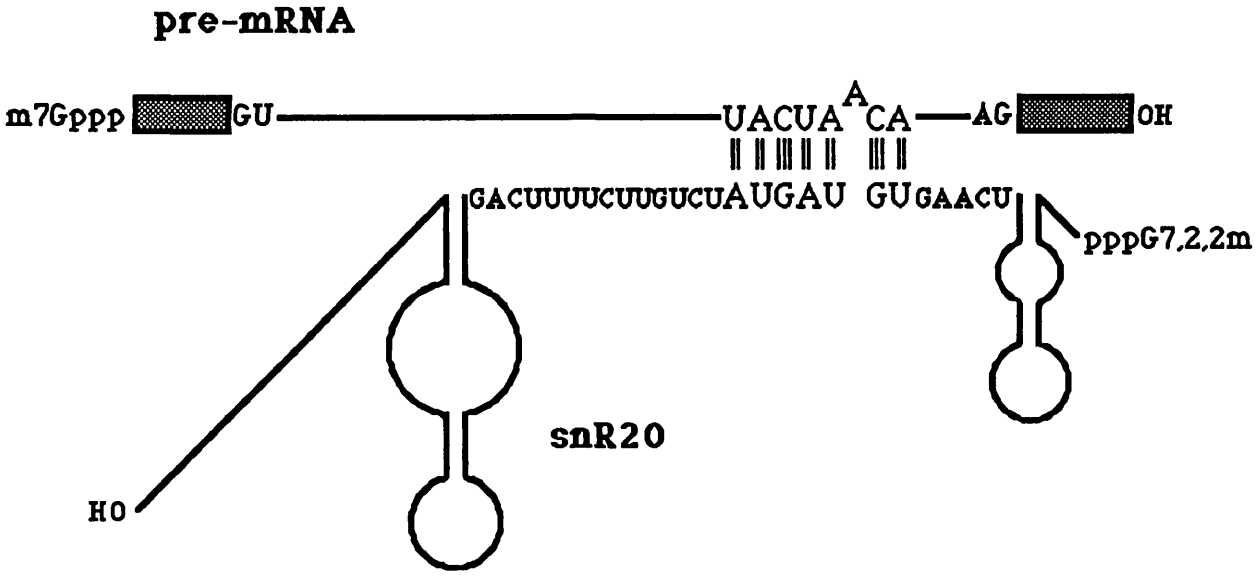
In vitro splicing was also performed in extracts made from cells carrying wild type U1 RNA or mutant U1 RNA with a C to U change at position 4 as the only functional U1 (Séraphin *et al.*, 1988). With the 5'II mutant substrate RNA all three splicing complexes (Pikielny *et al.*, 1986) were poorly assembled in wild type extract. In contrast, the mutant substrate was efficiently assembled into complexes and spliced in the mutant extract. Hence, proper pairing of U1 snRNA with position +5 of the intron is required at an early step in

spliceosome assembly. However, slightly less band I (the proposed second intermediate in spliceosome assembly) was formed than with wild type substrates. Thus, intron position +5 may be recognised a second time, subsequent to band III formation (the proposed first intermediate in spliceosome assembly). This observation is consistent with the results from the *in vitro* modification studies of Rymond and Rosbash (1988).

A similar strategy, of transforming yeast with plasmids containing a yeast centromere, a selectable marker (LEU2), and a mutant U2 snRNA gene has been used to investigate potential base pairing between the yeast TACTAAC box and snR20 (yeast U2) (Parker *et al.*, 1987). However, in this case the genomic U2 gene was not inactivated. The authors proposed the interaction shown in Figure 1.2 (Ares, 1986) rather than that proposed by Black *et al.* (1985), because the A residue that participates in the 2'-5' phosphodiester branch structure is bulged, a feature thought to be essential for self-splicing of group II mitochondrial introns, which also occurs via a lariat intermediate (Schmelzer and Schweyen, 1986). Second, this alignment allows an extra Watson-Crick interaction between U2 and the 90% conserved purine immediately 3' of the UACUAAC box (Guthrie *et al.*, 1986). Splicing actin-HIS4 chimeric transcripts with a mutation of the internal consensus from UACUAAC to UAC<u>A</u>AAC is abolished *in vivo* (Vijayraghavan *et al.*, 1986). However, production of mRNA was substantially restored in the presence of U2 containing the appropriate complementary mutation *in vivo*. A UACUAAC to U A <u>A</u> UAAC mutation inhibited splicing, but was much more leaky (Vijayraghavan *et al.*, 1986). However, efficiency of splicing *in vivo* was increased by the presence of U2 containing the appropriate complementary mutation. Much of this leakiness was due to the use of a cryptic UACUAAC as a branch point (Cellini *et al.*, 1986). When the cryptic UACUAAC element was deleted, leakiness was much reduced and the presence of U2 containing the appropriate complementary mutation could be seen to markedly suppress the mutation in the substrate RNA. The U2 with the suppressor mutation could use the UA<u>A</u>UAAC box in this construct with at least 10-fold greater efficiency than could wild type U2 alone. The two mutant U2 snRNAs were also tested for their ability to suppress mutations at other positions in the UACUAAC box than those at which they were proposed to base pair. No suppression was seen, hence, it was concluded that the mutations in U2 RNA do indeed suppress the splicing defects of UACUAAC mutations in an allele specific manner, as predicted in the base pairing model in Figure 1.2. The conclusion that base pairing occurs between the C at position 3 of the UACUAAC box and U2 is supported by several previous observations. The fact that mutation of this position to a U has no (Langford *et al.*, 1984; Vijayraghavan *et al.*, 1986) or little (Jacquier *et al.*, 1985) impact on splicing efficiency is in keeping with the

Figure 1.2. **Base pairing between snR20 and the UACUAAC box.**

The diagram shows the model for a base pairing interaction between snR20 and the UACUAAC box preferred by Parker *et al.* (1987). The pre-mRNA is shown as the upper molecule; exon sequences are represented as boxes, and the intron is shown as a line with the 5' junction sequence, the UACUAAC box, and the 3' junction as indicated. The yeast U2 analogue (snR20) is the lower molecule; the sequence of a portion is shown, with nucleotides 33 to 39, which are proposed to interact with the UACUAAC box, shown in larger type. The secondary structure shown is from Ares (1986).


pre-mRNA
m7Gppp
GU
UACUA A CA
AG
OH
GACUUUUCUUGUCUAUGAU GUGAACU
pppG7,2,2m
snR20
HO

effects expected for the substitution of a G-U base pair. In contrast, splicing was severely inhibited by the mutation of this C to a G (Fouser and Friesen, 1986) consistent with complete disruption of a base pair at this position. Recognition of the UACUAAC box, thus relies, at least in part, on Watson-Crick base pairing with the yeast U2 analogue snR20. This pairing may well occur only after a stable U2 snRNP-substrate complex has formed, as suggested by some experiments discussed above, which indicate that selective degradation of U2 snRNA in the pairing region does not abolish early splicing complex assembly in mammalian systems.

1.4. THE SPLICEOSOME

Eukaryotic pre-mRNA splicing takes place in a multicomponent complex designated the "spliceosome" (Brody and Abelson, 1985; Grabowski *et al.*, 1985). A variety of techniques have been used to study splicing complexes and these are discussed below.

1.4.1. Density gradient sedimentation

Rapidly sedimenting complexes containing the two RNA intermediates of splicing, the 5' exon and lariat intron-3' exon, were originally detected by density gradient sedimentation of yeast (Brody and Abelson, 1985) and mammalian (Frendewey and Keller, 1985; Grabowski *et al.*, 1985) *in vitro* splicing reactions. Sucrose (Frendewey and Keller, 1985; Bindereif and Green, 1986) and glycerol gradients (Brody and Abelson, 1985; Grabowski *et al.*, 1985) have been used, and the complexes containing splicing intermediates were found to sediment at 40S in yeast and 50-60S in mammalian systems. In addition to pre-mRNA and intermediates the spliceosome peak also contains the lariat intron produced by the second step of splicing. Spliced exons are much less abundant and seem to be released or more loosely associated with spliceosomes (Brody and Abelson, 1985; Frendewey and Keller, 1985; Bindereif and Green, 1986). Further evidence that the complex containing the intermediates was the functional splicing substrate came from the fact that pre-mRNA is also found in the peak which contains the intermediates (Brody and Abelson, 1985; Frendewey and Keller, 1985; Grabowski *et al.*, 1985). Like splicing, formation of the spliceosome complex requires ATP (Brody and Abelson, 1985; Frendewey and Keller, 1985; Grabowski *et al.*, 1985) and in the case of the yeast system substituting ATP with the non-hydrolysable analogue β,γ-methylene adenosine 5'-triphosphate (AMPPCP), both prevents step 1 of

splicing and and abolishes the gradient peak at 40S (Brody and Abelson, 1985). Glycerol density gradient centrifugation has also revealed that *in vitro* assembly of a 40S spliceosome on a truncated human metallothionein pre-mRNA was strongly inhibited by preincubation of the reaction mixture in the presence of m^7GTP (at concentrations as low as 5μM) before addition of the substrate RNA (Patzelt *et al.*, 1987). Furthermore, formation of this complex was also strongly reduced when substrate RNA lacking the 7-methylated cap structure was spliced *in vitro*. Formation of the spliceosome also requires U snRNPs, as does splicing (see section 1.3). Spliceosome formation was inhibited either by direct addition of anti-U1 snRNP antiserum to the extract (Grabowski *et al.*, 1985), or by pre-treating an extract with anti-Sm antiserum to deplete it of U snRNPs (Frendewey and Keller, 1985). Many mutations which inhibit splicing also lead to inefficient complex formation. In yeast, mutation of the branch point A to a C in the $CYH2^m$ gene intron, which prevents both steps of splicing (Newman *et al.*, 1985) also removed the 40S spliceosome peak (Brody and Abelson, 1985). An A to C branch nucleotide mutation in the yeast actin intron had the same effect (Vijayraghavan *et al.*, 1986). In contrast, mutating the 5' G of the $CYH2^m$ intron to A, which yields an RNA that is a substrate for step 1, but not step 2 of splicing (Newman *et al.*, 1985), only reduced the size of the 40S peak. Interestingly, the same mutation in the yeast actin intron, along with mutation of the first nucleotide of the intron to a C, the fifth from a G to an A and three mutations of the TACTAAC box (not at the branch residue) blocked formation of the 40S complex (Vijayraghavan *et al.*, 1986). These observations are consistent with spliceosome formation being essential for the first step of splicing to proceed. Mutations of the yeast actin gene 3' splice site sequence AG/AG to AC/AC, which yields an RNA which can undergo the first step of splicing both *in vivo* and *in vitro*, gave rise to a 40S complex indistinguishable from that formed on wild type RNA (Vijayraghavan *et al.*, 1986). Frendewey and Keller (1985) have also used various altered and truncated adenovirus 2 major late pre-mRNAs as substrates in a mammalian extract system and found 50S complex formation to require a 5' splice site, a polypyrimidine tract and possibly a branch site. A substrate with a wild type 5' splice site, branchpoint and polypyrimidine tract, but no conserved 3' AG, will form the 50S complex, but very inefficiently. This mutant (in contrast to the other mutations) can carry out step 1 of splicing (with reduced efficiency). In mammalian systems it would seem, therefore, that the conserved AG dinucleotide at the 3' splice site is important, but not essential for spliceosome formation and cleavage at the 5' splice site (see section 1.2.2). The same applies to exon 2 sequences, where the presence of as few as 4 nucleotides following the 3' cleavage site greatly enhances splicing complex assembly and step 1 of splicing. Deletion of the 5' splice site was found to

prevent 50S complex formation, but not 35S complex formation (Frendewey and Keller, 1985). Grabowski *et al.* (1985) found that although no 60S spliceosome complex formed on a bacterial RNA lacking authentic 5' and 3' splice sites, a 50S complex, slightly smaller than the 60S spliceosome, formed on a truncated RNA containing only a 5' splice site. However, some or all of this reduction in size must be due to the shorter pre-mRNA used. This result is now largely believed to be artifactual: other groups have failed to detect the formation of any splicing complex on substrates containing only a 5' splice site (Frendewey and Keller, 1985; Perkins *et al.*, 1986). In mammalian, as with yeast systems, it would seem that a peak of spliceosome mobility is only found when at least the first step of splicing can occur.

Other evidence for the spliceosome is that appearance of the 50S complex coincided with the appearance of intermediates and products of the splicing reaction following a lag period (Frendewey and Keller, 1985). More directly, when supplemented with nuclear extract and ATP, the pre-mRNA in the 60S complex can be chased into spliced RNA with accelerated kinetics and higher efficiencies than newly added precursor RNA (Grabowski *et al.*, 1985).

Smaller RNP complexes can also be identified. With a mammalian system, Grabowski *et al.* (1985) observed a 40S complex on glycerol gradients. They believed this complex not to be directly related to splicing, and instead probably due to non-specific binding of cellular components (probably hnRNP proteins) because it formed at 4°C, without ATP, when pre-incubated with anti-(U1 RNP) antisera and on RNA with neither 5' or 3' splice sites. Splicing of RNA in the 40S fraction showed similar efficiency and kinetics of splicing to exogenously added substrate RNA. However, assembly of some fraction of the 40S complex could still be an important step in 60S splicing complex assembly.

Frendewey and Keller (1985), on their sucrose gradients, identified in addition to the 50S splicing complex, a 22S and a 35S complex. The 22S complex is in many ways similar to the 40S complex of Grabowski *et al.* (1985) in that it forms immediately on mixing RNA and nuclear extract, in the absence of ATP or Mg^{2+} and does not require snRNPs. However, although a 5' splice site deletion mutant does form a 22S complex, a substrate lacking the polypyrimidine tract and 3' exon did not. The 22S complex disappears by 15 minutes into the reaction. The 35S complex is observed transiently between 5 and 20 minutes into the *in vitro* splicing reaction. They proposed this to be a "pre-splicing complex" because it appeared to be converted with time into the 50S complex; it required snRNPs for assembly; did not form with substrates lacking the polypyrimidine tract (and lacking the 3' exon); but would form with substrates from which the 5' splice site has been deleted. Wlthout ATP, only the 35S complex was observed, implying that conversion of the 35S to the

50S complex requires ATP and an intact 5' splice site. It was also found that the 5' splice site sequence must reside upstream of the 3' splice site sequences for 50S complex formation and splicing. Precursor RNA, lariat, lariat-IVS-2 and exon 1 were all immunoprecipitable from the spliceosome peak by anti-(U1 RNP) and anti-(Sm RNP) antisera, implying that snRNP specific polypeptides are present in the spliceosome (Bindereif and Green, 1986; Grabowski *et al.*, 1985). However, more detailed characterisation of the composition of the rapidly sedimentable splicing complexes awaited the use of affinity chromatography (see section 1.4.2).

Sucrose gradient sedimentation analysis has also been used to show that two spliceosomes can form simultaneously and independently on pre-mRNA containing two introns *in vitro* (Christofori *et al.*, 1987).

1.4.2. Affinity Chromatography

Pikielny and Rosbash (1986) carried out oligo(dT)-cellulose affinity chromatography of yeast spliceosomes on a polyadenylated substrate. The splicing substrate, with a poly(A) tail of 100 nucleotides, was incubated for 20 minutes in a yeast *in vitro* splicing system and then fractionated by oligo(dT)-cellulose affinity chromatography; the RNAs in the flow-through and bound fractions were analysed by denaturing PAGE. The bound fraction contained only very small amounts of RNAs if a poly(A)$^-$ control substrate was used. However, with poly(A)$^+$ RNA a fraction of substrate, lariat intermediate, spliced exons and 5' exon were present in the bound fraction, but there was no specific binding of free intron. Hence, it would seem that both products of step 1 are present in spliceosomes. Retention of the 5' exon is particularly significant since it has no poly(A) tail. Free lariat must not be associated to a significant extent with any particle containing the 3' exon.

TMG-capped snRNAs present in the bound fraction were visualised by [^{32}P]pCp labelling, immunoprecipitation using anti-(TMG) antibody (Lührmann *et al.*, 1982), and denaturing PAGE. Four snRNAs were specifically enriched in this fraction when wild type poly(A)$^+$ substrate RNA was used, but not with poly(A)$^-$ control substrate or poly(A)$^+$ mutants defective in splicing as a result of deletion of either the 5' splice site or TACTAAC box. One of these snRNAs was very large (approximately 1000 nucleotides) and almost certainly corresponds to snR20 (Ares, 1986, see below). The other three snRNAs, on the basis of their mobilities were probably snR14 (yeast U4), snR7S and snR7L (two forms of yeast U5).

Affinity chromatography of spliceosomes based on the high affinity of the biotin-streptavidin interaction has also been performed (Grabowski and

Sharp, 1986; Bindereif and Green, 1987). In these studies the substrate RNA was biotinylated by *in vitro* transcription in the presence of a biotin-UTP analogue. Bindereif and Green (1987) selectively recovered splicing complexes directly from *in vitro* splicing reactions by using biotinylated RNA as substrate and binding this RNA to streptavidin agarose beads. A non-physiological salt concentration of 300mM or higher was needed for efficient recovery of biotinylated pre-mRNA from a crude nuclear extract. After 30 minutes incubation in nuclear extract U1, U2, U4, U5 and U6 snRNAs were all associated with the pre-mRNA, as judged by their electrophoretic mobility, immunoprecipitability by anti-TMG antibody, their partial RNase T1 digestion patterns, and cleavage by RNase H in the presence of snRNA-specific oligonucleotides. These snRNPs did not associate with a prokaryotic control RNA. Binding of all snRNPs, except U1, was ATP dependent. Biotinylated human β-globin cDNA transcript, in contrast to the normal β-globin substrate, bound snRNPs poorly. With substrates in which the polypyrimidine tract, the last 4 nucleotides of the intron, including the highly conserved AG, or both these regions were deleted, the binding of U1 snRNP was unaffected, U2 snRNP binding was reduced to about 40% of normal levels, and U4, U5 and U6 snRNP binding fell to 10-40% of normal levels. A truncated substrate, lacking the entire branch point/3'splice site region and the 3' exon, bound U1 snRNP normally, but all other snRNP binding was reduced to low, levels. A substrate lacking the 5' splice site but containing the branch point/3' splice site region bound U2 snRNP at 50-100% of normal levels, but binding of U1, U2, U4, U5 and U6 snRNPs was much reduced. Assembly of U2, U4, U5 and U6 into all partial splicing complexes required ATP. A time course of snRNP binding revealed U1 snRNP to bind almost immediately (on ice, no incubation at 30°C) and to reach maximal levels after one minute at 30°C. U2, U4, U5 and U6 snRNPs all bound with identical kinetics, first associating with the pre-mRNA after 10 minutes and reaching maximal levels by 30 minutes. At 60 minutes, when a substantial portion of the pre-mRNA had undergone step 1, the levels of all bound snRNPs decreased dramatically.

Biotin-streptavidin affinity chromatography has also been used to determine the snRNP composition of complexes fractionated on sucrose or glycerol gradients (Grabowski and Sharp, 1986; Bindereif and Green, 1987, Lamond *et al.*, 1988). Splicing complexes incubated with heparin prior to gradient fractionation sedimented as discrete 35S, 25S and 15S complexes. Only the 35S complex contained RNA intermediates. The 35S complex was found to contain U2, U4, U5 and U6 snRNPs, the 25S complex only U2 snRNP, and the 15S complex no snRNPs (Grabowski and Sharp, 1986; Bindereif and Green, 1987). U1 snRNP was absent from all particles. With an adenovirus RNA substrate affinity chromatography showed the 35S spliceosome complex to contain also

an uncharacterised RNA species (X) of approximately 75 nucleotides (Grabowski and Sharp, 1986; Lamond *et al.*, 1988) but this RNA was not not found if the substrate was rabbit β-globin RNA (Lamond *et al.*, 1988). RNA X appears not to be a subfragment of U1, U2, U4, U5 or U6 snRNAs, or the adenovirus substrate, and is present in HeLa cells not infected by adenovirus. It may correspond to a minor snRNA species that can modulate spliceosome activity (Lamond *et al.*, 1988).

If sucrose gradient fractionation was performed without pre-treatment with heparin, three complexes were also seen, now sedimenting at 20S, 40S and 60S (Bindereif and Green, 1987). The 20S complex contained U1 snRNP only, and the 60S complex contained U1, U2, U4, U5 and U6 snRNPs. The level of U1 snRNP in the 60S complex was low unless the salt concentration was dropped, indicating that it is not strongly bound in this complex. The existance of a 40S complex containing only U1 and U2 snRNPs was postulated on the basis of indirect evidence but not not detected directly. Most complexes in the 40S region were thought to be non-productive and due to interaction with non-specific RNA binding factors.

The absence of U1 snRNP from the heparin pre-treated gradients was shown to be due to the heparin treatment, as splicing complexes affinity purified from crude extract after heparin treatment lacked U1 snRNP, in contrast to complexes affinity purified using high salt (Bindereif and Green, 1987). Based on the observed sites and times of binding of snRNPs to pre-mRNA obtained from combined RNase T1 digestion and immunoprecipitation experiments (Black *et al.*, 1985; Chabot and Steitz, 1987; see section 1.4.5), Bindereif and Green (1987) interpreted their results in terms of the following model for an ordered pathway of spliceosome assembly. First, U1 snRNP binds to the 5' splice site and forms an approximately 20S complex in an ATP independent manner. Second, U2 snRNP binds to the branch point of the 20S complex forming an approximately 40S complex in an ATP dependent step. The evidence for this step is, however, indirect. Third, U4, U5 and U6 snRNPs associate with the 40S complex to form the functional 60S spliceosome. There is no direct evidence for the precursor-product relationship described, the 20S and 40S complexes could be dead-end products, although there is some kinetic evidence for a stepwise assembly of spliceosome complexes from sucrose gradient sedimentation (Frendewey and Keller, 1985; see above) and fractionation on native gels (Pikielny *et al.*, 1986; Konarska and Sharp, 1986; see below).

Biotin-avidin affinity chromatography has also been used to investigate yeast spliceosome assembly (Rymond *et al.*, 1987; Kretzner *et al.*, 1987). snR19 (yeast U1) and snR20 (yeast U2) associated rapidly (within 30 seconds) and stably with the biotinylated substrate. snR7L and snR7S (yeast U5), and snR14

(yeast U4) were not stably associated with the substrate until 5 minutes after the start of incubation. A comparison of the kinetics of assembly of splicing complexes on native gels (as Pikielny *et al.*, 1986), and the kinetics of snRNA association with biotinylated substrate RNA, suggested that snR19 and snR20 were components of band III (the first to form), while snR7 and snR14 were absent from band III. This description was consistent with that previously presented for yeast spliceosome assembly and snRNA composition (Pikielny *et al.*, 1986), with the exception that snR19 would appear to be complexed with the substrate from an early stage in the splicing reaction onwards.

1.4.3. Gel filtration

Reed *et al.* (1988) have purified mammalian spliceosomes by fractionation on a sephacryl S-500 column. To enrich for a relatively homogeneous population of spliceosomes splicing was performed using heat-treated extract in which a factor required for step 2 of splicing had been inactivated (Krainer and Maniatis, 1985). Four major peaks of ^{32}P radioactivity were detected. One peak contained the spliceosomes since this is where the splicing intermediates were detected. The void volume of the column appeared to contain aggregated complexes, including spliceosomes, and the total column volume fractions contained free ^{32}P-labelled nucleotides generated from RNA turnover. There was also an additional peak containing pre-mRNA (peak X) and this may correspond to a putative pre-spliceosome complex previously identified by density gradient sedimentation (Frendewey and Keller, 1985; Grabowski *et al.*, 1985) and native gel electrophoresis (the A complex; Konarska and Sharp, 1986). This technique gave functional splicing complexes which contained the products of the first step of splicing. Spliced RNA could be efficiently generated when the purified spliceosomes were incubated with magnesium chloride, ATP, and either micrococcal nuclease-treated nuclear extract or cytoplasmic S100 extract which are not themselves competent in splicing (Krainer and Maniatis, 1985). End-labelling RNA in the peak fractions with [^{32}P]pCp revealed snRNAs U1, U2, U4, U5 and U6 to be highly enriched in the spliceosome peak. U2, U4, U5 and U6 were present in roughly equimolar amounts, but U1 was less abundant. Only snRNAs U1 and U2 were highly enriched in peak X. That snRNPs were specifically associated with purified ^{32}P-labelled spliceosomes was shown by the fact that pre-mRNA and splicing intermediates in the spliceosome peak were specifically immunoprecipitated with antisera specific for the TMG cap structure or the Sm antigen. The same result was found using anti-U1 snRNP monoclonal antibody, and U1, U2, U4, U5 and U6 snRNAs were quantitatively immunoprecipitated, indicating that at

least one U1 snRNP is associated with each spliceosome. This result is interesting, since certain other techniques have suggested that U1 snRNP may not be part of the spliceosome (Konarska and Sharp, 1986; see section 1.4.4). In addition, Lamond *et al.* (1988), using native gel electrophoresis believed U4 to be lost on formation of the rabbit β-globin IVS-2 spliceosome. This is inconsistent with the data from gel filtration. These discrepancies are probably explicable by a model in which U1 and U4 snRNAs are associated, but not tightly with the spliceosome and are dissociated by certain techniques but not others.

Few differences were observed between the SDS-PAGE patterns from the total spliceosome peak and proteins immunoprecipitated from this fraction with anti-U1 or anti-TMG antibodies. Hence, most of the proteins in this peak were snRNP associated. The spliceosome peak was examined by electron microscopy and found to consist of a relatively homogeneous population of 40-60nm ellipsoid-shaped particles. These particles were identified as spliceosomes by their sensitivity to micrococcal nuclease, their ATP-dependent assembly, and their immunoprecipitation with anti-TMG monoclonal antibody. Using a substrate RNA with a long 3' end, 10% of the particles in the spliceosome peak were visualised to contain this RNA after using bacteriophage T4 gene 32 protein to coat the RNA tail. The remaining 90% of the particles may be spliceosomes assembled on endogenous pre-mRNA in the nuclear extract, or multi-snRNP complexes assembled in an ATP dependent manner in the absence of pre-mRNA. Although complexes formed in the absence of pre-mRNA which contain all the snRNPs have not been reported, the majority of U4, U5 and U6 snRNAs have been found, by native gel electrophoresis, to be present as a single U4/5/6 complex in mammalian and yeast splicing extracts (Konarska and Sharp, 1987; Lossky *et al.*, 1987). However, under optimal conditions for *in vitro* splicing, i.e. in the presence of ATP and at 30°C, this complex dissociates to release free snRNPs. Analysis by glycerol gradient centrifugation and native gel electrophoresis has also revealed that a U2/4/5/6 complex (the pseudospliceosome) forms in the absence of precursor RNA during incubation of the extract at high salt concentrations, and this association was typically stimulated by the addition of ATP (Konarska and Sharp, 1988).

The size of the particles correlates well with the 40nm particles presumed to be spliceosomes found on nascent transcripts *in vivo* by using the Miller chromatin spreading technique (Osheim *et al.*, 1985; Beyer and Osheim, 1988).

1.4.4. Electrophoresis on native polyacrylamide gels

The earliest attempts at fractionating mammalian spliceosomes on native, low-percentage (4%) polyacrylamide gels revealed two distinct complexes, A and B, to be assembled specifically on an RNA precursor containing authentic 5' and 3' splice sites (Konarska and Sharp, 1986). In this study, the splicing reaction was pre-incubated with heparin before electrophoresis and a Tris-borate, EDTA buffer was used. The assembly of two complexes of low mobility, A and B, was ATP dependent, whereas a higher mobility smear was not. This higher mobility smear was presumed to be due to non-specific binding of components in the extract to RNA since a similar complex also formed on bacterial transcripts. The kinetics of assembly suggest that complex A may be being converted with time to the slower migrating complex B. By non-denaturing electrophoresis of heparin-treated complexes previously fractionated on glycerol gradients, it was found that complexes A and B corresponded to ATP-dependent structures of 25S and 35S respectively, and the faster migrating complex corresponded to the ATP-independent 15S complex. The 35S peak was the only one to contain splicing intermediates and presumably corresponded to the 60S spliceosome complex formed without heparin treatment (Grabowski *et al.*, 1985). There was, however, only a single complex (40S) formed without heparin pre-treatment, and this was ATP-independent. Substrate RNA containing only the 3' splice site could form the smaller complex A, but not complex B, whereas the 5' splice site alone could form neither complex A or B. Complex A protected a fragment from digestion by RNase T1, which consisted of 46 nucleotides upstream of the 3' splice site, including the polypyrimidine tract and the branch point. Northern hybridisation analysis indicated that complexes A and B contained U2 snRNA, but not U1 snRNA. This technique also revealed a large complex containing U2 snRNP which was present in native nuclear extracts and which on incubation with ATP broke down to a faster migrating form, probably the 8-10S snRNP particle.

Switching to a Tris-glycine buffer made native gel electrophoresis of splicing complexes possible without heparin addition (Konarska and Sharp, 1987). A two dimensional separation procedure involving fractionation on glycerol gradients followed by native gel electrophoresis of the gradient fractions was used to reinforce much of the data obtained from electrophoresis alone. A similar pattern of complexes A, B and H (the non-specific high mobility complex) was observed to that seen with heparin treatment but mobilities were slightly reduced. Both complexes A and B contained U2 but not U1 snRNA as determined by Northern hybridisation, as was the case with the equivalent heparin treated complexes (Konarska and Sharp, 1985; see above). Northern hybridisation also indicated that U4, U5 and U6 snRNAs were not present in complex A, but were all present in the larger complex B. This is

consistent with data from affinity chromatography of spliceosomes, which indicates that the mature spliceosome contains all these snRNAs (Grabowski and Sharp, 1986; Bindereif and Green, 1987; see section 1.4.2) and in the case of the study of Bindereif and Green (1987), also U1 snRNA. The excised intron was released in a complex containing U5, U6 and probably U2 snRNPs. Since U4 snRNA was not part of the intron containing complex, it would seem that the U4/U6 snRNP (see section 1.3) must assemble and disassemble during splicing. U1 snRNP was not detected in any of the splicing complexes. In contrast to the situation after heparin treatment, U2 snRNP from untreated HeLa cell nuclear extracts was found to run as a cluster of bands of relatively low mobility, with or without incubation with ATP.

Multiple forms of endogenous snRNPs have been found to pre-exist in mammalian nuclear extract. The majority of U4, U5 and U6 snRNAs were present in a single complex termed U4/5/6 (Konarska and Sharp, 1987). This large complex was stable under a variety of conditions, including a wide range of pH, temperature, and salt concentration. However when incubated in the presence of ATP under standard *in vitro* splicing conditions, this U4/5/6 complex disassembled to generate a U4/6 particle and finally separate U4, U5 and U6 snRNPs. The authors suggest that complex A, which contains only U2, may be converted to complex B by binding of a single snRNP particle containing U4, U5 and U6 snRNAs (the U4/5/6 complex). As described in section 1.4.3, analysis by glycerol gradient centrifugation and native gel electrophoresis has also revealed that during incubation of nuclear extract at high salt concentrations a U2/4/5/6 complex (the pseudospliceosome) forms in the absence of precursor RNA, and this association was typically stimulated by the addition of ATP (Konarska and Sharp, 1988).

The above studies were on RNA containing the leader 1 and 2 exons of adenovirus, but the same technique has also been applied to the study of spliceosome assembly on rabbit β-globin IVS-2 pre-mRNA from which much of the non-essential intronic sequence had been deleted (Lamond *et al.*, 1987). In contrast to the work with adenovirus RNA (Konarska and Sharp, 1987; see above), three, (rather than two) kinetically related, ATP-dependent complexes, denoted α, β and γ, were resolved. The fastest mobility complex (α) formed first, followed by complex β, and then finally the slowest migrating complex (γ). RNA intermediates of splicing were found predominantly in the γ complex, which is, therefore, believed to constitute the spliceosome. This fits well with the fact that 5' splice site cleavage was first seen at the same time as the γ complex. A small amount of excised intron was enriched in the γ band, but most was found in a large particle migrating between α and β (consistent with Konarska and Sharp, 1987). A smaller complex, migrating with the non-specific (sigma) complexes contained the spliced exons, suggesting that

upon completion of splicing, the spliceosome breaks down in a specific way, releasing two types of post-splicing complexes that contain the separate reaction products. It is thought that on adenovirus RNA the equivalent of the globin β and γ complexes comigrate in the adenovirus B complex band (Konarska and Sharp, 1987).

Northern hybridisation and electrophoresis of [^{32}P]pCp end-labelled RNA eluted from native gels revealed that the smallest complex (α) contained only U2 snRNA, complex β contained U2, U4, U5 and U6 snRNAs and the slowest migrating complex (γ) contained U2, U5 and U6 snRNAs, but not U4 (Lamond *et al.*, 1988). None of the complexes contained U1 snRNA. Thus, it was proposed, on the basis of this and affinity purification data presented in the same paper (Lamond *et al.*, 1988; see section 1.4.2) that spliceosome assembly occurs in a stepwise fashion. First, U2 snRNP binds in an ATP dependent manner to form complex α. This is followed by the apparently simultaneous binding of U4, U5 and U6 snRNPs, possibly as a single multi-snRNP complex (Konarska and Sharp, 1987) to form complex β. The subsequent release of U4 snRNA generates the mature spliceosome γ containing the splicing intermediates. The intron in lariat form is released in a complex containing U2, U5 and U6 snRNPs (Konarska and Sharp, 1987; Lamond *et al.*, 1988).

Mutating the conserved GU at the 5' splice site to AU or GA reduces the rate of 5' splice site cleavage *in vitro* and causes an accumulation of lariat intermediate due to step 2 being blocked (Aebi *et al.*, 1986 and 1987; see section 1.2.1). Both mutants, relative to wild type, showed a reduced rate of α formation, after which the β and γ bands formed quite rapidly (Lamond *et al.*, 1987). The kinetics of formation and accumulation of the γ band matched those of the splicing intermediates. The γ complex continued to accumulate throughout incubation, in contrast to the situation with the wild-type substrate. A deleted substrate, lacking the 5' splice site region, including the two major cryptic sites also formed three bands, α", β" and γ" with very similar kinetics to the above point mutations. These complexes migrated slightly faster than those on wild-type RNA, but this is at least in part due to the reduced pre-mRNA length. These deletion mutants utilise novel cryptic 5' splice sites to form lariats, but at very low levels, therefore it is difficult to know if some form of 5' splice site interaction is necessary for spliceosome formation or not. In contrast, RNA containing only the 5' splice site region showed no sign of specific complex formation. Mutating the conserved 3' splice site AG to UG or AU resulted in only trace amounts of RNA intermediates and no fully spliced products *in vitro*, in agreement with earlier work (Ruskin and Green, 1985b; Reed and Maniatis, 1985; Aebi *et al.*, 1986). Even at the longest incubation times, only a small fraction of the pre-mRNA was assembled into ATP-dependent complexes, the rest remained in the sigma region. The

three ATP-dependent complexes which did form also had different mobilities to those formed on wild-type substrates. α' and β' approximately comigrated with the α and β bands, whereas the third band, μ, migrated in a novel position between α' and β'. μ may correspond to an authentic intermediate in spliceosome formation, or could be a completely aberrant complex. The amount of ATP-dependent complex was reduced still further if these 3' splice site mutants were combined with 5' splice sites carrying either the G to A or U to A point mutations above, and ATP-dependent complex formation was abolished completely if the 3' splice site mutants were combined with deletions of the 5' splice site. Hence, formation of the α complex seems to be facilitated by recognition of both the 5' and 3' splice sites.

Zillmann *et al.* (1988) have used an alternative electrophoresis system to analyse splicing complex formation on adenovirus 2 RNA substrates. In this system splicing reactions were pre-treated with heparin before electrophoresis on a composite acrylamide-agarose gel, cast and run in a Tris-glycine buffer containing 10mM EDTA. The banding pattern seen was very similar to that found previously with heparin treated spliceosomes. However, the authors report that in this system radiolabelled RNA within the A and B complexes could be immunoprecipitated by U1-specific antibodies from gel-isolated complexes. This implies that U1 snRNP, or at least the U1 RNP antigen was present in these complexes. The authors also reported that an RNA with the 5' splice site deleted formed a single, unique complex (denoted α) which migrated slightly faster than complex A, despite the deleted substrate being longer than the wild-type RNA, but if electrophoresis of this substrate was performed in the absence of EDTA the complex formed co-migrated with complex A. Therefore, it was argued that a factor(s) involved in 5' splice site recognition may be missing from α, U1 snRNP being a particularly obvious candidate. Significantly, Konarska and Sharp (1986), found no U1 snRNP in the spliceosome bands and also found normal formation of complex A with a mutant substrate in which the 5' splice site had been deleted. A substrate lacking all 3' splice site and branch point sequences formed no ATP-dependent complexes.

Oligonucleotide directed RNase H cleavage of the 5' ends of U1 or U2 snRNAs prevented all specific complex assembly (Zillmann *et al.*, 1988). Therefore both U1 and U2 were required for complex A formation. α complex formation on substrates with the 5' splice site deleted also required intact U1 and U2 snRNAs (Zillmann *et al.*, 1988). When U2 snRNA was cleaved to remove nucleotides 28 to 42, complex A was formed in significant amounts, but no complex B was formed. The region cleaved in this case is that thought to be involved in base pairing with the branch point in an analogous manner to snR20(yeast U2) interacting with the TACTAAC box (Parker *et al.*, 1987). This

data, along with that of Krämer (1987b) discussed in section 1.4.5, would seem to imply that base pairing of U2 snRNA with the pre-mRNA may occur only after a stable U2 snRNP-substrate complex has already been formed. Cleavage of U4 snRNA allowed formation of complex A, but not complex B, although, strangely, complex A', which contains the lariat intron, was also observed. Hence, two complexes could be comigrating in the A' position, one of them being the spliceosome breakdown product which contains the intron, the other a pre-spliceosome complex.

Yeast spliceosomes have also been fractionated by electrophoresis on composite acrylamide-agarose gels (Pikielny and Rosbash, 1986). In this study no polyanions were added before electrophoresis and a single ATP-dependent splicing complex was observed which contained the intermediates of splicing and the excised lariat intron. Transcripts from genes in which the 5' splice site or TACTAAC box had been deleted, and which were completely defective as splicing substrates, did not form the ATP-dependent spliceosome band, in agreement with sedimentation data which indicated that the 40S spliceosome did not form if the substrate contains a mutation converting the branch point A to C (Brody and Abelson, 1985). RNA was extracted from the spliceosome band, 3' end labelled with [^{32}P]pCp, and immunoprecipitated with anti-TMG antibody. Three snRNAs were specifically associated with the spliceosome band. These three species are almost certainly snR14 (yeast U4), snR7S and snR7L (two forms of yeast U5).

The addition of a polyanion before electrophoresis, in this case non-specific competitor RNA, not heparin, increased the mobility of both the spliceosome complex and the higher mobility unspecific complex on the gels (Pikielny *et al.*, 1986). The spliceosome band could now be resolved into three seperate complexes, called I, II and III in order of increasing electrophoretic mobility. Bands I and II appeared very early in the reaction, being maximal at 5 minutes and decreasing thereafter. They therefore appeared well before splicing intermediates could be detected. Band II was not visible until 5 minutes of incubation, after which time it continued to increase in intensity. Its appearance coincided with the appearance of the intermediates. An RNA with an A to C mutation at the branch point, which gives rise to no detectable intermediates or products *in vitro* did, however, give rise to band III, but not to bands I or II, supporting the hypothesis that complex III is the initial complex formed, and that subsequent formation of complexes I and II is inhibited by the branch point mutation. The products of the first step of splicing were found almost exclusively in complex II. Lariat intron was, however, distributed over bands II and III. A very low level of spliced exons was visible in band II.

Four yeast snRNAs were found to be associated with the splicing

complexes by elution and [^{32}P]pCp labelling (Pikielny *et al.*, 1986). snR20 (LSR1) (yeast U2) was associated with all three complexes and band III appeared to contain only this species. Band I contained snR20 (yeast U2), snR7S and snR7L (yeast U5) and snR14 (yeast U4), but band II (the mature spliceosome) contained snR20, snR7S and snR7L, but lacked snR14. It was proposed, therefore, that complexes form in an ordered manner, first snR20 (U2) binds and forms complex III, then snR7S and/or snR7L (U5) and snR14 (U4) bind to form complex I. SnR14 (U4) is then lost to form complex II. Loss of snR14 may occur before, or coincidently with, the first step of splicing. The analogy with mammalian systems is striking, U4 being lost in each case in formation of the mature spliceosome.

Using a slightly different electrophoresis system (including heparin pre-treatment) and substrate RNA containing the yeast actin intron, four splicing-dependent complexes, A1, A2-1, A2-2 and B were observed (Cheng and Abelson, 1987). Complex B was the first to form. Complex A2-1 followed complex B and was detected after 2 minutes incubation. A1, the next kinetic intermediate was difficult to detect unless heat treated *rna2* extracts (Lustig *et al.*, 1986) were used, in which case A1 accumulated. Alternatively A1 could be seen if splicing was performed in the presence of low concentrations of EDTA such that only a portion of the Mg^{2+} was chelated. At higher concentrations of EDTA assembly was completely blocked at complex A1, which accumulated. Complex A1 is thought to be a functional intermediate, since when this complex was isolated from splicing reactions performed at high concentrations of EDTA or from heat treated *rna2* extracts, it could be chased into the active form by addition of an extract fraction and ATP. Complex A2-2 is the active form of the spliceosome, containing RNA intermediates, and accumulated when a substrate with the 3' splice site mutation AG/AG to AC/AC (Vijayraghavan *et al.*, 1986; see section 1.2.2) was used. This substrate can form spliceosomes, as detected by glycerol gradient sedimentation (Vijayraghavan *et al.*, 1986). ATP was shown to be absolutely required for formation of all four complexes. Northern hybridisation was performed to identify the snRNA components of the splicing complexes using DNA probes from clones of snR19 (yeast U1), snR20 (yeast U2), snR7 (yeast U5) and snR14 (yeast U4), and a 60 nucleotide oligomer complementary to residues 21-80 of *Drosophila* U6 snRNA. Complex B was found to contain snR20 (U2) alone. Complex A2-1, the putative next intermediate in the spliceosome assembly pathway, contained snR20 (U2), snR7 (U5), snR6 (U6) and snR14 (U4). Complex A1, the next complex in the assembly scheme had the same snRNP composition as A2-1, except it lacked snR14 (U4). Complex A2-2, the functional spliceosome, had the same snRNP composition as A1. SnR19 (yeast U1) was not found in any of the complexes. This pathway, therefore, matches very well, both that proposed from earlier

studies in yeast (Pikielny *et al.*, 1986), and the pathway proposed for mammalian splicing (Lamond *et al.*, 1988).

Northern hybridisation also revealed that, in yeast whole cell extracts, all of the snRNAs exist in several forms (Cheng and Abelson, 1987). Significantly, snR14 (U4), snR7 (U5) and part of the snR6 (U6) population appeared to form a single large (slowly migrating) complex which could be dissociated by ATP to give snR7 (U5) and snR6 (U6)-snR14(U4) complexes. A similar complex, containing U4, U5 and U6 snRNAs can also be detected in HeLa nuclear extracts and is thought to possibly assemble directly onto the spliceosome (Konarska and Sharp, 1987). Incubation with ATP also causes U5 to dissociate from this particle. SnR19 and snR20 were present in distinct RNPs, but the mobility of these was unaffected by incubation with ATP.

1.4.5. RNA footprinting

The association of splicing factors with specific pre-mRNA regions inherent in spliceosome assembly has been studied using a variety of nuclease protection assays.

DNA oligonucleotides complementary to specific RNA regions have been added to splicing reactions, along with RNase H, at various times during *in vitro* processing. A RNase H-directed site specific cleavage reaction should occur unless a factor(s) associated with the RNA prevents hybridisation of the oligonucleotide and/or cleavage by RNase H (Ruskin and Green, 1985a). This technique indicated that, on human β-globin IVS-1, prior to pre-mRNA cleavage, factors specifically associate with the 5' splice site and branch point. The association of a factor(s) with the branch point preceded the association of a factor(s) with the 5' splice site. Both IVS-1-containing processing products displayed a pattern of protection from RNase H-digestion analogous to that found with the pre-mRNA, i.e. protected against cleavage directed by oligonucleotides complementary to the 5' splice site and branch point. Hence, factors must be associated with the branchpoint region of the RNA lariats and this could explain the resistance of these RNAs to debranching by the endogenous debranching activity found in crude nuclear extracts and S100 unless thay are first deproteinised (Ruskin and Green, 1985c). Cleavage with an oligonucleotide directed against the 3' splice site was reproducibly less efficient than with the other oligonucleotides immediately after addition of nuclear extract (time zero) and remained essentially constant throughout the time course. Therefore, interaction of a factor(s) with the 3' splice site may be one of the earliest events in splicing.

Ruskin and Green (1985a) also analysed interactions of factors with

human β-globin IVS-1 using protection from RNase A cleavage as an assay. In contrast to the situation with the RNase H assay, no protection of the 5' splice site was observed, presumably because the 5' splice site binding factor does not interact tightly enough to resist the high concentrations of RNase A used. In the presence of ATP, five characteristic RNase A-resistant fragments were found, ranging from 18 to 60 nucleotides in length. All five fragments contained an unmodified 10 nucleotide RNase T1 fragment, 28 to 37 nucleotides upstream of the 3' splice site, which contains the site of branch point formation. These fragments were first detectable at 10 minutes and increased in abundance during the time course. In contrast to the regions around the branch point, the 16 nucleotides at the 3' end of IVS-1 were not present in any of the ATP-dependent protected fragments. This is consistent with the RNase H cleavage data. Some of the minor RNase A-resistant fragments were found to contain a branched nucleotide, and were, therefore, derived from lariat RNA. The protection of the branchpoint was dependent upon hydrolysis of ATP; other NTPs, or ATP analogues with non-hydrolysable β-γ or α-β bonds would not substitute. However, dATP could substitute for ATP. Hence, the nucleotide requirements for branch point protection were identical to those for the first step of splicing (Krainer *et al.*, 1984; Ruskin *et al.*, 1984; Grabowski *et al.*, 1984). A truncated substrate, lacking the 5' splice, was still protected around the branch point; hence, association of a factor(s) with the branch point is not dependent on the presence of a 5' splice site. Deletion of the 3' splice site consensus sequence (both the polypyrimidine tract and the AG dinucleotide) abolished branch point protection, despite all the sequences protected in normal pre-mRNA being present. Thus, association of a factor(s) with the branch point is dependent upon the 3' splice site consensus sequence. The RNase H cleavage assay also revealed that the protection of 5' splice site apparent with normal pre-mRNA did not occur with this same deletion mutant, indicating that the stability of the association of a factor(s) with the 5' splice site is directly, or indirectly, dependent upon the 3' splice site consensus sequence. Protection of the branch point from RNase A digestion was abolished by both pre-treatment of the nuclear extract by heating (75°C for 10 minutes) or micrococcal nuclease treatment, indicating that the factors involved in branch point protection probably contain an essential protein and an essential nucleic acid component.

A RNase A protection assay has also been used to investigate interactions of factors with pre-mRNA derived from the adenovirus 2 major late transcription unit (Krämer, 1987b). After short incubation times (5 minutes) at 30°C, fragments mapping <u>upstream</u> of the branchpoint, about 30 to 55 nucleotides upstream of the 3' splice site were obtained. This protection appeared to be equivalent to the protection of the region 20 to 75 nucleotides

upstream of the 3' splice site and encompassing the branch site found in human β-globin IVS-1 (Ruskin and Green, 1985a) with respect to several parameters. First, the interaction only occured under splicing conditions i.e. at elevated temperature and in the presence of ATP. Second protection was observed early in the splicing reaction, before RNA cleavage. Third, the interaction was independent of the presence of a 5' splice site, but required the polypyrimidine tract. Fourth, binding to the β-globin intron was micrococcal nuclease sensitive, and protection of the AdML intron was dependent on an intact end of U2 snRNA.

However, in contrast to the situation with β-globin pre-mRNA, in the case of the AdML intron, none of the protected oligonucleotides included the branch point. Hence, given that the distance between the branch point and 3' splice site is longer in the β-globin intron (37 nucleotides) than in the AdML intron (24 nucleotides), it might be fortuitous that the branch point was protected in the β-globin pre-mRNA. It is possible that the recognition sequence for the component binding in this region is not specified by the branch point sequence itself, but by another, as yet unidentified feature. However, Ruskin and Green (1985a) noted that, with human β-globin IVS-1 RNA, when the normal branch point was mutated RNase A protection was seen at the cryptic site to be used.

After 10 minutes or more into the splicing reaction, additional oligonucleotides derived from the AdML intron 5' splice site were protected from RNase A digestion. This is in contrast to the human β-globin gene IVS-1, where no protection from RNase A cleavage was observed at the 5' splice site (Ruskin and Green, 1985a). However, protection from oligonucleotide directed RNase H cleavage was observed at the 5' splice site of the β-globin intron with similar kinetics to protection of the AdML intron 5' splice from RNase H cleavage. Stable binding of factors to the 5' splice site of the AdML intron also occured only on molecules with intact 3' splice sites as determined by RNase A protection (Krämer, 1987b) in agreement with the RNase H cleavage data of Ruskin and Green (1985a) on β-globin IVS-1.

Pre-incubation of the extract with excess unlabelled RNA lacking the 5' splice site led to efficient competition for factors interacting with the branch point region and at the 5' splice site of the AdML intron, whereas a competing RNA that contained only the 5' splice site sequences had no effect on the interactions with the pre-mRNA (Krämer, 1987b). Hence, stable association with the 5' splice site seems to require prior binding of components in the branch point region.

When splicing complexes fractionated on sucrose gradients were digested with RNase A, only the branch point region was protected in the 35S complex. The 5' splice site only became resistant to RNase A in the mature 50S

spliceosome (Frendewey and Keller, 1985).

Degradation of specific regions in U1, U2 and U4 snRNA with complementary oligonucleotides and RNase H has been used to analyse involvement of the U snRNPs in substrate RNA protection (Krämer, 1987b). Removing the 5' end of U2 RNA abolished protection of the 5' splice site and greatly reduced branch point region protection. In contrast, after degradation of internal loops in U2 or U4 RNA, or after removal of the 5' end of U1 RNA, branch point protection was unaffected. Protection of the 5' splice site was, however, abolished. The requirement for an intact 5' end of U1 for 5' splice site protection is not unexpected, as this region has been shown to interact, at least in part, by base pairing with the 5' splice site (Zhuang and Weiner, 1986). The observation that degradation of U4 RNA does not prevent an association in the branch point region implies that the U4/U6 snRNP is essential for a step that occurs after binding of U2 snRNP to the branch point region, and is consistent with data from spliceosome fractionation on density gradients and native gels (Bindereif and Green, 1987; Konarska and Sharp, 1987; Zillmann *et al.*, 1988). Since the 5' splice site is not protected after degradation of the U4 RNA loop this particle would seem to be involved in the process leading to stable interactions with the 5' splice site. The results with U2 are more surprising, given that the internal loop of U2 digested in this assay includes the site proposed to interact by base pairing with the branch point region (Keller and Noon, 1984; Black *et al.*, 1985) in a manner analogous to the base pairing interaction between snR20 (yeast U2) and the TACTAAC box in yeast introns (Parker *et al.*, 1987), and yet digestion of this region did not prevent interaction with the branch point region. Surprisingly, Black *et al.* (1985) had previously reported that degrading this loop destabilised the entire U2 snRNP particle.

Sucrose density gradient centrifugation studies (D. Frendewey and W. Keller, cited in Krämer, 1987b) support this data. Formation of the pre-splicing complex was prevented by degradation of the 5' end of U2 RNA, whereas the pre-splicing complex, (but not the mature spliceosome) formed in the absence of an intact 5' end of U1 RNA and after cleavage of the internal loop sequences of U2 and U4 RNAs. Hence, these sequences must be involved in the conversion of the pre-splicing complex to the splicing complex. Native gel electrophoresis data are also consistent: Zillmann *et al.* (1988) reported that RNase H cleavage of the 5' end of U2 snRNA prevented all splicing complex assembly, but if U2 nucleotides 28 to 42 were cleaved complex A was formed, but not the mature splicing complex B.

A combination of protection from RNase T1 digestion and immunoprecipitation using snRNP specific antibodies has also been used. This technique was first used to show that purified U1 snRNPs could selectively

bind the 5' splice site of mouse β-globin IVS-1 (Mount *et al.*, 1983). A 15-17 nucleotide region, including the 5' splice site and extending 3 nucleotides into exon-1, remained undigested by RNase T1 and complexed with the snRNP such that it could be co-precipitated by antibodies directed against the U1 snRNP.

Sites on a human β-globin IVS-1 substrate RNA associated with U1 and U2 snRNPs have also been investigated by adding RNase T1 along with anti-(U1)RNP or anti-(U2)RNP antibodies to *in vitro* splicing reactions at various times and characterising the immunoprecipitated RNase resistant fragments (Black *et al.*, 1985; Chabot and Steitz, 1987). At time zero, anti-(U1)RNP antibodies immunoprecipitated a single 15 nucleotide fragment containing the 5' cleavage site (Black *et al.*, 1985) identical to that found with purified U1 snRNP by Mount *et al.* (1983). Only at later times (60 minutes) were other fragments precipitated by anti-(U1)RNP antibodies (see below).

At time zero, anti-(U2)RNP antibodies do not detectably immunoprecipitate any specific fragments, despite immunoprecipitation without RNase digestion revealing a substantial amount of human β-globin pre-mRNA to be associated with U2 snRNP at time zero (Black *et al.*, 1985). At 60 minutes, however, a variety of fragments were immunoprecipitated. These fragments corresponded to protection of a region stretching from the 3' splice site to upstream of the branch point and to an extended region around the 5' splice site. Interestingly, at 60 minutes all of the fragments immunoprecipitated by anti-(U2)RNP antiserum were also precipitated by anti-(U1)RNP antiserum, although at a lower level. The simplest explanation of these results is that U1 snRNP binds to the 5' splice site very early in splicing (at time zero), even at 0°C, and can do so even in the absence of binding of other components. This is, therefore, probably one of the first events in splicing. Stable U2 snRNP binding occurred with a region of pre-mRNA that included the branchpoint at later times in splicing reactions, although a weaker, RNase T1 sensitive, or possibly non-specific interaction must occur with the pre-mRNA at earlier stages. The stable interaction of U2 snRNP was dependent on both elevated temperature and ATP and did not occur with isolated U2 snRNPs. Hence, some other factor(s) must bind to the RNA before U2 itself can bind.

A direct, or indirect interaction between U1 and U2 snRNPs probably occurs during spliceosome assembly as evidenced by the observation that anti-(U1)RNP and anti-(U2)RNP antibodies immunoprecipitated a common set of protected pre-mRNA fragments from active splicing reactions.

In a more detailed analysis Chabot and Steitz (1987) mapped interactions with a truncated human β-globin IVS-1 containing transcript which possessed only 3 nucleotides of the 3' exon and gave rise to accumulation of splicing intermediates *in vitro*. Using this substrate, U2 snRNP has been shown to

associate with the branch point in the lariat intermediate. Two branched RNase T1 resistant fragments were immunoprecipitated by anti-(U2)RNP antibody and weakly by anti-(U1)RNP antibody. Both included 16 nucleotides upstream of the branch site and extended to the 3' splice junction, but included different amounts of sequence from the 5' end of the intron (12-13 nucleotides or 24-25 nucleotides respectively). It was evident that, following lariat formation, less protection upstream, but more downstream of the branch site was detected. This implies a conformational change of U2 snRNP or associated components at the branch site, or addition or loss of factors on forming the branch. Since anti-(U1)RNP antibody immunoprecipitated fragments containing the branch, U1 snRNP would seem, on the basis of this data, to be an integral component of the spliceosome. Anti-(U2)RNP antibody also immunoprecipitated a 28 nucleotide fragment mapping to exon 1, which terminated 3 nucleotides upstream of the 5' splice site. As the kinetics of the appearance of this fragment followed those of the two branched fragments it was thought to derive from the exon 1 splicing intermediate. This interaction may be at least in part responsible for holding the exon 1 intermediate into the spliceosome. Many of the minor bands produced from the truncated substrate by RNase T1 protection and immunoprecipitation were also characterised. Most of the minor fragments mapped to either the 5' splice site region or the capped 5' end of the transcript. One of the protected fragments mapping to the capped 5' end of the RNA co-sedimented with the spliceosome and could represent an interaction with a spliceosome component. Interestingly, however, pre-incubation of the extract with 12μM cap analogue (GpppG) did not prevent immunoprecipitation of the 5' terminal fragments.

RNase T1 digestion and immunoprecipitation was also performed on fractions from glycerol gradients (Chabot and Steitz, 1987). The branched fragments were immunoprecipitated by anti-(U2)RNP antibody exclusively from the spliceosome fractions. One of the major branch site fragments was found to be at the top of the gradient, in addition to in the spliceosome, hence, U2 snRNP must be associated with the branch site in a relatively small complex. In contrast, most other fragments, corresponding to protection of the 5' splice site (both local and extended), the 3' splice site, the 5' exon immediately upstream of the splice site consensus, and the capped 5' end of the transcript were all immunoprecipitated by anti-(U2)RNP antibody primarily from the spliceosome fractions.

If immunoprecipitation was with anti-(U1)RNP antibody, immunoprecipitation of fragments from the branch point region was found only in spliceosome fractions and in lower yield than with anti-(U2)RNP antibodies. Also, in contrast to with anti-(U2)RNP antibodies, a short protected fragment covering the 5' splice site, which was present only in the

spliceosome region with anti-(U2)RNP antibodies, was also found in fractions from the top of the gradient with anti-(U1)RNP antibodies.

Comparable results were obtained using this technique to map interactions with human β-globin IVS-1 in fractions from sucrose gradients (Bindereif and Green, 1987). RNase T1 digestion of the 20S peak generated only a small (15 nucleotide) fragment mapping to the 5' splice site which was precipitated by anti-Sm and anti-(U1)RNP antibodies. In the 60S spliceosome complex, RNase T1 resistant fragments, immunoprecipitated by anti-Sm or anti-(U1)RNP antibodies mapped to the branch point/3' splice site region and to the 5' splice site, extended protection of this site now being visible.

Oligonucleotide directed RNase H cleavage of the 5' end of U2 RNA led to complete abolition of branch point protection and extended 5' splice site protection whether assayed with anti-(U1)RNP or anti-(U2)RNP antibodies (Chabot and Steitz, 1987). The short, but not the extended 5' splice site fragment was, however, still detected with anti-(U1)RNP antibody, but not with anti-(U2)RNP antibody. Oligonucleotide directed RNase H cleavage of U4 and U6 snRNAs resulted in an identical banding pattern to normal, except for the absence of fragments derived from splicing intermediates. Hence, the snRNPs must be associating normally with the 5' splice site, 3' splice site and branch point.

A 25 nucleotide deletion, removing the 3' splice site (from -23 to +3 relative to the 3' splice site) drastically reduced protection of the branch point. However, in contrast to the RNase H protection assay of Ruskin and Green (1985a), local protection at the 5' splice site by U1 snRNP was still found, although extended protection was not seen. Hence, it seems likely that the RNase H assay only detected the interaction giving rise to extended RNase T1 protection and immunoprecipitation. An AG to GG mutation in the conserved 3' splice site dinucleotide also reduced branch point binding to about the same extent as it reduces step 1 of splicing.

Use of a mutant which exclusively splices to a cryptic AG created 21 nucleotides upstream of the normal 3' splice junction gave a protection pattern which strongly indicated that the position of U2 snRNP was unaltered on this mutant transcript.

The length of intron protected by snRNPs totalled about 65 nucleotides. This estimate is very close to the minimum intron length determined for efficient and correct splicing of a rabbit β-globin transcript *in vivo* (Wieringa *et al.*, 1984) and this probably represents the minimum sequence necessary for snRNP interactions with the pre-mRNA substrate.

This technique has also been used to show that another snRNP interacts with the 3' splice site (Chabot *et al.*, 1985). Fragments corresponding to sequences extending upstream from the 3' cleavage site of human β-globin

IVS-1 were found to be immunoprecipitable by either anti-(Sm) or anti-(TMG) antibodies at time zero and 0°C. Hence, the interacting factor must bind very quickly and be an snRNP, but not U1 or U2, since neither anti-(U1)RNP or anti-(U2)RNP antibodies would precipitate these fragments at 0°C. The binding of this snRNP was also not ATP dependent. U5 was suggested to be the snRNP interacting with the 3' splice site, since binding was surprisingly insensitive to treatment of the extract with micrococcal nuclease and U5 snRNP is the only abundant Sm snRNP with a TMG cap that is equally resistant to micrococcal nuclease. This suggestion of the involvment of U5 snRNP at such an early step contrasts starkly with the evidence from the use of native gels, which implies that U5 snRNP binds after U2, not before, in an ATP dependent manner and probably in the form of a U4/5/6 multi-snRNP particle (Konarska and Sharp, 1987).

1.4.6. Electron microscopy

The Miller chromatin spreading technique has been applied to the analysis of the RNP structure of nascent hnRNA transcripts by electron microscopy (E.M.). In these preparations transcripts appear as smooth RNP fibrils with a width of about 5nm. Stable RNP particles, about 25nm in diameter, interupt this fibril and occur at non-random and specific sites on the transcripts of a given gene (Beyer *et al.*, 1980). In some cases, pairs of adjacent particles stably associate to form approximately 40nm particles, looping out the RNA between them (Beyer *et al.*, 1981), and in some cases the loops were specifically removed in a process that may represent splicing. The 40nm particle was thought to be the spliceosome. In support of this hypothesis, 25nm particles have been observed to occur specifically at splice junctions on 3 genes identified in chromatin spreads (Osheim *et al.*, 1985; Beyer and Osheim, 1988). Beyer and Osheim (1988) indicated additional evidence that these particles were indeed spliceosome and pre-splicing complexes, including 1) frequent observation of the phenomenon; 2) a range of loop sizes consistent with known intron lengths; 3) RNP particles were found at the ends of segments to be excised, but not in between, consistent with the results of Black *et al.* (1985) and Ruskin and Green (1985a); 4) the 3' splice site was required for 40nm particle assembly; 5) more than one spliceosome could form per transcript consistent with the results of Christofori *et al.* (1987); 6) there was a preferred, but non-obligatory order of loop excision (intron removal) consistent with the work of Zeitlin and Efstratiadis (1984); 7) the bulk of the splicing particle was removed with the excised intron, consistent with the data of Konarska and Sharp (1987); 8) heat shock both disrupted splicing (Yost and

Lindquist, 1986) and caused a paucity of loops on transcripts from heat shocked *Drosophila* embryos.

By analysing unidentified transcription units from *D. melanogaster* early embryos, Beyer and Osheim (1988) identified a series of events which were in agreement with results from *in vitro* splicing systems, but differed in the more rapid completion of *in vivo* splicing as observed by E.M. The first observable structure related to splicing was a small, approximately 10nm particle at the 5' splice site, seen in about 50% of possible cases, before the 3' splice site was synthesised. This particle may be the U1 snRNP whose association with the 5' splice site is very rapid *in vitro* (Mount *et al.*, 1983; Black *et al.*, 1985). The next structure observed was an approximately 25nm particle at the 3' splice site junction/branch point. This was the first stable particle and could be seen within about 13 seconds of 3' splice site synthesis. A 25nm particle formed at the 5' splice site within about 40 seconds of 3' splice site particle formation. The authors speculated that the dependence of the 5' splice site particle on 3' splice site synthesis probably indicates that they are in contact in the cell and only become separated on spreading. This would correlate well with the extended 5' splice site protection from Rnase T1 observed by Chabot and Steitz (1987) to be dependent upon 3' splice site sequences. Intron loop formation occured within 2 minutes of 3' splice site synthesis and may represent covalent lariat formation stabilising the loop structure under spreading conditions. Loop formation was followed within a minute by loop removal, the bulk of the spliceosome particle being removed from the nascent transcript as suggested by Konarska and Sharp (1987).

In some cases, introns were removed from the 5' end of the transcript before the introns at the 3' end were synthesised, supporting the idea of a "first-come-first-served" principle of splice site selection as proposed by Aebi *et al.* (1986), and a role for splice site proximity (Reed and Maniatis, 1986). An example of exon skipping was also observed, hence alternative splicing can be fitted into the framework of this model.

1.4.7. Chemical modification/interference study of spliceosome assembly

A chemical modification/interference assay has been used to determine the pre-mRNA sequence requirements for yeast *in vitro* spliceosome assembly and splicing (Rymond and Rosbash, 1988). The assay involved labelling the 3' ends of substrate RNA, synthesised *in vitro* and containing the rp51A intron, with [^{32}P]pCp and RNA ligase, and then introducing on average less than one chemical modification per molecule using diethyl pyrocarbonate (to modify

purines by N-6,7 carbethoxylation of A and N-7 carbethoxylation of G) or hydrazine (to modify pyrimidines by base removal at C and U).

Spliced product, lariat intermediate, or precursor RNA from each of the three ATP-dependent splicing complexes found on native gels (Pikielny *et al.*, 1986) was then isolated. Aniline was used to introduce strand cleavage at the sites of modification and the RNA was fractionated by PAGE to determine the site of modification. Modifications which inhibited the assembly of splicing complexes or either step of splicing could thus be identified.

Modification of any of the nucleotides within the 5' splice site consensus and branch point (TACTAAC box) was found to inhibit spliceosome assembly and/or splicing. The interference pattern of the 5' splice site and TACTAAC box lesions increased as spliceosome assembly proceeded from complex III, to complex I, to complex II (Pikielny *et al.*, 1986). This implied that these sequence elements may play multiple roles at different stages of spliceosome assembly and splicing.

Interestingly, hydrazine lesions (which remove the pyrimidine base) in the 5' splice site strongly inhibited complex III formation, whilst hydrazine lesions within the TACTAAC box were relatively weak inhibitors of complex III formation. This may reflect the fact that snR19 (yeast U1) base pairing with the 5' splice site is important for early complex formation (Séraphin *et al.*, 1988), whilst snR20 (yeast U2) base pairing with the TACTAAC box (Parker *et al.*, 1987) is not, base pairing only occuring at a later stage. This idea has been supported by some work in metazoa (Krämer, 1987b; Zillmann *et al.*, 1988).

Modifications at less conserved positions also inhibited spliceosome assembly and/or splicing. Analysis of the spliced exons revealed that modifications of the last three nucleotides of exon-1 and the first two nucleotides of exon-2 inhibited mRNA formation, despite there being no effect of modifications of 3' splice site nucleotides on all previous stages. Lesions in exon-1 could inhibit either step of splicing, however, the modifications in exon-2 must affect step 2 alone, since these modifications did not inhibit formation of the lariat intermediate.

Modification of a TACTAAC-like sequence upstream of the branchpoint was found to inhibit the rate of spliceosome assembly, and it was suggested that this sequence may play a role in modulating the efficiency of splicing by facilitating spliceosome assembly. Modulation could be the result of splicing factors binding to this site stabilising similar interactions at the authentic TACTAAC box through cooperative interactions, or, as suggested by Cellini *et al.* (1986), by acting as a sink for limiting splicing factors: factors initially bound at this sequence subsequently migrating along the RNA to the authentic TACTAAC box.

1.5. NON-SMALL NUCLEAR RIBONUCLEOPROTEIN SPLICING FACTORS IDENTIFIED BY BIOCHEMICAL FRACTIONATION AND *IN VITRO* COMPLEMENTATION

A protein, designated IBP (Tazi *et al.*, 1986) or 70kd protein (Gerke and Steitz, 1986) has been identified which specifically recognises the 3' splice site of mammalian pre-mRNA introns. Tazi *et al.* (1986) used SDS-PAGE to separate proteins from partially purified HeLa cell snRNPs, fractionated by isopycnic centrifugation in caesium chloride at high Mg^{2+} concentrations. The proteins were electrophoretically transfered to nitrocellulose filters and probed with radioactive RNA. An intron binding protein (IBP) of approximately 100kd was detected, that specifically recognised the 3' splice site region of mammalian pre-nRNAs. Blocking binding by hybridising oligonucleotides complementary to the human β-globin pre-mRNA revealed the binding site to consist of both the polypyrimidine tract and the 3' splice site AG dinucleotide, but not the branch point or lariat structure. Although IBP could bind the 3' splice site in isolation, not relying upon snRNA-pre-mRNA interactions, it was found to co-elute with U5 snRNP when fractionated on DEAE-sepharose, and be immunoprecipitated by anti-(Sm) antibodies.

Gerke and Steitz (1986) fractionated HeLa cell nuclear extracts, and found a component which could be separated from all major snRNAs, and that allowed immunoprecipitation of a 19 nucleotide RNase T1 protected from the 3' splice site by anti-(Sm), but not anti-(TMG) antibodies. Protein blotting, as above, revealed that a 70kd pre-mRNA-binding protein was present in all fractions that exhibit 3' splice site binding activity, and immunoblotting showed this protein to have an Sm epitope. The 70kd protein was found to preferentially bind sequences upstream of the 3' splice junction and downstream of the branch point. RNase T1 digestion and immunoprecipitation of the 3' splice site fragment was found to be reduced by mutation of the 3' splice site AG dinucleotide to GG even when the 3' splice site binding component was separated from snRNPs. Under certain conditions (low Mg^{2+} concentrations), the 3' splice site binding component associated with snRNPs as judged by cofractionation during chromatography and immunoprecipitation by anti-(TMG) antibodies. The IBP (Tazi *et al.*, 1986) and the 70kd protein (Gerke and Steitz, 1986) are generally believed to be one and the same.

Hence, this factor would seem to be an Sm polypeptide, possibly associated with the U5 snRNP, which has an integral 3' splice site binding activity. An Sm snRNP, possibly U5, has previously been reported to bind very early in splicing reactions to the 3' splice site region in an ATP independent manner

(Chabot *et al.*, 1985). This protein may be the factor that mediates this interaction, rather than direct RNA-RNA base pairing.

A protein factor, designated U2AF, required for U2 snRNP binding to the branch point of mammalian introns, and for spliceosome assembly, has also been identified (Ruskin *et al.*, 1988). U2 snRNP, partially purified either by caesium chloride density equilibrium centrifugation, or by use of S100 fractions of lysed HeLa cells, did not bind the branch site, as assayed by RNase A protection and anti-(Sm) immunoprecipitation, or by visualisation of splicing complexes A or B on native gels as Konarska and Sharp (1986). However, U2 snRNPs in these fractions could bind the branch point and form splicing complexes A and B if complemented with micrococcal nuclease (MN) treated nuclear extract, which is itself inactive in branch point protection. The MN-resistant factor did not appear to be a known U2 snRNP structural polypeptide.

U2AF could thus be assayed by its ability to enable binding of partially purified U2 snRNPs to the branch point. U2AF was found to have the properties of a protein by the criteria of its resistance to MN-digestion, its buoyant density, heat lability and sensitivity to N-ethylmaleimide. It is not known if U2AF is a single component. U2AF could be distinguished from certain other proteins implicated in splicing, namely hnRNP C protein (Choi *et al.*, 1986), and the 70kd RNA binding protein (Tazi *et al.*, 1986; Gerke and Steitz, 1986). Therefore, U2AF seems to be distinct from these two species. Using RNA substrate exclusion and competition assays, U2AF was shown to bind to the 3' splice site region. The polypyrimidine tract was found to be essential for efficient U2AF binding, and with mutant RNAs lacking the last 4 nucleotides of the intron (which includes the highly conserved AG) or with the 3' AG of the intron mutated to GG, U2AF binding was also significantly reduced, indicating that the 3' AG dinucleotide is also important for U2AF binding. U2AF, and not 70kd RNA binding protein, was found to be the limiting component in these competition experiments, and addition of 70kd RNA binding protein could not increase the level of U2 snRNP binding.

U2AF may explain, why in mammalian pre-mRNA splicing, the major constraint for branch point selection appears to be distance from the 3' splice site, and why the 3' splice site is necessary for stable interaction of U2 snRNP with the branch point (Ruskin and Green, 1985a; Chabot and Steitz, 1987). U2 snRNP may need to interact with both the pre-mRNA branch site and U2AF bound to the 3' splice site.

In yeast the major determinant of branch point selection is the UACUAAC box at the branch site, and not distance from the 3' splice site. The yeast 3' splice site is not required for splicing complex formation (Rymond *et al.*, 1987) or for step 1 of splicing (Rymond and Rosbash, 1985). Hence, Ruskin *et al.*

(1988) speculate that yeast may lack a functional U2AF equivalent. However, Legrain *et al.* (1988) have performed experiments which imply that yeast splicing also requires the assembly of a *trans*-acting factor(s) that stably binds to the pre-mRNA prior to snR20 (U2) snRNP binding. These experiments involved depletion of U2 snRNA in yeast extract by oligonucleotide-directed RNase H digestion. This depleted extract could not support splicing or splicing complex formation. Using this extract chase experiments were performed which showed that a pre-mRNA substrate is committed to the spliceosome assembly pathway in the absence of functional U2 snRNP. Both the 5' splice site consensus sequence and the TACTAAC box sequence were necessary for this commitment step.

Additional splicing factors have also been identified by biochemical fractionation and *in vitro* complementation studies (Krainer and Maniatis, 1985; Furneaux *et al.*, 1985; Krämer and Keller, 1985; Perkins *et al.*, 1986; Krämer *et al.*, 1987). However, none of these activities have been substantially purified, and their roles in splicing have not been determined. Krämer *et al.* (1987) obtained 6 fractions from chromatographic fractionation of a HeLa cell nuclear extract which, when combined, support efficient *in vitro* splicing. Each of the fractions contains a component, or components, essential for the first step of splicing. Two of the fractions were enriched in snRNPs U1, U2, U4/U6 and U5. The remaining four fractions appear to contain protein factors, as suggested by their resistance to micrococcal nuclease, but could conceivably contain a micrococcal nuclease resistant snRNA e.g. U5 (Chabot *et al.*, 1985). One of these fractions (SF2) was purified earlier (Krämer and Keller, 1985) and is a protein with a molecular weight of approximately 50kd.

Three fractions (designated Ia, Ib and II), isolated from HeLa cell nuclear extracts by chromatographic separation, have also been shown to be required for splicing *in vitro* (Furneaux *et al.*, 1985; Perkins *et al.*, 1986). Fraction II alone is inactive in any step of splicing, but an ATP dependent 30S complex sediments through sucrose gradients when it is incubated with precursor RNA. Fractions Ib and II combined, result in production of a 55S complex containing splicing intermediates, but no products. Fraction Ia must also be present for production of spliced RNA. Fraction II is micrococcal nuclease sensitive, and therefore, contains essential essential RNA components, whereas fractions Ia and Ib are not, and so are probably protein factors. Fraction Ia must contain one or more components essential for step II of splicing.

Krainer and Maniatis (1985) identified five distinct factors essential for splicing by complementation studies. The first of these (SF1) was found to be sensitive to micrococcal nuclease and, therefore, to contain essential snRNAs. A second factor (SF2) was defined as a micrococcal nuclease resistant factor present in nuclear extract, but absent from an S100 extract. A third factor

(SF3) could be preferentially inactivated by relatively mild heat-treatment. Two further distinct factors (SF4A and SF4B) were identified by chromatographic fractionation of nuclear extract. SF1, SF2 and SF4B all contained factors essential for step 1 of splicing. However, SF3 and SF4A were only required for step 2. The nomenclature does not imply identity with factors of the same name identified by Krämer *et al.* (1987).

Mayeda *et al.* (1986) have also chromatographically fractionated HeLa cell nuclear extract and assayed the fractions for their ability to bind short RNA transcripts carrying splice junction and branch point sequences. Binding of all three sequences was found in the snRNP containing fraction. However, even stronger binding activities for the 3' splice site were found in the flow-through fraction where snRNAs were absent, and this fraction could stimulate binding of 5' splice site RNA by purified U1 snRNP.

Because of the different fractionation procedures used by different groups, correlation of the different factors identified by the different groups is difficult, and will, for the most part, have to await further experimentation.

Cheng and Abelson (1986) fractionated a yeast whole cell extract by ammonium sulphate fractionation and chromatography on heparin-agarose. Three fractions were obtained, each alone being devoid of any splicing activity. Fractions I and II, when mixed, could support the first step of splicing. Addition of fraction III allowed the splicing reaction to go to completion. Hence, fraction III must contain a component essential for step 2 of yeast splicing. Micrococcal nuclease treatment of either fraction I or II abolished splicing activity, indicating that these fractions contain essential RNA moieties.

1.6. GENETICALLY DEFINED FACTORS

Hartwell (1967) first identified a set of temperature-sensitive *rna* mutants (*rna1-rna11*) in yeast, in the course of screening for temperature-sensitive strains defective in macromolecule synthesis. Ten complementation groups (*rna2-rna11*) were initially described, that exhibited a rapid cessation of net RNA accumulation following a shift to the restrictive temperature, but continued to accumulate protein for several hours (Hartwell *et al.*, 1970). However, *rna10* and *rna11* were subsequently found to belong to the same complementation group (Lustig *et al.*, 1986). These mutants were found to be defective in rRNA synthesis, but not net tRNA or mRNA synthesis, and the defect in rRNA synthesis was at the level of processing the pre-rRNA, not transcription (Warner and Udem, 1972). Levels of nearly all ribosomal protein(rp) mRNAs were also reduced (Gorenstein and Warner, 1976; Warner

and Gorenstein, 1977). Most other mRNAs were unaffected, and it was proposed that ribosomal RNA must be assembled into a pre-ribosome before processing, and defects in rp synthesis could therefore affect pre-rRNA processing (Gorenstein and Warner, 1976). Many yeast rp genes contain introns, and rp pre-mRNA, containing introns, accumulated at the non-permissive temperature in the *rna2-rna11* mutants (Fried *et al.*, 1981; Rosbash *et al.*, 1981; Larkin and Woolford, 1983). Accumulation of pre-mRNA from the intron containing actin and *mat* a1 genes also occured (Teem *et al.*, 1983; Miller, 1984). These are two of the very few non-rp genes to contain introns in yeast. Thus, the apparent specificity of the effect on rp genes is a consequence of the rarity of introns within yeast genes.

More direct evidence that many of the *RNA* gene products are likely to code for products directly involved in and essential for mRNA splicing was found by Lustig *et al.* (1986), who made extracts from wild type cells and cells carrying each of the nine *rna* mutants *rna2-rna10/11* (hereafter referred to as *rna11*). Extracts from *rna2*, *rna3*, *rna4*, *rna5*, *rna7*, *rna8* and *rna11* mutant strains were more sensitive to heat inactivation of their *in vitro* splicing activity than extracts from wild type cells. Heat inactivation usually resulted in loss of an exchangeable component, as demonstrated by the fact that most combinations of heat-inactivated extracts from differerent mutants could complement each other. Also, heat-inactivated extracts could be complemented by at least one of the three extract fractions described by Cheng and Abelson (1986), and which were individually inactive in splicing. In three cases (*rna2*, *rna5* and *rna11*), confirmation that heat inactivation of *rna* mutant extracts resulted in the specific loss of a splicing component was obtained by reversion, tetrad, and *in vitro* complementation analyses.

Heat treatment of *rna6* and *rna9* extracts did not inhibit splicing *in vitro*, but this does not necessarily preclude a direct role for these genes in splicing.

The products of the *RNA* genes must be required for early stages of splicing since none of the heat-inactivated extracts could carry out step 1 of splicing. In addition, most heat-inactivated *rna* mutant extracts were found not to form spliceosomes, as judged by fractionation on glycerol gradients (Lin *et al.*, 1987), and therefore, these *RNA* gene products are likely to be involved in spliceosome formation. In contrast, heat inactivated *rna2* extracts did form a splicing dependent 40S complex, which contained uncleaved pre-mRNA exclusively. However, the pre-mRNA in the 40S complex could be spliced in the presence of ATP and complementing extracts. Hence, the *RNA2* gene product would seem to be a factor extrinsic to the spliceosome. Thus, it appeared that splicing could be divided into three distinct steps A) spliceosome formation, B) the 5' splice site reaction, C) the 3' splice site reaction. Spliceosome formation (step A) requires the products of the *RNA3*, *RNA4*,

RNA5, *RNA7*, *RNA8* and *RNA11* genes, and step B requires the *RNA2* gene product. The *RNA2*, *RNA3*, *RNA4*, *RNA8* and *RNA11* genes have been cloned and have been shown to encode proteins (see Lossky *et al.*, 1987, and references therein). The roles of the products of two of these genes, *RNA8* and *RNA11* have been studied in more detail (Lossky *et al.*, 1987; Chang *et al.*, 1988).

The *RNA8* gene encodes an unusually large (260kd) protein (Jackson *et al.*, 1988). Immunoprecipitation with antibodies raised against this protein was used to show that it exists stably associated with snRNPs containing snR7L and snR7S RNAs (yeast U5) (Lossky *et al.*, 1987). SnR7L and snR7S differ only in the length of their 3' termini. It is not known whether snR7L and snR7S are in the same particle. When yeast extract was incubated in the presence of ATP, antibodies against the RNA8 protein, in addition to precipitating snR7, also co-precipitated snR6 (yeast U6) and snR14 (yeast U4). The formation of an snRNP complex containing snR7, snR6 and snR14, which binds as a single entity in the course of spliceosome formation, is consistent with models for spliceosome assembly in yeast and metazoa proposed previously (Pikielny *et al.*, 1986; Konarska and Sharp, 1987). Indeed, Konarska and Sharp (1987) detected an analogous U4/U5/U6 snRNP complex in HeLa nuclear extracts. However, this mammalian complex dissociated on incubation with ATP. Anti-(RNA8) antisera also precipitated splicing intermediates and product RNAs from yeast *in vitro* splicing reactions, indicating the presence of RNA8 protein in spliceosomes.

It has also been shown that the RNA11 protein is associated with the yeast spliceosome (Chang *et al.*, 1988). *In vitro* transcription and translation of the *RNA11* gene *in vitro* yielded a 32kd protein. This protein contains a potential zinc finger motif, as frequently observed in nucleic acid-binding proteins (Evans and Hollenberg, 1988). A heat treated *rna11* extract could be complemented by RNA11 protein, but not by truncated RNA11 or *in vitro* translated globin. The complementation assay, however, only worked when heat inactivation was performed in the presence of the RNA11 protein. During *in vitro* complementation, the RNA11 protein was associated with the 40S spliceosome on glycerol gradients, and remained associated despite stringent treatments with either heparin or high salt. Anti-(Sm) and anti-(TMG) antibodies were also found to coprecipitate the RNA11 protein and the RNA splicing intermediates, and immuno-electronmicroscopy of purified spliceosomes also suggested that RNA11 protein is part of the spliceosome. During *in vitro* complementation, the RNA11 protein was found to be associated with a 30S particle. Two seperate complexes were, in fact, found to sediment at 30S. One of these, the [^{35}S]30S complex, was identified using ^{35}S labelled RNA11 protein, and formed only when *rna11* extract (not wild type, *rna2* or *rna5*) was used. This suggests that wild type RNA11 protein in this 30S complex may not

exchange with exogenously added wild type RNA11 protein, whereas mutant protein will exchange due to the decreased stability of this complex. Formation of the [^{35}S]30S complex was ATP-independent and did not require exogenously added substrate RNA. Using ^{32}P-labelled substrate RNA, a distinct, ATP-independent, 30S complex, the [^{32}P]30S complex, was detected. The [^{32}P]30S complex formed in active yeast extracts and, in contrast to the [^{35}S]30S complex, required the presence of a functional pre-mRNA (Brody and Abelson, 1985; Vijayraghavan *et al.*, 1986; Chang *et al.*, 1988). The [^{32}P]30S complex may be analogous to the mammalian 35S, ATP-independent complex, identified by Frendewey and Keller (1985). Both 30S complexes are distinct from the 25S complex which form on substrates lacking all required splicing signals and probably represents the yeast hnRNP particle.

The [^{35}S]30S complex may represent a multi-snRNP complex formed in the yeast extract. This complex may be that containing snR14 (yeast U4), snR6 (yeast U6) and snR7 (yeast U5) identified by Cheng and Abelson (1987), and analogous to the 25S multi-snRNP particle identified in HeLa cell nuclear extracts containing U4, U5 and U6 snRNAs (Konarska and Sharp, 1987). Lossky *et al.* (1987), as described above, have also found evidence for an association of snR7 with snR14 and snR6, but this association was ATP-dependent.

Splicing of wild type extracts has also been found to be inhibited by anti-(RNA11) antibodies, and the same antibodies have been used to localise the RNA11 protein to the periphery of the yeast nucleus, and this is possibly the region of the nucleus where splicing occurs. Immunofluorescence studies indicated that hnRNPs and snRNPs also predominate in the periphery of rat liver nuclei (Fakan *et al.*, 1984). Transcription of pre-mRNA may also be localised in this region as evidenced by the apparent localisation of RNA polymerase II (Clark and Hamkalo, 1987) and yeast heat shock transcription factor (Wiederrecht *et al.*, 1987) to peripheral regions of mouse and yeast nuclei respectively.

1.7. OTHER FACTORS INVOLVED IN, OR POSSIBLY RELATED TO SPLICING

1.7.1. Heterogeneous nuclear ribonucleoprotein polypeptides

Much data has been presented to indicate that pre-mRNAs in nuclei do not normally occur as naked polynucleotides, but are found complexed with specific proteins to form so called heterogeneous nuclear ribonucleoprotein (hnRNP) complexes (Reviewed in Dreyfuss, 1986; Chung and Wooley, 1986).

HnRNP is probably assembled during transcription on the nascent pre-mRNA and the RNA is probably maintained in this form until the fully processed messenger-RNP (mRNP) is exported to the cytoplasm. Electron microscopy reveals hnRNP to appear as a "beads-on-a-string" structure of 200-300Å diameter, repeating monomer units (monoparticles), linked by RNase sensitive strands. The constituents of the monomer particles are approximately 700 nucleotides of pre-mRNA and numerous proteins (Conway *et al.*, 1988). Immunopurification of hnRNP particles from HeLa cells revealed at least 24 polypeptides, ranging from 34-120kd (Piñol-Roma *et al.*, 1988). All the RNP proteins sequenced to date, including hnRNA, mRNA, snRNA, and pre-rRNA binding proteins, contain repeating 100 amino acid domains, which probably constitute RNA binding domains. Approximately in the middle of these repeating domains, is a segment of 8 amino acids, which is especially well conserved, and is termed the RNP consensus sequence (RNP1) (Adam *et al.*, 1986; Swanson *et al.*, 1987). A second, less strictly conserved element (RNP2), which is mostly aliphatic and aromatic in character, is located about 30 amino acids, amino terminal to RNP1.

Each monoparticle contains 6 major core proteins of approximately 30-40kd, which are, in fact, 3 pairs of closely related polypeptides, A1 and A2, B1 and B2, C1 and C2 (Beyer *et al.*, 1977). These proteins represent a high proportion of 40S particle protein mass.

In addition to a structural role in packaging hnRNA to perform functions such as preventing tangling of transcripts, compacting hnRNA, facilitating RNA strand displacement and release from template DNA, hnRNPs are more directly involved in the processing of pre-mRNA.

An essential role for at least two hnRNP proteins has been indicated by making use of monoclonal antibodies directed against hnRNP proteins (Choi *et al.*, 1986). *In vitro* splicing in HeLa nuclear extract was inhibited by a monoclonal antibody directed against the hnRNP C proteins (antibody 4F4). This inhibition was at an early stage in splicing, as 5' splice site cleavage and lariat formation were inhibited. In contrast, a different monoclonal antibody (2B12) had no inhibitory effect, and neither did antibodies to two other hnRNP proteins (the 120kd and 68kd proteins). Sedimentation experiments showed that the 4F4 antibody diminished, but did not prevent 60S splicing complex formation. In addition, depletion of C proteins from the extract by immunoadsorption with 4F4 or 2B12 resulted in loss of both splicing activity and 60S splicing complex formation.

It has been suggested that the hnRNP monoparticle may serve as an "operating table" for RNA splicing (Dreyfuss, 1986). The spliceosome may perhaps be a modified hnRNP particle, or monoparticle subdomain, that contains, in addition to hnRNP proteins, the other components identified

belonging to the spliceosome e.g. snRNPs. As very short pre-mRNA can be spliced *in vitro*, the spliceosome may not be a complete monoparticle, but only one of its subdomains, containing one or more C proteins.

1.7.2. Lariat-debranching enzyme

Extracts from HeLa cells (Ruskin and Green, 1985c) and yeast (N. Schneider, referenced in Green, 1986) contain an activity that specifically hydrolyses the 2'-5' phosphodiester bond of the lariat structure. The debranching activity was only observed after deproteinising the RNA and adding it back to the extract, suggesting that during the normal *in vitro* splicing reaction the 2'-5' phosphodiester bond is protected from cleavage by this activity. This enzymatic activity most likely functions in the intron degradation pathway *in vivo*, and a linear form of the intron, resulting from cleavage at the 2'-5' phosphodiester bond can be detected *in vivo* in both yeast (Domdey *et al.*, 1984) and mammalian cells (Zeitlin and Efstratiadis, 1984).

It is possible, however, that the lariat-debranching activity is actually an enzymic activity involved in pre-mRNA splicing. If 5' splice site cleavage and ligation at the branchpoint is an isoenergetic transesterification, catalysed by the debranching activity, then debranching itself could be the microscopic reversal of
this reaction, with H_2O or OH^- functioning as the nucleophile in the absence of the 3'-OH group of the 5' exon.

1.7.3. RNA duplex unwinding (helicase) activity

The presence of an ATP-dependent RNA duplex unwinding activity was first suggested by the observation that two RNA molecules, joined by intermolecular base pairing, could undergo ATP-dependent dissociation in a crude HeLa cell nuclear extract (Konarska *et al.*, 1985b). A helicase activity was also proposed as an explanation for the ATP facilitated increase in sensitivity of substrate RNA to degradation by RNase A, when incubated in HeLa cell nuclear extract (Ruskin and Green, 1985a).

Failed attempts to use antisense RNA techniques to inhibit mRNA translation and so study early embryogenesis in Xenopus, have revealed the presence of an activity that specifically unwinds or denatures RNA:RNA duplexes (Rebagliati and Melton, 1987; Bass and Weintraub, 1987). The denaturation of a specific RNA:RNA hybrid in S100 extracts could be inhibited by the addition of a second duplex RNA, but not by the addition of

single-stranded RNA, single stranded DNA, or double-stranded DNA (Bass and Weintraub, 1987). Furthermore, this unwinding activity was found to be developmentally regulated. It was present at only low levels in *Xenopus* oocytes, increased during oocyte maturation, existed at high levels in eggs and early embryos, and was absent, or very much diminished in late blastula embryos.

An RNA duplex unwinding activity has also been found to be present in crude cell extracts from a variety of mammalian tissue cell lines, including HeLa, mouse plasmacytoma and Burkitt lymphoma (Wagner and Nishikura, 1988). This activity appears identical to that reported from *Xenopus* oocytes and early embryos (Rebagliati and Melton, 1987; Bass and Weintraub, 1987), being specifically inhibited by addition of double-stranded RNA as a competitor. This activity does not, however, require ATP and is, therefore, not identical with that proposed by Konarska *et al.* (1985b) or (Ruskin and Green, 1986a). The products of unwindase activity appeared to be modified in some way, since although they were susceptible to RNases A and T1, they migrated differently on native polyacrylamide gels from the original RNA before annealing, and even after deproteinisation will not renature. The level of unwindase activity in mouse fibroblast 3T3 cells was also found to depend on the stage in the cell cycle. Unwindase activity was very low in 3T3 cells arrested into quiescence, but increased when the cells were released into renewed growth by serum. Using a gel retardation assay, a protein(s) which binds specifically to the RNA duplex was detected, and this protein(s) may well prove to be the unwindase itself.

An RNA duplex unwinding activity may be important for exposing critical regions of pre-mRNA and allowing assembly of the splicing complex

1.8. USE OF NUCLEOSIDE PHOSPHOROTHIOATES IN THE STUDY OF PRE-mRNA SPLICING

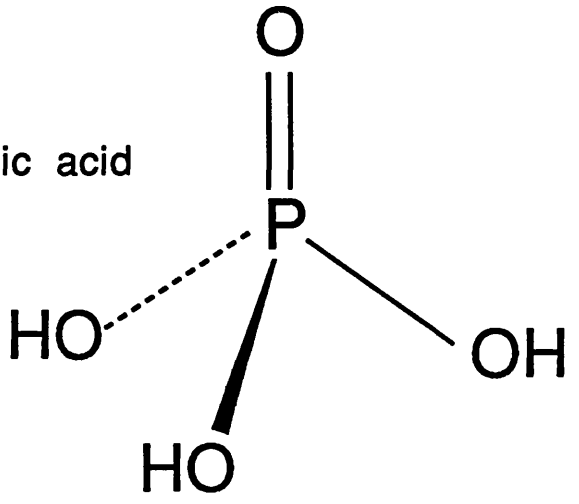
Knowledge of the stereochemical course of any chemical reaction provides a useful insight into the mechanism of that reaction, since mechanisms are defined by transition states and the stereochemical course of reactions are also determined by the structure of the transition state, or in multiple step reactions by the structure of the transition states and intermediates.

The phosphoric acid molecule is tetrahedral (Figure 1.3A) but not a perfect tetrahedron because of the inequivalence of the double bonded oxygen and the hydroxyl groups. Proton exchange, however, results in the oxygen atoms in a population of phosphoric acid molecules becoming equivalent. In contrast, the oxygen atoms do not themselves exchange rapidly, and their rates of exchange are

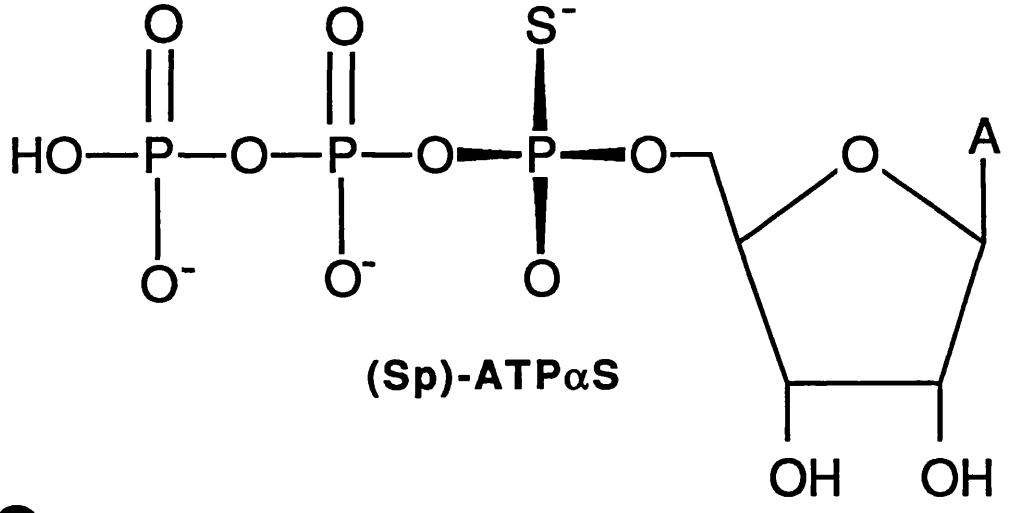
Figure 1.3. **Structural formulae of phosphoric acid (*panel A*), and the two diasterioisomers of adenosine 5'-O-(1-thiotriphosphate) (ATPαS) (*panels B and C*).**

A.

Phosphoric acid

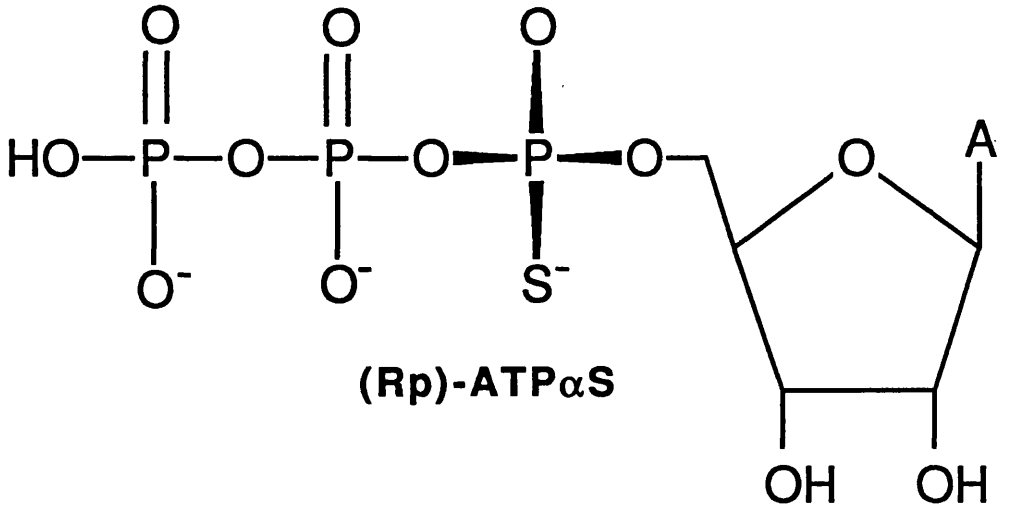


B.



(Sp)-ATPαS

C.



(Rp)-ATPαS

so small as to be negligible for most purposes. This is also true for phosphorothioates and phosphoramidates. Phosphates can, therefore, be chiral molecules when the substituents of tetrahedral phophorus are physically different. Exchange of protons, fractional bond orders in anionic forms, and the involvement of double bonds which shift positions in fully protonated species complicate the assignment of R and S configurational symbols to chiral phosphates. To avoid this problem, protons, negative charges and double bonds are ignored when assigning the configurational symbols Rp and Sp to chiral phosphorus centres. Double bonds are also not included in most structural formulas depicting chiral phosphorus centres. Figure 1.3B and C shows the structural formulae of the two diastereoisomers of adenosine 5'-O-(1-thiotriphosphate) (ATPαS).

Using chiral phosphates, detailed studies of non-enzymic reactions have resulted in three associative and one dissociative mechanism of nucleophilic substitution at phosphorus in phosphates being proposed (Westheimer, 1968; Benkovic and Schray, 1971; Westheimer, 1981). The dissociative mechanism (Figure 1.4, Eqn. 1) is analogous to the S_N1 reaction in carbon chemistry. The phosphate expels the leaving group in the rate limiting step, producing a planar electrophilic metaphosphate as an intermediate. Since the planar metaphosphate can capture the attacking nucleophile at either face, epimerisation should occur (if R_1 and R_2 are themselves chiral).

Of the three associative mechanisms, the first (Figure 1.4, Eqn. 2) is analagous to the concerted S_N2 mechanism. It is an "in-line" attack, the displacing nucleophile attacking from the side opposite the leaving group and displacing it in a single step. This mechanism should result in inversion of configuration at a chiral phosphorus centre.

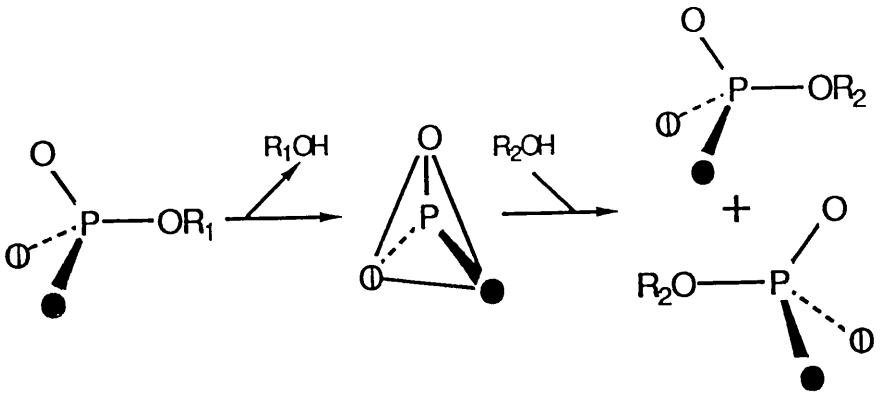
The other two associative mechanisms differ in that they involve intermediates. In the mechanism shown in Figure 1.4, Eqn. 3 is also an "in-line" attack, the attacking nucleophile approaching from the side opposite the leaving group, forming a bond to the phosphorus atom. The pentacovalent trigonal bipyramidal intermediate that results has the displacing nucleophile and the leaving group in the apical positions and decomposes to products by the departure of the leaving group from its apical position resulting in inversion of configuration.

The mechanism depicted in Figure 1.4, Eqn. 4 differs from that in Eqn. 3 in that it involves "adjacent" attack on the phosphoryl residue. The attacking nucleophile is in an apical position of the first intermediate, whereas the leaving group is in the equatorial plane. Since leaving groups only depart from apical positions, a pseudorotation (or turnstile rotation) must occur before product formation so as to form a second intermediate in which the two apical groups of the first intermediate become equatorial and two of the equatorial groups (including the leaving group), become apical. The leaving group then departs from its apical position in the second intermediate. Pseudorotatory rearangements of pentacovalent phosphorus

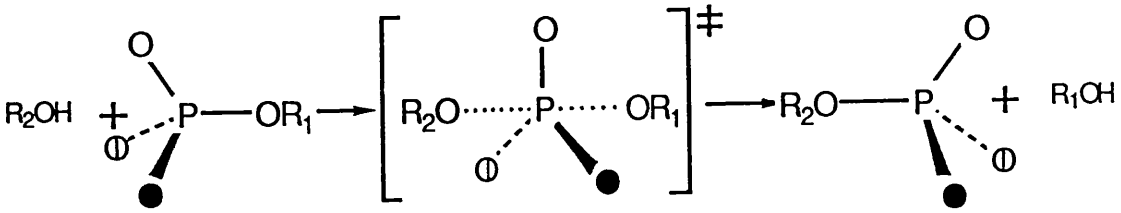
Figure 1.4. **The four possible mechanisms for phosphoryl transfer.**

The three peripheral oxygens of the phosphoryl group are distinguished so that the stereochemical consequence of each mechanism is evident: ^{16}O, ●; ^{17}O, O; ^{18}O, ⊕.

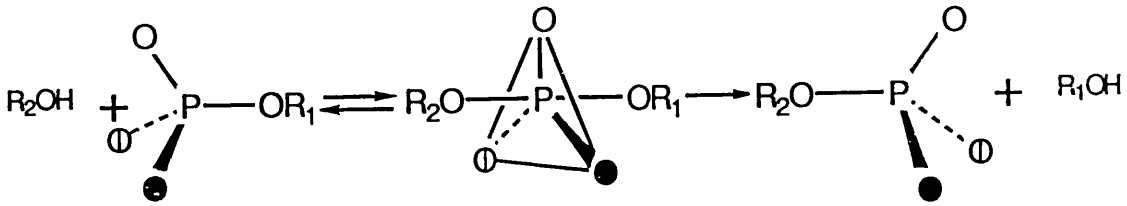
(1)



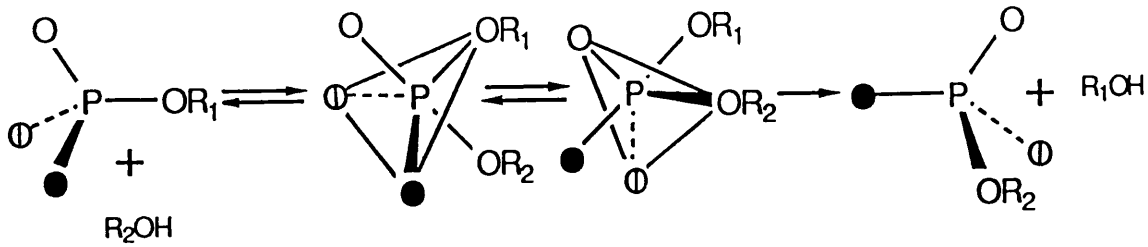
(2)



(3)



(4)



compounds are governed by rules based on kinetic and spectroscopic evidence from many phophorus compounds (discussed in Westheimer, 1968 and Westheimer, 1981). In Eqn. 4, ^{18}O serves as the pivot, remaining equatorial in the second intermediate. However, the ^{16}O may also serve as the pivot in which case the ^{16}O remains equatorial. Retention of configuration results whichever pathway is followed.

Of the enzymes studied with respect to the stereochemical course of the phosphoryl or nucleotidyl transfer reactions catalysed by them, most (but not all) show inversion of configuration (reviewed in Eckstein, 1985). This is consistent with a single step in-line substitution which inverts the configuration at the chiral phosphorus centre as in the mechanisms depicted in Eqns. 2 and 3, but not with the adjacent attack mechanism of Eqn. 4.

Retention of configuration could be the result of two successive substitution reactions, each proceeding with inversion of configuration and involve the formation of a covalent enzyme (or cofactor) intermediate (a double-displacement reaction). Alternatively, retention could result from adjacent attack and pseudorotation of the pentacoordinate intermediate (Figure 1.4, Eqn. 4). However, for most of the enzymes which show retention of configuration a covalent enzyme intermediate has been postulated either from kinetic arguments or from isolation. Hence, for these enzymes, a double-displacement mechanism would seem the most simple explanation. To date there is no evidence that pseudorotation is involved in any enzymatic phosphoryl or nucleotidyl transfer reactions. In addition, there is no indication of a dissociative mechanism involving formation of metaphosphate (Figure 1.4, Eqn. 1), since no epimerisation has been been observed in any enzyme reaction studied to date. It is possible, however, that inversion of configuration could result if the transient metaphosphate were to be tightly bound and not allowed to rotate. However, it may be meaningless to postulate a mechanism in which the acceptor substrate is juxtaposed in the active site so as to immediately react with the metaphosphate since bond breaking and formation would be almost simmultaneous, thus approaching the situation of an associative mechanism.

Although more complicated mechanistic interpretations of stereochemical data can be considered, for example, any odd number of in-line substitutions would result in a net inversion of configuration, in the absence of data to suggest a need to complicate the argument the simplest explanation seems to suffice. Staphylococcal nuclease catalyses a reaction which occurs with inversion of configuration, and the first step of RNase A catalysed cleavage (cyclisation) also occurs with inversion of configuration. Single step in-line substitution mechanisms were therefore proposed for these processes (Usher *et al.*, 1970; Usher *et al.*, 1972; Saenger *et al.*, 1974; Mehdi and Gerlt, 1982). The X-ray crystal structures of the active sites of RNase A and staphylococcal nuclease have been defined with high enough precision to enable a mechanism to be delineated (Cotton *et al.*, 1979; Wlodawer *et al.*, 1982) and this structural analysis confirms the mechanistic

predictions from the stereochemical studies.

The stereochemical course of enzymatic phosphoryl or nucleotidyl transfer reactions seems, therefore, to provide strong evidence to distinguish between a one step and a two step substitution reaction in which a covalent enzyme intermediate occurs. In the former case inversion, and in the latter case retention of configuration at the phosphorus atom should be observed.

The relevance of stereochemical data obtained using phosphorothioates has been established by analysing enzymatic reactions using both a phophorothioate substrate and a substrate in which the phosphorus at the reaction centre is made chiral by the presence of the different isotopes of oxygen (see Eckstein, 1985, and references therein). All enzymes studied by both methods to date show identical stereochemical courses with phosphates and phosphorothioates (see Eckstein, 1985, and references therein).

Phosphorothioates have the advantage over substrates made chiral with isotopic oxygens in that they are easier to manipulate as product analysis is simpler and much smaller amounts of material are required. Complications can arise, however, from the slower reaction rates often encountered with phosphorothioate substrates, especially when competing secondary reactions occur, for example, desulphurisation with snake venom phosphodiesterase or nuclease S1 (Burgers and Eckstein, 1978; Potter *et al.*, 1983a) or transphosphorylation with spleen phosphodiesterase (Mehdi and Gerlt, 1981).

Substrate analogues in which a chiral phosphorothioate group is the reaction centre also have the useful property that frequently one configuration is a much better substrate than the other in a particular enzymatic reaction. The reactivity of one diastereoisomer frequently approaches zero and is often found to act as a competitive inhibitor. Thus, stable enzyme-substrate complexes can be formed, or intermediates revealed. An interesting example of substrate specificity is seen with valyl-tRNA synthetase with (Rp)- and (Sp)-ATPβS. The enzyme uses only the (Rp) diastereoisomer as an aminoacylation substrate but also catalyses the interconversion of (Sp)-ATPβS and ATPγS (Rossomando *et al.*, 1979). The two processes apparently occur at different sites since neither diastereoisomer inhibits the other. Enzymatic reactions can also be influenced by chiral phophorothioate groups in the substrate which are not themselves involved in phosphoryl or nucleotidyl transfer. This is the case with the restriction endonuclease *Eco* RI, where an (Sp)- but not an (Rp)-thiophosphodiester link in the substrate DNA, one nucleotide upstream of the cleavage site, prevents cleavage at the normal site. Functional selectivity also extends to interactions at allosteric sites, for example, the regulatory subunit of cyclic AMP-dependent protein kinases bind both (Sp)- and (Rp)-cyclic AMPS, but only the the (Sp) epimer induces dissociation of the regulatory subunit from the catalytic subunit (de Wit *et al.*, 1982; O'Brien *et al.*, 1982; Rothermel *et al.*, 1983).

The different reactivity of phosphorothioate diastereoisomers, both from each other and from normal substrates, is a function of both the increased size of a sulphur atom compared to an oxygen atom (the Van der Waal's radius of sulphur is 1.85Å compared to 1.4Å for oxygen), and of the different bond order and charge localisation on a phosphorothioate group compared to a phosphate group. Although the bond order and charge localisation in phosphorothioate groups is still a subject of debate, evidence based on bond lengths obtained from X-ray crystallographic data and electron diffraction, the magnitude of the effects of ^{18}O on the ^{31}P-NMR chemical shifts of phosphorus in nucleoside [^{18}O]phosphorothioates, the pH-dependence of ^{17}O-NMR chemical shifts in [^{17}O]phosphate and [^{17}O]thiophosphate, the vibrational spectra of thiophosphate di- and trianions and the pKa values for phosphoric and triphosphoric acids, suggests that in phosphorothioate anions the P-S bond probably approximates to a single bond with a negative charge localised on the sulphur (Frey and Sammons, 1985).

This thesis describes the use of phosphorothioate analogues of RNA as substrate analogues for mammalian *in vitro* splicing reactions.

Some studies have been performed previously on polynucleotides containing thiophosphodiester linkages. Deoxynucleoside 5'-0-(1-thiotriphosphates) (NTPαS) have been used to determine the stereochemical course of reactions catalysed by DNA polymerases (Burgers and Eckstein, 1979b; Romaniuk and Eckstein, 1982; Brody *et al.*, 1982; Bartlett and Eckstein, 1982; Eckstein and Jovin, 1983), and internucleotidic thiophosphodiester linkages have been used to analyse the mechanism of reaction of nucleases and site-specific restriction endonucleases (Bryant and Benkovic, 1979; Burgers and Eckstein, 1979a; Potter *et al.*, 1983a; Potter *et al.*, 1983b; Connolly *et al.*, 1984) and for physical studies, for example on the B→Z transition (see Eckstein, 1985).

Phosphorothioate analogues of RNA have, on the other hand, been little exploited. *E. coli* RNA polymerase and *Micrococcus luteus* polynucleotide phosphorylase can utilise the (Sp)-diastereoisomers of NTPαS and NDPαS respectively to produce RNA containing (Rp)-thiophosphodiester links (Burgers and Eckstein, 1978; Burgers and Eckstein, 1979c). Phosphorothioate RNA has itself been used to investigate the stereochemistry of the reactions catalysed by RNase A, RNase T1, RNase T2 and *Enterobacter aerogenes* phosphohydrolase (Eckstein *et al.*, 1972; Usher *et al.*, 1972; Saenger *et al.*, 1974; Burgers and Eckstein, 1979a; Gerlt and Wan, 1979).

1.9. AIM OF THIS WORK

In the study presented in this thesis phosphorothioate RNA for *in vitro* splicing was generated by "run-off" transcription using T7 RNA polymerase. A study of the

stereochemical course and substrate specificity of the reaction catalysed by T7 RNA polymerase was undertaken. The configuration of the phosphorothioate RNA used in the *in vitro* splicing system was, therefore, known.

Splicing involves a minimum of two phosphoryl transfers, the first occuring at the 5' splice site, the second at the 3' splice site. It was considered possible that the presence of thiophosphodiester groups at the 5' and/or 3' splice sites would not prevent the normal phosphoryl (in this case thiophosphoryl) transfer reactions from taking place. Hence, determining the configuration of the transfered thiophosphoryl group in the product ,i.e. in the 2'-5' link at the branch point or the 3'-5' link joining the spliced exons, would reveal whether the reactions occured with inversion or retention of configuration. As described above, this information on the stereochemical course of the two steps of splicing should reveal whether each step is a single direct transesterification, or a more complex, multi-step reaction involving sequential phosphoryl transfers and a covalent enzyme (or cofactor) intermediate.

Alternatively, it was thought possible that the presence of phosphorothioates, not necessarily at the splice junctions, might block, or perturb, certain steps of splicing, revealing hitherto undetected intermediates in the splicing pathway or informative side reactions. Particularly interesting would be a double displacement reaction in which a thiophosphodiester link at a splice site was in a favourable configuration for the first phosphoryl transfer, but the covalent phosphoryl-enzyme (cofactor) intermediate generated was in an unfavourable configuration for the second phosphoryl transfer. In this way, a normally transient phosphoryl-enzyme (cofactor) intermediate could be caused to accumulate.

Chapter 2.

Methods

2. METHODS

2.1. Synthesis of ATPαS and purification of diastereoisomers

Adenosine 5'-phosphorothioate (AMPS) was synthesised as Murray and Atkinson (1968). This procedure involves reaction of adenosine with thiophosphoryl chloride in triethyl phosphate (to prevent reaction with 2' and 3' hydroxyls) followed by hydrolysis. AMPS was purified on a 28 by 2.2 cm column of Sephadex A25-DEAE eluted with a 50 to 300mM gradient of triethylammonium bicarbonate (TEAB) pH7.

Synthesis of adenosine 5'-(O-1-thiotriphosphate) (ATPαS) from AMPS was as Eckstein and Goody (1976). In this procedure AMPS is converted to the tri-*n*-octylammonium salt and reacted with diphenyl phosphorochloridate. The resulting complex is reacted with the tri-*n*-butylammonium salt of pyrophosphate to generate both diastereoisomers of ATPαS. The ATPαS was purified on a 20 by 2.2 cm Sephadex A25-DEAE column eluted with a 50 to 600mM gradient of TEAB pH7. Reactions and column fractions were assayed throughout by low field ^{31}P nuclear magnetic resonance (NMR) spectroscopy. Purified ATPαS was analysed by high field ^{31}P NMR spectroscopy with broad band proton decoupling on a Bruker AM300 spectrometer operating at 121.5 MHz.

Purification of diastereoisomers by HPLC was performed on a Shimadzu LC-4A system with an Apex 5μ octadecyl-silicon reverse phase column. The eluant was 50mM TEAB pH6.45 containing 2% acetonitrile.

2.2. Cloning

The *Pvu*II-*Taq*I fragment of the rabbit β-globin gene, which contains the small intron (IVS-1) and some flanking sequences, was inserted into mICE10 (Eperon, 1986) cleaved with *Sma*I and *Acc*I, generating **mICE10[IVS-1]**.

The *Taq*I-*Eco*RI fragment of the rabbit β-globin gene containing the large intron (IVS-2) and some flanking sequences was inserted into **mICE11** (Eperon, 1986) cleaved with *Acc*I and *Eco*RI, generating **mICE11[IVS-2]**.

2.3. Mutagenesis

mICE11[IVS-2] 3'A was created using a mismatched oligonucleotide 17

residues long according to the principle of Zoller and Smith (1983). 0.05 to 0.6 pmole kinased oligonucleotide was annealed to 0.2 pmole single stranded **mICE11[IVS-2]** DNA and then extended around the circle and covalently closed overnight at 25°C in a 12 µl reaction comprising all four dNTPs, each at 0.25 mM, 0.5 mM rATP, 10 mM Tris-HCl (pH7.5), 5 mM magnesium chloride, 0.6 units DNA polymerase I, Klenow subfragment (Pharmacia) and 0.6 units T4 DNA ligase (Pharmacia). Closed circles were purified on a 0.8% low melting point agarose gel, gel slices being heated at 65°C for 15 minutes in a ten fold volume of water and 1% and 10% of each slice used to transfect directly into competent *E. coli* JM101.

mICE10[IVS-1]Δ5'ss and **mICE10[IVS-1]ΔP(y)n** were made by mutating **mICE10[IVS-1]** as above, but using 30 residue long oligonucleotides designed to delete the 5' splice site sequence AG/GTTGGT (where / indicates the site of cleavage) and the polyprimidine tract from -30 to -5 (inclusive) relative to the 3' splice site.

To create **mICE10[IVS-1]3'A** an enrichment procedure was employed. A 17 residue long mismatched oligonucleotide was annealed to single stranded mICE10[IVS-1] DNA, extended around the circle and ligated as above. The copied DNA was then cut at a unique site with *Bal* I, and the *Bal* I inactivated by heating. An approximately equimolar quantity of M13mp19 single stranded DNA, previously cleaved with *Eco* RI and *Hind* III after annealing two appropriate 17 residue long oligonucleotides, was mixed with the copied DNA in a solution of 10mM Tris-HCl (pH7.5), 5mM magnesium chloride. This mixture was boiled in an Eppendorf tube for 60 seconds and transferred directly to a 65°C water bath for a further 5 hours, after which time it was placed on ice and used to transfect *E. coli* W71-18 (Winter and Fields, 1980). This procedure took advantage of the absence, in M13mp19, of the amber mutations in mICE10. *E. coli* W71-18 lacks an amber suppressor. Plaques were screened by sequencing (Eperon, 1986b).

2.4. Preparation of plasmid DNA or M13 replicative form (RF) DNA from 100ml cultures

A standard alkaline lysis method for recovering DNA from *E. coli* was employed.

M13 RF was produced by picking a plaque into 10ml TY broth and growing it with vigorous shaking for 4 hours at 37°C. This culture was then added to 80ml TY broth along with 10ml exponentally growing *E. coli* JM101 and shaking at 37°C continued for a further 3 to 4 hours before harvesting.

Cells were harvested by spinning at 6000 rpm for 5 minutes at 4°C. The

pellet was resuspended in 4ml ice-cold GTE (50mM glucose, 10mM EDTA, 25mM Tris-HCl (pH8.0)). 8ml 0.2M sodium hydroxide, 1% SDS were then added and the mixture shaken vigorously before leaving on ice for 5 minutes, after which time 6ml 5M potassium acetate (pH4.8) was added to neutralise, and the mixture shaken. after leaving on ice for a further 10 minutes the mixture was spun at 4000 rpm for 15 minutes at 4°C, the supernatant decanted and filtered through a 0.45µm non-absorbent filter (Gelman "acrodisc"). Early preparations were purified by caesium chloride density gradient centrifugation, in which case 1.22g caesium chloride (Koch-Light) was added per ml of supernatant. A one-fifth volume of 0.75mg/ml ethidium bromide was then added and the mixture spun overnight at 55000 rpm in a Beckmann vTi 65.2 rotor. The band was withdrawn and the ethidium extracted with water-saturated butan-2-ol. After adding 2 volumes of water the DNA was precipitated with ethanol. This procedure involved freezing the ethanol mixture on dry ice and then warming to room temperature to dissolve salt before centrifugation.

Latterly, however, a purification procedure has been used that negates the need for caesium chloride gradients. In this method, the supernatant from the alkaline lysis, having been filtered through a 0.45µm filter as above, was added to 17ml propan-2-ol and stored on dry ice for 15 minutes, after which time it was spun at 4300 rpm for 20 minutes. The pellet was dried and resuspended in 2ml TE.1 and 2.5ml 4.4M lithium chloride added. After storage at 4°C for one hour or more, this mixture was spun at 8000 rpm for 5 minutes at 4°C. The supernatant, which should now be considerably depleted of protein and RNA, was decanted, precipitated with ethanol and washed. The dried pellet was resuspended in 400µl TE.1 and incubated for 15 minutes at 37°C with 10µl 10mg/ml RNase A (Pharmacia) which had been incubated for 10 minutes in a boiling water bath before use. 20µl 10% SDS were then added and the mixture heated at 70°C for 10 minutes (to aid removal of an otherwise troublesome nuclease activity) then extracted twice with phenol, once with phenol : chloroform : isoamylalcohol in a 25 : 24 : 1 (v : v : v) ratio, and twice with ether. Finally, the DNA was precipitated with ethanol at room temperature and washed.

Yields were typically 0.5 to 2mg.

2.5. Preparation of single-stranded M13 DNA from 100ml cultures

After growing cells as for RF, 10ml of the culture was spun at 8000 rpm for 10 minutes and the supernatant filtered through a 0.45µm Gelman "acrodisc". 200µl 20% PEG (polyethylene glycol, 6000 grade), 2.5M sodium

chloride was added per ml supernatant, and after 15 minutes at room temperature the mixture was spun at 4000 rpm for 25 minutes at room temperature. The pellet was resuspended in 1ml TE.1 and precipitated with PEG as above, and spun at 13000 rpm for 5 minutes in a microcentrifuge. The supernatant was aspirated, the tube respun, and any remaining supernatant removed. The pellet was resuspended in 200µl TE.1, extracted once with phenol : chloroform : isoamyl alcohol in a 25 : 24 : 1 ratio, twice with phenol and twice with ether, and finally precipitated with ethanol at room temperature and washed.

Yields were typically 0.5 to 1mg DNA.

2.6. Small scale preparation of single-stranded M13 DNA in 96 well microtitre plates

The method is as Eperon (1986b) with the modifications in Eperon *et al.* (1988).

2.7. Preparation of competent *E. coli* using calcium chloride

The method is modified from Cohen *et al.* (1972).

A single colony from a fresh minimal plate was used to inoculate a suitable volume of TY broth and grown with vigorous shaking at 37°C. Cells were harvested when the culture had an OD550 of 0.6 to 0.9 by centrifugation at 2500 rpm for 5 minutes. The pellet was resuspended in ice-cold 50mM calcium chloride at half the original volume and left on ice for at least 20 minutes (up to overnight). The cells were recentrifuged as before, and resuspended in ice-cold 50mM calcium chloride at one-tenth the original volume. The cells were either used immediately, or made 16% with respect to glycerol and stored at -70°C for up to a month.

2.8. Transfection of competent *E. coli* JM101 or W71-18 with M13 DNA

200µl of competent cells were mixed with approximately 1ng M13 RF or 10ng single-stranded M13 DNA, left on ice for 40 minutes, and then heat-shocked at 42°C for 2 minutes. The cells were then added to 3ml molten top-agar at 42°C and poured onto SB agar plates. When employing the blue/white colour selection to cloning in mICE vectors 50µl IPTG (24mg/ml in water) and 50µl

BCIG (20mg/ml in dimethylformamide) were added to the top-agar (i.e. slightly more than is required for M13 mp vectors). Plates were incubated at 37°C overnight. Yields of 10^5 to 10^7 plaques per µg M13 RF were recovered.

If *E. coli* W71-18 was used then 200µl dilute exponential W71-18 was added after the heat shock to aid good lawn production.

When transfecting closed circular DNA from mutagenesis copying reactions, 200µl TY broth was added after the heat-shock and the mixture incubated at 37°C for one hour wthout shaking to allow a round of M13 replication and segregation of mutant and wild type alleles before plating in top-agar.

2.9. Preparation of competent *E. coli* using MOPS-rubidium chloride

The method is modified from that of Kushner (1978).

A single colony from a fresh minimal plate was used to innoculate a suitable volume of TY broth and grown with vigorous shaking at 37°C until an OD550 of about 0.5 was reached. For each transformation/transfection 1.4 ml of culture was aliquoted to an Eppendorf tube, and pelleted by a 30 second spin at 13000g. The supernatant was removed and the pellet resuspended in 0.5 ml 10mM MOPS (pH7.0), 10mM rubidium chloride, and spun again for 15 seconds. The pellet was then gently resuspended in 0.5 ml 100mM MOPS (pH6.5), 10mM rubidium chloride, 50mM calcium chloride and held on ice for 30 to 90 minutes (optimum). After centrifuging for 10 seconds and removing the supernatant, the pellet was gently resuspended in a further 150µl of 100mM MOPS (pH6.5), 10mM rubidium chloride, 50mM calcium chloride and 3µl dimethylsulphoxide (Aldrich, "Gold Label").

2.10. Transfection of competent *E. coli* JM101 prepared by the MOPS-rubidium chloride method

To an Eppendorf of competent cells from above, 1ng M13 RF or 10ng single-stranded M13 DNA were added, held on ice for 60 minutes and then heat-shocked in a 55°C water-bath for 30 seconds and then immediately plunged into an ice-water bath for 1 minute. The procedure from this point on was exactly as with calcium chloride prepared cells after heat-shock. This method more consistently gave high yields than the calcium chloride method, usually 10^6 to 10^7 plaques per µg M13 RF.

2.11. Sequencing by the dideoxy (chain termination) method

The method is that of Sanger *et al.* (1977).

The following nucleotide mixes were required (stored at -20°C).

ddCTP : 0.25mM in TE.1
ddATP : 0.20mM in TE.1
ddGTP : 0.16mM in TE.1
ddTTP : 0.50mM in TE.1

C° : 10µl 0.5mM dCTP + 200µl 0.5mM dGTP + 200µl 0.5mM dTTP
A° : 200µl 0.5mM dCTP + 200µl 0.5mM dGTP + 200µl 0.5mM dTTP
G° : 200µl 0.5mM dCTP + 10µl 0.5mM dGTP + 200µl 0.5mM dTTP
T° : 200µl 0.5mM dCTP + 200µl 0.5mM dGTP + 10µl 0.5mM dTTP

Sequencing was occasionally performed in Eppendorf tubes, but more usually in flat-bottomed 96 well microtitre plates (Nunc). For each reaction 0.01 to 0.04 pmoles of template and 0.03 pmoles of primer were annealed in 4µl of 2.5seq (25mM Tris-HCl (pH7.5), 12.5mM magnesium chloride) by heating to 65°C for 5 minutes, and then cooling at room temperature for 15 minutes. When using a 5' labelled primer the amount of primer and template used was increased ten-fold. To each reaction was added 2µl of a mixture of ddNTP, N° and 10µM dATP in a ratio of 4:4:1 (v:v:v), although the ratio of ddNTP : N° was varied depending on the length of extesion required. The 10µM dATP was omitted from reactions using 5' labelled primers. 2µl ice-cold 25% glycerol, 25mM potassium phosphate (pH7.5) containing 1µCi [α^{32}P] dATP (3000 Ci/mmol, Amersham) and 0.25 units DNA polymerase I, Klenow subfragment (Pharmacia) were added and incubated at room temperature for 20 to 35 minutes, or latterly at 37°C for 5 minutes. When using 5' labelled primer the [α^{32}P] dATP was omitted and the Klenow solution instead made 5µM with respect to dATP. 2µl of a mix of all four dNTPs, each at 0.5mM, was then added and incubation continued at room temperature for 15 minutes, or latterly at 37°C for 5 minutes. 2µl formamide-dye mix (90% formamide, 10% 0.5M EDTA with bromophenol blue and xylene cyanol) was then added and the mixture heated, uncovered, at 80°C for 8 minutes before loading 2µl onto a denaturing polyacrylamide gel. All solutions were conveniently added to the walls of microtitre plate wells and mixed by a brief centrifugation at 1800 rpm.

2.12. Maintenance of HeLa cells in suspension

HeLa cells in suspension (JS1000, a gift from Prof. J.A. Steitz) were grown in 500ml or 1.5l microcarrier stir bottles (Techne) in DMEM containing 10% newborn calf serum, 2g/l sodium bicarbonate, 10^5 u/l penicillin and 10^5 µg/l streptomycin. Cells were stirred continually at 60 rpm and maintained at a density of between 3 and 6 x 10^5 per ml by dilution every 24 hours.

2.13. Frozen cell stocks

HeLa cells lines were stored under liquid nitrogen in 20% foetal calf serum, 10% dimethylsulphoxide in DMEM. Once thawed and washed in fresh medium cells were grown initially in medium containing 20% foetal calf serum.

2.14. S1 nuclease mapping

The probe used for mapping the four G-A mutants spliced *in vivo* was an *Nco* I restriction fragment generated by digestion of 1µg of the appropriate construct with 2 units *Nco* I for 2 hours at 37°C in a 5µl volume comprising 100mM Tris-HCl (pH7.5), 50mM sodium chloride, 10mM magnesium chloride, 1mM DTT.

The probe for mapping *in vitro* spliced , unlabelled RNA, was an *Eco* RI-*Hind* III fragment of **mICE10[IVS-1]**, generated by digestion of 1µg RF DNA with 2 units *Eco* RI and 2 units *Hind* III for 1 hour at 37°C in TMS. The fragment was 3' end-labelled by a 2 nucleotide extension at the staggered end by adding 1.25 units DNA polymerase I, Klenow subfragment, in a 4µl volume comprising 60mM magnesium chloride, 3mM dTTP and containing 20µCi [α^{32}P]dATP (3000Ci/mmol, Amersham). Incubation was at room temperature for 15 minutes.

Labelled probes were purified on 5% polyacrylamide, 20% formamide, 7M urea gels. S1 mapping was as described by Favaloro *et al.* (1980), except the S1 digestion buffer was 280mM sodium chloride, 50mM sodium acetate (pH4.6), 4.5mM zinc sulphate, 20µg/ml denatured salmon sperm DNA. Duplexes were denatured at 80°C for 15 minutes and hybridisation was overnight at 54°C.

2.15. Primer extension

3pmole oligonucleotide were labelled in a 10µl reaction volume comprising 50mM Tris-HCl (pH7.5), 10mM magnesium chloride, 1mM DTT, 25µCi [γ^{32}P]ATP (3000Ci/mmol, Amersham) and 1 unit polynucleotide kinase for 1 hour at 37°C, after which the reaction was extracted with phenol and ether, spun through a Sephadex G50 column (Maniatis *et al.* , 1982) and precipitated with ethanol. Approximately 0.3pmole oligonucleotide was annealed to 20% of the yield of E2* made preparatively as above in a 14µl volume containing 30mM PIPES (pH6.7), 0.30M sodium chloride, 0.70mM EDTA, 0.14% SDS by placing at 80°C for 4 to 6 minutes, then cooling slowly (over about 45 minutes) to 37°C, at which temperature the solution was maintained for 40 minutes. After precipitation with ethanol the RNA was resuspended in 13µl 50mM Tris-HCl (pH8.3), 6mM magnesium chloride, 40mM potassium chloride, 0.38mM dNTPs and 12.5 units AMV reverse transcriptase (Anglian Biotechnology). Extension was for 30 minutes at 37°C. Labelled RNA was removed by adding 3µl 1M sodium hydroxide and heating for 5 minutes at 80°C. 12µl 5M potassium acetate (approximately pH4.8) was added to neutralise, and, after precipitation with ethanol, the DNA was loaded on a gel comprising 7M urea, 15% polyacrylamide. Sequencing markers were obtained by using the same oligonucleotides for dideoxy sequencing reactions (Sanger *et al.* , 1977) on **mICE10[IVS-1]** single-stranded DNA.

For analysis of kethoxal and DMS modification of RNA 0.3 pmole ^{32}P-labelled oligonucleotide was annealed to ethanol precipitated, modified RNA, in a total volume of 5µl consisting of 50mM potassium HEPES (pH7.0), 100mM potassium chloride (Moazed *et al.* ,1986) by heating to 80°C for 60 seconds and cooling immediately to 42°C, then continuing incubation at this tmperature for 15 minutes. The reaction volume was increased to 10µl for extension, and made to 50mM Tris-HCl (pH8.3), 6mM magnesium chloride, 40mM potassium chloride, 10mM DTT, 2.5mM each dNTP. Extension was at 42°C for 30 minutes with 9 units AMV reverse transcriptase (Promega). After ethanol precipitation the DNA was fractionated on a 10% polyacrylamide, 7M urea gel.

2.16. Chromatography

Thin-layer chromatography (TLC) was performed on 6.5cm high polyethyleneimine plates (PEI-Cellulose F, Merck) pre-treated and developed as Volckaert and Fiers, (1977). Two dimensional TLC was on 10cm x 10cm Polygram Cel 300 cellulose thin-layer plates (Macherey-Nagel). In the first

dimension, the solvent system used was isobutyric acid : 0.5M ammonium hydroxide, 5:3 (v:v); the second dimension solvent was propan-2-ol : concentrated HCl : water, 70:15:15 (v:v:v) (Saneyoshi *et al.* , 1972). Homochromatography was performed on 20cm high PEI-Cellulose F plates (Merck). The plates were pre-treated, and developed in homomixture β according to Volckaert *et al.* (1976). Samples were applied over widths of 1cm.

2.17. Nuclease digestions for stereochemistry studies

Serial dilutions of each enzyme were performed in reaction buffer : snake venom phosphodiesterase (SVPD) from *Crotalus durissus terrificus* (Boehringer) in 100mM Tris-HCl (pH8.75), 2mM magnesium chloride, 2mM DTT; nuclease S1 (Pharmacia) in 0.28M sodium chloride, 50mM sodium acetate (pH4.8), 4.5mM zinc sulphate; nuclease P1 (Pharmacia) in 30mM sodium acetate (pH4.8), 0.1mM zinc sulphate. The volumes of the products of transcription by *E. coli* and T7 RNA polymerases were adjusted to give equal counts per unit volume when digestions were performed to compare the two. 1µl of the purified preparative transcription products were added to 5µl volumes of each concentration of nuclease; when ^{32}P and ^{35}S-labelled materials were digested together, total reaction volumes were 10µl. When digestion of ^{35}S-labelled DNA was performed in parallel with that of mixed ^{32}P and ^{35}S-labelled DNA (Fig.), this too was performed in 10µl. Digestions were incubated for 30 minutes at 37°C and quenched in dry ice. The results were analysed by TCA precipitation (Maniatis *et al.* , 1982) or thin-layer chromatography on polyethyleneimine (PEI) plates (Volckaert and Fiers, 1977). In the latter case, 1µl of each digest was mixed with 2µg of *E. coli* $tRNA^{Phe}$ that had been digested to completion with nuclease P1.

2.18. Nuclease digestions of E2* RNA

All digestions of E2* RNA were performed in 10µl reaction volumes for 1 hour at 37°C. Reaction buffers were as above and that for RNase T2 was 50mM sodium acetate (pH4.6). Each digestion used 1µg SVPD, 1 unit nuclease P1 or 1 unit RNase T2 (Sigma).

2.19. Phosphatase and kinase treatment of E2* RNA

E2* was treated in a 10µl volume of 0.05M Tris-HCl (pH9.0), 1mM magnesium chloride, 0.1mM zinc chloride, 1mM spermidine, containing 30 units RNAguard (Pharmacia) and 1 unit calf intestinal phosphatase (Boehringer) for 30 minutes at 37°C after which time an extra 0.5 units of phosphatase was added and the incubation continued for a further 10 minutes. 200µl 0.5M sodium acetate (pH5), 1mM EDTA, 0.2% SDS was added, and after incubation for 20 minutes at 68°C , the reaction was extracted three times with phenol, twice with ether and precipitated with ethanol.

After phosphatase treatment, E2* was labelled using polynucleotide kinase as described for oligonucleotide substrates (Materials and Methods: Primer Extension) except that 30 units RNAguard were also present in the reaction mixture. The reaction was terminated by extracting once with phenol and once with ether and precipitating with ethanol. The RNA was fractionated by running on a 15% polyacrylamide, 7M urea gel, and full length RNA was excised, eluted and precipitated with ethanol prior to RNase T2 digestion.

2.20. Synthesis of pNp and pNp[S] markers

T7 RNA polymerase was used to transcribe *in vitro* 20µg each of RNA and an RNA analogue with all phosphodiesters replaced by thiophosphodiesters. Each RNA was cleaved with 500 units RNase T1 (Sigma) in a 60µl reaction volume comprising 50mM Tris-HCl (pH7.5), 10mM magnesium chloride, 1mM DTT, 30µCi [γ^{32}P]ATP (3000Ci/mmol, Amersham) for 1 hour at 37°C . 5 units of polynucleotide kinase were added , and the incubation was continued for a further hour at 37°C . The entire reactions were loaded onto a 25% plyacrylamide, 7M urea, 2 x TBE gel. Electrophoresis was stopped after the bromophenol blue marker had migrated 8cm. The T1 fragments were excised, leaving the [γ^{32}P]ATP on the gel. Once eluted the T1 fragments were precipitated with ethanol (in the presence of 200µg $tRNA^{Phe}$ carrier at 40000 rpm for 45 minutes), resuspended in 10µl 50mM sodium acetate (pH4.6) and digested at 37°C for 2 hours by 1 unit of RNase T2.

[^{32}P]pCp was from Amersham.

2.21. Oligonucleotide directed RNase H cleavage of snRNAs

The method was essentially as Black *et al.* (1985).

48µl splicing reactions, lacking substrate RNA, were incubated at 30°C

for 45 minutes with 1.2 units RNase H (BRL) and 10μg of the appropriate oligonucleotide before splitting into one 24μl aliquot and two 12μl aliquots. *In vitro* transcripts (RNA 50 or [sA]RNA 50) were added to each of the 12μl aliquots and splicing performed as normal. The 24μl aliquots were treated with proteinase K, extracted with phenol and precipitated with ethanol as for splicing reactions. Electrophoresis of these samples was on 1.14mm thick 8% polyacrylamide, 7M urea gels, and the extent of the snRNA cleavage was determined by staining with ethidium bromide. Oligonucleotide U1m is identical to that used to cleave U1 by Krainer and Maniatis (1985). Oligonucleotides U2a, U4a and U6b are as the oligonucleotides of the same name used by Black and Steitz (1986).

2.22. Oligonucleotides

Oligonucleotides were kindly synthesized in the department by Mr. J. Keyte and Mr. J. Turner. The oligonucleotides were either synthesized on paper discs (Matthes *et al.* ,1984) and purified on 20% polyacrylamide, 7M urea gels, or synthesized on an Applied Biosystems 380B DNA synthesizer and purified as above, or ethanol precipitated and used directly.

2.23. *In vitro* transcription for nuclease stereospecificity and T7 RNA polymerase stereochemistry studies

Preparative transcriptions by T7 RNA polymerase (Boehringer) were performed in 5μl volumes comprising 34mM Tris-HCl (pH7.5), 5mM magnesium chloride, 1.7mM spermidine, 8.8mM DTT, 8 units RNasin (P & S Biochemicals Ltd.), 0.44mM unlabelled NTPs or NTPαS, for 1 hour at 37°C with 20 units T7 RNA polymerase (Boehringer) and approximately 0.5μg template DNA. To prepare ^{32}P-labelled RNA, the mix was supplemented with 20μCi [α^{32}P] CTP at 800Ci/mmole (Amersham); to prepare ^{35}S-labelled RNA, unlabelled UTP was replaced with 10μCi [α^{35}S] UTP at 400Ci/mmole (Amersham). In all cases the template was *Hind* III-cut **mICE10[tet]** (a gift from Dr. I.C. Eperon), a recombinant of **mICE10** (Eperon, 1986) containing an insert of eukaryotic DNA, giving rise to a transcript of 470 nucleotides. After transcription, the reaction mixture was extracted with phenol and ether, and spun through a column of sephadex G50 (Maniatis *et al.* , 1982) to remove nucleoside triphosphates. After precipitation of the RNA with ethanol and solution in water, incorporation of radiolabelled nucleotides was measured by scintillation counting. For kinetic measurements of

transcription, 2µl reactions of the same composition were used, with [α^{32}P] CTP in all cases. The template was **mICE11[IVS-2]** RF, which, after cleavage with *Eco* RI, gave rise to a transcript of 830 nucleotides. The concentration of one nucleotide was varied as shown in Chapter 3. The reaction was initiated by addition of unlabelled nucleotides to a mix of all the other components, 2µl volumes were dispensed into tubes, and these were placed at 37°C for the times noted. Reactions were quenched in dry ice and followed by TCA precipitation (Maniatis *et al.* , 1982) and scintillation counting. Assuming the specific activity of the enzyme supplied was 6 x 10^5 units per mg (Chamberlin and Ryan, 1982), these assays used about 0.1pmole enzyme and 0.04pmole template.

Preparative transcriptions by *E. coli* RNA polymerase were performed in 50µl 32mM Tris-HCl (pH7.5), 120mM potassium chloride, 8mM magnesium chloride, 0.08mM EDTA, 0.08mM DTT, 0.2mM unlabelled NTPs with 8pmole *E. coli* RNA polymerase (Pharmacia) and 0.5µg uncut **mICE11[IVS-2]** RF DNA. 80µCi [α^{32}P] CTP or 40µCi [α^{35}S] UTP were used as supplements or replacements respectively, as noted above. Transcription was for 1 hour at 37°C. Incorporation was assessed as described for T7 RNA polymerase. CTPαS (an epimeric mixture), GTPαS and UTPαS were gifts from Prof. F. Eckstein, Göttingen.

2.24. *In vitro* transcription for splicing studies

The template for transcription was generated in two ways. First, replicative form **mICE10[IVS-1]** DNA was cleaved with *Hind* III in TMS. Alternatively, the 3' end of the transcript was defined by using extension from an oligonucleotide on single stranded **mICE10[IVS-1]** DNA (Turnbull-Ross *et al.* ,1988; Eperon, 1986a). 0.15pmole oligonucleotide was mixed with 0.5µg single stranded **mICE10[IVS-1]** DNA in a total volume of 3µl, heated at 65°C for 5 minutes and then left at room temperature for 15 minutes. The complementary strand was then extended past the T7 promoter region in a 5µl reaction comprising TMS, 0.5mM each dNTP and 2 units DNA polymerase, Klenow subfragment (Pharmacia) for 5 minutes at 37°C.

Transcription using T7 RNA polymerase was performed with 0.5µg of either template, in a 10µl volume comprising TMS, 5mM DTT, 1mM GpppG (Pharmacia) with 15 units RNAguard (Pharmacia), 10 units T7 RNA polymerase (Boehringer) and the appropriate concentrations of labelled and unlabelled NTPs. Unlabelled RNA was prepared in incubation mixtures which were 0.4mM in each NTP. ^{32}P labelled transcripts were prepared by supplementing the mix with 10µCi [α^{32}P] CTP or [α^{32}P] UTP at 800Ci/mmole,

or 10μCi 3000Ci/mmole [α^{32}P] ATP (all from Amersham) and in each case the concentration of the corresponding unlabelled NTP (or NTPαS) was dropped to 10μM. Phosphorothioate RNA was made by replacing 0.4mM NTP with 0.4mM NTPαS, either of the Sp configuration or an epimeric mixture. CTPαS, GTPαS and UTPαS were gifts from Prof. F. Eckstein and Amersham International.

Transcripts were purified by electrophoresis on 5% polyacrylamide, 20% formamide, 7M urea gels.

2.25. *In vitro* splicing

HeLa cell nuclear extracts were prepared from 3 litres of JS1000 cells (approximately 10^9 cells) according to Dignam *et al.* (1983), with the modifications of Heintz and Roeder (1984); triethanolamine was used in more recent preparations (Tazi *et al.*, 1986). The S100 fraction was prepared as Dignam *et al.* (1983) except buffers A and B contained, instead of HEPES, 10mM and 0.3M triethanolamine (pH7.9) respectively Splicing reactions were performed as Krainer *et al.* (1984) using 10μl of nuclear extract and 5μl duffer D.

2.26. DNA synthesis for T7 RNA polymerase stereochemistry studies

Labelled DNA was prepared by primed synthesis with *E. coli* DNA polymerase I, Klenow subfragment. 0.25pmole universal primer was annealed to 0.5μg single stranded mICE11[IVS-2] DNA in 20mM Tris-HCl (pH7.5), 10mM magnesium chloride, in a total volume of 32μl by heating to 65°C for 5 minutes and then leaving at room temperature for 15 minutes. Half of this mixture was used for ^{32}P-labelled DNA synthesis by making each dNTP to 30μM, with 10μCi [α^{32}P] dATP (3000Ci/mmole, Amersham), increasing the volume to 32μl and adding 4 units DNA polymerase I, Klenow subfragment (Pharmacia). ^{35}S-labelled DNA was synthesised likewise using the other half of the mixture in the presence of dCTPαS, dGTPαS and dTTPαS (all at 30μM) and 40μCi [α^{35}S] dATP (1000Ci/mmol, Amersham). Incubation was at room temperature for 15 minutes. Reaction products were purified as described for transcriptions.

2.27. Fractionation of spliceosomes on native polyacrylamide

gels

12.5µl splicing reactions (without PVA) were either loaded directly onto nondenaturing 4% polyacrylamide gels (acrylamide:bis-acrylamide weight ratio of 80:1) containing 0.5xTBE, or first pretreated by the addition of sodium heparin (BDH) to 5mg/ml and incubation continued for 10 minutes at 30°C (Konarska and Sharp, 1986). The gel (18 x 20 x 0.19cm) was pre-electrophoresed for 30 minutes at 200V in the same buffer. Electrophoresis was carried out at room temperature and the same voltage for about 5 hours (until xylene cyanole dye had run to the bottom). Gels of identical composition to above were also run, except the buffer was 50mM Tris-glycine (50mM Tris base, 50mM glycine [pH8.8 approx.]) (Konarska and Sharp, 1987). Pre-electrophoresis and electrophoresis were also as above, but the running buffer was Tris-glycine. Acrylamide, bis-acrylamide and ammonium persulphate were from BioRad (the quality of these reagents was found to be very important). Gels were autoradiographed at -70°C without drying.

2.28. Fractionation of spliceosomes by sucrose gradient centrifugation

The method is essentially that of Frendewey and Keller (1985).

75µl splicing reactions were stopped by quick freezing on dry ice and stored at -70°C for under 24 hours. The splicing reactions were thawed and diluted to 200µl with 1.5mM magnesium chloride, 100mM sodium chloride, 20mM Tris-HCl (pH7.5), 0.1mM EDTA before loading onto 4ml, 5%-20% sucrose gradients. The gradients were made by freezing at -70°C 12.5% sucrose, 100mM potassium chloride, 20mM HEPES-KOH (pH7.9), 1.5mM magnesium chloride, 0.1mM EDTA. Before use this solution was thawed at room temperature or 4°C, to allow the gradient to form, and pre-cooled to 4°C before loading. Centrifugation was at 45000 rpm in a Beckman SW50.1 rotor at 4°C for 75 minutes using a Sorvall OTD65B ultracentrifuge. Approximately 0.2ml fractions were collected and analysed by Cerenkov counting. RNA was prepared from pooled gradient fractions for analysis on 6% polyacrylamide, 20% formamide gels, 7M urea gels by extraction with phenol and precipitation with ethanol.

2.29. Immunoprecipitation and T1 protection analyses

Direct immunoprecipitation was performed on 25µl aliquots of splicing reaction as Black *et al.* (1985) except NET-2 buffer contained 0.05% Triton X-100 (v:v) instead of Nonidet P-40. Protein A-Sepharose CL-4B was from Sigma. Labelled RNA fragments were resolved on a 6% polyacrylamide, 20% formamide, 7M urea gel.

An identical procedure was used for RNase T1 protection experiments (Black *et al.* , 1985) except 3µl RNase T1 (500 units/µl, Sigma) was added immediately after the antiserum and the splicing reactions lacked PVA. High specific activity RNA transcribed with 20µCi [α^{32}P] UTP and no unlabelled UTP was used in the splicing reaction. Protected RNA fragments were either immediately fractionated on a 20% polyacrylamide, 7M urea, 2xTBE gel, or resuspended in 5µl 50mM sodium acetate (pH4.6) and digested to completion by 250 units RNase T1 at 37°C for 15 minutes before electrophoresis.

The anti-(TMG) antibody used was a rabbit polyclonal (Lührmann *et al.*, 1982), a gift from Dr. B. Lührmann. Anti-(U1)RNP (Ag) was a patient antiserum, and anti-(Sm) (Y12) was a mouse monoclonal antibody: both Ag and Y12 have previously been used in combined RNase T1 digestion and immunoprecipitation experiments (Black *et al.*, 1985; Chabot *et al.*, 1985; Chabot and Steitz, 1987b). Anti-(U2)RNP was a patient antiserum, which, like Ag and Y12, was a gift from Prof. J.A. Steitz.

3.30. Biotin-streptavidin affinity chromatography

Biotinylated RNA was synthesised by including biotin-11-UTP (BRL) (15% of total UTP) in the standard T7 transcription reaction (Bindereif and Green, 1987). The transcript was labelled with [α^{32}P] CTP.

Biotinylated RNA was incubated in nuclear extract under standard conditions (without PVA) in 150µl volumes. Biotin-streptavidin affinity chromatography was performed exactly as Bindereif and Green (1987), but volumes of all solutions used were increased 6-fold ,with the exception of the NET-2 buffer (0.5M potassium chloride) washes. Streptavidin-agarose was from BRL. Bound snRNAs were released from the streptavidin-agarose by adding 250µg proteinase K (Boehringer),150µl 0.2M Tris-HCl (pH7.5), 25mM EDTA, 0.3M sodium chloride, 2% SDS, and incubating at 37°C for 15 minutes. This mixture was extracted twice with phenol : chloroform : isoamylalcohol in a 25 : 24 : 1 ratio (v:v:v) and three times with ether and precipitated with ethanol in the presence of sonicated salmon sperm DNA at 10µg/ml.

[^{32}P]pCp was synthesised in a 50µl reaction consisting of 400µCi [γ^{32}P]ATP (3000Ci/mmol, Amersham), 1mM 3'-CMP, 25mM potassium-CHES (pH9.5), 5mM magnesium chloride, 3mM DTT and 4 units T4 polynucleotide

kinase (England *et al.* ,1980). Incubation was at 37°C for 90 minutes after which time the kinase was inactivated by heating at 70°C for 10 minutes.

3'-terminal labeling was essentially as England *et al.* (1980). The ethanol precipitated RNA from the affinity chromatography was resuspended in 10µl [^{32}P] pCp, synthesised as above, and 10µl 0.2mM ATP, 75mM HEPES (pH7.5), 30mM magnesium chloride, 4.5mM DTT, 20% dimethylsulphoxide. 4 units T4 RNA ligase (Pharmacia) were added and the reaction incubated overnight at 4°C after which time it was extracted with phenol, then chloroform : isoamylalcohol , 24 : 1 (v:v) and precipitated with ethanol. 3'-terminal labelled RNAs were fractionated on 6% polyacrylamide, 20% formamide, 7M urea gels and autoradiography performed.

2.31. Affinity chromatography using mercury-agarose

Phosphorothioate RNA was separated from normal RNA using Affi-Gel 501(BioRad), an organomercurial agarose gel, essentially as outlined by Reeve *et al.* (1977) for the affinity purification of transcripts initiated with nucleoside 5'[γ-S]triphosphates.

50µl of Affi-Gel 501 were suspended in 1ml Hg-agarose buffer in an Eppendorf tube, centrifuged at 13000g for 15 seconds, resuspended in a further 1ml Hg-agarose buffer and the process repeated for a total of four times. 2fmole ^{32}P-labelled RNA54 or [sA]RNA54 in 200µl Hg-agarose buffer was used to resuspend the gel pellet, and binding allowed to proceed for 10 minutes with constant mixing on a turntable. The gel was sedimented as above and washed 3 times with 1ml Hg-agarose buffer to remove unbound RNA. Elution was performed by resuspending the pellet in 200µl Hg-agarose buffer made to 10mM DTT and mixing for 10 minutes on a turntable. The gel was sedimented as above and washed with 1ml Hg-agarose buffer (+DTT). All supernatants were saved and Cerenkov counted. All manipulations were at room temperature.

Several Hg-agarose buffers were used:

α) 100mM sodium chloride, 40mM EDTA, 0.1% SDS, 50mM sodium acetate (pH4.9), (Marzluff and Huang, 1984).

β) 100mM sodium chloride, 10mM EDTA, 0.1% SDS, 10mM Tris-HCl (pH7.9), (Reeve *et al.* , 1977).

γ) 100mM potassium chloride, 0.2mM potassium EDTA, 0.5mM PMSF, 20% glycerol, 20mM triethanolamine (pH7.9) i.e. buffer D of Dignam *et al.* (1983) with triethanolamine as Tazi *et al.* (1986).

δ) As b) plus 0.5µg/µl yeast RNA, sodium salt (BDH).

ε) 100mM potassium chloride, 0.2mM EDTA, 3.2mM magnesium chloride,

20mM sodium acetate (pH4.9).

2.32. Oligonucleotide directed RNase H cleavage of RNA54 and [sA]RNA54

The total yield of gel purified RNA from a transcription reaction was cleaved in a 60µl reaction comprising 10mM HEPES-KOH (pH7.9), 50mM potassium chloride, 10% glycerol, 0.1mM potassium EDTA, 0.25mM PMSF, 0.25mM DTT, 8mM magnesium chloride, 2µg oligonucleotide ("170-185 cut", a 15 residue long oligonucleotide complementary to the RNA sequence 170 to 185 nucleotides downstream of the rabbit β-globin *in vivo* cap site), 40 units RNaguard (Pharmacia), 2 units RNase H (BRL). Incubation was at 37°C for 20 minutes. The reaction products were precipitated with ethanol, fractionated on 6% polyacrylamide, 20% formamide, 7M urea gels and cleaved RNA fragments (RNA 54-5'f and RNA 54-3'f) excised and eluted.

2.33. Chemical modification of RNA

Unlabelled RNA 50 (5% of an non-gel purified transcription reaction) was reacted with 20µl kethoxal (2-keto-3-ethoxybutyraldehyde, U.S. Biochemical Corp; 37mg/ml in 95% ethanol) or 5µl dimethyl sulphate (DMS; Aldrich) in 200µl N-1 buffer (50mM sodium cacodylate (pH6.8), 20mM magnesium acetate, 300mM potassium chloride, 5mM DTT) for 15 minutes and 5 minutes respectively at 30°C (Mougel *et al.* , 1987; Moazed *et al.* , 1986). The kethoxal reaction was stopped by adding 14µl 0.5M potassium borate (pH7.0), precipitating with ethanol and the RNA redissolved in 25mM potassium borate (pH7.0). The DMS reaction was stopped by ethanol precipitation and the RNA redissolved in water.

2.34. Treatment of extracts with micrococcal nuclease

Nuclear extract was pre-treated treated with micrococcal nuclease by making to 1mM calcium chloride adding 16 units micrococcal nuclease (Pharmacia) and incubating for 5 minutes at 30°C. The nuclease was then inactivated by making the extract to 2mM potassium EGTA (pH8.0).

Chapter 3.

Stereospecificity of Nucleases Towards Phosphorothioate-Substituted RNA: Stereochemistry of Transcription by T7 RNA Polymerase

3. Stereospecificity of Nucleases Towards Phosphorothioate-Substituted RNA: Stereochemistry of Transcription by T7 RNA Polymerase

This chapter describes the use of the two diastereoisomers of the substrate ATPαS and nucleases of known stereospecificity (previously characterised with respect to small molecules) to stereochemically characterise both the process of transcription by T7 RNA polymerase and the stereochemical discrimination of the nucleases with respect to large nucleic acids.

Most of the data presented and discussed in this chapter have previously been published in Griffiths *et al.* (1987).

3.1. RESULTS

3.1.1. Synthesis of Diastereoisomers of ATPαS

The epimeric mixture of ATPαS was examined by ^{31}P NMR spectroscopy (Fig. 3.1). The α-thiophosphate signal showed three peaks, representing overlap of two doublets (Fig. 3.1, *panel B*): each diastereoisomer gives rise to a doublet. One of the diastereoisomers was more abundant than the other, as reported from a previous synthesis (Burgers and Eckstein, 1978), and this more abundant product was assigned an Rp configuration on the basis of its chemical shift (Sheu and Frey, 1977). After separation of the two diastereoisomers using reverse phase HPLC (Fig. 3.2), the products were checked by a repeated HPLC separation. No cross-contamination of the two diastereoisomers was observed, indicating that less than 5% of the Rp-ATPαS isolate might represent Sp-ATPαS and that less than 10% of the Sp-ATPαS fraction might comprise Rp-ATPαS.

3.1.2. Substrate Preference and Transcription Kinetics

Each diastereoisomer was tested as a substrate for T7 RNA polymerase in a preliminary reaction. Sp-ATPαS was observed to act as a substrate whereas Rp-ATPαS at a much higher concentration does not seem to increase the rate of transcription above the background level seen in the absence of ATP

Figure 3.1. **^{31}P NMR spectrum of the mixture of Rp- and Sp-ATPαS produced by organic synthesis.**

The ^{31}P NMR spectrum of purified ATPαS (both diastereoisomers) is shown (*panel A*), and the peaks corresponding to the α, β, and γ phosphorus atoms are indicated. *Panel B* shows the signals from the α phosphorus in more detail, peaks deriving uniquely from the Sp or Rp diastereoisomers are labelled appropriately. *Panel C* and *panel D* show, in more detail, the signals from the β and γ phosphorus atoms respectively. NMR spectroscopy was performed with broad band proton decoupling on a Bruker AM300 spectrometer operating at 121.5 MHz.

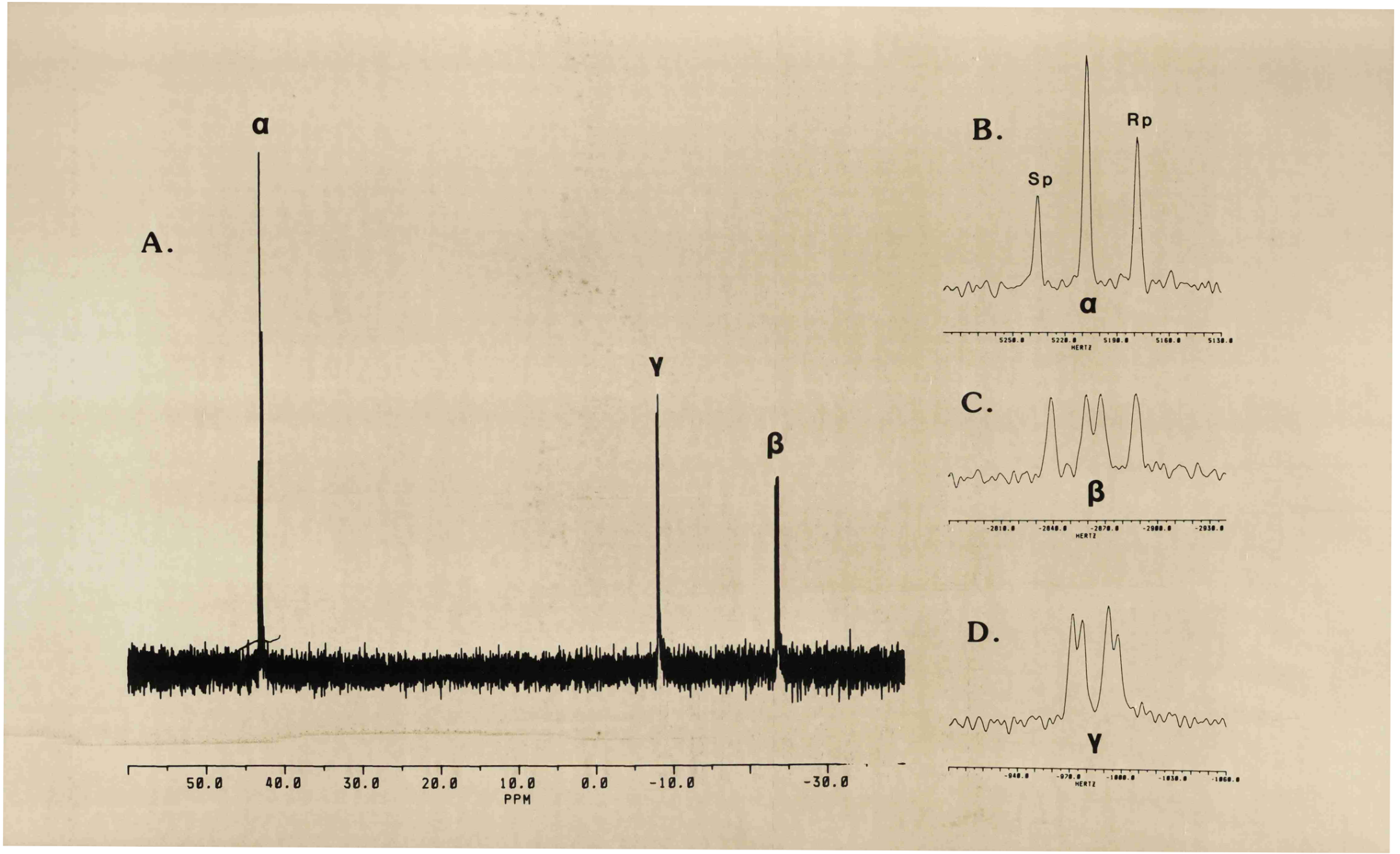
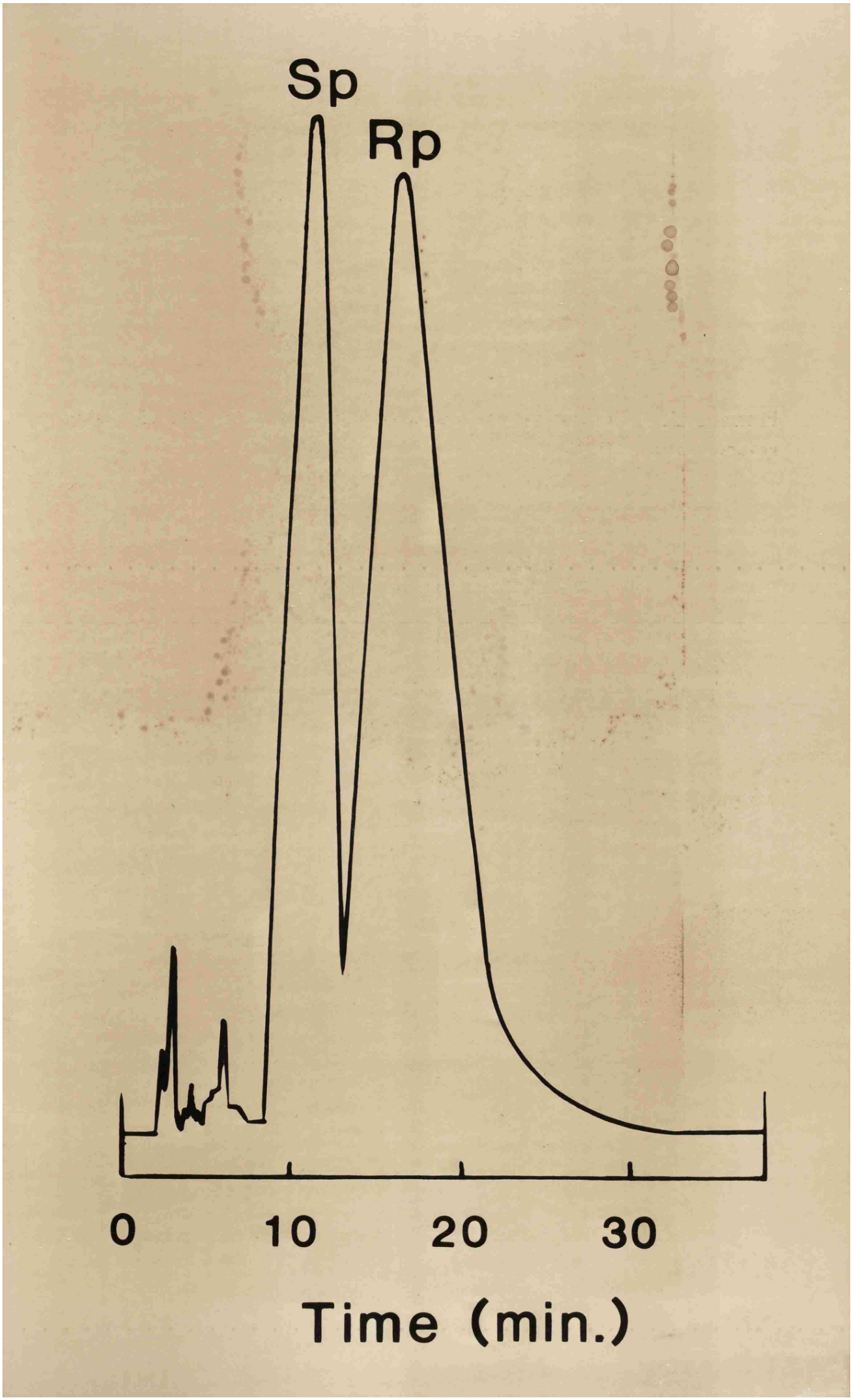

A.
α
γ
β
50.0
40.0
30.0
20.0
10.0
0.0
-10.0
-30.0
PPM
B.
Sp
Rp
α
HERTZ
C.
β
HERTZ
D.
γ
HERTZ

Figure 3.2. **Separation of the diastereoisomers of ATPαS using reverse phase HPLC.**


Sp
Rp
0
10
20
30
Time (min.)

(Table 3.4).

A series of determinations of the apparent K_M (K_{app}) were undertaken, each determination relying on single values for each of the time points used for estimation of the initial rate at each substrate concentration (see Discussion). Sample results are shown in Fig. 3.3 and all are summarised in Table 3.1. Although the significance of the kinetic constant may be doubtful (see Discussion), it represents a practical guide in assessing the requirements for optimal transcription efficiency.

It was not possible to determine reproducible values for the maximum velocity. However, duplicate rates were measured with ATP or ATPαS at identical concentrations (about six times K_{app}) in the course of a single experiment. Table 3.2 shows that on the two occasions when the measurements were performed there did not appear to be a substantial difference between the two substrates in the rate of reaction.

The poor transcription observed with Rp-ATPαS may be attributed to weak binding or to interference with the chemical reaction, i.e. a destabilised transition state. In order to test these possibilities, initial rates were measured with Sp-ATPαS at a concentration approximately corresponding to the K_{app} value in the presence or absence of the highest practicable concentration of Rp-ATPαS (Table 3.3). As with the previous comparison of rates, the use of single substrate concentrations allowed duplicate initial rates to be determined for each condition. No inhibition was detected, suggesting that Rp-ATPαS was not binding to a significant level. This is a little surprising, given the similarity in structure between Sp- and Rp-ATPαS, and the fact that Rp-ATPαS (or Rp-dATPαS) is a competitive inhibitor of all other polymerases investigated to date (Burgers and Eckstein, 1978; Burgers and Eckstein, 1979b; Romaniuk and Eckstein, 1982; Brody *et al.*, 1982; Bartlett and Eckstein, 1982; Eckstein and Jovin, 1983).

3.1.3. Configurational Analysis of Transcripts

Three nucleases have been shown to discriminate strongly between the two diastereoisomers of phosphorothioate linkages in dinucleotides and between these linkages and normal phosphodiesters. Snake venom phosphodiesterase (SVPD) has been shown to prefer Rp- to Sp-thiophosphodiester linkages and to act most rapidly on phosphodiester linkages (Bryant and Benkovic, 1979; Burgers and Eckstein, 1979a); nuclease S1 cleaves Sp-thiophosphodiester linkages at a rate similar to phosphodiester linkages (Potter *et al.*, 1983a) and, as with nuclease P1 (Potter *et al.*, 1983b), no cleavage of Rp-thiophosphodiester linkages has

Table 3.1. **Results of experiments to determine K_{app}.**

Substrate	K_{app} values (μM)	Mean (μM)
ATP	16.5 , 29	23
(Sp)-ATPαS	23 , 12 , 10 , 16	15

Table 3.2. **Comparisons of initial rates of transcription with ATP or ATPαS.**

Experiment	Substrate	Initial Rates (c.p.m./min)	Means (c.p.m./min)
1	ATP	931 , 1092 , 999	1007
1	(Sp)-ATPαS	959 , 1431	1195
2	ATP	2127, 2153 , 1727	2002
2	(Sp)-ATPαS	2041 , 1378	1710

Table 3.3. **Initial rates of transcription with 16.9μM (Sp)-ATPαS in the presence or absence of 597μM (Rp)-ATPαS.**

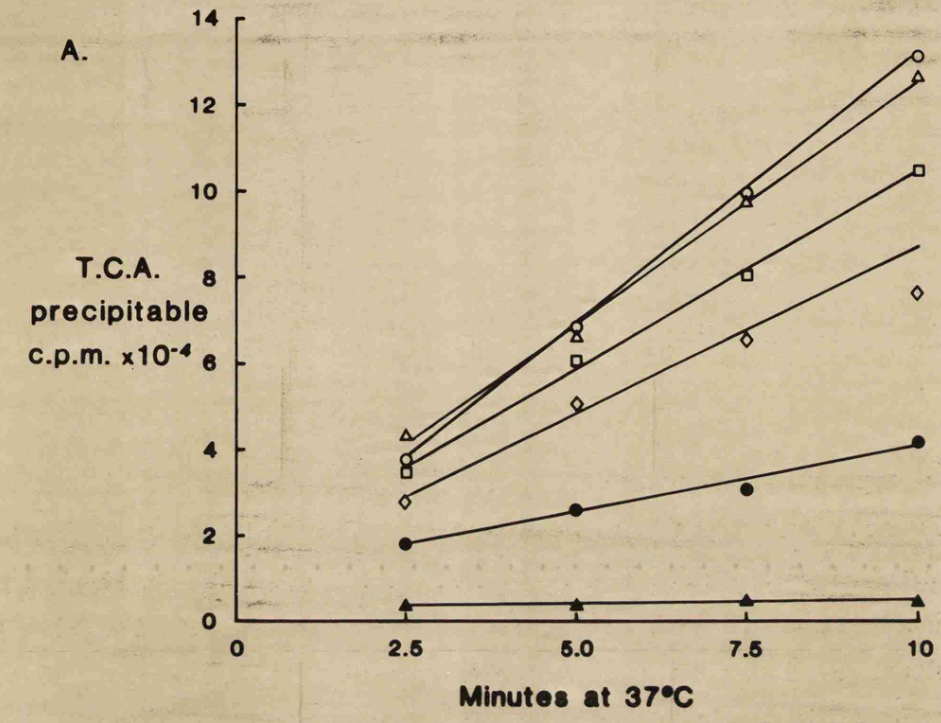
Experiment	Initial Rates (c.p.m./min)	Means (c.p.m./min)
+ (Rp)-ATPαS	977 , 1277 , 1105	1120
- (Rp)-ATPαS	1082 , 1048	1065

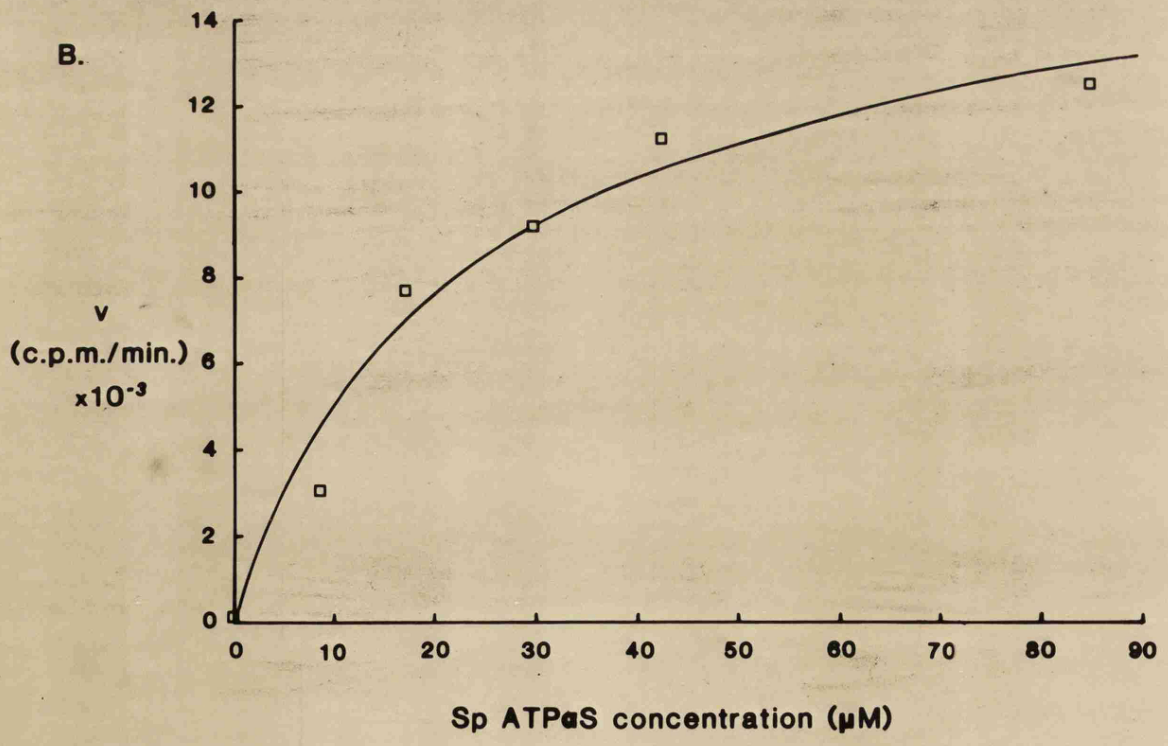
Table 3.4. **Initial rates of transcription with ATP, (Sp)-ATPαS, (Rp)-ATPαS, or no ATP (or ATPαS).**

Substrate	Initial Rates (c.p.m./min)	Means (c.p.m./min)
No ATP (or ATPαS)	30, 26	28
ATP	990, 1045	1018
(Sp)-ATPαS	1005, 1050	1028
(Rp)-ATPαS	10, 15	13

Figure 3.3. **Determination of K_{app} for T7 RNA polymerase with Sp-ATPαS.**

One of the experiments summarised in Table 4.1 is shown. Initial rates were calculated from the upper graph at the following ATPαS concentrations: ▲, OμM: ●, 8.5μM; ◇, 17μM; □, 30μM; △, 42μM; ○, 85μM. In all cases the incorporation of ^{32}P from [α-^{32}P]CTP was followed. The slopes were calculated by linear regression. The lower graph shows the initial rate plotted against Sp-ATPαS concentration, with a rectanbular hyperbola fitted by an iterative least squares procedure (Millar, 1984).


A.
T.C.A. precipitable c.p.m. x10⁻⁴
14
12
10
8
6
4
2
0
0
2.5
5.0
7.5
10
Minutes at 37°C


B.
v (c.p.m./min.) x10⁻³
14
12
10
8
6
4
2
0
0
10
20
30
40
50
60
70
80
90
Sp ATPαS concentration (μM)

been detected.

In order to use these nucleases for configurational assignments in RNA, it would be desirable to characterise the rates of digestion of RNA containing only Sp or Rp-thiophosphodiester internucleotide linkages (Sp- or Rp-[S]RNA). However, only Rp-[S]RNA can be made with RNA polymerases characterised to date. Nonetheless, stereospecificity could be inferred if there were consistent differences in rates of digestion between RNA with and without thiophosphodiester linkages: Rp-[S]RNA would be expected to be an extremely poor substrate for nuclease S1 but to be digested by SVPD, whereas Sp-[S]RNA would be expected to be a comparatively poor substrate for SVPD but to be almost comparable with normal RNA as a substrate for nuclease S1.

The two substrates used were synthesised by T7 RNA polymerase on the same template, giving rise to transcripts with major products of the same length and similar distribution of size for the minor products, as estimated by polyacrylamide gel electrophoresis. One substrate contained only normal phosphodiester linkages and [^{32}P]C whereas the other was expected to contain only thiophosphodiester linkages and [^{35}S]U. The two classes of transcript were mixed in a molar ratio of approximately 25:1 and aliquots digested for a fixed time with a range of concentrations of each enzyme. The assumption was that if no discrimination took place then, regardless of the ratio of mixed substrates, a single curve should be produced; alternatively, a high degree of discrimination would result in rapid digestion of the [^{32}P]RNA, and the digestion profiles of each substrate would be almost unaltered by the presence of the other substrate.

Figure 3.4 shows that there is comparatively little difference between the SVPD digestion profiles of the two substrates, the [^{35}S]-labelled RNA being digested slightly more rapidly than the [^{32}P]-labelled RNA, whereas nucleases P1 and S1 act more slowly on thiosubstituted RNA. The digestion by nuclease P1 of these thiophosphodiester linkages is so slow that the presence of thiosubstituted RNA does not affect the activity on normal RNA, and *vice-versa*; in seperate experiments, the curves from digestion of similar initial levels of the two RNAs, digested individually, were found to be almost perfectly superimposable on those shown in figure 3.4c. Thus, it is very probable that the nucleases retain stereospecific preferences with long RNA molecules and that T7 [S]RNA has an Rp configuration.

This conclusion was reinforced by comparing digestion of T7 [S]RNA with [S]RNA known to be of Rp configuration, i.e. synthesised by *E. coli* RNA polymerase (Eckstein *et al.*, 1976; Burgers and Eckstein, 1978). Transcripts were synthesised by T7 and *E. coli* RNA polymerases in the presence of ATPαS, CTPαS, GTPαS and [^{35}S]UTPαS. Approximately equal molar yields, in

Figure 3.4. **Experiments to determine whether nucleases discriminate between RNA and [S]RNA.**

Transcripts of 830 nucleotides were labelled with ^{32}P (■) or, if all internucleotidic linkages were phosphorothioate substituted, with ^{35}S (○). Mixed transcripts were digested at the enzyme concentrations shown. The proportion of labelled material rendered soluble in T.C.A. at each enzyme concentration was determined in triplicate; for each value the mean is plotted together with the standard error of the mean (where error bars are not shown, the dimensions of the point exceed those of the error bar). Enzyme concentrations were plotted on a logarithmic scale.

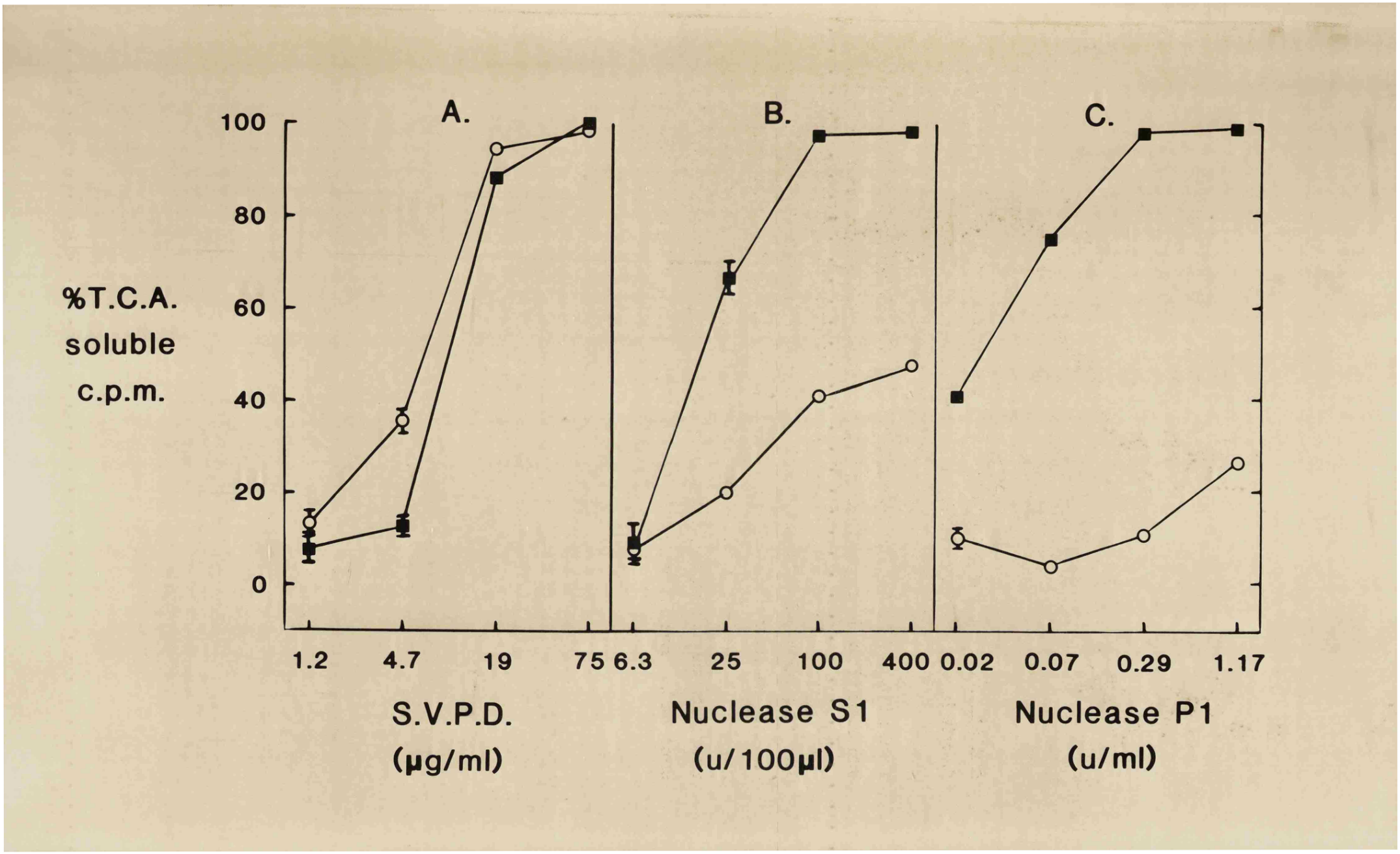
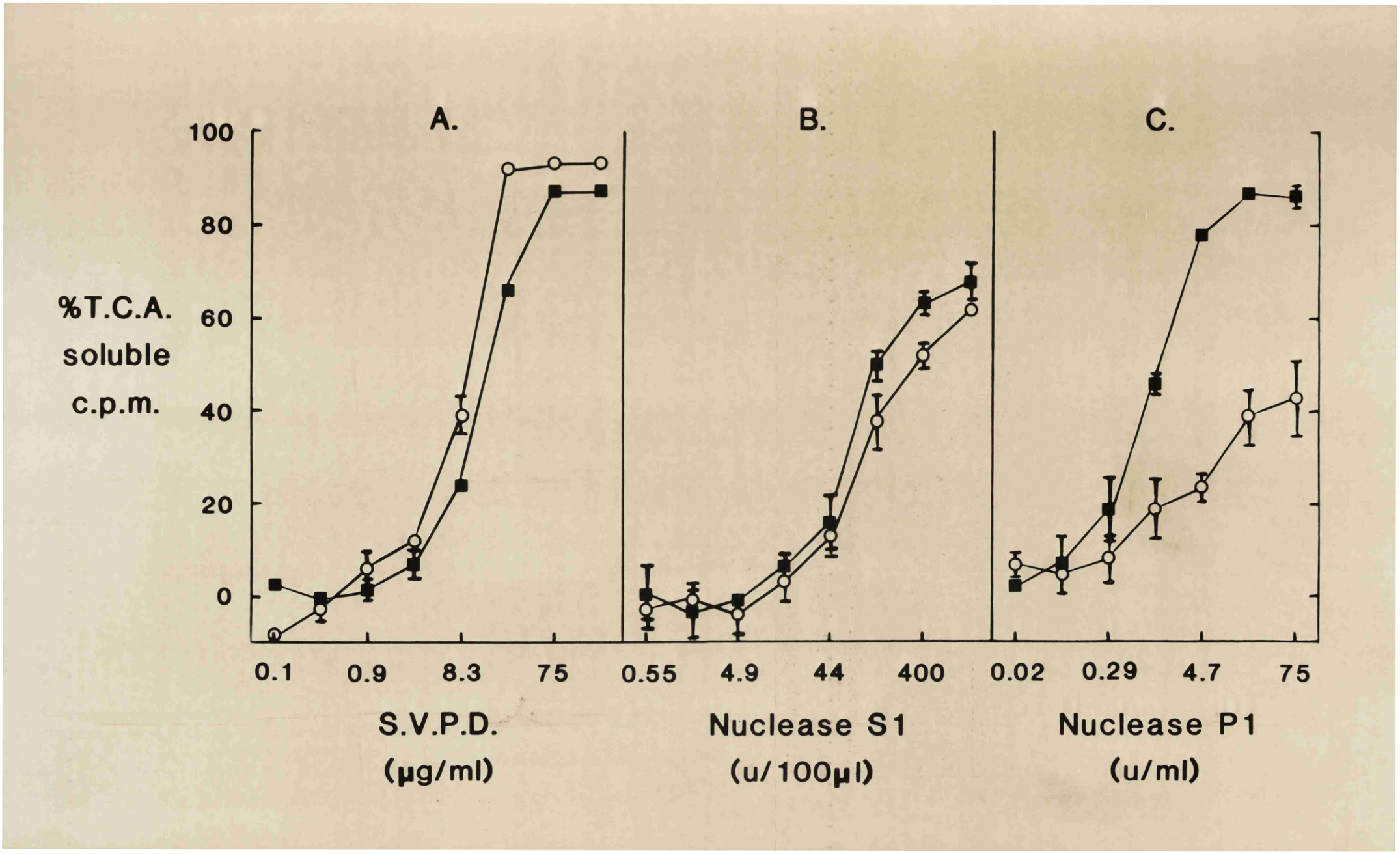

%T.C.A.
soluble
c.p.m.
100
80
60
40
20
0
A.
B.
C.
1.2
4.7
19
75
6.3
25
100
400
0.02
0.07
0.29
1.17
S.V.P.D.
(µg/ml)
Nuclease S1
(u/100µl)
Nuclease P1
(u/ml)

Figure 3.5. **Comparisons of nuclease susceptibility of [S]RNA transcribed by T7 RNA polymerase (○) with that of Rp-[S]RNA transcribed by *E. coli* RNA polymerase (■).**

Both transcripts were labelled with ^{35}S, and were digested in separate reactions. The means of determinations in triplicate are shown together with their standard error.


A.
B.
C.
100
80
60
40
20
0
%T.C.A.
soluble
c.p.m.
0.1
0.9
8.3
75
0.55
4.9
44
400
0.02
0.29
4.7
75
S.V.P.D.
(µg/ml)
Nuclease S1
(u/100µl)
Nuclease P1
(u/ml)

terms of nucleotides incorporated, were incubated under standard reaction conditions with a series of dilutions of each enzyme. Undigested material was measured by TCA precipitation. The results shown in figure 3.5 demonstrate that both RNAs are almost equally susceptible to digestion by SVPD and nuclease S1, even though there is more double-stranded template DNA present in the samples transcribed by *E. coli* RNA polymerase. However, nuclease P1 digests *E. coli* RNA polymerase transcripts more readily than those produced by T7 RNA polymerase. It should be noted that on the basis of studies with dinucleotides (Potter *et al.*, 1983b), the *E. coli* Rp-[S]RNA should be resistant to nuclease P1; if the T7 transcripts contained linkages of the Sp configuration they should be digested more readily. Therefore, the difference in rates of digestion cannot be attributed to the effects of thiophosphodiester configuration; instead, it was anticipated that the enzyme might be sensitive to differences in length or secondary structure of the transcripts. Both these aspects will differ substantially: the products of transcription of a cloned eukaryotic gene by T7 RNA polymerase have a sharply defined length, whereas those of M13 DNA by *E. coli* RNA polymerase are heterogeneous and shorter (data not shown). However, the control experiment in figure 3.6a shows that the difference in the rates of digestion by nuclease P1 is confined to thio-substituted RNA.

The analysis of TCA-precipitable material does not define the extent of digestion very precisely. Oligonucleotides as well as mononucleotides will escape precipitation. This allows the possibility that NTPs, if they were present as a contaminant in the stocks of NTPαS, will have been incorporated during transcription and will be digested preferentially by the stereospecific nucleases. With a sufficiently high level of contamination this would undermine the interpretation of the results with SVPD and nuclease S1 in Figures 3.4 and 3.5. In order to prove that thiophosphodiester linkages are being cleaved, we carried out the experiment shown in Figure 3.7. Transcripts were prepared using [^{35}S]UTPαS as the only source of thiophosphodiester linkages. These were digested by the same three nucleases, and the products were distinguished by thin-layer chromatography. Only the results from nuclease P1 are shown; at high concentrations of nuclease S1 and SVPD the lanes became very smeary. The appearance of [^{35}S]UMP confirmed that the thiophosphodiester linkages were being hydrolysed in the case of all three enzymes.

The rate of appearance of [^{35}S]UMP from both transcripts was identical with nuclease P1. Since we did not observe the discrepancy seen with TCA precipitation, the same experiment was carried out with transcripts which contained only thiophosphodiester linkages. The result (Figure 3.7) now matched those obtained with analysis by TCA precipitation: the T7

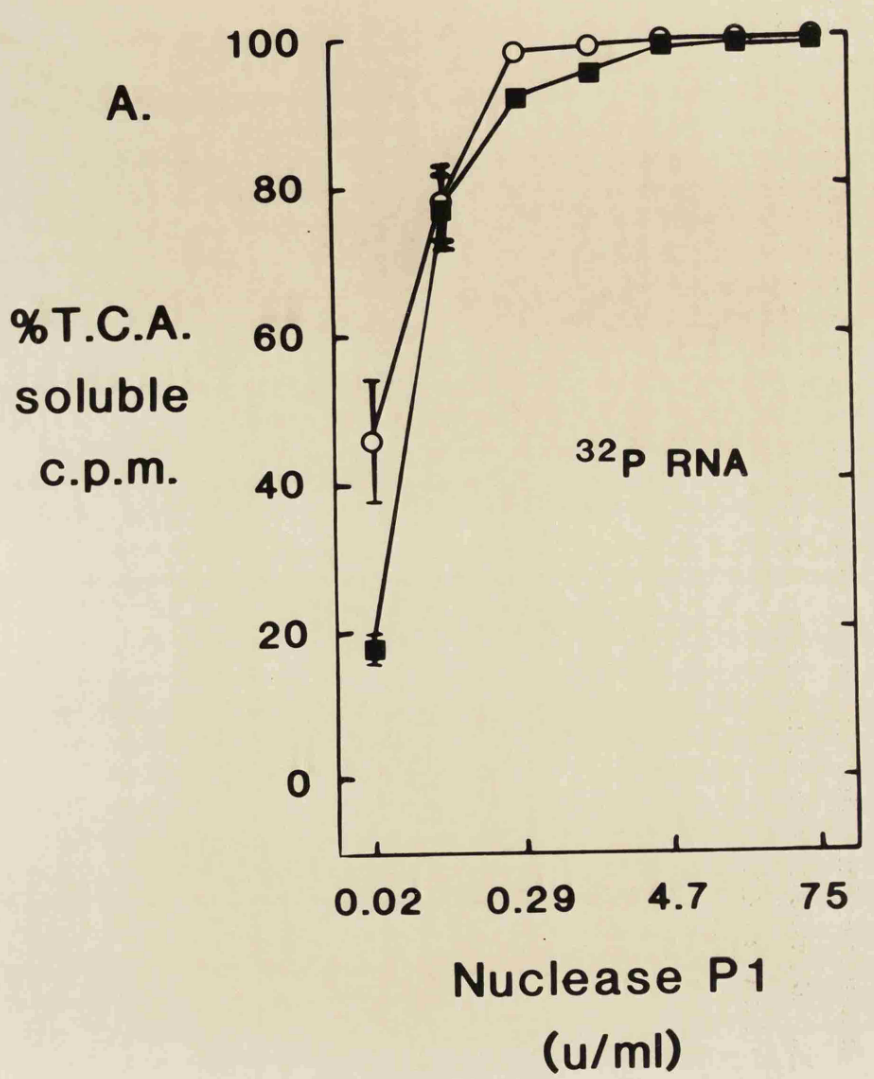
Figure 3.6. **Investigations of nuclease P1 activity.**

A. Comparison of nuclease P1 activities on RNA transcribed by T7 and *E. coli* RNA polymerases.

^{32}P-labelled RNA, lacking internucleotidic phosphorothioates, was transcribed by T7 RNA polymerase (○) and *E. coli* RNA polymerase (■); after digestion, the results were plotted as described in the legend to figure 3.5.

B. Experiments to determine whether nuclease P1 discriminates between DNA and [S]DNA.

DNA synthesis products were labelled with ^{32}P (■) or, when all internucleotidic linkages were phosphorothioate-substituted, with ^{35}S (○, △). Digestions were performed with a mixture of the two substrates (■, ○) or with one (△).


A.
%T.C.A.
soluble
c.p.m.
100
80
60
40
20
0
^{32}P RNA
0.02
0.29
4.7
75
Nuclease P1
(u/ml)

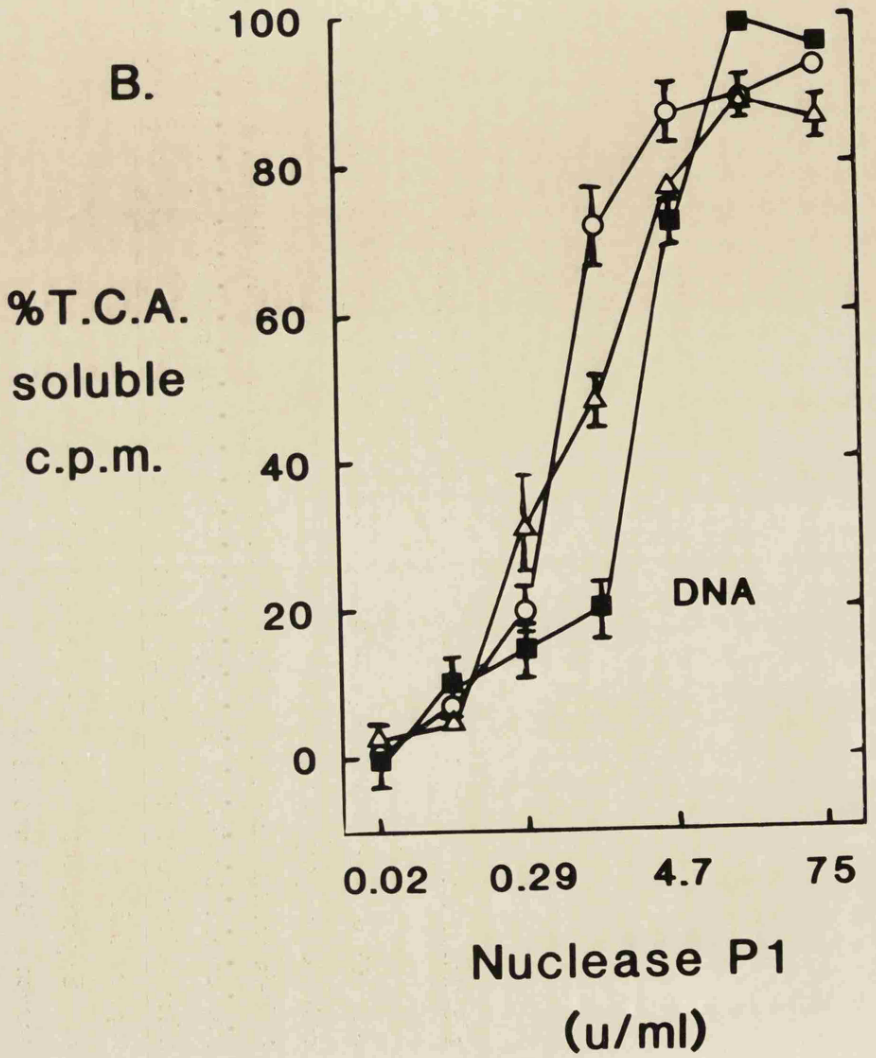
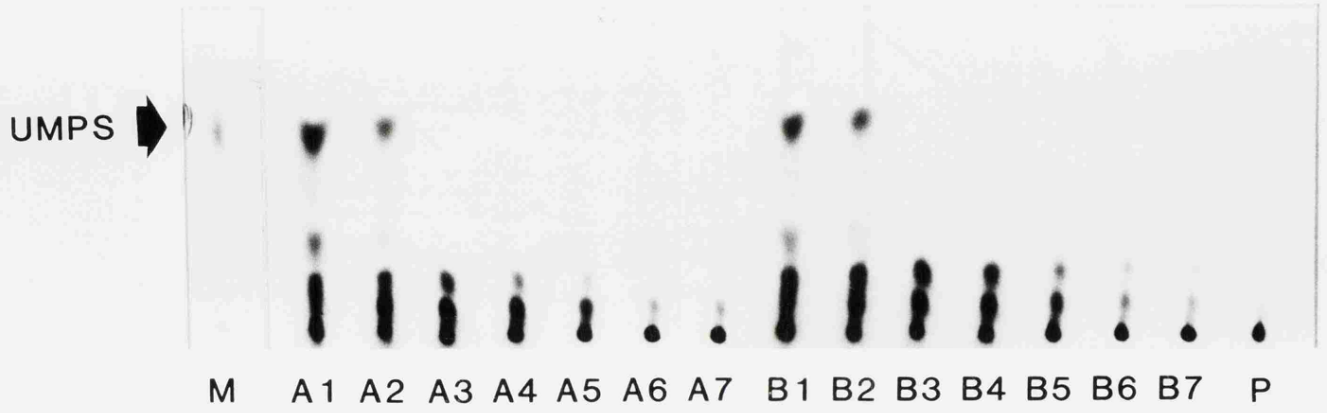
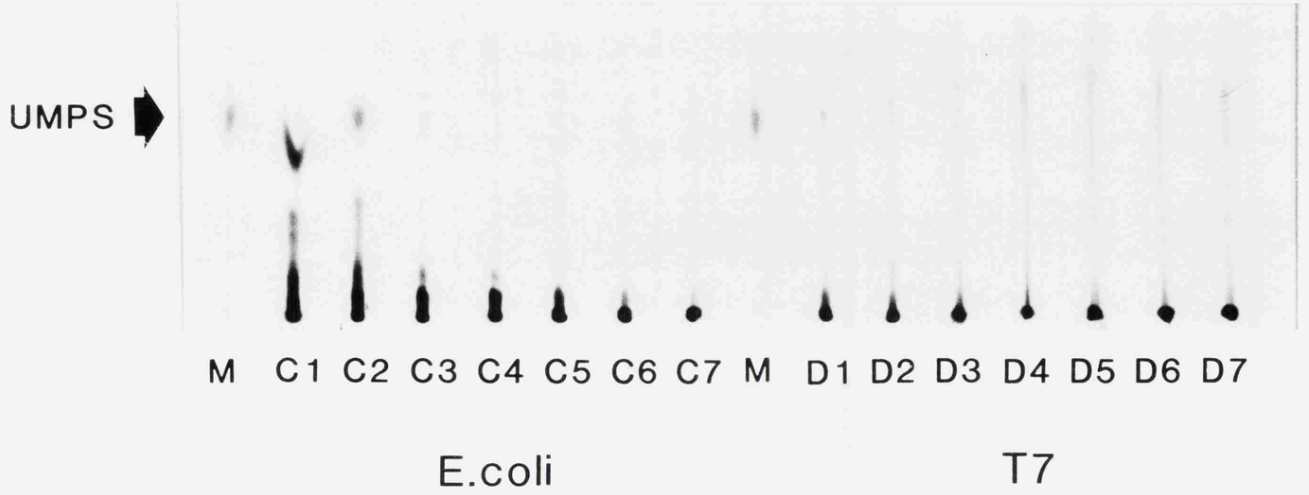

B.
%T.C.A.
soluble
c.p.m.
100
80
60
40
20
0
DNA
0.02
0.29
4.7
75
Nuclease P1
(u/ml)

Figure 3.7. **Chromatography on P.E.I.-cellulose to show cleavage at internucleotidic phosphorothioate linkages by nuclease P1.**

Samples marked A and B were the products of transcription by *E. coli* and T7 RNA polymerase respectively in the presence of ATP, CTP, GTP and [α-^{35}S]UTP; those marked C and D were transcribed in the presence of ATPαS, CTPαS, GTPαS and [α-^{35}S]UTP. Nuclease P1 digestions were performed with successive ten-fold dilutions of enzyme from sample 1 (0.83u/μl) to sample 6; sample 7 in each series was incubated without enzyme. Sample P represents digestions of [α-^{35}S]UTP with nuclease P1 at 0.83u/μl (showing that contamination of the transcript would not generate [^{35}S]UMP; M is [^{35}S]UMP generated by SVPD digestion of [α-^{35}S]UTP.


UMPS
M A1 A2 A3 A4 A5 A6 A7 B1 B2 B3 B4 B5 B6 B7 P
E.coli
T7


UMPS
M C1 C2 C3 C4 C5 C6 C7 M D1 D2 D3 D4 D5 D6 D7
E.coli
T7

transcripts were more refractory to nuclease P1 than those synthesised by *E. coli* RNA polymerase.

Nuclease P1 is regarded as showing no activity at all towards Rp-thiophosphodiester linkages in dinucleotides (Potter *et al.*, 1983b). Three possible reasons for digestion of Rp-[S]RNA are that contaminating phosphodiester linkages are being cleaved (shown above to be an insufficient explanation), that digestion of dinucleotides is more stringent than that of polynucleotides, or that RNA and DNA are digested with different stringencies. To test the last explanation, parallel syntheses were performed to produce DNA and Rp-[S]DNA (Burgers and Eckstein, 1979b) by primed synthesis with DNA polymerase 1. Mixed products, as well as Rp-[S]DNA alone (in this case, in the presence of less than half as much unlabelled template DNA), were digested by nuclease P1. Figure 3.6b shows that nuclease P1 does not discriminate against these Rp-thiophosphodiester linkages; it is clear that nuclease P1 stringency depends on the difference between dinucleotides and polynucleotides.

3.2. DISCUSSION

The work described above sought to establish whether nucleases could be used to assign the configuration of diastereomeric internucleotidic phosphorothioates in thio-substituted RNA and thus, knowing which diastereoisomer of ATPαS was a substrate for T7 RNA polymerase, whether the reaction proceeded with inversion of configuration.

After synthesis of ATPαS the ratio of Sp:Rp diastereoisomers was approximately 1:2, shown by ^{31}P NMR measurements based on known assignments (Burgers and Eckstein, 1978; Sheu and Frey, 1977). The diastereoisomers were separated by reverse phase HPLC and the peaks were assigned by their intensities. The Sp diastereoisomer was found to elute before the Rp. Separation had previously been performed using ion-exchange HPLC (Burgers and Eckstein, 1978).

The initial kinetic analysis of transcription showed that Sp-ATPαS was incorporated efficiently, but there appeared to be slower incorporation in the presence of Rp-ATPαS. Much, if not all, of the latter can be attributed to contamination of the other NTPs, rather than to desulphurisation of ATPαS or incorporation of Rp-ATPαS. Later experiments suggested that Rp-ATPαS did not bind significantly to the enzyme.

In order to establish whether Sp-ATPαS could be regarded as a good or a poor substrate, the values of K_{app} for this substrate and for ATP were

determined. Precise determinations of kinetic constants proved to be difficult. It was observed that incubations on ice during preparation of the reactions gave extremely unreliable results. Thus, precise incubation times could be measured for only a limited number of samples in parallel reactions with different concentrations of substrate. In practice, in one experiment it was possible to measure about 24 reactions, i.e. four time points for each of six concentrations of substrate (0 and five appropriate concentrations). It was not possible to generate duplicate values for each point, since the actual rates varied from day to day. However, the K_{app} values obtained did not vary very widely, and the mean values for Sp-ATPαS and ATP were very similar.

The variability of rates may have arisen from the known instability of the enzyme (Chamberlin and Ryan, 1982) and from the decay of the [α-^{32}P]CTP substrate, for, despite attempts to correct for the decrease in activity of the labelled substrate with time, the observed rate declined disproportionately (data not shown). Thus, any quantitative estimate of maximal velocity was impossible; furthermore it would have reflected the substantial period of each round of transcription involved in re-initiation (Chamberlin and Ring, 1973). However, it was possible to demonstrate that, in so far as the elongation rate affects the observed reaction rate, the Sp-ATPαS does not contribute to a reduction in rate. It should be noted that there are difficulties in interpreting the apparent K_M. Reactions with T7 polymerase show a lag of 10 to 15 seconds before the steady-state is reached (Chamberlin and Ring, 1973). With a maximum rate of elongation of 200 nucleotides per second (Golomb and Chamberlin, 1974), only about 4 seconds will be required for elongation of each transcript. Longer transcripts will require longer elongation times, and the contribution of initiation will be correspondingly reduced; thus it would be predicted that longer transcripts will give rise to higher values of K_{app}, which approach the K_M for elongation. It is of interest, and consistent with the above, to note that the value for the K_{app} of ATP determined in this study is 16-30μM, whereas with transcription of T7 DNA (where the average length of transcripts should be higher) a value of 47μM has been reported (Chamberlin and Ring, 1973). This uncertainty about the significance of the value of K_{app} has been circumvented in studies with T3 RNA polymerase or *E. coli* RNA polymerase where preinitiated complexes have been studied during elongation (McAllister *et al.*, 1973; Rhodes and Chamberlin, 1974). This approach was not feasible with the short transcripts used in this study.

The assignment of configuration by nuclease sensitivity yielded unexpected results. An early example of the application of "stereospecific" nucleases involved the assignment of configuration of DNA containing thiophosphodiester linkages synthesised by *E. coli* DNA polymerase (Burgers

and Eckstein, 1979b) where, as in this thesis, it was not possible to synthesise a polymer with the opposite configuration. It was shown that the relative rates of digestion with SVPD of [S]DNA and DNA were similar to those of Rp-[S]RNA compared with RNA; results from dinucleotides were also more consistent with an Rp than an Sp assignment. This was confirmed by an independent method (Brody and Frey, 1981). These experiments did not establish whether SVPD was sensitive to configuration in polynucleotides, but the assumption had been reasonable.

Based on this precedent, experiments were performed to determine whether [S]RNA synthesised by T7 RNA polymerase behaved like [S]RNA of known configuration with respect to nucleases of opposite stereochemical preference. The comparison of the two RNAs demonstrated that, with respect to two nucleases reported to have opposite preferences, the two RNAs behaved as if they were identical. The use of thin-layer chromatography confirmed that the rates of cleavage of thiophosphodiesters were the same. In all these experiments a constant time of incubation was used and the enzyme concentration varied in order to keep the background rate of chemical hydrolysis constant.

To demonstrate that the nucleases retained any stereochemical preferences on long [S]RNA, or on RNA rather than the DNA dinucleotides tested previously, the rates of digestion of [S]RNA and normal RNA were compared. The two transcripts were mixed in an arbitrary ratio (1:25), and digestion was followed by TCA precipitation. The concentration of phosphodiester linkages of each transcript was less than one-thousandth of the K_M reported for dinucleotides with nucleases P1, S1 and SVPD (Potter *et al.*, 1983a; Potter *et al.*, 1983b, Burgers and Eckstein, 1979a). A plot of low values of percentage of reactant solubilised in a given time against enzyme concentration is equivalent to a plot of V/S against enzyme concentration. If simple Michaelis-Menten kinetics are followed, this plot should be linear, or, with a semi-logarithmic plot as in Figures 3.4 to 3.6, exponential.

The exponential curves for two substrates should be superimposable by a displacement along the X-axis. The divergence of the lines for the two substrates seen in Figure 3.4 for nucleases S1 and P1 cannot be explained by the artefacts of an assay based on TCA precipitation. It is possible that thiosubstitution of RNA results in a shift of the relative rates of endonucleolytic and exonucleolytic activity, which would give rise to altered sensitivities to TCA precipitation; alternatively, if the stereospecificity of the nucleases is greater with smaller substrates, discrimination may increase with the extent of endonucleolytic digestion. This divergence means that measurements and comparisons of rates at any given nuclease comparison cannot be generalised. However, a comparison of rates can be made by

determining the enzyme concentrations required to achieve a fixed but arbitrary extent of digestion, when the distribution of sizes of material in the TCA-soluble and TCA-precipitable fractionswill be similar for any two substrates that are subject to the same enzymatic activities. In order to render 25% of the substrate soluble in TCA (when initial steady-state rates should still apply) the ratios of the required enzyme concentrations with [S]RNA and normal RNA are: SVPD, 0.4; nuclease S1, 4; nuclease P1 ≥100.

When characterised with dinucleotides, SVPD has been shown to digest normal linkages, Rp-[S] linkages and Sp-[S] linkages with approximate relative values for Vmax as follows: 185,000 : 1700 : 1 (Burgers and Eckstein, 1979a). Even given that SVPD digests Rp-[S]RNA and T7 [S]RNA equally well, i.e. that T7 [S]RNA has an Rp configuration, our results suggest that SVPD may not discriminate sufficiently well (if at all) to be used alone for configurational assignments of [S]RNA. In previous experiments, a much reduced level of discrimination has been found with Rp-[S] poly(A), which was digested 10-fold less rapidly than poly(A) (Burgers and Eckstein, 1979b). $Vmax/K_M$ values have been determined (Potter *et al.*, 1983a) for nuclease S1 digestion (at 4 units per 100µl) of normal, Sp-[S] and Rp-[S] linkages in dinucleotides, producing relative values: 10 : 1 : (0). Thus, although nuclease S1 discriminates against T7 [S]RNA (Figure 3.4), an assignment of Rp configuration could only be justified by the identical behaviour of known Rp-[S]RNA (Figure 3.5). Rates of digestion have not been published for nuclease P1, but Rp-[S] dinucleotides have been shown to resist digestion (Potter *et al.*, 1983b); from the results of this study, nuclease P1 seems to discriminate sufficiently against Rp-[S]RNA to suggest that this nuclease will be able to distinguish Sp from Rp configurations in RNA.

It is very striking that nucleases S1 and P1 digest thiophosphodiester linkages which, in the case of *E. coli* RNA, were certainly in the Rp configuration, whereas this had not been observed for either enzyme with studies on dinucleotides (Potter *et al.*, 1983a and 1983b). The thin layer chromatographic data are important, excluding the possibility that the observed digestion was due to cleavage of normal phosphodiester bonds. The high level of digestion seen in figure 3.5 suggests that the presence of occasional Sp-linkages in the transcript (produced by a low level of incorporation of contaminating Rp-NTPαS during transcription) can be excluded as an alternative explanation of this phenomenon.

In conclusion, it seems that discrimination between normal and thio-phosphodiester linkages is retained with RNA, but that the thiophosphodiester linkages are digested more readily in RNA than in dinucleotides. Thus, even the unfavourable configuration is digested, but at a lower rate than normal phosphodiester linkages. Since nuclease P1 does

not show even this level of discrimination with the products of DNA polymerase 1 incorporation, it is possible that the length of the substrate determines the reduced discrimination. This phenomenon might be related to binding of substrates, if subsidiary binding sites adjacent to the catalytic centre are able to bind extra nucleotides in longer molecules and therefore partly compensate for possible weak binding of substrates containing unfavourable thiophosphodiester linkages. Although the evidence is circumstantial, this study seems to infer that some discrimination between different configurations of thiophosphodiester linkages in [S]RNA is retained and that polymerisation by T7 RNA polymerase proceeds with inversion of configuration.

The reduced sensitivity of [S]RNA to nuclease P1 when the RNA was synthesised by T7 rather than *E. coli* RNA polymerase, in the presence of all four NTPαS substrates, was remarkable. In contrast, identical sensitivities were seen with the digestion of unsubstituted RNA, assayed by TCA precipitation, and with RNA wherein only one in four internucleotidic links was thio-substituted, with an assay based on thin-layer chromatography. Because the stereochemistry of the internucleotidic link in RNA synthesised by the *E. coli* RNA polymerase is of the Rp configuration it should be fully resistant to nuclease P1 (see above); the assays with normal or only partially thio-substituted RNAs showed that neither length nor RNA structure affected the digestion patterns. One possible explanation is that endonucleolytic activity was affected more than exonucleolytic activity by complete thio-substitution, and that therefore differences in transcript length became important. This cannot, of course, be estimated in studies with a dinucleoside phosphorothioate substrate. However, it can be seen that SVPD, an obligate exonuclease, did not show marked preferences for RNAs according to the polymerase used for transcription.

The application of stereospecific nucleases to the study of RNA processing reactions is, therefore, far from straightforward. In order to assign the configuration of internucleotidic RNA phosphorothioates it will probably prove necessary to rely on stereospecific digestion by nucleases such as those used in this study. To analyse the configuration of a particular phosphorothioate linkage it must probably first be isolated in an RNA dinucleotide or at the very least a small oligonucleotide. It is possible, however, that full stereospecificity will not be seen with RNA dinucleotides at very low concentrations and hence it is important to investigate the reasons for the reduced stereospecificity in the digestions seen here: it may not simply be a result of the use of polynucleotide substrates. Hence, to be certain of making a correct assignment of configuration, synthesis of pure diastereoisomers of the dinucleotide or small oligonucleotide being analysed

would seem to be essential. Parallel kinetic studies of digestion of the control di/oligonucleotide and the di/oligonucleotide under investigation should indicate whether or not the stereospecificity of the analytical enzymes used is sufficient, and if so, provide an accurate configurational assignment.

Chapter 4.

Substitution of Pre-Messenger RNA with Phosphorothioate Linkages Reveals a New Splicing Related Cleavage Reaction

4. Substitution of Pre-Messenger RNA with Phosphorothioate Linkages Reveals a New Splicing Related Cleavage Reaction

This chapter describes the investigation of splicing using pre-mRNA substituted during transcription with internucleotidic phosphorothioates. Certain of these substrates gave rise to an unexpected and novel exon 2 product, the characterisation of which is described.

Most of the data presented and discussed in this chapter have previously been published in Griffiths *et al.* (1988).

4.1. RESULTS

In vitro transcripts were prepared from a portion of rabbit β-globin gene, from 9 nucleotides preceding the normal transcription start to +310 (i.e., all of exon 1, IVS-1 and 39 nucleotides of exon 2), which had been cloned into **mICE10** (Figure 4.1A; Turnbull-Ross *et al.*, 1988; Eperon, 1986). Transcription with T7 RNA polymerase was terminated by a *Hind* III cleavage of the replicative form, 50 nucleotides to the 3' side of the 3' splice site on the RNA-sense strand. In addition, single-stranded viral DNA was prepared for transcription by extension of a complementary strand from an oligonucleotide primer with a 5' terminus 50 nucleotides beyond the 3' splice site (Turnbull-Ross *et al.*, 1988; Eperon, 1986). Experiments with templates which lack the insert in **mICE10** have shown that transcripts will extend to the 5' terminus of the staggered end of the cleavage site in *Hind* III-cut template, and to the 5' terminus of the oligonucleotide when primed viral DNA is used as a template (Turnbull-Ross *et al.*, 1988). Therefore, the two transcripts will give rise to transcripts ending 54 and 50 nucleotides, respectively, beyond the 3' splice site. These transcripts will be referred to as RNA 54 and RNA 50 respectively.

Transcripts were also prepared from a mutant gene in which the sequence AG/G at the 3' splice site (the slash indicates the splice site) had been changed to AG/A. Transcription of the mutated gene in the presence of ATPαS would place a phosphorothioate linkage at the 3' splice site, whereas this would be absent after similar transcription of the wild-type sequence.

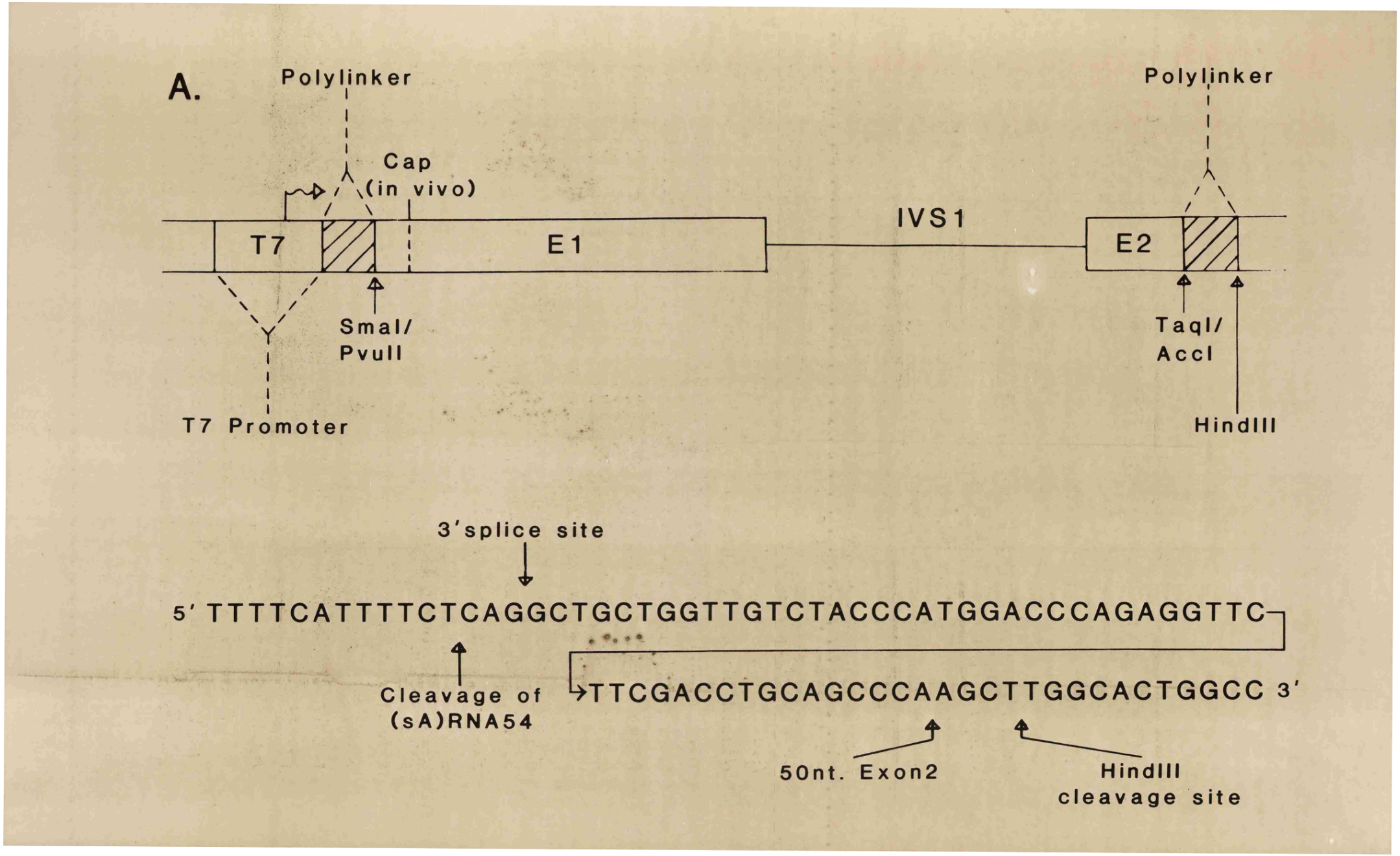
Transcripts were prepared in reactions containing, in each case, one of the four nucleoside α-thiotriphosphates. These transcripts were used for

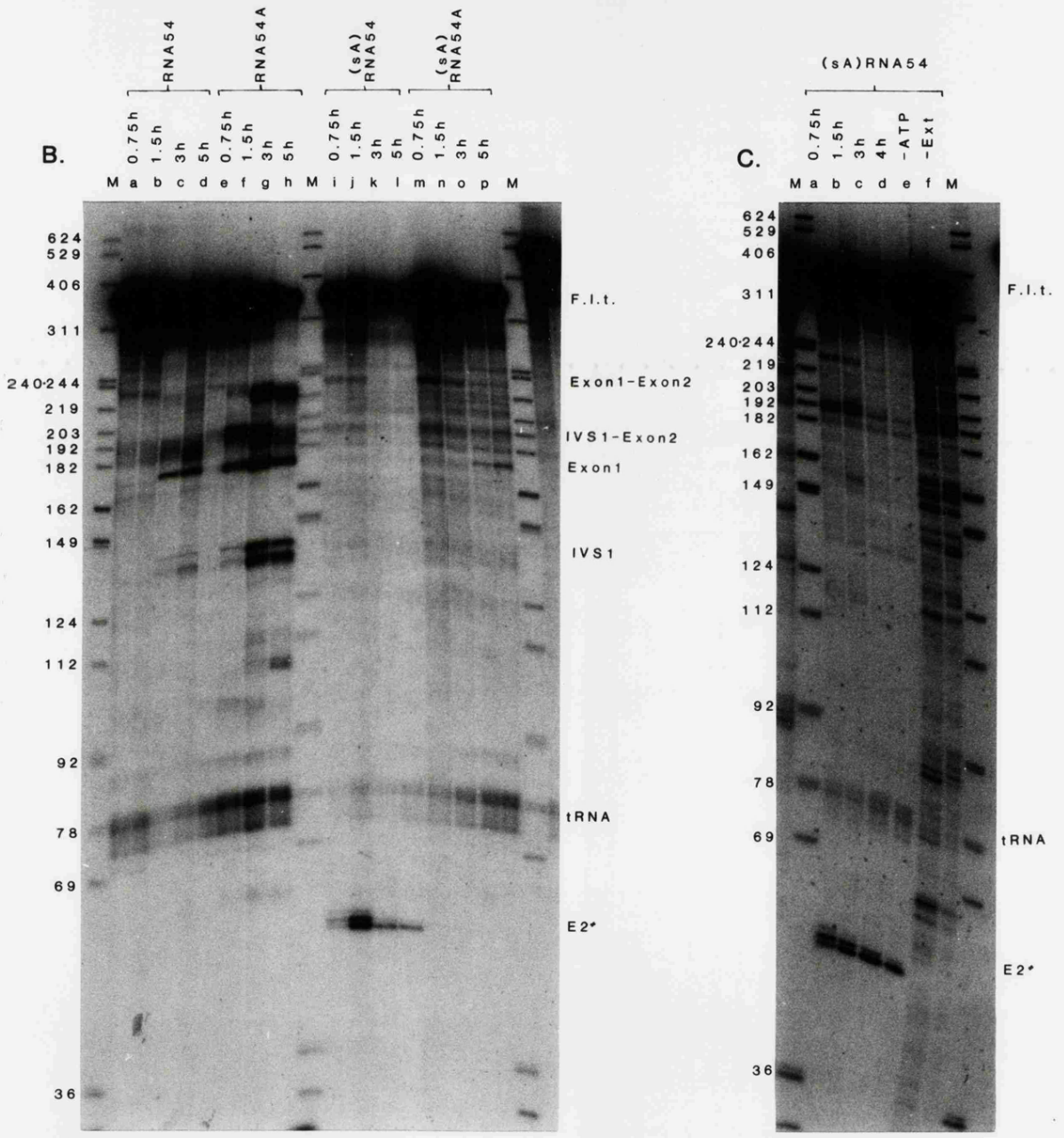
Figure 4.1. **Diagram of the structure of the cloned rabbit β-globin gene fragment used for preparation of transcripts (*panel A*) and the products of splicing with phosphorothioate-substituted transcripts (*panels B and C*).**

Panel A shows a *Pvu*II-*Taq*I fragment of a rabbit β-globin gene inserted into *Sma*I and *Acc*I sites of the polylinker (*striped*) in **mICE10**. The promoter for T7 RNA polymerase, the cap-addition site after transcription *in vivo* and the first and second exons (E1 and E2, respectively) of the rabbit β-globin gene are marked. IVS-1 of β-globin is shown as a *thin line*. The sequence of the RNA sense strand of the gene past the 3' splice site of IVS-1 shows the expected 3' terminus of transcripts (RNA54 and RNA50, respectively) produced after cleavage of plasmid DNA with *Hind*III (transcripts will end at the *arrow*, which marks the position of *Hind*III cleavage on the DNA strand complementary to that shown) or after priming with an oligonucleotide on single-stranded DNA (50-nucleotide exon 2). The 3' splice site and the 5' terminus of the RNA fragment (E2*) produced by a putative cleavage of [sA]RNA, deduced in this work, are shown.

Panel B shows the autoradiogram of a 7M urea, 20% formamide, 6% polyacrylamide gel used for analysis of *in vitro* splicing reactions of RNA54, RNA54A (as RNA54 but the nucleotide to the 3' side of the 3' splice site was mutated to A) and the corresponding RNAs transcribed with ATPαS instead of ATP, [sA]RNA54 and [sA]RNA54A. The times of reaction are shown in hours. *F.l.t.* is the full-length transcript, unprocessed; splicing reaction intermediates and products are marked. *E2** is the novel RNA fragment. tRNAs are labelled in the reaction, presumably by recycling of [^{32}P]pC from these transcripts prepared in the presence of [α-^{32}P]CTP. Size markers (M) used in this and subsequent figures are *Hpa*II fragments of pBR322, filled in with [α-^{32}P]dCTP and dGTP.

Panel C shows an autoradiogram from a similar reaction with [sA]RNA54, with extra lanes where the splicing reaction took place in the absence of ATP and creatine phosphate (*-ATP*) or extract (*-Ext*).


A.
Polylinker
Cap
(in vivo)
T7
E1
IVS1
E2
Polylinker
SmaI/
PvuII
TaqI/
AccI
T7 Promoter
HindIII
3'splice site
5' TTTTCATTTTCTCAGGCTGCTGGTTGTCTACCCATGGACCCAGAGGTTC
Cleavage of
(sA)RNA54
TTCGACCTGCAGCCCAAGCTTGGCACTGGCC 3'
50nt. Exon2
HindIII
cleavage site


B.
RNA54
RNA54A
(sA) RNA54
(sA) RNA54A
0.75h
1.5h
3h
5h
M a b c d e f g h M i j k l m n o p M
624
529
406
311
240·244
219
203
192
182
162
149
124
112
92
78
69
36
F.l.t.
Exon1-Exon2
IVS1-Exon2
Exon1
IVS1
tRNA
E2*
C.
(sA)RNA54
0.75h
1.5h
3h
4h
-ATP
-Ext
M a b c d e f M
624
529
406
311
240·244
219
203
192
182
162
149
124
112
92
78
69
36
F.l.t.
tRNA
E2*

splicing reactions *in vitro*, and the reaction products were analysed by polyacrylamide gel electrophoresis (PAGE). Normal splicing products were assigned on the basis of comparisons between substrates with different lengths of 3' exon (the mobilities of exon 1 intermediate and lariat product are unaltered) and on the shifts in mobility, relative to markers, of lariat-containing species on gels with different concentrations of acrylamide. Unsubstituted RNA was spliced efficiently, but splicing of thio-substituted RNA was severely reduced. Strikingly, Figure 4.1B shows that splicing of the wild-type RNA with phosphorothioate linkages 5' to adenosine ([sA]RNA 54) gave rise to novel products with lengths of about 54 to 55 nucleotides, and these were absent in the corresponding reaction with the mutant transcript ([sA]RNA 54A). Formation of these products requires ATP (Figure 4.1C). Further experiments showed that two bands (and sometimes a faint third band) were often produced in reactions with transcripts derived from *Hind* III-cut templates, whereas a single strong band and , occasionally, a fainter upper band were produced when the substrate was transcribed from primed viral DNA (as, for example, in Fig. 4.4). The multiple bands in the former case were gradually replaced by a single, stronger band with increasing incubation times. The major band was of the same length whether it was derived from [sA]RNA50 or [sA]RNA54. In some experiments, both wild-type and mutant [sA]RNA can be seen to undergo also one or both steps of splicing.

Transcripts prepared with phosphorothioate linkages 5' to guanosine, uridine and cytidine were also tested. It can be seen from Figure 4.2 that [sG]RNA54 was inert, [sU]RNA54 was a poor substrate for a normal splicing reaction, and [sC]RNA54 gave rise to small fragments very similar in length to those formed from [sA]RNA.

Transcripts with phosphorothioate linkages 5' to both cytidine and adenosine, i.e., [sC/sA]RNA54, gave rise to small fragments of similar, but not identical length to [sA]RNA54 and [sC]RNA54 (Fig. 4.3). The transcripts of the mutant template **mICE10[IVS-1] 3'A**, [sC]RNA54A and [sC/sA]RNA54A, did not give rise to low molecular weight fragments (Fig. 4.3). This is especially interesting in the case of [sC]RNA54A, where the G to A mutation in the first nucleotide downstream of the 3' splice junction does not cause any phosphorothioate linkages in the RNA to be gained or lost.

The novel products have been detected in every nuclear extract preparation that is competent in splicing. In order to determine whether the 54 nucleotide bands appeared with the same characteristic lag phase as the splicing reactions a comparatively inefficient extract was used. The new bands appeared with the same time dependence as did normal splicing intermediates (Figure 4.4); the lag phase has been correlated with the

Figure 4.2. **Splicing *in vitro* of transcripts with phosphorothioate linkages 5' to G ([sG]RNA54), U ([sU]RNA54), C ([sC]RNA54), or A ([sA]RNA54]).**

Transcripts were synthesised in the presence of GTPαS, UTPαS, CTPαS, ATPαS, or without analogues (RNA54) replacing the corresponding NTP. [α-^{32}P]UTP was present in the transcription reactions except for that with UTPαS, where [α-^{32}P]CTP was used. Splicing reactions were incubated for the times shown in hours. Analysis by gel electrophoresis was as described for Fig. 4.1. Molecular weight markers and the designation of molecular species was as in Fig. 4.1. IVS-1 (tailless) describes the lariat intron product from which the 3' linear tail has been removed by 3' exonuclease activity.

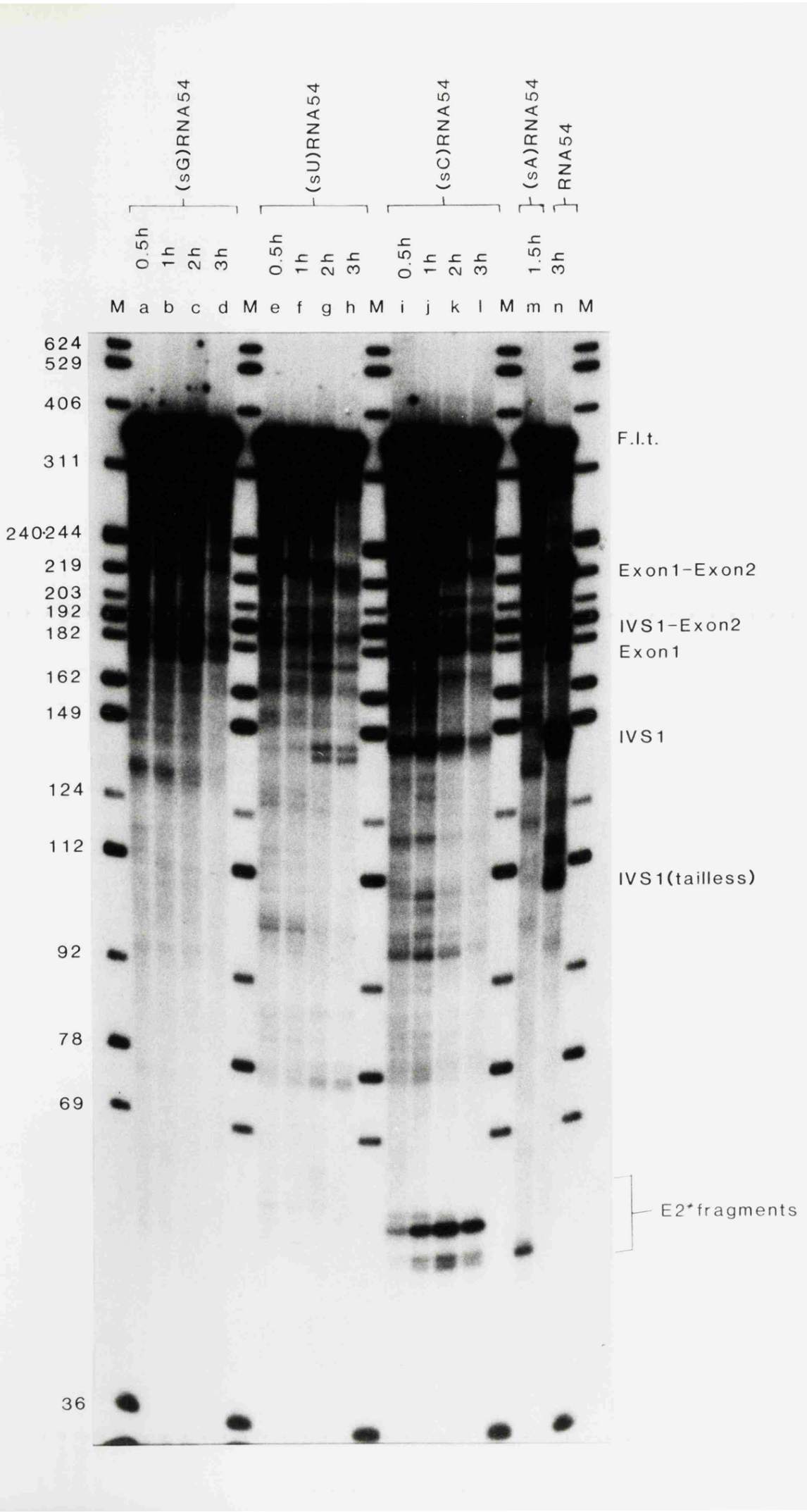

(sG)RNA54
(sU)RNA54
(sC)RNA54
(sA)RNA54
RNA54
0.5h
1h
2h
3h
0.5h
1h
2h
3h
0.5h
1h
2h
3h
1.5h
3h
M a b c d M e f g h M i j k l M m n M
624
529
406
311
240-244
219
203
192
182
162
149
124
112
92
78
69
36
F.l.t.
Exon1-Exon2
IVS1-Exon2
Exon1
IVS1
IVS1(tailless)
E2*fragments

Figure 4.3. **Splicing *in vitro* of wild type and mutant transcripts with phosphorothioate linkages 5' to C ([sC]RNA54 and [sC]RNA54A), both C and A ([sC/sA]RNA54 and [sC/sA]RNA54A), and A ([sA]RNA54).**

Transcripts were synthesised in the presence of CTPαS, ATPαS, CTPαS and ATPαS, or without analogues (RNA54) replacing the corresponding NTP. All transcription reactions contained [α-^{32}P]UTP. Splicing reactions were incubated for the times shown in hours. Analysis by gel electrophoresis was as described for Fig. 4.1. Molecular weight markers and indication of reaction products are as in Fig. 4.2.

(sC)RNA54
(sC)RNA54A
RNA54
(sA)RNA54
(sC/sA)RNA54
(sC/sA)RNA54A
0.5h
1h
2h
3h
M a b c d e f g h M i j k l M m n o p q r s t M u v w x M
F.l.t.
Exon1-Exon2
IVS1-Exon2
Exon1
IVS1
IVS1(tailless)
E2*

Figure 4.4. **Time dependence of splicing and formation of the novel RNA fragment (E2*).**

Transcripts synthesised in the presence of ATP (RNA50) or ATPαS ([sA]RNA50) were incubated in splicing reactions for the times shown in minutes. Analysis by gel electrophoresis was as described for Fig. 4.1. Molecular weight markers and indications of reaction products are as in Fig. 4.2.

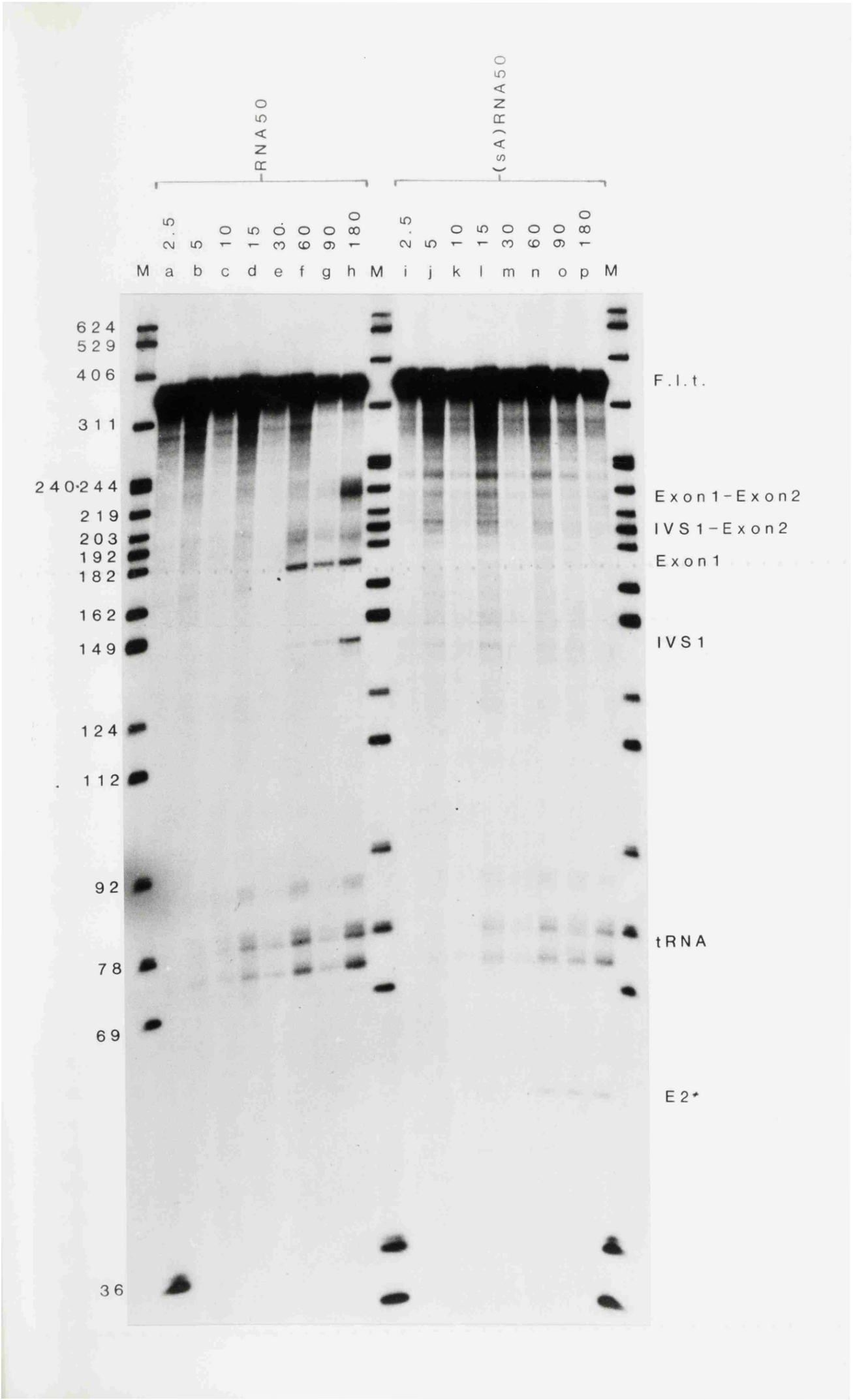

RNA50
(sA)RNA50
M a b c d e f g h M i j k l m n o p M
2.5 5 10 15 30. 60 90 180 2.5 5 10 15 30 60 90 180
624
529
406
311
240·244
219
203
192
182
162
149
124
112
92
78
69
36
F.l.t.
Exon1-Exon2
IVS1-Exon2
Exon1
IVS1
tRNA
E2*

Figure 4.5. **Primed synthesis to determine the 5' terminus of the RNA fragments, E2*, from [sA]RNA, [sC]RNA and [sC/sA]RNA54.**

Two oligonucleotides were used as primers, oligo.36 and oligo.28. The *number* indicates the distance from the 3' splice site to the nucleotide to which the 5' nucleotide of the primer is complementary. *Lanes d-g, i-l, o-r,* and *t-w* are sequencing reactions performed with the same primers on a DNA template with the structure shown in Fig. 5.1A, and *C, A, G,* and *T* refer to the dideoxynucleoside triphosphate (ddNTP) used in each reaction. *Lanes c, h, n* and *s* show primer extensions, in the absence of ddNTPs, with AMV reverse transcriptase on the gel-purified E2* fragments from [sA]RNA (*lanes c* and *h*), from [sC]RNA (*lane n*), or from [sC/sA]RNA (*lane s*). *Lane b* shows a similar reaction with oligo.36 in the presence of ddGTP, to demonstrate that any bands seen did not result from partial degradation of the RNA fragment (see Chapter 3, "Methods"). *Lane a* contains a sample of the E2* fragment used as a template for the reactions in *lanes b* and *c*. *Lane m* contains a sample of the E2* fragment used for the reaction in *lane h*. Markers (M) are as before. The autoradiogram shows the analysis of these reactions on a 7M urea, 15% polyacrylamide gel.

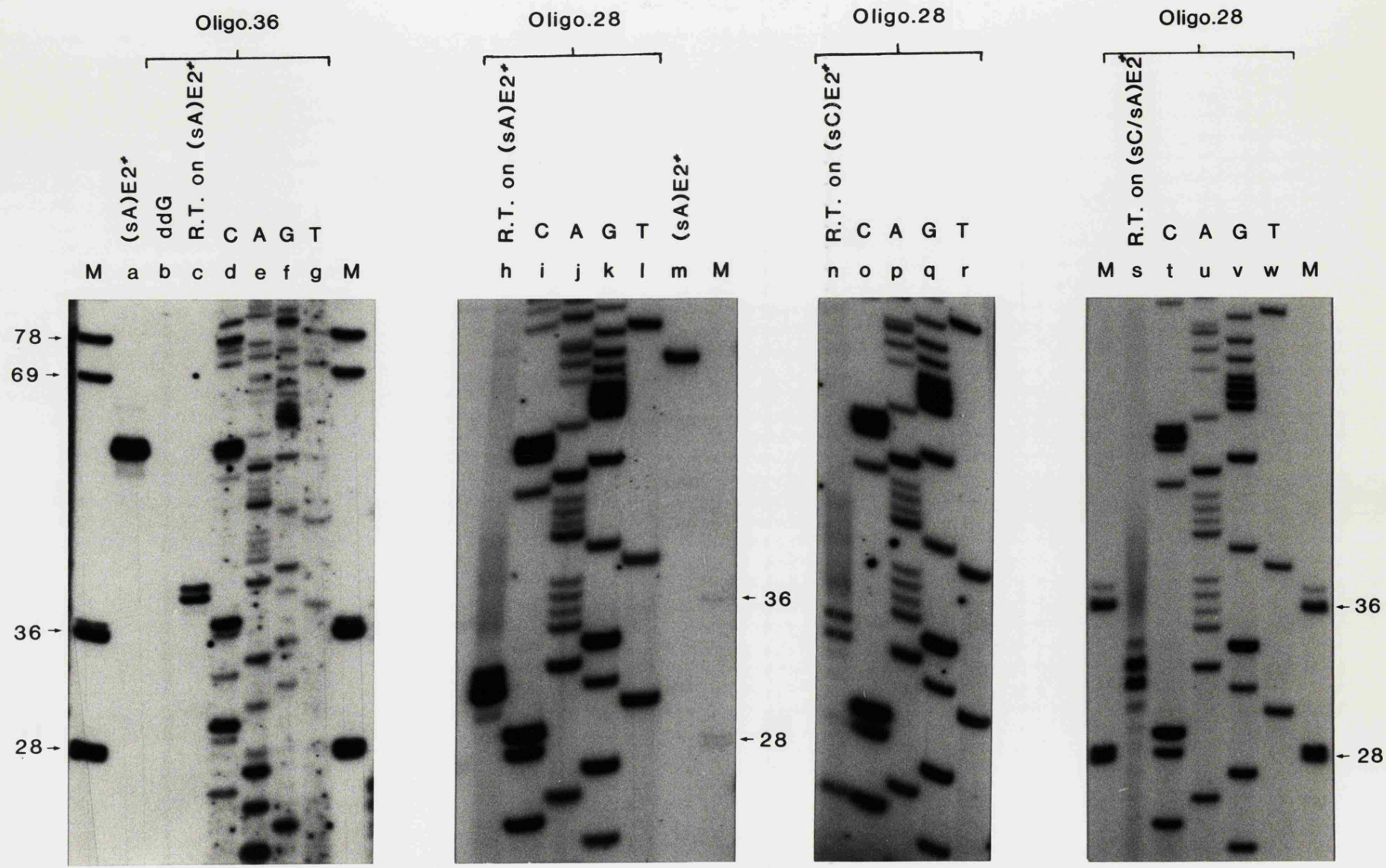
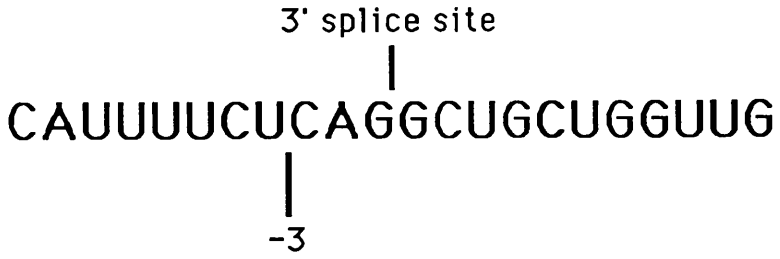

Oligo.36
Oligo.28
Oligo.28
Oligo.28
(sA)E2*
ddG
R.T. on (sA)E2*
C A G T
R.T. on (sA)E2*
C A G T
(sA)E2*
R.T. on (sC)E2*
C A G T
R.T. on (sC/sA)E2*
C A G T
M a b c d e f g M
h i j k l m M
n o p q r
M s t u v w M
78
69
36
28
36
28
36
28

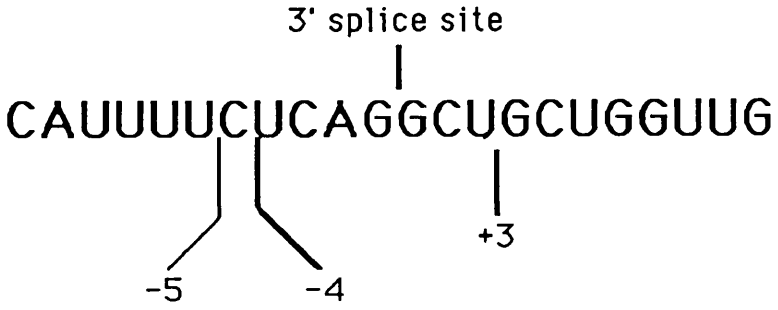
Figure 4.6 **Diagram illustrating the 5' termini of E2* fragments derived from [sA]RNA54 ([sA]E2*), [sC]RNA54 ([sC]E2*), and [sC/sA]RNA54 ([sC/sA]E2*), as determined by primer extension.**

Each panel shows the sequence of the rabbit β-globin pre-mRNA around the 3' splice site of IVS-1: the position of the 3' splice site is indicated. The locations of the putative 5' termini of each E2* RNA, as deduced from primer extension data, are indicated by lines extending beneath the sequence. Thick lines indicate the positions of termini which are probably genuine, whereas thin lines indicate termini which could have been falsely assigned due to template independent addition of an extra nucleotide by reverse transcriptase. The positions of the termini are labelled relative to the 3' splice site.

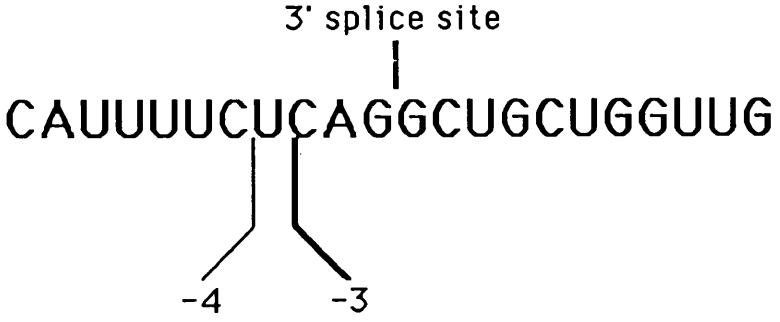
A. [sA]E2*



B. [sC]E2*



C. [sC/sA]E2*



assembly of the spliceosome, and this time-course suggests that a similar process may be required for generation of the new molecules.

4.1.1. Characterisation of the RNA fragment

For several seasons, the new fragments were thought to be derived from the 3' exon. The lengths of the fragments appeared to correspond with that of the 3' exon rather than of the other products which might conceivably arise from extensions of known reactions. Furthermore, the [sA]RNA fragment did not appear when a phosphorothioate linkage was placed at the 3' splice site (Figure 4.1B). One or two principal bands were seen with [sA]RNA: if the fragments represented the portion of a transcript protected against degradation by the binding of a site-specific factor (Ruskin *et al.*, 1985), such a clean result would not be expected. However, a site-specific cleavage at or near the 3' splice site would be expected to give rise to a doublet of major bands, the heterogeneity of which would result from the template-independent addition by T7 RNA polymerase of (predominantly) one nucleotide to about half of the transcripts which reached the expected 3' end during run-off transcription (Turnbull-Ross *et al.*, 1988). A separate explanation is required to account for the identical molecular weights of the major fragments derived from [sA]RNA 50 and [sA]RNA 54.

Two tests were made of this reasoning. Primed synthesis on the RNA fragments from primers complementary to the 3' exon should position the 5' end precisely, and splicing reactions of substrates with different lengths of 3' exon should give rise to fragments of appropriately different sizes if the fragments extend from the 3' end of the transcript. The results of the primed synthesis experiments on the gel-purified fragments are shown in Figure 4.5. Each extension was run on a gel adjacent to a sequencing reaction from the same primer on the same template as that used for transcription. Both primers on [sA]RNA fragments terminate in a doublet at the same position in the sequence. This doublet is not thought to indicate 5' heterogeneity because in the extension with oligo.28 the [sA]RNA fragment used as a substrate was derived from a single band; AMV reverse transcriptase is known to frequently add an extra nucleotide to the 3' end of the extended chain beyond the 5' end of the RNA template (Jacquier *et al.*, 1985; Inoue *et al.*, 1986). The common site of termination corresponds to the third nucleotide preceding the 3' splice site, indicating an RNA structure CAG/3' exon (Fig. 4.6), or a modification, such as a 2'-5' branched linkage, of the fourth nucleotide before the 3' splice site. Oligo.28 was also used as the primer for reverse transcription on gel purified E2* fragments derived from the processing of [sC]RNA54 and [sC/sA]RNA54 ([sC]E2* and [sC/sA]E2*

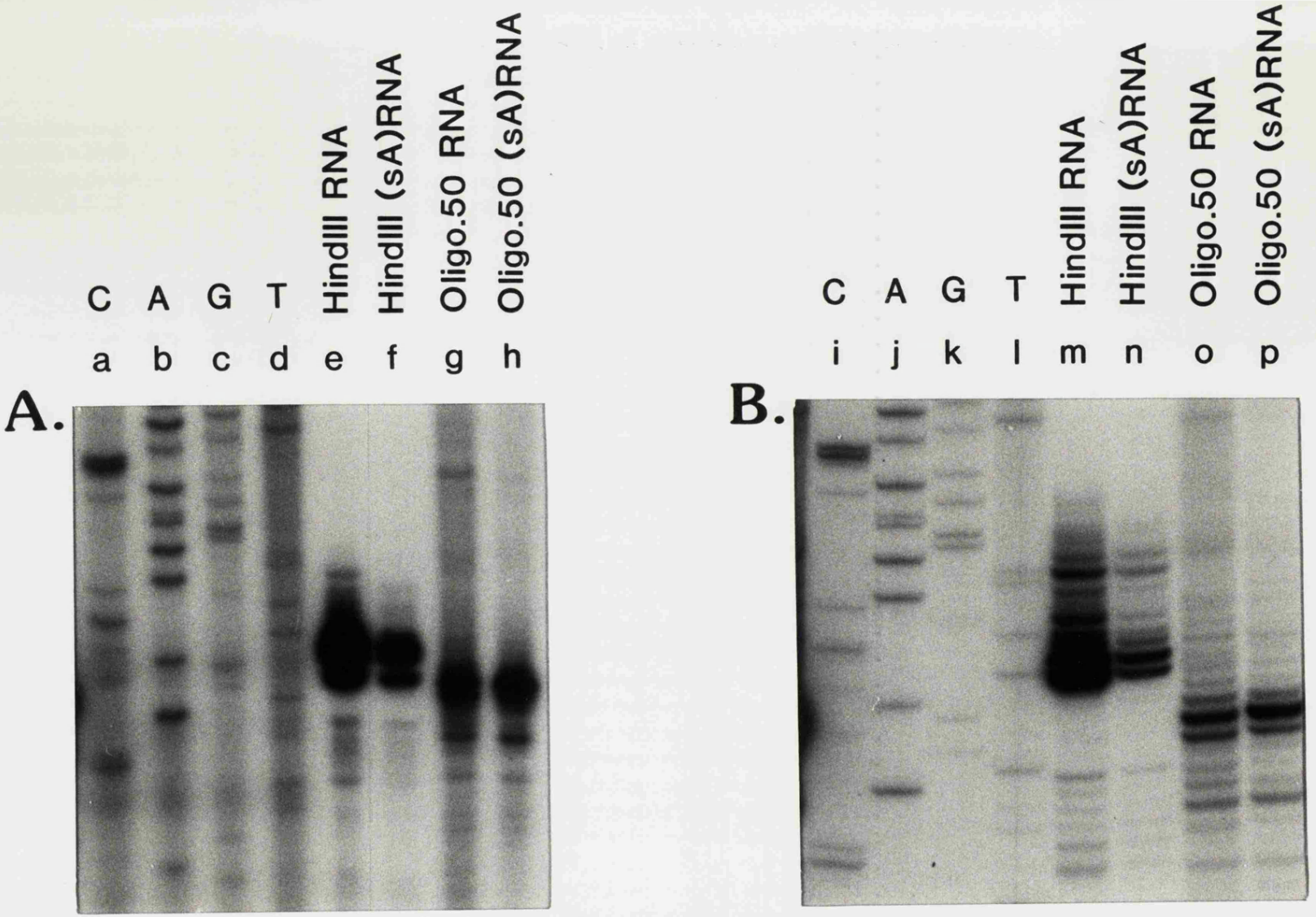
respectively). Three strong bands were seen from [sC]E2*, corresponding to reverse transcription stops 4 and 5 nucleotides upstream, and 4 nucleotides downstream of the 3' splice junction (Fig. 4.6). Hence, the 5'-terminus of E2* RNA must be heterogeneous, some molecules terminating 4 nucleotides upstream of the 3' splice junction, some 4 nucleotides downstream of the 3' splice junction, and possibly others terminating 5 nucleotides upstream of the 3' splice junction (Fig. 4.6), since the abundance of the cDNA corresponding to termination at the latter site suggests that it could arise from template independent addition of an extra nucleotide by reverse transcription, or it may correspond to a real 5'-terminus. Two strong bands were seen with [sC/sA]E2*, corresponding to reverse transcription stops 3 and 4 nucleotides upstream of the 3' splice junction. This indicates either that [sC/sA]E2* fragments have a homogeneous 5'-terminus, lying 3 nucleotides upstream of the 3' splice junction, or that some fragments have an additional 5'-terminus 4 nucleotides upstream of the 3' splice junction (Fig. 4.6).

In the same experiment, the gel-purified RNA fragments were loaded adjacent to the sequencing ladder. Here, the fragments derived from restriction enzyme cut template appear to run with a mobility of 56 and 57 nucleotides, although the CAG/3' exon sequence would predict fragments of 53 and 54 nucleotides (based on co-migration of the major [sA]RNA 54 bands with the fragments derived from [sA]RNA 50). However, on such gels with high percentages of acrylamide, increases are seen in the apparent molecular weight of RNA molecules relative to DNA markers.

Such an increase was demonstrated in a comparison of two short transcripts run on gels of different acrylamide concentrations. Figure 4.7 compares the mobility of a T7 RNA polymerase transcript, terminated by oligonucleotide limitation 51 nucleotides downstream of the transcription start site, with that of DNA markers (in the form of a sequencing ladder) on 6% and 15% denaturing polyacrylamide gels. The oligonucleotide used was oligo.50 and the template DNA was **mICE10** (Eperon, 1986). This transcript runs with an apparent mobility of 56 nucleotides with respect to the DNA markers on the 15% gel and with an apparent mobility of 54 nucleotides on the 6% gel. The failure to co-migrate with the DNA markers was partly a consequence of the markers being DNA, not RNA, and partly because the RNA possessed a 5'-terminal triphosphate whereas the DNA markers possessed a 5'-terminal hydroxyl group. The higher percentage gel can be clearly seen to cause the apparent molecular weight of the RNA to increase. This is in accordance with previous work (Grabowski *et al.*, 1984). RNA and [sA]RNA transcripts can also be seen to comigrate. Figure 4.7 also shows transcripts produced from **mICE10** R.F. cleaved with *Hind*III. Such

Figure 4.7 **Comparison of the mobilities of RNA and [sA]RNA relative to DNA on 6% and 15% denaturing polyacrylamide gels.**

DNA markers were produced by dideoxy sequencing reactions performed on viral **mICE11[IVS-2]** DNA using universal primer (*lanes a-d* and *i-l*). The sequencing markers were labelled with ^{32}P, incorporated by the inclusion of [α-^{32}P]dATP in the extension reaction. *C*, *A*, *G*, and *T* refer to the dideoxynucleoside triphosphate (ddNTP) included in each reaction. Transcriptions were performed using T7 RNA polymerase, either on single-stranded viral **mICE10** DNA to which an oligonucleotide primer with a 5' terminus 51 nucleotides downstream of the transcription start site (oligo.50) had been annealed and the complementary strand extended using DNA polymerase I (Klenow subfragment) (*lanes g, h, o*, and *p*), or on *Hind*III cleaved **mICE10** replicative form (*lanes e, f, m*, and *n*). In *lanes f, h, n*, and *p* the transcription reactions were performed with ATPαS replacing ATP, whereas in *lanes e, g, m*, and *o* no analogues were used. *Panel A* shows the autoradiogram of a 7M urea, 6% polyacrylamide gel, and *panel B* shows the autoradiogram of a 7M urea, 15% polyacrylamide gel.


A.
C A G T HindIII RNA HindIII (sA)RNA Oligo.50 RNA Oligo.50 (sA)RNA
a b c d e f g h
6% gel
B.
C A G T HindIII RNA HindIII (sA)RNA Oligo.50 RNA Oligo.50 (sA)RNA
i j k l m n o p
15% gel

transcripts should be of identical length to those produced by oligo.50 limitation if transcription terminates at the site of the (+) strand cleavage and 4 nucleotides longer if transcription terminated at the (-) strand cleavage site. The transcripts clearly terminated at the (-) strand cleavage site (Fig. 4.7) in agreement with previous studies (Turnbull-Ross *et al.*, 1988).

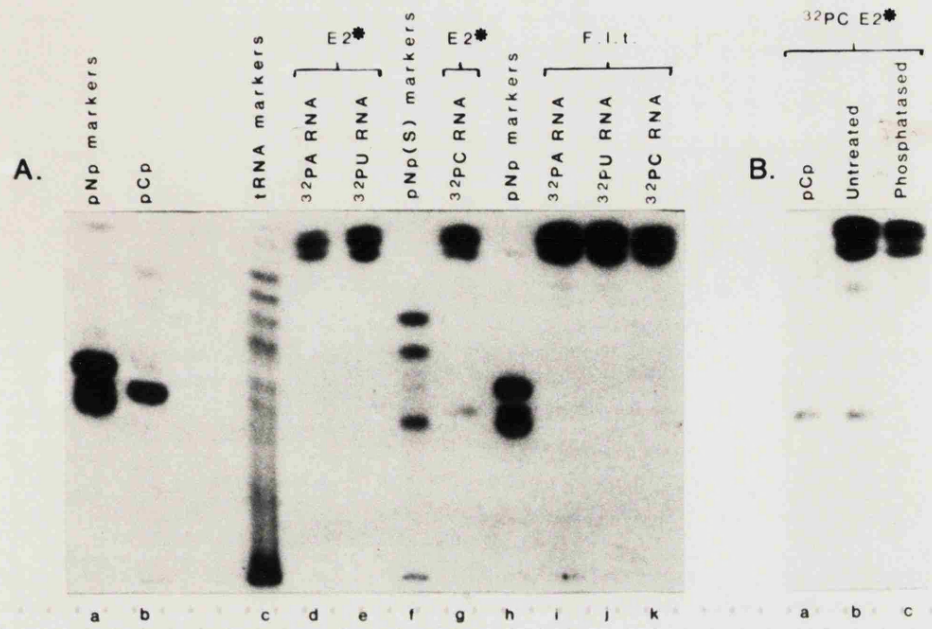
If the [sA]RNA fragments were produced by a specific processing event, rather than by an unusual activity of a degradative nuclease, a 5' phosphate might be expected on the 5'-terminal C. This could be verified by RNase T2 digestion of the [^{32}P]pC universally-labelled RNA fragment. Furthermore, identification of the 5' nucleotide as C would support the primer extension data and argue against very short capped 5' portions of the transcript being joined to the 3' exon in a splicing reaction.

RNA substrates were prepared in transcription reactions containing ATPαS with [α-^{32}P]UTP, [α-^{32}P]CTP or [α-^{32}P]ATP as labelled nucleoside triphosphates. After incubation in a splicing extract, the 54 nucleotide RNA fragments (E2*) were purified by gel electrophoresis. Digestions of E2* with RNase T2 were analysed by homochromatography. Figure 4.8A shows that a product other than [^{32}P]Np was detected in the reaction labelled with [α-^{32}P]CTP. This product disappeared when E2* was treated with calf intestinal phosphatase (Figure 4.8B), indicating that the product was derived from a phosphorylated 5' terminus. Attempts to recover the products by subsequent kinase treatment failed because the RNA was degraded and many otherwise internal nucleotides were phosphorylated. The extra product did not appear when E2* was labelled with [α-^{32}P]UTP or [α-^{32}P]ATP. This indicates that the 5' terminus of E2* is homogeneous, and that the phosphorylated 5' nucleotide is not pU or pA; it could be p[S]A, which would not be labelled by [α-^{32}P]ATP. However, the sequence prior to the 3' splice site is UCU/CAG/ (the first slash indicates the site of termination by reverse transcriptase (Figure 4.6) and the second slash shows the normal 3' splice site). In this region, [α-^{32}P]CTP would not label the RNase T2 digestion product containing A. Thus, that novel product was inferred to be pCp[S]. Furthermore, because it was not labelled by [α-^{32}P]ATP, it is apparent that the formation of E2* requires p[S]A at the second nucleotide preceding the 3' splice site, and that, where [^{32}P]pA was incorporated instead, E2* did not form.

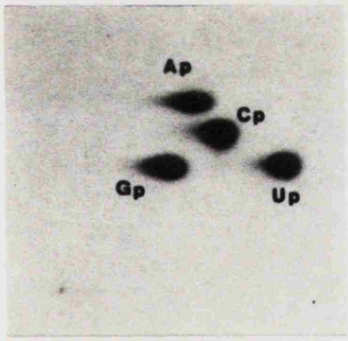
In order to confirm that pCp[S] was being formed, [^{32}P]pNp[S] markers were produced. RNA was transcribed in the presence of all four nucleoside thiotriphosphates, and digested with RNase T1. The products were end-labelled with [γ^{32}P]ATP, purified by PAGE, and digested with RNase T2. This procedure separated free [γ^{32}P]ATP from the desired products, i.e. [^{32}P]pNp[S] where N is C,A or U. [^{32}P]pNp markers were produced likewise.

Figure 4.8. **Direct analysis of the [sA]RNA fragment, E2*, to determine the 5' terminus.**

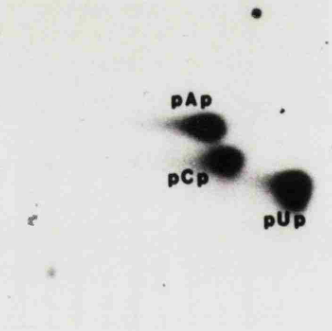
Panel A shows an autoradiogram of a polyethyleneimine plate on which samples were subjected to homochromatography from bottom to top. The preparation of [^{32}P]pNp and [^{32}P]pNp[S] markers is described in the text (Chapter 3, "Methods", and Chapter 5, "Results"). *Lanes a* and *h* contain [^{32}P]pNp markers; *lane b*, [^{32}P]pCp; *lane c*, [^{32}P]pCp-labeled tRNA hydrolysed in the presence of formamide; *lanes d, e*, and *g*, ribonuclease T2-digested E2* fragments prepared in splicing reactions from transcripts labeled with [^{32}P]pA, [^{32}P]pU, and [^{32}P]pC, respectively; *lane f*, [^{32}P]pNp[S] markers; *lanes i, j*, and *k*, ribonuclease T2-digested full-length transcripts labeled as for *lanes d, e*, and *g*. *Panel B* shows an autoradiogram of a polyethyleneimine plate (as *panel A*). Samples are:*lane a*, [^{32}P]pCp; *lane b*, as *panel A lane g*; *lane c*, as *lane b* except that the sample was incubated with calf intestinal phosphatase before digestion with ribonuclease T2. *Panel C* shows autoradiograms of samples fractionated by two-dimensional chromatography on cellulose thin-layer plates. Chromatography was from top to bottom in the first dimension, and left to right in the second dimension. *Plate a* shows the fractionation of all four nucleoside 3'-monophosphates, produced by ribonuclease T2 digestion of [^{32}P]pC-containing RNA; *plate b* shows the fractionation of ribonuclease T2 digested E2* fragment labeled during transcription with [^{32}P]pC; *plate c* contains [^{32}P]pNp markers; *plate d*, samples used for *plates a* and *c* were run together, producing four Np spots and three pNp spots; *plate e*, [^{32}P]pNp[S] markers; *plate f*, autoradiograms from *plates a* and *e* (run on the same day) were superimposed to show the relative positions of four Np spots and three pNp[S] spots. In *plates e* and *f*, pCp[S] and pAp[S] probably comigrate.



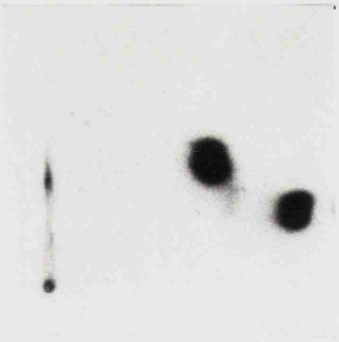
C.



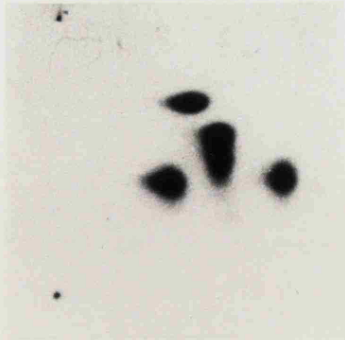
a: RNA54/T_2



c: pNp markers

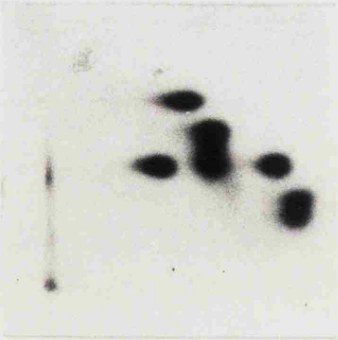


e: pNp(S) markers



b: E2*/T_2

d: a&c



f: a&e

Figure 4.8A shows that the novel product migrated with pCp rather than pCp[S].

For further analysis of the RNase T2 digestion products, two-dimensional chromatography on thin-layer plates was used. Figure 4.8C shows that pCp[S] or pAp[S], rather than pCp, run in the same position relative to Cp as does the spot derived from the 5' terminus of [α-^{32}P]CTP-labelled E2*. This result prompted a re-investigation of the use of homochromatography. Analysis of RNase T2 digestion products of RNA labelled with [α-^{35}S]ATP and [α-^{32}P]UTP during transcription led to the conclusion that homochromatography caused almost 90% of the sulphur content of nucleoside 3'-thiophosphates to be deposited on the line of application to the plate. RNA54 and [sA]RNA54 was transcribed in the presence of [α-^{35}S]ATP alone, or [α-^{35}S]ATP and [α-^{32}P]UTP. The RNA and [sA]RNA was then digested to completion with RNase T2 and the products of digestion fractionated by homochromatography (Fig. 4.9). The bands at the point of application (the origin), corresponding to nucleoside 3'-monophosphates (Np) and those thought to correspond to free inorganic phosphate (Pi). were scratted off the plate and scintillation was counted. 99% of the total ^{32}P counts were found in the Np bands, whereas only 12% of the total ^{35}S counts were found in the Np bands. This experiment did not show directly that nucleoside 3'-phosphorothioates are predominantly desulphurised on homochromatography to form nucleoside 3'-phosphates which migrate as normal, but this is highly likely. It is not understood why the pNp[S] markers in Figure 4.8A appear to be unaffected. However, it was concluded that the structure of E2* from [sA]RNA is almost certainly pCp[S]AG/3'exon.

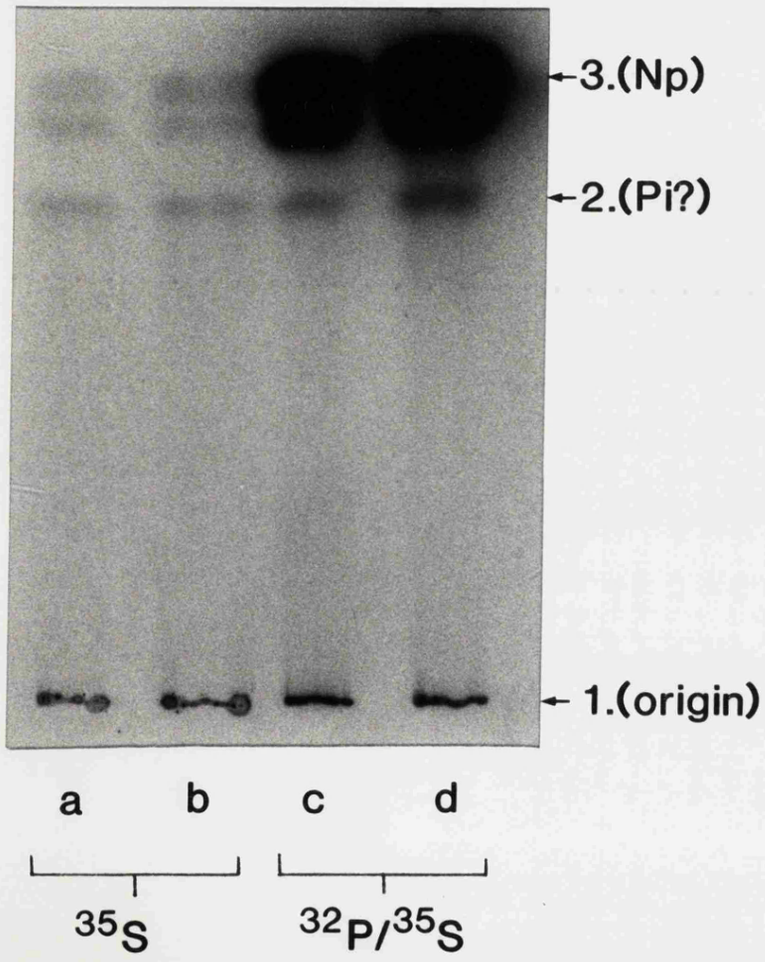
4.1.2. Phosphorothioate linkages inhibit 3' exonuclease

Substrates with different lengths of 3' exon sequence were produced in order to confirm that E2* included the 3' terminal sequences of the transcript. These substrates were produced by using various oligonucleotides as primers on viral DNA (Eperon, 1986). The original [sA]RNA substrates (3' exon lengths of 50 or 54 nucleotides) gave rise to E2* as before, whereas shorter substrates (3' exon lengths of 42, 46 or 48 nucleotides) gave rise to no E2* products; longer substrates gave rise to no E2* products, to fragments of the same size as did [sA]RNAs 50 and 54 or to longer fragments, varying with the extract (data not shown).

An explanation for these observations is that [sA]RNA 50 possessed a 3'-terminal structure that was resistant to 3' exonuclease activity. Of the

Figure 4.9. **Investigation of the behaviour of ^{35}-labelled RNase T2 digestion products when subjected to homochromatography.**

Transcripts were synthesised using T7 RNA polymerase on a template of *Hind*III cut **mICE10[IVS-1]**. The transcripts were digested to completion with RNase T2, and the products of digestion fractionated on a polyethyleneimine plate by homochromatography as in Fig. 4.8. and the location of the radiolabel identified by autoradiography. *Band 1. (origin)* is the line of application; *band 2. (Pi?)* possibly corresponds to free phosphate; and *band 3. (Np)* corresponds to nucleoside 3'-monophosphates. The digested transcripts run in *lanes a* and *b* were labelled with ^{35}S alone, by transcription in the presence of [α-^{35}S]ATP. The digested transcripts run in *lanes c* and *d* were labelled with both ^{32}P and ^{35}S by transcription in the presence of both [α-^{32}P]UTP and [α-^{35}S]ATP. The digested transcripts in *lanes b* and *d* were transcribed with ATPαS replacing ATP.


3.(Np)
2.(Pi?)
1.(origin)
a
b
c
d
35S
32P/35S

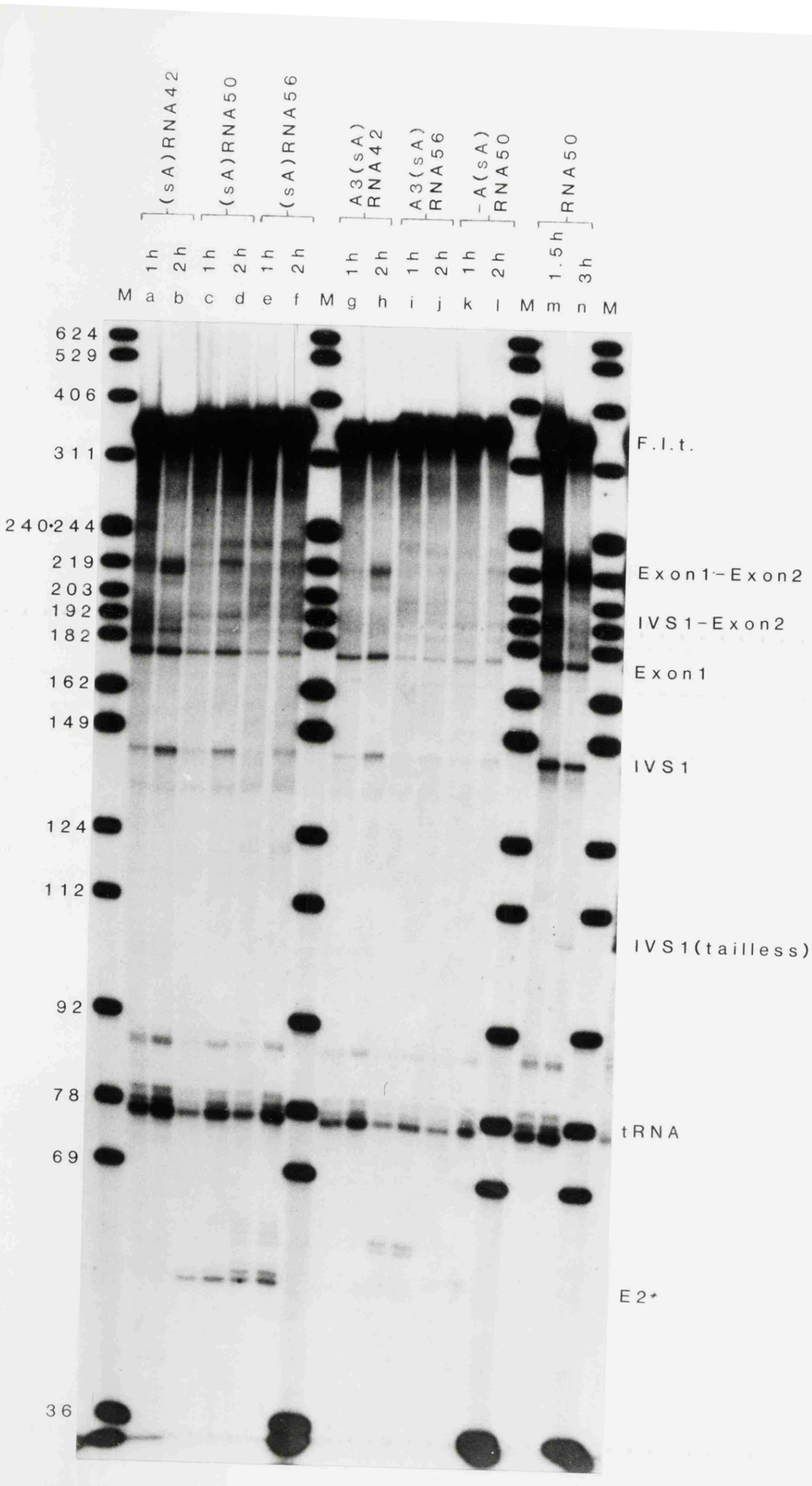
substrates tested above, only this substrate possessed a predicted 3'-terminal coded adenosine, and consequently a 3'-terminal phosphorothiate linkage. This explanation is consistent with the single E2* band produced from [sA]RNA 50, and the strong doublet produced from [sA]RNA 54: only a minority of the [sA]RNA 50 transcripts would be expected to have incorporated a resistant p[S]A nucleotide as the template-independent addition adduct, whereas [sA]RNA 54 will have incorporated two adjacent p[S]A linkages (Figure 4.1.A); if these linkages can be removed slowly, [sA]RNA 50 E2* will be fainter than [sA]RNA 54 E2*, which will appear as a doublet.

In order to test the proposal above, oligonucleotides with mismatched 5' termini were used as primers on viral DNA before transcription. One oligonucleotide directed a product with a 3' exon of 42 nucleotides ending in three As, another directed a transcript with a 3' exon of 56 nucleotides ending in three As, and another substituted C for A at the 3' end of the transcript with a 3' exon of 50 nucleotides. Figure 4.10 shows the result of splicing after transcribing these templates with ATPαS. Even without the substutution of AAA at the 3' end, the 56 nucleotide 3' exon transcript gave rise to appropriate products, although stronger bands are seen to co-migrate with those from the 50 nucleotide exon transcript, and the substitution of AAA inverts the intensities of these bands such that the longer products are now dominant. It was concluded that the presence of 3'-terminal phosphorothioate linkages does substantially stabilise this product, and that the range of suitable substrates is less narrow than it at first appeared. However, it should be noted that a substitution of AAA at the 3' end of the 42 nucleotide 3' exon transcript resulted in the loss of the very faint E2* band. The low yield of E2* with this exon length might be correlated with the relatively high level of normal splicing seen with this substrate in Figure 4.10. Substitution of C for A at the end of the 50 nucleotide 3' exon does result in a weakening of the signal, but there is also a one nucleotide increase in length; in part this can be attributed to the removal by an exonuclease of full length products ending in pC, whereas template-independent addition of one nucleotide to the 3' terminus of a transcript (Turnbull-Ross *et al.*, 1988) will give rise to a low proportion of molecules ending in p[S]A, which will be protected.

Thus, although there is no direct verification of the sequences present between those hybridising with the +36 primer and position 42 of the exon, it would seem that beyond reasonable doubt the E2* fragment produced from [sA]RNA extends from three nucleotides prior to the 3' splice site to the end of the transcript. A corresponding 5' fragment has not been observed in any of these experiments, and it was not detectable by S1 nuclease mapping

Figure 4.10. **The dependence of E2* formation on protection of the 3' terminus of the 3' exon by phosphorothioate linkages against 3' exonuclease activity.**

Transcripts synthesised in the presence of ATPαS and [α-^{32}P]CTP were incubated in splicing reactions for 1 and 2 hours, as shown, and subjected to electrophoresis as in Fig. 4.1. Markers and descriptions of reaction products are as in Fig. 4.2. The distance of the 3' nucleotide of the transcript from the 3' splice site was specified by an oligonucleotide (see Chapter 2, "Methods") and is described by the number of the RNA (i.e. 42, 50, or 56). In A3[sA]RNA42 and A3[sA]RNA56 (*lanes g-j*), the three nucleotides at the 3' end of the transcipt were changed to adenosine (see Chapter 2, "Methods"); in -A[sA]RNA50, the 3'-most adenosine in the transcript was changed to cytidine. All transcriptions were performed in the presence of ATPαS in place of ATP, except for the control reactions (*lanes m* and *n*).


(sA)RNA42
(sA)RNA50
(sA)RNA56
A3(sA) RNA42
A3(sA) RNA56
-A(sA) RNA50
RNA50
1h
2h
1.5h
3h
M a b c d e f M g h i j k l M m n M
624
529
406
311
240·244
219
203
192
182
162
149
124
112
92
78
69
36
F.l.t.
Exon1-Exon2
IVS1-Exon2
Exon1
IVS1
IVS1(tailless)
tRNA
E2*

(Data not shown).

Studies *in vitro* on the cleavage and polyadenylation reactions involved in 3'-terminal processing of pre-mRNAs revealed that, in the absence of free Mg^{2+}, the 3'-fragment produced by endonucleolytic cleavage was stabilised against degradation (Moore and Sharp, 1985). Hence, incubation of RNA54 and [sA]RNA54 was performed in a nuclear extract under splicing conditions, except for the absence of magnesium, to explore the possibility that 3'-proximal cleavage might still occur and the 5' fragment produced be sufficiently stable to be visualised (Fig. 4.11). No splicing of RNA54 was observed in the absence of magnesium, in agreement with previous work (Hardy *et al.*, 1984; Hernandez and Keller, 1983), and there was also no production of E2* from [sA]RNA54.

4.1.3. Reaction requires U1, U2 and U6 snRNAs

Selective degradation of snRNAs U1, U2, U4 and U6 in the nuclear extract by oligonucleotide-directed RNase H cleavage has been used to show that splicing requires these snRNAs (Krämer *et al.*, 1984; Krainer and Maniatis, 1985; Black *et al.*, 1985; Black and Steitz, 1986; Berget and Robberson, 1986; Chabot and Steitz, 1987). This technique was used to establish whether the same snRNAs are required for production of the novel 3' exon-derived fragment (E2*). In the first experiment, cleavage of U1 snRNA was incomplete (Figure 4.12A), and splicing of a normal transcript was reduced but not eliminated (Figure 4.13A). The same treated extract gave a five-fold reduction in the level of E2* compared with untreated extract, and an even greater reduction when compared with the activity of an extract treated without an oligonucleotide (Figure 4.13A). In a second experiment, U1 snRNA cleavage was still incomplete but the production of E2* was eliminated (Figures 4.12B and 4.13B). It has been observed previously that the extent to which U1 snRNA is depleted in extracts does not correlate well with the inhibition of splicing activity (Krämer *et al.*, 1984; Krainer and Maniatis, 1985; Black *et al.*, 1985; Chabot and Steitz, 1987).

Treatment with oligonucleotides directed against U2 and U4 + U6 snRNAs resulted in almost complete removal of the U2 and U6 snRNAs from the nuclear extracts (Figures 4.12A and 4.12B), and these extracts were inactive in both splicing and in E2* production (Figures 4.13A and 4.13B). Residual anti-U4 or U6 oligonucleotides cleaved some of the substrate RNA, but this is only a small proportion of the substrate and does not jeopardise the interpretation of the results. Oligonucleotides directed against U4 RNA failed to cleave the snRNA, despite the inclusion of ATP in the reaction.

Figure 4.11. **The dependence of E2* formation on the presence of Mg^{2+}.**

Transcripts synthesised in the presence of ATP (RNA54) or ATPαS ([sA]RNA54) were incubated in splicing reactions for the times shown in hours. The splicing reactions in *lanes a-d* contained $MgCl_2$ at the normal, optimal concentration for efficient *in vitro* splicing (3.2mM). The splicing reactions in *lanes e-l* were of normal composition except for the absence of $MgCl_2$ and the presence of 0.52mM (K^+) EDTA instead of 0.12mM (K^+) EDTA. Analysis by gel electrophoresis was as described for Fig. 4.1. Molecular weight markers and indications of reaction products are as in Fig. 4.2. The region of the gel where E2* fragments would migrate is marked (*E2* mobility*).

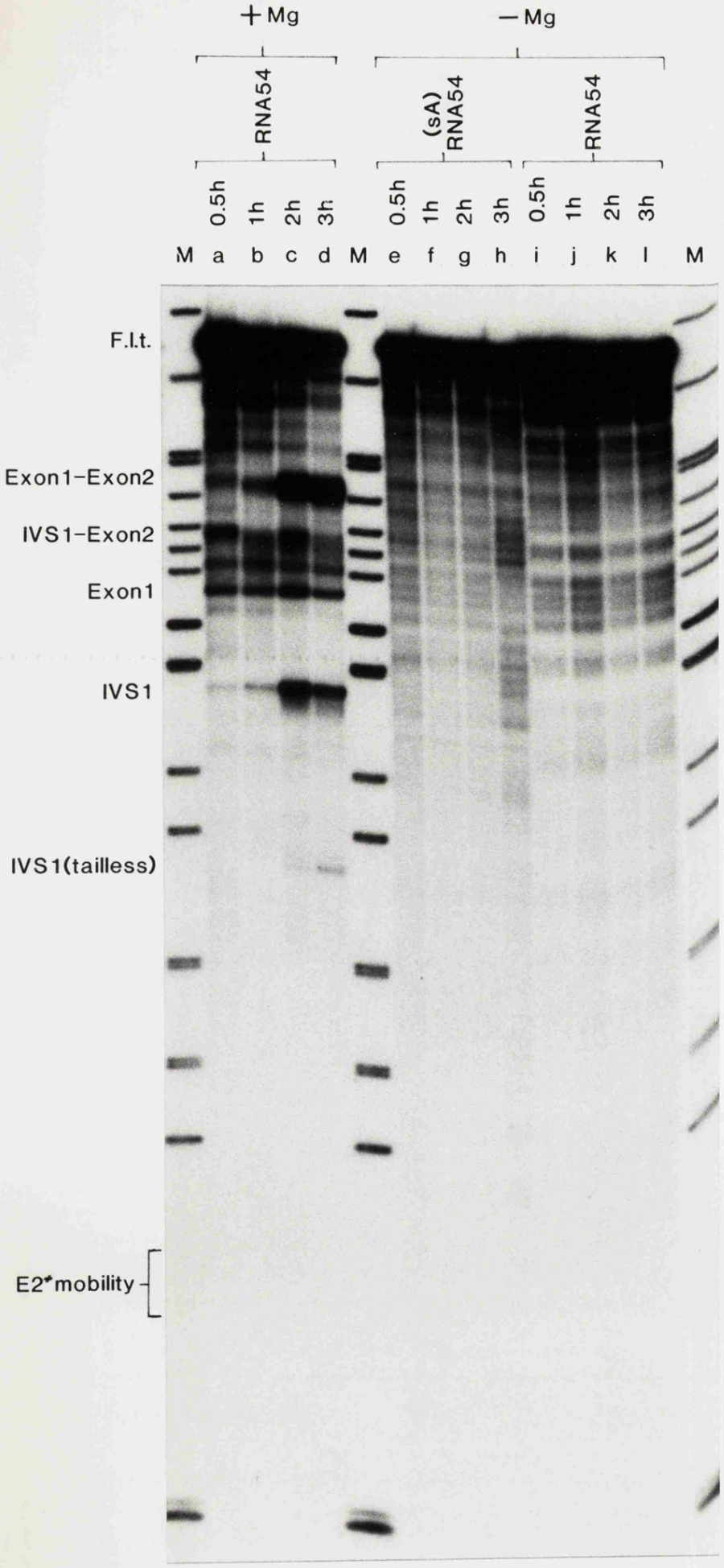

+Mg
−Mg
RNA54
(sA) RNA54
RNA54
0.5h
1h
2h
3h
0.5h
1h
2h
3h
0.5h
1h
2h
3h
M a b c d M e f g h i j k l M
F.l.t.
Exon1−Exon2
IVS1−Exon2
Exon1
IVS1
IVS1(tailless)
E2*mobility

Fig 4.12. **Site-directed RNase H cleavage of snRNAs.**

Aliquots of extract cleaved with the oligonucleotides U1m, U2a, or U4a + U6b in the presence of RNase H were fractionated on 7M urea, 8% polyacrylamide gels. The gel was stained with ethidium bromide. In parallel, an extract was incubated without oligonucleotide, and another sample was run without incubation. Tracks are labeled according to the treatment, and the appropriate RNAs are marked. *Panels A* and *B* show seperate experiments. The unfractionated remainder of each extract was used for the experiments shown in Fig. 5.8.

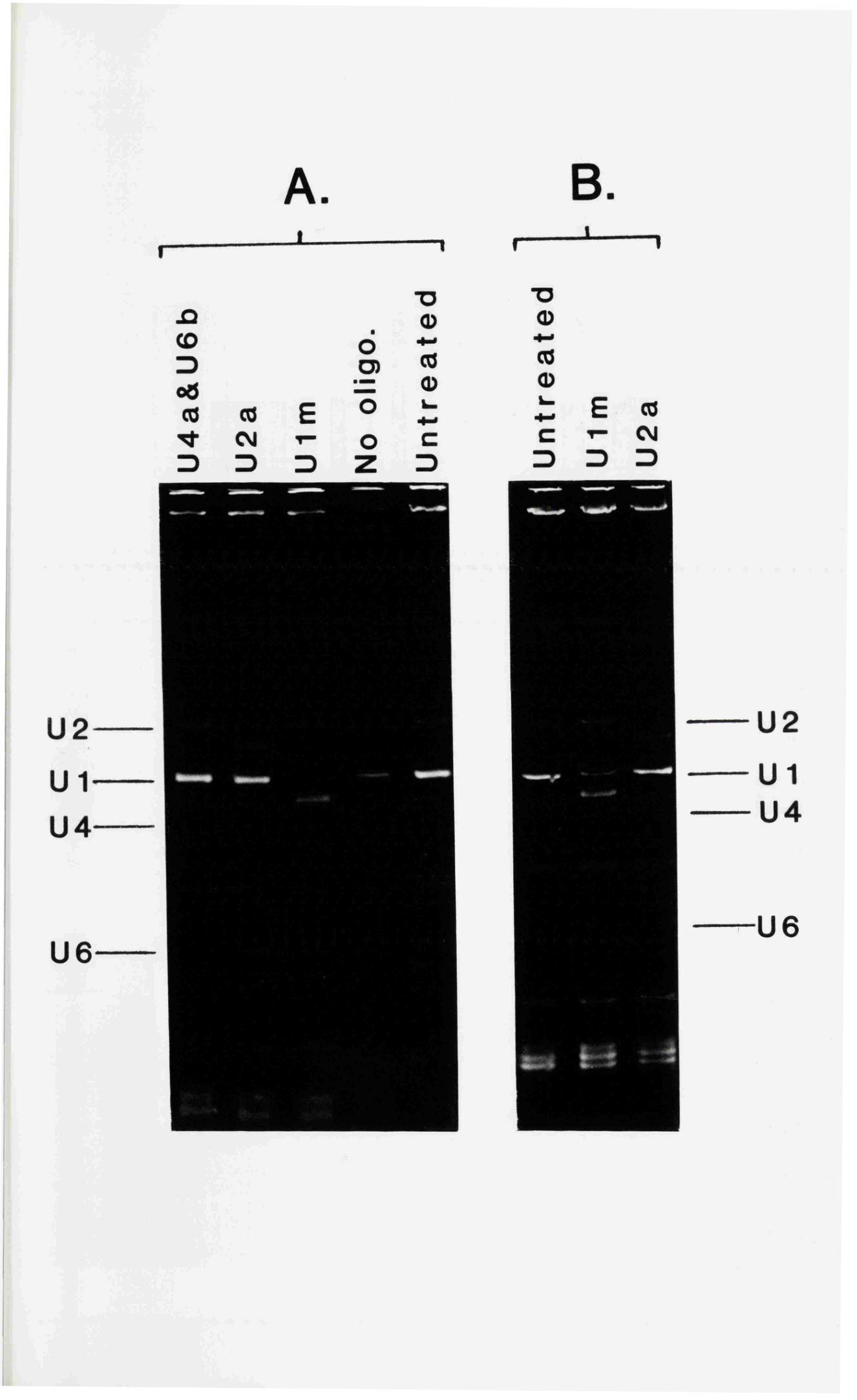
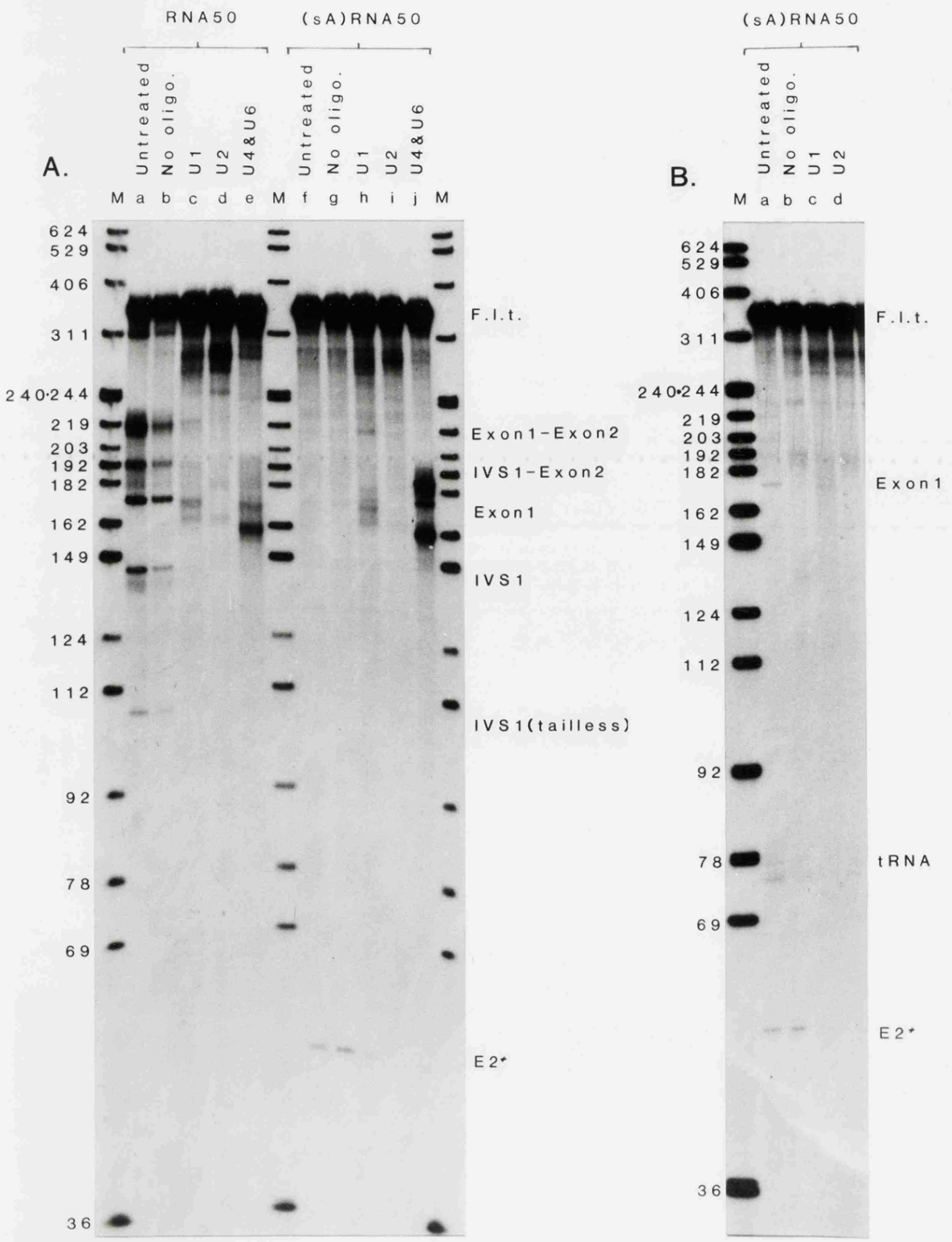

A.
B.
U4a&U6b
U2a
U1m
No oligo.
Untreated
Untreated
U1m
U2a
U2
U1
U4
U6
U2
U1
U4
U6

Figure 4.13. **Dependence of E2* production on the presence of snRNAs U1, U2, and U4/U6 in the nuclear extract.**

Nuclear extracts, treated (or otherwise) with RNase H and oligonucleotides complementary to specific snRNAs, were used for splicing reactions with substrates RNA50 and [sA]RNA50. Analysis was by gel electrophoresis as in Fig. 5.1. The experiments shown in *panels A* and *B* were seperate experiments, and each lane corresponds to use of the extract of which an aliquot had been analysed in Fig. 5.7 under the same lane headings.


A.
RNA50
(sA)RNA50
Untreated
No oligo.
U1
U2
U4&U6
Untreated
No oligo.
U1
U2
U4&U6
M a b c d e M f g h i j M
624
529
406
311
240·244
219
203
192
182
162
149
124
112
92
78
69
36
F.l.t.
Exon1-Exon2
IVS1-Exon2
Exon1
IVS1
IVS1(tailless)
E2*
B.
(sA)RNA50
Untreated
No oligo.
U1
U2
M a b c d
624
529
406
311
240·244
219
203
192
182
162
149
124
112
92
78
69
36
F.l.t.
Exon1
tRNA
E2*

4.2. DISCUSSION

The work described in this chapter shows that phosphorothioate-substituted RNAs are poor substrates for splicing *in vitro*. However, although [sG]RNA was inert and [sU]RNA spliced poorly, [sC]RNA, [sA]RNA and [sC/sA]RNA gave rise to fragments which extend to the 3' terminus of the transcript from near the 3' splice site. The [sA]RNA products were characterised in detail and shown to begin three nucleotides prior to the 3' splice site. The reaction was found to require ATP and snRNAs U1, U2 and U6, which supports a role for the splicing apparatus in this process.

The relationship of this novel reaction to the normal processes of splicing is not clear. The disappearance of the upstream portion of the substrate after E2* formation might suggest an a apparently attractive hypothesis wherein E2* is formed by a 5'-exoribonuclease which is arrested by complexes bound at the 3' splice site and stops prior to a phosphorothioate linkage. The existence of a 5'-exoribonuclease in HeLa cell extracts was inferred from the production during splicing of 3'-terminal fragments from uncapped RNA (Ruskin and Green, 1985; Noble *et al.*, 1986), and this activity has been characterised recently (Stevens and Maupin, 1987). However, there are several reasons why this may not be the correct explanation. The activity is known to be blocked by transcription of the RNA *in vitro* under conditions in which a cap analogue is incorporated (Ruskin and Green, 1985), as in this work. The 5' termini of fragments produced by the endogenous exonuclease during the splicing reaction are located 10-20 nucleotides upstream of the 5' splice site and 7-16 nucleotides upstream of the branch site (Ruskin *et al.*, 1984; Ruskin and Green, 1985; Noble *et al.*, 1986), and no fragments have been seen with termini near the 3' splice site. The presence of a phosphorothioate linkage to the 3' side of the 5' terminus does not imply that the exonuclease hypothesis is sufficient: for [sA]RNA it would have been necessary to excise 65 phosphorothioate linkages prior to the E2* start, and in [sC]RNA there are other phosphorothioate linkages near to the E2* 5' end which are not associated with E2* accumulation (see Fig. 4.6).

Another possible explanation is that a sub-reaction of step 2 of splicing has been isolated, i.e. that 3' cleavage normally takes place three nucleotides upstream of the 3' splice site and that this is followed by exonucleolytic removal of the three remaining nucleotides before the exons are ligated. This possibility would, however, seem to be ruled out by the results of early

RNase T1 digestions of the lariat product, which appeared to show that the large 3' fragment from the intron was different from the equivalent fragment from the precursor only in that it lacked a 3' phosphate (Padgett *et al.*, 1984; Ruskin *et al.*, 1984), i.e. that it extended to the 3' splice site.

A more probable explanation of the data from this study is that E2* has arisen by site-specific hydrolysis, possibly followed by local thiophosphate-arrested exonuclease activity. The hydrolysis reaction might be a result of the activation of the 3' splice site such as would be expected to precede step 2 of splicing in normal circumstances. The presence of phosphorothioate linkages in the transcript might cause a slight shift in the site of strain, causing the incorrect phosphodiester link to be activated. The normal 3' cleavage site when activated must be well shielded from attack by H_2O or OH^- as free 3' exon production is never detected in normal *in vitro* splicing reactions. However, if the presence of phosphorothioates in the substrate RNA alters interaction with the 3' splice site region so as to cause straining at a misplaced site, the activated phosphodiester linkage may now be accessible to attack by H_2O or OH^-, hence giving rise to E2* production. An obvious candidate for the phosphorothioate responsible for such a shift in the site of cleavage is that preceding the highly conserved penultimate A of the intron. This is supported by the experiment where transcription took place in the presence of [α-^{32}P]ATP and ATPαS. Although E2* was formed (with reduced efficiency), RNase T2 digestion gave rise to no spot corresponding to pCp on fractionation. This is consistent with incorporation of p[S]A at position -2 relative to the 3' slice site being necessary for E2* production.

Altered reactions with phosphorothioate-substituted molecules have been reported with simpler systems. It has been shown that the presence of one configuration of a phosphorothioate linkage near, but not at, the site of cleavage of an oligonucleotide by *Eco* RI will inhibit the reaction (Koziolkiewicz *et al.*, 1986). Furthermore, the presence of phosphorothioate linkages at the cleavage site of Ap_4A by Ap_4Aase of *Artemia* causes a shift in the site of cleavage (Blackburn *et al.*, 1987). The splice site-proximal cleavage postulated to generate E2* combines both elements, i.e., a phosphorothioate linkage may inhibit cleavage at a nearby site and cause the site of cleavage to shift.

Chapter 5.

INVESTIGATION OF THE BIOCHEMISTRY OF E2* FORMATION

5. INVESTIGATION OF THE BIOCHEMISTRY OF E2* FORMATION

This chapter describes an investigation of the biochemistry of E2* formation which was undertaken with the aim of ascertaining the nature of the relationship between E2* formation and splicing.

5.1. RESULTS

5.1.1. COMPLEX FORMATION ON PHOSPHOROTHIOATE RNA SUBSTRATES

Several experimental techniques have been employed to investigate the question of what sort of complex (if any) forms on phosphorothioate RNA substrates under *in vitro* splicing conditions.

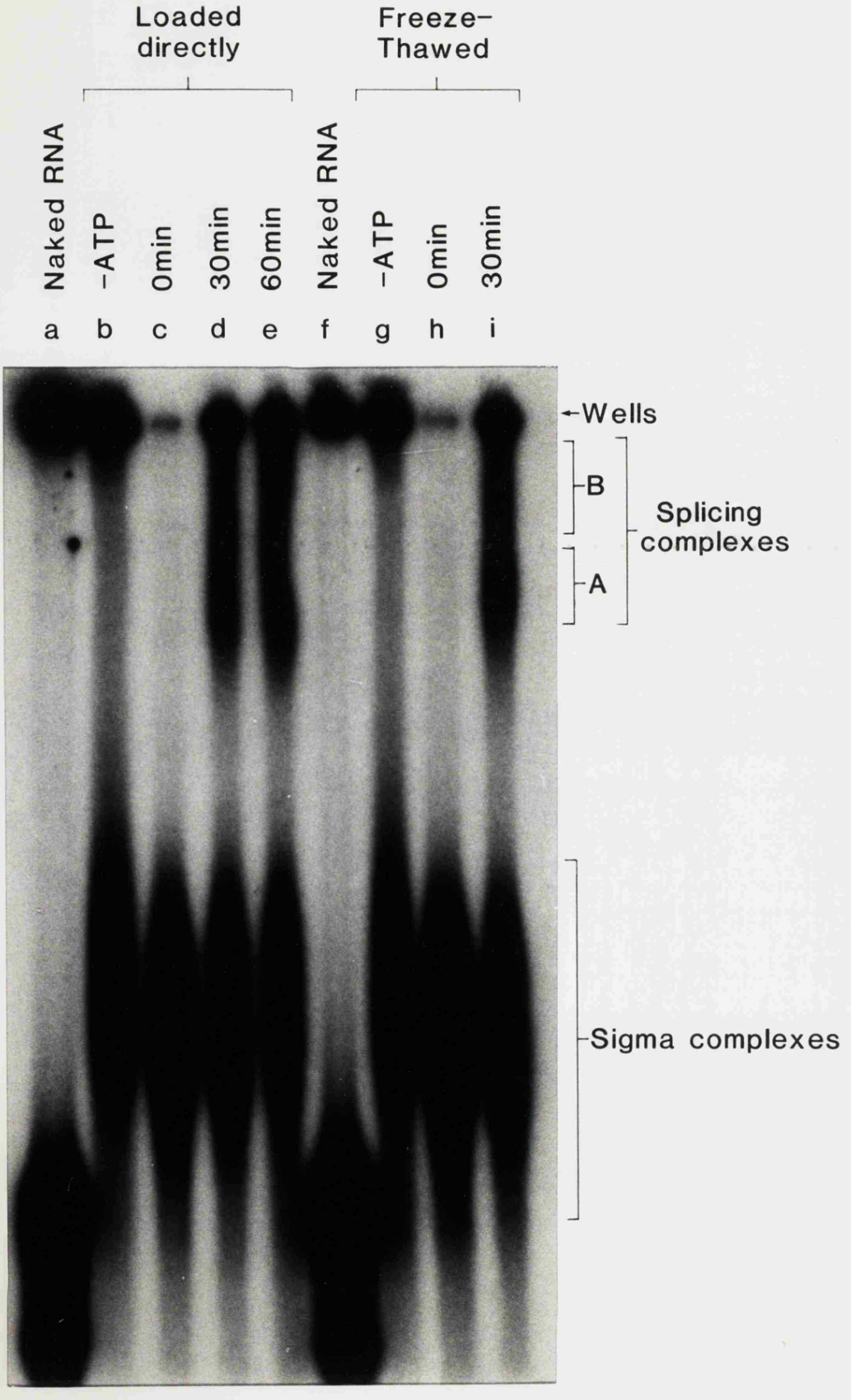
5.1.1.1. Electrophoresis on non-denaturing gels

Splicing complex formation was analysed by electrophoresis on native low-percentage polyacrylamide gels.

Initial experiments were concerned with optimising electrophoresis conditions to visualise splicing complexes formed on non-phosphorothioate RNA. RNA54 was incubated in a HeLa cell nuclear extract under splicing conditions, and treated with the polyanion heparin, prior to electrophoresis in a Tris-borate buffered gel as described by Konarska and Sharp (1986). Two distinct, ATP dependent complexes, A and B, of low mobility, were observed to form (Fig. 5.1), which presumably corresponded to the two complexes of the same name identified by Konarska and Sharp (1986) to form on an adenovirus RNA substrate. These two complexes were proposed to represent the mature spliceosome (complex B) and a pre-splicing complex (complex A). A number of faster migrating complexes also formed in a non-ATP-dependent manner, and these "sigma complexes" formed a smear in the lower part of the gel. Similar complexes were also observed by Konarska and Sharp (1986). Hence, this technique would seem to enable visualisation of splicing complexes on RNA54. No difference in resolution of complexes was observed when splicing reactions were loaded directly onto the gel after heparin treatment, or after splicing reactions were stored frozen on dry ice after heparin treatment (Fig. 5.1).

Figure 5.1. **Splicing complex formation on RNA54 assayed by electrophoresis on a native low-percentage polyacrylamide gel buffered with Tris-borate.**

Lanes a and *f* contain naked RNA54 which was not incubated with nuclear extract. *Lanes d*, *e*, and *i* contain RNA54 which had been incubated for the times shown with HeLa nuclear extract under optimal *in vitro* splicing conditions (without polyvinyl alcohol) for the times shown in minutes. *Lanes c* and *h* contain RNA54 which had been added to a complete *in vitro* splicing mixture on ice, and the the reaction stopped immediately by the addition of heparin. *Lanes b* and *g* contain RNA54 which had been incubated under splicing conditions for 60 minutes but in the absence of ATP and creatine phosphate. All the reactions were stopped by the addition of heparin, and the heparin treated samples were subjected to electrophoresis at room temperature on a 4% polyacrylamide gel buffered with 0.5xTBE. The positions of the complexes believed to correspond to splicing complexes A and B identified by Konarska and Sharp (1986) are indicated, as is the position of the higher mobility, ATP independent, Sigma complexes.


Loaded directly
Freeze-Thawed
a Naked RNA
b -ATP
c 0min
d 30min
e 60min
f Naked RNA
g -ATP
h 0min
i 30min
Wells
B
A
Splicing complexes
Sigma complexes

Splicing complexes formed on RNA54 were also resolved by electrophoresis on low percentage polyacrylamide gels using a 50mM Tris-glycine buffer for electrophoresis as described by Konarska and Sharp (1987). Using this buffer, addition of a polyanion such as heparin has been reported to be unnecessary. However, when heparin treated and untreated complexes were electrophoresed using Tris-glycine buffer, the best resolution of splicing complexes A and B was found after heparin treatment (Fig. 5.2). Indeed, complex B in the heparin treated tracks may be resolving into two distinct bands. Lamond *et al.* (1987) have previously resolved three, rather than two, kinetically related, ATP-dependent complexes, denoted α, β and γ, in order of decreasing mobility, which form on rabbit β-globin IVS-2 pre-mRNA. Complex γ is thought to represent the mature spliceosome. It may be resolution of the equivalent of the β and γ complexes which can be seen in band B of lanes i and j in Fig. 5.2. The resolution of heparin treated splicing complexes was superior when electrophoresis was performed using Tris-glycine buffer (Fig. 5.2) to when performed using Tris-borate buffer (Fig. 5.1).

The time course of splicing complex assembly on RNA54 (rabbit β-globin IVS-1) was found to be very similar to that observed in previous studies using different substrate RNAs (Konarska and Sharp, 1986; Lamond *et al.*, 1987), further reinforcing the assignment of bands A and B as pre-splicing complex and spliceosome respectively (Fig. 5.3). Sigma complexes alone were formed at time zero. By 5 minutes into the incubation, complex A, the putative pre-splicing complex could be seen. By 10 minutes the level of complex A had increased, and complex B was visible, and as in the case of Fig. 5.2, band B may in fact consist of two complexes. The level of complex A remained constant for the rest of the incubation, whereas the level of complex B continued to increase right up to the final time point, 90 minutes. In contrast, although the non-specific sigma complexes formed on [sA]RNA54 incubated in nuclear extract under splicing conditions, no low mobility bands corresponding to splicing complex assembly could be seen at any time (Fig. 5.3).

The failure to see splicing complex assembly on [sA]RNA54 could be due to large splicing complexes failing to form on [sA]RNA, or could be because complexes formed on [sA]RNA are unusually unstable, and do not remain attached to [sA]RNA during native gel electrophoresis. The splicing reactions electrophoresed on the gel in Fig. 5.3 were pre-treated with heparin before electrophoresis to improve resolution. However, it has previously been noted that heparin causes an apparent decrease in size of splicing complexes as judged both by native gel electrophoresis, and by glycerol gradient centrifugation (Konarska and Sharp, 1987). Hence,

Figure 5.2. **Splicing complex formation on RNA54 assayed by electrophoresis on a native low-percentage polyacrylamide gel buffered with Tris-glycine.**

Lanes a and *f* contain naked RNA54 which was not incubated with nuclear extract. *Lanes d, e, i* and *j* contain RNA54 which had been incubated for the times shown with HeLa nuclear extract under optimal *in vitro* splicing conditions (without polyvinyl alcohol) for the times shown in minutes. *Lanes c* and *h* contain RNA54 which had been added to a complete *in vitro* splicing mixture on ice and loaded directly (*lane c*) or the reaction stopped by the addition of heparin (*lane h*). *Lanes b* and *g* contain RNA54 which had been incubated under splicing conditions for 60 minutes but in the absence of ATP and creatine phosphate. The reactions in *lanes g-j* were stopped by the addition of heparin, whereas those in *lanes b-e* were loaded directly onto the gel without heparin treatment. All the samples were subjected to electrophoresis at room temerature on a 4% polyacrylamide gel buffered with 50mM Tris-glycine. The positions of the complexes are indicated as in Fig. 5.1.

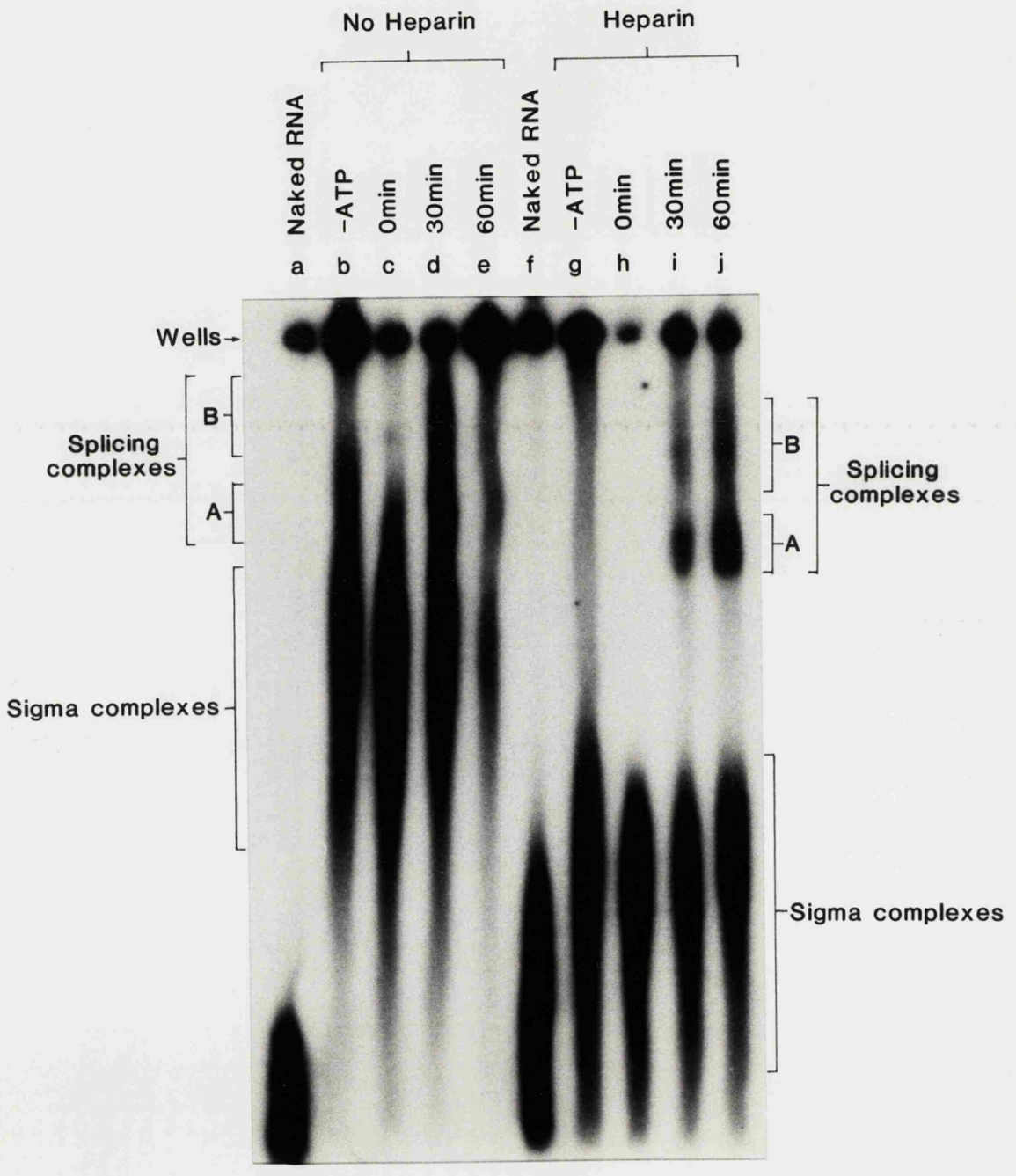
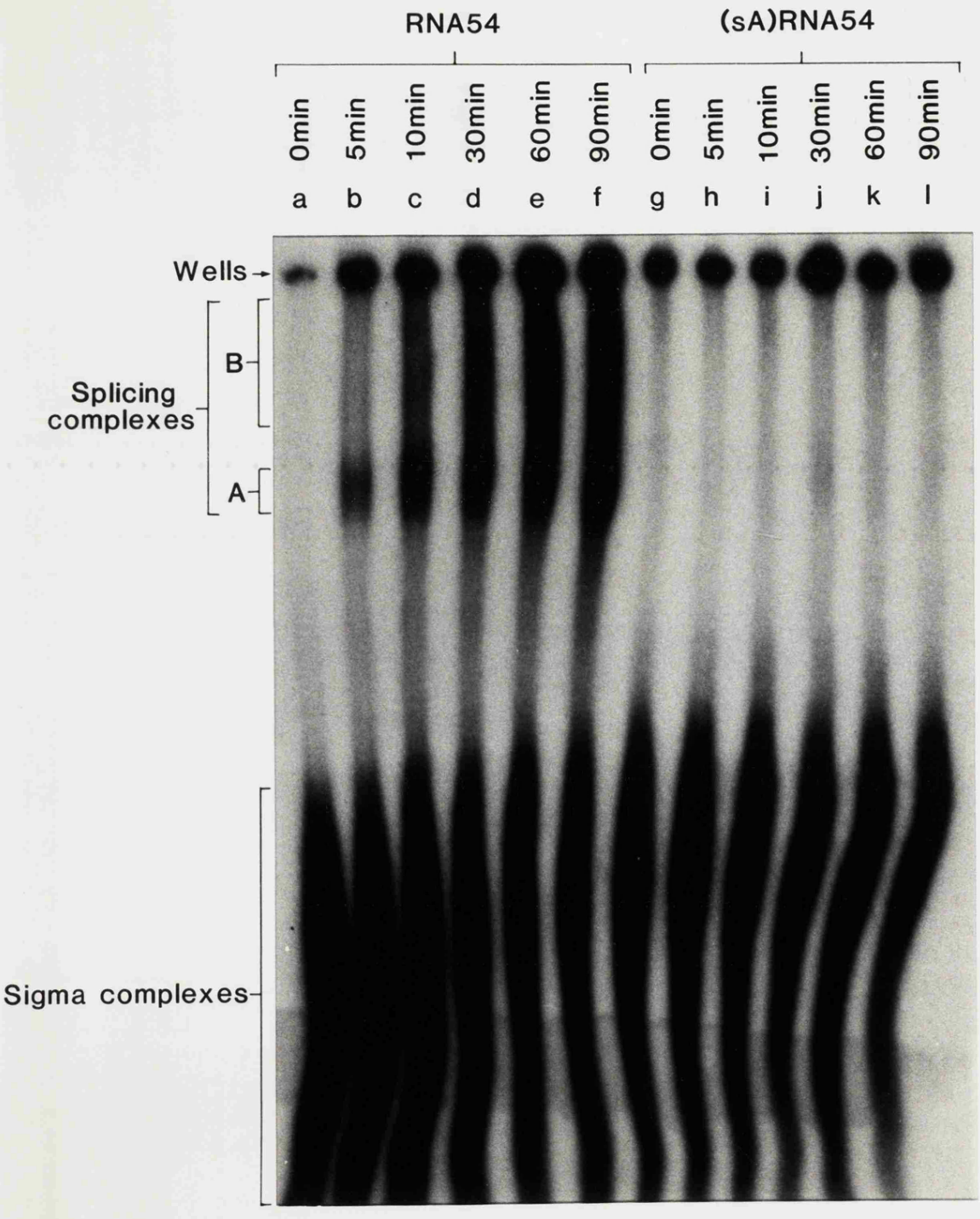

No Heparin
Heparin
Naked RNA
-ATP
0min
30min
60min
Naked RNA
-ATP
0min
30min
60min
a
b
c
d
e
f
g
h
i
j
Wells
B
A
Splicing complexes
Sigma complexes
B
A
Splicing complexes
Sigma complexes

Figure 5.3. **Time course of the assembly of splicing complexes on RNA54 and [sA]RNA54 assayed by electrophoresis on a native low-percentage polyacrylamide gel.**

RNA54 (*lanes a-f*) and [sA]RNA54 (*lanes g-l*) was incubated with HeLa nuclear extract under optimal *in vitro* splicing conditions (without polyvinyl alcohol) for the times shown in minutes. All the reactions were stopped by the addition of heparin, and after heparin treatment the splicing complexes were resolved by electrophoresis at room temperature on a 4% polyacrylamide gel buffered with 50mM Tris-glycine. The positions of the complexes are indicated as in Fig. 5.1.


RNA54
(sA)RNA54
0min
5min
10min
30min
60min
90min
0min
5min
10min
30min
60min
90min
a
b
c
d
e
f
g
h
i
j
k
l
Wells
Splicing complexes
B
A
Sigma complexes

factors associated with the spliceosome are definitely destabilised by heparin treatment. Indeed, affinity chromatography studies indicated that the association of U1 snRNP with the spliceosome was destabilised by heparin (Bindereif and Green, 1987). It was thought, therefore, that any low mobility complex associated with [sA]RNA may also be destabilised by the heparin treatment. To test this hypothesis, after incubation in nuclear extract under splicing conditions, [sA]RNA was fractionated by native electrophoresis on a Tris-glycine buffered gel, both with and without heparin treatment (Fig. 5.4). After incubation with ATP, both with and without heparin pre-treatment, splicing complexes A and B formed on RNA54, but no low mobility complexes formed on [sA]RNA54. Hence, if any large complex can form on [sA]RNA54 it is unstable to native electrophoresis even in the absence of heparin treatment. The mobility of the sigma complexes is lower with [sA]RNA54 than RNA54, both with and without heparin treatment, and independent of whether incubation in nuclear extract was in the presence or absence of ATP (Fig. 5.4). This difference in mobility must be due to a different composition, or different stability, of sigma complexes formed on [sA]RNA substrates to on normal RNA. This difference in mobility of sigma complexes on RNA54 and [sA]RNA54 may not be directly related to E2* formation, because E2* formation is ATP-dependent, whereas the sigma complex mobility shift was not.

When pre-treated with heparin, no complexes of low mobility were seen if RNA54 or [sA]RNA54 were incubated in extract in the absence of ATP (Fig. 5.4). However, if the heparin treatment was omitted, a complex of mobility intermediate between that of splicing complexes A and B could be seen with both RNA54 and [sA]RNA54 (Fig. 5.4). The nature of this ATP-independent, low mobility complex (denoted ϕ) is not known.

Splicing complex assembly on other phosphorothioate RNA substrates has also been investigated by native gel electrophoresis (Fig. 5.5). After 60 minutes incubation in nuclear extract under splicing conditions, heparin-resistant complexes were fractionated by native gel electrophoresis. RNA54 gave rise to high levels of splicing complexes A and B, as before. [sG]RNA54 gave rise to a band of the same mobility as the putative pre-splicing complex A, but no mature spliceosome band B. [sG]RNA54 also gave rise to a band of unknown composition (X), which ran only slightly higher than the sigma complexes. The majority of [sU]RNA54, for unknown reasons, failed to leave the well, making interpretation of the gel difficult. However, there was no obvious sign of splicing complex formation despite the fact that, of all the phosphorothioate substituted RNAs, this was the best substrate for a normal splicing reaction (see Chapter 4). [sC]RNA54 gave rise to a band of the same mobility as pre-splicing complex

Figure 5.4. **Analysis of splicing complex assembly on RNA54 and [sA]RNA54 by electrophoresis on a native low-percentage polyacrylamide gel with and without heparin treatment.**

RNA54 (*lanes a, c, e,* and *g*) and [sA]RNA54 (*lanes b, d, f,* and *h*) were incubated with HeLa nuclear extract under optimal *in vitro* splicing conditions (without polyvinyl alcohol) for 60 minutes either in the presence (*lanes c, d, g,* and *h*) or absence (*lanes a, b, e,* and *f*) of ATP (and creatine phosphate). The reactions in *lanes a-d* were loaded directly onto the gel without heparin treatment, whereas the reactions in *lanes e-h* were treated with heparin before loading. The splicing complexes were resolved by electrophoresis in an apparatus cooled with a water jacket on a 4% polyacrylamide gel buffered with 50mM Tris-glycine. The positions of the complexes are indicated as in Fig. 5.1, as is the position of the complex denoted ϕ (see text).

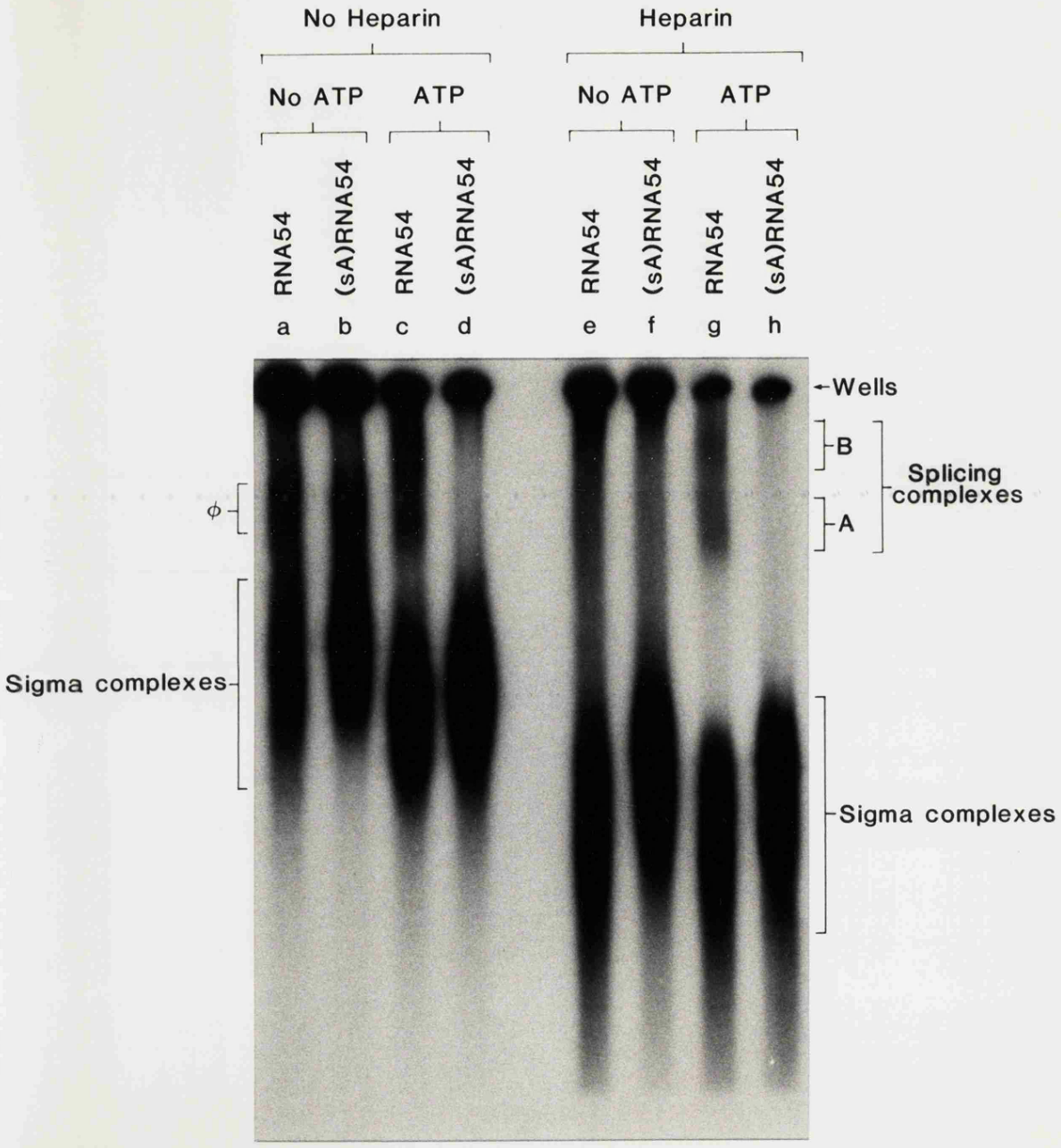
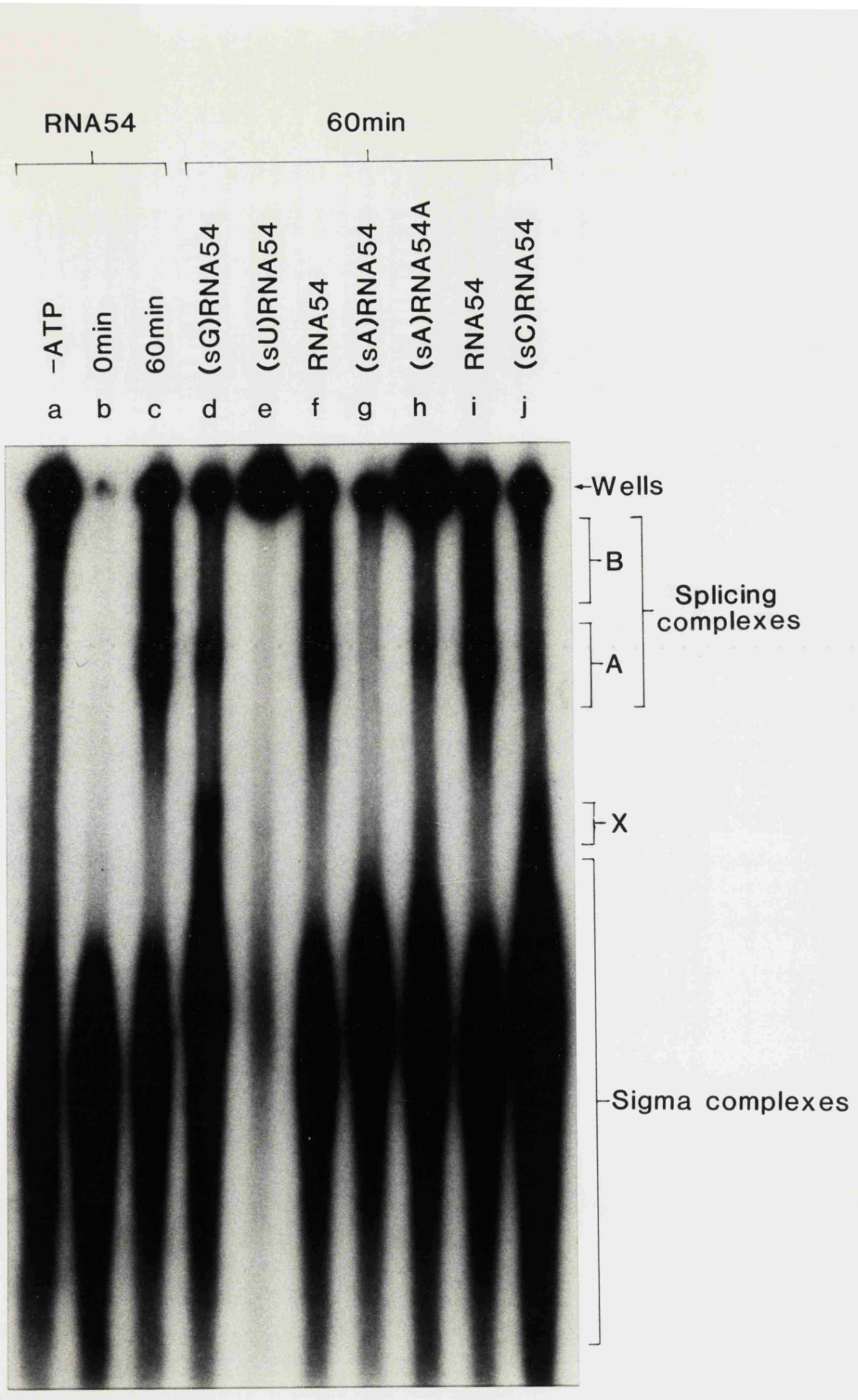

No Heparin
Heparin
No ATP
ATP
No ATP
ATP
RNA54
(sA)RNA54
RNA54
(sA)RNA54
RNA54
(sA)RNA54
RNA54
(sA)RNA54
a
b
c
d
e
f
g
h
Wells
B
A
Splicing complexes
ϕ
Sigma complexes
Sigma complexes

Figure 5.5. **Analysis of splicing complex assembly on [sG]RNA54, [sU]RNA54, [sC]RNA54, [sA]RNA54 and [sA]RNA54A by electrophoresis on a native low-percentage polyacrylamide gel.**

RNA54 was incubated in the presence (*lanes b, c, f,* and *i*) and absence (*lane a*) of ATP (and creatine phosphate) in HeLa nuclear extract under optimal splicing conditions (without polyvinyl alcohol) for 0 min (*lane b*) or 60 minutes (*lanes a, c, f,* and *i*). All phosphorothioate RNAs were incubated for 60 minutes under optimal splicing conditions (in the presence of ATP). *Lane d* contains [sG]RNA54, *lane e* [sU]RNA54, *lane g* [sA]RNA54, *lane h* [sA]RNA54A, and *lane j* contains [sC]RNA54. All the reactions were stopped by the addition of heparin, and after heparin treatment the splicing complexes were resolved by electrophoresis at room temperature on a 4% polyacrylamide gel buffered with 50mM Tris-glycine. The positions of the complexes are indicated as in Fig. 5.1, and the position of the band of unknown composition (band X, see text) found in *lane d* (which contains [sG]RNA54) is also shown.


RNA54
60min
a -ATP
b 0min
c 60min
d (sG)RNA54
e (sU)RNA54
f RNA54
g (sA)RNA54
h (sA)RNA54A
i RNA54
j (sC)RNA54
Wells
B
A
Splicing complexes
X
Sigma complexes

A, but no band appeared of the same mobility as complex B. The results with [sA]RNA were intriguing. [sA]RNA54, as before, gave rise to no low mobility complexes. In contrast, [sA]RNA54A (which differs from [sA]RNA54 only in that the sequence around the 3' splice site has been changed from AG/G to AG/A) formed a significant amount of complex of the same mobility as pre-splicing complex A. Once again there was no complex of the same mobility as complex B, the mature spliceosome. This result may be significant, given that [sA]RNA54A has been found to be a consistently better substrate for a normal splicing reaction than [sA]RNA54 (see for example Fig. 4.1B).

5.1.1.2. Sucrose density gradient centrifugation

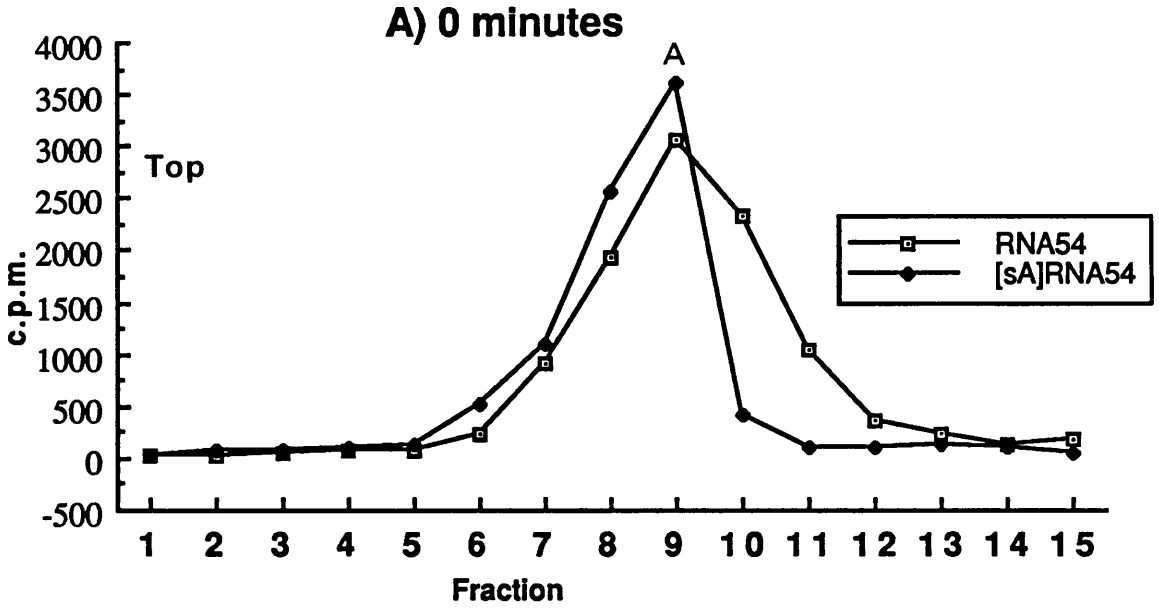
In vitro splicing reactions were performed with ^{32}P-labelled RNA54 or [sA]RNA54 as substrate, for 0 minutes, 30 minutes and 2 hours, and the reactions were fractionated by sedimentation on sucrose density gradients, in order to investigate complex formation (Frendewey and Keller, 1985). The sedimentation behaviour of the RNA was followed by Cerenkov counting of fractions from the gradient (Fig. 5.6).

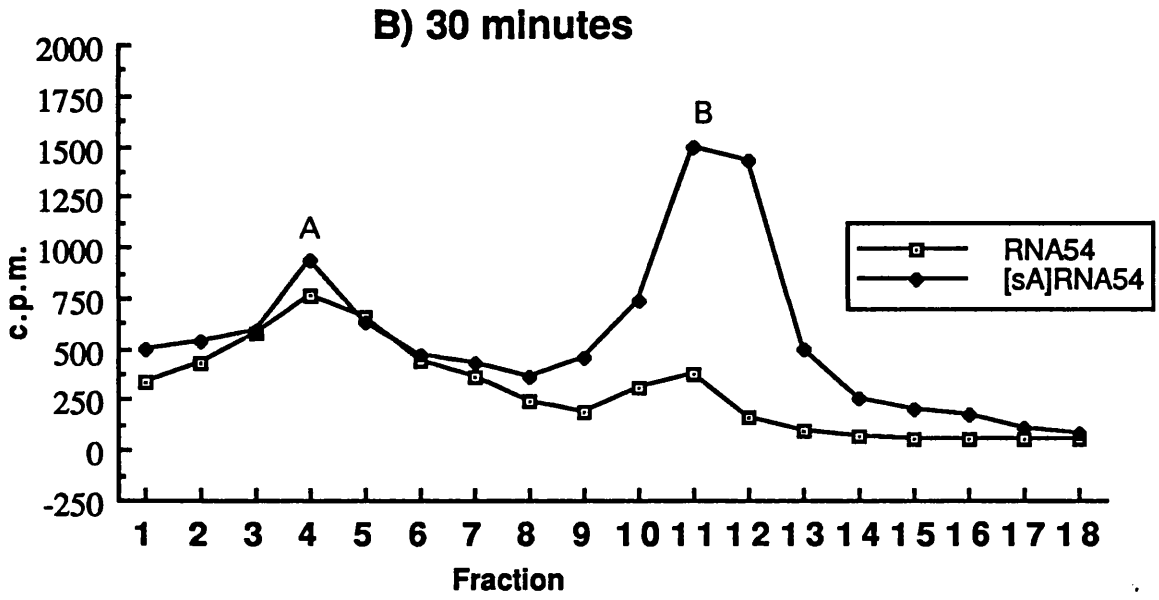
At 0 minutes, both RNA54 and [sA]RNA54, gave rise to a single peak with a nearly identical sedimentation rate (Oh Pk.A). This almost certainly corresponds to the 22S complex of Frendewey and Keller (1985). Pooled fractions from this peak, when deproteinised and electrophoresed on a denaturing polyacrylamide gel, were found to contain precursor RNA54 alone (Fig. 5.7). In the case of [sA]RNA54, this is probably also the case, however, the RNA was badly degraded, probably during fractionation.

The sedimentation profiles of RNA54 and [sA]RNA54 after incubation for 30 minutes in splicing conditions were similar (Fig. 5.6). Both substrates gave rise to 2 peaks which had similar sedimentation coefficients with both substrates. The faster sedimenting peak (0.5h Pk.B) had a similar sedimentation rate to the 0 minute peak, and, likewise, contained only precursor RNA (or [sA]RNA) with both substrates (Fig. 5.7). However this peak was markedly smaller with RNA54 than with [sA]RNA54, and even in the case of [sA]RNA54 the peak was smaller than at 0 minutes. This peak probably also corresponded to the 22S complex of Frendewey and Keller (1985), despite the fact that Frendewey and Keller (1985) could not detect the 22S complex by 15 minutes. With RNA54 the slower sedimenting peak (0.5h Pk.A) contained no detectable RNA of any description (Fig. 5.7). With [sA]RNA54, E2*, but no other RNA, was present in 0.5h Pk.A (Fig. 5.7). All of the counts in 0.5h Pk.A with RNA54 substrate, and most of the counts in 0.5h Pk.A with [sA]RNA54 substrate are probably due to free nucleotides or small oligonucleotides generated by degradation of RNA in the nuclear extract,

Figure 5.6. **Sedimentation behaviour of the ^{32}P-labelled products of incubation of RNA54 and [sA]RNA54 in nuclear extract under splicing conditions.**

RNA54 or [sA]RNA54, radiolabelled by transcription in the presence of [α-^{32}P]UTP, were either added to HeLa nuclear extract, on ice, and immediately frozen using dry ice (*panel A*), or incubated at 30°C in HeLa nuclear extract for 30 minutes (*panel B*), or 2 hours (*panel C*), before freezing. The reactions were centrifuged through 5%-20% sucrose gradients and the fractions were analysed by Cerenkov counting (see Chapter 2). The peaks from which RNA was taken, and analysed by electrophoresis on a denaturing gel (Fig. 5.7), are labelled. Sedimentation is from left to right in all profiles, i.e., the top of the gradient is on the left.


A) 0 minutes
A
Top
RNA54
[sA]RNA54
c.p.m.
4000
3500
3000
2500
2000
1500
1000
500
0
-500
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Fraction


B) 30 minutes
B
A
RNA54
[sA]RNA54
c.p.m.
2000
1750
1500
1250
1000
750
500
250
0
-250
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
Fraction

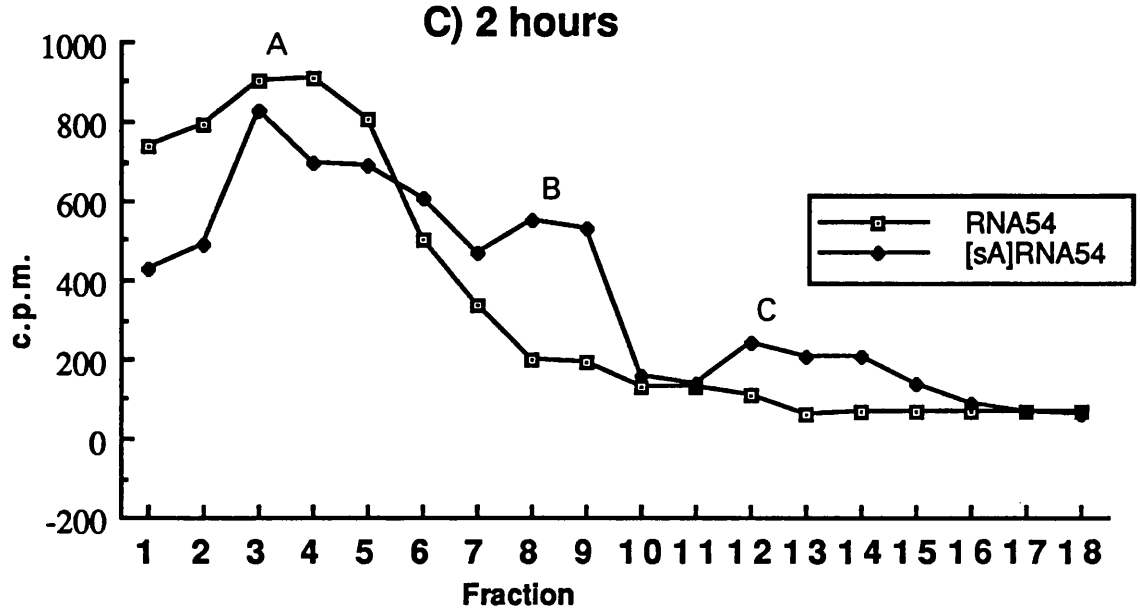
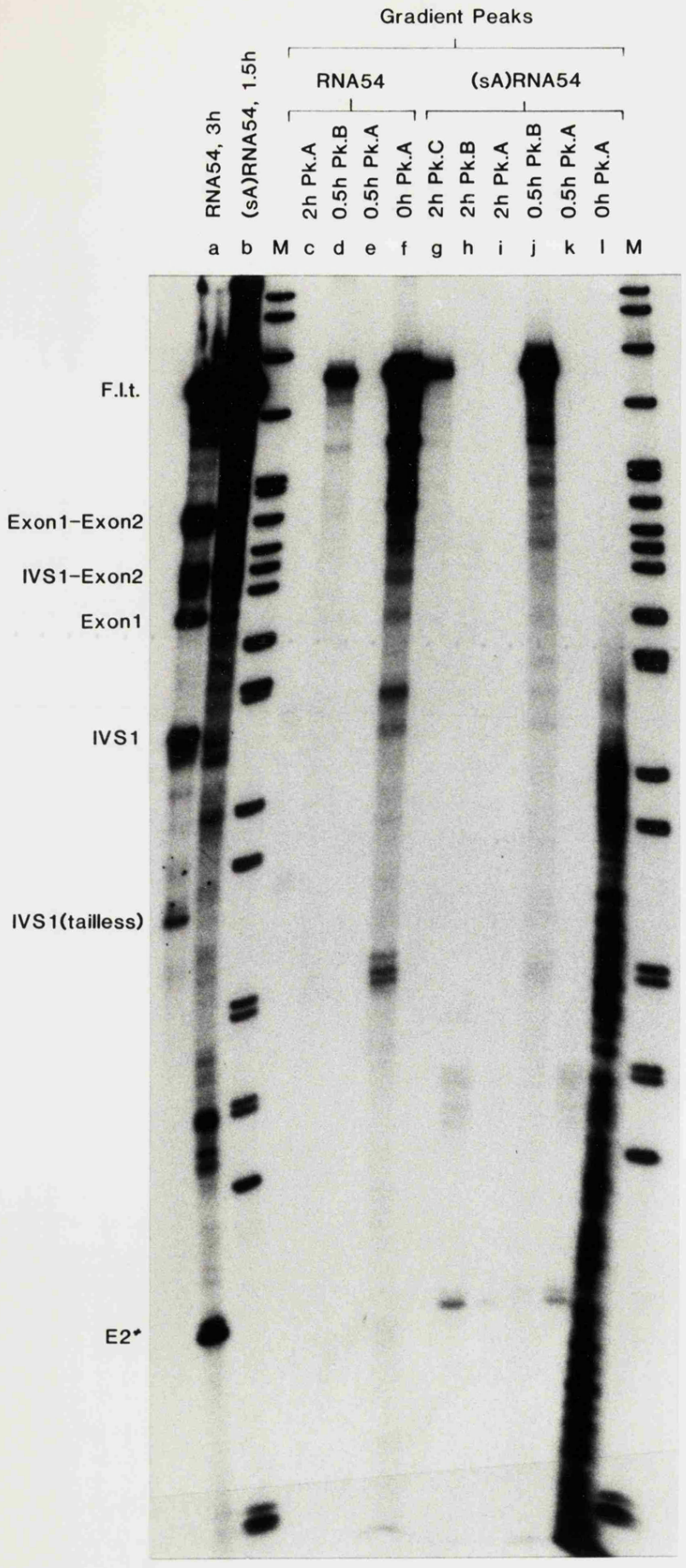

C) 2 hours
A
B
C
RNA54
[sA]RNA54
c.p.m.
1000
800
600
400
200
0
-200
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
Fraction

Figure 5.7. **Distribution of ^{32}P-labelled RNA species in the sucrose gradient.**

Fractions from each of the ^{32}P-peaks labelled in Fig. 5.6 were pooled, and when deproteinised, analysed by gel electrophoresis as described in Fig. 4.1. Molecular weight markers and indications of reaction products are as in Fig. 4.2. *Lanes c-f* contain the RNA species from the ^{32}P-peaks obtained from sucrose gradient centrifugation of splicing reactions performed with RNA54. *Lanes g-l* contain the RNA species from the ^{32}P-peaks obtained from sucrose gradient centrifugation of splicing reactions performed with [sA]RNA54. The lane nomenclature matches that of the ^{32}P-peaks labelled in Fig. 5.6. *Lane a* contains an unfractionated 10µl aliquot of the RNA54 splicing reaction which was fractionated by sucrose gradient centrifugation (shown in Fig. 5.6), except the incubation of this reaction was continued for a total of 3 hours. Similarly, *lane b* contains an unfractionated 10µl aliquot of the [sA]RNA54 splicing reaction which was fractionated by sucrose gradient centrifugation (shown in Fig. 5.6); incubation of this reaction was continued for a total of 1.5 hours.


Gradient Peaks
RNA54
(sA)RNA54
RNA54, 3h
(sA)RNA54, 1.5h
2h Pk.A
0.5h Pk.B
0.5h Pk.A
0h Pk.A
2h Pk.C
2h Pk.B
2h Pk.A
0.5h Pk.B
0.5h Pk.A
0h Pk.A
a b M c d e f g h i j k l M
F.l.t.
Exon1-Exon2
IVS1-Exon2
Exon1
IVS1
IVS1(tailless)
E2*

which sediment near the top of the gradient.

With RNA54 substrate, after 2 hours incubation in splicing extract, all the counts sedimented in a single peak (2h Pk.A) near the top of the gradient (Fig. 5.6). When pooled fractions from this peak were fractionated on a denaturing polyacrylamide gel, no RNA was visible (Fig. 5.7). Hence, this peak probably contained only free nucleotides, or small oligonucleotides, generated by degradation of RNA in the nuclear extract. In contrast, with [sA]RNA54 as substrate, after 2 hours incubation in nuclear extract 3 peaks were visible (Fig. 5.6). The peak of lowest sedimentation rate (2h Pk.A) was almost certainly analogous to the single peak with RNA54 substrate i.e. it contained only free nucleotides, or small oligonucleotides, since both peaks had similar sedimentation values, and the peak with RNA54 substrate contained no RNA, whilst that with [sA]RNA54 contained no RNA but for a trace of E2*. With [sA]RNA54, the peak with the highest sedimentation coefficient (2h Pk.C; Fig. 5.6) had a sedimentation value similar to that of the time zero peak (Oh Pk.A), and like this peak contained only precursor [sA]RNA54. 2h PkC was, however, much smaller. The complex in 2h Pk.C may, therefore, like that found in 0h Pk.A, be equivalent to the 22S complex identified by Frendewey and Keller (1985). With [sA]RNA54, the 2h Pk.B had a sedimentation value intermediate between that of the 2 other peaks. It contained no precursor RNA, but did contain E2*.

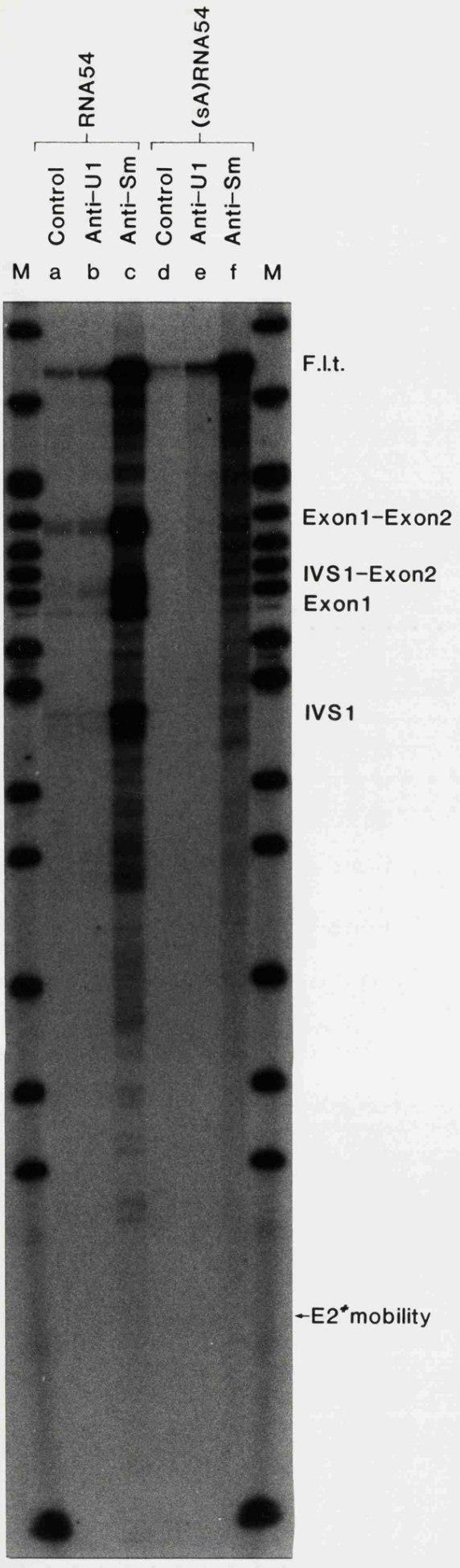
E2* would, therefore, seem to be in a relatively small complex, or no complex at all, as judged by its relatively low sedimentation coefficient. These experiments must, however, be interpreted with caution, since no complexes corresponding to the 35S putative pre-spliceosome or the 50S spliceosome complex (Frendewey and Keller, 1985) could be seen to form on RNA54, despite the *in vitro* splicing reaction itself proceeding efficiently (Fig. 5.7, track a). These complexes may have been destabilised under the centrifugation conditions used. Hence, this technique was not helpful in ascertaining whether large complexes form on [sA]RNA54 substrates under *in vitro* splicing conditions.

5.1.1.3. Immunoprecipitation studies

In order to investigate the interaction of U snRNPs with [sA]RNA54, complete *in vitro* splicing reactions were immunoprecipitated using anti-(U1)RNP and anti-(Sm)RNP antisera. Anti-(U1)RNP or anti-(Sm)RNP antibodies were added directly to the complete splicing system containing labeled RNA54 or [sA]RNA54 substrate 1.5 hours after starting the splicing reaction, and the RNA species recovered by binding to protein A Sepharose were analysed on a gel as Black *et al.* (1985) (Fig. 5.8).

Figure 5.8. **Immunoprecipitation of RNA species during incubation of RNA54 and [sA]RNA54 in HeLa nuclear extract under splicing conditions.**

RNA54 and [sA]RNA54, ^{32}P-labelled by transcription in the presence of [α-^{32}P]UTP, were incubated for 2 hours in HeLa nuclear extract under *in vitro* splicing conditions in 25µl reactions. The mixtures were then chilled on ice and immunoprecipitated as in Chapter 2 "Methods". The deproteinised RNAs were resolved on a denaturing gel as in Fig. 4.1. Immunoprecipitation was performed with 10µl of non-immune human serum (*control, lanes a* and *d*), 10µl anti-(U1)RNP antibody (*lanes b* and *e*), or 10µl anti-(Sm)RNP antibody (*lanes c* and *f*). Immunoprecipitated reactions with RNA54 as substrate are in *lanes a-c*, and reactions with [sA]RNA54 as substrate are in *lanes d-f*. Markers and descriptions of reaction products are as in Fig. 4.1.


RNA54
(sA)RNA54
Control
Anti-U1
Anti-Sm
Control
Anti-U1
Anti-Sm
M a b c d e f M
F.l.t.
Exon1-Exon2
IVS1-Exon2
Exon1
IVS1
←E2*mobility

Anti-(Sm)RNP antiserum immunoprecipitated precursor RNA54 and all the RNA products and intermediates of the splicing reaction indicating each of these species is associated with a species containing an Sm epitope (Fig. 5.8). In contrast, although precursor [sA]RNA54 was immunoprecipitated by anti-(Sm)RNP antiserum, and, therefore, would seem to interact with Sm snRNPs in active splicing extract, no immunoprecipitated E2* was visible, and so E2* would not seem to interact stably with an Sm snRNP.

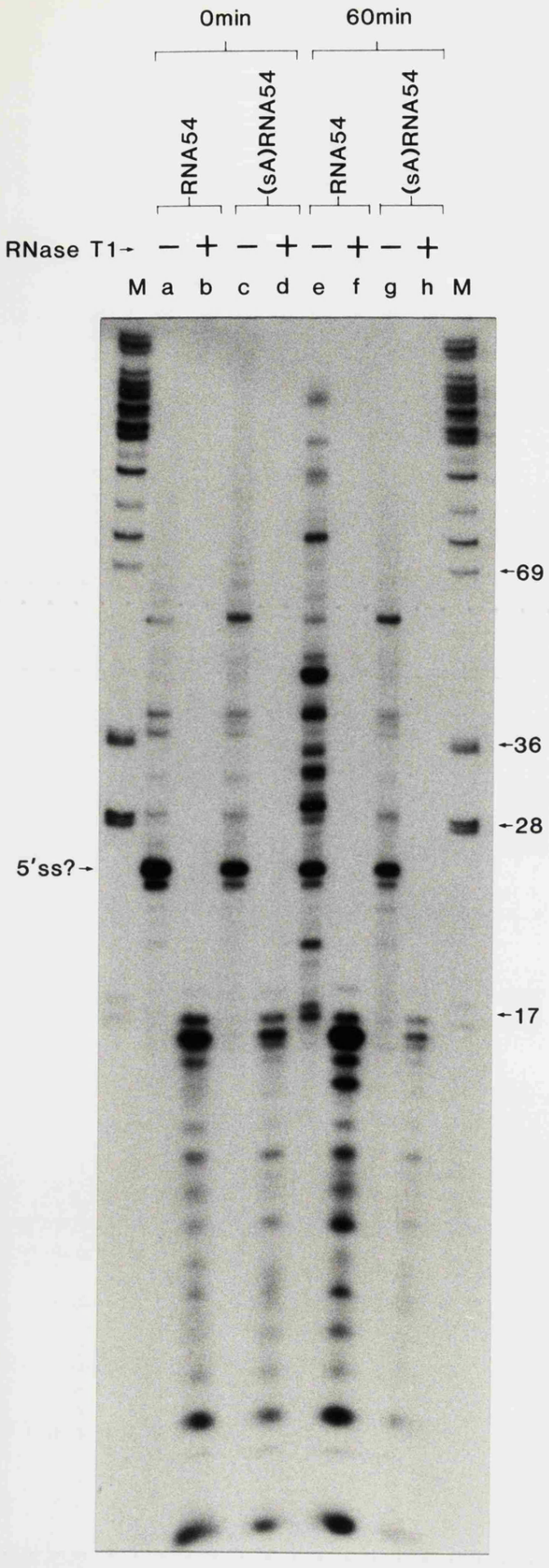
Anti-(U1)RNP antiserum, in contrast, barely precipitated more precursor RNA54, precursor [sA]RNA54, and RNA intermediates and products of splicing than did a control non-immune human serum (Fig. 5.8). This effect was not due to a deficiency in the concentration of anti-(U1)RNP antibody used, since this concentration caused quantitative immunoprecipitation of U1 snRNP from the volume of nuclear extract used in this experiment (data not shown). This is consistent with Black *et al.* (1985) and Grabowski *et al.* (1985) who also found levels of RNA precipitated by anti-(U1)RNP antibody to be low, the efficiency of precipitation decreasing as the splicing reaction proceeded. Black *et al.* (1985) suggested this to be due to sequestration of (U1)RNP determinants on formation of the larger splicing complexes, whereas Konarska and Sharp (1986) argued that this effect may be due to dissociation of the U1 snRNP during spliceosome assembly.

A combined RNase T1 digestion and immunoprecipitation assay, as described by Black *et al.* (1985), has been employed to elucidate the location and requirements for interactions between snRNPs and [sA]RNA54, and to compare these interactions with those between snRNPs and RNA54. Sites on the subtrate RNA associated with U1, U2, or Sm snRNPs were examined by adding RNase T1 along with anti-(U1)RNP, anti-(U2)RNP, anti-(Sm)RNP, or anti-(TMG) antibodies to the *in vitro* splicing reaction. The immunoprecipitated RNase resistant fragments were fractionated on gels.

The protected fragments of RNA54 and [sA]RNA54 associated with Sm snRNPs at 0 and 60 minutes are shown in Figure 5.9. Almost all these fragments were found to be genuinely protected fragments by subjecting half of the immunoprecipitated material to a second RNase T1 digestion after deproteinisation. With the exception of only one band, the immunoprecipitated RNA fragments were further reduced in size by this treatment (Fig. 5.9). At 0 minutes, the pattern of anti-(Sm)RNP immunoprecipitated fragments derived from both RNA54 and [sA]RNA54 was effectively identical, indicating that the interactions of Sm snRNPs with the two substrates was identical at time zero. In contrast, after 60 minutes incubation in splicing extract the pattern of anti-(Sm)RNP

Figure 5.9. **T1 ribonuclease protected fragments of RNA54 and [sA]RNA54 in splicing extract immunoprecipitated by anti-(Sm)RNP antibody.**

Immunoprecipitation of RNase T1 resistant fragments of RNA54 (*lanes a, b, e,* and *f*) and [sA]RNA54 (*lanes c, d, g,* and *h*) with anti-(Sm)RNP antibody. T1-resistant fragments were immunoprecipitated from extracts after incubation under splicing conditions for 0 minutes (*lanes a-d*), or 60 minutes (*lanes e-h*). After immunoprecipitation and treatment as described in Chapter 2 "Methods", the T1-resistant fragments were either digested to completion with RNase T1 (*lanes b, d, f,* and *h*), or electrophoresed directly (*lanes a, c, e,* and *g*). All fragments were resolved on a 7M urea, 20% polyacrylamide gel, buffered with 2xTBE. Size markers (*M*) were *Hpa*II fragments of pBR322, filled in with [α-^{32}P]dCTP and dGTP. The fragment thought to possibly correspond to the protected 5' splice site region is indicated (*5'ss?*).


0min
60min
RNA54
(sA)RNA54
RNA54
(sA)RNA54
RNase T1→ − + − + − + − +
M a b c d e f g h M
←69
←36
←28
5′ss?→
←17

immunoprecipitated fragments differed depending on whether the substrate was RNA54 or [sA]RNA54. With [sA]RNA54 the pattern of immunoprecipitated fragments was identical at 60 minutes and 0 minutes. However, with RNA54, at 60 minutes additional fragments were immunoprecipitated compared to those precipitated at time zero, some of these fragments being now relatively large (over 69 nucleotides). The increase in number and average size of protected RNA54 (rabbit β-globin IVS-1) fragments with time is consistent with previous experiments of this kind performed on human β-globin IVS-1 (Black *et al.*, 1985; Chabot and Steitz, 1987), and has been attributed to the assembly of a spliceosome on the RNA.

The RNA54 fragments immunoprecipitated in this study have not been characterised yet (although this is under way at present). However, this does not affect the major conclusions from this experiment i.e. the stable interactions formed between Sm snRNPs and RNA54 or [sA]RNA54 are identical at time zero, but although more extensive interactions occur between snRNPs and RNA54 after incubation under splicing conditions, the stable interaction between snRNPs and [sA]RNA54 remains unaltered by incubation.

The protected fragments immunoprecipitated by anti-(U1)RNP, anti-(U2)RNP, and anti-(TMG) antibodies are shown in Figure 5.10. At 0 minutes the pattern of immunoprecipitated fragments obtained was very similar with both anti-(U1)RNP and anti-(TMG) antibodies, and irrespective of whether the substrate was RNA54 or [sA]RNA54. The banding pattern was also very similar to that found using anti-(Sm)RNP antiserum at zero time (Fig. 5.9). This data is most easily explained if U1 snRNP, or a complex containing U1 snRNP, is the only stable interaction with both substrates at time zero. Unfortunately, once again, the fragments have not yet been characterised. However, it seems likely that the fragment labelled 5'ss? (Figs. 5.9, 5.10 and 5.11) is derived from the region including the 5' splice site. This tentative assignment is drawn from the observation that Black *et al.* (1985) found a 15 nucleotide fragment, encompassing the human β-globin IVS-1 5' splice site, to be the only fragment immunoprecipitated by anti-(U1)RNP antibody at time zero. The equivalent fragment from the rabbit β-globin IVS-1 5' splice site would be 22 nucleotides long, a very similar size to the major fragment precipitated by anti-(U1)RNP antibody in these experiments at 0 minutes. In contrast, the pattern of fragments immunoprecipitated by anti-(U2)RNP antibody at 0 minutes differs from the pattern obtained with anti-(U1)RNP or anti-(TMG) antibodies. At 60 minutes the pattern of RNA54 fragments immunoprecipitated by anti-(TMG) antibody increased in complexity and average length to give effectively the

Figure 5.10. **T1 ribonuclease protected fragments of RNA54 and [sA]RNA54 in splicing extract immunoprecipitated by anti-(U1)RNP, anti-(U2)RNP, or anti-(TMG) antibodies.**

Immunoprecipitation of RNase T1 resistant fragments of RNA54 (*lanes a, c, e, g, i,* and *k*) and [sA]RNA54 (*lanes b, d, f, h, j,* and *l*) with anti-(U1)RNP (*lanes a-d*), anti-(U2)RNP (*lanes e-h*), or anti-(TMG) (*lanes i-l*) antibodies. T1-resistant fragments were immunoprecipitated from extracts after incubation under splicing conditions for 0 minutes (*lanes a, b, e, f, i,* and *j*), or 60 minutes (*lanes c, d, g, h, k,* and *l*). After immunoprecipitation and treatment as described in Chapter 2 "Methods", the T1-resistant fragments were resolved on a gel as described in Fig. 5.9. Size markers were as in Fig. 5.9. The fragment thought to possibly correspond to the protected 5' splice site region is indicated (*5'ss?*).

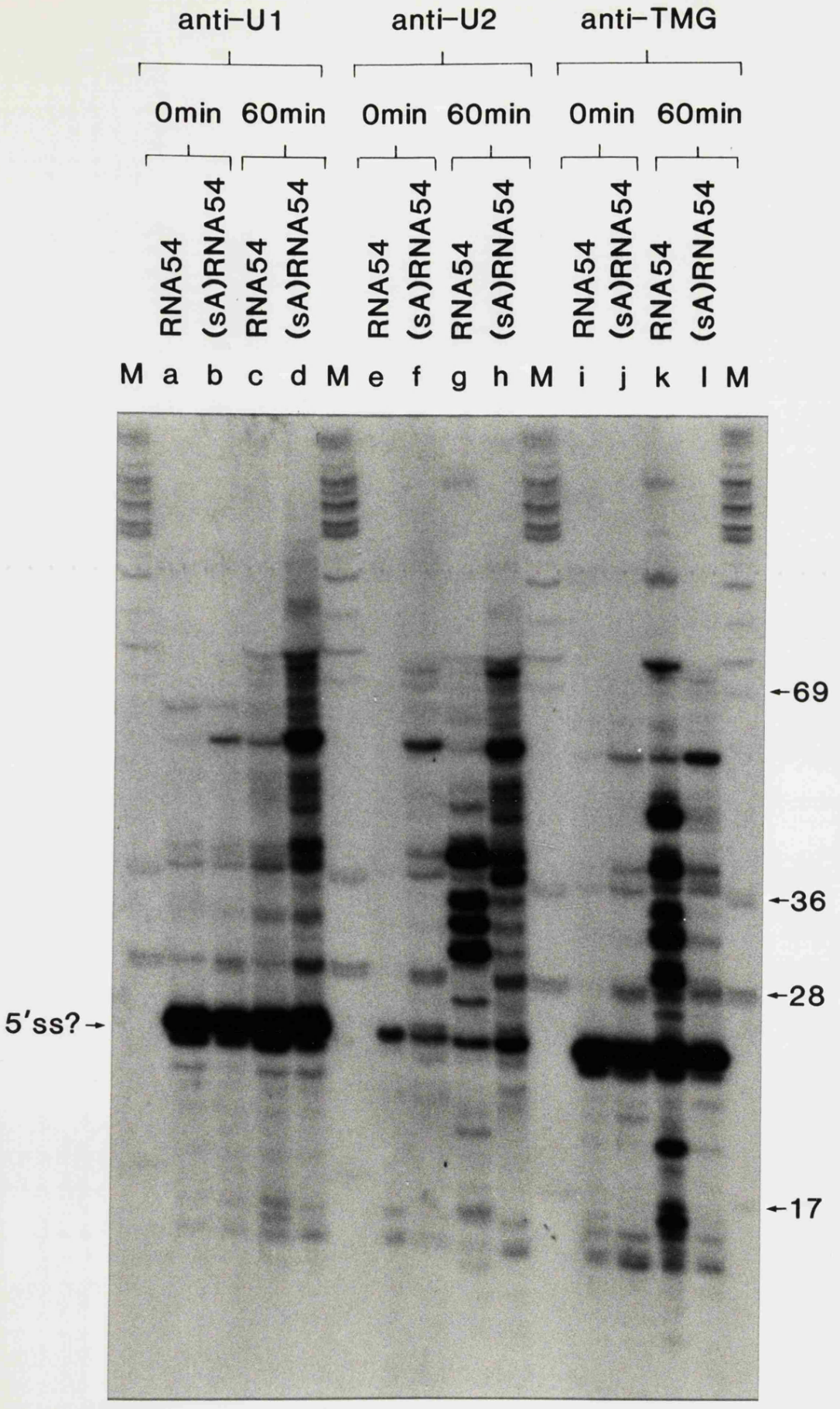

anti-U1
anti-U2
anti-TMG
0min
60min
0min
60min
0min
60min
RNA54
(sA)RNA54
RNA54
(sA)RNA54
RNA54
(sA)RNA54
RNA54
(sA)RNA54
RNA54
(sA)RNA54
RNA54
(sA)RNA54
M a b c d M e f g h M i j k l M
5'ss?
69
36
28
17

Figure 5.11. **T1 ribonuclease protected fragments of the deleted substrates RNA54Δ(Py)n, [sA]RNA54Δ(Py)n, RNA54Δ5'ss, and [sA]RNA54Δ5'ss in splicing extract immunoprecipitated by anti-(TMG) antibody.**

Immunoprecipitation of RNase T1 resistant fragments of RNA54Δ(Py)n (*lanes a* and *c*), [sA]RNA54Δ(Py)n (*lanes b* and *d*), RNA54Δ5'ss (*lanes e* and *g*), [sA]RNA54Δ5'ss (*lanes f* and *h*), and RNA54 (*lanes i* and *j*) with anti-(TMG) antibody. T1-resistant fragments were immunoprecipitated from extracts after incubation under splicing conditions for 0 minutes (*lanes a, b, e,* and *f*), or 60 minutes (*lanes c, d, g, h, i,* and *j*). After immunoprecipitation and treatment as described in Chapter 2 "Methods", the T1-resistant fragments were resolved on a gel as described in Fig. 5.9. Size markers were as in Fig. 5.9. The fragment thought to possibly correspond to the protected 5' splice site region is indicated (*5'ss?*).

0min
60min
0min
60min
60min
RNA54ΔP(y)n
(sA)RNA54ΔP(y)n
RNA54ΔP(y)n
(sA)RNA54ΔP(y)n
RNA54Δ5'ss
(sA)RNA54Δ5'ss
RNA54Δ5'ss
(sA)RNA54Δ5'ss
RNA54
M a b c d M e f g h M i j
5'ss?
69
36
28
17

same banding pattern as seen with anti-(Sm)RNP immunoprecipitation at this time (Fig. 5.9). Also, as with anti-(Sm)RNP antibody, the pattern of [sA]RNA fragments precipitated with anti-(TMG) antibody at 60 minutes was almost identical to that found at time zero.

The pattern of RNA54 fragments immunopecipitated by anti-(U1)RNP antibodies was, in contrast, little different at 60 minutes to at 0 minutes, although a few new fragments were visible by 60 minutes. Somewhat surprisingly, fragments of [sA]RNA54 of greater length than the 5'ss fragment were somewhat more efficiently precipitated than the equivalent RNA54 fragments. The amount of these higher molecular weight [sA]RNA54 fragments precipitated was significantly higher at 60 minutes than at 0 minutes and a few new species were brought down. Anti-(U1)RNP antibody noticeably precipitated a larger amount of fragments of higher molecular weight than the 5'ss? fragment when compared with anti-(TMG) antibody.

Anti-(U2)RNP antibody precipitated much lower amounts of the putative 5' splice site fragment (5'ss?) at all times than did the other antibodies (Fig. 5.10). With RNA54 substrate, very little additional RNA of higher molecular weight was precipitated at time zero, but with [sA]RNA54 the pattern of higher molecular weight fragments precipitated by anti-(U2)RNP antibody was very similar to that produced by the other antibodies at 0 minutes. At 60 minutes, anti-(U2)RNP antibody precipitated a series of extra RNA54 fragments to those precipitated at 0 minutes, indicating that during the course of incubation in the splicing extract more areas of RNA54 were interacting stably with complexes containing U2 snRNPs. At 60 minutes, with [sA]RNA54 as substrate, anti-(U2)RNP antibody precipitated greater amounts of fragments than at time zero, however the fragments precipitated were, for the most part, the same as those precipitated at time zero. However, a few seemed to correspond to fragments of RNA54 precipitated at 60 minutes.

The fragments immunoprecipitated by anti-(TMG) antibody at 0 minutes and 60 minutes from two deleted substrates are shown in Figure 5.11. RNA54Δ(Py)n and RNA54Δ5'ss were transcribed from *Hind*III cut **mICE10[IVS-1]Δ(Py)n** or **mICE10[IVS-1]Δ5'ss** R.F. DNA respectively. In RNA54Δ(Py)n the 5' splice site sequence AG/GUUGGU was deleted (where / indicates the site of cleavage), and in RNA54Δ5'ss the polypyrimidine tract was deleted from -30 to -5 (inclusive) relative to the 3' splice site, but both are otherwise as RNA54. Most of the tracks are somewhat smeared, probably due to salt in the samples. RNA54Δ5'ss is spliced *in vitro* by using a cryptic 5' splice site (see below), but RNA54Δ(Py)n is not spliced (see below). Both [sA]RNA54Δ5'ss and [sA]RNA54Δ(Py)n give rise to E2* (see below). The 5'ss? band was still precipitated at all times with both RNA54Δ(Py)n and

[sA]RNA54Δ(Py)n, yet was not precipitated at any time with either RNA54Δ5'ss or[sA]RNA54Δ5'ss, consistent with its assignment to the 5' splice site. Although detailed analysis of these results will require assignment of the protected fragments some information can still be obtained from this experiment. For example, certain bands were clearly visible with RNA54, RNA54Δ5'ss, [sA]RNA54Δ5'ss, RNA54Δ(Py)n and [sA]RNA54Δ(Py)n as substrates, indicating that they probably corresponded to fragments derived from protection of regions other than the 5' splice site or the polypyrimidine tract.

5.1.1.4. Affinity chromatography

Two methods of affinity chromatography have been attempted in order to isolate components stably associated with [sA]RNA54 after incubation under splicing conditions, and to compare them with those complexed with RNA54 under the same conditions.

a) Biotin-streptavidin affinity chromatography

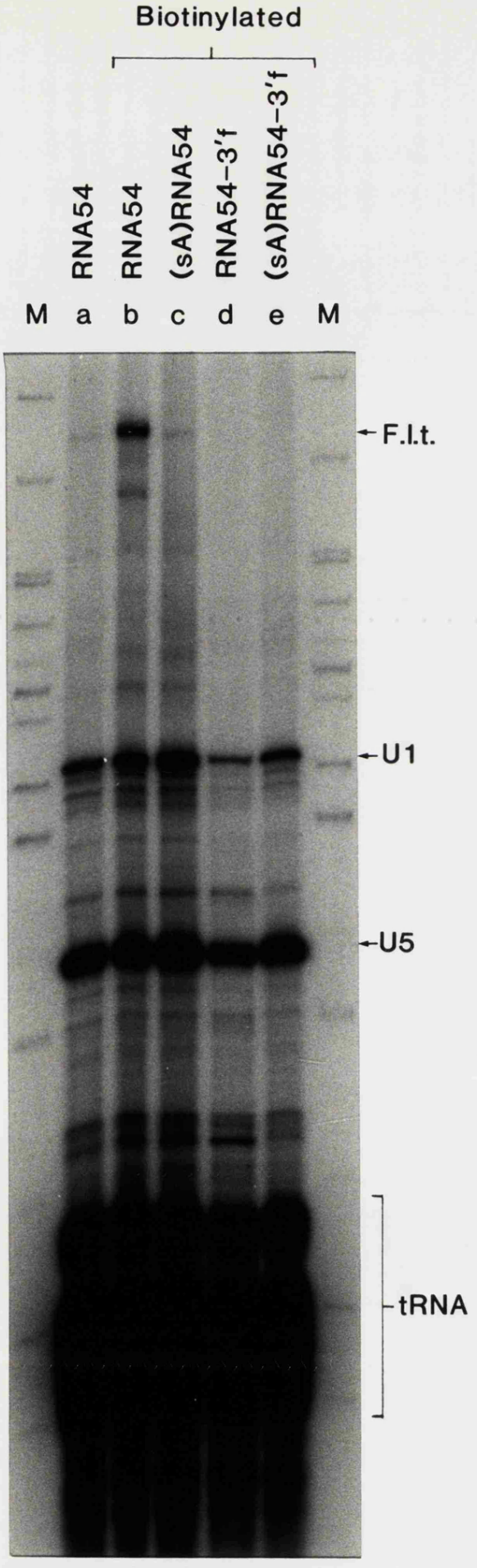
The strategy used was that described by Bindereif and Green (1987). The method involves synthesising ^{32}P-labelled pre-mRNA *in vitro* in the presence of a biotin-UTP analogue (biotin-11-UTP). Following incubation of the biotinylated pre-mRNA in nuclear extract under splicing conditions, stable complexes associated with the pre-mRNA can be selectively recovered on streptavidin immobilised on agarose beads (streptavidin-agarose).

^{32}P-labelled RNA54 and [sA]RNA54 were transcribed *in vitro* in the presence and absence of biotin-11-UTP. After incubation in a nuclear extract under splicing conditions only biotinylated species were recovered by binding to streptavidin-agarose, and both splicing and E2* formation was efficient (data not shown).

To identify the snRNPs complexed with [sA]RNA after incubation in nuclear extract under splicing conditions, the complexes bound to streptavidin-agarose were digested with proteinase K. The snRNPs released were then labelled at their 3' ends using T4 RNA ligase and [^{32}P]pCp, and identified by denaturing gel electrophoresis. The result of 3' end labelling the RNAs bound to streptavidin-agarose after incubating various subtrate RNAs for 30 minutes in nuclear extract is shown in Figure 5.12. Four biotinylated substrate RNAs were used, RNA54, [sA]RNA54, RNA54-3'f and [sA]RNA54-3'f. Production of the latter two species is described in the methods section. They consist of the downstream 86-101 nucleotides of RNA54 IVS-1 and the whole of exon-2 (the heterogeneity indicated at for the 5' end of these species is defined by the boundaries of the oligonucleotide used to direct RNase H cleavage). In addition, non-biotinylated RNA54 was

Figure 5.12. **RNAs recovered from splicing extract by biotin-streptavidin affinity chromatography.**

Non-biotinylated RNA54 (*lane a*), along with the following biotinylated substrates; RNA54 (*lane b*), [sA]RNA54 (*lane c*), RNA54-3'f (*lane d*), and [sA]RNA54-3'f (*lane e*), were incubated for 30 minutes in HeLa nuclear extract under splicing conditions (without PVA) in 150µl reaction volumes. Affinity chromatography was then performed using streptavidin-agarose as described in Chapter 2 "Methods", and the RNAs recovered were 3'-terminally labelled using [^{32}P]pCp and T4 RNA ligase. The labelled RNAs were fractionated on a 20% formamide, 7M urea, 6% polyacrylamide gel. Size markers were as Fig. 4.1. The bands thought to correspond to U1 and U5 snRNAs were assigned on the basis of their mobilities relative to the DNA markers and the published lengths of these snRNA species. The position of tRNAs and unprocessed RNA54 (*F.l.t.*) is also indicated.


Biotinylated
M
a
RNA54
b
RNA54
c
(sA)RNA54
d
RNA54-3'f
e
(sA)RNA54-3'f
M
F.l.t.
U1
U5
tRNA

spliced as a control substrate. As can be seen from Figure 5.12, the pattern of [^{32}P]pCp labelled RNAs is effectively identical for all the substrates, independent of whether or not they were biotinylated. Strong bands with the mobility of U1 snRNA (164 nucleotides), U5 snRNA (115 nucleotides), and the tRNAs were clearly visible. Hence, despite using the same conditions as those succesfully used by Bindereif and Green (1987), there was an excessively high level of non-specific binding of abundant RNA species in the nuclear extract, which obscured any specific binding of snRNAs by the biotinylated substrates.

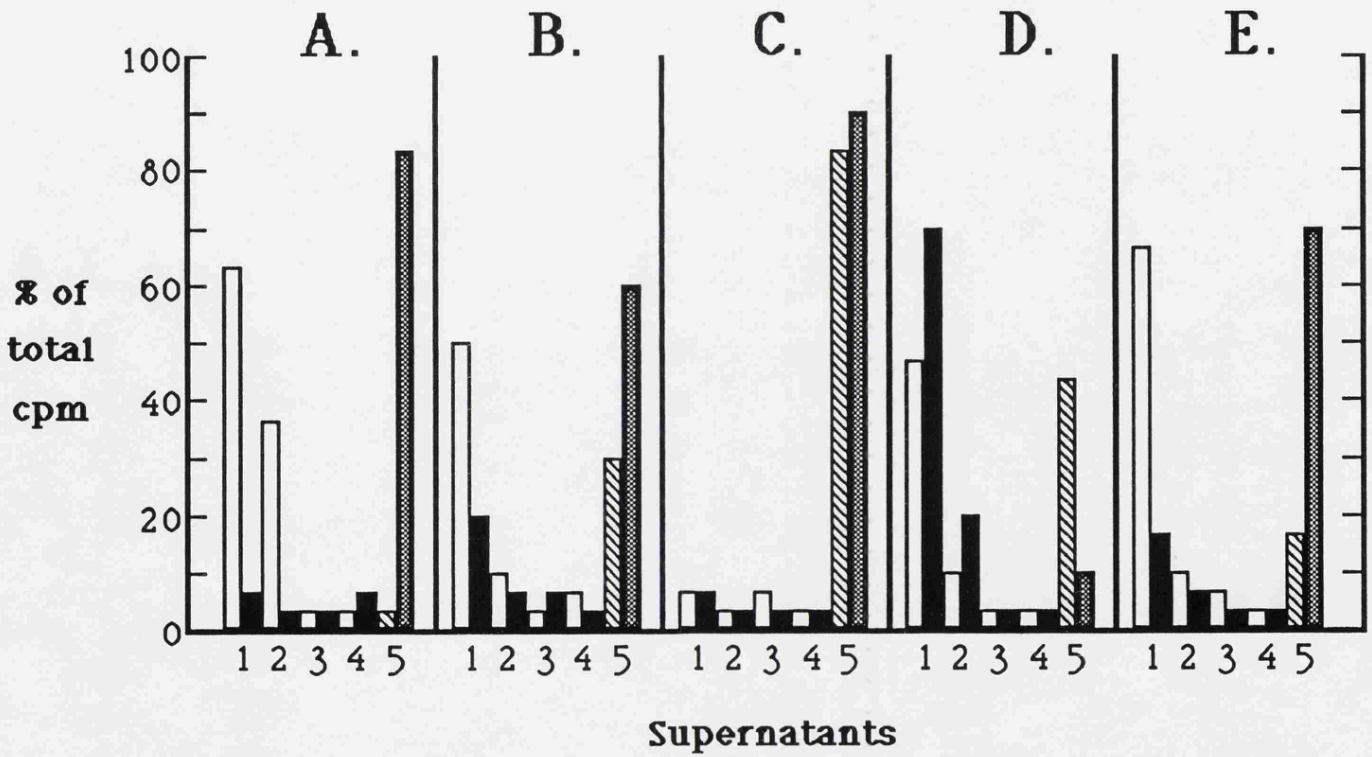
b) Mercury-agarose affinity chromatography

Nucleoside 5'-[γ-S]triphosphates (NTPγS) have previously been used to specifically label RNA initiated *in vitro*. This affinity labelled RNA can subsequently be isolated by affinity chromatography on an organomercurial matrix, namely Hg-agarose or Hg-cellulose (Reeve *et al.*, 1977; Smith *et al.*, 1978a). Nucleoside 5'-[γ-S]triphosphates and nucleoside 5'-[β-S]triphosphates have also been successfully used to study the initiation of transcription in nuclei from a variety of sources (see Marzluff and Huang, 1984, and references therein). The separation of newly-synthesised thiolated RNA (α-S-RNA), produced by transcription in the presence of NTPαS, from endogenous RNA, by affinity chromatography on Hg-cellulose or Hg-agarose has also been described (Marzluff and Huang, 1984).

Experiments were performed to determine whether Hg-agarose affinity chromatography could be used for the purification of factors complexed with [sA]RNA54. It was hoped that affinity purification could be performed in a buffer very similar to that used for *in vitro* splicing, since elution of phosphorothioate substituted RNA bound to the organomercurial matrix is performed simply by using an elution buffer containing 10mM DTT (Reeve *et al.*, 1977). Affi-gel 501 (BioRad) was the organomercurial agarose gel used: Affi-gel 501 has been successfully used to purify RNA initiated with NTPγS (Smith *et al.*, 1978b; Carroll and Wagner, 1979). Each binding assay utilised 2fmole of RNA54 or [sA]RNA54, transcribed in the presence of [α-^{32}P]UTP. Separation was tested using a variety of buffers (see section 2.31). Transcript which was unbound, and that released by elution with the appropriate buffer, made to 10mM with respect to DTT, was assayed by counting scintillations (Fig. 5.13). Marzluff and Huang (1984) reported that although binding of γ-S-RNA was efficient at pH7.9 (the optimal pH for splicing), binding of α-S-RNA to Hg-cellulose was very inefficient at this pH. They recommended that separation of α-S-RNA should be performed using a pH4.9 buffer, and using this buffer (buffer α)[sA]RNA54 was

Figure 5.13. **Bar graph showing the affinity selection of [sA]RNA54 using organomercurial agarose.**

RNA54 and [sA]RNA54, radiolabelled by transcription in the presence of [α-^{32}P]UTP, was subjected to affinity chromatography using Affi-Gel 501 (BioRad) as described in the chapter 2. Each assay used 2fmole RNA54 (or [sA]RNA54). Supernatants 1 to 4 (inclusive) were from the first four washes, which were performed, with the appropriate buffer, in the absence of DTT. Supernatant 5 was the supernatant after elution with the appropriate buffer made to 10mM DTT. RNA in the supernatants was assayed by Cerenkov counting and the results are plotted (to the nearest 3%) in terms of the percentage of the total counts in the five supernatants. The bars are infilled as follows: RNA54 eluted with buffer minus DTT (☐); [sA]RNA54 eluted with buffer minus DTT (■); RNA54 eluted with buffer plus 10mM DTT (▨); [sA]RNA54 eluted with buffer plus 10mM DTT (▩). The buffers used for elution are described in Chapter 2: *panel A* shows elution with buffer α; *panel B* shows elution with buffer β; *panel C* shows elution with γ; *panel D* shows elution with buffer δ; *panel E* shows elution with buffer ε.


A.
B.
C.
D.
E.
100
80
60
40
20
0
% of total cpm
1 2 3 4 5
Supernatants

specifically retained on the matrix, and released by treatment with 10mM DTT, whereas RNA54 was neither specifically bound, nor released by DTT (Fig. 5.13, *panel A*). The use of the pH7.9 buffer (buffer β) recommended for separation of γ-S-RNA (Reeve *et al.*, 1977) resulted in a marginally less efficient retention of [sA]RNA54 than did buffer α: RNA54 was also retained, and subsequently released by DTT, to an unacceptable extent (Fig. 5.13, *panel B*). This retention and DTT induced release of unsubstituted RNA was unnexpected, and was not seen in previous studies using Affi-gel 501 (Smith *et al.*, 1978b; Carroll and Wagner, 1979). A similar retention and DTT induced release of RNA54 was seen when chromatography was performed in a column (as in previous studies), rather than by centrifugation (data not shown). Using the same buffer as for *in vitro* splicing (pH7.9) (termed buffer γ in this study) both RNA54 and [sA]RNA54 were retained on the matrix and eluted by DTT (Fig. 5.13, *panel C*). Making buffer γ 0.5μg/μl with yeast RNA (now termed buffer δ) also failed to improve the selective retention of [sA]RNA54 (Fig. 5.13, *panel D*). A buffer of similar composition to splicing buffer, but made to pH4.9 using sodium acetate (buffer ε), not pH7.9 (with triethanolamine), and which lacked PMSF and glycerol, proved to specifically retain, and subsequently release, [sA]RNA54 quite efficiently, when treated with DTT (Fig. 5.13, *panel E*). However, although in this buffer only a relatively small amount of RNA54 was released by DTT, the amount released was still unnacceptably high (Fig. 5.13, *panel E*). The main difference between buffer α (which was effective) and buffer ε (which was not) is that buffer α contained 0.1% SDS and 40mM EDTA, whereas buffer ε contained no SDS and only 0.2mM EDTA. For the purposes of affinity purification of factors bound to [sA]RNA54 during splicing reactions addition of SDS to improve specificity is not practical. However, splicing complexes are stable when electrophoresed through native gels containing 10mM EDTA (Zillmann *et al.*, 1988); hence, since mercuric ions are known to form complexes with nucleic acids (Gruenwedel and Davidson, 1967) purification specificity for binding of [sA]RNA54 could perhaps be increased by increasing the EDTA concentration in buffer ε.

However, since the buffer needed for efficient affinity purification of [sA]RNA54 seems to be of necessity pH4.9, rather than pH7.9 (the pH at which splicing is optimal), and as other changes from the conditions optimal for splicing also seem necessary, it was decided not to pursue this line of experimentation any further.

5.1.2. PRE-MESSENGER RNA SEQUENCES NECESSARY FOR E2* FORMATION

The regions of [sA]RNA54 involved in interactions which are obligatory for E2* formation have been investigated, mainly by the use of deleted or truncated transcripts.

5.1.2.1. Deletion of the 5' splice site and polypyrimidine tract

Two deleted versions of RNA54 (or [sA]RNA54) were produced. The first of these RNAs, RNA54Δ5'ss (or [sA]RNA54Δ5'ss), was as RNA54, but the 5' splice site sequence AG/GUUGGU (where / indicates the site of cleavage) was deleted. The second RNA, RNA54Δ(Py)n (or [sA]RNA54Δ(Py)n), was as RNA54, but the polypyrimidine tract from -30 to -5 (inclusive) relative to the 3' splice site was deleted.

A time course of *in vitro* splicing of the non-phosphorothioate and [sA] versions of these deletion mutants is shown in Figure 5.14. RNA54 spliced efficiently, and [sA]RNA54 gave rise to E2*, as usual.

RNA54Δ5'ss, despite lacking the normal 5' splice site region, was also spliced efficiently (Fig. 5.14). The mobilities of the products and intermediates of this splicing reaction are consistent with the use of a potential cryptic 5' splice site 16 nucleotides upstream of the normal 5' splice site. This would seem likely, since the sequence of the human and rabbit β-globin genes are identical in the region immediately upstream of the 5' splice site, including this potential cryptic site (Lawn *et al.*, 1980; van Ooyen *et al.*, 1979), and if the highly conserved GT of the human β-globin IVS-1 wild type 5' splice site is mutated to AT the major splicing event *in vitro* is splicing using a cryptic 5' splice site 16 nucleotides upstream of the wild type cleavage site (Krainer *et al.*, 1984; Chabot and Steitz, 1987a). [sA]RNA54Δ5'ss efficiently gave rise to E2* (Fig. 5.14). This is, however, only to be expected if RNA54Δ5'ss is efficiently spliced.

RNA54Δ(Py)n was spliced very ineficiently, if at all (Fig. 5.14). No products of step 2 of splicing were visible, and there was only a very small amount of exon 1, indicating that step one was very ineficient. This is in agreement with *in vitro* splicing experiments performed with deleted human β-globin IVS-1 RNA (Ruskin and Green, 1985b). [sA]RNA54Δ(Py)n, however, gave rise to E2* with similar efficiency to [sA]RNA54. Interactions with the polypyrimidine tract are, therefore, not required for the production of E2*.

5.1.2.2. Splicing of truncated subtrate RNAs

Primed mICE10[IVS-1] viral DNA was used as the template for the production of 3'-truncated versions of RNA54. One oligonucleotide directed the production of RNA54-5'f, an RNA terminating 40 nucleotides downstream

Figure 5.14. **Splicing *in vitro* of transcripts with deletions of either the 5' splice site or polypyrimidine tract with and without phosphorothioate linkages 5' to adenosines.**

Transcipts were synthesised from *Hind*III cut **mICE10[IVS-1]** R.F. to generate RNA54, from *Hind*III cut **mICE10[IVS-1]Δ5'ss** R.F. to generate **RNA54Δ5'ss**, or from *Hind*III cut **mICE10[IVS-1]Δ(Py)n** to generate **RNA54Δ(Py)n**. Phosphorothioate analogues of these transcripts were also synthesises by transcription in the presence of ATPαS (generating [sA]RNA54, [sA]RNA54Δ5'ss, and [sA]RNA54Δ(Py)n respectively). [α-^{32}P]UTP was present in all the transcription reactions. Splicing reactions were incubated for the times shown in hours. Analysis by gel electrophoresis was as described in Fig. 4.1. Molecular weight markers and the designation of molecular species are as in Fig. 4.2.

RNA54
(sA)RNA54
RNA54Δ5'ss
(sA)RNA54Δ5'ss
RNA54ΔP(y)n
(sA)RNA54 ΔP(y)n
0.5h 1h 2h 3h
M a b c d e f g h M i j k l m n o p M q r s t u v w x M
F.l.t.
Exon1-Exon2
IVS1-Exon2
Exon1
IVS1
IVS1(tailless)
E2

of the 5' splice site i.e. in IVS-1. Another oligonucleotide, this time with a mismatched 5' terminus, directed the synthesis of RNA54-5'trans, which was identical to RNA54-5'f except for possessing an additional 25 nucleotides at the 3' terminus. These 3' terminal nucleotides were antisense RNA, complementary to the rabbit β-globin IVS-1 nucleotides 40 to 65 nucleotides downstream of the 5' splice site.

A 5' truncated version of RNA54 was produced by DNA oligonucleotide directed RNase H cleavage of RNA54. The oligonucleotide used was complementary to the the IVS-1 nucleotides 25 to 40 nucleotides downstream of the 5' splice site. After cleavage, the 3' fragment (RNA54-3'f) was purified by electrophoresis on a denaturing gel. An example of such a gel, used to purify [sA]RNA54-3'f, is shown in Fig. 5.15. The lower bands are fragments produced by a fraction of the oligonucleotide annealing to a second site.

The result of incubating RNA54-5'f, RNA54-5'trans, RNA54-3'f and their [sA]RNA analogues in nuclear extract under splicing conditions for 3 hours is shown in Figure 5.16. The only one of these RNAs to generate E2* was [sA]RNA54-3'f. This indicates that E2* formation in all probability does not depend on interactions with 5' splice site sequences. The possibility that E2* formation does still require some level of interaction with an unidentified sequence resembling that of a 5' splice site still present in [sA]RNA54-3'f cannot be totally excluded. This would, however, seem unlikely, since there seem to be no obviously good potential cryptic sites in [sA]RNA54-3'f, and as there is no sign of splicing of RNA54-3'f there would certainly appear to be no functional cryptic sites.

It can also be seen from Figure 5.16, that much RNA54-5'f is converted to an RNA of lower molecular weight, (approximately 184 nucleotides) denoted RNA Y. As exon 1 is 175 nucleotides long, this fragment could be explained by a 3' exonuclease activity which was blocked by a complex protecting the 5' splice site region. [sA]RNA54-5'f also gave rise to RNA Y, but the yield of Y was somewhat less than with RNA54-5'f. This reduced yield could be due to the inhibition of the putative 3' exonuclease activity by thio-phosphodiester linkages. The amount of RNA Y produced by RNA54-5'trans is even lower. If the exonuclease hypothesis is correct, this must result from the extra 25 nucleotides at the 3' terminus of RNA54-5'trans conferring added resistance to the 3' exonuclease, perhaps even in a sequence dependent manner. [sA]RNA54-5'trans gives rise to even less RNA Y than any of the previous substrates, further reinforcing the possibility that phosphorothioates inhibit the putative 3' exonuclease activity.

The kinetics of E2* formation with [sA]RNA54 and [sA]RNA54-3'f as

Figure 5.15. **Preparation of [sA]RNA54-3'f.**

The figure shows the autoradiogram of a 6% polyacrylamide, 20% formamide, 7M urea gel on which an [sA]RNA54 transcript, ^{32}P-labelled by transcription in the presence of [α-^{32}P]UTP, had been electrophoresed after RNase H cleavage directed to within the intron by the oligonucleotide "170-185 cut". The positions of the major products of RNase H cleavage (at the target sequence, 170-185 nucleotides downstream of the rabbit β-globin *in vivo* cap site) *[sA]RNA54-5'f*, corresponding to the 5'-fragment, and *[sA]RNA54-3'f*, corresponding to the 3'-fragment, are indicated, along with the position at which uncleaved [sA]RNA54 migrates (*Mobility of [sA]RNA54*). [sA]RNA54-3'f was subsequently excised and eluted.

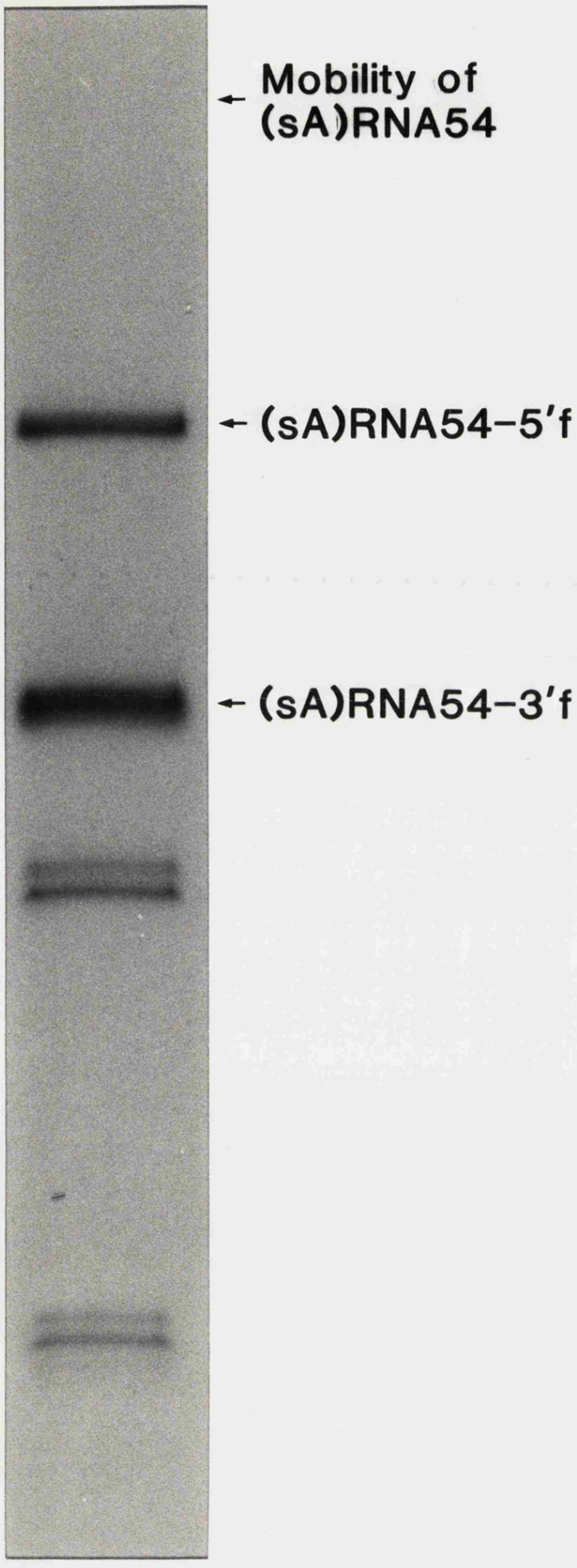
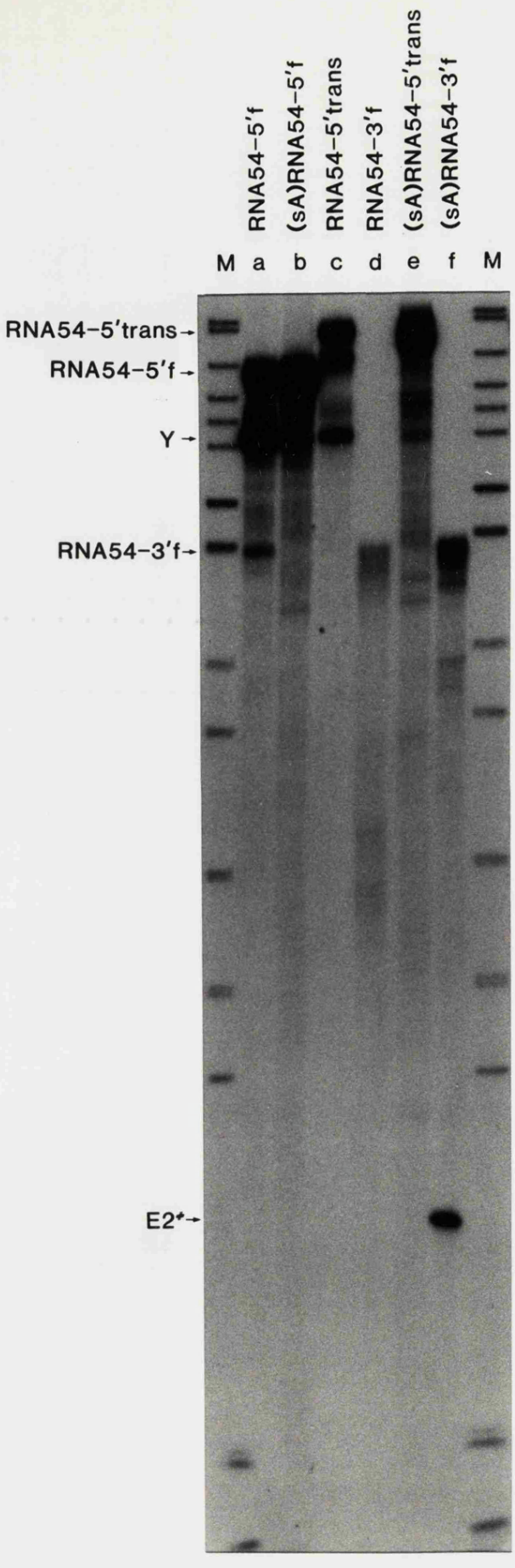

Mobility of
(sA)RNA54
(sA)RNA54-5'f
(sA)RNA54-3'f

Figure 5.16 **Splicing *in vitro* of 5'- and 3'-truncated transcripts with and without phosphorothioate linkages 5' to adenosine.**

RNA54-5'f (*lane a*) and [sA]RNA54-5'f (*lane b*) were 3'-truncated RNAs terminating 40 nucleotides downstream of the 5' splice site, synthesised as described in the text. RNA54-5'trans (*lane c*) and [sA]RNA54-5'trans (*lane e*) were RNAs identical to RNA54-5'f and [sA]RNA54-5'f except for possessing an additional 25 nucleotides at the 3'-terminus. These 3'-terminal nucleotides were antisense RNA, complementary to the rabbit β-globin IVS-1 nucleotides 40-65 nucleotides downstream of the 5' splice site. Synthesis of these RNAs is also described in the text. RNA54-3'f (*lane d*) and [sA]RNA54-3'f (*lane f*) were 5'-truncated RNAs whose 3' termini lay 25-40 nucleotides downstream of the 5' splice site, and which were generated by oligonucleotide directed RNase H cleavage (see text and Chapter 2 "Methods"). All transcriptions were performed in the presence of [α-^{32}P]UTP. Each of the above RNAs was incubated in HeLa nuclear extract under splicing conditions for 3 hours. Analysis by gel electrophoresis was described for Fig. 4.1. Molecular weight markers were also as in Fig. 5.1. The positions of E2*, unprocessed RNA54-5'trans, unprocessed RNA54-5'f, unprocessed RNA54-3'f, and fragment Y (see text) are indicated.


RNA54-5'f
(sA)RNA54-5'f
RNA54-5'trans
RNA54-3'f
(sA)RNA54-5'trans
(sA)RNA54-3'f
M a b c d e f M
RNA54-5'trans
RNA54-5'f
Y
RNA54-3'f
E2*

substrates are compared in Figure 5.17. As seen previously (Fig. 4.4), with [sA]RNA54 as substrate, the kinetics of appearance of E2* closely paralleled the kinetics of appearance of intermediates of splicing with RNA54 as substrate. A trace of E2* can though, be seen as early as 10 minutes, when no splicing intermediates can be seen. However, a similarly small quantity of splicing intermediates would not be visible above the background. With [sA]RNA54-3'f as substrate, E2* also first became visible at 10 minutes. However, the yield of E2* at these early times was significantly greater than with [sA]RNA54.

[sA]RNA54-3'f can also be seen to be degraded much more rapidly on incubation under splicing conditions than either RNA54 or [sA]RNA54 (Fig. 5.17). Whilst the precursors RNA54 and [sA]RNA54 were still present even after a 3 hour incubation, no [sA]RNA54-3'f was detectable after 1.5 hours incubation, and the vast majority had been degraded even by 30 minutes. The fact that E2* was still present after a 3 hour incubation, whereas [sA]RNA54-3'f had been completely degraded after only 1.5 hours indicates that E2* was more stable under *in vitro* splicing conditions than [sA]RNA54-3'f: this could be due to [sA]RNA54-3'f being particularly unstable (perhaps it contains sequences targeting it for degradation), or that E2* is particularly resistant to degradation (perhaps it is found in a complex), or a combination of these factors.

As E2* was still produced from [sA]RNA54-3'f, a substrate lacking a 5' splice site sequence, it was thought worthwhile to examine whether oligonucleotide directed RNase H cleavage of U1 snRNA had a similar inhibitory effect on E2* production with [sA]RNA54-3'f as substrate to with [sA]RNA54 as substrate (Fig. 5.18). Cleavage of U1 and U2 snRNAs was performed exactly as in Chapter 4. Cleavage of U1 snRNA was, as before, incomplete (data not shown), but still resulted in substantial inhibition of RNA54 splicing (Fig. 5.18), and inhibited E2* production from both [sA]RNA54 or [sA]RNA54-3'f. Cleavage of U2 snRNA completely inhibited both steps of splicing with RNA54 substrate, and inhibition of E2* formation from [sA]RNA54 and [sA]RNA54-3'f was similar to that seen after cleavage of U1 snRNA. Some E2* was visible, even after cleavage of U1 or U2 snRNAs, when the substrate was [sA]RNA54-3'f, but not when the substrate was [sA]RNA54. Some of this effect may, however, not be due to decreased inhibition of E2* production when the substrate was [sA]RNA54-3'f rather than [sA]RNA54, but a consequence of the overall higher yield of E2* from [sA]RNA54-3'f. With [sA]RNA54-3'f as substrate, after cleavage of U1 or U2 snRNAs, a distinct ladder of bands of unknown significance was visible immediately above the E2* band.

The other significant observation from Figure 5.18 was that

Figure 5.17 **Time dependence of splicing and formation of E2* from RNA54, [sA]RNA54 and the 5'-truncated substrate [sA]RNA54-3'f.**

RNA54 (*lanes s-α*), [sA]RNA54 (*lanes a-i*), and [sA]RNA54-3'f, whose 3'-terminus lies 25-40 nucleotides downstream of the 5' splice site (*lanes j-r*), were incubated in HeLa nuclear extract under splicing conditions for the times shown in minutes. Analysis by gel electrophoresis was as described in Fig. 4.1. Molecular weight markers and indications of reaction products are as in Fig. 4.2.

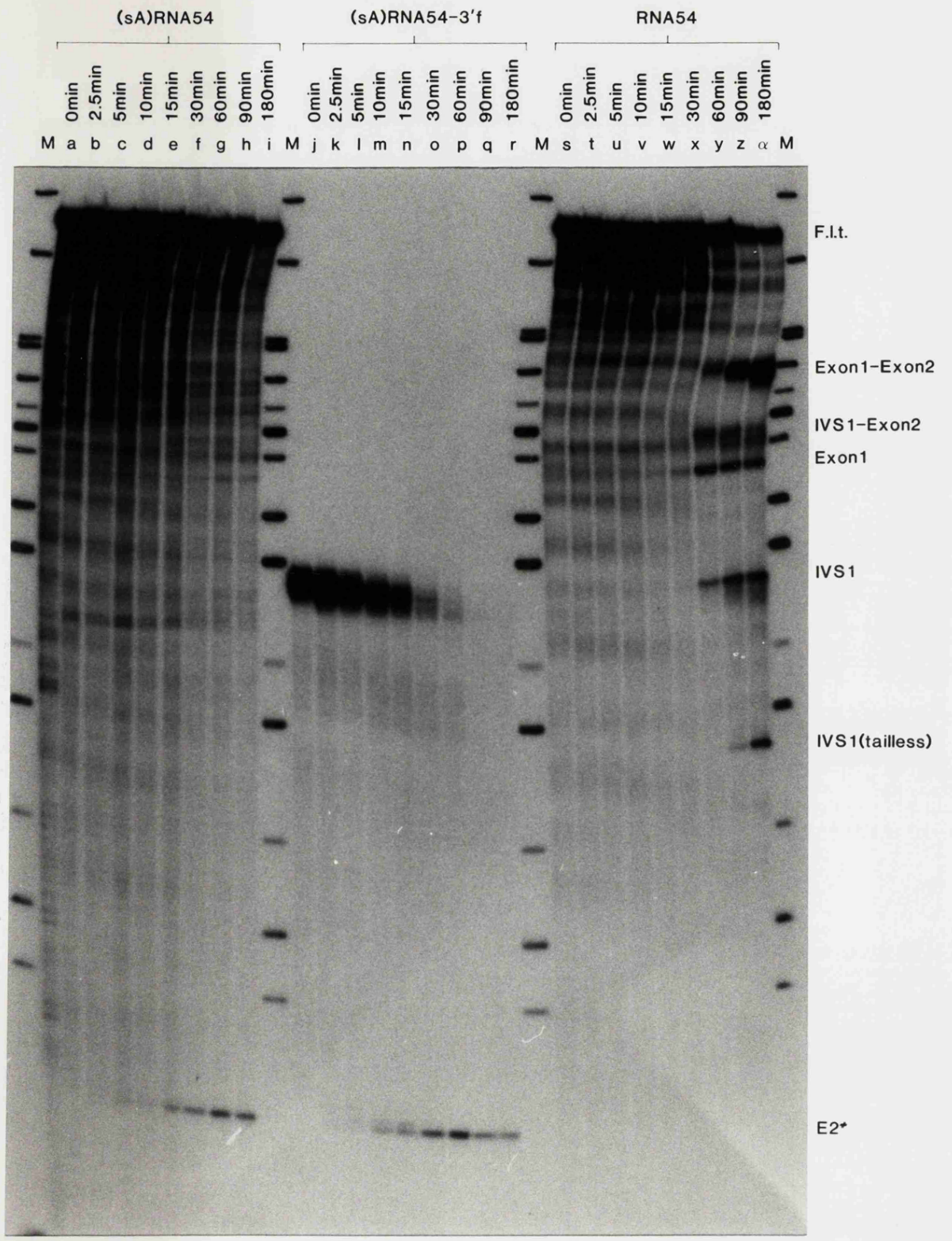
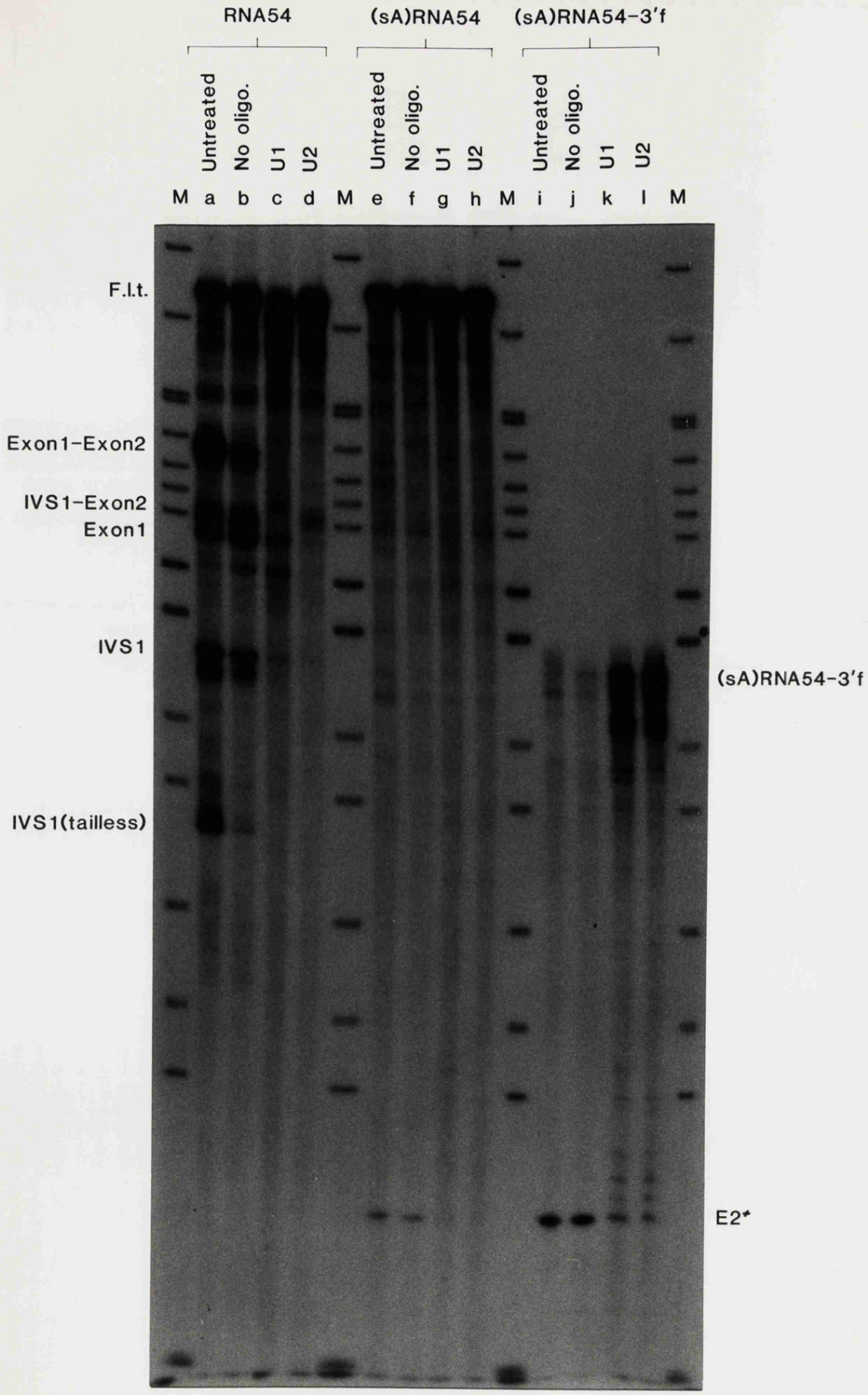

(sA)RNA54
(sA)RNA54-3'f
RNA54
0min
2.5min
5min
10min
15min
30min
60min
90min
180min
M a b c d e f g h i M j k l m n o p q r M s t u v w x y z α M
F.l.t.
Exon1-Exon2
IVS1-Exon2
Exon1
IVS1
IVS1(tailless)
E2*

Figure 5.18 **Dependence of E2* production from the 5'-truncated substrate [sA]RNA54-3'f on the presence of intact snRNAs U1 and U2.**

Nuclear extracts, treated (or otherwise) with RNase H and oligonucleotides U1m (complementary to U1 snRNA), or U2a (complementary to U2 snRNA), were used for splicing reactions with substrates RNA54, [sA]RNA54, and [sA]RNA54-3'f. Analysis was by gel electrophoresis as in Fig. 4.1. Molecular weight markers and designations of molecular species are as in Fig. 4.2. The position of unprocessed [sA]RNA54-3'f is also indicated.


RNA54
(sA)RNA54
(sA)RNA54-3'f
Untreated
No oligo.
U1
U2
Untreated
No oligo.
U1
U2
Untreated
No oligo.
U1
U2
M a b c d M e f g h M i j k l M
F.l.t.
Exon1-Exon2
IVS1-Exon2
Exon1
IVS1
IVS1(tailless)
(sA)RNA54-3'f
E2*

degradation of [sA]RNA54-3'f on incubation under splicing conditions was significantly reduced as a result of cleavage of snRNAs U1 or U2. Hence, the nuclear extract must contain a nuclease activity (or activities) specifically dependent on both U1 and U2 snRNPs being intact.

Fig. 5.19 shows time courses of E2* formation with [sA]RNA54 and [sA]RNA54-3'f substrates when no snRNAs are inactivated, and after U1 or U2 snRNAs have been cleaved using oligonucleotides and RNase H as above. In the absence of snRNA cleavage, the accelerated kinetics of E2* formation with [sA]RNA54-3'f compared to with [sA]RNA54 were as observed previously (Fig. 5.17). Cleavage of U1 snRNA was, as before, incomplete (data not shown), and once again it inhibited, but did not completely abolish, splicing of RNA54. Cleavage of U2 snRNA completely abolished splicing of RNA54. No E2* formation was, however, visible at any time point with [sA]RNA54 as substrate after cleavage of U1 or U2, and with [sA]RNA54-3'f as substrate, although E2* formation was not abolished, levels of E2* were reduced at all times. Once again, the apparent incomplete inhibition of E2* formation with [sA]RNA54-3'f as substrate may be due to the higher overall yield of E2* from [sA]RNA54-3f than from [sA]RNA54, rather than resulting from decreased inhibition of E2* formation. It is probably significant that the kinetics of E2* formation from the [sA]RNA54-3'f substrate were very similar when U1 or U2 snRNAs ahd been cleaved , and when they had not: it was only the yield of E2* that differed.

The time course also shows clearly that cleaving U1 or U2 snRNAs significantly reduced the rate of degradation of [sA]RNA54-3'f (Fig 5.19). With no snRNA inactivation almost no [sA]RNA54-3'f remained after incubation for 1.5 hours in splicing extract. However, when U1 or U2 snRNAs had been cleaved there was still a substantial amount of [sA]RNA54-3'f remaining after incubating for 1.5 hours. Many more bands, correponding to fragments of [sA]RNA intermediate in size between E2* and [sA]RNA54-3'f, were visible at all incubation times after cleavage of U1 or U2 snRNAs than when these snRNAs were left intact. These bands included the ladder, previously observed just above E2*. The intensity of the bands in this ladder varied with time, roughly in parallel with the intensity of the E2* band.

5.1.2.3. Chemical modification/interference study of sequence requirements for E2* formation

An attempt was made to set up a chemical modification/interference assay to investigate the bases required for E2* formation. Assays of this type have previously been employed by Rymond and Rosbash (1988) to study yeast pre-mRNA spliceosome assembly and splicing, and by Conway and

Figure 5.19 **The effect of cleavage of snRNAs U1 and U2 on the time dependence of E2* production from [sA]RNA54 and [sA]RNA54-3'f.**

Nuclear extracts, treated (or otherwise) with RNase H and oligonucleotides U1m (complementary to U1 snRNA), or U2a (complementary to U2 snRNA), were used for splicing reactions with substrates RNA54, [sA]RNA54, and [sA]RNA54-3'f. Reactions were performed for the times indicated in hours, and reactions where the extract was not pre-treated in any way (*lanes a, e* and *r*) were for 3 hours. Analysis was by gel electrophoresis as in Fig. 4.1. Molecular weight markers and designations of molecular species are as in Fig. 4.2.

RNA54, 3h
(sA)RNA54
(sA)RNA54-3'f
No oligo.
U1
U2
Untreated
15min
30min
1.5h
3h
M a b c d M e f g h i j k l m n o p q M r s t u v w x y z α β γ δ M
F.l.t.
Exon1-Exon2
Exon1
IVS1
IVS1(tailless)
E2*

Wickens (1987) to study 3' end cleavage, polyadenylation and polyadenylation complex formation.

The strategy was to introduce a limited number of chemical lesions into [sA]RNA54 and use this modified [sA]RNA as the substrate for E2* formation. The location of modified bases in E2* [sA]RNA formed in this reaction was then to be identified by a primer extension assay which makes use of the fact that reverse transcriptase is blocked by the sort of RNA modifications to be introduced. Modified nucleotides present in reduced amounts in E2* compared to in the substrate [sA]RNA54 must be inhibitory to the formation of E2*, hence interactions with these bases must be important for E2* formation.

Chemical lesions were introduced into RNA54 by treatment with kethoxal (2-keto-3-ethoxybutyraldehyde) or with dimethyl sulphate (DMS). Kethoxal reacts with guanine at N1 and N2, and DMS reacts with adenine at N1, guanosine at N7, and (more slowly) cytosine at N3. All the above modifications terminate reverse transcription, with the exception of the methylated N7 of guanosine. The extent of modification was optimised so as to generate a good ladder of bands corresponding to primer extension products terminated by lesions in the RNA over the region of interest (data not shown). The optimised modification conditions are described in the Methods chapter, and an example of a primer extension assay performed on RNA54 after modification with kethoxal or DMS under optimal conditions is shown in Figure 5.20. The primer used was Oligo.28 (see Chapter 4) which is complementary to the RNA54 3' exon, 11 to 28 nucleotides downstream of the 3' splice junction. It was found, however, that if this level of kethoxal or DMS induced lesions was introduced into RNA54 or [sA]RNA54, and these modified substrates were incubated under splicing conditions, then both splicing of RNA54 and E2* formation from [sA]RNA54 was severely inhibited, to the extent that there was not enough E2* with which to proceed. This strategy has not been further pursued.

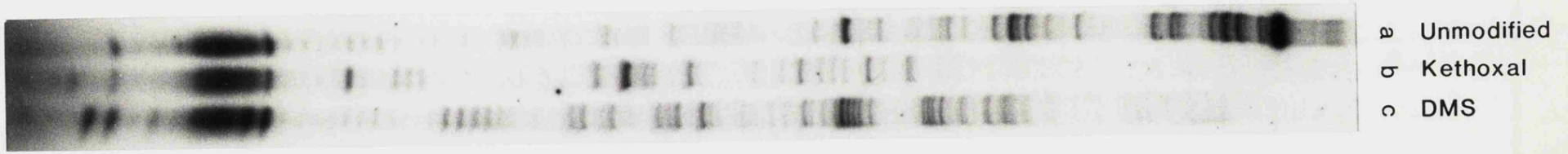
5.1.3. FACTORS REQUIRED FOR E2* FORMATION

Experiments described in Chapter 5 have already shown that E2* formation requires the presence of ATP, Mg^{2+}, and intact snRNAs U1, U2 and U6. All of these factors are also essential for *in vitro* splicing. Further experiments are described below, which attempt to determine whether other factors involved in splicing are also involved in E2* formation.

5.1.3.1. Heat inactivation

Figure 5.20 **Primed synthesis to identify the sites of kethoxal and DMS induced lesions in RNA54.**

^{32}P-labelled oligo.28 (see Chapter 4), which is complementary to the RNA54 exon 2 nucleotides 11 to 28 nucleotides downstream of the 3' splice junction, was used as the primer for DNA synthesis by AMV reverse transcriptase on unmodified RNA54 (*lane a*), kethoxal modified RNA54 (*lane b*), and dimethyl sulphate (DMS) modified RNA54 (*lane c*). Modification conditions were as described in Chapter 2 "Methods". The autoradiogram shows the analysis of these reactions on a 10% polyacrylamide, 7M urea gel.


a Unmodified
b Kethoxal
c DMS

Krainer and Maniatis (1985) identified a splicing factor (denoted SF3) in their HeLa cell nuclear extracts which was preferentially inactivated by relatively mild heat treatment (10 minutes at 45°C). SF3 was only required for the second step of splicing, the mild heat treatment of the extract resulting only in inhibition of step 2, not step 1. As one possible explanation for E2* formation is that it is the result of an aberrant step 2 cleavage event, it was particularly interesting to know if SF3 was also required for E2* formation.

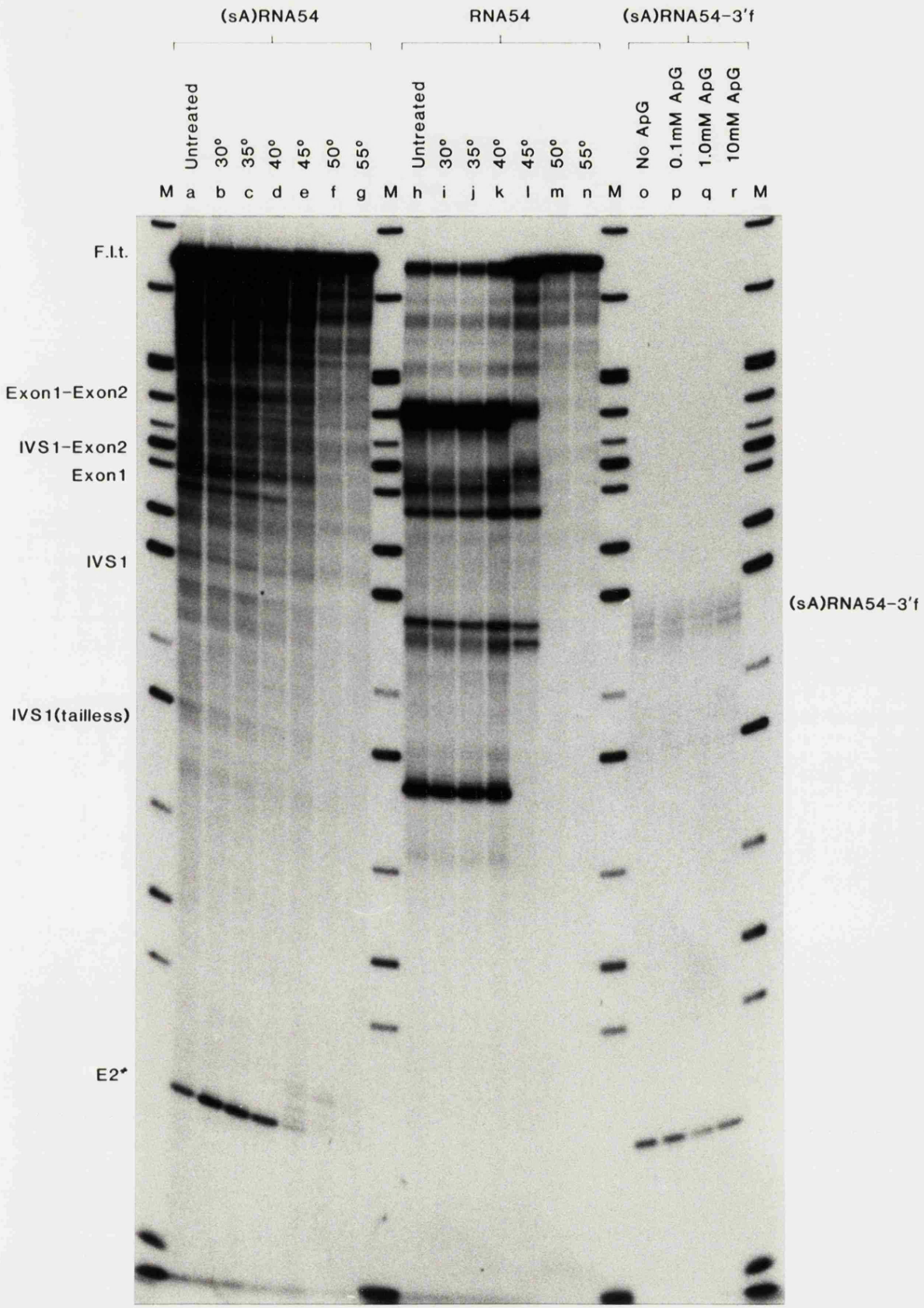
Figure 5.21 shows the effect of pre-incubating nuclear extract for 10 minutes at a range of different temperatures. Both splicing and E2* formation were efficient after pre-incubation of extracts at temperatures up to and including 40°C. After pre-incubation at 45°C, step 2 of splicing was reduced in efficiency, but by no means abolished, presumably due to partial heat inactivation of SF3. The most marked effect of this treatment was, however, a complete inhibition of the 3' exonuclease activity responsible for the removal of the 3' tail of the free lariat intron, as shown by the complete dissapearance of tailless IVS-1. Such treatment did not cause a disappearance of E2* fragments, but merely caused the appearance of a short ladder of E2* bands above the normal position on the gel. The largest of these E2* bands was about 4 nucleotides longer than the predominant E2* band found without heat treatment. Such a pattern is most simply explained by the inhibition of the 3' exonuclease activity noted above. Without heat treatment, as discussed in Chapter 4, the RNA54 transcripts, which extend predominantly 54 nucleotides past the 3' splice junction, have the last 3 or 4 nucleotides removed on incubation in extract by an endogenous 3' exonuclease activity, which is blocked by the presence of a pair of thio-phosphodiester links 50 and 51 nucleotides upstream of the 3' splice junction. When this exonuclease activity is inhibited removal of the 3' terminal 3 or 4 nucleotides is less efficient, giving rise to the pattern of E2* fragments seen in Figure 5.21 after 45°C pre-incubation of the extract.

Both steps of splicing of RNA54 were completely abolished by pre-incubation of the extract at 50°C. In contrast, E2* fragments were still formed from [sA]RNA54 after this pre-treatment of the extract (Fig. 5.21). Hence, SF3 is not required for E2* formation, even though it is required for step 2 of splicing.

If E2* formation is indeed due to an aberrant step 2 cleavage event, where the nucleophile is H_2O (or OH^-) rather than the 3'-OH of exon 1, it was thought possible that the dinucleotide ApG_{OH} (adenylyl-(3'-5')-guanosine), which is identical to the 3'-terminus of rabbit β-globin exon 1, might be able to substitute for the unavailable exon 1 and be ligated to the 5' end of E2*. The dinucleotide CpU_{OH} (cytidyl-(3'-5')-uridine), when incubated with

Figure 5.21. **Heat sensitivity of splicing and E2* formation: attempted *trans* splicing using ApG as the 5'-exon analogue.**

HeLa nuclear extract was either used directly for *in vitro* splicing (*lanes a*, *h*, and *o-r*), or was first incubated for 10 minutes at the temperatures indicated in °C (*lanes b-g*, and *i-n*). 1.5 hour splicing reactions containing [sA]RNA54 are shown in *lanes a-g*, and 3 hour splicing reactions containing RNA54 are shown in *lanes h-n*. *Lanes o-r* are 1.5 hour splicing reactions containing [sA]RNA54-3'f, in the presence of the concentrations of ApG (adenylyl-(3'-5')-guanosine) shown. Analysis was by gel electrophoresis as in Fig. 4.1. Molecular weight markers and designations of molecular species are as in Fig. 4.2. The position of undigested [sA]RNA54-3'f is also indicated.


(sA)RNA54
RNA54
(sA)RNA54-3'f
Untreated
30°
35°
40°
45°
50°
55°
Untreated
30°
35°
40°
45°
50°
55°
No ApG
0.1mM ApG
1.0mM ApG
10mM ApG
M a b c d e f g M h i j k l m n M o p q r M
F.l.t.
Exon1-Exon2
IVS1-Exon2
Exon1
IVS1
IVS1(tailless)
E2*
(sA)RNA54-3'f

self-splicing *Tetrahymena* pre-rRNA in the absence of GTP, has been shown to function as a 5' exon (Inoue *et al.*, 1985). CpU_{OH} resembles the CUCUCU sequence at the 3' end of the 5' exon, and cleaves the precursor exactly at the 3' splice site, becoming covalently ligated to the 3' exon. However, if a substrate lacking exon 1 (to increase the chance that the normal exon 1 binding site was free) i.e. [sA]RNA54-3'f, was incubated under splicing conditions in the presence of the highest attainable concentration of $ApG_{O\,H}$ (10mM), there was no evidence of ligation of this dinucleotide onto the 5'-terminus of E2* (Fig. 5.21).

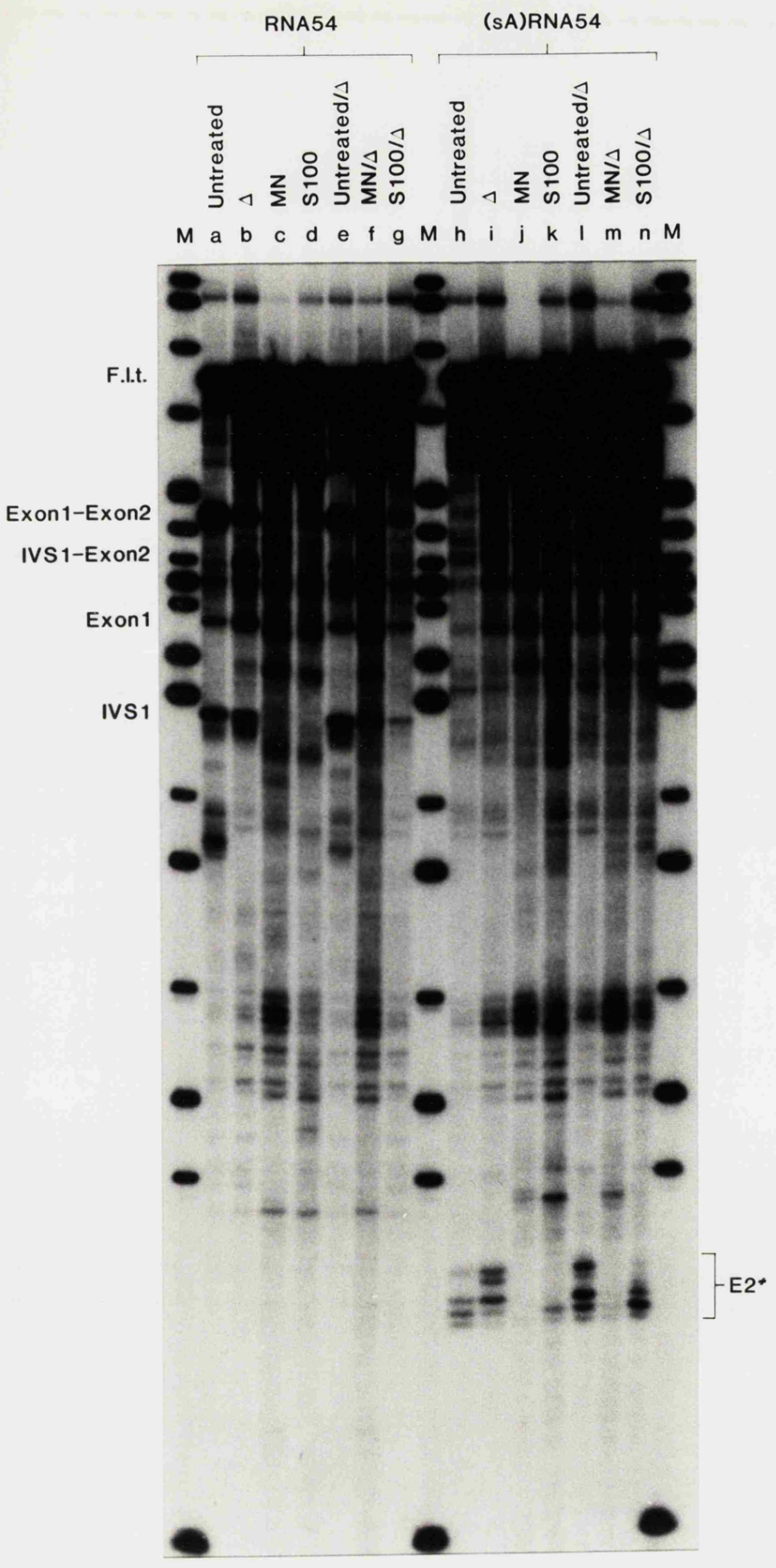
5.1.3.2. The effect of micrococcal nuclease treatment on E2* formation, and the ability of cytoplasmic S100 fractions to support E2* formation

Micrococcal nuclease treatment of nuclear extract completely inhibited both steps of splicing of RNA54 (Fig. 5.22), in accordance with previous results (Krainer and Maniatis, 1985). The essential RNA components inactivated by micrococcal nuclease activity presumably include the snRNAs U1, U2, U4 and U6, each of which must be intact for splicing to proceed, as shown by oligonucleotide directed RNase cleavage of these individual species (discussed in Chapter 4). Since E2* formation was also inhibited by oligonucleotide directed RNase H cleavage of snRNAs U1, U2 and U6, it was expected that micrococcal nuclease treatment of extract would also inhibit E2* formation, and Figure 5.22 shows that micrococcal nuclease treatment does indeed completely prevent E2* production from [sA]RNA54. Pre-treatment of the extract by incubating for 10 minutes at 45°C once again resulted only in partial inhibition of step 2 of splicing with RNA54 as substrate. This treatment failed to inhibit E2* formation, as before, and merely resulted in a greater predominance of longer E2* species, due to the inhibition of inhibition of the endogenous 3' exonuclease activity, as discussed above. A HeLa cell S100 fraction was found to be inactive in both steps of splicing (Fig. 5.22) as described previously (Krainer and Maniatis, 1985). Interestingly, however, the S100 fraction catalysed the formation of some fragments of E2* mobility with [sA]RNA54 as substrate, although the yield was lower than with nuclear extract. If these bands really do correspond to E2*, then E2* formation must not require all the components necessary for splicing.

Mixing of micrococcal nuclease treated and heat treated extract did not result in an increase in the efficiency of either step of splicing compared to that in heat treated extract alone; the efficiency was, if anything, slightly reduced. Similarly, the efficiency of E2* formation, when micrococcal nuclease treated and heat treated extract were mixed, was lower than in heat

Figure 5.22. **The ability of micrococcal nuclease treated HeLa nuclear extract and cytoplasmic S100 fraction to support E2* formation.**

RNA54 (*lanes a-g*) or [sA]RNA54 (*lanes h-n*) were spliced *in vitro* for 3 hours as described in Chapter 2 "Methods" but using the normal volume of the following extracts: untreated HeLa nuclear extract (*lanes a* and *h*), HeLa nuclear extract pre-treated by incubation at 45°C for 10 minutes (heat treated extract)(*lanes b* and *i*), HeLa nuclear extract treated with micrococcal nuclease as in Chapter 2 "Methods" (*lanes c* and *j*), HeLa S100 extract (*lanes d* and *k*), a 1:1 mixture of untreated nuclear extract and heat treated nuclear extract (*lanes e* and *l*), a 1:1 mixture of micrococcal nuclease treated extract and heat treated nuclear extract(*lanes f* and *m*), and a 1:1 mixture of S100 and heat treated nuclear extract (*lanes g* and *n*). Analysis was by gel electrophoresis as in Fig. 4.1. Molecular weight markers and designations of molecular species are as in Fig. 4.1.


RNA54
(sA)RNA54
Untreated
Δ
MN
S100
Untreated/Δ
MN/Δ
S100/Δ
Untreated
Δ
MN
S100
Untreated/Δ
MN/Δ
S100/Δ
M a b c d e f g M h i j k l m n M
F.l.t.
Exon1–Exon2
IVS1–Exon2
Exon1
IVS1
E2*

treated extract alone. This inhibitory effect was probably due to the buffer in which the micrococcal nuclease was dissolved (50% glycerol, 50mM Glycine/KOH (pH 9.2), 5mM $CaCl_2$), which was later found to inhibit splicing, though incompletely at the concentrations employed in this study. Mixing of S100 and heat-treated extract did not result in any significant increase in the efficiency of step 2 of splicing as compared to that in heat treated nuclear extract alone, but inhibition of step 2 by heat treatment was not very effective anyway, neither did it increase the overall efficiency of E2* formation from [sA]RNA54 compared to in heat treated extract alone. The main effect of mixing S100 and heat treated nuclear extract on E2* formation was to decrease the average length of the E2* fragments compared to those produced by processing in heat treated nuclease alone. This is simply explained by the previously discussed 3' exonuclease activity being active in the S100 fraction (as evidenced by the short length of E2* fragments produced using this extract alone), but inactivated by heat treatment.

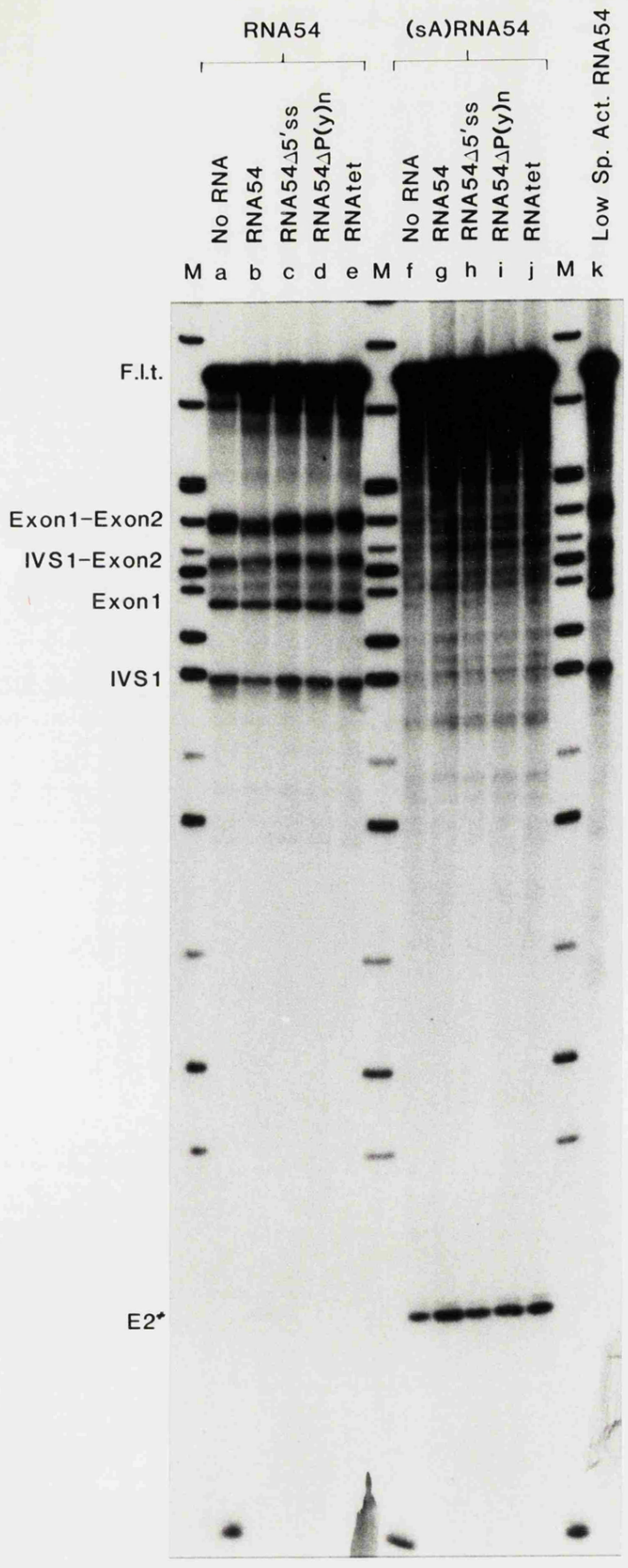
5.1.3.3. Competition assays

A competition assay which involves pre-incubating nuclear extract with an excess of unlabelled RNA (containing selected sequence motifs) before adding the radio-labelled substrate RNA, has previously been employed to investigate protection of RNA from digestion by RNase A (Krämer, 1987b), and to investigate the requirement for *trans*-acting factors at early stages in yeast splicing (Legrain *et al.*, 1988). In a further attempt to define the factors involved in E2* formation, and their relationship to normal splicing, 10fmole unlabelled RNA was pre-incubated for 20 minutes at 30°C in nuclear extract under *in vitro* splicing conditions before the addition of 0.5fmole of radiolabelled RNA54 or [sA]RNA54 (Fig. 5.23). The unlabelled, competing RNAs used were RNA54, RNA54Δ5'ss, RNA54Δ(Py)n and RNAtet (a 468 nucleotide long T7 RNA polymerase transcript from *Hind*III cut **mICE10[tet]** R.F., which consists of the *Tetrahymena* 26S rRNA intron and some flanking exon sequence). None of these competing RNAs, however, gave rise to any inhibition of splicing of RNA54 or E2* formation from [sA]RNA54 (Fig. 5.23). This was due to the quantity of unlabelled competitor RNA used not being saturating, since 10fmole of radiolabelled RNA54 (low specific activity) was found still to be efficiently spliced in the volume of extract used in the competition assays (Fig. 5.23).

5.2. DISCUSSION

Figure 5.23. **Competition of unlabelled RNA, containing selected sequence motifs, for factors required for splicing and E2* formation.**

HeLa nuclear extract was pre-incubated for 20 minutes at 30°C under *in vitro* splicing conditions with either no RNA (*lanes a* and *f*), 10fmole RNA54 (*lanes b* and *g*), 10fmole RNA54Δ5'ss (which has the 5' splice site deleted) (*lanes c* and *h*), 10fmole RNA54Δ(Py)n (which has the polypyrimidine tract deleted) (*lanes d* and *i*), or 10fmole RNAtet (which should contain no pre-mRNA splicing signals) (*lanes e* and *j*). After treatment these extracts were used to splice 0.5fmole ^{32}P-labelled RNA54 (*lanes a-e*) or 0.5fmole [sA]RNA54 (*lanes f-j*) for 1.5 hours. *Lane k* shows the result of splicing 10fmole of low specific activity ^{32}P-labelled RNA54 in nuclear extract that had undergone no pre-treatment. Analysis was by gel electrophoresis as in Fig. 4.1. Molecular weight markers and designations of molecular species are as in Fig. 4.1.


RNA54
(sA)RNA54
No RNA
RNA54
RNA54Δ5'ss
RNA54ΔP(y)n
RNAtet
No RNA
RNA54
RNA54Δ5'ss
RNA54ΔP(y)n
RNAtet
Low Sp. Act. RNA54
M a b c d e M f g h i j M k
F.l.t.
Exon1–Exon2
IVS1–Exon2
Exon1
IVS1
E2*

5.2.1. COMPLEX FORMATION ON PHOSPHOROTHIATE RNA SUBSTRATES

Formation of E2* from [sA]RNA54 substrates, as has been discussed in Chapter 5, would seem to be related to the process of splicing, based on the fact that both E2* formation and splicing are dependent on ATP, Mg^{2+}, and intact U1, U2 and U6 snRNPs, and that E2* is formed with the same kinetics as the intermediates of a normal splicing reaction. This data seemed to point strongly to E2* formation being catalysed by a high molecular weight complex identical, or akin to the "spliceosome" responsible for the catalysis of splicing (Frendewey and Keller, 1985; Grabowski *et al.*, 1985). A variety of techniques were, therefore, employed to directly investigate the question of what sort of complex (if any) form on phosphorothioate RNA substrates under *in vitro* splicing conditions.

Using native gel electrophoresis, splicing and pre-splicing complexes were found to form on RNA54 substrates which corresponded to those found in previous studies using different substrates (Konarska and Sharp, 1986; Lamond *et al.*, 1987). In contrast, no such low mobility complexes were found when [sA]RNA54 was incubated in nuclear extract under *in vitro* splicing conditions. This could indicate that no large complex is formed on [sA]RNA54, and hence that a complex related to the spliceosome or a pre-splicing complex is not involved in E2* formation. Alternatively, it could be that a large complex is assembled on [sA]RNA54 and is responsible for E2* formation, but is peculiarly unstable and dissociates from the [sA]RNA during electrophoresis.

The only complexes to form on [sA]RNA were the high mobility non-ATP dependent "sigma complexes" which form at time zero even on normal RNA transcripts lacking splice site signals (Konarska and Sharp, 1986). These complexes are thought to be due to non-specific binding of components in the extract to RNA, probably hnRNP proteins (Konarska and Sharp, 1986). The mobility of Sigma complexes was lower with [sA]RNA54 than with RNA54, whether or not the samples were heparin treated before electrophoresis, and independent of whether incubation in the nuclear extract was in the presence or absence of ATP. This difference in mobility could reflect a genuine difference in the composition of the Sigma complexes on these two substrates, or could be due to differing stabilities of the sigma complexes. The nature of the differences between the sigma complexes formed on RNA54 and [sA]RNA54 is not known, but is probably

not directly related to E2* formation since E2* formation is ATP dependent, whereas the difference in sigma complex mobility on RNA54 and [sA]RNA54 substrates was not.

The appearance of an ATP-independent complex (complex ϕ) of mobility intermediate between that of splicing complexes A and B when either RNA54 or [sA]RNA54 was used is of some interest, since no such complex has previously been reported using native gel electrophoresis as the assay for splicing complex assembly. This complex may correspond to the Sm-dependent 35S complex identified on sucrose gradients (Frendewey and Keller, 1985) which was found to assemble in the absence of ATP, and which was thought to be an intermediate in the pathway of 50S spliceosome assembly, and to the [^{32}P]30S ATP-independent complex identified in yeast systems by glycerol gradient centrifugation (Chang *et al.*, 1988). The composition of complex ϕ is unknown.

Although no complexes other than the sigma complexes were detected by native gel electrophoresis of [sA]RNA54, this was not the case with all phosphorothioate substituted RNAs tested. None of the phosphorothioate RNA analogues tested gave rise to an equivalent abundance of low mobility complexes as did RNA54. Indeed, no mature splicing complex B (Konarska and Sharp, 1986) was visible after incubation of any of any phosphorothioate RNA analogue under splicing conditions. Hence , the absence, or ineficiency, of splicing seen with phosphorothioate RNAs is, at least in part, due to ineficient splicing complex assembly. However, [sG]RNA54 gave rise to a complex which may correspond to the pre-splicing complex A (Konarska and Sharp, 1986), but no mature splicing complex B. The absence of complex B is not unnexpected, since [sG]RNA54 undergoes neither step of splicing when incubated in *in vitro* splicing conditions (see Chapter 5). [sG]RNA54 also gave rise to a complex (complex X), of unknown composition, which migrated just above the sigma complexes. This complex could be a normal intermediate in the pathway of splicing complex assembly, which is never seen with normal RNA substrates, but it could also be a completely abberant complex, either related or unrelated to splicing. Thio-phosphodiester linkages in [sG]RNA seem, therefore, to have pleiotropic effects, affecting more than one stage of splicing. [sC]RNA54, a transcript which is spliced ineficiently *in vitro* (see Chapter 5), also gave rise to a complex which may be equivalent to pre-splicing complex A, but no visible complex B. [sA]RNA54A, which differs from [sA]RNA54 only in that the 3' splice site sequence has been changed from AG/G to AG/A, in contrast to [sA]RNA54, formed a significant amount of pre-splicing complex A, but there was no mature spliceosome band B visible. The increase in the level of complex A with the mutant, compared to the wild type transcript correlates

quite well with the finding that [sA]RNA54A splices consistently more efficiently than [sA]RNA54. As [sA]RNA54A is not processed to give E2*, it would seem that the assembly of splicing complexes on [sA]RNA more closely parallels the efficiency of splicing than E2* formation.

Immunoprecipitation of whole *in vitro* splicing reactions with anti-(Sm)RNP antibody revealed that precursor [sA]RNA54 was, like precursor RNA54 and all the products and intermediates of RNA54 splicing, associated with a species containing an Sm epitope i.e. almost certainly a U snRNP. In contrast, E2* was not precipitated by anti-(Sm)RNP antibody and hence does not seem to be stably associated with an Sm snRNP. E2* not being stably associated with a snRNP particle would correlate with the low sedimentation coefficient of E2* containing particles on sucrose gradients.

A combined RNase T1 digestion and immunoprecipitation assay, as described by Black *et al.* (1985), was used to elucidate the location and requirements for interactions between snRNPs and [sA]RNA54, and to compare these interactions with those between snRNPs and RNA54. Although full interpretation of the results of these experiments awaits the characterisation of the precipitated fragments, certain conclusions can be drawn even in the absence of this data.

The pattern of protected fragments of RNA54 and [sA]RNA54 precipitated by anti-(U1)RNP, anti-(Sm)RNP and anti-(TMG) antibodies was essentially identical at time zero. This indicates that in splicing extract, at 0°C, the stable interactions formed between U1 snRNP and the Sm snRNPs as a whole are identical on both RNA54 and [sA]RNA54 substrates. At zero time, anti-(U2)RNP antibody also immunoprecipitated the same pattern of fragments from [sA]RNA54 as the other antibodies, with the exception that the level of the putative 5' splice site protected fragment (5'ss?) was much lower. However, at zero time, with RNA54 the main fragment precipitated by anti-(U2)RNP antibody was a relatively low level of the putative 5' splice site protected fragment (5'ss?), fragments of the same mobility as those precipitated by anti-(U2)RNP antiserum at time zero did, however, became visible on longer exposures. After incubation for 60 minutes at 30°C under splicing conditions, the pattern of protected [sA]RNA54 fragments immunoprecipitated by all four antibodies was essentially identical to that found at time zero, apart from the appearance of a limited number of new minor bands and in most cases a general increase in the yield of immunoprecipitated fragments. In contrast, when RNA54 was incubated at 30°C for 60 minutes, the pattern of protected bands immunoprecipitated by anti-(U2), anti-(Sm)RNP and anti-(TMG) antibodies was distinctly different to that seen at time zero. The pattern of protected bands immunoprecipitated by anti-(U1)RNP antibody after incubation of RNA54 at 30°C for 60 minutes

was little changed from the pattern seen at time zero. The protected fragments immunoprecipitated increased in both number and average size. This is consistent with previous studies using this technique on human β-globin IVS-1 RNA (Black *et al.*, 1985; Chabot and Steitz, 1987; Bindereif and Green, 1987), where the increased protection of the RNA on incubation has been equated with the assembly of splicing complexes. It would seem, therefore, that incubation at 30°C does not give rise to stable splicing complexes on [sA]RNA54 to any significant extent. Some of the minor bands precipitated by anti-(U2)RNP antibody after 60 minutes are of the same mobility as the major bands precipitated by the same antibody after a 60 minute incubation of RNA54, and this may imply some limited splicing complex assembly, although not necessarily assembly of the mature spliceosome.

The pattern of protected fragments immunoprecipitated from RNA54 (rabbit β-globin IVS-1) in this study show some interesting differences between those found in previous studies on human β-globin IVS-1 RNA in splicing extract, (Black *et al.*, 1985; Chabot *et al.*, 1985; Chabot and Steitz, 1987). With the human RNA, at time zero, anti-(U2)RNP antibody was not found to precipitate any RNA fragments (Black *et al.*, 1985; Chabot *et al.*, 1985). In this study, anti-(U2)RNP antibody did precipitate fragments of RNA54 at time zero. The most prominent band probably corresponds to the protected 5' splice site. This is interesting, since Black *et al.* (1985) found anti-(U2)RNP antibodies to precipitate uncleaved precursor RNA at zero time and yet this interaction was obviously unstable to RNase T1 digestion. If the main fragment precipitated by anti-(U2)RNP antibody at time zero does indeed include the 5' splice site, then it would seem most likely that U1 snRNP is directly responsible for protection of this fragment, and that at least a fraction of the U1 snRNP bound to the 5' splice site interacts with U2 snRNP even at time zero. This time zero interaction of U1 and U2 snRNPs (if this is indeed what it is) must be more stable in this study than in the earlier work. It could be argued that the precipitation of an apparently identical set of fragments by both anti-(U1)RNP and anti-(U2)RNP antisera at time zero may be due to contamination of one of the antisera with antibodies with specificity for snRNP proteins found on the other snRNP particle. However, both of these antisera, which were gifts from Prof. J.A. Steitz, have been extensively characterised in her laboratory, and the anti-(U1)RNP antiserum (Ag) was used in previous experiments by this group (Black *et al.*, 1985; Chabot *et al.*, 1985; Chabot and Steitz, 1987b). In addition, although the pattern of bands precipitated by anti-(U1)RNP and anti-(U2)RNP antibodies was identical at time zero, the ratio of intensities of the bands differed. The 5'ss? band was, relative to the other bands, much more intense when

immunoprecipitation was performed with anti-(U1)RNP antiserum than with anti-(U2)RNP antiserum. If U2 snRNP had no stable interaction at time zero with the substrate RNA, as found by Black *et al.* (1985), and the bands precipitated by the anti-(U2)RNP antiserum in this study were merely due to contaminating antibodies against proteins in the U1 snRNP, then the ratio of bands would be expected to be the same as that found when immunoprecipitation was performed with anti-(U1)RNP antiserum. In this study the pattern of protected fragments precipitated by anti-(U1)RNP, anti-(Sm)RNP and anti-(TMG) antibodies was effectively identical at time zero, consisting of one major band, probably corresponding to the protected 5' splice site and several minor bands. In contrast, with the human RNA at time zero, the pattern of fragments immunoprecipitated by anti-(Sm)RNP and anti-(TMG) antibodies was different from that precipitated by anti-(U1)RNP antibody (Chabot *et al.*, 1985). Anti-(U1)RNP antibody only precipitated a single fragment corresponding to the protected 5' splice site, whereas the other two antibodies also precipitated fragments from the 3' splice site region of the intron. Protection of the 3' splice site fragments must therefore be due to an snRNP other than U1 or U2, since antisera against either of these snRNPs failed to precipitate these fragments. U5 snRNP was suggested as the most likely candidate. It is tempting to speculate that the minor bands precipitated in this study at time zero correspond to protected 3' splice site fragments, and that in this study U1 snRNP was stably interacting with the snRNP which protects the 3' splice site region at time zero, and hence anti-(U1)RNP antibody caused co-precipitation of these fragments along with that from the 5' splice site, whereas in the work of Chabot *et al.* (1985) this interaction was unstable. In this study, the pattern of protected RNA54 fragments immunoprecipitated by anti-(U1)RNP antibody was largely identical at time zero and after incubation under splicing conditions for 60 minutes. The pattern of fragments precipitated by anti-(U1)RNP antibody after 60 minutes incubation was, furthermore, substantially different from that immunoprecipitated by anti-(U2)RNP, anti-(Sm)RNP, or anti-(TMG) antibodies. This would seem to suggest that U1 snRNP is not incorporated into the spliceosome, in accordance with the model proposed by Konarska and Sharp (1986) in which the U1 snRNP functions at an early stage in splicing to recognise the 5' splice site, but is not incorporated into the spliceosome. Alternatively, this effect may be due to sequestration of (U1)RNP determinants on formation of larger splicing complexes as proposed by Black *et al.* (1985) such that spliceosomes are not immunoprecipitated by anti-(U1)RNP antibody. This data contrasts with that found using human β-globin IVS-1 RNA (Black *et al.*, 1985; Chabot and Steitz, 1987) where it was claimed that at 60 minutes all of the fragments

immunoprecipitated by anti-(U2)RNP antibody were also immunoprecipited by anti-(U1)RNP antibody, merely in lower amounts. Furthermore, since some of the fragments precipitated by anti-(U1)RNP antibody contained branched RNA (Chabot and Steitz, 1987), U1 snRNP would seem to be an integral component of the spliceosome. U1 snRNP has also been shown to be a component of functional spliceosomes purified by gel filtration (Reed *et al.*, 1988).

5.2.2. PRE-MESSENGER RNA SEQUENCES NECESSARY FOR E2* FORMATION

RNA54Δ(Py)n, which lacks the polypyrimidine tract sequences found from nucleotides -30 to -5 relative to the 3' splice site in, was found to undergo step 1 of splicing, *in vitro*, with a drastically reduced efficiency compared to the wild type RNA54, and exon ligation was probably completely abolished. Ruskin and Green (1985b) found a similar effect on *in vitro* splicing when the polypyrimidine tract was deleted from human β-globin IVS-1 RNA. Frendewey and Keller (1985) also found that the polypyrimidine tract of an adenovirus 2 major late pre-mRNA intron was required for 50S spliceosome complexes to be seen on sucrose gradients. [sA]RNA54Δ(Py)n was, however, processed to give E2* with a similar efficiency to [sA]RNA54. Hence, interactions of factors with the polypyrimidine tract would not seem to be necessary for E2* formation. Since the polypyrimidine tract is apparently required for functional spliceosome assembly this is strong evidence that E2* formation is not catalysed by a large spliceosome complex. It would also seem that the protein designated IBP (Tazi *et al.*, 1986) or 70kd protein (Gerke and Steitz, 1986), and the seperate protein factor designated U2AF (Ruskin *et al.*, 1988) are unlikely to be the 3' splice site binding factors responsible for either cleaving the [sA]RNA or blocking a 5' exonuclease activity so as to generate E2*, since the efficient binding of both these factors is dependent upon a polypyrimidine tract. U2AF has been found to be required for U2 snRNP binding to the branch point of mammalian introns, and for spliceosome assembly (Ruskin *et al.*, 1988). Since E2* formation occurs from [sA]RNA54Δ(Py)n which lacks a polypyrimidine tract, this would seem to further point to the fact that the assembly of a spliceosome complex on the [sA]RNA is not required for spliceosome formation.

Furthermore, the truncated substrate [sA]RNA54-3'f, which lacks all of exon 1 and the 5' most 25 to 40 nucleotides of IVS-1, was efficiently processed *in vitro* to yield E2*, despite the fact that it lacked a functional 5' splice site.

Hence, it would seem that interactions with 5' splice site sequences are unnecessary for E2* formation. This does not, however, imply that a splicing complex of some description is not involved in E2* formation. A 35S complex was found to form on substrate RNA from which the 5' splice site had been deleted (Frendewey and Keller, 1985), but this complex was ATP-independent and therefore it would seem unlikely that this complex is directly involved in E2* formation, which is ATP-dependent. Three ATP-dependent complexes (α", β" and γ") with similar mobilities to the three splicing complexes (α, β and γ) on native ploy acrylamide gels, have been found to form on a deleted substrate rabbit β-globin IVS-2 RNA which lacks the 5' splice site region, including the two major cryptic sites (Lamond *et al.*, 1987). Adenovirus 2 RNA substrates from which the 5' splice site has been deleted have been reported to be incorporated into a complex with a mobility similar to that of complex A on native gels (Konarska and Sharp, 1986), and also to run as a single, unique complex, when assayed by the slightly different native gel electrophoresis system of Zillman *et al.* (1988). It would seem, therefore, that large, splicing related complexes can form on pre-mRNAs from which the 5' splice site sequences have been deleted.

The appearance of greater yields of E2* at early times with [sA]RNA54-3'f substrates than with [sA]RNA54 substrates may indicate that a 5' exonuclease activity, blocked by some factor bound to the 3' splice junction region, is responsible for E2* formation, since in [sA]RNA54-3'f there is neither a cap structure nor such a long sequence of RNA to be removed in order to generate E2*. The lag seen before the appearance of E2* may, therefore, merely represent the time taken for the nuclease to digest from the 5' terminus of the transcript to the 3' splice site, and the appearance of E2* from [sA]RNA54 substrates simultaneously with the intermediates of the splicing reaction may be purely coincidental. An alternative explanation is that a factor(s) may bind to the 3' splice site region, and catalyse an abberant endonucleolytic cleavage near to the 3' splice site, so as to generate E2*. This factor(s) may cleave the transcript with different kinetics dependent on whether it is, or is not, interacting with a factor bound to the 5' splice site. It was hoped to distinguish between these models by examining whether E2* formation also occured with accelerated kinetics when the substrate was [sA]RNA54Δ5'ss, which differs from [sA]RNA54 only in possessing an 8 nucleotide deletion of the 5' splice site. If the 5' exonuclease model is correct the kinetics of E2* formation from [sA]RNA54Δ5'ss should match those of E2* formation from [sA]RNA54 since both these substrates possess 5'-terminal caps and are very nearly the same length (and therefore would require removal of an almost identical number of nucleotides by the 5' exonuclease to generate E2*). In contrast, if

the endonuclease model is correct the kinetics of E2* production from the two substrates should differ. However, if the kinetics of E2* production from [sA]RNA54Δ5'ss and [sA]RNA54 were found to differ, this would not prove the endonuclease model, since it is possible that the difference could be due to blocking of the 5' exonuclease by factors bound to the 5' splice site of [sA]RNA54. However, this strategy did not prove possible, since RNA54Δ5'ss was found to splice efficiently by using a cryptic 5' splice site.

Since E2* formation could occur from a substrate lacking a functional 5' splice site sequence ([sA]RNA54-3'f), and U1 snRNP binds to the 5' splice site at least in part by base pairing (Zhuang and Weiner, 1986), it was slightly surprising that cleavage of the 5' terminus of U1 snRNA inhibited E2* formation from [sA]RNA substrates either containing or lacking 5' splice sites. U1 snRNP must, therefore, play a role in some process essential for E2* production which does not depend on interaction with a 5' splice site sequence, but does depend on the 5' terminus of U1 snRNA being intact.

[sA]RNA54-3'f was degraded under splicing conditions much more rapidly than [sA]RNA54. At least some of this increased sensitivity to nuclease activity is probably due to the absence of a cap structure on [sA]RNA54-3'f. It may be, however, that "incomplete" pre-mRNAs such as [sA]RNA54-3'f are specifically targeted for degradation, or conversely that "complete" pre-mRNAs are particularly protected from degradation, perhaps by the splicing machinery. It is very noticeable that the background smear of bands found at time zero in an *in vitro* splicing experiment due to degradation of the substrate RNA tends to disappear with increasing time of incubation under *in vitro* splicing conditions (see, for example Fig. 5.17). However, the most significant aspect of the rapid degradation of [sA]RNA54-3'f was that the rate of degradation could be significantly reduced by oligonucleotide directed RNase H cleavage of U1 or U2 snRNAs. This must indicate that there is within HeLa cell nuclear extracts a ribonuclease activity that is dependent on both U1 and U2 snRNAs being intact. This presumably implies that U1 and U2 snRNPs must both be present in a nuclease complex. Mattaj *et al.* (1986) reported that a small fraction of the U2 snRNPs in *Xenopus laevis* oocytes exists in a complex with U1 snRNPs, as evidenced by the precipitability of either U1 or U2 snRNA by monospecific antibodies directed against proteins specific for the other U snRNP. They argued that, because of their assay conditions, it was more likely that they were observing a direct U1-U2 snRNP interaction than immunoprecipitation of splicing complexes. The experiments with [sA]RNA54-3'f, described above, add weight to the hypothesis that E2* formation is catalysed by a 5' exonuclease activity which is blocked by a component(s) bound in the region of the 3' splice junction. A nuclease

which requires intact U1 and U2 snRNAs could explain the dependence of E2* formation on intact U1 and U2 snRNAs, and would overcome the otherwise difficult problem of how cleaving U1 or U2 snRNAs can inhibit E2* formation if no splicing complex is forming on the [sA]RNA. If this explanation for E2* formation is correct, then one must also suppose that U6 snRNA (probably in the form of the U4/U6 snRNP) must also be a component of the putative snRNP containing nuclease, since cleaving U6 snRNA also resulted in the inhibition of E2* formation (see Chapter 5). This would make the snRNP composition of the putative nuclease remarkably similar to that of a spliceosome, leading to the speculation that a spliceosome complex, or something closely related to it, may have a role in the degradation of incomplete pre-mRNAs in the nucleus.

The only mutation known to prevent E2* production is thus a mutation of the G immediately downstream of the 3' splice junction to an A, which prevents E2* formation from the substrates [sA]RNA54A and [sC]RNA54A (see Chapter 5), and this is consistent with a factor which binds to the 3' splice junction being involved in E2* formation. Strenuous attempts have been made to mutate the 3' splice site AG dinucleotide of **mICE10[IVS-1]** to GG and to AC in order to investigate further the sequence requirements for E2* formation in the 3' splice site region. Unfortunately, for unknown reasons neither mutation has ever been recovered.

5.2.3. FACTORS REQUIRED FOR E2* FORMATION

In Chapter 4, experiments are described which show E2* formation, like splicing, to require the presence of ATP, Mg^{2+}, and intact snRNAs U1, U2 and U6. However, experiments described in this chapter indicated that not all the components necessary for splicing are also required for E2* formation.

Micrococcal nuclease treatment of nuclear extracts abolished E2* formation, as expected, given that specific cleavage of U1, U2 and U6 snRNAs has already been shown to inhibit E2* formation (see Chapter 4). This merely goes to reinforce the important role of snRNAs in E2* formation.

The splicing factor designated SF3 by Krainer and Maniatis (1985) is preferentially inactivated by relatively mild heat treatment of splicing extracts, and is only required for the second step of splicing. Mild heat treatment of nuclear extract inhibits only step 2 of splicing, whilst allowing step 1 to proceed. Since one possible explanation for E2* formation is that it results from an aberrant step 2 cleavage event, it was of interest to know if SF3 was also required for E2* formation. In fact, SF3 was found not to be required for E2* formation: E2* formation could still be seen, even after a

heat treatment which completely inactivated both steps of splicing (50°C, 10 minutes). This also implies that there must be a factor essential for step 1 of splicing, but not for E2* formation, which is inactivated by the 50°C heat treatment. This result does not necessarily preclude an aberrant step 2 cleavage event from being the cause of E2* formation, since it is not known at what stage after step 1 SF3 has its effect, or what its role in the splicing process is, and this role may be unnecessary for E2* formation, especially since the kinetics of E2* production from [sA]RNA50 and [sA]RNA54 match those of step 1 of splicing better than those of step step 2. The most marked effect of mild heat treatment (45°C, 10 minutes) was to inhibit the previously reported endogenous nuclease (probably 3' exonuclease) activity responsible for removal of the 3' linear tail of lariat introns (Ruskin and Green, 1985c). This same mild heat treatment of nuclear extract also resulted in an average increase in the length of E2* fragments produced using these extracts. This can be correlated with inhibition of the 3' exonuclease activity, described in Chapter 4, which is responsible for trimming back the 3' terminus of E2* molecules to inhibitory thio-phosphodiester linkages. The identical heat sensitivity of the process of lariat 3' tail removal, and E2* 3' terminus trimming, may indicate that both processes are catalysed by the same 3' exonuclease activity.

Although HeLa cell S100 fractions were found not to be able to catalyse either step 1 or step 2 of splicing, they were found to catalyse the production of bands of the same mobility as E2*. Hence, if these bands correspond to E2* fragments, then nuclear extracts must contain a factor (or factors) which is absent from the S100 ,and which although essential splicing is not required for E2* formation. Krainer and Maniatis (1985) previously described a factor (SF2) which was micrococcal nuclease resistant and present in nuclear extract but absent from an S100 extract, and which could confer splicing activity if added to an S100 extract. SF2, whatever it may be, may therefore, be the factor(s) required for splicing, but not for E2* formation. However, there is no guarantee that the composition of the S100 fraction used in this study was identical to that of Krainer and Maniatis (1985).

Hence, E2* formation depends upon some of the factors essential for splicing, but not all of them, and on balance, it would seem that a large splicing complex does not form on [sA]RNA54 substrates which give rise to E2* on incubation under splicing conditions.

Chapter 6.

Conclusions

6. CONCLUSIONS

An attempt has been made to probe the mechanism of mammalian pre-mRNA splicing by using phosphorothioate analogues of RNA. *In vitro* transcription by T7 RNA polymerase in the presence of nucleoside 5'-O-(1-thiotriphosphates) was used to prepare pre-mRNA analogues for splicing *in vitro*.

The stereochemistry of the reaction catalysed by T7 RNA polymerase was investigated. The Sp diastereoisomer of adenosine 5'-O-(1-thiotriphosphate) (ATPαS) was found to be incorporated into RNA transcribed by T7 RNA polymerase with an apparent K_M similar to that for ATP; the Rp diastereoisomer was neither a substrate nor a competitive inhibitor. The configuration of the thiophosphodiester link in the RNA produced was analysed with stereospecific nucleases. Surprisingly, the nucleases exhibited reduced discrimination compared to their activity on dinucleotides. The results show that phosphorothioate linkages in T7 RNA polymerase transcripts are in the Rp configuration. Thus, the transcription reaction proceeds with inversion of configuration at phosphorus.

In vitro transcription by T7 RNA polymerase in the presence of a nucleoside 5'-O-(1-thiotriphosphate) has been used to prepare pre-mRNA analogues of the small intron of a rabbit β-globin gene and flanking exon sequences. Incubation of transcripts prepared with ATPαS in a HeLa cell nuclear extract showed that the presence of the thionucleotide in a transcript inhibited splicing, but a novel product (termed E2*) was formed which consisted of the 3' exon and the adjacent 3 nucleotides of the intron.

E2* formation from phosphorothioate analogues of RNA is still something of an enigma. It is definitely related to splicing, as shown by its dependence on ATP, and intact U1, U2 and U6 snRNAs. E2* formation does not, however, require all the factors necessary for *in vitro* splicing, and mutagenic studies, retardation gel assays and combined RNase T1 digestion and immunoprecipitation experiments would seem to indicate that no large splicing complex needs to be assembled for E2* formation to occur. Neither absence of a 5' splice site, or polypyrimidine tract sequence prevents E2* formation, despite both these sequences being required for splicing. The only mutation known to prevent E2* formation is that converting the 3' splice site sequence from AG/G to AG/A (where the / indicates the site of cleavage). This mutation even prevents E2* production from [sC]RNA54A in which case there is no change in the number and distribution of thio-phosphodiester linkages in the transcript. This probably implies that

E2* formation is dependent upon the binding of a factor to the region of the 3' splice junction. However, if this factor is a normal component of the splicing machinery it is somewhat surprising that its binding is affected by the AG/G to AG/A mutation, since this mutation enhances the rate of splicing *in vitro* of non-phosphorothioate substituted RNA54.

One possible explanation for E2* formation is that E2* has arisen by site specific hydrolysis, possibly followed by local thiophosphate-arrested exonuclease activity. A component, or group of components, bound to the 3' splice junction region might on normal RNA be responsible for the activation of the 3' splice site which must precede step 2 of the splicing reaction. The presence of phosphorothioate linkages in the transcript could cause a slight shift in the site of strain, causing the incorrect phosphodiester link to be activated. This strained bond might not be so well shielded from attack by H_2O or OH^- as the correct cleavage site must be, and so would be hydrolysed, giving rise to E2* production. The lag seen before the appearance of E2* could, in the case of this model, be explained if the lag represented the time required to assemble a collection of factors (possibly small factors, given the apparent lack of a stable large complex) at the 3' splice site, and the dependence of E2* production upon ATP could be explained if any of the steps leading to straining of the RNA backbone was ATP-dependent. However, reconciling this model with the dependence of E2* production on intact U1, U2 and U6 snRNPs is more difficult. One would have to speculate that the factor(s) bound at the 3' splice site interacts in a functionally important way, at some stage before straining occurs, with each of these snRNPs, possibly in the form of a single complex. Such a complex could not, however, be interacting with the 3' splice site/branch site region in a very stable manner, akin to a spliceosome, since no large complex could be detected by native gel electrophoresis after incubation under splicing conditions, and neither was any significant change in RNase T1 protected and immunoprecipitated fragments observed after time zero. Another possible explanation is that 3' splice site straining was dependent upon recycling of a factor that interacts dynamically with snRNPs in solution under splicing conditions. Assembly and disassembly of snRNPs and complexes has previously been suggested to be a very dynamic process (Konarska and Sharp, 1987).

A second model for E2* formation is that it is the result of a 5' exonuclease activity that is blocked by a factor(s) bound around the 3' splice junction. There are several features of the E2* formation process which make this model attractive. First, it would explain why the 5' fragment of [sA]RNA54, which would be generated if an endonucleolytic cleavage generated E2*, is never seen. It could be, however, that this 5' fragment is

particularly succeptible to degradation, especially as the particular sensitivity of a molecule containing only the downstream half of IVS-1 and exon 2 ([sA]RNA54-3'f) to nuclease activities in the nuclear extract was very noticeable. Second, the above mentioned molecule, [sA]RNA54-3'f, lacking the upstream portion of IVS-1 and all of exon 1 including the 5' cap structure, gave rise to E2* with accelerated kinetics, compared to untruncated [sA]RNA54, a phenomenon most easily explained by a 5' exonuclease activity. Third, the nuclear extract contained a powerful nuclease activity that was specifically inhibited by cleaving U1 or U2 snRNPs. This snRNP dependent nuclease activity could be the one responsible for the generation of E2*, and so could explain why E2* formation is inhibited by cleaving U1, U2 or U6 snRNAs, whilst no large (snRNP containing) complex seems to form on the substrate [sA]RNA. A small component, maybe even a single polypeptide, could bind to the 3' splice site and block the progress of this snRNP dependent nuclease activity if it is a 5' exonuclease. However, this model makes it difficult to explain why E2* formation from [sA]RNA54 is dependent upon the presence of a phosphorothioate linkage to the 3' side of the 5' terminus. The presence of a phosphorothioate linkage at this position would be thought to, if anything, destabilise the association of a factor with the 3' splice site, not to stabilise it.

Whatever the explanation for E2* formation, it seems fairly certain that there is a factor (or factors) associated with the 3' splice site of [sA]RNA54. Furthermore, on the basis of its sequence requirements, this factor does not seem to be the previously characterised 3' splice site binding proteins 70kd IBP (Tazi *et al.*, 1986; Gerke and Steitz, 1986) or U2AF (Ruskin *et al.*, 1988).

It also seems certain that the nuclear extract contains a nuclease activity which is dependent upon both U1 and U2 snRNPs being intact. Hence, the nuclease probably contains both U1 and U2 snRNPs. If this nuclease is responsible for E2* formation it must also be dependent on U6 snRNP, and hence probably contains in addition to U1 and U2 snRNPs, also U6 snRNA, probably in the form of the U4/U6 snRNP. This would make the snRNP composition of this nuclease remarkably similar to that of a apliceosome. It seems unlikely that the nuclease is indeed a complete spliceosome, however, since E2* formation does not require all the factors required for splicing. It is possible that the snRNP-containing nuclease is formed on endogenous pre-mRNA found in the nuclear extracts, and therefore perhaps represents some form of splicing complex; however, bands which probably represent E2* were still found after processing in S100 fractions, which are incapable of splicing complex assembly, probably due to the absence of U2AF (Ruskin *et al.*, 1988). Hence, this model would

seem to be unlikely, but cannot be excluded, since it is possible that the fragments of the same mobility as E2* formed after incubation in S100 are not in fact related to E2*, and that the S100 fractions used in this study were not identical in composition to those used by Ruskin *et al.* (1988). Could a multi-snRNP complex, which forms in the absence of substrate RNA, be the snRNA dependent nuclease? There is evidence from native gel electrophoresis that in nuclear extract the vast majority of the three snRNPs U4, U5, and U6 are present in a single complex (the U4/5/6 complex), whilst U1 and U2 snRNPs exist as separate particles (Konarska and Sharp, 1987). Under optimal *in vitro* splicing conditions, however, the U4/5/6 particle dissociates to release free snRNPs. During incubation under high salt conditions, in the absence of precursor RNA, a so-called pseudospliceosome complex forms, containing U2, U4, U5, and U6 snRNPs (Konarska and Sharp, 1988). This complex has the same snRNP composition as one of the forms of the spliceosome (complex B), as detected using the same gel system, in which U1 snRNP is never detected as part of the spliceosome. Konarska and Sharp (1988) suggested that the pseudospliceosome probably assembles as a result of a conformational change induced in snRNPs by the incubation at high temperature and ionic strength, and that an equivalent change in conformation might be induced by recognition of precursor RNA during spliceosome assembly. It seems equally possible, however, that the process seen here does not represent assembly of a multi-snRNP structure but merely different stabilities of pre-existing complexes. A complex containing U1, U2, U4, U5, and U6 snRNPs may be present even in the absence of precursor RNA and incubation in high salt. The U1 snRNP would be dissociated during native gel electrophoresis (it is never detected in spliceosomes using this system), as perhaps would U2 snRNP. Incubation in the presence of high salt or precursor RNA may result in a conformational change which causes stabilisation of the U2 snRNP association with the particle, so enabling its detection in spliceosome and pseudospliceosome complexes. The ATP-enhanced dissociation of the U4/5/6 particle under optimal splicing conditions could also merely represent a destabilisation, not a dissociation of this particle. Reed *et al.* (1988) purified functional spliceosomes by gel purification. Assembly of these particles was analysed by electron microscopy and found to be ATP dependent. However, by using a substrate RNA with a long 3' end, and visualising the tail by coating it with bactereophage T4 $g^{32}p$, only 10% of the particles in the spliceosome peak were found to contain this RNA. It was suggested that the remaining 90% of particles (which were morphologically indistinguishable from the other 10%) might be either complexes assembled in the absence of pre-mRNA, or spliceosomes assembled on endogenous pre-mRNA in the nuclear extract.

The spliceosome-containing peak was highly enriched in the snRNAs U1, U2, U4, U5, and U6, suggesting that the particles in the peak contained all these snRNPs.

A complex, which is pre-mRNA independent, and contains U1, U2, U4, U5, and U6 snRNPs, would not only provide a satisfactory explanation for how the nuclease reported in this study can be dependent on both functional U1 and U2 snRNPs (and possibly U6 snRNP too), it would also explain certain other conundrums about spliceosome assembly thrown up by the work of others. Using a gel retardation assay Konarska and Sharp (1987) found a putative pre-splicing complex (complex A), which contained only U2 snRNP. This they suggested was converted with time to the mature spliceosome (complex B) by the binding of a U4/5/6 complex in a single step. Reed *et al.* (1988), using gel filtration also identified a putative pre-spliceosome peak (peak X) which was highly enriched only in snRNAs U1 and U2. Assembly of all these complexes was dependent on time and ATP . This data is difficult to reconcile with the combined T1 digestion and immunoprecipitation data of Chabot *et al.* (1985). They found that fragments could be precipitated which extended from upstream of the 3' cleavage site of human β–globin IVS-1 by either anti-(Sm)RNP or anti-(TMG) antibodies at time zero, on ice, and in the absence of ATP. The interacting factor was not U1 or U2 snRNP as antisera against either of these factors did not immunoprecipitate fragments from the 3' splice site. The interacting factor was suggested to be U5 snRNP. However, a large, but relatively unstable multi-snRNP complex containing U1, U2, U4, U5, and U6 snRNPs could bind at time zero in an ATP-independent manner, the U5 snRNP interacting with the 3' splice site region at least in part by means of the 70kd IBP. RNase T1 digestion and immunoprecipitation would lead only to immunoprecipitation of 3' splice site fragments by anti-(Sm)RNP or anti-(TMG) antibody and not by anti-(U1)RNP or anti-(U2)RNP antibodies because association with these snRNPs would be unstable. The binding of such an unstable complex would also explain why Black *et al.* (1985) found a substantial amount of human β-globin pre-mRNA to be immunoprecipitable by anti-(U2)RNP antiserum at time zero, and yet no fragments of RNA were precipitated by the same antibody after RNase T1 digestion. This would also explain the fact that, in this study using rabbit β-globin, at time zero the same pattern of RNase T1 resistant bands were immunoprecipitated by either anti-(U1)RNP or anti-(U2)RNP, implying that U1 and U2 snRNPs were interacting at this stage. The multi-snRNP complex, being unstable, would not be detected by fractionation on native gels, or by gel filtration, unless a degree of stabilisation could occur as a result of some interaction with ATP. Only after incubation in the presence of ATP would the complex become more stable

and interactions with the precursor RNA become more extensive, for example extending to the region of the branch point. As a series of conformational changes progressed, first a complex containing U2 snRNP would become stable on native gels, and later a complex containing U2, U4, U5, and U6 would become stable. In the case of rabbit β-globin IVS-2 precursor RNA (and probably all pre-mRNAs) it would seem that an additional conformational change must occur to create the mature spliceosome, as the functional spliceosome lacks U4 snRNA when assayed by native gel electrophoresis (Lamond *et al.*, 1987) but U4 snRNA is definitely found in functional spliceosomes prepared by gel filtration (Reed *et al.*, 1988).

BIBLIOGRAPHY

Adam, S.A., T. Nakagawa, M.S., Swanson, T.K. Woodruff, and G.Dreyfuss (1986). Mol. Cell Biol. *6*, 2932-2943.

Aebi, M., H. Hornig, R.A. Padgett, J. Reiser, and C. Weissmann (1986). Cell *47*, 555-565.

Aebi, M., H. Hornig, and C. Weissmann (1987). Cell *50*, 237-246.

Ares, M. (1986). Cell *47*, 49-59.

Auffray, C., and F. Rougeon (1980). Nucleic Acids Res. *12*, 7035-7056.

Bartlett, P.A., and F. Eckstein (1982). J. Biol. Chem. *257*, 8879-8884.

Bass, B.L., and H. Weintraub (1987). Cell *48*, 607-613.

Benkovic, S.J., and K.Y. Schray (1971). In *The Enzymes*. Ed. P.D. Boyer, Vol. 8, p. 201. Academic Press, New York.

Benyajati, C., A.R. Place, N. Wang, E. Pentz, and W. Sofer (1982). Nucleic Acids Res. *10*, 7261-7272.

Berget, S.M., and B.L. Robberson (1986). Cell *46*, 691-696.

Beyer, A.L., and Y.N. Osheim (1988). Genes and Dev. *2*, 754-765.

Beyer, A.L., M.E. Christensen, B.W. Walker, and W.M. LeStourgeon (1977). Cell *11*, 127-138.

Beyer, A.L., O.L. Miller, and S.L. McKnight (1980). Cell *20*, 75-84.

Beyer, A.L., A.H. Bouton, and O.L. Miller (1981). Cell *26*, 155-165.

Bindereif, A., and M.R. Green (1986). Mol. Cell Biol. *6*, 2582-2592.

Bindereif, A., and M.R. Green (1987). EMBO J. *6*, 2415-2424.

Black, D.L., and J.A. Steitz (1986). Cell *46*, 697-704.

Black, D.L., B. Chabot, and J.A. Steitz (1985). Cell *42*, 737-750.

Blackburn, G.M., G.E. Taylor, G.R.J. Thatcher, M. Prescott, and A.G. McLennan (1987). Nucleic Acids Res. *15*, 6991-7004.

Breathnach, R., and P. Chambon (1981). Ann. Rev. Bioch *50*, 349-383.

Breathnach, R., C. Benoist, K. O'Hare, F. Gannon, and P. Chambon (1978). Proc. Natl. Acad. Sci. U.S.A. *75*, 4853-4857.

Breitbart, R.E., and B. Nadal-Ginard (1986). J. Mol. Biol. *188*, 313-324.

Bringmann, P., and R. Lührmann (1986). EMBO J. *5*, 3509-3516.

Bringmann, P., B. Appel, S. Rinke, R. Reuter, H. Theissen, and R. Lührmann (1984). EMBO J. *3*, 1357-1363.

Brody, E., and J. Abelson (1985). Science *228*, 963-967.

Brody, R.S., and P.A. Frey (1981). Biochem. *20*, 1245-1252.

Brody, R.S., S. Adler, P. Modrich, W.J. Stec, Z.J. Leznikowski, and P.A. Frey (1982). Biochem. *21*, 2570-2572.

Bryant, F.R., and S.J. Benkovic (1979). Biochem. *18*, 2825-2828.

Burgers, P.M.J., and F. Eckstein (1978). Proc. Natl. Acad. Sci. U.S.A. *75*, 4798-4800.

Burgers, P.M.J., and F. Eckstein (1979a). Biochem. *18*, 592-596.

Burgers, P.M.J., and F. Eckstein (1979b). J. Biol. Chem. *254*, 6889-6893.

Burgers, P.M.J., and F. Eckstein (1979c). Biochem. *18*, 450-454.

Burnett, L. (1982). J. Theor. Biol. *97*, 351-366.

Carroll, A.R., and R.R. Wagner (1979). J. Biol. Chem. *254*, 9339-9341.

Cellini, A., R. Parker, J. McMahon, C. Guthrie, and J. Rossi (1986). Mol. Cell

Biol. *6*, 1571-1578.

Chabot, B., and J.A. Steitz (1987). Mol. Cell Biol. *7*, 281-293.

Chabot, B., D.L. Black, D.M. LeMaster, and J.A. Steitz (1985). Science *230*, 1344-1349.

Chamberlin, M., and J. Ring (1973). J. Biol. Chem. *248*, 2235-2244.

Chamberlin, M., and T. Ryan (1982). In *The Enzymes*. Ed. P.D. Boyer, Vol 15, pp 87-108. Academic Press, New York.

Chang, T.H., M.W. Clark, A.J. Lustig, M.E. Cusick, and J. Abelson (1988). Mol. Cell Biol. *8*, 2379-2393.

Chen-Kiang, S., J.R. Nevins, and J.E. Darnell Jr. (1979). J. Mol. Biol. *135*, 733-552.

Cheng, S.C., and J. Abelson (1986). Proc. Natl. Acad. Sci. U.S.A. *83*, 2387-2391.

Cheng, S.C., and J. Abelson (1987). Genes and Dev. *1*, 1014-1027.

Choi, Y.D., P.J. Grabowski, P.A. Sharp, and G. Dreyfuss (1986). Science *231*, 1534-1539.

Christofori, G., D. Frendewey, and W. Keller (1987). EMBO J. *6*, 1747-1755.

Chu, G., and P.A. Sharp (1981). Nature *289*, 378-382.

Chung, S.Y., and J.C. Wooley (1986). Proteins *1*, 195-210.

Clark, R.F., and B.A. Hamkalo (1987). J. Cell Biol. *105*, 71a.

Cohen, N.N., A.C.Y. Chang, and L. Hsu (1972). Proc. Natl. Acad. Sci. U.S.A. *69*, 2110-2114.

Collins, F.S., and S.M. Weissman (1984). Prog. Nucleic Acids Res. Mol. Biol. *31*, 315-462.

Connolly, B.A., F. Eckstein, and A. Pingoud (1984). J. Biol. Chem. *259*, 10760-10763.

Conway, G., J.Wooley, T. Bibring, and W.M. LeStourgeon (1988). Mol. Cell Biol. *8*, 2884-2895.

Conway, L., and M. Wickens (1987). EMBO J. *6*, 4177-4184.

Cotton, F.A., E.E. Hazen, and M.J. Legg (1979). Proc. Natl. Acad. Sci. U.S.A. *76*, 2551-2555.

Darnell, J.E., Jr. (1982). Nature *297*, 365-371.

deWit, R.J.W., J. Hoppe, W.J. Stec, J. Baraniak, and B. Jastorff (1982). Eur. J. Biochem. *122*, 95-99.

Dignam, J.D., R.M. Lebovitz, and R.G. Roeder (1983). Nucleic Acids Res. *11*, 1475-1489.

Dodgson, J.B., and J.D. Engel (1983). J. Biol. Chem. *258*, 4623-4629.

Domdey, H., B. Apostol, R.J. Lin, A. Newman, E. Brody, and J. Abelson (1984). Cell *39*, 611-621.

Donahue, T.F., P.J. Farabaugh, and G.R. Fink (1982). Gene *18*, 47-59.

Dreyfuss, G. (1986). Ann. Rev. Cell Biol. *2*, 459-498.

Eckstein, F. (1985). Ann. Rev. Biochem. *54*, 367-402.

Eckstein, F., and R.S. Goody (1976). Biochem. *15*, 1685-1691.

Eckstein, F., and T. Jovin (1983). Biochem. *22*, 4546-4550.

Eckstein, F., H.H. Schulz, H. Rüterjans, W. Haar, and W. Maurer (1972). Biochem. *11*, 3507-3512.

Eckstein, F., V.W. Armstrong, and H. Sternbach (1976). Proc. Natl. Acad. Sci. U.S.A. *73*, 2987-2990.

Edery, I., and Sonenberg (1985). Proc. Natl. Acad. Sci. U.S.A. *82*, 7590-7594.

England, T.E., A.G. Bruce, and O.C. Uhlenbeck (1980). Meth. Enzymol. *65*,

65-74.

Eperon, I.C. (1986a). Nucleic Acids Res. *14*, 2830.

Eperon, I.C. (1986b). Anal. Biochem. *156*, 406-412.

Eperon, L.P., J.P. Estibeiro, and I.C. Eperon (1986). Nature *324*, 280-282.

Eperon, L.P., I.R. Graham, A.D. Griffiths, and I.C. Eperon (1988). Cell *54*, 398-401.

Erbil, C., and J. Niessing (1983). EMBO J. *2*, 1339-1343.

Esumi, H., Y. Takahashi, S. Sato, S. Nagase, and T. Sugimura (1983). Proc. Natl. Acad. Sci. U.S.A. *80*, 95-99.

Evans, R.M., and S.M. Hollenberg (1988). Cell *52*, 1-3.

Fakan, S., G. Leser, and T.E. Martin (1984). J. Cell Biol. *98*, 358-363.

Favaloro, J., R. Treisman, and R. Kamen (1980). Meth. Enzymol. *65*, 718-749.

Fischer, H.D., J.B. Dodgson, S. Hughes, and J.D. Engel (1984). Proc. Natl. Acad. Sci. U.S.A. *81*, 2733-2737.

Fisher, D.E., G.E. Conner, W.H. Reeves, G. Blobel, and H.G. Kunkel (1983). Proc. Natl. Acad. Sci. U.S.A. *80*, 6356-6360.

Fisher, D.E., W.H. Reeves, G.E. Conner, G, Blobel, and H.G. Kunkel (1984). Proc. Natl. Acad. Sci. U.S.A. *81*, 3185-3189.

Fouser, L.A., and J.D. Friesen (1986). Cell *45*, 81-93.

Frendewey, D., and W. Keller (1985). Cell *42*, 355-367.

Frey, P.A., and R.D. Sammons (1985). Science *228*, 541-545.

Fried, H.M., N.J. Pearson, C.H. Kim, and J.R. Warner (1981). J. Biol. Chem. *256*, 10176-10183.

Fu, X.Y., H. Ge, and J. Manley (1988). EMBO J. *7*, 809-817.

Fukumaki, Y., P.K. Ghosh, E.J. Benz, V.B. Reddy, P. Lebowitz, B.G. Forget, and S.M. Weissman (1982). Cell *28*, 585-593.

Furneaux, H.M., K.K. Perkins, G.A. Freyer, J, Arenas, and J. Hurwitz (1985). Proc. Natl. Acad. Sci. U.S.A. *82*, 4351-4355.

Galli, G., H. Hofstetter, H.G. Stunnenberg, and M.L. Birnsteil (1983). Cell *34*, 823-828.

Gallwitz, D. (1982). Proc. Natl. Acad. Sci. U.S.A. *79*, 3493-3497.

Gattoni, R., P. Schmitt, and J. Stevenin (1988). Nucleic Acids Res. *16*, 2389-2409.

Gerke, V., and J.A. Steitz (1986). Cell *47*, 973-984.

Gerlt, J.A., and H.Y. Wan (1979). Biochem. *18*, 4630-4638.

Goldenberg, C.J., and S.D. Hauser (1983). Nucleic Acids Res. *11*, 1337-1348.

Golomb, M., and M. Chamberlin (1974). J. Biol. Chem. *249*, 2858-2863.

Gorenstein, C., and J.R. Warner (1976). Proc. Natl. Acad. Sci. U.S.A. *73*, 1547-1551.

Grabowski, P.J., and P.A. Sharp (1986). Science *233*, 1294-1299.

Grabowski, P.J., R.A. Padgett, and P.A. Sharp (1984). Cell *37*, 415-427.

Grabowski, P.J., S.R. Seiler, and P.A. Sharp (1985). Cell *42*, 345-353.

Graham, F.L., and A.J. van der Eb (1973). Virology *52*, 456-467.

Green, M.R. (1986). Ann. Rev. Genetics *20*, 671-708.

Green, M.R., T. Maniatis, and D.A. Melton (1983). Cell *32*, 681-694.

Griffiths, A.D., B.V.L. Potter, and I.C. Eperon (1987). Nucleic Acids Res. *15*, 4145-4162.

Griffiths, A.D., B.V.L. Potter, and I.C. Eperon (1988). J. Biol. Chem. *263*, 12295-12304.

Grosveld, G.C., E. deBoer, , C.K. Shewmaker, and R.A. Flavell (1982). Nature *295*, 120-126.

Gruenwedel, D.W., and N. Davidson (1967). Biopolymers *5*, 847-861.

Guthrie, C., N. Riedel, R. Parker, H. Swerdlow, and B. Patterson (1986). In *Yeast Cell Biology*. Ed. J. Hicks, pp 301-322. Liss, New York.

Guthrie, C., N. Riedel, R. Parker, H. Swerdlow, and B. Patterson (1986). UCLA Symp. Mol. Cell. Biol. New Series *33*, 301-321.

Hardy, S.F., P.J. Grabowski, R.A. Padgett, and P.A. Sharp (1984). Nature *308*, 375-377.

Harris, S.G., S.O. Hoch, and H.C. Smith (1988). Biochem. *27*, 4595-4600.

Hartwell, L.H. (1967). J. Bacteriol. *93*, 1662-1670.

Hartwell, L.H., C.S. McLaughlin, and J.R. Warner (1970). Mol. Gen. Genet. *109*, 42-56.

Hashimoto, C., and J.A. Steitz (1984). Nucleic Acids Res. *12*, 3283-3293.

Heintz, M., and R.G. Roeder (1984). Proc. Natl. Acad. Sci. U.S.A. *81*, 2713-2717.

Hernandez, N., and W. Keller (1983). Cell *35*, 89-99.

Hinterberger, M., I. Petterson, and J.A. Steitz (1983). J. Biol. Chem. *258*, 2604-2613.

Hornig, H., M. Aebi, and C. Weissman (1986). Nature *324*, 589-591.

Horowitz, M., C.L. Cepko, and P.A. Sharp (1983). J. Mol. Appl. Gen. *2*, 147-159.

Inoue, T., F.X. Sullivan, and T.R. Cech (1985). Cell *43*, 431-437.

Inoue, T., F.X. Sullivan, and T.R. Cech (1986). J. Mol. Biol. *189*, 143-165.

Jackson, S.P., M. Lossky, and J.D. Beggs (1988). Mol. Cell Biol., in press.

Jacquier, A., and M. Rosbash (1986). Proc. Natl. Acad. Sci. U.S.A. *83*, 5835-5839.

Jacquier, A., J.R. Rodriguez, and M. Rosbash (1985). Cell *43*, 423-430.

Jongeneel, C.V., R. Sahli, G.K. McMaster, and B. Hirt (1986). J. Virol. *59*, 564-573.

Kedes, D.H., and J.A. Steitz (1987). Proc. Natl. Acad. Sci. U.S.A. *84*, 7928-7932.

Keller, E.B., and W.A. Noon (1984). Proc. Natl. Acad. Sci. U.S.A. *81*, 7417-7420.

Khoury, G., P. Gruss, R. Dhar, and C.J. Lai (1979). Cell *18*, 85-92.

King, C.R., and J. Piatigorsky (1983). Cell *32*, 707-712.

Kinlaw, C.S., B.L. Robberson, and S.M. Berget (1983). J. Biol. Chem. *258*, 7181-7189.

Kole, R., and S.M. Weissman (1982). Nucleic Acids Res. *10*, 5429-5445.

Konarska, M.M., and P.A. Sharp (1986). Cell *46*, 845-855.

Konarska, M.M., and P.A. Sharp (1987). Cell *49*, 763-774.

Konarska, M.M., and P.A. Sharp (1988). Proc. Natl. Acad. Sci. U.S.A. *85*, 5459-5462.

Konarska, M.M., R.A. Padgett, and P.A. Sharp (1984). Cell *38*, 731-736.

Konarska, M.M., P.J. Grabowski, R.A. Padgett, and P.A. Sharp (1985a). Nature *313*, 552-557.

Konarska, M.M., R.A. Padgett, and P.A. Sharp (1985b). Cell *42*, 165-171.

Koziolkiewicz, M., W. Niewiarowski, B. Uznanski, and W.J. Stec (1986). Phosphorus and Sulphur *27*, 81-92.

Krainer, A.R., and T. Maniatis (1985). Cell *42*, 725-736.

Krainer, A.R., T. Maniatis, B. Ruskin, and M.R. Green (1984). Cell *36*, 993-1005.

Krämer, A. (1987a). Proc. Natl. Acad. Sci. U.S.A. *84*, 8408-8412.

Krämer, A. (1987b). J. Mol. Biol. *196*, 559-573.

Krämer, A., and W. Keller (1985). EMBO J. *4*, 3571-3581.

Krämer, A., W. Keller, B. Appel, and R. Lührmann (1984). Cell *38*, 299-307.

Krämer, A., M. Frick, and W. Keller (1987). J. Biol. Chem. *262*, 17630-17640.

Kretzner, L., B.C. Rymond, and M. Rosbash (1987). Cell *40*, 593-602.

Kuhne, T., B. Wieringa, J. Reiser, and C. Weissman (1983). EMBO J. *2*, 727-733.

Kushner, S.R. (1978). In *Genetic Engineering*. Ed. H.B. Boyer and S. Nicosia, pp 17-23. Elsevier/North-Holland Publishing Co., Amsterdam.

Lamond, A.I., M.M. Konarska, and P.A. Sharp (1987). Genes and Dev. *1*, 532-543.

Lamond, A.I., M.M. Konarska, P.J. Grabowski, and P.A. Sharp (1988). Proc. Natl. Acad. Sci. U.S.A. *85*, 411-415.

Lang, K.M., and R.A. Spritz (1983). Science *220*, 1351-1355.

Langford, C.J., and D. Gallwitz (1983). Cell *33*, 519-527.

Langford, C.J., W. Nellen, J. Niessing, and D. Gallwitz (1983). Proc. Natl. Acad. Sci. U.S.A. *80*, 1496-1500.

Langford, C.J., F.J. Klinz, C. Donath, and D. Gallwitz (1984). Cell *36*, 645-653.

Larkin, J.C., and J.L. Woolford (1983). Nucleic Acids Res. *11*, 403-420.

Lawn, R.M., A. Efstratiadis, C. O'Connell, and T. Maniatis (1980). Cell *21*, 647-651.

Lee, S.I., S.C.S. Murthy, J.J. Trimble, R.C. Desrosiers, and J.A. Steitz (1988). Cell *54*, 599-607.

Leff, S.E., R.M. Evans, and M.G. Rosenfeld (1987). Cell *48*, 517-524.

Legrain, P., B. Seraphin, and M. Rosbash (1988). Mol. Cell Biol. *8*, 3755-3760.

Lerner, M.R., and J.A. Steitz (1979). Proc. Natl. Acad. Sci. U.S.A. *76*, 5495-5499.

Lerner, M.R., J.A. Boyle, S.M. Mount, S.L. Wolin, and J.A. Steitz (1980). Nature *283*, 220-224.

Lin, R.J., A.J. Newman, S.C. Cheng, and J. Abelson (1985). J. Biol. Chem. *260*, 14780-14792.

Lin, R.J., A.J. Lustig, and J. Abelson (1987). Genes and Dev. *1*, 7-18.

Lossky, M., G.J. Anderson, S.P. Jackson, and J. Beggs (1987). Cell *51*, 1019-1026.

Lowery, D.E., and B.G. van Ness (1988). Mol. Cell Biol. *8*, 2610-2619.

Lührmann, B., B. Appel, P. Bringman, J. Rinke, R. Reuter, and S. Rothe (1982). Nucleic Acids Res. *10*, 7103-7113.

Lustig, A.J., R.J. Lin, and J. Abelson (1986). Cell *47*, 953-963.

Maniatis, T., E.F. Fritsch, and J. Sambrook (1982). In *Molecular Cloning, a Laboratory Manual.* Cold Spring Harbor Laboratory, New York.

Manley, J.L., A. Fire, A. Cano, P.A. Sharp, and M.L. Gefter (1980). Proc. Natl. Acad. Sci. U.S.A. *77*, 3855-3859.

Marzluff, W.F., and R.C.C. Huang (1984). In *Transcription and Translation, a Practical Approach.* Ed. B.D. Hames and S.J. Higgins, pp 89-129. IRL press limited, Oxford/ Washington D.C.

Mattaj, I.W., and E.M. DeRobertis (1985). Cell *40*, 111-118.

Mattaj, I.W., W.J. Habets, and W.J. van Venrooij (1986). EMBO J. *5*, 997-1002.

Matthes, H.W.D., W.M. Zenke, T. Grundstrom, A. Staub, M. Wintzerith, and P. Chambon (1984). EMBO J. *3*, 801-805.

Mayeda, A., K. Tatei, H. Kitayama, and Y. Ohshima (1986). Nucleic Acids Res. *14*, 3045-3057.

Mehdi, S., and J.A. Gerlt (1981). J. Am. Chem. Soc. *103*, 7018-7020.

Mehdi, S., and J.A. Gerlt (1982). J. Am. Chem. Soc. *104*, 3223-3225.

Melton, D.A., P.A. Krieg, M.R. Rebagliati, T. Maniatis, K. Zinn, and M.R. Green (1984). Nucleic Acids Res. *12*, 7035-7056.

Millar, N.C. (1984). PhD. thesis, University of Bristol, U.K.

Miller, A.M. (1984). EMBO J. *3*, 1061-1065.

Mimori, T., M. Hinterberger, I. Petterson, and J.A. Steitz (1984). J. Biol. Chem. *259*, 560-565.

Mitchell, P.J., G. Urlaub, and L. Chasin (1986). Mol. Cell Biol. *6*, 1926-1935.

Moazed, D., S. Stern, and H. Noller (1986). J. Mol. Biol. *187*, 399-416.

Montell, C., and A.J. Berk (1984). Nucleic Acids Res. *12*, 3821-3827.

Montell, C., E.F. Fisher, M.H. Caruthers, and A.J. Burke (1982). Nature *295*, 380-384.

Moore, C.L., and P.A. Sharp (1985). Cell *41*, 845-855.

Mougel, M., F. Eyermann, E. Westhof, P. Romby, A. Expert-Bezançon, J.P. Ebel, B. Ehresmann, and C. Ehresmann (1987). J. Mol. Biol. *198*, 91-107.

Mount, S.M. (1982). Nucleic Acids Res. *10*, 459-472.

Mount, S.M., I. Petterson, M. Hinterberger, A. Karmas, and J.A. Steitz (1983). Cell *33*, 509-518.

Murray, A.W., and M.R. Atkinson (1968). Biochemistry *7*, 4023-4029.

McAllister, W., H. Kupper, and E. Bautz (1973). Eur. J. Biochem. *34*, 489-501.

Nelson, K.K., and M.R. Green (1988). Genes and Dev. *2*, 319-329.

Newman, A.J. (1987). EMBO J. *6*, 3833-3839.

Newman, A.J., R.J. Lin, S.C. Cheng, and J. Abelson (1985). Cell *42*, 335-344.

Noble, J.C.S., C. Prives, and J.L. Manley (1986). Nucleic Acids Res. *14*, 1219-1235.

O'Brien, C.A., S.O. Roczniak, N. Bramson, J. Baraniak, W.J. Stec, and E.T. Kaiser (1982). Biochem. *21*, 4371-4376.

Ohshima, Y., Y. Gotoh (1987). J. Mol. Biol. *195*, 247-259.

Osheim, Y.N., O.L. Miller, Jr., and A.L. Beyer (1985). Cell *43*, 143-151.

Padgett, R.A., S.F. Hardy, and P.A. Sharp (1983a). Proc. Natl. Acad. Sci. U.S.A. *80*, 5230-5234.

Padgett, R.A., S.M. Mount, J.A. Steitz, and P.A. Sharp (1983b). Cell *35*, 101-107.

Padgett, R.A., M.M. Konarska, P.J. Grabowski, S.F. Hardy, and P.A. Sharp (1984). Science *225*, 898-903.

Padgett, R.A., M.M. Konarska, M. Aebi, H. Hornig, C. Weissmann, and P.A. Sharp (1985). Proc. Natl. Acad. Sci. U.S.A. *82*, 8349-8353.

Padgett, R.A., P.J. Grabowski, M.M. Konarska, S. Seiler, and P.A. Sharp (1986). Ann. Rev. Bioch. *55*, 1119-1150.

Parker, R., and C. Guthrie (1985). Cell *41*, 107-118.

Parker, R., and B. Patterson (1987). In *Molecular Biology of RNA: New Perspectives*. Ed. B. Dudock, and M. Inouye, pp. 133-149. Academic Press, New York.

Parker, R., P.G. Siliciano, and C. Guthrie (1987). Cell *49*, 229-239.

Patterson, B., and C. Guthrie (1987). Cell *49*, 613-624.

Patzelt, E., E. Thalmann, K. Hartmuth, D. Blaas, and E. Kuechler (1987). Nucleic Acids Res. *15*, 1387-1399.

Perkins, K.K., H.M. Furneaux, and J. Hurwitz (1986). Proc. Natl. Acad. Sci. U.S.A. *83*, 887-891.

Petterson, I., M. Hinterberger, T. Mimori, E. Gottlieb, and J.A. Steitz (1984). J. Biol. Chem. *259*, 5607-6914.

Pikielny, C.W., and M. Rosbash (1985). Cell *41*, 119-126.

Pikielny, C.W., and M. Rosbash (1986). Cell *45*, 869-877.

Pikielny, C.W., J.L. Teem, and M. Rosbash (1983). Cell *34*, 395-403.

Pikielny, C.W., B.C. Rymond, and M. Rosbash (1986). Nature *324*, 341-345.

Piñol-Roma, S., Y.D. Choi, M.J. Matunis, and G. Dreyfuss (1988). Genes and Dev. *2*, 215-227.

Potter, B.V.L., P.J. Romaniuk, and F. Eckstein (1983a). J. Biol. Chem. *258*, 1758-1760.

Potter, B.V.L., B.A. Connolly, and F. Eckstein (1983b). Biochem. *22*, 1369-1377.

Rautmann, G., and R. Breathnach (1985). Nature *315*, 430-432.

Rautmann, G., H.W.D., Matthes, M.J. Gait, and R. Breathnach (1984). EMBO J. *3*, 2021-2028.

Rebagliati, M.R., and D.A. Melton (1987). Cell *48*, 599-605.

Reddy, R., D. Henning, and H. Busch (1985). J. Biol. Chem. *260*, 10930-10935.

Reed, R., and T. Maniatis (1985). Cell *41*, 95-105.

Reed, R., and T. Maniatis (1986). Cell *46*, 681-690.

Reed, R., J. Griffith, and T. Maniatis (1988). Cell *53*, 949-961.

Reeve, A.E., M.M. Smith, V. Pigiet, and R.C.C. Huang (1977). Biochem. *16*, 4464-4469.

Rhodes, G., and M. Chamberlin (1974). J. Biol. Chem. *249*, 6675-6683.

Riedel, N., J.A. Wise, H. Swerdlow, A. Mak, and C. Guthrie (1986). Proc. Natl. Acad. Sci. U.S.A. *83*, 8097-8101.

Riedel, N., S. Wolin, and C. Guthrie (1987). Science *235*, 328-331.

Rinke, J., B. Appel, M. Digweed, and R. Lührmann (1985). J. Mol. Biol. *185*, 721-731.

Rio, D.C. (1988). Proc. Natl. Acad. Sci. U.S.A. *85*, 2904-2908.

Rodriguez, J.R., C.W. Pikielny, and M. Rosbash (1984). Cell *39*, 603-610.

Rogers, J., and R. Wall (1980). Proc. Natl. Acad. Sci. U.S.A. *77*, 1877-1879.

Romaniuk, P.J., and F. Eckstein (1982). J. Biol. Chem. *257*, 7684-7688.

Rosbash, M., P.K.W. Harris, J.L. Woolford, and J.L. Teem (1981). Cell *24*, 679-686.

Rossomando, E.F., L.J. Smith, and M. Cohn (1979). Biochem. *18*, 5670-5674.

Rothermel, J.D., W.J. Stec, J. Baraniak, B. Jastorff, L.H. Parker-Botelho (1983). J. Biol. Chem. *258*, 12125-12128.

Rothstein, R.J. (1983). Methods Enzymol. *101*, 202-211.

Ruiz-Opazo, N., and B. Nadal-Ginard (1987). J. Biol. Chem. *262*, 4755-4765.

Ruskin, B., and M.R. Green (1985a). Cell *43*, 131-142.

Ruskin, B., and M.R. Green (1985b). Nature *317*, 732-734.

Ruskin, B., and M.R. Green (1985c). Science *229*, 135-140.

Ruskin, B., A.R. Krainer, T. Maniatis, and M.R. Green (1984). Cell *38*, 317-331.

Ruskin, B., J.M. Greene, and M.R. Green (1985). Cell *41*, 833-844.

Ruskin, B., C.W. Pikielny, M. Rosbash, and M.R. Green (1986). Proc. Natl. Acad. Sci. U.S.A. *83*, 2022-2026.

Ruskin, B., P.D. Zamore, and M.R. Green (1988). Cell *52*, 207-219.

Rymond, B.C., and M. Rosbash (1985). Nature *317*, 735-737.

Rymond, B.C., and M. Rosbash (1988). Genes and Dev. *2*, 428-439.

Rymond, B.C., D.D. Torrey, and M. Rosbash (1987). Genes and Dev. *1*, 238-246.

Saenger. W., D. Suck, and F. Eckstein (1974). Eur. J. Biochem. *46*, 559-567.

Salditt-Georgieff, M., M. Harpold, S. Chen-Kiang, and J.E. Darnell, Jr. (1980). Cell *19*, 69-78.

Saneyoshi, M., Z. Ohashi, F. Harada, and S. Nishimura (1972). Biochem. Biophys. Acta *262*, 1-10.

Sanger, F., S. Nicklen, and A.R. Coulson (1977). Proc. Natl. Acad. Sci. U.S.A. *74*, 5463-5467.

Schibler, U., D.E. Kelly, and R.P. Perry (1977). J. Mol. Biol. *115*, 695-714.

Schmeltzer, C., and R.J. Schweyen (1986). Cell *46*, 557-565.

Séraphin, B., L. Kretzner, and M. Rosbash (1988). EMBO J. *7*, 2533-2538.

Shapiro, D.J., P.A. Sharp, W.W. Wahli, and M.J. Keller (1988). DNA *7*, 47-55.

Shapiro, M.B., and P. Senapathy (1987). Nucleic Acids Res. *15*, 7155-7174.

Sharp. P.A. (1981). Cell *23*, 643-646.

Sharp, P.A., and M.M. Konarska (1986). Cell *44*, 211.

Sheu, K.F.R., and P.A. Frey (1977). J. Biol. Chem. *252*, 4445-4448.

Siliciano, P.G., D.A. Brow, H. Roiha, and C. Guthrie (1987a). Cell *50*, 585-592.

Siliciano, P.G., M.H. Jones, and C. Guthrie (1987b). Science *237*, 1484-1487.

Smith, M.M., A.E. Reeve, and R.C.C. Huang (1978a). Cell *15*, 615-626.

Smith, M.M., A.E. Reeve, and R.C.C. Huang (1978b). Biochem. *17*, 493-500.

Solnick, D. (1981). Nature *291*, 508-510.

Solnick, D. (1985a). Cell *42*, 157-164.

Solnick, D. (1985b). Cell *43*, 667-676.

Solnick, D. (1986). Cell *44*, 211.

Solnick, D., and S.I. Lee (1987). Mol. Cell Biol. *7*, 3194-3198.

Somasekhar, M.B., and J.E. Mertz (1985). Nucleic Acids Res. *13*, 5591-5609.

Stevens, A., and M.K. Maupin (1987). Nucleic Acids Res. *15*, 695-709.

Swanson, M.S., T.Y. Nakagawa, K. LeVan, and G. Dreyfuss (1987). Mol. Cell Biol. *7*, 1731-1739.

Tatei, K., K. Takemura, H. Tanaka, T. Masaki, and Y. Ohshima (1987). J. Biol. Chem. *262*, 11667-11674.

Tazi, J., C. Alibert, J. Temsamani, I. Reveillaud, G. Cathala, C. Brunel, and P. Jeanteur (1986). Cell *47*, 755-766.

Teem, J.L., J.R. Rodriguez, L. Tung, and M. Rosbash (1983). Mol. Gen. Genet. *192*, 101-103.

Teem, J.L., N. Abovich, N.F. Kaufer, W.F. Schwindinger, J.R. Warner, A. Levy, J. Woolford, R.J. Leer, M.M.C. van Raamsdonk-Duin, W.H. Mager, R.J. Planta, L. Schultz, J.D. Friesen, H. Fried, and M. Rosbash (1984). Nucleic Acids Res. *12*, 8295-8312.

Thimmapaya, B., and T. Shenk (1979). J. Virol. *30*, 668-673.

Tollervey, D., and C. Guthrie (1985). EMBO J. *4*, 3873-3878.

Tollervey, D., J.A. Wise, and C. Guthrie (1983). Cell *35*, 753-762.

Treisman, R., N.J. Proudfoot, M. Shander, and T. Maniatis (1982). Cell *29*, 903-911.

Treisman, R., S.H. Orkin, and T. Maniatis (1983a). In *Globin Gene Expression and Hematopoietic Differentiation.* Ed. G. Stamatoyannopoulos and A.W. Nienhuis, pp. 99-121. Liss, New York.

Treisman, R., S.H. Orkin, and T. Maniatis (1983b). Nature *302*, 591-596.

Turnbull-Ross, A.D., A.J. Else, and I.C. Eperon (1988). Nucleic Acids Res. *16*, 395-411.

Usher, D.A., D.J. Richardson, and F. Eckstein (1970). Nature *228*, 663-665.

Usher, D.A., E.S. Erenrich, and F. Eckstein (1972). Proc. Natl. Acad. Sci. U.S.A. *69*, 115-118.

van Ooyen, A., J. van den Berg, N. Mantel, and C. Weissmann (1979). Science *206*, 337-344.

van Santen, V.L., and R.A. Spritz (1985). Proc. Natl. Acad. Sci. U.S.A. *82*, 2885-2889.

Vijayraghavan, U., R. Parker, J. Tamm, Y. Iimura, J. Rossi, J. Abelson, and C. Guthrie (1986). EMBO J. *5*, 1683-1695.

Volckaert, G., and W. Fiers (1977). Anal. Biochem. *83*, 222-227.

Volckaert, G., W. Min Jou, and W. Fiers (1976). Anal. Biochem. *72*, 433-446.

Volckaert, G., J. Feunteun, L. Crawford, P. Berg, and W. Fiers (1979). J. Virol. *30*, 674-682.

Wagner, R.W., and K. Nishikura (1988). Mol. Cell Biol. *8*, 770-777.

Wallace, J.C., and M. Edmonds (1983). Proc. Natl. Acad. Sci. U.S.A. *80*, 950-954.

Warner, J.R., and S.A. Udem (1972). J. Mol. Biol. *65*, 243-257.

Warner, J.R., and C. Gorenstein (1977). Cell *11*, 201-212.

Weber, S., and M. Aebi (1988). Nucleic Acids Res. *16*, 471-486.

Westheimer, F.H. (1968). Accts. Chem. Res. *1*, 70-78.

Westheimer, F.H. (1981). In *Rearrangements in Ground and Excited States.* Ed. P. deMayo, Vol. 2, pp. 229-271. Academic Press, New York.

Wiederrecht, G., D.J. Shuey, W.A. Kibbe, and C.S. Parker (1987). Cell *48*, 507-515.

Wieringa, B., F. Meyer, J. Reiser, and C. Weissmann (1983). Nature *301*, 38-43.

Wieringa, B., E. Hofer, and C. Weissmann (1984). Cell *37*, 915-925.

Wigler, M., A. Pellicer, S. Silverstein, R. Axel, G. Urlaub, and L. Chasin (1979). Proc. Natl. Acad. Sci. U.S.A. *76*, 1373-1376.

Winter, G., and S. Fields (1980). Nucleic Acids Res. *8*, 1965-1974.

Wise, J.A., D. Tollervey, D. Maloney, H. Swerdlow, E.J. Dunn, and C. Guthrie (1983). Cell *35*, 743-751.

Wlodawer, A., R. Bott, and L. Sjölin (1982). J. Biol. Chem. *257*, 1325-1332.

Yang, V.W., M.R. Lerner, J.A. Steitz, and S.J. Flint (1981). Proc. Natl. Acad. Sci. U.S.A. *78*, 1371-1375.

Yost, H.J., and S. Lindquist (1986). Cell *45*, 185-193.

Zeitlin, S., and A. Efstratiadis (1984). Cell *39*, 589-602.

Zhuang, Y., and A.M. Weiner (1986). Cell *46*, 827-835.

Zhuang, Y., H. Leung, and A.M. Weiner (1987). Mol. Cell Biol. *7*, 3018-3020.

Zillmann, M., M.L. Zapp, and S.M. Berget (1988). Mol. Cell Biol. *8*, 814-821.

Zoller, V., and M. Smith (1983). Meth. Enzymol. *100*, 468-500.

Abstract

Investigation of the mechanism of pre-mRNA splicing using phosphorothioate analogues of RNA.

Andrew David Griffiths
1988

The Sp diastereoisomer of adenosine 5'-O-(1-thiotriphosphate) (ATPαS) was found to be incorporated into RNA transcribed by T7 RNA polymerase with an apparent K_M similar to that for ATP; the Rp diastereoisomer was neither a substrate nor a competitive inhibitor. The configuration of the thiophosphodiester link in the RNA produced was analysed with stereospecific nucleases. Surprisingly, the nucleases exhibited reduced discrimination compared to their activity on dinucleotides. The results show that phosphorothioate linkages in T7 RNA polymerase transcripts are in the Rp configuration. Thus, the reaction proceeds with inversion of configuration at phosphorus.

In vitro transcription by T7 RNA polymerase in the presence of a nucleoside 5'-O-(1-thiotriphosphate) has been used to prepare pre-mRNA analogues of the small intron of a rabbit β-globin gene and flanking exon sequences. Incubation of transcripts prepared with ATPαS in a HeLa cell nuclear extract showed that the presence of the thionucleotide in a transcript inhibited splicing, but a novel product (termed E2*) was formed by cleavage three nucleotides upstream of the 3' splice site. This product was formed with the same kinetics as the intermediates of a normal splicing reaction, and its formation depended on the presence of ATP, Mg^{2+}, and intact small nuclear RNAs U1, U2, and U6. Hence, E2* formation would seem to be a splicing related phenomenon.

However, E2* formation does not require all the components necessary for *in vitro* splicing. E2* formation was supported by S100 extracts and mildly heat treated nuclear extracts, both of which were inactive in splicing. Retardation gel asssays and combined RNase T1 digestion and immunoprecipitation experiments indicated that large spliceosome-like complexes did not form on transcripts prepared with ATPαS. Furthermore, neither the absence of a functional 5' splice site nor polypyrimidine tract sequence prevented E2* formation, despite both these sequences being required for splicing. In addition, a nuclease activity has been identified in HeLa nuclear extract which is dependent on intact small nuclear RNAs U1 and U2. This activity may be involved in E2* formation.