FACTORS INVOLVED IN 3' SPlice SITE SELECTION
IN EUKARYOTIC PRE-MESSENGER RNA.

J. Peter Estibeiro BSc. (Hons. Leicester).

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

Department of Biochemistry

October 1987
Statement.

This thesis is based on work conducted by the author in the Department of Biochemistry of the University of Leicester mainly during the period between October 1934 and July 1987.

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........................

Date......................15/10/87
ACKNOWLEDGMENTS.

I acknowledge the receipt of a Science and Engineering Research Council quota award for the duration of this work.

I thank my supervisor, Dr. Ian Eperon, for his help and advice during the time I spent in his laboratory.

I thank all the past and present members of Lab. 216 for their friendship and encouragement over the last three years.

Most of all I want to thank Kath for her help, encouragement and support over the last three years, and for accurately and efficiently typing, reading and re-reading the manuscript over the last three months.
Factors Involved in 3' Splice Site Selection in Eukaryotic Pre-Messenger RNA.

The work presented in this thesis is an investigation of factors involved in 3' splice site selection.

To try to determine the intrinsic strengths of 3' splice site sequences, a cis-competition assay system was used. This system was based on the large intervening sequence (IVS-2) of the rabbit β-globin gene. Synthetic 3' splice site sequence oligonucleotides were inserted into the EcoRI restriction site, forty nine nucleotides downstream of the authentic rabbit β-globin IVS-2 3' splice site. The oligonucleotides conformed to the established 3' splice site consensus sequence and allowed for variations within this sequence. The authentic site served as a constant reference site against which the strengths of the synthetic sites could be measured.

When spliced in HeLa cells in vivo, all constructs tested were seen to choose the authentic 3' splice site over the synthetic 3' splice site under test.

A series of mutageneses was carried out to try to decrease the intrinsic strength of the authentic site and/or improve the environment of the synthetic site such that the overall strengths of the two sites might be balanced. An AGaCG mutation at the authentic 3' splice site caused the synthetic 3' splice site to be activated as a cryptic site in vivo and in vitro. In this case lariat formation was mapped to an artificially created branch point within exon 3.

Splicing component binding to both 3' splice sites was investigated by looking at protection of the RNA from oligonucleotide directed cleavage by RNase H. Initial protection of both 3' splice sites was independent of the final choice of site. However, branch point protection was dependent on the 3' splice site chosen.

Components bound to the authentic 3' splice site could be immunoprecipitated whether that site was chosen or not. The synthetic 3' splice site was poorly precipitated even when it was chosen. This data tends to suggest that the synthetic 3' splice site directs inefficient complex assembly, and that at least partial complex assembly occurs at a 3' splice site which has been inactivated by an AGaCG mutation.

Preliminary work was carried out to develop a method for the analysis of splicing component binding to either or both 3' splice sites of material within fully and partially assembled splicing complexes (spliceosomes) isolated by sucrose gradient sedimentation.
INTRODUCTION.

Chapter 1 Introduction.

1.1 Introduction to Post Transcriptional Processing.
1.1.1 Polyadenylation.
1.1.2 An Introduction to Splicing.
1.2 Pre-mRNA Sequences Involved in Splicing.
1.3 The Mechanism of RNA Splicing.
1.3.1 Origin of the Phosphate Between Ligated Exons.
1.3.2 The Lariat Structure.
1.3.3 The Splicing Reaction.
1.3.4 The Splicing Reaction in vivo
1.4 The Effects of Mutation within the Splice Site Consensus Sequence.
1.4.1 Mutations at the 5' Splice Site.
1.4.2 Mutations at the 3' Splice Site.
1.4.3 Mutations at the branch point.
1.5 Factors Determining Splice Site Selection.
1.5.1 A Simple Scanning Mechanism?
1.6 Object of Investigation.

Chapter 2 Introduction.

2.1 Cellular Components Involved in Splicing.
2.2 Assembly of the Spliceosome.
2.3 Object of the Investigation.

Chapter 3 Materials and Methods.

3.1 Buffers.
3.1.1 Reaction Buffers.
3.1.2 Electrophoresis Buffers.
3.1.3 Electrophoresis Dyes.
3.2 DNA.
3.2.1 Bacteriophage M13 DNA.
3.2.2 Plasmid DNA.
3.2.3 Oligonucleotides.
3.3 Strains of Escherichia coli.
3.4 Bacterial Growth Media.
3.5 Transfection of Escherichia coli.
3.5.1 Preparation of Competent Escherichia coli.
3.5.2 Transfection of Competent HB101.
3.5.3 Transfection of Competent JM101 or BMH 71-18.
3.6 Preparation of DNA.
3.6.1 Preparation of Large Amounts of Single Stranded M13 DNA for Mutagenesis.
3.6.2 Preparation of Large amounts of Plasmid DNA.
3.6.3 Preparation of M13 Replicative Form (RF) DNA.
3.6.4 Preparation of Small Amounts of M13 Single Stranded DNA for Mutagenesis Screening and Sequencing. 42
3.6.5 Micropreps - Analytical Scale Microtitre Plate Plasmid Preps. 44
3.6.6 Plasmid Minipreps. 45
3.6.7 Synthetic Oligonucleotides. 46
3.6.7.1 Preparation. 46
3.6.7.2 Purification. 46
3.7 Ethanol Precipitation of DNA or RNA. 47
3.8 Phenol Extraction of DNA or RNA solutions. 47
3.9 Subcloning Small DNA Fragments into Plasmid (or M13RF) Vectors. 48
3.10 Restriction Digests. 49
3.11 Colony Hybridisation. 50
3.11.1 Preparation of Probe. 50
3.11.2 Preparation of Filters for Probing. 51
3.12 Electrophoresis. 53
3.12.1 Polyacrylamide Gels. 53
3.12.1.1 Acrylamide Preparation. 53
3.12.1.2 Denaturing Gel Preparation. 53
3.12.1.3 Sample Preparation. 54
3.12.1.4 Gel Handling. 54
3.12.2 Agarose Gels. 54
3.12.2.1 Gel Preparation. 54
3.12.2.2 Sample Preparation. 55
3.12.2.3 Gel Handling. 55
3.12.2.4 Low Melting Point (LMP) Agarose. 55
3.12.2.5 Hot Phenol Extraction. 55
3.13 Dideoxy Sequencing of DNA Cloned into Bacteriophage M13. 56
3.14 Oligonucleotide Mutagenesis. 58
3.14.1 Phosphorylation of Oligonucleotide Primer 5' Termini. 59
3.14.2 Second Strand Synthesis of M13 DNA ("Copying Reactions"). 59
3.15 Screening for and Determination of Orientation of Oligonucleotide Cloned into pJsv or Psv3 61
3.15.1 Screening for Insert. 61
3.15.2 Checking for Orientation of Insert. 62
3.15.2.1 Phosphorylation of Primer. 63
3.15.2.2 Preparation of Template DNA. 63
3.15.2.3 Procedure. 63
3.16 Cell Culture. 64
3.16.1 Maintenance of Monolayer Cells. 64
3.16.2 Maintenance of Suspension Cells. 65
3.16.3 Cell Stocks. 65
3.17 Transfection of HeLa Cells for in vitro Splicing. 66
3.17.1 Solutions. 66
3.17.2 Procedure. 66
3.18 Harvesting of Whole Cell RNA from HeLa Cells. 67
3.18.1 Solutions. 67
3.18.2 Harvesting. 67
3.19 Nuclease SI Mapping of HeLa Cell RNA. 68
3.19.1 Preparation of Probes. 68
3.19.2 Annealing of Probe to RNA. 69
3.19.3 Nuclease SI Digestion. 70
3.19.3.1 SI Incubation Mix. 70
3.19.3.2 Procedure. 70
3.20 In Vitro Transcription. 71
3.20.1 Preparation of Transcription Template. 71
3.20.2 Transcription of Linear Template by T7 RNA Polymerase. 71
3.20.3 Gel Purification of Transcripts. 72
3.21 Preparation of Splicing Extract. 73
3.21.1 Buffers. 73
3.21.2 Preparation of Extract. 74
3.21.3 Preparation of SI100 Fraction. 75
3.22 In vitro Splicing. 76
3.23 Mapping the Position of a Branched Nucleotide in RNA by Primed Synthesis. 77
3.24 RNase H Protection. 78
3.25 Immunoprecipitation. 79
3.26 Sucrose Gradients. 80
3.26.1 Preparation of Gradients. 80
3.26.2 Preparing Ribosome Subunits for Calibration of Gradients. 80
3.26.3 Preparing Splicing Reactions for Analysis on Sucrose Gradients. 81
3.26.4 Unloading Gradients and Analysis of Fractions. 81
3.26.5 Analysis of Spliceosome Gradient Fractions by Polyacrylamide Gel Electrophoresis. 82

RESULTS.

Results Section 1.

Chapter 4 3' Splice Site Usage in vivo. 83
4.1 Cloning of Synthetic 3' Splice Sites into pBSV. 83
4.1.1 Indirect Strategy. 83
4.1.2 Direct Strategy. 87
4.1.3 Plasmid Micropreps. 90
4.1.4 Microprep Screen. 91
4.1.5 Checking Orientation of Cloned Oligonucleotides. 92
4.1.6 Preparation of Material for Experimentation. 93
4.2 Transient Expression of pBSV Constructs in HeLa Cells and Isolation of Whole Cell RNA for SI Nuclease Mapping. 95
4.2.1 Transfection of HeLa Cells. 95
4.2.2 Harvesting of RNA. 96
4.2.3 SI Mapping of Transfected HeLa RNA. 96

Chapter 5 Investigation of Sequence Involvement in Splice Site Selection by In Vitro Mutagenesis. 98
5.1 Mutagenesis. 98
5.2 Subcloning Mutants Back into pBSV3. 99
<table>
<thead>
<tr>
<th>Section/Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>5.3</strong></td>
<td>3' Splice Site Usage in Modified pBSV3 Constructs <em>In Vivo.</em></td>
<td>101</td>
</tr>
<tr>
<td><strong>Results Section 2.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Chapter 6.</strong></td>
<td><em>In Vitro</em> Splicing.</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>Splice Site Usage <em>In Vitro</em> Analysed by S1 Nuclease Mapping.</td>
<td>104</td>
</tr>
<tr>
<td><strong>Chapter 7.</strong></td>
<td>Investigation of Spliceosome Components Binding to Authentic and Cryptic 3' Splice Sites by Ribonuclease H Protection Mapping.</td>
<td>112</td>
</tr>
<tr>
<td><strong>Chapter 8.</strong></td>
<td>Investigation of Spliceosome Assembly Using Immunoprecipitation.</td>
<td>119</td>
</tr>
<tr>
<td><strong>Chapter 9.</strong></td>
<td>Isolation of Spliceosome Complexes by Velocity Gradient Sedimentation.</td>
<td>124</td>
</tr>
<tr>
<td><strong>Chapter 10.</strong></td>
<td>Discussion.</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>Efficiency of Cloning Oligonucleotides.</td>
<td>129</td>
</tr>
<tr>
<td></td>
<td>The Microprep Screen.</td>
<td>129</td>
</tr>
<tr>
<td></td>
<td>Use of Alternative 3' Splice Sites <em>In Vivo.</em></td>
<td>130</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>10.4</td>
<td>The Influence of Secondary Structure on Splice Site Selection.</td>
<td>130</td>
</tr>
<tr>
<td>10.5</td>
<td>The Importance of Position and Context in Splice Site Selection.</td>
<td>132</td>
</tr>
<tr>
<td>10.6</td>
<td>The Effects of 3' Splice Site Point Mutations on 3' Splice Site Selection <em>In Vivo</em>.</td>
<td>135</td>
</tr>
<tr>
<td>10.7</td>
<td>Future Work to be Done <em>In Vivo</em>.</td>
<td>139</td>
</tr>
<tr>
<td>10.8</td>
<td>Future Development of the Splice Site Assay System.</td>
<td>141</td>
</tr>
<tr>
<td>10.9</td>
<td>Nuclease Protection of Authentic and Cryptic 3' Splice Sites <em>In Vitro</em>.</td>
<td>144</td>
</tr>
<tr>
<td>10.10</td>
<td>Complex Assembly at the Authentic and Cryptic 3' Splice Sites <em>In Vitro</em>.</td>
<td>150</td>
</tr>
<tr>
<td>10.11</td>
<td>Protection of Splice Sequences within Isolated Spliceosomes.</td>
<td>152</td>
</tr>
<tr>
<td>10.12</td>
<td>Concluding Remarks.</td>
<td>155</td>
</tr>
<tr>
<td></td>
<td>References.</td>
<td>158</td>
</tr>
</tbody>
</table>
INDEX OF FIGURES AND TABLES.

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

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>5</td>
</tr>
<tr>
<td>Figure 2</td>
<td>9</td>
</tr>
<tr>
<td>Figure 3</td>
<td>10</td>
</tr>
<tr>
<td>Figure 4</td>
<td>18</td>
</tr>
<tr>
<td>Figure 5</td>
<td>18</td>
</tr>
<tr>
<td>Figure 6</td>
<td>20</td>
</tr>
<tr>
<td>Figure 7</td>
<td>21</td>
</tr>
<tr>
<td>Figure 8</td>
<td>31</td>
</tr>
<tr>
<td>Figure 9</td>
<td>83</td>
</tr>
<tr>
<td>Figure 10</td>
<td>87</td>
</tr>
<tr>
<td>Figure 11</td>
<td>89</td>
</tr>
<tr>
<td>Figure 12</td>
<td>91</td>
</tr>
<tr>
<td>Figure 13</td>
<td>92</td>
</tr>
<tr>
<td>Figure 14</td>
<td>93</td>
</tr>
<tr>
<td>Figure 15</td>
<td>93</td>
</tr>
<tr>
<td>Figure 16</td>
<td>96</td>
</tr>
<tr>
<td>Figure 17</td>
<td>96</td>
</tr>
<tr>
<td>Figure 18</td>
<td>97</td>
</tr>
<tr>
<td>Figure 19</td>
<td>98</td>
</tr>
<tr>
<td>Figure 20</td>
<td>101</td>
</tr>
<tr>
<td>Figure 21</td>
<td>104</td>
</tr>
<tr>
<td>Figure 22</td>
<td>105</td>
</tr>
<tr>
<td>Figure 23</td>
<td>107</td>
</tr>
<tr>
<td>Figure 24</td>
<td>107</td>
</tr>
<tr>
<td>Figure 25</td>
<td>108</td>
</tr>
<tr>
<td>Figure 26</td>
<td>109</td>
</tr>
<tr>
<td>Figure 27</td>
<td>109</td>
</tr>
<tr>
<td>Figure 28</td>
<td>113</td>
</tr>
<tr>
<td>Figure 29</td>
<td>114</td>
</tr>
<tr>
<td>Figure 30</td>
<td>114</td>
</tr>
<tr>
<td>Figure 31</td>
<td>115</td>
</tr>
<tr>
<td>Figure 32</td>
<td>116</td>
</tr>
<tr>
<td>Figure 33</td>
<td>117</td>
</tr>
<tr>
<td>Figure 34</td>
<td>119</td>
</tr>
<tr>
<td>Figure 35</td>
<td>119</td>
</tr>
<tr>
<td>Figure 36</td>
<td>121</td>
</tr>
<tr>
<td>Figure 37</td>
<td>121</td>
</tr>
<tr>
<td>Figure 38</td>
<td>122</td>
</tr>
<tr>
<td>Figure 39</td>
<td>124</td>
</tr>
<tr>
<td>Figure 40</td>
<td>124</td>
</tr>
<tr>
<td>Figure 41</td>
<td>126</td>
</tr>
<tr>
<td>Figure 42</td>
<td>130</td>
</tr>
<tr>
<td>Figure 43</td>
<td>133</td>
</tr>
<tr>
<td>Figure 44</td>
<td>134</td>
</tr>
</tbody>
</table>
**Abbreviations.**

DNA  deoxyribonucleic acid.
RNA  ribonucleic acid.
ss  single stranded.
ssDNA  single stranded DNA.
RF  replicative form.
dN; dNTP  2'-deoxyribonucleotide 5'-triphosphate.
ddN; ddNTP  2', 3'-dideoxyribonucleoside 5'-triphosphate.
N  one of the five bases:
A  adenosine.
C  cytosine.
G  guanine.
T  thymine.
U  uracil.
N; NTP; rNTP  ribonucleoside 5'-triphosphate.
Y; Py  pyrimidine.
R; Pu  purine.
mRNA  messenger RNA.
tRNA  transfer RNA : phenylalanine specific.
bp  base pairs.
AMV  avian myeloblastosis virus.
SDS  sodium dodecyl sulphate.
EDTA  ethylene diaminetetraacetic acid.
DTT  dithiothreitol.
PMSF  phenylmethylsulphonylfluoride.
Tris  Tris (hydroxymethyl) - aminomethane.
LMP  low melting point.
TEMED  N, N, N', N'-tetramethylenediamine.
IPTG  isopropyl-β-D-thiogalactopyranoside.
BCIG  5-bromo, 4-chloro, 3-indolyl, β-galactoside.
RNase  ribonuclease
diGTP; cap  diguanosine triphosphate G(5')ppp(5')G.
RNP  ribonucleoprotein.
snRNP  small nuclear RNP.
hnRNP  heterogeneous nuclear RNP.
PCA  phenyl/chloroform/isomylalcohol.
PNK  polynucleotide kinase.
NaOAc  sodium acetate.
XC  xylene cyanol.
BPP  bromophenol blue.
Klenow polymerase  *E. coli* DNA polymerase I, Klenow fragment.
1.1 Introduction to Post-Transcriptional Processing.

It is a characteristic feature of eukaryotic cells that the primary transcription products of their structural genes (pre-mRNAs) are processed, post-transcriptionally, in a variety of ways to create the functional messenger RNAs (mRNAs).

The two post-transcriptional modifications to pre-mRNAs which involve site specific cleavage are polyadenylation and splicing. This thesis is a study of splicing and therefore polyadenylation will only be mentioned briefly.

1.1.1 Polyadenylation.

An endonucleolytic cleavage reaction may occur on the nascent RNA transcript, generating a 3' end to which a poly(A) tail of about two hundred nucleotides is added. There is a very highly conserved consensus sequence AAUAAA which is found 10 to 30 nucleotides upstream from the poly(A) addition site (Proudfoot and Brownlee 1976). Deletion or alteration of this sequence prevents cleavage of the pre-mRNA (Fitzgerald and Shenk 1981; Montell et al. 1983; Higgs et al. 1983; Wickens and Stephenson 1984) although polyadenylation may still occur at normal and abnormal sites.

There are many cases of transcribed AAUAAA sequences that do not serve as cleavage sites, and selection among multiple poly(A) sites would be conceptually and mechanistically difficult if based solely on this
simple sequence. In the last three or four years several groups have
defined sequence elements downstream of several poly(A) sites. These
elements are critical for poly(A) site formation (Gil and Proudfoot 1984;
McDevitt et al. 1984; McLauchlan et al. 1985). The conservation of these
downstream sequences is much less than that of the AAUAAA element
suggesting the possibility that there may be sites for regulatory
interactions. Indeed, it has been shown recently (Gil and Proudfoot 1987)
that a GU rich sequence element and a U-rich sequence element which are
found 3' of the AAUAAA in many eukaryotic structural genes have the
potential to regulate the efficiency of 3' end processing by the presence
of either or both of the elements. This ultimately results in the
modulation of the level of stable functional mRNA.

The mechanism and components involved in polyadenylation are very
poorly understood. It was originally suggested (Berget 1984) that U4
snRNP might be involved in the polyadenylation complex. Sequences within
the RNA of U4 are complementary to the sequence CAYUG which is found 3'
of the AAUAAA before or after the polyadenylation site of many genes.

The involvement of U4 small ribonucleoprotein (snRNP) in
polyadenylation is now thought to be unlikely. U4 has been shown to be
involved in splicing (Black and Steitz 1986; Berget and Robberson 1986)
but oligonucleotide directed cleavage of U4 RNA does not inhibit
polyadenylation in vitro (Berget and Robberson 1986). It is still possible
that parts of the molecule not inactivated by this procedure may have a
role in polyadenylation. The absence of any U4-specific antibodies at the
moment makes this a very difficult hypothesis to test.
Polyadenylation is similar to splicing in that the site of cleavage is determined by sequences within the pre-mRNA; the process, like splicing, probably involves small RNAs in the catalytic complex.

1.1.2 An Introduction to Splicing.

It was discovered in 1977 that many eukaryotic genes, unlike bacterial genes, were not continuous (Berget et al. 1977; Goldberg et al. 1977; Klessig 1977; Chow et al. 1977; Jeffreys and Flavell 1977), but rather were interrupted by intervening sequences (IVS), or introns, which were transcribed and so included in the nascent RNA. These intervening sequences are accurately removed by a post-transcriptional processing event to create the mature mRNA. This post-transcriptional event is, in very simple terms, a precise cleavage-ligation reaction and has been termed splicing.

Splicing demands remarkable accuracy as even a small error would be intolerable, resulting in an RNA that could not encode a functional polypeptide. Another related problem is how a pre-mRNA with up to fifty introns can be efficiently and accurately spliced without accidentally deleting exons.

In the following chapters which make up the introduction I will discuss many of the factors which are thought to confer this remarkable accuracy and selectivity on the splicing machinery.

All of the early work on splicing was done using in vivo systems. This involved the analysis of processing of either endogenous cellular genes, or normal and mutant cellular or viral genes reintroduced into cells. Although much progress was made using this approach, it is limited
because only the steady state levels of the spliced mRNAs can be measured. This means that the detailed biochemical mechanisms and kinetics of the reaction or reactions cannot be easily elucidated.

Over the last three years, with the development of in vitro splicing systems, it has been possible for a tremendous amount of progress to be made towards understanding the mechanism of the splicing reaction.

Data from these new in vitro systems has complemented data generated in vivo and in many cases the complementary data has led to a fuller understanding of the mechanism. For this reason, where applicable, I will discuss both sets of data together rather than discuss the progressive understanding of splicing in chronological order.

1.2 Pre-mRNA sequences involved in splicing.

The idea that cis-acting sequence elements play a major role in splicing was first given weight by the work of Breathnach et al. (Breathnach et al. 1978). They compared ovalbumin genomic and cDNA sequences and identified short, conserved-sequence elements at the exon/intron boundaries.

By analysing the exon/intron boundaries of a large number of nuclear and virally encoded genes, more extensive conserved sequences at the two splice sites have been identified (Mount 1982). In higher eukaryotes almost all naturally occurring introns have the dinucleotide GU at their 5' ends. The only known exceptions all have a GC and are an alternatively spliced intron in a murine αA-crystallin gene (King and Piatigorsky 1983) and introns of a chicken (Dodgson and Engel 1983) and duck (Erbil and Niessing 1983) α-globin gene. The remainder of the 5' splice site
consensus sequence is less well conserved, the best consensus is 
(C/A)AG^GU(A/G)AG.

All higher eukaryotic introns have the dinucleotide AG at their 3'
ends. There have been no reported exceptions. The consensus sequence
includes a pyrimidine-rich region of variable length (the polypyrimidine
tract), a nonconserved position, another pyrimidine and then the absolutely
conserved AG at the exon/intron boundary. The 3' splice site consensus
is (U/C)ₙ(N(C/U)AGₙG (Mount 1982).

The percentage of occurrence of each base at each position in both
the 5' and 3' consensus sequences is shown in Table 1. This is a
tabulation of the sequences at assigned 5' and 3' splice sites in
approximately four hundred vertebrate genes in the GenBank Data Base
(Padgett et al. 1986).

It was discovered that in yeast there is an additional signal
(Pikielny et al. 1983; Longford and Gallwitz 1983). The sequence UACUAAC
which is found within the intron near the 3' splice site and which is
required for splicing. This signal in yeast is very highly conserved
(Longford et al. 1984) but is only rarely seen in metazoan introns.

It is possible that all eukaryotes share a common basic mechanism
for intron splicing that requires a 3' splice signal comprising of a 3'
splice site and another element, i.e. the UACUAAC sequence. The exact
sequence of this other element may have been modified in different
organisms in the course of evolution. Making this assumption Keller and
Noon used a computer program to search intron sequences from various
organisms searching for homologues to the yeast signal (Keller and Noon
1984). Introns from sea urchin, rat and human genes were searched and in
Table 1.

Consensus sequences of vertebrates

5' splice site

<table>
<thead>
<tr>
<th>exon</th>
<th>intron</th>
</tr>
</thead>
<tbody>
<tr>
<td>%G</td>
<td>23</td>
</tr>
<tr>
<td>%A</td>
<td>34</td>
</tr>
<tr>
<td>%U</td>
<td>15</td>
</tr>
<tr>
<td>%C</td>
<td>29</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5' splice site</th>
</tr>
</thead>
<tbody>
<tr>
<td>_ _ A G: G U A A G U _ _</td>
</tr>
</tbody>
</table>

3' splice site

<table>
<thead>
<tr>
<th>intron</th>
<th>exon</th>
</tr>
</thead>
<tbody>
<tr>
<td>%G</td>
<td>18</td>
</tr>
<tr>
<td>%A</td>
<td>17</td>
</tr>
<tr>
<td>%U</td>
<td>37</td>
</tr>
<tr>
<td>%C</td>
<td>28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3' splice site</th>
</tr>
</thead>
</table>

A tabulation of the sequences at assigned 5' and 3' splice sites in approximately 400 vertebrate genes in the Gen Bank Data Base. All examples were included where intervening sequences began with a GU dinucleotide and terminated with an AG dinucleotide.
all cases homologues to the yeast signal were found. The consensus sequences from each species are shown below:

<table>
<thead>
<tr>
<th>Species</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yeast</td>
<td>UACUAAC</td>
</tr>
<tr>
<td>Sea Urchin</td>
<td>A(C/U)UAU</td>
</tr>
<tr>
<td>Mouse and Rat</td>
<td>CTGAC</td>
</tr>
<tr>
<td>Human</td>
<td>CTGAC</td>
</tr>
</tbody>
</table>

from these a more general consensus is YURAY. Y and R represent pyrimidine and purine nucleotides respectively.

1.3 The Mechanism of RNA splicing.

The roles of the above sequences have been studied in detail since the development of efficient cell free splicing systems. A very efficient in vitro splicing system was developed (Krainer et al. 1984). This system was used to splice RNA precursors synthesized from plasmids in which a promoter from the bacteriophage SP6 was fused to the human β-globin structural gene (Green et al. 1983; Kramer et al. 1984).

In addition to accurately spliced RNA, this system also generated a number of RNA processing products, one of which had the characteristics of the cleaved 5' exon and was thought to be an intermediate of the reaction. The existence of another possible intermediate was shown by a primer extension product, the 3' end of which mapped to a discrete site within the intron. However, analysis by S1 nuclease mapping failed to reveal the expected RNA cleavage site so it was suggested that the termination of
reverse transcription was due to the presence of a structure that blocks the enzyme, rather than RNA cleavage.

Work was undertaken to characterise fully the products and intermediates of the in vitro splicing reaction, mainly by the groups of Sharp (Padgett et al. 1984) and Maniatis (Ruskin et al. 1984). Although Sharp's group were using a precursor derived from the major late transcription unit of adenovirus 2 and Maniatis's group were using an SP6/β-globin RNA precursor containing the first two exons and the first intron (IVS1), their conclusions were the same.

1.3.1 Origin of the Phosphate between ligated exons.

Ribonuclease T1 cuts at the intron/exon junction of an adenovirus 2 major late transcript (Padgett et al. 1984). In a two dimensional fingerprint analysis of ribonucleasese T1 oligonucleotides from splicing intermediates of an adenovirus 2 major late transcript, the T1 products from the intervening sequence adjacent to the 3' splice site were different from those in digests of precursor RNA. These differences could also be resolved by polyacrylamide gel electrophoresis where the T1 products from the intermediates migrated more slowly than the equivalent oligonucleotides generated by T1 digestion of the precursor. This effect was explained by showing that if the T1 products of both intermediates and precursor were treated with phosphatase to remove 3' terminal phosphate moieties, then, when analysed by polyacrylamide gel electrophoresis the products adjacent to the 3' splice site, which had previously shown different mobilities, were seen to comigrate. The conclusion from this was that the intron was excised in a form lacking
the 3' terminal phosphate group. The T1 analysis therefore showed that processing at the 3' splice site involves cleavage after the AG dinucleotide and that the terminal G residue has a 3' hydroxyl group.

The above conclusion implies that the junction phosphate linking the two exons must have been retained from the phosphate on the 3' side of the G residue at the 3' splice site. In the adenovirus 2 major late transcription unit this newly formed phosphodiester bond is contained in the unique ribonuclease A tetranucleotide GpGpGp*Cp (* denotes the phosphate group at the junction). When α²³P-CTP labelled splice products were digested with ribonuclease A and analysed by two dimensional fingerprinting then the GpGpGp*Cp tetranucleotide is seen to be labelled. This confirms that the phosphate moiety in the bond joining the two exons originates from the α position of the 5' most residue (in this case a C) of the 3' exon.

1.3.2 The Lariat Structure.

It had been previously suggested (Grabowski et al. 1984; Krainer et al. 1984) that splicing products and intermediates containing the intron included a branched RNA structure similar to that described by Wallace and Edmonds (Wallace and Edmonds 1983) who first identified an RNA branch structure in HeLa nuclear but not cytoplasmic RNA.

Ruskin et al. (1984) observed that the mobilities of the two splicing specific, intron containing moieties on polyacrylamide gels were inconsistent with them being simple linear molecules. These discrepancies were substantial when the RNA species were analysed on high percentage polyacrylamide gels. The abnormal electrophoretic mobilities were
consistent with the hypothesis that the RNA species were in a circular form. Cleavage of the intron containing material by oligonucleotide directed RNase H confirmed that the molecules were circularised at their 5' end but not at their 3' ends.

As mentioned earlier it had been noted that reverse transcription from a primer annealed to the intron containing molecules was blocked by a structure within the RNA (Krainer et al. 1984). Furthermore, the primer extension stops on both intron containing moieties generated by the in vitro splicing reaction were mapped to the same position near the 3' end of the intron (Ruskin et al. 1984).

Both RNA species contain a component which is resistant to digestion by several nucleases: RNase T1, RNase A, nuclease P1 and nuclease T2 (Padgett et al. 1984; Ruskin et al. 1984). This component is, however susceptible to digestion by snake-venom phosphodiesterase (Konarska et al. 1985a).

These observations led to the proposal that the RNA was in the form of a lariat (Padgett et al. 1984; Ruskin et al. 1984). The 5' end of the intron is linked to an adenosine residue upstream of the 3' splice site through a 2'-5' phosphodiester bond (Fig. 1) (Konarska et al. 1985a). It is the 2'-5' phosphodiester bond that produces a nuclease resistant trinucleotide which constitutes the branch point of a lariat structure.

This branch point is within the 3' splice signal proposed by Keller and Noon (1984) confirming that this sequence plays an important role in splicing.
Figure 1. - The RNA branch structure. The core structure of an RNA branch is shown. The bases typically found in the branch structure are abbreviated and circled. Abbreviations: A, Adenosine; G, guanosine; Py, pyrimidine. The A and G are the same as shown in the branch point of Figure 2.
fig 1.
1.3.3 The Splicing Reaction

A diagramatic representation of the proposed splicing reaction is shown in Fig. 2. In the first step it is possible that the 2' hydroxyl group of the branch point A directly participates in a transesterification reaction producing the RNAs shown in the figure. If this is the case then 5' splice site cleavage and lariat formation are concerted reactions. Alternatively, cleavage at the 5' splice site and ligation of the free 5' end of the intron to the 2' position of the branch point nucleotide could occur through a series of reactions too rapid to produce detectable intermediates.

The second step in the reaction is the formation of a bond between the 3' hydroxyl group of the first exon and the phosphate group at the 3' splice site. This step could also proceed either by a concerted transesterification reaction or by a series of rapid reactions the intermediates of which are not detectable.

1.3.4 The Splicing Reaction in vivo

Analysis of the rabbit β-globin transcripts present in the steady state RNA population of fetal liver (Zeitlin and Efstratiadis 1984) and of the products of a modified β-globin gene introduced into HeLa cells (Rautmann and Breathnach 1985) have shown that splicing in vivo proceeds via a lariat intermediate and that the branch point maps to the predicted 'A' residue. This is evidence that the mechanism in vivo is the same as that shown in vitro.

It should be noted that parallel work on yeast, in vivo, has shown that splicing also proceeds through a lariat intermediate. Analysis of the
Figure 2. - A schematic representation of the pre-mRNA splicing reaction as described in the text. Boxes represent exons; the line represents an intron. The GU at the 5' splice site, the AG at the 3' splice site and A, at the the branch point are indicated. The phosphates at the 5' and 3' splice junctions are designated by a small 1 and 2 respectively.
major species of excised actin intron, labelled with \( ^{32}P \) in vivo, revealed the presence of a branch at the third A in the UACUAAC (TACTAAC box) sequence. This A was shown to be linked via a 3'-5' phosphodiester bond to the downstream C, and via a 2'-5' phosphodiester bond to a G, presumably from the 5' end of the intron (Domdey et al. 1984). This lariat intermediate was shown to contain a complete intron and the 3' exon but to be missing the 5' exon. Other intermediates found were the excised intron and the free 5' exon (Rodriguez et al. 1984). This striking structural resemblance between the intermediates suggests that splicing must proceed by fundamentally similar mechanisms in yeast and mammals, despite the stricter adherence to a specified primary sequence at the TACTAAC box in yeast.

1.4 The Effects of Mutation Within the Splice Site Consensus Sequences.

1.4.1 Mutations at the 5' Splice Site.

At this stage the effect of mutations within the consensus sequences will be discussed with respect to the two step mechanism. The effect of mutations on splicing complex assembly, and the use of cryptic splice sites will be discussed later.

It has been shown by deleting and replacing intron sequence (Wieringa et al. 1984), that the six 5'-terminal nucleotides of the intron, i.e. those within the proposed consensus sequence (Mount 1982), are sufficient for 5'-splice site use.

If the almost invariant GU dinucleotide at the 5' splice site is changed by mutation either of the G (Wieringa et al. 1983; Treisman et al.
1982) or of the U (Montell et al. 1982), then authentic splicing is completely abolished although cryptic sites (see later) may be activated. Mutations at positions 5 and 6 within the consensus have variable effects and usually cryptic sites are activated (Treisman et al. 1983). Purine transitions three and four nucleotides downstream or one, two or three nucleotides upstream from the rabbit β-globin IVS-2 5' splice site do not inactivate the splice site. Normal efficiency of authentic splicing is maintained and no 5' cryptic sites are activated (Wieringa et al. 1983).

Generally, mutations which change the 5' splice site away from the consensus sequence reduce the efficiency of splicing.

14.2 Mutations at the 3' Splice Site.

The effects of mutations within the 3' consensus sequence are probably even more surprising and interesting than those within the 5' consensus sequence.

Initial experiments in vivo showed that progressive deletion of the polypyrrimidine tract resulted in reduced splicing efficiency and eventually led to a cryptic 3' splice site being used (Wieringa et al. 1984). These in vivo experiments looking at steady state levels of RNA gave no clue as to which step of splicing was inhibited. It was predicted that as the 3' splice site participates in step two of the reaction, then if the two steps of the reaction could be uncoupled by mutations of the splice site sequences, the intermediates would be generated but no products. This was found to be true for only some of the mutants tested.

Following the development of efficient in vitro splicing systems, it was observed that if the 3' splice site comprising of the polypyrrimidine
tract and the AG dinucleotide was deleted, either experimentally (Ruskin and Green 1985a) or in the naturally occurring thalassaemia β⁰⁺ mutation (Reed and Maniatis 1985), then splicing was inhibited and only RNA precursor could be detected by polyacrylamide gel electrophoresis. To try to separate the functions of the polypyrimidine tract from those of the invariant AG dinucleotide, constructs were made either by a small deletion (Ruskin and Green 1985a) or by a single base change (Reed and Maniatis 1985), to remove or modify the AG whilst leaving the polypyrimidine tract intact.

Surprisingly, the polypyrimidine tract alone allowed cleavage at the 5' splice site and lariat formation to the authentic branch point (Ruskin and Green 1985a; Reed and Maniatis 1985). The splicing intermediates could be detected by gel electrophoresis and the reaction was seen to be at a slightly reduced efficiency when compared to the splicing of a normal transcript. This implies that although the 3' splice site participates in the second step of splicing and the AG dinucleotide is required for this step, the polypyrimidine tract is essential for 5' splice site cleavage and lariat formation. The polypyrimidine tract may therefore be important as a recognition signal.

1.4.3 Mutations at the Branch Point.

In yeast, efficient lariat formation has a stringent requirement for the strictly conserved TACTAAC box and, as mentioned earlier, deletions or mutations of this sequence abolish splicing. In contrast, the branch acceptor sequence of higher eukaryotes is poorly conserved (Keller and Noon 1984).
Wieringa et al. (1984) found that most of the intervening sequence including the authentic branch point could be deleted or replaced by junk sequence without affecting splicing in vivo. Closer examination of some of Wieringa's constructs, following the development of efficient cell free splicing systems, showed that splicing was still proceeding via a lariat intermediate although at a reduced efficiency compared to the unmodified transcripts in vitro. In the cases examined, the branch was found to be formed by a 2'-5' phosphodiester bond between the G residue at the 5' end of the intron and an A residue about thirty nucleotides upstream of the 3' splice site. Interestingly, the sequences at the branch site were noted to bear almost no resemblance to the consensus branch site sequences as proposed by Keller and Noon (1984). The only consistent feature was the presence of a purine nucleoside immediately upstream of the branch point A (Padgett et al. 1985). Even this R-A sequence is not invariant; similar experiments (Ruskin et al. 1985) have demonstrated the use of a TAT branch point (branched A is underlined) in a deleted human β-globin transcript.

The position of the branch site relative to the 3' splice site is important. All authentic and cryptic branch sites are twenty-two to thirty-seven nucleotides upstream of the 3' splice site (Padgett et al. 1985; Ruskin et al. 1985; Hornig et al. 1986). The spacing of the branch site from the 3' splice site is probably more important than the primary sequence of the branch site. It is possible to force lariat formation to nucleotides other than adenosine, in a very significant percentage of transcripts, when the authentic branch A is changed by site directed mutagenesis rather than removed by deletion (Hornig et al. 1986). When a lariat is forced to a G or a U the second step of splicing is strongly inhibited.
In summary, it can be said that when the authentic branch site is deleted then a cryptic branch site twenty-two to thirty-seven nucleotides upstream of the 3' splice site is utilized. The transcript is spliced at a reduced efficiency in vitro. Normal levels of spliced RNA are detected in vivo suggesting that the rate of the splicing reactions are not limiting in vivo. When the branch is made to a G or a U then only the second step of the reaction is inhibited with the rate of step one being independent of the branch point nucleotide (Hornig et al. 1986).

1.5 Factors Determining Splice Site Selection.

Perhaps the most fundamental question in splicing is how a 3' and a 5' splice site are correctly spliced together without accidentally skipping exons.

It has been shown by Chu and Sharp (1981) that a chimaeric gene containing the 5' splice site region of the first intron of the SV40 T antigen gene and the 3' splice site region of the second intron of the mouse β-globin gene can be accurately spliced. Also, synthetic splice sites cloned into an RNA pol. II transcription unit can be accurately spliced together in HeLa cells when separated by several different prokaryotic or eukaryotic RNA fragments to provide bulk intron sequence (Rautmann et al. 1984). The above work dismisses the possibility of splice site pairing by an obligate functional relationship between a splice site and its natural counterpart.
1.5.1 A Simple Scanning Mechanism?

A second, very attractive, hypothesis was one of simple linear scanning. A splice site is recognised by a component of the splicing machinery and then the intron is scanned in a linear fashion by tracking along the sugar-phosphate backbone until a corresponding counterpart is found. Both sites are recognised on the basis of their sequence resemblance to the consensus. The two splice sites are then joined together by the mechanism discussed above.

This hypothesis falls down for a number of reasons. First, there are sequences within both exons and introns which conform to the splice site consensus but which are not normally used. Mutations within the authentic splice site sequence can abolish authentic splicing and sometimes allow one or more of these otherwise unused sequences to be used as splice sites. These sequences are known as cryptic sites (reviewed briefly by Mount and Steitz 1983). Mutations which make sequences resemble more closely the splice site consensus sequences can allow the mutated sequences to be used as splice sites. Generation of a new 5' site can interfere with authentic splicing even when the new site is not used. There is an example of this in one of the human thalassaemias where creation of a new 5' site gives rise to a predominant mRNA species in which the normal 5' site is spliced to a cryptic 3' site.

Cryptic 3' splice sites can also be activated by mutations which delete the polypurimidine tract of the authentic site (Wieringa et al. 1984) or which make the mutated sequence resemble a splice site by the creation of an AG (Fukamaki et al. 1982).
As mentioned earlier, deletion of the authentic branch site can activate cryptic branch sites. However, there are no reports of abnormal branch site usage resulting from a sequence being mutated to more closely resemble an authentic branch point sequence.

The evidence of cryptic sites tend to discredit a scanning mechanism because many would be used if selection was brought about simply by choosing the nearest complementary site. Cryptic sites also provide evidence that the primary sequence of the transcript is involved in splice site selection as changing the sequence can influence this selection.

Secondly, the simple scanning model is further precluded by the phenomenon of trans splicing (Konarska et al. 1985b; Solnick 1985a). In this case there is discontinuity of the phosphodiester bonds between the 5' and 3' splice sites.

Solnick (1985a) argues that trans splicing cannot be detected unless the two RNAs are associated by a secondary structure within the intron (i.e. base pairing between complementary strands). The phosphodiester bond can be cleaved before or after the secondary structure has formed.

Konarska et al. (1985b) argue against the necessity for association of the two RNAs through secondary structure even in the presence of a vast excess of other heterosequence RNAs in the nuclear extract, compared to the level of substrate. They argue that each RNA molecule might be recognised as a potential but incomplete splicing substrate and become incorporated into a partial splicing complex. The association of two such incomplete complexes could then lead to the formation of a complete complex and result in splicing. There are conflicting reports about the extent of complex assembly on partial splicing substrates (see section on
complex assembly and references therein). Future work may clarify this point.

There is now evidence (Zaita et al. 1987; Koller et al. 1987) to support trans splicing as a mechanism for pre-mRNA processing in vivo. The gene for the chloroplast ribosomal protein S12 in tobacco consists of three exons. The second and third exons are located 28Kbp downstream from the first exon. All three exons are transcribed in chloroplasts and reverse transcription analysis revealed that trans splicing occurs between the first and second exons in vivo. In this system there are long complementary regions which could form quite extensive secondary structure between the two RNAs. This supports Solnick's (1985a) requirement that the RNAs be associated.

Although trans splicing precludes a simple scanning model for splice site selection, it should be noted that if there is a requirement for secondary structure, then the distance between the two phosphates at the base of a stem is about 17Å which is about three times the the distance between the phosphates of adjacent nucleotides. This perturbation may be insufficient to block tracking if the apparatus moves in steps of more than one nucleotide. Similarly, a stem may not affect the packaging of RNA with proteins. This RNA protein complex could then serve as a supramolecular conduit for tracking by a more complex mechanism.

Thirdly, simple scanning is not consistent with evidence which suggests that splice site selection is influenced by the position of a splice site. It has been shown in vivo that if restriction fragments of about 150 nucleotides are cloned adjacent to their authentic counterparts, either upstream or downstream, in the second intron of the rabbit β-globin gene, then the sites most distal to the intron are always used (Kühne et
Figure 3. - A schematic representation of the alternative splicing patterns of the calcitonin gene. Boxes represent exons; lines represent the splicing patterns; dotted lines represent the two alternative tissue specific splicing patterns as discussed in the text. Two alternative start sites lead to the inclusion of either exon A or exon B, these are non coding exons. Two alternative polyadenylation sites are indicated.

Splicing to include exon CALC leads to the synthesis of calcitonin in the thyroid. Exon CGRP is included by alternative splicing to make calcitonin gene related peptide in the hypothalamus.

Figure 4. - A schematic representation of the developmentally regulated alternative splicing patterns of the troponin T gene (TnT). Part of the gene including the two mutually exclusive α and β exons is shown. Boxes represent exons, lines represent the splicing pattern.

α-TnT is induced in later stages of rat skeletal muscle development and is primarily an adult muscle form of TnT. β-TnT is expressed throughout skeletal muscle development.
fig 3.

fig 4.
al. 1983). Other work in vitro (Reed and Maniatis 1986) also supports the hypothesis that the position of a splice site is an important factor in site selection. This work will be discussed at greater length later.

Finally, simple scanning could not account for alternative splicing. In many genes there is regulation at the level of pre-mRNA processing. This, in some cases, can be due to alternative use of splice sites. This alternative usage may be regulated in a tissue specific manner. An example of this is the tissue specific splicing of the calcitonin gene (Amara et al. 1982; Leff et al. 1987). Alternative splicing and polyadenylation results in the production of two distinct mRNAs (Fig. 3), one, encoding the hormone calcitonin, is predominant in the thyroid, whereas the other encoding, calcitonin gene related peptide (CGRP), predominates in the hypothalamus.

Alternative splicing may also be developmentally regulated, for example the troponin T (TnT) gene (Medford et al. 1984; Breitbart et al. 1985). TnT is a major regulatory protein in the striated muscle that exhibits developmentally regulated structural heterogeneity. Two TnT mRNAs, α and β, are derived from a single fast skeletal muscle TnT gene (Fig. 4). These two RNAs differ by a small internal sequence coding for amino acids 229 to 242. this isoform specific region of the mRNA is coded for by one of two alternatively spliced and mutually exclusive mini exons.

α-TnT is induced in later stages of rat skeletal muscle development and so is primarily an adult muscle form of TnT. β-TnT is expressed throughout skeletal muscle development and is the major form of TnT mRNA in skeletal muscles early in development as well as a major form in adult skeletal muscle.
This regulated alternative splicing may be brought about by some trans acting factor(s) which allows a weak site to be used, perhaps by sequestering a strong site such that it is no longer recognised by the splicing machinery.

There are also instances of alternative splicing which appear to be unregulated, i.e. in the same cell type at the same developmental stage more than one mRNA resulting from an alternative splice of the same transcript type can be detected. An example of this is the α-crystallin gene in some rodents (King and Piatigorsky 1983). In this case a small optional exon is sometimes included in the mature mRNA to give a third α-crystallin polypeptide (αA\(^\text{â""â""â""â""â""")\) (Fig. 5).

It is very likely that even "regulated" alternative splicing becomes unregulated for part of the time. For example, supposing the TnT gene α-splice was repressed early in development, derepression later in development would result in unregulated choice of splice sites in adult stages.

Evidence from cryptic sites, splice site duplications, and naturally occurring alternative splicing suggests that, when unregulated, there is a fine balance between the sequence of splice sites and their positions which determines which site will be chosen.

1.6 Object of Investigation.

The work presented in the first part of this thesis is an investigation into the importance of sequence in 3' splice site selection in mammalian pre-mRNA using a competitive assay system. To this end synthetic 21bp 3' splice site sequences (see materials and methods) were
Figure 5. - A schematic representation of the alternative splicing patterns of the α-crystallin gene as discussed in the text. Part of the gene including the alternatively spliced exon, which is included in αA\textsuperscript{iso}, is shown. Boxes represent exons, lines represent the splicing pattern.

This alternative splicing pattern is neither developmentally regulated nor tissue specific.
fig 5.
inserted into the third exon of the rabbit β-globin gene, in a plasmid derived from pBS'5'SVAg1 II; (pBSV) (Grosveld et al. 1982), at a constant position 49bp downstream of the authentic rabbit β-globin IVS-2 3' splice site. The synthetic sites conform to the 3' splice site consensus sequence (Mount 1982) and allow for variations within it (Fig. 6).

The relative strengths of the synthetic splice sites were to have been measured as percentage use compared to the authentic rabbit β-globin 3' splice site forty-nine nucleotides upstream. This serves as a constant reference site. However, I show here that when transiently expressed in HeLa cells, the twenty-one nucleotide synthetic splice sites are insufficient to generate an alternative splicing pattern, i.e. the site being tested is never chosen when it is in this position relative to the authentic site.

I have further shown that following mutagenesis to the invariant AG at the authentic 3' splice site, the synthetic site is activated at a low level both in vivo and in vitro. This is further investigated in the second part of this thesis.
Figure 6. - The oligonucleotide synthetic 3' splice site sequences are shown. The reasons for each sequence being chosen are stated.
3' Splice Site Oligonucleotides.

3' splice site consensus: (C/T)ₙ(N(C/T))AGₙG

A A T T T T C T T T T T C C T A C A G + C T
p8Aut - A copy of the authentic rabbit β-globin IVS-2 3' splice site.

A A T T T T T T T T T T T T A T A G + G T
p8Acc1 - An all 'T' polypyrimidine tract.

A A T T C C C C C C C C C C C A C A G + G T
p8Acc2 - An all 'C' polypyrimidine tract.

A A T T T T T T T T T T T T A G A G + G T
p8Acc3 - A double 'AG' and a purine at the conserved pyrimidine position.

A A T T T T T T T T T T T C A G + G T A A G T
p8Acc4 - Conforms to both the 5' and 3' splice site consensus sequences.

A A T T G C A A C G T T T T T A C A G + G T
p8Acc5 - Has a short polypyrimidine tract proximal to the 'AG'.

A A T T T T T T G C A A C G A C A G + G T
p8Acc6 - Has a short polypyrimidine tract distal to the 'AG'.

Chapter 2.

2.1 Introduction.

This second introductory chapter is a review of the components involved in, and the assembly of, the splicing complex or "spliceosome". It is an introduction to the second part of this thesis which is a preliminary investigation into the assembly of the spliceosome on a transcript where splicing is to a synthetic, cryptic 3' splice site. This cryptic site is activated by an A+C mutation at the invariant AG dinucleotide of the authentic site.

2.2 Cellular Components Involved in Splicing.

Small nuclear ribonucleoprotein particles (snRNPs) (Bringmann and Lührmann 1986, and references therein.) have long been implicated in splicing, since it was observed that there was sequence complementarity between the 5' end of U1 snRNA and pre-mRNA splice junctions (Lerner et al. 1980). Since the development of cell free splicing systems it has been possible to investigate the role of U1 snRNP and other snRNP and non-snRNP factors in splicing, and gain some insight into the relationships between the spliceosome components and the mechanism of splicing.

U1 snRNP can selectively bind pre-mRNA splice sites in vitro (Mount et al. 1983) and if the first eight nucleotides of U1 snRNP are removed by oligonucleotide directed ribonuclease H (RNase H) cleavage then splicing activity is lost (Kramer et al. 1984; Black et al. 1985). This data is
consistent with the earlier hypothesis, based on sequence complementarity, of a base pairing interaction between the 5' terminus of the U1 snRNA and the 5' splice site of an mRNA precursor. Furthermore, if a HeLa cell nuclear extract is depleted by immunoprecipitation of U-type snRNPs then splicing is abolished (Kramer et al. 1984).

U2 snRNP has also been shown to be involved in splicing (Black et al. 1985). By selectively degrading U2 snRNA by the addition of complementary oligonucleotides and RNAse H Black et al. (1985) showed that the splicing activity of a HeLa nuclear extract was abolished. Splicing activity could be restored in a complementation experiment by mixing the degraded U2 extract with a second extract in which the U1 snRNA had been selectively degraded in the same way.

When a splicing extract was probed using antibodies specific for U1 and U2 snRNPs and then digested with ribonuclease T1, the immunoprecipitated fragments which were inaccessible to T1 included the 5' splice site region and intron branch point region respectively. This is evidence that U2 has a role which includes an association with the branch site. Similarly, U1 snRNP probably associates with the 5' splice site. U1 snRNP is thought to associate with the 5' splice site at zero time and in the absence of ATP. U2 snRNP is not thought to be involved in the initial recognition of the splice sites because the interaction with the branch site is not detected at zero time or in the absence of ATP, implying that U2 is only bound during active assembly of the spliceosome. Purified U2 snRNP will not specifically bind to the branch site and during splicing the region of pre-mRNA protected from T1 digestion is about twice as long as that protected at the 5' splice site by U1 snRNP. These two observations could reflect the binding of some other tightly associated
component(s) to the pre-mRNA or to the U2 snRNP before the association between U2 and the branchpoint can be detected (Black et al. 1985).

By fractionating splicing extracts using various chromatographic techniques (Krainer and Maniatis 1985; Kramer and Keller 1985) it has been shown by complementation analyses that there are at least four components in addition to U1 and U2 snRNPs required for splicing. At least U1 and U2 together constitute what has been termed by Krainer and Maniatis (1985), "splicing factor 1" (SF-1). SF-1 is found both in the nuclear extract and in the S100 fraction. The second component, SF-2, is nuclease resistant implying that it may not have an RNA component, or that if it is a snRNP then the RNA component is not accessible to the micrococcal nuclease. SF-2 is found in the nuclear extract but not in the S100 fraction.

A third component, SF-3, was discovered by differential sensitivity to heat inactivation. When the extract was incubated at 45°C for 10 mins prior to addition of the substrate and cofactors, cleavage at the 3' splice site and exon ligation did not occur. Step one of the splicing reaction was unaffected. Step two activity could be restored by complementation with S100 fraction. SF-3 must therefore be present in both nuclear and S100 fractions and be required for step two of splicing. SF-1, SF-2 and SF-3 are not separable in active form by chromatographic techniques, suggesting they may normally exist as a complex. From the complementation analysis they must be freely exchangeable in the splicing reaction.

Following spermidine-agarose chromatography, SF-1, SF-2 and SF-3 were all found in a snRNP containing 0.3M salt eluted fraction. This fraction alone is inactive in splicing, but can be complemented by the flowthrough fraction. The factor in the flowthrough has been termed SF-4 and is also
present in the S100 fraction. Further fractionation of SF-4 by CM-Sepharose chromatography showed that it consists of at least two separable activities, SF-4A and SF-4B. SF-4B is required for the first step of splicing and SF-4A for the second.

Factors which may correspond to SF-2 and SF-4B have been discovered and further characterised.

A 3' splice site binding component has been discovered (Chabot et al. 1985) which has been shown immunologically to be a U type snRNP other than U1 or U2. Furthermore, its binding site has been shown by T1 nuclease protection to be a nineteen nucleotide region which maps to the 3' end of the intron and includes the 3' splice site polypyrimidine tract and AG dinucleotide. A deletion removing the 3' splice site but not activating any cryptic 3' splice sites also eliminated binding of the component to the intron or exon sequences; this is consistent with no cryptic 3' sites being activated. A mutation at the 3' AG, changing it to GG, reduces efficiency of step one and abolishes step two of the splicing reaction; consistent with this, immunoprecipitation of the component bound 3' splice site was strongly reduced.

Like SF-2 (Krainer and Maniatis 1985), the snRNP which is associated with the 3' splice site is highly resistant to digestion by micrococcal nuclease. The only snRNA which is retained and protected within the snRNP from digestion by micrococcal nuclease is U5. Although this is indirect evidence, it does support the hypothesis that U5 is involved in association with the 3' splice site during splicing (Chabot et al. 1985). It is likely that U5 is SF-2.

A protein or proteins have been assayed by a protein blotting method (Tazi et al. 1986) and by DEAE-Sepharose chromatography (Kramer
and Keller 1985). Kramer and Keller say that their protein has a molecular weight of 50 K\(\text{d}\) and have shown by complementation analysis that it is required both for the generation of the intermediates and the products of the splicing reaction in vitro. It is likely, therefore, that this protein is necessary for the first step of the splicing reaction.

Tazi et al. (1986) have detected their protein by a blotting technique involving electrophoretic transfer of protein from SDS-polyacrylamide gels to nitrocellulose filters followed by analysis by probing the filters using radiolabelled RNA. This protein, termed intron binding protein (IBP), has not been tested for splicing activity by complementation analysis but has been shown to have a molecular weight of 100 K\(\text{d}\) and to bind specifically to the 3' splice site. Furthermore, IBP may have a loose association with U5 snRNP at low Mg\(^{2+}\) concentrations (5mM) which is more evidence for a possible role in splicing.

It is not known whether the two proteins discussed above are the same. Their molecular weights are different but this could possibly be an artefact generated by the different detection procedures. Could the 50K\(\text{d}\) protein be a degradation product or perhaps a subunit of the 100K\(\text{d}\) protein? It is a pity that there is no immunological data to compare the two proteins. If both are the same then taken together the results suggest that this protein may be SF-4B.

This additional factor required for step one of splicing may act by initial recognition of the 3' splice site and then perhaps by stabilisation of the binding of U5 snRNP to the 3' splice site (see discussion).

Finally, it has been shown by specific degradation of U4 and U6 snRNA that U4/U6 snRNP is involved in splicing (Black and Steitz 1986; Berget and Robberson 1986). Cleavage of U4 or U6 using oligonucleotide
directed RNAse H inhibits the appearance of both spliced products and reaction intermediates. The splicing activity can be restored by mixing an extract in which U4 or U6 has been cleaved with one in which U1 or U2 has been cleaved.

Whereas U1, U2 and U5 snRNPs appear to interact directly with the pre-mRNA substrate, the U4/U6 snRNP might not. Cleavage of U4 or U6 snRNA by RNAse H inhibits splicing at an early stage but does not alter the pattern or protected fragments precipitated with anti-Sm antibodies (Chabot and Steitz 1987a). The U4/U6 particle may therefore interact with other snRNPs and perhaps has a role in directing the assembly of the spliceosome complex.

2.3 Assembly of the Spliceosome.

The association of the splicing components with the pre-mRNA results in the formation of a large splicing complex which has been termed the spliceosome.

The assembly of spliceosomes has been investigated by velocity gradient sedimentation (Frendewey and Keller 1985; Grabowski et al. 1985). Frendewey and Keller (1985), using sucrose gradients, showed that spliceosome assembly is probably a multistep process. In a time course experiment they showed that, at zero time, a 22S complex is formed on the RNA, independently of ATP and not requiring the presence of any snRNPs in the extract. After five minutes the level of the 22S complex is greatly reduced and has been replaced by a 35S complex which requires snRNPs but is still ATP independent. By fifteen minutes an ATP dependent 50S complex has formed, there is only a trace of the 22S complex remaining and the
35S complex has greatly reduced. Between fifteen and sixty minutes the
35S complex becomes undetectable; the 50S complex level peaks at twenty
minutes and then slowly decreases over the course of the experiment.

Analysis by polyacrylamide gel electrophoresis of the RNA within the
complexes isolated from gradients showed that the 22S and 35S complexes
contain only pre-mRNA whereas the 50S complex contains pre-mRNA and also
splicing intermediates and products. These results suggest that the
assembly pathway is:

22S→35S→50S.

The 50S complex is probably the fully assembled active spliceosome.

Grabowski et al. (1985) have shown a 40S and a 60S complex by
glycerol gradient sedimentation. The 40S peak contained only precursor
RNA and the 60S peak contained precursor, intermediates and products. It
is very likely that the 60S peak is the active spliceosome and corresponds
to the 50S peak of Frendewey and Keller (1985). Surprisingly, the 40S
complex (Grabowski et al. 1985) was generated at zero time in the absence
of ATP and in the presence of anti-U1 snRNP antibodies indicating that it
may correspond to the 22S (Frendewey and Keller 1985) and not the 35S
peak. Alternatively, Grabowski et al. (1985) suggest that their 40S peak
may be a non-specific association of substrate RNA and cellular material
and may not be related to the splicing process at all. If this is true
then perhaps the 60S peak (Grabowski et al. 1985) may be a mixture of the
35S and 50S peaks (Frendewey and Keller 1985).

To test signals on the transcripts required for spliceosome assembly,
Grabowski et al. (1985) showed that in the absence of any sequences known
to be required for splicing no complex was formed. Both groups tested the
requirement of a 3' splice site for complex assembly using run-off
transcripts which did not include the 3' splice site polypyrimidine tract or the AG dinucleotide. The data is contradictory. Grabowski et al. (1985) report the formation of a 50S complex which, they say, may correspond to their 60S complex, the difference in size being due either to the shorter substrate RNA or to 3' splice site specific components not being present.

Frendewey and Keller (1985) and also Bindereif and Green (1986) report that 3' splice site mutations which prevent splicing also prevent complete assembly. Therefore, in the total absence of a 3' splice site or if the polypyrimidine tract is deleted, then either no complex assembly is detected (Frendewey and Keller 1985) or a 40S complex is detected which probably corresponds to the 35S complex of Frendewey and Keller (1985) (Bindereif and Green 1986). There are no reports of the generation of any splicing intermediates or products from any of the above mutant constructs.

Frendewey and Keller (1985) reported the inefficient formation of a 50S complex on a transcript which had an intact polypyrimidine tract but lacked the AG dinucleotide. In this case, splicing intermediates but no products were found within the 50S complex. Finally, Frendewey and Keller (1985) have shown that on a construct lacking a 5' splice site there is formation of the 35S complex.

Taking all the data together it seems likely that the 3' splice site is initially recognised and a partial complex is formed which sediments at 35S to 40S and does not require ATP. The initial recognition may be by a non snRNP component to give a 20S to 22S complex which is then rapidly built up to the 35S to 40S complex, probably by the addition of components including some snRNPs. This sub-assembly is then built up to the active
spliceosome (50S to 60S) by a process which requires the presence of a 5' splice site and ATP. Disassembly or degradation of the 50S to 60S complex must be by a pathway other than the reverse of assembly as no partially disassembled complexes are detected as the 50S peak is decaying (Frendewey and Keller 1985).

To investigate the components within a spliceosome, material isolated from gradients has been probed with antibodies against the Sm class of U-type snRNPs (U1, U2, U5, U4/U6) and against U1 snRNP on its own (Grabowski et al. 1985; Bindereif and Green 1986). Grabowski et al. (1985) were unable to detect any association between U1 snRNP and labelled RNA in the total extract but were able to precipitate material isolated from a gradient with both anti-Sm and anti-U1. Bindereif and Green could efficiently precipitate material from both their 40S and 60S peaks using both antibodies implying that snRNPs including U1 are present in the spliceosome. Furthermore, Bindereif and Green showed equimolar recovery of the two splicing intermediates from the immunoprecipitations, implying that they are held together in a single RNP complex (the 60S).

Spliceosome components and the interactions between the spliceosome and the pre-mRNA have been further examined by nuclease protection experiments (Ruskin and Green 1985c; Chabot and Steitz 1987a) and by immunoprecipitation of the protected fragments following isolation of spliceosomes from glycerol gradients (Chabot and Steitz 1987a). Prior to cleavage of the pre-mRNA, factors associate specifically with the 3' splice site and branch point and then with the 5' splice site. The factor associating with the branch site has been shown by specific immunoprecipitation to include U2 snRNP. Although the interaction continues after lariat formation, the pattern of ribonuclease protection at

- 30 -
the branch site changes when the pre-mRNA is converted to the RNA lariat molecule. This suggests that either the U2 snRNP or associated components undergo a conformational change during the first step of the splicing reaction, or that after formation of the branch factors may be added to or lost from the spliceosome (Chabot and Steitz 1987a). Deletion or mutation of the 3' splice site greatly reduces the interaction between U2 and the branch site (Chabot and Steitz 1987a; Ruskin and Green 1985c; see also discussion).

The data is contradictory about whether the association of U1 with the 5' splice site is dependent on the presence of a 3' splice site. Chabot and Steitz (1987a) can detect the presence of U1 bound to the 5' splice site independently of whether there is a 3' splice site or not. They point out that the stability of the U1 snRNP-substrate interaction may be sensitive to the presence of the 3' splice site but that the different binding assays respond in different ways. Finally, Chabot and Steitz (1987a) have shown by their immunoprecipitation of protected fragments from isolated spliceosomes that there is interaction between U1 and U2 and possibly U5 snRNPs within the spliceosome.

Fig. 7. shows the model postulated by Chabot and Steitz (1987a) for the interactions of snRNPs with the substrate RNA during the splicing reaction.

2.4 Object of the Investigation.

The work presented in the second part of this thesis is an investigation of the interactions and spliceosome assembly on two rabbit \(\beta\)-globin transcripts. Both these transcripts have a synthetic 3' splice
Figure 7. - A scheme showing how RNP components interact with a splicing substrate \textit{in vitro}. This scheme is as proposed by Chabot & Steitz (1987a).
fig 7.
site, a twenty-one nucleotide copy of the authentic rabbit β-globin IVS-2 3' splice site, inserted at a position forty-nine nucleotides downstream of the IVS-2 3' splice site. In addition, both constructs have a modification to the sequence between the two 3' splice sites which generates a branch point identical to the authentic branch point. This can be utilised during splicing to the synthetic 3' splice site.

I show here that transcripts from one construct (pβ30) are spliced to the authentic 3' splice site in vivo and in vitro. The second construct (pβ30/8/9) has an AG→CG mutation at the authentic 3' splice site and in this case, in vivo and in vitro, I show activation of the synthetic "cryptic" 3' splice site. I also show, in vitro, that when the cryptic site is used, the lariat is formed to the new branch point. I also present immunoprecipitation and RNase H protection data which shows the interaction of factors with the splice sites of both transcripts, and discuss the assembly of spliceosomes as related to cryptic 3' splice site utilisation.
Chapter 3 Materials and Methods.

3.1 Buffers.

3.1.1 Reaction Buffers.

Special reaction buffers are as described in the text. Commonly used buffers were made as a concentrated stock.

10 x H - 500 mM NaCl, 100 mM Tris/Cl pH 7.5, 100 mM MgCl₂, 10 mM DTT.
10 x C - 500 mM Tris/Cl pH 7.5, 100 mM MgCl₂, 10 mM DTT.
10 x RT - 500 mM Tris/Cl pH 8.3, 60 mM MgCl₂, 400 mM KCl.
10 x Seq - 300 mM Tris/Cl pH 7.5, 50 mM MgCl₂.
5 x TMS - 200 mM Tris/Cl pH 7.5, 30 mM MgCl₂, 10 mM Spermidine.
10 x BRL polymerase buffer - (BRL catalogue)
10 x STE - 100 mM Tris/Cl pH 8.0, 1 M NaCl, 10 mM EDTA pH 8.0.

TE.1 - 10 mM Tris/Cl pH 7.5, 0.1 mM EDTA.

3.1.2 Electrophoresis Buffers.

Acrylamide gels contain and are run in 1 x TBE.

10 x TBE - 0.89 M Tris/borate, 0.89 M boric acid, 9.3 g EDTA pH 8.3

(per litre - 108 g Trizma base, 55 g boric acid, 9.3 g EDTA).

Agarose gels contain and are run in 1 x E

50 x E - 2 M Tris/acetate, 0.05 M EDTA pH 8.0 (per litre - 242 g Trizma base, 57.1 ml glacial acetic acid, 100 ml 0.5 M EDTA pH 8.0)
3.1.3 Electrophoresis Dyes.

Acrylamide gels, F-dyes - 90% deionised formamide, 50mM EDTA pH 8.5, 0.25% xylene cyanol (XC), 0.25% bromophenol blue (BPB).

Agarose gels, E-dyes - 6 x E buffer, 30% glycerol, 0.25% XC, 0.25% BPB. 1/6 volume added to sample prior to electrophoresis.

3.2 DNA.

DNA stocks were stored at -20°C in TE-1.

3.2.1 Bacteriophage M13 DNA.

| mp10  | Messing 1983 |
| mp11  | Messing 1983 |
| mICE10| Eperon 1986b |
| mICE11| Eperon 1986b |
| mICE18| Eperon 1986b |
| mICE19| Eperon 1986b |

Other M13 constructs, based on those listed above are as described in the text.
3.2.2 Plasmid DNA.

pBSV Bgl II (pSV) Grosveld et al. (1982)
pBSV3 } Derivatives of above-
pBSV4 } see text.
pBR322

Other plasmid constructs, based on those listed above are as described in the text.

3.2.3 Oligonucleotides.

Universal primer - 5' GTAAAACGACGGCCAGT 3' (Pharmacia).
Other oligonucleotides used are as described in the text.

3.3 Strains of Escherichia coli.

HB101 F−, hsdS20 (r−m−), recA13, ara-14, proA2,
lacY1, galK2, rpsL20 (S+m), xyl-5, mtl-1, supE44,
λ−.
(Maniatis et al. 1982).

JM101 SupE, thi, Δ(lac pro), F′traD36, proAB,
lacI=ΔK15.
(Yamisch-Peron et al. 1985).
BMH71-18  mutLK12, Δ(lac pro), SupE, tri/F', proAB,
lacI^ZAM15, mutL::Tn10.

3.4 Bacterial Growth Media.

TY broth:— 16g bacto-tryptone
10g yeast extract
5g NaCl

made up to 1l distilled water.

SB or standard broth:—
10g bacto-tryptone
5g yeast extract
5g NaCl

Top agar:— broth was made up to 0.7% agar.

Plate agar:— broth was made up to 2% agar.

Broth and agar were sterilised by autoclaving before storage.
E. coli JM101 was maintained on glucose minimal agar made as follows:-

90ml distilled water, 1.5g agar (autoclaved)
10ml 10xM9 salts (autoclaved)
0.1ml 1M MgCl₂ (autoclaved)
0.1ml 100mM CaCl₂ (autoclaved)
1ml 40% glucose (autoclaved)
0.2ml 1mg/ml thiamine (filter sterilised)
0.1ml 10mM FeCl₂ (filter sterilised)
0.2ml 1mg/ml biotin (filter sterilised)

3.5 Transfection of Escherichia coli.

3.5.1 Preparation of Competent Escherichia coli (modified from Cohen et al., 1972).

A single colony was picked into a flask of TY broth and incubated at 37°C with shaking at 300rpm until the culture had reached an A₆₅₀ of 0.5 to 0.7. The bacteria were harvested by centrifugation in sterile 50ml tubes in an IEC Centra 4X at 2.5K rpm for 5mins. The supernatant was discarded and the pellet was gently resuspended in half the original volume of ice cold 50mM CaCl₂. The cells were stored on ice for at least 20mins (up to 5 hours works). The cells were centrifuged as above and the supernatant discarded. The pellet was gently resuspended in one tenth of the original volume of ice cold 50mM CaCl₂.

Cells were now ready for transfection or could be made to 16% glycerol and stored at -70°C. Stored cells were simply thawed before use.
3.5.2 Transfection of Competent HB101.

200µl of competent cells were mixed with plasmid DNA (using 1µg of a known standard as a control). The cells were held on ice for 40mins and then heat-shocked at 42°C for 2mins. 1ml of TY was added and the cells were incubated at 37°C for 1 hour. The bacteria were pelleted by centrifugation for 20secs in a microfuge and the supernatant was discarded by aspiration. The pellet was resuspended in 200µl TY and then spread onto SB-ampicillin (100µg/ml) agar plates using a glass spreader. The plates were incubated at 37°C overnight. Yields of up to $10^6$ transformants were recovered per µg of plasmid.

3.5.3 Transfection of Competent JM101 or BMH71-18.

200µl of competent cells were mixed with M13 DNA (using 1µg of a known standard as a control). The cells were held on ice for 40mins then heat-shocked at 42°C for 2mins. The cells were added to 3ml soft agar at 42°C with 30µl IPTG (24mg/ml in water) and 30µl BCIG (20mg/ml in dimethylformamide) and poured onto SB agar plates. When set, the plates were inverted and incubated overnight at 37°C. Yields of up to $10^7$ plaques were recovered per µg of M13 RF.

Sometimes, when transfecting with closed circular DNA from mutagenesis copying reactions, 200µl SB were added after the heat shock and the mixture incubated at 37°C for 1 hour before plating in the soft agar. This was to allow a round of replication of M13 before plating so that mutant and wild type alleles could segregate.
3.6 Preparation of DNA.

3.6.1 Preparation of Large Amounts of Single Stranded M13 DNA for Mutagenesis.

A single plaque was picked into 10 ml of standard broth and grown up with shaking at 300 rpm at 37°C for 7 hours. The bacteria were removed by centrifugation at 8,000 rpm for 5 mins in an MSE 8x50 rotor at 4°C. Any residual cells were cleared by passing the phage-containing supernatant through a 0.45 μm 'Gelman' acrodisc. A volume of 20% polyethylene glycol (PEG), 2.5M NaCl was added to the filtrate to precipitate the phage and left standing at room temperature for 15 to 20 mins. The phage precipitate was pelleted by centrifugation at room temperature for 5 mins at 8,000 rpm in an MSE 8x50 rotor and then taken up in 400μl of water or TE.1 and transferred to a microfuge tube. The phage were re-precipitated by the addition of a volume PEG/NaCl and subjected to centrifugation for 10 mins in a microfuge after standing at room temperature for 5 to 10 mins. All traces of PEG/NaCl were removed by brief recentrifugation and aspiration. The pellet was taken up in 400μl of water or TE.1 before extracting twice with phenol, chloroform and ether to remove phage protein, leaving the DNA in the aqueous phase. The DNA was recovered by ethanol precipitation. This method usually yielded between 400 and 600μg of single stranded M13 DNA as estimated by agarose gel electrophoresis against known standards.
A colony was picked into a 500ml flask containing 100ml of rich tryptone yeast broth (TY) containing 100µg/ml ampicillin in a 500ml flask and incubated with shaking at 300rpm for about 6 hours until cloudy. The plasmid yield was amplified by the addition of chloramphenicol to a final concentration of 120µg/ml and the incubation with shaking continued for 16 to 20 hours. The bacteria were harvested by centrifugation for 5 minutes at 6,000rpm in an MSE 6x300 rotor at 4°C and the medium discarded. The pellet was resuspended in 1ml of ice cold 50mM glucose, 10mM EDTA, 25mM Tris/Cl (pH 8.0) (early preparations also included 4mg/ml lysozyme but this was later found to be unnecessary) and left at room temperature for 5mins. To this was added 2ml of 200mM NaOH, 1% SDS which was mixed in by vigorous shaking and then left on ice for 5mins. 1.5ml of potassium acetate were added (60ml of 5M potassium acetate, 11.5ml glacial acetic acid, 28.5ml water) and mixed by shaking before standing on ice for 5mins. The mixture was centrifuged at 10,000rpm for 5mins in an MSE 8x50 rotor and the plasmid containing supernatant was passed through a 0.45µ 'Gelman' acrodisc. At this stage one of two procedures was followed, both are detailed below.

a) The volume of the supernatant was measured and CsCl was added at 1.22g/ml. Likewise, 200µl of 750µg/ml ethidium bromide was added per ml of supernatant.

The plasmid preparation was sealed into a 5ml heat seal ultracentrifuge tube and centrifuged overnight at 55,000rpm in a Beckman
Vit 65.2 rotor at 15°C. At the end of the run the rotor was allowed to coast down to stationary.

Plasmid bands were visualised using long wave ultraviolet light and extracted from the tubes with a syringe and 22 gauge needle. The ethidium bromide was extracted from the sample, using isopropanol saturated with CsCl and water, before adding two volumes of water. The DNA was recovered by ethanol precipitation at room temperature for 30mins. Under these conditions the DNA will precipitate but the CsCl will remain in solution. Following ethanol precipitation, the DNA was resuspended in TE.1.

b) The plasmid was precipitated from the supernatant by the addition of 0.6 volumes of isopropanol and standing at room temperature for 30mins, followed by centrifugation at room temperature at 8,000rpm for 10mins in an MSE 8x50 rotor. The supernatant was discarded and the pellet was resuspended in 340µl of TE.1 and transferred to a microfuge tube. To this were added 113µl of 10M ammonium acetate and the mixture was left on ice for 15mins before centrifugation for 5mins in a microfuge. The supernatant was transferred to another microfuge tube and the pellet discarded. The plasmid containing supernatant was precipitated by the addition of two volumes of ethanol and standing at room temperature for 15mins. After centrifugation in a microfuge for 10mins, the supernatant was discarded and the pellet resuspended in 200µl of TE.1. 1µl of 10mg/ml RNase A were added and incubated at 37°C for 15mins. The prep was extracted twice with phenol/chloroform/isoamyl alcohol 25 : 24 : 1 (PCA), once with chloroform and twice with ether before being ethanol precipitated and resuspended in TE.1.
These methods gave up to 200μg of DNA as measured by agarose gel electrophoresis against known standards.

3.6.3 Preparation of M13 Replicative Form (RF) DNA.

A single plaque was picked into 10ml TY broth and grown for 4 hours at 37°C with shaking at 300rpm. This culture was added to 100ml of very dilute JM101 in TY broth and grown for a further 3 to 4 hours. The bacteria were harvested as for the plasmid prep described above then the pellet was washed by resuspension in TY and recentrifugation to remove any contaminating single stranded 'phage.

The procedure was then identical to the plasmid prep described above.

3.6.4 Preparation of Small Amounts of M13 Single Stranded DNA for Mutagenesis Screening and Sequencing (Eperon 1986a).

Small preparations of M13 single stranded (ss) DNA were grown in Nunc 96 well flat bottomed microtitre plates in 250μl of TY broth. Single plaques were picked into individual wells and the plate (without lid) was placed securely in a plastic lunch box. The cultures were incubated at 37°C with shaking at 300rpm for 7 hours. In order to separate the bacteria from the M13 containing broth, the cultures were filtered through a millititre-GV 96 well microtitre filter plate (Millipore) into a recipient flat bottomed microtitre plate (Nunc). To prevent cross contamination of the individual 'phage preparations, an intermediate transfer plate was made from a Nunc 96 well round bottomed microtitre plate with edges cut off and a hole pierced at the bottom of each well.
This was fixed to the bottom of the filter plate with a lattice of silicon rubber adhesive. This assembly was fitted into a millipore vacuum manifold and the filtration carried out by vacuum.

Following filtration, any residual filtrate still in the intermediate transfer plate could be centrifuged into the recipient plate using the appropriate rotor in an IEC Centra 4X bench top centrifuge at 500rpm for 2mins. The intermediate transfer plate was prised from the filter plate and, after cleaning, could be reused several times.

M13 phage from the filtrate were then precipitated by the addition of 50μl of 20% PEG 6,000, 2.5M NaCl (PEG/NaCl) to each well. The wells were sealed with white Sellotape, inverted to mix, and then left at room temperature for 20mins. The precipitate was pelleted by centrifugation at 2,500rpm in an IEC Centra 4X for 20mins. The supernatant was discarded by aspiration and any remaining PEG/NaCl removed by brief re-centrifugation and aspiration.

The precipitated phage were lysed by resuspending them in 40μl TE.1 and 0.5% SDS, covering the plate with Nescofilm and floating it on a 80°C non-circulating waterbath for 10mins.

The ssDNA was recovered by ethenol precipitation for 30mins on dry ice. The precipitate was pelleted by centrifugation at 2,500rpm in an IEC Centra 4X for 30mins. The pellets were washed and then dried briefly under vacuum before being resuspended in 50μl TE.1.

This method yielded sufficient single stranded DNA in each well for 25 sequencing tracks.
3.6.5 Micropreps - Analytical Scale Microtitre Plate Plasmid Preps.

For the analysis of large numbers of recombinant plasmids, a method was developed for simultaneously preparing DNA from a large number of clones in a Nunc 96 well, flat bottomed microtitre plate in 200μl TY broth containing 100μg/ml ampicillin.

Single colonies were picked into individual wells and the plate (without lid) was placed securely in a plastic lunch box. The cultures were incubated for 7 hours at 37°C with shaking at 300rpm. The bacteria were harvested by centrifugation at 2,000rpm for 5mins in an IEC Centra 4X bench top centrifuge and the supernatant discarded by aspiration. The pellet was resuspended in 50μl of ice cold 50mM glucose, 10mM EDTA, 25mM Tris/Cl (pH 8.0) by vortexing the plate. The plate was left at room temperature for 3mins before adding 100μl of 200mM NaOH, 1% SDS and mixing gently. The plate was left on ice for 5mins then 75μl ice cold potassium acetate (60ml 5M potassium acetate, 11.5ml glacial acetic acid, 28.5ml water) were added, mixed gently, and left on ice for 5mins. The chromosomal DNA was removed by filtration through a Millititre GV 96 well microtitre filter plate (Millipore) into a recipient plate (Nunc) fixed to the bottom of the filter plate by a lattice of silicon rubber adhesive. The filtration was achieved by centrifugation at 2,500rpm in an IEC Centra 4X until all the plasmid containing filtrate had collected in the recipient plate. The recipient plate, containing the filtrate, was prised from the filter plate and the filter plate discarded.

The DNA was precipitated from the filtrate by the addition of 100μl cold isopropanol and freezing on dry ice for 10mins. The precipitate was pelleted by centrifugation at 2,500rpm in the IEC Centra 4X for 30mins and
the supernatant removed by aspiration. Residual protein was removed by resuspending the pellets in 50\(\mu l\) TE.1, 0.5% SDS, covering the plate with Nescofilm and floating it on an 80°C non-circulating water bath for 15 minutes. The plasmid DNA was recovered by ethanol precipitation as described for the M13 ssDNA microtitre plate preps. Finally, the pellets were resuspended in 25 to 50\(\mu l\) TE.1. This method yields at least 0.5\(\mu g\) DNA per well.

Contaminating RNA could be reduced by including 20\(\mu g/ml\) RNase A in the TE.1 used for the final resuspension and incubating at room temperature for 15mins.


Occasionally, DNA was prepared by a conventional miniprep method.

A single colony was picked into 1ml TY broth in a 2ml culture tube and incubated at 37°C for 7 hours, shaking at 300rpm. The culture was transferred to a microfuge tube and the bacteria harvested by centrifugation for 1min. The medium was removed by aspiration and the pellet resuspended in 100\(\mu l\) of ice cold 50mM glucose, 10mM EDTA, 25mM tris/Cl (pH 8.0) and stored for 5mins at room temperature. 100\(\mu l\) of 200mM NaOH, 1% SDS were added and mixed by shaking the tube two or three times. The tube was left on ice for 5mins before the addition of 150\(\mu l\) ice cold potassium acetate (60ml 5M potassium acetate, 11.5ml glacial acetic acid, 28.5ml water). This was mixed in by gentle vortexing then stored on ice for 5mins. The chromosomal DNA was removed by centrifugation for 5mins in a microfuge and 400\(\mu l\) of the supernatant
containing the plasmid DNA was transferred to a new tube. The material was extracted twice with phenol, chloroform and ether then the plasmid DNA was recovered by ethanol precipitation at room temperature for 20 mins. The precipitate was pelleted by centrifugation for 10 mins in a microfuge and finally resuspended in 50μl TE.1 containing 20μg/ml RNase A.

This method yields 2 to 5μg of plasmid DNA per prep.

3.6.7 Synthetic Oligonucleotides.

3.6.7.1 Preparation. Synthetic oligonucleotides were prepared either by John Keyte and Jim Turner using the paper disc method (Matthes et al. 1984) or by Jim Turner using an Applied Biosystems 380B automatic DNA synthesiser.

3.6.7.2 Purification. The oligonucleotides prepared by the paper disc method were dissolved in 100μl TE.1 and ethanol precipitated. The pellets were resuspended in 10μl TE.1 and 10μl F-dyes, heated at 80°C for 10 mins and then centrifuged briefly to pellet any residual paper. The supernatants were electrophoresed on a 15% polyacrylamide, 7M urea gel until the bromophenol blue (BPB) was at least 1/3 down but not off the bottom.

The gels were wrapped in Saranwrap®, placed over a fluorescent thin-layer chromatography (TLC) plate, and the oligonucleotide bands were detected by shadowing with short wave ultra-violet light. The bands were excised and the DNA eluted from the gel slices in 400μl distilled water at 4°C overnight. The purified oligonucleotides were recovered from the water
by ethanol precipitation and resuspended in 10 to 100μl of TE.1. The yield ranged from 0.25 to 2.9nMoles.

The oligonucleotides made on the DNA synthesiser were diluted to 600μl with TE.1 then recovered by ethanol precipitation. The pellets were resuspended in 300μl TE.1 to give a final concentration of 150 to 200 pMoles/μl as determined by measuring the A$_{260}$.

3.7 Ethanol Precipitation of DNA or RNA.

Nucleic acids were precipitated from aqueous solution by the addition of one tenth volume of 3M sodium acetate and 2.5 volumes of cold 100% ethanol (-20°C). The solution was mixed well and stored for at least 20mins either on dry ice, or at room temperature if it was necessary to avoid precipitation of salts or SDS from the solution. The precipitate was pelleted by centrifugation in a microfuge for 10mins or in an MSE 8x50 rotor for 15mins at 8,000rpm. Ethanol precipitates in microtitre plates were centrifuged in an IEC Centra 4X centrifuge at 2,000 to 2,500rpm for 30mins. The supernatant was aspirated and the pellet washed in ethanol and then dried under vacuum.

Sometimes, when very small amounts of material were to be precipitated, yeast tRNA$^{sp}$ was added as a carrier to aid precipitation.

3.8 Phenol Extraction of DNA or RNA solutions.

Phenol was either freshly redistilled (Fisons) or HPLC grade (BDH) and was equilibrated twice with 100mM Tris/Cl pH7.5 and then with TE.1 and stored at -20°C. DNA solutions were phenol extracted to remove
contaminating proteins. In general, volumes were made up to 200µl with water or TE.1 and were extracted by the addition of an equal volume of phenol. The DNA solution/phenol mixture was vortexed and then centrifuged in a microfuge for 5mins. The aqueous phase (upper) was removed and could be re-extracted with phenol and then with chloroform/isoamylalcohol (24:1) (chloroform extraction). Extractions were repeated until any protein interphase had disappeared. Residual phenol was usually removed by extracting twice with ether saturated with TE.1. Excess ether was evaporated by blowing air over the surface of the solution.

Sometimes, where mentioned in the text, DNA solutions were extracted by addition of an equal volume of phenol, chloroform, isoamylalcohol (25 : 24 : 1) (PCA extraction). The procedure was identical to that for phenol extractions.

3.9 Subcloning Small DNA Fragments into Plasmid (or M13RF) Vectors.

DNA fragments with cohesive ends were subcloned into plasmid vectors using T4 DNA ligase.

200ng of linearised plasmid (or M13 RF) DNA (phosphatase treated if desired) was mixed with equimolar to three fold molar excess of the fragments to be subcloned in a total volume of 10 to 20µl of 1×C buffer and 1mM ATP. 0.1 units of ligase was used per ligation and incubation was at room temperature for at least 3 hours or at 16°C overnight. The ligated material was directly transfected into competent cells and the recombinants were detected by microprep screening or by colony hybridisation.
When oligonucleotides were to be cloned into vectors, they were first phosphorylated at the 5' termini by incubation of an equal volume of each strand of the oligonucleotide (approx. 60 pMoles each) with 1 mM ATP and 1 unit of T4 polynucleotide kinase in a total volume of 10 µl of 1xC buffer for 45 mins at 37°C. The reaction was stopped by incubation at 70°C for 10 mins and then cooled to room temperature for 20 mins to allow the strands of the oligonucleotide to anneal together. The prepared oligonucleotide was then cloned as described above.

3.10 Restriction Digests.

Restriction digests were carried out in a suitable reaction buffer depending on the enzymes' salt requirements. Common reaction buffers were made as 10x stocks and stored at -20°C. Usually, 1% to 2 units of enzyme were used to digest 1 µg of DNA. Incubation was usually at 37°C for 1 to 2 hours unless specified otherwise in the text. Reactions were stopped either by phenol extraction or by incubation at 70°C for 10 mins.

If it was required to remove the 5' terminal phosphate groups of the digested fragments, then 12 units of calf intestinal phosphatase were added after 1 hour and incubation at 37°C continued for a further 40 mins. In this case, the reaction was stopped by the addition of:

2.5 µl 10% SDS
5.0 µl 10x STE
x µl TE.1 to a total volume of 50 µl

and incubation at 70°C for 15 mins. The solution was then extracted twice with phenol, chloroform and ether and the DNA recovered by ethanol precipitation.
3.11 Colony Hybridisations.

When the Bam H1 Sal I fragment from any mICEβ construct was cloned into pBSV4, positive recombinants were often detected by colony hybridisation. A universally labelled probe was made by second strand synthesis of the single stranded mICEβ construct, labelled with α²³²PdATP, followed by digestion with Bam H1 and isolation of the probe from LMP agarose. The exact procedure is detailed below.

3.11.1 Preparation of Probe.

5µg (approx. 2pMoles) of mICEβ ssDNA were annealed to an equimolar amount of universal primer by incubation in a total volume of 10µl of 2×H buffer at 80°C for 3mins and then at room temperature for 10mins.

To this was added:

3µl of a mix of dCTP, dGTP, dTTP each at 2.5mM
2µl (20µCi) α²³²PdATP
5 units (0.6µl) Klenow polymerase
2µl TE.1

This was incubated for 15mins at room temperature and then the extension was chased by the addition of 0.5µl of 10mM dATP and incubation at 37°C for 10mins. The Klenow polymerase was then inactivated by incubation at 70°C for 10mins. The double stranded material was cleaved by the addition of 50 units (2µl) of Bst I (an isoschizomer of Bam H1 and incubation at 37°C for 30mins. The probe was denatured by the addition of 5µl of F-dyes and incubation at 100°C for 2mins. To prevent reannealing, the
probe was immediately transferred to ice for 2 to 3mins before electrophoresis on a 1% LMP agarose gel.

The probe band, the band below the smear, was localised using a short wave, hand held ultraviolet lamp and excised from the gel. The gel fragment containing the probe was placed in a microfuge tube with 500µl hybridisation buffer (5×SSC, 10×Denharts, 33% formamide) and heated to 80°C for 10mins. The probe was now ready for immediate use.

3.11.2 Preparation of Filters for Probing.

To lift colonies, circular nitrocellulose filters were laid on the culture plates for 30 to 60 seconds and then removed and kept face up. The culture plates were stored at 4°C.

Solutions prepared for colony hybridisation:

1) 10% SDS
2) 0.5M NaOH, 1.5M NaCl
3) 1.5M NaCl, 0.5M Tris/Cl pH8.0
4) 2×SSC

To use these a sheet of Saranwrap™ was laid on the bench and a little of each solution pipetted onto it. A 0.5ml drop of the first solution was used and 0.75ml drops of the other three. The filter was laid face up on each drop in turn, care being taken that the liquid did not spill onto the upper, colony bearing face.
The filter was placed in solution:

1) for 3 mins, then to
2) for 5 mins, then to
3) for 5 mins, then to
4) for 5 to 15 mins

then dried in air.

The filter was baked for 3 hours at 80°C between two sheets of Whatman 3mm paper and then prehybridised at 65°C in 20ml of 6xSSC, 10xDenharts, 0.2% SDS, 20µg/ml sonicated salmon sperm DNA. The prehybridisation was carried out with gentle shaking for about 4 hours in a perspex hybridisation box.

The probe was added to 9ml of hybridisation buffer (5xSSC, 10xDenharts, 33% formamide) in a second hybridisation box. The prehybridised filters were washed in 5xSSC (three times 1 min washed to remove excess salmon sperm DNA) then lowered gently into the hybridisation box with the probe. Hybridisation was carried out overnight with gentle shaking at 40 to 42°C.

The probe was removed from the filters by aspiration and the filters were washed to remove excess probe as follows:

2 washes in 6xSSC at room temperature for 1 min each.
1 wash in 6xSSC at 65°C for 30 mins
2 washes in 6xSSC at room temperature for 1 min each.
1 wash in 0.2xSSC at 42°C for 1 hour.
2 washes in 0.2xSSC at room temperature for 1 min each.

Finally, the filters were dried in air and analysed by autoradiography.
3.12 Electrophoresis.

3.12.1 Polyacrylamide Gels.

3.12.1.1 Acrylamide Preparation.

380g acrylamide, 20g N,N'-methylenebisacrylamide were made up to 1l with deionised water. Deionised with amberlite MB-1 resin (BDH) for 1 hour and then filtered through Whatman 3mm paper and stored at 4°C.

3.12.1.2 Denaturing Gel Preparation (containing 7M urea and, sometimes, 20% formamide).

For an x% acrylamide gel (up to 22% or 14% with formamide), xml 40% acrylamide, 16.8g urea, (8ml deionised or electrophoresis grade formamide), 4ml 10×TBE, 320μl 10% ammonium persulphate and deionised water up to 40ml. Cover and stir until dissolved then add 50μl TEMED (N,N,N',N'-tetramethylenediamine, Serva). 100μl of TEMED are used if the gel contains formamide. Mix gently before pouring between gel plates (20cm×40cm×4mm) separated by spacers (0.38mm thick plasticard™ along the sides and taped together with white Sellotape™). Gels could be stored, wrapped in Saranwrap™ for up to two days. Before use, the gel was untaped and the slots were flushed out with 1×TBE. Gels were normally run at 1,350 to 1,700 volts.
3.12.1.3 Sample Preparation.

Sample preparation is described in the relevant sections. The denatured samples were loaded using a drawn out capillary tube.

3.12.1.4 Gel Handling.

Gels could be transferred to the absorbant side of Benchkote™ covered in Saranwrap™ and exposed to x-ray film (Fuji RX or Kodak XAR) at -70°C.

Alternatively gels could be fixed by soaking the gel in 10% ethanol, 10% glacial acetic acid for 15mins. Gels were then transferred to Whatman 3mm paper, covered in Saranwrap™ and dried at 80°C under vacuum before being exposed to x-ray film at -70°C or at room temperature. Gels containing material labelled with ³⁵S had to be fixed and dried before exposure to x-ray film.

3.12.2 Agarose Gels.

3.12.2.1 Gel Preparation.

Agarose gels were used at 0.8% to 1.3% in 1×E buffer. Agarose was melted in 1×E buffer then ethidium bromide (EtBr) was added to 50µg/ml before the gel was poured. Gel plates were taped round the edges with white Sellotape™. BRL mini gel apparatus was used routinely and required about 20ml of gel mix. The gels were allowed to set for at least 1 hour before use. Minigels were run at 13mA and 75ml gels were run at 30 to 50 mA.
3.12.2.2 Sample Preparation.

Samples in 1×E buffer were loaded with a drawn out capillary tube.

3.12.2.3 Gel Handling.

DNA bands were visualised using an ultraviolet transilluminator or a hand held ultraviolet lamp and were photographed using Polaroid type 57 film with the following conditions:

Transilluminator - F4.5/0.5secs.
Hand held lamp - F4.5/10 to 30secs

3.12.2.4 Low Melting Point (LMP) Agarose.

LMP gels were used when the DNA bands were to be excised from the gels. The gels (BRL ultrapure LMP agarose) were poured at 4°C and allowed to set for at least two hours. Bands were observed using a hand held long-wave (300 to 360nm) ultraviolet light to minimise damage to the DNA.

Gel slices were cut as small as possible with a clean scalpel blade and placed into microfuge tubes. For transfection of the DNA into E. coli, the gel slice was diluted, melted and used directly. If required for manipulations, the DNA was hot phenol extracted from the agarose.

3.12.2.5 Hot Phenol Extraction.

An equal volume of TE.1 was added to the gel slice which was then incubated at 65°C for 15mins. An equal volume of phenol was added and
immediately mixed by vortexing before standing at room temperature for 5 mins. The solution was again vortexed and centrifuged for 5 mins in a microfuge. The aqueous phase was removed and further phenol extracted, then extracted with PCA until no interphase remained. The aqueous phase was then ether extracted twice and the DNA recovered by ethanol precipitation.

3.13 Dideoxy Sequencing of DNA Cloned into Bacteriophage M13.

Bacteriophage M13 DNA was sequenced by the dideoxy method of Sanger et al. (1977). This involved annealing an oligonucleotide "primer" to the single stranded form and extending from the 3' end of the primer using DNA polymerase 1 large fragment (Klenow polymerase). Four reactions were performed, namely C, A, G and T, each containing the four dNTPs, one of which was at a low concentration. The corresponding dideoxyribonucleoside triphosphate (ddNTP) was included and competed for incorporation on the extending 3' end. Once incorporated, the ddNTP prevented further extension and was resistant to the 3'-5' exonuclease activity of the Klenow polymerase. The products were made radioactive either by the use of a phosphorylated primer labelled with $^{32}$P at the 5' end, or by incorporating either $\alpha^{32}$PdATP or $\alpha^{35}$SdATP into the reaction mixes and excluding unlabelled dATP. The four reactions were analysed by electrophoresis on denaturing polyacrylamide gels.
Mixes required for dideoxy sequencing (stored at -20°C):

ddATP 0.04mM in TE.1
ddCTP 0.25mM in TE.1
ddGTP 0.16mM in TE.1
ddTTP 0.50mM in TE.1

A* 200μl 0.5mM dCTP + 200μl 0.5mM dGTP + 200μl 0.5mM dTTP
C* 10μl 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

α<sup>32</sup>PdATP 4,000 Ci/mMole
α<sup>32</sup>SdATP 1,000 Ci/mMole

Each reaction required 1μl of N* + 1μl ddNTP + 1μCi α<sup>32</sup>PdATP or 2μCi α<sup>32</sup>SdATP where appropriate. To terminate the elongation close to the primer, correspondingly higher ratios of ddNTP : N* were used. When using 5' labelled primers, unlabelled dATP was added to the reaction mixes at the same concentration as the usual labelled dATP.

Sequencing reactions were performed in Nunc flat bottomed, 96 well microtitre plates. When sequencing templates that had been prepared in microtitre plates, 2μl of ssDNA were used for each of the four sequencing reactions. For each reaction 0.01 to 0.04pMoles of template and 0.03pMoles of primer were annealed at 65°C in 2μl of 2xsequencing buffer for 5mins. This was performed by covering the wells of the microtitre plate and placing the plate above a 70°C water bath. The primer/template mixes were allowed to cool to room temperature for at least 15mins.
To each reaction was added:

\[ 2\mu l (\text{ddNTP} + \text{ddNTP}) \]

plus either \[ 2\mu l (\alpha^{32}\text{PdATP 0.5}\mu \text{Ci}/\mu l + 0.25\text{units Klenow polymerase}) \]
or \[ 2\mu l (\alpha^{35}\text{SdATP 1.0}\mu \text{Ci}/\mu l + 0.25\text{units Klenow polymerase}) \]

The Klenow polymerase and labelled dATP (where appropriate) were diluted in ice cold 25% glycerol, 25mM KPO\(_4\) pH 7.5 (Klenow diluent).

The sequencing reactions were incubated at room temperature for 20mins, 2\(\mu\)l of a 'chase' mix (0.5mM each dNTP) was added and then the reactions were incubated at room temperature for a further 15mins.

The reactions were stopped by adding 4\(\mu\)l of F-dyes and incubating the plate for 10mins at 80°C with the lid off. This denatured the DNA and reduced the volume by evaporation. The samples were then analysed by electrophoresis on a 7\(\times\) urea, polyacrylamide gel.

### 3.14 Oligonucleotide Mutagenesis

Mutations were made to the rabbit \(\beta\)-globin gene fragment cloned in M13 M1CE10 using a variation of the method of Zoller & Smith (1983).

A mismatched oligonucleotide was annealed to the single stranded template and extended around the circle using Klenow polymerase. Ligase seals the growing 3' end to the 5' end of the oligonucleotide. The closed circles were isolated by eluting the band, following electrophoresis through ethidium containing agarose gels, and transfected into JM101. The plaques were screened for mutants by single track sequencing and looking for the extra band or bands on the autoradiograph when compared to the wild type sequence.

Oligonucleotide primers were supplied with 5’ hydroxyl ends and required phosphorylation before use in mutagenesis. The phosphate donor was the γ-phosphate of rATP and the reaction was catalysed by T4 polynucleotide kinase in 1×C buffer. The reaction was:

- 2.5mMoles oligonucleotide
- 1µl 10×C buffer
- 1µl 10mM rATP
- 1unit kinase
- 7µl water

Incubation was for 1 hour at 37°C.

3.14.2 Second Strand Synthesis of M13 DNA ("Copying Reaction").

By varying the amount of phosphorylated oligonucleotide primer used in different annealing and copying reactions, conditions could be established where all the available ssDNA template was copied.

Annealing conditions: 0.2mMoles ssDNA template

- 0 to 0.6mMoles phosphorylated primer

Total volume was 3µl and 2×Seq. buffer.
Incubation was for 5 mins at 65°C and then cooled to room temperature for 20 mins. To this was added 6 μl of a "copy mix":

1 μl of a mix of all dNTPs, each at 2.5 mM
0.5 μl 10 mM rATP
0.5 μl 10×Seq. buffer
0.5 units T4 ligase
0.5 units Klenow polymerase
distilled water to 6 μl

This mix was added to the 3 μl primer/template and incubated at 25°C overnight. More recently, copy reactions were carried out in 1×TMS buffer at 37°C for 1 hour. The reactions were stopped by the addition of 0.5 μl 0.5 M EDTA and 2 μl E-dyes. The closed circles were isolated by electrophoresis through a 0.8% low melting point (LMP) agarose gel containing ethidium bromide. The gel was examined under short wave ultraviolet light and the closed circle band excised. The gel slice was placed in a microfuge tube with 10 volumes of distilled water and heated to 65°C for 15 mins to melt and dilute the agarose. The melted agarose was vortexed for 5 seconds and then allowed to cool to room temperature. 1% and 10% of the dilute agarose containing the DNA was used to transfect competent JM101 or BMH71-18. Plaques were picked and screened for mutants by single track sequencing.
3.15 Screening for and Determination of Orientation of Oligonucleotides cloned into pBSV or pBSV3.

3.15.1 Screening for insert.

Possible recombinants were grown up in a Nunc Flat bottomed, 96 well microtitre plate and plasmid DNA was prepared by the microprep method. The DNA was analysed for the presence of an inserted oligonucleotide by restriction mapping.

5µl of DNA from each well were transferred to a second microtitre plate and digested with Bgl II so each well contained:

- 5µl DNA
- 1µl (2 units) Bgl II
- 1µl 10×H buffer
- 4µl TE.1

The wells were covered with Nescofilm™ and incubated for 1 hour at 37°C. The reaction was stopped by incubation at 70°C for 10mins. In order to detect the restriction fragments by autoradiography, the Bgl II sticky ends were labelled by filling in using Klenow polymerase in the presence of cold dCTP, dGTP, dTTP and α²³PdATP. The following mix was made for n reactions and 2µl per well were added to the solutions in the microtitre plate:

- nµl of a mix of dCTP, dGTP, dTTP each at 1mM
- n×µl α²³PdATP (at 10µCi/µl)
- n units Klenow polymerase

make total volume to 2n µl with TE.1
Incubation was for 5 mins at room temperature followed by 70 °C for 10 mins to stop the reaction.

*Bgl II* has a unique site in pBSV or pBSV3, so to make the fragments small enough to resolve on a polyacrylamide gel, the DNA was further digested with *msp 1*.

The *msp 1* was diluted in TE.1 to about 0.5 units/µl and 1 µl was added to each well of the microtitre plate and incubated for 1 hour at 37 °C. The reaction was stopped by the addition of 4 µl F-dyes and incubation at 70 °C for 10 mins. The digests were analysed by electrophoresis on a 6% polyacrylamide, 7 M urea gel until the xylene cyanol (XC) was at the bottom.

### 3.15.2 Checking for Orientation of Insert

The orientation of the inserted oligonucleotide was determined by primed synthesis using one strand of the oligonucleotide as a primer annealed to its complementary strand in the recombinant. The primer was extended by Klenow polymerase and dCTP, dATP, dTTP were incorporated. The mix contained no dGTP but instead contained ddGTP. The ddGTP was incorporated opposite the first C in the template and the chain elongation was terminated. If the position of the first C in either direction was known then the orientation of the insert could be determined by sizing the extended products on a sequencing gel. In order to examine the gel by autoradiography, the primer was 5' labelled with ³²P from Y³²PATP in a reaction catalysed by T4 polynucleotide kinase.
3.15.2.1 Phosphorylation of Primer.

100nMoles of primer were phosphorylated using 10µCi of $^32$PATP by 1 unit of polynucleotide kinase in a total volume of 10µl of 1xC buffer. Incubation was at 37°C for 45mins.

3.15.2.2 Preparation of Template DNA.

Positive recombinants from the initial screen were transfected into competent HB101. Transfectants were cultured and DNA was prepared by the alkaline lysis miniprep method described elsewhere in this section. The miniprep DNA from a 1ml culture was resuspended in 200µl TE.1.

3.15.2.3 Procedure.

The DNA first had to be linearised and the strands melted to allow access to the primer. 2µl of DNA was transferred to a 1ml microfuge tube and digested by 1 unit of Bal1 in a total volume of 5µl of 1xC buffer for 1¾ hours at 37°C. To this was added 2µl of $^{32}$P labelled primer and the mixture was incubated at 100°C for 3mins, then immediately placed on ice for 10mins before being allowed to return to room temperature for 20mins. An extending mix for n tubes was made and 4µl of this mix were added to
each tube and incubated at room temperature for 7 mins. The extending mix is as follows:

- 2×n μl of a mix of dCTP, dATP, dTTP, each at 0.5 mM
- 1.1×n μl 10×Seq buffer
- n units Klenow polymerase
- n μl 0.16 mM ddGTP
- make total volume to 4×nμl with distilled water.

2 μl of a "chase" mix (dCTP, dATP, dGTP, dTTP, each at 0.5 mM) was added and incubation at room temperature continued for a further 10 mins. The reaction was stopped by the addition of 6 μl F-dyes and the samples were denatured by incubation at 80°C for 10 mins before electrophoresis on a 12% polyacrylamide, 7 M urea gel until the BPB was at the bottom.

3.16 Cell Culture.

3.16.1 Maintenance of Monolayer Cells.

Monolayer HeLa cells (Flow Laboratories) were maintained for in vivo splicing experiments. Plastic ware was all Nunclon®, media and sera were from GIBCO.

The cells were grown in 25 cm² or 80 cm² tissue culture flasks in Dulbecco's Modified Eagle Medium (DMEM) containing 2.5% fetal calf serum and 2.5% newborn calf serum, 2g/l sodium bicarbonate, 10,000u/l penicillin and 10,000μg/l streptomycin. The cells were fed by replacing the medium every 2 to 3 days and split up to a forty fold dilution when they became confluent. Prior to transfecting, the cells were split into 9 cm diameter dishes for ease of manipulation.
3.16.2 Maintenance of Suspension Cells.

HeLa cells in suspension (JS1000, a gift from Dr. J.A. Steitz) were maintained for making nuclear extracts for in vitro splicing experiments. The cells were grown in 250ml or 1l 'Techne' microcarrier stir bottles in DMEM containing 10% fetal calf serum, 3.7g/l sodium bicarbonate, and antibiotics as above. The cells were stirred continuously at approx. 60rpm on a 'Techne' microcarrier stirrer and were maintained at a density of between 3 and 6x10^6/ml by daily splitting. The cells were not permitted to fall below the critical density of 1x10^6/ml. Prior to making an extract, the cells were split by daily doubling of the volume to a final volume of 3l.

3.16.3 Cell Stocks.

Stocks of both cell lines were stored frozen in 20% fetal calf serum, 10% DMSO in DMEM, under liquid nitrogen. Frozen cells, once thawed and washed in fresh medium, had to be grown for the first 2 to 3 days in medium containing up to 20% fetal calf serum to allow them to recover. The suspension cells were initially grown in 25cm^2 and 80cm^2 tissue culture flasks until they had reached approx. 1.5 to 2x10^5/ml and could be transferred to a 250ml stir bottle.
3.17 Transfection of HeLa Cells for in vivo splicing.

3.17.1 Solutions.

2×HBS - 4g NaCl
- 0.185g KCl
- 0.05g Na₂HPO₄ (anhydrous)
- 0.5g glucose
- 2.5g Hepes

in 250ml distilled water; pH to 7.5, filter sterilise

CC⁺ - 100ml (1mM Tris/Cl pH 7.5, 0.1mM EDTA)
15ml 2M CaCl₂

3.17.2 Procedure

Cells to be transfected were seeded in a 9cm diameter dish.

The cells were split at least 24 hours before use so that they were about 50% confluent at the time of transfection. The medium was changed 4 hours prior to transfection so that the pH at the time of transfection would be optimal. The transfection solution was set up in a universal as follows:

400µl ice cold CC⁺
20µl (20µg) DNA to transfect
420µl room temperature 2×HBS

The mix was swirled then left on ice for 8 to 10mins before dripping onto the cells freshly taken from 37°C. The cells were immediately
returned to the incubator. 24 hours later the cells were shocked with DMSO, the medium was removed, replaced with about 10ml of 25% DMSO in DMEM and left to stand for 3mins before being removed. The cells were washed twice with fresh medium, new medium was replaced and the cells were returned to the 37°C incubator.

The RNA was harvested 16 to 24 hours following the DMSO shock.

3.18 Harvesting of Whole Cell RNA from HeLa Cells.

3.18.1 Solutions.

Trypsin - Trypsin-EDTA (GIBCO) contains 0.5g trypsin (1:250) and 0.2g EDTA/litre.

LiCl/Urea - 12.7g LiCl + 36g Urea in 100ml water.

3.18.2 Harvesting.

The medium was removed and 400μl of trypsin were added to each 9cm diameter dish and incubated at 37°C until the cells were seen to "round up". The cells were taken up in 2ml of ice cold LiCl/urea and placed in a 15ml 'Corex' tube on ice. The dish was washed in a further 2ml of ice cold LiCl/urea. Following sonication in an MSE sonicator set to medium power, amplitude 5, with a 2μ peak; for 3x40 second bursts with 20 second rests in between, the cells were left standing on ice overnight. The RNA was pelleted by centrifugation at 10,000rpm for 30mins in an MSE 8x50 rotor at 4°C and then washed by partial resuspension and recentrifugation before being resuspended in 250μl of 0.5% SDS in TE.1. The RNA was
chloroform extracted twice, ether extracted twice and then stored after the 
addition of one tenth volume sodium acetate + 2½ volumes ethanol, at 
-70°C.

3.19 Nuclease S1 Mapping of HeLa Cell RNA.

3.19.1 Preparation of Probes.

The S1 probes used for this study were 5' end labelled restriction 
fragments made as follows. The restriction fragment used was the small 
Bam HI Bgl II fragment from the relevant mICEβ construct e.g., mICEβ30. 
5µg of DNA was used for each probe. The restriction digest was set up as:

5µl DNA
1µl (10u) Bam HI
1µl (10u) Bgl II
1µl 10×H
2µl TE.1

After digestion for 1 hour at 37°C, 12u of calf intestinal 
phosphatase (Boehringer Mannheim) were added and incubation continued for 
a further 40mins in order to remove the 5' terminal phosphate groups from 
the fragments. The reaction was stopped by the addition of:

40µl TE.1
2.5µl 10% SDS
5µl 10×STE

and incubation for 15 minutes at 70°C. The DNA was then extracted twice 
with phenol, chloroform and ether; ethanol precipitated and resuspended in 
10.5µl TE.1.
The probes were labelled by replacing the 5' terminal phosphates with \(^{32}\)P phosphate, from \(^{32}\)P-ATP, with polynucleotide kinase (PNK).

PNK prefers to label single stranded DNA so the DNA in TE.1 was melted by heating to 100°C for 30secs and cooled on dry ice for 2mins then labelled by the addition of:

\[
\begin{align*}
3\mu l \ (30\mu Ci) &\quad \gamma^{32}\text{P-ATP} \\
1.5\mu l &\quad 10\times \text{C} \\
0.5\mu l \ (5u) &\quad \text{PNK}
\end{align*}
\]

After incubation for 1 hour at 37°C, the reaction was stopped by the addition of 15μl F-dyes and then heated to 80°C for 10mins before electrophoresis on a 4% acrylamide, 20% formamide, 7M urea gel until the BFB was at the bottom. The probe band was detected by autoradiography and excised from the gel. The probe was eluted into 400μl of RNA elution buffer (0.5M NaOAc pH 4.8, 1mM EDTA, 0.2% SDS) either for 1 to 2 hours at 55°C or overnight at 4°C. The gel slice was discarded and 2% volumes of ethanol were added to the probe in the elution buffer. The probe was always used immediately but storage at -70°C should be possible at this stage.

3.19.2 Annealing of Probe to RNA.

To 200μl of whole cell RNA in ethanol (approx. one third of RNA harvested from a 9cm diameter dish), 200μl of hot probe in ethanol were added together with 1μl of 5mg/ml yeast tRNA\(^{32}\text{RNA}^{\text{Nuc}}\) as carrier. After freezing on dry ice for 10 to 15 minutes, the samples were centrifuged for 10mins in a microfuge, the pellets were washed in absolute ethanol then resuspended in 30μl of S1 hybridisation mix (80% formamide, 0.4M NaCl,
40mM PIPES pH 6.4, 1mM EDTA). The samples were heated to 80°C to denature the probe then immediately transferred to 49°C (temperature optimised experimentally) to allow the probe to anneal to the RNA overnight.

3.19.3 Nuclease S1 Digestion.

3.19.3.1 S1 incubation mix: 0.28M NaCl

- 0.05M NaOAc (pH 4.6)
- 4.5mM ZnSO₄
- 20µg/ml carrier, sonicated salmon sperm DNA.

3.19.3.2 Procedure: The tubes were opened with the tips still immersed in the 49°C water bath. To each, 250µl of S1 incubation mix containing 250 units of S1 were added and mixed by pipetting. Tubes were immediately taken from the 49°C water bath, mixed further by vortexing, then stored on ice. The S1 digestions were for 30mins at 37°C and were then stopped by the addition of 4µl of 5mg/ml yeast tRNA<sup>sup</sup> and two volumes of 100% ethanol, followed by freezing for 10 to 15 minutes on dry ice. The samples were centrifuged for 15mins in a microfuge and the pellets were washed in 100% ethanol by recentrifugation before being resuspended in 6µl F-dyes and heated to 80°C for 5mins. This was followed by electrophoresis on a 20% formamide, 7M urea polyacrylamide gel.
3.20 *In vitro* Transcription.

3.20.1 Preparation of Transcription Template.

DNA template was linearised prior to transcription by digestion with *Bgl II* as follows:

2µl (2µg) DNA
2µl 5xTMS
1µl (10u) *Bgl II*
5µl TE.1

The digestion was for 1 hour at 37°C. The reaction was stopped by incubation at 70°C for 10mins. Linearised template can be stored indefinitely at -20°C.

3.20.2 Transcription of Linear Template by T7 RNA Polymerase.

rNTP mixes: Low C mix - 2.5mM A, G, U, 0.0625mM C.

Low C, low G mix - 2.5mM A, U, 1.25mM G, 0.0625mM C.
Approximately 0.6µg of linear template was transcribed at a time. Transcription reactions were set up at room temperature and then incubated at 37°C for 45 to 60 minutes. Each transcription reaction was as follows:

- 2µl 5×TMS or 5×BRL polymerase buffer.
- 1µl RNase inhibitor, RNasin (P&G) or RNAguard (Pharmacia)
- 2µl low C rNTP mix or low C, low G rNTP mix
- 0.5µl (15u) T7 RNA polymerase
- 1µl 10mM dIGTP
- 3µl linearised DNA template
- 1µl (20µCi) or 0.5µl (10µCi) α-32PCTP

Varying the buffer, or the rNTP mix, made no significant difference. 20µCi α-32PCTP were used to increase the specific activity of transcripts which were to be used for immunoprecipitation or sucrose gradient experiments.

3.2.0.3 Gel Purification of Transcripts.

After transcription, 20µl F-dyes were added to the reactions and heated to 80°C for 3mins. Samples were run on a 4% acrylamide, 20% formamide, 7M urea gel until the BPB was at least half way down. The transcript band was detected by autoradiography and excised from the gel. The RNA was eluted from the gel fragment by soaking in 400µl of RNA elution buffer (0.5M NaOAc pH4.8, 1mM EDTA, 0.2% SDS), either overnight at 4°C or for 1 hour at 55°C. The transcripts were recovered from the RNA elution buffer by ethanol precipitation at room temperature and resuspended in 20µl TE.1. Transcripts could be stored for up to three days frozen at -20°C.
3.21 Preparation of Splicing Extract.

The HeLa cell extract was prepared by a modified version of the method described by Dignam et al..

3.21.1 Buffers.

PBS: 0.02M K₂PO₄/KH₂PO₄ pH 7.4.
0.13M NaCl.

Buffer A: 10mM Hepes pH 7.9
1.5mM MgCl₂
10mM KCl
0.5mM DTT - added just before use.
pH should be adjusted using KOH or HCl

Buffer B: 0.3M Hepes pH 7.9
1.4M KCl
0.03M MgCl₂

Buffer C': 20mM Hepes pH 7.9
20% glycerol
0.6M KCl
0.2mM EDTA
0.5mM PMSF added just before use.
0.5mM DTT
Buffer D: 20mM Hepes pH 7.9
100mM KCl
20% glycerol
0.2mM EDTA
0.5mM PMSF added just before use
0.5mM DTT added just before use

Buffer C', described above, was used instead of Dignam's buffer C (Heintz & Roeder 1984). In later extracts the Hepes was replaced by triethanolamine (Tazi et al. 1986), this was found to increase the splicing efficiency.

3.2.1.2 Preparation of Extract.

The HeLa cells were harvested by centrifugation at room temperature at 2,000rpm in a MSE 6x300 rotor for 10mins. The cell pellets were resuspended in 10ml of PBS (per 250ml centrifuge tube) and transferred to 50ml centrifuge tubes for centrifugation at 1,500rpm at 4°C for 5mins in an MSE bench top centrifuge. The supernatant was discarded and the cell pellet resuspended in 5 volumes of buffer A (6x10⁶ cells = 2ml packed cell volume). The cells were centrifuged again at 1,500rpm for 5mins as described above, and then resuspended in 2 volumes of buffer A, for homogenisation on ice in a dounce homogeniser until at least 90% of the cells were lysed (usually 11 slow strokes of the 'B' plunger). The nuclei were spun down at 2,000rpm for 10mins in an MSE benchtop centrifuge at 4°C and the supernatant, containing cell debris and cytoplasm was discarded. The nuclear pellet was resuspended in buffer A in the same
volume as was used previously and was subjected to centrifugation at 15,000 rpm for 23 mins in an MSE 8x50 rotor at 4°C. The supernatant was decanted and stored at -70°C to make the S100 fraction (see below). The pellet was resuspended in buffer C', 2 ml per 10⁶ cells, added drop by drop and gently mixed using a cut off pasteur pipette. The mixing was continued by repeated inversion for 30 mins at 4°C. Following centrifugation at 15,000 rpm for 34 mins in an MSE 8x50 rotor at 4°C, the supernatant was dialysed against at least 50 volumes of buffer D for 5 hours. The extract was then recentrifuged for 10 to 15 mins at 15,000 rpm in an MSE 8x50 rotor at 4°C to clear any remaining debris. The extract was stored in 125 µl aliquots at -70°C until required.

The protein concentration of the extracts was determined from the following formula:

\[ [P] = 184 A_{230} - 81.7 A_{280} \, \mu g/ml. \]

A concentration of 6 mg/ml is ideal, 5 to 10 mg/ml is adequate. If the protein concentration was too high a precipitate was observed in the splicing reactions. This was not thought to affect the efficiency of splicing.

3.21.3 Preparation of S100 Fraction.

The decanted supernatant (see above) was mixed with 0.11 volumes of buffer B and centrifuged for 60 mins at 100,000 g*. The supernatant was dialysed for 5 hours against 20 volumes of buffer D then aliquoted and stored frozen at -70°C.
3.22 In Vitro Splicing.

Transcripts made in vitro were spliced in a HeLa cell nuclear extract. The splicing reactions were set up on ice. Each reaction consisted of:

- 1µl 75mM MgCl₂
- 1µl 10mM rATP
- 1µl 0.5M Creatine Phosphate
- 5µl 13% Polyvinylalcohol (PVA)
- 0.5µl RNase inhibitor (RNAsin(P&G) or RNAguard (Pharmacia)
- 5.5µl D buffer
- 1µl in vitro transcript
- 10µl HeLa cell nuclear extract.

Usually, all reaction components were pre-mixed on ice and added to the transcript. The reactions were started by transferring the tubes to a 30°C water bath. Following incubation for the required time, the reactions were stopped by freezing on dry ice. The reactions were phenol extracted, ethanol precipitated and resuspended in 6µl F-dyes before electrophoresis on a 20% formamide, 7M urea polyacrylamide gel. The gel was exposed to x-ray film.

With some extracts it was found to be advantageous to substitute the 5.5µl of D buffer in the reaction mix with 5.5µl S100 fraction. This often increased the efficiency of splicing.
3.23 Mapping the Position of a Branched Nucleotide in RNA by Primed Synthesis.

Mapping the position of a splicing branch point could be achieved using an oligonucleotide which was annealed to the RNA and then extended by reverse transcriptase in the presence of three cold deoxyribonucleoside triphosphates and one $^{32}$P deoxyribonucleoside triphosphate. The reverse transcriptase extension is unable to continue past a branched nucleotide in the RNA so is terminated. Comparing the extended product, on a sequencing gel, to sequencing reactions from the same primer on the equivalent DNA, will show the exact position of the branch.

Branched RNA was isolated from a preparative three fold scaled up splicing reaction. The splicing reaction was stopped by phenol extraction followed by ethanol precipitation. The pellet was resuspended in 20µl F-dyes and then subjected to electrophoresis on a 5% polyacrylamide, 20% formamide, 7M urea gel until the BPB was just off the bottom. The lariat bands were located by autoradiography and excised from the gel. The RNA was eluted from the gel into 400µl of RNA elution buffer (0.5M NaOAc pH4.8, 1mM EDTA, 0.2% SDS) at 55°C for 45mins. The buffer was transferred to a fresh tube and 2% volumes of cold ethanol were added to precipitate the RNA. 300nMoles of primer were added at this stage to co-precipitate with the RNA. After standing at room temperature for 20mins, the precipitate was pelleted by centrifugation in a microfuge for 10mins. The pellet was washed in 100% ethanol then dried under vacuum. The pellet was resuspended in 2µl TE.1 and 10µl (40mM PIPES pH 6.7, 0.4M NaCl, 1mM EDTA, 0.2% SDS) were added. The mixture was incubated for 4mins at 80°C and then 1½ hours at 37°C to anneal.
The primer/template mix was ethanol precipitated for 20 mins at room temperature. The pellet was resuspended in 9 μl TE.1 and 1.3 μl 10×RT buffer was added. The following mix was made for n reactions:

\[
\begin{align*}
1/6n & \mu l \text{ 10 mM dCTP} \\
1/6n & \mu l \text{ 10 mM dGTP} \\
1/6n & \mu l \text{ 10 mM dTTP} \\
1/2n & \mu l (5 μCi) \alpha^{32}PdATP \\
2/3n & \mu l \text{ 10 μM dATP (to reduce the specific activity of the label)} \\
12.5n & \text{ units AMV reverse transcriptase (Biorad)}
\end{align*}
\]

3 μl of this mix were added to the 10.3 μl of primer/template, and incubation was for 30 mins at 37°C. The reaction was stopped by the addition of 5 μl F-dyes and incubation at 80°C for 10 mins. The samples were analysed alongside appropriate sequencing reactions on an 8% polyacrylamide, 7M urea gel.

3.24 RNase H Protection.

RNase H degrades the RNA of an RNA:DNA duplex. Cleavage of an RNA transcript by RNase H can therefore be directed by DNA oligonucleotides complementary to the RNA. Protection from RNase H implies that the RNA is inaccessible to the oligonucleotide and this can be taken as an indication that a splicing component may be bound at that position.

Splicing reactions were set up on ice then incubated at 30°C for the required time. Following incubation, 400 pMoles of each oligonucleotide and 2.3 μ of RNase H (BRL) were added and incubation at 30°C was continued for a further 15 mins. The reactions were stopped on dry ice (at this point they can be stored at -70°C) then thawed, phenol extracted and ethanol
precipitated. Samples were resuspended in 6μl F-dyes, heated to 80°C for 2mins then run on a 7% polyacrylamide, 20% formamide, 7M urea gel until the BPB was just off the bottom.

3.25 Immunoprecipitation.

**Buffer - NET-2:** 50mM Tris/Cl pH7.5

- 150mM NaCl
- 0.05% Triton x-100

Splicing reactions were set up on ice and incubated at 30°C for the required time. After incubation, the tubes were transferred briefly to the bench and 400pMoles of each of oligonucleotides *size* and *imp*, and 2.3 units of RNase H (BRL) were added. Immediately 10μl of the appropriate antibody was added and the tubes were re-incubated at 30°C for 5mins to allow the oligonucleotides to anneal to the RNA and the RNase H digestion to begin. The tubes were transferred to ice for 20mins, then to each was added 50μl (4mg) protein A Sepharose (Sigma) in NET-2. The tubes were left on ice for a further 15mins, then centrifuged for 15 seconds at 12,000g in an MSE Microcentaur. The supernatant was removed by aspiration. The pellets were washed vigorously four times in NET-2 + 0.5mM DTT by resuspension, centrifugation and aspiration. Finally the pellets were resuspended in 200μl NET-2 + 0.5mM DTT and stored frozen on dry ice for up to two hours before they were phenol extracted, ethanol precipitated, resuspended in 6μl formamide dyes and run on a 7% acrylamide, 20% formamide, 7M urea gel until the BPB was just off the bottom.
3.26 Sucrose Gradients.


Gradient mix was prepared as:

- 12.5% Sucrose
- 100mM KCl
- 20mM Hepes-KOH pH 7.9
- 1.5mM MgCl₂
- 0.1mM EDTA

and stored at 4°C.

Gradients were prepared by aliquoting 4ml gradient mix into 5ml 'Kontron' cellulose centrifuge tubes and then freezing at -70°C until required. Before use, the gradients were thawed on the bench. The freezing and thawing of a 12.5% gradient mix allows a gradient of approximately 5 to 20% sucrose to form.


For each marker gradient 5A₂₅₀ units of 70S E. coli ribosome particles (200μl) were dialysed for about 1 hour at 4°C against 250ml of ribosome dialysis buffer which is:

- 1.5mM MgCl₂
- 100mM NaCl
- 20mM Tris pH 7.5
- 0.1mM EDTA
This dialysis causes the 70S ribosomes to dissociate to give 30S and 50S subunits which are ideal to optimise conditions for spliceosome gradients.

Mixed ribosome subunits were then layered onto gradients.

Gradients were subjected to centrifugation at 45,000rpm in a Beckman SW50L rotor at 4°C for varying times.

Centrifugation times of between 1½ to 2 hours, under these conditions, were found to be optimal.

3.26.3 Preparing Splicing Reactions for Analysis on Sucrose Gradients.

Three fold scaled up splicing reactions were set up on ice, incubated at 30°C for the required time, then stopped by freezing on dry ice. Reactions were thawed and then diluted to 200μl with ribosome dialysis buffer before being centrifuged at 12,000g for 20secs in a microfuge to remove any precipitate. The supernatants were layered onto the gradients which were then subjected to centrifugation as described above.

3.26.4 Unloading Gradients and Analysis of Fractions.

The gradients were unloaded from the top by displacement with 50% sucrose which was introduced at the bottom by a home made gradient unloading device (Fig. 40). 10 drop fractions (approx. 180μl) were collected for analysis.

Ribosome gradients were analysed by diluting the fractions to 1ml in water and measuring the A$_{260}$ of each.

Spliceosome gradients were analysed by Cerenkov counting.
3.26.5 Analysis of Spliceosome Gradient Fractions by Polyacrylamide Gel Electrophoresis.

Individual, or pooled fractions were diluted by the addition of two volumes of TE.1 before phenol extraction. Failure to dilute the fractions before extraction resulted in inversion or mixing of the layers due to the high sucrose concentration. Fractions were then ethanol precipitated, resuspended in 6μl F-dyes, heated to 80°C for 2mins before electrophoresis on a 7% acrylamide, 20% formamide, 7M urea gel.
Results Section 1.

Chapter 4.

3' Splice Site Usage in vivo.

4.1 Cloning of Synthetic 3' Splice Sites into pBSV.

4.1.1 Indirect Strategy.

Two different strategies were adopted to clone the synthetic oligonucleotides into pBSV.

The first strategy was an indirect approach in which the oligos were first cloned into a special M13 vector consisting of a slightly modified version of mpl1 which has had the EcoRI restriction site in the polylinker inactivated by site directed oligo mutagenesis. The vector also has the small BamHI Bgl II fragment of pBSV cloned into the polylinker. This vector is known as mpl1Rmut/BG6.

As can be seen from the map of pBSV (Fig. 8) the Bam Bgl fragment consists of the large intron of rabbit β-globin and quite a lot of the 3' and 5' flanking exon sequence including the EcoR1 site into which the oligos are cloned.

The advantage of the above strategy is that, once cloned, the oligo can be checked for correct sequence and orientation by dideoxy sequencing and, if positive, then the Bam Bgl fragment of rabbit β-globin from mpl1Rmut/BG6 can be cloned back into pBSV. The net result is that the oligo has been inserted into the EcoR1 restriction site 49bp downstream of
Figure 8. - This shows a schematic representation of the plasmid pJSV (see text) drawn to scale. Appropriate restriction enzyme recognition sites are indicated.

The plasmid contains SV40 sequences including the origin of replication and the gene encoding the large T antigen, these are essential for replication in HeLa cells. In addition the plasmid has sequences from pBR328 including the origin of replication so that it will replicate in E. coli, and the ampicillin resistance gene as a selectable marker. The plasmid contains the rabbit β-globin gene with 5' and 3' flanking sequences as shown (Grosveld et al. 1982).
SV40 sequence
pBR328 sequence
Rabbit \( \beta \)-globin sequence

\( p\beta_{SV} \)
9.7 kb
the authentic 3' splice site of the β-globin gene in pBSV and its sequence and orientation have been checked.

The EcoRI site in mpl1 had to be inactivated because the oligos were to be cloned into an EcoRI site in the rabbit β-globin gene and so this should be a unique site in the construct.

A mismatch primer was annealed to single stranded (ss) mpl1 DNA which changed a single base in the EcoRI site and thus inactivated it. The DNA was copied from the primer to give a double stranded circle and then ligated. Closed circle DNA was isolated from a low melting point (LMP) agarose gel and used to transfect E. coli JM101. DNA templates were prepared from the resulting plaques and were screened for mutants by single track dideoxy sequencing. 9% of the plaques screened were mutant. One mutant DNA was used to retransfect JM101 and the replicative form (RF) was prepared. This vector is mpl1Rmut.

The Bam Bgl fragment of pBSV was then cloned into the Bam site of mpl1Rmut as follows. Two approaches were used.

The first approach was to cut pBSV with BamH1 and Bgl II, isolate the small fragment from an LMP agarose gel and clone it into the Bam cut mpl1Rmut. The orientation of the fragment could then be checked by single track dideoxy sequencing. This method failed.

The second approach was a shotgun cloning approach which involved cutting pBSV with Sau3A (which cuts at Bam and Bgl sites but also at many other sites within pBSV). The fragments thus generated were cloned into the Bam site of mpl1Rmut and transfected into E. coli JM101. The resulting plaques were lifted onto nitrocellulose and probed using the nick-translated Bam Bgl fragment of pBSV. Potential positives were grown up and screened both by size selection on agarose gels and by single track
dideoxy sequencing. Two clones, 3 and 6, were positive and the inserts were in opposite orientations. This was confirmed by the ability of the two single stranded DNAs to hybridise (C-test). These new constructs were designated mpl1Rmut/βG3 (βG3) and mpl1Rmut/βG6 (βG6). βG6 was chosen to clone the oligos into (see later) as its EcoRI site is closer to the universal primer complementary region and it is therefore easier to check the orientation and sequence of the oligo.

To decide on the ratio of oligo to vector to use and also whether or not to use phosphatase treated vector, an oligo with an in-frame stop codon was cloned into the EcoRI site of mp8 to give a blue to white colour change. mp8 was chosen because βG6 or βG3 could not easily be used for this work as they always give white plaques and so the presence of an insert has to be detected by hybridisation or sequencing. A stop codon was used because the oligos are each 21bp long and so would not interrupt the reading frame of the lacZ gene in M13. One of the oligos fortuitously has an in-frame stop codon and so this was used for the preliminary experiments.

These experiments showed that the best conditions for cloning oligos are to use 0.2pMoles of oligo with 0.1μg phosphatase treated vector; this gives an efficiency of 25% to 30%.

As mentioned above, to facilitate checking the oligo insertion it is best to clone into βG6 and then sequence using the universal primer. I also had made a second primer which was to enable sequencing of oligos cloned into βG3, for this reason the first experiment was a pilot experiment where oligos were cloned into βG3 and transfected into JM101 to prepare ssDNA for analysis. DNA from each plaque was spotted onto nitrocellulose and probed using a radiolabelled mixture of both strands of
the oligos which were thought to be inserted. The 3G3 template is such that if an oligo is in the correct orientation then the B strand (an arbitrary assignment) of the oligo will be present in the ssDNA and so will hybridise to the A strand probe. Anything which is complementary to the A strand is therefore potentially positive and in the correct orientation.

This preliminary screen showed that of the plaques examined approximately 25% were positive and of these 50% were in the correct orientation. To confirm these results I tried single track dideoxy sequencing of the potential positives using the oligo primer mentioned above. This primer starts at nucleotide 1037 in the rabbit \( \beta \)-globin gene and primes towards the \textit{EcoRI} site. As a control against spurious priming by the oligo at other sites I also tried sequencing some of the same template with the universal primer. The results showed that the oligo was probably priming at spurious sites and the predominant sequence was not recognisable. The sequence from the universal primer was as expected.

The ability to sequence the oligos is essential to check their sequence exactly and for this reason the hybridisation screen above is inadequate. Inserts in the \textit{EcoRI} site of 3G6 can be sequenced quite easily using the universal primer so, whilst waiting for a new primer for 3G3, I went on to clone oligos into 3G6 with some success. Screening by single track dideoxy sequencing showed positives in the correct orientation for all but one of the oligos cloned.

To reclone the \( \beta \)-globin \textit{Bam Bgl} fragment from 3G6 to pBSV several approaches were tried.

The 3G6 was cut with \textit{Sau3A} and \textit{TaqI}. \textit{Sau3A} will cut out the \textit{Bam Bgl} fragment but will also generate other fragments, one of which is about the
same size as the *Bam Bgl* fragment. *TaqI* has a site within the other fragment so will remove any chance of contamination when the fragments are resolved on an agarose gel.

I intended to cut the *Bam Bgl* band from an LMP agarose gel and to extract the DNA for cloning into pBSV. Unfortunately, there was a group of four bands which ran very close together on the gel and this made it very hard to isolate the correct bands without contaminants.

A second, very similar approach was to cut either side of the *Bam Bgl* fragment in mpl1Rmut/βG6 using unique sites in the polylinker. This fragment was then isolated from an LMP agarose gel, the ends were trimmed using *Sau3A* and it could then be cloned into pBSV which had been cut with *Bam Bgl* and treated with phosphatase as shown in Fig. 9.

If the fragment is recloned in the correct orientation then it is possible to excise it again with *Bam* and *Bgl*.

For various reasons this recloning method never really worked. I suspected that something in the enzyme digests was degrading the DNA so tried religating the cut pBSV without insert and also ligating together the *Sau3A* cut insert fragments and looking at it by agarose gel electrophoresis. This seemed to show that the *BamH1* was not always cutting cleanly which, combined with the low recovery rate of the βG6 *Hind III* *Smal* fragment from the LMP agarose gel meant that at the time this strategy would not work.

4.1.2 Direct Strategy.

The second strategy for cloning oligos into the *EcoR1* site of the β-globin gene in pBSV was a direct approach in which the oligos were to
Figure 9. - Shows the initial proposed strategy for the subcloning of the rabbit \( \beta \)-globin *Bam HI-* *Bgl II* fragment, containing a synthetic 3' splice site, from mpl1Rmut/\( \beta \)G6 back into p\( \beta \)SV. The fragment (between the two *Sau 3A* sites) was to have been excised from mpl1Rmut/\( \beta \)G6 by cutting with *Hind III* and *Sma I* which are unique sites within the construct (step 1). This *Hind III-Sma I* fragment was to have been isolated from a low melting point (LMP) agarose gel and then trimmed using *Sau 3A* (step 2) to yield the rabbit \( \beta \)-globin *Bam HI-Bgl II* fragment. Finally this fragment was to have been cloned into a p\( \beta \)SV vector (step 3).
mp11Rmut/β G6

1

pβ SV + oligo

Bam HI

Bgl II

pβ SV vector (phosphatase treated.)

2

3

Hin DIII

Sau 3A

Sma I

Bam HI

Hin DIII

Sau 3A

Sma I

Sau 3A

Sma I

Sau 3A

Sma I

Sau 3A

Sma I
be cloned straight into the site in question. The insertion was to be confirmed by restriction analysis and then orientation and sequence checked by double stranded sequencing. This strategy was complicated by the presence of a second *EcoRI* site in pBSV and by the lack of an appropriate rapid method for the screening of large numbers of transfectants for the presence of an oligo in pBSV. The orientation and sequence of the oligo must then be checked.

To do a pilot experiment I partially digested pBSV with *EcoRI*. It has been observed that when partially digested with *EcoRI* more than 90% of the linear pBSV is cut at the site within the globin gene (I. C. Eperon, observation). By titrating with different amounts of enzyme, I found that the best conditions for the partial digest were to use one unit of enzyme per 5μg of DNA and to incubate for 10 mins at room temperature (21°C). The reaction was stopped by the addition of E-dyes and run on an LMP agarose gel to isolate the linear DNA.

To make the *EcoRI* partial cut vector I had to digest 20μg of DNA at a time. After isolating linears from LMP agarose, with an efficiency of about 10%, only 0.5μg of vector was recovered. The vector was then phosphatase treated so, from the initial 20μg, only about 0.4μg was recovered as usable vector. This was only enough for four reactions.

Having made the *EcoRI* partial cut vector I then cloned in two oligos as a pilot experiment to see whether positives could easily be screened and verified (see later). Following the success of these pilot experiments I decided to modify pBSV to enable me to clone straight into the *EcoRI* site without having to isolate partial digest products (which is time consuming and wasteful).
As mentioned above, more than 90% of linears generated by cutting pJSV with EcoRI are cut within the β-globin gene and not at the other site. For this reason I could not just fill in and religate an EcoRI partial digest product because it would probably be the wrong one.

A pJSV vector, a product of one of the pilot experiments designated pβAcc4 had an oligo cloned into the EcoRI site in the β-globin gene, inactivating that site. pβAcc4 was cut with EcoRI and the cut site filled in with dA and dT using Klenow polymerase. Subsequent blunt end ligation gave a version of pJSV with no EcoRI sites, called pβSV2. pβSV2 was verified by restriction analysis and agarose gel electrophoresis.

The second stage of this cloning strategy was to replace the β-globin gene in pβSV2 with a wild type rabbit β-globin gene from pBSV. This restored the EcoRI site within the gene to give a final version of pβSV with a unique EcoRI site within the β-globin gene, into which oligos can be cloned.

This reconstruction was performed by cutting both pBSV and pβSV2 using BstI (an isoschizomer of BamHI) and SalI. The digests were confirmed to be complete by agarose gel electrophoresis and then approximately 0.2pmoles of each were ligated together and used to transfect E. coli HB101. Colonies were grown up in a microtitre plate and DNA was prepared by the microprep method (see later, also Materials and Methods). This strategy of cloning by equimolar mixing of the two parental plasmids can potentially give rise to four recombinants (including re-formation of the parental types). Fig. 10 shows the possible recombinants. All four could be distinguished by restriction analysis using BamHI, SalI, and EcoRI. Of the thirty-two samples analysed by agarose gel electrophoresis, at least two were of the correct type. One
Figure 10. Shows the four possible recombinants generated during the making of the vector pSSV3. Appropriate restriction sites used to distinguish between these recombinants are indicated.
Rabbit $\beta$-globin gene fragment
of these was colony purified and re-analysed to confirm that it had a unique EcoRI site within the β-globin gene. The new vector was designated pβSV3.

4.1.3 Plasmid Microprep

As well as designing a suitable vector (pβSV3), I also had to devise a method for screening large numbers of recombinants following insertion of oligos into pβSV3. Initially the DNA was prepared for screening by a variation of the alkaline lysis method of Birnboim and Doly (1979). This method, although excellent in many respects, was found to be very laborious when dealing regularly with over twelve samples at a time (the maximum capacity of a microfuge). Due to the inefficiency of cloning the oligos, I often had to screen forty or more colonies in order to ensure one positive in the correct orientation. For the reasons outlined above I developed the microprep method (see Materials and Methods for latest protocol). Colonies were grown up in a ninety-six well microtitre plate which greatly facilitated screening.

The cells were grown up and alkali treated. They were then neutralised with potassium acetate and the chromosomal DNA removed by filtering through a ninety-six well microtitre filter plate. The filtrate was treated with 0.5% SDS at 80°C to destroy any remaining protein. The plasmid DNA was recovered from the SDS by ethanol precipitation and the pellet resuspended to give, on average, 0.5μg DNA per well. The DNA was usually free from chromosomal contamination and could be restricted.

The major technical problem with this method was the filtering. It was found to be preferable to drive the filtrate into the recipient plate.
by centrifugation. The use of a vacuum apparatus as described by Eperon (1986a) caused the SDS in the samples to froth. The recipient plate had to be bonded very firmly to the filter plate to prevent cross-contamination of the samples under centrifugation. The best way of sealing the plates together was found to be a lattice of a silicon rubber compound applied at least one hour before use and allowed to cure at 4°C under a light weight.

4.1.4 Microprep Screen.

The prepared DNA had then to be screened for the insertion of an oligonucleotide which, in turn, had to be checked for orientation. I experimented with two methods for detecting the presence of the oligos.

The first method tried was primer extension. The DNA was cut at the Bal I sites (see map: Fig. 8) and melted in a boiling water bath for three minutes. A 5' radiolabelled primer was annealed 5' of the EcoRI site and extended with unlabelled dNTPs by Klenow polymerase. This extension is very similar to hot primer dideoxy sequencing except that the nucleotide mix contains all four dNTPs and no dideoxy NTPs. The primer was extended through the EcoRI site and beyond to the end of the template at the Bal site. A band corresponding to this extended primer was detected by denaturing polyacrylamide gel electrophoresis. The presence of an insert in the EcoRI site led to a higher molecular weight band being detected. Multiple insertions of the oligo were also detected (Fig. 11).

The second method used to detect the presence of oligos was a restriction analysis method in which the DNA was cut with Bgl II, which has a unique site in pβSV and pβSV3. The cut Bgl II end was then filled
Figure 11. - Shows that the presence of an oligonucleotide cloned into pBSV can be detected by primed synthesis.

The DNA was cut at the Bal I sites (see map, Fig. 8) and melted in a boiling water bath for three minutes. A 5' radiolabelled primer was annealed upstream of the Eco RI site and extended with unlabelled dNTPs by Klenow polymerase. The primer was extended downstream through the Eco RI site and beyond to the end of the template at the Bal I site. A band corresponding to the extended primer was detected by denaturing polyacrylamide gel electrophoresis. Lanes 3, 4, 5 and 6 show the band detected in the absence of an oligonucleotide cloned into the Eco RI site. The presence of an oligonucleotide inserted into the Eco RI site led to a higher molecular weight band being detected (lanes 1, 2, 7 and 8) multiple insertions of the oligonucleotide were also detected (lane 2).
in by Klenow polymerase using $\alpha^{32}$FdATP and the other three dNTPs unlabelled. The DNA was then cut with MspI, which has multiple sites in pBSV and pBSV3, and analysed by denaturing polyacrylamide gel electrophoresis. As the Bgl II ends were labelled a band was detected, the size of which increased with the presence of an insert in the EcoRI site. As with the primer extension method, multiple insertions of the oligo could be easily detected (Fig. 12).

Although both methods worked well on CsCl purified DNA and on phenol/chloroform extracted mini-prep DNA, only the restriction analysis method worked reliably on microprep DNA, and could be carried out in a second microtitre plate. This may have been because of the higher levels of residual protein present in microprep DNA. This restriction analysis screen is now regularly used in conjunction with the microprep in the laboratory and has become known as the 'microprep screen'.

4.1.5 Checking Orientation of Cloned Oligonucleotides.

I tried checking the orientation and sequence of the inserted oligonucleotides by dideoxy sequencing of double stranded template. This was accomplished by linearising the plasmid, melting the strands, annealing a primer 5' of the EcoRI site and extending as for normal dideoxy sequencing. I found that phenol/chloroform extracted miniprep or microprep DNA could be accurately sequenced using a radiolabelled primer and extending using cold dNTPs and ddNTPs. Unfortunately, on miniprep DNA the sequence could only be read for about fifty nucleotides which was not far enough to read the sequence of an oligo inserted into the EcoRI
Figure 12. - Shows the result of a microprep screen to detect the presence of an oligonucleotide inserted into the Eco RI site of pBSV3 to make pBAcc3 (lanes 1 to 24) or pBAut (lanes 25 to 48).

The DNA was restricted with Bgl II which has a unique site in pBSV3, downstream of the Eco RI site in exon 3 of rabbit β-globin, (see Fig. 8). The cut Bgl II ends were filled in by Klenow polymerase using α³²P dATP and unlabelled dCTP, dTTP and dGTP. The DNA was then cut with Msp I which has multiple sites in pBSV3. The restriction pattern was analysed by electrophoresis on a 6% polyacrylamide denaturing gel. As the Bgl II ends were labelled a band was detected, the size of which increased with the presence of an insert in the Eco RI site. Track 13 is an example of a single oligonucleotide inserted into the Eco RI site of pBSV3 to give pBAcc3. Track 14 is an example of a negative result. The second labelled fragment is too large for resolution on this gel. The double bands seen in all tracks except 18 cannot be explained by the presence of restriction sites or by strand separation. The DNA may be partially renatured so bands corresponding to single and double stranded species may be seen.
<table>
<thead>
<tr>
<th>p\textsuperscript{\textbeta}Acc3</th>
<th>p\textsuperscript{\textbeta}Acc3</th>
<th>p\textsuperscript{\textbeta}Aut</th>
<th>p\textsuperscript{\textbeta}Aut</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 12</td>
<td>13 - 24</td>
<td>25 - 36</td>
<td>37 - 48</td>
</tr>
</tbody>
</table>
site. Sequencing using $\alpha^{32}$PdATP in the extension from an unlabelled primer gave a very high background and was unreadable.

To overcome these problems, I devised an indirect orientation checking method in which the DNA was linearised and melted as before. One strand of the oligo was radiolabelled and used as a primer. The oligo was hybridised to its complementary strand in the plasmid and then extended by Klenow polymerase copying one or the other of the strands depending on the orientation of the insert (Fig. 13). The primer was extended using dA, dC, dT and ddG. The extension stopped when a ddG was inserted opposite a C in the template giving one of two distinct bands depending on the orientation of the oligo in the template. A single base deletion in the oligo was easily detected by this method (Fig. 14).

In practice there was some read-through, probably due to some contaminating dG in one of the other nucleotides, but the extra bands were as predicted. Spurious priming was also occasionally detected but this was not found to be a problem.

4.1.6 Preparation of Material for Experimentation.

Positives in the correct orientation were prepared on a large scale and purified on a CsCl gradient. The Bam Sal fragments from each of these constructs were isolated from an LMP agarose gel and cloned into M13, mICE18 or mICE10 (Eperon 1986b) and both single and double stranded DNA was prepared for three reasons:

1) Final confirmation of the sequences of the oligonucleotides had to be by dideoxy sequencing of the inserts in mICE.
Figure 13. Explains the method adopted to determine the orientation of oligonucleotides inserted into pGSV3.

The plasmid was linearised by cutting with Bal I and the strands were melted apart by boiling for three minutes. One strand of the inserted oligonucleotide was labelled with $^{32}$P at its 5' end and was then used as a primer. The oligo was hybridised to its complementary strand in the plasmid and extended by Klenow polymerase copying one or the other strands depending on the orientation of the insert. The primer was extended using dA, dG, dT and ddG. The extension stopped when a ddG was inserted opposite a C in the template giving one of two distinct bands depending on the orientation of the insert in the plasmid.

Fig. 14 shows the result of this method used experimentally.
5' GGGAAAAGAATTXXXXXCACTCCTCAGG3'
ddTdTCTTTAAYYYYY5'  orientation 1,

3' CCGTTTCXXXXXTTAAGTCCAAGGTAC5'
5' YYYYYAAAATCTCCTCAddG  orientation 2,

XXXXX - Synthetic oligonucleotide which can be inserted in either of two orientations as shown above.

YYYYY - Second strand of the synthetic oligonucleotide, this was used as a primer on denatured plasmid DNA in order to determine the orientation of the oligonucleotide as described in the text.
Figure 14. - Shows an orientation check of the oligonucleotides indicated inserted into pBSV3 (the method used is as explained in figure 13). A 35 nucleotide band (as seen in the control track, 12) indicates that the orientation is correct; (tracks 7, 8, 9 and 12). A 29 nucleotide band indicates that the orientation is incorrect; (tracks 6, 10, 11). The 28 nucleotide band seen in tracks 1 to 5 suggests that oligonucleotide Accl is one base too short. The extra bands in tracks 9, 10 and 11 probably result from spurious priming.
2) mICE10 and mICE18 are such that \textit{in vitro} RNA transcripts of the Bam Sal fragment made by T7 polymerase were in the correct sense for \textit{in vitro} splicing in a HeLa cell nuclear extract.

3) Any \textit{in vitro} mutagenesis of the splice sites under investigation required a single stranded DNA.

The following is a list of constructs made:

- p\beta SV3
- p\beta Aut
- p\beta Acc2
- p\beta Acc4
- p\beta Acc5
- p\beta Acc6

- mICE18\beta Glo (no oligo)
- mICE18\beta Aut
- mICE18\beta Acc2
- mICE18\beta Acc4
- mICE18\beta Acc5
- mICE18\beta Acc6

- mICE10\beta Aut
4.2. Transient Expression of pBSV Constructs in HeLa Cells and Isolation of Whole Cell RNA for S1 Nuclease Mapping.

4.2.1 Transfection of HeLa Cells.

20µg of each of the pβ constructs were transfected into HeLa cells by calcium chloride precipitation as described in the Materials and Methods.

Initially, conditions for the transfection were optimised in collaboration with Dr. Lucy Eperon by detecting the presence of the SV40 large T antigen (T) in the HeLa nuclei by immunofluorescence (T is expressed from the SV40 fragment of all pβ constructs). The cells were fixed with acetone/methanol and then antibodies were added, first rat anti-T (a gift from Dr. P. Gallimore, CRC, Birmingham) and then FITC conjugated rabbit anti-rat (a gift from Dr. J.P.S. Banga, St. Thomas’s Hospital Medical School). The fluorescent nuclei were visualised by microscopy under ultraviolet light and the transfection efficiency was estimated to be between 30% and 50%.

In earlier experiments sonicated salmon sperm DNA was used as carrier in the transfections but this was later shown to be unnecessary and indeed reduced the efficiency of transfection in some controlled experiments (Dr. L.P. Eperon, personal communication).

DNA prepared by both large scale methods described in the Materials and Methods could be transfected satisfactorily. It is possible that DNA purified on a CsCl gradient transfects marginally more efficiently but there is no conclusive data to support this observation.

The cells were DMSO shocked 24 hours after transfecting and the RNA was harvested 24 hours after that.
4.2.2 Harvesting of RNA.

The RNA was harvested as described in the Materials and Methods. The only problem encountered with this was that the sonication step had to be optimised. This was achieved by harvesting untransfected cells which were then subjected to three different sonication conditions (Fig. 15). The RNA preps were continued and finally an aliquot of each was examined by agarose gel electrophoresis to look at the ribosomal RNA (rRNA) against *E. coli* rRNA markers (Fig. 15). Sonication conditions '2' were found to be best for shearing the DNA without seriously reducing the amount of RNA. These are the conditions which are now used in the final protocol.

4.2.3 S1 Mapping of Transfected HeLa RNA.

The splicing patterns of the pβ constructs were determined by S1 nuclease mapping of the RNA isolated from HeLa cells. The conditions for the S1 mapping were optimised by test mapping of RNA transcribed *in vitro* from a MICE construct. This was to reduce the number of variables and to confirm that the mapping procedure would work on RNA known to be present.

As I was mapping the use of 3' splice sites, I had to use either a universally labelled probe made by copying the *Bam Sal* fragment of β-globin in MICE with one or more α²³P labelled dNTPs, or a 5' end labelled probe. I chose to use as a 5' labelled probe, the *Bam Bgl* fragment of the β-globin gene 5' end labelled by using γ²³P-ATP as a substrate for polynucleotide kinase (Fig. 16-C1).
RNA was harvested from untransfected HeLa cells and subjected to three different sonication conditions (A). The RNA preps were continued and finally an aliquot of each was examined by agarose gel electrophoresis (0.8% gel) to look at the ribosomal RNA (rRNA) against *E. coli* rRNA markers. Sonication conditions '2' were found to be best for shearing the DNA (high molecular weight material in track 1 is much reduced in tracks 2 and 3. This band was identified as DNA by its resistance to digestion by RNase A.) without seriously reducing the amount of RNA. This method assumes that the effect of sonication on rRNA reflects the effect of sonication on mRNA.
A

<table>
<thead>
<tr>
<th>Sample</th>
<th>Power</th>
<th>Amplitude set</th>
<th>Amplitude peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Medium</td>
<td>5</td>
<td>2u</td>
</tr>
<tr>
<td>2</td>
<td>Medium</td>
<td>4</td>
<td>6u</td>
</tr>
<tr>
<td>3</td>
<td>Medium</td>
<td>3</td>
<td>10u</td>
</tr>
</tbody>
</table>

B

m123

m – E.coli rRNA

1)
2) 
3) 

Sonication conditions
Figure 16. Shows the *Bam HI-Eco RI* restriction fragment probe used for the S1 nuclease mapping (C1). The possible protected fragments are shown (C2, C3).
The advantage of using the 5' labelled probe was that only one S1 protected fragment was labelled and so it was much easier to tell exactly which 3' splice site had been used. Also, this probe was not susceptible to radiolytic decay, unlike the universally labelled probe, so the autoradiographs were very clean with no obscuring background smear.

As I was using a restriction fragment as a probe, I had to carefully optimise the hybridisation temperature. I showed that when hybridised at 45°C overnight, the two strands of the probe tended to reanneal. At 49°C the DNA probe no longer annealed to its complementary DNA strand but was still able to protect the RNA from S1 nuclease digestion. 52°C was too hot for the RNA : DNA duplex and no protected bands could be detected when the S1 digested material was analysed by denaturing polyacrylamide gel electrophoresis.

The p8 constructs were transiently expressed in HeLa cells and the transcripts were S1 nuclease mapped to determine the choice of 3' splice site. The probe (as described above) and possible protected fragments are shown in Fig. 16. The expected protected fragments were 149 nucleotides if the authentic 3' splice site had been used (128 nucleotides in the control of p8SV3 without an oligonucleotide inserted in the EcoR1 restriction site),(Fig. 16-C2) or 82 nucleotides if the alternative 3' splice site had been used (Fig. 16-C3).

My results (Fig. 17) show that in all the constructs tested, the synthetic sites are never chosen in vivo. This implies that the synthetic 3' splice sites alone are insufficient to generate an alternative splicing pattern. This suggested to me that other sequences, such as the branch point or the length and/or composition of the polypyrimidine tract may be involved in the selection of 3' splice sites.
Figure 17. Shows an S1 nuclease map of HeLa cell RNA. The cells were transfected with pβSV, pβSV3, pβAut, pβAcc2, pβAcc4, pβAcc5, pβAcc6. The RNA was harvested after transient expression of the plasmids (see materials and methods).

The probe used was the 5' end labelled Bam HI-Bgl II fragment of the relevant pβ construct (Fig. 16, C1). A different probe must be used for mapping each construct as the probe DNA must include the same synthetic splice site sequence as the RNA to be mapped, otherwise there will be a non-complementary bubble in the RNA/DNA hybrid which will be digested by S1 nuclease. The gel shown is 4% polyacrylamide, 20% formamide (see Materials and Methods). Markers (tracks M) are 5' end labelled, Hae III cut pBR322 fragments (see appendix 1).

The figure shows that β-globin RNA from pβSV and pβSV3 protects a 128bp probe fragment. β-globin RNAs from pβAcc2, pβAut, pβAcc4, pβAcc5, pβAcc6 protect a 149bp probe fragment. These must represent protected fragment C2 (Fig. 16) and imply that in all the constructs tested splicing in vivo in HeLa cells is to the authentic 3' splice site and that the synthetic 3' splice site is not chosen. possible reasons for the low levels of β-globin RNA detected in RNA harvested from cells transfected with pβAcc4 are discussed in the text.
Chapter 5.
Investigation of Sequence Involvement in Splice Site Selection by in vitro Mutagenesis.

5.1. Mutagenesis.

To investigate the sequence involvement in determining why, in my constructs, the authentic 3' splice site is always chosen, I undertook a series of mutageneses. The base changes and the rationale for each is shown in Fig. 18.

The mutants were made by oligonucleotide mismatch mutagenesis on mICEβAut. mICEβAut was chosen because the synthetic site is a copy of the authentic site and this was thought to eliminate any unpredictable effect of the sequences within the consensus. Constructs with double and triple mutations were made by progressive rounds of mutagenesis. Not all the mutants were successfully made, this was probably because of the inefficiency of the mutagenesis method.

Following annealing of the mutagenesis primer to mICEβAut and copying, closed circles were isolated from an LMP agarose gel and transfected into E. coli JM101. Plaques were grown up in a microtitre plate and single stranded DNA was prepared, and screened by single track dideoxy sequencing. One positive of each mutant was selected at random and retransfected into JM101 to plaque purify and to prepare large amounts of double and single stranded DNA for cloning or subsequent mutagenesis.
Figure 18. The proposed mutations to the sequences surrounding the 3' splice sites of p\textsuperscript{Aut} are shown:

1: This is an A \rightarrow G at the authentic branch point. It is possible that when a choice of splice sites must be made, both the 3' splice site and the branch point are together important in determining splice site selection. This mutation removes the advantage of a good branch point from the authentic splice site and this may be enough to alter the splicing pattern in favour of the alternative site.

2: This disrupts the authentic site polypyrimidine tract so may weaken the authentic site sufficiently to direct splicing to the alternative site. It has been reported that a single nucleotide change in a polypyrimidine tract can alter the splicing pattern of a gene (Archibald et al. 1986).

3: This disrupts only the 5' end of the authentic site polypyrimidine tract for the same reason as mutant 2 above.

28: This disrupts only the 3' end of the authentic site polypyrimidine tract, again the argument for this is as for 2 and 3 above.

8: This changes the invariant AG at the authentic site to a CG which will abolish splicing to this site and will show whether the alternative site can be used when there is no choice. This is an important control.

7: This removes the conserved pyrimidine in the authentic 3' splice site. Such a change has been reported to reduce the efficiency of splicing (Aebi et al. 1986), and may therefore weaken the authentic 3' splice site sufficiently to allow the alternative site to be used.

30: This creates a good branch point for the alternative splice site. The argument for this is as for mutant 1 above.

9: This destroys the AG immediately upstream of the alternative site polypyrimidine tract. This AG may interfere with splicing to the AG at the alternative site.

In some cases two or more of these mutants were combined in one construct in order to assess the relative importance of each in determining choice of splice site.

Nucleotide numbers within the rabbit \( \beta \)-globin gene are shown in green and are referred to in the discussion.
The mRNA sequence surrounding the two 3' splice sites of the rabbit β-globin / AAT construct is shown. Proposed base changes are shown, numbered above or below the sequence.

The mutants are described in the figure legend opposite.
The final plaque purified mutant DNA was checked by four track dideoxy sequencing. The following is a list of mutants made:

- mICE\$28
- mICE\$30
- mICE\$30/8
- mICE\$30/8/9
- mICE\$30/9

5.2. Subcloning Mutants Back into pBSV3.

In order to test the effects of my mutageneses in vivo, I had to subclone the mutated Bam Sal fragment from my mICE\$Aut constructs back into pBSV3. This could have been achieved by cutting the pBSV3 with Bam and Sal and then phosphatase treating to reduce the reformation of the parental recombinant. Alternatively, the wild type Bam Sal fragment could be separated from the large vector fragment by isolating the large fragment from an LMP agarose gel prior to cloning in the modified Bam Sal fragment from mICE\$Aut.

Both of the methods mentioned above would have been unsatisfactory because of the difficulty involved in screening for the mutant recombinants. The only difference between the pBSV3 and the mutant recombinant would be a few base changes. These would be very time consuming to detect and therefore verifying recombinants would be slow and laborious.

I needed a deleted pBSV3 vector which would allow me to detect recombinants by a large size change and, more importantly, guarantee that anything which was the right size contained the correct insert.
I decided to make a vector in which the β-globin Bam Sal fragment was replaced by the 276bp Bam Sal fragment of pBR322. This size difference is sufficient to detect the new vector or recombinants by agarose gel electrophoresis of microprep or miniprep DNA.

pβSV3 and pBR322 were cut with Bam and Sal and about 0.2pMoles of each were mixed and ligated before transfecting into E. coli HB101. The colonies were replica plated onto SB plates containing tetracycline (tet.) or ampicillin (amp.). Colonies which grew on both amp. and tet. (30%) were reconstituted pBR322 and were discarded. The ampicillin resistant, tetracycline sensitive colonies were grown up and DNA prepared by the microprep method before screening by size selection using agarose gel electrophoresis. Potential positives were re-transfected into HB101 to colony purify. DNA was prepared from 1ml cultures. The DNA was restriction analysed and, from three samples, one consisted of two large fragments of pBR322 joined together, one was the β-globin Bam Sal fragment of pβSV3 inserted into pBR322, and one was the pBR322 Bam Sal fragment inserted into pβSV3 to give the desired recombinant which I called pβSV4.

Having made pβSV4, I went on to clone into it, the Bam Sal fragments of mICEβ Aut, mICEβ30, mICEβ30/8/9, mICEβ28, mICEβ30/9 and mICEβ30/8. This was achieved by cutting both the pβSV4 and the mICE constructs with Bam and Sal and mixing 0.2pMoles of cut pβSV4 with a 2 to 3 fold molar excess of cut mICE and then ligating.

Transfectants having the large fragment of pβSV4 were selected for on ampicillin plates (mICE does not confer ampicillin resistance so any mICE recombinants are eliminated). Transfectants were screened for the presence of the β-globin Bam Sal fragment by colony hybridisation using a
universally α³²PdATP labelled Bam Sal fragment probe derived from mICEβAut. Possible positives from each were miniprepped and their size was checked by agarose gel electrophoresis before colony purification. DNA was prepared by a large scale method and finally checked by agarose gel electrophoresis. The following is a list of pBSV constructs made:

pBSAut
pBS28
pBS30
pBS30/8
pBS30/9
pBS30/8/9

5.3 3' Splice Site Usage in Modified pBSV3 Constructs in vivo.

The new mutant globin genes were transiently expressed in HeLa cells and the transcripts were analysed for 3' splice site usage by S1 nuclease mapping. As previously, the probe used was a 5' α³²P labelled Bam Bgl restriction fragment derived from the appropriate construct (see Fig. 16).

The results of this mapping are shown in Fig. 19. Mutations other than 28 seem to reduce the steady state level of mature mRNA detected. It can be seen that mutants 28 and 30 are both spliced to the authentic 3' splice site implying that a slight disruption of the authentic site polypyrrimidine tract is not enough to weaken the site sufficiently to activate the synthetic cryptic site downstream.

Creation of a good branch point for the cryptic site is also insufficient to allow activation of this site. However, the new branch site reduces the efficiency of splicing to the authentic site. This is
Figure 19. - Shows an S1 nuclease map of HeLa cell RNA. The cells were transfected with p828, p830, p830/9, p830/8/9 and p830/8. The RNA was harvested after transient expression of the plasmids (see materials and methods).

The probe used was the 5' end labelled Bam HI-Bgl II fragment of the relevant pβ construct (see Fig. 16 C1 and Fig. 17 legend). Probes are shown before (Probes -S1) and after (Probes +S1) S1 nuclease digestion. This shows that the probe is not protecting itself during the course of the experiment. The gel shown is 7% polyacrylamide, 20% formamide (see Materials and Methods). The markers (tracks M) are 3' end labelled Hpa II cut pBR322 fragments (see appendix 1).

The figure shows that β-globin RNA from p828 and p830 protects a 149bp probe fragment (tracks 28 and 30) implying splicing to the authentic 3' splice site (see fig. 16 C2). β-globin RNA from p830/8/9 protects an 82bp fragment (track 30/8/9) implying splicing to the synthetic 3' splice site (see fig. 16 C3). No β-globin RNA can be detected from constructs p830/8 and p830/9 (tracks 30/8 and 30/9). Possible reasons for the low levels of β-globin RNA detected in RNA harvested from cells transfected with p830/8, p830/9, p830/8/9 are discussed in the text.

In conclusion: of the mutants tested, only β-globin RNA from p830/8/9 (an AG→CG mutation at the authentic 3' splice site, see Fig. 18) is seen to splice to the synthetic 3' splice site.
intriguing, it could be that the unused branch point is targeting the transcript for degradation, or perhaps the cryptic site has become more competitive such that the "site choosing process" has a more difficult choice and therefore slows down, thus reducing the overall rate of splicing.

Spliced RNAs from 30/9 and 30/8 are not detectable by my methods. 30/8 has a AG→CG change at the authentic 3' splice site and as such was not expected to splice to the authentic site. The cryptic site is probably not activated because the AG at the EcoRI site immediately upstream of the cryptic site polypyrimidine tract is in some way interfering with and reducing the efficiency of that site, to the point where it is undetectably low. This hypothesis is supported by the observation that when the AG in question is changed to an AT (mutant 30/8/9) then the cryptic site is activated although the efficiency of splicing is at a very low level.

Why no spliced products are detectable from 30/9 is hard to say. It is possible that the sites are interfering with each other so that neither is favoured and neither is used. This is discussed further elsewhere.

It can be concluded from these results that the synthetic cryptic site can be activated by a mutation inactivating the authentic site. This implies that the cryptic site is available for splicing in vivo and is presumably neither involved in stable secondary structure nor tightly complexed with proteins. Possible future experiments based on this system in vivo are outlined in the discussion.
Section 2.

The second part of this thesis seeks to clarify the observations made in vivo from the experiments with pβ30, pβ30/9 and pβ30/8/9. To this end, I performed complementary experiments in vitro and studies of spliceosome assembly on transcripts where either the authentic site (30 or Aut) or the cryptic site (30/8/9) is used.

Chapter 6.

6.1 In Vitro Splicing.

I helped to set up in vitro splicing in the laboratory in collaboration with Dr. A.J. Else. Initially problems were encountered with growing HeLa cells to a sufficient density to make a nuclear extract. Experiments in growing cells in a 1l glass bottle containing a magnetic stirrer were unsuccessful as the cells either were not adequately stirred and so would not grow, or were disrupted and died.

Following a visit to Dr. J.A. Steitz's laboratory at Yale to learn extract making, I was able to successfully grow HeLa cells (JS1000, a gift from Dr. Steitz) to a density of greater than 6x10⁶/ml by daily doubling, using a 'Techne' microcarrier stirrer and microcarrier flasks. It is important that the cells should be doubling every 24 hours at least, and should have reached a density of 5x10⁶/ml before an extract is made.

The method we use for making the HeLa nuclear extracts and S100 fractions is essentially a variation of the method of Dignam et al. (1983) with modifications (Tazi et al. 1986; Heintz and Roeder 1984).
6.2 Splice Site Usage In vitro Analysed by S1 Nuclease Mapping.

Initially the splicing extract was inefficient and generated various artifacts when given α^32P-CTP labelled substrate RNA.

To investigate any authentic spliced product which may have been present below background, I tried making an unlabelled transcript from each of mICEßAut, mICEß28, mICEß30. These transcripts were spliced in vitro. The splicing reactions were stopped by phenol extracting after 0mins, 60mins and 120mins and the RNA was S1 mapped using the same 5' labelled Bam Bgl probes as had been used to analyse the material spliced in vivo. Fig. 20 shows that among the common bands (probably probe or mapping artifacts) specific bands can be seen for all three transcripts. The intensity of these bands increases with time and the size, 149 nucleotides, corresponds to predicted spliced products using the authentic 3' splice site. There is no evidence of splicing to the cryptic site in any of these three constructs.

Of the other expected bands generated by the reaction, free exon 1 and free lariat product would not be detected by a 5' labelled probe, and the lariat exon 2 intermediate would probably be indistinguishable from undigested probe. The results from transcript 30 are faint because much of the probe was lost during the course of the experiment. Nonetheless, after two hours a band is seen which corresponds to use of the authentic splice site and there is no evidence for use of the cryptic site.
Figure 20. - Shows an S1 nuclease map of an unlabelled in vitro splice of transcripts Aut, 28 and 30.

mICEβAut, mICEβ28 and mICEβ30 were linearised with Sal I (see appendix 2) and transcribed in vitro using T7 RNA polymerase and unlabelled ribonucleotides C, A, G, U (see Materials and Methods).

Approximately 0.2µg of each transcript was spliced in an in vitro splicing reaction for 0hrs, 1hr, or 2hrs (see Materials and Methods) using extract B9. The reactions were stopped by phenol extraction followed by ethanol precipitation. The RNA was S1 nuclease mapped using a 5' end labelled Bam HI-Bgl II restriction fragment probe (see Fig 16 C1). The gel shown is a 7% polyacrylamide, formamide denaturing gel with Hae-III cut pBR322 fragments as markers (see appendix 1).

The common background bands may result from either degradation of the transcript or from transcription products which are not full length. In addition, RNA from transcripts Aut and 28 protect a 149bp probe fragment which appears with the kinetics expected for the spliced product. This implies splicing to the authentic 3' splice site in vitro (see Fig. 16 C2). RNA was lost from transcript 30 but subsequent experiments have shown that this also splices to the authentic 3' splice site in vitro (see for example Fig 22).
6.3 Direct Analysis of \textit{in vitro} splicing.

Subsequent splicing extracts, especially those buffered with triethanolamine instead of Hepes (Tazi et al. 1986) were very efficient and it was then possible to use $\alpha^{32}$P-CTP labelled RNA transcribed \textit{in vitro} by T7 RNA polymerase as substrate for the splicing reactions. Splicing products and intermediates were then analysed directly by polyacrylamide gel electrophoresis.

6.4 \textit{In vitro} Splicing of Transcripts 28, 30, 30/8, 30/9, 30/8/9.

This was an initial experiment to show that the bands supposed to be the splicing intermediates and products were, in fact, splicing specific.

Fig. 21 shows a time course experiment of 90mins and 180mins for each of the five transcripts spliced using extract LX2. From the control lanes, in the absence of extract there is little except the full length transcript to be seen. The few extra bands are presumably degradation products. All transcripts were purified by polyacrylamide gel electrophoresis prior to use as splicing substrates.

When extract is added but no ATP then again no splicing specific bands can be seen. The bands seen are presumably extract specific but are nothing to do with splicing. They can be disregarded.

In the full splicing reactions, two high molecular weight bands (arrowed) appear above the full length transcripts, increasing in intensity over the time course. As the transcript cannot have become so much longer, these bands must be moving aberrantly through the gel and are
Figure 21. - Shows an in vitro splice of transcripts 28, 30, 30/8, 30/9 and 30/8/9. Controls are shown without extract (-E) and without ATP (-A). Active splicing reactions are labelled +. Markers (M) are Hpa II cut pBR322 (see appendix 1) and the gel is a 5% polyacrylamide,formamide gel (see materials and methods).

mICE β constructs were linearised with Bgl II (see appendix 2) and transcribed in vitro (see Materials and Methods). 1/10th of each reaction was spliced in vitro using extract LX2. The reactions were stopped by phenol extraction after 90 or 180 minutes. The control lanes show that there are no splicing specific bands in the absence of extract or ATP.

In active splicing reactions (lanes +) two high molecular weight bands (arrowed) appear above the full length transcripts increasing in intensity over the time course. The transcripts cannot have increased in length, therefore these bands must be moving aberrantly through the gel and are strong candidates for the lariat species generated by the splicing reaction (see Introduction). The presence of two lariat bands (product and intermediate) implies that both steps of the reaction are working (see Fig. 2). Ligated exon bands cannot be seen (a band of 188 nucleotides would be expected to indicate splicing to the authentic site or a band of 120 nucleotides would indicate splicing to the synthetic site). There may be an exonuclease activity in extract LX2, this is discussed in the text.

The lariat bands of transcripts 30/9 and 30/8/9 are moving slower through the gel than those of transcripts 28 and 30. Transcripts 30/9 and 30/8/9 may be using a cryptic branch point (see Fig.18 mutation 30) which could imply use of the synthetic splice site in vitro.
therefore strong candidates for the two lariat species generated by the splicing reaction (see Introduction).

Although no spliced products can be seen in Fig. 21 (a band of 188 nucleotides would be expected if the authentic splice site was used, or a band of 120 nucleotides if the cryptic site was used), the two lariat bands imply that both the lariat intermediate and the lariat product are being generated and therefore both steps of the splicing reaction are working. Perhaps there is an activity present in LX2 which degrades the spliced product once it is released from the protection of the spliceosome.

The most interesting point to note about Fig. 21 is that the lariat bands of 30/9 and 30/8/9 are moving slower through the gel than the corresponding lariat bands of 28 and 30 which have already been shown to splice to the authentic site \textit{in vitro}. This could be explained by the branch point of 30/9 and 30/8/9 being at the cryptic site. 30/8/9 has been shown to splice to the cryptic site \textit{in vivo} and 30/9 has the same AG\textsuperscript{+}CG mutation to inactivate the authentic 3' splice site.

\textbf{6.5 Use of Alternative Branch Points \textit{in vitro}.}

To investigate the possible use of the cryptic branch point \textit{in vitro}, I decided to use transcripts Aut, 30 and 30/8/9. Aut as a control, 30 because it has 'normal' looking lariats and has been shown to splice to the authentic site \textit{in vivo} and \textit{in vitro}, and 30/8/9 because it shows abnormal mobility of the lariat bands and has been shown to splice to the cryptic site \textit{in vivo}. 
I performed a time course experiment of 60mins and 120mins on each of the three transcripts using extract LX5. The 'no extract' controls have 'D' buffer instead of LX5 to make up the volume.

Fig. 22 shows that the intensity of the splicing specific bands increases over the time course. The figure confirms that 'Aut' and '30' both splice to the authentic site in vitro and 30/8/9 splices to the cryptic site. These results are qualitatively the same as those obtained in vivo. The gel also confirms that the 30/8/9 lariat bands are different from transcripts which splice to the authentic splice site and therefore the branch point of 30/8/9 may be at the cryptic branch site.

The apparent difference of the splicing efficiencies of these three constructs may be artifactual. The substrate RNA concentration may vary from transcript to transcript due to different labelling efficiencies of different transcription reactions. This is further investigated below.

6.6 Relative Efficiencies of Splicing Different Transcripts in vitro

To investigate more thoroughly the splicing efficiencies of transcripts Aut, 30 and 30/8/9 in vitro, a four point time course was performed (1,2,3,4 hours) splicing with extract LX8. The activity of each transcript was estimated by Cerenkov counting and approximately 10^3cps were used per reaction.

Both exposures of the gel are shown in Fig. 23. The efficiencies of splicing of Aut and 30 are virtually identical judging by the appearance of the lariat bands. 30/8/9 is a little slower but not dramatically so, perhaps 25% to 50% less efficient. The kinetics of appearance of the
Figure 22. Shows an in vitro splice of transcripts Aut, 30 and 30/8/9. The splicing reaction intermediates and products are indicated.

MICE8 constructs were linearised with Bgl II (see appendix 2) and transcribed in vitro (see Materials and Methods). 1/10th of transcription reactions were spliced in vitro by extract Lx5 (see Materials and Methods). The splicing reactions were stopped after 60 minutes and 120 minutes and the products analysed by electrophoresis on a 5% polyacrylamide formamide gel (see Materials and Methods). Markers are Hpa II cut pBR322 fragments (see appendix 1). The 'no extract' controls have 'D' buffer added instead of extract Lx5 and were incubated for 120 minutes.

This figure confirms that transcripts Aut and 30 both splice to the authentic site (188 nucleotide ligated exon band) and that transcript 30/8/9 splices to the synthetic site (120 nucleotide ligated exon band). The gel also confirms that the 30/8/9 lariat bands are different from transcripts which splice to the authentic splice site and therefore the branch point of 30/8/9 may be at the cryptic branch site.

Intensities of bands on this gel are not necessarily indicative of the relative splicing efficiencies of the transcripts (see text, also Fig. 23)
Figure 23. Shows the relative efficiencies of splicing of transcripts Aut, 30, 30/8/9. Two exposures of the same gel are shown. mICE constructs were linearised with Bgl II (see appendix 2) and transcribed *in vitro* (see Materials and Methods). The activity of each transcript was estimated by Cerenkov counting and approximately $10^{5}$ cps were used in each *in vitro* splicing reaction with extract LX8. The reactions were stopped after 1, 2, 3 or 4 hours and analysed by electrophoresis on a 5% polyacrylamide, formamide gel (see Materials and Methods). Markers are *Hpa II* cut pBR322 (see appendix 1). The splicing efficiencies of transcripts Aut and 30 are virtually identical judging by the appearance of the lariat bands. The splicing efficiency of transcript 30/8/9 is less efficient. The kinetics of appearance of the lower lariat band relative to the higher band are consistent with it being the lariat product and the higher band being the intermediate as indicated.

The spliced products are very faint on this gel implying that LX8 may have the same nuclease activity as LX2 (see Fig. 21); this is discussed in the text.
<table>
<thead>
<tr>
<th>Exp 1</th>
<th>Exp 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aut 30 30/8/9</td>
<td>Aut 30 30/8/9</td>
</tr>
<tr>
<td>1-4 hrs 1-4hrs</td>
<td>1-4hrs 1-4hrs</td>
</tr>
</tbody>
</table>

- The images on the left and right show gel electrophoresis results for two different experiments. Each experiment has a distinct set of samples labeled with specific autolysis times and dates.

- The gel patterns display varying bands and intensity levels, indicating different sample compositions or reaction outcomes.

- The table summary highlights the consistency in sample preparation across both experiments, focusing on autolysis times and dates.
lower lariat band relative to the higher band are consistent with it being the lariat product and the higher band being the intermediate.

Interestingly, the spliced products are very faint on this gel (as in Fig. 21). This could imply that LX8 has the same nuclease activity as LX2 and degrades the RNA once it is released from the spliceosome. If this were postulated to be an exoribonuclease activity it would explain why the lariat product is still detected. Even if the lariat product's short tail was removed, this would probably make almost no difference to its mobility on a polyacrylamide gel. The branched nucleotide is nuclease resistant, so degradation by an exoribonuclease would be stopped once the tail was removed.

6.7 Mapping of Branch Points Used in vitro.

It has been shown that the presence of a branched nucleotide in an RNA template causes the termination of the extension of a DNA primer by reverse transcriptase (Krainer et al. 1984; Ruskin et al. 1984). The position of the branch point can, therefore, accurately be determined by primer extension and polyacrylamide gel electrophoresis. Sequencing extensions from the same primer, on a DNA corresponding to the RNA, can be used as markers. The branch point can then be determined by comparing the size of the primer extension product to the sequencing tracks.

I had made two primers complementary to β-globin RNA; RpriAut and RpriEco. These anneal within the exon just downstream of the authentic and cryptic 3' splice sites respectively as shown in Fig. 24. These primers can only be used on the lariat intermediates and therefore the lariat products can be used as negative controls.
Figure 24. Shows extension and predicted termination of primers RpriAut and RpriEco on lariat intermediate material.

The β-globin exon 3 is indicated as a box, the intron sequence is indicated by a line. Proposed lariat intermediates are drawn and the used 3' splice sites are indicated by arrows. Primers RpriAut and RpriEco (see appendix 3) are shown opposite their complementary RNA sequence. Proposed extension to the used branch site is indicated by a dotted line terminating at the branch site (X). The branched nucleotide causes termination of the extension of a DNA primer by AMV reverse transcriptase (Krainer et al. 1984; Ruskin et al. 1984).
Two strategies were employed to try to map the branch points of transcripts Aut, 30 and 30/8/9.

The first strategy involved mapping minute quantities of gel purified RNA. In a three fold scaled up splicing reaction, transcripts Aut, 30 and 30/8/9 were spliced for 2% hours in LX8. Following polyacrylamide gel electrophoresis on a 5% gel, the bands shown in Fig. 25 were excised and the RNA eluted. The two lariat bands of 30/8/9 were insufficiently resolved to be excised as discrete bands so they were excised together.

Primer RpriEco was annealed to band 1 material and primer RpriAut was annealed to the material eluted from bands 2, 3, 4 and 5. The primers were extended using α32P-ATP and the other three unlabelled dNTPs by AMV reverse transcriptase (Biorad). The extended primers were denatured and resolved by polyacrylamide gel electrophoresis against sequencing markers made by sequencing mICEβ30/8/9 with RpriEco and mICEβAut with RpriAut.

Fig. 26 shows that although the signals are quite low there are definite bands in lanes 1, 2 and 4 which, when compared to the sequencing ladders, are seen to map to the predicted branch point nucleotide.

Band 1 (30/8/9) lariat is branched to the cryptic branch point and there is no evidence of any lariat formation to the authentic branch point. The signals are very weak so low levels of lariat formation to the authentic branch point cannot be ruled out but are unlikely.

Lanes 2 and 4 show that transcripts 30 and Aut both form lariats to the authentic branch point as expected. The faint band, one nucleotide above the major band in lane 4 could be indicative of a low level of branch formation to the A immediately upstream of the branch point A at the authentic branch site. This has been reported previously (Hornig et
Figure 25. - Shows the bands excised for branch point mapping.

In a three fold scaled up reaction, transcripts Aut, 30 and 30/8/9 were spliced for 2%hrs in extract LX8 (see Materials and Methods). The reactions were electrophoresed on a 5% polyacrylamide, formamide preparative gel and the lariat bands indicated 1, 2, 3, 4, 5 were excised and the RNA eluted. The lariat bands of 30/8/9 were not sufficiently resolved to be excised as discrete bands.

This figure suggests that splicing of transcript Aut is more efficient than splicing of transcript 30 (after 2%hrs more of transcript Aut has become converted to the lariat species). This is not consistent with the data presented in Fig. 23.
Figure 26. - Shows the mapping of the branch sites on gel purified RNA, by primer extension from the two primers indicated.

Primer RpriEco was annealed to band 1 material and primer RpriAut was annealed to the material eluted from bands 2, 3, 4 and 5 (see Figs. 24 and 25). The primers were extended using α32PdATP and unlabelled dCTP, dGTP and dTTP by AMV reverse transcriptase. The extended primers were denatured and resolved by electrophoresis on an 8% polyacrylamide sequencing gel (see Materials and Methods) against markers made by sequencing mICE30/8/9 with RpriEco and mICEAut with RpriAut. The expected branch points A (seen in the 'T' tracks of the sequencing reactions) are indicated. There are faint but discernable bands in lanes 1, 2, and 4 (indicated by dots to the right) which, allowing for the smile of the gel, when compared to the sequencing ladders, are seen to map to the predicted branch point nucleotide.

In conclusion, the band 1 lariat (30/8/9, see Fig. 25) in track 1 is branched to the cryptic branch point (Fig. 18, mutation 30). Bands 2 and 4 (30 and Aut respectively, see Fig 25) in tracks 2 and 4, are both branched to the authentic branch site as expected. The primers cannot anneal to lariat product material and therefore no bands were expected in lanes 3 and 5.
al. 1986). Alternatively, the faint band could be an artifact of the reverse transcription.

The absence of any bands in lanes 3 and 5 confirms that the primers cannot anneal and therefore these bands are the lariat products rather than intermediates. Incidentally, the results of these experiments confirm the assignations of these two lariat bands. No other splicing specific material could give these predictable results.

The second branch point mapping strategy involved trying to map larger quantities of RNA from an unlabelled splicing reaction. The disadvantage of this was that to avoid spurious incorporation of label by unintended priming by RNA present in the reaction, 5' $^{32}$P labelled primers had to be used (being unlabelled, the lariat species could not be isolated from the bulk RNA). Major bands appeared which could not have been generated by priming on lariat material. Presumably these were the products of spurious priming and they obscured the predicted band. Therefore, this strategy was unsuccessful.

6.8 Conclusions

In conclusion, both in vitro and in vivo I have shown that in my competitive system the authentic 3' splice site is used exclusively, unless the AG dinucleotide at this site is changed to a CG, thus disabling the site. In this case, splicing is to the new, synthetic cryptic site but at
a reduced level. This was a very interesting finding for the following reasons:

1) One may predict that, in a situation where there has to be a choice between 3' splice sites, if one site is disabled the other site will be used and the overall level of splicing will not decrease. This hypothesis is particularly attractive given that both sites are identical over a region of 21 nucleotides, and both have identical branch site sequences (by mutagenesis). One could argue that if neither site is disabled, then the authentic site is used because it is at a special advantage.

2) Previous work (Reed and Maniatis 1985) has shown that if a human \( \beta \)-globin 3' splice site is inactivated by an AG-GG mutation, step 1 of splicing still goes on. Frendewey and Keller (1985) have shown that the AG at the 3' splice site is not required for splicing complex formation. However, its presence may enhance the efficiency of assembly. It may therefore be expected that if the exclusively used site is mutated, this will lead to dead-end lariat intermediates branched to the authentic branch site.

Contrary to the evidence outlined above, I have shown that an AG-CG change at the 3' splice site causes lariat formation and splicing to the cryptic branch point and splice site respectively. There is no evidence for dead-end lariat intermediates branched to the authentic branch point. This result implies that factors influencing splice site choice may differ from those governing whether a splice site can be used at all.

The above conclusions are enlarged upon in the Discussion.
Chapter 7.

Investigation of Spliceosome Components Binding to Authentic and Cryptic 3' Splice Sites by Ribonuclease H Protection Mapping.

7.1 Introduction.

It is possible that the reduced level of splicing to the cryptic 3' splice site in transcript 30/8/9 could reflect competition from the inactivated authentic 3' splice site. Frendewey and Keller (1985) have shown that complex assembly still occurs at a 3' splice site which has been mutated at the AG dinucleotide. It has also been shown that if this AG dinucleotide is mutated or deleted then the first step of splicing occurs, but at a reduced efficiency (Reed and Maniatis 1985; Ruskin and Green 1985a). Part of the 3' splice site (the polypyrimidine tract) is required for step one of the splicing reaction (Reed and Maniatis 1985; Ruskin and Green 1985a). These results suggest that the authentic 3' splice site in 30/8/9, although disabled, may still be recognised as a splice site and there may be some abortive complex assembly as a result. This recognition of the authentic 3' splice site and partial complex assembly may compete with the complex assembly at the cryptic 3' splice site which leads to splicing at that site. The net result of this would be a reduced rate of splicing to the cryptic site.

I decided to investigate splicing factor binding at both the splice sites on transcripts where different sites are exclusively used. These experiments were performed using a method originally developed by Ruskin and Green (1985c) which assumes that binding of splicing components to
the RNA will protect regions of the RNA from oligonucleotide directed cleavage by ribonuclease H (RNase H). These experiments were all performed in vitro.

Initially I had made a series of oligonucleotides complementary to regions of the rabbit β-globin transcripts which were used as substrates for the in vitro splicing experiments. The regions chosen were those 3' splice site regions which had already been shown to be protected from cleavage by RNase H during in vitro splicing (Ruskin and Green 1985c). Fig. 27 shows the regions of the RNA against which the oligonucleotides can direct cleavage by RNase H. The oligonucleotides are discussed below.

The oligonucleotide designated *size* is complementary to a region of unprotected RNA within the intron. It causes the pre-spliced transcript to be cleaved to give a 275 nucleotide fragment which is easily resolved by polyacrylamide gel electrophoresis on a 7% gel. *Size* was always included as an extra oligo in all experiments so that RNase H cleavage products could easily be resolved and sized by polyacrylamide gel electrophoresis.

The oligonucleotides *Aut* and *New* are 18 base oligonucleotides complementary to the authentic and cryptic (new) branch points respectively. The region of complementarity is centred on the branch point A in both cases. These oligonucleotides were used to investigate the association of factors with the branch points on transcripts where one or the other branch point is used.

*Sp* is an 18 base oligonucleotide complementary to the authentic and cryptic splice sites of transcripts *Aut* and 30. The oligonucleotide extends from the AG dinucleotide into the polypyrimidine tract. Interestingly, *Sp* cannot direct cleavage against the mutated authentic 3'
Figure 27. - Shows the sites on the rabbit β-globin transcripts against which RNase H is directed by the oligonucleotides indicated.

Oligo 26 does not direct cleavage to the authentic 3' splice site of transcript 30/8/9 so is shown directing cleavage only to the synthetic 3' splice site.

In the following experiments (figs.28 to 32), in vitro transcripts are made by transcribing mICEβ constructs linearised with Bgl II (see appendix 2, also Materials and Methods). The total length of the transcript is 745 nucleotides. The region upstream of the site of cleavage by oligo size is 470 nucleotides. The region downstream is 275 nucleotides and this can be further cleaved by RNase H digestion following binding of the oligonucleotides as shown, the cleavage pattern can then be analysed by polyacrylamide gel electrophoresis as described in the materials and methods.

The sequences of the oligonucleotides are as shown in appendix 3.
Rabbit β-globin transcript used for *in vitro* splicing.

- **T7 promoter and mlCE10 polylinker sequences**
- **Alternative exon region**
- **Oligo with sequence AATTTTCTTTTTCTACAGGT cloned into EcoR1 site. This is a copy of the authentic 3' splice site**
- **Positions of mutations**
  - 30- Generates a branch point 5' to the new splice site. CTCAT→CTAAC.
  - 8- Inactivates the authentic splice site AG→CG
  - 9- AG→AT at the EcoR1 site 5' to the new splice site oligo in case this AG interferes with splicing to the new site.
  - 30/8/9- All three mutations on the same transcript
- **Oligos added to direct RNaseH cleavage to give products a-d of lengths:**
  - a ≤ 94nt
  - b ≤ 66nt
  - c ≤ 25nt
  - d ≤ 90nt
- **Mapped branch points.**
splice site of transcript 30/8/9 even though this only differs by a single base from the authentic 3' splice site of transcript Aut. (Fig. 28; track sp-)

7.2 Protection of Authentic and Cryptic 3' Splice Sites of Transcripts Aut, 30 and 30/8/9.

Transcripts Aut, 30 and 30/8/9 were set up on ice as for a normal splicing reaction with extract LX8. The reactions were incubated at 30°C for 5mins and then 3' splice site protection was mapped by the addition of RNase H and oligonucleotide sp as described in Materials and Methods. The transcripts were also cleaved at both branch sites as a control. Fig. 28 shows that after 5mins a sub-population of the ad fragment of transcript 30 is protected from RNase H cleavage directed against both the authentic and cryptic 3' splice sites by oligo sp. This shows that both 3' splice sites on transcript 30 are protected by 5 minutes although splicing is only to the authentic site. sp does not direct cleavage at the mutated authentic 3' splice site of transcript 30/8/9 so Fig. 28 does not show directly whether both 3' splice sites are protected. However, comparing the ratios of the bands ad : ac in sets 30 and 30/8/9 suggests that the cryptic site is protected to about the same extent, at least initially, whether it is ultimately spliced to or not.

To see whether both splice sites are protected on a single transcript of 30/8/9, a second sp oligo, sp2, was ordered. sp2 is complementary to the mutated authentic 3' splice site of 30/8/9 and directs RNase H cleavage to that site. Fig. 29A shows the results of a time course experiment over 120mins in which protection of the authentic and cryptic
Figure 28- Shows an RNase H protection experiment. Components bound to both 3' splice sites and branch points of the transcripts indicated are investigated after incubation for 5 minutes under splicing conditions. The control transcript is from mICEAut, this does not have a second branch point so is resistant to cleavage directed by oligonucleotide new. The cleavage products and protected species indicated are as described in Figure 27. Markers (M) are HpaII cut pBR322 fragments (see appendix 1).

**Left hand plate:** 'oligo' tracks are uncleaved transcript incubated for 5 minutes under splicing conditions. In addition to the oligonucleotides indicated all RNaseH cleavage reactions also included oligonucleotide sp. The positions to which the oligonucleotides direct cleavage and the possible cleavage products a, b, c, d, are shown in fig. 27. In this figure the fragments are abbreviated as shown:

\[
\begin{align*}
&ad = a+b+c+d \quad (275 \text{ nucleotides}) \\
&\text{transcript-ad} = \text{transcript-(a+b+c+d)} \quad (470 \text{ nucleotides}) \\
&ac = a+b+c \quad (185 \text{ nucleotides}) \\
&bd = b+c+d \quad (181 \text{ nucleotides}) \\
&ab = a+b \quad (160 \text{ nucleotides}) \\
&cd = c+d \quad (115 \text{ nucleotides})
\end{align*}
\]

Coeavage by oligonucleotides aut and new were included as controls as cleavage of both transcripts 30 and 30/8/9 by both oligonucleotides was expected after five minutes. The presence of a faint protected band (ad) in the 'aut' and 'new' tracks implies little or no protection at either branch point after 5 minutes. (Branch site protection is investigated in section 7.3; branch sites are protected after 5 minutes but detection is always much less than 3' splice site protection. Perhaps the oligonucleotides directed against the branch sites only partially cover the region of complex binding or perhaps complex binding is weaker at the branch sites and is displaced by the oligonucleotides).

The conclusion from this figure is that after 5 minutes a sub-population of the ad fragment of transcript 30 (see fig. 27) is protected from RNaseH cleavage directed against both the authentic and cryptic 3' splice sites by oligo sp (see the presence of the ad band in 'sp' tracks). This shows that both 3' splice sites on transcript 30 are protected by 5 minutes although splicing is only to the authentic site (see fig. 22). Oligonucleotide sp does not direct cleavage at the mutated authentic 3' splice site of transcript 30/8/9 (data shown in right hand plate, see below) so this figure does not show directly whether both 3' splice sites are protected. However, comparing the ratios of bands ad:ac (i.e. protection:cleavage of the cryptic site) when cleavage of transcripts 30 and 30/8/9 is directed by oligonucleotide sp (tracks 30sp and 30/8/9sp) suggests that the cryptic site is initially protected to the same extent whether it is ultimately chosen or not.

**Right hand plate:** Shows cleavage directed by the oligonucleotides indicated, of transcript 30/8/9 under the same conditions as above. Track 'sp-' shows cleavage by oligonucleotide sp in the absence of splicing extract. The presence only of bands ac and d indicate no cleavage at the authentic 3' splice site by oligonucleotide sp. Track 'sp+' is identical to track '30/8/9 sp' on the left hand plate.

The remaining two tracks are double cleavage experiments, directed by the oligonucleotides shown, to confirm the assignation of bands. They show that the predicted bands are generated.
Figure 29A. - Shows RNase H protection of both 3' splice sites of transcript 30/8/9 over a 120 minute time course.

Transcript 30/8/9 was set up as for a splicing reaction in extract LX8 and then RNase H cleavage was directed to both the authentic and synthetic (cryptic) 3' splice sites by oligonucleotides sp2 and sp respectively after incubation at 30°C for the indicated times (see Materials and Methods). The gel shown is a 7% polyacrylamide, formamide gel (see Materials and Methods) with Hpa II cut pBR322 markers (see appendix 1).

The presence of a 275 nucleotide band in the sp2 tracks (4, 6, 8, 10, 12) implies protection of the authentic site and a 275 nucleotide band in the sp tracks (2, 5, 7, 9, 11) implies protection of the cryptic site. Fig. 29B shows the expected cleavage products and lanes 1 to 4 show the generation of these products (sizes indicated, see Fig. 29B) after incubation for 5 minutes in the presence and absence (-) of extract LX8.

sp2 directed cleavage is incomplete on unprotected RNA (seen by the presence of a 275 nucleotide band in lane 1). However, oligonucleotide sp does direct cleavage to the synthetic site of transcript 30/8/9 (see Fig. 30A) and protection after incubation for 5 minutes in a splicing reaction is greatly enhanced (275 nucleotide band in track 2 is more intense than that in track 1. Also see Figs. 30A and 30B) and so the result is valid. Examination of the cleavage products (see Fig. 29B) confirms that sp and sp2 are directing cleavage specifically to the synthetic and authentic 3' splice sites respectively.

At the 5 minute time point both 3' splice sites are protected to about the same extent. After the 30 minute time point and beyond, the site which is ultimately chosen, i.e. the synthetic (cryptic) site remains well protected whereas protection of the unused, authentic site is reduced.

Track 13 is an uncleaved splice control. On a 7% gel the two uncleaved lariat species do not resolve and run just below the origin. The presence of exon 1 implies that at least the first step of the splicing reaction is working. No spliced products can be seen because LX8 has an exonuclease activity (see text). The bands above the transcript in lanes 7 to 12 are probably cleaved lariat bands (see Fig 29B) and are able to move faster through the gel than the uncleaved lariats in lane 13.
Figure 29B. Shows the predicted sizes of cleavage products when transcript 30/8/9 is cleaved by RNase H directed by oligonucleotides $sp$ and $sp^2$.

The transcripts are drawn as lariat intermediates to indicate the possible configuration of the cleaved lariat bands in lanes 7 to 12 of Fig. 29A.
3' splice sites of transcript 30/8/9 was examined. The presence of a 275 nucleotide band in the \textit{sp2} tracks implies protection of the authentic site and a 275 nucleotide band in the \textit{sp} tracks implies protection of the cryptic site. Fig. 29B shows the cleavage products and Fig. 29A lanes 1 to 4 shows the generation of these products in the presence and absence of extract LX8. In this experiment it can be seen (Fig. 29A lane 1) that \textit{sp} directed cleavage is not complete on unprotected RNA. Protection after incubation in a splicing reaction for 5mins is greatly enhanced and so the result can be considered valid. Examination of the cleavage products confirms that \textit{sp} and \textit{sp2} are directing cleavage specifically to the cryptic and authentic 3' splice sites respectively.

Fig. 29A shows that after the 5 minute time point both 3' splice sites are protected to the same extent. After the 30 minute time point and beyond, the site which is ultimately chosen, ie. the cryptic site, remains well protected whereas protection of the unused authentic site is greatly reduced.

The same 'used site' protection is seen on transcript Aut. In this case the evidence is more indirect as \textit{sp2} does not direct cleavage to the authentic 3' splice site (Fig. 30A). Comparing the intensities of the arrowed bands in Fig. 30A with those in Fig. 30B it can be seen that over a 60 minute time course, both 3' splice sites are protected to a certain extent and that where cleavage does occur, it is predominantly at the cryptic site. This implies that on transcript Aut, initially both 3' splice sites are protected. After 30mins, protection of at least one site is reduced and the RNase H cleavage products generated suggest that it is the authentic 3' splice site that is still predominantly protected ie. protection at the unused 3' splice site again is reduced.
Figure 30A. Shows the products when transcripts Aut and 30/8/9 are cleaved by RNase H directed by oligonucleotides sp and sp2 in the absence of protection by any splicing components.

The gel shown is a 7% polyacrylamide, formamide gel (see materials and methods). The tracks labelled ' - oligo' are controls showing negligible breakdown of the transcripts in the absence of the oligonucleotides. The sizes of the cleavage products are indicated.

Transcript 30/8/9 - Cleavage directed by sp: bands of 90 nucleotides and 185 nucleotides indicate cleavage only at the synthetic site. (See fig. 29B).

Cleavage directed by sp2: bands of 159 nucleotides and 117 nucleotides indicate cleavage only at the authentic site. (See Fig. 29B).

Transcript Aut - Cleavage directed by sp: The absence of the 185 nucleotide band (arrowed) which is replaced by smaller bands, indicates cleavage of both 3' splice sites.

Cleavage directed by sp2: The presence of an intense 275 nucleotide band indicates only inefficient cleavage (sp2 is directed against the nutated authentic 3' splice site of transcript 30/8/9). The presence of bands of 159 and 117 nucleotides indicates that the inefficient cleavage is to the authentic site.
Figure 30B. - Shows protection of both 3' splice sites of transcripts Aut and 30/8/9 over a 60 minute time course. The arrowed band indicates cleavage predominantly at the cryptic site rather than the authentic site.

The transcripts indicated were incubated under splicing conditions for the times indicated. RNase H cleavage was then directed against the two 3' splice sites by the oligonucleotides sp and sp2 as indicated (see materials and methods). The cleavage pattern was analysed by electrophoresis on a 7% polyacrylamide, formamide gel.

**Transcript Aut** - Oligo sp2 does not direct efficient cleavage of transcript Aut (Fig. 30A) so the Aut sp2 tracks may be disregarded. Oligo sp directs cleavage against both 3' splice sites of transcript Aut. The presence of a 275 nucleotide band in tracks Aut sp therefore indicates protection of both 3' splice sites on the same molecule. The presence of a 185 nucleotide band (arrowed) indicates preferential cleavage of the synthetic (unused) site (compare the Aut sp cleavage pattern to the Aut sp2 cleavage pattern of unprotected transcript in Fig. 30A, particularly the 185 nucleotide band arrowed in both figures.)

**Transcript 30/8/9** - Oligo sp directs cleavage almost exclusively to the synthetic (cryptic) 3' splice site, oligo sp2 directs cleavage almost exclusively to the authentic 3' splice site. The presence of a 275 nucleotide band indicates protection of the appropriate splice site. It can be seen that by 5 minutes both 3' splice sites are equally protected. By 30 minutes the unused synthetic site, (track sp) is slightly less well protected; by 90 minutes this difference is more pronounced.

In conclusion: In my model system both 3' splice sites are initially protected from oligonucleotide directed cleavage by RNase H. From 30 minutes complex binding at the unused 3' splice site is seen to begin to dissociate.
These results suggest that initially both 3' splice sites are recognised and bound by a component or components of the spliceosome complex. The site which is not chosen then becomes dissociated from the splicing component. This is discussed further in the discussion section.

7.3 Branch Point Protection on Transcripts 30 and 30/8/9.

The pattern of branch point protection when transcripts 30 and 30/8/9 were spliced in extract LX9 was not as expected. Fig. 31 shows a two point time course experiment to investigate branch point protection on both transcripts. The presence of a 275 nucleotide band in the 'A' tracks implies protection of the authentic branch site, a similar band in the 'N' tracks implies protection of the cryptic (new) branch site. The specificities of the oligonucleotide for the correct sites are shown in the '-extract' control lanes. The presence of a faint 275 nucleotide band in these lanes compared to the absence of such a band in, for example, the track 'A' of transcript 30/8/9 after 5 mins, implies that the RNase H cleaves more efficiently in the presence of extract or is perhaps augmented by the presence of endogenous RNase H in the extract.

Fig. 31 shows that on transcript 30/8/9 only the used branch point is ever protected. Protection has reached a maximum by 30 mins. In other experiments no trace of protection of the authentic branch point was ever detected between 5 and 90 minutes. The pattern of branch point protection for transcript 30 is different. After 5 minutes both the authentic and cryptic sites are protected to about the same extent. By 30 minutes, the protection at the unused site has decreased by about 60%, protection of the used authentic branch point has increased slightly to a
Figure 31. - Shows RNase H protection of both the authentic (A) and new (N) branch sites of transcripts 30 and 30/8/9. The presence of a 275 nucleotide band indicates protection.

Transcripts 30 and 30/8/9 were incubated under splicing conditions in extract LX9 for the indicated times. RNase H cleavage was then directed against the authentic (tracks A) and the cryptic (tracks N) branch sites of both transcripts. The -Ex tracks show cleavage directed by oligonucleotides Aut (tracks A) and New (tracks N) against the authentic and cryptic branch points respectively in the absence of splicing extract. The methodology was as described in the materials and methods, the gel shown is a 7% polyacrylamide, formamide gel.

Transcript 30/8/9 - The presence of a 275 nucleotide band only in the 'N' tracks and not in the 'A' tracks shows that only the cryptic (N) branch site is ever protected. The authentic site can be cleaved by RNase H directed by oligo Aut (A) at all times tested.

Transcript 30 - The presence of a 275 nucleotide band in both sets of tracks A and N at 5 minutes shows that at this time both branch sites are protected to a similar extent. After 30 minutes the 275 nucleotide band in the N track is less intense than that in the A track implying that efficient protection at the unused (in this case the cryptic) branch site is not maintained.

The high molecular weight bands (arrowed) present in all tracks from 5 minutes are running as would be expected for lariat or cleaved lariat molecules. However, it is unusual to see these species after a splicing reaction of only 5 minutes and therefore they have not been satisfactorily explained.
maximum level comparable with protection of the used site of transcript 30/8/9 at an equivalent time.

To investigate more thoroughly the pattern of branch site protection on transcript 30 a time course experiment over 90 mins was carried out (Fig. 32A). The results suggest that protection of both sites is initially the same after 5 minutes. Maximum protection is achieved at the used site after 30 mins and is maintained over the course of the 90 minute experiment. Protection at the unused site has decreased by 30 minutes and the site is virtually unprotected by the 90 minute time point.

7.4 Summary.

In summary, the following points can be made:

1) Both 3' splice sites are initially protected to about the same extent regardless of which site is to be used.

2) After about 30 minutes, protection of the unused 3' splice site has decreased but is still easily detectable. This could imply further assembly or increased commitment, perhaps by enhanced binding, of the complex at the 3' site which is to be used.

3) On a transcript where the cryptic site is exclusively used (i.e. an AG→CG mutation at the authentic site) there is protection only of the new cryptic branch point from 5 minutes. No protection of the authentic
Transcript 30 was incubated under splicing conditions for the times indicated. RNase H cleavage was then directed against both the authentic (A) and cryptic (N) branch sites by oligonucleotides Aut and New respectively (see Materials and Methods), the gel shown is a 7% polyacrylamide,formamide gel. Unfortunately this transcript had been stored for three days and had begun to decay which led to a high background on the gel.

The absence of a 275 nucleotide band at 0 minutes indicates that neither branch site is initially protected on ice. This observation is consistent with those of other groups (see Introduction and Discussion). By 5 minutes, protection of both branch sites is seen and this protection is maintained at the used (authentic, tracks A) branch site up to at least 90 minutes. Protection at the unused (tracks N) branch site is seen to begin to decrease by 30 minutes and this cryptic branch site is virtually unprotected by 90 minutes.
branch point ie. that which is not used, is ever detected between 5 minutes and 90 minutes.

4) On a transcript where the authentic site is exclusively used there is, after 5 minutes, protection of both branch points to the same extent. After 30 minutes protection of the unused branch point is seen to decrease rapidly until it is barely detectable after 90 minutes.

My results led to the following conclusions which are discussed further in the discussion section:

A) Choice of 3' splice sites, in this model system, is probably made after initial binding of a splicing component to the 3' splice site but immediately before branch point selection ie. the ultimate choice of branch point is dependent of which 3' splice site is chosen and not the other way round.

B) Although Frendewey and Keller (1985) have shown that the AG is not required for complex formation when there is only one 3' splice site, my results suggest that when there is a choice of 3' splice sites there is a requirement for the AG for any sort of complex formation beyond initial protection of the 3' splice site.
Chapter 8.

Investigation of Spliceosome Assembly Using Immunoprecipitation.

As discussed in the Introduction, it has been shown that transcript regions bound to the splicing complex can be precipitated by antibodies directed against U-type snRNPs. The antibody Y12 is directed against all the U-type snRNPs (anti-Sm) and the antibody AG is directed against U1 snRNP only (both were a gift from Dr. J.A. Steitz).

A further oligonucleotide (Imp) was made to direct cleavage by RNase-H to a region of the transcript immediately downstream of the authentic 3' splice site. As shown in Fig. 34 this oligo, when used in conjunction with oligo size cleaves the transcript into three fragments: a 5' splice site fragment which can be precipitated by both anti-sm and anti-U1, and two 3' splice site fragments each containing a complete 3' splice signal comprising of: the polypurimidine tract, the AG dinucleotide and the branch point. I predicted that both the 3' splice site fragments would initially be precipitated by anti-Sm and then, once spliceosome assembly was complete, U1 snRNP would associate with the complex at the used 3' splice site thus enabling anti-U1 to bring down that site.

8.1 Branch Point Protection and Immunoprecipitation of Transcript 30.

Although both 3' splice sites of transcript 30 are initially protected from cleavage by RNase-H, only the used 3' splice site fragment is initially immunoprecipitated by anti-sm. Fig. 35 shows that this ability to be precipitated must be attributable to a component bound to the 3' splice
Figure 34. Shows the three fragments of the *in vitro* β-globin transcript generated by *size* and *imp* directed RNase H cleavage.

In the following experiments, the mICEβ construct was linearised by *Bgl II* (see appendix 2) and transcribed *in vitro* (see Materials and Methods). The 5' fragment upstream of the region of complementarity to oligonucleotide *size* is 470 nucleotides and the total transcript length is 745 nucleotides.
1 - The authentic 3' splice site
2 - The synthetic 3' splice site
(1) - 3' fragment one, 145 nucleotides
(2) - 3' fragment two, 140 nucleotides
Figure 35. - Shows, by branch site protection and 3' splice site immunoprecipitation, that the authentic 3' splice site fragment of transcript 30 (see Fig. 34) is precipitated, at least initially, because of Sm precipitable factors bound to the 3' splice site rather than to the branch point.

Transcript 30 was incubated under splicing conditions for the times indicated and then processed as follows: - Markers are Hpa II cut pBR322 fragments. Tracks 1 and 6 are unprocessed splice reaction controls showing unspliced transcript (lane 1) and after 90 minutes (lane 6) the spliced products and intermediates can be seen (indicated). The two lariat species cannot be resolved on this 7% polyacrylamide gel and are seen as a single band below the origin.

Branch site protection - RNase H cleavage was directed against the Authentic and Cryptic branch points by oligos aut and new respectively at 0 minutes (tracks 4 and 5 respectively) and 90 minutes (tracks 9 and 10 respectively). The presence of a 275 nucleotide band indicates branch site protection. The absence of an intense 275 nucleotide band in tracks 4 and 5 indicates minimal branch site protection at 0 minutes on ice (see also Fig. 32). After 90 minutes, both branch sites are protected (tracks 9 and 10). The slightly more intense 275 nucleotide band in track 9 indicates that the authentic (used) branch site is better protected after 90 minutes.

Immunoprecipitation - Following incubation under splicing conditions, the transcript was cleaved into three fragments (see Fig. 34) and immunoprecipitated by anti-Sm antibodies. Tracks 3 and 8 have been cleaved following incubation but not immunoprecipitated. These show the expected positions of the three fragments (indicated 1, 2, and 5' frag, see Fig. 34) and confirm that the transcript is accessible to oligos imp and size under splicing conditions. Track 2 shows that at 0 minutes the authentic 3' splice site fragment (fragment 1) and the 5' fragment are efficiently precipitated. This is as expected (see text). After 90 minutes incubation under splicing conditions (track 7), the authentic 3' splice site fragment (fragment 1) and the 5' fragment are precipitated more efficiently this is also as expected. Fragment 2 is possibly precipitated inefficiently (see text). The presence of a 275 nucleotide band indicates incomplete cleavage directed by oligo imp. The other bands which can be seen may imply some non specific cleavage directed by one of the oligos, or they may be degradation products. They have not been satisfactorily explained.
site because at 0mins on ice, neither branch point is protected. The authentic branch point is protected at 90 minutes. The unused 3' splice site fragment is not precipitated at 0mins which is surprising as I have shown that the unused 3' splice site is protected from RNase H cleavage at 0mins. After 90mins, a fragment is precipitated (lane 7) which could be the unused 3' splice site fragment; fragment 2, it could also be the spliced product or lariat intermediate of transcript 30 cleaved by oligonucleotide imp and RNase H. For this to be precipitated it must remain bound to the complex and for this reason it is most likely to be either genuine complex binding to the unused site or an artefact generated by RNase H cleavage of the lariat intermediate. It is unlikely that spliced product would remain bound to the complex tightly enough to be precipitated.

Fig. 35 shows that material which can be cleaved by RNase H is complex bound. If this were not so then cleaved material would not be precipitated. It is likely that as the unused splice site is initially protected from RNase H cleavage but cannot be precipitated by anti-Sm antibodies, that in the first instance, protection of the 3' splice site is by a non-snRNP component. If the site is chosen to be used it is almost immediately bound by a second component (probably U5 snRNP) which can be precipitated.

8.2 Pattern of Spliceosome Assembly on Used 3' Splice Site.

It was predicted that if a complete spliceosome was assembled on either 3' splice site fragment, then that fragment would be precipitated by anti-U1 antibodies.
Following incubation in a splicing reaction for 90 mins, material from transcript 30 was cleaved by oligonucleotides imp and size and RNase H and immunoprecipitated using either anti-Sm or anti-U1 antibodies. It can be seen from Fig. 36 that both antibodies precipitate fragment 1 after 90 minutes, implying complete spliceosome assembly at the used 3' splice site. There is a trace of a band below fragment 1 in lanes 6 and 7. This could indicate some complex assembly at the cryptic site. After 90 mins, the cryptic branch point is not protected from RNase H cleavage and protection of the cryptic 3' splice site is reduced relative to protection of the used 3' splice site. This suggests that the band in question is more likely to be RNase H cleaved spliced product or lariat intermediate as discussed elsewhere.

Precipitation of the used splice site fragment by anti-U1 at zero minutes is unclear (Track 3). Chabot and Steitz (1987) have shown no precipitation of a 3' splice site fragment by anti-U1 at zero minutes. All my data shows some precipitation of the 3' splice site fragment 1 by anti-U1 at zero minutes. Sometimes, as by comparing lanes 2 and 3 in Fig. 37A there is more precipitation by anti-Sm than by anti-U1. At other times (Fig. 37B; lanes 1 and 2) there is no difference between the amounts of fragment 1 brought down by anti-Sm or anti-U1. It is clear, however, that at zero minutes there is never any tight association between U1 snRNP and the cryptic 3' splice site.
Figure 36. - Shows immunoprecipitation of transcript 30 by anti-Sm and anti-U1 antibodies following incubation under splicing conditions for 0 or 90 minutes.

Transcript 30 was incubated under splicing conditions for the times indicated. The RNA was then cleaved into three fragments (indicated 1, 2 and 5', see Fig. 34) and immunoprecipitated with either anti-Sm or anti-U1 antibodies (tracks 2, 3, 6, 7). Tracks 1 and 5 are material which has been incubated under splicing conditions and then cleaved into three fragments but not immunoprecipitated. These serve as markers and confirm that the RNA is accessible to oligos size and imp and RNase H. Tracks 4 and 8 are splice reaction controls to show that the extract is working. The spliced product after 90 minutes is indicated. The two lariat species cannot be resolved on this gel (7% polyacrylamide) and in this experiment the free exon 1 band cannot be seen, it may have run off the bottom of the gel. Markers on this gel are Hpa II cut pBR322 fragments (the methodology was as described in the materials and methods).

At 0 minutes anti-Sm precipitates 3' fragment 1 and the 5' fragment (track 2) anti-U1 precipitates only the 5' fragment (track 3). This is as expected (see text). After 90 minutes anti-Sm precipitates 3' fragment 1 and the 5' fragment very efficiently. A small amount of fragment 2 may also be precipitated but very inefficiently and there are alternative explanations for this band (see text). Anti-U1 also precipitates the 5' fragment and 3' fragment 1 indicating association of U1 snRNP with the authentic (used) 3' splice site. A band which could indicate association between U1 snRNP and the synthetic 3' splice site (fragment 2) is also precipitated (see text).

The presence of a 275 nucleotide band in tracks 6 and 7 indicate incomplete cleavage directed by oligo imp. Other bands have not been explained.
Figure 37. - Shows two experiments (A and B) which investigate the immunoprecipitability of transcript 30 by anti-Sm and anti-U1 antibodies at 0 minutes under splicing conditions.

Transcript 30 was set up under splicing conditions on ice and immediately cleaved into three fragments (see Fig. 34) and immunoprecipitated with anti-Sm and anti-U1 antibodies. The methodology was as described in the materials and methods. The gels shown are 7% polyacrylamide, formamide gels.

Track A1 shows transcript 30 cleavage directed by oligos size and imp. The two 3' fragments and the 5' fragment are indicated. This track is overexposed. Tracks A2 and B1 are immunoprécipitations at 0 mins by anti-Sm. The 5' fragment and 3' fragment 1 are precipitated as expected (see text, also Figs. 35 and 36). In both experiments shown here anti-U1 is also able to precipitate 3' splice site fragment 1 at 0 minutes (lanes A3, B2). This implies association between U1 snRNP at the 3' splice site at 0 minutes on ice. This data is inconsistent with the findings of other groups (see text) and also that presented in Fig. 36. However, a longer exposure of Fig. 36 may have indicated an association between U1 snRNP and the 3' splice site at 0 minutes.
8.3 Spliceosome Assembly on a Transcript with an Activated Cryptic 3' Splice Site.

Comparing the anti-Sm and anti-U1 precipitated fragments of transcripts 30 and 30/8/9 following incubation in a splicing reaction for 0 minutes and 180 minutes, gave a very unexpected result. From Fig. 38 it can be seen that both anti-Sm and anti-U1 antibodies precipitate only the 3' splice site fragment 1 of both transcripts at 0 minutes. This means that there are no snRNP components tightly associated with the cryptic splice site at zero minutes even when that site is used. After 180 minutes, fragment 1 is still the predominant 3' splice site fragment precipitated from either transcript by either antibody. Fragment 2 may also be precipitated after 180 minutes but again this band could be accounted for in other ways; i.e. in lanes 5 and 6 it could be the spliced product or lariat intermediate cleaved by RNase H. The presence of a second small band (arrowed) is consistent with the hypothesis that cleaved spliced product has been detected. The band in lanes 7 and 8 could be transcript 30/8/9 spliced product as seen when compared to the 180 minute splice control (lane 10).

8.4 Summary.

This data is not consistent with the RNase H protection data presented in chapter 7. The immunoprecipitation data presented in this chapter suggests that, whether or not the cryptic site is used, it cannot be immunoprecipitated efficiently and perhaps not at all. One explanation of this is that the position of complex formation on the transcript is not
Figure 38. Shows an immunoprecipitation of transcripts 30 and 30/8/9 by both anti-Sm and anti-U1 antibodies after incubation under splicing conditions for 0 or 180 minutes.

Transcripts 30 and 30/8/9 were incubated under splicing conditions for the times indicated. The RNA was then cleaved into 3 fragments (see Fig. 34) and immunoprecipitated by either anti-Sm or anti-U1 antibodies as indicated. The gel is 7% polyacrylamide with Hpa II cut pBR322 markers. Also shown, to the left of track 1 is transcript Alt cleaved using oligos size and imp and RNase H. the material in this track has not been immunoprecipitated. The two 3' fragments (1 and 2) and the 5' fragment (5') are indicated. Tracks 9 and 10 show splicing reaction controls of transcripts 30 and 30/8/9 after incubation under identical conditions to those used prior to immunoprecipitation. Products and intermediates of the splicing reaction are indicated. The methodology was as described in the materials and methods.

At 0 minutes (tracks 1, 2, 3, 4) both anti-Sm and anti-U1 precipitate the 5' fragment and 3' fragment 1 of both transcript 30 and 30/8/9. This implies that there are no snRNP components tightly associated with the synthetic (cryptic) splice site at 0 minutes even when that site is used. This is contradictory to the data generated by the RNase H protection experiments (see chapter 7). Factors may associate weakly with the cryptic 3' splice site and be dislodged by the immunoprecipitation procedure. After 180 minutes (tracks 5, 6, 7, 8), fragment 1 is still the predominant 3' splice site fragment precipitated from either transcript by either antibody. Fragment 2 may also be precipitated after 180 minutes but this band may be accounted for in other ways. In tracks 5 and 6, the presence of the small band arrowed is consistent with the hypothesis that cleaved spliced product has been detected. The "fragment 2 band" in lanes 7 and 8 could be transcript 30/8/9 spliced product as seen when compared to the 180 minute splice control in track 10.

At best the data presented here implies only a very weak association of snRNP components with the cryptic 3' splice site. This association can only be detected at the 180 minute time point.

The pattern of complex association with the 3' splice site is unaffected by the AG+CG mutation in transcript 30/8/9 or by the subsequent choice of splice site. This data does not necessarily reflect the association of components with either branch site (see Discussion, also Fig. 35)
affected by which site is used. The pattern of RNase H protection however varies predictably depending on whether a 3' splice site is used or not. These observations are discussed in the Discussion section.
Chapter 9.

Isolation of Spliceosome Complexes by Velocity Gradient Sedimentation.

9.1 Introduction

To determine whether the RNase H protection pattern obtained from whole splicing reaction RNA (Chapter 7) and the immunoprecipitation data presented in chapter 8 accurately reflect events within the spliceosome, it was decided to isolate assembled or partially assembled spliceosome complexes by sucrose gradient sedimentation.

The sucrose gradients were set up by the freeze thaw method described in the Materials and Methods section to be approximately 5% to 20% gradients. The system was initially calibrated to resolve 30S and 50S E. coli ribosomal subunits, the centrifugation time being adjusted until both peaks were resolved but not forced to the bottom of the gradients (Fig. 39A).

Scaled up three fold splicing reactions were used to generate material for sedimentation. These reactions were diluted to 200μl in ribosome dialysis buffer and subjected to brief centrifugation to remove precipitated material which tended to sink through the sucrose. The material was then layered onto the gradients. Following centrifugation the gradients were displaced by the introduction of a 50% sucrose solution stained slightly with bromophenol blue, and collected from the top in ten drop fractions. Fig. 40 shows the home made apparatus which I developed to unload material from the gradients.

-124-
Figure 39. A: Shows a profile of a marker gradient. 30S and 50S peaks of *E. coli* ribosome subunits are shown.

B: Shows a profile of a sucrose gradient of a partially assembled splicing complex on transcript 30 after a 45 minute splicing reaction.

C: As B above, but for transcript 30/8/9.
Figure 40. Shows the home made apparatus designed to unload sucrose gradients from ultra-centrifuge tubes.
50% Sucrose pumped from reservoir by micro perpex pump

Rubber bung

Sucrose gradient
Stainless steel cannula

50% Sucrose stained with bromophenol blue

Gradient fractions collected
Spliceosome gradients were analysed by Cerenkov counting of the fractions and the ribosome subunit gradients were analysed by reading the absorbance of the fractions at 260nm.

9.2 Assembly of Splicing Complex.

Fig. 39 B and C, suggests that complex formation is only partially completed after a 45 minute incubation under splicing conditions of transcripts 30 and 30/8/9. The data implies that complex assembly on transcript 30 is slightly more efficient than that on transcript 30/8/9. This is consistent with the observation that the in vitro splicing efficiency of transcript 30 is greater than that of transcript 30/8/9. It is also apparent that as there is no material with a sedimentation coefficient greater than 50S, it is likely that only a single complex is assembling on transcript 30.

9.3 Analysis of Material Within Partially Assembled Splicing Complexes.

For analysis, fractions were pooled as shown below:

<table>
<thead>
<tr>
<th>Sample</th>
<th>Transcript</th>
<th>Fractions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>13 &amp; 14</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>15 &amp; 16 &amp; 17</td>
</tr>
<tr>
<td>3</td>
<td>30/8/9</td>
<td>11 &amp; 12 &amp; 13</td>
</tr>
</tbody>
</table>

Pooled fractions were diluted in 'D' buffer and divided into aliquots for analysis. One aliquot of each sample was phenol extracted and then
analysed by polyacrylamide gel electrophoresis. Fig. 41B; Tracks 4, 8 and 12, show that in all samples only unspliced precursor RNA was present. This confirms that the splicing complex was only partially assembled. Intermediates and some products of the splicing reaction can be detected in active spliceosomes isolated from sucrose gradients (Frendewey and Keller 1985).

9.4 Nuclease Protection of Material Within Assembling Splicing Complexes

The fractions were tested for protection from cleavage by RNase H directed against both branch points and both 3' splice sites by oligonucleotides aut, new, and sp. The sizes of possible cleavage products are shown in Fig. 41A. Fig. 41B shows that all the predicted fragments are generated. The RNase H is not digesting to completion but examination of the partial digest products shows that neither 3' splice site is preferentially protected. The presence of a particular band does not necessarily mean that a particular site is protected within the spliceosome.

It is probable that no protection can be detected because the spliceosome complex has become dissociated during the course of the experiment. Conditions must be found under which the complex remains integral and tightly bound to the RNA so that splice site protection within the spliceosome can be properly investigated.
**Figure 41A.** Shows the sizes of possible fragments generated by oligonucleotide directed RNase H cleavage of transcript 30 and 30/8/9 material isolated from spliceosome gradients.

The oligonucleotides used to direct cleavage are indicated. Only fragments downstream of the region complementary to oligonucleotide size are shown.
Figure 41B. - Shows RNase H protection analysis of partially assembled splicing complexes on transcripts 30 and 30/8/9 isolated from the sucrose gradients shown in Figure 39.

Transcripts 30 and 30/8/9 were incubated under splicing conditions for 45 minutes before being subjected to velocity gradient sedimentation. The gradient profiles are shown in Fig. 39 B and C. Gradient fractions were pooled as described in section 9.3 and summarised here:

<table>
<thead>
<tr>
<th>Sample</th>
<th>Transcript</th>
<th>Fractions</th>
<th>Tracks on gel</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>13, 14</td>
<td>1, 2, 3, 4</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>15, 16, 17</td>
<td>5, 6, 7, 8</td>
</tr>
<tr>
<td>3</td>
<td>30/8/9</td>
<td>11, 12, 13</td>
<td>9, 10, 11, 12</td>
</tr>
</tbody>
</table>

One aliquot from each sample was analysed simply by gel electrophoresis on a 7% polyacrylamide gel with Hpa II cut pBR322 markers. As shown in lanes 4, 8 and 12 the gradient fractions analysed contain only unspliced full length RNA transcript.

The gradient fractions were tested for protection from cleavage by RNase H directed against both branch points by oligonucleotides \textit{aut} (against authentic branch point) and \textit{new} (against cryptic branch point), and the 3' splice sites by oligonucleotide \textit{sp}. \textit{sp} directs cleavage against both 3' splice sites of transcript 30 but predominantly the synthetic 3' splice site of transcript 30/8/9 i.e. the site which is used. This is confirmed by the greater intensity of the 176 nucleotide band in track 11 compared to that in tracks 3 and 7. The sizes of possible cleavage products are shown in Fig. 41A and the sizes are indicated on the gel. This figure shows that all the predicted fragments are generated. The RNase H has not digested to completion but examination of the partial digest products shows that neither 3' splice site nor branch point seems to be preferentially protected from cleavage (see Discussion).
9.5 Summary.

The work presented in this chapter is a preliminary investigation of RNA:complex interactions within fully and partially assembled spliceosomes. It is likely that a single splicing complex is assembled on each of the transcripts and that the assembly is more efficient on a transcript which splices to the authentic 3' splice site. A method to examine protection from RNase H cleavage of material within isolated spliceosomes has been devised but not yet perfected.

This work is further discussed in the Discussion.
Chapter 10 Discussion.

This work describes the preliminary development of a model system to investigate the selection of alternative splice sites in vivo and in vitro. As a basis on which to build the system, the rabbit β-globin gene was chosen. The rabbit β-globin gene has been well characterised and its splicing pattern in vivo has been well established in HeLa cells.

It was decided to use the 5' and 3' splice sites of the second intron of the rabbit β-globin gene as constant reference sites against which other splice site sequences could be measured.

We have successfully used our model system to investigate the intrinsic strengths of 5' splice site sequences. Several synthetic splice sites were inserted into the Bam HI restriction site eighteen nucleotides upstream of the authentic 5' splice site (Eperon et al. 1986). This strategy could control against the effects of splice site position and so enable the 5' splice site sequences tested to be ranked in a hierarchy of preferential use.

The work presented in the first part of this thesis is an attempt to investigate the intrinsic strengths of 3' splice site sequences. To this end, synthetic oligonucleotides conforming to the 3' consensus sequences, and allowing for variations within this sequence, were inserted into the Eco RI restriction site forty-nine nucleotides downstream of the authentic 3' splice site of the second intron of the rabbit β-globin gene.
10.1 Efficiency of Cloning Oligonucleotides.

The efficiency of insertion of the synthetic oligonucleotides into the Eco RI site of the rabbit β-globin gene was low. A maximum insertion efficiency of 30% was recorded, about 15% was more typical. The optimum efficiency of insertion was achieved by using 0.2 pmoles of oligonucleotide and 0.1 μg of phosphatase treated vector in the ligation reactions (see Results). Using less oligonucleotide resulted in a reduced efficiency of insertion. Doubling the amount of oligonucleotide present in the reaction also reduced the efficiency of insertion. This is probably due to having reached a critical concentration at which the oligonucleotides preferentially form multimers rather than inserting into the vector. I predict that further increasing the concentration of oligonucleotide would lead to a higher proportion of multiple inserts; this was not investigated.

10.2 The Microprep Screen.

The microprep screen had to be developed because of the low efficiency of oligonucleotide insertion. A method was required for screening large numbers of recombinant plasmids for a small insertion for which there was no phenotypic marker. The development and use of the microprep screen and of the primer extension method employed to check the orientation of inserted oligonucleotides are described in the results so will not be repeated here.
10.3 Use of Alternative 3' Splice Sites in Vivo.

The results presented show that the 21 bp synthetic 3' splice sites are insufficient to direct splicing away from the authentic 3' splice site. This could be due to a number of reasons. Perhaps the synthetic 3' splice sites are not accessible to be used as splice sites in vivo. They may be complexed within a heterogeneous nuclear ribonucleoprotein particle (hnRNP), or perhaps the synthetic sites are involved in some secondary structure which precludes their use as splice sites.

10.4 The Influence of Secondary Structure on Splice Site Selection.

Secondary structure has been shown to influence splice site selection in vivo. Our own work (Eperon et al. 1986) has shown that if a copy of the rabbit β-globin IVS-2 5' splice site is inserted into the Bam HI site of the rabbit β-globin gene (Fig. 42A) then this synthetic site is chosen in 64% of spliced transcripts. When a second oligonucleotide, complementary to the first, was inserted (Fig. 42B) the site usage dropped from 64% to 0%, and the reference site was used exclusively. The relative positions of the test and reference sites had remained unchanged. We postulated that the test site was now involved in forming a secondary structure with the complementary region and so was unavailable for splicing.

In my constructs it is unlikely that the synthetic 3' splice sites are directly involved in forming secondary structure. Several different synthetic sequences were tested and none were chosen over the authentic reference site for use as splice sites. Any direct involvement of the
Figure 42. Shows the rabbit β-globin gene construct used to test the intrinsic strength of 5' splice site sequences (Eperon et al. 1986). The boxes represent exons. If a copy of the authentic 5' splice site is inserted into the Bam HI site (A) then this synthetic site (site 1) is chosen in 64% of transcripts. When a second oligonucleotide, complementary to the first, was inserted (B) then the site usage fell to 0%. The authentic site (site 2) was exclusively used.
synthetic splice site in forming secondary structure would probably only
preclude one of the sequences from availability as a splice site. Even
complementarity to the synthetic site polypyrimidine tract can largely be
ruled out. pβAcc2 has a C rich tract whereas pβAcc3 and pβAcc4 have T
rich tracts.

Secondary structure not directly involving the splice sites can also
influence splice site selection in vivo. Solnick (1985b), modified a
splicing precursor such that it could form simple and stable secondary
structures. An exon was looped out in a structure not directly involving
the splice sites. This led to a jump splice in vitro and in vivo. Although
the results were qualitatively the same, only 14% of spliced RNA
in vivo lacked the looped exon compared to 96% of that spliced in vitro.

Solnick (1985b) goes on to suggest that an mRNA precursor with two
tandem splice sites, either 5' or 3', might sometimes fold such that one of
the sites is in a loop. The precursor would generate alternatively spliced
RNAs as a consequence of the alternative RNA structures.

To use this model to explain my observations in vivo would require
that the synthetic 3' splice site was always looped out of the way.
Solnick's constructs had extensive stable secondary structure, the stems of
the hairpin loop consisted of 105 nucleotide restriction fragments. With
this extensive secondary structure, only 14% of the mature mRNA was
alternatively spliced. There are no suitable regions of extensive
complementarity within my constructs which could form appropriate
secondary structure.

Therefore, it is unlikely that my observations in vivo will be explained by either secondary structure directly involving the splice site
or affecting the juxtaposition of the 5' site and synthetic 3' site by looping the synthetic 3' site out of reach of the splicing mechanism.

10.5 The Importance of Position and Context in Splice Site Selection.

As mentioned earlier, the sequence and position of two or more splice sites are thought to be in a delicate balance. The evidence for this is detailed in the Introduction. What is not really understood is how the context within which potential splice site sequences are located can influence the balance and so determine which splice site is chosen.

In my constructs the synthetic 3' splice sites are inserted into the Eco RI restriction site 49 nucleotides downstream of the authentic IVS-2 3' splice site. This position is in the middle of exon 3, flanked on both sides by exon sequences.

It is possible that the 21 nucleotide synthetic splice site does not contain enough sequence to constitute a complete splice site. This is not necessarily inconsistent with my findings that transcript 30/8/9 is spliced to the synthetic site in vivo and in vitro. 30/8/9 is a modified transcript with an AG→CG mutation at the authentic 3' splice site. Splicing to the synthetic or cryptic site is at a reduced level in vitro and in vivo. The immunoprecipitation data is identical whether the authentic or the synthetic 3' splice sites are chosen and indicates complex formation at the authentic 3' splice site. This evidence could imply that splicing to the cryptic site involves sequence elements at the authentic site for complex formation. Perhaps there is some poorly conserved sequence element or subtle secondary structure which is required
for efficient splice site usage and which is missing from the synthetic sequence.

Reed and Maniatis (1986) have shown that exon sequences can influence the selection between duplicated splice sites in vitro. The model system they use is a cis-competition assay very similar to mine.

By tandem duplication of the 5' and 3' splice sites of the human β-globin IVS-1, Reed and Maniatis (1986) demonstrate the effects of truncations and substitutions on the relative use of the competing sites (Fig. 43 shows the strategy employed). Efficient splicing is seen to site 1 in all cases in the absence of a competing site, i.e. the truncated site is a full splice site but the strength is in part determined by exon sequences.

To show that the observations could not be attributed to the spatial separation of the sites, Reed and Maniatis (1986) replaced the internal exon by other material derived from exons, introns and human β-globin 3' flanking material in both orientations. In most cases, splicing was to the external site: site 2. In these constructs, if the distance between the two 3' splice sites was the important factor, splicing would be predicted to be to site 1.

That the internal 3' splice site is exclusively used in precursors containing the normal full length exon adjacent to both duplicated sites, while most other sequences at this position inhibit the use of the internal site, suggests that the proximity of the 5' and 3' splice sites to one another is also an important criterion for choosing splice sites in these precursors. The data suggests that the proximity of the 5' and 3' sites is of secondary importance compared with the flanking exon sequences.
Figure 43. Shows the cis-competition splice site assay system used by Reed and Maniatis (1986). Exons are represented by boxes. Duplicated sequences are represented by shaded areas. The sizes of 3' duplications and their effects on 3' splice site usage are shown.
size of duplication

3' site usage

14  □  □
55  □  □
115 □  □
205 □  □

increasing splice site usage
Similar experiments involving duplicated 5' exons showed that most internal exon substitutions resulted in the use of the external site. However, the internal site was used when the full length duplicated exon was adjacent to both sites. This is evidence that for 5' splice site selection, as for 3' splice site selection, the proximity of a 5' to a 3' site is of secondary importance compared with the flanking exon sequence.

When both 5' and 3' exon duplications were made in the same construct, the results were as predicted from the 5' and 3' exon duplications alone. This implies that the effects of exon sequence are localised to the adjacent splice site.

Even small changes in the exon sequence have been shown to result in significant differences in splice site selection (Reed & Maniatis 1986). A human β-globin transcript consisting of exon 1, IVS-1 and two tandem repeats of exon 2 (Fig. 44A) has been shown to splice to the internal 3' splice site; as predicted from the work discussed above. When the internal exon was replaced by the highly homologous mouse exon (Fig. 44B), a decrease in usage of the internal 3' splice site was observed.

This detailed study of the involvement of exon sequences in splice site selection was made in vitro. Therefore, it is not known whether flanking exon sequences have so pronounced an influence on splice site selection in vivo.

There is some evidence that the context in which an intron is located can influence splice site usage in vivo. The herpes simplex virus thymidine kinase gene normally contains no introns i.e. it is not processed by splicing. When the human β-globin IVS-1 was placed within the 3' untranslated region of the herpes virus thymidine kinase gene, it was not spliced in vivo (Greenspan & Weissman 1985). This could imply that the
Figure 44. Shows that in a human β-globin construct consisting of exon 1, IVS-1 and two tandem repeats of exon 2. The internal 3' splice site is always used in vitro (A). When the internal exon is replaced by the highly homologous mouse exon (B), a decrease in the internal 3' splice site usage is observed. The exons are represented by boxes. The splicing patterns in vitro are shown.
sequence or structural environment within the non splicing thymidine
kinase gene is unsuitable for its use as a splicing substrate. Therefore,
the human \(\beta\)-globin IVS-1 placed within this gene may not be spliced
because the context is wrong.

If the flanking exon sequences are important for splice site
selection \textit{in vivo}, then this influence must be considered when developing
a splice site sequence assay system. My splice site sequence assay system
does not take the flanking exon sequences into consideration. This may be
why the rabbit \(\beta\)-globin IVS-2 authentic 3' splice site is always chosen
over the synthetic 3' splice site sequence under test.

10.6 The Effects of 3' Splice Site Point Mutations on 3' Splice Site
Selection \textit{in vivo}.

To investigate the contribution of specific nucleotides to the
intrinsic strength of the rabbit \(\beta\)-globin IVS-2 3' splice site (the
reference site in my competitive assay system) a series of mutants was
made (see results, Fig. 18). Of the mutants tested only an AG\textendash CG mutation
at the authentic 3' splice site abolished splicing to that site \textit{in vivo} or
\textit{in vitro}.

In addition, all the mutants tested, except mutant 28 had two
mutations at positions 1101 and 1103 within exon 3 of the rabbit \(\beta\)-globin
gene construct. These mutations changed the sequence from CTCAT to CTAAC.
The newly created sequence was a copy of the IVS-2 authentic branch site.

It was thought that perhaps the synthetic 3' splice site sequences
were never used because they were not seen as an intact splice signal. It
was hoped that the introduction of a suitable branch point would allow the synthetic sequence to be recognised as a splice site.

A G→T mutation at position 1116 within the third β-globin exon was to eliminate the AG dinucleotide between the newly created branch site and the AG dinucleotide of the synthetic sequence inserted in the Eco RI restriction site. This mutation was to remove the possibility that a mutation at the authentic 3' splice site might activate a cryptic 3' splice site at position 1116. The use of this AG dinucleotide as a cryptic 3' splice site in vivo has been reported previously (Aebi et al. 1986).

Aebi et al. (1986) showed that a mutation at the AG of the authentic IVS-2 3' splice site of rabbit β-globin led to the activation of a cryptic 3' splice site in vivo. This cryptic site was at the AG dinucleotide at position 1115/1116. Splicing to the cryptic site was at 40% to 80% of the wild type level. However, my construct pβ30/8, when transiently expressed in HeLa cells and the splicing patterns determined by S1 mapping, gave no detectable products.

The differences between pβAut30/8 and the construct of Aebi et al. (1986), discussed above, are only slight: pβ30/8 has a 21 nucleotide copy of the authentic rabbit β-globin IVS-2 3' splice site cloned into the Eco RI restriction site in exon 3; and two mutations at positions 1101 and 1103 to generate a copy of the authentic IVS-2 branch site.

Bearing in mind the work of Reed & Maniatis (1986), it is possible that the slight sequence differences between pβ30/8 and the Aebi et al. (1986) 3' splice site mutant might account for the different splicing efficiencies and site usage observed in vivo. However, I wish to propose an alternative explanation for the differences observed by myself and by Aebi et al. (1986). I suggest that the important difference between my
construct, p830/8, and that of Aebi et al. (op cit) is that p830/8 has a good branch sequence between the mutated 3' splice site and the cryptic site observed by Aebi et al. (op cit).

I have shown that the new branch point of p830/8 is recognised and used when spliced in vitro. The polyacrylamide gel mobilities of the lariat species generated from transcript 30/8 in vitro are identical to the mobilities of the lariat species generated from transcript 30/8/9. The lariat bands of transcript 30/8/9 were excised from the gel and the branch point mapped by primer extension to the A at position 1102. This branch is at the predicted residue and indicates that the newly created branch sequence is recognised and used in vitro. It may be assumed that the same branch site is used on both transcripts.

I have also shown by RNase H protection analysis that on transcript 30/8/9 the new branch site is protected from cleavage whereas the authentic branch site is not. Even on transcript 30, which has a new branch site but does not have a mutated authentic 3' splice site, the new branch site is initially protected from cleavage by RNase H in vitro.

The construct of Aebi et al. (1986) has no candidate branch point sequence between the mutated AG at the 3' splice site and the AG which is activated as a cryptic 3' splice site. Aebi et al. (op cit) showed that their 3' splice site mutant abolished splicing in vitro and that the small amount of lariat intermediate produced had the same gel mobility as the wild type lariat intermediate.

It is likely that step 1 of splicing as observed by Aebi et al. (1986) involved lariat formation to the authentic IVS-2 branch site. In vitro, this lariat intermediate was formed at a very reduced level and no spliced products were detected; in vivo a cryptic 3' splice site was
activated. Aebi et al. (1986) suggest splicing to the cryptic site is so ineffective that it requires natural (presumably higher) concentrations of splicing components, or perhaps the conformation of the RNA is different in the nascent and the purified states.

Although the AG→CG mutation in transcript 30/8 seemed to abolish the generation of spliced product in vivo, it did not abolish splicing in vitro; both lariat molecules were detected by polyacrylamide gel electrophoresis implying that both steps of the splicing reaction were being carried out. No spliced product was detected in vitro. Possible reasons for this are discussed in the results section.

The usual distance between a 3' splice site and a branch point is between twenty-two and thirty-seven nucleotides (see introduction). To use the cryptic 3' splice site observed by Aebi et al. (1986) means that the AG dinucleotide is separated from the branch point by an additional forty-nine nucleotides. This results in a very inefficient reaction which requires the high concentrations of splicing components present in vivo before splicing can proceed at all. A partial splicing complex probably assembles at the new branch point of pβ30/8 (see Results). This might be sufficient to disrupt the inefficient splicing reaction so it can no longer proceed.

I propose that splicing of transcript 30/8 in vivo is not detected because the AG at position 1115/1116 and the new branch point together are not recognised as a 3' splice signal in their own right. Alternatively, the synthetic 3' splice site may be recognised but the presence of the second AG dinucleotide at position 1115/1116 may interfere with splicing to the AG dinucleotide in the synthetic sequence cloned into the Eco R1 site. The second alternative is consistent with the
observation that a construct with an additional mutation eliminating the AG dinucleotide at position 1115/1116, pβ30/8/9, is spliced to the synthetic splice sequence in vivo.

However, transcript 30/8 is spliced in vitro. This could imply that certain requirements for a splice site to be used differ in vivo and in vitro.

10.7 Future Work to be Done In vivo.

The work done in vivo for the first part of this thesis has presented some interesting questions.

When transiently expressed in HeLa cells, some of my constructs were seen to splice at a very reduced level. Some mature mRNAs could not be detected at all by my S1 nuclease mapping procedure. This inefficient splicing was seen from the following constructs:- pβ30, pβ30/8/9, pβ30/8 and pβ30/9. Some speculations as to why these particular constructs may splice inefficiently are made elsewhere in the discussion and the results section. Other possible reasons for a low level of detectable mature mRNA, and experiments to test these hypotheses are discussed here.

It is possible that either the transfection of the HeLa cells with DNA or the harvesting of the RNA following transient expression was very inefficient in some experiments. This is unlikely as experiments with pβAut, pβ30 and pβ30/8/9 gave reproducible results. However, it is possible that particular preparations of DNA may be inherently worse than others for transfection. This can be put down to slight variations in the pH and/or the salt and RNA concentrations of individual preparations. If this is true then multiple transfections of DNA from the same batch might
yield artefactual results which could be interpreted as a reduced splicing efficiency.

This can be controlled against in two ways. Firstly the experiments could be repeated several times using different DNA preparations for each transfection. The second and perhaps more elegant control is to map a marker on the same plasmid as the \( \beta \)-globin gene construct under test. As mentioned in the results, the \( p\beta SV \) plasmids also encode the SV40 T antigen. This was used in immunofluorescence experiments to initially set up the transfection system. One can assume that the steady state level of SV40 T mRNA is relatively constant in transfected HeLa cells. Making this assumption, one can measure the steady state level of \( \beta \)-globin mRNA as a ratio against the level of SV40 T mRNA. This experiment could be done by Northern blotting the HeLa RNA using a universally labelled probe comprising of part of the SV40 T gene and part of the rabbit \( \beta \)-globin gene. Northern blotting must be used rather than the much simpler technique of dot blotting for the reasons discussed below.

A genuine low level of detectable spliced mRNA is easy to explain but the explanations are hard to test experimentally \textit{in vivo}.

The transcript may splice inefficiently. If this is so then one might expect an accumulation of precursor mRNA \textit{in vivo}. In my experiments unspliced pre-mRNA would be indistinguishable from undigested S1 mapping probe. To investigate the accumulation of precursor mRNA by S1 mapping, a different probe would have to be used.

Accumulation of precursor could also be investigated by Northern blotting. By using an appropriate probe, the 'normal' ratio of SV40 T mRNA : \( \beta \)-globin pre-mRNA could be established. The RNA levels in HeLa cells transfected with modified \( p\beta SV \) constructs could also be measured.
The levels of β-globin pre-mRNA and mRNA relative to the level of SV40 T mRNA might give some insight into the reason for a low level of detectable β-globin spliced mRNA. The transcript might be degrading before splicing in vivo because of some of the modifications. In this case there would be no accumulation of precursor or of spliced product. A second possibility is that the precursor is splicing almost normally but then the spliced product is unstable and is immediately degraded. To distinguish between these two possibilities, one would have to establish the stability of the spliced product. To achieve this, one could transfect HeLa cells with a plasmid containing a β-globin cDNA. This cDNA would be transcribed to give an RNA equivalent to the spliced β-globin mRNA. If the RNA is stable then a decrease in the level of precursor without a corresponding increase in the level of spliced product, indicates that the precursor is being degraded without being spliced. However, if the control RNA is unstable one cannot show in vivo whether the pre-mRNA is also unstable and therefore cannot ascertain at what level the RNA is being degraded.

10.8 Future Development of the Splice Site Assay System.

The future work outlined in the previous section consists of experiments to help clarify the results obtained from the model system as it is now. However, the initial aim of my project was to establish a hierarchy of the intrinsic strengths of 3' splice site sequences. My model system as it is will not do this. Splicing to the synthetic site is only achieved by an AG+CG mutation at the authentic 3' splice site.
It is possible that the context within which my synthetic 3' splice sites are located is wrong for an efficient splice signal. The flanking sequences of the authentic site differ from those of the synthetic site. It is not known what, if any, elements within the intronic or exonic flanking sequences might influence splice site choice. Reed & Maniatis (1986) have shown that flanking exon sequences can influence splice site selection. In order to investigate the intrinsic strength of a 21 nucleotide synthetic splice site sequence, relative to the constant reference site, all other influencing factors must balance.

Reed and Maniatis (1986) have shown that it is possible to balance 3' splice site usage in a human β-globin gene construct in vitro. Their constructs consisted of tandem duplications of a truncated second exon of the human β-globin gene (Fig. 43). The constructs both have duplications which include 55 nucleotides of IVS-1 including the branch sequence, in addition the constructs have either 55 nucleotides or 115 nucleotides of exon 2 duplicated. These constructs were designated 3'D-55 and 3'D-115 respectively.

In vitro, 3'D-55 is spliced exclusively to the external site and 3'D-115 is spliced predominantly to the internal site but also detectably to the external site. Constructs with larger exon duplications were spliced exclusively to the internal site (Fig. 43). Presumably, a human β-globin construct could be made with between 55 and 115 nucleotides of exon 2 duplicated. Such a construct might be made such that both 3' splices balance to give approximately 50% of splicing to each site.

A construct could be made, based on the rabbit β-globin gene IVS-2 which I have used. Intron and exon sequence would be duplicated such that there were two 3' splice sites in tandem and these two sites balanced.
This system could then be used to analyse the effects on the splicing pattern \textit{in vitro}, and perhaps \textit{in vivo}, of discrete changes at either 3' splice site. Site directed mutagenesis could be used to create restriction sites to enable the substitution of either or both of the 3' splice sites with synthetic sequences. Site directed mutagenesis could also be used to alter sequences within the 3' splice site consensus, at the branch points etc. Restriction sites might be created to test the effects of the spatial separation of the two 3' sites either by enabling the insertion of extra sequence or the deletion of existing sequence.

It has been shown by us (Eperon \textit{et al.} 1986) and by others (Solnick 1985b) that secondary structure can influence splice site selection \textit{in vivo} and \textit{in vitro}. The biological significance of secondary structure in splice site selection is still unclear. However, a computer search has revealed regions of complementarity within both the human tropomyosin nm30 and the immunoglobulin heavy chain genes. This complementarity could form secondary structures which might influence splice site selection (A. Turnbull-Ross, personal communication). Work is currently being undertaken in this laboratory to investigate the importance of secondary structure on splice site selection \textit{in vivo} and \textit{in vitro}.

We have a system which is balanced for 5' splice site selection \textit{in vivo} (Eperon \textit{et al.} 1986) and \textit{in vitro} (unpublished results). The effects of introducing secondary structure into this system are being investigated \textit{in vivo} and \textit{in vitro}. The development of a balanced 3' splice site selection system as discussed above will enable complementary experiments to be undertaken to assess the influence of possible secondary structure on 3' splice site selection. The data generated from these model systems, in conjunction with data generated following mutagenesis of the naturally
occuring complementary regions within the human tropomyosin nm30 and immunoglobulin heavy chain genes, might give some valuable insights into the biological significance of secondary structure in splice site selection.

10.9 Nuclease Protection of Authentic and Cryptic 3' Splice Sites in vitro.

It has previously been shown (Ruskin & Green 1985c) that splice specific regions of a human β-globin IVS-1 transcript are protected from oligonucleotide directed cleavage by RNase H when undergoing splicing in a HeLa cell nuclear extract in vitro. Oligonucleotides directed against the 5' and 3' splice sites and the branch point failed to cause complete cleavage by RNase H. This could mean that either the RNA is within a complex and so the oligonucleotides cannot anneal, or that the oligonucleotides can anneal to the transcript but the duplex is inaccessible to RNase H.

Ruskin & Green (1985c) showed that in the absence of extract, all the oligonucleotides used directed cleavage. In a splicing reaction, they showed that the 5' splice site and the branch point are largely unprotected at zero time. However, cleavage of the 3' splice site is less efficient at zero time than cleavage of either the 5' splice site or the branch point. These results were a little unexpected; it has since been shown that U1 snRNP binds to the 5' splice site immediately and in the absence of ATP (Bindereif & Green 1987; also see Introduction and references therein). This may initially be a loose association and U1 snRNP might be displaced by the oligonucleotide directed against the 5' splice site.

-144-
Ruskin & Green (1985c) showed that protection from cleavage directed by the branch site oligonucleotide is first detected by 5 minutes and has reached a maximum by 10 minutes. This branch site protection is consistent with my data (see later and Results section). They also showed that protection of the 5' splice site is first detected by 10 minutes and has reached a maximum level by 20 minutes. This is consistent with my earlier comment that initial binding of U1 snRNP to the 5' splice site may be a loose association which can be disrupted by the 5' splice site oligonucleotide. The 5' splice site protection detected after 10 minutes may be due to the extended 5' splice protection which is seen in the fully assembled spliceosome (Bindereif & Green 1987; Chabot & Steitz 1987a). This extended protection requires ATP and may involve U2 and other snRNP in addition to U1 (Bindereif & Green 1987; Chabot & Steitz 1987a). This complex is presumably stabilised in some way and can no longer be displaced by the annealing of the oligonucleotide. Consistent with the idea that the detectable 5' site protection is indicative of the assembly of an active spliceosome, the splicing products began to appear only after 5' splice site and branch point protection had reached a maximum.

I have used the same RNase H protection technique as Ruskin & Green (1985c) to investigate the pattern of complex binding in a situation where there are two 3' sites splice in tandem. The external, synthetic site was not normally used but was activated as a cryptic 3' splice site following an AG+CG mutation at the internal, authentic or reference site.

My results show that at zero time on ice, both 3' splice sites are protected; after 30 minutes and beyond, it can be seen that there is a slight decrease in the level of protection at the unused 3' splice site. It is possible that the interaction between the used 3' splice site and a
splicing component is stabilised by the binding of further splicing components after the site is chosen.

On transcript 30, which was spliced to the authentic 3' splice site, there is protection of both branch sites by 5 minutes. By 30 minutes, protection of the authentic i.e. the used branch site had reached a maximum. However, protection of the unused branch site began to decrease by 30 minutes and was virtually undetectable by 90 minutes.

On transcript 30/8/9 protection of only the used branch point, was seen by 5 minutes and had reached a maximum by 30 minutes. Protection of the unused branch point was never seen. This transcript has an AG+CG mutation at the authentic 3' splice site and the synthetic cryptic 3' splice site located 68 nucleotides downstream is activated in vivo and in vitro. That the unused branch point was not protected was a little unexpected. Frendewey & Keller (1985) showed that the AG at the 3' splice site was not required for complex assembly although assembly was at a reduced level in the absence of the AG. Also, other groups have reported that in the absence of the 3' splice site AG dinucleotide, step 1 of splicing (ie. 5' splice site cleavage and lariat formation) proceeds but at a reduced level. In these experiments the lariat intermediates were not processed further (Reed & Maniatis 1985; Ruskin & Green 1985a).

In light of the work mentioned above, I expected to find protection of both used and unused branch points following the AG+CG mutation at the authentic 3' splice site of my construct. I also expected to be able to detect some dead-end lariat intermediates branched to the authentic branch point. My results led me to conclude that although Frendewey & Keller (1985) have shown that the AG is not required for complex formation when there is only one 3' splice site, perhaps when there is a choice between 3'
splice sites there is a requirement for the AG for any sort of complex formation beyond initial protection of the 3' splice site.

Very recent work (Aebi et al. 1986) has led me to an alternative hypothesis. Working with the large rabbit β-globin intron, Aebi et al. (1986) have shown that a mutation at the 3' splice site AG dinucleotide drastically reduces 5' splice site cleavage and lariat formation. This is contrary to previous findings (Reed & Maniatis 1985; Ruskin & Green 1985a; Frendewey & Keller 1985; also see Introduction). The main difference between the systems used by Reed & Maniatis (1985), Ruskin & Green (1985a) and Frendewey & Keller (1985) and those used by myself and by Aebi et al. (1986), is the length of the intron.

Reed & Maniatis (1985) and Ruskin & Green (1985a) both used a β-globin IVS-1 based system. The β-globin IVS-1 is 126 nucleotides long. Frendewey and Keller (1985) used model substrates derived from the adenovirus 2 major late transcription unit. This system consisted of two short exons separated by a 113 nucleotide intron. This intron is approximately the same length as the β-globin IVS-1. My model system and that of Aebi et al. (1986) is based on the rabbit β-globin IVS-2. This intron is 573 nucleotides long. This is over four and a half times longer than the β-globin IVS-1 and over five times longer than the deleted adenovirus intron used by Frendewey & Keller (1985).

My results confirm the findings of Aebi et al. (1986) that the 5' splice site cleavage is sensitive to the presence of the AG diucleotide at the 3' splice site and I propose that this sensitivity is dependent on intron length. The evidence suggests that 5' splice site cleavage and lariat formation is more inhibited by the lack of a 3' splice site AG if the separating intron is longer. The data supporting this hypothesis is
at present minimal. However, the hypothesis can easily be tested experimentally in vivo and in vitro by progressive deletion of non-essential intron sequence from a rabbit β-globin IVS-2 construct lacking a 3' splice site AG dinucleotide. If my hypothesis is correct, one would detect increased efficiency of step 1 of the splicing reaction as the intron length was decreased.

Assuming it is true for the rabbit β-globin IVS-2 that 5' splice site cleavage and lariat formation are very inefficient in the absence of an AG at the 3' splice site, then my observations that on transcript 30/8/9 the authentic branch site is unprotected and that splicing is to the cryptic 3' splice site, can be explained in kinetic terms. For this explanation, it must be assumed that the synthetic cryptic 3' splice site together with its branch point has an intrinsic strength or efficiency which is unaffected by the mutation at the authentic 3' splice site. This assumption is supported by the observation that initial protection of the synthetic 3' splice site and its branch point is the same whether that site is subsequently used or not.

I propose that on transcript 30, the authentic 3' splice site is always used because it is intrinsically more efficient than the synthetic 3' splice site and therefore becomes committed to splicing first. In a splicing time course, intermediates from a transcript using the authentic 3' splice site can be seen slightly earlier than intermediates from transcript 30/8/9, which splices to the cryptic 3' splice site.

It is probably the mutation from AG→CG at the authentic 3' splice site which drastically reduces the efficiency of 5' splice site cleavage and lariat formation to the authentic branch site. This drastic reduction of step 1 involving the authentic 3' splice site may mean that the
intrinsic strength or efficiency of this site is now less than that of the cryptic 3' splice site. Therefore, the cryptic 3' splice site is able to become committed and to be spliced to the 5' splice site much more rapidly than the inefficient step 1 reaction to the authentic branch point. The net result of this is that splicing is seen only to the cryptic 3' splice site, at a reduced efficiency. No dead end lariats are seen to the authentic branch site because they are kinetically unfavourable.

The lack of authentic branch point protection when the cryptic 3' splice site is used might be indicative of very inefficient splicing complex assembly at the authentic splice site. It has been shown that in the absence of an AG dinucleotide at the 3' splice site, binding of U2 to the branch point is at a reduced rate (Bindereif & Green 1987; Chabot & Steitz 1987a). Cryptic branch point protection is detected when the authentic splice site is used.

Protection of both the cryptic 3' splice site and its branch point is seen to begin to decrease at about the same time as intermediates generated by splicing to the authentic 3' splice site are detected. This suggests that when the partial splicing complex at the authentic 3' splice site interacts with the partial splicing complex at the 5' splice site to form the active spliceosome, the partially assembled complex at the cryptic 3' splice site becomes destabilised and begins to dissociate.

The model discussed above could also explain splicing to the cryptic 3' splice site on the transcript 30/8/9 which has an AG+CG mutation at the authentic 3' splice site. On transcript 30/8/9, as on transcript 30, protection of the unused 3' splice site is seen to decrease slightly after about 30 minutes. However, the unused branch site of transcript 30/8/9 is never seen to be protected from oligonucleotide directed cleavage by
RNase H. This apparent lack of protection could be indicative of very inefficient complex interaction at the branch point (see above). This inefficient complex interaction might be a result of the mutation at the AG dinucleotide at the authentic 3' splice site combined with or compounded by the length of the intron.

The various hypotheses discussed above can be tested experimentally. For instance, if the theory about inefficient branch point protection being a function of the mutant AG at the 3' splice site and of intron length is true, an AG→CG mutation at the cryptic 3' splice site will reduce the level of protection at that branch point. If, as I have hypothesised, deleting the non-essential sequences of a rabbit β-globin IVS-2 with a mutant 3' splice site leads to an increased efficiency of step 1 of the splicing reaction, then I would expect a corresponding increase in the level of protection detected at the branch point. A final test of the hypotheses discussed above would be to show that when transcript 30/8/9 is progressively deleted, a balance can be reached between an abortive step 1 reaction to the authentic 3' splice site branch point, and a two step splicing reaction to the cryptic 3' splice site.

10.10 Complex Assembly at the Authentic and Cryptic 3' Splice Sites in vitro

To investigate the composition of the protecting species bound to transcripts 30 and 30/8/9, the RNA was cleaved into three discrete fragments using oligonucleotide directed cleavage by RNase H, and immunoprecipitated by anti-Sm and anti-U1 antibodies at various stages during an in vitro splicing reaction. The three RNA fragments were; a 5'
splice site fragment, an authentic 3' splice site fragment, and a cryptic 3' splice site fragment. Both 3' splice site fragments also included their branch sites.

The 5' splice site fragment was precipitated at all times by both the anti-Sm and the anti-U1 antibodies. This was as expected. The authentic 3' splice site fragment of both transcripts 30 and 30/8/9 was precipitated at zero time by the anti-Sm antibody, and also sometimes by the anti-U1 antibody. However, although the cryptic 3' splice site of both transcripts was well protected from RNase H digestion at zero time, it was very poorly precipitated by either anti-U1 or anti-Sm antibodies. This suggests that initial protection may be by a non-snRNP component. The poor precipitability of the site might imply that subsequent complex assembly is very inefficient, presumably because the cryptic site is intrinsically weak. Unexpectedly, precipitation of the cryptic 3' splice site fragment was poor even at later time points when the branch site was known to be protected. A band was precipitated by 90 minutes which might be the cryptic splice site fragment. However, this band could also be the 30/8/9 spliced product, or the 30 splice product or lariat intermediate cleaved by RNase H and oligonucleotide imp. Spliced products are not normally precipitated by antibodies directed against the Sm class of U-type snRNPs so it is probable that the majority of the band is the cryptic 3' splice site fragment.

It is surprising that the cryptic 3' splice site fragment is less well precipitated than the authentic 3' splice site fragment although both are equally protected from oligonucleotide directed cleavage by RNase H. A possible explanation involving initial binding of a non-snRNP component to the 3' splice site is discussed above. However, both branch sites of
transcript 30 seem to be well protected up to about 30 minutes and this is thought to be by U2 snRNP which is Sm precipitable. (See Introduction and references therein; Bindereif & Green 1987). This observation could reflect a difference in sensitivity between the two techniques.

The data generated from the immunoprecipitation experiments is consistent with the splice site selection model which I presented in the previous section. It suggests that splicing complex assembly is more efficient at the authentic 3' splice site than at the cryptic 3' splice site. The kinetics of precipitation of the authentic 3' splice site mirror those of 3' splice site protection not those of branch site protection. Therefore, my proposal that the authentic branch site is very poorly protected when there is an AG→CG mutation at the authentic 3' splice site has neither been confirmed nor denied. It would be very useful to examine complex binding at the branch point using anti-U2 antibodies.

Finally, my nuclease protection data and my immunoprecipitation data, when interpreted together suggest that an AG→CG mutation at the authentic 3' splice site does not interfere with splicing component binding to that site. However, it does inhibit subsequent interaction of splicing components with the branch site. U1 snRNP interacts normally with the 5' splice site but this could be due to the presence of the cryptic 3' splice site.

10.11 Protection of Splice Sequences Within Isolated Spliceosomes.

As discussed earlier, I have proposed that a partially assembled splicing complex at the unused 3' splice site dissociates once the complex at the used site becomes an active splicing substrate. This is probably
because the partial complex at the 5' splice site is no longer available.

To investigate this, I decided to isolate fully assembled spliceosomes from a splicing reaction by sucrose gradient sedimentation. The technique used was very similar to that described by Frendewey & Keller (1985).

Preliminary experiments were performed in order to determine the best way of handling the gradients. Unloading the gradients by dripping the sucrose out through a small hole in the bottom of the centrifuge tube was found not to be ideal. The flow rate varied with the volume of sucrose solution remaining in the tube and the last 500μl would not come out at all. The best way of unloading the gradients was found to be by slowly pumping in 50% sucrose solution to the bottom of the gradient tube to displace the gradient, which could then be collected from the top. The apparatus which I designed to accomplish this is described elsewhere in this thesis.

The gradient fractions were initially analysed by Cerenkov counting and by polyacrylamide gel electrophoresis to determine the state of assembly of the spliceosome and the RNA species present. The presence of unspliced precursor RNA only within the spliceosome peaks isolated from the gradients led me to conclude that the spliceosomes were still not fully assembled after the 45 minute splicing reaction. Fully assembled active spliceosomes contain both precursor RNA and splicing intermediates (Frendewey & Keller 1985).

To isolate fully assembled active spliceosomes, the in vitro splicing reaction must be allowed to run for longer before being subjected to velocity gradient sedimentation. Frendewey & Keller (1985) used a short intron (113 nucleotides compared to my 573 nucleotide intron) and this could explain the more rapid spliceosome assembly which they observed.
RNase H digestion was directed against both 3' splice sites and both branch points of transcripts 30 and 30/8/9 which had been isolated as partially assembled spliceosome complexes. From my model, I predicted that within a spliceosome only the used 3' splice site and branch point would be protected. Unfortunately, my preliminary experiment showed that none of the splice sites were protected from RNase H cleavage of material isolated from a sucrose gradient. This was probably an artefactual result. Other groups have shown that the splice sites on RNA isolated from spliceosome gradients are protected from nuclease digestion (Bindereif & Green 1987; Chabot & Steitz 1987). Both these groups show protection from digestion by RNase T1. Perhaps the RNA/complex interactions are weakened during the spliceosome isolation procedure such that bound complexes can be displaced from the RNA by the oligonucleotides which direct cleavage by RNase H.

I believe that the spliceosome complexes became dissociated from the RNA because the gradients were allowed to warm to room temperature as the fractions were collected; also the scintillation counter used in the analysis of fractions, counts at room temperature.

In future, gradient fractions should be collected at 4°C ie. in the cold room. Two parallel gradients should be run for each sample, one for the preliminary analysis by Cerenkov counting at room temperature and the second to be kept cold for subsequent protection studies or immunoprecipitation studies. Alternatively, each fraction could be split into two aliquots, or only alternate fractions could be counted.
10.12 Concluding Remarks.

The work presented in this thesis documents the preliminary development and characterisation of a 3' splice site assay system. Using the results obtained from this system, I have presented a model for 3' splice site selection based on the relative efficiencies of complex assembly at two 3' splice sites. Further experimentation to test the validity of this model has been suggested.

We have shown previously (Eperon et al. 1986) that 5' splice site selection is dependent to a large extent on the 5' splice site sequence. We have also shown that the position of a 5' splice site can be balanced against its sequence to determine the overall strength of the site. It has not yet been shown whether 5' splice site sequence influences the efficiency of complex assembly. Chabot & Steitz (1987b) have shown that mutant 5' splice sites are not detectably bound by U1 snRNP in complete splicing extracts. Furthermore, a cryptic 5' splice site located 16 nucleotides upstream from the mutated 5' splice site has been shown to bind U1 during active splicing. The sequences contained in protected fragments which are immunoprecipitable with anti-U1 antibodies correlate with the position of the cryptic site and is an argument for U1 snRNP interacting with this region directly. Initial binding of U1 to the cryptic site was not detected but is probably direct and at a very low level since cryptic sites can be used after deletion of the authentic 5' splice site. Chabot & Steitz (1987b) propose that mutations that decrease the affinity of a 5' splice site for a U1 snRNP without interfering with other processes might allow splicing while permitting cryptic sites to compete for complex formation and thus splicing. In other cases, mutated
sites could remain strong binding sites and block splicing by preventing complex formation at potential cryptic sites.

The model proposed by Chabot & Steitz is consistent with our findings that 5' splice sites with sequence changes away from a match with U1 snRNA tend to be inherently weaker sites. The model suggests that 5' splice site selection is a very early event as the association of U1 snRNP with the 5' splice site is one of the first events in the splicing process. The 5' splice site sequence probably does not directly influence complex assembly other than that complex assembly is favoured at a site with a strong affinity for U1 snRNP. This may simply reflect the fact that at a given time U1 snRNP is more likely to be bound to a site with a high affinity for it, than to a site with a low affinity for U1 snRNP. The affinity of a 5' splice site for U1 snRNP is probably also influenced by factors such as the position of the site relative to the 3' splice site and to other 5' splice sites, the context of the site and the flanking sequences, and any secondary structure which may directly or indirectly involve the site.

3' splice site selection may be a slightly more complex process than 5' splice site selection. Less is known of the sequences which influence the strength of a 3' splice site. Investigation of the inherent strengths of 3' splice site sequences must await the development of a balanced 3' splice site assay system as discussed earlier.

I have suggested that 3' splice site selection may be a kinetic process reflecting the efficiency of complex assembly directed by that site. I have shown that initial binding of splicing components to the 3' splice site is not noticeably reduced by an AG+CG mutation at that site. However, subsequent branch point binding is drastically reduced and cannot
be detected by my RNase H protection technique. This suggests that sequences at a 3' splice site can influence the interaction of splicing components with sequences remote from that site. This influence might be brought about by subtle conformational changes in the 3' splice site binding component dependent on the exact sequence of the 3' splice site. Such conformational changes might reduce or enhance the affinity of the 3' splice site/factor complex for components which mediate the binding of U2 snRNP to the branch site.

The factors influencing splice site selection in mammalian pre-mRNA are still unclear. Choice of splice site seems to be determined by a complex balance between the splice site sequence, the flanking sequence and the position of the site relative to its competitors and its counterparts.

My work, and that of others, suggests that splice site selection is finally determined by factors which affect the efficiency of spliceosome assembly. Therefore, splice site selection is not an all or nothing event, rather a progressive kinetic process; a race towards spliceosome assembly.
Appendix 1. Shows the sizes of the molecular weight markers used for the polyacrylamide gel electrophoresis.

_Hpa II_ cleaves pBR322 into 26 fragments. The sizes in nucleotides are shown below.

<table>
<thead>
<tr>
<th>Letter</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>622</td>
<td>527</td>
<td>404</td>
<td>309</td>
<td>242</td>
<td>238</td>
<td>217</td>
<td>201</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>J</td>
<td>K</td>
<td>L</td>
<td>M</td>
<td>N</td>
<td>O</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>190</td>
<td>180</td>
<td>160</td>
<td>160</td>
<td>147</td>
<td>147</td>
<td>122</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>Q</td>
<td>R</td>
<td>S</td>
<td>T</td>
<td>U</td>
<td>V</td>
<td>W</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>76</td>
<td>67</td>
<td>34</td>
<td>34</td>
<td>26</td>
<td>26</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>Z</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

_Hae III_ cleaves pBR322 into 17 fragments. The sizes in nucleotides are shown below.

<table>
<thead>
<tr>
<th>Letter</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>587</td>
<td>540</td>
<td>504</td>
<td>458</td>
<td>434</td>
<td>267</td>
<td>234</td>
<td>213</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>J</td>
<td>K</td>
<td>L</td>
<td>M</td>
<td>N</td>
<td>O</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>192</td>
<td>184</td>
<td>124</td>
<td>123</td>
<td>104</td>
<td>89</td>
<td>80</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>Q</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Value  | 57 |    |    |    |    |    |    |    |
|        |    |    |    |    |    |    |    |    |
Appendix 2. Shows the size of the in vitro transcription products used in the experiments discussed in this thesis.

The relevant mICE β construct was linearised with Bgl II and transcribed in vitro as described in the methods to give the RNA product shown below.

The total transcript length is 745 ribonucleotides, including the synthetic oligonucleotide 3' splice site sequence inserted into the Eco RI within exon 3.
Appendix 3. Shows the sequences of the oligonucleotides used in the RNase protection experiments and the immunoprecipitation experiments discussed in the text.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>RpriEco</td>
<td>CCT GCA CCT GAG GAG TG</td>
</tr>
<tr>
<td>RpriAut</td>
<td>CAG CAC GTT GCC CAG GA</td>
</tr>
<tr>
<td>size</td>
<td>CCA GGA TGT AGT TGT TTC</td>
</tr>
<tr>
<td>new</td>
<td>CAA AAT GGT TAG ACA GCA</td>
</tr>
<tr>
<td>aut</td>
<td>GAA CAT GTT TAG CAG AGG</td>
</tr>
<tr>
<td>sp</td>
<td>CTG TAG GAA AAA GAA AA</td>
</tr>
<tr>
<td>sp2</td>
<td>CGG TAG GAA AAA GAA GAA</td>
</tr>
<tr>
<td>imp</td>
<td>GGA CCC GTT GCA CGA CCA</td>
</tr>
</tbody>
</table>

The oligonucleotides described above are drawn 5' to 3'


