INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book. These are also available as one exposure on a standard 35mm slide or as a 17" x 23" black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700  800/521-0600
Analysis of the structure and function of intron A13 of
Saccharomyces cerevisiae

Hoffman, Peter William, Ph.D.
The Ohio State University, 1989
ANALYSIS OF THE STRUCTURE AND FUNCTION OF INTRON AI3 OF
SACCHAROMYCES CEREVISIAE

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of the Ohio State University

By

Peter William Hoffman, B.S., M.S.

* * * * *

The Ohio State University

1989

Dissertation Committee:
P. S. Perlman
P. Fuerst
E. F. Vanin

Approved by
Advisor
Department of Molecular Genetics
VITA

March 26, 1952 ........................ Born - Stroudsburg, Pennsylvania

1974 ................................. B.S., St. Lawrence University, Canton, New York

1978 - 1981 .......................... Research Associate, Laboratory of C. Richard Wrathall, American University, Washington, D.C.

1981 - 1987 .......................... Research Associate, Laboratory of Philip S. Perlman, The Ohio State University, Columbus, Ohio

1986 ................................. M.S., The Ohio State University, Columbus, Ohio
# TABLE OF CONTENTS

VITA ........................................... ii  
LIST OF TABLES ................................ vii  
LIST OF FIGURES ................................. viii  

## CHAPTER  PAGE

### I. INTRODUCTION ......................... 1  
A. Introductory Remarks ...................... 1  
B. A Word on Nomenclature ................... 2  
C. Yeast Mitochondria as an Experimental  
   System ..................................... 3  
   1. The Yeast Mitochondrial Genome ....... 3  
   2. Advantages of the System ............. 5  
D. Splicing Pathways .......................... 10  
   1. Nuclear Pre-mRNA Splicing Pathway ... 10  
   2. Splicing of Pre-tRNA ................... 12  
   3. Splicing of Group II Introns ........... 13  
E. Group I Introns ............................ 15  
   1. Introduction ............................ 15  
   2. Autocatalysis of Group I Intron Splicing  
   and the Tetrahymena rRNA Intron ........ 16  
      a. Autocatalytic Reactions of the  
      Tetrahymena rRNA Intron ............. 17  
      b. Reactions of Other Group I Introns .. 21  
F. The Secondary Structure of Group I Introns  
   1. The Internal Guide and Its Interactions. 24  
   2. E, E' and P3 Pairing ................... 31  
   3. P, Q, and the P4 Pairing ............... 32  
   4. R, S, and the P7 Pairing ............... 34  
G. Proteins Encoded by Group I Introns ..... 38  
   1. Group I Maturases ........................ 39  
   2. The S-5 Protein ........................ 42  
   3. Endonucleases .......................... 42  
H. Proteins Involved in Splicing Group I  
   Introns .................................... 46  
I. aI3 ........................................ 53  
   1. The Sequence and Structure of aI3 .... 54
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Mutants of aI3</td>
<td>56</td>
</tr>
<tr>
<td>3. In Vitro Reaction of aI3</td>
<td>58</td>
</tr>
<tr>
<td>4. Summary</td>
<td>61</td>
</tr>
<tr>
<td>II. EXPERIMENTAL TECHNIQUES</td>
<td>63</td>
</tr>
<tr>
<td>A. Strains</td>
<td>63</td>
</tr>
<tr>
<td>B. Growth Media</td>
<td>63</td>
</tr>
<tr>
<td>1. Yeast Growth Media</td>
<td>63</td>
</tr>
<tr>
<td>2. Bacterial Growth Media</td>
<td>64</td>
</tr>
<tr>
<td>C. Nucleic Acid Preparation</td>
<td>65</td>
</tr>
<tr>
<td>1. Mitochondrial DNA Isolation</td>
<td>65</td>
</tr>
<tr>
<td>2. Mitochondrial RNA Isolation</td>
<td>65</td>
</tr>
<tr>
<td>3. Rapid Plasmid Preparation</td>
<td>65</td>
</tr>
<tr>
<td>4. Plasmid Preparation</td>
<td>66</td>
</tr>
<tr>
<td>5. M13 Phage Preparation</td>
<td>66</td>
</tr>
<tr>
<td>D. DNA Manipulations</td>
<td>66</td>
</tr>
<tr>
<td>1. Restrictions</td>
<td>66</td>
</tr>
<tr>
<td>2. Ligations</td>
<td>67</td>
</tr>
<tr>
<td>3. Transformations</td>
<td>67</td>
</tr>
<tr>
<td>E. Electrophoresis Techniques</td>
<td>67</td>
</tr>
<tr>
<td>1. Agarose Gels</td>
<td>67</td>
</tr>
<tr>
<td>2. Acrylamide Urea Gels</td>
<td>67</td>
</tr>
<tr>
<td>3. Sequencing Gels</td>
<td>67</td>
</tr>
<tr>
<td>F. Northern Blots</td>
<td>68</td>
</tr>
<tr>
<td>G. Yeast Mitochondrial Genetic Analysis</td>
<td>68</td>
</tr>
<tr>
<td>H. Sequencing Reactions</td>
<td>69</td>
</tr>
<tr>
<td>1. Single Lane Screening Reactions</td>
<td>69</td>
</tr>
<tr>
<td>2. Standard Sequencing Reactions</td>
<td>70</td>
</tr>
<tr>
<td>I. Transcription and Purification of RNA</td>
<td>70</td>
</tr>
<tr>
<td>III. RESULTS</td>
<td>71</td>
</tr>
<tr>
<td>A. Genetic Mapping and Plasmid Construction</td>
<td>72</td>
</tr>
<tr>
<td>1. Cloning of the Wild-type aI3</td>
<td>72</td>
</tr>
<tr>
<td>2. Genetic Mapping and Plasmid Construction with Mutants M44, C1072, and C1085</td>
<td>75</td>
</tr>
<tr>
<td>3. Summary</td>
<td>82</td>
</tr>
<tr>
<td>B. Sequencing Data</td>
<td>83</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>83</td>
</tr>
<tr>
<td>2. Sequence of the aI3 Region of Strain IΔ1-6/161</td>
<td>86</td>
</tr>
<tr>
<td>3. Sequence of Mutants M44, C1072, and C1085</td>
<td>87</td>
</tr>
<tr>
<td>4. Summary</td>
<td>89</td>
</tr>
<tr>
<td>C. In Vitro Processing Analysis</td>
<td>90</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>90</td>
</tr>
<tr>
<td>a. Initial Time Course</td>
<td>91</td>
</tr>
<tr>
<td>b. BamHI Linearization</td>
<td>92</td>
</tr>
<tr>
<td>c. Summary</td>
<td>93</td>
</tr>
</tbody>
</table>
4. Experiments for Further Analysis of Mutant C1072 .................................................. 158

F. Future Directions for Intron aI3 and Group I Intron Research ................................. 160

1. Future Directions for aI3 Research ................................................................. 160

2. Future Directions for Group I Intron Research ..................................................... 162

TABLES ......................................................................................................................... 163

FIGURES ....................................................................................................................... 169

BIBLIOGRAPHY ........................................................................................................ 272
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Yeast and bacterial stock strains</td>
<td>163</td>
</tr>
<tr>
<td>2. Mit- and petite strains</td>
<td>164</td>
</tr>
<tr>
<td>3. Petite restoration of M44, C1072, C1085</td>
<td>165</td>
</tr>
<tr>
<td>4. Oxi3 synthetic primers</td>
<td>166</td>
</tr>
<tr>
<td>5. Strain polymorphisms in the aI3 region: D273 vs 161</td>
<td>167</td>
</tr>
<tr>
<td>6. Basepair changes in M44, C1072, C1085</td>
<td>168</td>
</tr>
<tr>
<td>FIGURES</td>
<td>PAGE</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>1. The nuclear mRNA splicing process</td>
<td>170</td>
</tr>
<tr>
<td>2. The tRNA splicing pathway</td>
<td>172</td>
</tr>
<tr>
<td>3. Typical group II intron structure</td>
<td>174</td>
</tr>
<tr>
<td>4. Model for splicing of group II introns</td>
<td>176</td>
</tr>
<tr>
<td>5. Autocatalysis of the Tetrahymena rRNA intron</td>
<td>178</td>
</tr>
<tr>
<td>6. The general transesterification reaction</td>
<td>180</td>
</tr>
<tr>
<td>7. Binding of guanosine to a group I intron</td>
<td>182</td>
</tr>
<tr>
<td>8. The general structure of group I introns</td>
<td>184</td>
</tr>
<tr>
<td>9. The structure of Tetrahymena L-rRNA intron</td>
<td>186</td>
</tr>
<tr>
<td>10. The pertinent sequence of the coxl gene from strain D273</td>
<td>188</td>
</tr>
<tr>
<td>11. The sequence of the cis-acting elements of intron aI3</td>
<td>192</td>
</tr>
<tr>
<td>12. Reaction pathway and products of aI3 as deduced by Tabak et al. (1987)</td>
<td>194</td>
</tr>
<tr>
<td>13. The physical map of petite A17-17</td>
<td>196</td>
</tr>
<tr>
<td>14. The extent of deletion C245</td>
<td>198</td>
</tr>
<tr>
<td>15. Maps of cloning vectors used in this study</td>
<td>200</td>
</tr>
<tr>
<td>16. The physical map of plasmid BS17-17</td>
<td>202</td>
</tr>
<tr>
<td>17. Physical map of petites with breakpoints in the middle of cox3</td>
<td>204</td>
</tr>
<tr>
<td>18. Mitochondrial translation products from M44, C1072, and C1085</td>
<td>206</td>
</tr>
<tr>
<td>19. The coxl BglII/HindIII fragment</td>
<td>208</td>
</tr>
<tr>
<td>20. Plasmid constructs used in analysis of mutants M44, C1072 and C1085</td>
<td>210</td>
</tr>
<tr>
<td>21. M44 sequencing strategy</td>
<td>212</td>
</tr>
<tr>
<td>22. C1072 and C1085 sequencing strategy</td>
<td>214</td>
</tr>
<tr>
<td>23. Primary structure of the GC clusters from the a2 family</td>
<td>216</td>
</tr>
<tr>
<td>24. IGS binding in strain D273-10B/A1 and strain ID4-6/161</td>
<td>218</td>
</tr>
<tr>
<td>25. Initial time course experiment</td>
<td>220</td>
</tr>
<tr>
<td>26. Transcripts used from plasmid BS17-17</td>
<td>222</td>
</tr>
</tbody>
</table>
27. Reaction of BamHI linearized transcript 224
28. GTP concentration response ................. 226
29. (NH₄)₂SO₄ concentration response .......... 228
30. MgCl₂ concentration response ............... 230
31. Temperature response ........................ 232
32. AI3 self-splicing products .................... 234
33. BS17-17 course gel and band identification .............................................. 236
34. Early time course gel and band identification .............................................. 238
35. PvuII vs HindIII linearizations of BS17-17 ............................................. 240
36. Probes used in this study ......................... 242
37. Northern blots of in vitro reaction products .............................................. 244
38. Overexposure demonstrating full-length intron circle .................................. 247
39. Mutant C1085 in vitro processing time course ........................................... 249
40. Mutant C1072 in vitro processing time course ........................................... 251
41. Mutant M44 in vitro processing time course .............................................. 253
42. Overexposure of M44 in vitro processing time course ................................... 255
43. The secondary structure of intron aI3 from strain 161 .................................. 257
44. Comparison of I5 region of aI3 from strains D273 and 161 .............................. 259
45. Autocatalytic pathways of intron aI3 .... 261
46. Involvement of the IGS in autocatalytic reactions of aI3 ............................... 263
47. Basepair alterations in mutant C1085 .... 265
48. Potential disposition of group I cis-acting sites in mutant C1085 .............. 267
49. IGS binding in strain ID4-6/161 and mutant M44 ........................................ 269
50. Autocatalytic reactions of aI3 in intron aI3 .............................................. 271
I. INTRODUCTION

A. Introductory Remarks

In the last decade two fundamental discoveries have revolutionized our understanding of biological processes: first, the discovery that some coding sequences in eukaryotes are interrupted by intervening sequences or introns (Padgett et al., 1986) which must be removed prior to translation; and, second, the discovery that some RNA molecules have catalytic activity (Cech and Bass, 1986), something long thought to be reserved exclusively for proteins. It has been my privilege to have been involved in research which dealt with both of these topics and to have been involved in that research at a time when new techniques, such as M13 sequencing and in vitro RNA transcription, made possible rapid progress in the study of RNA structure and function.
This thesis will describe experiments on the third intron of the coxI gene of the mitochondrial genome of the yeast *Saccharomyces cerevisiae*. This intron belongs to the family of introns known as group I. As will be discussed below, this intron is autocatalytic; that is, under suitable conditions it can catalyze its own splicing, without the addition of protein. It has an unusual structure for introns of this type, making it particularly worthy of study. The wildtype intron has been cloned and the sequence determined. Self-splicing conditions have been optimized and the reaction analyzed in light of recently published reports. Mutants mapping to this intron, which had previously been well characterized *in vivo*, have been cloned, sequenced and their *in vitro* effect studied.

B. A Word on Nomenclature

Throughout this thesis I will make use of nomenclature and abbreviations common to molecular biologists working in fungal mitochondrial RNA processing, but less so to others. The genes encoding the cytochrome b apoprotein and cytochrome oxidase subunit I are referred to as cob and coxI, respectively. These genes are multiply split in yeast mitochondria. The conventional nomenclature for the introns and exons
of these genes is that introns or exons of coxl are designated aI or aE while those of cob are designated bI or bE. This designation is then followed by a number representing the order of appearance. For example, the subject of this thesis, the third intron in coxl is called aI3 and exists between the exons aE3 and aE4. Some introns which, for historical reasons, have been given the same order number are differentiated by a greek letter. Examples are aI5alpha and aI5beta. Sequences that have the potential to encode proteins (i.e., contain no stop codons in a reading frame for a reasonable distance) are referred to as open reading frames or ORFs. Those sequences which contain stop codons in all frames and could not encode a protein are referred to as closed reading frames or CRFs.

C. Yeast Mitochondria as an Experimental System

1. The Yeast Mitochondrial Genome

The mitochondrial genome of Saccharomyces cerevisiae codes for components of the respiratory chain and the mitochondrial translation apparatus. The respiratory chain proteins are:

a) The genes for cytochrome oxidase subunits 1, 2, and 3. These proteins are referred to
as Cox1, Cox2, and Cox3.
b) Three subunits of the mitochondrial ATPase
complex, subunits 6, 8, and 9, are encoded by
the oli2, aapl, and oll1 genes respectively.
c) The cytochrome b apoprotein, coded for by
the cob gene.
Mitochondrially encoded components of the translation
apparatus are the 21S rRNA, the 15S rRNA, 25 tRNAs
constituting a complete set, and a ribosome associated
protein, var1. The genome also contains the region
coding for the RNA subunit of the mitochondrial RNaseP
which has been termed the tRNA synthesis locus. Five
free-standing unassigned reading frames (URFs) are
known to exist but have, as yet, no protein product or
function associated with them. Additionally, many
introns in the genome contain open reading frames (ORFs)
some of which are known to encode protein products
related to intron processing and dispersal. These are
discussed in more detail in following sections.

The primary structure of the Saccharomyces
cerevisiae mitochondrial genome is unusual. It is
circular in form and 78 to 85 kilobasepairs in length.
The size heterogeneity stems mainly from the presence or
absence of various introns in the cob, coxl, and 21s-
rRNA genes. In addition, small optional inserts may exist in the 15S, 21S, and var1 genes. The genome is very AT rich, approximately 82% overall. Coding sequences are, on average, 22% to 25% GC, with these pairs distributed fairly evenly throughout the sequence. The exception is the var1 gene which is 85% AT. The coding regions, however, make up, at most, 35% of the genome. The non-coding regions average about 15% GC, with most of the GC basepairs located in short (50-80 bp), GC rich sequences known as GC clusters (de Zamaroczy and Bernardi, 1985). The group headed by Bernardi (de Zamaroczy and Bernardi, 1986) has studied these clusters and found that they fall into families based on their primary sequence. The existence of a particular GC cluster is strain dependent and the size heterogeneity of the 15S and, in part, the var1 and 21S genes can be attributed to the presence or absence of GC clusters.

2. Advantages of the System

The yeast mitochondrial genome offers many advantages to the researcher wishing to study intron structure and function. The genome is small when compared to nuclear or bacterial genomes; this fact has
made it possible to completely restriction map many strains and to sequence approximately 80% of the genome. This simplifies cloning since one does not need to screen large libraries. Three genes are interrupted by introns and two of the four known types of introns are represented, group I and group II, making the genome a rich source of raw material for studies of RNA processing.

The most valuable aspect of yeast mitochondria as an experimental system is, however, the ease with which mutants can be isolated and characterized. Yeast are facultative anaerobes; therefore, while mutations which incapacitate mitochondrial function do not support growth on nonfermentable substrates and slow growth on fermentable substrates somewhat, they are not lethal and can be maintained easily. The importance of in vivo mutants to the study of RNA processing should not be overlooked. While in vitro analysis is useful in dissecting the fine points of known interactions or verifying suspected interactions, in vivo generated mutants often uncover unsuspected interactions, sometimes directing attention to a site or a protein, previously unknown, where in-depth analysis should be undertaken. A system where in vivo splicing mutations
Several classes of mutants affecting mitochondrial function exist. They are: petite (or rho\textsuperscript{-}), mit, syn, and pet. Petite mutants deleted for a large portion of the mitochondrial genome. The portion which remains is iterated many times until the size of each individual petite DNA circle is approximately the size of the wildtype mtDNA. The repeating units of a petite may be arranged head to head, head to tail, or in a more complex fashion. Petite mutants are characterized by the complete loss of all mitochondrial function. Mit mutants are point or deletion mutations in the mitochondrial genome. These mutants, in general, affect only the production of one mitochondrial product. The phenotype of a mit mutant may be pleiotropic because of interactions within the mitochondria. Syn mutants are mitochondrial point or deletion mutants which impair mitochondrial protein synthesis; as such, they are really a subset of mit mutants. They have the same pleiotropic deficiencies as petite mutants, but can be differentiated from them by their genetic properties. The phenotype of syn mutants can be corrected in crosses.
with appropriate petite tester strains (see below), while that of petites cannot.

Pet mutants are yeast nuclear mutations which disrupt mitochondrial function. Nuclear mutants may be distinguished from mitochondrial mutants by their classic Mendelian mode of inheritance as opposed to the cytoplasmic inheritance of mitochondrial traits (Dujon, 1981).

Petite mutations can be used to accurately map point or deletion mutations in the mitochondrial genome. When yeast mate, the mitochondria fuse and recombination takes place between the mitochondrial genomes. In matings of the type mit (or syn) x petite, the mit genome can be restored to respiration sufficiency if the petite retains the region of the genome where the mit lesion resides. The technique is analogous to classical phage deletion mapping except that the location of the mutation is defined by the regions retained by the tester strain rather than that deleted from it.

Mutations can be selected for study based upon their petite restoration pattern. A group of mit mutants can be mated against petites retaining the region of interest and those which are restored to respiration sufficiency subjected to further study. The
availability of this simple and efficient mapping
technique has allowed the isolation of many splicing-
defective mutants which were the first clue to
understanding the intramolecular interactions which are
the basis of RNA catalyzed reactions.

Once a mutation has been selected for study, the
researcher has a full range of experimental techniques
available to analyze it completely. Techniques exist
for the isolation and study of mitochondrial DNA, RNA,
and proteins. Using strains deficient in karyogamy
(kar⁻), one can switch the nuclear background of any
mitochondrial genome. A technique also exists to
observe complementation, thus allowing the
identification of cis and trans mutants. Procedures are
presently being perfected for the transformation of
yeast mitochondria using a system that injects DNA into
the cell on pellets shot from a gun-like apparatus
(Johnston et al., 1988). Furthermore, the yeast nuclear
system is probably the best characterized and most
easily manipulated eukaryotic genetic system, allowing
elegant genetic analysis and cloning of genes affecting
mitochondrial function.
D. Splicing Pathways

There are presently four known intramolecular splicing pathways: nuclear pre-mRNA, tRNA, group I, and group II. This section will provide a brief outline of the nuclear, tRNA and group II pathways. A more extensive discussion of group I intron splicing is presented below.

1. Nuclear Pre-mRNA Splicing Pathway

The distinguishing feature of nuclear pre-mRNA introns is the sequence at the splice sites. The nuclear mRNA precursor has the following consensus sequence at the intron exon boundaries:

\[
5'\text{-AG:GU(A/G)AGU-intron-(U/C)\_N\_AG:G-3'}
\]

The cleavage ligation sites are denoted by colons, and \(N\) equals any base. The underlined bases are essentially invariant, while the other bases vary somewhat (Padgett et al., 1986). The pathway of nuclear intron splicing is shown in Figure 1. The first step is the cleavage at the 5' splice site coupled with the linking of the first base in the intron to an adenosine residue within the intron in a 2' to 5' phosphodiester bond. The
adenosine is found as the last A within a consensus sequence UACUAAC located 20 to 50 basepairs upstream of the 3' splice site. This sequence is highly conserved in yeast introns, but much more variable in mammalian introns. Following branch formation, a cleavage occurs at the 3' splice followed by ligation of the 5' and 3' exons. Two products are formed by this reaction: the ligated exons and the excised intron in the form of a lariat. It should be noted that the phosphate moieties at the 5' and 3' splice sites are conserved in the products; therefore, the reactions could occur by concerted transesterifications (Cech, 1986).

Splicing of nuclear pre-mRNA requires the formation of a multicomponent RNA-protein complex known as the spliceosome. This complex has been shown in mammalian cells to include the pre-mRNA and five small nuclear RNAs (snRNAs), U1, U2, U4, U5, and U6. The snRNA U3, which has been shown to be required for splicing, has not been found in the spliceosome. Analogous snRNAs have recently been identified in yeast (Cheng and Abelson, 1987). Spliceosome assembly requires hydrolyzable ATP (Padgett et al., 1986). Disassociation of the complex has been shown to also require ATP in yeast (Cheng and Abelson, 1987). The protein
components of the spliceosome are not well understood; however, in yeast, the products of the RNA genes are required for activity of in vitro splicing extracts (Lustig et al., 1987).

2. Splicing of Pre-tRNA

The processing of pre-tRNAs has been most extensively studied in yeast. Approximately 40 of the 400 yeast tRNA genes contain introns, which range in size from 14 to 60bp (Peebles et al., 1983). The intron is invariably located one base 3' of the anticodon (Cech, 1983). There are no known consensus sequences or common structures for these introns; rather, the secondary structure of the pre-tRNA itself is thought to be responsible for the specificity of the reaction (Padgett, 1986). The reaction mechanism is shown in Figure 2. The first step is a cleavage by an endonuclease, resulting in the release of the intron and the production of a 2'-3' cyclic phosphate at the end of the 5' 1/2 molecule and a 5' hydroxyl terminal on the 3' 1/2 molecule (Peebles et al., 1983). The ligation step that follows is rather complex; it involves the addition of a phosphate to the 3' 1/2 molecule, followed by the adenylation of the same molecule, the resolution of the 2'-3' cyclic phosphate to a 2' phosphate and, finally,
the ligation of the two halves. The entire step is thought to require only a single enzyme (Peebles et al., 1983). Finally a 2'-phosphatase generates the 2'-hydroxyl group found in the mature tRNA (Greer et al., 1983). Both the endonuclease and ligase have been identified and isolated in the laboratory of John Ableson (Peebles et al., 1983) (Greer et al., 1983).

3. Splicing of Group II Introns

Group II introns are a family of organelle introns related by a common secondary structure (Michel and Dujon, 1983). This structure includes six conserved basepaired stems. Typical group II structures, of aI1 and aI2, are shown in Figure 3. Group II introns have been found in the mitochondria of yeast (both Saccharomyces and Schizosaccharomyces species), Podospora anserina, and higher plants, as well as the chloroplasts of Euglena, Chlamydomonas, and higher plants (Peebles et al. 1986). In Saccharomyces cerevisiae, the excised introns accumulate as lariat molecules (Arney et al., 1980).

Some group II introns have been shown to self-splice in vitro in a buffer containing 100mM MgCl$_2$ at 45 degrees (Peebles et al., 1986) (van der Veen et al.,
The proposed reaction mechanism is shown in Figure 4. Similar to nuclear pre-mRNA introns, group II introns form a lariat intermediate, with the 5' end of the intron joined to an internal adenosine residue by a 2'-5' phosphodiester bond (Peebles et al., 1986). The branch point forms at an adenosine, which is located in a bulge in a hairpin structure near the 3' end of the intron. The lariat is formed at the first step of the reaction, a nucleophilic attack on the phosphodiester bond at the 5' splice site by the 2' OH of the adenosine residue. The second step is an attack by the 3' OH of the released 5' exon on the 3' splice junction which results in the ligation of the exons and the release of the intron. Alteration of the salt conditions of the reaction yields additional products. In particular, when the reaction is carried out in the presence of 500mM KCl, linear intron and free exons are produced (Jarrell et al., 1988a). Interaction of the 5' exon and intron sequences is important for the accuracy of the reaction. Mutant constructs that retain as many as 18 nucleotides in the 5' exon produce aberrant products in vitro (van der Veen et al., 1987).

Recently, group II introns have been shown to have activity in trans, both in vitro and in vivo. In vitro,
a trans-splicing reaction has been demonstrated for the yeast mitochondrial intron a1Sgamma (Jarrell et al., 1988a). When free 5' exon and a molecule containing intron plus 3' exon are incubated together in vitro, ligated exons are produced (Jacquier and Rosbash, 1986); furthermore, the free intron is able to split the ligated exons in trans (Jarrell et al., 1988a).

Chloroplast trans-splicing pathways are thought to be mediated by group II introns. The chloroplast psaA gene of Chlamydomonas consists of three exons which are transcribed discontinuously and ligated in a trans-splicing reaction. The exons are flanked by sequences which are partial group II introns. In vivo, these partial sequences can pair to form a full group II intron structure which is thought to direct the ligation of the exons (Choquet et al., 1988).

E. Group I Introns

1. Introduction

Group I introns are a widespread class of intervening sequences defined by a set of conserved sequence elements, which allow the folding of all group I introns into a common secondary structure. At least 40 group I introns are known to exist (Collins, 1988). While group I introns are most commonly found in the
mitochondrial genomes of fungi, such as *Saccharomyces*, *Schizosaccharomyces*, *Neurospora* or *Aspergillus*, they have also been found in the nuclear rRNA genes of *Tetrahymena* and *Physarum*, and the chloroplast genomes of higher plants, and *Chlamydomonas*. The size of group I introns varies widely from 258bp to over 2kb (Collins, 1988). This diversity is for the most part due to the presence in some group I introns of long open reading frames (ORFs). Several of these ORFs encode proteins with known functions (see below).

2. Autocatalysis of Group I Intron Splicing and the Tetrahymena rRNA Intron

Group I introns were the first known class of autocatalytic RNAs (Cech, 1987). The original *in vitro* self-splicing experiments were carried out by Tom Cech and co-workers at the University of Colorado on the macronuclear rRNA gene intron of *Tetrahymena thermophila* beginning in 1981 (Grabowski et al., 1981). It was almost two years before it was realized that there is a structural relationship between the *Tetrahymena* rRNA intron and group I mitochondrial introns (Cech et al., 1983). The *Tetrahymena* rRNA intron has been the most intensively studied example of catalytic RNA. The reactions of this intron will be outlined briefly as
prototypical for all autocatalytic group I introns. There are three major reaction pathways of the Tetrahymena rRNA intron: self-splicing, trans-splicing, and 3' hydrolysis. In addition, a portion of the linear intron has been shown to have true enzymatic activity.

a. Autocatalytic Reactions of the Tetrahymena rRNA Intron

Self-splicing of group I introns occurs in a buffer containing only a monovalent cation (usually 50 to 200mM NH$_4^+$ or Na$^+$), a divalent cation (5 to 60mM Mg$^{++}$ or Mn$^{++}$), and a guanosine compound (at approximately 0.1 mM). Reactions are generally carried out between 30 and 42 degrees. The precise conditions for the Tetrahymena rRNA intron are 200mM NH$_4$Cl, 10mM MgCl$_2$, and 0.1 mM GTP at 37 degrees (Garriga et al., 1986). The reaction mechanism is shown in Figure 5. The reaction proceeds via a series of transesterifications of the general form shown in Figure 6. Since there is no net change in the number of ester linkages, the reaction can take place without an external energy source such as the hydrolysis of ATP or GTP.

The first transesterification reaction is the attack of the guanosine 3' hydroxyl group on the
phosphate at the 5' intron exon boundary. This results in the covalent addition of the guanosine residue to the intron in a 3' to 5' bond and the release of the 5' exon from the intron. The interaction of the guanosine with the pre-rRNA has been analyzed using reaction kinetics and guanosine analogues. These studies led Bass and Cech (1984) to propose the model for guanosine binding shown in Figure 7. Modifications of the guanosine that interfere with the hydrogen bonds in this figure were shown to inhibit or slow splicing. In addition, nucleotides modified at the 2' or 3' hydroxyl groups were also inactive. Since the 3' hydroxyl participates directly in the splicing reaction, it is not surprising that modifications to this group eliminate splicing. The role of the 2' hydroxyl is less clear. The authors suggest that the presence of this group could enhance the reactivity of the 3' hydroxyl through electronic effects. The kinetics of the reaction were unchanged when GDP or GMP was substituted for GTP; however, ATP, CTP, and UTP do not permit splicing. The second transesterification is the attack of the newly formed 3' hydroxyl of the 5' exon on the bond at the 3' intron/exon boundary. This ligates the exons precisely and liberates the intron.
The free intron can undergo an intramolecular transesterification that results in the circularization of the major portion of the intron and the release of a 15 nucleotide fragment from the 5' end, including the G added in the first step. Cyclization requires the G that is the terminal nucleotide of the linear intron. When that base is selectively removed, the intron can no longer react with itself (Tanner and Cech, 1987). However, free guanosine can make a nucleophilic attack at the normal site of cyclization releasing the 15mer in a reaction analogous to cyclization. Tanner and Cech (1987) have used this fact to study the G binding required for cyclization. Those authors found that the binding is similar to that used for splicing, since the same sites on the guanosine are found to be important for this reaction as are important for splicing. The circle can reopen via a site specific hydrolysis at the bond formed during circle formation and reclose with the loss of an additional 4 basepairs from the 5' end. The rate of this reaction increases with increasing pH within the range 7.5 to 9.5 (Zaug et al., 1985). The end product molecule produced is known as the L-19 IVS.

The L-19 IVS molecule has been shown to have true enzymatic activity (Zaug and Cech, 1986). When
incubated with the oligonucleotide pC5, the L-19 IVS acts as a nucleotidyltransferase, catalyzing the conversion of the pC5 to both longer and shorter molecules. The longer molecules are chains of C linked in 3',5'-phosphodiester bonds, with sizes to pC30. Shorter molecules are pC4 and pC3. This activity is sequence specific, since no activity is shown on oligo(A) or oligodeoxyribonucleotides. Oligo(U) yields a weak reaction. The L-19 IVS can also act as an RNA restriction endonuclease (Zaug et al., 1986). When the L-19 IVS is incubated in a buffer containing 10mM Mg++ and 1mM GTP and large RNA molecules, such as a fragment of mouse beta-globin pre-mRNA, those molecules are cleaved specifically following the sequence CUCU and a GTP added to the 5' end of the downstream cleavage product.

A trans-splicing activity of the Tetrahymena rRNA intron was demonstrated by Inoue et al. (Inoue et al., 1985). That study found that, when pre-rRNA was incubated under self-splicing conditions with the dinucleotide CpUh in place of GTP, two major products were produced; one consisted of the 5' exon and the full intron and the other was the 3' exon with CpU covalently attached to the 5' end in a 5'-3' phosphodiester bond.
The reaction was sequence dependent; no other dinucleotides were able to ligate to the 3' exon. The sequence specificity was attributed by the authors to the similarity to the sequence of the 5' exon at the splice site (CUCCUCU). This sequence is thought to basepair with the internal guide sequence so as to align the 5' exon for ligation (see below). The 5' exon + intron molecule can undergo two separate cyclizations, one at the normal site 15bp 3' of the 5' intron/exon boundary and another at the 5' splice site. The second cyclization site yields a circle that contains the entire intron (Inoue et al., 1986).

When the *Tetrahymena* pre-rRNA is incubated at 42 degrees in the standard solution without GDP, the molecule undergoes site specific hydrolysis at the 3' intron/exon boundary, thus releasing the 3' exon and the 5' exon + intron molecule (Inoue et al., 1986). As described above, the 5' exon + intron can undergo two cyclization reactions. The 3' exon hydrolysis increases with increasing pH in the range 7.5 to 9.0.

b. Reactions of Other Group I Introns

All self-splicing group I introns appear to follow the same transesterification mechanism used by the
Tetrahymena rRNA intron with GTP being used as the co-factor. Of the carefully studied self-splicing group I introns, most, but not all, appear to form circles, the notable exception being cob intron 1 of Neurospora mitochondria (Garriga and Lambowitz, 1984). One group I intron, the omega intron of yeast mitochondria, has been reported to form lariats as well as circles (Arnberg et al., 1986). The lariats contain exon sequences as well as intron sequences and this reaction, therefore, may represent a special case where particular exon sequences cause unique products. A complex series of cyclizations has been described for the third intron of the yeast coxI gene (aI3) (Tabak et al., 1987). These reactions and other reported activities of this intron are discussed in detail below. The trans-splicing reaction has been described for cob intron I of Neurospora, but with a different sequence specificity (Garriga et al., 1986). The trans reactions of the L-19 IVS molecule have not been published for other group I introns.
F. The Secondary Structure of Group I Introns

In RNA autocatalysis, the activity and specificity of the reaction is determined by the folded structure of the RNA molecule. In group I introns, this structure is defined by the interaction of a set of well-characterized sequence elements. These elements always occur in the same order in the intron; however, the spacing of some elements varies widely. Following the nomenclature of Davies et al. (Davies et al., 1982), the elements are, in the order of their occurrence from 5' to 3': the internal guide, E, P, Q, R, E', and S. Taken together, they constitute a set of cis-acting elements that have been shown to be required for autocatalytic activities of group I introns (see below). The internal guide is a sequence that can basepair with the 5' and 3' exons and apparently acts to align the exons. The rest of the elements can pair as follows: E and E', P and Q, R and S. When the secondary structure of a group I intron is drawn using these standard pairings, a typical core structure is produced. This structure is shown in the standard form in Figure 8. There are nine stem-loops in the basic structure, four of which are produced by pairings of the conserved elements: P1 is formed by the internal guide and the 5' exon, P3 is formed by E
and F', P4 is formed by P and Q, and P7 is formed by R and S. Additional stem-loops may be present, depending on the size of the particular intron. Recently, the nomenclature of the structure has been standardized (Burke et al., 1987). The nine characteristic pairings are termed P1 through P9, numbered according to their occurrence 5' to 3' on the RNA. Other stems which may occur between conserved stems m and n are called Pm.1, Pm.2, etc. Some stems are interrupted by bulges; in such cases the stem segments are termed Pm, Pma, Pmb, etc. Loops are termed Lm, Ln, based on the number of the stem at its base. As an example, the Tetrahymena rRNA intron is shown folded into this structure in Figure 9. Evidence for the importance of the cis-acting sites and their interaction has come from phylogenetic comparisons and from the study of in vivo and in vitro mutations. In the following sections, each interaction will be discussed separately.

1. The Internal Guide and Its Interactions

The existence of the internal guide was first suggested by Davies et al. (Davies et al., 1982) based on the observation that all group I introns possess a sequence that could basepair to both the 5' and 3' exons, thereby aligning the exons for ligation. The
pairing of the internal guide sequence (IGS) with the 5' exon forms the PI stem. The potential basepairing begins within the 5' exon and often continues several bases beyond the 5' intron/exon boundary. The pairing invariably pairs a G in the IGS with the last base of the 5' exon, which is always a U. This interaction appears to make the bond at the intron/exon junction susceptible to nucleophilic attack.

Several groups have studied the PI interaction. Perea and Jacq (1985) studied revertants of a cis-acting mutant of intron 4 of the yeast cob gene G2457. This mutant is a G to A transition at the base two basepairs 5' of the intron/exon boundary, which alters the 5' exon IGS basepairing. Several classes of revertants were found, one of which proved to be a second site revertant in the IGS that was a C to U transition restoring the basepairing of the IGS with the 5' exon.

Chandry and Belfort (1987) have isolated a spontaneous splicing defective mutant of the T4 td intron, which proved to be in the 5' exon three bases from the intron/exon boundary. The mutation is a G to A transition, which disrupts a G:C basepair in PI. This mutant processes td pre-mRNA at a cryptic site 29bp
upstream of the normal 5' splice site. The 3' splice site was not affected. The sequence around the cryptic 5' splice site showed complementarity to the IGS. Mutant RNA was also processed at the cryptic site in vitro, but at a much slower rate than the wild-type reaction.

Two groups have studied the role of the IGS in the self-splicing of the Tetrahymena pre-rRNA. Both groups coupled accurate splicing of the intron to production of beta-galactosidase activity in mutant E. coli strains in a system similar to that used to show insertion in pUC-based plasmids. Colonies carrying splicing defective introns were detected as white or light blue colonies or plaques on X-gal plates. Waring et al. (1986) studied one spontaneous mutant and several in vitro generated mutants and second site suppressors in the P1 stem. The spontaneous mutation was picked as a light blue colony. This mutation proved to be a C to U change four basepairs upstream of the 5' splice junction. This base is predicted to bind with a G in the IGS. The mutation thus retains a weak G:U basepair in the stem, probably accounting for the light blue phenotype. Using in vitro mutagenesis techniques, the authors altered the G involved in the pairing to a U. When assayed
separately, this change completely blocked splicing based on colony color. The double mutant, in which the G:C basepair was replaced with an A:T basepair, was found to splice at a rate slightly below that of the wildtype. The rate of splicing, therefore, seems to be dependent on the strength of basepairing at this stem position. In addition, these authors generated two complementing double mutants. The first changed the second and third bases upstream from the 5' splice from UC to AG, and the second changed the bases in the IGS thought to bind to them from GA to CU. Each mutant eliminated splicing; the double mutant which restored the P1 basepairing also restored splicing. The rate of splicing was reduced, however, suggesting that sequence as well as pairing may play a role in splicing efficiency.

Been and Cech (1986) performed a similar study with similar results. Six single base changes were generated that disrupted the P1 stem; three of the changes were in the 5' exon and three in the IGS base thought to bind with the mutated exon bases. All single base changes reduced or eliminated splicing as judged by colony color and in vitro assay. One single base change in the exon, a C to G transversion two bases 5' of the intron/exon
boundary, produced normal amounts of ligated exons in vitro when Mg"²⁺ concentration and temperature were raised. When sequenced, the exons were found to be incorrectly ligated. The 5' splice site was moved two bases upstream; the 3' splice site remained unaltered. Double mutations, which restored Pi basepairing but not exact sequence, restored splicing, but not to wild-type levels.

Other activities of group I introns are also affected by the IGS. In the study discussed above, it was found that modification of the IGS could change the sequence specificity of the intermolecular splicing and nucleotidyltransferase activity of the *Tetrahymena* rRNA intron (Been and Cech, 1986). When the G of the IGS, which pairs with the final C of the CUCUCU terminal bases of the 5' exon, was changed to a C, the specificity of the intermolecular splicing activity was changed from CpU to GpU. These researchers also altered the region of the IGS which binds with the 5' exon from GGAGGG to GAAAAG. The L-19 IWS with this modification retained its nucleotidyltransferase activity, but the specificity of the reaction was changed. No activity was observed on the normal C₅ substrate; however, there was activity on a new substrate, U₆. The authors
concluded that IGS binding was responsible not only for the specificity of splicing, but for these two additional activities as well. This conclusion is strengthened by the result of Garriga et al. (1986) who found that the Neurospora cob intron 1, another self-splicing group I intron, also has transsplicing activity but with a different specificity, GpU. This sequence is similar to the terminal sequences of the first exon of the Neurospora cob gene, UGGGU, and has complementarity to the Neurospora IGS.

Until recently, it was assumed that the same binding site in the IGS that was responsible for the specificity of the splicing, intermolecular splicing, and nucleotidyltransferase activities of the Tetrahymena rRNA intron was also responsible for the specificity of the cyclization reaction as well. This assumption was driven by the observation that, for those group I introns that circularize, the circularization sites are similar to the two or three bases preceding the 5' splice site and could bind to the IGS in a similar manner. This binding would then make the bond at the circularization site susceptible to nucleophilic attack from the G at the end of the intron (Zaug et al., 1983). Been and Cech (1987) have recently found this to be an
oversimplification. Those researchers found that base changes in the IGS that stopped the splicing reaction did not stop the cyclization reaction. Other mutations of the IGS and the base immediately 3' to it have no effect on splicing, but interfere with the cyclization activity. A mutation in the circularization site which abolishes circularization could be suppressed by two different second site suppressors in the IGS. These findings show that the exact relationship of the IGS to the cyclization activity of the *Tetrahymena* rRNA intron remains to be elucidated.

The pairing of the IGS with the 3' exon forms a stem, termed P10, which is not generally shown in the standard structure. The complementarity of this pairing is well conserved in most group I introns, suggesting some role in splicing (Waring and Davies, 1984). This stem has not been studied as rigorously in experimental systems as the P1 stem. Been and Cech (1985) deleted portions of the internal guide thought to pair with the 3' exon. Exon ligation was not affected *in vitro*. No *in vivo* or *in vitro* point mutants have been reported in this stem; therefore, the reason for the conservation of this pairing remains obscure at this time.
2. E, E' and the P3 Pairing

The E/E' pairing forms the P3 stem. The sequence of this pairing is poorly conserved; only a single base, a G at the 5' end of E', is regularly found. This pairing was originally proposed by Davies et al. (1982) because this sequence could always basepair with a sequence immediately 5' to the P sequence (E). Two groups have isolated mutants in the yeast cob intron 4 (bI4) which altered this sequence (Anziano et al., 1982; De La Salle et al., 1982). These groups isolated a total of three mutants, all of which alter the conserved G and disrupt the proposed P3 stem. Subsequently, Weiss-Brummer et al. (1984) reported five mutants in E'; again all mutants affected the conserved G and disrupted the E/E' pairing. Three of the mutants changed this base to an A, while the other two changed it to a T. Six revertants of these mutants were isolated. Four were true revertants which returned the mutated base to the original G. The other two were second site revertants in E of mutants where the conserved G in E' was replaced by an A. Both these revertants restored the P3 stem by replacing the C thought to pair with the conserved G with a T.
Williamson et al. (Williamson et al., 1987) used site specific mutagenesis to construct mutants in E and E' of the *Tetrahymena* rRNA intron. Two bases were changed in both sequences; E was altered from GACGUC to GACUCU, and E' was altered from GUOGGUC to AUGCUC. Individually, each of these alterations eliminates the splicing, trans-splicing, and 3' hydrolysis activities of the intron. The splicing activity is blocked at the G addition step, the first step in splicing. The splicing deficiency could be overcome by increased Mg**++** concentration at 42 degrees. When the mutant E and E' were combined in a double mutant, the potential for basepairing is restored. The authors found that the intron was fully active in this case. The kinetics of self-splicing were found to be similar to that of the wildtype pre-rRNA.

3. P, Q, and the P4 Pairing

The P-Q pairing, P4, was proposed by Davies et al. (1982) on purely phylogenetic criteria. Those authors noted that the sequence of each element was well conserved in the introns and species they studied; however, some variations existed. What the authors found striking was that, in the nine introns they looked at, each base change in one element was compensated for
by a corresponding change in the other element. Waring and Davies (1984) later extended this comparison to 17 additional group I introns. They found that, despite variations from the consensus sequences, P and Q could always basepair. Only two introns, the Tetrahymena rRNA intron and the Saccharomyces cerevisiae a15alpha intron, were found to have bulged bases in the P4 stem and these contain additional GC basepairs that increased the overall stability of the structure. Based on this observation and on the lack of in vivo mutants in the P4 stem, it has been suggested that single base changes may be tolerated by this structure so long as the overall integrity is not compromised (Waring et al., 1985).

Using the system described above that linked intron splicing to beta-galactosidase activity, Waring et al. (1984) studied the effect of two in vitro produced mutants in the P4 stem on the self-splicing of Tetrahymena pre-rRNA. A single base change in Q disrupting a GC pairing in P4 produced white colonies on X-gal plates, indicative of a strong splicing block. The authors suggest that the tight phenotype of this mutant may be related to the fact that this intron already contains a mismatch in the P4 stem, at a point next to the mutation site. A mutation which created a
GU pair at this same point produced light blue colonies, suggesting reduced splicing efficiency.

4. R, S, and the P7 Pairing

R and S are cis-acting elements whose sequence is well conserved. Part of both elements can basepair to form the P7 stem in the standard group I folded structure. Davies et al. (1982) noted the phylogenetic conservation of the sequence and basepairing. Early studies of mRNA processing in yeast mitochondria described mutants defective in the splicing of the cob intron 4 and oxi3 intron 4 which altered either R or S (Anziano et al., 1982; De La Salle et al., 1982; Weiss-Brummer et al., 1982). The bulk of these mutants mapped to bases involved in the P7 pairing, highlighting the importance of this interaction. One mutant studied in cob intron 4 of S. cerevisiae (cobI4) by Anziano et al. (1982), S118, proved to be in a region of R which could not basepair with S. This result was an early suggestion that the non-pairing bases of these elements have a direct function in splicing. Holl et al. (1985), working with cob I4, isolated a second site suppressor of a G to A transition mutant, G2590, in S which interferes with the P7 stem. This suppressor is a C to
T transition in R which re-establishes the pairing lost by the original mutation.

Several in vitro mutagenesis studies have been performed on the R and S elements of the Tetrahymena rRNA intron. Waring et al. (1985) reproduced the G2590 mutation of yeast bI4 (Holl et al., 1985) in their system which linked intron splicing to beta-galactosidase production. E. coli colonies carrying this mutation were light blue on X-gal plates, indicating that splicing of the intron was impaired but not totally blocked. Burke et al. (1986) created complementing mutants in the P7 stem by altering the pairing bases of R from GACUA to GAGUC and the pairing bases of S from UAGUC to GACUC. These mutants were studied in vitro singly and as a double mutant which reconstructed the P7 stem. In the standard conditions used by the authors (30mM Tris-HCl (pH 7.5), 2mM MgCl2, 100mM (NH4)2SO4 and 0.2mM GTP at 42 degrees), the single mutants showed no splicing, cyclization, or hydrolysis activities. However, when the Mg++ concentration was increased to 10mM, full activity of these mutants was restored, although the rate of splicing was lower than wildtype. The R mutant was fully active at 5mM Mg++. The S mutation showed splicing, but not cyclization
activity, at this concentration. At 50mM Mg**, both mutants were slightly more active than wildtype. The double mutant was inactive when assayed at 30 degrees, a temperature that supports full activity in wildtype. At 42 degrees, however, this mutant was active in intron excision, G addition, and exon ligation, albeit at a slower rate than wildtype, but deficient in cyclization and hydrolysis at the 3' splice site. These deficiencies could be overcome by increasing Mg** concentration to 10mM.

The most complete study of the role of R, S and the P7 stem was undertaken by Williamson et al. (1987) in an extension of the study by Burke et al. (1986) mentioned above. They tested the hypothesis that the increased temperature requirement of the compensatory double mutant mentioned above was due to the increased stability of the P7 stem due in turn to the mutant bases. To this end, they created new mutations in R and S. The stem forming bases of R were altered from GACUA to GAGUU. The stem forming bases of S were altered from UAGUC to GACUC. The P7 stem in the compensatory double mutant was predicted to have the same stability as the wildtype stem. As in the first study, the S mutant was inactive under standard conditions. Splicing, but not
cyclization or 3' hydrolysis, could be restored when Mg** concentration was increased from 2 to 10mM. Contrary to the hypothesis, the double mutant required 42 degrees to self-splice and at this temperature was defective in cyclization and 3' hydrolysis. Therefore, the stability of the stem is not the only factor affecting splicing in this interaction.

When the R mutant alone was analyzed, there was an even more surprising result. The mutant was more active than the wildtype at both 30 and 42 degrees. There was both an increase in $k_{cat}$ and a decrease in $K_M$. The authors also studied mutations in the conserved nonpairing bases of R and S. R was mutated from UCAOGACUA to AOOGACUA, and S was mutated from AAAGAUAGUC to OGAUAUGUC. Neither mutation alters the P7 stem. When the R mutant was assayed in vitro, it was found to have reduced affinity for GTP. At 200mM GTP, the rate of splicing of this mutant was the same as wildtype. At $\leq 50$uM GTP, however, the rate of splicing was significantly slower than wildtype. The $K_M$ for GTP was calculated to have increased four fold. The S mutant was found to be defective in splicing and cyclization due to an inability to carry out transesterifications associated with these activities.
This mutant was able to bind GTP and retained the 3' hydrolysis activity.

G. Proteins Encoded by Group I Introns

Many group I introns contain long open reading frames (ORFs). These may be free standing or contiguous with the upstream exon. In most group I introns where ORFs are present, the cis-acting elements IGS, E, P, Q, E' and R are within the ORF. S is never found in the reading frame. The interrelationship of the ORF and the cis elements argues for a long evolutionary association between the two. This may not be the case for some other introns where the reading frame is located between the elements (such as the yeast, Neurospora, and Aspergillus mitochondrial rRNA introns) or preceding the elements (such as yeast aI3). These cases may represent the recent acquisition of a reading frame through a transposition event. Several ORFs are known to be translated and the function of the proteins known. The proteins encoded by group I introns include both those with structural (a ribosomal protein) and enzymatic (maturases and endonucleases) functions. Both cis-dominant and trans-recessive mutants can be isolated from yeast introns which encode a functional protein. In this case, trans mutants are those that alter the
activity of the protein and may be compensated for by an active protein encoded elsewhere and cis mutants are those that alter the structure of the intron and cannot be compensated for by a diffusible factor encoded elsewhere.

1. Group I Maturases

One group of proteins encoded by group I introns has been shown to be required for processing the intron encoding it in vivo. These proteins are referred to as maturases. Maturase function has been assigned to the proteins encoded by yeast mitochondrial group I introns bI2 and bI4. Maturase activity is not limited to group I introns; two group II introns have been shown to encode maturases as well. The maturase model was first advanced in conjunction with the ORF of bI2, by Lazowska et al. (1980). The model was based on genetic, phenotypic, and sequence analysis of trans-acting mutants of that intron. The bI2 maturase is translated from the initiation codon of the cob gene. The authors noted that the protein is self-regulatory, since it destroys its own transcript. There must be an equilibrium between the unspliced bI2 message and the active maturase encoded by the intron.
The most thoroughly studied maturase is the protein encoded by cob intron 4 (bI4). The protein product of this intron is responsible for the BOX phenomenon in yeast mitochondria. This refers to a group of pleiotropic mutants which map to several sites in the cob gene. Lesions in these regions result in mitochondria deficient in cytochrome oxidase as well as cytochrome b. Several groups (Dhawale et al., 1981; De La Salle et al., 1982; Weiss-Brummer et al., 1982; Anziano et al., 1982) have analyzed trans-recessive mutants mapping to bI4. Those studies showed that the maturase encoded by bI4 is not only required for the processing of that intron but also for the processing of the fourth intron of the coxI gene (aI4). Thus, any mutation which affects the expression of the bI4 maturase results in the loss of cytochrome oxidase function as well as the loss of cytochrome b function. Since the the maturase is translated from the initiation codon of the cob transcript through exons 1, 2, 3, and 4 to the stop codon within the intron, such mutations could include not only those within the intron but also any that might affect mRNA processing or translation upstream. The initial translation product, which includes the translation product of exons 1-4 and the intron, is cleaved to release a 27 kd intron encoded
protein from the C terminus that is thought to be the active form of the maturase \textit{in vivo}. The precise location of the cleavage is not known.

To allow more careful study and \textit{in vitro} manipulation of the bI4 maturase, Banroques et al. (1986) have modified the reading frame of bI4 so that it could be translated on cytoplasmic ribosomes. This sequence was then cloned into a high copy number plasmid in-frame with the presequence of subunit 9 of the \textit{Neurospora} ATPase, which targets the translated product to the mitochondria. When transformed into the yeast nucleus, this construct has been shown to correct bI4 maturase deficiencies. The fact that this construct contains only intron sequences from yeast argues strongly that the active maturase is the 27kd proteolytic fragment, not the precursor. In a recent study (Banroques et al., 1987), the same group showed that a maturase that was deleted of 45 intron encoded amino acids from the amino terminus was even more active than the original construct. This protein has been shown to have a mobility similar to that of the 27kd protein found \textit{in vivo}. This suggests that the cleavage may be required to activate the I4 maturase.
2. The S-5 Protein

The mitochondrial L-rRNA gene of Neurospora mtDNA contains a group I intron which encodes a protein of known function. Based on the size and amino acid composition of the encoded protein, Burke and RajBhandary (1982) suggested that it is probably the mitochondrial ribosomal protein, S-5. S-5 is a subunit of the small ribosomal subunit of the Neurospora mitochondrial ribosome. This protein was originally characterized by LaPolla and Lambowitz (1979; 1981). Lending credence to this identification are the observations that the S-5 protein is translated on mitochondrial ribosomes (Collins et al., 1979) and that mutants that are blocked in the splicing of the L-rRNA precursor lack S-5 and the small ribosomal subunit.

3. Endonucleases

Another function which has been shown to be encoded by ORFs of group I introns is that of a site specific double strand endonuclease. This activity has been carefully studied in the intron of the L-rRNA gene of yeast mitochondria, but it has been recently shown to be an activity of a protein encoded by aI4 as well (P. S. Perlman personal communication).
The optional group I intron of the L-rRNA of yeast mitochondria has been termed the omega intron. The omega intron contains a long free-standing ORF that encodes a protein of 235 amino acids (Dujon, 1980). Strains carrying this intron are termed omega+ and those not carrying it are termed omega−. In crosses of the type omega+ x omega−, the omega− alleles are converted in a site specific gene conversion process to omega+. A double strand cut at the site of insertion has been shown to be an intermediate step in this process (Zinn and Butow, 1985). A mutation of the L-rRNA gene, termed omegaN, which modifies the sequence near the insertion site, prevents the conversion, demonstrating the sequence specificity of the activity. The conversion is extremely efficient. In a cross where the input of parental genomes is equal, the output frequency of genetic markers should be equal. In omega+ x omega− crosses, however, over 99% of the progeny are found to be omega+ (reviewed by Dujon, 1981; Butow, 1985). Flanking markers are also converted, at an efficiency inversely proportional to their distance from the intron insertion site. This phenomenon has been termed "polarity of recombination" (Dujon, 1981).
The protein product of the omega intron is required for polarity of recombination. A mutant termed omega<sup>d</sup>, which carries a one base deletion leading to premature chain termination of the omega protein, does not convert omega<sup>-</sup> alleles to omega<sup>+</sup> (Jacquier and Dujon, 1985; Macreadie et al., 1985). Furthermore, this phenomenon is not seen in matings of omega<sup>+</sup> x omega<sup>-</sup> petites, which due to their restricted mitochondrial genomes do not carry on mitochondrial protein synthesis (Jacquier and Dujon, 1985; Zinn and Butow, 1985). The sequence coding for the omega protein has been modified to allow its translation in *E. coli*. Using this system, Colleaux et al. (1986) have shown that the protein encoded by this sequence is a site-specific double stranded DNA endonuclease that cuts DNA at a sequence identical to the insertion site of the omega intron (Zinn and Butow, 1985). While the sequence at the insertion site is known, the exact recognition requirements for the endonuclease are not known. The endonuclease probably has a rather long and specific recognition sequence because, while the active protein was expressed in the *E. coli* cell, the host genome was not affected, suggesting that there are no sites on the entire chromosome.
Recently, the ORF of yeast coxI intron four (aI4) has been shown to encode an endonuclease analogous to the omega protein (Wenzlau et al., 1989). In crosses between aI4+ and aI4− strains, a site specific gene conversion event similar to omega conversion takes place. This conversion, like the omega conversion, is initiated by a double strand cut at the insertion site by the endonuclease encoded by the aI4 ORF. The sequence specificity of the aI4 endonuclease is different from the omega endonuclease. Conversion of aI4 also shows conversion of flanking sequences. In strains containing the intron aI3 gamma upstream of the conversion site, this intron is lost in the conversion reaction with strains that contain aI4 but not aI3 gamma.

The aI4 protein is remarkably versatile; not only does it have site specific endonuclease activity, but it can also be recruited to replace the bI4 maturase which, as stated above, is required for splicing of bI4 and aI4. Dujardin et al. (1982) isolated a second site suppressor of bI4 trans mutants, mim2-1, which proved to be a single G to A transition in the aI4 ORF, thus allowing the protein product to act as a maturase. Two groups (Dujardin et al., 1980; Labouesse et al., 1985;
Anziano et al., in preparation) have isolated nuclear suppressors of bI4 trans mutants. These suppressors, called NAM2-1 by Dujardin et al. and B2 by Anziano et al., are probably allelic since their mode of action is identical and they both map to the same nuclear chromosome (H. Conrad-Webb personal communication). They suppress bI4 trans-recessive mutations by interacting in some way with the protein product of aI4 so as to allow it to assume a maturase function. Doubly mutant mitochondrial genomes with a trans mutation in bI4 coupled with an additional lesion which prevents translation of the aI4 ORF do not support B2 suppression (Anziano et al., in preparation).

Many other group I introns contain ORFs from which no protein product has been observed or function assigned. Some of these will almost surely be found to code for functional proteins. It seems unlikely that so many reading frames could be maintained if there was not a positive selective advantage associated with them.

H. Proteins Involved in Splicing Group I Introns

It is generally assumed that the splicing of most, if not all, group I introns is facilitated by proteins. Even those introns which have been found to self-splice
in vitro probably do not do so in vivo. This is, at least in part, related to the absence of the ionic and temperature conditions required for efficient self-splicing within the cell. Even under the most conducive conditions, the Tetrahymena pre-rRNA has been calculated to self-splice at only 1/60 the rate of the in vivo reaction (Kruger et al., 1982). As noted above, several of these introns encode maturases required for their own splicing; however, some others do not. Therefore, other factors must exist which aid in in vivo splicing.

Besides maturases, several genes have been shown to directly affect splicing of group I introns. These are the cyt-18 gene of Neurospora and the CBP2, MSS18, and MRS1 genes of yeast. As described above, the NAM2 gene can suppress trans mutants in bI4 by interacting with the protein product of the aI4 ORF and, therefore, may be involved in group I splicing. Another gene product of yeast which may be involved in group I intron splicing is that of the MSS51 gene. All of these genes are nuclear encoded, but their target introns are mitochondrial.

Cyt-18 is a nuclear gene that has been shown to be required for splicing of most Neurospora mitochondrial group I introns (Garriga et al., 1986). The group I
intron of the L-rRNA gene of *Neurospora* mitochondria does not self-splice \textit{in vitro}. It will, however, splice \textit{in vitro} in mitochondrial lysates. When these lysates are prepared from a cyt-18 mutant background, no splicing is observed. Recently, Akins and Lambowitz (1987) have cloned and sequenced this gene. The surprising result of that study is that the cyt-18 gene codes for mitochondrial tyrosyl-tRNA synthetase. That protein acts as both a splicing protein and a tRNA synthetase \textit{in vivo}. The splicing activity of cyt-18 is not dependent on the tRNA synthetase activity. Partial second site revertants in the cyt-18 gene which allow splicing retain the tRNA synthetase deficiency (Akins and Lambowitz, 1987). The exact nature of the splicing function of the cyt-18 gene has yet to be elucidated.

The CBP2 gene has been shown to be required for the excision of the last intron of the cob gene, bI5, in yeast. McGraw and Tzagoloff (1983) found that mutants mapping to this gene were specifically blocked in the splicing of this intron leading to a specific deficiency of cytochrome b. The wild-type gene was cloned and sequenced. It was found to code for a protein of 630 amino acids with a molecular weight of 73,678. The exact activity of the protein product has not yet been
determined. The specificity of the CBP2 gene has been
demonstrated by the finding that mutations in CBP2 can
be suppressed by a mitochondrial mutation which
precisely excises bI5 (Hill et al., 1985).

The MSS18 has been shown by Seraphin et al.
(Seraphin et al., 1988) to be required for the excision
of the group I intron aI5beta. This gene has been
cloned, sequenced, and disrupted. MSS18 disrupted
strains grow slowly on glycerol, suggesting that the
defect is leaky. These strains accumulate a 3.8kb
mitochondrial RNA which hybridizes to aE1, aE6, and
aI5beta probes and is the correct size to be the coxl
message plus aI5beta. Mitochondrial revertants of MSS18
disrupted strains have been isolated and found to be of
two types, large deletions that cleanly removed aI5beta
either alone or in combination with aI5alpha and
aI5gamma, and point mutations located in aE5beta. Two
different point mutations were found; both were A to T
transitions. They were located 10 and 13 bases upstream
of the 5' intron/exon boundary. This is well beyond the
recognized PI stem of this intron. The authors suggest
that these mutations may suppress the splicing defective
phenotype by altering the local secondary structure.
Further, they speculate that the MSS18 protein may play a role in the 5' cleavage of αJ5βeta.

Kreike et al. (Kreike et al., 1986) have shown that the MRS1 gene product is necessary for the processing of the yeast mitochondrial intron bI3. Mutations in this gene have a pleiotropic phenotype; bI3, bI4, and aJ4 are not spliced. However, in strains which lack bI3, but retain bI4 and aJ4, mutations to MRS1 do not result in respiratory incompetence. Therefore, the incapacity to splice bI4 and aJ4 derives from a failure to produce the bI4 maturase in MRS1 mutant strains retaining aI3, not from the direct lack of MRS1 function. These authors localized the MRS1 gene to chromosome IX.

The yeast nuclear NAM2 gene has been studied by two groups, Slonimski's and Perlman's. The NAM2 gene was originally isolated as a dominant nuclear suppressor of bI4 maturase mutants (Groudinsky et al., 1981). As explained above, the mode of suppression involves the activation of the aJ4 protein product to maturase activity. The action of the gene product has been difficult to ascertain. Null alleles created by gene disruption result in the loss of the mitochondrial genome. This led to speculation that the NAM2 protein
is somehow involved in mitochondrial translation (Labouesse et al., 1987). Mutations in yeast which interrupt translation are known to result in increased petite production. The NAM2 gene and three suppressive alleles have been sequenced (Labouesse et al., 1987). The wild-type protein is 894 amino acids long with a molecular weight of 102,000 daltons. All three suppressor mutants alter the same amino acid, a glycine at residue 240, to either a serine or a cysteine.

Recently, Herbert et al. (Herbert et al., 1988) have conclusively demonstrated that the NAM2 gene encodes a mitochondrial leucyl tRNA synthetase. These authors assayed tRNA synthetase activity in mitochondrial extracts from three yeast strains; the first contained the wild-type NAM2 gene, the second contained a disrupted NAM2 gene and the third contained a disrupted chromosomal NAM2 gene but carried the wild-type gene on a plasmid under the control of the GAL10 UAS so that transcription was inducible by galactose. The level of leucyl tRNA synthetase was three times higher in the wild-type strain than the disrupted strain. The strain with the NAM2 gene under control of the GAL10 UAS was found to have much higher levels of leucyl tRNA synthetase when transcription was induced by
galactose. When each of these strains was assayed for other tRNA synthetases no differences were found. Furthermore, antiserum raised against a chimeric beta-galactosidase/NAM2 protein was found to inhibit leucine charging activity. This study also noted that the NAM2 protein has considerable homology to the E. coli leucyl tRNA synthetase. The NAM2 protein was found to have 34\% identical amino acids. The highest homology was found in the first 220 amino acids, where 46\% were identical in the two proteins.

The MSS51 gene product is required for the processing of yeast coxl pre-mRNA (Simon and Faye, 1984). Strains carrying mutant alleles of this gene can only excise aI3 correctly. Processing of other introns is either completely or partly blocked. Introns aI1 and aI2 are completely blocked, while the excision of aI4 and aI5 is severely retarded. The MSS51 gene product may be required for general processing of the coxl transcript. Alternatively, it may be only required for one or two introns, and the blockage of the others stems from a nonproductive secondary structure of the transcript caused by the retention of the target intron.
It is worth noting that aI3 may be removed by self-splicing.

In reviewing the data presented above, one is left with the surprising conclusion that there is no general protein complex which is responsible for splicing all group I introns. While general group I splicing proteins such as cyt-18 exist, they probably work in conjunction with intron specific factors such as maturase and the MRS1 gene product. One possible mechanism is that there is a general core complex for all group I introns in an organism whose action is directed and modified by the association of intron specific factors. This would be analogous to the modification of Bacillus subtilis RNA polymerase by various sigma factors to allow recognition of different promoters (Losick and Pero, 1981). Alternatively, there may be separate families of group I introns, each of which uses a distinct set of proteins for splicing.

I. aI3

There are several reasons to study coxl intron 3 (aI3). First, the intron has an unusual structure, with the IGS separated from the 5' intron/exon boundary by a 1kb ORF. Second, our laboratory has isolated the only
mutants known to exist in this intron. Several of these mutants, though shown to be cis mutants, map to regions outside the known location of the cis-acting sites. The \textit{in vivo} phenotypes of these mutants had been well characterized by Mecklenberg (1986), but the sequence change for each was unknown. Third, when this study began there was an indication in the literature that aI3 is autocatalytic. During this study, Tabak et al. (1984) carefully characterized the aI3 self-splicing products so it was of interest to confirm their results and to assess the effects of the mutants on the autocatalytic activities of the intron.

1. The Sequence and Structure of aI3

The sequence of the coxl gene from \textit{S. cerevisiae} strain D273-10B/A1 (referred to as D273), which contains aI3, was published by Bonitz et al. (Bonitz et al., 1980). The coxl sequences germane to this study are shown in Figure 10. Proposed cis-acting sites in aI3 are shown in Figure 11. To avoid confusion the numbering system of Bonitz et al. has been retained and will be used throughout this dissertation.

Davies et al. (1982) included aI3 in their original analysis of the conserved sequences of group I introns.
They noted that aI3 has an unusual structure. The intron is 1514 bp long. There is a 1008 bp ORF, which is contiguous with the upstream exon, followed by 506 bp of CRF. The IGS of this intron is separated from the 5' splice site by 1097 bp, creating a large L1 loop. All of the known cis-acting sites are located in the final 418bp of the intron. This is considerably different from most group I introns, where the L1 loop is less than 50bp, and sometimes as small as 5-6bp (Davies et al., 1982; Waring and Davies, 1984). Another example of a group I intron with this type of structure is the Neurospora crassa cob intron 2, where the L1 loop is 965bp (Waring and Davies, 1984).

The distance between the IGS and the 5' intron/exon boundary means that the L1 loop must have a stable secondary structure if the PI stem is to form. Davies et al. (Davies et al., 1982) stated that they doubted that the IGS was used to align the exons of aI3. This was, in part, based on misinformation. Those authors misread the published sequence and deleted a crucial G from the IGS. Furthermore, the published sequence of the D273 coxI gene at the IGS is at variance with that of other yeast strains. Tabak et al. (1987) have sequenced this region in strain KL14-4A and found that
the IGS sequence is AATAAGTGCG rather than AATAAGATGCG. This observation has also been made in this study for strain ID41-6/161 (referred to as 161 hereafter). These changes improve the proposed PI pairing. Recent data on the autocatalyzed reactions of this intron by Tabak et al. (1987) strongly suggest that the IGS is functional.

Another structural feature of aI3 is that it belongs to a subfamily of group I introns that have long L5 loops. Collins (Collins, 1988) studied 40 group I introns and found that they could be divided into two categories based on the length of the L5 loop, either long (59 to 295bp) or short (21 to 38bp). Nineteen introns, including aI3, were found to be in the long L5 category. The L5 loops of these introns shared a common secondary structure, featuring an adenine rich bulge a fixed distance from the P4 stem formed by the P-Q pairing.

2. Mutants of aI3

The major portion of this thesis focuses on the mutants M44, C1072, and C1085, which have been shown to map within or near aI3 (Mecklenberg, 1986) and to block its splicing in vivo (Sass, 1984). M44 and C1072 map to
the ORF, while C1085 maps to the CRF. Mecklenberg (1986) has analyzed the in vivo effect of these mutations. Complementation tests showed that each of them is cis-acting. Labeling of mitochondrial proteins showed protein patterns unique to aI3 mutants. All three mutants lack coxl, the protein product of the coxI gene. M44 and C1085 accumulate two novel proteins estimated to be 40 and 44kd. These are termed p40 and p44, respectively. Mutant C1072 accumulates these proteins as well as an additional one estimated to be 42kd and termed p42. These proteins have been shown to be the same in all mutants by partial proteolytic analysis (Sass, 1984). This same analysis failed to reveal that they are related to coxl. However, these proteins must contain some amino acid sequences from coxl, since they are precipitated by a coxl specific polyclonal antibody. This suggests that the proteins are produced from both the coxI and the aI3 coding sequences.

The RNA phenotypes of M44 and C1072 were also investigated (Sass, 1984). Both mutants produce the same pattern when northern blots are probed with an aE4 probe. These mutants do not accumulate the 2.2kb mature coxI message; rather, two other species are detected.
The major species is approximately 3kb. This is too small to be message plus aI3, but it is the correct size to be aI3 plus downstream exons. A minor species of the same size is detected in wildtype blots, suggesting that this is a normal splicing intermediate or side product. A less abundant 3.7kb species is also detected. This is the correct size to be message plus aI3. Both species do not hybridize to probes from aI1, aI2, aI4, or aI5.

3. In Vitro Reaction of aI3

Tabak's group has also studied the reactions catalyzed by aI3. Before this study began, that group published a report that suggested that aI3 was autocatalytic. Amberg et al. (1986) incubated total isolated mitochondrial RNA with 32p-GTP under group I self-splicing conditions. In that experiment, any group I autocatalytic introns should be labeled by G addition during self-splicing. Several species were labeled, one of which was shown to be related to aI3.

Late in this study, Tabak et al. (1987) published experiments detailing the in vitro catalyzed reactions of this intron. Those authors cloned a 1984bp MspI-BclI fragment of the coxI gene (see Figure 10) which contains part of aI2, all of aE3, all of aI3, and part of aE4,
behind a SP6 promoter. When $^{32}$P-UTP labeled \textit{in vitro} RNA produced from this clone was incubated under group I self-splicing conditions, a number of discrete products were produced. The reactions as described in this study are illustrated in Figure 12. A typical group I catalyzed reaction takes place as outlined above, resulting in the release of the linear intron with a G added to the 5' end and the ligation of the exons. The linear intron can give rise to several additional products. Three circular products can be formed: a full intron circle that contains all of the original intron sequences, but not the G added during splicing; a circle containing the first 1083nt of the intron; and a circle containing the final 431nt of the intron. Both these circles can linearize or break to form linear molecules of the same length. The 1083nt molecule contains the intron CRF plus 75nt of the CRF. The 431nt molecule derives entirely from the CRF and contains all the known cis-acting sites. The breakpoint between these two species is a sequence, which is similar to the intron/exon boundary and which could pair with the IGS. This sequence is GGIGIAA. The break is after the first T at base 6286 in Figure 10.
These authors also found a reaction product of the linear intron that proved to be the two smaller circles interlocked. The linear intron was found to undergo a second G attack at the same internal site that is the breakpoint between the 1083nt species and the 431nt species. This results in the production of the same two linear molecules mentioned above, except that they are 1nt longer, because they each have an added G at the 5' end. The linear ORF molecule retains the G added during the splicing reaction, and the CRF linear molecule retains the G from the second G attack. The added G makes the sequence at the end of this molecule GGTAAC which is complementary to the IGS sequence. The linear CRF molecule can circularize with the loss of the added G, probably due to the ability of the IGS to interact with the 5' end of the molecule. The linear ORF molecule does not undergo further reactions, because it lacks the group I secondary structure elements necessary for autocatalysis.

Additional autocatalytic reactions of aI3 have been described by van der Horst and Tabak (1987). These researchers have demonstrated that under self-splicing conditions guanosine can attack at the 3' intron/exon boundary releasing the 3' exon from the full-length
transcript with a guanosine residue added to its 5' end. This reaction is analogous to the G attack that initiates the self-splicing pathway. In the same study it was shown that aI3 was also capable of a 5' exchange reaction, in which a GTP in the reaction buffer was exchanged for a 5' terminal G on the full-length transcript. This latter reaction was sequence specific in that it only occurred when the transcript was produced from an SP6 promoter which causes the first three nucleotides to be GAA. When the same transcript was produced using a T7 promoter, which causes the first three nucleotides to be GGG, the reaction was not seen. Both these same reactions were also shown for the omega intron. The authors suggested that these reaction are the result of interactions with the IGS.

4. Summary

As can be understood from the discussion presented above, aI3 is a unique group I intron. Its unusual structure had led some researchers to doubt that it used the IGS to align its exons for ligation. Yet, since it is capable of self-splicing, this unusual structure must fold in such a way as to catalyze these activities. I have chosen to study the structure and function of aI3 through a sequence and in vitro splicing analysis of the
wild-type intron, derived from strain 161, and three mutants, M44, C1072, and C1085, also derived from that strain. The mutants are all cis-acting, yet two, M44 and C1072, map to regions 5' of the known conserved sequence elements of this intron. By investigating the wild-type sequence for strain polymorphisms and the mutants for their sequence and effect on in vitro splicing, I hope to gain some insight into the intramolecular relationships which are required for the activity of this intron in vivo and in vitro.
II. EXPERIMENTAL TECHNIQUES

A. Strains

All strains used during the course of this study are listed in Tables 1 and 2.

B. Growth Media

1. Yeast Growth Media

Yeast media were prepared as described by Birky et al. (1978). The recipes are as follows:

2% YD:

- Bacto peptone: 10 grams/liter
- Bacto yeast extract: 5 grams/liter
- Glucose: 20 grams/liter

for agar plates:

- Bacto agar: 20 grams/liter

2% YGal:

made as YD except that Galactose replaces Glucose
RG:

- Bacto peptone 10 grams/liter
- Bacto yeast extract 5 grams/liter
- Glycerol 32 ml/liter

for agar plates:
- Bacto agar 20 grams/liter

2. Bacterial Growth Media

YT:

- Bacto tryptone 10 grams/liter
- Bacto yeast extract 5 grams/liter
- NaCl 5 grams/liter

supplements:
- ampicillin 50 mg/liter
- IPTG 5 ml/liter of 100mM soln
- X-gal 25 ml/liter of 2% soln in dimethyl-formamide

for agar plates:
- Bacto agar 20 grams/liter

for top agar:
- Bacto agar 7 grams/liter
C. Nucleic Acid Preparation

1. Mitochondrial DNA Isolation

Mitochondrial DNA was isolated as described by Hudspeth et al. (1980) except that spheroplasts were broken in a homogenizer rather than with glass beads.

2. Mitochondrial RNA Isolation

Mitochondrial RNA was isolated as described by Conrad-Webb (1988).

3. Rapid Plasmid Preparation

Five ml cultures of *E. coli* were grown overnight at 37°C. Cells were harvested by centrifugation at 6000 RPM in a JA-20 rotor for 5 minutes or by centrifugation in a tabletop centrifuge at full speed for 10 minutes. The supernatant was decanted and the pellet resuspended in 50 ul of 28% sucrose, 50mM tris pH 8. Then 300 ul STET buffer (8% sucrose, 0.5% Triton X-100, 50mM EDTA, 10mM Tris pH8) was added to each tube and the contents transferred to a microcentrifuge tube. Twenty ul lysozyme (10mg/ml in distilled water) was added to each tube. The solution was placed in boiling water for one minute. The tubes were then placed on ice and 50 ul distilled water (dH2O) was added. Cell debris was removed by centrifugation in a microfuge for 20 minutes
at 4°C. The supernatant was saved and the proteins were removed by phenol and chloroform extractions as described by Maniatis et al. (1982). DNA was precipitated by ETOH precipitation (Maniatis et al. 1982) and resuspended in 50 ul TE buffer (10mM Tris-HCl pH7.5, 1mM EDTA). Generally, 5 ul of this solution was used for each restriction digest.

4. Plasmid Preparation

Plasmids were prepared for use in cloning and transcription reactions by the Alkali Lysis method (Maniatis et al., 1982).

5. M13 Phage Preparation

Single stranded M13 phage was prepared following the protocol in the BRL "M13 Cloning/Dideoxy Sequencing Manual".

D. DNA Manipulations

1. Restrictions

DNA restriction enzymes were purchased from Bethesda Research Laboratories (BRL), New England Biolabs, Boehringer Mannheim, and International Biotechnologies Inc. All reactions were performed according to the manufacturer's recommendations.
2. Ligations

Ligations were performed with T4 DNA ligase from BRL under conditions recommended by the manufacturer.

3. Transformations

Bacterial transformations were performed using the CaCl$_2$ protocol as published by Maniatis et al. (1982).

E. Electrophoresis Techniques

1. Agarose Gels

Agarose gel electrophoresis was performed in horizontal gel apparatus manufactured by BRL. The running buffer was 1X TBE.

2. Acrylamide Urea Gels

Acrylamide gels were prepared as described in Maniatis (1982), except that 8M urea was added as a denaturant in gels intended for the analysis of RNA. Gels for the analysis of aI3 processing products were 3.5% acrylamide. The running buffer was 1X TBE.

3. Sequencing Gels

Sequencing gels were 7% acrylamide, 7M urea and were prepared as described in the BRL sequencing manual.
Stock solutions of 7% acrylamide, 7M urea instagel were prepared as described in the BRL manual and stored at room temperature in the dark. Under these conditions instagel was found to remain stable for up to one month. Sequencing gels were poured as described in the BRL manual except that Gel Sealing Tape was not used; rather acrylamide plugs were poured in the bottom of the gel plates using the same instagel used for the rest of the gel system. This system proved to be a superior method of sealing the plates. Gels were run at constant power of 30-45mA. Under these conditions, samples were loaded every 2-2.5 hours. Runs of six hours were sufficient to allow the reactions to be read for 300-400bp.

F. Northern Blots

Gels were electroblotted, and nick translations and hybridizations were performed as described by Conrad-Webb (1988).

G. Yeast Mitochondrial Genetic Analysis

Genetic analysis was performed as described by Mecklenberg (1986). The theoretical basis of mitochondrial mapping has been discussed by Tzagoloff (1982).
H. Sequencing Reactions

1. Single Lane Screening Reactions

Single stranded M13 phage DNA was isolated following the protocol in the BRL sequencing manual from a 2.5ml overnight culture and resuspended in 18 ul M13 TE buffer (10mM Tris base pH7.5, .1mM EDTA). Three ul of 100mM Tris base pH8, 50mM MgCl and 10ng of primer were added to the DNA solution. The primer was hybridized by heating the solution to 75°C. for 2 minutes and allowing it to cool to less than 35°C. One ul of the annealed M13 DNA was mixed with 3ul of the T sequencing mix (33um each of A, C, and G and 1.66um T) and 1ul 32P ATP (400 Ci/m mole). The reaction was started by the addition of 1ul of Klenow fragment (.5 units/ul). The reaction was incubated at room temperature for 15 minutes. Then 1ul of chase solution (2mM of each dNTP) was added and the reaction incubated for an additional 15 minutes at 36°C. The reaction was stopped by the addition of 5ul stop buffer (95% Formamide, 20mM EDTA, 0.05% Bromophenol Blue, 0.05% Xylene Cyanol) and analyzed on sequencing gels. When a clone specific primer was used, the incorporation of radioactivity in high molecular weight bands was sufficient to identify a clone of interest; otherwise, clones could be recognized by the pattern of T residues.
2. Standard Sequencing Reactions

The Sequenase sequencing kit from United States Biochemical Corp. was used for all reactions. The protocol published by the manufacturer was followed with the following modifications. All reactions were performed using the dITP labeling mix in which dITP replaces dGTP. This mix was used because better resolution of bands in all size ranges was noted using this base. Also, the labeling mix was diluted 10:1 rather than 5:1 to prevent the reaction from running beyond the resolving capability of the gel system.

I. Transcription and Purification of RNA

RNA was transcribed and purified as described in Peebles et al. (1986).
III. RESULTS

This chapter is divided into three sections: A) Genetic Mapping and Plasmid Construction, B) Sequencing Data, and C) In Vitro Processing Analysis. The first deals with preliminary experiments needed to verify the genotype of the subclones of each strain used and to produce the plasmids required for all subsequent experiments. The second section details the sequencing strategy and results. The sequence of both the wild-type (strain 161) and mutant strains (M44, C1072, and C1085) are presented and compared to the previously published sequence of strain D273 (Bonitz et al., 1980). The final section presents experiments leading to an optimized in vitro splicing reaction for the wild-type aI3 clone, experiments identifying reaction products, and experiments showing the in vitro effect of each mutant.
A. Genetic Mapping and Plasmid Construction

1. Cloning of the Wild-type aI3

The wild-type DNA for this study was derived from a petite, termed 17-17. As can be seen in Figure 13, this petite contains a region of about 15 kb from within HhaI band 7 to within HhaI band 8, totally encompassing the coxl gene. A17-17 is structured as a head-to-tail repeat of 30 kb. Several subclones of A17-17 were isolated and tested for functional integrity of the coxl gene sequences by mating against a coxl deletion mutant C245, shown in Figure 14. The sequence deleted in C245 is approximately 13 kb long. The deletion starts 2-2.5kb 5' of aEl and extends into aI5b past the PvuII site but not past the HhaI site in the same intron. Several subclones of C245 were used in the test mating as well. An isolate of A17-17 that restores all subclones of C245 strongly was selected as the mtDNA source for further experiments.

Mitochondrial DNA from A17-17 was isolated and digested with BglII and BclI. This releases a 2.2 kb fragment which contains 337 bp of intron 2, all of exon 3 and intron 3, and 319 bp of exon 4. Following agarose
gel electrophoresis and band-isolation, this fragment was ligated into the Bam site of the Bluescribe vector.

Bluescribe is a multi-purpose cloning plasmid which was used in all constructions intended for in vitro processing experiments. It is diagrammed in Figure 15. This vector contains the beta-lactamase gene that allows for selection of transformants on ampicillin plates. It also contains a portion of the lacZ gene with the multiple cloning site (MCS) (see Figure 15) imbedded within it. This feature allows for selection of recombinant plasmids based on colony color on Xgal IPTG plates. Plasmids without an insertion in the MCS express the lacZ gene and have a blue colony phenotype when resident in an appropriate E. coli strain; those with inserts generally do not express lacZ and have a white colony phenotype. In addition, phage T3 and T7 promoters are located on either side of the MCS allowing in vitro transcription of any insert off either strand. Bluescribe may also be used to produce single stranded phage DNA for sequencing by virtue of an M13 origin located just beyond the lacZ gene. It should be noted that the construction described here destroys the vector Bam site, making the Bam site in the closed reading frame (CRF) of aI3 unique.
E. coli strain JM101 was transformed and plated to ampicillin, Xgal IPTG plates. White colonies were picked and plasmid DNA isolated as described in EXPERIMENTAL TECHNIQUES. Two diagnostic restriction digests were performed to verify and characterize the isolated plasmids. The first, an EcoRI+HindIII double digest, has the effect of releasing anything cloned into the multiple cloning site. This digest confirms that the insert is the correct size, and since there are no HindIII or EcoRI sites within the mtDNA fragment, this digest also will determine if the insert has a multimeric structure. The second, an EcoRI BamHI double digest, further verifies the identity and shows the orientation of the clone by determining the existence of the BamHI site within aI3 and whether it is proximal or distal to the EcoRI site. If the BamHI site is proximal, a 412 bp fragment will be released; if it is distal, a 1794 bp fragment will be released. One clone which, by the above criteria, had a single insert and was oriented with the BamHI site distal to the EcoRI site was selected and termed BS17-17. The nomenclature for this construct combines a letter designation denoting the vector (BS for Bluescribe) and the name of the insert DNA source. This clone is shown in Figure
16 and was used for all subsequent in vitro analysis of the wild-type intron.

2. Genetic Mapping and Plasmid Construction with Mutants M44, C1072 and C1085

The identities of the yeast strains used as sources of mtDNA for the cloning were tested in two ways: first, by petite mapping and second, by conformation of the mutant protein pattern. The petites used and their map positions are shown in Figure 17. Table 3 summarizes the results of these matings. A brief discussion of the critical petite matings will serve to localize the mutants as closely as possible with the current petite collection. Mutants M44 and C1072 may be discussed together since they cannot be separated on the basis of petite mapping. There are three petites that are key to delimiting the region where these may exist. These are 1) 401, 2) 422/C2 and 3) 422/N1 (see Figure 17). Petite 401 does not restore these mutants and delineates their 5' limit. It has been sequenced and found to extend to position 5061 (Bonitz et al., 1980). The smallest petite which restores mutants M44 and C1072 is 422/C2. This petite retains the HpaII site at 5052 but can extend no further 5' than 5025 since it is a derivative of 422 (Mecklenberg, 1986) whose boundaries have been
sequenced to this point. Restriction mapping has shown that 422/C2 retains the Hinfl site at 6240 but not the HaeIII site at 6395. M44 and C1072 are not restored by 422/N1. The 5' end of this petite contains the HaeIII site at position 6395 but not the Hinfl at 6240. Taken together, these data show that M44 and C1072 must be between positions 5061 and 6395. C1085 is not restored by 422/C2 but is restored by 422/N1. It is not restored by 422/P3, which retains the Hinfl site at 6395 but not the DdeI at 6629. These data position C1085 between bases 6240 and 6835. These results are completely consistent with the data of Mecklenburg (1986), who originally mapped these mutants.

As a further test of the identity of these mutants, mitochondrial proteins of the subclones selected from the above tests were labeled with $^{35}$S with the assistance of Ken Chambliss of our laboratory and separated on an SDS polyacrylamide gel. The results are shown in Figure 18 and are again concordant with Mecklenburg's analysis. In M44 and C1085 the cytochrome oxidase band is replaced with two novel proteins of approximately 40 and 44 kd. Mutant C1072 produces these two proteins as well as a third of 42 kd. It should be noted that, while M44 and C1072 cannot be distinguished
by petite mapping, they can be distinguished by protein phenotype. Mitochondrial DNA from these subclones was isolated and purified and used in the clonings described below.

The mtDNA from each of the mutants was restricted with BglII and HindIII. This releases a 3.3 kb fragment which contains the coxI sequences from 4798 to 8072, including 337 bp of the 3' end of intron 2, all of exon 3, all of intron 3 all of exon 4 and 874 bp of intron 4 (see Figure 19). This fragment runs as a doublet on agarose gels with another fragment containing the 3' region of the Oxi2 gene. The fragment carrying the coxI gene can be identified by the BamHI site at position 6625. The 3.3 kb doublet from each mutant was band purified from agarose gels and ligated into the pUC18 vector cut with BamHI and HindIII. This plasmid is a pBR322 derivative carrying the origin of replication and ampicillin resistance gene from that plasmid, as well as the lacZ gene and MCS. It again should be noted that this cloning strategy eliminates the vector BamHI site. E. coli strain JM101 was transformed and plated on media containing ampicillin, IPTG and Xgal. White colonies indicating a fragment inserted into the MCS were picked and plasmid DNA isolated. This DNA was
restricted with BamHI and EcoRI and run on agarose gels. Clones containing the correct fragment should have two bands, one of 4.1 kb composed of the entire pUC18 vector and 1.4 kb from the 3' end of the insert DNA and another of 1.8 kb composed of the 5' end of the insert DNA and a short segment of the MCS. Colonies carrying plasmids containing each of the mutants based on the above criteria were selected and large scale plasmid preps performed. The plasmids were then purified on EB-CsCl gradients and used as the main source of insert DNA for all further subclonings required for sequencing and in vitro studies. These plasmids will be hereafter referred to as the p-series plasmids and are individually termed pM44, pC1072, and pC1085.

Sequencing of the mutants required that they be subcloned into M13 vectors. The subcloning of C1072 and C1085 for sequencing was handled in the same manner and will be discussed together. Since the sequencing of the mutants was to be done using custom-made DNA primers for the most part, the subclonings could be straightforward. Plasmids pC1072 and pC1085 were digested with EcoRI and HindIII to release the entire insert. This DNA was used in ligation reactions with the M13 based vectors Mpl8 and Mpl9. These plasmids are a modified form of the
replicative form of the male-specific bacteriophage M13. They, like pUC18 and Bluescribe, carry a portion of the lacZ gene with the MCS to allow for the selection of recombinants based on the loss of beta-galactosidase activity. They differ only in the orientation of the MCS and hence the strand of any DNA inserted in it that is packaged during the phage life cycle. The resultant DNA was used to transfect competent E. coli which were then plated with a lawn of host cells in top agar containing Xgal and IPTG. White plaques indicating a recombinant phage were picked and the single stranded phage DNA isolated. To identify those plaques carrying coxI sequences, single lane screening reactions were performed, as described in EXPERIMENTAL TECHNIQUES, using a primer specific for the coxI sequences. This primer, H1, is shown in Table 4 and is specific for the 3' end of intron 2. When this reaction is performed using specific primers, one may select recombinants carrying the insert of interest based solely on the incorporation of counts into bands on a sequencing lane. Based on this test, phage containing C1072 and C1085 in both orientations were selected. C1072 and C1085 cloned into M13mp19 were termed Mp1972 and Mp1985, respectively. Similarly, clones in M13mp18 were termed Mp1872 and Mp1885.
An additional set of clones using C1072 and C1085 sequences was made to allow sequencing of the region downstream of intron 3 without requiring the production of additional custom primers. BamHI and HindIII were used to release the 3' 1.4 kb of the insert, which was then ligated into Mp19. This piece contains the final 93 bp of intron 3 and all exon 4 and intron 4 sequences cloned into the p-series plasmids. Recombinant phage were screened by sequence analysis using the M13 primer and compared to the published sequence. These plasmids were called Mp19B72 (derived from p1072) and Mp19B85 (derived from p1085).

The M44 insert was cloned for sequencing using a different approach for several reasons. First, several attempts at cloning the EcoRI+HindIII piece from pM44 were unsuccessful, suggesting that it may for some reason be incompatible with either the vector or the host strain. Second, when an alternate digest, EcoRI+BamHI, was tried, it was found that pM44 carries a BamHI site so near the plasmid EcoRI site that the difference between the expected size of the EcoRI+BamHI fragment and the actual size of what was later shown, by BamHI digestion, to be the BamHI fragment containing the
5' region of the insert, could not be demonstrated on agarose gels. This restriction site was not investigated further. Based on these results, it was decided to clone M44 using the Bam fragment. Plasmid pM44 was digested with BamHI and mixed in a ligation reaction with similarly digested M13mp19. Recombinant phage were isolated and screened using the single lane screening reaction with the same coxI specific primer (HI) used above. In this case it should be noted that the use of the specific primer not only shows which of the recombinants have the insert of interest, but also which ones have the insert in the correct orientation to be sequenced. One recombinant which was positive by the single lane screen test was selected for sequence analysis and named Mpl944.

Plasmids for in vitro analysis of mutants were produced in the same way for M44, C1072, and C1085. Each of the p-series was digested with EcoRI and HindIII and ligated into the Bluescribe vector described in the wildtype cloning section. DNA from each ligation reaction was used to transform E. coli strain JM101. Recombinant plasmids were isolated and screened by restriction analysis. Plasmids were digested with EcoRI and HindIII and run on agarose gel. Clones were
selected that showed a doublet running at 3.2 kb, one band of which was the insert, the other the Bluescribe vector. These plasmids could be distinguished from reconstitutions of the original p-series plasmid by the vector band which is 500 bp smaller in the latter case. Recombinant Bluescribe plasmids carrying M44, C1072 and C1085 inserts were selected and termed BSM44, BSC1072 and BSC1085, respectively. The nomenclature for these constructs combines a letter designation denoting the vector (BS for Bluescribe) and the name of the insert DNA source. Each plasmid was used to transform E. coli strain GM161, a dam- strain. Plasmid DNA was isolated using the alkali lysis technique and purified on EB-CsCl gradients. A dam- strain was used to allow linearization at the BclI site within the fourth exon required for some \textit{in vitro} analysis. These plasmids were used in all \textit{in vitro} processing experiments discussed below. All constructions described in this section are summarized in Figure 20.

3. Summary

The genetic identity of all strains to be used in this study was verified by genetic mapping and, in the case of mutants M44, C1072, and C1085, protein labeling. Mitochondrial DNA was isolated from each and a region
containing aI3 cloned into a series of plasmids required for sequencing and in vitro splicing experiments.

B. Sequencing Data

1. Introduction

Sequencing of mutants M44, C1072 and C1085 was undertaken with two basic objectives in mind: 1) to pinpoint the lesions responsible for the mutant phenotype, and 2) to define any strain polymorphisms between our laboratory strain, ID41-6/161, and that used by Tzagoloff's group, D273-10B/A1, in the initial sequencing studies of the coxl gene. The data obtained from the accomplishment of each objective tend to shed light on the structure and function of group I introns in different ways. Loss of function mutations, such as the ones in the present study, disclose changes which are nonpermissible in terms of the continued function of the intron and, by extension, basepairs or regions crucial to that function. Strain polymorphisms reveal changes which are permissible in terms of the continued function of the intron; this knowledge may establish basepairs or regions which are not crucial to that function. Additionally, by highlighting conserved basepairs and relationships, one may infer important
features of related structures in evolutionarily separate strains.

The published sequence of the coxI gene of strain D273 (Bonitz et al., 1980) greatly simplified this project, since it allowed the use of synthetic oligonucleotides as primers, reducing the subcloning required to a minimum. Table 4 lists the primers used in this study and the relevant D273 sequence. All primers were synthesized at Ohio State University in the laboratory of Dr. Richard Swenson. Primers were positioned 250-300bp apart. This is close enough to assure overlap, yet far enough to optimize the amount of data from each sequencing reaction with the technology employed. The primer nomenclature used designated each primer as H plus a numerical subscript based on its location in the sequence. Two primers were manufactured that proved to be not useful in sequencing. One, H0lf, is complementary to the GC-cluster, starting at 5025; as explained below, that cluster is missing in strain 161. A second primer, H6, is complementary to basepairs 6400 to 6417; this primer was inefficient, probably due to local secondary structure in the template. The bases 6386 to 6421 can form an 18 bp stem which includes all bases in the primer. This stem probably forms so
quickly and stably that the primer never binds. In general, however, the synthetic oligonucleotides proved to be an efficient and easily processed sequencing tool.

The general sequencing strategy for each mutant is shown in Figures 21 and 22. Mutant M44 was sequenced from 5017 to 6502, C1072 from 4809 to 6841 and C1085 from 4800 to 6852. The wild-type sequence of strain 161 was not sequenced directly, but was deduced from the following logic. If a base was found to deviate from the published sequence, it was compared with the same position in another mutant; if two or more mutants carried the same change, this was considered to be a strain polymorphism. This required that at least two mutants be sequenced through the entire region in which any of the mutants in the study could exist; this was done with C1072 and C1085. Since C1072 and C1085 are genetically distinct, any change common to both must be silent. A further control was provided by M44. That mutant was isolated seven years earlier than C1072 and C1085; any change common to all of the mutants, therefore, could not be the result of a sequential mutation in the same mutagenesis. A situation could have arisen where all the mutants sequenced in a particular basepair or region could have differed from
D273 and from each other. This unlikely circumstance never arose, but it could have been resolved by sequencing the wild-type clone in a restricted region.

2. Sequence of the aI3 Region of Strain ID41-6/161

The deduced sequence differences between D273 and 161 in the area between positions 4809 and 6841, the bases where sequence is available from at least two mutants, are shown in Table 5. Overall there are 58 basepair changes over the region, giving a 97% homology. The amino acid homology in the 1008 bp intron reading frame is 98%. Several of the changes are noteworthy. The deletion which begins at 5026 removes most of a GC-cluster, number 87 in the nomenclature of de Zamaroczy and Bernardi (1986). The bases left from the cluster, TACAT, are part of an element conserved throughout the family, a2, to which this cluster belongs. Furthermore, part of this sequence, ACT, is repeated at each end of this cluster and most of the clusters in this family. The a2 GC-cluster family is shown in Figure 23. Three polymorphisms completely change more than one contiguous codon; these are the alterations beginning at 5611, 5702, and 5767. These are rather large and surprising, since there is no single-step way in which the D273 sequence could be modified to give the 161 sequence.
This seems unusual in sequences which are otherwise nearly identical. These changes also alter all but one of the amino acids. One single basepair change deserves mention; this is the T to A transversion at 6307, which modifies the internal guide sequence and improves the basepairing of the internal guide with the 3' exon (see Figure 24 and DISCUSSION). This change is identical to that noted by Tabak et al. (1987) for strain KL14-4A. Other alterations are either single basepair transitions or transversions or small insertions. The most obvious effect of any of these is to change a single amino acid in the reading frame. None of the changes in the intron ORF changes the reading frame or produces a stop codon. None of the changes in the CRF (6212 to 6717) modifies any known cis-acting site except for the internal guide change already mentioned.

3. Sequence of Mutants M44, C1072, and C1085

Basepair alterations unique to each mutant are shown in Table 6. Mutant M44 contains only one deviation from the 161 deduced wildtype sequence; this is a G to T transversion at basepair 5202, the penultimate base in aE2. It exists well within the limits set by the petite mapping data for this mutant. This directly affects the pairing of that part of the 5'
exon with the internal guide; it also changes the amino acid encoded from a glycine to a valine.

Mutant C1072 has two sequence changes relative to the wildtype sequence, one at 5196 and one at 6620. The T to C transition at 6620 is well beyond the map limits of the mutant and therefore cannot be responsible for the observed phenotype. The G to A transition at 5196 then must be the responsible lesion, since it is the only basepair change in this mutant in the region indicated by the mapping data. This basepair, like the alteration in M44, is within exon 3 and changes an amino acid in the oxidase protein sequence. This change is too far upstream to affect internal guide binding, yet it has been shown to have its primary effect on intron 3 splicing (Sass, 1984). Since the same studies showed that the mutant is cis-acting, this mutant either alters an as yet undetected RNA secondary structure interaction or alters a protein (or RNA) binding site important in splicing.

The situation in C1085 is much more complex; there are nine modifications of the wild-type sequence, eight of which could be responsible for the observed phenotype, based on mapping data alone. The one change
noted which maps beyond the known limits of C1085 is the insertion of an A at position 6218, which is 5' of the Hinfl site at 6395, the 5' limit of C1085. There are four basepair alterations that interrupt known cis-acting sites; these are located at 6310, 6528, 6534 and 6643. The C to T transition at 6310 is at the very 5' edge of where C1085 could exist. It is, however, located in the internal guide sequence and alters the base that pairs with 5202, the base mutated in M44. Two changes are located in Q, those at 6528 and 6534. The first is a C to T transition; the second a G to T transversion. E' is altered by a G to T transversion at base 6643. All these mutations exist well within the map limits of C1085. While other sequence changes in C1085 could, in principle, have an effect on splicing, these four mutations are, due to their location, the prime candidates to be the cause of the mutant phenotype. Whether one of these is the sole or primary cause of the mutant phenotype or whether the phenotype is the result of some combination acting in concert cannot be ascertained from the current data.

4. Summary

Sequencing of M44, C1072, and C1085 was accomplished using strategically placed synthetic
primers. Several strain polymorphisms were revealed which distinguish strain 161 from D273. Furthermore, the effects of M44 and C1072 were shown to be the result of single basepair changes. In contrast, the sequence of C1085 disclosed a plethora of changes, several of which are in known cis-acting sites, making positive identification of a single responsible change difficult.

C. In Vitro Processing Analysis

1. Introduction

As discussed in INTRODUCTION, previous to the undertaking of this study, van der Horst and Tabak (1985) had shown that if total \textit{in vivo} mtRNA were incubated in a typical group I self-splicing buffer with $^{32}$P labeled GTP, aI3 related transcripts were labeled. This G-addition reaction is indicative of group I self-splicing and suggested that aI3 self-splices \textit{in vitro}. Based on these data, it was decided to investigate this reaction with several questions in mind: 1) Does aI3 self-splice \textit{in vitro}? 2) What are the optimum conditions for this reaction? 3) What are the products generated by this reaction? and 4) What are the effects of the mutants M44, C1072, and C1085 on the \textit{in vitro} reaction? Two preliminary experiments were performed which convinced me that aI3 could undergo autocatalytic
reactions that required the group I cis-acting sites. These experiments are discussed below.

a. Initial Time Course

The result of the first test of aI3 self-processing performed is shown in Figure 25. In this reaction plasmid BS17-17 was linearized at the HindIII site and 32p labeled RNA was transcribed from the T7 promoter in vitro. As shown in Figure 26, this transcript covers the entire cloned sequence plus some vector sequences. This transcript was phenol extracted, ethanol precipitated and incubated in a self-processing buffer suggested by the Tabak experiments. This buffer contains 100mM (NH4)2SO4, 50mM Tris-HCl(pH 7.5), 60mM MgCl2, and 0.15mM GTP. Five reaction vials were set up and incubated at 36 degrees C. Reactions were stopped at 0, 5, 10, 20, and 30 minutes by the addition of 1/5 volume of 100mM EDTA and placing the reactions on ice. Samples were run on an acrylamide-urea gel as described in EXPERIMENTAL TECHNIQUES. It is apparent from this gel that over time several species accumulate over time in the absence of protein. These bands will be identified below as the products of aI3 autocatalytic activity. One band is key in this preliminary experiment in suggesting that this is a true self-
processing reaction and not the result of degradation of the transcript. This is the species which migrates more slowly than the full-length transcript. Bands which migrate anomalously in acrylamide gels have been seen in many RNA processing reactions (Bruce and Uhlenbeck, 1978; Sanger et al., 1979; Grabowski et al., 1981; Tabak et al., 1984; Peebles et al., 1986) and are identified as RNA lariats or circles. As discussed in INTRODUCTION, several excised group I introns have been shown to form circles. The accumulation of a topologically constrained molecule over time in the absence of protein strongly suggested that aI3 could undergo auto-catalysis.

b. BamHI Linearization

To test if the observed activity requires the typical group I cis-acting sites, plasmid BS17-17 was linearized at the unique BamHI site within aI3, and used as a template for RNA transcription by T7 polymerase. This has the effect of producing a runoff transcript that is deleted for several cis-acting sites important in splicing, R, E', S, and the 3' intron\exon boundary (see Figure 26). Aliquots of this RNA were incubated in processing buffer for 0, 15, and 30 minutes and the samples were then analyzed on a polyacrylamide gel. As
demonstrated in Figure 27, no reaction occurred. The reaction, therefore, requires that regions of the transcript containing sites known to be required for self-splicing of other group I introns be present.

c. Summary

Taken together, these data show that 1) a time-dependent reaction takes place, under conditions known to be suitable for the self-splicing of other group I introns, 2) the reaction requires no protein, and 3) regions containing the group I cis-acting sites are required. The most straightforward interpretation of these results is that the reaction is a genuine \textit{in vitro} self-processing reaction. Based on these results, further characterization of the reaction was undertaken.

2. Standardization of Reaction Conditions

a. Introduction

Initial studies of the self-splicing of the \textit{Tetrahymena} pre-rRNA (Cech et al., 1981) showed that three components were necessary for the reaction, a monovalent cation (\(\text{NH}_4^+\)), a divalent cation (\(\text{Mg}^{++}\)) and GTP. These requirements have proved universal for splicing of all self-processing group I introns. To further understand the aI3 self-processing reaction and
to improve product yield, a series of experiments was undertaken to standardize the reaction conditions. Starting with the conditions mentioned above, 100mM (NH₄)₂SO₄, 50mM Tris-HCl (pH 7.5), 60mM MgCl₂, and 0.15mM GTP at 36 degrees C, the concentration of both salts and GTP and the temperature were each varied, one at a time, while other conditions were held constant. The range of each test was selected to include known standard conditions for other self-processing reactions as well as upper and lower limits. The substrate for these experiments was ³²P labeled RNA produced in vitro (see EXPERIMENTAL TECHNIQUES) using T7 polymerase, with plasmid BS17-17 linearized at the HindIII site as the template (see Figure 26). Reactions were initiated by the addition of RNA to the reaction buffer, incubated for .5 to 1 hour, and stopped by the addition of 1/5 volume of a .5M EDTA stock solution. Results were analyzed by polyacrylamide-urea gel electrophoresis and autoradiography.

b. GTP Concentration Response

Added GTP concentrations of 0.00mM, 0.075mM, 0.15mM, 0.3mM, and 0.6mM were tested; the results are displayed in Figure 28. Even when no GTP is added, some reaction products are generated; however, consistent
with the results of Cech et al. (1981), full activity requires GTP in micromolar amounts. The low level activity at 0mM is, at first, puzzling; however, Cech et al. (1981) reported some activity with the Tetrahymena rRNA intron at GTP concentrations as low as 10nM. It is possible that some GTP is present in the reaction mix derived from degradation of RNA or as a contaminant from the transcription reaction. Alternatively, the intron may retain some activity without exogenous GTP. Inoue et al. (1984) have reported accurate splicing of Tetrahymena pre-rRNA in the absence of added GTP. That activity was enhanced at 50mM MgCl$_2$, similar to the 60mM MgCl$_2$ used in this experiment and, as demonstrated below, required for aI3 activity. Based on these results, GTP concentration was held at .15mM. It should be noted that the requirement for GTP is a hallmark of group I self-splicing reactions and that the above results, therefore, tend to further characterize this reaction as genuine group I self-processing activity.

c. (NH$_4$)$_2$SO$_4$ Concentration Response

Next, the involvement of a monovalent cation was tested. For this experiment full-length BS17-17 RNA was incubated in buffers containing 0mM, 25mM, 50mM, 100mM, and 200mM (NH$_4$)$_2$SO$_4$. As demonstrated in Figure 29, most
bands are produced at 0mM; some, particularly the slowly migrating band above the full length transcript and the fastest migrating band, are not produced. These results led to the standardization of (NH₄)₂SO₄ concentration at 100mM as the concentration leading to peak production of all bands and the concentration most consistent with the results of other researchers. This is the concentration used by Tabak et al. (1987). This concentration is also comparable to that used by Cech et al. (1981), who found that the maximum amount of total intron was released at between 100mM and 160mM (NH₄)₂SO₄.

d. MgCl₂ Concentration Response

MgCl₂ concentration was examined at 0mM, 10mM, 20mM, 60mM, and 100mM. Figure 30 shows the result of this experiment. Based on this result, this reaction has an absolute requirement for MgCl₂, since no reaction is observed at concentrations below 60mM. This is a much higher concentration than reported by Cech et al. (1981) for the *Tetrahymena* rRNA intron. In initial studies of self-splicing of the *Tetrahymena* rRNA intron, MgCl₂ concentration was optimized at 10mM. Higher concentrations, up to 50mM, were tested and found to be inhibitory. The current finding is consistent with the study of Garriga and Lambowitz (1984) who established
that the \textit{in vitro} self-splicing of the first intron of the \textit{Neurospora} cytochrome b gene requires a MgCl$_2$ concentration greater than 30mM. A similar optimum (50mM) has also been noted for intron bI5 by both Gampel and Tzagoloff (1987) and Partono and Lewin (1988). At 100mM MgCl$_2$, some nonspecific degradation is noted, especially in the higher molecular weight bands. Based on these results, the MgCl$_2$ concentration was standardized at 60mM.

\textbf{e. Temperature Response}

Equal amounts of HindIII linearized BS17-17 transcript were incubated at 22, 30, 36, and 42 degrees C. The result, shown in Figure 31, indicates a clear temperature dependence of the reaction. All product bands are clearly enhanced at higher temperatures. Some products appear at the lower temperatures, while others do not; this is probably the result of slower kinetics of the overall reaction rather than the result of differential production. Cech's laboratory has reported using between 30 and 42 degrees C. in self-splicing reactions using the \textit{Tetrahymena} intron (Cech et al., 1981; Been and Cech, 1987). Other researchers routinely use 37 to 60 degrees for group I self-splicing
reactions (Garriga and Lambowitz, 1984; Tabak et al., 1987).

f. Summary

Based on the data presented above, the conditions for all future reactions were set at 100mM (NH₄)₂SO₄, 50mM Tris-HCl (pH 7.5), 60mM MgCl₂, and 0.15mM GTP at 42 degrees C. These conditions are, in general, consistent with those used for other self-splicing reactions, and, except for temperature, are the same as the initial conditions. The enhanced activity at higher temperatures may reflect a need to melt out competing, nonproductive secondary structures. The requirement for 60mM Mg⁺⁺ distinguishes this reaction from that of the Tetrahymena pre-rRNA; however, other group I introns, notably the Neurospora cytochrome b intron 1 and yeast bI5, have similar requirements. The aI3 reaction, therefore, seems to have the requirements of a typical group I intron for self-splicing.

3. Product Characterization

a. Introduction

During the course of this study, Tabak et al. (1987) published a paper describing the in vitro reactions of the aI3 intron. That publication is
discussed in full in the INTRODUCTION. In brief, those researchers found that aI3 contains two sites that undergo nucleophilic attack by GTP. The first is at the true intron/exon boundary. The second is at an internal site in the closed reading frame, at position 6286; that region has the potential to basepair with the internal guide sequence and, so, could mimic the actual intron/exon boundary. Furthermore, Tabak et al. (1987) established that the excised intron sequences could produce a complex series of circular and linear products. These products are: 1) the complete intron (1514 nt) in both linear and circular form; 2) the 5' product of the internal guanosine attack, a 1083 nt piece containing the entire intron reading frame plus 75 nt of the CRF, in both circular and linear forms; 3) the 3' intron product of the internal guanosine attack, a 431 nt piece that includes all of the known cis-acting sites, again in both circular and linear forms; and 4) a molecule formed from the 1083 nt circle and the 431 nt circle interlocked. Those researchers only considered that the guanosine attacks would take place in an ordered fashion, that is that the internal guanosine attack would only take place on the released full-length intron. If, however, one assumes that the alternative guanosine attack can also take place on the full-
length transcript, some additional products may be envisioned. These are the 5’ exon plus the 1083 nt of the intron 5’ of the secondary guanosine addition site and the 3’ exon plus the 431 nt of intron 3’ of the secondary guanosine addition site. The order of potential products on a gel appears below with these additional products inserted. Additional bands are marked *.

i. interlocked circles (1083 nt + 431 nt)

ii. full intron circle (1514 nt)

iii. ORF circle (1083 nt)

iv. full-length transcript (2267 nt)

v. intron + 5’ exon (1915 nt)

vi. intron + 3’ exon (1866 nt)

vii. linear intron (1514 nt)

*viii. 5’ exon + ORF (1484 nt)

ix. linear ORF (1083 nt)

*x. CRF + 3’ exon (784 nt)

xi. ligated exons (755 nt)

xii. CRF circle (431 nt)

xiii. linear CRF (431 nt)

xiv. 5’ exon (402 nt)

xv. 3’ exon (353 nt)
These products are diagrammed in Figure 32. Several experimental approaches were taken to identify the products of the BS17-17 transcript in light of those products identified by Tabak et al. (1987). Results of these experiments are detailed below.

b. Time Course and Gel Mobilities

In order to determine gel mobilities and to observe any time dependent precursor product relationships, two time course experiments were performed. Both experiments used $^{32}$P labeled full-length transcript, which was transcribed from plasmid BS17-17 linearized with HindIII as a substrate. Samples were incubated under standard conditions as described above. In the first experiment, equal aliquots of the reaction mix were removed at time points of 0, 10, 20, 40, 60, and 80 minutes. In the second experiment, equal aliquots were removed at 0, 1, 2, 4, 6, 8, and 10 minutes. Aliquots, in each case, were mixed with $1/5$ volume of .5M EDTA to stop the reaction. The samples were analyzed by urea-polyacrylamide gel electrophoresis and autoradiography. The gels employed were longer than used previously to improve band separation, especially higher molecular weight species and bands of similar size. These gels are displayed in Figures 33 and 34. The molecular
weight estimates of each band or doublet is also given, derived from the compilation of data from several gels.

The first time course (Figure 33) shows the accumulation of a complex pattern of products. At time 0 there are three bands pre-existing. One band with a mobility of 2300 nt is the full-length transcript. Another band appears just below the full-length transcript with a mobility of approximately 1800 to 2000 nt. At the bottom of the gel a third band with a mobility corresponding to 350 nt also pre-exists. The two additional bands are products of the full-length transcript based on several findings. First, they increase in intensity over time. Second, they are observed in reactions where the full-length transcript has been band isolated prior to incubation. Third, as will be demonstrated below, they hybridize to transcript specific probes.

Over time several additional bands are produced. As discussed previously, a band which migrates more slowly than the full-length transcript is produced over time. This suggests that it possesses a topological constraint that impedes migration through the gel system. Below the full-length transcript a doublet with
a mobility of 1800 to 2000 nt accumulates. This doublet includes the pre-existing band plus a slightly faster migrating band. A second doublet is produced below the first which migrates at 1500 nt. There is a band which accumulates slowly at a gel position corresponding to 1000 nt. Next, there is a doublet migrating at 750 to 800 nt. Below this doublet three single bands accumulate with mobilities corresponding to 550, 450 and 350 nt.

The second time course (Figure 34) shows the accumulation of the same set of products described above, except for the band at 1000 nt which accumulates too slowly to be seen in this time frame. The rest of the bands seem to appear almost immediately in the reaction except for the larger band in the 750 to 800 nt doublet which is not apparent until the third time point (four minutes).

c. PvuII Linearization of BS17-17 to Identify 3' Related Species

Plasmid BS17-17 contains two PvuII sites near the ai3 insert. When this plasmid is digested with PvuII and RNA transcribed by T7 polymerase, a transcript is produced which includes the entire cloned sequence plus
213 nt of the bluescribe vector at the 3' end. Therefore, any processing product which contains the 3' end of the transcript will be lengthened by this amount. This is very useful, especially in the analysis of lower molecular weight products. This experiment has been performed with both labeled and unlabeled substrate. In either case RNA was transcribed from BS17-17 linearized with HindIII or PvuII and incubated under standard conditions. Products were separated on urea-polyacrylamide gels and visualized by autoradiography. The results are shown in Figure 35.

When the product pattern is compared between transcripts produced from BS17-17 linearized at HindIII or PvuII, two bands appear to be affected. The first is the band at 750 to 800 nt. As discussed above, this band is a doublet on gels where the products are resolved more fully. In the reaction using PvuII derived transcript this band does not appear, but is replaced by a band with a mobility of 950 to 1000 nt. It is important to note that, since the band is not present, both bands in the doublet are affected. The second band which is clearly affected by the PvuII linearization of the template is the band at 350 nt. This band does not appear in the reaction using PvuII
derived RNA but is replaced by a band with a mobility of approximately 600 nt. The species which have mobilities of 1000, 550, and 450 nt are not altered by the PvuII digest. The larger species in the reaction are not resolved well enough on this gel system to determine if their mobility is altered in the PvuII digest reaction.

d. Northern Blots of the In Vitro Reaction Products

To further characterize the auto-catalytic products of aI3 two northern blot experiments were performed. Unlabeled transcript from BS17-17 linearized with HindIII was processed under standard conditions. The products were separated on urea-polyacrylamide gels, then electroblotted to Zeta-probe. Blots were hybridized with two probes, one intron-specific, the other 3' exon-specific. The intron-specific probe was nick-translated plasmid pMR-31 provided by Kirk Mecklenburg. This plasmid contains a 450 bp segment of aI3 from the MboI site at 6174 to the BamHI site at 6624 cloned into the BamHI site of pBR322. The 3' exon probe was produced by cutting BS17-17 with Sau96I and using the resultant DNA as a template for T3 polymerase. When this is done, a 181 nt antisense transcript is made that extends from the T3 promoter to the Sau96I site in the 3' exon at base 6898. The regions covered by these
probes are diagrammed in Figure 36. The results of the blots are displayed in Figures 37A and 37B. The 3′ exon probe hybridizes strongly to four species: 1) the full-length transcript, 2) a species at 1800 to 2000 nt, 3) a species at 750 to 800 nt, and 4) the species at 350 nt. The intron probe hybridizes to: 1) the full-length transcript, 2) three bands which migrate more slowly than the full-length transcript, 3) a band at 1800 to 2000 nt, 4) a band at 1500 nt, 5) a band at 1000 nt, 6) a band at 550 nt and 7) a band at 450 nt. The two most slowly migrating bands which hybridize to the intron probe are generally not visible on normally exposed processing gels. They are probably visible in this experiment due to the sensitivity of the northern blot assay and because much higher amounts of RNA were used on gels intended for northern blots than in other experiments.

e. Product Identification

Using the data generated by the experiments outlined above and the data of Tabak et al. (1987), one may deduce the identity of the bands produced by the self-processing of aI3. Most of the products discussed in INTRODUCTION are accounted for. Figure 33 shows a time course experiment with the product identities
assigned. Those products which do not appear are shown in the position where they would run relative to other bands. Their existence is discussed in light of other evidence. In this section the rationale for each assignment will be discussed in the order of appearance on the gel.

i. **Interlocked circles (1083 nt + 431 nt).** This band is not visible on this gel; however, on the northern blot probed with the pure intron probe a slow moving band appears near the top of the gel. This band does not respond when probed with exon probe (see Figures 37A and 37B). Based on gel mobility and northern blot results, this species is present in the reaction products, at least in those reactions undertaken prior to northern blotting. It should be noted that processing reactions for northern blots were performed at much higher RNA concentrations than other reactions.

ii. **Full-length intron circle (1514 nt).** Again, this species does not appear on this exposure but does appear in the intron northern blot as the intermediate band of the three which migrate more slowly than full-length. This band is not found in the northern probed
with the 3' probe. Also, Figure 38 shows an overexposure of a short time course gel with this band appearing at low concentrations.

iii. **ORF circle (1083 nt).** This band is the only one of the three bands which migrate more slowly than full-length which is present in sufficient quantity to be visible on normal exposures. It hybridizes to the intron probe, but not to the 3' exon probe. Based upon this and its observed mobility relative to the other slow moving bands and the full-length transcript, this band most probably represents the ORF circle.

iv. **Full-length transcript (2267 nt).** This band has an observed mobility of approximately 2300 nt. This is the largest band produced by the transcription reaction.

v. **Intron + 5' exon (1915 nt); intron + 3' exon (1866 nt) doublet.** This doublet has an observed mobility of 1800 to 2000 nt. Due to the large and similar size of the bands, this doublet is poorly resolved on this gel; however, close examination shows that two bands are present. Both bands hybridize with the intron probe. One of the bands in the doublet.
hybridizes to the 3' exon probe. Only one product of this size range could have this property, the intron + 3' exon 2/3 molecule. The existence of the alternate 2/3 molecule is confirmed by other criteria. First, there is no other product described by Tabak et al. that has this mobility other than the previously identified intron + 3' exon molecule. Second, in the 0 minute time point three bands are present: the full-length molecule, the larger band of this doublet and a low molecular weight band which will be identified below as the 3' exon. Some 3' exon RNA is released during the transcription reaction. This release must also generate the 5' exon + intron 2/3 molecule. The presence of the band in question in concert with the appearance of the 3' exon confirms this identification.

vi. Linear excised intron (1514 nt): 5' exon + ORF (1484 nt) doublet. The mobility of this band is 1500 nt. It hybridizes with the pure intron probe but not with the 3' exon probe. Therefore, this band most probably represents the linear excised intron. Additionally, in some gels another band which has almost the same size can be seen. This can best be seen on the short time course experiment (Figure 34). The only other molecule in this system that could have a mobility
of approximately 1500 nt is the upstream product of the guanosine attack at the internal site in the 5' exon + ORF molecule; that RNA is 1484 nt long. This species may be poorly resolved on other gels or simply present in very low amounts. The band at this region on the long time course gel appears to be a doublet that is insufficiently separated.

vii. Linear excised ORF (1083 nt). This band appears more slowly than most others and, therefore, may be a secondary product. It hybridizes with the intron probe, but not with the 3' exon probe. Further, it does not change position in the PvuII linearization experiment. The calculated mobility of this band is 1000 nt. These data correlate with the results expected for the linear intron ORF based on Tabak's results. The lag in the appearance of this band could have two explanations; first, it may be a product of the breakage of the circular ORF with the circle being a required intermediate. Alternatively, the formation of the circle may be favored over the linear ORF under these reaction conditions so that, because a higher percentage of the ORF produced at any time is circular, it becomes visible first, mimicking a precursor-product relationship.
viii. CRF + 3' exon (784 nt) : ligated exons (755 nt). This is another doublet of bands having a size of 750 to 800 nt. In some experiments, only one product band is seen at this position. This is true of the RNA blot experiments (see Figure 37). In that case, only one band appeared, did not hybridize to the intron probe but did hybridize to the 3' exon probe. This product is also seen in the experiment comparing the products of RNA derived from HindIII and PvuII linearized template (see Figure 35). In this experiment, the band at 750 nt produced with RNA derived from template linearized at the HindIII site is replaced by a product of 950 nt in the reaction with RNA derived from PvuII linearized template. These data suggest that this band is the ligated exons. The size of that product should be 755 nt in the HindIII linearized construct and 968 nt in the PvuII linearized construct.

In other reactions a doublet is clearly seen at this position (see Figure 34). The logical candidate to be the second molecule in the doublet is the downstream product of the internal guanosine addition, the CRF + 3' intron molecule (784 nt). This means that this molecule would be the upper band in the doublet. It is not the
band seen in the northern blot, since it does not hybridize to the intron probe. It should be noted that this molecule is an intermediate capable of further processing into circular or linear CRF plus 3' exon. The efficient conversion to these products may explain the absence of this band from some reactions.

ix. **CRF circle (431 nt).** Northern blot analysis shows that this band contains intron sequences, but not 3' exon sequences. This band is not affected by linearizing the template at the PvuII site. The observed mobility is 550 nt. Tabak et al. (1987) have identified a molecule with similar properties in their system as a circular form of the 431 nt CRF portion of a13. No linear molecule identified for this system is 550 nt. As noted previously, retarded mobility on PAGE gels is a characteristic of circular molecules. The gel position of this band is thus consistent with the interpretation that this band is the circular CRF.

x. **Linear CRF (431 nt).** This gel position corresponds to approximately 450 nt. Two molecules envisioned for this system could run at this position, the linear intron CRF (431 nt) or the 5' exon (402 nt). At this position in the gel these two bands should be
well resolved. This band hybridizes to the intron probe and, therefore, must be the linear intron CRF. This suggests that in the processing reaction of this construct the ligation of the 5' exon to the 3' exon is very efficient allowing no free 5' exon. A doublet is occasionally seen at this position with a second band running below the first; this is clearly seen in the mutant C1072 time course experiment described below. This additional band most probably corresponds to the free 5' exon.

iii. 3' exon (353 nt). This band displays all the characteristics expected of the 3' exon. It has a mobility of 350 nt, hybridizes to the 3' exon probe, and is replaced by a band of approximately 600 nt when the reaction is performed with PvuII linearized BS17-17. This molecule can be released without any reaction at the 5' end of the intron. This conclusion is supported by the observation of the 3' molecule in low levels of the 0 minute time point, before any products related to the splicing pathway are present. Also, the presence of the 5' exon + intron molecule in this system and in that of Tabak et al. requires that the independent release of the 3' exon take place. This reaction accounts for the appearance of the free 3' exon but the lack of the free
5' exon. As will be discussed fully in the DISCUSSION these products could either be the result of a G attack or hydrolysis at the 3' intron/exon boundary. Guanosine attack at the 3' intron/exon boundary has previously been described for aI3 under self-splicing by van der Horst and Tabak (1987). Hydrolysis at the 3' intron/exon boundary has been described for the *Tetrahymena rRNA* intron by Inoue et al. (1986).

f. Summary

Three experimental approaches, northern blots, gel mobilities, and linearization at the *PvuII* site, have been used to correlate the products seen with the processing of BS17-17 with the products identified by Tabak et al. Using these data, it is possible to distinguish all these products except for the 5' exon which is not a product of this construct, because it is ligated with high efficiency to the 3' exon. Additionally, two other products appear to be present in some experiments with the BS17-17 transcript not discussed by these researchers. These are the products derived from a guanosine addition attack at the internal site of the intron on the full-length transcript, the 5' exon + ORF and the ORF + 3' exon. The internal guanosine attack on the full-length
molecule was not documented by Tabak et al., but that reaction was shown to occur on the released intron. This demonstrates that all the structural information exists within aI3 for this reaction to occur, and the reaction is, therefore, suggested as possible by that publication. The pathways which lead to the products identified in this section will be described fully in DISCUSSION.

4. Effect of Mutations on In Vitro Reactions

a. Introduction

To further understand mutants M44, C1072, and C1085 and the in vitro processing reaction of aI3, the consequence of each mutant on the reaction was tested. Plasmids BS44, BS1072, and BS1085 were linearized at the BclI site within exon four. Labeled run-off transcripts were made in vitro utilizing T7 RNA polymerase. This results in a correctly oriented full-length transcript of 2238 nt. This is 29 nt shorter on the 3' end than the BS17-17 full-length transcript, but it contains all the mitochondrial sequences present in that transcript and lacks only MCS sequences. BS44 and BS1072 transcripts were incubated under standard conditions. Equal aliquots were removed at time points of 0, 10, 20, 40, 60, and 80 minutes and the reactions stopped by the
addition of 1/5 volume of .5M EDTA. BS1085 transcripts were treated in the same manner except that incubation times of 0, 10, 20, 40, and 60 minutes were analyzed. Results were analyzed by urea-polyacrylamide electrophoresis and autoradiography. The results of each reaction are shown in Figures 39, 40 and 41. Each result is considered separately.

b. The Effect of the C1085 Mutation

The result of the BS1085 transcript incubated in vitro is shown in Figure 39. Although this autoradiograph is overexposed, no time dependent species are observable. None of the bands in this gel, with the exception of the full-length transcript, co-migrates with known self-processing products. These data show that the basepair changes in C1085 completely inhibit all autocatalytic activities, under these conditions.

c. The Effect of the C1072 Mutation

In sharp contrast to C1085, the results of the C1072 self-processing experiment show that all the bands produced by the wild-type are present when one takes into account the 29 nt difference at the 3' exon. Figure 40 shows these results.
The alteration of gel migration of some species due to the 29 nt difference allows independent verification of some of the product identification presented in the previous section. There are three bands which migrate faster when compared to the BS17-17 control reaction. The first two may be considered together. These are the two bands that run as the doublet at 750 nt. These two bands were interpreted above as being the CRF + 3' exon and the ligated exons. In wild-type these molecules are 784 nt and 755 nt long, respectively, while in the mutant experiment they are 755 nt and 726 nt, respectively. When the C1072 and wild-type results are compared, one finds that one band in the C1072 doublet migrates with the lower band of the 17-17 doublet, the ligated exon band at 755 nt. Another band is present just below this one, with a distance separating the two equal to that separating the bands in the 17-17 control lane.

The most likely explanation for this is that both molecules have become smaller by 29 nt, consistent with each containing the 3' exon. Therefore, the upper band in the doublet is still the CRF + 3' exon, which in the mutant construct migrates exactly with the ligated exons in the wild-type. The lower band is, of course, the
ligated exons, with a length of 726 nt in the mutant. The third band with altered mobility in the mutant samples is the smallest species, which was determined above to be the 3' exon; again, this is consistent with the current data. There is an additional band present in this gel not seen in the wild-type time course, which runs just below the band at 430 nt, a band identified above as the linear intron CRF. It appears slowly over time, suggesting that it is either a secondary product or that it represents a product that is formed only in a small percentage of reactions. The size of the new band is approximately 400 nt. This is exactly the size expected for the 5' exon, a product not seen in the processed BS17-17 transcript. Given the known position of the C1072 lesion near the 5' intron/exon boundary, it is not surprising that this mutation should affect the reaction at this point. The most likely identification of this band is that it is the free 5' exon.

d. The Effect of the M44 Mutation

As demonstrated in Figure 41, in vitro processing of the BS44 transcript results in a third pattern of products where a large subset of the wild-type products is present. Initially, the most striking feature of this reaction is the complete lack of the topologically
constrained species running above the full-length transcript, the complete intron circle and the ORF circle. To further verify this fact, an overexposure of this region of the gel is provided (Figure 42). At the 1800 to 2000 nt region of the gel, the wild-type doublet is replaced with a single band. This is also true of the doublet at 1500 nt. The 1083 nt linear intron ORF is missing entirely. The next band on the gel co-migrates exactly with the 755 nt ligated exon band on the wild-type control. As discussed above, this is the mobility expected for the CRF + 3' exon piece in the mutant experiments. This band decreases in intensity over time, which is consistent with its undergoing further reactions. The intron CRF circle is present in the pattern and appears over time in a manner coordinated with the disappearance of the CRF + 3' exon molecule. Both the linear intron CRF and the 3' exon are present in this pattern. A proposed mechanism of the autocatalytic activity of the BS44 transcript will be presented in DISCUSSION.

e. Summary

This section has presented the effects of the M44, C1072, and C1085 mutations on the in vitro autocatalytic activity of a13. Each mutant appears to have a unique
consequence. C1085 completely inhibits all autocatalytic activity; C1072 has only a minimal effect allowing the production of all bands seen in the BS17-17 self-processing with the important addition of the 5' exon band; and M44 produces an intriguing subset of the wild-type bands. Using these data, the sequence of each mutant, and the 161 wild-type sequence, the DISCUSSION section will focus on integrating these results with the general body of knowledge on the coxl sequence, group I intron splicing, and aI3, in particular, to present a picture of the normal splicing pathway and the effect of each mutant on it.
IV. DISCUSSION

A. Introduction

This study has investigated the processing of yeast mitochondrial intron aI3 by DNA sequence analysis and in vitro processing analysis of both the wildtype intron and three in vivo mutants. As noted in INTRODUCTION, aI3 is a group I intron with an unusual configuration which made it particularly worthy of study. The wildtype sequence of aI3 from strain 161 was determined and found to contain some differences from the published sequence (Bonitz et al., 1980) of strain D273. These polymorphisms have been noted in Table 5. When in vitro transcribed RNA from the wild-type aI3 was incubated under conditions known to promote self-splicing of other group I introns, a set of products was obtained. These products are, in general, the same as those reported by Tabak et al. (1987) in a paper describing the autocatalytic reactions of this intron. There are important differences, however, that will be discussed below. The sequence of each of the mutants was determined for those regions where petite mapping had
determined that the lesion could exist. Mutants M44 and C1072 were each found to differ from the wild-type sequence by a single basepair change in the region where these mutants map. In each case, this change must be responsible for the mutant phenotype. Mutant C1085 was found to harbor several changes, some of which were located in the known group I cis-acting sites. The changes found for the mutants have been summarized in Table 6. Finally, the effect of the mutants on the autocatalytic reactions of aI3 was investigated. Each mutant was found to affect the reaction in a different manner. Mutant C1085 completely blocked autocatalysis. Mutant C1072 produced all the autocatalytic products noted for the wild-type reaction plus one additional species. Mutant M44 produced a subset of the species which accumulate in the wild-type reaction.

In this section, these data will be analyzed in order to shed some light on basic questions concerning aI3 and the mutants which affect its processing. These questions include: 1) What is the probable secondary structure of the intron in the region of the cis-acting sites? 2) How do the differences between strain D273 and strain 161 affect this structure? 3) Does this intron employ a pairing of the IGS and the 5' exon to
align the 5' exon for splicing? 4) What are the reaction pathways which could be responsible for the complex pattern of autocatalytic products formed by this intron in vitro? 5) How do the basepair changes found in each mutant strain cause the mutant phenotype? 6) How are the autocatalytic pathways of this intron altered by each mutation? 7) How does the in vitro phenotype of each mutant relate to the in vivo phenotype?

B. Analysis of aI3 in Strain 161

1. Secondary Structure of aI3

Using the sequence of aI3 from strain 161, I have put the intron into the standard group I secondary structure. This structure is shown in Figure 43. The P1, P3 and P7 stems were taken from Davies et al. (1982). The P4 and P5 stem regions are after Collins (1988). Polymorphisms found in this study were incorporated in both cases; the effect of these polymorphisms will be discussed below. All other stems were constructed using PCFOLD:version 3.0, an RNA secondary structure program by Zucker and Stiegler (1981) that has been modified to run on IBM compatible computers.
2. Sequence Analysis

Strain 161 lacks the GC cluster found in strain D273 at bases 5024 to 5058. Other GC clusters have been found to be polymorphic. Three optional GC clusters are found in the varl region. One of these occurs within the varl ORF and, when present, is translated as part of the active protein. The other two GC clusters occur outside the ORF. Optional GC clusters are also found near the olil gene, within the 15S and 21S rRNA genes and bI5 (Butow et al., 1985). Inspection of the sequence at the cluster in D273 and 161 reveals that the cluster was inserted with the duplication of the sequence ACT (see Figure 10). This GC cluster has been grouped in a family of related clusters (the a2 family) by de Zamaroczy and Bernardi (1986) that contains 26 clusters from various yeast strains. Twenty-one of these clusters contain the element ACTAT at their 3' end. Fourteen of these clusters contain the element TACT or a closely related element at their 5' end. Three more contain this element within their sequence at some distance from the 5' end. Thus, it appears that these elements may insert at a defined sequence related to ACT and that this insertion takes place with the duplication of the ACT.
A similar mechanism has been noted for other
optional GC clusters (Butow et al., 1985). A 46 bp GC
cluster is optionally present at three locations in the
genome; that is, they are not found in all yeast
strains. These sites are: the "common cluster" 5' end
of the var1 reading frame, which is found in all
standard laboratory strains; a site within the var1 ORF;
and a site 3' to the olil gene. In each case it is
flanked by the same sequence. Comparison of the
flanking sequences in strains with and without these
clusters suggests that insertion may occur with the
duplication of two basepairs in the putative target
sequence. A transposition-like event may be responsible
for the insertion of some GC clusters; however, other
evidence suggests that asymmetrical gene conversion may
also play a role in the spread of GC clusters. In
crosses of two mitochondrial genomes which do not carry
the optional GC cluster in the var1 coding sequence, no
progeny are produced which have this cluster at this
site, even though the cluster is found in other sites in
the genome. However, in crosses where one strain
carries the cluster in the var1 gene, conversion of the
var1 allele without the cluster takes place. If
transposition of GC clusters does take place, it may be
an extremely rare event or may require specific conditions to trigger the transposition pathway.

Several functions have been suggested for GC clusters including acting as ori sequences, as sites of mRNA processing, and as structural elements in the genome (de Zamaroczy and Bernardi, 1986). The existence of optional GC clusters precludes an obligatory function for those optional elements and suggests that GC clusters may be examples of what has been called "selfish DNA" (Dawkins, 1976); it further suggests that any function they currently have may be secondary and fortuitous.

The reading frame of aI3 from strain 161 contains several polymorphisms when compared to that of strain D273 (see Table 5). Overall, the reading frame is 98% identical to that of strain D273. This is actually low homology when compared to other studies. Anziano et al. (1982) found >99% homology when they compared the reading frame of bI4 in the same strains. Each of the changes alters the amino acid encoded by the relevant codon; however, none of the changes leads to a stop codon. This suggests that this reading frame may have a function which is selected for. Since there are no
known trans-recessive mutants in aI3, no function can be assigned to this protein. Splicing mutants in aI1 and aI2 which prevent the translation of the aI3 protein from the normal coxl AUG still process aI3, suggesting that a protein encoded by aI3 is not absolutely required for the splicing of that intron.

This does not eliminate the possibility that the protein is translated from an internal translation start site. Nor does this eliminate the possibility that the aI3 protein facilitates splicing in vivo, but that the self-splicing activity of the intron can compensate in the absence of a maturase. This could explain why no trans-recessive mutants have ever been found in aI3; they may not have a readily assayable mutant phenotype. Alternatively, the reading frame may be a recent acquisition of this intron which has not yet had time to degenerate. This view is supported by the structure of the intron itself. While many reading frames of group I introns contain most of the cis-acting sites, suggesting an intimate and long standing relationship with the intron, the aI3 ORF is completely separate from the cis-acting sites, thus allowing the possibility that the cis-acting sites and the reading frame are associated
through evolutionary accident rather than through any functional necessity.

The polymorphism in the IGS (at position 6307) leads to improved stability of the P1 stem. As discussed in INTRODUCTION, the published sequence of strain D273 predicts poor binding at the P1 stem leading Davies et al. (1982) to question the existence of this interaction. The altered sequence improves the interaction and makes it more likely. Tabak et al. (1987) have noted the same sequence change in their studies of strain KL14-4A. It is possible that the sequence originally reported for D273 was in error and that the correct sequence for the aI3 IGS is the one reported in this study.

Four polymorphisms are located in the region between the elements P and Q. Collins (1988) has studied the secondary structure of this region of aI3 and other introns which have relatively long distances between P and Q. In his study he proposed a secondary structure for this region of aI3 based on the D273 sequence. Figure 44 shows the secondary structure of this region in strain 161 using the Collins structure as a guide. The sequence changes in D273 which are
located in this region are also shown, along with the original form of the Collins structure at the point where they are located. Two of the polymorphisms noted in 161 clearly improve regions of this structure. The first is the insertion of a C between bases 6412 and 6413. This C can pair with a G which is bulged in the Collins model. Using the calculation method of Salser (1979), this improves the stability of the stem by -2.8 kcal/mol. The second is the insertion of three G's between bases 6502 and 6503, increasing the run of G residues from 9 to 12 at this point. This insertion stabilizes a stem by removing three bulged U's which can now basepair with G's in this stem. This change improves the stability of the stem by -2.3 kcal/mol.

One change is neutral; this is the T to A change at base 6422. In either case, the base remains bulged. The A insertion between 6426 and 6427 destabilizes the structure by 2.8 kcal/mol. by creating a new one base bulge.

The above discussion has supposed that the sequence differences noted between strain D273 and strain 161 are real polymorphisms. An alternative explanation exists, however. Some or all of the differences could be due to sequencing errors. I believe that the
sequence I have proposed for strain 161 to be correct. Each basepair was read at least twice from separate reactions. Most of the basepairs were read from many more reactions than that. Also, I had the advantage of using synthetic oligonucleotides to prime my sequencing reactions. This means that I could place my oligos so that no sequences would have to be read over 300 basepairs from any primer where the resolving power of sequencing gels begins to fall off. Bonitz et al. (1980) did not have this luxury. There are few good restriction enzyme sites for subcloning in al3 and that group may have had to stretch their reading of some sequencing experiments, resulting in some errors. As mentioned above, and in INTRODUCTION, this is almost certainly the case for the difference in the IGS. Which, if any, of the other strain differences are the result of sequencing errors would require a careful resequencing of al3 from strain D273.

3. Analysis of the In Vitro Processing Reaction of al3

The conditions for the in vitro splicing of al3 as optimized in this study are not unusual for group I introns. The requirement for 60mM Mg\textsuperscript{++} is much higher than the 10mM required by the Tetrahymena rRNA intron but is the same as for the Neurospora cob intron 1. It
has been suggested that Mg$^{++}$ is required for self-splicing to stabilize the secondary structure of the intron (Chu et al., 1987). The unusually large distance between the 5' intron/exon boundary and the IGS in aI3 may destabilize the active structure of the intron and may be the reason for the high Mg$^{++}$ requirement. The 42$^\circ$ C. optimum temperature found for this intron is higher than that for all group I self-processing introns except the T4 group I intron (Chu et al., 1987). That intron reaches optimum reactivity at 60$^\circ$ C. That temperature was not tested for aI3. The effect of temperature on the splicing of aI3 suggests that at lower temperatures the intron can exist in alternate non-productive structures. Alternatively, lower temperatures may inhibit a secondary structure transition required for splicing. The other standard conditions deduced for aI3 are in line with those published for other self-splicing group I introns. All the conditions used except for the temperature are the same as those used by Tabak et al. (1987) as "average" group I processing conditions.

The complex pattern of products produced by in vitro processing of aI3 is probably the result of three reaction pathways: the normal self-splicing pathway
initiated by the G attack at the 5' intron/exon boundary, the pathway initiated by the G attack at the internal site, and the breakage of the full-length molecule at the 3' intron/exon boundary. These pathways are diagrammed in Figure 45.

The G attack at the 5' intron/exon boundary (Figure 45A) releases the 5' exon from the transcript. The 5' exon can then attack at the 3' intron/exon boundary, resulting in the ligation of the exons and the release of the linear intron. The ligation of the exons is very efficient. No free 5' exon is found among the wildtype reaction products. This suggests that all the 5' exon released by this G attack is ligated to a 3' exon. The linear intron is the precursor to the three circular forms, the full length intron circle, the ORF circle, and the CRF circle, as well as the linear ORF and CRF forms. This molecule has also been shown by Tabak et al. (1987) to be the precursor to the interlocked circles.

The second pathway (Figure 45B) begins with a G attack at an internal site of the intron located at base 6287. The products identified in the wild-type pattern suggest that this attack can occur on either the full-
length molecule or the excised intron. When the attack occurs on the full-length molecule, it is split into the 5' exon + ORF and CRF + 3' exon molecules. The 5' exon + ORF molecule is a self-processing dead end, since it does not contain the group I cis-acting elements to drive further reactions. This point has been tested by the in vitro processing experiment using transcript from the wild-type clone (BS17-17) linearized at the BamHI site. This transcript is structurally analogous to the 5' exon + ORF molecule as it, too, lacks some of the cis-acting sites. As demonstrated by Figure 27, this molecule is completely unreactive. The CRF + 3' exon molecule does contain these elements and can undergo further reactions to produce the linear and circular CRF molecules as well as the released 3' exon. The evidence for these reactions comes from the autocatalytic reactions of mutant M44, which will be discussed below. When the internal G attack occurs on the excised intron, the linear CRF and the linear ORF are produced. Experiments by Tabak et al. (1987) have shown that the linear ORF is another self-processing dead end, while the linear CRF can circularize.

The third pathway (Figure 45C) is the release of the 3' exon from the full-length molecule. This results
in the production of the 5' exon + intron molecule as well as the free 3' exon. Potentially, the 5' exon + intron molecule could react further since it contains all the sites necessary for autocatalysis. A guanosine could attack at the intron internal site producing the linear CRF and 5' exon + ORF molecules. Additionally, the G residue at the 3' end of this molecule could attack at the 5' intron/exon boundary in a reaction analogous to the circle formation of the linear intron producing the free 5' exon and the full intron circle.

There are two possible mechanisms for the reaction which releases the 3' exon from the full-length molecule. The first is hydrolysis. Hydrolysis at the 3' intron/exon boundary has been described for the Tetrahymena intron. The second possible mechanism is a guanosine nucleophilic attack at this site. This reaction has been reported by van der Horst and Tabak (1987) for aI3 and the omega intron and would result in the addition of a G residue to the 5' end of the 3' exon. This reaction can take place during transcription of the full-length molecule, since the reaction products are seen at t=0 in time course experiments. The reaction conditions for in vitro transcription are 6mM MgCl₂, 2mM spermidine, 50mM NaCl and 2mM of each rNTP.
These conditions do not support other reactions of AI3 and, therefore, the secondary structure requirements for the 3' exon release may be different from those for other reactions. The reaction can also take place in the standard self-splicing conditions used in this study, since the amount of the 5' exon + intron molecule increases over time in time course reactions.

A reaction mechanism which explains how all intron-derived products are formed has been proposed by Tabak et al. (1987). The mechanism suggests that all products are the result of various sequences interacting with the IGS element. Those interactions are diagrammed in Figure 46. The first interaction diagrammed (Figure 46A) is that of the IGS with the 5' and 3' exon in the normal self-splicing pathway. The nucleophilic guanosine attack releases the 5' exon, creating a free 3' OH group on the final U residue of the exon, which then attacks the 3' intron/exon boundary. The circularization reactions of the linear intron are thought, by Tabak et al. (1987) to be produced by further interactions at the IGS. The G residue added to the 5' end of the intron makes the sequence at this point GAAC, which is the sequence at the 3' intron/exon boundary, and which can bind to the same point within
the IGS thought to be involved in the proposed P10 stem (see INTRODUCTION). The G residue at the 3' end of the intron can then attack at the bond between the G and A residue to form the full-length intron circle and release the added G. Given the recent results of Been and Cech (1987) circularization mechanisms which rely exclusively on the IGS as a binding site for circularization may prove to be simplistic. Those researchers showed that the set of bases in the IGS required for formation of the P1 stem of the Tetrahymena rRNA intron to be different from the set required for circularization. Whether this is generally true for all group I introns remains to be tested.

Reactions that produce the linear and circular forms of the intron ORF and CRF molecules from the excised intron involve the intron internal G addition site. This sequence is GGUGUA and begins at base 6284. The GGU can bind with the IGS in a manner similar to the 5' exon as shown in Figure 46B. The G residue at the 3' end of the intron can attack at the bond following the GGU splitting the molecule and creating a free 3' OH at the end of the piece containing the ORF and a G residue at the 5' end of the piece containing the CRF. The G residue at the 3' end of the intron can bind to the G at
the 5' end of the CRF containing fragment, thus forming the CRF intron circle. The first step in this reaction is analogous to the G attack, which is the first step in self-splicing. The free 3' OH group can, in a reaction analogous to exon ligation, attack the bond between G and A at the 5' end of the intron to form the ORF intron circle. This reaction can produce either free or interlocked circles. These circles can be the progenitors of the CRF and ORF linear molecules. There are two possible ways that the linear molecules can be formed from the circular molecules: site specific hydrolysis and mechanical breakage. Site specific hydrolysis has been demonstrated for the Tetrahymena rRNA intron circle (Inoue et al., 1986). However, this reaction is thought to require the group I cis-acting sites and is, therefore, an unlikely mechanism for the formation of the linear ORF from the circular ORF.

When the intron internal site is attacked by guanosine in solution rather than the guanosine at the 3' end of the molecule, the released CRF molecule has the sequence GGURAC at its 5' end. This sequence can bind to the IGS as shown in Figure 46C. A nucleophilic attack by the G residue at the 3' end of the molecule on the bond between the two G's of this sequence would form
a circle with the primary sequence structure which Tabak et al. (1987) has found at the cyclization junction of the CRF circle and release the added G residue. This reaction is unusual in that it does not involve a U residue paired with a G in the IGS.

The full-length intron circle can undergo a similar set of reactions. These reactions can be initiated by the attack of either a guanosine nucleotide or a hydroxyl ion. The reaction has been shown to proceed in buffers without added guanosine (Tabak et al., 1987). However, when unlabeled full-length circle is incubated with $^{32}$P GTP, some end-labeled linear CRF molecules are formed (Tabak et al., 1987). Therefore, both reactions may take place. In the case of G addition to the full intron circle (Figure 46D), the G residue attacks at the intron internal site in a reaction similar to the initial transesterification in the splicing pathway. This produces a free 3' OH at the U residue in the GGU of the intron internal site, which can then attack at the GAAC sequence formed by the circularization of the full-length intron. This reaction is analogous to exon ligation and forms the ORF circle. These reactions release the 431 nt linear CRF molecule with an additional G at the 5' end, which can undergo the
cyclization reaction described above. Hydrolysis of the intron is thought to occur at two possible sites: the intron internal G addition site and the cyclization site (Figures 47D and 47E). A single transesterification following hydrolysis results in the production of a circle and a linear molecule. When the hydrolysis takes place at the intron internal site, the ORF circular and the CRF linear molecules are formed. When the hydrolysis takes place at the cyclization site, the CRF circular and the ORF linear molecules are formed.

The reactions of aI3 seen by Tabak et al. (1987) with strain KL14-4A and those seen in this study with strain 161 are very similar. Some differences exist, however. Tabak et al. (1987) saw no evidence of a G attack at the intron internal site on the full-length molecule. I have found evidence of such a reaction. The processing reaction of the wild-type in vitro transcript produces molecules which I have identified as the 5' exon + ORF and the CRF + 3' exon. These are the species which would be formed by a guanosine attack at the internal site. As will be discussed below, mutant M44 also shows that this reaction can occur.
One of the major products of aI3 processing described by Tabak et al. (1987) is the interlocked ORF and CRF circles. In this study, that product is rarely seen, except in those reactions done at high concentrations in preparation for northern blotting. There are several possible explanations for this discrepancy. First, this may be a strain-related difference. The sequence of aI3 in strain KL14-4A is unpublished; there may be some difference that encourages interlocked circle formation. However, this seems unlikely, since the sequences of the elements most intimately involved in circle formation (see Tabak mechanism below), the IGS and the intron internal G addition site, have been published for KL14-4A and are identical to those of 161. Another possibility is that in my hands the interlocked circle is subjected to more mechanical breakage, which would result in the release of the ORF and CRF linear molecules. This could be caused by minor differences in handling the reaction or in loading and the running the analytical gel.

A third possibility is that the interlocked circle molecule is an artifact resulting from high RNA concentrations. If this were the case, the interlocked molecules are formed simply because ORF and CRF circles
are being formed in the reaction mix and occasionally they become interlocked. In this case one would expect to see at least three different populations of interlocked circles: 1) two ORF circles interlocked; 2) two CRF circles interlocked; and 3) one ORF and one CRF circle interlocked. The expected ratios of these species would be 1:1:2. Additionally, one might also expect to see a small number of more complex interlocked circles. Tabak et al. (1987), however, never reported two interlocked circles of the same species, even though aI3 reaction products were viewed under an electron microscope.

Tabak et al. (1987) also mention the free 5' exon as one of their reaction products. The relative amount of free 5' exon observed by them cannot be determined because all their published gels were run so far as to have lost the 5' exon. However, they state that it is found in gels that are run for shorter times. In the aI3 wild-type reaction in that study, this species is not seen. The predicted size of the 5' exon molecule is such that it should appear on gels just below the linear CRF species. A band of that size is seen in the C1072 reaction as discussed below. The ligated exons molecule is formed, however, suggesting that the ligation of the
5' exon to the 3' exon is highly efficient. It should be noted that a band is produced that has all the characteristics of the intron + 3' exon molecule (see RESULTS). A possible explanation for this is that the 5' exon is produced in such small amounts in the wild-type reaction that it is not visible on the autoradiographs. In the wild-type reaction, the intron + 3' exon proves to be a minor species; the band representing this molecule is of lower intensity than other bands in this size range. Since the intron + 3' exon is 4.6 times larger than the 5' exon, the molar amount of this molecule which is visible on a gel might not be visible with a molecule the size of the 5' exon.

An important difference between my data and those of Tabak et al. (1987) which may have an effect on the reaction products is the difference in constructs. I have used a construct which, when transcribed in vitro, produces RNA that starts at the BglII site in aI2 and ends at the BclI site in aE4. The in vitro construct is considerably different at both the 5' and 3' ends. They chose to clone a 1984 bp MspI to BclI fragment. The BclI site is the same one used in construction of BS17-17, the plasmid used for all wild-type in vitro
processing reactions in this study. The MspI site is at base 5052 within the polymorphic GC cluster in aI2 and thus does not exist in strain 161. The construct used by Tabak et al. is, therefore, 225 bp shorter at the 5' end than the construct used in this study. At the 3' end Tabak et al. linearized at the PvuII site in the plasmid. This adds over 200 bp of plasmid sequence to the transcript. I have noted in transcription reactions of the BS17-17 template linearized at the PvuII site that significant splicing takes place. Linearization at PvuII seems to alter the conditions required for splicing since no splicing activity is noticed in transcription reactions of the same plasmid linearized at the HindIII site. The effect of additional upstream or downstream sequences on group I autocatalytic activity is not well understood. It is possible that the differences in the templates have led to the minor differences noted in reaction products between the two studies.

There is no evidence that any reactions take place at the intron internal site in vivo in wild-type mitochondria. This suggests that the productive structure of the intron is stabilized in vivo in a way that it is not in the in vitro reaction. Several
proteins involved in the splicing of group I introns are known (see INTRODUCTION) but none are known to stabilize the group I secondary structure or to discriminate between true and cryptic splice sites. Such a role has been recently postulated for the MSS18 protein (Seraphin et al., 1988). This activity remains to be conclusively proven for that protein, however. The MSS18 protein is not required for the splicing of aI3, since strains carrying a disruption of this gene continue to splice this intron. What parameters such a protein (or other trans-acting factor) would use to recognize the productive structure from the non-productive structure is an open question at this time.

C. Analysis of Mutant C1085

1. Sequence Analysis

Mutant C1085 maps between bases 6240 and 6835, the region of the intron that contains all the recognized group I cis-acting sites. C1085 has eight basepair alterations in this region. These changes along with their placement in the standard group I secondary structure are shown in Figure 47. Four of these are located in cis-acting sites. The sites affected are the IGS, Q and E'. While the other changes could, in theory, cause a splicing defect, these changes are the
ones most likely to inhibit autocatalytic activity. The IGS mutation is a C to T transition at base 6310. The altered C is thought to bind with the G at base 5202 in the P1 stem. This binding is retained in weakened form by the GU pairing that can form in the mutant. Since this pairing can form, it is questionable that this change alone could produce a tight splicing defective phenotype. Waring et al. (1986) isolated a similar mutation in the IGS of the Tetrahymena rRNA intron. That mutant replaced a GC pair in the P1 stem with a GU pair and was found to have a leaky splicing defect (see section I.F.1.). It should be noted that the pairing affected is the same one altered in mutant M44 (see below). Two alterations are located in Q. The first is located at base 6528; this base participates in the P4 stem. Disruption of the P4 stem has been shown to block splicing of the Tetrahymena rRNA intron (Waring et al., 1986) (see section I.F.3.). The second is located at base 6528; this is the first base in the P6 stem. In C1085, E' contains a G to T transversion at base 6643. This G residue is the most highly conserved base in this element. Mutations to this conserved G have been previously found to inhibit splicing in bI4 (see section I.F.2.). Based on these findings this change alone could possibly inhibit splicing of C1085.
Based on the data presented above, it is difficult to assign the splicing defect in C1085 to any particular basepair change. As discussed above, several of the changes are related to analogous changes that have been shown to block splicing in other introns. The splicing block in C1085 is, therefore, most likely due to the combined effects of two or more of these mutations.

To better understand the effect of these changes on the secondary structure of ai3, I have analyzed the CRF region using the PCFOLD program developed by Zucker and Stiegler (1981). This program produces the lowest energy structure of RNA sequences up to 425 bp long. This distance is large enough to encompass all the cis-acting sites except for the IGS and the 5' intron/exon boundary. When this program is used to analyze the wild-type ai3 sequence, it correctly predicts the P3 and P4 stems and the pairings of E and E', and of P and Q, respectively. When this program is used to analyze the C1085 sequence, the structure produced no longer contains the P3 and P4 pairings. Instead, a stem is formed that includes S and E' interacting with Q. This means that P and E are left to become involved in alternate pairings. These interactions are shown in
Figure 48. The changes specific to C1085 are denoted as clear letters on a black background. This demonstrates the potential effect of the C1085 changes on the secondary structure of this intron and suggests a mode of action for this mutation.

It has recently been shown (John Moran personal communication) that C1085 does not revert to wild-type function. This suggests that more than one basepair change is responsible for the phenotype of this mutant.

2. Analysis of the In Vitro Reaction of C1085

When $^{32}$P labeled C1085 RNA is incubated under conditions which promote the self-splicing of wildtype aII3, no product species are formed over time. Based on the sequencing of C1085, this in vitro processing phenotype is not surprising. The interactions of three important cis-acting elements, which are essential for group I intron splicing both in vivo and in vitro, are disrupted in this mutant. This mutant serves as a negative control, which demonstrates the importance of these elements to the activities demonstrated by the other mutations analyzed in this study, as well as those activities demonstrated by the wild-type intron.
3. Analysis of the In Vivo Phenotype of C1085

In vivo C1085 fails to splice aI3 and produces two protein bands characteristic of aI3 splicing mutants, p44 and p40. As discussed in the Introduction, these proteins are not closely related to coxl based on their proteolytic digest patterns, but they do react with coxl specific polyclonal antibodies (Sass, 1984). The proteolytic digestions reveal that these proteins are highly related to one another. That is appropriate if they are translated from a reading frame made up of exons 1, 2 and 3 and the aI3 ORF. This reading frame is 1245 basepairs long and, therefore, would encode a protein of 415 amino acids. The size of the largest protein, approximately 44 kd, is consistent with its being encoded by a reading frame of this size. The smaller protein is probably derived from p44, possibly by a site specific proteolysis near one end or the other.

4. Experiments Which Could Determine the Mutations Leading to the Splicing Defect in Mutant C1085

A series of experiments could be performed to pin down the exact basepair responsible for the splicing defect in C1085. The most straightforward approach
would be to engineer each of the basepair changes found in C1085 into a wild-type aI3 time using in vitro mutagenesis techniques. In this way the effect of each change could be analyzed separately by testing each construct for splicing in vitro.

D. Analysis of Mutant M44

1. Sequence Analysis

Mutant M44 has been genetically mapped between bases 5061 and 6395. There is only one basepair change in this region from the deduced wild-type sequence, a G to T transversion at base 5202. Therefore, this change must be responsible for the splicing defective phenotype. An analogous splicing mutation, G3457, in bI4 has been described by Perea and Jacq (1985). This mutation changes a G to an A one basepair 5' of the intron/exon boundary.

Since the original paper proposing the pairing of the IGS with the 5' exon in group I introns (Davies et al., 1982), there has been some doubt as to whether the PI pairing was functional in aI3 (see the INTRODUCTION). Mutant M44 helps to prove that this interaction does take place. The M44 lesion destabilizes the proposed PI pairing, which aligns the 5' intron/exon boundary in
the group I secondary structure. The effect of this mutation on the P1 stem is shown in Figure 49. It is important to note that other than the P1 stem no important interaction in aI3 is modified. As has been demonstrated in this study and by Tabak et al., aI3 contains a sequence that can form an alternate P1 stem, including the intron internal G addition site in place of the 3' end of the 5' intron.

2. Analysis of In Vitro Processing of M44

Under self-splicing conditions, M44 RNA accumulates a set of products which appear to be a subset of those that accumulate in wild-type (see Figure 41). No bands migrating more slowly than the full-length transcript are produced. Single bands accumulate with mobilities corresponding to 1800 nt, 1500 nt, 755 nt, 550 nt, 450 nt, and 320 nt. As mentioned in the RESULTS, the 755 nt molecule has the correct length to be the CRF + 3' exon molecule. The 550 nt molecule co-migrates with the CRF circle, and the 450 nt molecule co-migrates with the linear CRF in the wild-type reaction. The 320 nt molecule is the correct length to be the 3' exon in the linearizations used for the mutant transcriptions.
The in vitro processing phenotype of mutant M44 is consistent with the sequence of the mutation. All the species produced following incubation of M44 RNA under self-processing conditions (shown in Figure 41) can be explained by assuming that all reactions described for the wild-type intron take place except those which require the formation of the normal P1 stem. The proposed in vitro processing pathway for M44 is shown in Figure 50. Since the M44 mutation weakens the normal P1 stem, no nucleophilic guanosine attacks can take place at the 5' intron/exon boundary. However, M44 retains all the other group I interactions intact; therefore, all the activities associated with the intron internal G addition site should take place. As diagrammed in Figure 50, the reactions begin with the nucleophilic G addition at the intron internal site. This releases the 5' exon + ORF (1484 nt) and the CRF + 3' exon (755 nt) molecules. The CRF + 3' exon molecule can further react to yield circular and linear CRF and free 3' exon. The band which runs at the 1800 to 2000 nt region of the gel is most probably the intron + 5' exon (1915 nt) that results from hydrolysis (or a nucleophilic G attack, see above) at the 3' intron/exon boundary. While the intron internal site is capable of supporting G addition, it does not seem to be efficient in intron ligation in
vitro. The molecule which replaces the true 5' exon in this reaction, the 5' exon + ORF species, is produced in abundance. This is in contrast to the situation in the wild-type reaction, where no free 5' exon is produced. Furthermore, no band of the appropriate length (1808 nt) to be the 5' exon + ORF molecule ligated to the 3' exon is produced over time.

3. Analysis of the In Vivo Phenotype of M44

The in vivo phenotype of M44 has been investigated by Mecklenberg (1986). As presented in the INTRODUCTION, this mutant produces two novel proteins, p44 and p40. These are identical to those produced by C1085. They are in the correct size range to be encoded by the reading frame formed by aE1, aE2, aE3 and the ORF of aI3. Northern blots of in vivo mitochondrial RNA probed with an aE4 probe reveal two RNA species of 3.7 and 3 kb (Sass, 1984). The 3.7 kb RNA has been identified by Sass (1984) as message plus aI3. The 3 kb species is the correct size to be aI3 plus downstream exons. This is consistent with the location of the M44 lesion within the P1 stem. In vivo M44 coxl pre-message may be cut at the 5' exon, but not ligated. This reaction would have to be facilitated by trans-acting factors, because the in vitro reaction shows no evidence
of any reaction at the 5' intron/exon boundary. If this model is true, a trans-acting factor would stabilize the P1 stem sufficiently to allow the 5' cut but not enough to support ligation.

An alternate explanation exists, however. In vivo, M44 may use the intron internal site as the intron/exon boundary and produce a message that is all the exons plus the aI3 ORF. That molecule has a predicted size of 3.2 kb and would be difficult to distinguish from a genuine 3 kb molecule on mobility alone. Also, it could encode the mutant related proteins, since it retains the upstream exons plus intron ORF reading frame. The ligation of the upstream to the downstream molecule would have to be a protein aided reaction, since this ligation does not take place in vitro. Chandry and Belfort (1987) have isolated an exon mutation in the T4 td gene which disrupted the true P1 stem of the group I intron of that gene. That mutation caused td pre-mRNA to be processed at a cryptic site which was homologous to the bona fide intron/exon boundary.
E. Analysis of Mutant C1072

1. Sequence Analysis

Mutant C1072 maps to exactly the same region as M44. In C1072, this region contains only one deviation from the wild-type sequence, a G to A transition at base 5196. This alteration must be responsible for the phenotype of the mutant. Base 5196 is 8 bp 5' of the 5' intron/exon boundary. That position is beyond the recognized P1 stem. The C1072 lesion is in a location analogous to that of mitochondrial revertants from strains disrupted for the MSS18 gene mentioned in the INTRODUCTION. This region, therefore, has been shown to be important for group I splicing. The reason for this importance is not yet clear, however. There are several possibilities which will be discussed below in light of their relationship to the in vitro and in vivo phenotypes of this mutant.

2. In Vitro Processing Analysis

The pattern of self-processing products formed by the C1072 aI3 mutant is identical to that formed by the wild-type intron with one important difference: some free 5' exon accumulates. There are two possible explanations for the appearance of this molecule. First, the mutation may change the site where the
initial nucleophilic attack takes place, leading to the production of a free 5' exon molecule with a 3' end different from that of the normal exon. This molecule would have a reduced ability to ligate to the 3' exon. Second, the mutation, while not altering the size of the 5' exon, may affect its ability to ligate to the 3' exon. If the first model were true, one would expect that the C1072 mutation would change the sequence at the point of the mutation to be more like the true 5' splice site. In fact, the mutation changes the site from GGTT to GATT, which is less homologous to GGTA, the true 5' splice site, than to the wild-type sequence. The second model suggests that the local sequence or secondary structure could impede the ligation of the 5' and 3' exons, but not the nucleophilic attack which releases the 5' exon from the intron. This would be accomplished if the C1072 mutation disrupted a stem, which could act to stabilize the free 5' exon. In vitro, this mutant is leaky, since some ligated exons molecules are formed. This suggests that the in vitro effect of this mutation is subtle.

3. Analysis of the In Vivo Phenotype of C1072

The in vivo phenotype of C1072 has been shown to be similar to that of M44 (Mecklenberg, 1986). This is
consistent with the location of both mutants, since they are situated within the 5' exon near the 5' splice site.

C1072 produces both the p44 and p40 proteins, as well as an additional protein, p42. This protein has been shown to be related to p44 and p40. The C1072 mutation alters an amino acid encoded by aE3 from GLY to ASP. This amino acid change may form a new proteolytic site in the protein encoded by the upstream exons and the intron ORF. If this hypothesis is correct, p44 is the full-length protein encoded by the upstream exons and the intron ORF, and p42 and p40 would result from proteolytic cleavages of that protein. The coxl related RNAs found in C1072 are the same as those found in M44, and exactly the same arguments apply as to their identification and origin. It should be noted that, while the two possible origins for the 3 kb RNA seen in M44 hold as well for C1072, the answer may prove to be different in each case. The in vitro phenotype of C1072 argues strongly for the structure of the 3 kb RNA to be aI3 plus the downstream exons in this mutant. This is exactly the species expected if the ligation of the 5' exon is inhibited in the mutant, as it is in the in vitro reaction.
In contrast to the observed in vitro phenotype, the C1072 mutation forms a very tight processing block in vivo. This is demonstrated by the lack of any coxl in the mitochondrial protein gel shown in Figure 18. Also, Sass (1984) has reported that no mature message is formed in this mutant. If the mutation were as leaky in vivo as it is in vitro, one would expect to see some coxl on such gels. Therefore, the effect of C1072 demonstrated in vitro may be only a secondary effect in vivo. Alternatively, in vivo the effect may be essentially the same, but it may be exacerbated by the requirements of in vivo splicing. In vivo C1072 may alter a protein binding site. Potential activities of this protein include stabilizing the P1 stem or holding the 5' exon in place following the cleavage at the 5' splice site. The second activity is consistent with the 3kb RNA being aI3 plus downstream exons. MSS18 gene product has been suggested to bind at the 5' end of aI5beta (see the INTRODUCTION). The C1072 mutation may disrupt an as yet undefined primary or secondary structure element recognized by a similar splicing protein. It should also be noted that the activities described above could be fulfilled by RNA molecules as well as proteins. This is certainly the case in the splicing of nuclear mRNA molecules.
4. Experiments for Further Analysis of Mutant C1072

The most straightforward approach to further study of mutant C1072 would be to isolate second site revertants of this mutant. Revertants of mitochondrial mutants are easily isolated by selecting for growth on glycerol media. Revertants isolated in this manner could then be screened to differentiate second site revertants from true revertants in two ways. First, one could test for temperature sensitivity of the revertants. Temperature sensitive revertants would not be back mutations, since the normal processing of aI3 is not temperature sensitive. Second, revertants could be screened by differential hybridization. In this experiment, revertant colonies would be transferred to nitrocellulose filters and lysed, and their total DNA bound to the filter. This filter would then be hybridized at low stringency with a $^{32}$P labeled oligonucleotide containing the C1072 sequence. A series of washes and exposures could then be done at increasing stringency (higher temperature) until a point was reached where the stability of the duplex between the probe containing the C1072 mutation and the mitochondrial sequence of a same site revertant caused the probe to be washed off those colonies. This
strategy would eliminate revertants which do not retain the C1072 basepair alteration.

Two types of second site revertants can be envisioned for C1072, mitochondrial revertants and nuclear revertants. Second site mitochondrial revertants of C1072 would involve the reformation of a secondary structure element disrupted by this mutation. Nuclear revertants could involve nuclear proteins involved group I intron processing. As discussed above, the C1072 lesion could alter the binding site of a trans-acting factor required for in vivo splicing of a13. If a nuclear suppressor of C1072 could be found, it would strongly suggest that this mutation affects a protein binding site.

Additionally, *in vitro* mutagenesis techniques could be used to engineer more changes in the 5' exon near the C1072 mutation. If an alteration could be found that blocked *in vitro* splicing of this mutant it would suggest that C1072 effects a secondary structure element.
F. Future Directions for Intron aI3 and Group I Intron Research

1. Future Directions for aI3 Research

The most important remaining questions to be resolved in the splicing of aI3 are those of the importance and function of the ORF. There are basically two questions: 1) What is the secondary structure of the ORF and what is its contribution to splicing? 2) Does the ORF encode a maturase or other enzymatic activity? Since aI3 can splice in vitro one could answer the first question by in vitro mutagenizing the wild-type intron. The ORF region of the intron could be deleted entirely either by removing a restriction fragment which contains the ORF or by exonuclease digestion. This construct could then be tested for autocatalytic activity. If such a construct were found to retain autocatalytic activity, it would suggest that structure of the ORF does not play an active role in the autocatalytic activity of aI3. The role of the secondary structure of the ORF region may be a passive one. That is, the ORF may be required to fold into a secondary structure simply to avoid obstructing the folding of the traditional group I interactions. This could also be tested using in vitro mutagenesis. One could predict a secondary structure for the ORF region
and then engineer in changes which would interrupt this structure and test for self-splicing. If a particular mutation disrupted splicing, the authenticity of the proposed interaction could be tested by making another construct, which reconstituted the original structure but not the original sequence, and testing for self-splicing.

The second question can be answered by testing aI3 containing petites for splicing of that intron. As mentioned in the INTRODUCTION, petites are incapable of mitochondrial protein synthesis because genes encoding portions of the translation apparatus are missing. If aI3 were found to splice in such a strain, this result would eliminate any possibility that it encodes a maturase required for its own excision. This does not eliminate the possibility that the ORF encodes whose function is unrelated to splicing. One possible activity for this protein would be as an endonuclease similar to that encoded by the omega ORF. This possibility could be tested by mating an aI3 containing strain to one with no intron at the aI3 position in the genome and looking for conversion.
2. Future Directions for Group I Intron Research

Though group I introns have been studied intensively in the past decade, they are far from being completely understood. The interaction of the 3' exon with the structure is poorly defined at this time. This could be investigated using in vitro mutagenesis techniques similar to those employed in the study of other interactions in the group I secondary structure. The roles of trans-acting factors need to be more fully investigated. As mentioned above, the C1072 mutation may shed some light on this problem with further inquiry. Experiments designed to isolate nuclear suppressors of this mutation could isolate factors which interact at the 5' splice site of group I introns. An interesting question raised by this research is how bona fide splice sites are discriminated from alternate sites. Alternate sites have been shown to exist in aI3, yet they are ignored in the natural situation. The evolution and dispersal of group I introns is also a potentially interesting area of study. Given what is currently known about transposition of the omega and aI4 introns and the trans-splicing activity of the Tetrahymena rRNA intron, it is possible that group I introns have evolved from a transposon-like ancestor.
### Table 1

** YEAST AND BACTERIAL STOCK STRAINS **

<table>
<thead>
<tr>
<th>Organism</th>
<th>Strain</th>
<th>Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yeast</td>
<td>ID41-6/161</td>
<td>$a, ade, lys$</td>
</tr>
<tr>
<td>Yeast</td>
<td>D273-10B/A1</td>
<td>$alpha, met$</td>
</tr>
<tr>
<td>E. <em>coli</em></td>
<td>HB101</td>
<td>$F-, hsdS20, recA13, ara-14, proA2, lacY1, galK2, rslP20, xyl-5, mtl-1, supE44k, lambda-</td>
</tr>
<tr>
<td>E. <em>coli</em></td>
<td>JM101</td>
<td>D(lac proA,B), supE, thi, (rk+, mk+)/F' traD36, proAB, lacIQZDM15</td>
</tr>
<tr>
<td>E. <em>coli</em></td>
<td>GM161</td>
<td>thr1, leuB6, dam4, thi1, hsdS1, lacY1, tonA21, supE44, lambda-</td>
</tr>
</tbody>
</table>
Table 2

MIT- AND PETITE STRAINS

<table>
<thead>
<tr>
<th>Strain</th>
<th>Mit-/petite</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>A17-17</td>
<td>petite</td>
<td>Vincent, 1982</td>
</tr>
<tr>
<td>401</td>
<td>petite</td>
<td>Bonitz et al., 1980</td>
</tr>
<tr>
<td>422/C2</td>
<td>petite</td>
<td>Anziano, 1984</td>
</tr>
<tr>
<td>422/N1</td>
<td>petite</td>
<td>Anziano, 1984</td>
</tr>
<tr>
<td>C245</td>
<td>mit-</td>
<td>Mecklenburg, 1986</td>
</tr>
<tr>
<td>C1072</td>
<td>mit-</td>
<td>Mecklenburg, 1986</td>
</tr>
<tr>
<td>C1085</td>
<td>mit-</td>
<td>Mecklenburg, 1986</td>
</tr>
<tr>
<td>M44</td>
<td>mit-</td>
<td>Mecklenburg, 1986</td>
</tr>
</tbody>
</table>
Table 3

**PETITE RESTORATION OF M44, C1072, C1085**

<table>
<thead>
<tr>
<th>petite tester</th>
<th>M44</th>
<th>C1072</th>
<th>C1085</th>
</tr>
</thead>
<tbody>
<tr>
<td>401</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>422</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>422/C2</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>422/N1</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>422/P3</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
**Table 4**

**COXI SYNTHETIC PRIMERS**

<table>
<thead>
<tr>
<th>Designation</th>
<th>Sequence</th>
<th>Position</th>
<th>Sense/Antisense</th>
</tr>
</thead>
<tbody>
<tr>
<td>H02</td>
<td>GGTAATATAATGGTCCA</td>
<td>4987-5004</td>
<td>Sense</td>
</tr>
<tr>
<td>H01</td>
<td>GGGTGCGCCCGGGGGCGG</td>
<td>5031-5048</td>
<td>Sense</td>
</tr>
<tr>
<td>H1</td>
<td>TTACATTGCAGATAGGTTT</td>
<td>5141-5158</td>
<td>Sense</td>
</tr>
<tr>
<td>H2</td>
<td>GATTTAGAATTCCTAAAT</td>
<td>5404-5422</td>
<td>Sense</td>
</tr>
<tr>
<td>H3</td>
<td>CCATTAGATACATCAGAT</td>
<td>5660-5677</td>
<td>Sense</td>
</tr>
<tr>
<td>H4</td>
<td>GTATCTTTAATAATACG</td>
<td>5921-5938</td>
<td>Sense</td>
</tr>
<tr>
<td>H5</td>
<td>AGAACTACGTTTACTTGA</td>
<td>6158-6175</td>
<td>Sense</td>
</tr>
<tr>
<td>H6</td>
<td>CACATGTGGGCCCTCGC</td>
<td>6400-6416</td>
<td>Sense</td>
</tr>
<tr>
<td>H7</td>
<td>AATTCTTGGAAATGCTGT</td>
<td>6771-6754</td>
<td>Antisense</td>
</tr>
<tr>
<td>Sequence</td>
<td>Change (in 161)</td>
<td>Effect</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>----------------------------------</td>
<td>---------------------------------</td>
<td></td>
</tr>
<tr>
<td>5027-5058</td>
<td>deletes these bases</td>
<td>lose a GC cluster</td>
<td></td>
</tr>
<tr>
<td>5327</td>
<td>TCT to ACT</td>
<td>ser to thr</td>
<td></td>
</tr>
<tr>
<td>5490</td>
<td>CCT to CTT</td>
<td>pro to thr</td>
<td></td>
</tr>
<tr>
<td>5562</td>
<td>GGA to GAA</td>
<td>gly to glu</td>
<td></td>
</tr>
<tr>
<td>5611-5617</td>
<td>ACC AAC CAA to ACA ACA ATT</td>
<td>thr asn gln to thr thr ile</td>
<td></td>
</tr>
<tr>
<td>5702-5707</td>
<td>ATC TTC to GGT ATG</td>
<td>ile leu to gly met</td>
<td></td>
</tr>
<tr>
<td>5767-5768</td>
<td>AGG ACA to AGA GCA</td>
<td>arg thr to arg ala</td>
<td></td>
</tr>
<tr>
<td>5814</td>
<td>ATT to AAT</td>
<td>ile to asn</td>
<td></td>
</tr>
<tr>
<td>6103</td>
<td>AAT to AAA</td>
<td>asn to lys</td>
<td></td>
</tr>
<tr>
<td>6307</td>
<td>A to T</td>
<td>changes internal guide</td>
<td></td>
</tr>
<tr>
<td>6412-6413</td>
<td>GCCT to GCCCT</td>
<td>unknown</td>
<td></td>
</tr>
<tr>
<td>6422</td>
<td>T to A</td>
<td>unknown</td>
<td></td>
</tr>
<tr>
<td>6426-6427</td>
<td>TAAAT to TAAAAAT</td>
<td>unknown</td>
<td></td>
</tr>
<tr>
<td>6502-6503</td>
<td>TGGGGGGGGGGT to TGGGGGGGGGGGGT</td>
<td>unknown</td>
<td></td>
</tr>
</tbody>
</table>
Table 6

**BASEPAIR CHANGES IN M44, C1072, C1085**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Sequence</th>
<th>Change</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>M44</td>
<td>5202</td>
<td>GGT to GTT</td>
<td>gly to val alters 5' exon binding with internal guide</td>
</tr>
<tr>
<td></td>
<td>6620</td>
<td>T to C</td>
<td></td>
</tr>
<tr>
<td>C1072</td>
<td>5196</td>
<td>GGT to GAT</td>
<td>gly to asp</td>
</tr>
<tr>
<td>C1085</td>
<td>6218</td>
<td>TATT to TAATT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6310</td>
<td>C to T</td>
<td>alters internal guide</td>
</tr>
<tr>
<td></td>
<td>6377</td>
<td>T to C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6425-6426</td>
<td>GTAAAAA to AAAAA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6500</td>
<td>C to T</td>
<td>alters Q</td>
</tr>
<tr>
<td></td>
<td>6528</td>
<td>C to T</td>
<td>alters Q</td>
</tr>
<tr>
<td></td>
<td>6534</td>
<td>G to T</td>
<td>alters Q</td>
</tr>
<tr>
<td></td>
<td>6643</td>
<td>G to T</td>
<td>alters E'</td>
</tr>
</tbody>
</table>
Figure 1. The nuclear mRNA splicing process. Exons are shown as thick lines. The intron is shown as a thin line. Conserved bases are shown as sequence. The fate of the phosphate moieties at the 5' and 3' splice sites can be deduced by following the p in the circle (5') and the p in the diamond (3'). From Padgett et al. (1986).
Figure 1
Figure 2. The tRNA splicing pathway. Exon sequences are shown as straight lines, intron sequences as a wavy line. From Padgett et al. (1986).
Figure 2
Figure 3. Typical group II intron structure. The proposed secondary structure of group II intron aI2 is shown. Conserved structures from intron aI1 are shown in lower case letters in boxes. Arrows point to the intron/exon boundaries. From Michel and Dujon (1983).
Figure 3
Figure 4. Model for splicing of group II introns. Cross-hatched blocks represent exons. The thin line represents the intron. The fate of the phosphate moieties at the splice site can be deduced by following the p in the circle (5') and the p in the diamond (3'). From van der Veen et al. (1987).
Figure 4
Figure 5. Autocatalysis of the *Tetrahymena* rRNA intron. Exons are shown as straight lines, the intron as a wavy line. Brackets indicate a transient intermediate. L IVS is the linear intron. C IVS is the circular intron. The phosphate at the 5' splice site is shown in a circle. The phosphate at the 3' splice site is shown in a square. The phosphate at the cyclization site is shown in a diamond. From Cech (1987).
Figure 5
Figure 6. The general transesterification reaction. In the case of group I introns, R is guanosine, R' is the 5' exon, R'' is the intron plus the 3' exon. From Cech (1987).
Figure 6
Figure 7. Binding of guanosine to a group I intron. The proposed binding of guanosine with the Tetrahymena rRNA intron is shown. The proposed hydrogen bonds are shown as dotted lines. The bonds could involve up to four different nucleotides in the RNA chain. The nucleophilic attack by the 3' oxygen at the phosphate of the 5' splice site is also shown. From Bass and Cech (1984).
Figure 7
Figure 8. The general structure of group I introns. Arrows indicate the location of the 5' and 3' splice sites. P1 is the pairing of the bases at the 5' splice site and the IGS. P3 is the pairing of E and E'. P4 is the pairing of P and Q. P7 is the pairing of R and S. Other stems are conserved but do not involve recognized cis-acting elements. Stems are numbered following Burke et al. (1987). From Burke et al. (1987).
Figure 9. The structure of *Tetrahymena* L-rRNA intron. Exon bases are indicated by lower case letters. Intron sequences are indicated by upper case letters. Arrows indicate the location of the 5' and 3' splice sites. Stems are numbered following Burke et al. (1987). From Burke et al. (1987).
Figure 9
Figure 10. The pertinent sequence of the cox1 gene from strain D273. Bases are numbered following Bonitz et al. (1980). Restriction sites are shown under the DNA sequence. Open reading frames are indicated by the bases being separated into codons above which are the amino acids encoded. Exons are noted by upper case amino acids. Intron reading frames are indicated by lower case amino acids. Intron aI3 runs from base 5204 to base 6717 (the sequence as originally published was in error by one codon at the 5' and 3' splice sites). Splice sites are shown by arrows. Group I cis-acting sites of aI3 are denoted by bold upper case letters above the sequence. From Bonitz et al. (1980).
Figure 10
Figure 11. The sequence of the cis-acting elements of intron aI3. The sequence of each of the cis-acting elements of intron aI3 is shown. Letters above the sequences indicate the element. Arrows point to the exact intron/exon boundary. Numbers between the sequences represent the number of base pairs which separate the elements. The numbers below each element indicate the basepair where the elements start in the D273 cox1 sequence.
CIS-ACTING ELEMENTS OF A13

Figure 11
Figure 12. Reaction pathway and products of aI3 as deduced by Tabak et al. (1987). Full-height dark box represents the 5' exon. Full-height cross hatched box represents the 3' exon. The intron ORF is shown by the half-height open box. The intron CRF is shown by a bold line. Guanosine is indicated by G. Guanosine attacks are shown by G's with arrows. The pathway and precursor products relationships are indicated by arrows.
Figure 13. The physical map of petite A17-17. The organization of the cox3 gene and surrounding loci is indicated above the horizontal line with the start of the cox3 coding sequence at 0. The coding regions are indicated by solid bars. The intron ORFs are represented by smaller open bars. Below the physical map is the restriction map as constructed by Mecklenberg (1986). The restriction fragments shown are those produced when wild-type rho+ mtDNA is cut with the enzyme indicated at the left. Fragments are numbered according to size. The extent of petite A17-17 is represented as the horizontal line at the bottom of the figure. The dotted portion of this line represents the sequences which may be retained by this petite but were not confirmed by restriction mapping. From Mecklenberg (1986).
Figure 14. The extent of deletion C245. The organization of the cox3 gene and surrounding loci is indicated above the horizontal line with the start of the cox3 coding sequence at 0. The coding regions are indicated by solid bars. The intron ORFs are represented by smaller open bars. Below the physical map is the restriction map as constructed by Mecklenberg (1986). The restriction fragments shown are those produced when wild-type rho+ mtDNA is cut with the enzyme indicated at the left. Fragments are numbered according to size. The extent of the deletion C245 is indicated by the open bar below the restriction map. The segments of the bar formed by dotted lines indicate sequences which may be included in the deletion but could not be confirmed by restriction mapping. From Mecklenberg (1986).
## C245

<table>
<thead>
<tr>
<th>Restriction Enzyme</th>
<th>Site 1</th>
<th>Site 2</th>
<th>Site 3</th>
<th>Site 4</th>
<th>Site 5</th>
<th>Site 6</th>
<th>Site 7</th>
<th>Site 8</th>
<th>Site 9</th>
<th>Site 10</th>
<th>Site 11</th>
<th>Site 12</th>
<th>Site 13</th>
<th>Site 14</th>
<th>Site 15</th>
<th>Site 16</th>
<th>Site 17</th>
<th>Site 18</th>
<th>Site 19</th>
<th>Site 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>BamHI</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BglII</td>
<td>2</td>
<td>9</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EcoRI</td>
<td>2</td>
<td>9</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HhaI</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HindIII</td>
<td>3</td>
<td>11</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HindIII</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PvuII</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XbaI</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 14**
Figure 15. Maps of cloning vectors used in this study. The four cloning vectors used in this study are represented as circles. The name of each vector is shown within the circle. The size of each vector, in basepairs, is indicated underneath the name. Genes carried by each are noted. The beta-lactamase gene confers ampicillin resistance. Each of the plasmids contains a fragment of the lac-Z gene which confers beta-galactosidase activity through alpha complementation. The multiple cloning site (MCS) is indicated for each vector and its sequence and restriction map shown at the bottom of the figure. The orientation of the MCS in the vectors is denoted by the placement of the EcoRI and HindIII sites in each vector. The T7 and T3 promoters noted on the Bluescribe vector are bacteriophage promoters oriented so that either strand of a fragment cloned into the MCS can be transcribed.
Figure 15
Figure 16. The physical map of plasmid BS17-17. The bluescribe plasmid is shown as a circle broken by the mtDNA insert. The mtDNA insert is shown below the plasmid. Intron and exon segments included in the insert are denoted by bold letters over the diagram. The length of each segment is shown directly beneath the designation. Exons are shown as dark bars. Intron open reading frames are shown as shorter open bars. Intron closed reading frames are shown as bold lines. The placement of the group I cis-acting sites of aI3 are designated by lines extending above the insert. Short lines extending down from the insert represent a distance of 200 bp. Longer lines indicate restriction sites as noted. Restriction enzyme designations connected by a slash indicate that the site in question was destroyed by the cloning protocol which ligated BglII and BclII insert ends into BamHI plasmid ends.
Figure 17. Physical map of petites with breakpoints in the middle of oxi3. The physical map of the middle of the coxl gene is shown above the numbered line. The distance from the translational start is shown in kilobase. The coding regions are indicated by solid bars. The intron ORFs are represented by smaller open bars. Below the physical map is the restriction map as constructed by Mecklenberg (1986). The restriction fragments shown are those produced when wild-type rho+ mtDNA is cut with the enzyme indicated at the left. Fragments are numbered according to size. The solid lines below the restriction map represent the portion of the wild-type genome retained by each petite. The petite strain designations are shown above the lines. The dotted lines represent sequences which could be retained by the petite, but which could not be confirmed by restriction mapping. The open boxes within some lines indicate DNA not present in petites because they were derived from a yeast strain which does not contain those sequences. From Mecklenberg (1986).
Physical map of petites with breakpoints in the middle of ori3

Figure 17
Figure 18. Mitochondrial translation products from M44, C1072, and C1085. Mitochondrial proteins were $^{35}$S labeled as described in EXPERIMENTAL TECHNIQUES and separated on SDS acrylamide gels. Mitochondrial proteins are indicated on the right margin. Novel polypeptides p44, p42, and p40 are indicated on the left margin. Lanes are: lane 1, C1085; lane 2, C1072; lane 3, M44; lane 4, 161 wt control.
Figure 19. The coxl BglII/HindIII fragment. Intron and exon segments are denoted by bold letters over the diagram. The length of each segment is shown directly beneath the designation. Exons are shown as dark bars. Intron open reading frames are shown as shorter open bars. Intron closed reading frames are shown as bold lines. The placement of the group I cis-acting sites of a13 are designated by lines extending above the insert. Short lines extending down from the insert represent a distance of 200 bp. Longer lines indicate restriction sites as noted.
OXI3 BGL/HIND FRAGMENT

Figure 19
Figure 20. Plasmid constructs used in analysis of mutants M44, C1072 and C1085. Each construct is denoted by the designation to the right or left. When constructs of the same nature were made for each mutant, the designation of each construct is given. Commercial vectors used in each construct are shown as circles broken by inserts. The name of each vector is given within the circle. The physical map of each insert is shown below the vector. The inserts are simplified versions of Figure 19. Exons are shown as dark bars. The shorter exon is exon 3 and the longer exon is exon 4. Intron open reading frames are shown as shorter open bars. Intron closed reading frames are shown as bold lines. Short lines extending down from the insert represent a distance of 200 bp. Restriction sites used in each cloning are indicated at the ends of each insert. The orientation of each insert relative to the MCS is also indicated.
Figure 20
Figure 21. M44 sequencing strategy. The region cloned for sequencing of M44 is shown. Exons are shown as dark bars. Intron open reading frames are shown as shorter open bars. Intron closed reading frames are shown as bold lines. Short lines extending down from the insert represent a distance of 200 bp. Longer lines indicate restriction sites as noted. Sequence numbers at the beginning and end of the figure refer to the Bonitz et al. (1980) sequence of strain D273 (see Figure 10). Horizontal lines above the insert map represent the extent and direction of the sequence derived from each primer used. Primer designations are given at the beginning of each line. The location of the mutation is indicated by an arrow below the diagram of the cloned region.
M44 SEQUENCING STRATEGY

Figure 21
Figure 22. C1072 and C1085 sequencing strategy. The region of C1072 and C1085 sequenced is shown. Exons are shown as dark bars. Intron open reading frames are shown as shorter open bars. Intron closed reading frames are shown as bold lines. Short lines extending down from the insert represent a distance of 200 bp. Longer lines indicate restriction sites as noted. Sequence numbers at the beginning and end of the figure refer to the Bonitz et al. (1980) sequence of strain D273 (see Figure 10). Horizontal lines above the insert map represent the extent and direction of the sequence derived from each primer used. Primer designations are given at the beginning of each line. The M13 primer designation refers to the primer supplied by the Sequenase kit which hybridizes to the M13 cloning vectors. The locations of base pair changes in each mutant are indicated by arrows beneath the diagram.
SEQUENCING STRATEGY FOR C1072 AND C1086

Figure 22
Figure 23. Primary structure of the GC clusters from the a2 family. Each line displays the sequence of a GC cluster of the a2 family. This is the family which includes the GC cluster found at 5023 in the coxl sequence of strain D273, but not in the 161 sequence. Numbers on the left margin refer to numbers assigned to each known GC cluster by de Zamaroczy and Bernardi (1986). Common sequence motifs at the 5' and 3' borders of the elements are underlined. The cluster found to be polymorphic between strain D273 and strain 161 in this study is number 87. From de Zamaroczy and Bernardi (1986).
Figure 23
Figure 24. IGS binding in strain D273-10B/A1 and strain ID4-6/161. The potential basepairing of the IGS elements of strains D273 and 161 is illustrated. The relevant sequences are shown. Other bases in the intron are represented as a curved line. Lower case letters represent exon sequences. Upper case letters represent intron sequences. The 5' and 3' orientation of each sequence is denoted by numbers at the ends of the sequence. Bases which could pair are denoted by a line connecting them. G-U wobble pairs are denoted by a dot between them. The change in the 161 IGS is shown as a white U outlined in black. The base number according to the D273 numbering of Bonitz et al. (1980) of the base at the 5' and 3' splice site is shown above (5') or below (3') the base.
Figure 24
Figure 25. Initial time course experiment. RNA derived from plasmid BS17-17 linearized at the HindIII site was incubated, at 36°C, in a buffer containing 100mM (NH₄)₂SO₄, 50mM Tris-HCl (pH 7.5), 60mM MgCl₂, and 0.15mM GTP. Reactions were stopped at the times indicated below by the addition of 1/5 volume of 500mM EDTA and separated on a polyacrylamide-urea gel as described in EXPERIMENTAL TECHNIQUES. Approximate sizes of bands are given on the left margin. Time points are: Lane 1, 0 minutes; Lane 2, 5 minutes; Lane 3, 10 minutes; Lane 4, 20 minutes; Lane 5, 30 minutes.
Figure 26. Transcripts used from plasmid BS17-17. The physical map of plasmid BS17-17 and the extent of various transcripts derived from this plasmid are shown. In the physical map of plasmid BS17-17, exons are shown as dark bars, intron open reading frames are shown as shorter open bars and intron closed reading frames are shown as bold lines. Short lines extending down from the insert represent a distance of 200 bp. Longer lines indicate restriction sites as noted. Restriction enzyme designations connected by a slash indicate that the site in question was destroyed by the cloning protocol which ligated BgIII and BclI insert ends into BamHI plasmid ends. Plasmid sequences are shown as stippled bars. The horizontal lines below the physical map represent the extent of transcripts produced by linearizing BS17-17 at the unique restriction sites. The restriction enzyme used for each linearization is noted in the center of the line. The promoter used for each transcription is identified at the 5' end of each transcript.
BS17-17 TRANSCRIPTS

Figure 26
Figure 27. Reaction of BamHI linearized transcript. RNA derived from plasmid BS17-17 linearized at the BamHI site was incubated, at 36° C., in a buffer containing 100 mM (NH₄)₂SO₄, 50 mM Tris-HCl (pH 7.5), 60 mM MgCl₂, and 0.15 mM GTP. Reactions were stopped at the times indicated below by the addition of 1/5 volume of 500 mM EDTA and separated on a polyacrylamide-urea gel as described in EXPERIMENTAL TECHNIQUES. Time points are: Lane 1, 30 minutes; Lane 2, 15 minutes, Lane 3, 0 minutes.
Figure 28. GTP concentration response. RNA derived from plasmid BS17-17 linearized at the HindIII site was incubated, at 36°C, in a buffer containing 100mM (NH₄)₂SO₄, 50mM Tris-HCl (pH 7.5), 60mM MgCl₂, and varying concentrations of GTP as listed below. Reactions were stopped by the addition of 1/5 volume of 500mM EDTA and separated on a polyacrylamide-urea gel as described in EXPERIMENTAL TECHNIQUES. Approximate sizes of bands are given on the left margin. GTP concentrations are: Lane 1, 0.0mM; Lane 2, 0.075mM; Lane 3, 0.15mM; Lane 4, 0.3mM; Lane 5, 0.6mM.
Figure 28
Figure 29. (NH₄)₂SO₄ concentration response. RNA derived from plasmid BS17-17 linearized at the HindIII site was incubated, at 36°C, in a buffer containing 50mM Tris-HCl (pH 7.5), 60mM MgCl₂, 0.15mM GTP and varying concentrations of (NH₄)₂SO₄ as indicated below. Reactions were stopped by the addition of 1/5 volume of 500mM EDTA and separated on a polyacrylamide-urea gel as described in EXPERIMENTAL TECHNIQUES. Approximate sizes of bands are given on the left margin. (NH₄)₂SO₄ concentrations are: Lane 1, 200mM; Lane 2, 100mM; Lane 3, 50mM; Lane 4, 25mM; Lane 5, 0mM.
Figure 29
Figure 30. MgCl₂ concentration response. RNA derived from plasmid BS17-17 linearized at the HindIII site was incubated, at 36°C, in a buffer containing 100mM (NH₄)₂SO₄, 50mM Tris-HCl (pH 7.5), 0.15mM GTP and varying concentrations of MgCl₂ as indicated below. Reactions were stopped by the addition of 1/5 volume of 500mM EDTA and separated on a polyacrylamide-urea gel as described in EXPERIMENTAL TECHNIQUES. Approximate sizes of bands are given on the left margin. MgCl₂ concentrations are: Lane 1, no incubation; Lane 2, 0mM; Lane 3, 10mM; Lane 4, 20mM; Lane 5, 60mM; Lane 6, 100mM.
Figure 30
Figure 31. Temperature response. RNA derived from plasmid BS17-17 linearized at the HindIII site was incubated in a buffer containing 100 mM (NH$_4$)$_2$SO$_4$, 50 mM Tris-HCl (pH 7.5), 60 mM MgCl$_2$, and 0.15 mM GTP at the temperatures indicated below. Reactions were stopped below by the addition of 1/5 volume of 500 mM EDTA and separated on a polyacrylamide-urea gel as described in EXPERIMENTAL TECHNIQUES. Approximate sizes of bands are given on the left margin. Incubation temperatures are: Lane 1, no incubation; Lane 2, 22° C.; Lane 3, 30° C.; Lane 4, 36° C.; Lane 5, 42° C.
Figure 31
Figure 32. AI3 self-splicing products. The self-splicing products of AI3 as reported in this study and by Tabak et al. (1987) are represented diagrammatically. In this diagram, exons are represented as dark bars, intron ORFs are represented as shorter open bars, and intron CRFs as bold lines. The sequences represented are those cloned in plasmid BS17-17 (see Figure 16). The linear sequences are depicted with the 5' end on the left.
SELF-SPLICING PRODUCTS

- Interlocked circles (1083nt + 431nt)
- Full intron circle (1514nt)
- ORF circle (1083nt)
- Full length transcript (2267nt)
- Intron + 5' exon (1915nt)
- Intron + 3' exon (1866nt)
- Linear intron (1814nt)
- 5' exon + ORF (1484nt)
- Linear ORF (1883nt)
- CRF + 3' exon (784nt)
- Ligated exons (768nt)
- CRF circle (431nt)
- Linear CRF (431nt)
- 5' exon (482nt)

Figure 32
Figure 33. BS17-17 time course gel and band identification. RNA derived from plasmid BS17-17 linearized at the HindIII site was incubated, at 42°C, in a buffer containing 100mM (NH₄)₂SO₄, 50mM Tris-HCl (pH 7.5), 60mM MgCl₂, and 0.15mM GTP. Reactions were stopped at the times indicated below by the addition of 1/5 volume of 500mM EDTA and separated on a polyacrylamide-urea gel as described in EXPERIMENTAL TECHNIQUES. The identification of each band, as elucidated in RESULTS, is given at the right. Product species which cannot be visualized on this gel but which have been reported by other groups (Tabak et al., 1987) or visualized in other experiments in this study are shown in parentheses. Time points are: Lane 1, 80 minutes; Lane 2, 60 minutes; Lane 3, 40 minutes; Lane 4, 20 minutes; Lane 5, 10 minutes; Lane 6, 0 minutes.
Figure 33
Figure 34. Early time course gel and band identification. RNA derived from plasmid BS17-17 linearized at the HindIII site was incubated, at 42°C, in a buffer containing 100mM (NH₄)₂SO₄, 50mM Tris-HCl (pH 7.5), 60mM MgCl₂, and 0.15mM GTP. Reactions were stopped at the times indicated below by the addition of 1/5 volume of 500mM EDTA and separated on a polyacrylamide-urea gel as described in EXPERIMENTAL TECHNIQUES. The identification of each band, as elucidated in RESULTS, is given at the right. Product species which cannot be visualized on this gel but which have been reported by other groups (Tabak et al., 1987) or visualized in other experiments in this study are shown in parentheses. Time points are: Lane 1, 0 minutes; Lane 2, 1 minute; Lane 3, 2 minutes; Lane 4, 4 minutes; Lane 5, 6 minutes; Lane 6, 8 minutes; Lane 7, 10 minutes.
Figure 34
Figure 35. PvuII vs HindIII linearizations of BS17-17. RNA derived from plasmid BS17-17 linearized at the HindIII site or PvuII site was incubated, at 42°C, in a buffer containing 100mM (NH₄)₂SO₄, 50mM Tris-HCl (pH 7.5), 60mM MgCl₂, and 0.15mM GTP. Reactions were stopped by the addition of 1/5 volume of 500mM EDTA and separated on a polyacrylamide-urea gel as described in EXPERIMENTAL TECHNIQUES. Linearization at the PvuII site increases the size of any species containing the 3' end of the full-length transcript by 213 bases over that of the same species derived from HindIII linearized template. Species for which altered mobility could be demonstrated under these conditions are noted in the margins. Lane identifications are: Lane 1, PvuII linearized BS17-17 incubated; Lane 2, PvuII linearized BS17-17 no incubation; Lane 3, HindIII linearized BS17-17 incubated; Lane 4, HindIII linearized BS17-17 no incubation.
Figure 35
Figure 36. Probes used in this study. The physical map of plasmid BS17-17 and the extent of probes used in identification of product species is shown. In the physical map of plasmid BS17-17, exons are shown as dark bars, intron open reading frames are shown as shorter open bars and intron closed reading frames are shown as bold lines. Short lines extending down from the insert represent a distance of 200 bp. Longer lines indicate restriction sites as noted. Restriction enzyme designations connected by a slash indicate that the site in question was destroyed by the cloning protocol which ligated BglII and BclI insert ends into BamHI plasmid ends. Plasmid sequences are shown as stippled bars. The horizontal lines below the physical map represent the region covered by each probe. The 3' probe was an RNA probe transcribed from the T3 promoter of plasmid BS17-17. The 5' end of this transcript is indicated by the T3 at one end of the horizontal line.
BS17-17 PROBES

Figure 36
Figure 37. Northern blots of in vitro reaction products. RNA derived from plasmid BS17-17 linearized at the HindIII site or PvuII site was incubated, at 42°C, in a buffer containing 100mM (NH₄)₂SO₄, 50mM Tris-HCl (pH 7.5), 60mM MgCl₂, and 0.15mM GTP. Reactions were stopped by the addition of 1/5 volume of 500mM EDTA and separated on a polyacrylamide-urea gel as described in EXPERIMENTAL TECHNIQUES. Products were blotted and probed as described in EXPERIMENTAL TECHNIQUES. Bands which hybridize to each probe are identified in the margin. A. Intron probe. Lane 1, PvuII linearized transcript; Lane 2, HindIII linearized transcript. B. 3' probe. Lane 1, HindIII linearized transcript.
INTRON PROBE

LANE 1 2

INTERLOCKED CIRCLES
FULL INTRON CIRCLE
ORF CIRCLE

FULL LENGTH TRANSCRIPT
INTRON + 5' EXON
INTRON + 3' EXON
LINEAR INTRON
5' EXON + ORF
LINEAR ORF

CRF CIRCLE
LINEAR CRF

Figure 37A
Figure 37B
Figure 38. Overexposure demonstrating full-length intron circle. RNA derived from plasmid BS17-17 linearized at the HindIII site was incubated, at 42°C, in a buffer containing 100 mM (NH₄)₂SO₄, 50 mM Tris-HCl (pH 7.5), 60 mM MgCl₂, and 0.15 mM GTP. Reactions were stopped at the times indicated below by the addition of 1/5 volume of 500 mM EDTA and separated on a polyacrylamide-urea gel as described in EXPERIMENTAL TECHNIQUES. The autoradiograph has been purposely overexposed to show the full-length intron circle which is identified in the margin. Time points are: Lane 1, 0 minutes; Lane 2, 1 minute; Lane 3, 2 minutes; Lane 4, 4 minutes; Lane 5, 6 minutes; Lane 6, 8 minutes; Lane 7, 10 minutes.
Figure 38
Figure 39. Mutant C1085 in vitro processing time course. RNA derived from plasmid BS1085 linearized at the BclI site in aE4 was incubated, at 42°C, in a buffer containing 100mM (NH₄)₂SO₄, 50mM Tris-HCl (pH 7.5), 60mM MgCl₂, and 0.15mM GTP. Reactions were stopped at the times indicated below by the addition of 1/5 volume of 500mM EDTA and separated on a polyacrylamide-urea gel as described in EXPERIMENTAL TECHNIQUES. The position of the full-length transcript is indicated in the margin. Time points are: Lane 1, 0 minutes; Lane 2, 10 minutes; Lane 3, 20 minutes; Lane 4, 40 minutes; Lane 5, 60 minutes.
Figure 40. Mutant C1072 in vitro processing time course. RNA derived from plasmid BSC1072 linearized at the BclI site within aE4 was incubated, at 42°C, in a buffer containing 100 mM (NH₄)₂SO₄, 50 mM Tris-HCl (pH 7.5), 60 mM MgCl₂, and 0.15 mM GTP. Reactions were stopped at the times indicated below by the addition of 1/5 volume of 500 mM EDTA and separated on a polyacrylamide-urea gel as described in EXPERIMENTAL TECHNIQUES. The identification of each band, as elucidated in RESULTS, is given at the right. Product species which cannot be visualized on this gel but which have been reported by other groups (Tabak et al., 1987) or visualized in other experiments in this study are shown in parentheses. Time points are: Lane 1, BS17-17 HindIII linearized control; Lane 2, 0 minutes; Lane 3, 10 minutes; Lane 4, 20 minutes; Lane 5, 40 minutes; Lane 6, 60 minutes; Lane 7, 80 minutes.
Figure 40
Figure 41. Mutant M44 in vitro processing time course. RNA derived from plasmid BS44 linearized at the BclI site within aE4 was incubated, at 42°C, in a buffer containing 100mM (NH₄)₂SO₄, 50mM Tris-HCl (pH 7.5), 60mM MgCl₂, and 0.15mM GTP. Reactions were stopped at the times indicated below by the addition of 1/5 volume of 500mM EDTA and separated on a polyacrylamide-urea gel as described in EXPERIMENTAL TECHNIQUES. The identification of each band, as elucidated in RESULTS, is given at the left. Time points are: Lane 1, 80 minutes; Lane 2, 60 minutes; Lane 3, 40 minutes; Lane 4, 20 minutes; Lane 5, 10 minutes; Lane 6, 0 minutes; Lane 7, BS17-17 HindIII linearized control.
Figure 41

LANE
1 2 3 4 5 6 7

FULL LENGTH TRANSCRIPT
INTRON + 5' EXON
5' EXON + ORF

CRF + 3' EXON

CRF CIRCLE

LINEAR CRF

3' EXON
Figure 42. Overexposure of M44 in vitro processing time course. This figure shows a drastic overexposure of the upper portion of the gel shown in Figure 41. This was done to demonstrate the complete lack of bands migrating more slowly than the full-length transcript in the M44 reaction.
Figure 43. The secondary structure of intron aI3 from strain 161. The group I cis-acting interactions of aI3 from strain 161 are shown in the standard form suggested by Burke et al. (1988). The typical group I cis-acting elements are underlined. Stems are numbered according to the current convention (Burke et al., 1988). The 5' and 3' splices are indicated by arrows. Sequence numbers at the 5' and 3' ends of the figure refer to the Bonitz et al. (1980) sequence of strain D273. (See Figure 10).
Figure 44. Comparison of L5 region of aI3 from strains D273 and 161. The L5 region from intron aI3 from strain 161 is shown. The differences in the structure in strain L5, as drawn by Collins (1988), are shown in boxes next to the part of the 161 structure that is altered.
Figure 44
Figure 45. Autocatalytic pathways of intron aI3. Full-height dark box represents the 5' exon. Full-height cross-hatched box represents the 3' exon. The intron ORF is shown by the half-height open box. The intron CRF is shown by a bold line. Guanosine is indicated by G. Guanosine attacks are shown by G's with arrows. The pathway and precursor products relationships are indicated by arrows. 

A. Pathway initiated by guanosine attack on the 5' intron/exon boundary. The attack releases the 5' exon which then ligates to the 3' exon. The freed intron can undergo circularization into any of the species shown. The circles can linearize to form the linear species shown. 

B. Pathway initiated by guanosine attack at the intron internal site. When the internal attack takes place on the full-length transcript, the 5' exon + ORF and the CRF + 3' exon molecules are produced. The CRF + 3' molecule can undergo further reactions to produce the 3' exon and the linear and circular CRF. When the internal attack takes place on the free intron, the linear ORF and CRF molecules are formed. The linear CRF then can circularize. 

C. Release of the 3' exon. The 3' exon can be released to form free 3' exon and the 5' exon + intron molecule. This release may take place either by G addition or hydrolysis.
Figure 45
Figure 46. Involvement of the IGS in autocatalytic reactions of aI3. A. Interaction of the IGS with the full-length molecule. The pairing of the 5' exon (first line) and the 3' exon (third line) with the IGS (second line) leading to intron release is shown. Exon sequences are enclosed in a box. Guanosine attack is indicated by a G with an arrow. B. Interaction of the intron internal G addition site with the IGS. The possible pairing of the intron internal site (second line) and the IGS (third line) leading to the formation of subintronic circles is shown. The G attack by the guanosine residue at the 5' end of the linear intron (first line) is indicated by an arrow. The attack of the free 3' OH group, formed by the G attack (fourth line), on the 5' end of the molecule is also indicated by an arrow. C. Interaction leading to the formation of the CRF circle from the CRF linear molecule. The pairing of the 5' end of the CRF molecule (second line) with the IGS (third line) is shown. The CRF circle is formed by the attack (arrow) of the 3' G (first line) on this pairing. D. Interaction of the intron internal G addition site (second line), the IGS (third line), and the circularization site of the full intron circle leading to the formation of subintronic circular and linear species. G attack or hydrolysis (indicated by upper arrow) is thought to occur at the intron internal G addition site. G addition at this site leads to the formation of the ORF circular molecule and the CRF linear molecule which can form the CRF circle. Hydrolysis at this site leads to the formation of the ORF circular molecule (indicated by lower arrow) and the CRF linear molecule. D. Hydrolysis at the circularization site. Hydrolysis (indicated by the lower arrow) is also thought to occur at the circularization site in the full intron circle. This leads to the formation of the ORF linear molecule and the CRF circular molecule (indicated by upper arrow).
Figure 47. Basepair alterations in mutant C1085. The group I cis-acting interactions of aI3 from strain 161 are shown in the standard form suggested by Burke et al. (1988). The typical group I cis-acting elements are underlined. Stems are numbered according to the current convention (Burke et al., 1988). The 5' and 3' splice sites are indicated by arrows. Sequence numbers at the 5' and 3' ends of the figure refer to the Bonitz et al. (1980) sequence of strain D273 (see Figure 10). Regions of the structure affected by the base changes of mutant C1085 are shown in boxes next to the affected region. The base in each case is shown as white on a black background.
Figure 48. Potential disposition of group I cis-acting sites in mutant C1085. The figure shows the lowest energy configuration of the group I cis-acting elements of mutant C1085 as deduced by the FCFOID program (Zuker and Stiegler, 1981). C1085 base changes are shown as white letters on a black background. Cis-acting elements are underlined.
Figure 48
Figure 49. IGS binding in strain ID4-6/161 and mutant M44. The potential basepairing of the IGS elements of strain 161 and the effect of mutant M44 are illustrated. The relevant sequences are shown. Other bases in the intron are represented as a curved line. Lower case letters represent exon sequences. Upper case letters represent intron sequences. The 5' and 3' orientation of each sequence is denoted by numbers at the ends of the sequence. Bases which could pair are denoted by a line connecting them. G-U wobble pairs are denoted by a dot between them. The change in the 161 IGS is shown as a white U outlined in black. The base number, according to the D273 numbering of Bonitz et al. (1980), of the base at the 5' and 3' splice site is shown above (5') or below (3') the base. The M44 change is shown as a white letter on a black background.
Figure 50. Autocatalytic reactions of aI3 in intron aI3. Full-height dark box represents the 5' exon. Full-height cross-hatched box represents the 3' exon. The intron ORF is shown by the half-height open box. The intron CRF is shown by a bold line. Guanosine is indicated by G. Guanosine attacks are shown by G's with arrows. The pathway and precursor products relationships are indicated by arrows.


Herbert, C., Labouesse, M., Dujardin, G. and Slonimski, P. (1988). The NAM2 proteins from S. cerevisiae and s. douglasii are mitochondrial leucyl-tRNA synthetases, and are involved in mRNA splicing. EMBO J. 7:473-483.


covalently closed circular RNA molecules existing as highly base-paired rod-like structures. PNAS 73:3852-3856.


