MUTATIONAL ANALYSIS OF THE TRANSCRIPT 3' END SIGNAL
OF THE CYC1 GENE OF YEAST

By
Andrew Michael Spence
B.Sc.(4 yr.), The University of Winnipeg, 1980

A Thesis Submitted in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy
in
The Faculty of Graduate Studies
Department of Biochemistry

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
November 1985
© Andrew Michael Spence, 1985
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Biochemistry

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date April 9, 1986
ABSTRACT

This thesis describes an attempt to define the nucleotide sequences involved in specifying the location of the 3′ ends of transcripts of the CYCl gene of the yeast Saccharomyces cerevisiae. The boundaries of the 3′ end signal were defined with the aid of two sets of plasmid-borne deletion variants of CYCl. Mutants in each set lacked sequences extending for varying distances toward the CYCl transcript 3′ end site from a fixed site either upstream or downstream. The properties of yeast strains carrying the altered genes suggested that the CYCl 3′ end signal occupied no more than 50 base pairs, 119-168 base pairs downstream of the coding sequence and therefore immediately upstream of the 3′ end site itself.

A fragment carrying the intact 3′ end signal was inserted between the CYCl promoter and the E. coli lacZ gene on an autonomously-replicating plasmid called a pAll plasmid. Sequences within the inserted fragment were capable of suppressing the expression of lacZ from the plasmid in yeast. The distal boundary of sequences responsible for this effect coincided with the distal boundary of the 3′ end signal itself. I tentatively concluded that the suppression of lacZ expression depended on 3′ end signal function.

Mutations were introduced throughout the 3′ end signal using a new procedure of in vitro mutagenesis relying on limited primer extension on a gapped heteroduplex and subsequent misincorporation of an excision-resistant α-thionucleotide. The mutant 3′ end signal fragments were introduced into pAll plasmids and assayed for their effects on lacZ expression in yeast. Several of the plasmids carrying
mutant fragments supported lacZ expression at greater levels than the parental, "wild-type" plasmid, suggesting that the mutations interfered with 3′ end signal activity.

Those 3′ end signal fragments which suppressed lacZ expression from pAll plasmids were also capable of causing the synthesis of truncated cycl transcripts when inserted into a site within the CYCl coding sequence, indicating that they indeed retained 3′ end signal activity. Those fragments which allowed elevated lacZ expression from pAll plasmids did not cause production of truncated cycl transcripts from analogous constructs. These observations confirmed that lacZ expression from a pAll plasmid could be a useful screening device for 3′ end signal defects. Some 3′-extended transcripts were produced from all of the truncated genes suggesting that maximally efficient 3′ end generation required sequences outside the region defined by deletion analysis.

Five of six mutations tested which introduced GC base pairs into the 3′ end signal impaired its activity, suggesting that a high overall AT content may be an important feature of the 3′ end signal.

"Terminator" sequences recognized by Zaret and Sherman in 1982 [Cell 28: 563-573] and Henikoff and colleagues in 1983 [Cell 33: 607-614] indeed seem to be involved in 3′ end generation in CYCl transcripts, although neither completely describes the sequence requirements of the 3′ end signal.
## TABLE of CONTENTS

**ABSTRACT**  
ii  
**LIST of TABLES**  
xi  
**LIST of FIGURES**  
xii  
**ACKNOWLEDGEMENTS**  
xv  
**LIST of ABBREVIATIONS**  
xvi  

Chapter I. **INTRODUCTION**  
1  

**TRANSCRIPT 3’ END GENERATION: PERSPECTIVES**  
1  
Why Have 3’ End Signals?  
2  

**TRANSCRIPT 3’ END GENERATION in PROCARYOTES**  
4  
Factor-Independent Transcript Termination  
4  
Factor-Dependent Transcript Termination  
5  
3’ End Processing in Procaryotes  
7  
Transcript Termination as a Regulatory Device in Procaryotes  
7  
Rho-Dependent Transcript Termination and  
Translational Polarity  
7  
Antitermination in Bacteriophage Lambda  
8  
Antitermination in rRNA Operons  
10  
Attenuation in Bacterial Biosynthetic Operons  
11  

**TRANSCRIPT 3’ END GENERATION in EUKARYOTES**  
12  
Large Ribosomal RNA Genes  
12  
Small Ribosomal and Transfer RNA Genes  
15  
Evidence for Transcript Termination  
15  
Termination Signals for RNA Polymerase III  
16  
5S Transcript 3’ End Generation in Other Systems  
19  
3’ End Generation in tRNA Transcription  
19  
Genes Transcribed by RNA Polymerase II  
23
3'-Terminal Poly(A), and the Origin of Eucaryotic mRNA

Transcript Termination vs. 3' End Processing

Termination Sites for RNA Polymerase II

Transcription

Termination Signals for RNA Polymerase II

Polyadenylation Signals in "Higher" Eucaryotes

Polyadenylation in Cell-Free Systems

Nonpolyadenylated mRNA

Signals Required for Histone mRNA 3' End Generation

Evidence for 3' End Processing of Histone Gene Transcripts

Histone Transcript Termination

The Function of Poly(A) in Eucaryotic mRNA

3' End Generation as a Regulatory Device in Eucaryotes

Control Over Transcript Termination

Control Over Polyadenylation Site Selection

Poly(A) Sequences in Yeast mRNA

Chapter II. MATERIALS AND METHODS

REAGENTS

Enzymes

Oligonucleotides

Nucleotides

Galactosides

Phenol

Glyoxal
Preparation of M13 RF
Preparation of Single-Stranded pA4 Plasmid DNA
Silanization of Containers for Use with Nucleic Acid Solutions

**AGAROSE GEL ELECTROPHORESIS of DNA**
Electrophoresis and Fragment Purification Using IMP Agarose

**ACRYLAMIDE GEL ELECTROPHORESIS**
Autoradiography

**RESTRICTION ENDONUCLEASE DIGESTION**

**PHOSPHATASE TREATMENT of DNA**

**LIGATIONS**
Linker Ligations
Assaying Linker Ligation and Removal

**PLASMID CONSTRUCTIONS**
Construction of CYClA{H5}′ Deletions
Construction of YEp213CYClA{H5}′ Plasmids
Construction of CYClA{K3}′ and CYClA{K5}′ Deletions
Construction of YEp13CYClA{K3}′ Plasmids
Determination of ΔK3′ Deletion Endpoints
Construction of pYeCYClA{K5′}/ΔK3′ Plasmids (Promoter/3′ End Signal Fusions)
Construction of Deletions in mp11CYClT2:
The pA4 Plasmids
Construction of placZ
Construction of YRp72, YRp73
Construction of YEp73
Construction of pA5s 141
Construction of pA6 Plasmids 143
Construction of pA7 Plasmids 145
Construction of pA10 Plasmids 145
Construction of pA11 Plasmids 145
Construction of mpl0Al 147
Construction of pA11s Carrying Point Mutations in 3' End Signal 149
Construction of pA12, pA12A, and pA12B Plasmids 150

OLIGONUCLEOTIDE-DIRECTED MUTAGENESIS 154
Screening for Mutations 157
Transfer of Oligonucleotide-Mutagenized CYCl Genes to Yeast Plasmids 161
Sequence Confirmation of Oligonucleotide-Directed Mutations in CYCl 162

SEGMENT-DIRECTED MUTAGENESIS 163
Annealing 163
Limited Primer Extension on a Gapped Duplex 165
Misincorporation 166
Gap Repair 167

DNA SEQUENCING 168
Chain Terminator Method 168
Primer:Template Annealing 168
Single-Stranded Template 168
Double-Stranded Template 169
Sequencing Reactions 169
Sequencing by Base-Specific Chemical Cleavage 171
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TABLE I</td>
<td>LIST OF OLIGONUCLEOTIDES</td>
<td>96</td>
</tr>
<tr>
<td>TABLE II</td>
<td>PLASMIDS AND BACTERIOPHAGE VECTORS</td>
<td>103</td>
</tr>
<tr>
<td>TABLE III</td>
<td>BUFFERS FOR RESTRICTION ENDONUCLEASE DIGESTION</td>
<td>121</td>
</tr>
<tr>
<td>TABLE IV</td>
<td>SUMMARY OF OLIGONUCLEOTIDE-DIRECTED MUTAGENESIS EXPERIMENTS</td>
<td>156</td>
</tr>
<tr>
<td>TABLE V</td>
<td>COMPOSITION OF ddNTP/dNTP TERMINATOR MIXES</td>
<td>170</td>
</tr>
<tr>
<td>TABLE VI</td>
<td>RESULTS OF MISINCORPORATION MUTAGENESIS EXPERIMENTS</td>
<td>234</td>
</tr>
<tr>
<td>TABLE VII</td>
<td>COMPARISON OF THE ZARET AND SHERMAN CONSENSUS SEQUENCE TO MUTANT 3' END SIGNALS</td>
<td>254</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1. Construction of ΔH5' Deletions 126
Figure 2. Construction of CYC1 ΔK5' and CYC1 ΔK3' Deletions 129
Figure 3. Construction of the pA4 Promoter:3' End Signal Fusions 135
Figure 4. Restriction Maps of pMC1403 and placZ 137
Figure 5. Restriction Maps of YRp73 and YEp73 140
Figure 6. Structure and Derivation of pA5 Plasmids 142
Figure 7. Restriction Maps of pA6a, pA6, pA7, and pA10 144
Figure 8. Structure of the pA11 Plasmids 146
Figure 9. Construction of mp10Al 148
Figure 10. Reconstruction of pA11 Plasmids Carrying Point Mutations in the CYC1 3' End Signal 151
Figure 11. A. Construction of pAl2 Plasmids 153
     B. Orientation of the Truncated cycl Gene in the pAl2A and pAl2B Plasmids 153
Figure 12. Oligonucleotide-Directed Mutagenesis 158
Figure 13. Outline of Segment-Directed Mutagenesis 164
Figure 14. Hybridization of a CYC1 Probe to Total Yeast RNA 185
Figure 15. Restriction Map of CYC1 and Flanking Regions 186
Figure 16. Hybridization of a CYC1 Probe to RNA from Strains Bearing Deletions Distal to CYC1 190
Figure 17. Endpoints of the CYC1ΔH5' and CYC1ΔK3' Deletions 193
Figure 18. Hybridization of a CYC1 Probe to RNA from Transformants of GM-3C-2 Carrying CYC1ΔH5' Genes 195
Figure 19. Hybridization of CYC1 Probe to RNA from GM-3C-2 Transformants Carrying CYC1A3' Genes

Figure 20. Sequence of the CYC1 3' End Signal and Flanking Regions

Figure 21. Hybridization of a CYC1 Probe to RNA from GM-3C-2 Transformants Carrying YEpl3CYC1(2.5) (WT), YEpl3CYC1GG462, YEpl3CYC1C482, or YEpl3CYC1GG462C482

Figure 22. Strategy for Screening for Mutations in the 3' End Signal of CYC1

Figure 23. A. Sequence in the Vicinity of the GT473 and C474 Mutations, and the Corresponding Wild-Type Sequence

B. Hybridization of a CYC1 Probe to RNA from GM-3C-2 Transformants Carrying YEpl3CYC1(2.5) (WT), YEpl3CYC1GT473, or YEpl3CYC1C474

Figure 24. Deletion Endpoints in the pA4 Promoter:3' End Signal Fusions, for use in Promoter: 3' End Signal: lacZ Plasmids

Figure 25. Level of β-Galactosidase Produced by GM-3C-2 Transformants Carrying pAll Plasmids

Figure 26. Annealing of mpl0A1 and mpl0/SmaI

Figure 27. Limited Primer Extension on a Gapped Heteroduplex

Figure 28. Single-Track (T) Sequences of 12 Clones Obtained in Experiment T1

Figure 29. Sequences of 3' End Signal Regions of Various pAll Plasmids
Figure 30. Levels of β-Galactosidase Activity in Yeast Transformants Carrying pAll Plasmids with 3' End Signal Mutations 243

Figure 31. Hybridization of a CYCl Probe to RNA from Yeast Transformants Carrying pAl2A and pAl2B Plasmids 246
I would like to thank my supervisor, Dr. Michael Smith, for providing me with a stimulating environment in which to work, for encouraging me to pursue my ideas, and for giving them sober second thought. I am also grateful for his patience during the rather protracted gestation period of my thesis. The members of my advisory committee, Drs. Pat Dennis and Caroline Astell, provided useful criticism at various stages of the work, and for this, I thank them. Dr. Steve McKnight served as a source of technical advice and continues to be a source of inspiration. I owe a special debt of gratitude to Caroline Beard, who performed some of the experiments described in this thesis and has been a sounding-board for ideas and a fount of technical information. I thank, though I cannot list all of them, the friends and colleagues who made my time in Vancouver the happiest imaginable. In particular, I would like to thank my fellow student, Susan Porter, for her very important friendship.

My sister, Janet Henrikson, very kindly undertook to type much of the written draft of my thesis, at considerable risk to her eyesight and sense of order. I am also grateful to Sharon Berg for typing part of the introduction. Dr. Tom Hatton deserves acknowledgement for having rescued my hapless thesis from a capricious computer on several occasions.

I am of course grateful to the Natural Sciences and Engineering Research Council of Canada for its support, in the form of a 1967 Science Scholarship.

Finally, and most of all, I thank my wife, Donna Henrikson, for enduring, and for continuing to love me and believe in me.
LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Adenosine</td>
</tr>
<tr>
<td>A$_{260}$</td>
<td>Absorbance at 260 nm</td>
</tr>
<tr>
<td>Ap</td>
<td>Ampicillin</td>
</tr>
<tr>
<td>$\theta$-gal</td>
<td>$\theta$-galactosidase</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>C</td>
<td>Cytidine</td>
</tr>
<tr>
<td>cc</td>
<td>closed circular (DNA)</td>
</tr>
<tr>
<td>ddNTP</td>
<td>2',3'-dideoxynucleoside triphosphate (nucleoside may be specified as A,C,G, or T)</td>
</tr>
<tr>
<td>DEP</td>
<td>diethyl pyrocarbonate</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>ds</td>
<td>double-stranded (DNA)</td>
</tr>
<tr>
<td>dNTP</td>
<td>2'-deoxynucleoside triphosphate (nucleoside may be specified as A,C,G, or T)</td>
</tr>
<tr>
<td>DTT</td>
<td>1,4-dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EGTA</td>
<td>ethyleneglycol-bis-((\theta)-aminoethyl ether) N,N,N',N'-tetraacetic acid</td>
</tr>
<tr>
<td>G</td>
<td>guanosine</td>
</tr>
<tr>
<td>GART</td>
<td>glycinamide ribonucleotide transformylase</td>
</tr>
<tr>
<td>HEPES</td>
<td>N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl $\theta$-D-thiogalactoside</td>
</tr>
<tr>
<td>LaRNP</td>
<td>ribonucleoprotein bearing antigenic determinant La</td>
</tr>
<tr>
<td>LMP</td>
<td>low melting point (agarose)</td>
</tr>
<tr>
<td>MOPS</td>
<td>3-(N-morpholino)-2-hydroxypropanesulfonic acid</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>NTP</td>
<td>ribonucleoside triphosphate (nucleoside may be specified as A,C,G, or U)</td>
</tr>
<tr>
<td>OAc</td>
<td>acetate</td>
</tr>
<tr>
<td>oc</td>
<td>open circular (DNA)</td>
</tr>
<tr>
<td>OD&lt;sub&gt;600&lt;/sub&gt;</td>
<td>optical density at 600 nm</td>
</tr>
<tr>
<td>ONPG</td>
<td>o-nitrophenyl β-D-galactoside</td>
</tr>
<tr>
<td>ori</td>
<td>origin of replication</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethylsulfonylfluoride</td>
</tr>
<tr>
<td>RF</td>
<td>replicative form</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>Rot</td>
<td>the product of initial RNA concentration (Ro) and time of hybridization (t)</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>SC</td>
<td>yeast synthetic complete medium</td>
</tr>
<tr>
<td>SC-X</td>
<td>yeast synthetic complete medium, lacking growth factor X</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecylsulfate</td>
</tr>
<tr>
<td>snRNP</td>
<td>small nuclear ribonucleoprotein</td>
</tr>
<tr>
<td>ss</td>
<td>single-stranded (DNA)</td>
</tr>
<tr>
<td>SSC</td>
<td>standard saline citrate (0.15 M NaCl, 0.015 M sodium citrate)</td>
</tr>
<tr>
<td>T</td>
<td>thymidine</td>
</tr>
<tr>
<td>TBE</td>
<td>Tris:borate:EDTA electrophoresis buffer</td>
</tr>
<tr>
<td>TE</td>
<td>Tris:EDTA (concentrations specified)</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N',N'-tetramethylethylenediamine</td>
</tr>
<tr>
<td>Tris</td>
<td>tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>-------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>U</td>
<td>uridine</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>VRC</td>
<td>vanadyl ribonucleoside complexes</td>
</tr>
<tr>
<td>WT</td>
<td>wild-type</td>
</tr>
<tr>
<td>XGAL</td>
<td>5-bromo-4-chloro-3-indolyl-β-D-galactoside</td>
</tr>
</tbody>
</table>

Please note: In figures showing nucleic acid sequences, the abbreviations for the nucleosides are used, and the phosphodiester linkages are omitted for clarity.
INTRODUCTION

TRANSCRIPT 3' END GENERATION: PERSPECTIVES

Benzer's (1955) analysis of genetic fine structure in the bacteriophage T4 portrayed the gene as a discrete functional element forming a contiguous segment of a linear array of mutational and recombinational sites. Although more recent work has shown that a gene product may be specified by several non-contiguous segments of the array (Berget et al., 1977; Klessig, 1977) and that more than one gene may occupy a given segment (Sanger et al., 1977b) it remains true that a gene is a functionally distinct subset of a structural continuum. The process of gene expression demands that signals of some sort identify the gene as a region which is to be transcribed, mark those portions of the transcript which are to be conserved through a variety of possible processing reactions, specify sites at which transcript modification reactions are to take place, and for the transcripts of most genes, determine what regions of the transcript are to be translated into protein. Since nucleotide sequence is the only intrinsic feature of the genome which distinguishes one region from another (Watson and Crick, 1953) all of these signals must ultimately reside in particular nucleotide sequences. The work to be described in this thesis represents an effort to identify the nucleotide sequences which constitute a signal for transcript 3' end generation in the yeast, Saccharomyces cerevisiae. Considerable effort has already been expended on understanding the signals and mechanisms involved in 3' end generation in both procaryotes and eucaryotes. Before reviewing the fruits of this effort, it is useful
to consider the importance of transcriptional 3′ end signals to the organization and function of the genome.

Why Have 3′ End Signals?

Promoter signals, which specify sites of transcript initiation and control the frequency of initiation, offer a potent way of controlling gene expression, but the possible roles of signals which specify sites of 3′ end generation are perhaps less readily appreciated. The minimum requirement that may be imagined for a functional transcript is that it contain all of the sequences which form a part of or encode the gene’s ultimate product. Since the uncontrolled synthesis of incomplete transcripts would in general interfere with gene expression and regulation, the transcriptional apparatus should have a low probability of randomly terminating transcription. On the other hand, 3′ end generation must take place if transcripts are to be released from their templates. At least in eucaryotes, transcript release is prerequisite for transcript function. Efficient transcript release would demand a signal-directed process of 3′ end generation, although it would not impose a requirement for site-specific 3′-end generation. For example, a 3′-end signal might simply establish a boundary beyond which 3′ end generation would be permitted to occur randomly, perhaps as a result of some modification in template organization or transcription complex structure.

A requirement for site-specific 3′ end generation might exist if sequences or structures at the 3′ terminus of a transcript could influence subsequent steps in its processing or its function. Either -2-
termination signals or 3′ end processing signals could serve the purpose of ensuring the release of functional transcripts with appropriate 3′-terminal sequences. A termination signal would at the same time define the downstream boundary of a transcriptional unit, thereby serving the additional purpose of limiting the domain of influence of a promoter and its associated regulatory sequences and allowing the independent control of genes located further downstream. A 3′ end processing signal could serve the same purpose only if for some reason transcribed sequences downstream of the processing site were non-functional and regulatory signals downstream of the processing site were not vulnerable to direct interference from polymerases "reading through" from the processing site. Such a processing signal would be "phenotypically equivalent to a terminator", as Birnstiel and colleagues put it in a recent review (Birnstiel et al., 1985).

A 3′ end signal which acted as a barrier to the expression of genetic information located downstream would offer a point at which the expression of that information could be controlled if the efficiency of the 3′ end signal could be regulated. Termination signals are widely used as regulatory genetic elements in procaryotes, and it is becoming apparent that 3′ end generation is also used as a regulatory device by eucaryotes. Differences in "procaryotic" and "eucaryotic" mechanisms of regulation at the level of 3′ end generation reflect differences in both genomic and cellular architecture.
TRANSCRIPT 3' END GENERATION IN PROCARYOTES

Factor-Independent Transcript Termination

Certain transcripts found in *E. coli* can be produced by RNA polymerase upon transcription of the appropriate template in vitro, in the absence of any other proteins. The 3' ends of such transcripts must be specified by signals which cause RNA polymerase to terminate transcription. These signals are commonly called factor-independent terminators, although their efficiency may be influenced by proteins other than RNA polymerase (e.g., see Rosenberg et al., 1975).

Transcripts produced by factor-independent terminators characteristically end in a series of template-encoded U residues, immediately preceded by a GC-rich sequence with hyphenated dyad symmetry. (See Rosenberg and Court, 1979, for examples.) Studies of the in vitro transcription of templates containing mutationally altered and synthetic terminators illustrate the importance of both of these sequence elements to factor-independent termination (Christie et al., 1981).

Several workers suggested that the GC-rich dyad might hinder the movement of RNA polymerase along its template, either by causing very stable base-pairing of transcript and template (Gilbert, 1976), or by allowing the formation of secondary structures in the transcript, the template, or both (Adhya and Gottesman, 1978). Mutations which reduce the point symmetry of the dyad interfere with termination, suggesting that secondary structure formation is important to the process (reviewed by Platt, 1981; Yanofsky, 1981). Studies of in vitro transcription using nucleotide analogs suggest that the important base-pairing interactions in the GC-rich dyad involve RNA, but not DNA (Farnham and Platt, 1982).
On the basis of the observation that rU-dA base pairs are unusually unstable, Martin and Tinoco (1980) proposed that weak pairing between the template and the 3′ terminal uridylylate tract of the transcript might facilitate factor-independent termination. Stabilizing transcript-template interactions in the vicinity of the oligouridylylate tract during in vitro transcription, whether by lowering the temperature, by incorporating base analogs into the transcript or template, or by reducing the number of U residues encoded by the template, interferes with termination (Farnham and Platt, 1980; 1982; reviewed by Platt, 1981). These studies support the idea that factor-independent termination depends in part on the lability of the interaction between the template and the 3′ end of the transcript.

Factor-Dependent Transcript Termination

In 1969, Roberts reported the discovery of a protein called rho factor which caused E. coli RNA polymerase to terminate transcription in vitro in response to specific signals in λ DNA which could not be recognized by the polymerase alone. Considerable genetic and biochemical evidence supports the involvement of rho in termination of transcription of many phage and cellular operons (reviewed by Adhya and Gottesman, 1978; Rosenberg and Court, 1979; Platt, 1981; Platt and Bear, 1983).

Rho protein is active as a hexamer of 46 Kd subunits (Finger and Richardson, 1979). It binds to single-stranded RNA with little sequence-specificity other than a preference for cytidylate-containing polymers (Richardson, 1982; Lowery and Richardson, 1977). Upon binding to RNA, rho exhibits nucleoside triphosphatase activity, preferentially hydrolysing ATP (Lowery-Goldhammer and Richardson, 1974). Both its...
RNA-binding and NTPase activities are necessary for rho-mediated transcript termination (Howard and DeCrombrugghe, 1976; Sharp and Platt, 1984).

Rho-mediated termination does not usually occur at only one site in a transcription unit, but rather at several sites, which may span a region of over one hundred nucleotides (Kupper et al., 1978; Morgan et al., 1983a). There is no obvious sequence homology between sites of rho-mediated termination (Platt, 1981; Morgan et al., 1985), yet the process is clearly sequence-dependent: deletion of a region of rho-dependent termination downstream of the trp operon causes readthrough into adjacent operons in vivo (Wu et al., 1981) and prevents activation of the rho NTPase in vitro (Sharp and Platt, 1984).

Morgan et al. (1983a,b; 1984) carried out a detailed analysis of in vitro transcription and termination in a fragment of λ DNA including the rho-dependent terminator, tRl. In the absence of rho, RNA polymerase paused at five sites corresponding to the sites of rho-dependent termination in the tRl region. Pausing also occurred at more promoter-proximal sites, although the addition of rho did not cause termination at the proximal pause sites. When randomly-paused transcription complexes were prepared and incubated with rho, only those transcripts longer than 290 nucleotides were released from the complexes. Ceruzzi et al. (1985) reported that transcripts of the tRl region which were longer than 290 nucleotides were preferentially bound by rho.

Morgan and colleagues (1985) have proposed that rho binds to certain regions in nascent transcripts, such as the region 290 nucleotides distal to the rightward promoter of λ, because these regions lack stable secondary structures for at least 60 nucleotides and contain cytidylate
residues. The authors compared the sequences of several regions of rho-dependent termination and in each case found a tract with these features. The same workers have suggested that once bound to the nascent transcript, rho is able to cause its release from a transcription complex paused at any point downstream. The manner in which rho causes transcript release is not understood, although a role for the rho NTPase seems likely, and genetic evidence suggests interactions between rho and the β subunit of RNA polymerase (Guarente and Beckwith, 1978).

3' End Processing in Procaryotes

Although it seems that most transcript 3' ends in bacteria are the direct products of transcript termination, there is evidence for 3' end processing. Certain transcripts of λ DNA bear short 3'-terminal tracts of adenylate residues which are not template-encoded (Smith and Hedgepeth, 1975; Rosenberg et al., 1975). Gopalakrishna et al. (1981) detected polyadenylate tracts in about 20% of the pulse-labelled RNA in E. coli, and other workers have since detected polyadenylated transcripts in a variety of eubacteria (Hussain et al., 1982; Crouch et al., 1983; Majumder and McFadden, 1984) and an archaeabacterium (Brown and Reeve, 1985). The polyadenylate tracts in procaryotes tend to be fairly short (less than 20 residues) and very unstable. No specific polyadenylated transcript has been characterized, and the significance of the processing reaction remains obscure.

Transcript Termination as a Regulatory Device in Procaryotes

Rho-Dependent Transcript Termination and Translational Polarity

Mutations which interfere with the translation of one gene often
exert an inhibitory "polar effect" on the expression of downstream genes in the same operon. In vitro transcription experiments revealed the existence of rho-dependent terminators within the lac and gal operons (De Crombrugghe et al., 1973). Richardson et al. (1975) and Adhya and Gottesman (1978) suggested that rho-dependent terminators residing within operons are normally masked by ribosomes engaged in translation, but that mutations which prevent translation allow rho to gain access to the nascent transcript and cause its premature release from paused transcription complexes. The discovery that suA mutations, which suppress polarity, cause the production of altered rho factor (Richardson et al., 1975; Ratner, 1976), substantiated the idea that polarity is a consequence of rho-dependent termination within operons. The existence of latent terminators within operons illustrates the importance of coupling between transcription and translation to gene expression in bacteria. The binding preferences of rho might be expected to have placed constraints on the length and structure of intergenic untranslated sequences within operons.

**Antitermination in Bacteriophage Lambda**

Transcript termination is a pivotal event in the reproductive cycle of bacteriophage λ. The product of gene N, pN, is required for the expression of all λ genes distal to the t_{L1} and t_{R1} terminators. It acts to prevent transcript termination at t_{L1} and t_{R1}, as Roberts proposed in 1969. Adhya et al. (1974) suggested that pN somehow converts RNA polymerase into a "transcriptional juggernaut", unable to respond to sequences which normally cause termination. The effect is not restricted to any particular terminator, but on first examination it seems to be
specific to transcription originating at the \( \lambda \) promoters, \( p_L \) and \( p_R \) (Adhya et al.,1974; Franklin, 1974). Salstrom and Szybalski (1978) identified a site between \( p_L \) and \( t_{L1} \) which is required in cis for pN-mediated antitermination, and which they called \textit{nutL}. A homologous \textit{nutR} site exists between \( p_R \) and \( t_{R1} \) (Rosenberg et al.,1978). The presence of a \textit{nut} site within a transcription unit is both necessary and sufficient for pN to act as an antiterminator of transcription at terminators distal to the \textit{nut} site (de Crombrugghe et al.,1979).

Friedman (1971) undertook to define the proteins involved in interactions with pN by selecting mutants of \textit{E. coli} which failed to support N-dependent growth of \( \lambda \). Several genes are now identified by \textit{nus} (N-utilization substance) mutations. Among them are the genes encoding rho protein (\textit{nusD}), the \( \sigma \)-subunit of RNA polymerase (\textit{nusC}), and ribosomal protein S10 (\textit{nusE}) (Friedman and Gottesman, 1983). The \textit{nusA} gene originally identified by Friedman encodes a protein which binds tightly and specifically to both pN and RNA polymerase core enzyme (Greenblatt and Li, 1981a,b). The \textit{nusA} protein is displaced from RNA polymerase by the \( \sigma \) subunit of the polymerase (Greenblatt and Li, 1981b). NusA protein causes an increase in the length of time for which RNA polymerase pauses at some terminators, suggesting a role in termination at these sites (Greenblatt et al.,1981; Farnham et al.,1982). A binding site for \textit{nusA} has been identified in the vicinity of the \textit{nut} sites of lambdoid phages and in those cellular operons in which \textit{nusA} is involved in termination (Olson et al.,1982; Friedman and Olson, 1983).

Cell-free extracts which support pN-mediated antitermination during transcription of added DNA templates promise to be useful in deciphering the interactions of pN and the various \textit{nus} products during
antitermination (Ghosh and Das, 1984; Das and Wolska, 1984; Goda and Greenblatt, 1985).

Expression of the late genes of λ depends not only upon pN-mediated antitermination, but also upon the product of gene Q, pQ. The coincidence of the genetically-defined site of action of pQ and the rho-independent "λ 6S" terminator led to the suggestion that Q, like N, encodes an antiterminator (Roberts, 1975). Forbes and Herskowitz (1982) demonstrated that pQ, like pN, can suppress the polar effects of mutations lying distal to its site of action. Grayhack and Roberts (1982) showed that purified pQ causes RNA polymerase to transcribe through the λ 6S terminator in vitro in the absence of other proteins. Comparison of the mechanisms of action of pQ and pN promises to be very interesting in view of their apparently quite different requirements.

Antitermination in rRNA Operons

Antitermination of transcription is not the sole property of bacteriophage λ and its cohort. Holben and Morgan (1984) reported that mutations which are strongly polar to transcription from the lac or ara promoters have no polar effect on transcription originating from the rrnC promoter. Li et al. (1984) found that a 67-base pair fragment from the rrnG leader region rendered transcription from any of three promoters termination-resistant. The active fragment contained sequences homologous to the nusA-binding sites and nut sites of the lambdoid phages. Characterization of the trans-acting factors involved in antitermination of rrn transcription promises to be interesting from both a biochemical and an evolutionary point of view.
Attenuation in Bacterial Biosynthetic Operons

Many biosynthetic operons in E. coli and related bacteria are regulated by means of transcript termination signals called attenuators, which are located upstream of the structural genes in their respective operons (Kolter and Yanofsky, 1982). The archetypal attenuator is that of the trp operon of E. coli, which has been intensively studied by Yanofsky and his colleagues. The fruit of their efforts is a detailed model of attenuation which was most eloquently reviewed by Yanofsky in 1981.

The transcript of the trp attenuator region contains several different potential hairpin structures. One of these is a typical rho-independent terminator, the attenuator itself. A second hairpin can form shortly upstream of the attenuator, and its formation precludes the formation of the attenuator hairpin. The efficiency of termination in the attenuator region depends upon which of these secondary structures forms each time the region is transcribed. Regulation of trp expression is achieved by regulating the formation of the two alternative structures.

The attenuator lies within a short open reading frame which encodes a 14-amino acid peptide containing two adjacent tryptophan residues. While transcription is underway, ribosomes bind to the nascent transcript and begin translation of the short open reading frame. If tryptophanyl-tRNA_{Trp} is in short supply, the ribosome stalls over the tandem Trp codons. In this position, it sterically prevents the formation of the attenuator hairpin but allows the competing hairpin to form. Transcription consequently continues into the structural genes of the operon. If on the contrary the cell is well-supplied with
tryptophan, ribosomes translate the entire leader peptide coding sequence. Formation of the attenuator hairpin is allowed, and transcription terminates without reaching the structural genes of the operon. This model of attenuation is supported in detail by a wealth of genetic evidence, for which the reader is directed to the review by Yanofsky (1981) and the references therein.

TRANSCRIPT 3' END GENERATION in EUCARYOTES

The nuclei of eucaryotic organisms contain three types of DNA-dependent RNA polymerase. Each is responsible for the transcription of a particular class of genes. RNA polymerase I produces precursors containing the sequences of 18S, 28S, and 5.8S rRNAs, while RNA polymerase II is responsible for the synthesis of mRNAs or their precursors, and RNA polymerase III transcribes genes encoding 5S RNA and tRNA. [For a review, see Roeder, 1976; Sentenac and Hall (1981) review the characteristics of yeast nuclear RNA polymerases.] The process of 3' end generation will be described separately for each class of transcripts.

Large Ribosomal RNA Genes

Genes encoding ribosomal RNAs are amplified to varying degrees in eucaryotes. The oocytes of amphibians have been particularly favourable subjects for the study of ribosomal RNA genes because the genes are highly reiterated and because cloned genes, including variants created by
in vitro manipulation, may be introduced into the cells by microinjection for analysis of their expression. Active ribosomal RNA transcription units from *Xenopus* oocytes have been directly observed in the electron microscope (Miller and Beatty, 1969). Nascent transcripts are visible as fibrils projecting radially from an axial fibre, which is the DNA template. The fibrils are clustered in very sharply-defined regions along the axial fibre, these regions defining the extent of the ribosomal RNA transcription units. Neighbouring transcription units are separated by regions which are completely free of radial fibrils and are therefore non-transcribed. Within each transcription unit, the fibrils continually increase in length from one end of the unit to the other. The short-fibril end of the transcription unit must represent the site of transcript initiation, while the long-fibril end corresponds to the site of transcript termination. The termination site has been located on the physical map of the rDNA repeat unit by digesting active transcription units with restriction endonucleases prior to visualization in the electron microscope. The location of the termination site corresponds, within the resolution of measurements made on electron micrographs (about 200 base pairs), to the location of the 3' end of the 40S rRNA precursor, which was accurately determined by S1 nuclease digestion of RNA-DNA hybrids (Sollner-Webb and Reeder, 1979). The mature 28S rRNA happens to share exactly the same 3' end as the 40S precursor, and it appears that this end is generated by termination of transcription, although rapid processing of a short 3'-terminal extension has not been formally excluded.

The 3' end of the 40S pre-rRNA of *Xenopus* maps immediately upstream
of a cluster of 4 T/A base-pairs. Bakken et al. (1982) provided evidence that this T/A cluster comprises at least part of the termination signal for transcription by RNA polymerase I. They injected Xenopus oocytes with plasmids carrying various rDNA fragments and mapped the extent of the polymerase I transcription unit on each plasmid using the electron microscope. They found that deletion of the last base-pair in the \((T/A)_4\) cluster and all distal rDNA sequences did not prevent transcription from terminating at the normal site. Deletion of the last two T/A base pairs, however, caused transcription to proceed beyond the normal 3' end site. These results located the distal boundary of the termination signal, but the position of its proximal boundary was not determined. The T/A cluster is preceded by a sequence having hyphenated dyad symmetry, which means that the termination signal for RNA polymerase I bears at least a superficial resemblance to factor-independent terminator signals in procaryotes.

The ribosomal DNA repeat unit in yeast is tandemly repeated about 120 times on chromosome XII. A single repeat unit includes one transcription unit encoding the 18S, 5.8S, and 25S rRNAs linked to a divergently transcribed 5S rRNA gene (reviewed by Warner, 1982). Klootwijk et al. (1979) argued that the 37S ribosomal rRNA precursor of yeast is a primary transcript because it is the largest pre-rRNA detectable in normal cells and its 5' end carries the hallmark of initiation, a nucleoside triphosphate. The authors isolated a unique 3' terminal oligonucleotide from the 37S molecule and showed that its sequence resembled those found in amphibian rRNA termination sites and in factor-independent terminators from bacteria. They suggested on these
grounds that the 3' end of the 37S pre-rRNA is the product of transcript termination. The 3' end is located exactly 7 nucleotides beyond the 3' end of the mature 26S rRNA (cited in Warner, 1982).

**Small Ribosomal and Transfer RNA Genes**

**Evidence for Transcript Termination**

Attempts to produce cell-free systems which support polymerase III-mediated transcription met with success earlier than similar efforts to make polymerases I and II function properly in vitro. Such systems proved to be powerful tools for analyzing the signals and interactions involved in transcription of 5S and tRNA genes. Brown and Gurdon (1977, 1978) demonstrated that *Xenopus* 5S RNA genes introduced into *Xenopus* oocytes as high molecular weight DNA or in recombinant plasmids were transcribed to produce authentic 5S RNA. Birkenmeier, Brown and Jordan (1978) and Ng, Parker and Roeder (1979) subsequently prepared extracts of *Xenopus* oocytes which carried out the accurate transcription of cloned 5S RNA genes. In each case the 3' end of 5S RNA mapped to a cluster of 4 T/A base pairs in the DNA, which if transcribed encoded uridylate residues. The majority of transcripts ended with two such residues, but some heterogeneity was observed. In addition to authentic 5S RNA, a related species bearing a 3' terminal extension was occasionally produced in the extracts, or in intact oocytes (Brown and Gurdon, 1978; Birkenmeier et al., 1978). In the cell-free extract, both 5S and the extended RNA accumulated at constant rates after an initial lag, which could be eliminated by preincubation of the extract with the DNA template in the absence of one nucleoside triphosphate. These results argued
against a precursor-product relationship for the extended molecule and 5S RNA and suggested instead that both were primary transcripts of the 5S gene. The 3' end of mature 5S RNA was held to be the product of site-specific termination of transcription, and the extended RNA was thought to be produced by polymerases which read through the normal termination site and terminated shortly downstream at a second cluster of 4 T/A base pairs in the DNA.

The possibility that rapid processing rather than termination gave rise to the observed 3' ends was effectively laid to rest by the results of Cozzarelli et al. (1983), who found that pure RNA polymerase III from X. laevis accurately terminated transcription of cloned 5S RNA genes in the absence of other proteins. Transcription was not initiated specifically at the beginning of the 5S gene in this system, but it was still possible to assay for correct termination by hybridizing the transcripts to a DNA probe which extended beyond the 3' end of the 5S coding sequence from a site within the gene, and digesting the hybrids with RNase T₁. The length of the RNA fragments protected from digestion provided a direct measure of the distance between the 3' ends of the transcripts and a fixed, well-defined site within the gene. Termination at the normal site immediately downstream of the 5S sequences and at sites further downstream occurred with the same efficiency as in the crude oocyte extract.

**Termination Signals for RNA Polymerase III**

Brown and his colleagues have analyzed in considerable detail the sequences that are required to bring about termination of transcription
by RNA polymerase III. In 1981, Bogenhagen, Sakonju, and Brown reported that a Xenopus 5S RNA gene lacking all but 4 nucleotides of its normal 3' flanking sequence could still direct the synthesis of authentic 5S RNA in a cell-free extract. On the other hand, a deletion which extended 9 base pairs further upstream, removing the last 5 nucleotides of the 5S coding sequence as well as the 3' flanking sequences, completely abolished termination at the normal site. These results formally demonstrate that the distal boundary of sequences required for transcript termination resides in the 9 base pair sequence GGCTTTTGC which spans the termination site.

The 5S gene and its normal flanking sequences directed the synthesis of an extended transcript, as well as authentic 5S RNA, in the oocyte nuclear extract. This transcript terminated 8 nucleotides beyond the main 5S 3' termination site, in a cluster of 4 U residues. Deletion of sequences downstream of the corresponding T/A tract in the DNA, to within 2 base pairs of the tract itself, did not affect synthesis of the extended transcript, but removal of 8 additional base pairs including the T/A cluster prevented the appearance of the extended transcript. The boundary region for the termination signal that gives rise to the extended transcript has the sequence ACTTTTGC. Taken together these results suggested that an important component of an RNA polymerase III termination signal is a cluster of T/A base pairs located at the site of termination itself. The possibility that sequences upstream of the T cluster also form part of the termination signal had not been excluded at this stage.

Bogenhagen and Brown (1981) confirmed the importance of the T
cluster to termination of 5S gene transcription. They showed that a gene which retained the T cluster at the main termination site and only 2 base pairs of normal downstream flanking sequence gave rise to normal 5S RNA, but a derivative lacking the last 3 Ts of the cluster could not support termination. They also localized the upstream boundary of the termination signal by joining fragments which included the T cluster, and varying lengths of upstream flanking sequence to fragments containing the signals required for transcript initiation. A fragment containing only 14 base pairs upstream of the T cluster caused termination within the cluster when joined to the first 83-115 base pairs of the 5S gene, suggesting that sequences 14-36 base pairs upstream of the T cluster are not required for termination. If sequences further upstream are involved in termination, it seems that their distance from the T cluster is of no consequence. Fragments containing only 3 or 4 base pairs upstream of the T cluster were joined to the first 97 base pairs of the 5S gene and found to cause termination within the T cluster. The efficiency of termination was reduced in constructs retaining only 3 base pairs upstream of the T cluster, which led the authors to conclude that the immediate context of the T cluster influenced the efficiency with which it could be recognized. Any cluster of 4 or more T/A base pairs (T residues in the sense DNA strand) flanked by GC-rich sequences was capable of functioning as a signal for termination of polIII transcription. Clusters of T/A base pairs flanked by a preponderance of A/T base pairs functioned inefficiently or not at all as termination signals.
5S Transcript 3′ End Generation in Other Systems

Although the 3′ end of the predominant primary transcript of Xenopus 5S genes coincides with that of mature RNA, precursors of 5S RNA bearing 3′ terminal extensions have been detected in other organisms. Hamada et al. (1979) found that isolated nuclei of rat liver or HeLa cells produce, in addition to authentic 5S RNA, a molecule with 8 extra nucleotides at its 3′ end. The extended 5S RNA was also detected by pulse-labelling cultured cells, and it was shown to give rise to mature 5S RNA during a chase period.

Tekamp et al. (1980) detected a similar 3′-extended "5S" RNA after transcription of a fraction of yeast chromatin enriched in ribosomal RNA genes with exogenous RNA polymerase III. In this case, the 3′ termini of the extended transcripts were slightly heterogeneous, spanning a 7 nucleotide region. The extended transcripts could be converted to authentic 5S RNA by a processing activity present in the supernatant of the chromatin fraction. The sequence of the yeast 5S gene suggests that the 3′ end of the 5S precursor is produced by transcript termination in response to the same signal as that recognized by Xenopus RNA polymerase III. The 3′ end of the precursor maps precisely to a cluster of 8 consecutive T/A base pairs in the DNA.

3′ End Generation in tRNA Gene Transcription

Genes encoding transfer RNAs are also accurately transcribed upon injection into Xenopus oocytes or incubation with cell-free extracts derived from various cell types. DeRobertis and Olson (1979) reported that a yeast tRNA^TYR gene is transcribed in Xenopus oocytes to produce a
molecule which comigrates with mature tRNA\textsuperscript{tyr} upon gel electrophoresis and a number of larger molecules. The time course of the reaction showed these molecules to be precursors of the mature species. A transcript of 108 nucleotides was identified as the primary transcript because it was the first and largest precursor to be synthesized and because it could be labelled with $\gamma^3\text{P}$ATP. The authors did not precisely locate the 3' end of this molecule, but it could have been no more than a few nucleotides beyond the tRNA sequence, and in this region a cluster of 6 T/A base pairs is present in the DNA. A similar or identical tRNA\textsuperscript{tyr}-related species has been identified in yeast cells which are defective in processing various precursor RNAs (Hopper and Kurjan, 1981), suggesting that RNA pol III from yeast recognizes the same termination signal as the Xenopus enzyme. Standring et al. (1981) made similar observations concerning the transcription of a yeast tRNA\textsuperscript{leu} gene in a HeLa cell extract. In this case the molecule presumed to be the primary transcript had heterogeneous 3' termini spanning a region of 6 nucleotides, 4-9 nucleotides distal to the tRNA sequence, and corresponding precisely to a series of 6 consecutive T/A base pairs downstream of the gene.

Further evidence that a cluster of T/A base pairs provides the signal for tRNA gene transcript termination has come from studies on the transcription of mutant yeast tRNA\textsuperscript{tyr} genes in Xenopus extracts. Koski et al. (1980) observed that two different mutations in the SUP4 gene caused the synthesis of transcripts which were considerably shorter than the normal tRNA\textsuperscript{tyr} precursor. Neither short transcript seemed to be the product of processing of a larger molecule. Sequence analysis of the
shortened transcripts indicated that both initiated at the normal site and terminated heterogeneously near the sites of the mutations in their respective genes. One mutation created a cluster of 6 consecutive T residues in the transcript-isoparallel strand of the gene, 40-45 base pairs downstream of the initiation site, and the other produced a T<sub>5</sub> cluster slightly further downstream. Termination at the T<sub>6</sub> cluster in the first mutant appeared to be about 80% efficient, while the T<sub>5</sub> cluster functioned as a termination signal with about 50% efficiency. The sequences flanking each cluster were not particularly GC-rich, which suggests that perhaps a cluster of 5 or 6 consecutive T/A base pairs is effective as a pol III termination signal in any context. Alternatively, terminator signal recognition in tRNA gene transcription may differ slightly from that involved in 5S gene expression. It is interesting to note that two mutations which created clusters of 4 T/A base pairs within the SUP4 gene did not cause premature termination, providing support for the notion that the length of a T cluster is crucial to its recognition as a terminator.

The transcription of the same mutant SUP4 genes in an extract of yeast cells has been analyzed more recently by Koski et al. (1982). Mutations which created clusters of 5 or 6 consecutive T residues within the gene cause transcription in the yeast extract to terminate upon reaching these clusters, just as had been observed in the Xenopus extract. In the same study, the authors showed that deletion of the T cluster immediately downstream of the coding sequence abolished the synthesis of the normal tRNA<sup>Tyr</sup> precursor. Termination occurred instead at the next cluster of T residues. This result adds force to the idea
that the mechanism of transcript termination by RNA polymerase III has been widely conserved amongst eucaryotes. A cautionary note must be added in light of the results of Hopper and Kurjan (1981), who failed to detect prematurely terminated transcripts of the same mutant SUP4 genes in yeast cells. A possible explanation for the discrepancy between results obtained in vivo and in vitro is that the prematurely terminated transcripts were rapidly degraded in yeast.

Further evidence in support of the idea that RNA polymerase III from any eucaryote recognizes the same type of termination signal was provided by Watson et al. (1984). They found that the RNA polymerase III in a preparation containing all 3 nuclear RNA polymerases from calf thymus terminated transcription of PvuII-linearized SV40 DNA specifically in a cluster of 8 consecutive T/A base pairs.

In mammalian cells, the primary transcripts of pol III genes are generally associated with a 50 Kd protein bearing an antigenic determinant known as La. Stefano (1984) purified the La-antigenic protein from HeLa cells and tested its ability to bind a variety of model pol III transcripts constructed by ligating the 3' end of yeast tRNA^Phe to various oligoribonucleotides. Model substrates bearing oligouridylate stretches of 2-6 nucleotides at the 3' end all bound to the protein if it was present in sufficient excess, but binding reactions carried out with lower quantities of protein showed that it preferentially bound substrates with 4 or more 3'-terminal uridylate residues. Protein-RNA complexes involving RNAs with 4-6 3'-terminal uridylate residues were somewhat more stable to salt and heparin than those involving RNAs with shorter oligouridylate tails. RNAs with tails ending in a 3'-OH formed
much more stable complexes than those bearing a 3′ phosphate group. Model substrates with tails of other nucleotides bound La antigen less efficiently than analogous oligo(U)-tailed substrates, and (U)₅OH dissociated complexes containing substrates with non-U tails more easily than those containing oligo(U)-tailed substrates. Normal termination by pol III produces primary transcripts bearing 3′-terminal oligo(U) tracts, and Stefano’s results suggest that such tracts may be important in allowing antigen La to bind to the transcripts. Stefano suggested that binding of La protein might be required for the processing or nuclear localization of pol III transcripts. It might also afford a means of protecting authentic primary transcripts from degradation. In any event, the apparent affinity of the 3′ ends of pol III transcripts for La antigen suggests that correct termination of transcription is of some consequence, not only in terms of defining boundaries of expression in the genome, but also for metabolism of the immediate products of gene expression.

**Genes Transcribed by RNA Polymerase II**

**3′-Terminal Poly(A), and the Origin of Eucaryotic mRNA**

Messenger RNA in eucaryotic cells was first recognized, by analogy with the properties of bacterial messengers, as an unstable class of polysome-associated RNA which could best be observed after brief labelling with radioactive RNA precursors, and which had a broad size distribution and a base composition like that of DNA (see Penman et al., 1963; Penman, Vesco, and Penman, 1968). Prior to the advent of cDNA cloning, the only individual mRNAs to be isolated were a few of those
produced in relative abundance by specific cell types, such as the globin mRNAs and, because of their unusually small size and characteristic pattern of synthesis, a few histone mRNAs (for review see Lewin, 1975). Most individual mRNAs are impossible to isolate without cloning because each constitutes a very small fraction of the total mRNA population in the cell. On the other hand, the complexity of the population as a whole makes its analysis complicated and means that the results of such an analysis do not necessarily describe the properties of any particular mRNA.

In 1971, three groups reported that messenger RNA isolated from human or mouse cells contained fairly long tracts of adenosine residues which could be detected by virtue of their resistance to the combined action of RNases A and T₁ (Edmonds, Vaughan and Nakagato, 1971; Lee, Mendecki and Brawerman, 1971; Darnell, Wall and Tushinski, 1971a). Similar poly(A) tracts were also found in the heterogeneous RNA of the nucleus (hnRNA).

Several experimental approaches led to the conclusion that poly(A) tracts in both mRNA and hnRNA are covalently attached to the 3’ ends of the molecule. Alkaline hydrolysis of isolated poly(A) released AMP and adenosine in the ratio of about 200:1, suggesting a 3’-terminal location and a length of about 200 residues for the poly(A) tract (Mendecki et al., 1972; Nakazato et al., 1973). The presence of a free 3’OH group on poly(A) isolated from either mRNA or hnRNA was confirmed by its susceptibility to digestion with nucleases specific for 3’OH termini (Molloy et al., 1972; Sheldon et al., 1972). Poly(A) isolated by digestion of hnRNA or mRNA with RNase T₁ was free of GMP residues, and
that released by RNase A contained no pyrimidines, suggesting that all cleavage sites for these nucleases resided upstream of the poly(A) (Molloy and Darnell, 1973). The rapid appearance of a [3H] adenosine label in cytoplasmic poly(A) suggested that poly(A) must be the last part of an mRNA molecule to be synthesized, which is also consistent with a 3′-terminal location (Mendecki et al., 1972; Sheldon et al., 1972).

The presence of 3′-terminal poly(A) tracts in mRNA and hnRNA added to a list of similarities between the two classes of RNA which collectively suggested that mRNA might be derived from hnRNA: both are unstable, they resemble total cellular DNA in their base composition, they exhibit broad size distributions, and the synthesis of both classes is inhibited by levels of actinomycin D about ten-fold higher than the levels required to inhibit ribosomal RNA synthesis. On the other hand, hnRNA is on average much larger than mRNA, and its sequence complexity exceeds that of the mRNA population, usually by a factor similar to the size difference (see Lewin, 1980, pp.728-750). The instability and high complexity of hnRNA implied that a large fraction of the hnRNA synthesized during any given time period suffered rapid degradation within the nucleus. This made it difficult to demonstrate that a significant fraction of a radioactive label in hnRNA found its way into mRNA. However, the kinetics of labelling of poly(A) with tritiated adenosine did suggest that poly(A) is synthesized in the nucleus and subsequently transported to the cytoplasm. Darnell et al. (1971b) observed that after a 15-minute period of labelling, most of the radioactive poly(A) was located in the nucleus and that the fraction present in the cytoplasm increased with the length of the labelling.
period. More detailed observations were made by Jelinek et al. (1973), who found that over 90% of the poly(A) labelled in a 1.5 minute period was located in the nucleus and that after labelling for 20 minutes or longer, the amount of nuclear poly(A) remained constant, while that in the cytoplasm increased linearly. The amount of poly(A) in the cytoplasm after a short period of labelling continued to increase for 30-40 minutes if further synthesis was inhibited with 3' deoxyadenosine triphosphate (cordycepin triphosphate). During the "cordycepin chase" the amount of nuclear poly(A) declined.

The same authors were unable to detect any free poly(A) in Hela cells, which suggested that poly(A) was not assembled prior to its attachment to RNA, but instead formed by the polymerization of adenylate residues onto an RNA primer.

Edmonds et al. (1971) had initially suggested that poly(A) was synthesized by template-directed transcription because its synthesis could be inhibited by actinomycin D. However, doses of actinomycin D sufficient to inhibit the synthesis of hnRNA and mRNA by over 90% had much less dramatic effects on the incorporation of labelled adenylate into poly(A), at least during short periods of drug treatment and labelling (Darnell et al., 1971b; Jelinek et al., 1973). Inhibition of poly(A) synthesis became more severe over longer periods, but this could have reflected the utilization or elimination of all available primers for poly(A) synthesis, rather than a direct effect on the synthetic process. Cordycepin triphosphate inhibited the incorporation of labelled adenosine into nuclear poly(A) over a 30 minute period by 90%, but during the same period, synthesis of hnRNA was inhibited by only 50%, again.
suggesting that poly(A) was synthesized by some process other than transcription.

Proof of a precursor-product relationship between hnRNA and mRNA required that the sequence of a particular mRNA be detected in a larger nuclear RNA, and that the mRNA be shown to arise from the larger species in a pulse-chase experiment. By hybridizing pulse-labelled nuclear RNA to a cDNA probe prepared against globin mRNA, Ross (1976) detected a 15S RNA containing globin sequences, as well as mature globin mRNA of about 10S. During a chase with unlabelled nucleotides, the labelled 15S material disappeared, and label accumulated in the 10S peak. Globin mRNA therefore seemed to be derived from a precursor about 3-fold larger than the message itself.

Goldberg et al. (1977) used the technique of UV transcription mapping to show that the average transcription unit in Hela cells is 4-7-fold larger than the average mRNA, suggesting that many mRNAs in mammalian cells might be derived from larger precursors.

**Transcript Termination vs. 3’ End Processing**

The demonstration that eucaryotic mRNAs may derive from larger precursors raises questions of whether the mRNAs and their precursors share the same 3’ termini and of whether these 3’ termini are generated by transcript termination or nucleolytic processing. The first question was answered in the affirmative for θ-globin mRNA and its 15S precursor, by Grosveld and colleagues (1981). They used the S1 nuclease mapping procedure of Berk and Sharp (1977) to show that the mRNA and its precursor were 3’-coterminal. The second question is more difficult to answer, because the failure to detect transcripts extending beyond a poly(A) site could reflect either transcript termination at that site or
the rapid degradation of distal portions of the transcript. Two approaches have been used to detect short-lived transcript sequences (reviewed by Darnell, 1982). The first employs very short pulse-labels of living cells. For this approach to be useful, the species of interest must be synthesized at high rates, to allow it to be detectably labelled in a short time interval. The second approach involves labelling nascent transcripts which have been initiated in vivo by allowing their elongation to continue in isolated nuclei in the presence of radioactive nucleoside triphosphates. This approach would not necessarily yield meaningful results if transcript elongation could continue indefinitely in isolated nuclei, because it might proceed beyond sites at which it would terminate in vivo. However, only a few hundred nucleotides are added to nascent transcripts after nuclear isolation, and new chains are not initiated (Weber et al, 1977). Therefore the distribution of labelled transcript ends should reflect their distribution prior to the isolation of nuclei, to within a few hundred nucleotides.

Viral transcripts are useful subjects for the study of transcription and RNA processing because in many cases they are synthesized at high rates during lytic infection. Darnell and his colleagues size-fractionated pulse-labelled nascent RNA synthesized late in adenovirus infection and tested the ability of each fraction to hybridize to various restriction fragments of adenovirus DNA. Their results indicated that most of the late adenovirus transcripts were derived from a single transcriptional unit, which began near map unit 16 and continued
at least as far as map unit 99, near the end of the genome (Weber et al., 1977). Five different sets of 3' coterminous mRNAs are produced from this transcription unit by polyadenylation at each of 5 different sites. Nevins and Darnell (1978) showed that restriction fragments including any of the 5 polyadenylation sites hybridized to equimolar amounts of pulse-labelled nuclear RNA from infected cells. In fact, transcription continued at the same level beyond the last polyadenylation site, which clearly indicated that none of the polyadenylation sites served to cause transcription to terminate.

Fraser et al. (1979a) provided corroborating evidence by showing that nonpolyadenylated nuclear transcripts contained oligonucleotides encoded downstream of the last polyadenylation site, although these oligonucleotides were absent from polyadenylated nuclear transcripts and cytoplasmic RNA. The sensitivity to UV irradiation of transcription downstream of the last poly(A) site suggested that this transcription did originate at the major late initiation site at map unit 16.4.

Transcription beyond polyadenylation sites is not a peculiarity of the adenovirus major late transcription unit. Nevins et al. (1980) showed that pulse-labelled RNA synthesized early in adenovirus infection included molecules derived from early transcription units 2 and 4 which extended beyond the respective poly(A) addition sites of those units. The size of the nascent transcripts and UV transcription mapping indicated that transcription beyond the poly(A) sites was not the result of termination and reinitiation, but rather was due to the extension of transcripts initiated at the promoters of early regions 2 and 4.

Ford and Hsu (1978) tested pulse-labelled RNA from SV40-infected
cells for its ability to hybridize with restriction fragments collectively representing the entire SV40 genome and found that transcription originating at the SV40 late initiation site regularly proceeded at least 1000 nucleotides beyond the 3' termini of the late SV40 mRNAs.

Transcription has since been shown to continue beyond the polyadenylation sites of several cellular transcription units. Hofer and Darnell (1981) labelled nascent transcripts in isolated nuclei of DMSO-induced erythroleukemia cells and tested for hybridization to cloned fragments of the \( \beta \)-major globin transcription unit. They found that the labelled transcripts hybridized in equimolar amounts to each of a series of restriction fragments spanning a region from the initiation site to 1 kb beyond the polyadenylation site. The hybridizing transcripts were all encoded by the sense DNA strand, and the minimum size of transcripts hybridizing to a particular DNA fragment reflected the distance of that fragment from the cap site, suggesting that all of the hybridizing transcripts originated at the cap site. Weintraub et al. (1981) reported that transcription of the \( \alpha \)-globin genes in chicken embryos proceed for several hundred nucleotides beyond the respective poly(A) sites, and similar observations regarding the \( \alpha \)-globin gene in erythroleukemia cells (Sheffery et al., 1984) and the rabbit \( \beta_1 \) globin gene in fetal liver cells (Rohrbaugh et al., 1985) have been published more recently. Hagenbuchle et al. (1984) found that transcription of the mouse \( \alpha \)-amylase gene Amy\( ^{2a} \) proceeded well beyond the polyadenylation site. In each of these studies, nascent transcripts labelled in isolated nuclei were hybridized to cloned fragments of genomic DNA.
Whether or not transcription proceeded beyond poly(A) sites in all genes was until recently a matter of controversy. Roop et al. (1980) had shown that the largest ovalbumin mRNA precursor detectable in steady-state RNA populations shared the same 3' end as the mature message, and Tsai et al. (1980) were unable to find evidence of transcription downstream of the poly(A) site using pulse-labelled RNA. The authors concluded that transcription terminated at or very near the poly(A) site. However, LeMeur et al. (1984) have recently shown that transcription proceeds at least 900 nucleotides beyond the poly(A) site, terminating within the next 170 nucleotides. It seems reasonable to conclude that poly(A) sites in higher eucaryotes are sites of nucleolytic processing and are not generally sites of transcript termination.

Earlier failures to detect transcripts of sequences distal to the poly(A) site of the ovalbumin gene must be attributed to the instability of such transcripts. Hofer and Darnell (1981) showed that transcribed sequences downstream of the globin poly(A) site are confined to the nucleus and are very unstable, exhibiting substantial turnover within a 3-minute labelling period. Those portions of the major late primary transcript of adenovirus which are distal to the last poly(A) site are similarly confined to the nucleus, but their rate of accumulation in the nucleus with increasing labelling periods suggests a half-life of about 20 min., similar to the average half-life of hnRNA in the host cells (Fraser et al., 1979a).

Termination Sites for RNA Polymerase II Transcription

Precise transcript termination at the end of any RNA pol II transcription unit has yet to be demonstrated.
Transcription of the major late adenovirus transcription unit appears to terminate within the last 2% of the genome length, a region of about 700 bp. Although nascent late transcripts hybridize in equimolar amounts to all regions of the genome between map units 16.4 and 98.2, much lower levels of hybridization to a fragment spanning map units 98.2-100 were detected by Fraser et al. (1979a). No more than a minority of the polymerase complexes transcribing the major late transcription unit can continue all the way to the end of the genome, but the data do not allow the precise site(s) of termination to be deduced. Electron microscopic analysis of transcription complexes in the major late transcriptional unit similarly suggested that termination occurs within the last 650 bp of the genome (Fraser and Hsu, 1980).

Hofer et al. (1982) attempted to localize the termination site for \( \beta \)-globin gene transcription. In agreement with their earlier study, they found that transcripts labelled in isolated nuclei of erythroleukemia cells hybridized in equimolar amounts to restriction fragments from within the coding region and from as far as 1400 bp downstream of the poly(A) site. Considerably lower levels of hybridization to a fragment located 1400-2000 bp downstream of the poly(A) site were observed, and hybridization to a fragment immediately downstream was almost undetectable. The results suggested that over 95% of the polymerases which transcribe the \( \beta \)-globin gene terminate in a region 1400-2000 nucleotides downstream of the poly(A) site, but no evidence could be found for a discrete termination site. A report of a discrete termination site for \( \beta \)-major globin transcripts (Salditt-Georgieff and Darnell, 1983) was later retracted.
(Salditt-Georgieff and Darnell, 1984). Citron et al. (1984) also failed to demonstrate termination at a discrete site distal to the β-globin gene.

Transcription of the mouse Amy2a α-amylase gene does not terminate at a well-defined site (Hagenbuchle et al., 1984). The level of hybridization of nascent transcripts, labelled by limited extension in isolated nuclei, to cloned DNA fragments indicated that the number of polymerases on the DNA template gradually drops over a 2 kb region, 2-4 kb downstream of the poly(A) site. Nuclear, non-polyadenylated transcripts spanning the Amy2a poly(A) site were detected in pancreatic nuclei, and their 3' termini were mapped by S1 nuclease digestion of the hybrids formed between these transcripts and 3' end-labelled DNA restriction fragments extending beyond the poly(A) site. Heterogeneous termini were detected throughout the region 2.5-4 kb downstream of the poly(A) site. If a discrete termination site existed for the Amy2a gene, the heterogeneous distribution of 3' ends could result from rapid exonucleolytic degradation of the terminated transcripts. However, the dilution of polymerases revealed by hybridization analysis of nascent chains is incompatible with termination at a single site, and together the data suggest that termination occurs at many sites over a region of about 2 kb.

The major late transcription unit of adenovirus is active at early stages of infection, though its activity is about 30-fold lower than after DNA replication begins (Fraser et al., 1979b). The same techniques that were used to analyze late transcription have been brought to bear upon the "early" transcripts of the major late transcription unit. Hybridization of pulse-labelled RNA to cloned restriction fragments and
UV transcription mapping, indicate that prior to viral DNA replication transcription from the major late promoter terminates before reaching map unit 70 (Nevins and Wilson, 1981). Although the site(s) of termination of these early transcripts clearly differ from those utilized later in infection, no evidence for precise termination at well-defined sites exists in either case.

Pulse-labelled nascent RNA chains from the major late transcriptional unit exhibit a bimodal size distribution indicative of some sort of discontinuity in transcription about 2000 bp downstream of the major late promoter (Evans et al., 1979). The authors attributed the discontinuity to premature termination. They tested their hypothesis by hybridizing briefly-labelled nuclear RNA to various restriction fragments under conditions of DNA excess. Transcripts from the first 2 kb of the late transcription unit were present in a 3-6 fold molar excess over transcripts of more promoter-distal sequences (Evans et al., 1979). Fraser et al. (1979b) detected a series of major late promoter-proximal transcripts 100-800 nucleotides long amongst the RNA labelled over a 5 hr period late in infection. Analysis of oligonucleotides released from the heterogeneous transcripts by RNase T1 suggested that many of the transcripts initiated at the same site as the major late transcript. Two-thirds or more of the transcripts initiated at the major late promoter apparently terminate within 2 kb of the promoter, but if a specific termination site exists within this region, its activity is obscured in these studies by subsequent degradation of the terminated transcripts. Mok et al. (1984) have found that premature termination of transcription in the major late transcription unit of adenovirus occurs
only at late stages of infection and generates two predominant transcripts of 120 and 175 nucleotides. The larger prematurely-terminated transcripts seen by Fraser et al. were not evident in this study.

Transcript termination has been reported to occur at a specific site in the late transcription unit of SV40. Hay et al. (1982) isolated the nuclei of SV40-infected cells and analyzed the RNA that became labelled when the nuclei were briefly incubated with $\alpha^{32P}$UTP. A prominent RNA fraction of sedimentation coefficient 5S was detected, which hybridized to restriction fragments proximal to the SV40 promoter region. The authors later demonstrated its release from the template (Hay and Aloni, 1984), thereby excluding the possibility that the RNA was a transitory species resulting from pausing of RNA polymerase. Pretreatment of the cells with DRB, which is believed to enhance premature termination, although not necessarily at termination sites utilized in its absence, (Fraser et al., 1978; 1979b) increased the levels of the 94 nucleotide, 5S species (Skolnik-David et al., 1982).

Various authors have suggested that premature termination of transcription is a common occurrence during transcription of nuclear genes. The suggestion is based on observations of a heterogenous population of short, capped transcripts in the nuclei of various cell lines (Salditt-Georgieff et al., 1983). These transcripts form a distinct peak upon size-fractionation of total nuclear RNA, suggesting that a discontinuity of some sort, perhaps a termination site, occurs within a few hundred base pairs of many cellular promoters. There has been no proof, however, that these transcripts arise because of termination at discrete sites in the genome. Tamm and Kikuchi (1979) found that most of
the [3H]Uridine incorporated by isolated HeLa cell nuclei entered short transcripts of 1-300 nucleotides, suggesting a numerical excess of short nascent transcripts over longer ones. Pulse-chase experiments gave ambiguous results, since the failure of a radioactive label in short RNA to enter longer RNA during a chase could have reflected either termination or the demise of the transcriptional apparatus in the isolated nuclei. Conversely, the production of larger labelled RNA molecules during the chase could have indicated that no promoter-proximal termination sites exist, or that they are not active in isolated nuclei.

Termination Signals for RNA Polymerase II

Since little evidence exists for sequence-specific termination of transcription by RNA polymerase II, it is reasonable to suspect that signals which specify exact termination sites are rare amongst polII transcription units. Termination does, however, occur reproducibly within broadly defined regions of the template, and signals might exist which establish the boundaries of such regions. Falck-Pedersen et al. (1985) introduced the 1.5 kb termination region of the mouse 0-major globin gene into the adenovirus ElA transcriptional unit. They found that transcription which initiated at the ElA promoter terminated within the inserted globin sequence, suggesting that the insert indeed carries some sort of termination signal. Hagenbuchle et al. (1984) suggested a variety of forms which a "sloppy" termination signal might take, including the presence of a downstream transcription unit, or a particular pattern of DNA modification or chromatin structure. Termination might result from the absence of some condition which
facilitates transcript elongation. For example the region between a promoter and the last polyadenylation site of a transcription unit might be maintained in a relatively open chromatin structure which offers little or no hindrance to the movement of the transcription complex. Beyond the polyadenylation site, a more compact chromatin structure might lead to more frequent and extended pausing of the polymerase and so increase the probability of termination and transcript release.

The synthesis of discrete prematurely terminated transcripts of SV40 suggests that RNA polymerase II is capable of site-specific termination in response to an appropriate signal. Hay et al. (1982) have estimated the location of the 3′ end of the 94 nucleotide prematurely terminated RNA, and they find that the 3′ terminal sequence bears considerable resemblance to a rho-independent bacterial terminator. The RNA appears to end in a series of U residues encoded by 4 consecutive T/A base pairs in the DNA. Immediately preceding the U tract in the RNA is a GC-rich sequence with dyad symmetry, which should be capable of forming a stable stem-and-loop structure. The authors suggest that termination at this site occurs by essentially the same mechanism as termination at a rho-independent terminator in bacteria, which is to say that RNA polymerase pauses upon transcribing the potential stem-and-loop structure and releases the transcript while paused, release being facilitated by the instability of the rU-dA base pairs between transcript and template. Pausing by RNA polymerase in SV40-infected cells at the SV40 attenuator site has since been demonstrated (Skolnik-David and Aloni, 1983). Experiments described by Hatfield et al. (1983) support the idea that eucaryotic RNA polymerases may respond in the same way as
the bacterial enzymes to sequences with the cardinal features of rho-independent terminators. These workers linked the major late promoter of adenovirus to a terminator from bacteriophage \( \lambda \) and allowed the resulting construct to be transcribed in a HeLa cell extract. "Runoff" from the end of the template produced a 311 nucleotide RNA, but a 273 nucleotide species was also produced. Its 3' end mapped to within a few nucleotides of the corresponding site of termination in \( \lambda \)-infected E. coli, in a series of base pairs immediately downstream of a GC-rich dyad symmetry. An analysis of the time course of the in vitro transcription showed that the polymerase paused upon reaching the inserted terminator sequence and after pausing, either continued synthesis of the runoff transcript or, about 30% of the time, terminated to produce the 273 nucleotide species. Neither of two other terminators from phage \( \lambda \) caused pausing or termination by RNA polymerase II in HeLa cells (unpublished work cited by Hatfield et al., 1983) although one of those tested is a typical rho-independent terminator. Eucaryotic RNA polymerase II can apparently respond to some of the terminators recognized by the E. coli enzyme, but features which are not found in these terminators may be required for efficient termination. Of course, efficient termination at a single site does not seem to be the rule amongst RNA polymerase II transcription units.

**Polyadenylation Signals in "Higher" Eucaryotes**

Signals which specify sites of polyadenylation for polymerase II transcripts serve functions analogous to those of the termination signals in other types of transcription unit inasmuch as they set downstream limits to the regions of the genome which can be expressed from.
particular promoters. Even though transcription seems to proceed beyond active polyadenylation sites, the transcripts so produced never find their way out of the nucleus and are rapidly degraded. Sites of polyadenylation are much more precisely specified than are sites of termination for polymerase II transcripts, and consequently the task of identifying the signals involved in polyadenylation is simpler in principle than that of finding polymerase II terminators.

In 1976, Proudfoot and Brownlee compared the sequences flanking the poly(A) tails of six eucaryotic messenger RNAs and noticed that the hexanucleotide AAUAAA was common to all six. It occurred about 20 nucleotides upstream of the poly(A) tail, and the authors suggested that it might constitute some sort of signal involved in mRNA processing or transport. Subsequent DNA sequencing studies and transcript mapping have shown that the sequence AAUAAA is indeed widely conserved amongst the 3' untranslated region of many mRNAs from higher eucaryotes. Wickens and Stephenson (1984) compared the sequences of 161 polyadenylated mRNAs from vertebrates and found that 90% of them contained the AAUAAA hexanucleotide in their 3' untranslated regions. Those which lack the canonical sequence generally have a variant sequence which differs by only a single nucleotide.

The first direct tests of the role played by the sequence AATAAA in the production of a eucaryotic mRNA were described by Fitzgerald and Shenk in 1981. By sequentially digesting SV40 DNA with a restriction enzyme and S1 nuclease, the authors produced deletions in the vicinity of the AATAAA hexanucleotide located 12 bp upstream of the polyadenylation site for the SV40 late mRNAs. Several of the viruses they recovered
after infecting monkey cells contained deletions on one side or the other of the AATAAA sequence, but none had lost the hexanucleotide itself. The procedures used to construct the deletions should not of themselves have prevented deletions from extending into the AATAAA hexanucleotide. The failure to isolate such deletions therefore implied that removal of the hexanucleotide, when it occurred, rendered the virus inviable.

The authors also constructed a derivative of SV40 containing a tandem duplication of a 240 bp fragment which included the late polyadenylation site and AATAAA hexanucleotide. The late mRNAs produced by this virus were polyadenylated with approximately equal efficiency at either of the duplicate sites. Deletion of a 16 bp fragment, including the sequence AATAAA, from either copy of the duplicated region abolished polyadenylation at the corresponding polyadenylation site. Instead, polyadenylation occurred efficiently 12 bp downstream of the remaining AATAAA hexanucleotide. These results suggest that the sequence AATAAA is at least part of a signal which is essential for polyadenylation.

In the same study, Fitzgerald and Shenk mapped the 3' termini of the late mRNAs produced by SV40 mutants with deletions immediately downstream of the AATAAA hexanucleotide near the late polyadenylation site. They found that the site of polyadenylation of these mRNAs was shifted downstream, with the result that the hexanucleotide and the polyadenylation site were always separated by 11-19 nucleotides. The hexanucleotide itself, or some sequence outside the deleted region might be responsible for maintaining this spacing. An indication that sequences upstream of the AATAAA hexanucleotide might be involved was provided by the analysis of a mutant which had short deletions on both
sides of the hexanucleotide. Transcripts produced by this mutant were polyadenylated immediately downstream of the hexanucleotide as well as at sites further downstream. No specific sequence occurred at all sites of poly(A) addition, although the sequence CA was a preferred site. These observations are consistent with the idea that some factor recognizes the sequence AATAAA (or its RNA homologue) and cleaves the transcript within a narrowly-defined region downstream, the specific site of cleavage within that region being one that fulfills a simple sequence requirement such as XA, where X is any nucleotide, C being preferred. The authors pointed out that the AATAAA hexanucleotide cannot be the only specific sequence requirement for polyadenylation, because the same sequence occurs within the coding region of the SV40 early transcriptional unit, where it does not serve as a signal for polyadenylation.

The importance of the AATAAA hexanucleotide as part of the polyadenylation signal in various eucaryotic genes has been well-established by several studies in which the effects of point mutations in the hexanucleotide were assessed. Wickens and Stephenson (1984) showed that a 220 bp fragment including the poly(A) site of the SV40 late transcription unit is sufficient, when ligated into pBR322 and introduced into Xenopus oocytes, to cause cleavage and polyadenylation of transcripts initiated in the vector. Four different mutations were introduced into the AATAAA hexanucleotide in the cloned SV40 fragment. Each of the mutant plasmids directed the synthesis of the same levels of RNA as the "wild-type" when injected into oocytes. However, most or all of the RNA produced from the mutant plasmids failed to undergo cleavage and polyadenylation. No correctly cleaved RNA could be detected
in oocytes carrying plasmids with the sequence AACAAA or AATGAA in place of AATAAA. Low levels of cleaved RNA were produced by plasmids carrying the sequence AATACA (10%) or AATTAA (<5%), and in each case, all of the cleaved RNA was polyadenylated. This suggests that mutations in the AATAAA hexanucleotide interfere with cleavage at the polyadenylation site, and that once cleavage has occurred, polyadenylation occurs normally. The AATAAA sequence may not be required for the polyadenylation reaction itself, but only for the (normally) prerequisite cleavage step. A similar effect was noted by Montell et al. (1983), who altered the sequence AATAAA near the polyadenylation site of the adenovirus EIA transcription unit to AAGAAA. Very low levels of correctly cleaved EIA mRNA were produced by the mutant virus in HeLa cells, but at least 80-90% of the correctly cleaved RNA was polyadenylated. Two types of evidence excluded the possibility that a substantial amount of correctly cleaved RNA was not polyadenylated but escaped detection because it was rapidly degraded. First, the levels of nuclear RNA containing EIA sequences were very similar in cells infected with the wild-type and mutant viruses. Second, Zeevi et al. (1982) had shown that inhibition of polyadenylation of adenovirus EIA and EIB transcripts did not alter their stability within the nucleus.

Higgs et al. (1983) found the first example of a naturally occurring cleavage/polyadenylation signal defect in a patient with a particular form of α-thalassemia. The patient’s α1 globin gene was completely inactive as a result of a frameshift mutation, and the α2 gene had undergone a single mutation which altered the sequence AATAAA near the poly(A) site to AATAAG. This mutation apparently reduced the level of
α₂ mRNA in the patient's reticulocytes by about 20-fold. When the defective α₂ gene was cloned and introduced into HeLa cells on an SV40 vector, it directed the production of normal levels of mRNA, but all of the transcripts were extended beyond the normal polyadenylation site into vector sequences. These results suggest that the α₂ mutation prevented transcript cleavage and therefore polyadenylation at the normal site. "Readthrough" transcripts were not detected in the patient's reticulocytes, perhaps because they were not sufficiently stable.

Orkin et al. (1984) described a β-thalassemia gene in which the AATAAA sequence of the normal polyadenylation signal had been altered to AACAAA. The effect of this mutation on the efficiency of polyadenylation at the normal site in erythroid cells could not be estimated because the mutation has been found only in heterozygotes. When expressed in HeLa cells, however, the mutant gene directed the synthesis of transcripts with the normal 3' end at levels 5-10 fold lower than did the normal β-globin gene. In both HeLa cells and erythroid cells in β-thalassemia patients, the mutant gene also caused the production of an extended transcript. This transcript was polyadenylated about 900 nucleotides 3' to the normal site, 10-15 nucleotides from the next AATAAA sequence.

Considerable effort has recently been devoted to identifying sequences or structures which might be required along with the AAUAAA hexanucleotide for transcript cleavage and polyadenylation. No other sequence in the 3' flanking regions of eucaryotic genes is as widely conserved as the sequence AATAAA, but sequence homologies have nonetheless been identified. Benoist et al. (1980) noted that the
sequence UUUUCACUGC is located immediately upstream of the polyadenylation sites of five different mRNAs. Berget (1984) compared the sequences around the polyadenylation sites of 61 different vertebrate mRNA precursors. (The sequences were deduced from the DNA sequences of the respective genes.) She identified the sequence CAYUG (Y=pyrimidine) as a conserved element located either immediately upstream or immediately downstream of the polyadenylation site. This sequence occurs within the conserved sequence noted by Benoist et al. (1980). The polyadenylation site itself was usually marked by the sequence NAA, in which N is most frequently C, less frequently G, and still less often U. Fitzgerald and Shenk (1981) observed the same sequence preference at sites of poly(A) addition in the mutants of the SV40 late transcription unit.

McDevitt et al. (1984) introduced deletions into the 3' flanking sequences of the adenovirus E2 gene and assayed their effects on the production of E2 mRNA and protein after introducing the altered genes into human cells. A gene retaining as few as 35 bp of the sequence normally found downstream of the polyadenylation site produced normal levels of functional E2 mRNA. In marked contrast, deletions which extended to within 20 or fewer base pairs of the polyadenylation site from downstream abolished the synthesis of E2 protein. No discrete polyadenylated RNA which could hybridize to an E2 probe was produced in cells transfected with E2 genes retaining 2, 12, or 20 bp of 3' flanking sequence. However, total RNA from these cells did protect part of an E2 probe from S1 nuclease digestion. The region protected by transcripts of each plasmid extended from the proximal end of the probe as far as the site of the deletion breakpoint in the plasmid, suggesting that the
deletions prevented transcript cleavage and caused the transcripts to be extended downstream for an unknown distance. Sequences required for correct cleavage and polyadenylation of adenovirus E2 transcripts must have a boundary 20-35 bp downstream of the polyadenylation site, but what feature of the boundary region is involved in transcript processing is not known. The authors note that a CAYUG-related sequence (CATG) occurs within the boundary region. A different CAYUG analog occurs immediately after the poly(A) site. A 12 bp sequence within the boundary region is an inverted repeat of a sequence which is located upstream of the poly(A) site and which includes the first 5 bp of the AAUAAA hexonucleotide. Pairing between the two inverted repeats might be involved in poly(A) site recognition (McDevitt et al., 1984), but this notion has yet to be tested.

Gil and Proudfoot (1984) produced deletions extending towards the AATAAAA poly(A) signal of the rabbit β-globin gene from a point 355 bp downstream. For each deleted variant, a fragment extending from the deletion breakpoint to a fixed site in the 3′ untranslated sequence was excised and introduced into an intact β-globin gene, upstream of its AATAAAA hexanucleotides. The resultant genes contained two AATAAAA hexanucleotides, the first of which retained a variable length of its normal 3′ flanking sequence. Polyadenylation occurred efficiently after the first AAUAAA sequence if it retained 51 bp of its 3′ flanking sequence. Transcripts of a gene in which the first AATAAAA was flanked distally by only 15 bp of a "wild type" sequence were polyadenylated exclusively near the second AAUAAA sequence. Sequences required for recognition of the β-globin poly(A) site therefore extend 15-51 bp.
beyond the AATAAA hexanucleotide. This region includes a CAYUG-related sequence, but again, whether or not it is specifically required is not known.

Woychik et al. (1984) have analyzed the sequence requirements for correct polyadenylation of transcripts of the bovine growth hormone gene. Fragments extending from a fixed point in the first exon of the gene to different positions in the 3′ flanking sequences were cloned into a eucaryotic expression vector in such a way as to place the gene under the control of the SV40 late promoter. The resulting plasmids were introduced into COS1 cells, and the 3′ ends of the growth hormone gene transcripts were located by S1 nuclease mapping. A plasmid which included 84 bp of the 3′ flanking sequence of the growth hormone gene produced transcripts which were polyadenylated exclusively at the normal site. One which lacked the entire 3′ flanking sequence of the growth hormone gene and 37 bp upstream of the normal poly(A) site produced transcripts which were polyadenylated in adjacent vector sequences about 200 bp downstream of the deletion breakpoint. This plasmid lacked the AATAAA sequence which is presumably involved in polyadenylation at the normal site. A plasmid containing the sequences normally found upstream of the poly(A) site and 1 bp of 3′ flanking sequence produced some transcripts which extended to the polyadenylation site in the vector, but most were polyadenylated in the immediate vicinity of the normal site. The inclusion of 10 or 13 bp of 3′ flanking sequences in the same type of plasmid eliminated the extended transcripts but did not completely restore polyadenylation to the normal site. Efficient polyadenylation at the normal site therefore requires sequences extending 13−84 bp beyond
the site itself. This region could theoretically base-pair with proximal sequences to produce an extended region of secondary structure, but the significance of this observation is unknown. A CAYUG sequence spans the poly(A) site of the growth hormone gene: whether or not it is required, it cannot account for the apparent requirement for more distal sequences for correct polyadenylation.

Conway and Wickens (1985) extended earlier analyses of the SV40 late polyadenylation signal by producing various deletions of sequences distal to the AATAAA hexanucleotide and assaying the ability of the remaining sequences to cause polyadenylation in frog oocytes. They found that sequences in the immediate vicinity of the polyadenylation site affected both the efficiency and the specificity of polyadenylation. Efficient polyadenylation also required sequences residing up to 26 base pairs distal to the normal poly(A) site.

Sadofsky and Alwine (1984) reported that some of the SV40 mutants constructed by Fitzgerald and Shenk (1981) produced higher levels of extended late transcripts than the wild-type virus. The data are difficult to interpret, in part because high molecular weight fragments of the probes used for S\textsubscript{1} nuclease analysis were apparently protected from the nuclease by all RNA samples, suggesting the presence of relatively abundant extended transcripts in all samples or of partial S\textsubscript{1} digestion products. No explanation for the presence of these large fragments was offered. Furthermore, in none of the figures were the levels of correctly cleaved and extended transcripts directly compared, because different probes were used to analyze the two classes of transcript. The authors suggest that extended transcripts comprise up to
6% of the late transcripts in some of the mutant viruses compared to 0.1% in the wild-type virus, and that sequences 3-60 bp downstream of the AATAAA hexanucleotide affect the efficiency of cleavage at the normal polyadenylation site. Of the bands which were attributed to protection of the probe by extended transcripts, only 1 was observed when polyadenylated RNA was used in the analysis, suggesting the presence of a cryptic polyadenylation signal downstream of the normal one.

Simonsen and Levinson (1983) have begun to characterize the signals involved in polyadenylation of transcripts of the hepatitis B surface antigen gene. The sequence AATAAA is absent from the region in which the transcript is cleaved and polyadenylated, but the variant TATAAA occurs 12-19 bp upstream of the poly(A) sites. Deletion of this sequence, along with downstream flanking sequences, prevents the synthesis of HBsAg mRNA in COS1 cells. In fact, 11-30 bp of the sequence downstream from the TATAAA hexanucleotide are required for production of the mRNA, and normal levels are not produced unless at least 45 bp of normal sequence are retained downstream of the hexanucleotide. The sequence CAYUG is not present in the region known to be required for polyadenylation.

The full extent of the sequences required for polyadenylation at any defined site have not been determined; nor is it clear to what extent the polyadenylation signals characterized to date depend for their activity upon their location with respect to other functional genetic elements.

Cole and Santangelo (1983) showed that certain sequences which in their normal location do not cause transcript cleavage or polyadenylation can do so if transferred to a different context. They constructed a derivative of the Herpes virus tk gene from which all sequences more than
23 bp downstream of the protein-coding region had been deleted. This
derivative, in contrast to one which retained 500 bp of 3′ untranslated
and flanking sequences, did not direct the synthesis of tk mRNA when
introduced into COS1 cells on an expression plasmid. Various DNA
fragments containing sequences known to be involved in polyadenylation of
their respective transcripts were able to restore the expression of the
defective gene when inserted downstream from it. Moreover, an 88 bp
fragment of the SV40 early transcription unit which included the sequence
AATAAA, but which does not cause SV40 transcripts to be cleaved or
polyadenylated in its vicinity, also acted as a polyadenylation signal
when joined to the defective tk gene. This result not only illustrated
the importance of the AATAAA sequence in transcript polyadenylation, but
also suggested that sequences which may be quite far removed from the
hexanucleotide are important in allowing it to function. The authors
point out two different possibilities which must be considered: that some
sequence in the SV40 early transcriptional unit prevents the unused
AATAAA from being recognized (formally, this includes the possibility
that the polyadenylation signal of the early transcription unit competes
with the upstream hexanucleotide for some essential factor), or that some
sequence in the tk gene might facilitate recognition of the
hexanucleotide.

Polyadenylation in Cell-Free Systems

Little is known of the mechanism by which primary transcripts are
cleaved at appropriate sites for polyadenylation. Nor is it known
whether polyadenylation is an automatic consequence of cleavage at the
correct site, or whether signals in addition to those which specify the
cleavage site are required for the polyadenylation reaction. Berget (1984) proposed that U4 snRNA might be involved in the recognition of polyadenylation sites by analogy with the role of U1 snRNA in splice site recognition (Rogers and Wall, 1980; Lerner et al., 1980). U4 contains two copies of a pentanucleotide complementary to the first 5 bp of the sequence AAUAAA, and two pentanucleotides complementary to CAUUG and CACUG, respectively. The potential recognition elements for AAUAAA and CAUCUG are appropriately ordered in U4 to allow them to pair simultaneously with their cognate sequences in an mRNA precursor. Such pairing was hypothesized to bring the polyadenylation site into an appropriate position and/or conformation for the cleavage reaction to occur. The cleavage enzyme in this model might be a component of U4 snRNP.

Several laboratories have recently succeeded in making cell sub-fractions or extracts which specifically polyadenylate mRNA precursors. These systems promise to be powerful tools in defining the steps in the process of polyadenylation, their relationships to each other, the nucleotide sequence requirements and trans-acting factor requirements for each step. Chen-Kiang et al. (1982) isolated nucleoprotein complexes engaged in transcription from adenovirus-infected HeLa cells. Under appropriate conditions these complexes produced transcripts of the major late transcription unit of adenovirus which were polyadenylated at the sites normally used in vivo. Manley et al. (1982) demonstrated correct polyadenylation of adenovirus late transcripts in nuclei isolated from infected HeLa cells. They showed that the polyadenylated transcripts were derived from larger precursors by cleavage.

Soluble extracts of HeLa cells capable of specifically initiating
transcription at some polII promoters had been developed in 1980 by Manley and colleagues. Transcript cleavage and polyadenylation did not occur under the conditions used for transcript initiation in these extracts, and specific 3' termini were produced only by "runoff" of the polymerase from the end of a linear template. In 1983, Manley reported that certain runoff transcripts prepared in cell extracts were efficiently polyadenylated if subsequently incubated in a mixture containing a different concentration of the extract. Of ten runoff transcripts tested, only two were efficiently polyadenylated (70% of input was retained on oligo(dT) cellulose after processing). One common feature of these two was that their 3' ends were located within 50 nucleotides of the 3' ends of the corresponding transcripts found in vivo. This suggested that polyadenylation in the cell extract depended upon the same signals as those required for polyadenylation in the cell. In each case, poly(A) was added to the end generated by runoff transcription, indicating that the endonuclease which cleaves transcripts at poly(A) sites was not active in the cell extract.

Transcripts truncated 237 nucleotides upstream of an efficiently-polyadenylated site were not polyadenylated at all, suggesting that an essential sequence resided within 237 bp of the efficiently-utilized 3' terminus, which itself was located near the poly(A) site of the SV40 late transcription unit. The nature of this essential sequence was investigated by Manley et al. (1985). They first showed that two different runoff transcripts containing the 237 nucleotide SV40-encoded segment at their 3' termini were polyadenylated efficiently in vitro, while the corresponding transcripts without the 237
nucleotide segment were not polyadenylated at all. Deletions extending into the 237 bp SV40 segment from either end were produced, and runoff transcripts were prepared from the altered plasmids (which were linearized at the promoter-distal end of the SV40 sequence). The transcripts were tested for their ability to undergo polyadenylation.

Two copies of the hexanucleotide AAUAAA are present within 50 nucleotides upstream of the normal poly(A) site of SV40 early transcripts. Runoff transcripts which retained one copy of this sequence were efficiently polyadenylated in vitro but transcripts lacking the AAUAAA hexanucleotide were not detectably polyadenylated. Formally, the results showed that either of two regions, one 22 nucleotides and the other 38 nucleotides, contained sequences required for in vitro polyadenylation. Similarity between the two regions was restricted to the sequence AAUAAA and 3 flanking nucleotides on either side. The same regions could direct polyadenylation of 3' termini located 275-375 nucleotides downstream, although these termini were processed less efficiently than termini only 50 nucleotides downstream.

These results provided the first indication that the AAUAAA hexanucleotide is specifically required for polyadenylation as distinct from transcript cleavage. The hexanucleotide alone is not sufficient for polyadenylation in vitro, since a runoff transcript terminated 60 nucleotides downstream of the internal AAUAAA in the SV40 early transcription unit is not a substrate (Manley, 1983). Poly(A) polymerases have been purified from a variety of cell types (reviewed by Edmonds and Winters, 1976). The purified enzymes show no endonucleolytic activity, nor do they show any requirement for specific sequences in the RNA.
primers. An attractive hypothesis is that the endonuclease and polyadenylate polymerase activities form part of a complex which includes a single sequence recognition factor. One must suppose that in the HeLa whole cell extract the endonuclease is inactive for some reason (eg. loss from complex, lack of cofactor). This hypothesis would account for the coupling between cleavage and polyadenylation that is observed in isolated nuclei (Manley et al., 1982) and for the fact that the same nucleotide sequence is required for both activities (Manley et al., 1985; and refs. cited earlier). Berget's (1984) hypothesis could be accommodated by supposing that U4 snRNP is the recognition factor in the hypothetical endonuclease/polymerase complex, but Manley et al. (1985) caution against this supposition by citing unpublished results which suggest that anti-snRNP antibodies have no effect on the in vitro polyadenylation reaction.

Moore and Sharp (1984) have also developed a HeLa whole-cell extract which polyadenylates mRNA "precursors" generated by in vitro transcription. Their extract, like that of Manley (1983), polyadenylated 3' termini of RNA generated by runoff transcription in a separate reaction. They also observed, however, that under appropriate conditions, transcripts generated in the whole cell extract could be polyadenylated in the same extract at a site normally used in vivo, namely the L3 polyadenylation site of adenovirus. Polyadenylation at the normal site was only observed when the transcript was generated in situ, yet the reaction did not seem to be directly coupled to transcription. Pulse-chase experiments showed that there was a considerable lag between the synthesis of a transcript and its polyadenylation. As well,
transcripts synthesized in a 50 min period were cleaved and polyadenylated over the next 4 hr in the presence of inhibitors which prevented further transcription. Moore and Sharp suggest that a structure which is subsequently required for transcript 3' end processing assembles during transcription. They note in this context the results of Economidis and Pedersen (1983), who found that polII transcripts synthesized in vitro could be assembled into ribonucleoprotein particles in the same extract, but that assembly of exogenous RNA into such particles was inefficient. The mechanism by which the 3' termini for polyadenylation were generated in the extract could not be directly determined because the efficiency of polyadenylation was relatively low and the background of heterogeneous non-polyadenylated transcripts was high. Nevertheless, given the evidence that endonucleolytic cleavage occurs at the L3 poly(A) site in vivo (Nevins and Darnell, 1978; Fraser et al., 1979) and in isolated nuclei (Manley et al., 1982) the authors concluded that the same mechanism probably operated in the cell extract.

More recently, Moore and Sharp (1985) have reported that a nuclear extract of HeLa cells accurately and efficiently cleaved and polyadenylated exogenous runoff transcripts. Antibodies to U1 snRNP, or to determinants shared by U1, U2, U4, U5 and U6 snRNP, or to LaRNP inhibited the in vitro polyadenylation reaction in nuclear or whole-cell extracts (Moore and Sharp, 1984;1985). Since such antibodies apparently fail to inhibit polyadenylation of exogenous RNA substrates (cited in Manley et al., 1985) it would seem that one or more snRNPs may be involved in cleavage at poly(A) sites, but are not required for polyadenylation per se. The relationships between the endonuclease which
generates 3' termini for polyadenylation, the poly(A) polymerase itself, and the factor(s) required to confer sequence specificity on each of these enzymes remain obscure for the time being, but experiments with in vitro polyadenylation systems offer the exciting prospect of unravelling these relationships.

Nonpolyadenylated mRNA

One ubiquitous class of eucaryotic mRNAs commonly lacks poly(A): the mRNAs which encode the major histones of chromatin. Histone mRNAs were originally recognized as components of small polysomes which appeared during S phase of the cell cycle and which directed the incorporation of radiolabelled amino acids in proportions expected for histone-synthesizing particles (Borun et al., 1967). Adesnik and Darnell (1972) showed that little if any of the radioactive adenosine incorporated into the histone mRNAs of HeLa cells could be recovered in the RNase-resistant form expected of poly(A) sequences. The level of nuclease-resistant material, and the approximate average length of the mRNAs suggested that no more than 8 consecutive A residues could be present per molecule. Histone mRNAs from a variety of organisms and cell types have since been shown to similarly lack poly(A) (see Henstschel and Birnstiel, 1981, for review). [It is also true, however, that histone mRNAs in Tetrahymena and yeast, (Bannon et al., 1983; Fahrner and Hereford, 1983) and a fraction of the histone mRNAs in amphibian oocytes, (Slater and Slater, 1974) are polyadenylated. Messenger RNAs encoding variant histones are also usually polyadenylated (Doenecke and Tonjes, 1984).]
Whether or not any mRNAs other than those encoding histones lack poly(A) has been a matter of some dispute. Greenberg and Perry (1972) found that 62% of the polysomal RNA of L cells which was labelled in the presence of a low concentration of actinomycin D (to prevent labelling of rRNA) bound to poly(U) in solution, suggesting that it contained poly(A). In the fraction which apparently lacked poly(A), they found histone mRNAs and a heterogeneous population which they described as non-messenger polysomal RNA. They concluded that the vast majority of cytoplasmic mRNAs are polyadenylated, histone messengers being the only species which completely lack poly(A). However, Greenberg (1976) showed that about 30% of the RNA in L cells lacks poly(A), and judging from the association with polysomes and its labelling in the presence of inhibitors of rRNA synthesis this poly(A){eq}^-\text{ fraction is genuine mRNA.} About 50% of the RNA in this fraction is larger than histone mRNA.

Two years earlier, Milcarek, Price, and Penman (1974) isolated HeLa cell mRNA from polysomes and showed that about 30% of it failed to bind to oligo(dT) cellulose. RNA which fails to bind to oligo(dT) cellulose may still have a short poly(A) tract of up to about 15 nucleotides (Brawerman, 1981). However, Milcarek and colleagues showed that the poly(A){eq}^-\text{ fraction of HeLa cell mRNA contained no more than 6-8 consecutive A residues per molecule, on average, and that these short tracts were internal rather than 3'-terminal in location. RNase digestion of the poly(A){eq}^-\text{ fraction released only short oligomers of adenylate residues, and the content of adenosine measured after alkaline hydrolysis was much lower than expected if all were 3' terminal in location. Even mRNA which genuinely lacks poly(A) could conceivably be derived from polyadenylated...
species. While measures which guard against RNA degradation during extraction procedures should prevent artifactual loss of the poly(A) segment, there is evidence that such loss can occur in the cell. Sheiness and Darnell (1973) showed that the average length of the poly(A) segments on mRNA declines after the mRNA arrives in the cytoplasm. While poly(A) in nuclear RNA or in newly-synthesized cytoplasmic mRNA has an average length of about 250 nucleotides, the most abundant size class of cytoplasmic poly(A) includes tracts of about 50 residues (Sheiness et al., 1975). Harpold et al. (1981) studied the synthesis and processing of nine specific mRNAs in CHO cells by testing the hybridization of their respective cloned cDNAs to poly(A)~ and poly(A)^+ mRNA fractions. They found that after a brief period of labelling with radioactive nucleoside, over 90% of the labelled RNA which hybridized to any of the 9 cDNA clones was polyadenylated. After longer labelling periods, up to 45% of the mRNA detected by certain of the cDNA probes failed to bind to oligo(dT) cellulose, suggesting the loss of poly(A) from this fraction of the steady-state mRNA. These observations were not defended as proof that all mRNAs are initially polyadenylated [after all, cDNA probes prepared against poly(A)^+ RNA would not be expected to detect RNAs which never have poly(A)], but they do show that nonpolyadenylated species may be derived from polyadenylated ones.

If all poly(A)^- RNAs are derived from poly(A)^+ mRNA, one would expect to find that all sequences present in the poly(A)^- RNA fraction are represented in the polyadenylated fraction. Milcarek, Price, and Penman (1974) showed that this is not the case in HeLa cells, because a cDNA probe prepared against the poly(A)^+ population hybridized to
poly(A)^- mRNA to less than 20% of the extent of its hybridization to the poly(A)^+ fraction. Milcarek (1979) reported that excess unlabelled poly(A)^+ mRNA could compete effectively with labelled poly(A)^+ mRNA for hybridization sites in genomic DNA, but that it had no effect on the hybridization to DNA of labelled poly(A)^- mRNA. This suggested that the two mRNA fractions were encoded by different DNA sequences. [A similar experiment with unlabelled poly(A)^- RNA gave ambiguous results, in that the unlabelled RNA reduced the extent of hybridization of either a poly(A)^- or poly(A)^+ tracer to DNA by only about 30%. The reasons for this result were not explained, but included amongst them may have been the fact that the unlabelled RNA used as competitor contained a vast excess of non-mRNA sequences as compared to poly(A)^- mRNA.] The kinetics of hybridization of poly(A)^+ and poly(A)^- mRNA fractions to a DNA probe made against poly(A)^+ mRNA suggested that some sequence overlap existed between the two fractions. Most abundant polyadenylated mRNAs were found to be represented in the poly(A)^- fraction, about 10% also being abundant in the latter fraction and the remainder occurring much less frequently. About 60% of the remaining "scarce" polyadenylated RNAs were found to be unique to that fraction, while 40% were also present amongst the nonpolyadenylated RNAs. Milcarek (1979) therefore suggested that HeLa cell mRNAs could be assigned to one of three classes based on their possession of poly(A): one class consisted of sequences which occurred mainly in polyadenylated form, a second of sequences which for the most part lacked poly(A), and a third of sequences which were represented in both fractions.

Similar approaches have been used to identify nonpolyadenylated
mRNAs in other cell types. Nemer et al. (1974) found that in sea urchin blastulae, about 50% of the mRNA failed to bind to oligo(dT) cellulose. The level and size-distribution of RNase A+T1-resistant material in adenosine-labelled, non-polyadenylated RNA indicated that no more than 2% of the molecules in the poly(A)- RNA fraction could contain tracts of 30 or more A residues, and that on average, a molecule in that fraction contained no more than 5 consecutive A residues. Histone mRNAs were the major nonpolyadenylated RNA species, but 70% of the poly(A)- fraction consisted of heterogeneous species larger than histone mRNA. Hybridization of the poly(A)- RNA with a cDNA probe prepared against poly(A)+ RNA suggested that very little sequence overlap existed between the two fractions.

Grady et al. (1978) reported that poly(A)+ and poly(A)- mRNAs in mouse liver comprise two essentially nonoverlapping sequence populations, with the nonpolyadenylated sequences accounting for about 43% of the total complexity. The same group, however, later found that about 20% of the total sequence complexity of the rat liver RNA was represented in both the poly(A)+ and poly(A)- populations, while about 35% was unique to the nonpolyadenylated fraction (Grady et al., 1981).

Attempts to identify specific nonpolyadenylated, nonhistone gene transcripts have not generally been very successful. One approach has been to compare the in vitro translation products of poly(A)+ and poly(A)- RNA fractions. Proteins encoded by poly(A)- RNA in these experiments seem to comprise a subset of the much larger array of proteins encoded by the poly(A)+ fraction (Kaufmann et al., 1977). Chikaraishi (1979) reported that 12 of 18 in vitro translation products
of a complex poly(A)^− fraction of brain mRNA were not found amongst the products of the poly(A)^+ fraction, but a skeptic would say that the data did not reliably distinguish these products, which were present in very low levels, from peptides of very similar mobilities encoded by the poly(A)^+ mRNA. Another approach is to search for cDNA clones which hybridize to poly(A)^−, and not to poly(A)^+ mRNA, although it should be noted that standard procedures for cDNA synthesis are applicable only to polyadenylated mRNAs. The failures of both types of approach have quite recently been cited as evidence against the existence of nonpolyadenylated, nonhistone mRNAs (Darnell, 1982).

Both approaches, however, are biased in favour of detecting abundant transcripts, and it is conceivable that their failure could result from the low abundance of any transcripts which exist only in nonpolyadenylated form. The skeptic would perhaps have been impressed with a recent report from Brilliant, Sueoka, and Chikaraishi (1984). The authors constructed a partial library of rat genomic DNA enriched in those sequences which hybridized at high R^T values to brain mRNA. From this library they retrieved nine clones which hybridized to rare transcripts in brain, or in brain and other tissues. Two of the clones hybridized to transcripts which could only be detected in the poly(A)^− fraction of polysomal RNA.

Although at least two nonhistone, nonpolyadenylated mRNAs exist in the steady state mRNA population in the brain, it remains possible that they are synthesized in polyadenylated form and rapidly lose their poly(A) tracts. It will be interesting to know, when the regions encoding the 3′ ends of these transcripts are sequenced, whether or not any of the signals associated with the generation of polyadenylated 3′
termini are present.

Signals Required for Histone mRNA 3' End Generation

A haze of uncertainty still surrounds the origins and identities of nonhistone, nonpolyadenylated RNAs, and nothing whatsoever is known of the signals and processes involved in generating their 3' ends. In striking contrast, the signals which specify the locations of the 3' ends of histone mRNAs can be described quite precisely, and recent reports suggest that a detailed understanding of the mechanisms of 3' end generation in these mRNAs will soon be at hand. The studies of Birnstiel and his colleagues on the embryonic histone genes of sea urchins have contributed much of our current understanding of histone gene organization and expression.

In 1979, Busslinger, Portmann, and Birnstiel identified two regions of nucleotide sequence conservation shortly downstream from the coding sequences of nine histone genes from two species of sea urchin. The first conserved sequence was 23 bp long and included a 6 bp inverted repeat: AACGGC(C/T)CTTTTCAG(G/A)GCCACCA. Hentschel et al. (1980a) mapped the 3' terminus of each of the mRNAs encoded by the major embryonic histone repeat unit of Psammechinus miliaris to within a few nucleotides of the 3' end of the 23 bp conserved sequence. The same sequence has since been found shortly downstream of various histone genes of Drosophila, Xenopus, chicken, mouse, and humans (reviewed by Hentschel and Birnstiel, 1981). The mRNA of a histone H4 gene from Xenopus is polyadenylated at the 3' end of the 23 nucleotide conserved sequence (Zernik et al., 1980), but the mRNAs of most genes flanked by the 23 bp
sequence are not polyadenylated. Conversely, those histone genes which are transcribed to produce polyadenylated RNAs generally lack the 23 bp conserved sequence (Doenecke and Tonjes, 1984).

Hentschel et al. (1980b) showed that all 5 histone genes of the sea urchin h22 repeat are transcribed upon injection into *Xenopus* oocyte nuclei to produce RNAs with the same 5' and 3' termini as those found in sea urchin embryos. (However, correct 5' ends of the H1 and H4 mRNAs and correct H3 mRNA 3' ends were generated very inefficiently. The latter observation proved to be very useful, as will be discussed later.) Birchmeier et al. (1982) made use of the *Xenopus* oocyte as a histone gene expression system to test the functional importance of the 23 bp conserved sequence downstream of the sea urchin H2A gene. They found that a deletion which removed 12 bp from within the 23 bp sequence completely prevented the H2A gene from directing synthesis of normal H2A mRNA in the oocyte nucleus. S1 nuclease mapping showed that the mutant gene encoded transcripts which had the same 5' end as authentic H2A mRNA, but which extended beyond the site of the normal 3' end. The removal of sequences extending from the 3' end of the conserved inverted repeat to a point within the coding sequence of the downstream H1 gene also prevented correct 3' end generation in H2A transcripts. On the other hand, deletion of the entire H2A coding region, and part of the 5' untranslated sequence, did not affect the production of normal transcript 3' ends. The deletion analysis pointed to the importance of the 23 bp conserved sequence, and possibly downstream intergenic sequences, in H2A transcript 3' end generation. The involvement of downstream spacer sequences was also suggested by the observation that a 60 bp fragment which included
the last 4 bp of the H2A coding sequence and extended 4 bp beyond the 23 bp conserved sequence could not cause 3' end generation when inserted into the H2B gene. The 3' ends of all the transcripts of the gene carrying the insertion mapped to the normal H2B 3' end site. In a second report, Birchmeier et al. (1983) showed that an H2A gene which retained 24 bp of spacer DNA downstream of the conserved sequence, but not one which had only 4 bp of spacer DNA, could direct the synthesis of a correctly "terminated" transcript. However, 80 bp of spacer DNA was required for the correct transcript 3' end to be produced with maximum efficiency. They also showed that all 3' untranslated sequences upstream of the 23 bp conserved element could be deleted without affecting 3' end generation.

To find out whether the inverted repeat within the 23 bp conserved sequence was a required feature for H2A transcript 3' end generation, Birchmeier et al. (1983) subjected the sequence to bisulfite mutagenesis in vitro. Three different mutations which reduced complementarity between the two halves of the inverted repeat either abolished or severely curtailed the production of correct H2A mRNA 3' termini from the mutant genes in Xenopus oocytes. Sequence complementarity was restored by replacing the unaltered half of the inverted repeat with a synthetic duplex oligonucleotide containing the appropriate mutant sequence. The two double mutant genes so prepared contained inverted repeats of different sequences from the wild-type gene, yet both directed the production of H2A mRNAs with correct 3' termini in oocytes. This result was interpreted to mean that some sequence alteration in the conserved inverted repeat could be tolerated as long as the point symmetry of the repeat was maintained. The importance of symmetry suggests that the two
halves of the repeat interact with each other either at the DNA or RNA level. Birchmeier and colleagues distinguished between the two possibilities by constructing and injecting into oocytes heteroduplex templates containing the wild-type inverted repeat in one strand and a variant sequence in the other. They found that mutations which abolished complementarity between the halves of the inverted repeat prevented 3' end generation if included in the RNA-antiparallel strand, but had little effect if present only in the RNA-isoparallel strand. If the postulated interaction between the halves of the inverted repeat occurred at the DNA level, mutations in either strand could have prevented the interaction. The observed results were more consistent with the idea of an interaction at the RNA level, since only the RNA-antiparallel DNA strand could influence the structure of the transcript.

The 10 bp sequence CAAGAAGAA was the second conserved sequence found by Busslinger and colleagues. It occurred 6-8 bp downstream of the 23 bp element. The first 9 bp of this sequence appear to be quite strictly conserved amongst sea urchin histone genes but only weakly so amongst the histone genes of other organisms (Hentschel and Birnstiel, 1981).

The studies of Birchmeier and colleagues clearly demonstrated that sequences downstream of the 23 bp conserved element were important for H2A transcript 3' end generation, but they did not establish whether or not the second conserved element was functionally important. Georgiev et al. (1985) have removed any doubts as to its role in 3' end generation by producing a variety of nested deletions and "linker scanner" mutations (McKnight and Kingsbury, 1982) in the 3' flanking sequences of the sea urchin H3 gene and testing their transcription in the Xenopus oocyte.
Deletions which extended into either the 23 bp or the 9 bp conserved sequence completely prevent the production of H3 transcripts with correct 3' termini. Furthermore, the spacing between these two elements appeared to be crucial to 3' end generation, because the insertion of as few as 6 extra base pairs between the two abolished 3' end generation at the normal site. This observation led the authors to suggest that the two elements should be considered portions of a single 3' end signal. The deletion of sequences downstream of the conserved nonamer also reduced the efficiency of the 3' end generation at the normal site, but it was not possible to attribute this effect to a particular sequence within the spacer DNA.

Evidence for 3' End Processing of Histone Gene Transcripts

The inverted repeat sequence near the 3' termini of histone mRNAs was reminiscent of rho-independent terminator sequences near the 3' ends of procaryotic transcripts. For this reason, it was considered likely that the 3' termini of histone mRNAs were the products of transcript termination (Busslinger et al., 1979; Hentschel and Birnstiel, 1981). Recent results have made this opinion untenable.

Birchmeier et al. (1984) reasoned that if histone mRNA 3' termini are produced by a processing reaction, it should be possible to uncouple their formation from transcription. They linked the sea urchin H2A gene to the P_L promoter of bacteriophage λ, linearized the recombinant molecule at a site downstream of the H2A gene and transcribed it in vitro using E. coli RNA polymerase. The runoff transcripts so produced included 230 nucleotides of spacer sequence downstream of the H2A transcript 3' end.
site. After being injected into Xenopus oocyte nuclei, the runoff transcripts were processed to molecules whose 3' termini mapped at the same site as that of authentic H2A mRNA. Analogous runoff transcripts of an H2A gene with two point mutations in the conserved inverted repeat were not processed at all upon injection into the oocyte, and nor were runoff transcripts which included only 4 nucleotides of spacer sequences. Transcripts which included 24 nt of spacer were processed very inefficiently, and the efficiency of processing increased as the amount of spacer sequence included in the transcript increased. Transcripts with 200 nt 3' spacer extensions were processed completely under the conditions of the assays. The fact that the processing of injected histone pre-mRNAs exhibited the same sequence requirements as the production of histone mRNAs from injected plasmids strongly suggested that processing is the normal mechanism for generating the 3' ends of histone mRNAs.

Krieg and Melton (1984) used a similar approach to study the mechanism by which the 3' ends of transcripts of the chicken H2B gene are produced in Xenopus oocytes. A plasmid including the complete gene with its promoter and 3' flanking regions was correctly transcribed in oocytes to produce authentic H2B mRNA. A 3'-extended H2B transcript, produced in vitro by runoff transcription from a linked phage SP6 promoter, was correctly processed to a molecule with the same 3' end as H2B mRNA upon injection into oocytes. A runoff transcript containing only 76 nucleotides downstream of the mature 3' end was processed, indicating that it contained all sequences required for processing.

Histone transcript 3' end processing is not a unique feature of the
Xenopus oocyte. Price and Parker (1984) found that it also occurs in a nuclear extract of cultured Drosophila cells. They prepared a DNA fragment including the 3' end of the Drosophila H3 gene and its downstream flanking sequences and added a single-stranded tail of dCMP residues to the 3' end of the RNA-antiparallel strand. Partially purified RNA polymerase II from Drosophila initiated transcription at the single-stranded tail and produced a runoff transcript which extended 55 nucleotides beyond the site of the mature H3 mRNA 3' end. If a nuclear extract of a Drosophila was added after a short preincubation of the template with polymerase, the runoff transcript gradually disappeared and was replaced by a shorter species whose 3' end coincided with that of mature H3 mRNA. The precursor-product relationship between the extended and mature RNAs was confirmed by showing that the purified extended RNA could be processed to the mature RNA upon addition to nuclear extract in the absence of ongoing transcription. The authors suggested that processing involved a specific endonucleolytic cleavage of the extended transcript, because transcripts with 55-nucleotide and 1 kb 3' extensions were processed with the same kinetics. The processing activity in the nuclear extract has been partially purified, fuelling the hope that its structure and mode of action will soon be understood in detail.

A factor responsible for the production of the 3' termini of mature transcripts of the sea urchin H3 gene was identified in 1982 by Stunnenberg et al. They made use of the earlier observation that H3 mRNA 3' termini are produced very inefficiently in Xenopus oocytes (Hentschel et al. 1980b) and searched for a component of sea urchin embryos which would allow efficient H3 mRNA synthesis in oocytes. Such a component was
found in a salt wash fraction of chromatin prepared from over $10^4$ embryos. The component had a sedimentation coefficient of about 12S, and Galli et al. (1983) showed it to be an snRNP. It contained a non-polyadenylated RNA of about 60 nucleotides which itself allowed efficient H3 mRNA production when injected into Xenopus oocytes prior to the injection of the H3 gene. Although the 12S snRNP was first referred to as a termination factor, it is now believed to be a 3' end processing factor. Birchmeier et al. (1984) showed that 3' extended H3 transcripts synthesized in vitro could be processed to molecules with mature 3' termini in Xenopus oocytes only if the oocytes were first injected with the 60-nucleotide RNA found in the 12S factor. The 60-nucleotide RNA, which is estimated to be present in no more than $10^5$ copies/cell in sea urchin embryos (Galli et al., 1983), is referred to as U7 RNA. Its 5' end is complementary to the conserved sequences found downstream of sea urchin histone coding sequences, and Strub et al. (1984) advanced the hypothesis that the 3' end processing activity of U7 rests on its ability to base pair with these sequences. It is not clear why, when the 23 bp and 9 bp sequences to which U7 RNA is complementary are common to all sea urchin histone mRNA precursors, the requirement for exogenous U7 RNA for processing in Xenopus oocytes is only apparent with H3 mRNA precursors.

Histone Transcript Termination

Birchmeier et al. (1984) investigated the question of histone transcript termination by mapping the transcripts produced from an H2A gene which lacked the conserved inverted repeat and retained varying amounts of downstream spacer. They reasoned that the absence of the inverted repeat would prevent H2A transcript processing, but that readthrough into the downstream H1 gene would occur only if any
terminator signals in the normal spacer DNA were also absent. They found that a plasmid containing 230 bp of spacer DNA produced almost no readthrough transcripts, but that a plasmid containing 130 bp of spacer DNA allowed higher levels of readthrough transcription, and maximal levels resulted from the deletion of all but 25 bp of spacer DNA. This suggested that most transcription originating at the H2A promoter terminates heterogeneously within 230 bp of the 3' end of the gene.

The Function of Poly(A) in Eucaryotic mRNA

The presence of 3'-terminal poly(A) tracts in most eucaryotic mRNAs naturally raises suspicions that the sequence is important to some aspect of mRNA synthesis or function. In 14 years since the discovery of polyadenylate in mRNA and hnRNA, many plausible roles for poly(A) have been proposed, but few have been proven. The mere existence of nonpolyadenylated transcripts, such as those of most histone genes, does not in any way argue against an essential role for polyadenylate in those transcripts in which it is found. The same essential role may simply be served by other sequences or structures in poly(A)− mRNAs. The effects of mutations in cleavage/polyadenylation signals in a variety of genes have clearly indicated the importance of the normal 3' end generation process to the production of stable, functional transcripts, but they have not singled out a role for polyadenylation itself.

Most hnRNA is degraded without ever leaving the nucleus (Soeiro et al., 1968). The idea that mRNA is derived from hnRNA must therefore imply that some mechanism selects those hnRNA molecules or
segments which are to give rise to mRNA. Darnell et al. (1971b) suggested that poly(A) might be involved in the selection mechanism on the basis of the kinetics of poly(A) accumulation in the nucleus and cytoplasm. Labelled poly(A) accumulated quite rapidly in the cytoplasm to levels which exceeded those in the nucleus, whereas the total labelling of hnRNA greatly exceeded that of cytoplasmic mRNA for several hours. This observation suggested that poly(A) was more efficiently conserved between nucleus and cytoplasm than bulk hnRNA sequences.

Jelinek et al. (1973) extended the earlier observations of Darnell and coworkers by showing that labelling of HeLa cell nuclear poly(A) reached a plateau after about 30 minutes, while the amount of label in cytoplasmic poly(A) continued to increase linearly. The authors (and see also Darnell et al., 1973) interpreted their findings as indicating that the rate of nuclear synthesis of poly(A) was matched by the rate of its transport to the cytoplasm, and therefore that poly(A) was efficiently, perhaps completely, conserved between the two compartments.

Perry and his colleagues disputed the idea of complete nucleocytoplasmic conservation of poly(A) after finding that the specific activity of the RNA precursor pool in the nucleus increased throughout labelling periods of several hours. The rate of cytoplasmic accumulation of labelled poly(A) under these circumstances would increase continually rather than remain constant if poly(A) was completely conserved. Perry et al. (1974) therefore argued that some poly(A) must turn over in the nucleus, and that poly(A) addition by itself could not guarantee the conservation of a sequence as mRNA. The argument rested in part on the notion that the half-life of eucaryotic mRNA is long in comparison to the
labelling periods used in the accumulation experiments. Rapid cytoplasmic degradation of some mRNAs could have prevented an increase in pool specific activity from being reflected in the accumulation curve for cytoplasmic poly(A).

Puckett et al. (1975) were able to demonstrate the existence in L cells of a class of mRNA with a short half-life by pulse-labelling with $[^3H]$-guanosine and chasing with excess unlabelled nucleosides. The chase became effective within about 2 hr, as revealed by a levelling-off in the total amount of radioactivity incorporated into RNA. After 4 hr the amount of labelled poly(A)$^+$ mRNA began to decrease. It dropped by 50% over the next 2-3 hr, suggesting that perhaps half of the mRNA belonged to a rapidly-turning over class. A specific member of a similar unstable mRNA class in CHO cells was detected by Harpold et al., (1981), who measured the rate of accumulation of radioactive mRNAs capable of hybridizing to nine cDNA clones. One of the nine clones detected an mRNA which had a half-life of only 1-2 hr. The others hybridized to more stable transcripts, but these transcripts became more abundant in relation to the rest of the labelled RNA as the period of labelling increased, implying that much of the mRNA in CHO cells must have been quite unstable. The existence of rapidly turning-over mRNA made it possible to propose a model which could account for the observed kinetics of accumulation of poly(A) and changes in pool specific activity while assuming complete conservation of poly(A) between nucleus and cytoplasm.

Nevins and Darnell (1978) addressed the question of whether poly(A) might play a role in marking mRNA sequences for conservation by comparing the rates of accumulation of specific adenovirus late cytoplasmic mRNAs to the rates of accumulation of the same sequences in poly(A)$^+$ and total poly(A).
nuclear RNA. The rate of accumulation in total nuclear RNA of newly-labelled sequences specifically found in a given mRNA family exceeded the rate of accumulation of the mRNAs themselves by three- to ten-fold, suggesting degradation of a substantial fraction of the nuclear adenovirus-specific RNA. However, the rate of appearance of the cognate sequences in polyadenylated nuclear RNA was similar to their rate of accumulation in cytoplasmic RNA, suggesting that polyadenylation might be sufficient to distinguish a nuclear RNA molecule destined for conservation as mRNA from those destined for degradation.

Early studies of the effects of cordycepin on RNA synthesis in eucaryotic cells (Penman et al., 1970) showed that the drug greatly inhibited the accumulation of labelled cytoplasmic mRNA without affecting the synthesis of hnRNA. The subsequent finding that cordycepin prevented polyadenylation of hnRNA (Darnell et al., 1971b) suggested that poly(A) addition was required for some step in the production of mRNA from hnRNA. It might play a role in some subsequent obligatory step in mRNA processing, or it could conceivably be directly involved in nucleocytoplasmic transport. Nevins and Darnell (1978) observed that late adenovirus transcripts were polyadenylated within about one minute of transcription of their poly(A) sites, and that the first polyadenylated transcripts to be produced were unspliced. Transcripts of SV40 (Lai et al., 1979), the ovalbumin gene (Tsai et al., 1980) and a globin gene (Grosveld et al., 1981) are also polyadenylated before being spliced. In view of these observations and of the fact that no nonpolyadenylated, spliced mRNA was known, Zeevi et al. (1981) asked whether or not polyadenylation might be a prerequisite for splicing.
They treated adenovirus-infected HeLa cells briefly with cordycepin so as to completely inhibit polyadenylation while allowing transcription to continue at near-normal rates for the duration of the experiment. Pulse-labelled nuclear RNA was isolated from the cells, and adenovirus ElB and E2 transcripts were selected by hybridization to filter-bound, cloned DNA. Poly(A) could not be detected in the hybrid-selected transcripts, but in both cases, normally spliced transcripts were detected. These experiments did not exclude the possibility that splicing was dependent on correct transcript cleavage at poly(A) sites, but they showed that polyadenylation itself was not a prerequisite for splicing. The more recent demonstration that runoff transcripts terminated at a site remote from known poly(A) sites were substrates for in vitro splicing (Padgett et al., 1983) further suggests that splicing is not obligatorily dependent upon normal 3' end generation. Zeevi, Nevins and Darnell (1982) went on to consider the possibility that polyadenylation was involved in nucleocytoplasmic transport of mRNA. They found that after brief cordycepin treatment of adenovirus-infected cells, nonpolyadenylated, spliced transcripts of the ElA, ElB and E2 adenovirus transcription units accumulated in the cell nuclei at about the same rate as the normal transcripts for about 40 minutes. These nonpolyadenylated adenovirus mRNAs could be detected in total cytoplasmic or polysomal RNA, but they accumulated in the cytoplasmic compartment at a much lower rate than normal, poly(A)^+ transcripts in untreated cells. The shorter the period of labelling, the smaller was the difference between ElB cytoplasmic mRNA levels in cordycepin-treated and untreated cells, suggesting that nucleocytoplasmic transport itself was not
inhibited by cordycepin treatment, but that the resulting poly(A)$^-$ transcripts were unstable in the cytoplasm. The authors were led to conclude that poly(A) has no obligatory role in the cell nucleus, and that its sole function may be to stabilize cytoplasmic mRNA. They went so far as to propose that the apparent selection of certain hnRNA sequences for transport to the cytoplasm may be a consequence of the rapid degradation of nonpolyadenylated sequences after transport. However, they could not exclude the possibility that non-polyadenylated nuclear RNA sequences are subject to intranuclear degradation, or that the transport apparatus normally exhibits a preference for polyadenylated sequences. Although nonpolyadenylated transcripts did enter polysomes, they appeared to do so about half as efficiently as polyadenylated molecules. Whether this was a consequence of the reduced stability of the poly(A)$^-$ transcripts, or whether it was indicative of a role for poly(A) in facilitating polysome assembly, was not clear.

Deadenylation of purified, polyadenylated mRNAs failed to prevent their translation in cell-free systems, which argued against an essential role for poly(A) in ribosome binding or protein synthesis. (For reviews see Brawerman, 1976; Littauer and Soreq, 1982.) However, Huez et al.(1974) demonstrated that poly(A) is important to the functional stability of purified globin mRNA injected into *Xenopus* oocytes. Deadenylation of a sample of the normal, poly(A)$^+$ globin message was carried out by treatment with polynucleotide phosphorylase. Both the polyadenylated mRNA and its deadenylated counterpart were translated upon injection into oocytes, but translation of the poly(A)$^+$ mRNA continued for longer periods than that of the poly(A)$^-$ message. The difference in
translational activity of the two mRNAs was detectable after one hour, and the functional half-life of the poly(A)− mRNA was estimated at 5-10 hours. Marbaix et al. (1975) measured the levels of poly(A)+ and poly(A) mRNA remaining 56 hours after injection into oocytes and found that while poly(A)+ mRNA remained at about the same level as immediately after injection, the amount of poly(A)− mRNA had declined to 15% of its original level. To prove that the reduced stability of the deadenylated RNA was a direct effect of the removal of its poly(A) tail, and not of the possible removal of adjacent sequences, Huez et al. (1975) readenylated the poly(A)− mRNA, adding about 30 adenosine residues per 3' end with E. coli ATP-RNA adenyl transferase. The readenylated sample, like authentic globin mRNA, supported globin synthesis in Xenopus oocytes for at least 48 hours after injection. The behaviour of authentic, deadenylated, and readenylated globin mRNAs in HeLa cells was later shown to be similar to their behaviour in oocytes (Huez et al., 1981).

The length of the poly(A) segment proved to be of importance to the functional stability of globin mRNA oocytes: while a partially deadenylated mRNA sample retaining on average 32 adenosine residues per 3' end was as stable as authentic globin mRNA, a sample with only 16 adenosine residues per 3' end was no more stable than poly(A)− mRNA (Nudel et al., 1976).

The discovery that poly(A) shortening takes place normally in the cytoplasm (Sheiness and Darnell, 1973) and that both the shortening process and mRNA turnover could be slowed by an inhibitor of translation (Sheiness et al., 1975) led to speculation that the rate of poly(A) shortening might control the rate of mRNA turnover. Darnell and his
colleagues (Sheiness et al., 1975) suggested a model in which poly(A) decayed by random endonucleolytic cleavage, and a poly(A) tract of at least a certain threshold length was required to protect an mRNA molecule from degradation. The above-mentioned data of Nudel et al. (1976) suggested that for globin mRNA in oocytes, the threshold length of poly(A) might be about 30 residues.

The results of Huez et al. (1977) also hinted at a relationship between translation and mRNA degradation, since deadenylated globin mRNA proved to be less stable in oocytes if injected with haemin, which increases the efficiency of translation of the mRNA, than if injected in the absence of haemin. Combined with the earlier results of the same group, this study suggested that the role of poly(A) may be to protect mRNA from translation-associated degradative processes. Poly(A) binding proteins which could conceivably be involved in the protective role of poly(A) have been found, but their true significance remains a matter of speculation (Brawerman, 1981).

Factors other than the mere presence of poly(A) must influence mRNA stability, because different polyadenylated mRNAs can display quite different half-lives (see, for example, Harpold et al., 1981). Whether or not these differences depend on different rates of poly(A) shortening is not known (Wilson et al., 1978). Brawerman (1981) and Littauer and Soreq (1982) invoked differences in primary or secondary structure of 3' untranslated regions to explain the differences in stability of different poly(A)+ mRNAs. If this explanation is correct, it requires further elaboration to account for changes which can occur in the stability of a given mRNA. Yeast histone mRNAs, which are polyadenylated, are more
stable in S phase than in other phases of the cell cycle (Hereford et al., 1981; Osley et al., 1981). Prolactin induces a 17-25-fold increase in the half-life of casein mRNA in breast tissue explant culture (Guyette et al., 1979), and mitogenic stimulation of arrested fibroblasts is accompanied by a dramatic increase in the half-life of c-myc mRNA (Blanchard et al., 1985).

In at least one case, neither the poly(A) tail nor the adjacent 3' untranslated sequences seem to affect mRNA stability, in Xenopus oocytes. Sehgal et al. (1978) reported that deadenylated fibroblast interferon mRNA exhibited the same functional stability as the intact mRNA in oocytes. Large stretches of 3' untranslated sequences could also be removed without affecting the stability of the interferon mRNA molecules (Weissenbach et al., 1980). At present, it seems that the only safe conclusion regarding the function of polyadenylate in mRNA is that it stabilizes the mRNA in the cytoplasm—sometimes.

3' End Generation as a Regulatory Device in Eucaryotes

The physical compartmentalization of the eucaryotic cell prevents coupling between transcription and translation, and it therefore excludes the possibility of regulating gene expression by means of polar effects or attenuators of the sort found in bacteria. Nonetheless, 3' end generation is important in regulating the expression of some genes in eucaryotes. Regulation is achieved either by controlling the efficiency of termination, or by controlling the efficiency of polyadenylation at two or more alternative sites. The first case is formally similar to bacterial attenuation in that the expression of genes downstream of a terminator depends upon the efficiency of that terminator. The second
case controls the relative level of two or more different pre-mRNAs from a given transcription unit. The transcripts may subsequently undergo different processing pathways to produce different mRNAs encoding different polypeptides.

Control Over Transcript Termination

At least two cases of gene regulation by means of transcript termination have been described in eucaryotes and their viruses. The adenovirus major late transcription unit is active early in infection as well as after the commencement of DNA synthesis. Transcripts initiated at the major late promoter at early stages of infection do not extend beyond map unit 99 on the genome, as do those produced at later stages. Nevins and Wilson (1981) mapped the extent of the major "late" transcription unit prior to DNA replication by measuring the sensitivity to UV irradiation of transcription from various regions of the genome. The results suggested that transcripts originating from the major late promoter did not extend much further than map unit 60. Akusjarvi and Persson (1981) isolated pulse-labelled nuclear RNA from infected cells in the presence of a DNA synthesis inhibitor and measured its ability to protect various fragments of the genome from S1 nuclease. A fragment containing the L3 polyadenylation site at about map unit 60 was not protected, suggesting that most of the nuclear transcripts terminated before reaching this point. The results of Nevins and Wilson differed slightly in that they suggested that some mRNAs of the L3 family were produced early in infection, but in any event, termination seemed to prevent the synthesis of the L4 and L5 families. No evidence concerning
the mechanism by which transcript termination is regulated during infection has been reported, but Nevins (1982) and Thomas and Matthews (1980) speculated that transcription from the oppositely-oriented E2 promoter might block transcription from the "late" promoter.

Transcript termination may also regulate the expression of the SV40 late genes during lytic infection. The discovery of a 94-nucleotide prematurely-terminated transcript from the SV40 late transcription unit was reported in 1982 by Hay and Aloni (see Termination Sites for RNA Polymerase II Transcription). At the time the authors noted that the 94-nucleotide transcript could potentially fold into either of two alternative secondary structures which were analogous to those available to the leader transcripts of bacterial biosynthetic operons. One structure contained a stem-and-loop followed by a tract of uridylate residues, which resembled a factor-independent terminator. The other structure contained an overlapping stem-and-loop, the formation of which would prevent formation of the terminator. The analogy between the 94-nucleotide RNA and the leader transcript of a bacterial biosynthetic operon could be extended further, in that the 94-nucleotide species could in principle encode a small peptide, the so-called agnoprotein. The authors suggested that the agnoprotein might serve to modulate the formation of the terminator hairpin, thereby regulating the expression of the structural gene downstream (VPl). They specifically postulated that in the nucleus, the agnoprotein would bind to nascent SV40 transcripts, stabilizing the terminator hairpin and preventing transcription of the VPl gene, and that in the cytoplasm, the same type of binding would enhance translation of VPl by masking the agnoprotein initiator codon. It is difficult to see how this mode of regulation would be of any advantage to the virus. Nonetheless, Hay and Aloni
(1985) have provided some circumstantial evidence which supports the idea of a relationship between the synthesis of agnoprotein and the synthesis of the 94-nucleotide prematurely-terminated RNA.

Synthesis of the 94-nucleotide RNA was followed in isolated nuclei, and synthesis of the agnoprotein was monitored by labelling infected cells for 3 hours with $[^{14}\text{C}]$-arginine. Nuclei isolated from cells 48 hours after infection produced the 94-nucleotide RNA, whereas nuclei isolated at the beginning of the late stage of infection (24 hours post-infection) did not. Synthesis of agnoprotein was detected 50 hours after infection, but not at earlier times. Nuclei isolated from several mutant viruses which do not encode agnoprotein exhibited reduced levels of prematurely-terminated RNA. However, since each of the mutations in the viruses tested could have affected the secondary structures available to nascent late transcripts, proof of a causal relationship between agnoprotein synthesis and premature termination is still lacking.

A case of regulated transcript termination with more obvious physiological significance was reported in 1984 by Mather and colleagues. They analyzed nascent transcripts of the immunoglobulin heavy chain locus after limited extension in isolated nuclei and found that the region of transcript termination differed according to the state of differentiation of the cells from which the nuclei were isolated. In B lymphoma lines arrested at early stages of B cell differentiation, transcription proceeded through the adjacent $\mu$ and $\delta$ heavy chain genes, but in IgM-secreting hybridomas and plasmacytomas, transcription terminated between the $\mu$ and $\delta$ coding regions.
Control Over Polyadenylation Site Selection

Selection of alternative polyadenylation sites is also important in regulating the expression of the immunoglobulin heavy chain gene. Immunoglobulin \( \kappa \) chains are produced either in secreted or membrane-bound form. The two differ in their C-terminal regions: the last 21 amino acids of the secreted form are replaced in the membrane-bound form by a 41-amino acid membrane-spanning domain. They are encoded by different mRNAs produced from the same transcriptional unit (Early et al., 1980). The mRNA encoding the secreted form is polyadenylated downstream of the fourth exon of the \( C_\kappa \) gene, while that encoding the membrane-bound \( \kappa \) chain is polyadenylated at the 3' end of the sixth exon. Somehow the 3' end processing apparatus is able to ignore the first polyadenylation site under appropriate conditions and allow the incorporation of the last two exons into the mRNA precursor. The fifth exon encodes the membrane-spanning segment of the membrane-bound \( \kappa \) chain. Selection of each of the alternative poly(A) sites is controlled inasmuch as B lymphocytes express mainly the membrane-bound \( \kappa \) chain, while plasma cells synthesize large amounts of secreted \( \kappa \) chains.

Mather et al. (1984) found that transcription proceeds through all six \( C_\kappa \) exons and the seven \( C_\delta \) exons downstream in B lymphomas expressing surface immunoglobulin. Control over the production of mRNAs encoding \( \kappa_s, \kappa_m, \) and \( \delta \) immunoglobulin heavy chains must therefore be exerted at the level of mRNA processing. In other systems cleavage and polyadenylation precede splicing (Nevins and Darnell, 1978; Lai et al., 1979) so it is assumed that the choice of one of several available
polyadenylation sites restricts the splicing pattern subsequently available to a given transcript. Consistent with this notion, Mather and colleagues observed that \( \delta \) sequences are much less abundant in polyadenylated nuclear RNA from lymphomas expressing only IgM than are \( \alpha \) sequences. It seemed that although transcription proceeded through the \( C_\delta \) exons, polyadenylation at sites adjacent to the \( C_\alpha \) exons precluded the possibility of producing \( \delta \) mRNA. In one lymphoma expressing both \( \alpha \) and \( \delta \) heavy chains, it seemed that both splice site selection and polyadenylation site selection may have been important in controlling the production of \( \alpha \) and \( \delta \) mRNAs. Nuclear transcripts polyadenylated distal to the \( \delta \) exons seemed to have already undergone splicing, suggesting that a rapid splicing event may have influenced selection of a poly(A) site.

Rosenfeld and colleagues have uncovered a clear case of regulation at the level of polyadenylation in the gene encoding calcitonin. Calcitonin is encoded by an mRNA polyadenylated downstream of the fourth exon in the gene. A different mRNA containing two additional exons is produced from the same gene and encodes a previously-undescribed neuropeptide (reviewed by Rosenfeld et al., 1984). The relative levels of the two mRNAs are differentially regulated in thyroid C cells and in various cells of the central nervous system. Amara et al. (1984) showed that transcription continued beyond the second of the two polyadenylation sites in the calcitonin gene in a thyroid tumour cell line which produced calcitonin, but not the related neuropeptide. The same pattern of transcription was detected in a cell line producing only the neuropeptide mRNA, indicating that the differential production of the two mRNAs resulted from differences in mRNA processing, not differences in
transcription. The existence of unspliced or partially spliced transcripts extending from the cap site to either of the two polyadenylation sites suggested that polyadenylation preceded splicing and that 3' end processing, not splicing controlled the relative levels of the two alternative mRNAs.

Poly(A) Sequences in Yeast mRNA

McLaughlin et al. (1973) reported that about 2-3% of the radioactive adenine incorporated into RNA by *Saccharomyces cerevisiae* in a 10-minute period could be recovered in the form of RNase (A+T₁)-resistant polyadenylate. Much of the labelled poly(A) sedimented with polysomes in a sucrose gradient, and it accounted for about 4% of the radioactivity in the polysomal region of the gradient. As the period of labelling increased, the proportion of the polysomal radioactivity which was due to poly(A) decreased; after several generations of labelling and a 2 hour chase, essentially no labelled poly(A) was associated with the polysomes. Together these observations suggested that the polysomal poly(A) was associated with an unstable type of RNA. Disruption of the polysomes in low-Mg buffer released ribosomal subunits and polydisperse ribonucleoprotein complexes which had previously been identified as containing mRNA. Almost all of the poly(A) sedimented with the polydisperse ribonucleoproteins, suggesting that the poly(A) in yeast, as in mammalian cells, was associated with mRNA. The possibility of artifactual association between poly(A) and polysomal mRNA was excluded by experiments with a mutant having a temperature-sensitive defect in
protein synthesis. Polysomes in the mutant were disrupted by a temperature shift and concomitant with their disruption, the sedimentation rate of the labelled poly(A) decreased to coincide approximately with that of free ribosomes. Unlabelled polysomes did not affect the sedimentation rate of the labelled poly(A) when added to the extract of temperature-shifted mutant cells prior to centrifugation.

Poly(A) which had been labelled for a 10-minute period in spheroplasts of normal yeast cells was isolated from the rest of the labelled RNA by digestion with RNases A and T1. Upon alkaline hydrolysis, it released AMP and adenosine in the ratio of about 50:1, suggesting that it was about 50 nucleotides long and located at the 3' ends of mRNA molecules. Its electrophoretic mobility also suggested a length of about 50-70 nucleotides. Groner et al. (1974) estimated the average length of pulse-labelled yeast poly(A) to be about 60 nucleotides, which is about one-quarter as long as newly-synthesized poly(A) in mammalian cells (Sheiness and Darnell, 1973). After a 100-minute period of labelling, which would be expected to completely label the mRNA population (Hutchison et al., 1969), the poly(A) had a broader size distribution, ranging from 60 nucleotides to 20 or perhaps fewer (Groner et al., 1974). [RNase-resistant poly(A) was purified by binding to oligo(dT) cellulose prior to electrophoretic analysis, and the authors cited control experiments showing that poly(A) containing fifteen adenylate residues failed to bind to oligo(dT) cellulose.] The finding that poly(A) tracts in steady-state mRNA tended to be shorter on average than those in pulse-labelled mRNA suggested that poly(A) may undergo gradual shortening after synthesis, as it does in mammalian cells.
About 64% of the polysomal RNA labelled in a 3-minute pulse was found to bind to oligo(dT) cellulose (Groner et al., 1974). The material which failed to bind included a substantial amount of 18S rRNA as well as a small amount of heterogeneous RNA. McLaughlin et al. (1973) also detected polydisperse polysomal RNA which would not bind to oligo(dT) cellulose, and they suggested that some yeast mRNAs may lack poly(A). The poly(A) content of the material which failed to bind to oligo(dT) cellulose was indeed very low, but the assay itself depended upon binding to oligo(dT) cellulose and therefore did not exclude the possibility that very short poly(A) chains may have been present. The possibility that molecules which genuinely lacked poly(A) were the products of rapid processing of polyadenylated molecules also remained open. Hereford and Rosbash (1977) determined the sequence complexity of yeast RNA by measuring the rate and extent of its hybridization to genomic DNA and to cDNA prepared against polyadenylated yeast RNA. Polyadenylated yeast RNA, isolated on oligo(dT) cellulose, saturated the same proportion (20%) of a single-copy yeast DNA tracer as did total yeast RNA, suggesting that essentially all of the sequence complexity of yeast RNA is represented amongst the polyadenylated sequences. If any RNA exists in exclusively non-polyadenylated form in yeast, it must constitute a low-complexity sequence class. Both polysomal polyadenylated RNA and total polyadenylated RNA hybridized at the same rate and to the same extent to cDNA probes prepared against poly(A)⁺ RNA, which further suggested that the mRNA population of the yeast cell contains all of the sequences found
in total polyadenylated yeast RNA.

The transcripts of many specific yeast genes have since been shown to bear poly(A) tails, but perhaps the most telling observations regarding the ubiquity of poly(A) in yeast mRNA were those of Fahrner et al. (1980). These authors found that yeast histone mRNAs bound to oligo(dT) cellulose almost quantitatively and were susceptible to reverse transcription from an oligo(dT) primer. The observation that yeast histone mRNAs contain poly(A) makes the supposition that all yeast mRNAs are polyadenylated seem quite plausible.

How the 3' ends of yeast mRNAs or their precursors are generated prior to their polyadenylation is not known. Sequence complexity measurements (Hereford and Rosbash, 1977) and the size distribution of rapidly-labelled yeast RNA (Groner et al., 1974) both suggest that large mRNA precursors are not common in yeast. However, neither type of result rules out the possibility that in yeast, as in higher eucaryotes, transcription may proceed beyond poly(A) sites, and that mature mRNA 3' ends may be the products of rapid processing reactions. Because the mechanism of mRNA 3' end generation in yeast is not known, signals involved in the process will simply be referred to as 3' end signals.

Comparison of the 3' untranslated and flanking sequences of various yeast genes gives an indication that 3' end signals in yeast might be different from those in other eucaryotes. The hexanucleotide AATAAA is found downstream from the coding sequences of some but by no means all yeast genes (Bennetzen and Hall, 1982; Zaret and Sherman, 1982), in contrast to its near-universal occurrence near poly(A) sites in the genomes of other eucaryotes.
One yeast gene which is not flanked by the sequence AATAAA is CYCl, which encodes the more abundant of the two iso-cytochromes c of yeast (Sherman et al., 1966). It has been subjected to more intensive study than any other yeast gene. CYCl was cloned in 1978 by Montgomery et al., who identified it in a library of cloned yeast DNA fragments by virtue of its hybridization to a synthetic oligonucleotide. Synthesis of an oligonucleotide complementary to 13 nucleotides of the CYCl coding region was possible because Stewart and Sherman (1974) had deduced 44 bp of the CYCl coding sequence from the amino acid sequences of the products of cycl frameshift alleles. The sequence of the entire coding region of the gene and 250 and 280 bp of its 5′ and 3′ flanking sequences, respectively, was reported in 1979 by Smith et al. The authors noted the absence of the hexanucleotide AATAAA from the 280 bp following the CYCl coding sequence but recognized the possibility that the 3′ end of the CYCl transcript might be encoded further downstream. However, Boss et al. (1981) reported the sequence of the CYCl transcript, which showed that the junction between transcribed sequences and polyadenylate tail was located 172-175 nucleotides beyond the coding region. If, as seemed likely, the CYCl 3′ end signal resided near the 3′ end site, then it could not include the sequence AATAAA.

Zaret and Sherman (1982) found that the cycl-512 allele, one of about 500 cycl alleles characterized by Sherman and his colleagues (Sherman et al., 1974), lacked a 38 bp sequence found downstream of the coding sequence in the wild-type allele. Mutants carrying the cycl-512 allele produced 5-10% of the normal amount of iso-1-cytochrome c. Whereas the transcript of the CYCl+ allele was about 630 nucleotides long,
including its poly(A) tail, cycl-512 mutants produced a family of discrete transcripts, ranging in length from 630 to about 2400 nucleotides, which could be detected by a CYCl probe. The aberrant transcripts extended beyond the normal poly(A) site of CYCl mRNA, which indicated that the 38 bp deletion in the cycl-512 allele disrupted a signal required for 3′ end generation at that site. All of the transcripts of the cycl-512 allele bound almost quantitatively to poly(U)-Sepharose, suggesting that they were polyadenylated. Zaret and Sherman proposed that the deletion in the cycl-512 allele impaired transcript termination at the normal poly(A) site, and that the extended transcripts were the products of coupled termination and polyadenylation at a series of downstream sites. The iso-1-cytochrome c deficiency of the cycl-512 mutant correlated with the lower abundance of CYCl transcripts in the mutant as compared to the wild-type strain. Zaret and Sherman suggested that the extended transcripts of the cycl-512 allele were less stable than normal CYCl mRNA. A somewhat different point of view is that the extended, polyadenylated transcripts seen in the steady-state RNA represent the few stable products of a gene with very inefficient transcriptional 3′ end signals. Several cases have been described of mutations in other eucaryotes which disrupt 3′ end signals and concomitantly limit the expression of the corresponding genes by preventing the efficient synthesis of stable transcripts (Higgs et al., 1983; Orkin et al., 1985). A third possibility is that the rate of transcription of the cycl-512 allele was actually lower than that of CYCl+. While it seems unlikely that a mutation in the 3′ flanking sequences of a polymerase II-transcribed gene would directly limit the
activity of its promoter, Zaret and Sherman noted that the cycl-512 deletion caused overlapping transcription of cycl-512 and a downstream, oppositely oriented gene which came to be called UTRl (Zaret and Sherman, 1984). The UTRl transcript in CYCl+ yeast is about 1450 nucleotides long, whereas that produced in cycl-512 cells is about 2000 nucleotides long and extends through the CYCl coding region. Transcripts of both CYCl and UTRl are less abundant in cells carrying the cycl-512 allele than in their CYCl+ counterparts, and the authors suggested that overlapping, convergent transcription of the two genes might limit the rate at which they could be transcribed.

Whatever the exact reason for the iso-1-cytochrome c deficiency of cells with the cycl-512 mutation, it proved to be very useful, in that it allowed Kotval et al. (1983) and Zaret and Sherman (1984) to select revertants which synthesized higher levels of cytochrome c. Amongst the mutations found in the revertants were several chromosomal rearrangements which were genetically linked to CYCl, unlinked suppressor mutations in two different genes, and many `strictly local' mutations which behaved genetically as intragenic CYCl mutations.

Zaret and Sherman (1984) analyzed the physical structure of the CYCl alleles in some of the revertants and compared the CYCl transcripts produced by the revertants to those of the cycl-512 mutant. In those revertants with gross genetic aberrations, the sequences flanking the 3' end of the cycl-512 allele had been replaced by other sequences, and the 3' ends of the major CYCl transcripts in each revertant mapped within these new flanking sequences. In each case it seemed that the genetic rearrangement downstream of the CYCl coding region had introduced new 3'
end signals which were stronger than those associated with the cycl-512 allele. Two insertions which had not been detected as rearrangements by genetic analysis similarly introduced new 3’ end signals downstream from the cycl-512 allele.

The remaining revertants produced various subsets of the transcripts found in the cycl-512 mutant or transcripts which were at least similar in size to those found in the mutant. Many of the revertants owed their increased levels of iso-1-cytochrome c to mutations which allowed the most CYCl-proximal 3’ end site to be used more efficiently than in the cycl-512 mutant. Others produced heterogeneous CYCl transcripts with various size distributions, suggesting that control over 3’ end site selection might be quite complex. Revertants which carried sut2 suppressor mutations accumulated the extended, 1650-2400 nucleotide transcripts in higher levels than the mutant, which suggested to the authors that sut2 mutations inhibited the degradation of these hypothetically unstable transcripts.

The complexity and variety of patterns of 3’ end generation observed amongst the revertants analyzed by Zaret and Sherman (1984) give cause for concern that the signals involved in 3’ end generation might exhibit parallel complexity and variety. In an attempt to identify signals involved in transcript 3’ end generation, Zaret and Sherman (1982) compared the sequences downstream of the coding regions of CYCl and a number of other yeast genes. They identified a conserved sequence which in the case of CYCl occurred in the 38 bp region that was lacking from the cycl-512 allele. The consensus for this conserved "terminator" sequence was TAG...TA(T)GT...TTT. Sequences flanking the central element
of this tripartite sequence tended to be A-T rich, with a bias towards having Ts in the mRNA-parallel strand.

No direct test of the role of the consensus in 3' end generation has yet been reported, but Zaret and Sherman (1984) reported the sequences of two revertant alleles derived from cycl-512. Both sequences differed from the mutant allele in a region with some homology to the tripartite "terminator" sequence. Each revertant allele differed at two adjacent positions from the cycl-512, and surprisingly, the differences did not improve homology to any of the three conserved elements of the consensus sequence. Instead they permuted the sequence of an A-T rich tract found between the homologues of the second and third conserved elements.

Henikoff and his colleagues have attempted to locate and identify the sequences required to specify the 3' end of transcripts of a Drosophila gene in yeast. A segment of the Drosophila glycinamide ribonucleotide transformylase (GAR transformylase) gene complements a yeast ade8 mutation, and its transcripts in yeast are polyadenylated at a site near an AAUAAA hexanucleotide, which raised the question of whether yeast recognized the same type of 3' end signal as other eucaryotes.

Henikoff et al. (1983) produced nested deletions which extended toward the GAR transformylase coding sequence for varying distances from a site downstream of the poly(A) site. Deletions which disrupted the 3' end signal caused the production of a 3'-extended 1.5 kb transcripts in place of the 1.0 kb transcript made from the intact gene. The downstream boundary of the 3' end signal was first mapped to a 29 bp region, and the effects of deletions with proximal endpoints within this region showed that the sequence TTTTTATA defined the boundary of the 3' end signal.
Deletions which left this sequence intact allowed normal 3′ end generation, but the removal of even the terminal A residue caused about 50% of the transcripts to extend beyond the normal 3′ end site. Deletions which extended still further upstream reduced the efficiency of correct 3′ end generation, to the point that a deletion extending 8 bp upstream of the TTTTATA sequence abolished correct 3′ end generation. The same octanucleotide occurs in the CYC1, CYC7 and GPD genes of yeast, in each case downstream of the coding region. Suspicions that it might be a general feature of 3′ end signals in yeast were tempered by the finding that 14 other yeast genes lack the sequence. In a continuation of the same study, Henikoff and Cohen (1984) analyzed the effects of nested deletions which approached the 3′ end site from a fixed site upstream. The results were surprising: a deletion which ended only 12 bp upstream of the last A in the sequence TTTTATA produced normal transcripts and only low levels of extended transcripts. Not until the entire octanucleotide and 23 bp downstream were deleted was the synthesis of correctly "terminated" transcripts abolished. In principle this result could be explained by the presence of a duplicated signal, one copy of which is sufficient to cause 3′ end generation. In fact, Henikoff et al. (1983) had noticed that a second copy of the sequence TTTTATA was present 18 bp downstream of the first. Deletions approaching the region from upstream inactivated the 3′ end signal only when they disrupted the second copy of the octanucleotide. This suggested that each octanucleotide comprised part of a separate, at least partly autonomous 3′ end signal, or as the authors had it, terminator signal.

Henikoff and Cohen tested the degree to which the first copy of the
signal could function autonomously by deleting its normal flanking
sequences on both sides, so that the signal resided on a 25-36 bp
fragment in a new sequence context. Surprisingly, the signal functioned
very inefficiently, allowing detectable 3' end generation in only one of
three constructs tested. The normal sites of 3' end generation were
absent from downstream of the signal fragment in each construct.
Re-introducing these sites allowed some improvement in the efficiency of
3' end generation. In general terms, these results indicated that the 3'
end signal as delimited by deletion analysis was not functionally
autonomous. At the very least, its activity depended upon the presence
of suitable 3' end sites downstream. In their earlier study, Henikoff et
al. (1983) had noted a distinct sequence preference in the precise sites
at which 3' end generation occurred. Sites with the sequence CAA...CTTTG
seemed to be preferred, although their use was not exclusive of 3' end
generation at other sites. Roughly constant spacing was maintained
between the 3' end signal and the sites of 3' end generation in a given
deletion mutant, suggesting that the factors responsible for 3' end
generation might recognize the signal and act if a suitable site could be
found within a certain range of signal. The range in the case of the
Drosophila GAR transformylase gene in yeast seemed to be 50-90
nucleotides.

Neither the signals involved in 3' end generation in yeast nor the
manner of their involvement can presently be described completely, yet
the facility of genetic manipulation in yeast offers at least a clear
approach toward defining the signals. The study to be described here was
undertaken in an effort to define the limits of the sequences required to
specify the 3' end site of CYCl mRNA, and to develop a genetic screen for
defects in 3' end generation. Coupled with an efficient means of mutagenizing the CYCl 3' end signal region, such a screen should make it possible to identify precisely the sequences which constitute the 3' end signal. Mutations which complement defects in 3' end generation should also be detectable using the same screen and amongst them might be mutations in genes whose products interact with 3' end signals. Gaining access to these genes and their products would be an important step in determining how the process of 3' end generation takes place in yeast.
MATERIALS AND METHODS

REAGENTS

Enzymes

Restriction endonucleases used in this work were from Bethesda Research Laboratories (BRL) or New England Biolabs (NEBL). DNA polymerase I (Klenow fragment), T4 DNA ligase, T4 polynucleotide kinase, bacterial alkaline phosphatase, and RNase T1 were also from BRL. Exonuclease III and nuclease BAL31 were from NEBL. S1 nuclease was from P-L Biochemicals, RNaseA and lysozyme were from Sigma, proteinase K was from Boehringer Mannheim, and Glusulase was supplied by Endo.

The conditions under which each enzyme was used are described later in this chapter. Units are those defined by the supplier.

Oligonucleotides

Table I lists the sequences and suppliers of the oligodeoxyribonucleotides used in this work.

Oligonucleotides synthesized in this laboratory were purified by electrophoresis in 20 x 40 cm x 0.2 mm thick gels containing 20% acrylamide, 0.67% bis-acrylamide, and 7 M urea in TBE electrophoresis buffer (50 mM Tris base; 50 mM boric acid; 1 mM EDTA; Maxam and Gilbert, 1980). Formamide was added to the crude oligonucleotide in water to a concentration of 50%, and the mixture was heated for 3 minutes at 90°. About 2 A260 units of crude material (in 10 μl) was loaded into each 1 cm wide slot. Electrophoresis was carried out at 1,500 V for 4 - 6 hours, and the bands were visualized by shadowing against fluorescent silica thin-layer plates under UV illumination. The band containing the full length oligonucleotide was excised with a
<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>oAS1</td>
<td>5'-(ATTTATTTGGTTATAG)-3'</td>
<td>P-L Biochemicals</td>
</tr>
<tr>
<td>oAS2</td>
<td>5'-(GTTCTTGATACCTA)-3'</td>
<td>&quot;</td>
</tr>
<tr>
<td>oAS3</td>
<td>5'-(TATAGTTAGTTAGTTTTA)-3'</td>
<td>T. Atkinson (UBC)</td>
</tr>
<tr>
<td>oAS4</td>
<td>5'-(TATAGTTATCTTAGTTAT)-3'</td>
<td>&quot;</td>
</tr>
<tr>
<td>oAS5</td>
<td>5'-(TTAATAATGACTGG)-3'</td>
<td>&quot;</td>
</tr>
<tr>
<td>oAS6</td>
<td>5'-(AATTCAGCTCATATTAAA)-3'</td>
<td>&quot;</td>
</tr>
<tr>
<td>oAS7</td>
<td>5'-(ATGCATGTGCTCTGTAT)-3'</td>
<td>&quot;</td>
</tr>
<tr>
<td>M1 3FP1</td>
<td>5'-(TCAGCGGTTGATACAC)-3'</td>
<td>R. Barnett (UBC)</td>
</tr>
<tr>
<td>M1 3RP1</td>
<td>5'-(TCACACGAAACAGCT)-3'</td>
<td>&quot;</td>
</tr>
</tbody>
</table>
scalpel, cut into small pieces, and transferred to a 1.5 ml Eppendorf tube. The oligonucleotide was eluted by overnight incubation in 500 μl of 0.5 M NH₄OAc, 10 mM MgCl₂. The eluate was transferred to a clean tube, the gel pieces were rinsed with 500 μl of the same solution, and the rinse was combined with the eluate. Repeated extraction with n-butanol reduced the volume to about 100 μl. The oligonucleotide was then precipitated by the addition of 1 ml of ethanol, followed by storage at -70° for 30 minutes. The precipitate was collected by centrifugation in an Eppendorf microfuge for 10 minutes at 4°.

**Nucleotides**

2'-Deoxyribonucleoside triphosphates, α-thiodeoxyribonucleoside triphosphates, and 2',3'-dideoxyribonucleoside triphosphates were from P-L Biochemicals. They were dissolved in water at an approximate concentration of 10 mM and the pH was adjusted to 7 by adding Tris from a 50 mM solution. The exact concentration of each stock was determined spectrophotometrically. Stocks were stored frozen at -20°.

α[^32p]-2'-deoxyribonucleoside triphosphates, and γ[^32p]-adenosine triphosphate were purchased from New England Nuclear. They had specific activities of about 3,000 Ci/mmol and were supplied as aqueous solutions containing 10 μCi/μl.

Vanadyl ribonucleoside complexes (VRC) were obtained as a 0.2 M solution from Bethesda Research Laboratories.

**Galactosides**

o-Nitrophenyl β-D-galactoside (ONPG), isopropyl β-D-thiogalactoside (IPTG) and 5-bromo-4-chloro-3-indolyl β-D-galactoside (XGAL) were from Sigma. All were stored at -20°. IPTG was dissolved in water at a concentration of 0.1 M and XGAL was
made up as a 2% solution in dimethylformamide. Both solutions were stored at -20°. ONPG was dissolved as needed.

Phenol

Phenol was purchased from Fisher as a 90% solution and was purified by distillation. It was stored frozen at -20° under an atmosphere of argon. Aliquots were thawed, saturated with TE10:1 (10 mM Tris-HCl, pH 7.5, 1 mM EDTA) and kept at 4° for periods of up to about a month.

Glyoxal

Glyoxal was purchased as a 6 M aqueous stock, which was deionized by repeated passage through BioRad AG501X8(D) resin before use. Deionization was considered complete when, after the glyoxal had been passed through a column of resin three times, the resin retained its blue colour (McMaster and Carmichael, 1980). If the resin changed colour, it was replaced and the procedure repeated.

Formamide

Formamide (99%) was purchased from Aldrich and deionized by stirring with BioRad mixed bed resin AG501-X8-D (1 g/20 ml) for 15 - 20 minutes, then filtering through glass wool to remove the resin. It was stored frozen at -20° in small aliquots.

Formaldehyde

Formaldehyde was purchased as a 37% solution from Fisher. The pH of this solution was about 4. It was stirred with AG501-X8-D mixed bed resin and filtered before being used. This treatment raised its pH to 6.5.

Agarose

Sigma Type 1 agarose was used for analytical agarose gel electrophoresis. Low melting point agarose for preparative gels was
from BRL.

Acrylamide

Acrylamide and bis-acrylamide from BioRad ("electrophoresis purity reagents") were used without further purification.

Components of Culture Media

Bacto-tryptone, Bacto-yeast extract, Bacto-peptone, and Bacto-yeast nitrogen base without amino acids were from Difco. Amino acids, purines and pyrimidines, and vitamins were from Sigma and Calbiochem. Ampicillin was from Ayerst, and chloramphenicol was from Calbiochem.

Others

Nuclease-free bovine serum albumin (BSA) was from BRL.

Technical grade CsCl was from KBI, a division of Cabot Corporation. Other chemicals were of reagent grade where applicable and were used without further purification.

Supplies for Autoradiography

Kodak XRP-1 film was used for autoradiography of gels or filters carrying 32p-labelled nucleic acids. Developer and fixer were from Kodak and were used in accordance with the manufacturer’s instructions. Dupont Cronex intensifying screens were occasionally used.

MICROBIAL STRAINS

Bacteria

The following strains of *E. coli* were used as hosts for recombinant DNA molecules:

*E. coli* RR1, F′, hsdS20, ara-14, proA2, lacY1, galK2, rpsL20, xyl-5, mtl-1, supE44, λ−, was constructed by Bolivar *et al.* (1977).

*E. coli* JM101, Δ(lac,pro), supE, thi, strA, sbcB15, endA, hspR4,
F'traD36, proAB, lacI, lacZ M15, was constructed by Messing (1981).

Yeast

Strain GM-3C-2, described by Faye et. al. (1981), was used as a host for plasmids carrying deletions of the CYCl 3’ end signal. Its genotype was \(\alpha\), leu2-3,112, trp1-1, his4-519, cycl-1, cyp3-1.

Strain RP123, which was used as a host for CYCl: lacZ fusion plasmids, was provided by S. Roeder (Yale). Its genotype was as follows: \(\alpha\), leu2, adel, trp1, met14, ura3.

Strain D311-3A (Sherman et. al., 1966) is a standard CYCl+ strain, with the following genotype: \(a\) lys2-1, his1, trp2.

CULTURE MEDIA AND CONDITIONS

E. coli

The following media were used for the growth of E. coli:

- LB-glucose 1.0% Bacto-tryptone, 0.5% Bacto-yeast extract, 0.5% NaCl, 0.1% glucose
- YT 0.8% Bacto-tryptone, 0.5% Bacto-yeast extract, 0.5% NaCl
- 2YT 1.6% Bacto-tryptone, 1.0% Bacto-yeast extract, 0.5% NaCl (Sanger et. al., 1980)
- M9S 50 mM Na₂HPO₄, 25 mM KH₂PO₄, 8.5 mM NaCl, 20 mM NH₄Cl, 1 mM MgSO₄, 0.1 mM CaCl₂, 10 mM glucose, 0.001% thiamine. (Miller, 1972)

For plates, Bacto-agar was added to the appropriate liquid medium at 20 g/l. Soft agar for overlays contained 7 g Bacto-agar/l.

Strains containing plasmids were grown in media containing 50 µg/ml
ampicillin.

To screen Ml3 recombinants for insertional inactivation of the lacZ α-peptide coding region, infected cells were plated in soft YT agar, supplemented with 0.03% XGAL and 0.3 mM IPTG.

E. coli was cultured at a temperature of 37°C. Growth was monitored by measuring OD600 on a Bausch and Lomb Spectronic 21 spectrometer.

Yeast

Media for the culture of yeast have been described by Sherman et al. (1981). The following were used in this study:

- **YPD**: 2% Bacto-Peptone, 1% Bacto-Yeast extract, 2% glucose, pH 5.8.
- **YPG**: Same as YPD except 3% glycerol substituted for glucose.
- **YPL**: Same as YPD except 3% lactic acid substituted for glucose.
- **SD**: 0.7% Bacto-Yeast Nitrogen Base without amino acids, 2% glucose, pH 5.8.
- **SC**: SD, with the following supplements: adenine, uracil, histidine, arginine, methionine, tryptophan (20 mg/1 each); lysine, isoleucine, tyrosine, leucine (30 mg/1 each); phenylalanine (50 mg/1); sodium glutamate, sodium aspartate (100 mg/1 each); valine (150 mg/1); threonine (200 mg/1); serine (375 mg/1).

For plates, the above media were solidified by adding agar to a concentration of 2%.
R (regeneration) agar
SD supplemented with 1 M sorbitol, 2% YPD, 3% agar.

RC (complete regeneration) agar
SC, supplemented with 1 M sorbitol, 2% YPD, 3% agar.

Yeast Ca-Free XGAL agar (Ruby et al., 1983)
0.1 M (KH₂PO₄ + K₂HPO₄, pH 7); 15 mM (NH₄)₂SO₄; 1 mM MgSO₄, 2uM FeCl₃; 0.11 M glucose; adenine, uracil, and amino acids as in SC; thiamine, pyridoxine, pantothenic acid (0.4 mg/l each); biotin (2 ug/l); myo-inositol (2 mg/l); XGAL (40 mg/l); 2% agar.

Yeast cultures were incubated at 30°C. Liquid cultures were shaken on a rotary platform at 200 rpm. Growth in liquid was monitored by measuring OD₅₅₀ on a Bausch and Lomb Spectronic 21 spectrometer.

PLASMIDS and BACTERIOPHAGE

Bacteriophage and plasmids constructed during the course of this study are described in detail in later sections of this chapter. Those provided by other workers are listed in Table II with references to a published description of each.

TRANSFORMATION of E. coli

Competent cells of E. coli were prepared and transformed as described by Dagert and Ehrlich (1979). A single colony of strain RRL or JM101 was used to inoculate 3 ml of LB-glucose or M9S, respectively, and the culture was incubated overnight at 37°C. The next morning, 50 ml of the same medium was inoculated with 0.1 - 0.5 ml of the overnight culture and incubated until the culture had an OD₆₀₀ of 0.1 to 0.15. The culture was chilled on ice for 10 minutes and the cells were
<table>
<thead>
<tr>
<th>Name</th>
<th>Reference</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBR322</td>
<td>Bolivar et al. (1977)</td>
<td>D.W. Russell (UBC)</td>
</tr>
<tr>
<td>pYeCYCl(2.5)</td>
<td>Faye et al. (1981)</td>
<td>S. Bektesh</td>
</tr>
<tr>
<td>YEpl3</td>
<td>Broach et al. (1979)</td>
<td>S. Bektesh</td>
</tr>
<tr>
<td>pYeCEN3(41)</td>
<td>Clarke and Carbon (1980)</td>
<td>S. Bektesh</td>
</tr>
<tr>
<td>2H26, 4H40</td>
<td>Faye et al. (1981)</td>
<td>G. Faye</td>
</tr>
<tr>
<td>PMC1403</td>
<td>Casadaban et al. (1982)</td>
<td>P. Dennis (UBC)</td>
</tr>
<tr>
<td>M13mp vectors</td>
<td>Messing (1983)</td>
<td>J. Messing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(U. of Minnesota)</td>
</tr>
<tr>
<td>pUC13</td>
<td>Messing (1983)</td>
<td>J. Messing</td>
</tr>
<tr>
<td>pEMBL vectors</td>
<td>Dente et al. (1983)</td>
<td>R. Cortese</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(EMBO)</td>
</tr>
<tr>
<td>IRL</td>
<td>Dente et al. (1983)</td>
<td>R. Cortese</td>
</tr>
</tbody>
</table>
harvested by centrifugation at 5,000 rpm in a SS-34 rotor at 4°C for 10 minutes. The cell pellet was resuspended in ice-cold 0.1 M CaCl₂ (50 mM CaCl₂ was sometimes used with strain JM101; either concentration seemed to be equally effective) and kept on ice for 20 - 30 minutes. The cells were harvested by centrifugation as before and resuspended in 1 - 2 ml of ice-cold 0.1 M (or 50 mM) CaCl₂. They could then be used in transformation, but they were usually kept on ice for 12 - 18 hours prior to transformation since this resulted in higher transformation efficiency, as reported by Dagert and Ehrlich (1979).

Plasmid or phage DNA (up to 0.1 μg in 10 μl or less) was added to 0.1 ml aliquots of competent cells in glass culture tubes and the cells were left on ice for 20 - 40 minutes. They were transferred to a 42°C water bath for 1 - 2 minutes and removed to room temperature. LB-glucose (1 ml) was added to each tube, and they were incubated for 30 - 60 minutes at 37°C on a shaker. Aliquots of 50 - 100 μl were then plated on appropriate selective or indicator media. The plates were incubated at 37°C overnight, and the remainder of the transformed cell suspension was saved at 4°C in case it was necessary to plate out more samples.

**ISOLATION of PLASMID DNA from E. coli**

Two different procedures were used for the preparation of plasmid DNA from *E. coli*, the first relying on cell lysis by Triton X-100, the second using alkali and SDS to lyse the cells. Both procedures are described below, as applied to cultures of 50 - 500 ml and as used for "minipreps" from 1 - 5 ml cultures. Both were satisfactory, but the alkaline lysis procedure was much faster and somewhat more reliable.
Large-Scale Plasmid Isolation

A single ampicillin-resistant colony was used to inoculate 5 ml of LB-glucose-Ap and the culture was incubated overnight at 37°. M9S medium (500 ml), supplemented with leucine (0.4 mg/ml) and proline (0.8 mg/ml) to allow growth of *E. coli* RRL derivatives, was inoculated with 2.5 ml of the saturated culture and incubated at 37° until it attained an OD600 of 0.2-0.25. Chloramphenicol (100 mg/ml in ethanol) was added to a concentration of 0.2 mg/ml, and incubation was continued for 14-18 hours. Cells were harvested by centrifugation at 4° in a Sorvall GSA rotor at 8,000 rpm for 10 minutes, resuspended in a total of 50 ml of TE (10 mM Tris-HCl pH 7.5: 1 mM EDTA) and centrifuged as before. The pellet was frozen at -20°.

Triton Lysis Procedure

This procedure is from Davis et al. (1980). The frozen cell pellet from a 500 ml culture was resuspended in 10 ml of ice-cold sucrose:TE (50 mM sucrose; 35 mM Tris-HCl pH 8; 100 mM EDTA), and 2 ml of a freshly prepared solution of lysozyme (10 mg/ml in sucrose:TE) was added. The suspension was kept on ice for 5-10 minutes, and 4 ml of 0.25 M EDTA was then added with gentle mixing. The suspension was put at room temperature, and 0.5 ml of a solution of RNaseA (2 mg/ml in 50 mM NaOAc, pH5) which had been boiled for 10 minutes and cooled to room temperature was added. After 5 minutes, the suspension was added slowly to 20 ml of Triton mix (1.0% Triton X-100; 50 mM Tris-HCl pH 8; 15 mM EDTA) with gentle mixing. The lysate was centrifuged at 30,000 rpm for 1 hour at 10° in a Beckman 35 rotor in a Beckman Model L3-40 centrifuge. The supernatant was carefully transferred to a clean, 150 ml Corex glass bottle. An equal volume of TE-saturated phenol was
added and after 10 minutes of mixing, the phases were separated by centrifugation in a Sorvall GSA rotor at 5,000 rpm for 10 minutes at 4°. The upper aqueous phase was removed to a new 150 ml bottle, and 2 - 2.5 volumes of ethanol were added. The mixture was kept at -20° overnight, and the precipitate was pelleted by centrifugation in a GSA rotor at 5,000 rpm for 20 minutes at 4°. After brief drying under vacuum, the pellet was dissolved in TE50:10 (50 mM Tris-HCl pH8: 10 mM EDTA). The plasmid DNA was further purified by CsCl gradient centrifugation, as will be described below.

**Alkaline Lysis Procedure**

This procedure was described by Maniatis et al. (1982) and is a modification of the method of Birnboim and Doly (1979). The frozen cell pellet from a 500 ml culture was resuspended in 10 ml of Glucose-TE (50 mM Glucose; 25 mM Tris-HCl pH 8.0; 10 mM EDTA) and powdered lysozyme was added to a concentration of 5 mg/ml. The suspension was kept at room temperature for 5 minutes; and 20 ml of an ice-cold solution of NaOH (0.2 M) and SDS (1%) was added. The suspension was mixed by inverting gently several times and kept on ice for 10 minutes before adding 15 ml of a ice-cold solution containing 3 M KOAc and 2 M HOAc (pH approximately 4.5). The lysate was mixed by inverting several times and kept on ice for 10 minutes. It was then centrifuged for 60 minutes at 4° in a Sorvall SS-34 rotor at 20,000 rpm. The supernatant was transferred to a Corex glass bottle and extracted with an equal volume of phenol/chloroform (1:1). The phases were separated by centrifugation as described above. The aqueous phase was transferred to a new bottle and one-half volume of isopropanol was added. The mixture was kept at room temperature for 30 minutes and...
then centrifuged at room temperature in a GSA rotor for 20 minutes at 5,000 rpm. The pellet was rinsed by adding about 50 ml of ethanol and repeating the last centrifugation step. The pellet was dried briefly under vacuum and dissolved in TE50:10 in preparation for CsCl gradient centrifugation.

**Purification of Plasmid DNA by Cesium Chloride Gradient Centrifugation**

The plasmid DNA isolated from a 500 ml culture by either the Triton lysis or alkaline lysis procedure was split into two equal portions. The volume of each portion was adjusted to 9.5 ml with TE50:10. Cesium chloride (9.4 g) was added to each portion. After the CsCl had dissolved, the solution was transferred to a Beckman quick-seal polyallomer tube. Approximately 0.9 ml of a solution of ethidium bromide (10 mg/ml in H2O) was added to each tube, allowing a small air space to remain beneath the stem of the tube. With plasmids prepared by the alkaline lysis method, ethidium bromide was added before transferring to a quick-seal tube, and the solution was centrifuged at 9,000 rpm in a Sorvall SS34 rotor for 5 minutes to remove the purple precipitate that formed. The tubes were sealed using the Beckman tube sealer, inverted a few times to mix the contents thoroughly and then centrifuged in a Beckman 50Ti rotor at 37,000 rpm for 40 hours at 10° in an L3-40 centrifuge, or in a 70.1 Ti rotor at 65,000 rpm for 16 hours at 20° in an L8-70 centrifuge. After centrifugation, the tubes were mounted on a retort stand and illuminated with an ultraviolet lamp. Two bands of DNA were generally visible, although the upper band was often very faint, especially when the alkaline lysis procedure had been used. The lower band, containing
supercoiled plasmid DNA, was withdrawn by means of an 18-guage hypodermic needle inserted through the side of the tube just below the position of the band. The needle was removed from the syringe before expelling the DNA into a 15 ml Corex glass tube. Ethidium bromide was removed from the DNA solution by adding an equal volume of n-butanol, mixing, allowing a minute or two for the phases to separate and removing the upper (butanol) phase with a Pasteur pipet. The extraction was repeated 5 or 6 times, or until all traces of ethidium bromide (pink) had disappeared. The solution was diluted with two or three volumes of water, and ethanol (2.5 times the total aqueous volume) was added to precipitate the DNA. The mixture was chilled at -20° for 4-24 hours and the precipitate was collected by centrifugation in a Sorvall SS-34 rotor at 9,000 rpm for 20 minutes at 4°. The supernatant was discarded, and the precipitate was rinsed by adding about 5 ml of cold 70% ethanol and repeating the centrifugation. The precipitate was dried briefly, dissolved in 0.4 ml of water and transferred to a 1.5 ml Eppendorf tube. Sodium acetate (40 µl of a 3 M stock) and ethanol (1.0 ml) were added, and the tube was chilled at -70° for 30 minutes. The precipitate was collected by centrifugation for 5 minutes at 4° in an Eppendorf centrifuge, rinsed with 1 ml cold 70% ethanol, centrifuged for 2 minutes at 4°, dried, and dissolved in 0.5 - 1.0 ml TE10:1. An aliquot of this solution was diluted and its ultraviolet absorption spectrum measured. The $A_{260}/A_{280}$ ratio, which was 1.8 - 2.0, showed the DNA to be free from protein. Its concentration was estimated from its $A_{260}$, assuming a concentration of 50 µg/ml to have an $A_{260}$ of 1.0 (Davis et al., 1980).
Plasmid "Minipreps": Triton Procedure

The following procedure was originally described by Ferguson et al. (1981).

Single colonies of ampicillin-resistant *E. coli* transformants were picked into 5 ml aliquots of LB-glucose containing ampicillin and incubated overnight at 37°. Cells were pelleted by centrifugation at 4° in a Sorvall SS-34 rotor at 5,000 rpm for 5 minutes. Each pellet was resuspended in 1.2 ml of cold TE10:1, transferred to a 1.5 ml Eppendorf tube, and pelleted again by brief (1 minute) centrifugation in an Eppendorf centrifuge. All subsequent centrifugations in this procedure were carried out with the Eppendorf centrifuge. The pellet was resuspended in 0.45 ml of cold sucrose:TE (50 mM sucrose; 35 mM Tris-HCl pH8; 100 mM EDTA), and 65 μl of a freshly-prepared solution of lysozyme (10 mg/ml in sucrose:TE) was added. The suspension was kept on ice for 10 minutes before adding 0.5 ml of cold Triton mix. After a further 10 minutes on ice, the lysate was heated at 80° for 5 minutes and chilled to 0°. It was kept on ice for 15 minutes and then centrifuged for 15 minutes at room temperature. The clear supernatant was decanted into a new microfuge tube and 2.5 μl of diethyl pyrocarbonate (DEP) was added. The mixture was heated for 15 minutes in a 65° water bath and then centrifuged for 3 minutes at room temperature. The supernatant was transferred to a clean tube and 1 ml of ethanol was added. After 1 hour at -20°, the precipitate was pelleted by centrifugation for 5 minutes at room temperature. The pellet was dried, dissolved in 0.4 ml of water, and 40 μl of 3 M NaOAc and 1 ml of ethanol were added. The mixture was again chilled for 1 hour at -20° and centrifuged for 5 minutes at 4°. The pellet was
rinsed by adding 1 ml of 70% ethanol and centrifuging for 2 minutes and removing the supernatant, dried briefly under vacuum, and dissolved in 50 ul TE10:1.

Alkaline Lysis Procedure

This procedure, described by Maniatis et al. (1982), is a modification of the method of Birnboim and Doly (1979). Aliquots of 2 ml of LB:glucose:ampicillin were each inoculated with a single colony of an ampicillin-resistant transformant and incubated overnight at 37°. A 1.5 ml portion of each culture was poured into a 1.5 ml centrifuge tube, and the cells were harvested by centrifugation for 1 minute in an Eppendorf centrifuge. The pelleted cells were resuspended in 0.1 ml of a freshly prepared solution of lysozyme (4 mg/ml) in glucose:TE (50 mM glucose;25 mM Tris-HCl, pH 8; 10 mM EDTA). The suspension was kept at room temperature for 5 minutes, and 0.2 ml of a freshly-prepared ice-cold solution of 0.2 M NaOH and 1% SDS was added. The suspension was mixed by inverting the tube sharply once or twice and then kept on ice. After 5 minutes, 0.15 ml of a solution containing 3 M KOAc and 2 M HOAc (pH 4.5) was added, and the suspension was mixed by vortexing briefly. After another 5 minutes storage on ice, the tube was spun for 5 minutes at 4°. The supernatant was decanted into another 1.5 ml tube and extracted with an equal volume of phenol/chloroform (1:1). The phases were separated by centrifugation for 1 minute, and the aqueous phase was transferred to a clean 1.5 ml tube. Ethanol was added to fill the tube, and after 5 minutes at room temperature, the tube was centrifuged for 5 minutes at room temperature. The pellet containing plasmid DNA and RNA was rinsed with 70% ethanol, as previously described, then dried briefly under vacuum and dissolved in 50 ul of TE10:1.
Preparation of Plasmid DNA for Sequencing

Plasmid DNA was isolated from 1 ml cultures using the alkaline lysis procedure as previously described. It was prepared for DNA sequencing using a procedure developed by Caroline Beard in this laboratory. Plasmids prepared by the alkaline lysis procedure were redissolved in 200 µl of TE10:1, and 100 µl of 7.5M NH₄OAc was added to each. The mixture was kept on ice for 1-16 hours, after which the precipitate of high molecular weight RNA was removed by centrifugation for 15 minutes at 4°C in an Eppendorf centrifuge. The supernatant was transferred to a clean tube and 0.6 ml of ethanol was added to precipitate the remaining nucleic acid. The precipitate was collected by centrifugation for 5 minutes at 4°C (Eppendorf), rinsed with 70% ethanol, dried and dissolved in 50 µl TE10:1 containing RNaseA (40 µg/ml) and RNase T₁ (40 u/ml). After incubation at 37°C for 2 hours, the solution was extracted once with an equal volume of TE-saturated phenol, once with phenol/chloroform (1:1), and twice with ether. Residual ether was evaporated, and the DNA was precipitated by adding 5 µl 3 M NaOAc, 125 µl ethanol, and chilling to -70°C for 30 minutes. The precipitate was collected by centrifugation for 5 minutes at 4°C, rinsed in 70% ethanol, dried, and dissolved in 50 µl TE10:1. A 5 - 8 µl aliquot of this solution was usually used for DNA sequencing.

Preparation of M13 DNA

Single Stranded Phage DNA

The procedure of Sanger et al. (1980) was followed for the isolation of single-stranded phage DNA from small phage-infected cultures. An overnight culture of E. coli JM101 was grown in M9S broth
from a single colony. An aliquot of this culture was used to inoculate 50-100 volumes of 2YT broth, and the new culture was incubated for about an hour at 37°. Aliquots of 1 ml were transferred to 13x100 mm culture tubes. Using a sterile capillary, a single plaque was picked from a fresh (1-2 days old) plate and transferred to a 1 ml aliquot of the 2YT culture. The infected cultures were incubated at 37° for 4-5 hours with vigorous shaking. Each culture was then poured into a 1.5 ml Eppendorf tube and the cells were pelleted by centrifugation for 1 minute in an Eppendorf centrifuge. The supernatant containing phage particles (up to 10^{12}/ml) was transferred to a clean tube and the pellet was discarded. To the supernatant, 0.2 ml of PEG/NaCl (20% PEG 6,000/2.5M NaCl) was added, and after mixing, the tube was left at room temperature for 15 minutes. The phage precipitated during this interval and were collected by centrifugation for 10 minutes at room temperature in an Eppendorf centrifuge. The supernatant was discarded, and the tube was given a brief (10 second) spin in the centrifuge to cause any supernatant which had been adhering to the walls of the tube to accumulate in the bottom of the tube. This residual liquid was removed with a drawn-out capillary and discarded. The phage pellet was resuspended in 100 μl TE10:1. This phage suspension could be used directly for certain applications, such as screening for hybridization with a \[^{32}\text{P}]-labelled oligonucleotide. If pure phage DNA was required, as for DNA sequencing, the phage suspension was extracted with an equal volume of TE-saturated phenol. The aqueous phase was transferred to a clean tube and residual phenol was removed by repeated ether extraction. Sodium acetate was added to 0.3M and the DNA was precipitated at -70° for 30 minutes after adding 2.5 volumes of
ethanol. The precipitate was collected by centrifugation, rinsed with 70% ethanol, dried and dissolved in 50 μl TE10:1. An aliquot of 5 μl was sufficient for DNA sequencing.

Clone Orientation

The procedure of Winter and Fields (1980) was used in some instances to identify M13 clones which carried the same inserts in opposite orientations. One clone was chosen as a reference, against which others were tested. Phage DNA of the clone to be tested (2μl) and that of the reference clone (2μl) were mixed in a total volume of 10 μl, containing 40 mM Tris-HCl pH 7.5, 4 mM MgCl₂, 0.2 M NaCl, 85% glycerol, 0.1% SDS, and 0.02% bromophenol blue. The mixture was heated at 68° for one hour and quickly chilled in an ice-water bath before loading on a 0.5% agarose gel. After electrophoresis, the gel was stained with ethidium bromide. If the test and reference clones carried the same inserted sequences in opposite orientations, they formed a hybrid of markedly lower mobility than either clone alone. If they did not hybridize, the mobility of each clone in the mixture was unaffected by the presence of the other.

Preparation of M13 RF

The replicative forms of M13 and its derivatives were isolated from the cells of 50 or 500 ml infected cultures which were prepared as follows: a single colony of E. coli JM101 was introduced into 5 ml of M9S medium and incubated overnight. A 50 μl aliquot of this culture was used to inoculate 2 - 3 ml of 2YT. One hour later, a phage plaque was picked into this culture. The remainder of the overnight culture was used to inoculate 500 ml of M9S. Both cultures were incubated at 37° for 5 hours. At this time, the large culture had an A₆₀₀ of
0.3 (about 3 x 10^8 cells/ml). The cells in the small, infected culture were pelleted by centrifugation, and 2 ml of the supernatant (about 10^{12} phage/ml) was added to the large culture. The large infected culture was incubated for 3 hours and the cells were pelleted by centrifugation at 5,000 rpm for 10 minutes in a Sorvall GSA rotor at 50. The supernatant was discarded, and the cell pellets were resuspended and combined in a total of about 100 ml of ice-cold TE10:1. The cells were again pelleted by centrifugation and the pellet was frozen at -200. Phage replicative form DNA was isolated from the cell pellet using either the Triton lysis or alkaline lysis procedure of plasmid purification and further purified by centrifugation in CsCl with ethidium bromide as described earlier.

Preparation of Single-Stranded pA4 Plasmid DNA

Miniprep pA4 DNA was introduced into E. coli JM101, selecting for ampicillin-resistant transformants. The procedure used to isolate single-stranded plasmid DNA was described by Dente et al. (1983). Aliquots of LB:glucose:Amp (2 ml) were inoculated with pA4 transformants of JM101 and incubated until slightly cloudy (A_{600} of 0.1 - 0.2, corresponding to a cell density of 1 - 2 x 10^8/ml). Each culture was then infected with about 10^9 pfu of helper phage IR1. (A stock of this helper phage had been prepared by PEG/NaCl precipitation of the supernatant of an infected 40 ml culture of JM101, in the manner normally used for preparation of M13 phage. The precipitate was resuspended in 20 ml of TE and stored frozen in 1 ml aliquots. It had a titre of 5 x 10^{11} pfu/ml on JM101.) The pA4 transformant cultures were incubated for 6 hours after infection with IR1, after which "phage" were precipitated from the culture supernatants with PEG and
NaCl as for M13. The phage pellets were resuspended, phenol-extracted and the DNA precipitated twice, first from 0.9 M NaClO₄ and 30% isopropanol, and then from 0.3 M NaOAc and 70% ethanol. The final precipitate, containing a mixture of IRl phage DNA and single-stranded pA4 DNA, was dissolved in 30 ul TE for sequencing. For reasons that remain unclear, about 40% of the cultures treated in this manner yielded no single-stranded DNA at all, or less often, yielded only the DNA of the superinfecting phage. Loss of the F' episome from pA4 transformants of JM101 would prevent superinfection with phage IRl and thus prevent production of single-stranded DNA. However, plating the transformants on proline-deficient medium before inoculation of the 2 ml cultures for superinfection, which should have selected for the maintenance of the F' episome, did not improve the success rate for isolating single-stranded DNA. The absence of pA4 DNA from some of the preparations of single-stranded DNA implies selective production or packaging of IRl DNA, not absence of pA4 from the cell, since all media contained ampicillin to select for the presence of pA4). Samples of all preparations were subjected to electrophoresis on agarose minigels to ensure that the preparations used for sequencing did in fact contain pA4 DNA. Preparations containing pA4 single-stranded DNA were sequenced exactly as M13 clones, using M13FP1 as sequencing primer. The IRl DNA present in each preparation did not interfere in the least with sequencing the pA4s.

Silanization of Containers for use with Nucleic Acid Solutions

Glass tubes, bottles and capillaries, and polypropylene microfuge tubes to be used for storing or transferring nucleic acid solutions were silanized by placing in a vacuum desiccator over a tray containing
a few milliliters of a 10% solution of dimethyldichlorosilane in toluene. The desiccator was partially evacuated on a water aspirator and left overnight. The tubes were then baked at 90° for a few hours, rinsed with distilled water, autoclaved and dried. Silanizing made the recovery of small volumes and small quantities of nucleic acid more efficient by preventing aqueous solutions from sticking to glass or plastic surfaces.

**AGAROSE GEL ELECTROPHORESIS of DNA**

Agarose gel electrophoresis was routinely used to display the products of restriction digests. Gels containing 0.5 - 2.0% agarose were prepared by weighing an appropriate quantity of agarose into TBE electrophoresis running buffer (50 mM Tris base; 50 mM boric acid; 1 mM EDTA; pH 8.3), heating to dissolve the agarose, cooling to approximately 50-55°, and pouring onto a clean, dry glass plate. "Minigels", about 2 mm thick, were poured on 5 x 7 cm lantern slides, while 10 x 17 cm plates were used for 100 ml gels. Plastic combs served as moulds for the sample wells, which measured 0.5 x 3 x 1.5 mm deep for minigels, or 1.5 x 5 x 4 mm deep for 100 ml gels. Gels were allowed to set at room temperature. Electrophoresis was carried out horizontally with the gel submerged in TBE buffer to a depth of about 2 mm, at a voltage gradient of 2-5 V/cm. Gels were stained in ethidium bromide (1 μg/ml in H2O) for 15-20 minutes at room temperature and photographed over a UV transilluminator using Polaroid film in a Polaroid MP-4 camera.

**Electrophoresis and DNA Fragment Purification Using LMP Agarose**

Specific DNA fragments were easily purified after electrophoresis of restriction digests in low melting point (LMP) agarose as described
by Maniatis et al., 1982. Gels containing 0.7-1.0% LMP agarose in TBE buffer were prepared in the same way as normal agarose gels. They were poured onto 10 x 17 cm glass plates at 4°, because the LMP agarose did not set well at room temperature. Sample wells of 1.5 x 5 x 4 mm deep to 1.5 x 15 x 4 mm deep were made, according to the quantity of DNA that was to be loaded. Gels were run submerged in TBE at either 4° or room temperature at approximately 2-3 V/cm. They were stained with ethidium bromide in the same way as ordinary agarose gels and DNA bands were observed under UV illumination either from a hand-held lamp or a transilluminator. The desired bands were immediately excised with a scalpel in an effort to minimize their exposure to UV. The excised gel fragments were placed in 1.5 ml Eppendorf tubes and melted by incubating in a water bath at 70° for 10 minutes. Sufficient TE was added to each tube to bring the total volume of liquid to 0.7 ml. If the melted gel slice had a volume greater than 0.35 ml, it was divided between 2 tubes and each portion was diluted to 0.7 ml. An equal volume (0.7 ml) of TE-saturated phenol was added to each tube. After vortexing for about 1 minute, the phases were separated by centrifugation for 1 minute in an Eppendorf centrifuge. The aqueous phase was transferred to a new 1.5 ml tube. Care was taken to avoid the white interfacial material, and if some was transferred with the aqueous layer, a second extraction with phenol was carried out. The aqueous solution was extracted with an equal volume of phenol:chloroform and then with chloroform, each time by adding the organic solvent, mixing, centrifuging to separate the phases, and transferring the aqueous layer to a clean tube. After chloroform extraction the aqueous solution was extracted several (2-5) times with
n-butanol to reduce its volume to about 0.2 ml. A single extraction with ether served to remove residual butanol, and residual ether was evaporated in air at room temperature. Sodium acetate was added to a concentration of 0.3 M, and 2.5 volumes of ethanol were added to precipitate the DNA. The tube was chilled at -20° overnight, and the precipitate was collected by centrifugation for 5 minutes in an Eppendorf centrifuge. The pellet was rinsed by adding 1 ml of cold (-20°) 70% ethanol, centrifuging for 2 minutes and removing the supernatant. The pellet was dried briefly and dissolved in 10-100 ul of TE10:1. The concentration of the purified fragment was estimated by electrophoresing an aliquot through an agarose gel, staining with ethidium bromide and comparing the fluorescence with that of markers of known concentration.

**ACRYLAMIDE GEL ELECTROPHORESIS**

A stock containing 43.5% acrylamide and 1.5% bis-acrylamide in deionized water was prepared and stored at 4° in the dark. Non-denaturing gels were prepared by mixing appropriate volumes of this stock and a ten-fold concentrate of TBE electrophoresis buffer with water, then adding ammonium persulfate to 0.05% from a 1.6% stock and N,N,N',N'-tetramethylethylenediamine (TEMED) to a concentration of 0.05%. The preparation of denaturing gels differed only in that urea was dissolved in the mixture at a concentration of 7M before the addition of catalysts.

Gel plates were washed with soap and water, rinsed and dried. Their inner surfaces were rinsed with 1% dimethyldichlorosilane in toluene, then with ethanol, then dried. Plastic spacers (0.2 - 1.0 mm thick) were placed along the long edges of the front (rectangular)
plate. The back (notched) plate was placed on top of the spacers and the bottom and sides of the assembly were sealed with 3M Type 56 electrical tape. The plates were held on an incline of about 30° while the gel solution was slowly poured down one side of the space between the plates. A plastic comb was clamped into the opening at the top of the gel plates. The teeth of the comb, which formed the sample wells of the gel, measured 6-15 mm wide. Polymerization was allowed to continue for at least 20 minutes with the gel slightly inclined from the horizontal before the comb was removed and the sample wells were thoroughly flushed out with water. The tape along the bottom edge of the gel plates was then removed and the gel was clamped into a vertical electrophoresis apparatus. Aluminum plates (20X30 cm X 3 mm thick) were clamped onto gels which were to be run at high voltage, such as those used for DNA sequencing.

Upper and lower reservoirs were filled with TBE electrophoresis buffer, and the sample wells were flushed immediately before loading. Non-denaturing gels were run at 2-10 V/cm, while denaturing gels were run at 30-40 V/cm.

After electrophoresis, the gel was placed on a benchtop, the tape was removed from the edges of the gel plates, and the plates were separated by inserting a knife between them. The gel usually remained on the bottom plate. It could then be stained in ethidium bromide (1 µg/ml in H2O) for about 30 minutes prior to photography under UV illumination. If the DNA had been radioactively-labelled, the gel was autoradiographed directly, or transferred to Whatman 3MM paper and dried, then autoradiographed. A BioRad gel dryer was used in the latter case.
Autoradiography

Autoradiography of dried gels was carried out at room temperature using Kodak XRP-1 film. Gels which had been neither fixed nor dried were frozen at -20° during autoradiography. Occasionally it was convenient to use an intensifying screen to shorten exposure time. In such cases, autoradiography was allowed to proceed at -70°.

Restriction Endonuclease Digestion

Restriction endonuclease digestion was performed using the buffers recommended by Davis et al. (1980). The composition of each buffer and the enzymes for which it was suitable are indicated in Table III.

Each buffer was made from sterile stock solutions as a 10-fold concentrate which was stored at -20°. Analytical digests were performed on samples of 0.2-1 μg DNA, while quantities of 5-100 μg were digested when it was necessary to purify particular restriction fragments or use the products of the digest in subsequent reactions. The volume of the reaction mixture was chosen to give a DNA concentration of 0.02-0.2 μg/ml and the reaction was carried out in Eppendorf centrifuge tubes. The DNA sample to be digested was generally in TE10:1 to begin with. The appropriate 10X buffer concentrate was added (one-tenth of the final volume of the mixture), along with an equal volume of a sterile solution of nuclease-free bovine serum albumin (BSA; 1mg/ml) and sufficient water to bring the mixture to the correct volume. Miniprep DNA contained large quantities of RNA and analysis of the DNA was facilitated if RNaseA was added to the mixture to a final concentration of 50 μg/ml (from a 10 mg/ml stock in 50 mM NaOAc which had been boiled for 10 minutes and stored in aliquots at -20°). About 0.5-2 u of the appropriate restriction

-120-
TABLE III

BUFFERS FOR RESTRICTION ENDONUCLEASE DIGESTION

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Composition (Final Concentrations, mM)</th>
<th>Enzymes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris-HCl, MgCl₂, NaCl, DTT, KCl</td>
<td>pH 7.5</td>
<td>KpnI, SstI, SacI</td>
</tr>
<tr>
<td>Kpn</td>
<td>10, 10, 1, 1</td>
<td>KpnI, SstI, SacI</td>
</tr>
<tr>
<td>Hin</td>
<td>10, 10, 50, 1</td>
<td>HindIII, BamHI, PstI, AvaI, Hinfl</td>
</tr>
<tr>
<td>Eco</td>
<td>50, 10, 100, 1</td>
<td>XhoI³, SalI³, EcoRI</td>
</tr>
<tr>
<td>Sma</td>
<td>10, 10, 1, 20</td>
<td>SmaI</td>
</tr>
</tbody>
</table>

³ These enzymes also worked in Hin buffer.
endonuclease were then added per μg of DNA using a drawn-out capillary micropipet. The reaction mixture was incubated at 37°. An incubation time of 1 - 4 hours was generally sufficient to ensure complete digestion; preparative digests were sampled periodically and the samples electrophoresed through agarose minigels and stained with ethidium bromide to check the extent of digestion. One-fifth volume of glycerol dye mix (50% glycerol; 0.1 M EDTA; 0.15% bromophenol blue; 0.15% xylene cyanol) was added to those samples or digests which were to be directly analyzed by electrophoresis. If the DNA was to be purified before further treatment or analysis, EDTA was added to the digest to a concentration of 12 mM. Proteins were removed by one or two extractions of the digest with an equal volume of TE-saturated phenol. The phases were separated by centrifugation for 1 minute in an Eppendorf centrifuge at room temperature, and the aqueous phase was transferred to a new tube. Residual phenol was removed by 2 or 3 extractions with ether, and residual ether was removed by evaporation in air at room temperature. Sodium acetate was added to 0.3 M, and the DNA was precipitated by adding 2-2.5 volumes of ethanol and chilling at -70° for 30 minutes. The precipitate was collected by centrifugation for 5 minutes at 4° in an Eppendorf centrifuge, then rinsed by adding cold (-20°) 70% ethanol, centrifuging for 2 minutes and discarding the supernatant. The pellet was briefly dried under vacuum and dissolved in water or TE10:1 as appropriate for the next step in its treatment.

Simultaneous digestion with more than one restriction enzyme was performed whenever necessary as long as all the enzymes had the same salt requirements, or if one enzyme was sufficiently "relaxed" with
respect to its salt requirement to function in the buffer required by another. Digestions involving enzymes with incompatible salt requirements were carried out sequentially, the enzyme with the lower salt requirement being used first in the appropriate buffer. When the first reaction was complete (after 1 - 2 hours, or as judged by minigel electrophoresis of aliquots) the salt concentration of the mixture was appropriately adjusted, and the second enzyme was added.

**PHOSPHATASE TREATMENT of DNA**

Restriction endonuclease-digested vectors were occasionally treated with bacterial alkaline phosphatase in order to prevent their subsequent recircularization during ligation reactions. DNA to be treated with phosphatase was ethanol-precipitated, washed in 70% ethanol as described, and redissolved in 10 mM Tris-HCl pH 8 at a concentration of 0.01-0.02 μg/μl. Fifty units of bacterial alkaline phosphatase were added per microgram of DNA, and the mixture was incubated at 65°C for one hour. The mixture was extracted three times with an equal volume of phenol to ensure complete removal of the enzyme. Ether extraction removed residual phenol, and the DNA could then be used directly in a ligation reaction. It was usually ethanol-precipitated and redissolved in TE10:1 at about 0.1 μg/μl before being used.

**LIGATIONS**

Ligation reactions were usually done in a volume of 10-20 μl with 0.05-1.0 μg of an appropriately-treated vector and a three to five-fold molar excess of insert fragment. Fragments with cohesive single-stranded ends were ligated in L buffer, containing 66 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 10 mM DTT and 1.0 mM ATP. About 0.1 u
of T4 DNA ligase was added, and the mixture was incubated at room temperature for 4-16 hours. Fragments with at least one blunt end were ligated in LK buffer, containing 66 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 15 mM DTT, 1 mM spermidine, 0.2 mg/ml BSA. Incubation proceeded at room temperature for 12 - 18 hours. Ligation mixes were used without further treatment to transform E. coli as described earlier.

**Linker Ligations**

Synthetic oligonucleotide linkers were phosphorylated at 37° in 5 μl of LK buffer (66 mM Tris-HCl pH 7.5; 1 mM spermidine; 10 mM MgCl₂; 15 mM DTT; 0.2 mg/ml BSA). T4 polynucleotide kinase (1-2 u) and γ[^32P]ATP were added to start the reaction, and after 10 minutes, unlabelled ATP was added to a final concentration of 1.0 mM. Incubation continued for 30 minutes.

DNA to which linkers were to be attached was rendered blunt-ended by incubation for 30 minutes at room temperature in 10-20 u of LK buffer containing 50 μM dNTPs and 1 u of DNA polymerase I (Klenow). The linker phosphorylation mixture was added directly to this reaction mixture. T4 DNA ligase (2.5 u) and ATP (1.0 mM) were added, and incubation was continued at room temperature for 12-18 hours.

**Assaying Linker Ligation and Removal**

To assess the efficiency of linker ligation and of the subsequent removal of excess linkers, samples were removed before and after the ligation reaction, and after restriction digestion of the ligated products. These samples were analyzed by electrophoresis through a 20 x 40 cm x 0.2 mm thick gel containing 7% acrylamide, 0.23% bis, at 10 V/cm for 2 hours. Autoradiography revealed a single band corresponding to monomeric linkers, and a ladder of bands corresponding to polymeric
ligation products. The ladder was generally still present after digestion with the cognate restriction enzyme, but much less extensive linker polymerization was evident.

**PLASMID CONSTRUCTIONS**

**Construction of CYC1AH5′ Deletions**

The construction of the pYeCYC1AH5′ plasmids is diagrammed in Figure 1. Ten micrograms of plasmid pYeCYC1(2.5) was linearized by digestion with HindIII. The DNA was purified from the reaction mix and incubated at 37° with 5 μ of exonuclease III in 100 μl of exo buffer, containing 10 mM Tris-HCl pH 7.5, 10 mM MgCl2, 1 mM DTT. Samples of 10 μl were removed at intervals of 30 seconds after starting the reaction and added directly to a tube containing 100 μl of 2 x SI buffer (0.1 M NaOAc pH 4.6, 0.3 M NaCl, 1 mM ZnSO4). (Exonuclease III is inactivated by Zn2+ in the SI buffer.) SI nuclease (12 u) was added and the mixture was kept at room temperature for 30 minutes. The DNA was purified by phenol extraction and ethanol precipitation and treated for 30 minutes at room temperature with DNA pol I (Klenow fragment) in 20 μl of LK buffer containing 50 μM dNTPs. The blunt-ended DNA was ligated to 30 pmol of phosphorylated, 8 bp PstI linkers in the manner described in the previous section. The DNA was purified by phenol extraction and ethanol precipitation and digested with BamHI (16 u) and PstI (150 u) in 100 μl. (A large quantity of PstI was used in an effort to ensure complete digestion of the large excess of PstI linkers present in the mixture.) The DNA was again purified by phenol extraction and ethanol precipitation and electrophoresed through 0.7% LMP agarose. BamHI-PstI fragments of 2.2-2.4 kb were purified and ligated into pUC13 which had been digested.
CONSTRUCTION OF ΔH5' DELETIONS

Figure 1. Construction of ΔH5' Deletions
The CYCl coding sequence is indicated by stippling. Open boxes represent flanking yeast DNA. pBR322 is represented by a single line. Recognition sites for restriction endonucleases are indicated as follows: B, BamHI; H, HindIII; P, PstI.
with BamHI and PstI. Transformants of E. coli JM101 were identified by plating on LB-glucose-Ap-XGAL medium, and plasmid DNA was prepared from small cultures of Ap^R lacZ^− transformants. Digestion with EcoRI and PstI allowed the approximate deletion endpoint of the CYC1A^H5′ derivative in each plasmid to be estimated. The exact deletion endpoints of those derivatives chosen for further study were determined by sequencing of EcoRI-digested plasmid DNA from the M13RF1 primer (Table I). [Some of the CYC1A^H5′ derivatives were originally isolated in pEMBL8(+) and were transferred as BamHI-HindIII fragments to M13mpl0 for sequencing from M13FP1 because of problems experienced in isolating single-stranded pEMBL DNA for sequencing.]

**Construction of YEp213CYC1A^H5′ Plasmids**

Selected pUC13CYC1A^H5′, or pEMBL8CYC1A^H5′ plasmids were digested with BamHI and HindIII, and the 2.2-2.4 kb CYC1A^H5′ fragments released from each was purified after electrophoresis in LMP agarose. The HindIII site of the pUC/pEMBL polylinker is located immediately downstream of the PstI site which marks the CYC1A^H5′ endpoint. Digestion with HindIII rather than PstI allowed the CYC1A^H5′ fragment to be ligated into a yeast vector more easily. Each BamHI-HindIII CYC1A^H5′ fragment was ligated to YEp213 which had been digested with BamHI and HindIII and treated with bacterial alkaline phosphatase. The ligation mixtures were used to transform E. coli JM101 to ampicillin resistance, and plasmids prepared from 1.5 ml cultures of individual transformants were screened by digestion with BamHI/HindIII and with EcoRI/PstI.

Suitable plasmids, denoted YEp213CYC1A^H5′ plasmids, were then introduced into yeast strain GM-3C-2 using the transformation procedure
Construction of CYC1ΔK3′ and CYC1ΔK5′ Deletions

The starting material for constructing deletions from the KpnI site of the CYC1 gene was plasmid pYeCYCl(2.5) which had been digested to completion with KpnI, purified from the reaction mixture by phenol extraction and ethanol precipitation, and dissolved in TE10:1 at 1 μg/μl. The procedure used to generate deletion plasmids is illustrated in Figure 2. Ten micrograms of the linearized plasmid was treated with 1 μ of nuclease Bal31 at 30°C in a 100 μl mixture containing 20 mM Tris-HCl pH8, 12 mM CaCl₂, 12 mM MgCl₂, 200 mM NaCl, 1 mM EDTA, and 250 μg/ml BSA. Samples of 25 μl, removed at 1 minute intervals beginning 5 minutes after the addition of enzyme, were immediately added to separate tubes containing 2.5 μl of 0.2 M EGTA to inactivate the nuclease. The reason for keeping the samples separate at this stage was to check the extent of nuclease digestion at each time point. A 1 μl aliquot of each sample was diluted to 10 μl with Kpn restriction buffer and digested with Hind III (2 μ) at 37°C for 2 hours. The products of the digest were labelled by adding α[³²P]dATP (10 μCi), DNA polymerase I (Klenow fragment ; 1 μ) and incubating at room temperature for 15 minutes. The labelled DNA was purified from each digest by phenol extraction and ethanol precipitation and redissolved in 4 μl water and 1 μl glycerol dyes. Aliquots (2 μl) were electrophoresed through a 7% acrylamide thin gel beside a labelled HinfI digest of pBR322 at 10 V/cm.

Digestion of pYeCYCl(2.5) with KpnI and HindIII releases a 360 bp fragment. In the Bal31-treated samples, this fragment was replaced by a series of fragments ranging in length from about 350 bp to less than
Figure 2. Construction of CYClΔK5' and CYClΔK3' Deletions
CYCl coding sequences are indicated by stippling. Restriction sites are indicated as follows: B, BamHI; H, HindIII; K, KpnI.
75 bp. The distribution of fragment sizes was quite heterogenous at each time point, though the shift towards smaller fragment sizes with increasing digestion time was obvious. Since all samples contained deletion endpoints in the region of interest (50-250 bp from the HindIII site), the remaining portions of all 4 samples were pooled. The pooled DNA was purified by phenol extraction and ethanol precipitation and redissolved in 5μl of water. After overnight incubation, reaction mixture was diluted to 50 μl and EDTA was added to 12 mM. The mixture was deproteinized by phenol-extraction, and residual phenol was removed by ether extraction. After the residual ether had evaporated, the reaction products were loaded onto a 1.5 ml Sephadex G-100 column. The column was eluted with TE10:1 and 0.1 ml fractions were collected. The first peak of radioactivity eluted in fractions 4-8, which were pooled and precipitated with ethanol. The DNA was then digested with KpnI (40u). After digestion EDTA was added to the reaction mixture to 12 mM, and the DNA was deproteinized by extraction with phenol, followed by ether. The plasmid DNA was separated from released linkers by gel filtration on Sephadex G-100 as before. Fractions of 0.1 ml were again collected, and fractions 6 - 9, containing the first peak of radioactivity, were pooled. The DNA was ethanol-precipitated and digested sequentially with KpnI and BamHI in 50 μl. The DNA was purified by phenol extraction and ethanol precipitation, redissolved, loaded on a 0.7% LMP agarose gel and electrophoresed alongside a HindIII digest of λ DNA at 5 V/cm for 5 hours. Two bands were visible after staining with ethidium bromide: one containing fragments of about 4.0-4.2 kb carried sequences downstream of the KpnI site of CYCl, while a population of fragments of
about 1.9 - 2.1 kb carried the 5' end of the gene. The DNA in each band was purified as described and dissolved in 20 μl of TE10:1 (approximately 0.02 μg/μl). The 2.2 kb and 4.3 kb BamHI/KpnI fragments of pYeCYCl(2.5) were also purified from 0.7% IMP agarose for ligation to the resected fragments as described below.

One quarter of the recovered 4 kb fragment pool (about 0.2 μg) was ligated to 0.4 μg of the 2.2 kb BamHI/KpnI fragment of pYeCYCl(2.5) to produce the pYeCYClA3' plasmids. Compared to pYeCYCl(2.5), these plasmids carry deletions extending from the KpnI site toward the 3' end of CYCl. Similarly, one quarter of the recovered 2 kb fragment pool was ligated to 0.7 μg of the 4.3 kb BamHI/KpnI fragment of pYeCYCl(2.5) to produce the pYeCYClA5' plasmids which, compared to pYeCYCl(2.5), carry deletions extending from the KpnI site toward the 5' end of CYCl. Both ligation mixes were used to transform E. coli RRL to ampicillin resistance. Plasmid DNA was prepared from 1 ml cultures of individual transformants. The extent of each A3' deletion was estimated from the size of the small KpnI/HindIII fragment released by digestion of the miniprep DNA. Digestion of the pYeCYClA5' plasmids with KpnI and XhoI allowed the extent of the A5' deletions to be similarly estimated. Digests were initially analyzed on agarose gels, and those of interest were also electrophoresed in 7% acrylamide gels to allow more accurate estimation of the size of the small fragments involved.

Construction of YEpl3CYClA3' Plasmids

E. coli transformants carrying pYeCYClA3' plasmids with suitable deletion endpoints were replated and a single colony of each was picked separately into 3 ml of LB-glucose-Ap. Plasmid DNA isolated from these
cultures was digested with BamHI and HindIII, and the 2.2-2.4 kb fragments carrying the 3' portion of CYC1 were purified from 0.7% LMP agarose. These fragments were ligated into YEpl3 which had been digested with BamHI and HindIII and treated with bacterial alkaline phosphatase. The ligation mixtures were used to transform E. coli RRL to ampicillin resistance, and miniprep plasmid DNA isolated from the transformants was screened by digestion with BamHI and HindIII. Suitable transformants, carrying plasmids with Bam/HindIII fragments of 10.4 and approximately 2.4 kb, were replated, and 50 ml cultures were grown from single colonies for plasmid DNA isolation. The plasmids were screened by digestion with several restriction enzymes to confirm their structure. They were then introduced into yeast strain GM-3C-2, selecting for LEU2+ transformants.

Determination of $\Delta$K3' Deletion Endpoints

Purified YEpl3CYClAK3' plasmids were digested with BamHI and Hind III, and the 2.2-2.4 kb CYClAK3' fragment from each was ligated into Ml3mpl0 RF which had been cleaved with BamHI and HindIII. The ligation mixtures were used to transfect E. coli JM101. Colourless plaques were picked into 1 ml 2YT cultures of JM101 for isolation of phage DNA, as described. The DNA was sequenced using dideoxynucleotide chain terminators from the universal primer, Ml3FP1 (see Table I). The sequence across the KpnI site marking the deletion endpoint was read.

Construction of pYeCYClAK5'/AK3' Plasmids (Promoter/3' End Signal Fusions)

As mentioned earlier, several pYeCYCl K5' plasmids, carrying deletions extending towards the CYC1 promoter from the KpnI site within the gene, were screened by digestion with XhoI and KpnI. Four out of
twelve of these plasmids had XhoI/KpnI fragments of about 230-250 bp, indicating that they lacked all or most of the coding sequence upstream of the KpnI site. Transformants carrying these plasmids were replated and plasmid DNA was purified from 50 ml cultures grown from single colonies. Each plasmid was digested with BamHI and KpnI, and the 1.9-2 kb BamHI/KpnI fragment carrying the 5' flanking sequences of CYCl was purified from 0.75% LMP agarose. Each of these fragments was ligated to the purified 4.2 kb BamHI/KpnI fragment of pYeCYClΔK3'449 carrying the 3' end of the CYCl gene. E. coli RR1 was transformed to ampicillin resistance with these ligation mixes. Plasmid DNA prepared from 3 ml cultures of individual transformants was digested with BamHI and HindIII, and the 2 kb fragment carrying a ΔK5'/ΔK3' deleted derivative of CYCl was purified from IMP agarose. Each fragment was ligated into BamHI-HindIII-digested M13mpl0 RF. The ligation mixtures were used to transfect E. coli JM101, and 2 colourless plaques derived from each ligation mixture were picked into separate 2 ml 2YT cultures of JM101. Phage DNA was prepared and a region including the fused ΔK5' and ΔK3' deletion endpoints was sequenced from M13FP1. In one of the constructs, mpl0CYClΔK-26/+449, the ΔK3' deletion endpoint at position +449 of the CYCl sequence was joined through a KpnI linker, to position -26 of the CYCl promoter region. The BamHI/HindIII fragment that had been inserted into mpl0CYClΔK-26/+449 was similarly introduced into BamHI/HindIII-digested M13mpl1. Phage DNA was prepared from several colourless plaques, and the presence of the CYClΔK insert was confirmed by hybridization to the DNA of mpl0CYClΔK-26/+449 using the procedure of Winter and Fields [(1980); see "clone orientation"]. One of these phage was taken as mpl1CYClΔK-26/+449 and used as a template for
mutagenesis with oligonucleotides oAS3 and oAS4. Clones were isolated which carried the GT473 and C474 mutations. They were referred to as mpl1CYClT1 and mpl1CYClT2, respectively. Double-stranded RF DNA of mpl1CYClT2 was isolated from a 500 ml culture of infected JM101 for use in the experiments described below.

Construction of Deletions in mpl1CYClT2: The pA4 Plasmids

The construction of deletion derivatives of mpl1CYClT2 is outlined in Figure 3. The replicative form of mpl1CYClT2 was linearized by digesting with HindIII and purified by phenol extraction and ethanol precipitation. Ten micrograms of the linear DNA was treated with 2 u of exonuclease III in 100 /al of exo buffer. Samples of 15 /al were removed at intervals of 30 sec after starting the reaction and transferred to S1 buffer to inactivate the exonuclease. The pooled samples were treated with 6 u of S1 nuclease in 300 /al of S1 buffer for 30 minutes at room temperature. The DNA was purified by phenol extraction and ethanol precipitation and made blunt-ended by treatment with DNA polymerase I (Klenow) and dNTPs as described earlier. Oligonucleotide oAS5 and oAS6 (150 pmol each) were mixed, phosphorylated, and ligated to the resected, blunt-ended DNA overnight at room temperature under the conditions described. After deproteinization and ethanol precipitation, the DNA was digested with BamHI (16 u) and EcoRI (100 u). The DNA was purified from the digest and electrophoresed through 0.8% LMP agarose. A region containing fragments of 1.9 - 2.0 kb was excised from the gel and the DNA within this region was purified. The fragment pool was ligated to pEMBL9(+) which had been digested with BamHI and EcoRI and purified after electrophoresis through 0.7% LMP agarose. The ligation mix was used to
Figure 3. Construction of the pA4 Promoter:3' End Signal Fusions
Restriction sites are labelled as follows: E, EcoRI; K, Kpnl; H, HindIII. The sequence of the ATG/EcoRI adapter is noted in the text, and in Table I.
transform E. coli RRL to ampicillin-resistance. The use of RRL rather than JM101 precluded the possibility of screening for plasmids with inserts on XGAL plates, but at the time, my stock of JM101 gave very low transformation efficiencies. Plasmid DNA was prepared from 1.5 ml cultures of transformants using the alkaline SDS lysis procedure. Digestion with SalI and EcoRI revealed that about 50% of the plasmids, referred to as pA4s, carried the CYCl insert. Deletion endpoints were estimated from the electrophoretic mobility of the smaller fragment in XhoI/EcoRI digests of the plasmids. The exact endpoints of deletions of interest were determined by sequencing single-stranded pA4 DNA, which was isolated as described earlier. As well as defining the deletion endpoint, the sequence confirmed the presence of the C474 mutation in each plasmid which still retained position +474 of the CYCl sequence.

Plasmids with deletion endpoints of interest were digested with EcoRI and BamHI, and the digested DNA was electrophoresed through 0.7% IMP agarose. The 1.9-2 kb BamHI/EcoRI fragment of each plasmid was purified for use in constructing promoter /3' end signal/ lacZ fusion plasmids.

**Construction of placZ**

Plasmid pMC1403 (Casadaban et al., 1980, see Figure 4), carrying lacZ'YA, was used to construct a plasmid carrying only lacZ'. The lacA gene and most of lacY can be deleted by eliminating the largest AvaI fragment from pMC1403 (Casadaban et al., 1983). Accordingly, 5 μg of the plasmid was digested with AvaI (8 u) in 50 μl, and samples of 10 μl were removed at 15-minute intervals. Aliquots were examined after electrophoresis on agarose. Samples removed after 60 minutes and 75...
Figure 4. Restriction Maps of pMC1403 and placZ
Restriction sites are denoted as follows: AvaI, A; BamHI, B; EcoRI, E; PstI, P; SalI, S; SmaI, Sm. The inset is a photograph of ethidium bromide-stained AvaI digests of pMC1403 (lane 1) and placZ (lane 2), showing that placZ lacks the largest AvaI fragment of pMC1403.
minutes contained very little full length linear plasmid, indicating that most molecules had been cut at 2 or more AvaI sites. These samples were digested to completion with SalI, which cuts pMC1403 once, within the large AvaI fragment. The object of this step was to reduce the "background" of plasmids retaining the large AvaI fragment but lacking other fragments. DNA was purified from each digest by phenol extraction and ethanol precipitation, and 0.05 μg was recircularized with T4 DNA ligase in a volume of 20 μl. The ligation mixes were used to transform E. coli RRL to ampicillin-resistance, and plasmids isolated from 1.5 ml cultures of individual transformants were screened by AvaI digestion. Eight of 36 transformants screened retained all but the largest of the AvaI fragments of pMC1403 and were shown to contain no extra fragments by digestion with SacI. One was chosen as placZ and was isolated from a 500 ml culture grown from a single colony. Digestion with seven restriction enzymes, alone and in all pairwise combinations, confirmed that placZ differed from pMC1403 only insofar as it lacked the largest AvaI fragment.

Construction of YRp72, YRp73

Plasmid pAAR6 (Ammerer, 1983) is a derivative of YRp7 (Struhl et al., 1979) lacking both of the EcoRI sites of that plasmid and carrying the promoter and 3' flanking sequences of the ADC1 gene on a BamHI fragment. A sample of pAAR6 was digested to completion with BamHI and purified by phenol extraction and ethanol precipitation, and 0.1 μg of the digested DNA was recircularized in a volume of 20 μl. E. coli RRL was transformed to ampicillin resistance with the ligation mix and plasmids isolated from several transformants were screened by restriction with HindIII, SalI, and EcoRI alone or in combination. A
plasmid lacking the BamHI fragment of pAAR6 was identified and referred to as YRp72 (see Figure 5 for map). The plasmid was purified from cells of a 50 ml culture grown from a single colony.

Plasmid YRp72 (Figure 5) has 2 PstI sites, one in sequences derived from pBR322, and one in the yeast ars1 sequence. To facilitate the isolation of BamHI/PstI or SalI/PstI fragments carrying both TRPl and ars1, the second of these PstI sites was eliminated as follows: five micrograms of the plasmid was digested with 5 u of PstI at 37°, and samples were removed at 10 minute intervals. The 20- and 30-minute samples were pooled because gel electrophoresis of an aliquot of each showed them to contain a substantial proportion of singly-cut molecules. The DNA was purified by phenol extraction and ethanol precipitation and treated with DNA polymerase I (Klenow) in a 20 µl mixture containing 50 µM dNTPs in LK buffer. An aliquot of this mixture was then treated with T4 DNA ligase, and the ligation mix was used to transform E. coli RRI to ampicillin-resistance. Plasmids prepared from the transformants were screened by digestion with BamHI and PstI. A plasmid lacking the PstI site in ars1 was chosen and subsequently referred to as YRp73 (For map, see Figure 5).

Construction of YEp73

About 3 µg of YRp73 was digested to completion with BamHI. The linearized DNA was purified by phenol extraction and ethanol precipitation, and 1 µg was treated with bacterial alkaline phosphatase. The DNA was purified once more, and 0.1 µg was ligated to a purified 1.5 kb Sau3AI fragment of YEpl3 which carries 2u circle sequences necessary for replication at high copy number in yeast (Broach, 1983). E. coli RRI was transformed to ampicillin-resistance
Figure 5. Restriction Maps of YRp73 and YEYp73
The single line denotes pBR322 sequences. Fragments of yeast DNA are labelled appropriately. Restriction sites are denoted as follows: BamHI, B; EcoRI, E; PstI, P. Restriction sites destroyed during construction of the plasmids are enclosed in parentheses. Plasmid YRp72 differs from YRp73 only in retaining the PstI site in arsI.

-140-
with the ligation mix and plasmid DNA isolated from 1.5 ml cultures of individual transformants was digested with BamHI and PstI. A plasmid was identified which carried the 2u circle insert and retained a BamHI site on the side of the insert distal to the TRPl-arsl fragment of the vector. [The Sau3A fragment involved extends from position 344 to 1915 of the B form of 2u circle, and the sequence at one end of the fragment (Hartley & Donelson, 1980) allows the formation of a BamHI site when that end is ligated to a BamHI fragment terminus.] This plasmid was referred to as YEp73. A single colony derived from a YEp73 transformant of RRL was used to inoculate a 50 ml culture, from which the plasmid was then purified for use in constructing the pA7s and pAlls (see below).

Construction of pA5s

The first set of promoter:3′end signal: lacZ fusion plasmids to be constructed was the pA5 series. Plasmids in the series had the structure shown in Figure 6. Each pA5 was constructed by ligating, in a single reaction, the 2 kb BamHI/EcoRI fragment of a pA4 to the 2.4 kb BamHI/PstI fragment of YRp73 and the 5.6 kb EcoRI/PstI fragment of placZ. All fragments were purified from 0.7% LMP agarose for use in the ligation reaction. Plasmids were isolated from small cultures of the ampicillin-resistant transformants of E. coli RRL produced by each ligation mix, and pA5 plasmids were identified by digestion with EcoRI and PstI. Each pA5 plasmid was then purified from a 50 ml culture grown from a single colony of a transformant. The purified DNA was used to transform yeast strain RPL23 to tryptophan prototrophy. This strain was convenient to use because it carried both trp1 and ura3 markers, allowing selection of transformants carrying pA5 plasmids.
Figure 6. Structure and Derivation of pA5 Plasmids
Restriction sites are as follows: BamHI, B; EcoRI, E; HindIII, H; KpnI, K; PstI, P; SmaI, Sm. Labels enclosed in parentheses denote restriction sites which were destroyed during plasmid construction. The CYCl 3' end signal fragment is indicated by the black box.
or pLG669Z (URA3+), the latter serving as a positive control for \(\beta\)-galactosidase assays.

**Construction of pA6 Plasmids**

As shown in Figure 7, the pA6 plasmids differ from the pA5s in that they carry in addition to the arsI origin of replication, sequences from the centromere of chromosome 3, CEN3. A 2.2 kb BamHI/BgII CEN3 fragment was purified from the plasmid pYeCEN3(41). One of the pA4 plasmids (Its specific identity doesn’t matter for this construction.) was digested with BamHI, treated with bacterial alkaline phosphatase and, after deproteinization and ethanol precipitation, ligated to the CEN3 fragment. *E. coli* RRL was transformed to ampicillin-resistance with the ligation products, and plasmids isolated from the transformants were screened for the presence and orientation of the CEN3 insert by digestion with BamHI and EcoRI. When a BamHI/BgII fragment is inserted into a BamHI site a BamHI site is restored at one end of the insert. A plasmid in which the restored BamHI site was at the end of the insert distal to the CYC1 region of the plasmid was chosen as pA6a (see Figure 7).

To construct a given pA6 plasmid, the 6.5 kb SalI/PstI fragment of the corresponding pA5 was ligated in a single reaction to the 2.4 kb BamHI/PstI fragment of YRp73 and the 3.2 kb BamHI/SalI fragment of pA6a. All fragments were first purified from IMP Agarose. The ligation mixtures were used to transform *E. coli* RRL to ampicillin-resistance, and plasmids isolated from the transformants were screened by digestion with BamHI and PstI.

Plasmids having the structure expected of the pA6 series were introduced into yeast strain RP123, selecting for TRP1+ transformants.
Figure 7.  Restriction Maps of pA6a, pA6, pA7, and pA10 Plasmids
The single thin line denotes pBR322 sequences. Other sequences, indicated by a thick line or by boxes, are labelled appropriately. Restriction sites are labelled as follows: B, BamHI; E, EcoRI; K, KpnI; P, PstI; S, SalI. Sites destroyed during plasmid construction are indicated by parentheses flanking the appropriate label.

-144-
Construction of pA7 Plasmids

The pA7 plasmids differ from the corresponding pA5 plasmids in that they carry a 1.6 kb Sau3A fragment of the yeast 2μ circle which contains an origin of replication and cis-acting sequences needed for maintenance at high copy number in yeast (Broach, 1983; see Figure 7). Each pA7 was constructed by ligating, in a single reaction, the 5.2 kb EcoRI/PstI fragment of placZ, the 4.0 kb BamHI/PstI fragment of YEp73, and the 2 kb BamHI/EcoRI fragment of pA4. Plasmids were isolated from a few ampicillin-resistant transformants of E. coli RR1, and pA7s were identified from the pattern produced by digestion with BamHI and PstI. Miniprep pA7 plasmid DNA was used to transform yeast strain RP123 to tryptophan prototrophy.

Construction of pA10 Plasmids

The pA10 plasmids (Figure 7) differ from the corresponding pA5s in that they carry a lacZYA fragment of pMC1403 rather than the lacZ fragment of placZ. Each was constructed by ligating the 9 kb EcoRI/PstI fragment of pMC1403 to the 4 kb EcoRI/PstI fragment of the corresponding pA5. The ligation mixes were used to transform E. coli RR1 to ampicillin-resistance, and plasmids isolated from the transformants were identified as pA10s from the patterns produced by digestion with SalI/EcoRI and EcoRI/PstI.

Construction of pAll Plasmids

An example of a pAll plasmid is shown in Figure 8. It carries the CYCl promoter:3' end signal:linker region of a pA5, joined to a lacI'Z fragment derived from pLG669Z, as well as the yeast TRPl-arsl fragment, 2μ circle sequences necessary for high copy-number maintenance in yeast, and pBR322.
Figure 8. Structure of the pAll Plasmids
Restriction sites are labelled as follows: B, BamHI; E, EcoRI; K, KpnI; P, PstI; S, SalI; Sm, SmaI. The various functional elements of the plasmid are labelled as such.
Plasmid pLG669Z (2 μg) was digested to completion with BamHI, and the resulting single-stranded ends were filled in by incubating with DNA polymerase I (Klenow) and 100 μM dNTPs. The plasmid was purified by phenol extraction and ethanol precipitation and digested with PstI. The 5 kb Blunt/PstI fragment carrying laci'Z was purified from the digest after electrophoresis through LMP agarose.

Each pA5 plasmid was similarly digested with BamHI and treated with DNA polymerase I (Klenow) and dNTPs, purified, and then digested with SalI. The 1.1 kb SalI/Blunt fragment carrying the CYCl promoter:3' end signal region was purified from IMP agarose. These 2 fragments were ligated in a single reaction with the 4.2 kb SalI/PstI fragment of YEp73, E. coli RRl was transformed to ampicillin-resistance with each ligation mixture, and plasmids isolated from small cultures of the transformants were digested with EcoRI to allow identification of the pAlls. Several independent representatives of each pAll were used to transform yeast strain RP123 to tryptophan prototrophy.

Construction of mp10Al

Plasmid pAll-71 was digested with XmaI and the 430 bp fragment carrying the CYCl promoter:3' end signal region was purified after electrophoresis through LMP agarose. This fragment (about 0.2 μg) was ligated to mpl0 RF DNA (about 0.05 μg) which had been linearized by digestion with XmaI. The ligation mix was used to transfect E. coli JM101. Single-stranded phage DNA was prepared from several small JM101 cultures, each infected with one of the colourless plaques produced by transfection. The phage DNA was screened for the presence of the 430 bp XmaI fragment by copying it from M13FP1 with DNA polymerase I (Klenow) and 50 μM dNTPs, then digesting with SalI and EcoRI. The 450
Recognition sites for restriction endonucleases are denoted as follows: B, BamHI; E, EcoRI; H, HindIII; K, KpnI; S, SalI; Xm, Xmal; X, XhoI. The priming site for oligonucleotide FPl (Table I) is indicated by a small arrow on the map of mpl0A1. The 3' end signal fragment is indicated by cross-hatching.
bp fragment so released from clones carrying the Xmal insert was easily seen after agarose gel electrophoresis of the digest, followed by ethidium bromide staining. One clone carrying the insert was chosen as a reference for checking the orientation of the insert in the other clones, using the procedure of Winter and Fields (1980).

Two clones carrying the insert in opposite orientations were replated on JM101, and a single isolated plaque of each was used to prepare a phage inoculum for infection of 50 ml culture of JM101. Phage DNA prepared from these cultures was sequenced from M13FP1. The clone in which the mRNA-parallel strand of the CYCl region was linked to the + strand of M13 mp10 was used in misincorporation mutagenesis experiments. It is referred to as mp10Al, and it is diagrammed in Figure 9.

Construction of pAlls Carrying Point Mutations in 3' End Signal

Single-stranded DNA (about 1 μg) of an mp10Al derivative carrying a mutation in the CYCl 3' end signal was annealed with M13FP1 and treated with DNA polymerase I (Klenow fragment) in 20 μl of Kpn buffer containing 0.1 mg/ml BSA and 0.2 mM dNTPs. After 2 hours at room temperature, the mixture was heated to 70° for 10 minutes to inactivate the polymerase. Potassium chloride was added to a concentration of 20 mM, and the DNA was digested with SmaI. The digest was electrophoresed through 0.8% LMP agarose and the 430 bp CYCl fragment was purified. The fragment was ligated to pAll.11 which had been digested with SmaI and treated with bacterial alkaline phosphatase. *E. coli* RRL was transformed to ampicillin resistance with the ligation mixture, and plasmid DNA was prepared from several of the transformants. Digestion with XhoI and EcoRI revealed the presence and

-149-
orientation of the 430 bp SmaI fragment and allowed it to be clearly
distinguished from the shorter fragment derived from pAll.11 (See
Figure 10). Treatment with alkaline phosphatase should have prevented
religation of the 2 SmaI fragments of pAll.11, and XhoI/EcoRI digestion
of the plasmids recovered from the transformants indicated that all
contained a small fragment of the same size as that in pAll.71, which
was the source of the SmaI fragment in mplOAl.

Each of the pAlls carrying point mutations in the 3' end signal was
introduced into yeast strain RP123, selecting for TRP1+ transformants.
The remaining miniprep DNA was purified for DNA sequencing as described
earlier, and a region including about 100 bp of the CYCl promoter
region, the 3' end signal fragment, and about 100 bp of the adjacent
lacI'Z gene was sequenced from oAS7 (see Table I for oligonucleotide
sequence).

Construction of pAl2, pAl2A, and pAl2B Plasmids

The single-stranded DNA of each derivative of mplOAl (1.0 µg) was
annealed with M13FP1 (12 pmol) in 20 µl of Kpn buffer containing 0.1
mg/ml BSA and 0.4 mM dNTPs. The mixture was incubated with 1 µl of DNA
polymerase I (Klenow fragment) for 1 hour at 23°, after which the
enzyme was inactivated by heating to 65° for 10 minutes. The DNA was
digested with KpnI and HindIII and purified by phenol extraction and
ethanol precipitation. One-fifth of the precipitated DNA was ligated
to 0.1 µg of pYeCYCl(2.5) which had been digested with KpnI and HindIII
and treated with BAP. After overnight incubation at 23°, the
ligation mixture was used to transform E. coli RR1 to
ampicillin-resistance. Plasmids were isolated from 1.5 ml overnight
cultures of individual transformants and screened by restriction
Figure 10. Reconstruction of pAll Plasmids Carrying Point Mutations in the CYCl 3' End Signal

The construction of pAll.Tl.36 is shown as an example. Restriction sites are labelled as follows: K, KpnI; Sm, SmaI. The inset is a photograph of an ethidium bromide-stained electropherogram showing the fragments produced by XhoI/EcoRI digestion of various plasmids: lane 1, pAll.Tl.36; lane 2, pAll.G3.7; lane 3, pAll.11; lane 4, pAll.71. The smallest XhoI/EcoRI fragment visible carries the 3' end signal region. It is apparent that the fragments produced by digestion of pAll.11 and pAll.71 are readily distinguished, and that the plasmids containing 3' end signal mutations yield fragments of the same size as pAll.71, not pAll.11.
mpIOTl.36
ssDNA

Ml3FPI, dNTPs
Klenow

partial
duplex

Sm Sm

450 bp

Sm K Sm

10.6 kb

Sm

400 bp

Sm K Sm

pAll.11
Smal fragments

1 2 3 4
Figure 11  

A. Construction of pAl2 Plasmids

Restriction sites are labelled as follows: B, BamHI; H, HindIII; K, KpnI; Sm, SmaI. The CYCl coding sequence is stippled, and the 3' end signal fragment is cross-hatched.

B. Orientation of the Truncated cycl Gene in the pAl2A and pAl2B Plasmids.

Restriction sites are labelled as in A, with the following additions: E, EcoRI; P, PstI.
digestion. Plasmids in which the 355 bp KpnI/HindIII fragment of pYeCYCl(2.5) had been replaced by the small (70 bp) KpnI/HindIII fragment of the mpl0A1 derivatives were chosen as pAl2 plasmids (Figure 11).

Each pAl2 plasmid was digested with BamHI, and the 2 kb fragment carrying the truncated CYCl gene and 3’ end signal fragment was purified from IMP agarose for ligation into BamHI, BAP-digested YEpl3. Plasmids were isolated from ampicillin-resistant transformants of E. coli RRL and screened by digestion with EcoRI and XbaI/HindIII. Those in which the inserted CYCl gene was oriented towards 2μ circle sequences in the vector were referred to as the pAl2A series. Those in which the insert had the opposite orientation were called the pAl2B series (Figure 11). Both series were introduced into yeast strain GM-3C-2, selecting for LEU2+ transformants. RNA isolated from these transformants was analyzed for hybridization to a CYCl probe after electrophoresis and transfer to nitrocellulose.

OLIGONUCLEOTIDE-DIRECTED MUTAGENESIS

Oligonucleotide-directed mutagenesis was carried out using the procedure described by Zoller and Smith (1982;1983). Each mutagenic oligonucleotide was 14-18 nucleotides long and was equivalent in sequence to a region of interest on one strand or the other of the 2.5 kb BamHI/HindIII CYCl fragment of pYeCYCl(2.5), except at certain positions. At those positions, the oligonucleotide sequence matched the desired mutant sequence. The template used for mutagenesis was the single-stranded DNA of an M13 clone carrying the CYCl strand complementary to the oligonucleotide. A particular mutagenic oligonucleotide was chosen on the basis of a computer search of the
sequences of the 2.5 kb BamHI/HindIII C\text{YC}l fragment and M13 phage DNA using the programme ADDL:SEQ\text{NC}E (Delaney, 1982), which indicated that the oligonucleotide was unlikely to form a stable duplex with the template molecule at any site other than the intended target in C\text{YC}l. The criterion used to determine which oligonucleotides would be useful was that an oligonucleotide:template duplex formed at any site other than the target should contain at least two more mismatched base pairs than the duplex formed at the target site. The specific oligonucleotides and templates used to produce particular mutations are listed in Table IV.

The mutagenic oligonucleotide was phosphorylated with T4 polynucleotide kinase and ATP in LK buffer and used for mutagenesis without further pretreatment. About 0.3 pmol of single-stranded template DNA was mixed with a 10-100 fold molar excess of the phosphorylated oligonucleotide in 10 \mu l of buffer containing 20 mM Tris-\text{HCl} pH 7.5, 10 mM MgCl$_2$, 50 mM NaCl, 1 mM DTT. The mixture was heated to 55° for 5 minutes and cooled to room temperature. An equal volume of a mixture containing 1 mM dCTP, dGTP, and dTTP, 0.1 mM $\text{\alpha}^{[32P]}$dATP (100 Ci/mmol), and 1 mM ATP in 20 mM Tris-Cl pH 7.5, 10 mM MgCl$_2$, 10 mM DTT was added with 1 u of T4 DNA ligase and 2 u of DNA polymerase I (Klenow enzyme). After 2-5 minutes at room temperature, dATP was added to 1 mM, and incubation was continued at 15° for 12-16 hours. The reaction mixture was diluted to 50 \mu l with TE10:1 and an equal volume of 13% PEG in 1.6 M NaCl was added. The mixture was chilled on ice for 15-40 minutes, and the DNA was pelleted by centrifugation for 5 minutes at 4° in an Eppendorf centrifuge. The pellet was rinsed by adding 100 \mu l of ice-cold 6.5% PEG in 0.8 M NaCl.
<table>
<thead>
<tr>
<th>Mutation</th>
<th>Oligonucleotide</th>
<th>Template</th>
<th>Efficiency</th>
<th>Name of Mutant Clone</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG462</td>
<td>oAS1</td>
<td>mp9CYCl(2.5)</td>
<td>6</td>
<td>mp9CYCl(GG462)</td>
</tr>
<tr>
<td>C482</td>
<td>oAS2</td>
<td>mp8CYCl(2.5)</td>
<td>2</td>
<td>mp8CYCl(C482)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mp8CYCl(GG462)</td>
<td>8</td>
<td>mp8CYCl(GG462/C482)</td>
</tr>
<tr>
<td>GT473</td>
<td>oAS3</td>
<td>mp9CYCl(2.5)</td>
<td>42</td>
<td>mp9CYCl(GT473)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mp11CYClΔK(-26/+49)</td>
<td>19</td>
<td>mp11CYClT1</td>
</tr>
<tr>
<td>C474</td>
<td>oAS4</td>
<td>mp9CYCl(2.5)</td>
<td>58</td>
<td>mp9CYCl(C474)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mp11CYClΔK(-26/+49)</td>
<td>3</td>
<td>mp11CYClT2</td>
</tr>
</tbody>
</table>
and centrifuging for 5 minutes. As much of the supernatant as possible was removed using a capillary micropipet. The pellet was dissolved in TE10:1 (180 μl), and 20 μl of 2M NaOH was added. The denatured DNA was loaded onto a 5 ml 5-20% sucrose gradient in 0.2 M NaOH, 1 M NaCl, and centrifuged at 37,000 rpm for 2 hours in a Beckman SW50.1 rotor.

Fractions of about 0.2 ml were collected from the bottom of the gradient and the Cerenkov radiation emitted by each was measured in a liquid scintillation spectrometer. An example of the results is shown in Figure 12A. Fractions containing the lower (faster-sedimenting) of the two peaks of radioactivity were pooled, neutralized by adding an appropriate volume of 1 M Tris-citrate, pH 5 and dialyzed at 4°C against 2 1.5 l batches of TE10:1 and 1.5 l of TE2:0.01 for at least 3 hours each. The dialysate was used to transfected competent cells of E. coli JM101. Plaques obtained in the transfection were used to infect separate 1 ml cultures of JM101, and phage were precipitated from the supernatants of the infected cultures as described earlier. The phage were resuspended in 50 μl of TE10:1 and screened for the presence of oligonucleotide-directed mutations.

Screening for Mutations

Clones carrying oligonucleotide-induced mutations were identified by hybridization to the mutagenic oligonucleotide, as described by Wallace et al. (1979) and Zoller and Smith (1983). A grid was drawn lightly in pencil on the surface of a sheet of nitrocellulose and the sheet was then wet with water, soaked in 6XSSC for a few minutes, and air-dried. A 1-2 μl aliquot of each 50 μl phage suspension was spotted onto the nitrocellulose. An aliquot of the template used for mutagenesis was also spotted onto the filter as a "wild-type" control.
Figure 12. Oligonucleotide-Directed Mutagenesis

A. Profile of alkaline sucrose gradient, showing separation of covalently closed circular DNA (cc) from open circular DNA (oc). The gradient was fractionated from bottom to top, and each fraction was assayed for radioactivity in a liquid scintillation counter.

B. Identification of clones carrying C474 mutation. (Top) Primary screen: 36 phage isolates were bound to nitrocellulose, subjected to denaturation, hybridized to 5'[^32]P-oAS4, and washed at the indicated temperatures. The rows marked "WT" contain DNA from the parental phage, mp9CYCl(2.5). (Bottom) secondary screen: one of the positive clones from the first screen was replated and phage prepared from 12 isolated plaques were screened as before. The rows marked "WT,m" contain DNA from mp9CYCl(2.5) on the left, and DNA from the presumptive mutant phage identified in the first screen on the right.
The filter was then baked in vacuo at 80° for 1-2 hours and prehybridized in 6XSSC and 10X Denhardt's solution (0.2% Ficoll, 0.2% polyvinylpyrrolidone, 0.2% BSA) at 25° for 45 minutes.

The mutagenic oligonucleotide was 5' end-labelled for use as a hybridization probe by incubating 50 pmol of oligonucleotide with 1-2 u of T4 polynucleotide kinase and 10-20 uCi [32P]ATP in 10 µl of LK buffer for 30-60 minutes at 37°. The labelled oligonucleotide was separated from unincorporated ATP by gel filtration over Sephadex G-25 in 0.1 M ammonium bicarbonate, pH 7.8, or by binding to a 1 ml column of DE-52 in 0.2 M NaCl in TE10:1. The DE-52 column was washed in the binding buffer to elute ATP, and then the oligonucleotide was eluted with 1 M NaCl in TE10:1. Approximately 10^7 cpm (Cerenkov) were usually incorporated into 50 pmol of oligonucleotide. The hybridization mixture contained 6XSSC, 10X Denhardt's solution, and 1-5X10^6 cpm of labelled oligonucleotide, in a volume of 3 ml for use with a 50 cm² filter. The filter was incubated in a sealed bag with the hybridization mixture for 1-4 hours at 15° or room temperature. The filter was then subjected to four 10-minute washes in 6XSSC at room temperature and autoradiographed for 1-4 hours. A series of 10-minute washes were performed, each at a temperature 5-7° higher than the previous one. After each wash, the filter was autoradiographed for 1-16 hours, longer exposures being necessary after the later washes in the series. Phage which gave a strong hybridization signal after washing at a temperature sufficient to remove the labelled oligonucleotide from the "wild-type" DNA on the same filter were tentatively considered to carry the oligonucleotide-induced mutation.

-160-
In the experiments described in this study, washing at a temperature of 25-37\degree was sufficient to remove the labelled oligonucleotide from "wild-type" phage DNA, while mutant phage gave a strong hybridization signal after a 37\degree wash (See Figure 12).

The closed circular DNA purified from the alkaline sucrose gradient and used to transfect JM101 should have consisted of heteroduplexes with one mutant and one "wild-type" strand. Each plaque produced upon transfection might therefore have contained a mixture of mutant and "wild-type" phage. Phage suspensions identified as containing mutants in the first round of hybridization screening were replated on JM101 in order to obtain pure stocks of mutant phage. Well-isolated plaques were used to infect 1 ml cultures of JM101, and phage prepared from the supernatants of these cultures were subjected to a second round of hybridization screening. One of the mutant phage identified in the second screen was replated on JM101 and one of the plaques so obtained was used to infect a large culture of JM101 for the isolation of RF DNA using the procedures described earlier.

Transfer of Oligonucleotide-Mutagenized CYC1 Genes to Yeast Plasmids

The RF DNAs of the phage mp9CYClGG462, mp8CYClC482, mp8CYClGG462C482, mp9CYClGT473, and mp9CYClC474 (Table IV) were all digested with BamHI and HindIII, and the 2.5 kb BamHI/HindIII fragment of each was purified after electrophoresis through LMP agarose. The purified fragments were separately ligated to BamHI/HindIII-digested, alkaline phophatase-treated YEpl3, and each ligation mix was used to transform E. coli RRL to ampicillin-resistance. Plasmids isolated from the transformants were screened for the presence of the 2.5 kb BamHI/HindIII fragment carrying the CYC1 gene. Plasmids carrying each
mutant CYCl region were referred to as YEpl3CYClGG462, YEpl3CYClC482, YEpl3CYClGG462C482, YEpl3CYClGT473, and YEpl3CYClC474, respectively. They were introduced into yeast strain GM-3C-2, selecting for LEU2+ transformants.

Sequence Confirmation of Oligonucleotide-Directed Mutations in CYCl

The plasmid YEpl3CYClGG462 was digested with XhoI and HindIII and radioactively labelled by incubating with $\alpha[32P]dATP$ and DNA polymerase I (Klenow fragment). The 850 bp XhoI/HindIII fragment carrying CYCl, specifically labelled at its HindIII site, was purified by electrophoresing the digest through a 5% acrylamide gel.

Plasmids YEpl3CYClC482 and YEpl3CYClGG462C482 were digested with KpnI and HindIII, labelled with $\alpha[32P]dATP$ and DNA polymerase I (Klenow), and the 355 bp KpnI/HindIII fragment of each was purified by electrophoresing through a 5% acrylamide gel. The three purified fragments were subjected to the chemical cleavage reactions of Maxam and Gilbert (1977;1980) and electrophoresed through 12% acrylamide/7M urea gels. After autoradiography, the sequence of a 100 bp region including the intended mutant site was read. Each fragment contained the desired mutation and no other differences from the "wild-type" sequence were noted in the region sequenced.

The phage DNA of mp8CYClC482 was sequenced from M13FP1 by the chain termination method (Sanger et al.,1977;1980). The C482 mutation was the only difference from the "wild-type" sequence in a region of about 200 bp. The GG462 mutation could not be sequenced from FP1 in mp9CYClGG462, which was the phage produced by mutagenesis with oligonucleotide oASl (Table IV). The BamHI-HindIII fragment of the RF
DNA of that phage was however cloned into mp8 to produce mp8CYClGG462. Sequencing of the DNA of this phage from M13FP1 confirmed the presence of the GG462 mutation and the absence of other changes in its vicinity. This phage was used as the template for mutagenesis with oAS2 to produce mp8CYClGG462C482, as noted in Table IV.

The purified BamHI/HindIII fragments of the RFs of phage mp9CYClGT473 and mp9CYClC474 were ligated into BamHI-HindIII-digested mp10, and E. coli JM101 was transfected with the ligation mixes. Phage DNA was prepared from a series of 1 ml cultures, each infected with a single plaque obtained upon transfection. The DNAs of two independent mp10 clones of each Bam/HindIII fragment were sequenced from M13FP1, and in each case the desired mutation was the only change from "wild-type" sequence evident in a region of about 250 nucleotides.

The product of mutagenesis of npllcYClAK5'-26/AK3'+449 with oligonucleotide oAS4 was called mpl1CYClT2 (Table IV). The RF DNA of the phage was used to produce deletions extending towards the CYCl 3' end signal from the HindIII site downstream, as described above. The plasmids so produced were the pA4s. The sequence of the deletion endpoint region of each pA4 confirmed the presence of the C474 mutation.

SEGMENT-DIRECTED MUTAGENESIS

An outline of the procedure is shown in Figure 13.

Annealing

M13mp10 RF DNA was digested with SmaI and purified by phenol extraction and ethanol precipitation. The gapped heteroduplex which was to serve as template for misincorporation mutagenesis was produced by mixing the single-stranded DNA of mp10A1 with a 5-fold weight excess
Figure 13. Outline of Segment-Directed Mutagenesis
The inserted 3' end signal fragment in mpl0Al is indicated by the thick line. Mismatched base pairs are indicated by carets.

-164-
of SmaI-cut mpl0 RF in Hin buffer, heating to 100°C for 3 minutes and holding at 65°C for 15 minutes before cooling to room temperature. A typical reaction included 0.5 μg of single-stranded DNA and 2.5 μg of RF in a volume of 10 μl. The production of gapped heteroduplexes was easily monitored by agarose gel electrophoresis of samples of single-stranded DNA, uncut and linear mpl0 RF, the annealed mixture, and the same mixture after ClaiI digestion.

The template nucleotide immediately adjacent to a primer terminus defines the target site for mutagenesis by misincorporation. In order to use misincorporation to generate mutations throughout a region of interest, primer termini must first be produced which correspond to each position in that region. The limited, random extension of the 3' end of the incomplete strand of a gapped heteroduplex serves to produce such a set of primer termini.

Limited Primer Extension on a Gapped Duplex

Deoxynucleoside triphosphates were added to the gapped duplex formed between ss mpl0 Al and SmaI-cut mpl0 in 5 μl of Hin buffer. (Final concentrations: dCTP, dGTP, dTTP, 33μM each; α[32P]dATP, 800 Ci/mol, 1.8 μM.) The mixture was cooled to 9°C, and DNA polymerase I (Klenow fragment: 1u) was then added. Samples were removed periodically during the subsequent incubation at 9°C. In early experiments designed to check the rate of primer extension, samples of 3 μl were removed at 30-second intervals and immediately added to separate tubes containing 20 μl of ice-cold 10 mM EDTA. The DNA was deproteinized, ethanol-precipitated and then digested with BamHI in 5 μl of Hin buffer. Formamide-dye mix was added to each digest, which was then heated to 90°C for 3 minutes, quick-chilled and loaded on a
thin 8% acrylamide/7 M urea gel for electrophoresis.

To prepare primer-templates for use in misincorporation, samples of 7.5 μl were removed from the primer extension reaction after 1 minute and 3 minutes and transferred directly to a single tube containing 50 μl of 10 mM EDTA. The DNA was deproteinized by phenol extraction and ethanol-precipitated three times from 2M ammonium acetate to free it of deoxynucleoside triphosphates. Ammonium acetate was added to the DNA solution from a 7.5 M stock, followed by 2 volumes of ethanol. The mixture was chilled in dry ice/ethanol for 15 minutes, then warmed to room temperature and centrifuged for 10 minutes in an Eppendorf centrifuge. The final precipitate was rinsed with 70% ethanol, dried, and dissolved in water. All of the radioactivity in samples of this solution was TCA-precipitable, suggesting that the repeated precipitations had effectively removed unincorporated nucleotides. Unfortunately, about 10 - 20% of the primer-template was also lost during the process, but the elimination of unincorporated nucleotides was seen as more important than quantitative recovery of the DNA, because free nucleotides would have interfered with misincorporation of nucleotide analogs in the next step in the procedure.

**Misincorporation**

The primer-template population was incubated for 16-24 hours at room temperature with 1 u of DNA polymerase I (Klenow fragment) in a volume of 50-75 μl, containing 130 mM NaHepes pH 7.5, 0.2 mM MnCl₂, 2 mM DTT, 0.1 mg/ml BSA, and either dGTP[S] or dTTP[S] at a concentration of 0.2 mM (Shortle et al., 1982). The DNA was deproteinized by phenol extraction and ethanol precipitated.
Gap Repair

To complete the repair of the gap remaining in the primer:template after misincorporation, the DNA was incubated at room temperature for 16-24 hours with 1 u of DNA polymerase I (Klenow fragment) and 0.5-1 u of T4 DNA ligase in a volume of 20-40 μl containing 50 mM Tris-HCl pH 7.5, 1 mM MgCl₂, 10 mM DTT, 2 mM MnCl₂, 0.1 mg/ml BSA, 1 mM ATP, and dATP, dGTP, dCTP and dTTP at 0.1 mM each. The reaction mixture could be used directly for the transfection of JM101. Alternatively the DNA was deproteinized by phenol extraction, ethanol-precipitated and redissolved in 20 μl of 10 mM Tris-HCl pH 8, and an aliquot of this solution was used to transfec JM101. It was observed at this time that the addition of 1 or 2 μg of tRNA to a DNA sample prior to transfection of JM101 caused a 2-10-fold increase in the efficiency of transfection.

In one experiment, an attempt was made to eliminate leftover mpl0Al DNA in the hope of reducing the background of "wild-type" plaques. DNA recovered from the gap repair reaction was treated with 0.002 u of SI nuclease in 20 μl of SI buffer for 30 minutes at room temperature. Control experiments with mpl0Al single-stranded DNA and circular mpl0 RF had shown that 0.001u of SI nuclease in 20 μl effectively degraded 0.5 ug of mpl0Al DNA in 30 minutes without detectably affecting either the electrophoretic mobility of RF DNA or the efficiency with which it transfected JM101. The SI-treated DNA was purified by phenol extraction and ethanol precipitation before being introduced into JM101.

The cells were plated in soft YT agar with XGAL and IPTG following transfection, and the colourless plaques which appeared on the plates were picked into 1 ml cultures of JM101 in 2YT. Phage DNA was prepared
as usual and subjected to "single-track" sequencing from M13FP1 with the dideoxynucleotide corresponding to the α-thiodeoxynucleotide used during the misincorporation reaction. Single-track sequences were displayed by autoradiography after electrophoresis in 6% acrylamide/7 M urea thin gels. Mutant clones were easily identified because most of the mutations produced by this procedure caused an extra band to appear in the single track sequence. Clones carrying mutations in the target region were used to construct the corresponding pAll plasmids. The complete sequence of the promoter/3′ end signal region, including all promoter/3′ end signal/lac fusion junctions was determined for each pAll so constructed.

DNA SEQUENCING

Chain Terminator Method

Most of the DNA sequencing required over the course of this study was performed using chain-terminating inhibitors as originally described by Sanger et al. (1977a). This method was applied to the single-stranded DNA of recombinant M13 clones (Sanger et al., 1980; Messing, 1983), single-stranded pEMBL DNA (Dente et al., 1983) and to double-stranded plasmid DNA (Smith et al., 1979). The preparation of each type of template has already been described. Different procedures of annealing template to sequencing primer were applied to single- and double-stranded templates, as will be described below.

Primer:Template Annealing

Single-Stranded DNA Template

The oligonucleotide M13FP1, prepared in this laboratory (Table I), was used as a sequencing primer on single-stranded M13 or pEMBL(+) templates. About 0.5 μg of template DNA was annealed to 6-10 pmol of
FP1 in a volume of 8-10 μl, containing 10 mM MgCl₂, 10 mM Tris-HCl, pH 8.5, by heating to 55°C for 5 minutes and cooling to room temperature.

**Double-Stranded Template**

Plasmid DNA, prepared for sequencing as described earlier, was first digested with suitable restriction enzyme(s) for 1-2 hours in a volume of 10 μl of either low or medium salt buffer. About 6-10 pmol of oligonucleotide primer was added directly to the digest, and the mixture was heated to 100°C for 5 minutes and quickly chilled in ice-water.

**Sequencing Reactions**

The reactions were carried out in 0.5 ml Eppendorf tubes. α[32P]dATP (7-10 μCi, in 1 μl) and unlabelled dATP (12 - 15 pmol, in 1 μl) were added to the primer:template solution, and 2 μl aliquots of the resulting mixture were transferred to 4 tubes, marked C, T, A, G. The same volume of the appropriate ddNTP/dNTP "terminator mix" was then added to each tube. The compositions of the terminator mixes are listed in Table V.

The tubes were then transferred to a 30°C water bath, and the chain extension reaction was started by adding to each tube 2 μl of a freshly prepared dilution of DNA polymerase I (Klenow fragment) (0.25 u/μl in 10% glycerol,50 μg/ml BSA,10mM Tris-HCl pH 7.5,1 mM DTT). The reaction was allowed to proceed for 15 minutes, at which time 2 μl of a solution containing all 4 deoxynucleoside triphosphates at a concentration of 0.5 mM each was added. Fifteen minutes later, the reaction was stopped by adding 4 μl of formamide dye mix (90% formamide/20mM EDTA/0.03% bromophenol blue/0.03% xylene cyanol). The
### TABLE V

**COMPPOSITION OF ddNTP/dNTP TERMINATOR MIXES**

<table>
<thead>
<tr>
<th>Component</th>
<th>Mixture</th>
<th>C</th>
<th>T</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>dCTP</td>
<td></td>
<td>5.5</td>
<td>110</td>
<td>110</td>
<td>110</td>
</tr>
<tr>
<td>dTTP</td>
<td></td>
<td>110</td>
<td>5.5</td>
<td>110</td>
<td>110</td>
</tr>
<tr>
<td>dGTP</td>
<td></td>
<td>110</td>
<td>110</td>
<td>110</td>
<td>5.5</td>
</tr>
<tr>
<td>ddCTP</td>
<td></td>
<td>50</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ddTTP</td>
<td></td>
<td>-</td>
<td>500</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ddATP</td>
<td></td>
<td>-</td>
<td>-</td>
<td>50</td>
<td>-</td>
</tr>
<tr>
<td>ddGTP</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>300</td>
</tr>
</tbody>
</table>

a. All mixes also contained 2.5 mM Tris-HCl, pH 7.5, and 0.05 mM EDTA.
mixture was heated to 90° for 3 minutes, chilled in ice-water, and loaded immediately onto a 20 X 40cm X 0.3 mm thick 6% acrylamide/7M urea gel. Electrophoresis was carried out at 1,400 - 1,500 V for 1-3 hours. Each set of sequencing reactions was usually loaded on two gels, the first being run for 2-1/2 - 3 hours, and the second for 1-1/2 hours to allow different regions of the nucleotide sequence to be read.

**Sequencing by Base-Specific Chemical Cleavage**

**Purification of Fragments Prior to Sequencing**

End-labelled DNA fragments to be sequenced by the chemical cleavage method of Maxam and Gilbert (1977;1980) were purified after electrophoresis in acrylamide gels. The gel was autoradiographed to allow regions containing the DNA fragments of interest to be identified, and those regions were excised with a scalpel. Each gel piece was placed inside a dialysis bag containing 0.5 X TBE electrophoresis buffer. The bag was placed between and parallel to the electrodes of a horizontal electrophoresis apparatus and submerged in the same buffer. Electroelution of the DNA from the gel piece was carried out at 5 V/cm for 1-2 hours. The solution in the dialysis bag was transferred to a 1.5 ml tube. The gel piece was rinsed with the same buffer, and this rinse was added to the eluate. The eluate was phenol-extracted, and the DNA was then precipitated with ethanol, rinsed, dried, and dissolved in 30 μl of water.

The DNA was divided into 4 aliquots, two of 5 μl each for the C and G sequencing reactions, and two of 10 μl each for the (C+T) and (A+G) reaction. Carrier DNA (1 μl of a 1 mg/ml solution of calf thymus DNA) was added to each tube.
Base-Modification Reactions

C Reaction

Twenty microliters of 5M NaCl were added, and the C modification reaction was started by adding 30 ul of hydrazine. The reaction was allowed to proceed for 12 minutes at room temperature, at which point 300 μl of pyrimidine stop mix (0.3M NaOAc, pH 6.0; 0.1 mM EDTA; 50 μg/ml tRNA) were added, followed by 1 ml of cold ethanol (-70°).

(C+T) Reaction

The reaction was exactly the same as the C reaction, except that 15 μl of water were added to the DNA instead of 5 M NaCl before starting the reaction.

G Reaction

The DNA was diluted with 300 ul of cacodylate buffer (50 mM Na-cacodylate; 10 mM MgCl₂; 0.1 mM EDTA; pH 8.0) and the reaction commenced with the addition of 2 μl of dimethylsulfate. Five minutes later, base modification was stopped by adding 50 ul of G stop mix (2.5 M β-mercaptoethanol; 3 M NaOAc, pH 6.0; 0.1 M Mg(OAc)₂; 0.1 mM EDTA; 0.5 mg/ml tRNA) and 1 ml of cold ethanol.

(A+G) Reaction

The DNA was diluted with 10 μl of water, and depurination was initiated by adding 3 μl of 10% formic acid. The reaction continued for 10 minutes at 37°, at which time 300 μl of A stop mix (pyrimidine stop mix, with ATP added to 0.5 mM) were added, followed by 1 ml of cold ethanol.

Purification and Hydrolysis of Modified DNA

All samples were chilled in a dry ice/ethanol bath for 15 minutes, after which the DNA was pelleted by centrifugation for 5 minutes.
After drying, the pellet was resuspended in 0.25 ml of 0.3 M NaOAc, and the ethanol precipitation was repeated. The dried pellet was again resuspended, this time in 10 μl of water, and precipitated for a third time. The pellet was rinsed in ethanol, dried, and dissolved in 20 μl of 1 M piperidine. Samples were heated at 90° for 30 minutes and then lyophilized. They were twice redissolved in 20 μl of water and lyophilized before being dissolved in formamide-dye mix. They were denatured at 90° for 3 minutes, chilled in ice-water and loaded on acrylamide/7 M urea thin gels. Electrophoresis was carried out at 30-40 V/cm until the marker dyes had migrated a suitable distance. The gels were subjected to autoradiography at -20° without pretreatment.

**TRANSFORMATION of YEAST**

The procedure used for yeast transformation was similar to those described by Hinnen et al. (1978), Beggs (1978), Sherman et al. (1981), and Orr-Weaver et al. (1983). A single colony of the strain to be transformed was used to inoculate 5 ml of YPD, and the culture was incubated at 30° overnight. One ml of this saturated culture was used to inoculate 100 ml of fresh YPD, and the new culture was incubated at 30° until an OD₆₀₀ of about 0.3, corresponding to a cell density of about 1.4 x 10⁷ cells/ml, was reached. The cells were harvested by centrifugation at 4,000 rpm for 5 minutes at 4° in a Sorvall SS-34 rotor. The pellets were resuspended in a total of about 30 ml of 1 M sorbitol, combined, and the cells were pelleted as before. The cells were resuspended in 10 ml of 1 M sorbitol, and 50 μl of 1M DTT and 100 μl of Glusulase were added. The suspension was incubated at 30° with gentle shaking, and 5-10 μl samples were removed periodically, diluted with 50 μl H₂O and examined under a
phase-contrast microscope to determine the approximate proportion of non-refractile spheroplasts and "ghosts". If over 50% of the cells remained intact and refractile after 40 minutes, a second 100 µl aliquot of Glusulase was added. This was often necessary with strain GM-3C-2 but rarely with other strains. When 90% or more of the cells had been converted to spheroplasts, they were harvested by centrifugation at 2,000 rpm for 5 minutes at room temperature in an ICN benchtop centrifuge. The pellet was washed by gently resuspending in 10 ml of 1 M sorbitol and centrifuging in the benchtop centrifuge as before. A second wash with 10 ml of 1 M sorbitol and a third with 10 ml of STC (1M sorbitol, 10 mM Tris-HCl pH 7.5, 10 mM CaCl₂) were carried out. The final pellet was resuspended in 0.5-1.0 ml of STC. Aliquots of 0.1 ml were transferred to sterile plastic tubes (Falcon) and 1-10 µg of plasmid DNA, in a volume not exceeding 10 µl was added. The spheroplast suspensions were kept at room temperature for 15-20 minutes, and 1 ml of PEG-T-C (40% PEG 3350, 10 mM Tris-HCl pH 7.5, 10 mM CaCl₂) was added to each. After 15-20 minutes at room temperature, the suspensions were centrifuged at 2,000 rpm for 10 minutes in a Sorvall SS-34 rotor at room temperature. The viscous supernatant was carefully removed from each pellet with a sterile Pasteur pipet, and the pellet was gently resuspended in 0.5 ml of STC.

An aliquot of 0.25 ml was removed and added to 10 ml of molten RC agar which lacked a growth factor X and which had been held at 55°. The missing growth factor, either tryptophan, leucine or uracil, was required by the parental yeast strain but not by cells which had acquired the plasmid used for transformation. Immediately after adding the spheroplast suspension, the molten agar was mixed briefly and
poured over a plate of SC-X agar. Transformants gave rise to colonies embedded within the regeneration agar, which were easily visible after 2-3 days of incubation at 30°. Transformants were streaked onto plates of SC-X, and single colonies from these plates were used to inoculate cultures for storage, RNA isolation, or β-galactosidase assays.

**ISOLATION of YEAST RNA**

A single colony of yeast was used to inoculate 3-5 ml of an appropriate medium and the culture was incubated for 1-2 days at 30°. Selective media were used for all yeast strains carrying plasmids, but YPD was suitable for other strains. An aliquot of the small saturated culture was diluted into 100-200 ml of the same medium, which was then incubated at 30° until its A_{530} was between 0.6 (for cultures in selective medium) and 2 (for cultures in YPD). Cycloheximide was then added to the culture to a concentration of 0.1 mg/ml from a freshly-prepared 20 mg/ml stock in ethanol. Incubation was continued for 5 minutes, and the culture was then poured into two 250 ml centrifuge bottles, each half-full of crushed ice. Cells were harvested by centrifugation at 3,000 rpm for 1 minute at 4° in the Sorvall GSA rotor. The cell pellets were resuspended in ice-cold water containing cycloheximide (0.1 mg/ml), transferred to a chilled 30 ml Corex tube and pelleted by centrifugation at 5,000 rpm for 3 minutes at 4° in an SS-34 rotor. The pellets were immediately frozen in dry ice/ethanol. They could be kept frozen at -70° for at least several days without affecting the quality of the RNA extracted from them, but the extraction was usually done the day that the cells were harvested. Silanized, acid-washed glass beads were added to the frozen cell pellets (3 g beads/g wet weight cells) followed by 3 ml/g of ice-cold
RNA extraction buffer (0.15 M NaCl, 0.1 M Tris-HCl pH 7.5) and 50 μl/g of vanadyl ribonucleoside complexes (VRC: 0.2 M). Cells were broken by vortexing hard for six 15-second intervals, each followed by 45 seconds of cooling on ice. Cell debris was pelleted by centrifugation at 9,000 rpm for 10 minutes at 4° in an SS-34 rotor. The supernatant was transferred to a clean tube and SDS and proteinase K were added to concentrations of 0.5% and 0.5 mg/ml, respectively. The mixture was incubated in a 37° water bath. The cell debris was extracted a second time by adding RNA extraction buffer and VRC and vortexing as before. After centrifugation, the second extract was added to the first, and the SDS concentration was adjusted to 0.5%. Incubation at 37° continued for 1 hour. The mixture was then extracted once with an equal volume of phenol/chloroform (1:1), and the phases were separated by centrifugation (9,000 rpm, 10 minutes, 4°, SS-34). The aqueous supernatant was transferred to a clean tube and nucleic acid was precipitated by adding sodium acetate to 0.3 M, 2.5 volumes of ethanol, and chilling at -20° for 1-16 hours. The precipitate was collected by centrifugation (9,000 rpm, 20 minutes, 4°, SS-34) rinsed in cold ethanol, dried and dissolved in 5 ml of 20 mM EDTA. An equal volume of 4M LiCl was added, and the mixture was chilled on ice overnight to precipitate high molecular weight RNA. The final precipitate was collected by centrifugation (9,000 rpm, 40 minutes, 4°, SS-34), rinsed in cold 2 M LiCl/10 mM EDTA and dissolved in water. A final ethanol precipitation from 0.3 M NaOAc, followed by rinsing with ethanol, served to desalt the RNA. The precipitate was dried, dissolved in water and stored frozen. The yield of RNA was estimated from its UV absorbance, assuming a solution of 40 μg/ml RNA
to have an $A_{260}$ of 1.0. Up to 6 mg of RNA was obtained from 1 g (wet weight) of cells, which equals or exceeds the yields obtained by the author using procedures relying on repeated phenol extraction. The $A_{260}/280$ ratio of the RNA prepared by this method was at least 1.8, indicating that it was substantially free of protein. The LiCl precipitation was very effective in freeing high molecular weight RNA of smaller RNA species and DNA.

RNA Minipreps

The same RNA isolation procedure was adapted to the preparation of RNA from small cultures as follows: 2-3 ml of a suitable medium was inoculated with a single yeast colony and incubated until saturation was reached. An aliquot (approximately 2 ml) of this culture was diluted into 8 ml fresh medium and incubated at 30° until a cell density of about $2 \times 10^7$/ml was reached. The cultures were chilled in an ice-water bath and transferred to chilled 40 ml centrifuge bottles. Cells were harvested by brief centrifugation (5,000 rpm, 3 minutes, 4°, SS-34), resuspended in 1 ml of ice-cold water, transferred to a 1.5 ml Eppendorf tube, and pelleted by a 10-second spin in an Eppendorf centrifuge at 4°. The cell pellets were frozen in dry ice/ethanol. Ice-cold RNA extraction buffer (0.2 ml), VRC (10 ul), and glass beads (to meniscus) were added, and the tube was vortexed hard for six 15-second periods, each followed by 45 seconds on ice. After 15 seconds' centrifugation to pellet cell debris, the supernatant was transferred to a clean tube, and SDS (10 μl of 10%) and proteinase K (10 μl of a fresh 10 mg/ml solution in RNA extraction buffer) were added. RNA extraction buffer (0.7 ml) and VRC (10 μl) were added to the pellet, and the vortexing and centrifugation were repeated. The
second supernatant was combined with the first and the SDS concentration was adjusted to 0.5%. The combined extract was incubated at 37°C for 1 hour, after which an equal volume of 4 M LiCl was added. The tube was chilled on ice for 4-16 hours, and the precipitate was collected by centrifugation for 15 minutes in an Eppendorf centrifuge. The pellet was redissolved in 0.5 ml of water, and the LiCl precipitation was repeated. The second precipitate was rinsed with 2 M LiCl, 10 mM EDTA and redissolved in water (0.4 ml), and the UV spectrum of the resulting solution was measured. A final precipitation was carried out, this time using ethanol and 0.3 M NaOAc. After rinsing with ethanol, the final precipitate was dissolved in water to give an RNA concentration of about 10 μg/μl, based on the UV absorbance measured earlier. Yields from 10 ml cultures at 2 x 10^7 cells/ml were generally 200 - 400 μg. The ratio of A_{260}/280 was approximately 2.0.

Preparation of Glassware and Solutions for Handling RNA

Glass beads used in the isolation of yeast RNA were washed with concentrated HCl, rinsed exhaustively with glass distilled water, silanized, and baked at 200°C for at least 12 hours. All glassware used in the preparation, analysis and storage was similarly baked after washing and rinsing in glass distilled water. Plasticware, such as Eppendorf tubes and micropipet tips, was used without pretreatment from previously unopened packages. Solutions which were to come into contact with RNA were prepared with autoclaved glass distilled water in baked glassware. They were treated with 0.1% diethylpyrocarbonate (DEP) at 37°C for 12-24 hours and autoclaved prior to use. (In the case of solutions containing Tris, the solution was made up without
Tris and DEP-treated, and the Tris was added from an autoclaved stock. The complete solution was then autoclaved. Plastic gloves were worn whenever RNA samples, or solutions to be used with RNA, were being handled.

GEL ELECTROPHORESIS of RNA

RNA was analyzed by electrophoresis in agarose gels after denaturation by glyoxal and dimethylsulfoxide (Carmichael and McMaster, 1980; Thomas, 1980) or by electrophoresis in formaldehyde-agarose gels after denaturation with formaldehyde and formamide (Lehrach et al., 1977; Maniatis et al., 1982).

Electrophoresis After Denaturation with Glyoxal

RNA samples of up to 15 μg were denatured in a volume of 20 μl containing 1 M deionized glyoxal, 50% dimethylsulfoxide, and 10 mM (NaH₂PO₄ + Na₂HPO₄), pH 7.0, by heating at 50° for 1 hour. Sample loading buffer (5 μl) containing 50% glycerol, 10 mM (NaH₂PO₄ + Na₂HPO₄) pH 7.0, and 0.02% bromophenol blue, was added before loading the samples on a 1-1.4% agarose gel. The gel was cast and run in 10 mM phosphate buffer, pH 7.0. The buffer was recirculated during electrophoresis by means of a peristaltic pump. Electrophoresis was carried out at about 2-4 V/cm for 6-12 hours. The gel was not pretreated in any way before transferring the RNA from the gel to nitrocellulose.

Electrophoresis After Denaturation with Formaldehyde

As much as 20 μg of RNA was denatured in a volume of 20 μl containing 2.2 M formaldehyde, 50% formamide and 1/2X MOPS buffer (1X MOPS buffer contained 40 mM NaMOPS, 10 mM sodium acetate, 1 mM EDTA, pH 7.0) by heating at 55° for 15 minutes. Sample loading buffer (2
μl) was added, and the samples were loaded on a 1-1.4% agarose gel which had been cast in 1X MOPS buffer containing 2.2M formaldehyde. The gel was run in 1X MOPS buffer at 0.5-1 V/cm for 6-12 hours.

After electrophoresis, the gel was rinsed for 5 minutes in distilled water and then soaked for 1 hour in 20X SSC (1X SSC is 0.15 M NaCl, 0.015 M Na citrate) prior to transferring the RNA to nitrocellulose as described in the next section.

Transfer of RNA to Nitrocellulose

The procedures of Thomas (1980) were used to transfer RNA from agarose gels to nitrocellulose and hybridize the immobilized RNA to radiolabelled probes. The gel was placed on top of two sheets of Whatman 3MM paper on a glass plate. The ends of the 3MM paper were submerged in a reservoir of 20X SSC. A sheet of nitrocellulose, slightly larger than the gel, was wet with distilled water, then soaked in 20X SSC for a few minutes before being placed on top of the gel. Two more pieces of 3MM were placed on top of the nitrocellulose, followed by a 6 cm stack of paper towels. Care was taken to ensure that neither the paper towels nor the 3MM paper beneath them touched either the gel or the 3MM wick at the base of the stack. A glass plate and a mass of a few hundred grams was placed on top of the assemblage, and transfer was allowed to proceed for about 16 hours. The nitrocellulose filter was then air-dried briefly and baked under vacuum at 80° for 2 hours.

Prehybridization/Hybridization

Nitrocellulose filters with bound RNA were prehybridized for 4-16 hours at 42° in sealed plastic bags containing 10 ml of a mixture of 50% formamide, 5X SSC, 1X Denhardt’s solution (0.02% Ficoll, 0.02%
polyvinylpyrrolidone, 0.02% BSA), 50 mM (NaH₂PO₄ + Na₂HPO₄) pH 7.0 and denatured, sheared salmon sperm DNA (250 μg/ml). The prehybridization mixture was then removed and replaced with a hybridization mixture containing a radioactively-labelled DNA probe. The hybridization mixture had a total volume of 6-10 ml and was composed of four parts of prehybridization mixture and one part of 50% dextran sulfate. The probe was denatured by heating in a boiling water bath for 5-10 minutes and chilling in ice-water before adding it to the hybridization mixture. Hybridization proceeded for 16-24 hours at 42°, after which the filter was washed at room temperature in four changes of 2X SSC, 0.1% SDS, for five minutes each. After being washed at 50° in two changes of 0.1X SSC, 0.1% SDS for 15 minutes each, the filter was subjected to autoradiography.

Radioactive Labelling of a CYCl Hybridization Probe

The hybridization probe used for analysis of CYCl transcripts was the 600 bp EcoRI-HindIII fragment of pYeCYCl(2.5), which was purified from LMP agarose and end-labelled with DNA polymerase I (Klenow fragment). About 0.2 μg of DNA was incubated in 10 μl of Hin buffer with 1 u of Klenow fragment and 20-30 μCi of α[³²P]dATP for 30 minutes at room temperature. The mixture was then passed over a 1.5 ml column of Sephadex G-100 to remove unincorporated nucleotides. Radioactivity eluting in the void volume was collected and used as a hybridization probe after denaturation.

QUANTITATIVE ASSAY of β-gALACTOSIDASE in YEAST TRANSFORMANTS

Assays of β-galactosidase activity were carried out as described by Ruby et al. (1983). Strains to be assayed were grown to saturation at 30° in 3 - 5 ml of appropriate selective media (SC-Trp for strains
bearing TRP1+ plasmids; SC-Uracil for strains bearing URA3+ plasmids; SC for untransformed strains). A 2 ml aliquot of each saturated culture was diluted with 8 ml of fresh medium, and the culture was incubated at 30° for 4-6 hours to allow the population to complete about two doublings. The culture was then chilled in ice water, and the cells from a sample of 4-8 ml were harvested by centrifugation (5,000 rpm, 5', 4°, SS-34). The supernatant was removed by aspiration, and the cell pellet was resuspended in 0.4 ml of 50 mM potassium phosphate, pH 7. An equal volume of Z buffer (60 mM Na₂HPO₄, 40 mM NaH₂PO₄, 10 mM KCl, 1 mM MgSO₄, 40 mM -mercaptoethanol) containing 0.025% SDS and 1 mM PMSF was then added. The tube was vortexed for 5 seconds and left at room temperature for 10 minutes, after which 2 drops of chloroform were added. The tube was vortexed hard for 5 seconds and incubated at 30° for 10 minutes, and the reaction was started by adding 0.2 ml of a solution of ONPG (4 mg/ml in 0.1 M potassium phosphate, pH 7), vortexing for 2 seconds, and continuing to incubate at 30°. The release of o-nitrophenol was linearly related to the time of incubation for at least two hours as long as the tube was agitated during the incubation, whether continually in a shaking water bath or by vortexing for 5 seconds at 10-minute intervals. The assay was stopped at a convenient time by adding 0.5 ml of 1 M Na₂CO₃. Cell debris was pelleted by brief centrifugation (5,000 rpm, 5', SS-34 or SA600) and the absorbance at 420 nm of the supernatant or a suitably diluted aliquot was measured. Assays were always performed on 2-4 independent cultures of a given strain in an experiment. The levels of -galactosidase measured in duplicate samples of a single culture were identical to within 5%, but
the levels measured in independent cultures of the same strain, or in samples removed at different times during the exponential growth of one culture, varied by up to 20%.

\( \beta \)-galactosidase activity in a culture was calculated using the expression:

\[
\text{\( \beta \)-gal Activity} = \frac{1,000 \ A_{420}}{A_{600} \ V \ t} \text{ units/ml, where}
\]

\( A_{420} \) = absorbance at 420 nm of the assay supernatant;
\( A_{600} \) = absorbance at 600 nm of the yeast culture at the time it was sampled for assay; \( V \) = volume in ml of the sample, and \( t \) = time, in minutes, between the addition of ONPG and the addition of Na2CO3 to the assay mix. The \( A_{600} \) was directly proportional to the cell number as determined by plating, up to an \( A_{600} \) of 1.0. At higher values of \( A_{600} \), the \( A_{600} \) underestimated cell number. The \( A_{600} \) of cultures used for \( \beta \)-galactosidase assay was between 0.5 and 0.8.
RESULTS and DISCUSSION

The CYC1 3' End Signal Resides Within 300 bp Downstream of the Coding Sequence

Yeast strain GM-3C-2, constructed by G. Faye, produces no functional cytochrome c whatsoever (Faye et al., 1981). It carries both the cycl-1 deletion, which completely removes the gene encoding iso-1-cytochrome c, (Sherman et al., 1975) and a nonsense mutation in CYC7, which leads to the production of an inactive fragment of iso-2-cytochrome c. As a result, the strain is able to grow only on media containing a fermentable carbon source. A second consequence of the cycl-1 deletion is that total RNA isolated from strain GM-3C-2 does not hybridize to a radioactively labelled fragment of the CYC1 coding region when hybridization and washing are carried out as described in Materials and Methods. (See Figure 14.)

The CYC1+ gene normally resides on a 2.5 kb BamHI/HindIII fragment of yeast chromosome X, as shown in Figure 15. This fragment may be cloned in an autonomously replicating plasmid such as YEpl3, to produce YEpl3CYC1(2.5), and introduced into strain GM-3C-2 by transformation. The transformants are capable of normal growth on glycerol or lactate, indicating that sequences on the recombinant plasmid complement the cytochrome c deficiency of the parental yeast strain.

Figure 14 illustrates the hybridization of a radioactively labelled CYC1 probe to RNA isolated from strain GM-3C-2, from a transformant of GM-3C-2 carrying the plasmid YEpl3CYC1(2.5), and from a strain carrying
Figure 14. Hybridization of a CYC1 Probe to Total Yeast RNA. Total yeast RNA was denatured with glyoxal and DMSO as described in Chapter II and was then electrophoresed through a 1% agarose gel. After electrophoresis, the RNA was transferred to nitrocellulose and probed with a $^{32}$P-labelled 600 bp EcoRI/HindIII fragment which included most of the CYC1 coding sequence (see Figure 15). The probe was end-labelled with $\alpha[^{32}P]dATP$ using the Klenow fragment of DNA polymerase I, as described in Chapter II. Each lane contained 15 $\mu$g of RNA. The RNA in lane 1 was from D311-3A (CYC1$^+$), that in lane 2 was from strain GM-3C-2 (cycl-1), and that in lane 3 was from a transformant of GM-3C-2 carrying plasmid YEpl3CYC1(2.5). (See Figure 15 for a map of the plasmid.)
Figure 15.

A. Restriction Map of CYCl and Flanking Regions
The map is from the sequence data of Smith et al. (1979) and D.W. Leung and M. Smith (unpublished results). Recognition sites for restriction endonucleases are labelled as follows: B, BamHI; E, EcoRI; H, HindIII; K, KpnI; S, SalI; Sm, SmaI; X, XhoI. The CYCl coding sequence is indicated by cross-hatching. The wavy line represents the CYCl transcript, with (A)n denoting the poly(A) tail. The transcript initiates heterogeneously, 10-150 nucleotides upstream of the coding sequence (McNeil and Smith, 1985). Its 3' end is located 172-175 nucleotides downstream of the coding sequence (Boss et al., 1981).

B. Structure of YEpl3CYCl(2.5)
The CYCl coding sequence is cross-hatched. Stippling is used to denote sequences from the yeast plasmid, 2μ circle. The single line represents pBR322, and the open box represents a region of yeast chromosome III carrying the LEU2+ gene. The filled circle indicates the normal 3' end site of CYCl transcripts, and the open circle represents the approximate position of the 3' end site of the transcripts of a 2μ circle gene called FLP [Broach et al. (1979)]. Restriction sites are labelled as in A, with the addition that P represents a PstI recognition site.
the "wild-type" chromosomal CYCl gene. While no hybridization to the RNA of GM-3C-2 can be detected, both of the other strains produce an RNA of about 650 nucleotides which is detected by the CYCl probe. This species is polyadenylated, although the proportion which binds to oligo(dT)-cellulose in 0.1 M NaCl may be as low as 50%, (not shown) presumably because of the short length of poly(A) sequences in yeast (Groner et al., 1974). The 650 nucleotide species must represent the CYCl mRNA. Other estimates of the length of the CYCl mRNA in wild-type yeast range from 630-700 nucleotides (Faye et al., 1981; Boss et al., 1980, 1981; Zaret and Sherman, 1982). Faye et al. (1981) and McNeil and Smith (1985) have shown that the mRNAs transcribed from the CYCl gene on the plasmid YEpl3CYCl(2.5) have the same distribution of 5' ends as those produced from a single chromosomal copy of the gene. Since the plasmid and chromosome-derived CYCl transcripts are indistinguishable in length, their 3' ends must also be generated at nearby, if not identical, sites. Boss et al., (1981) have shown that CYCl mRNA in a "wild-type" yeast strain is polyadenylated at a site 172-175 nucleotides downstream of the coding sequence. Utilization of the same 3' end site in transcripts of the plasmid-borne CYCl gene indicates that the 2.5 kb BamHI/HindIII fragment contains all of the sequences which are necessary for use of the normal CYCl mRNA 3' end site. These sequences are collectively referred to as the CYCl 3' end signal. A distinction is made in this work between the 3' end signal as a functional element determining where 3' end generation takes place, and the 3' end site as the position at which 3' end generation occurs. The nucleotide sequence at the 3' end site may or may not be an important component of the 3' end signal.
Deletion of the 3' End Signal Causes the Synthesis of Extended Transcripts

G. Faye provided two plasmids which were useful in preliminary experiments for designing the approach used in identifying the CYCl 3' end signal. The two plasmids, 2H26 and 4H40, differed from YEpl3CYCl(2.5) only in carrying deletions extending from the HindIII site downstream of CYCl toward the gene for different distances. The exact deletion endpoint of each plasmid was determined by subjecting a fragment of each plasmid, end-labelled at the HindIII site, to the chemical sequencing procedure of Maxam and Gilbert (1977; 1980). Plasmid 2H26 retained 179 bp of the normal CYCl 3' untranslated and flanking sequences, while 4H40 carried a deletion that extended 65 bp further upstream, to within 114 bp of the coding sequence. The two plasmids were used to transform GM-3C-2 to leucine prototrophy, and RNA isolated from the transformants was resolved by gel electrophoresis, transferred to nitrocellulose, and hybridized with a labelled CYCl probe. The result is shown in Figure 16. Transformants carrying plasmid 2H26 produced, in addition to the normal 650 nucleotide CYCl mRNA, low levels of an extended transcript, about 1,100 nucleotides long. Transformants bearing plasmid 4H40, in contrast, produced only extended transcripts of about 1,100 and 1,300 nucleotides. Transformants of both types were capable of growth on plates containing glycerol or lactate as principal carbon source, indicating that the extended transcripts were functional CYCl mRNAs. The extended transcripts might differ from the normal CYCl mRNA with respect to the positions of their 5' ends or 3' ends or both. However, Faye et al. (1981) and McNeil and Smith (1985) have shown that the
Figure 16. Hybridization of a CYCl Probe to RNA from Strains Bearing Deletions Distal to CYCl

Total yeast RNA was denatured with glyoxal and DMSO and electrophoresed through a 1% agarose gel. The RNA was transferred from the gel to a nitrocellulose filter, which was then probed with a [32P]-labelled 600 bp EcoRI/HindIII fragment including most of the CYCl coding sequence (see Figure 15). The probe was end-labelled with α[32P]dATP and the Klenow fragment of DNA polymerase I. The filter was washed and autoradiographed as described in Chapter II. Each lane contains 15 μg of RNA. Lane 1, RNA from GM-3C-2/YEp13CYCl(2.5); lane 2, RNA from GM-3C-2/2H26; lane 3, RNA from GM-3C-2/4H40. Other lanes contained RNA from transformants of GM-3C-2 carrying plasmids with more extensive deletions of CYCl sequences than plasmid 4H40. They are not relevant to this study. Numbers in the right-hand margin indicate the sizes (in bp) and positions of fragments produced by TaqI digestion of X174 RF DNA. These fragments were end-labelled with α[32P]dCTP using Klenow fragment. They were subsequently denatured in exactly the same way as the RNA samples and electrophoresed in the same gel.
positions of the normal 5' ends of CYCl mRNA depend on sequences located 20-150 bp upstream of the gene, and Guarente and co-workers have shown that a fusion of the CYCl promoter to lacZ or LEU2 coding and 5'-untranslated sequences have the same 5' end distribution as normal CYCl transcripts (Guarente and Mason, 1983; Guarente et al., 1984). Since the deletions in 4H40 and 2H26 remove only 3' untranslated and flanking sequences, it is unlikely that they would affect the 5' end distribution of the CYCl transcripts. Although the transcript ends were not mapped in this study, it is reasonable to conclude that the extended, 1,000 nucleotide transcript produced from 4H40 and 2H26 extends about 500 nucleotides beyond the 3' end site of the normal CYCl mRNA. The 3' end of the extended transcript therefore lies in the 2μ circle sequences of the plasmid, in a region known to correspond to the 3' end site of the FLP transcript of the intact 2μ circle. (Broach et al., 1979). The extended CYCl transcripts approach the FLP 3' end site from the same direction as the normal FLP transcript. It appeared from these results that disruption of the normal CYCl 3' end signal caused transcription to extend beyond the position corresponding to the normal 3' end site. Zaret and Sherman (1982) also noted the synthesis of 3'-extended CYCl transcripts from the cycl-512 allele, which carries a deletion of sequences normally found downstream of the gene.

Most of the CYCl mRNA in 2H26 transformants of GM-3C-2 is of normal size, which suggest that the normal CYCl 3' end site was used quite efficiently in these cells and that the 3' end signal remained functional. The presence of low levels of extended transcripts suggested a slight reduction in the "strength" of the 3' end signal,
perhaps caused by the encroachment of the deletion in 2H26 into sequences at the downstream boundary of the signal. Since the CYCl-proximal deletion endpoint was within 5 bp of the CYCl 3' end site, it seemed at this stage that sequences at the 3' end site might form part of the 3' end signal.

**Mapping the Boundaries of the CYCl 3' End Signal**

An effort was made to localize the boundaries of the CYCl 3' end signal by means of an approach first used to systematically define the boundaries of the promoter regions of the *Xenopus* 5S RNA genes (Sakonju et al., 1980; Bogenhagen et al., 1980) and the Herpesvirus tk gene (McKnight et al., 1981). Two sets of nested deletions were introduced into the 2.5 kb CYCl BamHI/HindIII fragment, and their effects on CYCl transcription were studied after introducing the modified fragments into yeast on high copy number plasmids. The two sets of deletions extended into the 3' untranslated region of the CYCl gene from opposite directions, all the members of one set sharing a common endpoint either upstream or downstream of the region of interest. Deletions in the CYClAH5' series were used to identify the downstream boundary of the 3' end signal. They were like the deletions in plasmids 4H40 and 2H26 in that they extended toward the CYCl coding sequence for varying distances from the HindIII site 275 bp downstream. Deletions in the CYClAK3' series extended toward the HindIII site for varying distances from the KpnI site within the CYCl coding sequence. They served to identify the upstream boundary of the 3' end signal. Figure 17 illustrates the relationship of the two sets of deletions to each other and to the intact CYCl gene. The construction of both sets of deletion derivatives of the 2.5 kb CYCl fragment is described in detail in
Figure 17. Endpoints of the CYClΔH5' and CYClΔK3' Deletions
The CYCl coding sequence is stippled. The endpoint of each deletion is indicated, taking the A of the CYCl initiation codon as +1. The normal CYCl transcript 3' end site spans positions +502-504, and it is indicated by the filled circle.
Chapter II. Selected derivatives in each set were transferred to YEpl3 as BamHI/Hindlll fragments, and the resulting YEpl3CYClA+ and YEpl3CYClA-K3' plasmids were used to transform GM-3C-2, selecting for LEU2+ transformants.

The hybridization of a CYCl probe to RNA isolated from YEpl3CYClA+ transformants of GM-3C-2 is illustrated in Figure 18. Deletions extending as far upstream as position +518 apparently do not interfere with the utilization of the normal CYCl 3' end site, since the only transcript detected by a CYCl probe in cells carrying such deletions in a plasmid-borne copy of CYCl is of normal size. The CYCl-proximal deletion endpoint in the YEpl3CYClA+(+497) plasmid is 21 bp further upstream, and in cells carrying this plasmid, an extended transcript about 500 nucleotides longer than the normal transcript is present at low levels, along with much higher levels of the normal CYCl transcript. This result is similar to that obtained with RNA from 2H26 transformants (Figure 16). The difference is that the CYCl-proximal deletion endpoint is 12 bp further upstream in YEpl3CYClA+(+497) than in 2H26, with the result that the former plasmid lacks the sequence of the normal CYCl 3' end site (positions +502-505). The efficient production of a normal CYCl transcript from YEpl3CYClA+(+497) suggests that the 3' end signal remains functional in the absence of the normal 3' end site. The signal must cause 3' end generation at an approximately equivalent position in sequences flanking the CYClA+(+497) deletion. The normal sequence of the CYCl 3' end site does not appear to be an important part of the 3' end signal, although the production of a low level of extended transcripts from both the CYClA+(+497) and 2H26 templates suggests that sequences in the
Figure 18. Hybridization of a \textit{CYCl} Probe to RNA from Transformants of GM-3C-2 Carrying \textit{CYClA}H5' Genes.

Total yeast RNA was denatured with formaldehyde and electrophoresed through a 1\% agarose gel. The RNA was transferred from the gel to nitrocellulose and was subsequently hybridized to a $^{[32P]}$-labelled \textit{CYCl} probe. The probe was made by copying the single-stranded DNA of mp8\textit{CYCl}(2.5) using oligonucleotide FPl (Table I) as a primer, $\alpha[^{32P}]dATP$, unlabelled dCTP, dGTP and dTTP, and the Klenow fragment of DNA polymerase I. Each marked lane contained 20 \textmu g of RNA. The lane marked WT contained RNA from a transformant of GM-3C-2 carrying plasmid YEpl3\textit{CYCl}(2.5). The other lanes contained RNA from transformants of GM-3C-2 carrying various YEpl3\textit{CYClA}H5' plasmids. The \textit{CYClA}H5' deletion in each plasmid extended from the HindIII site at position +605 to the position indicated at the top of the lane.

vicinity of the 3′ end site may play a minor role in making the 3′ end signal maximally efficient.

Figure 18 shows that a deletion extending from the HindIII site to a position 35 bp or more upstream of position +497 completely abolishes the activity of the 3′ end signal. The extended CYCl transcript is the only one detectable in cells carrying YEpl3CYClΔH5′ plasmids with CYCl-proximal deletion endpoints at position +464, or further upstream. From these results it may be concluded that one boundary of the functional CYCl 3′ end signal lies between positions +464 and +497.

Figure 19 illustrates the hybridization of a CYCl probe to RNA isolated from transformants of GM-3C-2 carrying various YEpl3CYClΔK3′ plasmids. The deletions in these plasmids removed sequences which are normally part of the CYCl transcription unit. Those deletions which did not interfere with CYCl mRNA 3′ end generation were therefore expected to cause the length of the mRNA to be reduced by an amount equal to the length of the deleted sequences. Those deletions which disrupted the 3′ end signal were again expected to cause the extension of CYCl transcripts into 2μ circle sequences on the plasmid.

Deletions extending downstream from the KpnI site in CYCl as far as position +449 apparently have little effect on the CYCl 3′ end signal because the predominant CYCl transcript detected in cells carrying CYClΔK3′ derivatives with such deletions is slightly shorter than the normal CYCl mRNA. The length difference in each case is approximately equal to the extent of the deletion.

Very low levels of an extended transcript are produced from CYClΔK3′ templates with deletions extending as far as position +390. Deletions extending to positions +407 or +429 cause a slight increase in the level of the extended transcripts, which still account for only
Figure 19. Hybridization of CYC1 Probe to RNA from GM-3C-2 Transformants Carrying CYC1ΔK3′ Genes.

Total yeast RNA was denatured with glyoxal and DMSO and was then electrophoresed through a 1% agarose gel. The RNA was transferred from the gel to a nitrocellulose filter. The filter was incubated with a $^{32}$P-labelled CYC1 probe, washed, and autoradiographed. The hybridization probe was the 600 base pair EcoRI-HindIII fragment of CYC1, end-labelled with $\alpha[^{32}\text{P}]\text{dATP}$ and the Klenow fragment of DNA polymerase I.

Each marked lane contained 15 µg of RNA. The lane marked WT contained RNA from a GM-3C-2 transformant carrying plasmid YEpl3CYC1(2.5). The other lanes contained RNA from GM-3C-2 transformants carrying YEpl3CYC1 ΔK3′ plasmids. A given plasmid carried a deletion extending from the KpnI site at position +247 in the CYC1 coding sequence to the position indicated at the top of the lane.
a small fraction of the RNA detected by a CYCL probe.

The CYCL AK3'(+449) derivative directs the production of a shortened CYCL transcript, but no extended transcript is observed. This may mean that the 3' end signal in the CYCL AK3'(+449) mutant is actually more efficient than in AK3' mutants with shorter deletions. A more plausible interpretation is that the extended transcript does not accumulate to detectable levels in the CYCL AK3'(+449) mutant because its stability is lower than that of the extended transcripts produced by mutants with shorter deletions.

A gradual drop in the level of the shortened transcript with increasing deletion length is evident in cells carrying the CYCL AK3'(+390), CYCL AK3'(+407) and CYCL AK3'(+429) genes, with quite a marked drop in transformants bearing the CYCL AK3'(+449) gene. This observation may indicate that sequences upstream of position +449 influence the stability of CYCL transcripts. Other explanations, some of them trivial, cannot be excluded at present. For example, the probe used in this study may have detected shorter transcripts less efficiently than full-length transcripts, because the length of the region of complementarity between probe and transcript decreased with transcript length.

The important point for the purposes of this study is that CYCL AK3' derivatives with deletion endpoints as far downstream as position +449 appeared to retain a functional CYCL 3' end signal. The synthesis of small quantities of extended CYCL transcripts from templates with shorter AK3' deletions was interpreted to mean that the context of the 3' end signal could influence its efficiency without preventing it from functioning altogether.
The extension of a ΔK3′ deletion as far as position +474 prevents the utilization of the normal \textit{CYCl} mRNA 3′ end site. Only extended transcripts are evident in cells carrying the \textit{CYClΔK3′(+474)} or \textit{CYClΔK3′(+497)} genes. This suggests that the deletion of sequences between positions +449 and +474 inactivates the \textit{CYCl} 3′ end signal.

Analysis of the \textit{CYCl} transcripts produced from \textit{CYClΔH5′} and \textit{CYClΔK3′} templates (Figures 18,19) allows the functional boundaries of the \textit{CYCl} 3′ end signal to be mapped: one lies between positions +449 and +474, and another maps between positions +464 and +497. Sequences within these boundary regions are essential to \textit{CYCl} 3′ end signal function and are illustrated in Figure 20. Sequences flanking the boundaries appear to have minor effects on the efficiency of 3′ end generation at the normal \textit{CYCl} 3′ end site. However, it would be premature to conclude that the 49 bp region within the boundaries constitutes or includes an autonomous functional element capable of causing 3′ end generation in any context. Each of its boundaries was defined by deletions which removed only sequences flanking that boundary, while the other boundary and its flanking sequences remained intact. It is possible that some general feature of sequences on both sides of the 49 bp region or a sequence repeated on either side of this region is required for 3′ end generation. This requirement might not become evident until the normal flanking sequences from both sides of the 49 bp region were replaced by other sequences.

\textbf{Point Mutations Within the 3′ End Signal Region of the Intact CYCl Gene}

Definition of the boundaries of the \textit{CYCl} 3′ end signal allows the question of the specific sequence requirements of transcript 3′ end
Figure 20. Sequence of the CYC1 3' End Signal and Flanking Regions

A. The CYC1 coding sequence and its immediate flanking regions are shown (Smith et al., 1979). The 3' end signal region, as defined by the CYC1ΔH5' and CYC1ΔK3' deletions, is bracketed. The filled circle indicates the most 5' site of polyadenylation of CYC1 transcripts as reported by Boss et al. (1981). Because there are three consecutive A/T base pairs at positions +503-505, polyadenylation at any of these positions would produce the same mature transcript.

B. The 3' end signal region of CYC1 is shown. Only the mRNA-parallel strand of the DNA sequence is presented. The sequence alterations introduced by the GG462, C482, GT473, and C474 mutations are indicated.
generation in yeast to be approached. As discussed earlier the sequence AATAAA had been shown to be required for polyadenylation in higher eucaryotes (Fitzgerald and Shenk, 1981, and references cited in Chapter I). It was therefore worth testing the importance of any analogous sequences in the CYCl 3’ end signal. The closest such analog to be found is the sequence ATTTAA, at positions +480 to +484. Accordingly, the C482 mutation, which altered this sequence to ATCTAA, was introduced into the 3’ end signal using a mutagenic oligonucleotide as described by Zoller and Smith (1982;1983). (The procedure is described in Chapter I, and the sequence of the oligonucleotide used, oAS2, is given in Table I.)

A striking feature of the 3’ end signal region and its neighbouring sequences is its high content of adenine:thymine base pairs. A tract of seven consecutive T/A base pairs is present at positions +459 to +465, and in light of the importance of 3’-terminal uridylate tracts in transcript termination by bacterial RNA polymerase and by eucaryotic RNA polymerases I and III, it was reasonable to suspect that this tract of T/A base pairs might be involved in 3’ end generation. The GG462 mutation was introduced by means of a mutagenic oligonucleotide, oASl (Table I), in order to disrupt the T/A tract. The mutation replaced the T/A base pairs at positions +462 and +463 with G/C base pairs. The 3’ end signal sequences created by the GG462 and C482 mutations are noted in Figure 20.

Both mutations were cloned into plasmid YEpl3 as part of the 2.5 kb BamHI-HindIII fragment and introduced into yeast strain GM-3C-2, selecting for LEU2+ transformants. RNA isolated from both types of transformant was probed for CYCl sequences after electrophoresis and
transfer to nitrocellulose, and the result is shown in Figure 21. The C482 mutation had no detectable effect on 3' end generation in CYCl transcripts inasmuch as no transcript other than the normal 650 nucleotide species was evident. Small changes in the exact location of transcript 3' termini would not have been detected by this type of analysis.

Transformants carrying the CYClGG462 allele synthesized, in addition to the normal CYCl transcript, small amounts of an 1100 nucleotide transcript, indicating that the GG462 mutation reduced the efficiency of 3' end generation at the normal 3' end site. The extended transcript accumulated to less than one-tenth the level of the normal mRNA, which suggests that unless the stability of the two transcripts differed markedly, the altered 3' end signal remained more than 90% efficient.

A derivative of YEpl3CYCl(2.5) was constructed in which the CYCl 3' end signal contained both the GG462 and C482 mutations, and RNA isolated from yeast transformants carrying this plasmid was also probed for CYCl sequences. The results, shown in Figure 21, suggest that the presence of both mutations may cause a modest increase in the relative level of the extended CYCl transcript. That a mutation such as C482, which by itself does not alter the efficiency of 3' end generation, can augment the effect of another mutation, such as GG462, suggests there might be some redundancy in the functionally important elements of the 3' end signal. If this is the case, then the mutational analysis of the 3' end signal might be simplified by restricting it to the smallest segment still capable of bringing about 3' end generation. The deletion analysis described earlier suggested that this segment would
Figure 21. Hybridization of a CYCl probe to RNA from Yeast Strains with Point Mutations in the CYCl 3’ End Signal. Yeast RNA was denatured with glyoxal and DMSO and electrophoresed through 1% agarose. The RNA was transferred from the gel to a nitrocellulose filter, which was subsequently incubated with a \( ^{32}P \)-labelled CYCl hybridization probe, washed, and autoradiographed. The probe was the 600 base pair EcoRI-HindIII CYCl fragment, end-labelled with \( \alpha[^{32}P]dATP \) using the Klenow fragment of DNA polymerase I. Lanes marked "a" contained 15 \( \mu \)g of total yeast RNA. Those marked "b" contained RNA which bound to oligo(dT)-cellulose in 0.1 M NaCl, while those marked "c" contained RNA which failed to bind. The amount of RNA loaded in each lane "b" or "c" corresponded approximately to the amount of each fraction found in 15 \( \mu \)g of total RNA. Lanes marked WT contained RNA from GM-3C-2 transformants carrying plasmid YEpl3CYCl(2.5). Other lanes contained RNA from GM-3C-2 transformants carrying plasmid YEpl3CYClGG462, YEpl3CYClC482, or YEpl3CYClGG462C482, as indicated by the name of the mutation at the top of each group of lanes.
reside within the 49 bp region between positions +448 and +497.

A Screening System for 3' End Signal Mutations

The identification of the precise sequence requirements of the CYCl 3' end signal would be facilitated if there was a way of rapidly screening for mutants with defects in the 3' end signal. Such defects could not necessarily be expected to have any appreciable effect on CYCl expression, because it had already been shown that functional CYCl mRNA could be produced in the complete absence of the normal 3' end signal, at least if another 3' end signal was present some distance downstream. The fact that disruption of the normal 3' end signal caused the production of extended transcripts suggested that if some gene was placed downstream of the CYCl 3' end signal, its expression might be dependent upon the inactivation of the 3' end signal. The E. coli lacZ gene promised to be a suitable indicator for 3' end signal inactivation, because it can be expressed in yeast and its product, $\beta$-galactosidase, is easily assayed. S. cerevisiae produces no $\beta$-galactosidase of its own which might interfere with the assay for lacZ expression (Guarente and Ptashne, 1981).

If the expression of an indicator gene is to be an unambiguous measure of the efficiency of a transcriptional 3' end signal, the gene should be transcribed only as part of an extended mRNA from a promoter located upstream of the 3' end signal. Neither the efficiency of transcript initiation at the promoter, nor the efficiency of translation of the coding sequence of the indicator gene, should vary. The functional arrangement envisioned here, promoter-3' end signal-indicator gene, is of course analogous to that found in many bacterial biosynthetic operons (see Chapter I). It has been used quite
successfully to detect and study procaryotic terminators.

Casadaban and Cohen (1980) demonstrated that transcription terminators in cloned bacterial DNA fragments could be detected by virtue of their ability to prevent lacZ expression when inserted between the gene and its promoter. A similar system using the tet gene instead of lacZ was described by Enger-Valk et al. (1981), and Rosenberg et al. (1983) were able to detect mutations interfering with transcript termination at a given terminator by selecting for increased expression of a galK gene located downstream of the terminator. More recently, Honigman et al. (1985) described a plasmid in which the lacZ or galK gene served as an indicator of terminator efficiency, while the cat gene provided an internal control for plasmid copy number.

The success of this approach to studying procaryotic terminators depends in part on the mechanism of translation initiation in procaryotes. Translation initiates at AUG triplets preceded by suitable ribosome binding sites, and the presence of other sequences upstream of these sites is not sufficient to prevent initiation. In yeast, as is usually true in other eucaryotes, translation of most mRNAs initiates at the AUG triplet closest to the 5′ end of the message (Kozak, 1983a; 1983b; Sherman and Stewart, 1981; Stiles et al., 1981). For this reason, mRNAs in eucaryotes, in contrast to those in procaryotes, are usually functionally monocistronic. [See Zitomer et al. (1984) for evidence of initiation at sites other than AUG triplets in yeast.]

For lacZ, or any other gene, to serve as an indicator for the function of an upstream 3′ end signal in yeast, it seemed likely to be important that no AUG triplets intervene between the 5′ end of the extended lacZ message and the lacZ initiation codon. The type of
construct envisioned, and the consequences of inactivating the CYCl 3′ end signal, are illustrated in Figure 22. The CYCl 3′ end signal would be linked directly to the CYCl promoter upstream and to the lacZ coding sequence downstream. Inactivation of the 3′ end signal would cause the production of an extended mRNA encoding β-galactosidase. The leader of the mRNA would include part of the normal CYCl leader sequence and the sequences of various defective 3′ end signals. One AUG triplet is present in the 49 bp region within the functional boundaries of the 3′ end signal, at positions +472 to +474 (see Figure 20). The scheme proposed in Figure 22 could only be useful if this AUG triplet could be eliminated without inactivating the 3′ end signal. It seemed unlikely that any other feature of the 3′ end signal, or any alteration which did not introduce an AUG triplet, would interfere with the translation of an extended lacZ mRNA. Comparisons of many eucaryotic mRNAs have revealed no other constraints on their leader sequences (Kozak, 1983b). Leaders ranging in length from 1-200 nucleotides are known, although very long leader sequences may reduce translation efficiency (Darveau et al., 1985). Zitomer et al. (1984) have recently shown that substantial alterations in the length and sequence of the 5′ untranslated region of a CYCl:galK fusion have little or no effect on the levels of galactokinase produced in yeast, unless out-of-frame AUG triplets are introduced into the leader region by the alterations. Johansen et al. (1984) made similar observations regarding the expression of galK in monkey and hamster cells.

Because the leader sequence of a eucaryotic mRNA seems to have little influence on translation efficiency as long as it is free of AUG triplets, it was judged to be worthwhile to eliminate the single AUG triplet from the CYCl 3′ end signal and, as long as this alteration did
Figure 22. Strategy for Screening for Mutations in the 3' End Signal of CYCl

A. The CYCl 3' end signal is inserted between the CYCl promoter and the lacZ coding sequence. If the 3' end signal is functional, it is expected to prevent expression of lacZ.

B. Mutational inactivation of the 3' end signal is expected to allow lacZ expression.
not inactivate the 3' end signal, include the altered signal in the
type of construct shown in Figure 22. Mutations could then be
introduced into the 3' end signal at random and those which inactivated
it detected through the consequent increase in \textit{lacZ} expression.

**Mutagenesis of Positions +472 to +474 Within the CYCl 3' End Signal**

It was first necessary to test the effects of altering the 3' end
signal sequence at positions +472 to +474. Oligonucleotides oAS3 and
oAS4 (Table I) were used to produce two mutant 3' end signal regions,
with the sequences AGT and ATC, respectively, at positions +472 to
+474. The mutations are referred to as GT473 and C474, and they were
created using the procedure of Zoller and Smith (1982;1983) as
described in Chapter II. Figure 20 shows the sequences of the mutant
3' end signal regions with the normal sequence of the region for
comparison. Each mutation was introduced into the 2.5 kb BamHI-HindIII
\textit{CYCl} fragment, and the altered fragments were ligated separately into
\textit{YEpl3} and introduced into yeast strain GM-3C-2, selecting \textit{LEU2+}
transformants. Figure 23 illustrates the pattern of hybridization of a
\textit{CYCl} probe to RNA isolated from transformants carrying each type of
plasmid. The transcripts of the \textit{CYCIGT473} and \textit{CYClC474} alleles are
identical in length to the transcript of the normal \textit{CYCl} gene. This
observation suggests that neither the GT473 nor the C474 mutation
interferes with the function of the 3' end signal. It seemed feasible
to use a 3' end signal carrying one of these mutations to construct a
promoter 3' end signal: \textit{lacZ} fusion plasmid which would provide a
phenotypic indicator of 3' end signal function.

**Testing Various Promoter:3' End Signal: \textit{lacZ} Fusion Plasmids**

The construction of promoter:3' end signal: \textit{lacZ} fusion plasmids is
Figure 23.

A. Sequence in the Vicinity of the GT473 and C474 Mutations, and the Corresponding Wild-Type Sequence.

The wild-type CYCl 3’ end signal region and its GT473 and C474 mutant derivatives were sequenced by the chain termination method of Sanger et al. (1977a) after cloning in M13mp8 (Messing, 1983). From left to right, the sequencing channels in each case are C, T, A, G. The sequence displayed is that of the mRNA-antiparallel strand of the 3’ end signal region, and consequently, the direction of reading (from bottom to top) is toward the CYCl coding sequence. The sequence at positions +474 to +472 is highlighted in each case for comparison.

B. Hybridization of a CYCl Probe to RNA from GM-3C-2 Transformants Carrying the GT473 and C474 Mutations of the 3’ End Signal.

Total yeast RNA was denatured with formaldehyde and electrophoresed through a 1.4% agarose gel. The RNA was then blotted to nitrocellulose. The filter was incubated with a $^{32}$P-labelled CYCl probe, washed and autoradiographed, all as described in Chapter II. The probe was the 600 bp EcoRI-HindIII fragment of CYCl, end-labelled with $\alpha[^{32}P]dATP$ using the Klenow fragment of DNA polymerase I. Lanes M contained EcoRI-HindIII fragments of X DNA which had been end-labelled with $\alpha[^{32}P]dATP$ using Klenow fragment, then denatured, electrophoresed and transferred in exactly the same way as the RNA samples. The positions to which these fragments migrated, and their sizes in kbp, are indicated at the right-hand side of the figure. The other lanes contained RNA from transformants of strain GM-3C-2 carrying plasmid YEpl3CYCl(2.5) (Lane ATG), plasmid YEpl3CYClGT473 (Lane AGT), or plasmid YEpl3CYClC474 (Lane ATC).
described in detail in Chapter II and outlined in Figure 6. A CYCl promoter fragment completely lacking CYCl coding sequences was joined by means of a KpnI linker to the upstream boundary of the CYCl 3′ end signal at position +449. The C474 mutation was then introduced into the 3′ end signal region using oligonucleotide oAS4. Sequences flanking the downstream boundary of the 3′ end signal remained intact as far as the HindIII site at this stage. Deletions extending from the HindIII site toward the 3′ end signal were produced, and to each deletion endpoint, a synthetic oligonucleotide duplex was ligated. The duplex consisted of oligonucleotides oAS5 and oAS6 (Table I) and it had the following structure:

oAS5 5′ d(pTTAATAATGACTGG) 3′
oAS6 3′ (AATTATTACTGACCTTAAp)d 5′

The blunt end of the duplex was ligated to the resected 3′ end signal fragments described above. The other end of the duplex could be ligated to fragment termini generated by EcoRI cleavage, which allowed the fused promoter:3′ end signal:adapter fragments to be ligated into pEMBL9(+) as BamHI/EcoRI fragments. The deletion endpoints of several such fragments were determined by DNA sequencing. Figure 24 illustrates the deletion endpoints chosen for further study.

The first set of promoter:3′ end signal: lacZ fusion plasmids was constructed by joining the EcoRI termini of selected promoter:3′ end signal:adapter fragments to the EcoRI terminus of a fragment of placZ. The fragment included all but the first 19 bp of the lacZ gene. The resulting plasmids also carried a fragment of yeast DNA including the TRPl gene and chromosomal replication origin, ars1, and they were referred to as the pA5s. (The map of a pA5 is shown in Figure 6.)
Figure 24.

Deletion Endpoints in the pA4 Promoter:3' End Signal Fusions, for use in Promoter:3' End Signal: lacZ Plasmids.

The sequence shown extends from the KpnI linker marking the promoter:3' end signal junction, through the 3' end signal region to the ATG triplet of the oAS5/6 adapter. Sequences derived from either the KpnI linker or the oAS5/6 adapter are underlined. Sequences identical to that of the wild-type 3' end signal region (except for the C474 mutation) lie between two vertical lines.
Joining the promoter:3′ end signal:adapter and lacZ fragments in this manner placed the ATG triplet in oAS5/6 (underlined in the structure given above) in the correct reading frame to serve as the initiation codon for translation of lacZ, should an extended lacZ mRNA be produced. Nucleotides adjacent to the ATG triplet of oAS5/6 matched those flanking the normal initiation codon of CYC1. It was hoped that for this reason, any extended mRNA produced from the pA5s would be translated about as readily as CYC1 mRNA. Schweingruber et al. (1981) demonstrated that mutational relocation of the AUG initiation codon of CYC1 over a 37-nucleotide region had little effect on the expression of the gene. However, Kozak (1984a) has shown that nucleotides immediately flanking an AUG initiation codon can have some influence on the efficiency of initiation of translation at that codon.

The pA5 plasmids were used to transform yeast strain RPI23 to tryptophan prototrophy. No β-galactosidase activity could be detected in any of the transformants after growth on XGAL plates for up to two weeks, or after quantitative assays of permeabilized cell suspensions. Transformants of the same strain carrying plasmid pLG669Z (Guarente and Ptashne, 1981) provided a positive control for the β-galactosidase assays: such transformants turned blue after overnight growth on XGAL plates, and quantitative assays detected 30 units of β-galactosidase per ml per OD600 in exponentially growing cultures.

The complete absence of β-galactosidase from any of the pA5 transformants apparently had nothing to do with the presence of the CYC1 3′ end signal. Plasmid pA5.3 lacked the 3′ end signal completely, retaining only 2 bp of a Kpnl linker between the promoter and oAS5/6, and plasmid pA5.11 had only a 6 bp fragment of the 3′ end signal.
During growth in the absence of tryptophan, only 1-20% of the cells in a population derived from a pA5 transformant actually carried the plasmid. Plasmids with chromosomal replication origins frequently segregate with the mother cell during mitotic growth (Murray and Szostak, 1983) and as a result they are usually found in a relatively low fraction of the cells in a population, even during growth under selective conditions. Martinez-Arias and Casadaban (1983) reported that LEU2 promoter: lacZ fusions carried on ars plasmids directed the synthesis of only very low levels of 3-galactosidase, but that plasmids carrying 2µ circle sequences produced higher levels of 3-galactosidase, presumably because 2µ plasmids have higher average copy number.

I thought that modifying the pA5s so as to make them less susceptible to mitotic segregation might result in the production of detectable levels of 3-galactosidase, at least in transformants carrying those plasmids with inactive 3' end signals. Accordingly, I constructed two series of modified plasmids. Plasmids in the pA6 series differed from the corresponding pA5s only in carrying a 2.2 kb fragment from the centromere of yeast chromosomes III, CEN3. Circular plasmids carrying CEN3 and an ars sequence have previously been shown to be much more stable, both mitotically and meiotically, than similar plasmids lacking CEN3 (Clarke and Carbon, 1980; Fitzgerald-Hayes et al., 1982; Murray and Szostak, 1983). Plasmids in the pA7 series differed from the pA5s in carrying a fragment of the 2µ circle including its origin of replication and the REP3 region, which is needed in cis for plasmid maintenance at high copy number (Jayaram et al., 1983). Plasmids carrying the 2µ origin and REP3 sequence are
substantially more stable than plasmids with only an *ars* sequence because they show no mother-daughter segregation bias and because they replicate to high copy number provided that the yeast strain has endogenous 2μ circles (Murray and Szostak, 1983; Jayaram et al., 1983).

Plasmids pA6.3, pA6.11, pA7.3 and pA7.11 were all introduced into strain RP123, selecting for TRP1+ transformants. β-galactosidase was not detectable in any of the transformants after plating them on XGAL medium.

The *lacZ* gene of the pA5s, pA6s, and pA7s was from placZ (the construction of which was described in Chapter II), which was derived from pMCl403 by deleting lacYA. Plasmid placZ had initially been constructed to allow more convenient manipulation of the *lacZ* fragment and the fusion plasmids, but to test whether the deletion of lacYA had removed some sequence important for *lacZ* expression in yeast, I replaced the *lacZ* fragment of pA5.3 and pA5.11 with the lacZYA fragment of pMCl403. The resulting plasmids, pAl0.3 and pAl0.11, did not direct the synthesis of detectable levels of β-galactosidase in yeast.

**The pAl0 Series of Plasmids**

It seemed possible that perhaps the pA5 plasmids and their derivatives encoded an inactive "β-galactosidase" because of some peculiarity of the amino acid sequence encoded by the oAS5/6:*lacZ* junction. To test this idea, I replaced the *lacZ* fragment of the pA5s with a fragment from pLG669Z which carried 114 bp of the *laci* coding sequence fused to the *lacZ* coding sequence 64 bp downstream of the *lacz* initiation codon. This fragment is known to encode an active β-galactosidase when fused to the first 4 bp of the CYCl coding sequence (Guarente and Ptashne, 1981).
Plasmids carrying the lacI'Z fragment are referred to as pAlls, and their construction is described in Chapter II. A map of a pAll plasmid is shown in Figure 8. They also differed from the pA5s in that they carried a fragment of 2μ circle including the origin of replication and REP3 sequence. In exponentially growing cultures of pAll transformants of strain RP123, 75-80% of the cells carried the plasmid in the absence of tryptophan. Transformants of yeast strain RP123 carrying the pAll plasmids were plated on XGAL medium to test for β-galactosidase synthesis, and quantitative assays of β-galactosidase in suspensions of permeabilized cells from exponentially growing cultures were carried out. The results are shown in Figure 25. It is clear that the level of expression of lacZ on a pAll plasmid in yeast is affected by the extent of the CYCl 3' end signal fragment. Plasmids in which this fragment extends at least as far as position +496 direct the synthesis of very little β-galactosidase in yeast. Plasmid pAll.71 carries a 3' end signal fragment extending from position +449 to +493, but the first 3 bp of the oAS5/6 adapter are identical to sequences normally present at positions +494 to +496 of the CYCl 3' end signal. Plasmid pAll.50 carries a 3' end signal fragment 20 bp shorter than that of pAll.71, and it directs 13-fold higher levels of β-galactosidase synthesis in yeast than does pAll.71. Sequences which prevent lacZ expression in the pAlls therefore exhibit a functional boundary between positions +476 and +496 of the CYCl 3' end signal fragment.

This boundary corresponds very closely to the boundary of the 3' end signal as defined by the analysis of transcripts of CYClAH5' derivatives, which suggests that the functional element required to
**β-GALACTOSIDASE ACTIVITY IN YEAST CARRYING pAll PLASMIDS**

<table>
<thead>
<tr>
<th>pAll Plasmid</th>
<th>Kpnl Sites</th>
<th>EcoRI Sites</th>
<th>β-Gal Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>pAll.12</td>
<td>-24/+448</td>
<td>521</td>
<td>0.3 (0.1)</td>
</tr>
<tr>
<td>pAll.20</td>
<td>518</td>
<td></td>
<td>0.1 (0.1)</td>
</tr>
<tr>
<td>pAll.19</td>
<td>507</td>
<td></td>
<td>0.1 (0.1)</td>
</tr>
<tr>
<td>pAll.30</td>
<td>502</td>
<td></td>
<td>0.2 (0.1)</td>
</tr>
<tr>
<td>pAll.71</td>
<td>496</td>
<td></td>
<td>0.3 (0.2)</td>
</tr>
<tr>
<td>pAll.50</td>
<td>476</td>
<td></td>
<td>4.2 (1.1)</td>
</tr>
<tr>
<td>pAll.11</td>
<td>453</td>
<td></td>
<td>13 (2.6)</td>
</tr>
<tr>
<td>pAll.3</td>
<td>449</td>
<td></td>
<td>11 (2.1)</td>
</tr>
</tbody>
</table>

Figure 25. Level of β-Galactosidase Produced by GM-3C-2 Transformants Carrying pAll Plasmids

Units are defined in Chapter II. Each value given represents the mean of 6-10 independent measurements. The standard deviation from the mean is listed in parentheses. The 3' end signal region is cross-hatched, flanking sequences are shown as open boxes, and lacZ sequences are drawn as boxes with horizontal lines in them. The letters K and E denote recognition sites for restriction endonucleases KpnI and EcoRI, respectively.
prevent lacZ expression in the pAll plasmids is the CYCl 3′ end signal itself.

A further 3-fold increase in β-galactosidase activity is seen in cells bearing plasmid pAll.11, which retains only 6 bp of the 3′ end signal. This may indicate that the 3′ end signal fragment in pAll.50 remains partially active. Plasmid pAll.3 carries no 3′ end signal sequences, and only a 2 bp remnant of a KpnI linker separates the CYCl promoter from oAS5/6 sequences in this plasmid. It directs lacZ expression at the same level as pAll.11 in yeast.

What Might lacZ Expression from pAll Plasmids Measure?

Analysis of transcripts of deletion derivatives of the CYCl gene revealed two boundary regions of the CYCl 3′ end signal. Sequences within these boundary regions bring about a 40-fold reduction in the expression of lacZ from the CYCl promoter when inserted between gene and promoter. These results are most easily explained by supposing that sequences within the boundary regions act autonomously as a 3′ end signal, and when located upstream of lacZ, prevent the synthesis of lacZ transcripts.

A second interpretation of the results presented in Figure 25 is that sequences within the 3′ end signal region interfere with transcript initiation at the CYCl promoter when placed immediately adjacent to it. There is no evidence that an essential component of the CYCl promoter resides within the 5′ untranslated region of the gene, and it therefore seems unlikely that replacing the normal 5′ untranslated sequence with the 3′ end signal sequence would of itself inactivate the promoter. The 3′ end signal might interfere with the promoter as a consequence of its 3′ end signal activity. For instance,
3′ end signal activity might normally require the binding of particular proteins to the +449 to +496 region. Such binding might potentially interfere with transcript initiation at sites immediately upstream. In that case, although the observed reduction in \( \text{lacZ} \) expression would not result from 3′ end generation, it would remain a valid indication of at least one aspect of 3′ end signal function. Inactivation of the 3′ end signal could occur through disruption of the protein binding site(s), which would at the same time relieve interference with transcript initiation. Brent and Ptashne (1984) have proposed that a similar mechanism might account for the inhibitory effect of two different 3′ end signal regions on \( \text{CYCl} \) promoter activity when either of those regions is placed between the upstream activator site and transcriptional start sites.

A third interpretation of the levels of \( \beta \)-galactosidase observed in pAll yeast transformants is that \( \text{lacZ} \) mRNA is produced from all of the pAll plasmids, but that the mRNAs produced from plasmids with longer \( \text{CYCl} \) 3′ end signal fragments are less efficiently translated because of their necessarily longer leader sequences. This interpretation could also allow that certain sequences present in the \( \text{CYCl} \) 3′ end signal region interfere with translation of the hybrid mRNAs. (The absence of certain sequences which are normally present in the 5′ untranslated region could conceivably interfere with translation, but this effect could not contribute to differences in \( \text{lacZ} \) expression between transformants bearing different pAlls.) The coincidence of the downstream functional boundary of the \( \text{CYCl} \) 3′ end signal with that of the sequences required to prevent \( \text{lacZ} \) expression in a pAll plasmid would be viewed as unfortunate but meaningless. Studies of translation
in eucaryotes argue against the idea that sequences in mRNA leaders dramatically influence translation efficiency. The leader regions of normal CYCl mRNAs vary in length from 25 to 100 nucleotides. Assuming that the same 5′ end sites are used in transcripts initiated at the CYCl promoter of a pAll plasmid, plasmid pAll.11 would produce hybrid transcripts with leaders 6 nucleotides shorter than the corresponding CYCl+ transcripts, while the leader regions of transcripts from pAll.71 would be 34 nucleotides longer than CYCl+ transcripts. A 40-fold difference in lacZ expression accompanies this 40 nucleotide difference in mRNA leader length, a difference smaller than the range of leader lengths observed amongst CYCl+ mRNAs, and much smaller than the range observed amongst eukaryotic mRNAs in general (Kozak, 1983b; 1984b).

The 5′ untranslated regions in the CYCl:galK leader sequence fusions constructed by Zitomer et al. (1984) were 2–39 bp longer than those of the corresponding CYCl+ mRNAs. The maximum difference in galactokinase expression between yeast carrying different fusions was 2.6 fold.

Pelletier and Sonenberg (1985) reported that regions of extensive secondary structure within the 5′ untranslated sequences of a eukaryotic mRNA can severely inhibit its translation. The hybrid mRNAs encoded by the pAlls contain no notable regions of potential secondary structure, other than the KpnI linker octanucleotide at the promoter:3′ end signal junction and the EcoRI/BamHI/SmaI linkers at the oAS5/6:lacI junction. None of these could form a duplex of more than 4 bp. The KpnI linker sequence clearly does not limit translation efficiency, since it is absent from pAll.3 and present in pAll.11, yet both plasmids encode the same levels of β-galactosidase. The other linker sequences are present in all of the pAlls and should influence equally
the translation of mRNAs transcribed from any of them.

The idea that effects on promoter activity or mRNA translation contribute to the differences in lacZ expression between different pAll transformants cannot be excluded without a direct demonstration of 3′ end generation upstream of lacZ in pAll plasmids which do not produce detectable β-galactosidase. In an effort to examine the transcription of lacZ sequences in pAll plasmids directly, RNA isolated from various pAll transformants of strain RP123 was electrophoresed in agarose, transferred to nitrocellulose, and hybridized to a lacZ probe. No discrete RNA species hybridized to the lacZ probe. When hybridization was observed, it appeared as a diffuse pattern spanning a wide range of RNA sizes. The pattern was not due to general degradation of the RNA samples, because discrete bands of ribosomal RNA were visible upon staining samples which had been electrophoresed with those transferred to nitrocellulose. Discrete bands of hybridization to a TRPl probe were observed after transfer of RNA from a duplicate gel, run together with the first, to nitrocellulose.

Saturation Mutagenesis of a CYCl 3′ End Signal Fragment

While I recognized that the phenotype conferred by any given pAll plasmid may have reflected effects on promoter activity or translation efficiency, it seemed that the process of 3′ end generation could most simply and completely account for the dependence of β-galactosidase production on the extent of the CYCl 3′ end signal fragment. Mutations were introduced into the 3′ end signal fragment carried by plasmid pAll.71 with the idea that mutations which allowed increased lacZ expression might identify sequences important for 3′ end generation. The plasmid pAll.71 was chosen because it carried the smallest 3′ end
signal fragment which still maximally suppressed $\beta$-galactosidase synthesis.

The goal of this approach to studying the CYC1 3' end signal was to produce every possible point alteration of a functional 3' end signal fragment and identify those mutations which impaired 3' end generation. Screening a large number of mutants for increased $\beta$-galactosidase production is certainly feasible because of the ease and speed of the assay.

The object of saturating the 3' end signal region with point mutations imposed several requirements on the mutagenic procedure: 1) it should be reasonably efficient; 2) it should be targeted to the region of interest; 3) it should produce mutations at random throughout that region; 4) it should produce any type of base substitution.

A plethora of methods of in vitro mutagenesis, allowing the efficient production of mutations in target regions of varying size, have now been described. (The reader is directed to the excellent reviews of Shortle et al.,1981, Smith, 1985, and Botstein and Shortle, 1985). I should like to describe briefly the type of consideration that guided my choice of mutagenic method for the 3' end signal of CYC1. Linker-scanning mutagenesis provides a systematic approach to identifying functionally important sequence elements within a region of interest (McKnight and Kingsbury, 1982). This method, however, represents an intermediate step between defining the region of interest and defining important nucleotides within it, and I hoped that the simple screening device of elevated lacZ expression would allow such a step to be omitted. Bisulfite mutagenesis has been used to produce mutations throughout defined segments of DNA, but it causes only GC-AT
transitions (Shortle et al., 1980; Ciampi et al., 1982; Weiher and Schaller, 1982; Folk and Hofstetter, 1983). It would be particularly unsuitable for the CYCl 3' end signal region, which has a GC content of only 20% (Figure 23). The mutagenic base analogue N4-hydroxymethylcytosine did not meet the required criteria because it induces only transitions (Muller et al., 1978).

Oligonucleotide-directed mutagenesis is capable of efficiently producing any desired base substitution at a target site specified by the sequence of the oligonucleotide (Smith and Gillam, 1981; Zoller and Smith, 1982; 1983). The method has very recently proven adaptable to the aim of saturating a target region with point mutations. Kalderon et al. (1984) showed that a mixture of homologous nucleotides could mutagenize a defined subset of the sites within a target region. McNeil and Smith (1985) and Wells et al. (1985) have devised an elegant extension of this method which allows the efficient production of any desired spectrum of point mutations within a target region of up to at least 50 bp. However, when I was designing my experiments on the 3' end signal of CYCl, I thought of oligonucleotides as strict site-specific mutagens and therefore turned to other approaches.

The approach taken in this study was based on nucleotide misincorporation. Several groups had demonstrated that nucleotide misincorporation during in vitro copying of DNA could be used to produce almost all types of point mutations. Conditions which had been used in various procedures to increase the frequency of misincorporation included the use of Mn2+ rather than Mg2+ as an activator (Kunkel and Loeb, 1979), the use of non-proofreading DNA polymerases (Zakour and Loeb, 1982), the omission of one or more
nucleotides from the copying reaction (infinite pool bias), and the use of excision-resistant α-thionucleotides (Kunkel et al., 1981; Shortle et al., 1982).

If a defined primer-template complex is incubated with a single nucleotide and DNA polymerase (an α-thionucleotide if the enzyme has a proofreading exonuclease activity), then the site immediately downstream of the 3' end of the primer becomes a target for efficient mutagenesis (Shortle et al., 1982). Misincorporation of one nucleotide prevents misincorporation at the next site. Therefore, synthesis stops until the other nucleotides are added. It should be possible in principle to introduce any mutation at the site by carrying out four reactions, each with a different nucleotide.

To subject a target region to misincorporation mutagenesis, it should simply be necessary first to produce a population of primers with 3' ends at every position in the region. The approach used in this study is illustrated in Figure 13 and described in detail in Chapter II. It was necessary to obtain the target region in single-stranded form for the in vitro copying reaction, and this was conveniently done by cloning a fragment including the target into M13. The small SmaI fragment of pAll.71, which included the CYCl 3' end signal and part of the CYCl promoter, was ligated into M13mpl0 such that the mRNA-isoparallel strand of the insert was linked to the + strand of M13. Single-stranded DNA of the recombinant phage, which was called mpl0Al, was annealed with the complementary strand of SmaI-linearized M13mpl0 replicative form DNA, to produce a gapped heteroduplex in which the target for mutagenesis (the 3' end signal region) lay at the 3' end of the single-stranded region. Figure 26
illustrates the results of the annealing reaction. Conditions for limited primer extension which allowed the production of random primer termini over a region of up to 150 nucleotides had been described by Brown and Smith (1977). Similar conditions were used to extend the 3' end of the mpl0Al::mpl0/SmaI gapped heteroduplex into the single stranded region for 20-70 nucleotides, as illustrated in Figure 27. The product of this step was a population of gapped heteroduplexes with 3' ends opposite essentially every position of the CYCl 3' end signal. Figure 27 shows that the distribution of 3' ends was not entirely random, 3' ends at certain positions being considerably more abundant than at others. Every position of the 3' end signal appeared to be targeted for mutagenesis, but it was expected that mutations would be recovered more frequently at some positions than at others.

A different approach to targeting misincorporation mutagenesis throughout a defined region was described while this work was in progress. Abarzua and Marians (1984) annealed circular, single-stranded DNA carrying the target region with a population of complementary linear fragments which collectively had 3' ends at all positions throughout the target. These fragments had been produced by exonucleolytic digestion of a fragment carrying the complete target region. The population of gapped heteroduplexes so produced was formally equivalent to the population produced by limited primer extension in this study. As observed in the present study, some primer termini occurred more frequently than others, but the authors were nevertheless able to recover mutants at over half of the positions in the target region.

Another approach to targeting, which was originally applied to
Figure 26. Annealing of mplOAl and mpl0/SmaI

A. Map showing the structure of the gapped heteroduplex formed by annealing mplOAl and SmaI-cut mplORF. The positions of recognition sites for ClaI are indicated by "C". The approximate sizes, in kbp, of the fragments expected from ClaI digestion of the linear RF and of the gapped heteroduplex are indicated.

B. Substrates and products of the annealing reaction.
Two samples of each of the substrates and of the products of the annealing reaction were taken: One sample of each pair was digested with restriction endonuclease ClaI. All samples were then electrophoresed through a 0.8% agarose gel. The gel was stained with ethidium bromide (1 μg/ml in water) and photographed under UV illumination. In each pair of lanes, the right-hand lane shows the products of ClaI digestion of the DNA in the left-hand lane. Lanes 1, mplOAl single-stranded DNA; Lanes 2, SmaI-cut mplORF; Lanes 3, mplOAl and SmaI-cut mplORF after mixing and annealing as described in Chapter II. Annealing is evidenced by (1) the appearance in lane 3 (left) of a band of lower mobility than the linear mplORF, and (2) the appearance of a new fragment in the ClaI digest of the products of the annealing reaction, lane 3 (right). The size of this fragment is about 4.8 kbp, as predicted in A.
Figure 27. Limited Primer Extension on a Gapped Heteroduplex
Lanes 1 and 2. Primer extension was carried out as described in Chapter II. An aliquot of the products was digested with BamHI prior to electrophoresis in a 7% acrylamide gel (lane 1). The remainder was used in a misincorporation reaction, after which a second aliquot was removed, digested with BamHI, and electrophoresed (lane 2).
Lanes 3 and 4. As in lanes 1 and 2, respectively, except that dATP[αS] was included in the primer extension reaction at a concentration of 15 μM.
Lanes M. ddGTP chain terminator sequencing reaction of mpl0Al ssDNA with oligonucleotide FPl as primer. Fragments produced in the sequencing reactions provided size markers. Fragment sizes in bp are indicated to the left of each panel.
-230-
bisulfite mutagenesis (Shortle et al., 1980), involves annealing the target region in a supercoiled duplex molecule to a complementary fragment to produce D-loops which can be randomly nicked by SI nuclease. Each nick can be extended into a short gap by exonucleolytic digestion to produce a population of randomly gapped heteroduplexes.

The gapped heteroduplex population produced by limited primer extension was incubated with DNA pol I (Klenow fragment) and a single α-thiodeoxynucleoside triphosphate. It was expected that during this incubation, the α-thionucleotide would be randomly misincorporated onto the available primer termini. The fraction of primers terminated by a misincorporated nucleotide was expected to increase with time because α-thionucleotides cannot be excised by the proofreading exonuclease of DNA polI (Kunkel et al., 1981).

The fourth step in the procedure involved filling in the remainder of each gap with Klenow fragment and normal deoxynucleoside triphosphates and then ligating to produce closed duplex circular molecules. DNA purified from this reaction was introduced into E. coli JM101. The majority of the plaques obtained were colourless, suggesting that they were derived from repaired heteroduplexes or from mpl0Al DNA "leftover" after the annealing reaction. (Leftover mp10 RF DNA should have produced blue plaques after recircularization and introduction into cells.) To assess the efficiency of the mutagenic procedure, single-stranded DNA was prepared from several plaques and subjected to single-track sequencing using the dideoxynucleotide corresponding to the mutagenic α-thionucleotide. Mutations were expected to be visible as extra bands in the single-track sequences as compared to that of the parent phage. An example of the results is

-231-
shown in Figure 28. Essentially all clones examined contained the pAll.71 fragment, confirming that they were derived from either heteroduplexes or mplOAl DNA. Of 100 clones screened in three experiments, 15 carried mutations, as summarized in Table VI. Twelve of the mutations recovered were within the target region. Those outside it were in adjacent vector or CYCI promoter sequences, as if some primer termini had been located in these regions.

Comments on the Mutagenic Procedure

Targeting of Mutations

Two problems limit the accuracy of targeting using the procedure described here: limited primer extension cannot produce primer termini which are precisely and completely confined to a target region within a larger single-stranded region, and redistribution of the primer termini by the proofreading exonuclease of DNA pol I would be expected to occur prior to misincorporation, producing some primer termini upstream of the target. Primer end redistribution did in fact occur during the misincorporation reaction, as shown in Figure 27.

The first problem could be overcome by restricting the extent of the single-stranded region within the gapped heteroduplex to the target itself. All primer termini initially produced would necessarily be in the target region. Several ways of preventing primer redistribution suggest themselves. The method of including an α-thionucleotide in the limited primer extension reaction seems fairly effective (Figure 27) but not ideal because it increases the chance of introducing mutations during limited primer extension and because it doesn’t fix the positions of all primer termini, but rather provides a series of "stops" to block their exonucleolytic digestion. A more effective
Figure 28. Single-Track (T) Sequences of 12 Clones Obtained in Experiment T1 (Table VI). Oligonucleotide FPI served as the primer for sequencing by the chain termination method of Sanger et al. (1977b). The sequence of the parental clone, mp10Al, is at the left. The extent of the 3' end signal region and the adapter region is indicated. T sequences of 12 clones occupy the remainder of the autoradiograph. Clones with differences from the parental sequence are numbered.
TABLE VI

RESULTS OF MISINCORPORATION MUTAGENESIS EXPERIMENTS

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Clones Screened</th>
<th>Mutants Screened</th>
<th>Name of Mutant</th>
<th>Position of Mutation</th>
<th>Mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gl</td>
<td>24</td>
<td>5</td>
<td>Gl.2</td>
<td>+473</td>
<td>T to C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-70</td>
<td>T to Gd</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Gl.5</td>
<td>M13c</td>
<td>T to C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Gl.10</td>
<td>oAS5;5</td>
<td>T to C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+478</td>
<td>G to Td</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Gl.13</td>
<td>+453 - +455</td>
<td>TAT to CCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Gl.16</td>
<td>+487</td>
<td>A to C</td>
</tr>
<tr>
<td>G3</td>
<td>36</td>
<td>3</td>
<td>G3.7</td>
<td>+496</td>
<td>A to C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>G3.13</td>
<td>+472</td>
<td>A to C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>G3.32</td>
<td>+487</td>
<td>A to C</td>
</tr>
<tr>
<td>Tl</td>
<td>36</td>
<td>7</td>
<td>Tl.17</td>
<td>-38</td>
<td>Ad</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tl.18</td>
<td>+495</td>
<td>T to A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tl.26</td>
<td>-30</td>
<td>G to A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tl.30</td>
<td>+473</td>
<td>T to A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tl.34</td>
<td>+481</td>
<td>T to A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tl.36</td>
<td>+465</td>
<td>T to A</td>
</tr>
</tbody>
</table>

a Positions in CYCl sequences are numbered with respect to the sequence of the intact gene, taking the first base of the coding sequence as +1. Positions upstream are denoted by a minus sign. Mutations in the oAS5/6 adapter are indicated by oAS5, followed by the position of the alteration.

b Sequence changes are shown as they affect the mRNA-parallel strand of CYCl. The original base is shown on the left, the one inserted by mutation on the right.

c The mutation was located in vector sequences upstream of the target.

d These mutations could not have resulted from misincorporation of the \( \alpha \)-thionucleotide provided.
method might be to use a non-proofreading DNA polymerase. Shortle and
Lin (1985) have found that primer terminus redistribution by DNA pol I
can be prevented by including an excess of α-thionucleotide over Mn\(^{2+}\)
in the misincorporation reaction. The proofreading exonuclease is not
active under these conditions because all available Mn\(^{2+}\) is chelated
by the nucleotide, and none is available to activate the nuclease.

Primer end redistribution, as well as reducing the accuracy of
targeting, might create "hot spots" for misincorporation.
Exonucleolytic digestion of any given primer would be expected to
continue until the 3' terminal nucleotide of the resected primer was
the same type as the α-thionucleotide used for misincorporation.
Replacement of the 3' terminal nucleotide with the α-thionucleotide
could then occur readily and produce a stable primer terminus. Primers
correctly terminated in the α-thionucleotide would therefore accumulate
over time, and positions immediately downstream of these abundant
primers might become hotspots for misincorporation. A possible example
of such a hotspot is evident from the data in Table VI. Two
independent isolates of an A/T:G/C transversion at position +487 were
recovered in experiments using dGTP[αS] as the mutagenic nucleotide.
(Three more isolates were recovered upon screening more clones from
experiment G3; C. Beard, personal communication.) The position
immediately upstream of +487 on the strand used for priming is position
+488, and the nucleotide normally present on the bottom strand at that
position is G. Stable primer termini at position +488 could have
accumulated, leading to frequent misincorporation of G at position
+487. Rendering all primer termini equally stable using one of the
methods described above would prevent this type of hotspot from arising
(Shortle and Lin, 1985, also mention the contribution of primer end redistribution to mutagenic hotspots.)

Accurate targeting of mutations by this procedure requires not only that limited primer extension produce primer termini confined to the target region, but also that the limited primer extension and the final, gap-repair steps proceed with high fidelity. Optimum fidelity during in vitro DNA synthesis with DNA pol I is achieved by using relatively low (approximately 20 \( \mu \)M) equal concentrations of all four deoxynucleoside triphosphates (Shi and Fersht, 1984). The limited primer extension step was carried out with dCTP, dGTP, dTTP at a concentration of 35 \( \mu \)M each and dATP at about 3 \( \mu \)M. A 10-fold bias against dATP might be expected to reduce fidelity somewhat, but the concentrations of the other nucleotides were sufficiently low as to allow proofreading by the 3' exonuclease of DNA pol I. The inclusion of dATP[S] during the limited primer extension, as in Experiment G3, would prevent proofreading of misincorporated A residues, but the 10-fold excess of other nucleotides would not favour misincorporation of A in the first place. The gap repair step was carried out in the presence of 100 \( \mu \)M of each deoxynucleoside triphosphate. It is possible that reducing this concentration would guard against misincorporation by allowing more effective proofreading. The gap repair reaction included Mn\(^{2+}\), as well as Mg\(^{2+}\), to improve the efficiency of extension of mispaired 3' ends. Its presence might lead to misincorporation during the gap repair reaction.

Efficiency of Mutagenesis

From the results presented in Table VI the efficiency of the mutagenic procedure is estimated to be about 15\% (or about 0.2\% per base pair over a region of about 80 base pairs). This is high enough
to make the identification of mutants by single-track sequencing feasible in the absence of any phenotypic screening device. A number of factors might limit the efficiency of mutagenesis. Transfection by leftover single-stranded DNA of mpl0Al would have increased the "background" of unaltered mpl0Al clones. It is unlikely that this had a serious effect on the efficiency of mutagenesis, since in one experiment (T1; Table VI) the products of the gap repair reaction were treated with Sl nuclease prior to being used in transfection. The conditions of Sl treatment had previously been shown to allow complete degradation of single-stranded DNA, in the amount present in the annealing reaction, without allowing degradation of duplex DNA. Although the small sample sizes compared prevent accurate statistical comparisons, the number of mutants recovered in this experiment was not markedly different from that recovered in either of the other experiments.

The efficiency of the misincorporation step itself may have limited the efficiency of mutagenesis, although Shortle et al. (1982) and Abarzua and Marians (1984) found that misincorporation could occur at frequencies higher than 40%. The presence of nucleotides other than one α-thionucleotide during the misincorporation step would reduce the efficiency of misincorporation. Such contaminants might be carried over from the limited primer extension step or introduced with the α-thionucleotide.

Misincorporation, followed by gap repair and ligation, would result in a heteroduplex containing a mismatched base pair. Mismatch repair following transfection of E. coli would reduce the efficiency of mutagenesis if the repair process was biased toward using the
"wild-type" strand of the heteroduplex as a template for correction of the "mutant" sequence. Abarzua and Marians guarded against the danger of mismatch repair by selectively replicating the "mutant" strand of each heteroduplex in vitro. One reservation which may be held about their procedure is that it is quite involved and requires at least one protein which is not readily available, the gene A protein of ϕX174.

Perhaps the most elegant way of avoiding mismatch repair of mutant sequences is that described recently by Kunkel (1985). As applied to misincorporation mutagenesis of the 3' end signal fragment of mpl0Al, the approach would involve preparing uridine-containing mpl0Al single-stranded DNA from a dut−ung− host. This DNA would be used in annealing, misincorporation and gap-repair reactions as described, and the products would be transfected into an ung+ host. The presence of U in the wild-type strand of each heteroduplex would direct repair processes to that strand. Alternatively, the template strand could be degraded prior to transfection by treatment with dUTPase, followed by alkali (Kunkel, 1985).

Types of Mutations Induced

A disturbing feature of one experiment in which dGTP[αS] was used for misincorporation was that three of five mutants contained multiple sequence changes. In each of two clones, Gl.2 and Gl.10, two mutations were present at nearby non-adjacent positions. Three of the four mutations in these clones were of the type expected from misincorporation of dGMP, while the fourth may have resulted from the misincorporation of dAMP. The continuation of DNA synthesis after one misincorporation event implies that nucleotides other than the one misincorporated were available. Other nucleotides may have been -238-
present in the dGTP[αS], or they may have been carried over from the limited primer extension reaction. According to Botstein and Shortle (1985), error-prone DNA synthesis in the presence of more than one nucleoside triphosphate leads to a high frequency of clustered multiple mutations.

The third multiple mutant recovered in the same experiment had the sequence CCCCCC instead of CCCTAT at positions +450-+455. It seems likely that the mutation resulted from some interaction involving the neighbouring sequence, but the nature of this interaction is obscure.

No multiple mutants were observed in the other misincorporation experiments, but the apparent difference between experiments may be entirely due to the small numbers of mutants in each.

The entire region subjected to the mutagenic procedure, or the portion of it subcloned for further analysis, should be sequenced to check for mutations outside the actual target region. In the pAlls reconstructed from mutants of mpl0A1, about 250 bp of the CYCl promoter, between the SmaI site at position -380 and the oAS7 priming site at -125, was not sequenced. As will be described, independent tests of the function of the mutant 3’ end signals were later carried out using fragments which had been completely sequenced.

Effects of Mutations in the 3’ End Signal Fragment on lacZ Expression from pAll Plasmids

Eight derivatives of plasmid pAll.71 carrying point mutations in the CYCl 3’ end signal were produced from mutants of mpl0A1 as described in Chapter II. The CYCl 3’ end signal and flanking regions of each plasmid were sequenced in order to confirm the identity of the point mutation(s) in the 3’ end signal and the structure of the
Plasmids carrying mutant derivatives of the CYC1 3′ end signal fragment of pAll.71 were introduced into yeast strain RPL23, and the β-galactosidase levels in the TRP1+ transformed strains were measured. The results are presented in Figure 30. The variability between independent assays of the same type of transformant was greater in several cases than had been observed in earlier experiments, and what this means is not clear. It is clear, however, that six of the eight plasmids tested directed β-galactosidase production at levels 4-15-fold higher than did the parent plasmid, pAll.71. The remaining two seemed to support even lower levels of lacZ expression than pAll.71.

The results presented in Figure 30 suggest that all of the mutations tested except Tl.30 and Tl.34 impaired the activity of the CYC1 3′ end signal and caused increased synthesis of functional lacZ transcripts. To argue against the involvement of 3′ end generation in determining the level of lacZ expression from the pAll plasmids, it would be necessary to suppose that six of eight mutations tested caused increased lacZ expression by quite dramatically improving the stability or translational efficiency of lacZ transcripts. Either interpretation would imply a strong dependence of the translational yield of an mRNA on the sequence of its leader region. As discussed previously, there is little support in the literature for the notion of such dependence, except with respect to AUG triplets and secondary structure.

3′ End Signal Function in Truncated CYC1 Genes

In an attempt to confirm that increased lacZ expression from pAll plasmids is indicative of 3′ end signal dysfunction, a fragment of
Plasmids in the pAll series were prepared for sequencing as described in Chapter II. They were digested with restriction endonucleases XhoI and SacI to release a 2.1 kbp fragment carrying part of the CYCl promoter, the 3' end signal region, oAS5/6 adapter, and part of laci'Z. The digestion products were denatured and annealed to oligonucleotide oAS7 (Table I), which served as a primer for sequencing the mRNA-parallel strand of the 3' end signal region by the chain termination method of Sanger et al. (1977b). Products of the sequencing reactions were denatured and fractionated on 6% acrylamide/7M urea gels, which were then autoradiographed. The panel at the left illustrates the sequence of part of the 3' end signal region carried on plasmid pAll.71. The lanes in this and all other panels contain, from left to right, products of the C, T, A, and G sequencing reactions. Other panels illustrate the sequences of 3' end signal regions carrying the point mutations named at the bottom of each panel. Any base which differs from the pAll.71 sequence is labelled by a black dot to the left of the corresponding band in the autoradiograph. To the left or right of each panel, the sequence in the vicinity of the mutation is written out, with the site of the mutation again identified by a black dot to the left of the altered base.
Figure 30. Levels of β-Galactosidase Activity in Yeast Transformants Carrying pAll Plasmids with 3' End Signal Mutations

Transformants of strain RP123 carrying pAll plasmids were grown in 10 ml cultures of selective medium and assayed for β-galactosidase activity as described in Chapter II. The name of the plasmid carried by a given transformed strain is indicated in the left-hand column. The centre of the figure shows the sequence of the CYCl 3' end signal region in each plasmid. The complete sequence of the CYCl 3' end signal region of plasmid pAll.71 is written out. The vertical line marks the distal endpoint of sequences derived from the 3' end signal. Sequences to the right of this line were derived from the oAS5/6 adapter. For each of the other plasmids, only those bases which differed from the sequence of the CYCl 3' end signal in pAll.71 are indicated. The horizontal lines indicate positions of sequence identity between pAll.71 and the other plasmids. The right-hand column lists the β-galactosidase activity in each transformed strain. Units are defined in Chapter II. Each number listed is the mean of measurements made on 4 independent cultures.
<table>
<thead>
<tr>
<th>PLASMID</th>
<th>SEQUENCE OF 3' END SIGNAL</th>
<th>β-GALACTOSIDASE ACTIVITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>pAll-71</td>
<td>GTCCCTATTTTATTTTTATAGTTATCTTAGTATTAAGAACGTTATTTATAATGACT</td>
<td>0.8</td>
</tr>
<tr>
<td>pAll-G12</td>
<td></td>
<td>3.9</td>
</tr>
<tr>
<td>pAll-G10</td>
<td></td>
<td>11.5</td>
</tr>
<tr>
<td>pAll-G16</td>
<td></td>
<td>3.2</td>
</tr>
<tr>
<td>pAll-Ti30</td>
<td></td>
<td>0.1</td>
</tr>
<tr>
<td>pAll-Ti34</td>
<td></td>
<td>0.2</td>
</tr>
<tr>
<td>pAll-Ti36</td>
<td></td>
<td>7.4</td>
</tr>
<tr>
<td>pAll-G37</td>
<td></td>
<td>3.5</td>
</tr>
<tr>
<td>pAll-G313</td>
<td></td>
<td>3.8</td>
</tr>
</tbody>
</table>
pAll.71 was joined to the first half of the CYC1 gene and the truncated CYC1 gene so produced was assayed for the production of truncated CYC1 transcripts. Truncated genes were also produced using the 3' end signal fragment of each mutant derivative of pAll.71. The structure of the truncated genes is diagrammed in Figure 11, and their construction is described in detail in Chapter II. The 355 bp KpnI/HindIII fragment of the normal 2.5 kb BamHI/HindIII CYC1+ fragment was replaced with a 70 bp fragment extending from the KpnI site of mpl0A1 or one of its derivatives to the HindIII site immediately downstream of the 3' end signal region. The truncated CYC1 gene was transferred to YEpl3 on a 2 kb BamHI fragment, and the resulting plasmids, referred to as pA12As or pA12Bs, depending on the orientation of the insert, were introduced into yeast strain GM-3C-2. Total RNA was isolated from the transformants, electrophoresed through agarose, transferred to nitrocellulose, and tested for hybridization to a CYC1 probe. Plasmids carrying a functional CYC1 3' end signal should produce an RNA of about 450 nucleotides which hybridizes to the probe. As shown in Figure 31, such a transcript is produced from plasmids carrying the "wild-type" 3' end signal or the Tl-30 or Tl-34 mutant 3' end signal. These are exactly the 3' end signal fragments which prevent lacZ expression in pAll plasmids. The position of the truncated transcript is indicated by the "C" in the margin of the figure. The level of the truncated transcript is greatly reduced in cells bearing plasmids with those mutant 3' end signals which allow elevated lacZ expression from pAll plasmids.

Whether any transcript other than the truncated one is produced from the truncated CYC1 genes depends upon the orientation of the 2 kb BamHI fragment (C. Beard, personal communication). If the fragment is
Figure 31. Hybridization of a CYCl Probe to RNA from Yeast Transformants Carrying pA12A and pA12B Plasmids

A. Total RNA from transformants of strain GM-3C-2 carrying pA12A or pA12B plasmids was denatured with formaldehyde and electrophoresed through 1% agarose. The RNA was then transferred to a nitrocellulose filter, which was subsequently incubated with a $^{32}$P-labelled CYCl hybridization probe, washed and autoradiographed. The probe was generated by copying the ss DNA of mp8CYCl(2.5) using oligonucleotide FPl (Table I) as a primer, the Klenow fragment of DNA polymerase I, $\alpha$[$^{32}$P]dATP, dCTP, dGTP, and dTTP. Each lane contained 20 $\mu$g of total RNA. The RNA was isolated from transformants carrying the plasmids indicated at the top of each lane. The transcripts produced by 3' end generation in the vicinity of the CYCl 3' end signal fragment are indicated by the line labelled C. Transcripts produced by 3' end generation further downstream in 2$\mu$ circle sequences are indicated by the line labelled RT. Transcripts from the wild-type CYCl gene carried on plasmid YEpl3CYCl(2.5) and the deletion derivative carried on plasmid 4H40 (see Figure 16) provided size markers (of approximately 650 and 1100 nucleotides, respectively). The positions to which they migrated are indicated by the lines labelled WT and 4H40.

B. Total RNA from transformants of strain GM-3C-2 carrying pA12B plasmids was denatured with formaldehyde and electrophoresed through 1% agarose. The RNA was transferred to a nitrocellulose filter, which was then incubated with a $^{32}$P-labelled CYCl hybridization probe, washed and autoradiographed. The probe was prepared as described in A. Each lane contained 20 $\mu$g of RNA. The lane marked 71 contained RNA from a transformant of GM-3C-2 carrying pA12B.71. Other lanes contained RNA from transformants carrying pA12B plasmids with the mutations indicated at the top of each lane. Transcripts produced by 3' end generation in the vicinity of the 3' end signal fragment are marked by the arrow labelled C. Those produced by 3' end generation in 2$\mu$ circle sequences are marked by the arrow labelled RT. The lines labelled WT and 4H40 indicate the positions to which transcripts from the CYCl genes borne on plasmids YEpl3CYCl(2.5) and 4H40 migrated.
oriented so that transcription of the **CYCl** gene proceeds away from 2μ circle sequences, as in the pAl2A series, then no discrete transcript other than the 400 nucleotide one is detected by the **CYCl** probe. If the truncated transcript could not be produced because of a defect in the **CYCl** 3′ end signal, transcription would be expected to proceed into pBR322 sequences. The presence of another yeast 3′ end signal in these sequences would be purely fortuitous, and in the absence of such a signal, 3′ end generation could not occur. The extended transcript might therefore be unstable and escape detection. Several studies have suggested that disruption of the only 3′ end signal downstream of a gene prevents the synthesis of a stable transcript of that gene (Higgs et al., 1983; Fitzgerald and Shenk, 1981; McDevitt et al., 1984).

In pAl2B plasmids containing the 2 kb fragment in the opposite orientation, transcription would proceed from the **CYCl** gene through 375 bp of pBR322 and into 2μ circle sequences if the **CYCl** 3′ end signal was defective. Termination or processing would be expected to occur about 500 bp distal to the junction with 2μ sequences to produce an extended transcript of about 1,300 nucleotides. Such a transcript is in fact observed in cells bearing any of the pAl2B plasmids. In cells bearing pAl2B.T1, pAl2B.T1.30, or pAl2B.T1.34, the truncated, 400 nucleotide transcript is also observed, in levels equaling or exceeding those of the extended transcript.

The production of truncated transcripts from the truncated **CYCl** gene is indicative of 3′ end signal function, and regardless of the orientation of the gene with respect to the vector, the synthesis of truncated transcripts correlates with the ability of the 3′ end signal to prevent **lacZ** expression in pAll plasmids. This observation supports
the idea that increased lacZ expression in pAll transformants is indicative of 3' end signal dysfunction.

The production of extended CYCl transcripts from all of the pAl2B plasmids implies that the 50 bp 3' end signal fragment, while retaining the ability to cause 3' end generation, is not completely autonomous. Some sequence outside this fragment is necessary to allow efficient 3' end generation. It has already been noted that sequences on either side of the boundary region have some influence on 3' end generation. The results obtained with the pAl2s suggest that simultaneous removal of the normal sequences flanking both boundaries causes a more dramatic drop in 3' end signal efficiency than the removal of sequences flanking only one boundary. Similar observations were made by Henikoff and Cohen (1984), who defined the functional boundaries of a yeast 3' end signal near a Drosophila gene segment but found that the 21 bp region within the boundaries could cause 3' end generation only when certain sequences were present downstream, and then only inefficiently. The question of what additional sequences must flank the 50 bp CYCl 3' end signal fragment in order for it to function efficiently might be addressed using the plasmid pAl2B.71. Sequences to be tested could be inserted at either end of the 50 bp 3' end signal fragment, and those which allow more efficient 3' end generation would be expected to prevent the synthesis of the extended CYCl transcript in yeast.

The production of extended transcripts from one series of pAl2s, but not the series in which the orientation of the truncated gene is reversed, may simply be due to a difference in the stability of the extended transcript produced in each instance. What is in some sense surprising is that although the extended transcripts from pAl2B
plasmids accumulate to the same level as the truncated transcripts, a high "background" of lacZ expression from pAll plasmids carrying functional 3′ end signal fragments is not observed. Two explanations suggest themselves. One is that the context of the 3′ end signal fragment in the pAlls allows it to function more efficiently than in the pAl2s. The other is that extended transcripts are in fact produced as a relatively low proportion of total transcripts pAl2B plasmids carrying the "wild-type" 3′ end signal, but that the extended transcripts are more stable than the truncated transcripts and are therefore "enriched" in steady state mRNA. It would be necessary to suppose that a similar enrichment for the extended lacZ transcripts produced from pAll plasmids does not occur, but since no known yeast 3′ end signal is present downstream of lacZ, such enrichment would not be expected.

The utility of the pAll plasmids in screening for 3′ end signal mutations should not be taken to imply that β-galactosidase levels in pAll-transformed cells provide a direct quantitative measure of the efficiency of a given 3′ end signal fragment. The level of β-galactosidase in cells carrying a pAll plasmid such as pAll.71, with an intact 3′ end signal, may be taken as a basal level. For reasons already discussed, it seems fair to say that inactivation of the CYC1 3′ end signal is prerequisite to the synthesis of β-galactosidase in excess of the basal level. However, if two pAll plasmids express lacZ at different levels above the basal level, it should be remembered that other factors, such as differences in the translation efficiency or stability of the lacZ mRNAs might contribute to the difference in lacZ expression.
The apparent requirement for sequences outside the functional boundaries of the CYC1 3′ end signal complicates the task of identifying those sequences within the boundaries which are important in 3′ end generation. Mutations within the 3′ end signal region of the intact CYC1+ gene would be expected to inactivate the signal if they alter a unique sequence which must be specifically recognized in some way during transcript 3′ end generation. Mutations affecting a repeated recognition sequence, or altering nucleotides which contribute to some general property of the signal, might be expected to have more moderate effects on its activity. When the 3′ end signal region is "trimmed down" to 50 bp, "general properties" of the region must be specified by fewer nucleotides, and extra copies of repeated recognition sequences may be lost in the trimming. The result is that the trimmed 3′ end signal, located outside its normal context, may be more dramatically affected by certain mutations than it would be if it were in its usual position within the CYC1 transcription unit.

The data presented here do not allow the sequence requirements of the 3′ end signal to be completely specified, but they do suggest certain of its features and they allow "terminator sequences" recognized on the basis of sequence homology to be assessed on the basis of functional importance.

Zaret and Sherman (1982) noticed that the sequence TAG...TAGT...TTT, which occurs in the 38 bp region deleted from the cycl-512 allele, is also present in the sequences flanking the 3′ ends of a number of other yeast genes. They proposed that this sequence might be a required component of at least some yeast transcription
terminators. At the same time they suggested that some other property of the 3′ flanking region of a gene, such as high AT content or a disproportionate thymidine content in the mRNA-parallel strand, might be important to 3′ end generation. A 3′ end signal sequence should be present near sites at which 3′ end generation is known to occur, but it is equally important that its essential features be absent from sites at which 3′ end generation does not occur. The consensus sequence of Zaret and Sherman (1982) meets the first of these criteria in several cases, but it does not seem to meet the second. The same authors (Zaret and Sherman, 1984) reported the sequence of a region flanking two revertant alleles of the \textit{cycl-512} allele. Although they did not establish what segment of DNA was responsible for 3′ end generation in each case, and they did not sequence the entire region in which transcript 3′ ends were found, they did find a region of homology to the consensus sequence in which each of the three alleles had a different sequence. The regions homologous to the "core" of the tripartate consensus sequence \[\text{TAC...ATGT...TT}\] were, however, identical in the defective \textit{cycl-512} allele and the \textit{CYC1-512-E} and \textit{CYC1-512-K} alleles in which 3′ end signal function had been restored. Differences between \textit{cycl-512} and the revertant alleles were confined to an AT-rich sequence between the last two elements of the consensus sequence. That such permutations of an AT-rich sequence should affect the efficiency of the 3′ end signal implies that a particular AT-rich sequence, and not simply a high AT-content, is required for 3′ end generation.

The results of my study suggest that at least one element of the Zaret and Sherman consensus sequence is part of the 3′ end signal, but
they also indicate that other specific sequence requirements exist.

Evaluation of the importance of the first two elements of the Zaret and Sherman consensus sequence is complicated by the fact that they are related and repeated within the \textit{CycI} 3' end signal region. Table VII compares the consensus sequence to corresponding regions of the \textit{CycI} 3' end signal and its mutant derivatives. It is clear that the sequence TATGT, which is one form of the second element of the consensus sequence, is not specifically required for 3' end generation. It can be altered to TAGTT or TATCT in the intact \textit{CycI} gene without affecting 3' end signal function. The latter mutation also does not inactivate the \textit{CycI} 3' end signal in plasmid pAll.71. Similarly, the T1.30 mutation, which creates the sequence TAACCT, does not impair 3' end signal function. Although some flexibility in the sequence at positions 471 to +475 is apparently compatible with 3' end signal function, certain sequences at these positions do seem to impair 3' end generation. The G1.2 and G3.13 mutations, which create the sequences TACCT and TCTCT, respectively, both interfere with 3' end generation to the extent that they cause a 4-fold increase in \textit{lacZ} expression from pAll plasmids, as compared to pAll.71. Taken together, the results suggest that the sequence TATGT itself is not an essential feature of the 3' end signal, but that it might contribute to some general property which is important in 3' end generation.

The tetranucleotide TAGT immediately follows the sequence TATGT in the intact \textit{CycI} 3' end signal and represents a second candidate for the second element of the Zaret and Sherman consensus sequence. The present study provides some indication that the tetranucleotide is required for 3' end generation. As shown in Table VII, the 3' end
### TABLE VII

**COMPARISON OF THE ZARET AND SHERMAN CONSENSUS SEQUENCE TO MUTANT 3' END SIGNALS**

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>3' End Signal</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consensus</td>
<td>TAG.TA(T)GTT....TTT</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CYC1+</strong></td>
<td>TAGTTATGTTAGT...TTT</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>GI473a</td>
<td>TAGTTAGTTAGT...TTT</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>C474</td>
<td>TAGTTATCCTAGT...TTT</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>(pA11.71)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pA11.50</td>
<td>TAGTTATCTTT</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Gl.2</td>
<td>TAGTTAQCCTAGT...TTT</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>G3.13</td>
<td>TAGTTCTCCTAGT...TTT</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>T1.30</td>
<td>TAGTTAQCCTAGT...TTT</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Gl.10b</td>
<td>TAGTTATCCTTT</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>T1.18c</td>
<td>TAGTTATCCTAGT...TTA</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

a Differences from the CYC1+ sequence are underlined.

b Gl.10 includes a second point mutation, a T/A:C/G transition at position 5 of the oAS5/6 adapter

c The effects of the T1.18 mutation were assayed only in the pA12A/B plasmids.
signal fragment in plasmid pAll.50 lacks this tetranucleotide, though it retains the other elements of the Zaret and Sherman terminator, and it is at best marginally functional. Unfortunately, the deletion which eliminated the TAGT tetranucleotide from plasmid pAll.50 also eliminated 17 bp of 3' end signal sequence from downstream of the tetranucleotide. It would be somewhat reckless to attribute the level of lacZ expression from pAll.50 entirely to the lack of the TAGT tetranucleotide.

The plasmid pAll.Gl.10 is interesting because it supports lacZ expression at about the same level as pAll plasmids lacking the CYCl 3' end signal entirely. It carries two mutations as compared to plasmid pAll.71, the first being a G/C to T/A transversion at position +478 which alters the sequence TAGT to TATT. The second is a T/A to C/G transition at position 5 of the oAS5/6 adapter. Although it will be important to obtain separate 3' end signal fragments carrying each of these mutations, it seems likely that both contribute to the phenotype of pAll.Gl.10 transformants. Four other mutations (G3.13, Gl.2, Gl.16, G3.7) which introduce C/G base pairs into the 3' end signal fragment of pAll.71 all cause about a 4-fold increase in lacZ expression. One of these, G3.7, affects position 4 of the oAS5/6 adapter, immediately upstream of the position affected by the T/A to C/G transition in the Gl.10 mutant. It is likely that the dramatic enhancement of lacZ expression in pAll.Gl.10 transformants is at least partly due to the mutation at position +478, suggesting that the TAGT tetranucleotide is in fact required for 3' end generation.

The trinucleotide TTT makes up the third element of the Zaret and Sherman consensus sequence. It occurs at positions +493 to +495 of the
intact 3' end signal and again at positions +499 to +501, although this copy as well as others further downstream can be eliminated without ill effect. The Tl.18 mutation produces the sequence TTA at position +493 to +495. Plasmid pAl2A.Tl.18 does not produce a truncated CYCl transcript, suggesting that it lacks a functional 3' end signal.

Henikoff and Cohen (1984) demonstrated that the sequence TTTTTATA is part of the 3' end signal for transcription of a fragment of the Drosophila GART gene in yeast. The sequence T7ATA lies within the functional boundaries of the CYCl 3' end signal, at positions +459 to +468, and it might therefore also form part of the CYCl 3' end signal. It is clearly not the only required component of the 3' end signal, because plasmid pAll.50 retains the sequence T5ATA and yet directs lacZ expression at levels 15-fold greater than pAll.71, which has an additional 20 bp from the 3' end signal region.

Point mutations which alter the octanucleotide provide evidence that it is important to the function of the 3' end signal. The GG462 mutation creates the sequence TTTGGTTATA and causes a fraction of the transcripts of the intact CYCl gene to be extended beyond the normal 3' end site. The fact that the altered 3' end signal remains almost maximally active (90% efficient) suggests that perhaps a tract of thymidine residues in the mRNA-parallel strand is not a specifically required feature of the 3' end signal but contributes to some "general property" which is necessary for efficient 3' end generation. Another striking tract consisting mostly of thymidine residues occurs in the mRNA-parallel strand just outside the boundaries of the 3' end signal, at positions +506 to +519 [T5CT8]. This tract, though clearly not essential to 3' end generation, may be involved in providing the proper
sequence context for the 3' end signal. The \textit{CYClH5'}(+497) deletion removes this thymidine tract and although it does not prevent 3' end generation, it does cause the production of a small proportion of extended \textit{CYCl} transcripts.

The Tl.36 mutation is a T:A transversion at position +465 which creates the sequence TgAATA. A pAll plasmid carrying the Tl.36 mutation allows \textit{lacZ} to be expressed at levels 10- to 25-fold higher than the corresponding plasmid carrying the "wild-type" 3' end signal, pAll.71. \textit{LacZ} expression from pAll plasmids completely lacking the 3' end signal exceeds that from pAll.Tl.36 by less than a factor of two. Its dramatic effect on \textit{lacZ} expression suggests that the Tl.36 mutation seriously interferes with the function of the 3' end signal. (As discussed earlier, it must be borne in mind that the difference in \textit{lacZ} expression between pAll.Tl.36 and other pAll plasmids may not be due entirely to differences in the process of 3' end generation.) The sequence in the immediate vicinity of position +465 must therefore be considered as a candidate for an essential component of the 3' end signal. This possibility should be tested by attempting to identify other mutations in the region using the pAll screening system. Mutations which allow greatly increased \textit{lacZ} expression from pAll plasmids could then be introduced into the flanking sequences of the intact \textit{CYCl} gene and tested directly for their effect on \textit{CYCl} transcription.

A sequence which is noteworthy for its absence from the \textit{CYCl} 3' end signal region is the AATAAA hexanucleotide, which is almost universally present in the 3' flanking regions of higher eukaryotic genes and is known to be an essential part of the 3' end signal in several of them.
The **CYCl** 3' end signal contains the sequence ATTAA, but it seems fairly certain that this sequence is not crucial to transcript 3' end generation. The C482 mutation changed this sequence to ATCAA and did not cause the production of an extended transcript from an otherwise unaltered CYCl transcription unit. The Tl.34 mutation produced the sequence AATAA, and not surprisingly, pAll.Tl.34 did not exhibit elevated *lacZ* expression compared to plasmid pAll.71. In fact, pAll.Tl.34 seemed to direct lower levels of β-galactosidase synthesis than pAll.71, but the level of *lacZ* expression from either plasmid was so low that it is doubtful that the difference is significant. Zaret and Sherman (1982) pointed out that sequences closely related to AATAAA do not occur in regions flanking many yeast genes and are therefore unlikely to be important in 3' end generation. Henikoff et al. (1983) found that although the sequence AATAAA occurred in the 3' untranslated sequences of the *Drosophila ADE8* gene fragment, this sequence could be deleted without interfering with 3' end generation in yeast. The results of the present study, then, are consistent with those reported by other workers in suggesting that sequences related to AATAAA are not required for 3' end generation in yeast.

The C474 mutation carried by pAll.71 and its derivatives apparently did not inactivate the CYCl 3' end signal fragment carried on these plasmids. However, five other mutations which introduced C/G base pairs into the 3' end signal fragment all impaired its activity. While each may have affected some specifically required sequence, it is more plausible that all exerted an effect on 3' end generation by altering some general property of the 3' end signal. One interpretation is that a high proportion of A/T base pairs is important to 3' end generation.
As noted by Zaret and Sherman (1982), putative 3' end signal regions in yeast tend to have high A/T contents. The 50 bp CYC1 3' end signal fragment, for example, contains 40 A/T or T/A base pairs. Alternatively, it might be supposed that C/G base pairs, with C in the mRNA-parallel strand, are particularly injurious to the 3' end signal. To distinguish these possibilities, it will be necessary to analyze the effects of many more mutations which introduce C/G and G/C base pairs throughout the 3' end signal region.

The specific sequence requirements of the CYC1 3' end signal are not fully described by either the terminator sequence studied by Henikoff and Cohen (1984) or the consensus sequence recognized by Zaret and Sherman (1982). The results of this study suggest the existence of three specifically required blocks of sequence information in the CYC1 3' end signal region. The first includes at least the 3' end of the "terminator" octanucleotide studied by Henikoff et al. (1983) and Henikoff and Cohen (1984), TATA. In the CYC1 3' end signal, this octanucleotide overlaps the first copy of a repeated tetranucleotide, TAGT (positions +467 to +470), but the question of whether this particular copy of the tetranucleotide is subject to specific sequence constraints will only be settled when mutations altering its sequence are obtained. The second region in which requirements for a particular sequence exist is in the vicinity of position +478, and it coincides with the location of a second copy of the TAGT tetranucleotide (positions +476 to +479). An AT-rich sequence which includes the trinucleotide TTT and lies near the downstream boundary of the 3' end signal is the third region in which the nucleotide sequence is constrained by the requirement for 3' end signal function.
Applications of the pAll Screening System

Definition of the 3' End Signal

In this study, mutations in the 3' end signal region were first identified by single-track sequencing so as to test the efficiency of the mutagenic procedure, and then they were tested for their effects on 3' end generation. It should be possible to introduce a pool of pAll plasmids with mutagenized 3' end signal fragments into yeast and identify, by means of galactosidase assays, clones carrying pAll plasmids with inactive 3' end signals. Given a fairly efficient procedure for selectively mutagenizing the 3' end signal region, it should be feasible to test most single point mutant derivatives of the region. Such an analysis will facilitate the definition of the exact sequence requirements of the region. Certain mutations which interfere with 3' end generation may escape detection using the pAll system. A mutation creating an ATG triplet within the 3' end signal region, for instance, might interfere with 3' end generation, but it could also prevent the translation of the lacZ coding sequence.

Sherman and colleagues have attempted to define the sequence requirements of the CYCl 3' end signal by selecting Cyc+ revertants of a cycl-512 yeast strain (Kotval et al., 1983, Zaret and Sherman, 1984). Reversion of the cytochrome c deficiency was found to occur by means of a variety of genetic alterations, some of which were described in Chapter I. The apparent absence of any striking similarity between the 3' end signal regions of various Cyc+ revertants of a cycl-512 strain led to the suggestion that 3' end generation in yeast can occur in response to a number of different signals (Zaret and Sherman, 1984). Although a detailed comparison of sequences flanking the CYCl
alleles in various revertants might yet reveal features common to the 3' end signals, it is clear that the approach would be usefully complemented by studying mutations which inactivate a single, well-defined 3' end signal. If the signal exhibits definite sequence requirements, mutational inactivation of the signal should reveal those requirements. Flexibility in the sequence requirements of the signal may also be revealed through mutations which leave the signal active though they occur at sites at which other mutations are not tolerated, or through pairs or groups of mutations which complement each other in cis to restore 3' end signal function. The work described here is an effort at exploiting the latter approach.

Identification of Genes Required for 3' End Generation

The pAll plasmids may be useful in identifying genes whose products interact with the CYCl 3' end signal in bringing about transcript 3' end generation. Kotval et al. (1983) and Zaret and Sherman (1984) have identified two extragenic suppressors of the cytochrome c deficiency brought about by a CYCl 3' end signal defect. Mutations in one of the suppressors appear to cause 3' end generation at cryptic sites downstream of the defective CYCl gene. Mutations in the other suppressor appear to affect mRNA stability and may not directly affect 3' end generation.

Mutations in genes required for transcript 3' end generation would be lethal if they prevented the synthesis of functional transcripts of other essential genes. Otherwise, such mutations would not normally be expected to confer upon cells a readily detectable phenotype. Two different approaches to identifying genes required for 3' end generation by means of their effects on lacZ expression in pAll
transformants are envisioned. In the first, a pAll plasmid carrying a
defective 3' end signal (preferably one differing from the functional
signal by a single base substitution) would be introduced into a
population of mutagenized diploid cells. Most of the transformants
would be expected to express lacZ at levels higher than the basal
level. Any which exhibited only basal levels of lacZ expression would
be candidates for carrying a mutant gene whose product is capable of
causing proper 3' end generation at the mutant 3' end signal of the
pAll plasmid. Screening for transformants with basal levels of
β-galactosidase could conveniently be carried out on indicator plates.

An alternative approach would involve looking for mutations which
prevent 3' end generation at an intact CYCl 3' end signal and thus
allow increased lacZ expression from a plasmid carrying the signal.
This type of search would have to be carried out in haploid cells
because the mutations would probably be recessive. One normal allele
of a gene required for 3' end generation in a heterozygote would
probably suffice to allow normal 3' end generation. Since mutations
interfering with transcript 3' end generation at the CYCl 3' end signal
might also interfere with the transcription of other genes, they might
be lethal. Therefore the search should be directed toward finding
conditional mutants. Temperature-sensitive mutants might be isolated
in the following manner: a pAll plasmid bearing the functional CYCl 3'
end signal would be introduced into a mutagenized population of haploid
cells, which would then be plated on indicator medium at a permissive
temperature. The entire population would be replica-plated to
indicator medium, grown at the low temperature to produce colonies,
then shifted to a higher (non-permissive)
temperature. Clones producing higher levels of $\beta$-galactosidase after the temperature shift might carry ts mutations interfering with 3' end generation, and they could be recovered from the original plates which had been maintained at the permissive temperature.

Other Screening Systems for 3' End Signal Defects

Brent and Ptashne (1984) recently reported that fragments containing yeast 3' end signals limited the expression of $\text{lacZ}$ from the GAL1 promoter when inserted between the upstream activation site and TATA site of the promoter. A 182 bp fragment carrying the 3' end signal cycl-512 allele, which has been shown to be very inefficient in bringing about 3' end generation (Zaret and Sherman, 1982) caused a 2-fold drop in $\text{lacZ}$ expression. The corresponding fragment of the CYC1-512-F revertant allele, which differed from the cycl-512 fragment by a single base substitution, caused a 6-fold reduction in $\text{lacZ}$ expression. A 324 bp fragment including the putative 3' end signal region of ADC1 was much more effective in limiting $\text{lacZ}$ expression, reducing it by a factor of 100. Two fragments without known 3' end signal activity were shown to have no effect on $\text{lacZ}$ expression in the same context. If the ability of a fragment to prevent $\text{lacZ}$ expression when inserted within a promoter sponsoring $\text{lacZ}$ transcription is correlated with its ability to serve as a 3' end signal, then constructs of this type may be used in the same ways as the pAll plasmids to identify mutations affecting 3' end signal activity.

The type of construct described by Brent and Ptashne places the 3' end signal upstream of the transcript initiation sites, in a region which is not known to be transcribed. Brent and Ptashne mention the possibility that an unstable transcript is synthesized in the region
between the upstream activation site and the TATA site, but there is no evidence for this and the role that such a transcript would play in promoter function and transcription of the structural gene is obscure. McNeil and Smith (1985) have provided evidence that transcription of the CYCl gene initiates at sites downstream of a series of TATA sequences and that CYCl mRNAs are not derived from longer transcripts by processing.

Placing the 3' end signal in a non-transcribed region offers the advantage that its sequence and length can have no influence on the stability or translational efficiency of the lacZ mRNA. Accordingly, the requirement that any lacZ mRNA produced be functional places no constraints on the length, base composition or sequence of the 3' end signal fragment. One might imagine that sequence changes between the upstream activator site and TATA site could alter promoter activity regardless of their effects on a 3' end signal within the region, but Brent and Ptashne (1984) showed that sequences normally present within this region are unimportant to the activity of the promoter. While it is possible that comparisons between the expression of constructs carrying completely different 3' end signal fragments between the UAS and TATA sites might be affected to some degree by differences in promoter activity, comparisons between genes carrying related 3' end signal fragments in the promoter region are likely to be valid comparisons of some property of the inserted fragments.

A more important concern is that, since the fragments are inserted in a non-transcribed region, the property which allows them to limit lacZ expression may be different from or may be only one of several properties that are required for 3' end signal function. For example,
in the same paper, Brent and Ptashne showed that a \textit{lexA} operator could block transcription from the \textit{GAL1} promoter if inserted between the upstream activation and TATA sites. It is possible that any fragment which specifically and tightly binds a protein would similarly block transcript initiation. Specific binding of some protein may well be a property of 3' end signal fragments, but it may not be sufficient to account for their activity as 3' end signals. Mutations which inactivate the 3' end signal but have at least one protein-binding site intact would not cause \textit{lacZ} expression to change and would escape detection. It will be interesting to learn whether or not the same features of the \textit{CYCl} 3' end signal which allow it to block \textit{lacZ} expression from a pAll plasmid are also responsible for its effect on \textit{lacZ} expression in the type of plasmid described by Brent and Ptashne (1984).

A third approach to screening for mutations affecting 3' end signal function might overcome the limitations of both the approach developed in the present study and that outlined by Brent and Ptashne. Teem and Rosbash (1983) have constructed a hybrid gene consisting of the \textit{E. coli} \textit{lacZ} gene joined to the second exon of the rp51 gene. The production of a functional \textit{lacZ} mRNA from this gene requires that the rp51 intron be correctly excised from the primary transcript. It should be possible to insert a 3' end signal with the intron of this or a similar hybrid gene in such a way that it would not interfere with splicing of the primary transcript. [Splicing in yeast has been shown to require specific sequences at exon/intron junctions and the sequence TACTAACA near the 3' end of the intron (Langford and Gallwitz, 1983; Langford et al., 1984; Pikielny et al., 1983). Insertions upstream of this]
octanucleotide would not be expected to interfere with splicing.] A functional 3' end signal would prevent $\text{lacZ}$ expression by preventing the completion of the primary transcript. The location of the 3' end signal within a transcribed region would offer some assurance that its effects on $\text{lacZ}$ expression were a direct reflection of its normal effect on transcription. An inactive 3' end signal would offer no impediment to transcription or, because it would not form part of the spliced mRNA, to message stability or translational efficiency.
REFERENCES


-269-


-G272-


-276-


-280-


-282-


