INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

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

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

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Expression of the mouse thymidylate synthase gene is regulated through synergy between the essential promoter and introns

Ash, John D., Ph.D.
The Ohio State University, 1994
EXPRESSION OF THE MOUSE THYMIDYLATE SYNTHASE GENE IS REGULATED THROUGH SYNERGY BETWEEN THE ESSENTIAL PROMOTER AND INTRONS

DISSERTATION

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

BY

John D. Ash, BA

*****

The Ohio State University

1994

Dissertation Committee:

Dr. Lee F. Johnson
Dr. Deborah Parris
Dr. Ing-Ming Chiu
Dr. Jeffrey Masters

Approved By:

Dr. Lee F. Johnson
Adviser

Ohio State Biochemistry Graduate Program
ACKNOWLEDGEMENTS

I would like to express my appreciation to my advisor Dr. Lee Johnson for the opportunity to study under his guidance. I would also like to thank my committee members for their valuable suggestions during the course of my research. I thank the past and present members of our laboratory for personal and scientific interactions, and for their cooperation in establishing a mutually productive environment in the lab. I thank my colleagues John Lucas, Depa Subramanian, and Allisa Winard for their support and friendship. Finally I would like to express my deepest appreciation to my wife Cindy for her understanding, love, and support.
VITA

March 31, 1964 .......................................................... Born, Richmond, Texas

1986 ................................................................................. BA Biology

Austin College
Sherman, Texas

1986-1988 ........................................................................ Research Assistant

Department of Research
The Veterans Administration
Medical Center, Dallas

1988-1990 ........................................................................ Teaching Assistant

Departments of General Biology and Biochemistry
The Ohio State University,
Columbus, Ohio

1990-1993 ........................................................................ Research Assistant

Departments of Biochemistry and Molecular Genetics
The Ohio State University
Columbus, Ohio

1993-1994 ........................................................................ N.I.H. Predoctoral Trainee

The Ohio State University
Columbus, Ohio
Field of Study:

Major Field: The Ohio State Biochemistry Program
TABLE OF CONTENTS

ACKNOWLEDGEMENTS .................................................................................................................. ii

VITA ........................................................................................................................................ iii

PUBLICATIONS ........................................................................................................................ iv

TABLE OF CONTENTS ........................................................................................................ vi

LIST OF FIGURES ................................................................................................................... x

LIST OF TABLES .................................................................................................................... xiii

INTRODUCTION ......................................................................................................................... 1

THE MOUSE THYMIDYLATE SYNTHASE (TS) GENE ............................................................... 1

Genomic Structure And cDNA Structure .............................................................................. 1

Interesting Aspects Of TS Gene Expression .......................................................................... 2

TS Growth-Dependent Regulation ....................................................................................... 6

REGULATION OF THE CELL-CYCLE ..................................................................................... 9

GROWTH STIMULATION ......................................................................................................... 13

REGULATION OF GROWTH-DEPENDENT GENES ............................................................... 16

E2F-1 ....................................................................................................................................... 17

TK ............................................................................................................................................. 19

DHFR ...................................................................................................................................... 20
Nuclear Extract Preparation .................................................................................................. 46
Probe Preparation ............................................................................................................. 47
Gel Mobility Shift assay ..................................................................................................... 49

RESULTS .................................................................................................................................... 51

MINIGENE NOMENCLATURE.................................................................................................. 51
THE TS POLYADENYLATION SIGNAL IS NOT REQUIRED FOR REGULATION ............... 52
INTRONS AS CIS-ACTING REGULATORY ELEMENTS ........................................................... 60
Minigenes Contain All Necessary Signals For S Phase Regulation ...................................... 60
Introns Are Essential For Growth Regulation ....................................................................... 64
The TS Promoter is Essential For Regulation ...................................................................... 73
The TS Promoter Regulatory Elements Are Portable .......................................................... 80
There Are No Unique Regulatory Elements
Located Within The Interior Of Intron Two ........................................................................... 82
REGULATORY ELEMENTS 5' OF THE AUG CODON .............................................................. 86
Defining the 5' Boundary ...................................................................................................... 86
Defining the 3' Boundary ...................................................................................................... 101
Analysis of The Gl Regulation of the -118 to 90 Promoter ................................................. 109
Regulation of the TS Promoter in the Reverse Orientation .................................................. 116
Alterations of the Promoter that Affect Regulation also Affect Transient Expression ........ 118

ALTERATIONS THAT AFFECT REGULATION EFFECT THE FORMATION OF DNA-PROTEIN
COMPLEXES ................................................................................................................................ 121
LIST OF FIGURES

FIGURE 1 REGULATION OF THE CELL CYCLE BY CYCLINS ........................................ 12

FIGURE 2. PROBES USED IN NUCLEASE S1 ASSAYS .................................................. 38

FIGURE 3. MINIGENES USED IN POLY(A) SIGNAL SUBSTITUTION STUDIES .............. 53

FIGURE 4. POLY(A) SIGNAL UTILIZATION ...................................................................... 58

FIGURE 5. GROWTH REGULATION OF T11,2T ............................................................ 62

FIGURE 6. GROWTH REGULATION OF T15,6T ............................................................. 65

FIGURE 7. GROWTH REGULATION OF TTS .................................................................. 67

FIGURE 8. GROWTH REGULATION OF TTP ................................................................. 68

FIGURE 9. GROWTH REGULATION OF TTP ................................................................. 69

FIGURE 10. GROWTH REGULATION OF T15,6P .......................................................... 71

FIGURE 11. GROWTH REGULATION OF T15,6S .......................................................... 72

FIGURE 12. MINIGENES USED IN PROMOTER SUBSTITUTION STUDIES .................. 74

FIGURE 13. GROWTH REGULATION OF GG11,2G .................................................... 77

FIGURE 14. GROWTH REGULATION OF SI5,6T .......................................................... 79

FIGURE 15. GROWTH REGULATION OF TG11,2G ...................................................... 81

FIGURE 16. GROWTH REGULATION OF T11,2dT ...................................................... 83

FIGURE 17. GROWTH REGULATION OF SI1,2dT ...................................................... 85
FIGURE 18. SUMMARY OF PROMOTER DELETIONS BETWEEN -1000 AND -250 .......... 87
FIGURE 19. MUTATIONS IN THE PROMOTER BETWEEN -250 AND -150 ................. 89
FIGURE 20. MUTATIONS IN THE PROMOTER BETWEEN -165 AND -11 ................. 91
FIGURE 21. GROWTH REGULATION OF M115 .............................................................. 92
FIGURE 22. GROWTH REGULATION OF M110 .............................................................. 93
FIGURE 23. GROWTH REGULATION OF mCR ............................................................. 94
FIGURE 24. GROWTH REGULATION OF -113 Tt1,2dT ............................................. 96
FIGURE 25. GROWTH REGULATION OF -105 Tt1,2dT ............................................. 98
FIGURE 26. GROWTH REGULATION OF -165 del130-105 .................................... 100
FIGURE 27. THE 3' BOUNDARY OF THE REGULATORY REGION ......................... 102
FIGURE 28. GROWTH REGULATION OF -118-TO-11 ............................................. 103
FIGURE 29. GROWTH REGULATION OF -118-TO-75 ............................................. 104
FIGURE 30. GROWTH REGULATION OF -118-TO-90 ............................................. 105
FIGURE 31. GROWTH REGULATION OF -105-TO-75 ............................................. 107
FIGURE 32. SUMMARY OF ESSENTIAL REGION MODIFICATIONS ....................... 108
FIGURE 33. MUTATIONS WITHIN THE ESSENTIAL REGULATORY REGION ............. 110
FIGURE 34. GROWTH REGULATION OF -118-TO-92 ............................................. 111
FIGURE 35. GROWTH REGULATION OF -118-TO-92 M100 .................................... 112
Figure 36. Growth Regulation of -118-to-92 M107 ................................................. 113

Figure 37. Growth Regulation of -118-to-92 MB0th .............................................. 114

Figure 38. Growth Regulation of 2x -118-to-92 ....................................................... 115

Figure 39. Growth Regulation of -118-to-11 Rev .................................................. 117

Figure 40. Protein Binding sites within the Essential Region .................................... 122

Figure 41. DNA-Protein Interactions within The Essential Region ............................ 124

Figure 42. DNA-Protein Interactions with Mutant Probes ........................................ 127

Figure 43. DNA-Protein Interactions With the -118 to -90 Probe ............................. 129

Figure 44. The Location of DNA-Protein Interactions ............................................. 132

Figure 45. Growth Stimulated DNA-Protein Interactions ....................................... 135

Figure 46. Conserved Promoter Elements ................................................................. 137
LIST OF TABLES

Table 1. Expression Levels of Minigenes with Substituted Poly(A) Signals .... 55
Table 2. Transient Expression of Promoter Alterations ............................... 120
INTRODUCTION

The Mouse Thymidylate Synthase (TS) Gene

Thymidylate synthase (TS) catalyzes the reductive methylation of deoxyuridylic acid to form deoxythymidylic acid in the de novo biosynthetic pathway. TS is therefore an essential enzyme (EC 2.1.45) in proliferating cells, and is often an important target enzyme in cancer chemotherapy. Since TS has an important role in DNA synthesis, it is not too surprising that the expression of this gene is at a maximum when cells are in S phase and actively synthesizing DNA. A fluorodeoxyuridine (FdUrd)--resistant 3T6 cell line was generated (LU3-7), in which the TS locus had been amplified more than 50 fold. This cell line proved to be a useful tool in not only the isolation of both cDNA and genomic clones, but also in the initial characterization of S phase regulation.

Genomic Structure And cDNA Structure

Analysis of the TS cDNA revealed that the TS mRNA contains a 921 nt. open reading frame, which predicts an evolutionarily conserved protein of 34.9 kDa. Analysis of overlapping genomic clones revealed that the mouse TS gene spans a 12 kb region, and that the coding region is interrupted by 6 introns. Due to the complex pattern of
transcription initiation the majority of TS mRNA has a short, but variable 5' untranslated region (1 to 75 nt.). TS mRNA is unusual in that it is polyadenylated at the translation termination codon, resulting in a 3' untranslated region composed entirely of the poly (A) tail.7,8,9

**Interesting Aspects Of TS Gene Expression**

TS minigenes were constructed and analyzed for their expression in mammalian cells. Utilizing DNA fragments from the cDNA clone and appropriate fragments from genomic clones, an intronless TS minigene was constructed, which contained the TS promoter and 5' flanking region linked to the TS coding region, (derived from the cDNA), and the TS 3' flanking region.10 By substituting genomic DNA fragments into the cDNA, intron containing minigenes were created which contained introns located at their normal position relative to the coding region.11 Extensive deletion and mutational analyses have been carried out on the promoter sequences, intron sequences, and 3' processing sequences. The transient expression analysis of these minigenes has revealed several interesting aspects of TS gene expression.

**Promoter Studies**

The TS promoter has been shown to be unusual in that it lacks TATA and CCAAT sequences, and that transcription is initiated at virtually every nucleotide over a 60 nt. region.12 Extensive mutation and deletion analysis has shown that the complex
pattern of transcriptional initiation is due to the lack of a specific initiator element, and that the core promoter lies within a 38nt region between -113 and -75 (relative to the AUG translation initiation codon).\textsuperscript{9,13}

The core promoter contains close matches to the consensus recognition sequences for several known transcription factors, including sites for Ets, E2F, and Spl. There are two GGAA elements that match the core recognition sequence for members of the Ets transcription factor family located at -100 and -85. A G/C box corresponding to the recognition motif for the transcription factor Spl is located at -80.\textsuperscript{14} A close match to the binding sequence for the E2F1 transcription factor is located just upstream of -105 (-114 to -105).\textsuperscript{15} In gel retardation assays, an unknown DNA-protein complex has been mapped to the -100 Ets site (referred to as CIII). Spl has been shown to bind to its recognition element at -80, although weakly.\textsuperscript{12} No complex has been observed for the -85 GGAA box. Mutations at both GGAA boxes, and at the Spl elements have resulted in reduction of promoter activity.\textsuperscript{12,13} No DNA-protein complex has been observed for the putative E2F1 element. An unknown complex was observed (referred to as CII), and mapped by mutation and footprinting analysis, to a region just downstream of the 5' GGAA box. The binding site for this protein does not match any known recognition motifs, and a mutation that abolishes the DNA-protein complex has no effect on TS gene expression. Therefore, the biological significance of this complex is unknown.
The TS core promoter element has been shown to be transcribed bidirectionally, raising the possibility of a second gene being expressed from the TS promoter.\textsuperscript{16,17} The transcription pattern of these opposite direction transcripts appears to be as complex as the TS transcription pattern.

\textit{Intron Studies}

The first TS minigenes constructed were intronless, and regardless of the length of the 5' flanking region, they were found to express at very low levels in transient transfection assays.\textsuperscript{10} Surprisingly, when introns were included within the minigenes the level of expression increased 4 to 30 fold depending on which introns were included.\textsuperscript{11} Only intron 4 did not stimulate gene expression. By mutation and deletion analysis it was determined that intron-stimulated gene expression was correlated with intron splicing efficiency.\textsuperscript{18} Intron 4 was shown to contain poor matches to consensus splicing signal sequences, and that it could be converted to a stimulatory intron by improving the U1-donor recognition sequence. Further, a stimulatory intron could be converted to a non-stimulatory intron by altering the splice donor signal sequence to the one found in intron 4. The biochemical mechanism for intron dependent gene expression is not known at this time; however, intron recognition is clearly an important step.
Polyadenylation Signal Studies

The TS mRNA is unusual in that it lacks a 3' untranslated region. The termination codon in the genomic sequence is TAG, while in the cDNA it is TAA. Therefore, TS mRNA is polyadenylated at the termination codon, and the last nucleotide in the termination codon TAA is created by the addition of the poly-A tail.\(^{19}\) Interestingly, the integration of an L1 retrotranposable element, and subsequent activation of a cryptic polyadenylation signal sequence is responsible for the unusual 3' end structure of the TS gene. There are two cis-acting elements that are important for the cleavage and polyadenylation reactions, including the hexanucleotide which in the TS gene is located in the coding region upstream of the cleavage site, and the G/U rich region which is located downstream of the cleavage site. The poly-A tail of the integrated L1 element created the G/U rich region.

The TS hexanucleotide element (AUUAAA) is unusual in that is a 5 out of 6 match to the consensus element (AAUAAA). In in vitro polyadenylation reactions, this is the only functional variant to the consensus sequence, and it has been shown to occur in about 10% of the known genes. In preliminary studies, perfecting the hexanucleotide signal was shown to increase gene expression 2 to 5 fold, suggesting that the TS polyadenylation signal is not utilized efficiently. When the TS polyadenylation signal and 3' flanking region were replaced with the SV40 early signal the level of expression
increased about 10 fold, further demonstrating that the TS polyadenylation signal in not utilized efficiently.\textsuperscript{19,20}

**TS Growth-Dependent Regulation**

While TS gene expression is marginally regulated (2 to 3 fold) in rapidly dividing cells, both the human and mouse genes are highly regulated (10 to 20 fold) in cells that have been growth stimulated.\textsuperscript{21,22,23,24} In either cycling cells or growth stimulated cells TS expression is at maximum when cells are in S phase and actively synthesizing DNA. The LU3-7 cell line was used to show that in growth stimulated cells the steady-state cytoplasmic level of TS mRNA increases 10 to 20 fold.\textsuperscript{24} During the same interval the rate of transcription increases only 2 to 3 fold, thus indicating that the majority of regulation is due to post-transcriptional events. One such event could be regulated mRNA turnover in the cytoplasm. This, however, was shown not to be important for TS gene regulation since the cytoplasmic half life of TS mRNA is fairly constant during the induction period (9.6 hours in resting cells and 7.3 hours in growing cells). RNA processing, particularly polyadenylation, was shown to be important for regulation by pulse-labeling experiments. When resting LU3-7 cells were pulsed labeled with tritiated uridine, only 30\% of the newly transcribed TS transcripts were polyadenylated. However, in S phase cells this fraction increased to 70\%.\textsuperscript{24} This data suggested that the efficiency of 3' end processing for TS RNA is regulated during growth stimulation.
Since the polyadenylation state of mRNA is important for stability, as well as other aspects of RNA metabolism, it seemed possible that TS gene regulation could be accomplished by differential 3’ processing.

To further study TS regulation, and to identify cis-acting regulatory elements in the mouse TS gene, chimeric minigenes were constructed, reintroduced into 3T6 cells, and analyzed in growth stimulation experiments. Minigenes were constructed to dissect the TS gene into three parts, the promoter and 5’ flanking region, coding region, and polyadenylation signal. In the initial experiments it was shown that minigenes containing 1kb of the TS genomic 5’ flanking region, introns 5 and 6 or introns 1 and 2, and 334bp of 3’ genomic flanking region, were regulated normally in response to growth stimulation. In further studies intron 5 and 6 containing minigenes were used to search for elements located in the 5’ and 3’ ends of the TS gene.

When the TS promoter and 5’ flanking region were replaced with the SV40 early promoter (creating the minigene Si5,6T) regulation was lost, demonstrating that the TS 5’ flanking region and promoter were essential for regulation. Minigenes with 5’ deletions were created and analyzed to locate regulatory elements within the 5’ flanking region. When the 1 kb 5’ flanking region in Ti5,6T was deleted to -250 (relative to the AUG codon), the magnitude of the minigene S phase response was significantly reduced. When the 5’ flanking region was further deleted to -150, regulation was apparently lost. Therefore, it seemed that the 5’ flanking region might contain two distinct regulatory
regions, one between -1000 and -250, and another between -250 and -150. Since regulation was appeared to be lost when the promoter was deleted to -150 it seemed unlikely that any regulatory elements would lie between the AUG codon and -150. Previously the essential promoter was determined to be located between -113 and -75. Therefore, the 5’ regulatory elements were thought to lie outside the TS promoter.

When the entire 1kb promoter was used to drive the expression of the CAT (chloramphenicol acetyl transferase) gene, no regulation was observed. Thus, even though the 5’ flanking region is essential for regulation, it can not function alone. Sequences downstream of the AUG codon must also be included within the minigene for regulation. These sequences might be located within introns, the coding region, or the polyadenylation signal. Preliminary growth stimulation analysis of intronless minigenes suggested that introns were also essential for regulation (unpublished observations of Yue Li and Linda Ritter). Since both intron combinations (1 and 2, or 5 and 6) appeared to function similarly in regulation, the interpretation of these experiments was that a unique regulatory element would not likely be redundant in each intron. Instead, it seemed more likely that intron splice site recognition and commitment to intron splicing were the important steps in TS growth stimulation. If this is hypothesis true, then it is possible that just about any functional intron might be sufficient.

Since regulated polyadenylation seemed to be involved in TS regulation, chimeric intron 5 and 6 containing minigenes were constructed. In one minigene the aberrant
hexanucleotide (AUUAAA) was modified to contain the consensus hexanucleotide (AAUAAA), creating Ti5,6P. In another minigene the entire TS 3' genomic flanking region, including an essential portion of the TS polyadenylation signal, was replaced with the SV40 early polyadenylation signal, creating Ti5,6S. Preliminary analysis indicated that either modification resulted in reduction of minigene response to growth stimulation. Since both polyadenylation signal modifications were shown to improve 3'end processing, and reduce minigene regulation, it was thought that an inefficient signal might be important for full regulation.

**Regulation of the Cell-Cycle**

In eukaryotic cells the process of cell division is divided into four distinct phases, M, G1, S, and G2. The last phase in the cycle, mitosis, is the most obvious under light microscopy, and is the process of physically dividing the parental cell into two daughter cells. Following mitosis each daughter cell enters a new cell division beginning with G1, in which cells grow in size and prepare for DNA synthesis. If growth factors or nutrients are limiting during G1 the cell can reversibly arrest growth and withdraw from the cell cycle entering a G0 phase. When conditions are favorable for cell division, the cell will exit G0 and reenter G1, proceeding through the remainder of cell division. However, if growth factors are not limiting, the cell will not enter the G0 state, but will proceed through the cell cycle unabated. Once dividing cells have passed through a point
in G1 (called the restriction point) they no longer require exogenous growth factors to proceed though the cell cycle. From this point throughout the remainder of the cell division, cells follow a tightly regulated progression of events that are controlled by the internal cell cycle clock.

Progression through the cell cycle is regulated by cyclin dependent kinases as illustrated in Figure 1. In mammalian cells there are five classes of cyclins (A, B, C, D, and E). All cyclin proteins and mRNAs have short half lives, and the expression of each is tightly regulated with the cell cycle. The D cyclins control events in G1, by responding to growth factor signals. Unlike the other cyclins, whose expressions are tightly linked to the cell cycle, the D cyclins are expressed in G1 only if exogenous growth factors are present. Cyclin D disappears rapidly once exogenous growth factors are removed from G1 cells, and increase sharply in G0 cells that have been stimulated with mitogens. Because of their involvement in G1 and their responsiveness to growth factors, D cyclins are strong candidates for the cellular factors that regulate entry and exit from G0 as well as the initial steps leading through the restriction point. The D cyclins, like all cyclins, are only functional if they form complexes with cyclin dependent kinases (cdk). Six different cdks (cdk1 through cdk6) have been identified, and all are activated only when bound to a cyclin. All cdks are expressed constitutively throughout the cell cycle but their catalytic activity is regulated by cyclins and by the phosphorylation state of the cdk. Cyclin-cdk complexes are believed to control cell cycle progression through phosphorylation of other regulatory molecules. The cyclin-cdk
complexes are themselves regulated by phosphorylation state. Additional kinases and phosphatases have been identified which contribute to the regulation of cyclin-cdk complexes.\textsuperscript{27} The D cyclins have been shown to interact with four different cdks during the G1 phase (cdk2, cdk4, cdk5, and cdk6).\textsuperscript{30} As cells enter S phase, cyclin D expression is turned off, and cyclin E expression is activated. During the time interval from when cells progress through restriction point to when they are well into S phase cyclin E forms a stable complex with cdk2, and together these molecules control events in early S phase. Throughout S phase cyclin A expression is activated, and cyclin A forms a complex with cdk2. Toward the end of S phase and during G2, cyclin A and B each form a stable complex with cdk1 (formerly cdc2) to regulate the progression into mitosis. Once mitosis is completed A and B cyclins are rapidly degraded. If growth factors are present, D cyclins are turned on, to start the process over again, or if growth factors are not present cyclin D expression is not turned on and the cell enters the G0 state.
Figure 1  Regulation of the Cell Cycle by Cyclins.
**Growth Stimulation**

With respect to the normal development of multicellular organisms, the G1 phase is perhaps the most important regulatory phase in the cell cycle. Most somatic cells are not continuously dividing, but instead are either in a terminally differentiated state and incapable of further cell division, or are in the reversible growth arrested G0 phase. In the body and in culture, when mitogenic factors are at limiting concentrations, fibroblast cells will exit the cell cycle and enter the G0 phase. Cells in the G0 phase will remain dormant until differentiation, death, or until they receive a mitogenic signal to reenter the cell cycle. During development and during wound healing, G0 fibroblasts can be stimulated by mitogens to progress through multiple cell divisions. In the laboratory, immortalized, non-transformed fibroblast cells will also arrest in the G0 state when growth factors are limiting. Resting cells can be stimulated to divide when cultured at modest densities, in media supplemented with mitogenic growth factors. The process of inducing non-growing cells to re-enter the cell cycle is called growth stimulation.26,28

The maintenance of the G0 state is accomplished by the action of tumor suppresser genes like Rb.39,40 The Rb protein is a member of a group of proteins that are referred to as "pocket proteins". The group includes Rb, p107, and p130. Each of these proteins shares a common structural feature that allows them to bind factors that are
important for cell division.\textsuperscript{41,42,43,44} The structural feature has been referred to as the protein binding pocket, hence the name pocket proteins.

Rb is a nuclear factor that undergoes changes in phosphorylation as cells progress through the cell cycle. In quiescent cells, Rb is hypophosphorylated, but in late G1, S, G2, and M phases it is found in a hyperphosphorylated form.\textsuperscript{45,46,39} At the end of mitosis it is believed that Rb is converted from an inactive hyperphosphorylated state to its active state by specific phosphatases, and when cells are growth stimulated Rb is inactivated by cyclin dependent kinases.\textsuperscript{47,48}

Consistent with the observations of Rb regulation is a model of how Rb might regulate G1 events. When growth factors are limiting, cyclin D levels will not increase and Rb will not be inactivated by phosphorylation. Rb can then repress late G1 gene expression, driving the cell into G0. When G0 cells are stimulated to reenter the cell cycle (typically by mitogen activated signal transduction pathways), cyclin D levels increase, leading to Rb inactivation through phosphorylation. Once inactivated, Rb is no longer able to block cell cycle progression. If, however, growth factors are not limiting at the end of mitosis the cellular levels of cyclin D increases in G1, and cyclin D (interacting with a cdk) phosphorylates Rb.

Rb is believed to block G1 gene expression through direct interactions with transcription factors. Recent evidence has shown that the active, hypophosphorylated form of the Rb protein, is able to enter into higher order protein complexes with several
cellular factors, including but not limited to, the transcription factors in the E2F, and DTRF families. E2F family members have been shown to transactivate the expression of late G1 and S phase genes, and it has been shown that interactions with Rb prevents E2F from activating its target promoters. From these observations it has been proposed that part of the mechanism by which Rb blocks progression to S phase, is by disrupting E2F directed transcription.

Recent observations have suggested the possibility that each member of the pocket protein group plays a specialized role in regulating cell cycle progression. As indicated above, when Rb binds to E2F its trans-activation properties are repressed; also indicated above was that Rb is in its active form during G0 phase. However, the association of Rb with E2F is more apparent during S phase than during G1, suggesting that Rb has a relatively small role to play in the G0-S regulation of E2F transcriptional activity. The fact that Rb homozygous knock-out mice, survive until about 12 days post-gestation indicates that during the periods of rapid proliferation in early embryogenesis Rb is not required to regulate the cell cycle. Instead, the abnormal differentiation observed in these embryos suggests that Rb might be more important for cellular differentiation than for cell cycle progression. Like Rb, the other pocket proteins can repress E2F-dependent transcription, and can prevent cell cycle progression. However, their temporal difference in E2F association implies that each may play a
unique role in cell cycle progression. In contrast to Rb-E2F, p107-E2F complexes occur throughout G1 and S phase, and p130-E2F complexes appear predominantly in G1.57,52,53

Another mechanism by which gene expression is controlled during G1, is by direct association of cyclin-cdk to transcription factors. In conjunction with p107, E2F forms complexes with cyclin A and the cdk2 kinase in extracts prepared from S phase cells, and with cyclin E and cdk2 in extracts from G1 cells.58,59,60,53,50 It is believed that these interactions form a DNA bound kinase complex that is capable of phosphorylating other DNA binding proteins, including the general transcription factors and RNA polymerase subunits. The possible phosphorylation of these factors is a potential regulatory mechanism of G1-S gene expression.

Regulation Of Growth-Dependent Genes

Following mitogen stimulation there is an increased expression of transcriptional regulatory proteins, including Fos, and erg-1. These factors in turn modulate the expression of other genes. Eventually cells progress through G1 and become committed to cell division, and once the cell is committed to begin DNA synthesis there is a dramatic increase in the expression of enzymes necessary for DNA replication. These S phase genes include transcription factors like E2F1 and cMyb, which are believed to influence the expression of other S phase genes. The expression of enzymes necessary for DNA synthesis are also induced during S phase, including the expression of thymidine
kinase (TK), dihydrofolate reductase (DHFR), the large subunit of ribonucleotide reductase (RNR), DNA polymerase α (DPα), proliferating cell nuclear antigen (PCNA), and thymidylate synthase (TS). Also induced is the expression of structural proteins like histones, which are necessary for the assembly of newly synthesized DNA into chromatin. In yeast the induction of S phase genes is the result of transcriptional activation by a single transactivator.\textsuperscript{61} In mammalian cells however, it has been difficult to demonstrate a common mechanism that is responsible for the regulation of all S phase genes. The observed regulation of many of these genes is the additive result of multiple regulatory mechanisms including, transcription, mRNA stability, RNA processing, mRNA translation, and post-translational regulation of the proteins. To further complicate the situation, not all genes are regulated by the same combination of mechanisms. The identification of cis-acting regulatory elements has given some information on how the coordinated expression of these genes is controlled, and has shed some light on how these controls are integrated with growth regulation and the cell cycle.

\textit{E2F-1}

E2F transcription factors were originally identified as factors binding to control elements in the E2 gene in the adenovirus genome.\textsuperscript{62} Later, it was found that E2F transcription factors also bind to the DHFR promoter as well as to the promoters of other cellular genes whose expression is regulated in growth dependent manner.\textsuperscript{63,64,65,66,67,68}
DNA tumor viruses, like adenovirus, must induce cells to enter S phase in order to have a productive infection. There is a large body of evidence that suggests that the E2F complex is recruited by these viruses to activate viral and cellular genes that function to establish an environment for DNA synthesis. It has also been demonstrated that the E2F complex is involved in promoting the G1 to S progression in both the unperturbed cell cycle and the growth stimulation cycle.

The transcriptionally active E2F complex is a heterodimer between E2F and DRTF1 subunits (reviewed in 69). To date there are at least four members of the E2F family, including E2F-1, -2, -3, and the YI protein. The factor DRTF1 was identified as a binding and transactivation partner with the E2F-1 subunit. Three genes have been found for DRTF1 and identified as DP-1, DP-2, and DP-3, with DP-1 being the most abundant. The heterodimer between these gene products and E2F-1 greatly enhances E2F DNA-binding and transcriptional activation capacities. To avoid confusion I shall refer to the E2F/DTRF complex as E2F, and refer to the subunits by their gene names from this point on. It is not known if these different gene products have specific functions, nor is it known whether particular DP-x/E2F-x interactions have specific functions. However, the possibility of such complex interactions has been proposed as a possible mechanism to regulate E2F activity. Cyclins A and E, and their respective cyclin dependent kinases, have also been observed in DNA bound complexes with E2F (see above). This association has the potential to anchor an active cyclin dependent kinase to the promoter.
region of genes containing E2F sites, thus concentrating a cell cycle kinase activity in the region of transcription initiation. This active kinase complex has the potential to regulate transcription by direct phosphorylation of the transcriptional machinery.

The binding and transactivation properties of the E2F complex are regulated during the G1 to S phase transition, in part by the association with Rb and other Rb-related proteins, and in part by the abundance of E2F1, which is expressed maximally during late G1-S phase. The gene product of the E2F1 is growth regulated at the late G1-S phase boundary due to transcriptional regulation of the E2F1 promoter. The E2F1 promoter, containing three CCAAT boxes, three SP1 sites and two E2F sites, is typical for a late G1-S expressed gene. Interestingly, the E2F-1 promoter is autoregulated, but mutation of the E2F sites did not completely abolish S phase regulation (see reference 72). This observation underscores the importance of cooperation between multiple promoter elements to achieve proper S phase expression.

TK

Thymidine kinase (TK) is a salvage enzyme that converts thymidine to thymidylate. In growth-stimulated cells, TK is regulated by transcriptional, post-transcriptional, and post-translational mechanisms. TK regulation in part is due to post-transcriptional events, because as cells progress through G0 to S phase transcription of the TK gene was observed to increase only 3 to 7 fold (depending on the
cell line), while the steady state levels of TK mRNA was observed to increase 20 to 50 fold. In myoblasts stimulated to differentiate, TK mRNA was shown to decrease rapidly due to changes in mRNA stability. The cis-acting sequence responsible for this has been mapped to sequences downstream of the TK promoter. Enhanced nuclear processing and export of TK mRNA, and increased cytoplasmic mRNA stability during S phase have been demonstrated to contribute to the post-transcriptional regulation of the TK gene. The mouse and human promoters have been shown to direct S phase regulation to the cat and neo reporter genes respectively, indicating that in the growth stimulation response the promoters can function independent of the downstream sequence. Regulatory elements have been located within intron two of the mouse gene. These intron elements have been shown to bind to Sp1. The promoters for the human, hamster, and mouse genes have been analyzed for regulatory elements. All three contain binding sites for the transcription factor SP1. The hamster and human promoters contain CCAAT boxes and TATA elements. In addition the human and mouse promoters contain binding sites for the E2F transcription factors. Recently, the immediate early gene Egr-1 has also been shown to bind the mouse TK promoter.

**DHFR**

DHFR catalyzes the conversion of dihydrofolate to tetrahydrofolate, and is therefore involved in the de novo pathways of both purine and thymidylate biosyntheses.
DHFR expression is growth regulated, although the mechanism affecting DHFR regulation is less complicated and less tightly controlled than TK. In continuously cycling cells there is a transient increase in DHFR gene transcription in cells entering S phase, but because of the long half life of the message this brief increase in transcription is too short to result in much change in the steady state levels of DHFR mRNA. In nutrient-starved cells DHFR expression is primarily during S phase, and is controlled primarily at the post-transcriptional level. In growth stimulated cells DHFR expression is S phase specific, but regulation is controlled at the transcriptional level. The DHFR promoter from the human, mouse, and hamster genes have been analyzed and show a notable degree of conservation. The hamster promoter is GC rich, containing four copies of a 48 bp repeat region that contains binding sites for Sp1. The promoters of the mouse and hamster genes also contain two overlapping sites for E2F, but lack TATA and CAAT boxes. The Sp1 and E2F sites are important for promoter activity, and are responsible for determining the position of transcription initiation. The mouse DHFR promoter from -270 to +20 (relative to the transcription initiation site) can direct S phase regulation of a chimeric luciferase reporter gene. Full regulation from the DHFR promoter requires an intact E2F site. The story on DHFR regulation is somewhat confusing at the moment. One group of researchers has found evidence that in the hamster promoter the Sp1 site alone is capable of conferring growth-dependent regulation, while another group has shown that the E2F site in the mouse promoter is
capable of growth regulation (see references 89, and 90). Perhaps both groups are correct, and full DHFR regulation is the combined effect of both sites.

**PCNA**

The proliferating cell nuclear antigen gene encodes a cofactor for DNA polymerase δ. In continuously dividing cells the gene is expressed constitutively in all stages of the cell cycle. In growth stimulated cells the steady state mRNA levels are much higher in S phase cells than in G0 cells. This regulated increase in expression is thought to be due to post-transcriptional mechanisms. The promoter elements responsible for expression are located within 73 bp upstream of the transcription initiation site. Further, it has been shown that the PCNA promoter alone is not sufficient for regulation; sequences within introns one and four of the PCNA gene also play an essential role in the regulation of this gene. Intron one contains a reverse CCAAT element that is conserved between human and mouse genes. This DNA element has been shown to bind proteins from nuclear extracts, and has been demonstrated to inhibit PCNA gene expression when located in the transfected minigene and when cotransfected in trans. The normal PCNA gene and minigenes containing intron 4 are expressed at very low levels in G0 cells and are expressed at much higher levels in growth stimulated cells that are in S phase. Minigenes lacking intron 4, have abnormal regulation. These minigenes show high level expression in G0, low level
expression from 4 to 16 hours after stimulation, then high level expression again after 16 hours when the cells are in S phase.\(^9\) This altered regulation is thought to be due to post-transcriptional mechanisms.

**DNA Polymerase-\(\alpha\)**

The DNA polymerase-\(\alpha\) gene product is involved in replicative DNA synthesis, by synthesizing the lagging strand at the replication fork. Although in continuously dividing cells the enzyme and mRNA levels are constant throughout the cell cycle, the subcellular localization of the enzyme is shifted from the cytoplasm to the nucleus at the G1 to S phase transition.\(^9\) In growth stimulated cells the gene expression is regulated at the RNA level. Cells in the G0 state have low steady state levels of DP-\(\alpha\) mRNA, but in growth stimulated cells that are synthesizing DNA, DP-\(\alpha\) mRNA levels was shown to increase 10 fold.\(^9\) This growth dependent regulation is thought to be due to increased transcription of the DP-\(\alpha\) gene. The promoter of the DP-\(\alpha\) gene has been cloned, and a region from -248 to +45 (numbers are relative to the transcription initiation site) has been shown to be sufficient to direct regulation of a luciferase reporter gene.\(^6\) Further analysis has indicated that multiple sites of DNA-protein interactions occur within this region, including binding elements for serum response factors (SRE), AP1, AP2, E2F, SP1, and cat box binding proteins. Deletion analysis has indicated that multiple elements are required for expression and for regulation.
Ribonucleotide Reductase

The mammalian ribonucleotide reductase (RNR) catalyses the reduction of ribonucleotides into the corresponding deoxyribonucleotides. The holoenzyme is composed of two non-identical subunits R1 and R2. At the protein level the R1 subunit is expressed at constant levels throughout the cell cycle, but the R2 subunit which is limiting and therefore the determinant of enzyme activity, is expressed primarily in S phase. In continuously cycling cells and in growth stimulated cells the mRNAs for both R1 and R2 are specifically expressed in S phase. It has been shown that in all phases of the cell cycle transcription initiation at the RNR promoter occurs at the same rate, but in non-S phase cell transcription is terminated at a unique location in intron one. S phase induction for this gene is believed to be the result of a release from a transcriptional block.

Dissertation Overview

The main focus of my dissertation research was to investigate the regulatory mechanism for TS gene growth dependent expression. In particular, I wished to identify the cis-acting elements affecting regulation. From the results of previous studies, TS gene regulation seemed very complex. It seemed possible that multiple regulatory elements, located in several different positions within the TS gene, including multiple
sites within the 5' flanking region, the presence of introns, and the polyadenylation signal.

Initially I attempted to locate regulatory elements in the 3' end of the TS gene and within TS introns. Contrary to previous experiments, I observed that not all TS introns function in the same manner for S phase regulation. I was able to demonstrate that the inclusion of introns 1 and 2 gave rise to minigene regulation that was identical to the endogenous TS gene. Minigenes with introns 5 and 6 had expression that was S phase specific, even though the induction was only 1/3 of the level of the endogenous TS gene. Further, I was able to demonstrate that introns from the human β-globin gene could function in regulation to the same extent as TS introns 5 and 6, but not as well as introns 1 and 2. In collaboration with Yunbo Ke, we have shown the presence of an intron to be absolutely required for regulation.

The weak TS polyadenylation signal was shown not to be important for regulation. I was able to demonstrate that the perfected polyadenylation signal was not utilized more efficiently than the normal TS signal, and that this mutation did not have an effect on regulation. Further, in collaboration with Yunbo Ke, it was shown that, contrary to previous results, substitution of more efficient polyadenylation signals had no effect on regulation.

Later, I focused on locating regulatory elements within the 5' flanking region. I was able to demonstrate that, contrary to previous results, no regulatory elements were
upstream of -113, and that full regulation could be observed with a promoter consisting of sequences between -113 and -75. In fact, I was able to demonstrate that elements essential for basal promoter activity are also important for regulation. I was also able to demonstrate sequence specific DNA-protein interactions occurring at the sequences that are important regulatory elements.
MATERIALS AND METHODS

**Subcloning**

Enzymes, and other reagents were purchased from major biotechnology companies.

**Bacterial Host Strains**

Bacterial hosts used in these studies include *E. coli* strains; XL1-Blue (Stratagene), DH5α (Gibco-BRL), CJ236, and HB101. Strains XL1-Blue and HB101 were grown in 2x YT media (16 gm/l Tryptone, 10 gm/l yeast extract, and 5 gm/l sodium chloride per one liter), and DH5α were grown in TB media (12 gm/l Tryptone, 24 gm/l yeast extract, 0.4% glycerol, 0.017 M KH₂PO₄, and 0.072 M K₂HPO₄). CJ236 were grown in 2x YT media supplemented with 10 μM uridine. The F' plasmid was selected for using the antibiotic tetracycline (10 μg/ml) in the XL1-Blue strain, and with chloramphenicol (25 μg/ml) in the CJ236 strain.

**Bacterial Cell Transformation**

In some cases ampicillin sensitive host bacterial cells were transformed by introducing plasmids bearing the ampicillin resistance gene to CaCl₂ induced competent
Competent cells were made by inoculating a 50ml culture with 0.5ml from a fresh overnight culture. The 50ml culture was incubated at 37°C with agitation until the cells reached early log phase (O.D. at 560nm = 0.5). The culture was chilled on ice for 15 minutes then pelleted at 4,000 rpm for 5 minutes at 4°C in a Sorval HS-4 rotor. The pellet was resuspended in 25ml of sterile 4°C, 100mM MgCl₂ solution, then repelleted as before. The pellet was then resuspended in 25 ml of sterile, 4°C, 100 mM CaCl₂. The cells were either used immediately or glycerol was added to a final concentration of 15%, and stored in 100μl aliquots at -70°C. The transformation procedure consisted of adding DNA to 100μl competent cells, and incubating on ice for 30 minutes. The cells were then heat-shocked at 42°C for 2 minutes then diluted in 1ml of 2x YT media. The culture was incubated for 1 hour at 37°C with agitation to allow the bacteria time to express the antibiotic resistance gene. Fractions of the transformation mixture were spread on LB plates containing ampicillin (50μg/ml) and incubated overnight in a 37°C incubator.

Competent DH5α were purchased from Gibco-BRL, and transformed following the protocol provided.

When high-efficiency competent cells were required (greater than 1x10⁸ transformants per μg plasmid DNA) electroporation-competent cells were used. Cells were made competent by inoculating a 1L culture of SOB media (5 gm/l yeast extract,
20 gm/l Tryptone, 0.186 g/l KCl, and 0.5 g/l NaCl) with a 1:50 dilution of an overnight culture, and grown with agitation at 37°C until the OD at 650 nm equals 0.8. The cells were pelleted in a sterile bottle using a Sorval HBS rotor spun at 3,000 RPMs for 20 minutes. The cells were resuspended in 500ml of sterile 4°C, 10% glycerol solution, then respun. This procedure was repeated two times for a total of 3 wash spins. After the final spin the supernatant was poured off and the pellet drained briefly. The pellet was dislodged from the bottom of the tube by agitation, creating a slurry of concentrated cells.

Then 80μl aliquots were transferred to sterile microcentrifuge tubes for quick freezing in liquid N₂. For transformation, 40μl of competent cells was placed in a sterile tube containing an appropriate amount of DNA (in less than 4μl), then transferred to a sterile electroporation cuvette (0.2mm gap cuvet from BioRad). The cells were electroporated at 2.5Kvolts, and 25μf capacitance. Immediately, 0.8ml of SOC media (SOB media plus 20 mM glucose, and 0.1M MgCl₂) was added, and the culture was transferred to a sterile culture tube and incubated at 37°C with agitation for 1 hour before plating on LB plates containing ampicillin.

**Oligonucleotide-Directed Mutagenesis**

For some mutations oligonucleotide-directed mutagenesis was performed by the method of Kunkel. A DNA fragment containing the target sequence was subcloned into the M13 vector mp18, or into the plasmid Bluescript KS (Stratagene). Single
stranded DNA was generated from either vector using standard procedures (see reference 101). To generate single stranded DNA from the Bluescript plasmid the helper-phage mcs M13 was used to induce the single-strand DNA replication. Uracil-containing single stranded DNA was generated from either vector using the dut lung E.coli strain CJ236. This strain incorporates deoxyuridine into the template strand instead of deoxythymidined. After synthesis the double stranded DNA is used to transform a wt E. coli strain. In the wt strain the uridine containing wt template strand is degraded, and the mutagenic strand is used to repair and replicate vector DNA. To create the site specific mutations 10 μl annealing reactions containing 20ng of phosphorylated mutagenic primer and 600ng of uracil-containing single-stranded DNA template are incubated at 70°C for 2 minutes then allowed to cool to 30°C over 30 minutes. Mutagenic strand synthesis and ligation reactions were performed in 20μl reaction volumes containing 10μl of the above annealing mix, 100mM Tris-Cl, pH8.8; 10 mM DTT; 50mM MgCl₂; 2.5 mM each dATP, dTTP, dGTP, dCTP; 5 mM ATP; 2.5 units of T4 DNA polymerase, and 2 units of T4 DNA ligase. The reactions were incubated on ice for 5 minutes, at 25 °C for 5 minutes, then at 37°C for 2 hours. A 2μl aliquot of the reaction mixture was analyzed in a 1% agarose gel to check for completion of second strand synthesis. The E.coli strain XL1-Blue was transformed with 1 to 2 μl of the synthesis reaction. If the M13 vector was used, the transformation reaction was plated in top agar, and after 20 hr. individual
plaques were picked for replicative-form miniprep analysis. If the bluescript vector was used, the transformation reaction was plated on LB-ampicillin plates. Individual colonies were picked for miniprep analysis.

Some oligo-directed mutations were generated using modifications of a standard Polymerase Chain Reaction (PCR) based mutagenesis strategy. Primers containing the desired mutation were used to amplify regions of the TS gene. To create deletions or mutations at the ends of the amplified fragment, the PCR fragment was directly subcloned back into the appropriate restriction sites within TS minigenes. To create mutations internal to the ends of the amplified fragment two step reactions utilizing four different primers were used (see reference 104). The specific conditions for each PCR reaction were the same for all reactions. Cesium chloride purified plasmid DNA was used as template DNA. The components in the 100µl PCR reactions included MgCl₂ (1.5 mM), dNTP (50 µM), Tris-HCl (20 mM, pH 8.3), 20pmol of each primer, 10ng DNA template, and 2units of Taq DNA polymerase (BRL). Reactions were overlaid with silicon oil and amplified in a Perken-Elmer Gene Amp machine using the following protocol. 95°C for 2 minutes (for the initial round only), 55°C for 30 seconds, 72°C for 45 seconds, 95°C for 30 seconds. This procedure was repeated for 25 cycles. After the final cycle, the aqueous fraction was removed, phenol-chloroform-isoamylalcohol extracted, and ethanol precipitated. The pellet was resuspended in 100 µl of TE (10mM
Tris pH 7.9, 1mM EDTA), and 5 µl was used to check the reaction product on a 1% agarose gel. If the reaction was successful 20µl samples were digested with the appropriate restriction enzyme and ligated into the appropriate sites in the TS minigene. If the PCR fragments were to be ligated as a blunt-end fragment, T4 DNA polymerase was used to repair the 3' A "tail" generated by the Taq DNA polymerase. Briefly, in a 30 µl reaction, 20µl of the PCR product, 0.1mM dNTP, mM Tris pH 7.5, 5 mM MgCl₂, and 1 unit T4 DNA polymerase were incubated at 11°C for 20 minutes. The reaction was stopped by heating to 75°C for 10 minutes, followed by phenol/chloroform extraction and ethanol precipitation.

**DNA Purification And Analysis**

To facilitate the identification and isolation of newly created subelones, miniprep analysis was performed on ligation-product transformants. Briefly, colonies were picked from plated transformation reactions and grown in 2ml overnight cultures. Miniprep DNA was isolated from 1 ml of this culture by either the alkali lysis method, or the boiling lysis method (see reference 101). After the final ethanol precipitation, the DNA was resuspended in 100µl H₂O. Either procedure yields DNA of suitable quality for restriction enzyme analysis, or sequencing.

To isolate DNA of suitable quantity and quality for minigene functional studies, DNA was prepared from one liter cultures using the alkali lysis method (see reference
The DNA was further purified by two rounds of band isolation from CsCl equilibrium density gradients. Confirmation of subclone identity and DNA integrity, was performed on all purified DNAs by restriction enzyme analysis, and in some cases by DNA sequencing.

Restriction enzyme digestions were performed on 5 µl of miniprep DNA, or 1µg of purified DNA, following the recommended conditions of the restriction enzyme used. Dideoxynucleotide chain termination sequencing was performed on 5 to 10 µl of miniprep DNA, or 5 µg of purified DNA, using a kit purchased from United States Biochemicals (USB). In some cases RNA had to be removed from the minipreparation DNA before sequencing. This was done by incubating the DNA to be sequenced with 100ng/µl RNAse A (previously boiled for 10 minutes to inactivate DNAses), for 30 minutes at 37°C. The DNA was then phenol/chloroform extracted, to remove contaminating proteins, and ethanol precipitated. Sequencing products were resolved by electrophoresis through a denaturing 6% polyacrylamide gel.

Cell Culture

Mouse 3T6 cells are immortalized contact inhibited fibroblasts that were isolated from mouse embryos. These cells were maintained on plastic petri dishes (falcon) in Dulbecco’s high glucose modified eagle’s medium (DMEM purchased from Gibco-BRL) supplemented with 10% calf serum (Colorado serum). TS deficient (ts-) V79
Chinese hamster cells were cultured in DMEM supplemented with 10% nuserum (collaborative research), and 10 µM thymidine.106

**Harvesting Cells and mRNA Analysis**

**Cytoplasmic RNA Isolation**

All solutions used for RNA isolation or analyses were made from 0.1% DEPC (diethylpyrocarbonate) treated water. All glassware was baked at 300°F for >6 hours to inactivate RNAses. Where possible, single use, sterile plastic tubes were used for RNA isolation and storage.

Total cytoplasmic RNA was isolated from tissue culture cells following a modification of a standard protocol.101 Cells attached to 100 mm tissue culture plates were placed on ice, and washed three times with ice-cold phosphate buffered saline (PBS, 140 mM NaCl, 3 mM KCl, 8 mM Na2 HPO4, 2 mM KH2PO4, pH 7.4). The cells were scraped from the plate with a rubber policeman in 3 to 5 ml of PBS, then centrifuged for 3 minutes at 2000 rpm in a Beckman TJ-6 table top centrifuge. The cell pellet was lysed in 0.4 ml of RSB (reticulocyte standard buffer, 10 mM NaCl, 3 mM MgCl2, 10 mM Tris-Cl pH 8.0) containing 0.5% NP-40 detergent by vortexing for 30 seconds. The lysates were transferred to 1.5 ml centrifuge tubes and nuclei were pelleted by centrifugation in an Eppendorf microcentrifuge at 8000 rpm for 3 minutes. The supernatant was transferred to a new 1.5 ml microcentrifuge tube containing 80 µl of 5x SDS Proteinase K digestion
buffer (0.5M NaCl, 50mM Tris-Cl pH 7.5, 5mM EDTA, and 2.5% SDS). Proteinase K was added to give a final concentration of 250 μg/ml, and the samples were frozen at -70°C.

To purify the RNA from the lysates, the frozen samples were thawed and incubated at 37°C for 60 minutes to complete proteolysis. The lysates were then extracted 3 times with equal volumes of PCI (buffered phenol, chloroform, and isoamyl alcohol mixed in a ration of 24:24:1), and once with chloroform. After the final extraction the chloroform was removed and the samples were spun down to pellet any remaining precipitates. The supernatant was carefully removed and transferred to a new microcentrifuge tube for ethanol precipitation. The RNA pellet was dried in a Savant speed vac for 30 minutes, and redissolved in 300μl of DEPC treated water. In a fresh tube 10μl of the RNA sample was diluted into 300μl of water and vortexed. The concentration of the diluted RNA was determined by measuring absorption at 260nm.

5' End Labeling And Isolation Of Probes

The generation of uniquely 5' end labeled double stranded DNA probes was accomplished following established procedures (see reference 102). Briefly, 20μg of plasmid DNA was digested with an appropriate restriction endonuclease in a 50μl reaction. After the digestion reaction, 5 units of calf intestine alkali phosphatase was added and incubated at 37°C for 60 minutes. The reaction was terminated by PCI
extraction and ethanol precipitation. The pellet was redissolved in 23μl of water, 3μl of 10x T4 polynucleotide kinase reaction buffer (final concentration is 0.5M Tris-Cl pH 7.5, 0.1M MgCl₂, and 50mM EDTA pH 8.0), 3μl [γ-³²P] ATP (final concentration of 400μCi/30μl, specific activity of 5000-7000 Ci/mmol), and 30 units of T4 polynucleotide kinase (USB). The reaction was incubated at 37°C for 60 minutes, then PCI extracted and ethanol precipitated in 0.3M sodium acetate. The pellet was dried, redissolved in 40μl of water, and digested with a second restriction enzyme that would generate a DNA fragment appropriate for the probe in question. The digested DNA fragments were resolved through a 1% agarose gel, and the radioactive bands were identified by a 3 minute exposure of Kodak XAR-5 film. The band containing the desired fragment was cut from the gel, and the DNA was eluted using either a Gene Clean kit (Bio 101), or by using an MC filter cartridge (Millipore). The incorporated radioactivity was determined by liquid scintillation counting. Generally, the specific activities were about 1×10⁷ cpm/μg DNA.

_Nuclease S1 Protection Assay_

Nuclease S1 protection assays were carried out as described in reference 102, with only minor modifications. Total cytoplasmic RNA (usually around 10μg to 50μg) dissolved in 0.3M sodium acetate, was co-precipitated in ethanol with 1×10⁵ cpm of the 5' end labeled double stranded DNA probe in 1.5ml microcentrifuge tubes. The pellet
was washed thoroughly with 70% ethanol to remove all traces of dissolved salts, then
dried, and redissolved in 30μl hybridization buffer (80% deionized formamide, 40 mM
PIPS pH 6.4, 400 mM NaCl, and 1mM EDTA pH 8.0). In order to have low lane
background it was critical to completely redissolve the pellet by repeated pipetting and
vortexing. I also found it helpful to let the pellets rehydrate in hybridization buffer for 30
minutes before pipetting. After the pellet was completely dissolved, the solution was
incubated at 85°C for 15 minutes to denature the DNA probe and RNA secondary
structures, then transferred to a 56°C water bath for DNA-RNA hybridization. The
hybridizations were carried out for 12 to 16 hours. Unhybridized DNA probe was
digested by adding ice-cold S1 digestion buffer (0.28 M NaCl, 50 mM sodium acetate pH
4.5, and 4.5 mM ZnSO₄) containing 100 units of S1 Nuclease (Pharmacia), to the 56°C
hybridization reaction. The tube was immediately transferred a 37°C water bath, and
incubated for 30 minutes. The digestion was terminated by the addition of 75μl of S1
termination buffer (4 M ammonium acetate, 20 mM EDTA pH 8.0, and 40 μg/ml yeast
tRNA), and ethanol precipitated by adding 1ml of ice-cold ethanol and incubating at -20
°C for 20 minutes. To recover the protected probe, the tubes were allowed to warm to
room temperature, then spun for 20 minutes in a microcentrifuge. The pellets were rinsed
thoroughly to remove all dissolved salts, dried, and redissolved in 5μl of denaturing
loading dye (80% formamide, 10 mM NaOH, 1 mM EDTA pH 8.0, 0.1% xylene cyanol,
Figure 2. Probes Used in Nuclease S1 Assays

The boxes indicate the minigene constructs that were used to generate the uniquely 5' end labeled DNA probes. For each probe the longest line indicates the size of the probe. The lines below the probe indicate the size of the bands that are protected in the S1 nuclease protection assays. 

A. The 1.6 kb TTT probe, made from the intronless TTT minigene, is labeled at the BglII site (Bg) and extends to the HindIII site (H) at the 5' end of the TS promoter. 

B. The STT probe is made from the intronless STT minigene in which the TS promoter has been replaced with the SV40 early promoter. This 5 kb probe is labeled at the BglII site and extends through the pUC18 vector to the SacI site (S) located at the 3' end of the minigene. 

C. The 1.4 kb TG1,2G probe is made from the human β-globin minigene, and is labeled at the BamHI (B) site that is located in the second exon. The probe contains intron one which is not present in globin mRNA, so the detected signal is protected to the intron one-exon two boundary. 

D. The 4 kb RPL-32 probe is made from the intronless RPL-32 minigene, and is labeled at the DraIII site (D) located in exon three. The probe extends through the pUC18 vector to the HindIII site located at the 3' end of the minigene. 

Other abbreviations: X, XbaI; E, EcoRI; L, Sall; A, AvaII; P, PstI; "H", HindIII site that was created in the minigene construct.
Figure 2. Probes Used in Nuclease S1 Assays
and 0.1% bromphenol blue). The samples were then denatured at 85°C, and protected fragments were resolved on a 6% denaturing polyacrylamide gel. Signals derived from various mRNAs were quantified using a Betagen Blot Analyzer. Autoradiograms were also made using Kodak XAR-5 film and Lightening Plus intensifying screens. The probes used in these studies, and the expected size of protected fragments are shown in Figure 2.

**Introduction Of DNA Into Mammalian Cells**

There are several techniques available to introduce DNA into mammalian cells. The efficiency at which cells will take up and express foreign DNA is highly cell line dependent. A method that works well in one cell line may not work at all in another. Hamster V79 cells readily take up DNA using a CaPO<sub>4</sub> co-crystallization technique, and is the method of choice for this cell line. Mouse 3T6 cells however, are somewhat refractory to this technique, so electroporation was established for these cells.

**Transfection Into V79 Cells**

The CaPO<sub>4</sub> procedure used to introduce DNA into V79 cells has been described previously, and was followed with minor revisions. Confluent plates of V79 cells were split to about 20 to 40% confluence, and were allowed to recover and reattach for about 5 hours before transfection. To transfet the cells CaPO<sub>4</sub>-DNA crystals were grown by mixing no more than 12 pmoles (about 40 μg) of sterile minigene DNA (test
minigene and control minigene), in 375µl of sterile water with 125µl of 1M sterile CaCl₂. This solution was then added drop wise to 0.5ml sterile 2x HBS (50mM HEPES, 280mM NaCl, 1.5mM Na₂HPO₄·2H₂O, pH 7.05), mixed gently, and incubated at room temperature for 20 to 45 minutes. The crystals were added drop wise into the media of the cell culture dish. The dishes were returned to the incubator and left undisturbed for 12 to 15 hours. After the incubation the transfection media was replaced with fresh media, and the cells were cultured for 36 to 48 hours before harvesting for minigene expression analysis.

Electroporation Into 3T6 Cells

The introduction of DNA into mammalian cells by electroporation, involves exposing cells, suspended in a buffered solution containing high concentrations of DNA, to a brief electric pulse. The electric current creates holes in the cell membranes through which molecules in the surrounding media can enter. The efficiency at which cells take up DNA is affected by many parameters, including, magnitude of electric field (as determined by, voltage, ionic strength, volume of the solution, distance between the electrodes, and the capacitance), the DNA concentration, cell density, and cell type. For any given cell line the optimum condition will be different, and must be determined empirically. The disadvantage of this technique is that it is extremely detrimental to cell viability, so when establishing the optimum conditions a compromise must be reached.
between DNA uptake efficiency and cell viability. I found that 3T6 cells had a broad range of tolerance to electroporation.

To establish the parameters for 3T6 electroporation I used the pCH110 vector (Pharmacia), which utilizes the control elements of the SV40 early genes to express the bacterial lac Z gene in mammalian cells. To determine the number of cells expressing the lac Z gene, I used a standard in situ staining technique. To determine the percentage of cells surviving I compared the number of cell on the plate the day after electroporation to the number of cells plated on control plates at 1/2, 1/4, and 1/10 densities.

I found that 3T6 cells had a broad range of tolerance to salt concentration from 100 to 150mM NaCl. There was a linear response to DNA concentrations from 100μg to 300μg. Less minigene DNA could be used if carrier DNA was added (either sonicated salmon sperm DNA or pUC18 DNA), and larger amounts of DNA was cytotoxic. There was a broad range of voltage-dependent DNA uptake, from 250 volts to 290 volts. However, at the upper voltage range, cell viability was dramatically reduced.

In the optimized protocol, 2x10⁷ exponentially growing cells were harvested by trypsinization, washed in 1x HeBS (20mM HEPES pH 7.05, 137 mM NaCl, 5mMKCl, 0.7mM Na₂HPO₄, 6mM dextrose), collected by centrifugation, and resuspended in 300μl of HeBS containing 100 to 200μg of minigene DNA. The cell suspension was transferred to a 0.4 cm gap electrode cuvett and electroporated at 270 V and 960 μf in a Bio-Rad Gene Pulser. The electroporated cells were then plated in cell culture media.
Under these conditions, about 20 to 30% of the cells survive, and of the surviving cells about 20 to 40% take up and express DNA. Using the CaPO₄ procedure on 3T6 cells resulted in less than 1% of the cells taking up and expressing the DNA.

*Establishing Stable Cell Lines*

Initially stable cell lines were made in 3T6 cells using the CaPO₄ procedure. Cell were cotransfected with 40 μg of minigene DNA and 1 μg pSV2neo DNA. pSV2neo minigene is used to confer resistance to the mammalian protein synthesis inhibitor G418. Prior to transfection, minigene and pSV2neo DNA were digested with the restriction endonuclease Sall, ethanol precipitated, and redissolved in 375μl of water. Following the standard transfection procedure, cells were allowed to recover for 48 hours, then split into four 100mm dishes containing 400μg/ml G418 (GIBCO-BRL) supplemented media.

In later experiments electroporation was used to establish cell lines. Typically 100μg to 200μg test minigene DNA, along with 1μg pSV2neo DNA was linearized with Scal. All minigenes used in these studies have only one site for Scal, which is located in the plasmids ampicillin resistance gene. Following digestion the samples were sterilized by ethanol precipitation, redissolved in 300μl 1x HeBS, and electroporated into 3T6 cells using the standard conditions. Surviving cells were plated into four dishes, and were allowed to recover for 24 hours. To select for cells which have stably integrated minigene DNA into their chromosomes, the cells were cultured in media containing
400µg/ml G418. The selective media was changed every third day until colonies of resistant clones became apparent (about 7 to 10 days). All colonies were pooled together and cultured as a single cell line. To avoid site of integration artifacts, all cell lines used in these studies were represented by at least 100 independent integration events. Cell lines derived from electroporation typically contain several hundred independent integration events.

**Transient Expression Assays**

To test the expression levels of minigenes in V79 cells, the test minigene DNA (12µg) was cotransfected with a control minigene (usually 1.2µg of Si5,6S). Cytoplasmic RNA was isolated from one 100mm transfected dish 36 to 48 hours after transfection, and the expression level of the two minigenes was determined in an S1 nuclease protection assay. The signal derived from the test minigene was normalized to the signal derived from the control minigene, so that variations due to transfection efficiency, RNA loading, and sample recovery could be factored out.

To test the expression levels of minigenes in 3T6 cells, test minigene DNA (usually 100µg) was co-electroporated into cells with a control minigene (usually 10µg of Si5,6S). Cytoplasmic RNA was isolated 36 to 48 hours after electroporation, and the expression level of the two minigenes was determined in an S1 nuclease protection assay. The signals from the test minigenes were again normalized to the control minigene.
Since equal mass was used in each electroporation experiment, the minigene signals had to be corrected for differences in minigene molecular weight.

**Growth Stimulation Assays**

Mouse 3T6 cells and all derived cell lines can be reversibly forced into the $G_0$ state by culturing in media containing 0.5% calf serum. These cells can be induced to reenter the cell cycle by changing the medium with media supplemented with 20% calf serum. In these studies, cells were synchronized by plating 1-3 x 10^6 cells per 100 mm dish in 0.5% calf serum. The media was replaced on the third day with fresh 0.5% calf serum media, and growth stimulated on the sixth day by replacing the media with media supplemented with 20% calf serum. Cells were harvested at 5 hour intervals following serum stimulation. Four plates were used for each time point before 20 hours and two plates each for the later time points. Cytoplasmic RNA was prepared, and an equal amount of RNA from each time point was analyzed in an S1 nuclease protection assay. Since minigenes were tagged with an internal 57nt deletion the signal derived from the endogenous gene and the integrated minigene could be determined simultaneously using the same DNA probe.

In later experiments I further improved the assay by normalizing TS signals to a signal derived from the ribosomal protein gene RPL-32. RPL-32 gene expression is not growth regulated, therefore, normalization to this signal would factor out error due to
RNA loading and sample recovery. This improvement resulted in significant reduction of experimental variation.

**DNA Protein Interactions**

**Nuclear Extract Preparation**

Nuclear extracts were prepared from 80% confluent exponentially growing mouse 3T6 fibroblasts. Extracts were essentially prepared by the method of Dignam et. al., following the modified protocol given in Ausubel et. al.\textsuperscript{110,102} Monolayer cultures were washed twice with PBS, then scraped with a rubber policeman to harvest cells from the culture dishes. From this point on all procedures were carried out at 4°C using pre-cooled equipment and reagents. The cells were transferred to a 50 ml centrifuge tube and were pelleted by centrifugation for 10 minutes in clinical centrifuge set at 70% of maximum. The pellet was resuspended in 30 ml of fresh PBS and respun. The volume of cells was estimated by measuring the volume of water needed to fill a second tube so that the water level equaled the cell pellet level in the tube. This volume was recorded as the cell pellet volume (pcv). The cells were quickly resuspended in 5 pcv of hypotonic buffer (10mM HEPES-KOH pH 7.9 at 4°C, 1.5 mM MgCl\textsubscript{2}, 10 mM KCl, 0.2 mM PMSF, and 0.5 mM DTT), and were pelleted to remove the salt from the PBS. The pellet was resuspended in 3pcv of hypotonic buffer and incubated at 4°C for 10 minutes to allow
the cells to swell. The cells were then homogenized in a glass Dounce homogenizer with ten up-and-down strokes using a type B pestle, then transferred to a 15ml conical centrifuge tube. Nuclei were pelleted in a Sorval HS-4 rotor for 15 minutes at 5000 rpm.

The volume of the nuclear pellet (npv) was measured from the graduations on the tube, and the pellet was resuspended in $1/2$ npv of low-salt buffer (20 mM HEPES-KOH pH7.9 at 4°C, 25% glycerol, 1.5 mM MgCl$_2$, 20mM KCl, 0.2 mM EDTA, 0.2 mM PMSF, 0.5 mM DTT). The total volume of sample was measured, and a $1/4$ volume of high-salt buffer (same solution as the low-salt buffer, except the KCl concentration is 1.6 M) was added in a drop wise fashion to gradually bring the final KCl concentration to 420 mM.

The nuclei were rehomogenized with three strokes to disrupt clumps, then were transferred to a 15ml conical centrifuge tube and placed on a rotating platform for 30 minutes to extract proteins from nuclei. The extracts were cleared by centrifugation in a Sorval HS4 rotor for 20 minutes at 6000 rpm. The supernatant containing soluble proteins was carefully removed so as not to disturb the pellet, and was transferred to storage vials. The aliquots were then stored in liquid nitrogen.

**Probe Preparation**

Some of the probes used in this study were generated by PCR amplification using 5' end labeled primers. To generate these probes, one or both of the oligo primers (20 pmol of each primer) was phosphorylated using T4 polynucleotide kinase and $\gamma^{32}$P-ATP.
Reactions were carried out in a 50 μl reaction volume containing oligos (20 pmol for each oligo), 400μCi of γ^{32}P-ATP, 50 mM Tris-HCl pH7.6, 10 mM MgCl₂, 10 mM 2-mercaptoethanol, and 30 units of T4 polynucleotide kinase. After a one hour incubation at 37 °C the reaction was transferred to a tube containing 10 μl of 10X PCR reaction buffer (BRL), 5 μl of 1 mM dNTP mix, 50 ng of the appropriate DNA template, an appropriate volume of water to bring the final volume to 100 μl, and 10 units of Taq DNA polymerase. If only one of the primers was end labeled then the other primer was added prior to PCR amplification. The probes were amplified 30 cycles using the following profile; 97 °C denaturation for 15 seconds, 56°C annealing for 30 seconds, and 72°C polymerization for 45 seconds. Following amplification, 10 μl of 3 M sodium acetate and 250 μl of 95 % ethanol were added, and the DNA probes were precipitated at 4°C for 20 minutes. The samples were pelleted by centrifugation, dried, then dissolved in 50 μl of low ionic strength loading dye (50mM KCl, 10% glycerol, 0.1% bromphenol blue, and 0.1% xylene cyanol), incubated at 45 °C for 10 minutes (to make sure probes were completely annealed), then loaded on an 8% polyacrylamide gel (non-denaturing) that had been pre-run for 30 minutes. The gel was run at 30 mA until the xylene cyanol had migrated 10 cm. The gel was exposed to x-ray film for 1 minute to locate the position of the probe. The gel area containing the probe was cut out and transferred to a 2 ml microfuge tube, 1.2 ml of elution buffer (0.5 M ammonium acetate, 10 mM
magnesium acetate, 1 mM EDTA (pH 8.0), and 0.1% SDS) was added, and the probes were eluted overnight on a rotating platform. Following elution, the buffer was transferred to two tubes (about 600 μl to each tube), and each tube was ethanol precipitated with 1.2 ml of 95% ethanol. The dried pellet was redissolved in 100 μl of 50 mM KCl, and was incubated at 45°C for 20 minutes to anneal DNA strands. A 1 μl aliquot of the probe was counted in scintillation fluid to determine the cpm per μl. The volume of the probe was adjusted so that 2 μl contained 1 X 10⁴ cpm.

Probes that were not generated by PCR were made in the same manner as the probes used in the nuclease SI assays (see above), except that the probe fragments were band isolated from 8% polyacrylamide gels instead of agarose gels, and were eluted from the gel in the same fashion as the PCR probes.

**Gel Mobility Shift assay**

The binding reaction in the gel mobility shift assays were carried out in 20 μl reaction volumes, consisting of 1x10⁴ cpm of the indicated probe, 20 mM HEPES-KOH (pH 7.9), 0.5 mM MgCl₂, 1 μg salmon sperm DNA as a non-specific competitor, 60 mM KCl, and 1 to 2 μl nuclear extract (about 12 to 15 μg of protein). The binding reactions were incubated at room temperature for 20 minutes before adding 2 μl of low ionic strength loading dye (see above), and loading onto a polyacrylamide gel. The complexes were resolved in a 4% polyacrylamide gel (with an acrylamide to bis-acrylamide ratio of
65 to 1), cast in 0.5X TBE (44.5 mM Tris-base, 44.5 mM boric acid, and 1 mM EDTA) running buffer. The gels were pre-run at 15 mA for one hour before loading, then run at 20 mA after samples were loaded. When the xylene cyanol dye had reached about one centimeter from the bottom, the gel was transferred to Whatman 3mm paper and dried before autoradiography.
RESULTS

Minigene Nomenclature

Mammalian minigenes, by definition are portable DNA constructs that contain all necessary signals for gene expression in mammalian cells. Minigenes typically contain promoters, coding regions, and polyadenylation signals. Minigenes used in these studies are shown in Figure 3, and Figure 12. Minigene names were designed to describe the gene expression components used in the construct. The first letter in the minigene name indicates the promoter; S indicates SV40 early, T indicates the TS promoter, G indicates the human β-globin gene promoter. The last letter indicates the source of the polyadenylation signal; T indicates the TS signal, S indicates the SV40 early signal, P indicates the perfected TS poly(A) signal, B indicates the bovine growth hormone signal, and G indicates the human β-globin signal. The letters between the first and last refer to the coding region; T indicates the intronless TS coding region, the small case i followed by numbers is used to indicate that TS introns were included and the numbers indicate which introns, G indicates the β-globin coding region including β-globin introns one and two. In minigenes with two polyadenylation signals, the last two letters refer to polyadenylation signals.
The TS Polyadenylation Signal Is Not Required For Regulation

Transient Expression Levels In Mouse 3T6 Cells.

In previous experiments transient expression was measured in ts-V79 hamster cells, while growth stimulation studies were measured in mouse 3T6 cells. V79 cells are a transformed cell line that will not growth arrest in 0.5% calf serum, and have an abbreviated G1 phase. Therefore, the G1 regulatory events in V79 cells are not the same as 3T6 cells. Because of these differences, it may be inappropriate to compare the effects of minigene alterations on gene expression observed in V79 cells with the effects on gene regulation observed in 3T6 cells. The purpose of these experiments was to determine the effect of minigene alterations on transient expression in 3T6 cells so that I could directly compare their effects on growth regulation response.

TS minigenes shown in Figure 3 were electroporated into mouse 3T6 cells, and expression levels were measured at the RNA level using the STT probe (Figure 2) in the nuclease S1 protection assay. The results of my analysis are shown in table 1. The mutation perfecting the aberrant hexanucleotide had little effect on minigene expression. Minigene TTP was expressed only about 1.2 times higher than TTT, and Ti5,6P was expressed about 1.4 times higher than Ti5,6T (see Table 1).
Figure 3. Minigenes Used in Poly(A) Signal Substitution Studies.

The construction of these minigenes has been described previously. The minigene TTT is an intronless TS minigene that contains 1kb of the TS 5' genomic flanking sequences, including the promoter (open boxes), driving the expression of the intronless coding region (dotted boxes). The TS polyadenylation signal in this minigene is contained in a 336 bp fragment from the TS 3' genomic flanking sequences (solid boxes). The Ti5,6T minigene is identical to TTT except that the coding region is interrupted with TS introns five and six (indicated by lines) located at their normal position relative to the coding region. The introns are not drawn to scale; intron five and six are 1kb and 0.8 kb in length respectively. The coding region in minigene Ti1,2T is interrupted by TS introns one and two. Intron one is approximately 1.5 kb in length, while intron two is approximately 2.9 kb. The polyadenylation signals of the minigenes in A. were replaced with the polyadenylation shown in B. The perfected (P) polyadenylation signal, is identical to the TS signal, except that the aberrant hexanucleotide (AUUAAA) has been changed to the consensus (AAUAAA). The TS signal was also replaced with the SV40 early (S (box with slanted lines)), bovine growth hormone (B (box with squares)), and the human β-globin (G (box with horizontal lines)) polyadenylation signals. Minigenes that were used in these studies lacked the 57 bp fragment located between the two BamHI sites (B) located in exon three. Other Abbreviations: H, HindIII; X, XbaI; Bg, BglII; S, SacI; E, EcoRI; "S", created SacI site; V, PvuII.
Figure 3. Minigenes Used in Poly(A) Signal Substitution Studies.
Table 1. Expression Levels of Minigenes with Substituted Poly(A) Signals

<table>
<thead>
<tr>
<th>Minigene</th>
<th>expression relative to Si5,6Tb</th>
<th>expression relative to TTTc</th>
<th>n\textsuperscript{a}</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>169 ±24</td>
<td>1.00</td>
<td>2</td>
</tr>
<tr>
<td>TTP</td>
<td>210 ±10</td>
<td>1.24</td>
<td>2</td>
</tr>
<tr>
<td>TTS</td>
<td>728 ±41</td>
<td>4.31</td>
<td>2</td>
</tr>
<tr>
<td>TTB</td>
<td>708 ±161</td>
<td>4.19</td>
<td>2</td>
</tr>
<tr>
<td>TTG</td>
<td>611 ±42</td>
<td>3.62</td>
<td>2</td>
</tr>
<tr>
<td>Ti5,6T</td>
<td>1759 ±13</td>
<td>10.41</td>
<td>2</td>
</tr>
<tr>
<td>Ti5,6P</td>
<td>2452 ±499</td>
<td>14.5</td>
<td>2</td>
</tr>
<tr>
<td>Ti5,6S</td>
<td>7907 ±143</td>
<td>46.79</td>
<td>2</td>
</tr>
<tr>
<td>Ti5,6B</td>
<td>6271</td>
<td>37.11</td>
<td>1</td>
</tr>
<tr>
<td>Ti5,6G</td>
<td>5941 ±867</td>
<td>35.00</td>
<td>3</td>
</tr>
<tr>
<td>Ti1,2T</td>
<td>6196 ±910</td>
<td>36.66</td>
<td>4</td>
</tr>
<tr>
<td>Ti1,2B</td>
<td>25775</td>
<td>152</td>
<td>1</td>
</tr>
<tr>
<td>Ti1,2G</td>
<td>15876 ±7077</td>
<td>94.94</td>
<td>3</td>
</tr>
</tbody>
</table>

\textsuperscript{a} n refers to the number of independent electroporation experiments.

\textsuperscript{b} Minigenes were coelectroporated into 3T6 cells with the Si5,6T minigene as described in the methods section. To account for variable electroporation efficiencies the expression of the test minigenes was normalized to the expression of the Si5,6T minigene. Since equal masses of test minigene DNA were electroporated, data was normalized for differences in minigene molecular weights. Data shown is the average (± observed range) of independent electroporations.

\textsuperscript{c} For the sake of direct comparison, minigene expression relative to the TTT minigene is shown. These data were calculated from column two, by dividing the expression of all minigene by the expression of TTT, which was set at 1.
Replacement of the TS 3' flanking region with either the SV40 early poly(A) signal, the human \( \beta \)-globin poly(A) signal, or the bovine growth hormone poly(A) signal led to a significant increase in minigene expression (about 3 to 5 fold). Since known enhancer elements were not included in any of the poly(A) signal substitutions, the increase in minigene expression observed with these substitutions is likely due to an increase in 3' processing of primary transcripts.

There are two possible explanations of why perfecting the TS hexanucleotide did not lead to an increase in processing efficiency. It is possible that the TS polyadenylation signal is processed as efficiently as the others, but when chimeric minigenes were constructed, a repressor element located in the TS 3' untranslated region was removed, leading to higher minigene expression. Another possibility is that the TS polyadenylation signal is not utilized as efficiently as the others, but perfecting the hexanucleotide does not increase processing efficiency because the aberrant signal is not the rate limiting step in the 3' processing reaction. Instead, the downstream GU region may be responsible for determining processing efficiency, as has been suggested for other polyadenylation signals.\(^{111}\)

To test for repressor elements located in the 3' untranslated region, I created minigenes containing an intact TS polyadenylation signal and 3' flanking region, immediately followed by the SV40 early signal (creating minigenes TTTS and Ti5.6TS). It has been shown that when two poly(A) signals of equal efficiency are located in
tandem, or when a more efficient poly(A) signal is located upstream of a weaker signal the most 5' signal is utilized almost exclusively.\textsuperscript{112} Likewise, it has been shown that when a weaker signal is located upstream of a more efficient signal, the downstream signal is utilized more frequently. If there is a repressor element located within the 3' untranslated region it should affect minigene expression regardless of which poly(A) signal is present. However, if the TS poly(A) is utilized less efficiently than the SV40 early signal, then the majority of the primary transcripts should be polyadenylated at the SV40 site, not at the TS site, and expression levels for the double poly(A) signal minigenes should be approximately the same as those for TTS and Ti5,6S.

When I transfected TTTS and Ti5,6TS into V79 cells, I observed that the double poly(A) signal minigenes expressed to the same level as TTS and Ti5,6S relative to TTT and Ti5,6T respectively. When I assayed for polyadenylation site utilization by 3' S1 nuclease mapping, I observed almost exclusive utilization of the downstream SV40 early poly(A) signal (Figure 4). The RNA from the minigenes Ti5,6T and Ti5,6P protect the probe to the TS polyadenylation cleavage site (lanes 1 and 2). The RNA from the minigenes with two poly(A) signals (lane 3, TTTS and lane 4, Ti5,6TS) protect bands that extend beyond the TS signal. The band labeled "T region" is an S1 artifact, that corresponds to S1 digestion of a hypersensitive site. At this position there is a region containing 14 U:A base pairs that is digested by S1 nuclease due to
Figure 4. Poly(A) Signal Utilization
Panel A. 3' S1 nuclease protection assay on RNA from ts-V79 cells transfected with the minigenes Ti5,6T, Ti5,6P, TTTS, and Ti5,6TS (lanes 1, 2, 3, and 4 respectively). The lower band (labeled: TS poly A) indicated the band protected to the cleavage site of the TS polyadenylation signal. The middle band (labeled: T region) indicates an S1 hypersensitive site located at a poly T region in the TS 3' flanking region. The upper band (labeled: SV40 poly A) indicates the probe protected to the cleavage site of the SV40 poly(A) signal. Both the SV40 poly(A) band and the T region band are derived from mRNA polyadenylated at the SV40 poly(A) signal. Panel B. Structure of the 3' end of minigenes containing two poly(A) signals. The open box indicates a portion of the TS gene, including the poly(A) signal and the 3' flanking region. The closed box indicates the SV40 early poly(A) signal. The probe is 3' end labeled at the BglII site in exon 6 of the TS coding region (shown below the boxes). The location of the TS and SV40 early poly(A) cleavage sites are indicated, (TS P(A) and SV40 P(A) respectively) as is the location of an S1 hypersensitive site located in the TS 3' flanking region (T region). The sizes of the expected protected fragments are shown below the probe. Abbreviations: Bg, BglII; S, SacI; E, EcoRI; “TT” the location of the S1
Figure 4. Poly(A) Signal Utilization
excessive "breathing" in the RNA-DNA hybrid. Therefore, RNA polyadenylated at the
SV40 signal protects 2 bands; the T-region band and the SV40 poly(A) band. The
exclusive use of the downstream signal is consistent with the hypothesis that the TS
poly(A) signal is utilized less efficiently than the SV40 early poly(A) signal, and that
there is no repression element located in the 3’ untranslated region.

The intron stimulatory effect observed for TS minigenes does not require an
inefficient poly(A) signal, an aberrant hexanucleotide, or the TS 3’ untranslated region.
Regardless of which polyadenylation signal was included, minigenes with introns 5&6
were shown to express about 10 times higher than the intronless minigene counterpart,
while intron 1&2 containing minigenes were shown to express about 25 to 35 times
higher (Table 1). These observations are consistent with those made in V79 cells,
although, the intron stimulatory effect is more pronounce in 3T6 cells than in V79 cells
(see introduction).

**Introns as Cis-Acting Regulatory Elements**

**Minigenes Contain All Necessary Signals For S Phase Regulation.**

When I began this project Yue Li (a former graduate student in the lab) had
observed that TS minigenes containing introns one and two or introns five and six,
displayed normal growth stimulated response when reintroduced into 3T6 cells (see
reference 25). I was able to confirm that intron one and two containing minigenes
responded normally to growth stimulation (Figure 5). The minigene Ti1,2T is a TS minigene that contains TS introns one and two located at their normal position relative to the coding region (see Figure 3). The autoradiogram (inset within the graph) is the result of a nuclease S1 analysis of total cytoplasmic RNA isolated from a serum stimulated 3T6 cell line containing the TS minigene Ti1,2T. Each lane represents RNA isolated at the indicated hours following serum stimulation. The band labeled END indicates the signal derived from the endogenous TS gene, and the band labeled MG indicates the signal derived from the integrated minigene. The radioactivity in each band was quantified in a Betagen blot analyzer (Betagen Co.). The fold increase for the endogenous gene and the minigene was calculated by dividing the signals from each time point by the radioactivity of the 0 hour time point, which was set at 1. The fold increase was then plotted vs. time of stimulation (in hours) as shown in the graph. From the autoradiogram and the graph, the Ti1,2T minigene is clearly induced with the same kinetics and magnitude as the endogenous TS gene. This result indicates that all cis-acting signals necessary for S phase regulation are located within the Ti1,2T minigene.
3T6 cells were stably transfected with the Ti1,2T minigene. The stable cell line was growth-arrested in 0.5% calf serum and then growth stimulated to reenter the cell cycle as indicated in the Methods section. Total cytoplasmic RNA was isolated from stimulated cells at the indicated times (hours after stimulation). Expression was analyzed using a constant amount of RNA for each time point (approximately 20 μg), and 1X10⁴ cpm TTT probe (shown in Figure 2) in nuclease S1 protection assays. The result of the S1 assay is shown in the inset autoradiogram. The minigene lacks a 57 bp BamHI fragment located in exon 3, therefore, the minigene mRNA protects a band that is 216 nt. long (labeled MG), while the endogenous TS gene protects a band that is 560 nt. long (labeled END). The radioactivity for each band is quantified in a Betascope Blot analyzer. In the graph, mRNA (fold increase) is plotted vs. time following serum addition. Fold increase for each gene is calculated by normalizing the radioactivity in each band by the radioactivity in the zero hour band, which is set at one. Symbols: Solid line and filled squares indicates the fold stimulation for the endogenous TS gene; Dotted line and open circles indicates the fold stimulation of the integrated minigene.
Figure 5. Growth Regulation of Ti1,2T.
When I reanalyzed intron five and six containing minigenes the results were somewhat different (the minigene is shown in Figure 3). The growth stimulation analysis of a cell line containing Ti5,6T is shown in Figure 6. From the autoradiogram and graph it is quite obvious that, while Ti5,6T responded with the same timing as the endogenous gene, the magnitude of induction is about one third. I have repeated the analysis of both Ti1,2T and Ti5,6T minigenes in three independent cell lines. Each minigene clearly responds to S phase induction, however, the magnitude of the response is consistently greater for Ti1,2T minigenes, than it is for Ti5,6T minigenes.

Introns Are Essential For Growth Regulation

Preliminary data from Yue Li and Linda Ritter on the analysis of intronless minigenes TTS and TTP had indicated that these minigenes were not regulated in growth stimulation experiments. To confirm this I thoroughly analyzed these minigenes and attempted to analyzed the TTT minigene. After many attempts I was unsuccessful in generating cell lines that expressed the minigene TTT. As indicated previously this
Figure 6. Growth Regulation of Ti5,6T

Stable cell lines were generated in 3T6 cell using the Ti5,6T minigene. The growth stimulation response was analyzed as indicated in Figure 5.
minigene is expressed at very low levels in transient assays. In stable cell lines, the minigene signal was too low to reliably distinguish from background. When I analyzed cell lines made with the minigene TTS, I was able to confirm Yue's observation that this intronless minigene is unregulated. I generated three independent cell lines with this minigene and a typical result is shown in Figure 7. The analysis of the data demonstrates that while the expression of the endogenous gene increased dramatically as cells enter S phase (around 12-15 hours), the expression of TTS does not increase significantly throughout the entire experiment. When I analyzed cell lines containing integrated copies of the intronless minigene TTP, a similar observation was made. In Figure 8 the result from one TTP cell line is shown. In this cell line the endogenous gene is induced about 24 fold while the minigene only increases about 3 fold (about 1/8 of the endogenous gene induction). In Figure 9 the result from another TTP cell line is shown. For some unknown reason the minigene is expressed at a very high level in this cell line, yet the minigene is still unregulated in response to growth stimulation.
Figure 7. Growth Regulation of TTS
Stable cell lines were generated in 3T6 cells using the TTS minigene. The growth stimulation response was analyzed as indicated in Figure 5.
Figure 8. Growth Regulation of TTP
Stable cell lines were generated in 3T6 cells using the TTP minigene. The growth stimulation response was analyzed as indicated in Figure 5.
Figure 9. Growth Regulation of TTP
Stable cell lines were generated in 3T6 cells using the TTP minigene. The growth stimulation response was analyzed as indicated in Figure 5.
Since I was unable to analyze the minigene TTT, and since the minigenes TTS and TTP contained altered polyadenylation signals, one possible explanation as to why these minigenes are not regulated is that the TS polyadenylation signal is important for regulation. To test this I reanalyzed minigenes Ti5,6S and Ti5,6P. The result of growth stimulation analysis of a cell line made with Ti5,6S is shown in Figure 10, and the analysis of a cell line made with Ti5,6P is shown in Figure 11. The data clearly demonstrate these minigenes are S phase regulated in the same manner as the minigene Ti5,6T (compare Figure 10 and Figure 11 with Figure 6). These results are in close agreement with those made by Yue Li, and indicate that the TS polyadenylation signal and 3' flanking region are not important for S phase regulation. The only difference between the minigenes TTS and Ti5,6S or TTP and Ti5,6P, is the presence or absence of introns five and six. Since the intron containing, but not the intronless minigenes, are growth regulated, it seems reasonable to conclude introns are required for regulation. Identical observations were made by Yunbo Ke (another graduate student in the lab), when he analyzed intronless, intron five and six, and intron one and two containing minigenes with the bovine growth hormone poly(A) signal and with the human β-globin
Figure 10. Growth Regulation of Ti5,6P

Stable cell lines were generated in 3T6 cells using the Ti5,6P minigene. The growth stimulation response was analyzed as indicated in Figure 5. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene.
Figure 11. Growth Regulation of Ti5,6S
Stable cell lines were generated in 3T6 cells using the Ti5,6S minigene. The growth stimulation response was analyzed as indicated in Figure 5. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene.
poly(A) signal. All intronless minigenes were unregulated, all intron five and six containing minigenes are regulated at about 1/3 the endogenous, and all intron one and two containing minigenes are regulated in nearly an identical fashion as the endogenous TS gene.

**The TS Promoter is Essential For Regulation.**

I have shown that TS introns are required for regulation, and that inclusion of introns one and two leads to stronger induction than introns five and six. Since introns are important for S phase regulation, and since most mammalian genes contain introns, one question that must be addressed is whether or not all intron containing genes are regulated during growth stimulation. To test this possibility I analyzed the expression of two intron containing genes whose products do not have a specialized function during the S phase of the cell cycle. To check the expression of an endogenous gene, I use probes in an S1 nuclease protection assay that would detect the ribosomal protein gene RPL-32. To check the expression of an integrated minigene, I established stable cell lines with a plasmid containing the human β-globin gene (GGi1,2G), and analyzed the expression of this minigene in the nuclease S1 protection assay (see Figure 12 for the description of GGi1,2G). In Figure 13, I show the analysis of both genes. In these experiments three probes were used simultaneously to detect the endogenous TS gene, the RPL-32 gene,
Figure 12. Minigenes Used in Promoter Substitution Studies

The GGi1,2G minigene was subcloned by ligating a 2.8 kb Sphl to Pstl fragment containing the human β-globin gene\textsuperscript{113} into the pUC18 cloning vector. The minigene contains the globin promoter (vertical striped box), and the globin coding region and 3' flanking region (horizontally striped boxes) interrupted by the globin introns one and two (lines between boxes). To clone TGi1,2G, HindIII sites ("H") were introduced into the globin gene 50 bp upstream of the AUG translation initiation codon and in the TS promoter 5 bp upstream of the TS AUG codon. A 1kb HindIII fragment containing the TS promoter (white box) was cloned into the HindIII sites of the globin gene. The construction of the Si5,6T minigene has been described previously.\textsuperscript{20} The solid boxes indicates the SV40 early promoter. The dotted boxes indicate the TS coding and 3' flanking regions. The lines between boxes indicates introns. The Ti1,2dT minigene is identical to the Ti1,2T minigene in Figure 3, except that in Ti1,2dT intron two consists of 100 bp of the 5' end and 75 bp of the 3' end of intron two. The interior 2.7 kb of intron two, between the HindIII and EcoRI sites, has been removed. The Si1,2dT minigene was subcloned by replacing the HindIII fragment of Ti1,2dT (between the HindIII sites located in the polylinker at the 5' end of the promoter and at the 5' end of intron two), with a HindIII fragment from Si1,2T\textsuperscript{20} (the fragment between the HindIII sites in the polylinker at the 5' end of the SV40 early promoter and the 5' end of intron two).

Abbreviations: H, HindIII; Sp, Sphl; B, BamHI; P, Pstl; E, EcoRI; X, XbaI; S, Sacl; L, SalI.
Figure 12. Minigenes Used in Promoter Substitution Studies
and the integrated globin gene (see Figure 2 for probes). I designed the DNA probes so that the protected bands would be readily distinguishable from one another when resolved on a 6% polyacrylamide gel (the expected sizes of protected bands are shown in Figure 2). In separate experiments I confirmed that each probe protected bands that did not overlap (data not shown). In the autoradiogram in Figure 13, the only band that increases in response to growth stimulation is the band corresponding to the TS gene. Since the expression of the RPL-32 gene is not regulated, I developed the analysis of this minigene as an internal negative control (see figure legend). By normalizing the signal for the endogenous and integrated minigenes to the RPL-32 signal, I am able to correct data for anomalies in RNA loading and sample recovery. The fold increase for each time point is calculated using the normalized data, and the 0 hour is set as one. The graph, like the autoradiogram, clearly indicates that the β-globin minigene not regulated in an S phase specific manner. Since both the RPL-32 and β-globin genes contain introns, the presence of introns alone is not sufficient to direct S phase regulation.

One possibility is that there is something special about the TS introns which stimulates growth induced expression. To test whether the TS introns could regulate TS gene expression independently of the TS promoter, Yue Li constructed and analyzed a TS minigene in which the 1kb TS promoter and 5' flanking region were replaced with the SV40 early promoter (creating Si5,6T). I have repeated the analysis of this minigene
Figure 13. Growth Regulation of GG1,2G

3T6 Cell lines were established with the GG1,2G minigene, and were analyzed in growth stimulation assays as indicated in Figure 5. The probes STT, RPL-32, and TG1,2G (shown in Figure 2) were used simultaneously to detect the endogenous TS gene, the RPL-32 gene, and GG1,2G minigene respectively. In the autoradiogram the band labeled END corresponds the endogenous TS gene, the band labeled RPL corresponds to the endogenous RPL-32 gene, and the MG band corresponds to the integrated globin gene. In the S1 analysis the same amount of total cytoplasmic RNA for each time point was analyzed. For each time point in the graph the signal from the endogenous TS gene and the integrated minigene is divided by the signal derived from the RPL-32 gene. Normalizing the expression of both the endogenous and integrated minigene to that of the unregulated RPL-32 gene would allow data adjustment to account for differences in RNA sample loading, and digestion product recovery errors. The fold stimulation was calculated from the normalized data instead of the raw data. Symbols: Solid lines indicate the fold stimulation of the endogenous TS gene (solid boxes) and the endogenous RPL-32 gene (solid triangles); the dotted line and open circles indicate the fold stimulation of the integrated minigene.
Figure 13. Growth Regulation of GGi1,2G
Figure 14. Growth Regulation of Si5,6T

Cell lines were established in 3T6 cells using the Si5,6T minigene, and were analyzed as indicated in Figure 13 using the STT, and RPL probes. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene; RPL, signal for the endogenous RPL-32 gene.
and my results (shown in Figure 14) agree with those made previously. In this experiment a single probe (STT) was used to detect the RNA derived from both the endogenous TS gene and the Si5,6T minigene. The probe was made from the intronless TS minigene which contained the same promoter substitution as Si5,6T. In the autoradiogram the minigene signal migrates slower than the endogenous gene, because the minigene signal protects the probe out to the transcription initiation sites of the SV40 promoter, while the endogenous TS signal is only protected to a divergent point in exon one (see Figure 2). It is clear from both the autoradiogram and the graph that the Si5,6T minigene is not regulated. Therefore, while TS introns are essential for regulation, they cannot function alone. The TS promoter must contain cis-acting sequences that cooperate with introns to direct growth-stimulated expression.

The TS Promoter Regulatory Elements Are Portable.

Whatever the mechanism, growth-stimulated regulation of the TS gene requires both introns and the promoter. Since two different sets of TS introns can cooperate with the promoter to direct regulation, it seemed a likely possibility that any functional intron could function similarly, as long as the TS promoter element was present. To test this possibility I constructed and analyzed a chimeric minigene (TGi1,2G), containing the 1kb TS promoter driving the expression of the intron containing globin gene (see Figure 12).
Figure 15. Growth Regulation of TGi1,2G

The TGi1,2G minigene was used to establish stable cell lines in 3T6 cells, and the cell line was analyzed as indicated in Figure 13 using the STT, RPL, and globin probes. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene; RPL, signal for the endogenous RPL-32 gene.
The growth stimulation response of this minigene is shown in Figure 15. From the autoradiogram it is clear that TGi1,2G mRNA is induced following growth stimulation, while the RPL-32 mRNA is not. By comparing the graph in Figure 15 to the graph in Figure 6, it is clear that the globin introns can direct a growth stimulated response similar that observed for TS introns five and six.

Since intron containing minigenes which lack the TS promoter are not regulated, and since TS promoter containing minigenes which lack introns are also not regulated, it seems likely that the TS promoter is responsible for growth regulation, but that the mechanism of regulation requires the presence of introns. The difference in growth stimulation response of minigenes with different intron combinations is still unclear and is currently being pursued by other workers in the lab.

**There Are No Unique Regulatory Elements Located Within The Interior Of Intron Two**

As demonstrated previously, the Ti1.2T minigene is induced to a much greater magnitude than the Ti5,6T minigene. Since the focus of my research shifted towards locating the regulatory elements within the TS promoter and 5' flanking region, I wanted to work with the intron one and two containing minigenes so that I could take advantage of the higher induction. Unfortunately, the size of Ti1.2T is about 9kb, with intron two accounting for about 3 kb. Since the overall size of this plasmid and the presence of many inconvenient restriction sites makes minigenes with intron two difficult to
Figure 16. Growth Regulation of Ti1,2dT

The minigene Ti1,2dT was used to establish 3T6 stable cell lines. The cell lines was analyzed as indicated in Figure 5 using the STT probe. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene.
subclone, I constructed a subclone of Ti1,2T from which most of the interior of intron two had been deleted, creating Ti1,2dT (see Figure 12). This minigene is identical to Ti1,2T, except that intron two is only about 180 bp long. The overall size of this plasmid is about 6 kb and all of the inconvenient restriction sites have been removed. When I checked the regulation of this minigene, I found Ti1,2dT maintained the same regulatory pattern as the parent minigene Ti1,2T (Figure 16). From both the autoradiogram and the graph, it can be seen that the response of Ti1,2dT is nearly identical to the endogenous TS gene, as was shown for the intact minigene Ti1,2T. The fact that Ti1,2dT has the same regulation pattern as Ti1,2T indicates that no essential cis-acting elements are located within the interior of intron two.

Since Ti1,2dT is regulated at much higher levels than Ti5,6T or TGi1,2G, it was possible that intron one and what remains of intron two might contain cis-acting regulatory elements that can function independently of the TS promoter. To test this I created and analyzed the minigene Si1,2dT (see Figure 12 for minigene structure, and Figure 17 for the growth-stimulation response). In this experiment the endogenous gene expression increases about 16 fold while Si1,2dT expression increases less than two fold. Therefore, Si1,2dT is clearly not growth regulated. From these observations it is clear that when the TS promoter is present, introns one and two do allow for higher growth-stimulated induction than intron five and six, or the globin introns. However, none of the
Figure 17. Growth Regulation of Si1,2dT

The minigene Si1,2dT was used to establish 3T6 stable cell lines. The cell lines was analyzed as indicated in Figure 5 using the STT probe. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene.
intron combinations can direct growth-stimulated induction independent of the TS promoter.

**Regulatory Elements 5' Of The AUG Codon.**

**Defining the 5' Boundary**

The TS promoter used in the preceding minigene studies extends to -1000 upstream of the AUG translation initiation codon (A = +1). The regulatory elements located within the promoter could potentially be located anywhere within this region. In preliminary studies, Yue Li constructed and analyzed intron five and six containing minigenes with 5' end truncated promoters. When the promoter was deleted to -250 (numbers are relative to the AUG codon), minigene response to growth stimulation was reduced to about one-third of the regulation seen with a -1000 promoter (Yue's analysis of Ti5,6T). When a minigene with a -150 promoter was analyzed the minigene was apparently unregulated. These results indicated that some elements seemed to be located upstream of -250, while others were located between -250 and -150. Since the core promoter had been defined as a region downstream of -113, it also seemed that the 5' regulatory elements were well outside the core promoter region.

To locate elements upstream of -250 I analyzed minigenes with the 5' end deletions indicated in Figure 18. Using intron five and six containing minigenes I found
Response to Stimulation

Figure 18. Summary of Promoter Deletions Between -1000 and -250

The position of the deletion end points are indicated by numbers that are relative to the AUG codon. The generation of these deletions has been described previously. The response of minigenes containing these deletions is shown to the right. Symbols: Solid boxes indicate the TS promoter; and open boxes indicate the first exon of the TS gene; a refers to analysis of minigenes with introns five and six; b refers to analysis of minigenes with introns one and two. Normal indicates that minigene response was the same as the minigene with the promoter extending to -1000.
that minigenes with deletion end points at -1000, -700, -500, -300, and -250 were all regulated in an identical fashion (results summarized in Figure 18). Therefore, no regulatory elements are located upstream of -250. As indicated above intron five and six containing minigenes are not induced to very high levels, therefore, the regulation assay using these minigenes is not sensitive. Due to the lack of sensitivity it is possible that by using intron five and six containing minigenes I might have missed a regulatory element upstream of -250. Since the demonstrated higher induction level of intron one and two containing minigene would allow for a more sensitive assay, I confirmed that no regulatory elements were located between -1000 and -250 in intron one and two based minigenes (summarized in Figure 18).

The results of these analyses and those of previous experiments indicated that the regulatory elements were located between -250 and -150. Within this region there are close matches to consensus binding sequences of transacting factors that are known to affect the regulation of other serum induced genes. There are two possible sites for the factor Yi located at -230 and -180. Yi has been shown to be involved in the regulation of the thymidine kinase gene. There is an eight out of nine match to the serum response element of the cFos gene (located at -204), and there is a possible Oct1 site located at -167. I created and analyzed minigenes with mutations at the binding sites shown in Figure 19. None of these mutations had a measurable effect on minigene regulation (data not shown). To further screen this region for regulatory elements, I created and analyzed
Figure 19. Mutations In The Promoter Between -250 and -150

The TS promoter sequence from -245 to -150 is shown. Numbers indicate nucleotide position relative to the A (+1) of the AUG translation initiation codon. The mutation positions are indicated by the heavy lines. The nucleotides changed in mutation are shown above the lines. The mUSYI and mDSYI mutations alter sequences that have partial homology to an element that is important for the regulation of the mouse TK gene. The mSRE and mOCT1 mutations change sequences that have homologies to the serum response element located in the cFos gene and to the oct1 element that is important for the regulation of histone genes.
intron one and two containing minigenes with deletion end points at -212, -179, and -150. None of these deletions had an effect on minigene regulation (data not shown). Since the -150 minigene was regulated as well as the -1000 minigene, the results of these growth stimulation analyses indicated that, contrary to previous data, there were no regulatory elements located upstream of -150.

To locate the regulatory element between -150 and the coding region, I created and analyzed -165 Ti1,2dT based minigenes with the mutations shown in Figure 20. Except for the mCR mutation, all mutations are located outside of the essential promoter region, which has been defined as the sequences between -105 and -75. Analysis of minigenes with these mutations has shown that only two mutations, m110 and mCR, had any effect on regulation. In Figure 21 the analysis of m115 is shown as an example of a -165 minigene with a mutation that does not affect regulation. From the figure it is clear that this minigene has a normal response to growth stimulation. The promoter in this minigene extends from -165 to -11 and contains the mutations m155, m24, and m115 (see Figure 20). The mutations m155 and m24 were used as cloning sites to generate all of the mutant promoter minigenes. Since the m115 minigene as well as other mutant promoter minigenes, which also contain the m155 and m24 mutations, are regulated normally, it is clear that the -165 deletion and these mutations have no effect on regulation. The mutation m110 lies outside of the essential promoter region and, as shown in Figure 22 the expression of this minigene displays normal S phase timing;
Figure 20. Mutations In The Promoter Between -165 and -11
The TS promoter from -160 to +1 is shown. Numbers indicate the nucleotide position relative to the A (+1) of the AUG codon. The position of mutations are shown by heavy lines. The nucleotides changed in each mutation are indicated above the lines. The previously defined essential region and transcription initiation window are shown above the sequence.
Figure 21. Growth Regulation of m115
Cell lines were established in 3T6 cells using the -165 m115 Ti1,2dT minigene, and were analyzed as indicated in Figure 15 using the STT, and RPL probes. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene; RPL, signal for the endogenous RPL-32 gene.
Figure 22. Growth Regulation of m110
Cell lines were established in 3T6 cells using the -165 m110 Ti1.2dT minigene, and were analyzed as indicated in Figure 15 using the STT, and RPL probes. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene; RPL, signal for the endogenous RPL-32 gene.
Figure 23. Growth Regulation of mCR
Cell lines were established in 3T6 cells using the -165 mCR Ti1.2dT minigene, and were analyzed as indicated in Figure 15 using the STT, and RPL probes. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene; RPL, signal for the endogenous RPL-32 gene.
however, the magnitude of induction is consistently lower than the induction of the endogenous TS gene (compare the graphs in Figure 22). Clearly the m110 mutation has an effect on regulation, and must be located on or near a regulatory element. The mutation mCR also has an effect on minigene regulation as shown in Figure 23. The induction pattern of the mCR minigene is significantly different from either the endogenous gene or the wild type minigene. Not only is the induction consistently early for this minigene, more than half of the minigene induction occurs by the 5 hour time point, but also the maximum expression is about one-third of the endogenous gene. The mCR mutation clearly affects an important regulatory element.

The region of the promoter from -165 through -100, as shown in Figure 20 has been saturated with mutations, and only those mutations in or near the essential promoter have been shown to affect regulation (m110 and mCR). To confirm that no other regulatory elements lie within this region, I used the restriction sites created in the mutations m150, m145, mSp1, m124, m115, and m110 to create minigenes with 5' deletions at -150, -145, -130, -118, -113, and -105 respectively (see Figure 20 for the location of the mutations). Analysis of -150, -145, -130, -118, and -113 have demonstrated these minigenes to have normal regulation (-113T1i1.2dT is shown as an example in Figure 24). Although minigenes with the -113 promoter do increase expression (less than two fold) at the 5 hour time point, minigene expression further
Figure 24. Growth Regulation of -113 Ti1.2dT
Cell lines were established in 3T6 cells using the -113 Ti1.2dT minigene, and were analyzed as indicated in Figure 15 using the STT, and RPL probes. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene; RPL, signal for the endogenous RPL-32 gene.
increases specifically during S phase (note the 15 hour time point), reaching maximum expression with nearly the same timing and magnitude as the endogenous gene. Therefore, this minigene is exhibits S phase specific expression. Minigenes with promoter deletions through -113 have normal regulation patterns, indicating that no unique regulatory elements are located 5' to -113.

Since the mutation m110 had an effect on regulation, it seems reasonable to predict that the deletion generated from this mutation would also affect minigene regulation. As shown in Figure 25 the promoter deletion to -105 does indeed have an effect on minigene regulation. The analysis of this minigene indicates that the deletion to -105 has a slightly more severe effect on regulation than the mutation m110. Note that the expression of the deleted minigene only slightly increases after the initial increase at 5 hours, while for the minigene with the mutation, expression did not increase until after 5 hours (compare Figure 22 and Figure 25). One possible reason for this slight discrepancy is that the flanking sequences 5' to -105, which are different between the deleted and mutated minigenes, are affecting regulation. To test if the flanking sequences might be influencing regulation in the -105 minigene, I used the mutations mSp1 (at -130) and m110 (at -105) to create a minigene that had a promoter extending from -165 through -10, but was lacking the sequences between -130 and -105. This internally deleted promoter is identical to the 5' deleted promoter to -105, except in the 5' deleted promoter the flanking sequences are derived from the plasmid, and in the internally
Figure 25. Growth Regulation of -105 Ti1.2dT
Cell lines were established in 3T6 cells using the -105 Ti1.2dT minigene, and were analyzed as indicated in Figure 15 using the STT, and RPL probes. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene; RPL, signal for the endogenous RPL-32 gene.
deleted promoter the flanking sequences are derived from the TS promoter region -165--130. It is doubtful that the sequences between -165 and -130 will have an effect on regulation themselves, since they are removed in the -130, -118, and -113 minigenes without any effect on minigene regulation. The analysis of the internally deleted promoter indicates that in this context the mutation and deletion have the same effect (Figure 26). In this analysis, the internally deleted minigene does not increase significantly at the 5 hour time point, and like the m110 mutant minigene, it reaches a maximum that is about 50% of the endogenous gene. From the combined analyses of these deleted and mutated promoter minigenes, it is clear that all 5' regulatory elements lie within the TS promoter sequences from -113 to -11. It is not possible to delete further in from the 5' end, since further deletions would remove an essential promoter element that has been shown to lie within a region from -113 to -95.
Figure 26. Growth Regulation of -165 del130-105
Cell lines were established in 3T6 cells using the -165 del130-105 Ti1,2dT minigene, and were analyzed as indicated in Figure 15 using the STT, and RPL probes. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene; RPL, signal for the endogenous RPL-32 gene.
Defining the 3' Boundary

To locate the smallest regulatory element within the region of -113 to -11, I generated the internally deleted promoters shown in Figure 27. To avoid any possible flanking sequence artifacts from the 5' end, I fixed the 5' boundary at -118, and created minigenes with promoters that contained sequences from -118 through -11, -118 through -75, -118 through -90, and -118 through -92. For comparison purposes the analysis of the -118 to -11 minigene is shown in Figure 28. As discussed above, all minigenes with the promoter region -118 to -11 had normal regulation. The growth response of the minigene with the -118 to -75 promoter is virtually identical to the response of minigene with the -118 to -11 promoter (compare Figure 29 with Figure 28), demonstrating that all 5' regulatory sequences necessary for growth stimulation response were contained within the region between -118 and -75. When I analyzed minigenes that contained a promoter extending from -118 to -90, I found that I had lost S phase specific expression. In Figure 30, the result of the analysis of one cell line containing the -118 to -90 minigene is shown. The autoradiogram and the graph show that the endogenous gene responded in a fairly typical fashion to growth-stimulation, and was induced 10 fold at 20 hours following stimulation. The minigene expression, however, is reliably detected only at the 5 hour time point. This expression pattern is not typical for an S phase gene, but is
Figure 27. The 3' Boundary Of The Regulatory Region.
Structure Of Internal Deletions To Define the 3' boundary of the regulatory region. All minigene contained introns on the deleted intron two. The 5' boundary of each minigene was located at -118. Open boxes and dashed lines represent TS sequences downstream of The AUG codon. Solid boxes represent the TS promoter region present in each minigene. Numbers indicate the boundaries of the deleted regions.
Figure 28. Growth Regulation of -118-to-11
Cell lines were established in 3T6 cells using the -118-to-11 Ti1,2dT minigene, and were analyzed as indicated in Figure 15 using the STT, and RPL probes. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene; RPL, signal for the endogenous RPL-32 gene.
Figure 29. Growth Regulation of -118-to-75
Cell lines were established in 3T6 cells using the -118-to-75 Ti1.2dT minigene, and were analyzed as indicated in Figure 15 using the STT, and RPL probes. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene; RPL, signal for the endogenous RPL-32 gene.
Figure 30. Growth Regulation of -118-to-90
Cell lines were established in 3T6 cells using the -118-to-90 Ti1,2dT minigene, and were analyzed as indicated in Figure 15 using the STT, and RPL probes. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene; RPL, signal for the endogenous RPL-32 gene.
instead more like the expression pattern of an immediately early gene, that is expressed early during G1 phase. I had difficulty establishing cell lines with this minigene. In five attempts only two cell lines expressed this minigene, and in both of those lines minigene expression was very weak and reached a maximum during early G1 phase. To confirm that the -113 to -75 region is the smallest possible regulatory region, I created and analyzed a minigene that contained a promoter from -105 to -75. The analysis shown in Figure 31 demonstrates that this minigene has the same regulation pattern as the original -105 minigene (see Figure 25). The endogenous gene is faint in the autoradiogram, but was reliably detectable on the Betagen blot analyzer. The endogenous gene expression increases about 18 fold as cells enter S phase. Although minigene expression does increase during S phase, it only increases about 5 fold (about 1/3 to 1/4 of the endogenous gene).

By comparing the results of all deletions and mutations, I can conclude that the smallest region of the TS promoter that is required for growth stimulated regulation is contained within the core promoter sequences from -113 to -75. In Figure 32, I show the sequence of the region between -118 and -75, and in addition I show the mutations m110 and mCR, the 3' deletion -118 to -90, and the 5' deletion -105 to -75. The full length sequence is necessary for regulation. Each of the sequence alterations led to abnormal or reduced regulation.
Figure 31. Growth Regulation of -105-to-75
Cell lines were established in 3T6 cells using the -105-to-75 Ti1,2dT minigene, and were analyzed as indicated in Figure 15 using the STT, and RPL probes. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene; RPL, signal for the endogenous RPL-32 gene.
<table>
<thead>
<tr>
<th></th>
<th>-119</th>
<th>-110</th>
<th>-100</th>
<th>-90</th>
<th>-80</th>
<th>-75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promoter</td>
<td>cacggattctggccggggaagtttcccagcaggaagaggcgggc</td>
<td>119-75</td>
<td>Normal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>gcta c</td>
<td>165-11 m110</td>
<td>Reduced</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>m110</td>
<td>165-11 mCR</td>
<td>Abnormal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>gct gc</td>
<td>119-89</td>
<td>Abnormal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mCR</td>
<td>105-75</td>
<td>Abnormal</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>-119</th>
<th>-110</th>
<th>-100</th>
<th>-90</th>
<th>-80</th>
<th>-75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promoter</td>
<td>cacggattctggccggggaagtttcccagcaggaagaggcgggc</td>
<td>119-75</td>
<td>Normal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>gcta c</td>
<td>165-11 m110</td>
<td>Reduced</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>m110</td>
<td>165-11 mCR</td>
<td>Abnormal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>gct gc</td>
<td>119-89</td>
<td>Abnormal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mCR</td>
<td>105-75</td>
<td>Abnormal</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 32. Summary of Essential Region Modifications Effecting Regulation. The promoter alterations effecting regulation are shown relative to their position in the essential regulatory region from -119 to -75. Terms: Normal, the regulation pattern was identical to the endogenous TS gene; Reduced, the timing of induction was the same as the endogenous gene, but the magnitude was reduced; Abnormal, both the timing and magnitude of induction are altered.
Analysis of The G1 Regulation of the -118 to 90 Promoter

Since sequences at either end of the essential region are important for regulation, it seemed that multiple elements are located within the 40 bp region. It is possible that the abnormal regulation seen with minigenes containing the -118 to -90 promoter is due to -100 element functioning alone. To investigate this I constructed minigenes with promoters shown in Figure 33. The -118 to -92 minigene contains the wild type promoter sequences from the indicated region of the TS promoter. The regulation of this minigene is identical to the regulation of the -118 to -90 minigene (Figure 34). The mutations m107 and m100 are located at the same positions as the mutations m110 and mCR respectively. The mutation mBoth contains both m107 and m100 mutations. Minigenes with either mutation m107 or m100, or minigenes with both mutations demonstrated the same G1 regulation pattern as the wild type -118 to -92 minigene (Figure 35, Figure 36, and Figure 37). These results indicate that elements located at the sites of the mutations do not function in regulating the expression of minigenes with the -118 to 92 promoters. To determine whether two tandem copies of the region between -118 to -92 would restore S phase regulation, I analyzed a minigene containing this promoter, and found that it also was regulated like a G1 phase gene and not like an S phase gene (Figure 38).
Figure 33. Mutations within the Essential Regulatory Region
The location and nucleotide changes for mutations m107, m100, and mBoth are shown
m107 changes the same nucleotides that m110 change, and m100 change the same
nucleotides as mCR.
Cell lines were established in 3T6 cells using the -118-to-92 Ti1,2dT minigene, and were analyzed as indicated in Figure 15 using the STT, and RPL probes. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene; RPL, signal for the endogenous RPL-32 gene.
Figure 35. Growth Regulation of -118-to-92 ml100
Cell lines were established in 3T6 cells using the -118-to-92 ml100 Ti1,2dT minigene, and were analyzed as indicated in Figure 15 using the STT, and RPL probes. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene; RPL, signal for the endogenous RPL-32 gene.
Figure 36. Growth Regulation of -118-to-92 m107
Cell lines were established in 3T6 cells using the -118-to-92 m107 T11,2dT minigene, and were analyzed as indicated in Figure 15 using the STT. and RPL probes. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene; RPL, signal for the endogenous RPL-32 gene.
Figure 37. Growth Regulation of -118-to-92 mBoth Cell lines were established in 3T6 cells using the -118-to-92 mBoth Ti1,2dT minigene, and were analyzed as indicated in Figure 15 using the STT, and RPL probes. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene; RPL, signal for the endogenous RPL-32 gene.
Figure 38. Growth Regulation of 2x-118-to-92
Cell lines were established in 3T6 cells using the 2x-118-to-92 Ti1,2dT minigene, and were analyzed as indicated in Figure 15 using the STT, and RPL probes. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene; RPL, signal for the endogenous RPL-32 gene.
Regulation of the TS Promoter in the Reverse Orientation.

As indicated in the introduction, the TS promoter has been shown to initiate transcription bidirectionally with almost equal efficiency. To determine if the promoter regulatory elements could lead to minigene regulation in the reverse orientation, I constructed a minigene that contained the TS promoter -165 to -11, in which the sequences from -118 to -11 were flipped in the reverse orientation (rev -118 to -11 Ti1.2dT). The analysis of a growth stimulated cell line expressing this minigene is shown in Figure 39. This minigene is regulated with the same timing and magnitude as the endogenous gene, indicating that the regulatory elements, as well as the promoter elements can function bidirectionally. To show that the same elements that are important in the normal orientation for regulation are also important in the reverse orientation, I have constructed and analyzed a minigene containing the TS promoter sequences from -118 to -90 in the opposite orientation. I could not analyze this minigene in growth regulation experiments because I was unable to establish a cell line that expressed this minigene.
Figure 39. Growth Regulation of -118-to-11 Rev
Cell lines were established in 3T6 cells using the -165(rev-118-to-11) Ti1,2dT minigene, and were analyzed as indicated in Figure 15 using the STT, and RPL probes. Abbreviations: END, signal for the endogenous TS gene; MG, signal for integrated minigene; RPL, signal for the endogenous RPL-32 gene.
Alterations of the Promoter that Affect Regulation also Affect Transient Expression.

The essential promoter region for the TS gene has been defined by 5' deletion and mutation analysis, as the sequences from -105 to -75 (see page 2). Mutations located within this region have been shown in ts-V79 hamster cells to reduce transient expression 3 to 10 fold, depending on the mutation. To determine whether the alterations that I have shown to affect regulation in mouse 3T6 cells would also affect transient expression in 3T6 cells, I analyzed the transient expression of minigenes with altered regulation by transient electroporation analysis. The results are shown in Table 2. To facilitate comparison of alterations affecting the 5' end of the regulatory region, the expression of the minigenes in the top group have been normalized to the expression of the normally regulated minigene -165 to -11, which has been set to one. The -105 deletion, and the m110 mutation both lie outside of the previously defined essential region, yet both alterations reduce minigene serum induction and reduce minigene transient expression by about half. Minigenes with the mCR mutation are expressed about 16% of wild-type indicating that promoter alterations which have a severe effect on regulation, also have a severe effect on minigene transient expression.

To facilitate comparison of alterations of the 3' end of the regulatory region, the expression of the minigenes in the bottom group has been normalized to the expression of
the minigene with the -118 to -11 promoter, which has been set to one. Deletion of the sequences between -75 and -11 had no effect on regulation, and in transient assays led to a 2 fold increase in minigene expression (compare the expression of the -118 to -11 promoter to the expression of the -118 to -75 promoter). The deletion of an additional 15 bp, to create the -118 to -90 minigene, resulted in the greatest effect on regulation, and nearly eliminated all minigene expression. The data in Table 2 indicate that this minigene is expressed about 13% of wild type. This value has a large error (100%), because the signal from this minigene was often near background in the S1 nuclease assays. Therefore, the actual expression of this minigene is likely to be even lower than 13% of wild type.
Table 2. Transient Expression of Promoter Alterations

<table>
<thead>
<tr>
<th>minigene</th>
<th>expression relative to -165-11</th>
<th>expression relative to -118-11</th>
</tr>
</thead>
<tbody>
<tr>
<td>-165 Ti1,2dT&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.00 ± 0.00</td>
<td>1.00 ± 0.00</td>
</tr>
<tr>
<td>-105 Ti1,2dT&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.48 ± 0.11</td>
<td>2.37 ± 0.12</td>
</tr>
<tr>
<td>-165 m110&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.57 ± 0.10</td>
<td>0.13 ± 0.12</td>
</tr>
<tr>
<td>-165 mCR&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.16 ± 0.05</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Minigene expression was measured at the RNA level using the nuclease S1 protection assay. The data is average of 7 experiments. Total cytoplasmic RNA was isolated from cells following transient electroporation. Test minigenes (100 µg DNA) were coelectroporated with the Si5,6S minigene (25 µg), the expression of which was used as a control to adjust data for electroporation efficiency variation.

<sup>b</sup> For comparison purposes, minigene expression is reported relative the expression of -165 Ti1,2dT, which was set at one.

<sup>c</sup> For comparison purposes, minigene expression is reported relative to the expression of -118 Ti1,2dT, which was set at one.
**Alterations That Affect Regulation Affect The Formation Of DNA-Protein Complexes.**

**DNA-Protein Interactions In The Regulatory Region**

Within the regulatory region of the TS promoter (-118 to -75), there are consensus or near consensus sequences to several trans-acting factor families including the E2F, Ets, Sp1, LSF, and Erg families (shown in Figure 40). Previous gel-retardation studies using crude nuclear extracts have shown that three DNA-protein complexes were resolved using a probe from -114 to -57 (see introduction). The ovals in Figure 40 represent the location of the three known binding sites labeled CIII, CII, and CIV. The sequence underlying the CIII oval is a very close match to the GGAAG core sequence of binding sites for members of the Ets transcription factor family. The identity of the protein that forms a complex at this site is unknown. The sequence underlying the CII oval does not match any known consensus sequence, and the protein that forms a complex at this site is also unknown. The sequence underlying the CIV oval contains regions that are close matches to binding sites for members of the Ets, Sp1, LSF and Erg transcription factor families. Purified Sp1 and partially purified LSF have been shown to bind to this sequence, although with relatively low affinity (see reference 12). The identity of the protein to the CIV site is not known, but the mobility of this protein-DNA complex is very similar to the mobility of the purified Sp1-DNA or LSF-DNA complexes.
Figure 40. Protein Binding sites within the Essential Region
The sequence of the promoter region between -119 and -75 is shown (numbers are relative to the A, of the AUG codon). The **known complexes** that are shown (CIII, CII, and CIV) are the result of previous studies (see reference 12). The **possible binding sites** were identified by sequence homology to known binding sites for E2F, Ets, Sp1, Lsf, and Erg-1 (see references 15, 168, 14, 115, and 83).
The sequences underlying the square box are a close match to the consensus sequence for the E2F transcription factor family. However no DNA-protein complex has been observed for this region.

**DNA-Protein Interaction In Promoters From Minigenes With Abnormal Regulation**

To investigate whether any of these DNA-protein complexes plays a role in the expression and growth regulation of the TS gene, I used DNA probes that contain the promoter sequences from -118 to -75 in gel mobility shift assays. As can be seen in Figure 41, when I used the wild-type probe -118 to -75, which contains the promoter sequences from -118 to -75, at least four DNA-protein complexes were resolved (complexes are labeled a, b, c, and d). This probe was generated from the -118 to -75 Ti1.2dT minigene, and in addition to the promoter sequence it contains an additional 25 bp from exon one. When I used probe -118 to -75, which contained only the promoter sequences from -118 to -75, complex-a was not present. To show that complex-a was binding to promoter sequences and not to exon one sequence, I analyzed the -118 to -75 probe which contains additional promoter sequence flanking the -75 end of the probe shown in Figure 41, but does not contain any exon one sequence. This probe forms the four DNA-protein complexes observed with the -118 to -75 probe, demonstrating that all complexes are binding within the TS promoter sequence. Since complex-a forms in
Figure 41. DNA-Protein Interactions Within The Essential Region.

Panel A. Autoradiograms of gel mobility shift assays using probes 118-75₁, 118-75₂, and 118-75₃ are shown in lanes 1 and 2; 3 and 4; and 5 and 6 respectively. In lanes 2, 4, and 6 probes were incubated as described in the methods section, with approximately 15 μg total protein isolated from exponentially growing 3T6 cells. Lanes 1, 3, and 5 are free probe lanes. The four complexes resolved (a, b, c, and d) are indicated to the left of the autoradiograms. The * symbol indicates a missing complex.

Panel B. The sequence of the probes used in panel A are shown. The numbers are relative to the A in the AUG codon (+1). The underlined sequences indicate sequences that are not found in the wild type promoter. The restriction sites located at these mutations were used to generate the probes. The 118-75₁ probe was generated as a restriction fragment from the 118-to-75 Ti1.2dT minigene, between the XbaI site at m124 and a PstI site 25 bp into exon one. The 118-75₂ probe was generated as a restriction fragment from the 118-to-75 minigene, between the XbaI site at m124 and the NheI site at m75. The 118-75₃ probe was generated as a PCR product between the primers used to generate the mutations at m124 and m75.
Figure 41. DNA-Protein Interactions Within The Essential Region.
promoter sequences, but does not bind when -75 is located at the end of the probe suggesting that complex-a is binding near -75.

To determine the effects of the mutations m110 and mCR on complex formation, I analyzed probes containing these mutations in gel shift assays (Figure 42). The probe containing the m110 mutation, like the wild type probe, forms four DNA-protein complexes with crude nuclear extract. The relative intensities of complexes-a, -c, and -d are the same with the wild type and m110 probes, however, for the m110 probe the relative intensity of complex-b is slightly reduced when compared to the other complexes (compare lane 2 with lane 4). The mCR probe is unable to form complex-b, and has a reduced relative intensity for complex-c (lane 6; affected bands are indicated with * symbols). The m14 probe contains a mutation in the sequences that were shown to abolish the binding of CII, and lane 8 shows that complex-d is not formed with this probe.

As demonstrated in Figure 41, complex-a is not present when the 3' end of the probe is very near -75, suggesting that the protein that forms complex-a cannot bind when its target sequence is located near the end of the probe. This localizes the binding site of complex-a near -75. When I analyzed the probe -118 to -90, which contains the same exon one flanking sequences used in the -118 to -75 probe, but lacks the sequences from -88 to -75 (Figure 43), complexes-b, -c, and -d, are resolved, but not
Figure 42. DNA-Protein Interactions with Mutant Probes

Panel A. The autoradiograms of gel mobility shifts using wild type, m110, mCR, and m14 probes (lanes 1-2, 3-4, 5-6, and 7-8 respectively). Lanes 1, 3, 5, and 7 are free probe lanes. Lanes 2, 4, 6, and 8 are incubated with approximately 15 μg protein from 3T6 exponentially growing nuclear extracts. DNA-protein complexes are labeled a, b, c, and d. The * symbol indicates complexes that are missing with mutant probes.

Panel B. The sequence of the wt probe is shown. The regions of the wt probe that are changed in the mutation are indicated in bold below the sequence. The wild type probe is the 118-75 probe. The mutant probes were generated by PCR using the primers m124 and m75 (the primers used to generate the 118-75 probe), and the DNA templates containing the mutations m110, mCR, and m14.
Figure 42. DNA-Protein Interactions with Mutant Probes
Figure 43. DNA-Protein Interactions With the -118 to -90 Probe.

Panel A. The autoradiograms of gel mobility shifts using the 118-75, and 118-90 probes (lanes 1-2, and 3-4 respectively) are shown. Lanes 1 and 3 are incubated without extract, and lanes 2 and 4 are incubated with 15 μg-protein of nuclear extract isolated from exponentially growing 3T6 cells. DNA-protein complexes are labeled a, b, c, and d. The * symbol indicates a missing complex.

Panel B. The sequences of the probes used in panel A are shown. The dashed line indicates the region that is present in the 118-75 probe, but is missing in the 118-90 probe. The 118-90 probe was generated as a restriction fragment between the Xbal site at m124 and a PstI site in exon one.
Figure 43. DNA-Protein Interactions With the -118 to -90 Probe.
complex-a. This confirms the location of complex-a formation to the sequences between -90 and -75.

In Figure 44, the location of the mutations and deletion used in these studies are shown relative to the location of known sites of DNA-protein interactions. The mutation m110 is located at the potential E2F site, and in gel shift assays this mutation had little effect, only slightly reducing the intensity of complex-b. The mCR mutation is located in sequences that comprise the CIII complex which is a potential Ets binding site. In gel shift assays this mutation greatly affected the formation of complexes-b and -c. The m14 mutation is located at the site of CII complex formation, and in gel shift assays it affects the binding of complex-d. The probe -118 to -90, lacks the sequences that are responsible for CIV binding as well as the potential binding sites for Ets, Erg, and LSF family members, and in gel shift assays complex-a is not formed with this probe.

By comparing the gel shift results of the mutant and deleted probes, complex-a is equivalent to CIV, complex-b is equivalent to CIII, and complex-d is equivalent to CII. Complex-c is a new complex not described in previous experiments, but since the mCR mutation affects both complex-b and -c, complex-c is the result of a DNA-protein interaction at or near the same site of complex-b (CIII) formation.
Figure 44. The Location of DNA-Protein Interactions
Panel A. The sequence of the regulatory region (-119--75) is shown. The locations of known DNA-protein complexes (CIII, CII, and CIV) are shown in the ovals, and the location of the possible E2F site is shown in the square box. The locations sequence alterations of the mutations m110, mCR, and m14 are indicated. The sequence of the -119 - 90 region is also shown.
Panel B. A summary of complexes effected by the mutations and deletion shown in panel A is shown.
Figure 44. The Location of DNA-Protein Interactions
Complex Formation In Growth Stimulated Cells

In order to correlate the DNA-protein complexes with growth regulation, I analyzed the complex formation of a wild type probe 118-751 with proteins from extracts made from exponentially growing cells, resting cells, or resting cells that had been stimulated with 20% calf serum for 17 hours prior to extraction. In comparing the complex formation in these extracts I observed that complex b formation is nearly the same in all extracts (Figure 45). In comparing the complex formation between growing and resting extracts, I observed the formation of complex-c was dramatically reduced in resting extracts as compared to the other complexes. The formation of complexes-a, -b, and -d are also slightly reduced in resting extracts, however, this reduction could be due to variation in the amount of extracted protein incubated with the probe. When comparing changes between resting versus 17 hour stimulated extracts, complex-c is greatly reduced in resting extracts and reappears in the 17 hour stimulated extracts. There was little difference observed between growing and 17 hour stimulated extracts. The slight difference in band intensities is proportional for all bands, and is likely due to different amounts of protein incubated with the probe (see figure legend).
Figure 45. Growth Stimulated DNA-Protein Interactions.
The autoradiogram of a gel mobility shift assay is shown. DNA-protein complexes that form on the 118-75 probe are labeled a, b, c, and d. The 118-75 probe was incubated with 0µg protein (lane 1), 11.3 µg protein from 17 hr serum-stimulated-resting 3T6 cell nuclear extract (lane 2), 15.4 µg protein from exponentially growing 3T6 cell nuclear extract (lane 3), and 15.6 µg protein from resting 3T6 cell nuclear extract (lane 4).
**Conserved Sequences Within the Regulatory Region**

One of the most common methods to find important regulatory elements is to search for regions of conservation between species. In comparing the sequences between the mouse and human TS promoters there is a 15 bp homologous region that is present in the mouse promoter between -77 and -91 (see Figure 46). This is the longest stretch of homology between these promoters. Using the 40 bp essential region sequence I searched for conserved elements in the Eukaryotic Promoter Database (release 35) using the blastn protocol.\(^\text{116}\) The search parameters are somewhat stringent since the smallest region of homology that would be considered a match was 13 bps. The search identified sequence homologies in the promoters of the nucleolar antigen p120 gene, mouse ferritin heavy chain gene, the rat ferritin heavy chain gene, and the mouse rpl32 gene (shown in Figure 46). Such conservation suggests a functional significance for this region.
Figure 46. Conserved Promoter Elements
Sequence homologies are shown between the mouse TS and the human TS, human nucleolar p120 antigen, mouse ferritin heavy chain gene, human heavy chain gene, mouse RPL-32 gene, human hsp70 gene promoters and the polyomavirus enhancer (see text for references). The * symbol indicates homologies. a indicates homologies that were identified manually. b indicates homologies identified using the blastn protocol. c indicates homologies identified using the signal scan software. In all searches, the entire regulatory region from -113 to -70 was used to search for homologies.
I also searched the essential region for matches to known binding site for transacting factors, using the Signal Scan software, and the results are shown in Figure 46.\textsuperscript{117} This program compares DNA sequences to a database of known transcription factor binding sites. Unfortunately, the software only recognizes a match if the sequences are identical to the sequences in the data base. Therefore, this software cannot identify potential binding sites based on divergent consensus sequences. The search is useful, however, to identify elements that have previously been shown to be important for the expression of other genes. This search revealed that the exact sequence for the weak Sp1 site (aggcggg) found in the mouse TS promoter, is also located in the hsp70 gene.\textsuperscript{118} The -85 Ets site was identified as the same binding site (aggaag) in polyomavirus enhancer for the Ets factor PEA-3.\textsuperscript{119} The -100 Ets site was identified as an EF-1A binding site (cggaagt) found in the adenovirus E1A core enhancer region.\textsuperscript{120}

Another method used to look for potential binding sites is the direct comparison of similar sequences between two promoters. When I compared the promoters of growth stimulated genes including RNR, hsp70, TK, and DHFR, the only element the mouse TS gene had in common with these other genes was an Sp1 site. The TK, DHFR, and possibly the mouse TS gene have in common an E2F site, although, the E2F site in the TS promoter is not a perfect match. Many non-growth regulated genes have an Sp1 site, including the SV40 early promoter, which has 4 sites, suggesting that Sp1 cannot be important for regulation. However, when I compared the TS essential region to the mouse RPL-32 gene promoter, I found two regions of similarity. Both promoters contain
a gccggaagt element that corresponds to the -100 Ets site or the CII binding site (see Figure 46). Both promoters also contain the element identified above (aggaagagccgg) that corresponds to the region in the mouse promoter from -76 to -90. As I will discuss below, both of these elements cooperate in the TS promoter to express and regulate the TS gene. However, in the context found in the RPL-32 gene these elements cannot specify S phase regulation. Therefore, because an element like Sp1 is also located in unregulated genes does not rule out the possibility that it can participate in regulation in another promoter context.
DISCUSSION

The studies described above clearly demonstrate that TS gene regulation requires both the TS core promoter and introns. Precise details of the mechanisms regulating TS gene expression are not known; however, as shown by nuclear run-on experiments transcriptional regulation accounts for one third (or less) of the total regulation and post-transcriptional events account for the remainder. RNA metabolism studies have indicated that the cytoplasmic stability of TS mRNA does not vary between resting and growing cells. The post-transcriptional regulation of the TS gene must, therefore, occur before the TS message is transported from the nucleus. Since the intronless minigenes did not demonstrate any regulation, both the transcriptional and post-transcriptional components of TS regulation appear to require the presence of introns. Further evidence for cooperation between promoter elements and introns, is the lack of regulation from reporter minigenes that use the TS promoter to drive the expression of either the CAT or Luciferase genes. Neither of these expression vectors contain an efficiently spliced intron.

There are several possible mechanisms that could account for intron dependent regulation. These include regulated transcription (either activation, repression, or attenuation), regulated hnRNA processing, or regulated hnRNA stabilization. These
mechanisms all require cis-acting elements to be located within introns. Possible cis-acting elements for these mechanisms include promoter elements for transcriptional activation, or elements for transcriptional silencing, elements for transcriptional attenuation, elements such as splicing enhancers, and elements for hnRNA stabilization. The PCNA gene is an example of a gene that is regulated by cis-acting sequences located in introns. A transcriptional repressor element is located in intron one of that gene, and the G0 down regulation of PCNA gene expression is accomplished by an unknown mechanism that is controlled by a cis-acting element located in intron four of that gene (see page 22). However, neither intron is necessary for growth dependent induction at S phase.

Transcriptional attenuation or transcriptional pausing has been shown to regulate the expression of a few cellular genes including cMyc, cMyb, cFos, adenosine deaminase (ADA), and ribonucleotide reductase (RNR). The expression of each of these genes is growth dependent, and the expression of cMyb, ADA, and RNR are late G1-S phase specific. The cis-acting elements that are required for transcriptional block are located in exons of cMyc, cFos, cMyb, and ADA, and in the first intron of RNR. The mechanism governing attenuation is not well understood, but as a result of attenuation short RNA transcripts are generated that are not polyadenylated and will usually fit into a stem loop structure.
There are few examples of genes that are regulated at the level of hnRNA stabilization, or enhanced RNA processing. It is possible that the G0 down regulation of the PCNA gene, which has been shown to require intron 4, is accomplished in this manner (see reference 95). The TK gene has also been shown to be regulated at the level of processing in growth-stimulated cells (see reference 74). Immediately following growth stimulation several intermediate spliced forms of TK hnRNA have been shown to accumulate in the nucleus. As cells progress through S phase the unprocessed TK transcripts are converted more rapidly to mRNA. The cis-acting sequences and the mechanisms controlling this process are not known, but because splicing intermediates are involved introns must be important for regulation.

The main focus of my research was to locate and identify cis-acting elements that are important for the growth dependent expression of the TS gene. The results of this research clearly demonstrate that TS regulation does not require the TS coding region, the 3' genomic flanking region, or the weak TS polyadenylation signal. I was able to demonstrate that TS regulation requires the cooperation of the promoter and the introns. I observed that TS introns one and two will direct minigene regulation that is identical to the endogenous TS gene regulation, but that other introns can also allow S phase regulation. Further, I determined that the elements in the TS promoter that are required for regulation coincide with core promoter elements. The significance of these
observations and a possible mechanism by which these diverse elements cooperate to express and regulate the TS gene are discussed below.

A Weak Polyadenylation Signal Is Not Essential For Growth Regulation

Previous experiments have suggested that the TS gene may be regulated at the level of polyadenylation (see page 6 and reference 24). Through pulse labeling experiments it was demonstrated that in G0 cells, 70% of newly transcribed TS RNA was not polyadenylated, while in S phase cells only 30% of the TS RNA lacked a poly(A) tail. Further experiments also demonstrated that the TS polyadenylation signal was utilized inefficiently. Taken together the two results suggested that modulation of the efficiency of the weak TS polyadenylation signal could account for S phase regulation. To test for the possibility that a weak polyadenylation signal was important for S phase regulation, the TS polyadenylation signal in Ti5,6T was replaced with the SV40 early signal (creating Ti5,6S), and with the perfected TS polyadenylation signal (creating Ti5,6P).

In previous analysis, Ti5,6S and Ti5,6P minigenes responded to S phase induction, with timing similar to the endogenous TS gene; however, the maximum expression for these minigenes was shown to have increased only one half to one third of the induction observed for the endogenous gene (see page 6). Since the original studies on Ti5,6T expression indicated that this minigene was stimulated to the same extent as
the endogenous gene, it appeared that the weak polyadenylation signal was contributing to minigene regulation.\textsuperscript{25}

When I reanalyzed the transient expression of minigenes with modified poly(A) signals, I found that the perfected mutation did not increase minigene expression, although other poly(A) signal substitutions did. With this in mind it seemed puzzling that a modification which did not affect expression could effect gene regulation. To clarify this I reanalyzed the growth stimulation response of the modified minigenes. I observed the same regulation pattern for modified minigenes that was demonstrated in previous experiments, i.e. that minigene response was 1/3 of the endogenous gene. Contrary to previous experiments, reanalysis of the growth stimulation response of the wild type Ti5.6T minigene demonstrated that the wild type minigene, like the modified minigenes, was also induced to about 1/3 of the endogenous gene. Therefore, the poly(A) signal modifications had no effect on minigene regulation, demonstrating that the weak polyadenylation signal and the TS 3' flanking region are not required for regulation.

In doing these experiments two major improvements were developed in the experimental procedures which allowed me to make more quantitative measurements of minigene response to growth stimulation. These improvements could account for the differences between my observations and those of Yue. First, was the development of electroporation as a means of introducing DNA into 3T6 cells. Electroporation allowed
me to generate stable cell lines represented by several hundred to several thousand independent integration events. In contrast, the CaPO₄ transfection protocol typically yielded no more than 30 to 50 independent integration events per cell line. Since the cell lines used in these studies are mass cultures of all integration events, a larger pool of different integrations reduces the possibility of site of integration artifacts. A second improvement was made in the way data was collected from S1 analysis. In previous experiments band intensities from nuclease S1 assays were measured by densitometric scans of autoradiograms. Films were scanned using a light densitometer, and peaks from the paper tracings were cut out and weighed. This method is severely limited by the linearity of film image intensity versus radioactivity. In addition, the cutting out and weighing of paper is inherently error prone. In my experiments, radioactivity for each band in the S1 nuclease protection assay is directly quantified in a Betagen blot analyzer. This instrument is able to detect radioactive decay in a linear fashion over a very broad range, so data are not compromised by the limitations of autoradiographic measurements.

The Significance of Introns as Regulatory Elements.

When TS introns one and two are present minigene induction during growth stimulation is comparable to the endogenous TS gene. Other introns, including TS introns five and six, or the human β-globin gene introns one and two also function in an S
phase specific induction, but at one third the level of TS intron pair one and two. Since TS introns one and two give rise to a greater minigene induction than any other intron, it is likely that the regulation observed with this minigene reflects both the transcriptional and post-transcriptional regulation seen in the endogenous TS gene. The reduced induction of minigenes containing other introns suggests that one or more components of the endogenous TS gene regulation is reduced when introns one and two are not present. Whether the reduction is in the transcriptional regulation or post-transcriptional regulation is not known.

The difference in induction between intron pairs suggests that a unique regulatory element is located within introns one or two. In the Ti1,2dT minigene I deleted all but 200 bp from intron two without affecting regulation. Therefore no regulatory element is located within the deleted interior of intron two. It is possible, however, that regulatory elements might be located within the remaining 200 bps. Observations made by Martin Korb indicated that intron one by itself exhibited minigene regulation that was comparable to introns one and two combined. When he deleted most of the interior of intron one he observed a dramatic reduction in minigene S phase response. These observations suggest that intron one might contain a unique regulatory element. The significance of this observation is currently being investigated.

Different introns can contribute to TS regulation, even introns from the globin gene. Since the globin gene is not an S phase gene, S phase regulatory elements located
in globin introns would have no function, and would not be conserved. Therefore, the only elements TS introns and globin introns are likely to have in common are intron splicing signals. This implies that TS regulation might involve hnRNA processing. Since hnRNA metabolism is a process vs. degradation mechanism, any transcript that is not processed quickly is degraded within the nucleus. Clearly, a potential post-transcriptional regulatory mechanism would be to regulate the rate of hnRNA processing. Such a mechanism would require intron splicing signals which any functional intron could supply.

Nuclear run-on assays demonstrated that during growth stimulation transcription of the TS gene increased 2 to 3 fold as cells entered S phase. In those experiments it was not determined whether the transcriptional component of TS regulation was due to regulated transcriptional initiation, as has been suggested for TK and DHFR, or due to attenuation, as has been shown for RNR and adenosine deaminase (ADA) (see pages 19, 20, 24, and reference 124). As indicated above introns one and two might contain a regulatory element not found in other introns. If there is a unique element located in intron one, then this element could function either as an enhancer to regulate transcriptional initiation or as a transcriptional block similar to the one located in the first intron of the RNR gene.
Mechanisms For Cooperativity Between The Promoter And Introns.

Most mammalian genes contain introns, yet only a select few of these genes are induced during S phase following growth stimulation. In the results presented above minigenes were fully regulated only when the essential region of the TS promoter was used to drive expression, and then only if introns one and two were also present. Other introns could also function, but substitution with these introns led to reduced induction. Substitution of the TS promoter with the SV40 early promoter or with alterations of the essential promoter region also resulted in loss of regulation. These observations clearly demonstrate that both the transcriptional and post-transcriptional regulation seen in the TS gene is the result of cooperation between promoter elements and introns. Similar observations have been reported for the human TS gene.\textsuperscript{126} Since TS post-transcriptional regulation requires transcription from the TS promoter and involves events in RNA metabolism, the trans-acting factors interacting with the promoter elements somehow must communicate with the trans-acting factors responsible for hnRNA metabolism. Since introns in addition to the promoter appear to be required for the transcriptional component of regulation, elements located in introns (either as DNA or RNA elements) must communicate with factors binding to the promoter elements.

This phenomena of cooperativity between promoter elements and downstream elements is unprecedented in the regulation of other S phase genes. The promoters alone of TK, DHFR, and DNA Pol-\(\alpha\) genes, are sufficient to regulate minigene expression
when linked to reporter genes such as CAT or luciferase (see pages 20, 23, and 19). For the PCNA gene it has been reported that the promoter is responsible only for basal transcription, and that the introns alone are sufficient for regulation (see page 22). For the TS gene, however, the promoter elements are not sufficient to confer regulation of CAT, luciferase, or to intronless TS minigenes (see page 6).

There are, however, several examples of communication between promoter elements and RNA metabolism. First, the elements that are required for transcription initiation of the human U1 and U2 snRNA genes are also required for correct processing at the 3' end of the RNA molecules. Substitution with pol II promoters resulted in incorrect processing. This demonstrates that by specifying an RNA polymerase complex, promoter elements can determine RNA processing events. As a second example, some class II genes are not processed efficiently when transcribed by either RNA polymerase I or III. This is another example of promoter elements affecting RNA processing events by specifying the appropriate transcriptional complex. Since in both examples inappropriate RNA processing occurs even though the appropriate cis-acting (RNA) elements are present, this suggests that cis-acting RNA sequences are not sufficient to determine RNA fate. Instead, it is likely that the transcriptional complex plays a direct role in determining the fate of RNA processing. Direct communication between RNA processing factors and transcriptional complexes is inherent in this hypothesis.
There are three examples of RNA binding proteins interacting with the transcriptional machinery. When binding to its RNA recognition element the HIV Tat protein is able to transactivate pol II transcription. The lupus autoantigen La protein is an RNA binding protein that interacts with the pol III transcriptional complex to affect the efficiency of pol III transcription and termination. In Xenopus the 5S RNA transcription is regulated by the RNA-binding transcription factor TFIIIA. A third example of communication between promoter elements and RNA metabolism is that efficient expression of the immunoglobulin μ mRNA normally requires introns. Intron dependency is eliminated when the gene is expressed by the cytomegalovirus or heat shock promoters. Intronic dependent gene expression for the μ gene is believed to be due to post-transcriptional events. If this is true then this observation is an example of specific pol II promoter elements affecting class II RNA processing events. As a fourth example, promoter elements have been shown to act in trans to effect the export of mRNA from the nucleus in microinjected Xenopus oocytes. Finally, nonsense mutations in exon one of the human β-globin gene have been shown to down regulate the cytoplasmic accumulation of β-globin mRNA. The mechanism of the downregulation is not known, but has been shown to be post-transcriptional and to take place in the nucleus. When the globin promoter was replaced with the herpes simplex virus TK gene promoter, post-transcriptional down regulation was lost, further demonstrating that mRNA metabolism can be influenced by promoter elements.
The mechanisms by which promoter elements control post-transcriptional processing is not known; however, recent observations concerning transcription and RNA processing have suggested several possibilities. First, DNA bound factors might directly attract RNA processing factors to the site of transcription. Second, promoter elements might mark or tag transcripts during transcription initiation by altering hnRNA packaging factors. Finally, transcription factors might modify the RNA pol II complex during transcription initiation. As the modified complex is transcribing the gene it might serve as an intermediate between transcription and processing factors (see below).

Attraction Of Processing Factors

Perhaps the simplest mechanism by which promoter elements might control post-transcriptional processing is the direct attraction of RNA processing factors to the site of transcription. In this model when DNA binding factors are bound to their respective promoter elements they activate transcription by attracting transcription factors, and activate post-transcriptional events by attracting processing factors. The increased local concentration of processing factors would facilitate enhanced processing of newly synthesized transcripts. This hypothesis is supported by recent observations. Upon infection with adenovirus, or after transfection of a plasmid expressing the β-tropomyosin gene, the localization of RNA pol II, hnRNP C, and splicing factors (snRNPs and SC-35) were shifted from their normal nuclear distribution to the sites of
adenovirus or tropomyosin gene transcription. As a model the authors suggest that at the time of transcription initiation, transcription, packaging, and processing factors are co-recruited to sites of active transcription. The mechanism of this coordinated recruiting is not known. It is possible that the DNA bound transcription factors that can attract RNA pol II to the site of transcription initiation can also attract hnRNA packaging and processing factors. This possibility is supported by the observation indicated above where correct RNA processing is determined by the type of polymerase transcribing the gene (i.e. pol I vs. pol II vs. pol III). Since transcription by the correct polymerase is determined by the promoter elements, correct processing may also be determined by promoter elements.

**Transcript Packaging**

An indirect method by which promoter elements could affect post-transcriptional events involves directing the attachment of specific packaging factors like cap binding proteins and hnRNP to the primary transcript, thus marking them for regulated processing. As nascent transcripts emerge from pol II they are immediately bound by hnRNA binding proteins to form a large hnRNP complex. A large number of unique hnRNP have been identified, and each appears to bind RNA with a loose sequence specificity. Due to differences in RNA sequences, specific transcripts are bound by a subset of all possible hnRNPs. Some hnRNPs compete with splicing factors to bind
intron splice signals. The hnRNP C protein, for example, binds to the 3' end of introns and blocks splicesome formation by preventing U2 snRNP from binding to the branch point. hnRNP C has also been shown to participate in the polyadenylation reaction. The ability of hnRNP C to bind RNA and block splicing has been shown to prevent splicing during mitosis, and is regulated by phosphorylation. This establishes the possibility for S phase specific regulation by an hnRNP. If this hypothetical mechanism were to play a role in the regulation of the TS gene, then the promoter elements might participate in regulation by signaling which transcripts are to be bound by this S phase hnRNP. It is important to note, however, that an S phase counterpart to hnRNP C has not been found, and promoter elements have not yet been determined to effect hnRNP associations with transcripts.

Another alteration of hnRNA packaging is to modify the 5' end of the transcript. Immediately following transcription of the 5' end of the transcript, a cap structure is added. This cap structure is bound by cap-binding proteins which have been shown to participate in RNA metabolism by protecting the transcript from 5' exoribonucleases, by defining the 5' end of the first exon in mRNA splicing, by facilitating nucleocytoplasmic transport, and by facilitating translation initiation. Alterations of the cap-protein complex could therefore have considerable effect on mRNA metabolism. DNA bound transcription factors are in close proximity to the cap site and could potentially interact with capping enzymes, or could modify the proteins binding to the
cap structure. By either of these mechanisms transcription factors could participate in transcript processing through tagging the 5' end of the transcript.

**RNA Pol II As An Intermediate Between Transcription And Processing Factors**

A third possible mechanism by which promoter elements could affect post-transcriptional events involves the RNA polymerase II complex as a physical link between transcription and splicing. The pol II complex must interact first with the DNA bound general transcription factors and activators before initiating transcription. Promoter elements have been shown to effect transcriptional activity by modifying the phosphorylation state of the carboxy terminal domain (CTD) of the largest subunit of RNA pol II, which in the mouse has 52 phosphorylation sites, and by specifying the association of elongation factors to the transcriptional complex. For most promoters the non-phosphorylated form of pol II preferentially associates with the preinitiation complex and is phosphorylated at the time of promoter clearance. In the transcription of other promoters, initiation can proceed without phosphorylation of the CTD. In in vitro transcription reactions of the DHFR promoter, the E2F site is responsible for the requirement of the phosphorylation of the RNA pol II complex for initiation. RNA pol II complexes that lacked the CTD could initiation transcription from promoters lacking the E2F site.
The possibility that some promoters can initiate transcription without phosphorylation of the RNA pol II CTD has been suggested as a mechanism to coordinate transcription with splicing.\textsuperscript{157} The Drosophila hsp70 gene is transcribed primarily by the unphosphorylated form in the uninduced state, but is transcribed by both forms of pol II following heat shock.\textsuperscript{158} In the uninduced state the hsp70 gene is only transcribed to a transcriptional block site in the 5' end of the gene. Upon heat induction the transcriptional block is relieved, and transcription proceeds throughout the entire gene. The heat-shock response of hsp70 has been shown to result from a factor binding to a cis-acting element located in the promoter, therefore, the interpretation of these experiments is that a trans-acting factor binding to a promoter element determines the form of RNA pol II transcribing the gene, and the form of the pol II complex determines transcriptional elongation. In situ immunolabeling has also shown that only after heat shock, when transcription of the hsp70 gene is predominately by the phosphorylated form of pol II, are hnRNP proteins and splicing components co-localized to the site of transcription. Before heat induction, transcription is by the unphosphorylated form of pol II and processing factors are not co-localized.\textsuperscript{158} Therefore, the phosphorylation state of the transcriptional complex, has postulated to facilitate mRNA processing by attracting packaging and processing factors to the sites of active transcription. By this mechanism promoter elements participate in both transcriptional and post-transcriptional regulation through modifications made to the transcriptional complex.
It is interesting to note that like the Drosophila hsp70 gene some growth stimulated genes are also regulated by transcriptional pausing. These genes include ADA, RNR, Fos, cMyc, and cMyb. There is some evidence that the regulated transcriptional block of mammalian genes is accomplished by a mechanism similar to the one found in Drosophila. Promoter elements including TATA box, and the inducer-of-short-transcripts (IST) element have been shown to be responsible for generating nonprocessive transcription of the HIV-1 genome. In the Drosophila hsp70 gene a GAGA element is responsible for producing short transcripts. The serum induction of the Fos, cMyc, and cMyb genes requires promoter elements to specify transcriptional elongation.

Recent observations have demonstrated that DNA bound transcriptional activators increase both the rate of initiation and the efficiency of elongation. Clearly, at the time of initiation promoter bound activators are able to modify the processivity of the transcriptional complex. Although it is not known what modifications are made to the polymerase in mammalian cells, phosphorylation of the pol II CTD, and association of additional factors with the polymerase complex have been suggested (see above).

**Cooperation Between Regulatory Elements Within the Essential Promoter Region.**

The studies presented above clearly demonstrate that both transcriptional and post-transcriptional regulation of the TS gene requires transcription initiation driven by
the core TS promoter. The core promoter of the mouse TS gene has been defined by extensive deletion and mutational analyses to occupy a 40 bp region from -115 to -75 relative to the translational start site (see page 2). This region of the promoter contains all cis-acting elements necessary to initiate transcription at levels comparable to promoters with more than 1 Kb of additional sequences.

**Organization Of The Core Promoter**

The in vitro binding studies presented in previous work and in this study show that there are four DNA-protein complexes that form in the essential region when excess probe is incubated with crude nuclear extract. The four complexes have been mapped to three elements within the core promoter, and include the -100 Ets site (labeled CIII), an unknown factor site at -95 (labeled CII), and the GC box at -80 (labeled CIV in Figure 44).

By sequence analysis the core region has been shown to contain potential binding sites for a number of transcription factors, including E2F, Ets (EF-1A, Erg-1, and PEA-3), Sp1, and LSF (see Figure 41, and Figure 46). The potential binding sites for many of these factors overlap so that it would be impossible for the two factors to bind simultaneously. The factors binding the promoter in vivo would compete for the available binding sites. The relative concentrations, binding affinities, and binding kinetics of binding factors would determine which factors will bind the promoter. One way to assure
the proper DNA-protein associations is through protein-protein interactions. Ets factors and Sp1 are known to act synergistically with each other and with other factors to activate transcription. As I will discuss in more detail below, synergism between factors binding to the core promoter is likely to be extremely important for TS expression and regulation.

A DNA-protein complex has not been observed for the E2F site in the TS promoter from either crude extracts or with partially purified E2F-1. In Dr. Farhnam's lab, under identical conditions that allowed E2F-1 to bind the E2F site in the DHFR promoter, E2F-1 did not bind the site in the TS promoter. From these results it is possible to conclude that if the mouse site is a true binding site for E2F-1 then it is a relatively low affinity site. As mentioned in the introduction, three E2Fs have been identified, and it is possible that one of the other E2Fs can bind the site to the TS promoter. From the gel-shifts of mutant probes no DNA-protein complex can be assigned to the E2F site. The mutation of the E2F site did, however, slightly reduce minigene regulation and expression. These observations can be explained by the reduced affinity of the protein binding to the adjacent -100 Ets element when the m110 mutation is present. The two sites are close enough to overlap, so the base changes of the m110 mutation are likely to interfere with CIII formation.

The CII complex observed in the gel shifts forms over an element that was not identified in any of the searches. All TS promoters used in this study contained the CII
binding site, including the altered promoters that dramatically affected expression and regulation. Therefore, if this element serves a function in the TS promoter it does not function alone to express or regulate the TS gene. In previous experiments, mutations of the binding site for CII had no effect on promoter activity (see references 13, and 12). The significance of this element and the factor that binds to it are not known at this time.

The GC box at -80 in the core promoter has been shown to be a weak binding site for Sp1. However, mutation of this weak site had a more severe effect on expression than did a mutation of a stronger Sp1 site located at -130. Since purified Sp1 had a 10 fold higher affinity for the -130 site, the implication of these results is that the location of Sp1 binding is more important for promoter activity than the relative affinity of Sp1 for a particular site. One reason for the importance of binding location is that protein-protein interactions between DNA-bound transcription factors is essential for transcription initiation. Synergistic activation by Sp1 has been shown to be important for activation of other promoters. The significance of Sp1 synergistic activation will be discussed in more detail below. Sp1 is a constitutive factor that is present in almost every mammalian cell type. The transcriptional activity of Sp1 has been suggested to be regulated by Rb and possibly by phosphorylation.

There are two GGAAG boxes in the core promoter, one located at -100 (CIII) and the other located at -85 (see Figure 44). DNA-protein interactions have been observed for the -100 site, but not for the -85 site. Nevertheless, the mutation of the -100 and -85
sites reduced minigene expression to 16% and 30% respectively.\(^9\) The sequence of these sites matches the core binding sequence for members of the Ets transcription factor family.\(^{168}\) The Ets family members are a diverse group of factors that share a DNA binding domain consisting of an 84 amino-acid sequence. Ets factors bind to DNA as monomers, but some members require tandem binding sites for transcriptional activation due to protein-protein interaction between Ets factors. Ets members are known to act synergistically with other transcription factors, including AP-1 and PEA-1, to activate transcription.\(^{169,170,171,172,173}\) The relevance of this will be discussed in more detail below. The activity of Ets factors is regulated by phosphorylation in response to growth modulators. A sub-family of Ets factors including SAP-1, SAP-2, Elk-1, and NET have been shown to be the ternary complex factors that are required by the serum response factor to activate transcription following serum stimulation.\(^{174}\) Because of their demonstrated response to mitogens, Ets factors are strong potential candidates for the factors regulating the TS gene.

**cooperativity between elements in the essential region.**

Mutations of the CIII site, the GC-80 site, and the -85 Ets site have previously been shown to reduce gene expression. In this study and in previous studies the mutation mCR, which eliminates CIII binding, reduces expression by a factor of ten. This mutation has also been shown to alter minigene regulation. Together these results indicate that the
-100 Ets element and possibly the CIII DNA-protein interaction are essential for expression and regulation of the TS gene. This element alone is not sufficient, however, for either expression or regulation since the -118 to -90 minigene (which retains the -100 Ets site, but lacks both the GC-80 and the -85 Ets sites) has altered regulation and is expressed at extremely low levels. These results clearly demonstrate that multiple elements within the essential region act synergistically to express and regulate the TS gene. Interestingly, in separate mutations of the GC-80 box and the -85 Ets site, each mutation reduced expression only by 3 fold (see reference 16). That the deletion which removes both elements has a more severe effect than the single mutations suggests cooperativity between these elements. The conservation indicated in Figure 46, combined with the fact that this region spans only 15 bp suggest that the -85 Ets site and the GC-80 site function as a single element in a variety of genes. A similar arrangement of two Ets sites and an Sp1 site has been observed in the LTR of the HTLV1 genome. Cotransfection studies with minigenes containing the HTLV1 promoter and plasmids expressing Ets-1, have shown that Ets-1 can transactivate the HTLV1 promoter only when both Ets binding sites and the Sp1 site are intact. Further, it was shown that Ets-1 transactivation also required the Sp1 factor. The observations made in this study closely mimic the observations made in the HTLV1 studies. Since the minigenes that contain only the -100 Ets site have abnormal regulation, and are expressed at very low levels, the -100 Ets site cannot activate expression without the -85 Ets and Sp1 sites. This
conclusion is supported by the observation that minigenes with promoters m100 and mBoth have the same altered regulation as the minigene containing the wild type -118 to -90 promoter (see figures 30, and 34 - 37). Since both mutations (which abolish the -100 Ets site) have no additional effect on regulation, it is apparent that the -100 Ets site cannot function without the -85 Ets and Spl sites. Further, the observation that promoters containing two copies of the -100 Ets site cannot restore regulation suggests that two copies of an Ets site cannot complement the loss of the Spl site. Taken together, these data demonstrate that multiple elements within the essential region function as a unit to direct efficient expression and regulation of the TS gene. Since in the gel shift assays the mCR mutation does not effect Spl binding, and the -90 deletion does not effect CIII binding, the loss of one binding site, does not effect the ability of the other factors to bind to their respective elements, at least in in vitro assays. This may not be the situation in vivo, however, where multiple factors would be competing for limited binding sites. The ability of transcription factors to interact with each other to form a more stable complex than either factor could form by interacting alone, is a potential mechanism to maintain the open complex that is necessary for transcription initiation. In fact, Ets factors have a loose sequence specificity for DNA binding. In some cases Ets factors recognize only a core GGA sequence. Synergistic binding has been suggested as a mechanism to drive the association of Ets factors to the most appropriate binding site for transcriptional activation (see reference 175).
As shown in Figure 46, the CIII binding site and the 13 bp sequence containing the -80 Ets site and -85 GC box are present in the unregulated RPL-32 gene. This observation suggests that the spatial arrangement of these elements is as important for function as their presence. There is, however, another interpretation of this observation. The expression of the RPL-32 gene is necessary at all times. It is possible that in the RPL-32 gene there are elements that allow expression during all phases of the cell cycle; the elements in common with the TS gene are responsible for expression during S phase. Clearly, more work is needed to determine the role of the trans-acting factors binding to these cis-acting sequences in the growth regulation of the TS gene.

**Possible mechanism for TS regulation.**

Any model for TS regulation must take into account what is known about the required cis-acting signals, and must also explain both the transcriptional and post-transcriptional regulation. The TS gene is induced at the G1-S phase boundary in growth stimulated cells by both transcriptional and post-transcriptional mechanisms. Both mechanisms require the TS core promoter region and TS introns one and two. The transcriptional component could be due to either regulated transcriptional initiation or regulated transcriptional elongation. As discussed above the post-transcriptional regulation is likely to involve hnRNA processing. Of the three possible mechanisms to coordinate promoter elements with post-transcriptional regulation (described on pages
151, 152, and 154), perhaps the mechanism which best fits what is known about TS regulation, involves RNA polymerase II functioning as an intermediate between transcription and processing.

In the RNA pol II intermediate model, the TS promoter is responsible for transcription initiation during all phases of the cell cycle. If TS transcriptional regulation is caused by regulated initiation, intron elements could participate by providing binding sites for transcription factors. On the other hand, if TS transcriptional regulation is by attenuation, introns could participate by providing transcriptional pause sites. During the transition from G0 to S phase, trans-acting factors binding to the promoter elements cause a shift in the form of RNA polymerase II that can initiate transcription at the TS promoter. The alteration to the transcriptional complex could be either a change in the phosphorylation state or the addition of ancillary factors. During S phase the altered polymerase complex would be more processive and could transcribe through the pause sites if attenuation is involved. The altered polymerase complex could participate in post-transcriptional regulation by attracting packaging and processing factors to the site of transcription. Attraction of packaging and processing factors would increase the local concentrations, thus facilitating conversion of primary transcripts into mature mRNA. In G0 cells TS expression would be shut off because the transcriptional would be reduced because either the transcriptional complex would not be initiated at the higher rate or the transcriptional complex would not be able to transcribe through pause sites.
Furthermore, transcriptional complexes that were initiated or that did proceed beyond the pause sites would be unable to attract processing factors to facilitate post-transcriptional events.

The regulation of the Ti1.2dT minigene is likely to be due to both transcriptional and post-transcriptional events. Minigene expression is at low levels during the G0 phase, suggesting that the TS promoter is not capable of expressing the gene in resting cells. Since the SV40 promoter can express this minigene at high levels during all phases of the cell cycle, it is likely that the SV40 promoter contains elements that can function in G0 cells. If the above model is correct then the promoters found in the RPL-32, the human β-globin, and the SV40 early genes (as well as other promoters) are expressed constitutively, because during all phases of the cell cycle, RNA polymerases initiating transcripts from these promoters are fully competent to elongate transcription and to attract processing factors. Minigenes with the TS promoter are regulated because they contain elements that can only specify an activated RNA polymerase during S phase. Therefore, the TS promoter elements would act as positive activators during S phase, and not as negative elements during G0. This hypothesis is supported by the observation that alteration in the promoter that affect regulation also reduces minigene expression. If these elements were negative elements, then expression should increase not decrease.

To investigate this model it is important to distinguish whether the transcriptional regulation is due to initiation or to attenuation, and to determine the nature of the
regulatory elements located in the introns. It would also be necessary to identify the trans-acting factors which interact with the regulatory elements. In order to test the model above it is essential to determine the state of the RNA polymerase transcribing the TS gene in G0 cell and in growth stimulated cells. In Drosophila the heat shock response of the hsp70 promoter was recently confirmed to be due to changes in the phosphorylation state of the CTD of RNA polymerase II.\textsuperscript{176} RNA polymerase was UV crosslinked to DNA in vivo, and immunoprecipitated with antibodies specific for the phosphorylation state of the CTD. DNA was purified from the precipitates and subjected to southern blot analysis. In the uninduced state only the antibody specific for the unphosphorylated CTD precipitated hsp70 DNA. In the induced state, both antibodies immunoprecipitated hsp70 DNA. Similar experiments could be done on the TS gene, in the resting and serum stimulated cells. It is unlikely that the expression of the TS gene is regulated by a unique mechanism. Once we understand the mechanism by which transcription factors, interacting with TS promoter elements affect both transcription initiation and posttranscriptional regulation, it will be a trivial matter to look for this mechanism regulating the expression of other genes.
LIST OF REFERENCES


167


78. Gross, M.K. and Merrill, G.F. Thymidine kinase synthesis is repressed in nonreplicating muscle cells by a translational mechanism that does not affect the polysomal distribution of thymidine kinase mRNA. *PROC. NATL. ACAD. SCI. USA* 86: 4987-4991. 1989.


115. Hanson, U *Personal communication*, 1994. (UnPub)


Sequence requirements for transcriptional arrest in exon 1 of the murine
adenosine deaminase gene. *MOLECULAR & CELLULAR BIOLOGY* 10:
1484-1491. 1990.

125. Ash, J., Ke, Y., Korb, M. and Johnson, L.F. Introns are essential for
growth-regulated expression of the mouse thymidylate synthase gene. *MOL.

126. Takayanagi, A., Kaneda, S., Ayusawa, D. and Seno, T. Intron 1 and the 5'-flanking
region of the human thymidylate synthase gene as a regulatory determinant of

127. Hernandez, N. and Lucito, R. Elements required for transcription initiation of the
human U2 snRNA gene coincide with elements required for snRNA 3' end

128. Hernandez, N. and Weiner, A.M. Formation of the 3' end of U1 snRNA requires

129. Neuman de Vegvar, H.E., Lund, E. and Dahlberg, J.E. 3' end formation of U1
snRNA precursors is coupled to transcription from snRNA promoters. *CELL* 47:

pathways: RNAs transcribed by RNA polymerase III are not substrates for

131. Smale, S.T. and Tjian, R. Transcription of herpes simplex virus tk sequences under
the control of wild-type and mutant human RNA polymerase I promoters. *MOL.

132. Neuberger, M.S. and Williams, G.T. The intron requirement for immunoglobulin
gene expression is dependent upon the promoter. *NUCLEIC ACIDS RES.* 16:

133. de la Pena, P. and Zasloff, M. Enhancement of mRNA nuclear transport by promoter


163. Li, Yue. and Farnham, P.J. [personal communication], 1994. (UnPub)


