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.

UMI
University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700  800/521-0600
Mechanisms of the mouse thymidylate synthase gene regulation: Identification of the regulatory elements

Li, Yue, Ph.D.
The Ohio State University, 1991
MECHANISMS OF THE MOUSE THYMIDYLATE SYNTHASE
GENE REGULATION: IDENTIFICATION OF
THE REGULATORY ELEMENTS

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

Yue Li, M. S.

The Ohio State University
1991

Dissertation Committee:
Dr. Lee F. Johnson
Dr. Ing-Ming Chiu
Dr. Caroline Breitenberger

Approved by

The Ohio State University
Biochemistry Program
To my parents, my husband Honggao, and my daughter Phyllis
ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to Professor Lee F. Johnson for his encouragement, support, and guidance throughout this dissertation research. I would also like to thank the other members of my graduate committee for their helpful advice on my future career. I am grateful to the past and present members of Professor Johnson's laboratory, especially Dawei Li, Chris Harendza, and Haichao Zhang, for their valuable discussions and cooperations. Finally, I like to offer deep thanks to my entire family for their endless love, care, and support.
VITA

March 26, 1958 ........................................... Born, Tianjin, P. R. China

1982 ...................................................... B. S. in Biology, Nankai University, Tianjin, China

1982-1984 ............................................... University Teacher, Beijing Agricultural University, Beijing, China

1987 ...................................................... M. S. in Chemistry, University of Oklahoma, Norman, Oklahoma

1987-present .......................................... Graduate Teaching and Research Associate, The Ohio State University, Columbus, Ohio

PUBLICATIONS


FIELDS OF STUDY

Major Field: Biochemistry

Studies in Molecular Biology
TABLE OF CONTENTS

DEDICATION ............................................................................................................ ii

ACKNOWLEDGEMENTS ...................................................................................... iii

VITA ......................................................................................................................... iv

LIST OF TABLES ................................................................................................... ix

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

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

General strategies of gene regulation ................................................................. 1
Regulation of transcription initiation ................................................................. 2
Regulation of RNA processing .......................................................................... 6
Transcription initiation of genes with TATA-less promoters ............................... 8
  Roles of GC boxes ......................................................................................... 9
  Roles of the transcriptional start site ........................................................... 11
  Roles of sequences downstream elements .................................................. 12
Regulation of growth-dependent gene expression ............................................ 14
  An overview of an eukaryotic cell cycle ....................................................... 14
  Dihydrofolate reductase (DHFR) ................................................................ 16
  Thymidine kinase (TK) ................................................................................ 17
Regulation of the mouse thymidylate synthase gene ....................................... 21
  Isolation and characterization of TS gene and the gene products ... 21
  Regulation of the mouse TS gene in growth-stimulated cells ............ 24

MATERIALS AND METHODS

Bacterial culture ..................................................................................................... 27
Enzymes and reagents .......................................................................................... 27
Bacterial cell transformation and transfection ..................................................... 27
Isolation of plasmid and M13 RF (replication form) DNA .................................. 28
Isolation of M13 single-stranded DNA ............................................................. 30
DNA sequencing ................................................................................................... 31
Oligonucleotide-directed mutagenesis .............................................................. 31
Subcloning ............................................................................................................. 33
Cell culture .......................................................................................................... 33
Mammalian cell transfection ............................................................................. 34
Isolation of cytoplasmic RNA .......................................................................... 35
Isolation of poly(A)+ RNA ................................................................. 36
5' end labelling ................................................................................. 37
S1 nuclease protection assays ....................................................... 38
Enzyme assays ................................................................................ 39

RESULTS ................................................................................................. 41

Analysis of the 5'-flanking sequences of the mouse TS gene .......... 41
  Analysis of the upstream GC box and USF elements ................. 44
  Analysis of the TS promoter by deletion mutagenesis ............... 55
  Analysis of the essential region by site-specific mutagenesis ... 59
Analysis of TS intron effect on the expression of the mouse TS gene 64
  Construction of intron containing TS minigenes ....................... 70
Identification of the regulatory elements of the mouse TS gene
  by chimeric gene analysis ................................................................. 75
  Construction of chimeric TS minigenes with alternative promoter 76
  Construction of chimeric TS minigenes with alternative
    polyadenylation signal ................................................................. 82
  Analysis of promoter strength ...................................................... 82
  Analysis of the efficiency of TS polyadenylation signal .......... 92
Study of the regulation of the mouse TS gene in response to
growth stimulation ........................................................................ 93
  Regulation of wild-type TS minigene expression in
    growth-stimulated cells .............................................................. 95
  Role of the TS 5'-flanking region in growth-regulated
    TS gene expression ................................................................. 99
  Role of the TS polyadenylation signal in growth-regulated
    TS gene expression ............................................................... 111

DISCUSSION ............................................................................................. 117

Analysis of the 5'-flanking sequences of the mouse TS gene .......... 117
  Roles of the GC boxes in transcriptional initiation of the TS gene .. 117
  Essential region of the mouse TS promoter is in close proximity
to the transcriptional start sites ..................................................... 123
The mouse TS gene expression is stimulated by some
(but not all) introns ........................................................................... 127
Identification of the regulatory elements of the mouse TS gene by
chimeric gene analysis .................................................................... 130
  The low level of expression of the mouse TS gene is probably
    not due to a weak promoter, it is instead due to an inefficient
    polyadenylation signal ............................................................. 131
  The stimulatory effect of introns 5 and 6 is not specific for the TS
    promoter or TS polyadenylation signal ...................................... 132
Study of the regulation of the mouse TS gene in response to growth
stimulation ......................................................................................... 133
The sequences responsible for growth-regulated TS gene expression are present and functional in the intron-containing TS minigenes ................................................................. 134
The 5'-flanking region of the mouse TS gene is necessary but not sufficient for normal regulation in growth-stimulated cells .......... 135
The 3'-flanking region is not essential for growth-regulated expression of the mouse TS gene ................................................................. 139

REFERENCES ..................................................................................................... 141
## LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Effects of promoter mutations on TS minigene expression</td>
<td>48</td>
</tr>
<tr>
<td>2. Effects of mutations in GC boxes on the pattern of transcriptional start sites</td>
<td>54</td>
</tr>
<tr>
<td>3. Effects of promoter deletions on TS minigene expression</td>
<td>56</td>
</tr>
<tr>
<td>4. Effects of TS introns on the expression of TS gene</td>
<td>69</td>
</tr>
<tr>
<td>5. Activity of chimeric TS minigenes</td>
<td>86</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>42</td>
</tr>
<tr>
<td>2.</td>
<td>45</td>
</tr>
<tr>
<td>3.</td>
<td>50</td>
</tr>
<tr>
<td>4.</td>
<td>52</td>
</tr>
<tr>
<td>5.</td>
<td>57</td>
</tr>
<tr>
<td>6.</td>
<td>61</td>
</tr>
<tr>
<td>7.</td>
<td>65</td>
</tr>
<tr>
<td>8.</td>
<td>67</td>
</tr>
<tr>
<td>9.</td>
<td>71</td>
</tr>
<tr>
<td>10.</td>
<td>73</td>
</tr>
<tr>
<td>11.</td>
<td>78</td>
</tr>
<tr>
<td>12.</td>
<td>80</td>
</tr>
<tr>
<td>13.</td>
<td>83</td>
</tr>
<tr>
<td>14.</td>
<td>88</td>
</tr>
<tr>
<td>15.</td>
<td>90</td>
</tr>
<tr>
<td>16.</td>
<td>97</td>
</tr>
<tr>
<td>Number</td>
<td>Topic</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>17</td>
<td>Expression of pTl\textsubscript{56}T(d) in growth-stimulated cells</td>
</tr>
<tr>
<td>18</td>
<td>Expression of pTl\textsubscript{12}T(d) in growth-stimulated cells</td>
</tr>
<tr>
<td>19</td>
<td>Expression of pSI\textsubscript{56}T in growth-stimulated cells</td>
</tr>
<tr>
<td>20</td>
<td>Expression of pTl\textsubscript{56}T(d)(-250) in growth-stimulated cells</td>
</tr>
<tr>
<td>21</td>
<td>Expression of pTl\textsubscript{56}T(d)(-150) in growth-stimulated cells</td>
</tr>
<tr>
<td>22</td>
<td>Expression of pTl\textsubscript{56}S in growth-stimulated cells</td>
</tr>
<tr>
<td>23</td>
<td>Expression of pSI\textsubscript{56}S in growth-stimulated cells</td>
</tr>
<tr>
<td>24</td>
<td>A proposed model for the role of GC boxes in transcriptional initiation from the mouse TS promoter</td>
</tr>
</tbody>
</table>
INTRODUCTION

The concept of gene expression generally refers to the process of converting genetic information stored in DNA into protein by RNA. This process is strictly controlled so that organisms develop, reproduce, and pass on the information to the future generation. During the past 20 years or so tremendous efforts have been made toward understanding the mechanisms of gene regulation. As a result, an enormous amount of information has been accumulated, primarily due to the use of molecular biological techniques in recent years. Much is known about the mechanisms of gene regulation in prokaryotic cells because of the ease of study due to their simple genome. However, to completely understand how genes are regulated in eukaryotic cells is a more challenging task.

General strategies of gene regulation

The ultimate results of gene regulation are reflected as differences in the amount and the kinds of proteins under various developmental and physiological conditions. Some genes are expressed constitutively in most cells since the proteins they encode are essential for cell viability. These genes are called housekeeping genes. In contrast, some genes are expressed in large amounts in selected types of cells because the functions of the proteins they encode are needed only for these particular cell phenotypes. These genes are called tissue-specific or specialized genes.
From a large amount of evidence it is known that gene expression can be regulated at all steps in the pathway leading from DNA to protein, including mRNA transcription, mRNA processing, mRNA transport, mRNA template selection, mRNA stability, and protein activity.

In a broad term, gene regulation is accomplished by a variety of interactions between macromolecules in the cells, especially between DNA and proteins. The combinatorial effects of these interactions can help to turn a gene either on or off under various developmental and physiological conditions.

**Regulation of transcriptional initiation**

It is now well established that gene activity is regulated in large part at the level of mRNA synthesis (transcription). The process of transcription consists of three steps: initiation, elongation, and termination. Theoretically all steps are subjected to control, but regulation at transcription initiation appears to be the most important (Darnell et al., 1987). There are two distinct levels of control in the initiation of transcription, namely the frequency of initiation and the site of initiation. The sequences required for the precise and regulated initiation of transcription are composed of various regulatory elements, namely the promoter, enhancer, and/or silencer, which may be dispersed over a large region of DNA 20 kb or more upstream or downstream from the start site of mRNA transcription (Maniatis et al., 1987; Müller et al., 1988; Wasylyk, 1988). These elements, in turn, are composed of discrete DNA sequence motifs (modules), which contain one or more binding sites for
sequence-specific DNA binding proteins (McKnight and Tjian, 1986; Jones et al., 1988; Dynan, 1989).

The promoter element consists of a number of transcriptional control motifs, usually spanning from -200 to +30 bp relative to the transcriptional start site (Dynan, 1989; Wasylyk, 1988). It is well documented that a highly conserved AT-rich region, designated as TATA box (TATAAA), about 25 to 30 bp upstream of the start site is important for both positioning the transcription start site and the basal level of transcription (Breathnach and Chambon, 1981; Nakajima et al., 1988; Buratowski et al., 1988). It has been demonstrated that besides RNA polymerase II, at least five cellular factors (TFIIA, TFIIB, TFIID, TFIIE, and TFIIF) are involved in transcriptional initiation (Nakajima et al., 1988). Among these, TFIID is the only protein which specifically binds to the TATA box. The detailed molecular steps in transcriptional initiation are not clear. However, it is generally accepted that the initiation process requires the formation of a preinitiation complex at the TATA box and this subsequently directs RNA polymerase II to start transcription 25 to 30 base pairs downstream (Nakajima et al., 1988). Since the TATA sequence is present in most of the promoters studied so far and can provide a basal level of transcription, it is regarded as the general transcription element, and therefore TFIID, which binds to it, is regarded as a general transcription factor. By definition, the general transcription machinery does not play an important role in regulating specific gene expression.

Apart from the TATA box, another promoter component is the so called upstream promoter element(s) which is usually located about 40 to 200 bp upstream of the mRNA start site (Wasylyk, 1988; Dynan, 1989). This upstream
element(s) can regulate the frequency of transcriptional initiation, and is a major target for the action of promoter-specific transcription factors (Dynan and Tjian, 1985a). Many DNA sequence motifs and the corresponding regulatory proteins have been identified in this region, such as the GC box/Sp1 (Benoist and Chambon, 1981; Dynan and Tjian, 1985b), the CAAT box/CTF/NF1 (Dierks et al., 1983; Jones et al., 1987), and the USF or MLTF sequence /USF or MLTF (Sawadogo and Roeder, 1985; Carhew et al., 1985). It has been demonstrated that the interactions of many DNA sequence motifs in this region with the corresponding proteins can stimulate transcription. Information accumulated over the past decade reveals some general features of the upstream promoter elements. First, the position and spacing of the DNA sequence motifs in this region are relatively flexible (Dynan and Tjian, 1985; Dynan, 1989). Second, many of the motifs can function in both orientations (Dynan and Tjian, 1985; Dynan, 1989). Third, each motif can interact with one or more transcription factors (Dynan, 1989). Finally, the combination of different numbers and types of these motifs constitute a wide variety of complex promoter structures of higher eukaryotic genes (Mitchell and Tjian., 1989). Thus, by differential interactions of transcription factors with these motifs, gene expression can be sufficiently regulated under different developmental and physiological conditions.

The mechanism of regulation of transcriptional initiation is complicated further by the finding that sequences distant (both downstream and upstream) from the transcription start site are required for efficient transcription. These sequences are called enhancers (Serfling et al., 1985; Müller et al., 1988; Wasylyk, 1988; Dynan, 1989). It has been demonstrated that enhancers can
stimulate transcription in a distance and orientation-independent, and in many cases, tissue-specific manner. The mechanisms for enhancer action are still not known, although several working models have been proposed (Müller et al., 1988; Dynan, 1989). It has been demonstrated that enhancers are also comprised of many DNA sequence motifs similar to the promoter organization (Wasylyk, 1988). These enhancer motifs may, in turn, consist of a number of subunits, called enhansons, which may bind to transcription factor(s) (Ondek et al., 1988). Interestingly, it has been found that enhancer elements and upstream promoter elements can bind to common transcription factors (Wasylyk, 1988). For example, the octamer-binding factors OTF-1 and OTF-2 can not only bind to the octamer motifs in the IgH enhancer and the SV40 72 nucleotide repeats, but also to the octamer in the histone H2B promoter (Müller et al., 1988; and references therein). These observations suggest that enhancer elements and upstream promoter elements may modulate transcription by a similar mechanism.

While most sequence-specific DNA binding proteins act as positive regulators, it is conceivable that negative regulation is as important as positive regulation. It has been demonstrated that overlapping or superimposed DNA sequence motifs can bind to positive or negative transcriptional factors under different developmental and physiological conditions (Mitchell and Tjian, 1989, and references therein). Similarly, a given transcription factor can both activate and repress gene expression. Examples of this type include the regulation of gene expression by steroid hormone receptors (Levine and Manley, 1989). It has been shown that the same glucocorticoid receptor that activates gene expression can also
negatively regulate expression of the human glycoprotein hormone α-subunit gene (Akerblom et al., 1988). The binding site of the glucocorticoid receptor was found to overlap the cAMP responsive element sites. Therefore, it has been suggested that the negative regulation by glucocorticoid receptor is accomplished by interference with a cAMP responsive enhancer (Akerblom et al., 1988).

**Regulation of RNA processing**

In eukaryotic cells, processing of primary RNA transcripts (5' end capping, RNA splicing, and 3' end cleavage-polyadenylation) is absolutely required for the formation of mature mRNA. An increasing body of evidence has suggested that RNA splicing and polyadenylation play important roles in regulating gene expression, although the mechanisms are not well understood. For example, both alternative splicing and polyadenylation of primary RNA transcripts have been found at different developmental stages, which results in production of different mature transcripts and therefore of different proteins from a single gene (Andreadis et al., 1987; Galli et al., 1988). Examples of this type include the IgM heavy-chain (μ) transcription unit (Galli et al., 1988). At the early stages of B-cell development, a downstream poly(A) site is used to produce the membrane-bound protein (μm). When the secreted protein (μs) is needed, B cells switch to use an upstream poly(A) site to eliminate the carboxy-terminal membrane binding domain.

Another aspect of gene regulation is the efficiency of RNA processing. It has been demonstrated that polyadenylation efficiency can be affected by a number of factors. First, it has been found that mutations in the
hexanucleotide AAUAAA (the upstream polyadenylation signal located 10 to 20 nucleotides upstream of the poly(A) site) result in a dramatic reduction in the level of cytoplasmic mRNA (Wickens and Stephenson, 1984). Similarly, when a variant AUUAAA was mutated to AAUAAA, the mRNA level is increased significantly (Harendza and Johnson, 1990), indicating that the sequence AAUAAA is important for not only locating the position but also the efficiency of polyadenylation. In addition, sequences both downstream and upstream of the AAUAAA signal also affect both the site and efficiency of polyadenylation (Ryner et al., 1989; Carswell and Alwine, 1989). Finally, sequences of the downstream polyadenylation signal, known as GU-rich and/or U-rich region about 50 nucleotides downstream of the poly(A) site, are important for both accurate and efficient polyadenylation (Manley, 1988).

Similarly, both the splicing efficiency and accuracy can be affected by the cis-acting splicing sequences, namely the 5', 3' splicing sites and the splicing branch point (Raumann and Breathnach, 1985; Krainer and Maniatis, 1988). Besides the cis-acting processing elements (polyadenylation signal, splicing sites and branch point), it was interesting to find that the efficiency of RNA processing can be influenced by sequences in different parts of the gene (Sisodia et al., 1987; Niwa et al., 1990; Manley et al., 1990). In the studies of in vitro processing of chimeric RNA, Niwa et al. (1990) have shown that the efficiency of polyadenylation is increased when an upstream intron is present. They have also demonstrated that the 3' splicing site is important for the stimulation effect. Therefore, it has been speculated that splicing and polyadenylation factors may interact to recognize the poly(A) site within the last exon. Furthermore, Sisodia et al. (1987) have shown that splicing and
polyadenylation are very inefficient when the promoter for an rRNA gene was used to direct the synthesis of a protein-coding gene. These observations suggest that the processes involved in mRNA formation (e.g. transcription initiation, RNA splicing, and polyadenylation), can communicate with each other to regulate the efficiency of individual processes, which may provide another means of gene regulation.

**Transcriptional initiation of genes with TATA-less promoters**

The promoters of many genes associated with cell growth have been found to lack the canonical TATA box. Some can still transcribe mRNA from a single site (Smale and Baltimore, 1988; Dudov and Perry, 1984), while others direct the transcription from multiple start sites (Dyan, 1986). Most of the TATA-less genes discovered so far have housekeeping functions (Azizkhan et al., 1986; Brown et al., 1985; Chen et al., 1984; Deng et al., 1986; Farnham et al., 1985; Ingolia et al., 1986; Melton et al., 1984; Mitchell et al., 1986; Patel et al., 1986; Reynolds et al., 1984; Takeishi et al., 1989; Valrio et al., 1985). As mentioned earlier, the housekeeping genes are expressed in all cell types at relatively low levels, which is in contrast with the specialized genes that are expressed at relatively high levels in selected cell types. It has been speculated that the distinctive pattern of gene expression exhibited by each type of genes may be due, at least in part, to the process by which the transcriptional initiation is specified.

While a large body of information has been accumulated for understanding the biochemical mechanism of transcriptional initiation from TATA-containing promoter, little is known about the molecular steps involved
in the transcriptional initiation of housekeeping genes. However, since the
discovery of the first TATA-less promoters, there has been much interest in
unravelling the mechanism of transcriptional initiation in the absence of a
TATA box.

Roles of GC boxes. Besides lacking the TATA box, one of the
common features of the promoters of many housekeeping genes is that they
are very GC-rich. This is usually associated with the presence of multiple GC
boxes in the promoter region (Dynan, 1986; Bird, 1986; 1987). The GC box
was originally discovered to be critical in directing efficient transcription from
some viral promoters, such as simian virus 40 (SV40) early promoter
(Kadonaga et al., 1986 and references therein). The transcription factor Sp1
was identified to interact with GC box and was isolated from HeLa cells on the
basis of its ability to bind selectively to the GC consensus sequence,
G/TGGGCGGGGC (Kadonaga et al., 1986).

Despite the high GC content in the TATA-less promoters of
housekeeping genes, little is known about the role of GC boxes in the
transcriptional initiation from this type of promoter. The question has not been
directly addressed until recently. Most of the information was obtained from
studies on the mammalian dihydrofolate reductase (DHFR) gene, which is
one of the best-characterized housekeeping genes.

The 5'-flanking regions of the murine (Farnham et al., 1985), hamster
(Azizkhan et al., 1986; Mitchell et al., 1986), and human (Chen et al., 1984)
DHFR genes have been cloned and sequenced. There is a striking degree of
sequence conservation in the 5’ flanking sequence among these species.
Like many other housekeeping genes studied so far, the DHFR promoter does not contain apparent TATA and CAAT boxes. There are four GC boxes located within 300 base pairs 5' of the ATG start codon. The sequence containing these GC boxes is sufficient to direct transcription, but whether more than one GC box is absolutely required to support transcription is still in debate (Mitchell et al., 1986; Farnham and Schimke, 1986; Means and Farnham, 1990, Swick et al., 1989). In vitro transcription analysis has shown that Sp1 can bind to these GC boxes, and activate transcription (Dynan et al., 1986; Swick et al., 1989).

In vivo RNA 5' end mapping studies have revealed that there are four clusters of start sites in the mouse DHFR gene (McGrogan et al., 1985; Sazer and Schimke, 1986). Two major start sites are proximal to the ATG translation start codon and two minor sites are far upstream of the gene. The GC boxes are located within a set of 45 to 48 bp sequence repeats and each GC box is about 40 to 50 nucleotides upstream of a start site. This spatial arrangement is conserved among mammalian DHFR genes, indicating that the relative distance of the start site to each upstream GC box is important for specifying the position or the relative utilization of the start sites. A recent study of the hamster DHFR gene has demonstrated that GC boxes not only are required for efficient transcription but also regulate start site utilization in the promoter (Blake et al., 1990). Transient expression and in vitro transcription analyses have shown that inactivation of the GC boxes not only causes a reduction in promoter activity but also changes the relative utilization of transcriptional start sites in a distance-dependent manner. Primer extension results indicate that although the GC boxes can regulate the relative utilization of
transcriptional start sites, it appears that they have no role in specifying the positions of initiation.

**Roles of the transcriptional start site.** Since the discovery of TATA-less promoters, it has always been of a great interest to identify the DNA sequence responsible for specifying the transcriptional start site of this type of gene. Recently, the region surrounding the start site has been demonstrated to position the start site of transcription of several mammalian genes with TATA-less promoters. It has been demonstrated that a 17 bp element (initiator, Inr) containing the transcriptional start site of the terminal deoxynucleotidyltransferase (TdT) gene is sufficient for accurate basal level of transcription (Smale and Baltimore, 1989). Mutation of the Inr can alter both the efficiency and accuracy of initiation. Similarly, it has been found that substitution of nucleotides around the major late start site of SV40 results in both a reduction in transcriptional efficiency and alteration of the position of the start site (Ayer and Dynan, 1988). More recently a protein, designated as HIP1 (Housekeeping Initiator Protein 1), has been found to bind to the transcriptional start site of the mouse DHFR gene and specify the position of the start site (Means and Farnham, 1990). It has been further demonstrated that apart from the sequences around the transcriptional start site, the proper initiation of the mouse DHFR gene also requires an upstream GC box.

Sequence comparison has revealed that the SV40 late major start site has striking sequence homology to the mouse DHFR start site (Means and Farnham, 1990). DNase I footprint analysis using HeLa nuclear extract demonstrated that the protected sequences of SV40 major late and the
mouse DHFR genes are very similar (Means and Farnham, 1990). Furthermore, addition of the concatemer of the DHFR start sites can not only decrease the transcriptional efficiency but also alter the position of start site of the SV40 major late gene in vitro (Means and Farnham, 1990). These results suggest that the same, or a similar protein may bind to both the SV40 major late and the mouse DHFR start site and specify the site of transcription. Further sequence comparison has revealed that the sequences around transcription start sites of many housekeeping genes exhibit strong homology to the start site sequence of the mouse DHFR gene (Means and Farnham, 1990). These sequence analyses suggest that HIP1, which binds to the start site consensus sequence, could be a general transcription factor for housekeeping genes with TATA-less promoters.

In contrast, the TdT Inr exhibits little homology to the start site region for the mouse DHFR promoter. In addition, the proper initiation of transcription of TdT gene does not depend on upstream promoter elements, which is different from the DHFR gene. TdT gene is not a housekeeping gene since its expression is limited to the precursor of B and T lymphocytes (Landau et al., 1984). Moreover, unlike many housekeeping genes, the promoter region of TdT gene is not GC-rich and has only one transcriptional start site. Thus, it is possible that the mechanism of transcriptional initiation of TdT is somewhat different from that of the SV40 major late and the mouse DHFR genes, or housekeeping genes in general, even though they all lack a TATA box.
Roles of sequences downstream of transcriptional start site.
Recent evidence suggests that transcription of many mammalian genes which lack a TATA box can be regulated by sequences downstream of the transcriptional start site. Ayer and Dynan (1990) has identified a promoter element centered 28 bp downstream of the SV40 major late transcription start site. The activity of this promoter element appears to be mediated by a DNA-binding protein, DAP (downstream activating protein). Interaction of this downstream element with DAP facilitates assembly of a functional preinitiation complex, and therefore increases the efficiency of transcription. An essential transcriptional element (δ element) has also been found in the first untranslated exon of various ribosomal protein genes (rpL30, rpL32, and rpL16) (Atchison et al., 1989; Hariharan et al., 1989). Gel mobility shift analyses indicated that the same trans-acting factor (δ factor) may bind to this common element (Atchison et al., 1989; Hariharan et al., 1989). Site-specific mutagenesis has demonstrated that the interaction of δ element with δ factor is critical for promoter activity (Hariharan et al., 1989). Moreover, deletion studies of the mouse DHFR promoter have shown that two regions downstream of the transcriptional start site are important for efficient transcription (Farnham and Means, 1990). One is located in the 5' untranslated region, the other spans exon 1 and part of intron 1. The relationship of these downstream elements to the transcriptional start sites of these genes is not known, although it has been suggested that the downstream element may influence the initiation pattern through interactions of its binding protein with the proteins which bind to the initiation site (Farnham and Means, 1990).
Sequence comparison has revealed striking similarity among SV40 major late downstream element, the δ element of ribosomal protein genes, and the downstream element in the 5' untranslated region of the mouse DHFR gene (Farnham and Means, 1990). Interestingly, all of these elements are in close proximity to the transcriptional start site. Therefore, it is possible that the same or similar proteins may bind to these downstream elements and modulate transcription by a similar mechanism.

Regulation of growth-dependent gene expression

The study of regulation of growth-dependent genes has become an important subject. Unraveling the mechanisms for this regulation not only will provide some insight on how gene expression is regulated in general, but it may also help us to understand better how cell replication is controlled.

An overview of the eukaryotic cell cycle. The eukaryotic cell cycle is a cascade of events that lead to cell division. The whole cycle consists of four phases. S phase is the period of time during which DNA synthesis takes place. It is preceded and followed by two gap periods, G₁ and G₂, respectively. Throughout these two gap periods there is continued synthesis of macromolecules such as protein and RNA. Cell division occurs in the M (mitotic) phase. When cells stop proliferating, however, they are generally in a quiescent state which is called the G₀ (resting) phase. Cells in G₀ phase are preparing for DNA synthesis, which is the major difference from cells in G₁ phase. However, cells in G₀ phase can be activated to reenter the cell cycle (usually G₁ phase) under appropriate conditions. In tissue culture, G₀ cells
can be prepared by reducing the concentration of serum in the culture medium. The cells are induced to reenter the cell cycle by addition of high concentration serum back to the culture medium. The $G_0$ phase has often been chosen as the state to synchronize cells due to the ease of preparation.

It has been known for some time that many proteins which are involved in DNA replication are produced in a cell cycle-dependent manner. Examples include histones and enzymes that participate in nucleotide metabolism (Denhardt et al., 1986 and references therein). In addition, the expression of some cellular proto-oncogenes can also be stimulated in response to growth stimulation (Denhardt et al., 1986 and references therein). Over the past ten years or so, tremendous effort has been made toward uncovering the molecular basis for this regulation. The information accumulated so far implicates a complex picture concerning the mechanisms for the regulation of these genes.

Recently, an increasing number of studies have demonstrated that the expression patterns of a number of genes are different when different cell synchronization methods are used (e.g. serum starvation/stimulation vs centrifugal elutriation or mitotic selection) (Feder et al., 1989; Hendrickson et al., 1980; Sherley and Kelly, 1988; Stuart et al., 1985; Steward et al., 1987; Coppock and Pardee, 1987; Navalgund et al., 1980; Jenh et al., 1985a; Ali Imam et al., 1988). It is generally accepted, however, that the former (serum starvation/stimulation) deals with growth-dependent gene regulation, whereas the latter (centrifugal elutriation or mitotic selection) deals with regulation in the normal cell cycle. Whether these two processes are interrelated is not entirely clear at the present time. This introduction only
concerns the growth-dependent gene expression. The two enzymes
discussed below are involved in DNA synthesis and are expressed in a
growth-dependent manner.

**Dihydrofolate reductase (DHFR).** Most of the growth-dependent
regulation studies of the DHFR gene have been performed in mouse cell lines
with amplified copies of the DHFR gene. It has been demonstrated that the
synthesis of DHFR enzyme is controlled relative to the growth state of the cell
(Alt et al., 1976; Frearson et al., 1966). The level of DHFR mRNA increases 10
to 20-fold as cells progress from G0 phase to S phase (Hendrickson et al.,
1980). The linear relationship between mRNA level, the amount and rate of
enzyme synthesis indicates that the expression of the DHFR gene is mainly
controlled at the level of mRNA production (Kellems et al., 1979).

Evidence for regulation of the DHFR gene at both transcriptional and
posttranscriptional levels has been reported. It has been demonstrated that
following serum stimulation of quiescent mouse 3T6 cells, the rates of
synthesis of both hnRNA and cytoplasmic DHFR RNA are increased (Wu and
Johnson, 1982; Hendrickson et al., 1980). Moreover, a recent study using a
differentiation-competent mouse muscle cell line showed that the transcription
rate is reduced during muscle cell commitment, indicating that the DHFR gene
is transcriptionally regulated (Schmidt and Merrill, 1989). However, studies
using fibroblasts released from nutrient starvation (an alternative way to arrest
cells) showed that there is no change in rate of the DHFR gene transcription,
and therefore it was speculated that the elevated level of DHFR mRNA is due
to increased stability (Leys and Kellems, 1981; Leys et al., 1984).
Additional evidence for posttranscriptional control of DHFR gene expression was mainly from the study of chimeric DHFR minigenes (Kaufman and Sharp, 1983). Minigenes containing the adenovirus late promoter, DHFR cDNA, and DHFR or SV40 late polyadenylation signal are regulated normally during growth stimulation. However, minigenes containing the same promoter and DHFR coding region but SV40 early or a cellular polyadenylation signals do not show increased expression during growth stimulation, suggesting that posttranscriptional control plays an important role in regulating DHFR gene expression.

Thymidine kinase (TK). The TK gene has been studied from a variety of sources, including chicken, hamster, mouse, and human. It has been demonstrated that the levels of both TK enzyme and mRNA are low in quiescent cells, and increase 10 to 40-fold following serum stimulation (Steward et al., 1987; Coppock and Pardee, 1987). The regulation of TK gene expression appears even more complicated than that of the DHFR gene, probably due to the fact that more cell types were involved in the studies. No consensus has been reached regarding the mechanisms for regulation of the TK gene. Many levels of control (transcriptional, posttranscriptional, translational, and posttranslational) have been implicated to play a major role in regulating growth-dependent TK gene expression.

Evidence has been accumulated supporting a transcriptional control mechanism. Steward et al. (1987) have shown that when time points of the nuclear run-on experiments were chosen to span the G1-S phase interface, a 6 to 7-fold increase in the rate of transcription was observed in growth-
stimulated CV1 cells, indicating that transcription control plays an important role in regulating TK gene expression. This result is consistent with the observation that TK hnRNA is maximally expressed in early S phase after serum stimulation of quiescent human fibroblasts (Lipson and Baserga, 1989). By chimeric gene analysis, it has been demonstrated that the human TK promoter can direct growth-dependent expression of several bacterial reporter genes (Travali et al., 1988; Kim et al., 1988).

To identify the locations of the regulatory sequences in the promoter region, deletion analysis has been carried out. Kim et al. (1988) have shown that the sequence between -441 to -63 relative to the transcriptional start site of the human TK gene is sufficient for normal regulation. By further detailed deletion and site-directed mutagenesis analyses, Lipson et al. (1989) have claimed that the sequence between -83 to -63 is crucial for growth-regulated expression of TK gene. Similarly, Roehl and Conrad (1990) have found that the region between -135 to -67 can provide growth-regulated expression of the bacterial neomycin resistance gene.

Some progress has been made toward identification of the trans-acting factor(s) which interact with the regulatory sequences. Gel mobility shift analysis has revealed that two inverted CAAT boxes, located at -36 and -67 upstream of the human TK promoter form complexes with nuclear DNA-binding proteins in a growth state-specific manner (Knight et al., 1987). Several recent studies on the mouse TK promoter have demonstrated that sequences between -174 to +159 relative to the transcriptional start site are sufficient to provide S-phase-specific regulation of two reporter genes (bacterial chloramphenicol acetyltransferase gene and human β-globin gene)
(Fridovich-Kei et al., 1991). Gel mobility shift and DNase I footprint analyses have revealed that the sequences between -174 to -4 contains three distinct protein binding sites with a consensus sequence CCCNCNNNCT (Dou et al., 1991). This binding activity, designated as Yi, is absent in G<sub>0</sub> and G<sub>1</sub> cell extracts but present in G<sub>1</sub>/S cell extracts (Dou et al., 1991), presumably due to the change in the stability of the Yi protein (Bradly et al., 1991).

The increase in the rate of transcription of the TK gene (7-fold) from G<sub>0</sub> to S phase, however, does not entirely account for the large increase (more than 20-fold) in the level of TK mRNA. Thus, posttranscriptional control must also play an important role in TK gene expression. Most of the evidence favoring posttranscriptional control is from chimeric gene analyses. Several studies using different sources of TK genes (e.g. chicken, mouse, and human) have shown that sequences within the mature TK mRNA (cDNA) can provide growth-dependent expression of the TK gene (Merrill et al., 1984; Lewis and Matkovich, 1986; Travali et al., 1988; Liberman et al., 1988; Steward et al., 1987). The mechanism for this observation is not clear, although a posttranscriptional event is the most plausible explanation. It has been suggested that the TK polyadenylation signal and most of the 3' untranslated sequences are not required for the regulation since substituting them with other polyadenylation signals does not abolish the regulation (Liberman et al., 1988; Steward et al., 1987). By investigating the metabolism of BALB/c3T3 cell nuclear RNA during growth stimulation, Gudas et al. (1988; 1990) have found a dramatic change in the nuclear posttranscriptional processing of TK hnRNA at the G<sub>1</sub>/S boundary, resulting in the appearance of TK mRNA precursors and the accumulation of mature TK mRNA in the nucleus. They
have also demonstrated that intron removal of mouse and hamster TK hnRNA occurs in an ordered fashion in the nuclei of S-phase cells. These observations suggest that the level of TK mRNA is controlled by a posttranscriptional event involving nuclear RNA processing. Furthermore, it has been demonstrated that the half-life of TK mRNA is about 8 to 12 hours in S and M phases and decreases as cells enter quiescence, suggesting that the content of TK mRNA can also be controlled at the level of stability (Coppock and Pardee, 1987).

The study of the regulation of TK gene is complicated further by the observation that, in some cases, TK enzyme level does not correlate with mRNA level during growth stimulation or growth arrest. In studies using nonreplicating mouse muscle cells transfected with multiple copies of the chicken TK gene, Gross and Merrill (1988; 1989) have reported that the decline in TK mRNA is only about 2-fold as the cells become committed to terminal differentiation, whereas TK enzyme activity decreased to a greater extent (20-fold) due to a reduction in the rate of TK protein synthesis. They have also shown that the polysomal distribution of mRNA is about the same in both proliferating and committed cells, indicating that synthesis of TK protein in nonproliferating muscle cells is inhibited by a translational mechanism that does not affect the polysomal distribution of TK mRNA. In line with these observations, by using rat cells transfected with a human TK cDNA under control of a number of different promoters, Ito and Conrad (1990) have demonstrated that the pattern of TK mRNA expression is uncoupled from the growth state of the cells, but rather is a function of promoter used. In contrast, TK protein level always remains low in G1 phase and increases as cells enter
S phase regardless of promoter used to direct mRNA synthesis. These results suggest that the growth-dependent expression of the TK gene is regulated at the level of translation and/or posttranslation, rather than transcription and/or posttranscription as suggested above.

Although no consensus conclusion has been reached in regard to the major control step, it is generally accepted that the regulation of TK gene expression in response to growth stimulation is controlled at multiple levels.

**Regulation of the mouse thymidylate synthase gene**

Thymidylate synthase (TS, EC 2.1.1.45) is a housekeeping enzyme that catalyzes the reductive methylation of deoxyuridylic acid to form thymidylic acid in the de novo biosynthetic pathway. Due to its direct involvement in DNA synthesis, TS is essential for the survival of all proliferating cells, and therefore an important target enzyme in cancer chemotherapy (Blakley, 1969; Danenberg, 1977). The expression of the TS gene is highly regulated both in response to growth stimulation (Conrad, 1971; Navalund et al., 1980) and, to a lesser extent, in the normal cell cycle (Nagarajan and Johnson, 1989). The mechanism for this regulation, however, is still not clear. During the past decade or so research in our laboratory has been focused on unravelling the biochemical mechanisms for controlling TS gene expression.

**Isolation and characterization of the TS gene and the gene products.** In order to study TS gene regulation, it was essential to isolate the TS cDNA and gene. TS represents only about 0.01% of total cell proteins in proliferating cells, which is typical for the product of a housekeeping gene.
Therefore it was difficult to isolate the TS cDNA by conventional methods. To facilitate the isolation of the TS gene, a fluorodeoxyuridine (FdUrd)-resistant mouse 3T6 cell line (LU3-7) was isolated, which overproduces TS enzyme by a factor of 50 to 100 (Rossana et al., 1982) due to a corresponding increase in the amount of TS mRNA (Geyer and Johnson, 1984) and the number of copies of the TS gene (Jenh et al., 1985b). It was further demonstrated that the expression of the TS gene is regulated in LU3-7 cells in a similar manner as in 3T6 cells (Navalgund et al., 1980; Jenh et al., 1985c). Therefore, the LU3-7 cell line not only facilitated the isolation of TS cDNA and gene (Geyer and Johnson, 1984; Deng et al., 1986), but also provided a convenient system to study TS gene expression (Jenh et al., 1985c; Jenh et al., 1985a).

TS cDNA sequence analysis revealed that the TS open reading frame is 921 nucleotides long and specifies a protein of 34.9 kDa (Perryman et al., 1986), which is highly conserved throughout evolution (Perryman et al., 1986 and references therein). To better understand the role of TS in metabolism and chemotherapy, a bacterial expression vector containing the TS cDNA has been constructed (Zhang et al., 1989). This paves the way to study the catalytic mechanism of TS enzyme by protein engineering techniques.

Sequence analysis of the mouse TS genomic clones revealed that the coding region is contained within a 12-kb fragment of DNA and is interrupted by 6 introns (Deng et al., 1986). Although the mouse TS gene has the overall structure typical of mammalian genes, it has several unusual features in the 5'- and 3'- flanking regions. It was found that the predominant TS mRNA does not have a 3' untranslated region (Perryman et al., 1986; Jenh et al., 1986). The termination codon, UAA, of TS mRNA is immediately followed by the
poly(A) sequence. Since the termination codon of the TS gene is TAG (Deng et al., 1986), the final A of the UAA is added by poly(A) polymerase. Two elements have been identified to be essential for efficient polyadenylation of TS mRNA. One is the variant hexanucleotide AUUAAA, which is located in the TS coding region. The other is an oligo(U) sequence which is 32 nucleotides downstream of the termination codon (Harendza and Johnson, 1990). Interestingly, it was found that the oligo(U) sequence corresponds to the poly(A) tail of a mouse L1 long interspersed repetitive element (LINE) that was inserted in opposite orientation to the mouse TS gene more than 5 million years ago.

Like many housekeeping genes, the TS 5'-flanking region is very GC-rich, does not contain the consensus TATAA and CCAAT sequences, and directs transcription from multiple initiation sites spanning a 60-nucleotide region. However, the 5'-flanking region does contain potential binding sites for several transcription factors. Examples include GC boxes (Benoist and Chambon, 1981; Dynan and Tjian, 1985b) at -130 and -80 relative to translational start codon ATG, a potential binding site for USF (Sawadogo and Roeder, 1985) or MLTF (Carthew et al., 1985) at -137, and a cluster of potential binding sites for NF-κB (Sen and Baltimore, 1986), E2F(Kovesdi et al., 1986), HSTF (Pelham, 1982), and NF1(Jones et al., 1987) in a region between -113 to -90. Deletion analyses have demonstrated that the region between -150 to -54 (containing all the potential binding sites mentioned above) is important for TS promoter activity (DeWille et al., 1988). Gel mobility shift and footprint analyses have demonstrated that multiple protein-DNA interactions occur in this region (Deng et al., 1989a; Jolliff et al., 1991).
To facilitate the study of mouse TS gene expression, several intronless TS minigenes have been constructed (DeWille et al., 1988). These minigenes contain various length of the 5'-flanking sequence and 0.25 kb of the 3'-flanking region of the TS gene linked to TS cDNA at common restriction sites in the first and last exons (DeWille et al., 1988). Transient expression analyses revealed that these minigenes are expressed normally, although at a relatively low level, when transfected into ts- hamster V79 cells. Many studies on the regulatory sequences mentioned above were carried out using these TS minigenes.

**Regulation of the mouse TS gene in growth-stimulated cells.**

As mentioned earlier, expression of the TS gene is tightly regulated during growth stimulation. The primary goal of the research in our laboratory is to unravel the mechanisms that are responsible for regulating TS gene expression in growth-stimulated cells. It was shown that when resting mouse 3T6 cells are stimulated with serum, the amount of TS enzyme increases about 10-fold between 10 to 25 hours (Navalgund et al., 1980). Northern blot analyses demonstrated that when LU3-7 cells progress from G0 to S phase, TS mRNA increases 20 to 40-fold (Jenh et al., 1985a). Pulse-chase studies have shown that the half-life of TS mRNA is about 8 hours and does not fluctuate in response to growth stimulation (Jenh et al., 1985a). In addition, no evidence was found for differential translation of TS message during serum stimulation (Jenh et al., 1985a). These results suggest that the growth-regulated expression of the TS gene is controlled at the level of transcription, or at the level of RNA processing, or a combination of both. In line with this
conclusion, it was demonstrated that there is a differential partitioning between relative stable poly (A)+ species and a labile poly (A)- species of TS RNA during growth stimulation, suggesting that the TS gene expression may be, at least in part, controlled at the level of the 3' processing.

Nuclear run-on analyses, at 5 hour-intervals, showed that there is only a small (3-fold) increase in the rate of TS transcription during S phase, indicating that the TS gene expression is mainly controlled at the level of RNA processing. However, studies on other S-phase genes showed that when time points of the nuclear run-on experiments were chosen to span the G1-S phase interface, a larger (6 to 7-fold) increase in the rate of transcription was observed in growth-stimulated cells (Steward et al., 1987). Therefore, it is possible that the TS gene expression may be controlled at both transcriptional and posttranscriptional levels in growth-stimulated cells.

The subject of my dissertation research was to investigate the regulatory mechanism for expression of the mouse TS gene. In particular, I wished to identify the regulatory elements which are responsible for both basal level and growth-regulated TS gene expression. Along with other members in the lab, I have identified and analyzed the promoter elements that are required for efficient expression of the mouse TS gene. We found that the essential sequences for TS gene expression are in close proximity to the first transcriptional start site. Three proteins have been found to bind this region, but only one appears to be important for promoter activity. I have also investigated the role of the TS introns in TS gene expression and found that most TS introns can stimulate the gene expression and that the stimulatory effect is not restricted to minigenes that contain the TS promoter or
polyadenylation signal. To determine the role of individual gene components in both basal level and growth-regulated expression of the TS gene, I have constructed a series of chimeric TS minigenes. From transient expression analyses I found that the low level of expression of the mouse TS gene is not due to a weak promoter. To study the growth regulation of the TS minigene constructions, I analyzed their expression in stably transfected 3T6 cells. The results indicate that the 5'-flanking region of the TS gene is necessary but not sufficient for normal regulation in growth-stimulated cells, whereas the 3'-flanking region does not appear to be essential for normal regulation.
MATERIALS AND METHODS

Bacterial culture

E. coli NM522 or XL1-Blue were grown in 2xYT medium (16 g Bacto tryptone, 10 g Bacto yeast extract, and 5 g NaCl per liter) (Ausubel et al., 1989). NM522 or XL1-Blue containing ampicillin resistant plasmid were grown in 2xYT medium supplemented with ampicillin (50 μg/ml). To ensure the presence of F factor, XL1-Blue was maintained on an agar plate containing tetracyclin (15 μg/ml) and E. coli CJ236 was maintained on an agar plate containing chloramphenicol (30 μg/ml).

Enzymes and reagents

All enzymes and reagents were purchased from established chemical and biotechnology companies.

Bacterial cell transformation and transfection

Transformation was performed according to the standard CaCl₂ procedure (Maniatis et al., 1982; Ausubel et al., 1989) with minor modifications. Competent cells were made as follows. 50 ml of 2xYT medium was inoculated with 0.5 ml of an overnight bacterial culture and shaken at 37 °C until the cells reached early log phase (2 to 4 hours). The culture was chilled on ice for 15 minutes and spun down at 4,000 rpm for 5 minutes at 4 °C. The pellet was then resuspended in 25 ml of sterile ice-cold 100 mM
MgCl₂ solution and spun down as before. The cells were resuspended in 25 ml of sterile ice-cold 100 mM CaCl₂ solution and kept on ice for 30 to 45 minutes. The cells were then spun down as before and resuspended again in 5 ml 100 mM CaCl₂ and used immediately (usually 100 µl of cells per transformation). The remaining cells were mixed with sterile glycerol to a final concentration of 15% and stored in 100 µl aliquots at -70 °C. An appropriate amount of DNA was added to the competent cells, mixed well, and kept on ice for 30 minutes. The cells were then heat-shocked at 42 °C for 2 minutes or 37 °C for 5 minutes. 1 ml of 2xYT was added to the cells and the culture was incubated for 45 minutes in a 37 °C incubator to allow the bacteria to express the antibiotic resistance. Portions of the transformation mixture were spread on an agar plate containing ampicillin (50 µg/ml). The plate was then incubated overnight in a 37 °C incubator.

Transfection was performed the same way as transformation until the heat-shock step. After the cells were heat-shocked, 0.2 ml of the overnight culture of the same bacteria and 3 ml of a soft YT agar were added to the transfection mixture, mixed, and poured onto a warm (room temperature) YT agar plate immediately. The plate was left at room temperature until the soft agar became solid and incubated overnight in a 37 °C incubator. If blue/white color selection was necessary, 10 µl of 100 mM IPTG and 30 µl of 2% X-gal were added to the transfection mixture before pouring onto the plate.

**Isolation of plasmid and M13 RF (replication form) DNA**

Plasmid DNA preparations were obtained by the alkaline lysis method as described by Maniatis et al., (1982) with minor modifications. Briefly, 3 ml
of 2xYT medium containing 150 μg of ampicillin was inoculated with a single
coli and grown overnight at 37 °C in a shaker with a speed of 250 rpm. 1
ml of the overnight culture was then transferred to a microtube and the cells
were collected by centrifugation in a microcentrifuge for 1 minute. The
supernatant was discarded by aspiration. The cell pellet was resuspended in
100 μl of suspension buffer (50 mM glucose, 10 mM EDTA, and 25 mM Tris-
Cl, pH 8.0). The cells were then lysed by gently mixing with 200 μl of freshly
prepared lysis buffer (1% SDS and 0.2 N NaOH) and kept on ice for 5
minutes. To separate the plasmid DNA from other macromolecules, 150 μl of
high salt buffer (5M KOAc, pH 4.8) was added to the mixture, mixed well, and
kept on ice for 10 minutes. The insoluble content was removed by
centrifugation for 5 minute and the supernatant containing the plasmid DNA
was transferred to a new microtube. The plasmid DNA was precipitated by
adding 2.5 volumes of 95% ethanol and spun down for 5 minutes. The pellet
was washed with 70% ethanol and dried in a Speed Vac concentrator. The
dried pellet was then dissolved in 30 μl of TE buffer (10 mM Tris-Cl, 1 mM
EDTA, pH 8.0) and used directly for restriction analysis or subcloning. If
necessary, RNA was removed by DNase-free pancreatic RNase A (20 μg/ml)
digestion for 1 hour at 37 °C.

M13 RF DNA was purified as follows. 3 ml of 2xYT medium was
inoculated with 30 μl of a fresh overnight culture (any E.coli strains containing
F factor) and allowed to grow to early log phase (approximately 1 to 2 hours
until slightly opaque). A single plaque was then transferred into the above
culture using a sterile toothpick and incubated with shaking at 37 °C for about
8 hours. The cells were collected for M13 RF DNA isolation, whereas the
supernatant was used for M13 single-stranded DNA isolation. For M13 RF DNA isolation, the rest of the procedures were the same as for preparation of plasmid DNA.

**Isolation of M13 single-stranded DNA**

M13 single-stranded DNA was isolated as described (Ausubel et al., 1989) with minor modifications. The inoculation and incubation were the same as for the preparation of M13 RF DNA. 1 ml of the culture was transferred to a microtube and the cells were pelleted by centrifugation for 1 minute. The supernatant was then transferred to a new microtube. To precipitate the M13 phage, 200 μl of 20% polyethylene glycol (PEG)-8000, 2.5 M NaCl was added to the supernatant, mixed well, and allowed to incubate at room temperature for 30 minutes. The phage was then precipitated by spinning for 5 minutes and the supernatant was removed by aspiration. A respinning is necessary to completely remove any traces of PEG. The phage pellet was resuspended in 100 μl of TE buffer and the phage protein was then removed by a single phenol extraction. Traces of phenol were removed by diethyl ether extraction and the remaining ether was evaporated for several minutes. The single stranded DNA was then precipitated by addition of 2.5 volumes of 95% ethanol and 10 μl of 3 M NaOAc, pH 4.8. The precipitation took place at -70 °C for 20 minutes or at -20 °C for more than two hours. The pellet was then collected by centrifugation, washed with 70% ethanol, and dried in a Speed Vac concentrator. The resulting DNA pellet was dissolved in 20 μl of TE buffer and used for DNA sequence analysis or site-specific mutagenesis.
**DNA sequencing**

DNA sequencing was carried out by the chain termination method of Sanger (1977) with a sequencing kit from United States Biochemical Corp. (USB). The experiment was done by following the protocol provided by the company.

**Oligonucleotide-directed mutagenesis**

Oligonucleotide-directed mutagenesis was performed by the method of Kunkel (1985) and was done according to the protocol provided by Ausubel et al., (1989). Uracil-containing single-stranded DNA template was prepared as follows. 3 ml of 2xYT medium containing chloramphenicol (15 µg/ml) was inoculated with a single colony of E.coli CJ236 and incubated with shaking at 37 °C overnight. 10 ml of 2xYT medium was inoculated with 100 µl of the overnight culture and allowed to grow to early log phase (approximately 1 to 2 hours until slightly opaque). A single plaque which contains the template for mutagenesis was transferred into 1 ml of 2xYT medium using a sterile Pasteur pipet, heated at 65 °C for 5 minutes to kill the original host bacteria, and vortexed to release the phage. The cells and agar were removed by centrifugation for 2 minutes. 100 µl of the supernatant was transferred to above 10 ml of CJ236 culture and incubated with shaking at 37 °C for about 8 hours. The uracil-containing single-stranded DNA was isolated as described in the previous section. For the annealing reaction, 600 ng of uracil-containing single-stranded DNA template generated from E. coli CJ236, and 20 ng of phosphorylated mutagenesis primer were mixed in 10 µl of annealing buffer,
incubated at 70 °C for 2 minutes, and allowed to cool to 30 °C for 30 minutes. The polymerization and ligation reactions were performed in a 20 µl mixture containing 10 µl of the above annealing mix, 4 µl of 5x polymerase mix (100 mM Tris-Cl, pH 8.8, 10 mM DTT, 50 mM MgCl₂, 2.5 mM each of dATP, dTTP, dGTP, dCTP, and 5 mM ATP), 2.5 units of T4 DNA polymerase, and 2 units of T4 DNA ligase. The mixture was first kept on ice for 5 minutes, then at room temperature for 5 minutes to stabilize the initial duplex between the template and mutagenic primer, and then incubated at 37 °C for 2 hours to complete the reaction. 1 to 2 µl of the reaction product was used to check the result of the reaction on a 1% agarose gel. If the reaction was successful (appearance of the relaxed closed circular double stranded DNA), another 1 µl of the reaction mixture was used to transform E. coli XL1-Blue. Single-stranded DNA was isolated from individual plaques for DNA sequence analysis.

To generate single-stranded DNA template for mutagenesis of the 5' flanking region of the mouse TS gene, the SphI-SphI fragment of pTTT (pTSMG2) was subcloned into the SphI site of RF M13 mp18. The mutations generated as described above were screened by both restriction analysis and DNA sequence analysis. After the correct mutations were generated, the mutated SphI fragments were then used to replace the corresponding SphI fragment of pTI₅₆T(-250).

To obtain single-stranded DNA templates for the mutagenesis to generate pTI₁T, the PstI-PstI fragment of pTTT containing the first four TS exons and the Sacl-HindIII fragment of pTI₁₂T containing part of intron 1, exon 2, and part of intron 2 were subcloned into RF M13 mp18 at the PstI site and the Sacl/HindIII sites, respectively. After the mutations were generated, the
mutated Bsu36I-Apal fragment within the PstI-PstI fragment was used to replace the corresponding fragment of pTTT and the mutated Sacl-HindIII fragment was used to replace the corresponding fragment of pTl12T.

The mutations in the final plasmids were confirmed by double stranded DNA sequence analysis (Chen et al., 1985; Ausubel et al., 1989) and restriction analysis.

All of the oligonucleotides were synthesized on an Applied Biosystem 381A DNA synthesizer at the Biochemical Instrument Center of The Ohio State University. The sequences of the mutagenic oligonucleotides will be described in the Results section.

**Subcloning**

All of the procedures involved in the DNA subcloning, including restriction analysis, electrophoresis, isolation of DNA fragments from low melting agarose gel, and ligation of DNA fragments were performed according to standard protocols (Maniatis et al., 1982; Ausubel et al., 1989). The detailed strategies for specific plasmid constructions are described in the Results section.

**Cell culture**

Mouse 3T6 fibroblasts (Todaro and Green, 1963) were maintained on plastic petri dishes (Falcon) in Dulbecco’s modified Eagle’s medium (DMEM, Gibco) supplemented with 10% calf serum (Colorado Serum). The thymidylate synthase-deficient (ts-) V79 Chinese hamster cells (Nussbaum et
were maintained in medium supplemented with 10% NuSerum (Collaborative Research) and 10 μM thymidine.

Cells were synchronized in the cell cycle using a serum-stimulation protocol (Navalgund et al., 1980). Resting (G₀) stably transfected 3T6 cells were prepared by seeding about 4 x 10⁶ cells per 100 mm plate in medium supplemented with 0.5 % calf serum. The medium was changed at 48-h intervals until day 7. The cells were then allowed to reenter the cell cycle by replacing the medium with 10% calf serum.

**Mammalian cell transfection**

1. **Transient transfection**

   The transient transfection experiment was done using the calcium phosphate method described by Wigler et al., (1979) with a few modifications. ts- hamster V79 cells were first seeded to about 30-50 % confluence on a 100 mm tissue culture plate. About four hours before the transfection, the cells were fed with fresh medium. 8.5 to 12 pmoles of TS minigenes (about 20 to 40 μg of DNA) along with 5 μg of pSV2cat (control plasmid), or 3.0 pmole of pTcat along with 10 μg of pTi₅₆T (control plasmid) in 375 μl of water were mixed with 125 μl of 1M CaCl₂. The mixture was then added dropwise to 0.5 ml of 2 x HEPES buffered saline (2 x HBS, 0.28 M NaCl, 0.05 M HEPES, and 1.5 mM Na₂HPO₄•2H₂O, pH 7.12) and mixed gently. The pH of the 2 x HBS is very critical for the transfection efficiency. Therefore, frequent checking of the pH of the solution is necessary. The mixture was incubated at room temperature for 20 to 45 minutes and added to the plate prepared previously. The DNA precipitates remained on the plate about 10 to 15 hours and were
removed by washing the plate with serum free medium. The cells were then fed with fresh complete medium and harvested two days after the transfection.

2. Stable transfection

Stably transfected 3T6 cells were selected by cotransfecting a plasmid containing the bacterial neo gene (that confers resistance to the antibiotic G418) along with the TS minigene in a ratio of about 1:40. Before the transfection, both TS minigene and the G418 resistance plasmid were linearized at a restriction site flanking the gene to increase the transfection efficiency. The digestion reaction was followed by ethanol precipitation to sterilize the DNA. The ethanol was then removed by air drying the pellet in a laminar flow hood. Transfection were performed as described before. About 16 h after the transfection, the medium was removed, and the cells were shocked with medium containing 15% glycerol for 60 to 90 seconds (Lopata, et al., 1984). The cells were then washed with serum free medium to remove the remaining glycerol. About one day later, the cells were split 1:4 and fed with medium containing 10% calf serum and G418 (Gibco, 400 µg/ml). The medium was changed twice a week until the appearance of G418 resistant colonies (usually takes about 10 to 14 days). Clones of cells that were resistant to G418 were pooled and maintained as a mass culture for the regulation study.

Isolation of cytoplasmic RNA

Isolation of cytoplasmic RNA was performed according to Maniatis et al., (1982). Cells from tissue culture plates were washed three times with ice-cold phosphate buffer saline (PBS, 140 mM NaCl, 3 mM KCl, 8 mM Na₂HPO₄, 2
mM KH$_2$PO$_4$ pH 7.4) and scraped from the plates with a rubber policeman. The cells were collected in a 12 ml conical plastic tube and pelleted by centrifugation at 2,000 rpm for 2 minutes at 4 °C. After removing the supernatant, the cells were lysed in 2 ml of reticulocyte standard buffer (RSB, 10 mM NaCl, 3 mM MgCl$_2$, 10 mM Tris-Cl pH 8.4, and 0.5 % NP 40) containing 50 µl of 200 mM vanadyl ribonucleoside complexes (VRC). The mixture was vortexed for about 30 seconds and kept on ice for 5 minutes to complete the lysis process. The nuclei and other cell debris were removed by centrifugation at 2,000 rpm for 2 minutes at 4 °C. The supernatant containing the cytoplasmic RNA was collected and mixed with 0.5 ml of 5x SDS buffer (0.5 M NaCl, 5 mM EDTA, 2.5% SDS, and 50 mM Tris-Cl, pH 7.5). The cytoplasmic RNA was purified first by proteinase K (100 µg/ml) digestion at 37 °C for 1 hour and then by three phenol-chloroform-isoamyl alcohol extractions and ethanol precipitation.

**Isolation of poly(A)+ RNA**

The poly(A)+ RNA was isolated by the protocol described by Berger et al. (1987). All of the solutions used in poly (A)+ RNA isolation were treated with DEPC (Maniatis et al., 1982). Cytoplasmic RNA was dissolved in 1 ml of elution buffer (10 mM Tris-Cl, pH 7.5, and 1 mM EDTA) in a 12 ml Corex tube and heated at 65 °C for 5 minutes. 1 ml of 2x binding buffer (20 mM Tris-Cl, pH 7.5, 1 M NaCl, 2 mM EDTA, and 1% SDS) was added to the RNA sample and applied to an oligo(dT)-cellulose column which had been washed with more than 10 volumes of binding buffer (10 mM Tris-Cl, pH 7.5, 0.5 M NaCl, 1 mM EDTA, and 0.5% SDS). The sample was reapplied to the column twice.
and washed once with 10 volumes of binding buffer and once with 10 volumes of washing buffer (10 mM Tris-Cl, pH 7.5, 0.2 M NaCl, and 1 mM EDTA). The poly(A)+ RNA was eluted with 2 ml of 65 °C elution buffer. The amount of poly(A)+ RNA was determined by measuring absorbance at 260 nm and then precipitated by addition of 0.2 ml of NaOAc, pH 4.8, 2.5 volumes of ice-cold 95% ethanol, and 100 µg of yeast tRNA. The precipitation was done at -20 °C for at least two hours or at -70 °C for more than 20 minutes.

5' end labeling

The 5' end-labeled DNA probes for the S1 analyses were prepared essentially as described by Ausubel et al., (1989). Briefly, 20 µg DNA was linearized by a suitable restriction enzyme. The linearized DNA was purified either by ethanol precipitation or by phenol extraction and ethanol precipitation, depending on the restriction enzyme used. The purified DNA was then dephosphorylated in a 50 µl mixture containing 5 µl 10x calf intestine phosphatase buffer (0.2 M Tris-Cl, pH 8.0, 10 mM MgCl₂, and 10 mM ZnCl₂) and 1 unit of calf intestine phosphatase, at 37 °C for 30 minutes. The reaction was terminated by heat inactivation of the phosphatase at 75 °C for 10 minutes and phenol extraction to remove the phosphatase. The labeling reaction was carried out in a 50 µl mixture containing the above dephosphorylated DNA fragment, 5 µl of 10x T4 kinase buffer (0.5 M Tris-Cl, pH 7.5, 0.1 M MgCl₂, and 50 mM EDTA), 400 µCi [γ-³²P] ATP (specific activity 5000-7000 Ci/mmol), and 30 units of T4 polynucleotide kinase at 30 °C for 1 hour. The reaction was stopped by addition of 30 µl of 4 M NH₄OAc. The labeled DNA was purified by phenol extraction followed by ethanol
precipitation. After the second restriction enzyme digestion, the desired DNA fragment was isolated by 1 % agarose gel electrophoresis and purified from the gel fragment by the Geneclean method as described by the manufacturer (Bio 101, CA). The amount of radioactivity in the probe was determined by a liquid scintillation counter. Specific activities generally were about 1 x 10^7 cpm/µg DNA.

**S1 nuclease protection assays**

S1 analysis was carried out as described by Favaloro et al., (1980) with minor modifications. Briefly, 1 to 8 µg of poly(A)+ RNA sample and 100 µg of yeast tRNA (carrier RNA) were coprecipitated with 1 x 10^5 cpm of a ^32P end-labeled double stranded DNA probe (about 10 ng). The probe was at least 100-fold excess relative to the TS mRNA. The dried pellet was resuspended in 30 µl of hybridization buffer (40 mM PIPES, pH 6.4, 400 mM NaOH, 1 mM EDTA, and 80 % deionized formamide). The mixture was incubated at 85 °C for 15 minutes and then quickly transferred to a 56 °C water bath. The hybridization reaction was carried out for 12 to 18 hours. S1 digestion was done by adding 300 µl of ice-cold S1 reaction buffer (280 mM NaCl, 30 mM NaOAc pH 4.4, and 4.5 mM ZnOAc₂) and 100 units of S1 nuclease to the above hybridization mixture while it was still submerged in the 56 °C water bath, and then incubating at 37 °C for 30 minutes. The reaction was stopped by addition of 70 µl of termination buffer (2.5 M NH₄OAc and 50 mM EDTA). The protected probe was precipitated by addition of 20 µg of yeast tRNA and 400 µl of isopropanol and incubated at -70 °C for 20 minutes. The dried pellet was dissolved in 6 µl of NaOH loading dye (80% formamide, 10 mM NaOH, 1
mM EDTA, 0.1% xylene cyanol, 0.1% bromophenol blue, and 0.1% orange G) and the S1 digestion results were analyzed on a 6% denaturing polyacrylamide sequencing gel.

Enzyme assays
1. [3H] fluorodeoxyuridylic acid (FdUMP) binding assay (TS assay)

The TS assay (measuring the formation of a covalent ternary complex between TS, FdUMP and 5,10-methylenetetrahydrofolic acid) was performed as described previously (Lockshin et al., 1979; Jenh et al., 1985c). The cell extract was prepared as follows. The cells from the culture plates were washed three times with ice-cold phosphate buffered saline (PBS, 140 mM NaCl, 3 mM KCl, 8 mM Na\(_2\)HPO\(_4\), and 2 mM KH\(_2\)PO\(_4\), pH 7.4) and scraped from the plates with a rubber policeman. The cells were transferred to a 10 ml Corex glass tube and collected by centrifugation at 2,000 rpm for 2 minutes at 4 °C. The cells were lysed in 500 µl of reaction buffer (50 mM Tris-Cl, pH 7.3 and 0.5 % NP 40). The mixture was vortexed for about 30 seconds and incubated on ice for 5 minutes to complete the lysis process. The nuclei and other cell debris were removed by centrifugation at 2,000 rpm for 2 minutes at 4 °C. The supernatant was either used for the enzyme assay immediately or stored in a -70 °C freezer. The enzyme reaction was carried out in a 100 µl mixture containing 90 µl of cell extract and 10 µl of substrate mixture (0.02 µCi [3H] FdUMP with 20 Ci/mmol, 740 µM tetrahydrofolate, 10 mM DTT, and 50 mM Tris-Cl, pH 8.0) at room temperature for 30 to 45 minutes. The ternary complexes were precipitated by addition of more than 10 volumes of 10% ice-cold TCA and incubated at 4 °C for more than 1 hour. The radioactive ternary
complexes were separated from the free $[^3H]$-FdUMP by nitrocellulose filtration. The radioactivity of the precipitates was then determined by a liquid scintillation counter.

2. Chloramphenicol acetyl-transferase (CAT) assay

The CAT assay (measuring the percentage conversion of $[^{14}C]$-chloramphenicol to acetylated derivatives) was performed as described by Gorman et al., (1982) with minor modifications. The preparation of cell extract was the same as described above. Prior to the reaction, 55 μl of cell extract plus 70 μl of 1 M Tris-Cl, pH 8.0 solution was heated at 65 °C for 10 minutes to increase the reaction sensitivity. The reaction was carried out by incubating the mixture containing 55 μl of cell extract in 70 μl of 1 M Tris-Cl, pH 8.0 solution, 20 μl of 4 mM acetyl-CoA, and 1 μl of $[^{14}C]$-chloramphenicol (0.05 μCi, 50 Ci/mmol) at 37 °C for about 30 to 60 minutes. The $[^{14}C]$-chloramphenicol products were then extracted with 1 ml of ethyl acetate. After a brief centrifugation, the upper (organic) phase was collected and evaporated under vacuum. The lower aqueous phase was discarded. The dried reaction products were redissolved in 20 μl of ethyl acetate. A 10 μl aliquot was spotted on a silica TLC plate and chromatography was carried out in chloroform-methanol (19:1) in a closed glass tank for about 30 minutes. The plate was then dried and autoradiographed. After autoradiography the spots were cut out from the plate and counted to obtain the conversion values. The conversion values were then used to correct for differences in transfection efficiency.
RESULTS

Analysis of the 5'-flanking sequences of the mouse TS gene

The promoter of the mouse TS gene, like that of many other housekeeping genes, is very GC rich, has no apparent TATA and CAAT boxes and has multiple transcriptional initiation sites (Deng et al., 1986). It has been speculated that the biochemical mechanism of transcriptional initiation of a gene with a TATA-containing promoter might be different from that with a TATA-less promoter (Dynan, 1986). Thus, studies of the role of the 5'-flanking sequences in mouse TS gene expression are not only important for elucidating the mechanism of mouse TS gene regulation but they can also provide a model system for studying the mechanism of transcriptional initiation of genes with TATA-less promoters.

To facilitate the study of mouse TS gene expression, James Dewille constructed an intronless TS minigene (pTSMG2) (Fig. 1). This minigene contains 1 kb of the 5'-flanking sequence and 0.25 kb of the 3'-flanking region of the TS gene linked to TS cDNA at common restriction sites in the first and last exons (DeWille et al., 1988). Transient expression analysis revealed that this minigene is expressed at a very low level when transfected into ts-hamster V79 cells. However, inclusion of introns 5 and 6 at their normal locations in the coding region of pTSMG2 (pI56) (Fig. 1) stimulates the level of expression about 8 to 10-fold (Deng et al., 1989b). Most of the following
Fig. 2. Point mutations and deletions of the mouse TS promoter. A portion of the 5'-flanking sequences is shown. Numbering starts at the A of the ATG initiation codon. The approximate location of the major (V) or minor (v) transcriptional start sites are indicated. The locations of the potential binding sites for various transcriptional factors are indicated. The nucleotide changes introduced by site-specific mutagenesis are indicated as (*). The deletion endpoints are shown as (:).
Figure 1
analyses were performed using the manipulated TS minigenes derived from both pTSMG2 and pI56.

**Analysis of the upstream GC box and USF element by site-specific mutagenesis.** A previous study has shown that when the 5'-flanking sequence of pTSMG2 was deleted to -150 relative to the ATG translational start codon, the activity of the minigene was the same as that of pTSMG2. But when it was deleted to -54, the minigene became inactive (DeWille et al., 1988). This result suggested that the region between -150 and -54 is essential for TS gene expression. DNA sequence analysis revealed that this region does contain several potential binding sites for a variety of transcriptional factors (Deng et al., 1986), including binding sites for the cellular transcription factors Sp1 and USF (Fig. 2). These transcription elements are known to be important for expression of a wide variety of cellular and viral genes (Dynan and Tjian, 1985a; 1985b; Kadonaga et al., 1986; Sawadogo and Roeder, 1985).

To determine whether these proteins interact with the TS promoter, Keith Jolliff performed DNase I footprint analyses and the results showed that Sp1 protected a 13-base-pair region from -123 to -136, whereas USF protected a 19-base-pair region from -125 to -144 (Deng et al., 1989a). To determine the biological significance of these results, Tiliang Deng generated several specific point mutations in these elements both in pTSMG2 and pI56 (Fig. 2). He then analyzed the effects of the mutations by measuring the TS enzyme level of the mutated pTSMG2 minigenes. His results showed that inactivation of the USF element led to about a 2-fold increase in the TS enzyme level,
Fig. 2. Point mutations and deletions of the mouse TS promoter. A portion of the 5'-flanking sequences is shown. Numbering starts at the A of the ATG initiation codon. The approximate location of the major (V) or minor (v) transcriptional start sites are indicated. The locations of the potential binding sites for various transcriptional factors are indicated. The nucleotide changes introduced by site-specific mutagenesis are indicated as (*). The deletion endpoints are shown as (:).
Figure 2
whereas inactivation of the GC box(-130) or both the GC box(-130) and the USF site led to a 3-fold reduction in TS enzyme level (Deng et al., 1989a), indicating that the GC box(-130) functions as a positive element in the TS promoter but is not essential for TS gene activity. However the significance of the two fold increase in the enzyme level by the inactivation of USF site was difficult to interpret. Since USF and GC box(-130) are adjacent to each other (Fig. 2), it is possible that inactivation of the USF binding site would eliminate the binding competition between the USF and GC box binding proteins to their binding sites. Alternatively, the two-fold increase in gene expression is simply within the experimental error (due to the fact that the expression level of pTSMG2 was very low).

To further determine the effects of these point mutations on TS promoter activity, I analyzed expression of mutated pl56. The expression level of pl56 is about 8 to 10 times higher than that of pTSMG2 (discussed in a later section). This was done by transiently transfecting the altered pl56 into ts-hamster V79 cells and harvesting the cells two days later. The expression of the minigenes was then determined by measuring the TS enzyme level. Table 1a shows that inactivation of the USF element had little effect on TS promoter activity, which is different from the result observed from pTSMG2. However, inactivation of the GC box(-130) led to a 3-fold reduction in promoter activity. A promoter with alterations in both the USF and GC box(-130) had about the same activity as the GC box-deficient promoter. These results suggest that neither element is essential for TS gene expression. The GC box(-130) functions as a positive element in TS promoter, whereas USF binding site does not appear to be important for TS promoter activity.
Table 1
Effects of promoter mutations on TS minigene expression

<table>
<thead>
<tr>
<th>Minigenes b</th>
<th>TS enzyme level</th>
<th>TS mRNA level c</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pl156 (-1 kb)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>-USF</td>
<td>0.88 ± 0.30</td>
<td>0.82 ± 0.06</td>
</tr>
<tr>
<td>-GC box(-130)</td>
<td>0.40 ± 0.12</td>
<td>0.30 ± 0.16</td>
</tr>
<tr>
<td>-USF and GC box(-130)</td>
<td>0.32 ± 0.12</td>
<td>0.37 ± 0.08</td>
</tr>
<tr>
<td>Part b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pl156 (-1kb or -250)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>-CII</td>
<td>1.37 ± 0.16</td>
<td>1.39, 1.04</td>
</tr>
<tr>
<td>-GC box(-80)</td>
<td>0.36 ± 0.07</td>
<td>0.35, 0.49</td>
</tr>
<tr>
<td>+GC box(-80)</td>
<td>0.67 ± 0.26</td>
<td>0.56, 0.91</td>
</tr>
<tr>
<td>-CII</td>
<td>0.08 ± 0.03</td>
<td>0.04 ± 0.01</td>
</tr>
</tbody>
</table>

a The minigenes that contain the mutated promoters were transiently transfected into ts- hamster V79 cells along with pSV2cat as a control gene. TS enzyme and mRNA levels were normalized to CAT enzyme level (to correct for differences in transfection efficiency). This value was normalized to the value obtained with the normal promoter, which was assigned as 1.00. Data are the average (± standard deviation) of at least three independent experiments except some of the mRNA levels in part b were determined twice, and both values are presented.

b The 5' flanking sequence of the minigenes extends either to -1kb or to -250 nucleotides upstream of the ATG codon.

c RNA was quantitated by densitometric analysis of S1 autoradiograms.
As mentioned earlier, one of the common features of the promoters of many housekeeping genes is that they are very GC-rich. This is usually associated with the presence of multiple GC boxes in the promoter region (Dynan, 1986; Bird, 1986; 1987). Thus, determining the functional role of the GC box is very important for understanding the transcriptional regulation of genes with TATA-less promoters. It has been proposed that GC boxes may have a role in specifying the start site for promoters lacking the TATA sequence. A recent study on hamster dihydrofolate reductase (DHFR) gene has actually demonstrated that GC box interactions not only are required for efficient transcription but also regulate the start site utilization in the TATA-less promoter of hamster DHFR gene (Blake et al., 1990).

To investigate the effect of the mutations on the pattern of transcriptional start sites in the TS promoter, and also to determine whether the differences in enzyme levels were due to differences in mRNA levels, I analyzed the content and 5'-terminal structures of TS mRNA produced from altered pl56 by S1 nuclease protection assay. Densitometric analyses indicated that the level of TS mRNA correlated well with the TS enzyme level (Table 1a), suggesting that the changes in TS enzyme level resulted from the changes in the amount of TS mRNA.

From a first look at the S1 result it seems that there was no change in the pattern of start sites by the inactivation of GC box at -130 when compared with that of pl56 (Fig. 3). However, a closer examination by densitometric analysis revealed that inactivation of the GC box at -130 not only decreases the amount of mRNA but also alters the relative utilizations of individual start sites without changing their positions (Fig. 3 and 4). Table 2 shows that the
Fig. 3. Effect of specific point mutations on the pattern of transcriptional start sites. Cytoplasmic poly (A)+ mRNA was isolated from ts- hamster V79 cells that were transiently transfected with indicated TS minigenes. S1 nuclease protection assays were performed with 4 µg of mRNA from cells transfected with TS minigenes containing the wild-type (WT) promoter or promoters with the indicated point mutations. The probe (probe A, see Fig 11) was derived from pTSMG2. It was 5' end labeled at the promoter-proximate BamHI site in the coding region and extended to the XbaI site in the 5'-flanking region. Digestion products were analyzed on a 6 % polyacrylamide sequencing gel and detected by autoradiography.
Figure 3

WT
- GC(-130)
- USF
- BOTH
Fig. 4. Densitometer scans of the S1 autoradiograms from the GC box mutations. The Data analysis is shown in Table 2. The first upstream major start site and the last downstream major start site are labeled as I and II, respectively.
Figure 4
Table 2
Effect of mutations in GC boxes on the pattern of transcriptional start sites

<table>
<thead>
<tr>
<th>Minigenes&lt;sup&gt;a&lt;/sup&gt;</th>
<th>The ratio of transcriptional start sites utilization&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>pl56(-1 kb or -250)</td>
<td>1.0</td>
</tr>
<tr>
<td>-GC box(-130)</td>
<td>0.74, 0.68</td>
</tr>
<tr>
<td>-GC box(-80)</td>
<td>9.2, 6.3</td>
</tr>
</tbody>
</table>

<sup>a</sup> The 5' flanking sequence of the minigenes extends either to -1 kb or to -250 nucleotides upstream of the ATG codon.

<sup>b</sup> The ratio is the utilization of the first upstream major start site to that of the last downstream major start site. The quantitation was obtained by densitometric analysis of S1 autoradiograms (Fig. 4). All values were normalized to those of pl56(-1 kb or -250), which was designated as 1.0 to facilitate the comparison. The experiments were performed twice and both values are presented.
inactivation of this GC box decreases the transcription more from the proximal start sites (about 40 to 50 bp from this GC box) than the distal start sites (about 80 to 100 bp from this GC box). Similar observations were made from the study of hamster DHFR gene (Blake et al., 1990). These observations indicate that this GC box is not only necessary for efficient expression of TS gene, but it is also important for specifying the frequency of transcriptional initiation at various sites.

**Analysis of the TS promoter by deletion mutagenesis.** To further identify the DNA sequences that are essential for the TS promoter activity, additional deletions in the region between -150 and -54 were generated both in pTSMG2 and pl56 by Tiliang Deng (Fig. 2). Again, he analyzed the effects of the deletions in pTSMG2 and I analyzed the effects in pl56 by transient expression assays as described in the previous section. The results from the two minigenes were in good agreement (Deng et al., 1989a). In addition, I also analyzed the effects of the deletions on the level of TS mRNA and the pattern of the transcriptional start sites by S1 nuclease protection assays. Table 3 shows that a promoter deleted to -113 (which eliminated both the GC box (-130) and the USF elements) had almost the same activity as that with the -150 deletion. Deletion to -105 led to a slight decrease in gene expression. However, when the sequences upstream of -85 were eliminated, the gene was almost completely inactivated. S1 nuclease protection analyses demonstrated that there was a good agreement between the relative amount of mRNA and the relative amount of TS enzyme derived from each minigenes (Table 3, Fig. 5).
Table 3
Effects of promoter deletions on TS minigene expression

<table>
<thead>
<tr>
<th>Minigenes</th>
<th>TS enzyme level</th>
<th>TS mRNA level</th>
</tr>
</thead>
<tbody>
<tr>
<td>pl56 (-150)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>-113</td>
<td>0.76, 0.74</td>
<td>1.2, 0.71</td>
</tr>
<tr>
<td>-105</td>
<td>0.63, 0.51</td>
<td>0.78, 0.65</td>
</tr>
<tr>
<td>-85</td>
<td>0.02, 0.05</td>
<td>0.0, 0.0</td>
</tr>
<tr>
<td>-54</td>
<td>0.01, 0.01</td>
<td>0.0, 0.0</td>
</tr>
</tbody>
</table>

a The experiments were performed and analyzed as described in Table 1, except that the values were normalized to those of the -150 deletion, which is expressed at about the same level as that of pl56 (date not shown).

b The experiments were performed twice, and both values are presented.

c When the amount of TS mRNA was being determined, the band corresponding to the deletion endpoint (Fig. 5) was ignored.
Fig. 5. Effect of deletion mutations on the pattern of transcriptional start sites. Cytoplasmic poly (A)+ mRNA was isolated from mouse LU3-7 fibroblasts (a cell line that overproduces TS by a factor of 50 [Rossana et al., 1982] because of gene amplification [Jenh et al., 1985]) or from ts- hamster V79 cells that were transiently transfected with the indicated TS minigenes. S1 nuclease protection assays were performed with 100 ng of mRNA from LU3-7 cells or with 4 μg of mRNA from cells transfected with TS minigenes containing the wild-type promoter or promoters with the indicated deletions. Probe A was used in this experiment (Fig.11).
Figure 5
Densitometric analyses revealed that, with the exception of an S1-resistant fragment that mapped to the deletion endpoint, there was no significant change in the pattern of transcriptional start sites for deletion to -150. This result is consistent with the observations on the effect of upstream GC box mutation (described in previous section) since deletions to -113 and -105 eliminate the upstream GC box, indicating that the upstream GC box does play a role in specifying the transcriptional initiation pattern of the TS gene. Further deletion to -85 and -54 resulted in the complete loss of the normal pattern of start sites (Fig. 5).

Above results confirmed that the USF element and GC box upstream of the -113 elements are not essential for promoter activity, and demonstrated that the sequence essential for TS promoter activity is present in the region between -105 and -85 (relative to the ATG translational start codon). This region is in the vicinity of the first transcriptional start site at about -90 and is referred to as the "essential region".

**Analysis of the essential region by site-specific mutagenesis.** Gel mobility shift and footprint analyses by Keith Jolliff with various sequences from the essential region identified three major protein-DNA interactions (Jolliff et al., 1991). One of these corresponds to Sp1 interacting with a non-consensus binding site downstream of the first transcriptional initiation site (at -80 relative to the ATG translational start codon). The proteins responsible for the other two complexes (CII and CIII) do not appear to correspond to any of the common transcriptional factors studied so far. The binding site for CII protein has been located between -99 to -91 which is very close to the first
transcriptional initiation site, whereas the CII protein binds between -99 to -109.

To confirm the locations of the recognition sites suggested by gel mobility shift and footprint analysis and to evaluate the biological significance of the protein-DNA interactions in the essential region of the TS promoter, I introduced several specific point mutations into the CII site and GC box(-80) by site-directed mutagenesis (Fig. 2). The oligonucleotides used for mutagenesis were 5' CGCCTCTTCCCTGTCGAAACTTCCGG 3' (for -CII) and 5' GCCCGCCTCTTCCCTGCTGGGAAC 3' (for -GC box at -80). The mutation in CII was created in pl56(-1 kb) by Tiliang Deng (Fig. 2). The altered promoters were inserted into pl56(-250), which is the same as pl56 except that it extends to the Sphl site 0.25 kb upstream of the ATG codon. The activities of these mutated minigenes were then determined by measuring both TS enzyme and mRNA levels. Table 1b shows that inactivation of the CII binding site had little effect on the expression of the minigene. S1 nuclease protection analysis revealed that there was no distinguishable change in the pattern of transcriptional start sites when compared with that of pl56(-250) (Fig. 6). These results demonstrated that this element does not play a major role in TS gene expression.

Inactivation of the GC box at -80 led to a 3-fold decrease in gene expression (Table 1b), which is similar to the result from the upstream GC box(-130) mutation. However, it was interesting to find that there was a more dramatic change in the pattern of transcriptional start sites when compared with the results from upstream GC box(-130) mutation (Fig. 6). The similarity
Fig. 6. Effects of mutations in the essential region on the pattern of transcriptional start sites. Cytoplasmic poly (A)+ mRNA from the transfected ts- hamster V79 cells was isolated and 2 µg of mRNA from each sample was analyzed by S1 nuclease protection assay. Probe A (Fig. 11) was used in this experiment.
between these two results is that the mutations in both GC boxes caused a
decrease in transcription from downstream start sites without altering the
locations of the start sites. However, in the case of the GC box(-80) mutation,
there is a striking increase in the frequency of utilization of the upstream start
sites (Table 2), indicating the mutation in this GC box can both decrease
transcriptional initiation from the downstream start sites and increase the
availability of upstream start sites to the transcriptional machinery. When this
non-consensus Sp1 site was changed to a consensus Sp1 site, both the
promoter activity and the initiation pattern of the altered minigene were almost
the same as that of the wild-type minigene (Table 1b, Table 2, and Fig. 6).

Inactivation of the CIII binding site had the most severe effect on gene
expression. The TS enzyme level was about 12-fold lower than that observed
with the wild-type minigene (Table 1b). However, the pattern of transcriptional
start sites was similar to that of the wild-type promoter except for the
appearance of a new start site approximately 25 nucleotides downstream of
the first T of the TATAA sequence created by this mutation (Fig. 6). This result
indicates that CIII interaction could be an important factor for controlling
transcription of the mouse TS gene.

The effects of the mutations on protein binding were determined by gel
mobility shift assay performed by Keith Jolliff using the probes generated from
the altered promoter. The results showed that the mutations at each binding
site indeed abolished the formation of the corresponding protein-DNA
complexes, which confirmed the locations of the binding site.
Analysis of the effect of introns on the expression of the mouse TS gene

As mentioned earlier, an intronless TS minigene, pTSMG2, was constructed to facilitate the study of TS gene expression (DeWille et al., 1988). However the expression level of this minigene was very low when it was transfected into ts- hamster V79 cells. There are a number of possible reasons for this low level of expression, including the elimination of introns. To test this possibility, Tiliang Deng constructed several intron-containing minigenes in which the introns were at their natural positions in the coding region (Deng et al., 1989b). He found that inclusion of introns 5 and 6 (p156) or internally deleted intron 3 (Fig. 7) increased the level of TS enzyme of pTSMG2 about 3 to 8-fold. However, inclusion of intron 4 (p14) (Fig. 7) led to a 2-fold reduction in TS enzyme level. These results indicated that TS gene expression can be stimulated by some but not all introns. In addition, he also found that the intron stimulatory effect was not additive since the inclusion of intron 3 into p156 did not lead to any significant increase in p156 expression. However, when partially deleted intron 3 as well as intact intron 4 were both included in the minigene (p13A4) (Fig. 7), the stimulatory effect of intron 3 was reduced about 2-fold.

All of his measurements of TS gene expression were based on the amount of TS enzyme. To determine whether the increased amount of TS enzyme was the result of a corresponding increase in content of TS mRNA, I transiently transfected minigenes p156, p13A4, p14 and the intronless minigene
Fig. 7. Structures of the intron-containing TS minigenes.
Constructions of pl56, pl3A4, and pl4 have been described by Tiliang Deng (Deng et al., 1989b). Constructions of pl12, pl1, and pl3456 is described in the text and the detailed construction strategy is shown in Fig. 8. Exons are indicated by the closed boxes. Thin lines represent flanking regions and intron sequences. Open boxes represent the vector sequence. // indicates that the drawing is not to scale. The restriction sites are indicated as: H, HindIII; X, XbaI; and Sc, SacI.
Figure 7
Fig. 8. S1 protection analysis of TS intron effect (I). S1 analysis were performed with: lane 1, 1 μg of total cytoplasmic mRNA from LU3-7 cells; lane 2, 20 μg of yeast tRNA. Analysis in lanes 3-9 were performed with indicated amounts of cytoplasmic poly (A)+ mRNA isolated from untransfected ts- cells (lane 3) or cells that had been transfected with the minigene as indicated. Size markers (in nucleotides) are indicated. Probe A was used in this experiment (Fig. 11).
Figure 8

<table>
<thead>
<tr>
<th>243-</th>
<th>310-</th>
<th>405-</th>
</tr>
</thead>
<tbody>
<tr>
<td>LU3-7</td>
<td>tRNA</td>
<td>ts^-</td>
</tr>
<tr>
<td>TSMG2</td>
<td>14</td>
<td>4 ug</td>
</tr>
<tr>
<td>13A4</td>
<td>156</td>
<td>2 ug</td>
</tr>
<tr>
<td>13A4</td>
<td>156</td>
<td>0.4 ug</td>
</tr>
</tbody>
</table>
### Table 4
Effects of TS introns on the expression of TS gene

<table>
<thead>
<tr>
<th>Minigenes</th>
<th>TS enzyme levels</th>
<th>TS mRNA levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>pTSMG2</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>pI56</td>
<td>8.1 ± 1.4</td>
<td>9.6 ± 0.6</td>
</tr>
<tr>
<td>pI3Δ4</td>
<td>2.4 ± 0.3</td>
<td>5.1 ± 1.7</td>
</tr>
<tr>
<td>pI4</td>
<td>0.4 ± 0.1</td>
<td>1.0 ± 0.3</td>
</tr>
<tr>
<td>pI12</td>
<td>20.8 ± 4.4</td>
<td>20.5</td>
</tr>
<tr>
<td>pI1</td>
<td>13.0 ± 3.5</td>
<td>14.0</td>
</tr>
<tr>
<td>pI3456</td>
<td>4.9 ± 1.4</td>
<td>10.3</td>
</tr>
</tbody>
</table>

Transfections and enzyme assays were performed as described in Table 1. All values were normalized to those of pTSMG2, which was designated as 1.0. Data represent the average (± standard deviation) of at least three independent experiments except the mRNA levels of pI12, pI1, and pI3456 levels were determined once. The values are normalized to pI56 and presented.
pTSMG2 into the ts- hamster cells and analyzed the cytoplasmic poly (A)+ mRNA using a 5' S1 nuclease protection assay. Fig. 8 shows that the pattern of the 5' termini of mRNA isolated from cells transfected with the intron-containing and intronless minigenes was the same as that observed in the mouse fibroblasts. Densitometric analyses revealed that there was good agreement between the increase in enzyme level and mRNA level for pl56 (Table 4). However, the level of TS mRNA was about 2-fold higher than the enzyme for pl3A4 and pl4. The reason for this discrepancy is not clear. It appears, however, that inclusion of intron 4 in the minigenes caused this reduction in TS enzyme activity.

Construction of Intron containing TS minigenes. To obtain a more complete picture of the intron effect on TS gene expression, I constructed pl12, pl1 and pl3456. The construction strategy is described in Fig. 9. pl12, which contains intact introns 1 and 2 at their normal locations in the coding region, was constructed by cloning the Xbal fragment of phage TSB-9, which contains 1 kb of the 5'-flanking region, exon 1, intron 1, exon 2, and part of exon 3 of the TS gene (Deng, et al., 1986), into the Xbal site of pBluescript M13+ (Stratagene) to create construct I. The HindIII-Apal fragment of pTSMG2 was replaced by the HindIII-Apal fragment of pTSPPHI, which contains most of intron 2 and part of exon 3 of the TS gene (Deng, et al., 1986), to form construct II. The HindIII fragment of construct I that contains the 5' portion of the TS gene was then inserted into the HindIII site of construct II to generate pl12. To construct pl11, a EcoRI site was created in exon 2 of both in pTSMG2 and pl12 by site-directed mutagenesis without changing the
Fig. 9. Constructions of intron-containing TS minigenes. The complete construction process is described in the text. Exons are indicated by closed boxes. Thin lines represent flanking region and intron sequences. Open boxes represent the vector sequence. Part A shows the construction of pl12, part B shows the construction of pl1, and part C shows the construction of pl3456. // indicates that the drawing is not to scale. The restriction sites are indicated as: A, Apal; C, Clal; E, EcoRI; H, HindIII; and X, Xbal.
Figure 9
Fig. 10. S1 protection analysis of TS intron effect (II). Cytoplasmic poly (A)+ mRNA from the transfected cells was isolated. S1 analysis was performed with 4 μg of mRNA from cells transfected indicated TS intron-containing minigenes and probe A was used in this experiment (Fig. 11).
amino acid identity. The oligonucleotide used for mutagenesis was 5' GAGCAGAGGGAATTTCATC 3'. These two plasmids were designated as pTSMG2(E) and pl12(E) respectively. The Xabl-EcoRI fragment of pTSMG2(E) which contains the 1 kb 5'-flanking sequence, exon 1 and part of exon 2 was replaced by the Xabl-EcoRI fragment of pl12(E) containing the same region and also intron 1 to generate pl1. To construct pl3456, the Apal-ClaI fragment of pTSMG2 was replaced by the Apal-ClaI fragment of pUTSB16 which contains 3' end of the TS gene (Deng et al., 1986). The structures of these intron-containing minigenes are summarized in Fig. 7.

To study the intron effect on TS gene expression, I transiently transfected these intron-containing minigenes into ts-hamster V79 cells and analyzed the activities of these minigenes by measuring both TS enzyme and mRNA levels. Table 4 shows that inclusion of introns 1 and 2 into pTSMG2 stimulates the expression about 20-fold. Intron 1 stimulates pTSMG2 expression more than 10-fold. These results indicates that the stimulatory effect is not intron specific. S1 analyses confirmed the results of TS enzyme analysis except for pl3456 (Fig. 10). Densitometric analysis revealed that the level of mRNA derived from pl3456 was almost the same as that derived from pl56, but the TS enzyme level of pl3456 was about two times less than that of pl56. This again supports the idea that the presence of intron 4 led to a reduction in TS enzyme activity.

Identification of the regulatory elements of the mouse TS gene by chimeric gene analysis

To further identify the regulatory elements of the mouse TS gene, I constructed a number of chimeric TS minigenes. The general design was to
dissect the mouse TS gene into three parts, the promoter region, coding region and the polyadenylation signal and analyze the role of each part on regulating mouse TS gene expression. The construction strategy is described in the following sections and the structures of the chimeric minigenes are summarized in Fig. 11.

**Construction of chimeric TS minigenes with alternative promoters.** The schematic representation of the construction strategy is shown in Fig. 12. To facilitate the construction of minigenes with alternative promoters, I created a HindIII site in pTSMG3 (DeWille et al., 1988)10 nucleotides upstream of the ATG translational start codon by site-directed mutagenesis using the mutagenic oligonucleotide 5' CCACCAGCATAACGAAGCTTTCTGGCAGCAGTGTAG 3'. This minigene was designated as pTSMG3(H). The rational for creating the HindIII site at this location was to eliminate as much mRNA noncoding sequence as possible but to keep the ATG translational start codon and the surrounding sequences of the TS gene since the TS initiation codon is in good agreement with the consensus sequences GXXAUGG which was proposed by Kozak (1984). To facilitate the subsequent constructions, the entire TS minigene from pTSMG3(H) was subcloned into the HindIII/Sacl site of pBluescript M13+ (Stratagene) in order to obtain several restriction sites (see Fig. 12). To construct pSTT, the SalI-HindIII fragment of above plasmid which contains the TS promoter was replaced by the 368-nucleotide SalI-HindIII fragment of the pDSP-1 (Pfarr et al., 1985) (kindly provided by M. Reff). This fragment contains the Pvull-HindIII fragment of the simian virus 40 (SV40) early region
which includes the promoter-enhancer region and transcripational start sites but lacks the translational start codon. To change the vector (from pBluescript M13+ to pUC18), the Sall-Bglll fragment of the above plasmid which contains the SV40 early promoter and part of the TS coding region was used to replace the Sall-Bglll fragment of pTTT (also named as pTSMG2). To create pSI_{56}T, the Sall-Bsu36l fragment of pTI_{56}T (also named pl_{56}) was substituted for the Sall-Bsu36l fragment of pSTT.

To construct pKTT, the 0.7-kb BamHI-Bglll fragment of the herpes simplex virus (HSV) thymidine kinase (TK) promoter (McKnight, 1980) was inserted into the BamHI site of pUC18. This fragment contains the promoter elements and transcripational start sites but lacks the translational start codon for the HSV TK gene (McKnight et al., 1981). To eliminate the ATG sequence in the Sphl site in the polylinker region of pUC18 (which might compete for translational initiation), the plasmid was linearized at the Sphl site, the protruding 3' ends were removed with T4 DNA polymerase, and finally the blund-ends were ligated. Elimination of the ATG sequence was confirmed by DNA sequence analysis. The Kpnl-HindIII fragment of this plasmid was then inserted into the Kpnl/HindIII sites of pBluescript M13+. The TS coding region and 3'-flanking region derived from the HindIII-SacI fragment of pTSMG3(H) were then inserted into the HindIII/SacI sites of the above plasmid to form pKTT. pKI_{56}T was generated by substituting the Bsu36l-SacI fragment of pKTT with the Bsu36l-SacI fragment of pTI_{56}T. In control experiments, I determined that the level of expression of pKTT cloned into pBluescript M13 was the same as that observed with pKTT that was cloned into pUC18.
Fig. 11. Structures of the chimeric TS minigenes. Construction of pTcat has been described (Li et al., 1991). Constructions of other chimeric TS minigenes are described in Fig. 11 and 12. The probes used for S1 nuclease protection assays are derived from the minigenes and are shown. Restriction sites: A, Apal; B, BamHI;Bg, BglII; Bs, Bsu36I; C, Clal; E, EcoRI; H, HindIII; K, Kpnl; S, SalI; Sc, SacI; X, Xbal. Symbols: , TS promoter or polyadenylation signal; , SV40 early promoter or polyadenylation signal; , HSV TK promoter; , CAT coding region; , TS coding region; , TS intron sequence or vector sequence.
Fig. 12. Construction of the chimeric TS minigenes with altered promoters. Constructions of the chimeric TS minigenes is described in the text. Restriction sites: B, BamHI; Bg, BglII; Bs, Bsu36I; H, HindIII; K, KpnI; S, SalI; Sc, SacI; Sp, SphI; X, XbaI. Symbols:

- , TS promoter or polyadenylation signal; ■ , SV40 early promoter or polyadenylation signal; □ , HSV TK promoter; □ , TS coding region; ——, TS intron sequence or vector sequence.
Figure 12
Construction of chimeric TS minigenes with an alternative polyadenylation signal. The schematic representation of the constructions in this section is showed in Fig. 13. A Sacl site was introduced by site-directed mutagenesis just after the TGA termination codon in pMTS which contains the TS coding region in the pDSP-1 vector (Pfarr et al., 1985) by Haichao Zhang. The Apal-Sall fragment from pMTS which contains part of the TS coding region and the SV 40 early polyadenylation signal was subcloned into the Apal/Sall sites of pBluescript 13+ to obtain more restriction sites. From the resulting plasmid, the BglII-EcoRI fragment was then used to replace the BglII-EcoRI fragment of pSTT to generate pSTS. pSI56S was produced by substituting the Bsu36I-ClaI fragment of pSTS with the Bsu36I-ClaI fragment of pT156T. To construct pTTS, the BglII-EcoRI fragment of pTTT was replaced by the BglII-EcoRI fragment of pSTS and finally the pT156S was created by substituting the Sall-Bsu36I fragment of pSI56S with the Sall-Bsu36I fragment of pT156T.

Analysis of promoter strength. TS represents only about 0.01% of total cell protein in all types of proliferating cells, which is typical for the product of a housekeeping gene. It has been suggested that the reason for this low level expression of many housekeeping genes is that the housekeeping promoters direct RNA synthesis at relatively low levels (Dynan, 1986). This speculation is probably supported by the fact that the promoter structure of many housekeeping genes (which are expressed at relatively low levels in all cell types) is very different from that of most specialized genes (which are expressed at relatively high levels in selected cell types) in the
Fig. 13. Construction of the chimeric TS minigenes with altered polyadenylation signals. Construction of the chimeric TS minigenes is described in the text. Restriction sites: A, Apal; Bg, BgIII; Bs, Bsul36I; C, CiaI; E, EcoRI; S, Sall; Sc, ScaI. Symbols:
- , TS promoter or polyadenylation signal; ■■ , SV40 early promoter or polyadenylation signal; , TS coding region; —— , TS intron sequence or vector sequence.
sense that it lacks apparent TATA and CAAT elements and tends to have multiple transcriptional initiation sites.

To determine whether the low level of expression of the TS gene is due to the fact that it has an exceptionally weak promoter, I compared the strength of the mouse TS promoter with that of two other well-characterized viral promoters, the HSV TK promoter and the SV40 early promoter. The TS promoter used in this study extended from the XbaI site 1 kb upstream of the open reading frame to the A of the ATG start codon. The various promoters were linked to the intronless TS coding region and polyadenylation signal and the activities were determined by transient expression analysis. Table 5 shows that the strength of the TS promoter was approximately the same as that of the SV 40 early promoter but about two times greater than that of the HSV TK promoter, indicating the TS promoter is not exceptionally weak. Inclusion of introns 5 and 6 led to a significant (4- to 11-fold) increase in the level of expression regardless of which promoter was used (Table 5), indicating that the stimulatory effect was not specific for the TS promoter.

To determine whether the same relative promoter strength would be observed with a different coding region and polyadenylation signal, I also analyzed the expression of pTcat constructed by Dawei Li. This minigene contains TS promoter (extending to -1kb) linked to the CAT coding region and SV 40 early polyadenylation signal (Fig. 11). The activity of pTcat was compared with that of pSV2cat (Gorman, 1982), which has the SV40 early promoter in place of the TS promoter but is otherwise identical to pTcat. As shown in Table 5, approximately the same amount of CAT activity was observed with both minigenes, again indicating that the TS and SV40 early
### Table 5
**Activity of chimeric TS minigenes**

<table>
<thead>
<tr>
<th>Minigenes</th>
<th>TS enzyme levels&lt;sup&gt;a&lt;/sup&gt;</th>
<th>CAT activity&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introns</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pTTT</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>pSTT</td>
<td>1.3 ± 0.5</td>
<td></td>
</tr>
<tr>
<td>pKTT</td>
<td>0.44 ± 0.15</td>
<td></td>
</tr>
<tr>
<td>pSTS</td>
<td>22.8 ± 9.8</td>
<td></td>
</tr>
<tr>
<td>pTTS</td>
<td>8.8 ± 3.3</td>
<td></td>
</tr>
<tr>
<td>Intron containing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pTl56T</td>
<td>11.4 ± 1.3</td>
<td></td>
</tr>
<tr>
<td>pSl56T</td>
<td>7.7 ± 2.3</td>
<td></td>
</tr>
<tr>
<td>pKl56T</td>
<td>1.6 ± 0.6</td>
<td></td>
</tr>
<tr>
<td>pSl56S</td>
<td>107 ± 51</td>
<td></td>
</tr>
<tr>
<td>pTl56S</td>
<td>16.4 ± 4.6</td>
<td></td>
</tr>
<tr>
<td>pTcat</td>
<td>1.0</td>
<td>0.87 ± 0.10</td>
</tr>
<tr>
<td>pSV2cat</td>
<td>0.87 ± 0.10</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> The transfections and enzyme assays were performed as described in Table 1. All values were normalized to those observed with pTTT. Data represent the average (± standard deviation) of 3 to 10 experiments.

<sup>b</sup> Normalized to TS enzyme level. The value obtained for pTcat was set to 1.0 to facilitate the comparison.
promoters had approximately same strength. These results also demonstrate that all of the essential TS promoter elements are located upstream of the ATG start codon.

To be certain that the differences in enzyme levels were due to differences in mRNA levels, cytoplasmic poly(A)+ mRNA was isolated from the transfected cells and analyzed by S1 nuclease protection assays. The probe used in comparing the strength of the TS promoter with that of the SV40 early promoter corresponded to the TS coding region linked to the SV40 early promoter (probe C, Fig. 11). Therefore, mRNA derived from minigenes with the SV40 early promoter protected a DNA fragment extending from the TS coding region to the transcriptional start sites of the SV40 early promoter, whereas mRNA derived from minigenes with TS the promoter protected shorter fragments that extended to the point of insertion of the SV40 sequences. The S1 analysis result shows that there was good agreement between the relative amount of mRNA and the relative amount of enzyme derived from each minigene (Fig. 14). Densitometric scanning of the autoradiograms revealed that approximately the same amount of mRNA was produced by minigenes with the SV40 early promoter as by minigenes with the TS promoter. Minigenes with the HSV TK promoter produced about two to three times less mRNA than did minigenes with TS promoter (Fig. 15). These results clearly demonstrate that that the low level expression of TS gene is not due to the fact that the TS promoter is exceptionally weak.
Fig. 14. Analysis of expression of TS chimeric minigenes by S1 protection assays (1). Cytoplasmic poly(A)+ mRNA from transiently transfected cells was isolated. 4 µg of mRNA from each sample was hybridized with probe C (Fig. 11) derived from pSTT. The probe was labeled at the 5' end with $^{32}P$ at a BamHI site in the TS coding region (256 nucleotides downstream of the ATG start codon) and extended to the Sall site 377 nucleotides upstream of the ATG start codon.
Fig. 15. Analysis of expression of TS chimeric minigenes by S1 nuclease protection assays (II). Cytoplasmic poly(A)+ mRNA was isolated from the transiently transfected cells, and 8 μg of mRNA from each sample was hybridized with probe A (Fig. 11) derived from pTTT(pTSGM2).
Analysis of the efficiency of the TS polyadenylation signal.
Since the low level of expression of the mouse TS gene is not due to the fact that it has an exceptional weak promoter, suggesting that posttranscriptional processes may play an important role in regulating the basal level expression of the mouse TS gene. A previous study by Chris Harendza has shown that when the non-consensus hexanucleotide AUUAAA (the upstream polyadenylation signal of the TS gene) was mutated to a consensus hexanucleotide AAUAAA, the expression was increased 2 to 7 fold. This indicates that the TS gene has a relatively inefficient polyadenylation signal (Harendza and Johnson, 1990). I took a different approach to investigate the efficiency of the TS polyadenylation signal. I compared the efficiency of the TS polyadenylation signal with that of the SV40 early polyadenylation signal by transient expression analysis. The results showed that when the TS polyadenylation signal from both pTTT and pSTT was replaced by the SV40 early polyadenylation signal, the levels of expression of both minigenes were increased approximately 10 times (Table 5), indicating that the TS polyadenylation signal is very inefficient when compared with the SV40 early polyadenylation signal, regardless which promoter is used to direct the TS RNA synthesis. This result implies that only small portion of the primary transcripts is polyadenylated due to the inefficient TS polyadenylation signal and this is at least partly responsible for the low level of expression of the TS gene.

Inclusion of introns 5 and 6 into the intronless TS chimeric minigenes with the SV40 early polyadenylation signal also led to a significant increase (2 to 5-fold) in the level of expression (Table 5), indicating that the stimulatory
effect of introns 5 and 6 is not specific for the TS polyadenylation signal, although the stimulatory effect of introns 5 and 6 is larger when coupled with TS polyadenylation signal than with the SV40 early polyadenylation signal. The S1 nuclease protection analysis demonstrated that the relative changes in TS enzyme level in each case was due to the corresponding changes in TS mRNA level (Fig. 14).

Finally, it was interesting to find that the expression level of pSI58S (which contains the SV40 early promoter, TS coding region with introns 5 and 6, and the SV40 early polyadenylation signal) is about 100 times higher than that of pTTT. From a practical point of view, this construction may provide a very high level mammalian expression system for detailed studies of the TS protein.

**Study of the regulation of the mouse TS gene in response to growth stimulation**

It has been demonstrated that the expression of the TS gene is highly regulated in growth-stimulated cells (Conrad, 1971; Navalgund, et al., 1980; Jenh, et al., 1985a). Both TS enzyme and mRNA levels are very low when cells are in the G₀ (resting) phase but reach a maximum level when cells are in S (DNA synthesis) phase. The mechanism for this regulation however is still not clear. One of my primary goals was to study the mechanism which is responsible for regulating the expression of the mouse TS gene in response to growth-stimulation, by determining the approximate location(s) of the regulatory sequences which are responsible for the growth-regulated expression.
Previous studies have shown that during serum stimulation of LU3-7 cells (a TS-overproducing cell line) (Rossana et al., 1982), the TS mRNA level increased 20-40 fold (Jenh et al., 1985a). Pulse-chase studies showed that the half-life of TS mRNA is about 8 h and does not fluctuate in response to growth-stimulation (Jenh et al., 1985a). In addition, Jenh et al. also demonstrated that the level of TS mRNA was not regulated at the level of differential translation of TS message during serum stimulation (Jenh et al., 1985a). These results suggested that the growth-regulated expression of the TS gene may be controlled at the level of transcription, or at the level of RNA processing, or the combination of both. Nuclear run-on experiments at 5 h intervals showed that there was only a small (3 to 4-fold) increase in the rate of TS the transcription during S phase. Therefore TS gene expression is mainly controlled at the level of RNA processing although transcriptional control may play a small role. However, studies on other S-phase genes showed that when time points of the nuclear run-on experiments were chosen to span the G1-S phase interface, larger (6 to 7-fold) increases in the rate of transcription were observed in growth-stimulated cells (Steward et al., 1987), indicating transcriptional control actually plays an important role in regulating the expression of these S-phase genes during growth stimulation.

I took a different approach to investigate TS gene regulation. Instead of studying the endogenous TS mRNA, I chose to study the expression of recombinant TS minigenes in order to determine the role of different components of the gene. The general approach I used in these studies was to stably transfec the TS minigenes, which may contain the potential regulatory elements, into the wild-type mouse 3T6 cells. A plasmid that confers
resistance to the antibiotic G418 was cotransfected with the TS minigene to permit selection of cells that had stably incorporated the minigenes into their genome. To avoid possible artifacts that may result from the site of integration of the minigenes into the host genome, the regulation studies were performed by using pools of cells representing at least 100 independent G418-resistant clones rather than single clones. To study the growth-regulation, the stably transfected cells were first allowed to rest at G₀ phase by reducing the serum concentration in the culture medium for 7 days and then inducing the cells to reenter cell cycle by feeding them with fresh medium containing 10% serum. S phase begins about 12 h after the stimulation. The levels of expression of both transfected and endogenous TS genes can be compared by determining the TS mRNA levels using an S1 nuclease protection assay. The following experiments were repeated several times with two different stably transfected cell lines, and similar observations were made each time.

Regulation of wild-type TS minigene expression in growth-stimulated cells. Using minigenes to study gene regulation is a very common and convenient approach. However, it has been found that regulatory elements can be present in intragenic sequence (Rosenthal, 1987 and references therein; Ottavio et al., 1990), and therefore it is possible that by deleting an internal portion of a gene, one may eliminate an important region of control. Thus, in order to characterize the sequences and mechanisms that are important for regulating TS gene expression, it is first necessary to demonstrate that the regulatory sequences are present and functional in the wild-type TS minigenes. As mentioned earlier, this was done
by stably transfecting the wild-type TS minigene into mouse 3T6 cells and determining the content of TS mRNA from the transfected cells by S1 nuclease protection assays.

Since the 3T6 cells were wild type, it was not possible to measure minigene expression by monitoring TS enzyme levels. However it was possible to distinguish TS mRNA derived from the endogenous TS gene from that derived from the transfected TS minigenes by using 5' S1 nuclease protection assays. For studies of minigenes in which the TS promoter was linked to the TS coding region, the minigenes were tagged by deleting a 57-nucleotide fragment from the middle of the coding region (between two BamHI sites). Minigenes that contained the deletion were designated by the suffix (d). The S1 probe did not contain the deletion and extended to the BgIII site 218 nucleotides 3' of the deletion. Thus, mRNA derived from the minigene protected the probe up to the site of the deletion, whereas mRNA derived from the endogenous TS gene protected the probe to the sites of transcriptional initiation. The schematic representation of this experimental strategy is shown in Fig. 16.

Fig. 17 shows that the regulation of pT156T(d) was very similar to that of the endogenous TS gene in serum-stimulated mouse 3T6 cells. TS mRNA levels from the minigene and from the endogenous TS gene remained low until about 10 h after serum stimulation and then increased dramatically. Densitometric analysis revealed that the amount of TS mRNA increased about 20-fold between 5 and 15 h both for the endogenous TS gene and for the minigene. Therefore, all of the sequences that are necessary and sufficient for normal growth regulation appear to be contained within this
Fig. 16. Schematic representation of the experimental design of tagging TS minigenes for growth regulation study. To distinguish the endogenous TS mRNA from mRNA derived from the transfected TS minigenes using S1 nuclease protection assays, minigenes in which the TS promoter (thin line) is linked to the TS coding region (open box) are tagged by deleting a 57-nucleotide fragment (closed box) from the middle of the coding region (between two BamHI sites in the exon 3). The S1 probe does not contain the deletion and extend to the BglII site 218 nucleotides 3' of the deletion. mRNA derived from the minigene protects the probe up to the site of the deletion, whereas mRNA derived from the endogenous TS gene protects the probe to the sites of transcriptional initiation, which generates several fragments around 560 nucleotide long. The two probes with different sizes can be easily distinguished on a 6% polyacrylamide gel.
Endogenous TS gene

Transcription

S1 probe

S1 nuclease digestion

NaOH hydrolysis

~ 560nt

electrophoresis on 6% polyacrylamide gel

Figure 16
minigene. To determine whether the regulatory sequences are contained in a specific intron, I also analyzed the expression of pTI12T(d). As shown in Fig. 18, pTI12T(d) was also regulated normally. Again, there was at least a 20-fold increase in TS mRNA derived from the endogenous and transfected minigenes between 5 and 15 h after serum stimulation. These results indicate that even though TS introns can stimulate the basal level expression of TS genes, they do not appear to contain the regulatory sequences for growth-regulated TS gene expression.

I also tried to analyze the expression of the intronless TS minigene pTTT to see whether the presence of an intron can, in general, provide the growth-regulated expression of TS gene. Unfortunately, I was not able to detect any signals corresponding to mRNA from pTTT in the S1 autoradiogram due to the extremely low level expression of pTTT in the stably transfected cell lines.

Role of the TS 5'-flanking region in growth-regulated ts gene expression. A previous study by Dawei Li showed that a minigene consisting the TS promoter (extending to 1kb upstream of ATG translational start codon) linked to the bacterial chloramphenicol acetyl-CoA transferase (CAT) coding region was expressed constitutively in growth-stimulated 3T6 cells (Li, et al., 1991). There are at least two possible explanations for this result. The first explanation is that the TS 5'-flanking sequence is not important for growth-regulated TS gene expression. The second explanation is that the TS 5'-flanking sequence alone is not sufficient but still necessary for the normal regulation of the TS gene. To distinguish these two possibilities, I replaced the TS promoter of pTI56T with the SV40 early
Fig. 17. Expression of pTI56T(d) in growth-stimulated cells. 3T6 cells that were stably transfected with pTI56T(d) were allowed to rest in 0.5% calf serum for 7 days and then induced to reenter the cell cycle by feeding them with fresh medium containing 10% serum at time 0. At the indicated times, cytoplasmic poly(A)+ mRNA was isolated and equal amounts of mRNA (approximately 1 μg) were analyzed by S1 nuclease protection assays. Probe B (derived from pTTT, Fig. 11) was used for this experiment. The probe was 5'-end labeled at the BglII site in the TS coding region 531 nucleotides downstream of the ATG start codon and extended to the XbaI site 1 kb upstream of the ATG codon. Fragments labeled E and M correspond to the fragment protected by TS mRNA derived from the endogenous TS gene and the transfected TS minigene, respectively. The autoradiograms were scanned with a densitometer, and the results are plotted below each autoradiogram.
Figure 17
Fig. 18. Expression of pTI_{12}T(d) in growth-stimulated cells. The experiment was performed as described in Fig. 17 except that the cells were stably transfected with pTI_{12}T(d).
Figure 18
promoter (which has been shown to be neutral in response to growth-stimulation [Stewart et al., 1987]) to create pSI{sub}se{T}{sub}T (Fig. 11). This minigene was stably transfected into 3T6 cells and mRNA derived from both the endogenous and transfected TS genes was analyzed by S1 nuclease protection assays. Fig. 19 shows that the amount of mRNA derived from the minigene increased only two- to three fold following serum stimulation, whereas the endogenous TS mRNA increased about 20-fold. This result indicates that the 5'-flanking sequence is indeed necessary, although not sufficient, for the normal regulation of mouse TS gene in response to growth-stimulation.

To further determine the approximate location of the regulatory sequences within the 5'-flanking region, I analyzed the regulation of several TS minigenes with a shortened 5'-flanking sequence. When the 5'-flanking sequence was deleted to -250, the minigene was partially regulated (Fig. 20). Densitometric analysis revealed that the endogenous TS mRNA increased about 20-fold but the amount of TS mRNA derived from the transfected minigene increased about 5-fold following serum stimulation. However, when the 5'-flanking sequence was deleted to -150, the growth-regulation of the minigene was greatly reduced (Fig. 21). Densitometric analysis revealed that the amount of mRNA derived from the endogenous TS gene increased about 20-fold, whereas the TS mRNA derived from the transfected minigene increased only about 2 to 3-fold after the simulation. These results indicated that the sequences responsible for the regulation of TS gene in serum-stimulated cells are present, or at least in part present, in the region between -1kb and -150 relative to the ATG translational start codon.
Fig. 19. Expression of pSI\textsubscript{56}T in growth-stimulated cells. The experiment was performed as described in Fig. 17 except that the cells were stably transfected with pSI\textsubscript{56}T and the probe C (Fig. 11) was used in this experiment.
Fig. 20. Expression of pTl56T(d)(-250) in growth-stimulated cells. The experiment was performed as described in Fig. 17 except that the cells were stably transfected with pTl56T(d)(-250).
Figure 20

TS mRNA (arbitrary units)

Time (min)

0 0.05 0.10 0.15 0.20 0.25 0.30

0 10 20 30

0 5 10 15 20 25
Fig. 21. Expression of pTls6T(d)(-150) in growth-stimulated cells. The experiment was performed as described in Fig. 17 except that cells were stably transfected with pTls6T(d)(-150). The values for the minigene were multiplied by 5 so that they could be more readily compared with the values for the endogenous gene.
Figure 21
Role of the TS polyadenylation signal in growth-regulated expression of TS mouse gene. The above observations suggest that TS 5'-flanking sequence alone is not sufficient but still necessary for normal regulation of mouse TS gene. This implies that sequences downstream of the ATG translational start codon are also important for proper regulation of TS gene in response to growth stimulation. Previous studies have suggested that the regulation of TS gene in growth-stimulated cells may be at the level of the efficiency of 3' processing (polyadenylation) (Jenh et al., 1985a).

To determine the role of the TS 3'-flanking sequence in growth-regulated expression of the TS gene, I replaced the TS 3'-flanking sequence of pTI56T with the SV40 early polyadenylation signal (which has been shown to be neutral during growth-stimulation [Kauman et al., 1983]) to create pTI56S minigene (Fig. 11). pTI56S was then stably transfected into 3T6 cells. The TS mRNA derived from both endogenous and transfected TS genes was compared during the growth-stimulation by S1 protection assay. Fig. 22 shows that the amount of TS mRNA derived from the endogenous TS gene increased about 15-fold, whereas the TS mRNA derived from the transfected gene increased about 7-fold after the serum stimulation. This indicates that the TS 3'-flanking sequence is not essential for the growth-regulated mouse TS gene expression.

In addition, I also analyzed the expression of pSI56S which contains SV40 early promoter, TS coding region with introns 5 and 6 at their normal position, and the SV40 early polyadenylation signal. Although the result from pSI56T already implied that the TS coding region is not sufficient for normal regulation of the TS gene, it is still very informative to analyze the expression
Fig. 22. Expression of pTl$_{56}$S(d) in growth-stimulated cells. The experiment was performed as described in Fig. 17 except that cells were stably transfected with pTl$_{56}$S(d).
Figure 22
Fig. 23. Expression of pSI_{56}S in growth-stimulated cells. The experiment was performed as described in Fig. 17 except that cells were stably transfected with pSI_{56}S and probe C (Fig. 11) was used in this experiment. The values for the endogenous were multiplied by 2 so that they could be more readily compared with the values gene for the minigene.
Figure 23

TS mRNA (arbitrary units) (E-1)

Time (min)

0 20 40 60 80 100

0 10 20 30
of the minigene containing only TS coding region under control of SV40 early promoter and polyadenylation signal. S1 analysis showed that the endogenous TS mRNA was regulated normally (more than 20-fold increase) in response to growth stimulation, whereas the growth regulation of the transfected minigene was greatly reduced (2 to 3-fold increase) (Fig. 23), proving that the TS coding region alone is not sufficient for normal regulation of the TS gene during growth stimulation.
DISCUSSION

Analysis of the 5'-flanking sequence of the mouse TS gene

Sequence analysis revealed that the promoter of the mouse TS gene has some unusual features. Like many other housekeeping genes, it is very GC-rich, lacks canonical TATA and CAAT sequence elements and has multiple transcriptional start sites (Deng et al., 1986). Earlier deletion analyses showed that the TS minigenes deleted to -150 relative to ATG translational start codon are just as active as the wild-type promoter but those deleted to -54 are inactive (DeWille et al., 1988). Gel mobility shift and footprint analyses demonstrated that multiple protein-DNA interactions occur in the region between -150 to -54. Site-specific point mutations have been made to evaluate the biological significance of these interactions and to confirm the locations of the binding sites of potential transcriptional factors.

Roles of the GC boxes in transcriptional initiation of the mouse TS gene. As mentioned previously, one of the common features of the promoters of many housekeeping genes is that they are very GC-rich. This is usually associated with the presence of multiple GC boxes in the promoter. Therefore, determination of the functional role of multiple GC boxes is crucial for understanding the transcriptional regulation of this important class of promoters.
Two GC boxes have been identified in the mouse TS promoter region by DNA sequence analysis. One is centered at -130 (relative to the ATG start codon) which is about 40 nucleotides upstream of the first transcriptional start site, and the other is centered at -80 which is in close proximity to the first transcriptional start sites. Mutations in either GC box led to a 3-fold reduction in both TS enzyme and mRNA levels. S1 nuclease protection analyses revealed that inactivation of the GC boxes also alters the relative utilization of transcriptional start sites. The effect of inactivation of the second GC box (-80) is more dramatic (Fig. 3 and Fig. 6). It appears that the mutation in this GC box not only causes decreased transcription from downstream sites, but also results in a increased utilization of the upstream start sites.

A reduction in transcription rate by both mutations is expected since GC boxes can stimulate transcription from downstream start sites, presumably by interacting with trans-acting factors such as Sp1 (Kadonaga et al., 1986). The increased utilization of the upstream start sites by inactivation of the second GC box (-80) is unexpected, but it is not unreasonable since this GC box is located between the start sites and interaction of this GC box with the corresponding trans-acting factor may interfere with transcription from the immediately proximal start sites. Thus, abolishing this GC box may decrease transcription from the downstream start sites and at the same time increase the availability of the immediately proximal start sites to the transcriptional machinery. Thus, the overall change in the initiation pattern by the mutation of this GC box is rather dramatic. In the case of the GC box at -130 (which is upstream of all start sites), inactivation of this GC box decreases transcription from all downstream start sites. But the effects are more significant on the
proximal start sites (about 40 to 50 nucleotides downstream) than on the distal ones (about 80 to 100 nucleotides downstream) (Table 2). Therefore, inactivation of this GC box can also change the initiation pattern but the change is not as dramatic. This observation is consistent with the notion that weak activators cannot function at a distance. A recent study on the promoter of murine terminal deoxynucleotidyl transferase (TdT) gene suggests that the optimal spacing of GC boxes relative to the start sites is 42 to 50 nucleotides upstream (Smale and Baltimore, 1989). It is interesting to note that the distance of each GC box in the TS promoter relative to the start sites it functions on is also about 40 to 50 nucleotides (Fig. 2). This explains the differential effects on the utilization of start sites by the mutation in the GC box at -130. A model for how the GC boxes function in TS promoter is proposed in Fig.24. It would be interesting to see the pattern of start sites when both GC boxes are mutated at the same time.

Recently, Blake et al. (1990) reported their study on the promoter of the hamster DHFR gene (another housekeeping gene). The results also demonstrated that GC boxes not only are required for efficient transcription but also regulate start site utilization in the TATA-less DHFR promoter. Their results are very similar to my observations with the mouse TS promoter. It is interesting to note that one of the GC boxes in the hamster DHFR promoter overlaps with an initiation site; a mutation in this GC box not only reduces the transcription from a downstream initiation site but also increases the utilization of the coincident start site. This is exactly the same as my observation with the GC box(-80) in the TS promoter.
Fig. 24. A proposed model for the role of GC boxes in transcriptional initiation from the mouse TS promoter. GC boxes stimulate transcription from downstream start sites with an optimal distance of 30 to 50 nucleotides. Inactivation of the GC box(-130) decreases transcription from all downstream start sites. But the effects are more significant on the proximal start sites (about 40 to 50 nucleotides downstream) than on the distal ones (about 80 to 100 nucleotides downstream). Abolishing the GC box(-80) decreases transcription from the downstream start sites and at the same time increase the availability of the immediately proximal start sites to the transcriptional machinery. Symbols: [ ], GC box; [ ], GC box binding protein; [ ], the major transcriptional start site; and [ ], the minor transcriptional start site.
Figure 24
Although the GC boxes can regulate the relative utilization of transcriptional start sites, it appears that they have no role in specifying the positions of initiation. The results of Blake et al. (1990) are similar in this aspect. As mentioned in Introduction section, the region containing the transcriptional start sites has been demonstrated to specify the pattern of transcriptional initiation of several genes with TATA-less promoters (Ayer and Dynan, 1988; Smale and Baltimore, 1989; Means and Farnham, 1990). An obvious question is that if a protein other than RNA polymerase II binds to the transcription initiation site(s), how can RNA polymerase II occupy the exact same site to initiate transcription. It would be interesting to find out how these transcription factors interact with the corresponding DNA elements, with RNA polymerase II, or even with other transcription factors to accomplish the initiation process in genes with TATA-less promoters.

Although gel mobility shift and footprint analyses demonstrated that Sp1 binds to both GC boxes in the mouse TS gene, the identity of the protein(s) which interact with the GC boxes in vivo to regulate the relative utilization of transcriptional start sites is still not clear. Several transcription factors besides Sp1 have been found to interact with GC boxes, including LSF (Kim et al., 1987), ETF (Kageyama et al., 1989), GCF-1 (Kageyama and Pastan, 1989) and AP2 (Mitchell et al., 1987). ETF might be a good candidate since it has been shown to specifically stimulate transcription from TATA-less promoters. Analysis using an in vitro transcription system with well defined components should provide insight on the identity of the transcription factor which interacts with the GC boxes to regulate the utilization of the start sites in the TS promoter.
In summary, interactions of GC boxes with the corresponding proteins can stimulate transcription from downstream start sites within a certain distance (probably about 40 to 50 nucleotides) without affecting the positions of initiation. The GC boxes are necessary for efficient expression of the TS gene and proper utilization of transcriptional start sites but each individual GC box is not absolutely essential for TS gene activity.

The essential region of the mouse TS promoter is in close proximity to the first transcriptional start site. Further deletion analyses indicated that the sequence between -105 (which eliminated both the GC box at -130 and the USF element) and -85 is necessary for TS promoter activity (Table 3). Gel mobility shift and footprint analyses, carried out by Keith Jolliff, with various sequences from this region identified three major protein-DNA interactions (Jolliff et al., 1991). One of these corresponds to Sp1 interacting with a non-consensus GC box between the first and second start sites or at -80 relative to ATG start codon. The proteins responsible for the other two complexes (CII and CIII) do not appear to correspond to any of the common transcriptional factors studied so far. The binding site for the CII protein is located between -99 to -91, which is very close to the first transcriptional initiation site, whereas sequences necessary for CIII binding are located between -99 to -109 (Fig. 2).

To determine the biological significance of these interactions, point mutations were made in these binding sites and the mutant promoters were then analyzed by transient expression assays. Table 1 shows that inactivation of the CIII binding site or the GC box caused a decrease in expression of the
TS gene by 12-fold or 3-fold, respectively, whereas inactivation of the CII binding site had no apparent effect.

It appears that the function of the CIII site is critical for TS gene expression. However, the mutation in the CIII site created a TATAAA sequence in this site. S1 nuclease protection analysis reveal that this artificial TATAAA box indeed directs initiation of RNA synthesis from a site about 25 nucleotides downstream. Thus, the decreased expression of the TS gene could be due to the binding of the transcriptional factor TFIID to the artificial TATAAA box and consequently interfering with the interactions of other protein(s) close to the CIII site. Due to the TATA box complication, it is necessary to confirm the function of the CIII site by other mutations.

Although the mutation in the CII binding site has no apparent effect on TS gene expression (a similar observation was made with the USF mutation), two alternative explanations should be considered. The first is related to the fact that TS gene expression is tightly regulated in response to growth-stimulation. The transient expression assay used in this study only reflects the expression of the gene constructs in exponentially growing cells. Thus, it cannot be excluded that the CII site and/or USF site might play important roles in regulating TS gene expression in response to growth stimulation. The second is related to the fact that the TS gene is an essential gene that must be expressed in a wide variety of proliferating cells. The binding sites that are irrelevant for expression in the hamster V79 cells may be more important for expression in other cells, especially those that may contain low levels of Sp1 or CIII proteins.
The proteins that interact with CII and CIII sites have not been characterized. Competition analysis by Keith Jolliff demonstrated that the CII and CIII proteins do not correspond to any trans-acting factors identified so far. However, a stretch of 9 nucleotide sequence (GCCGGAAGT) within the region encompassing the CIII binding site has a very high homology with some transcription elements identified in other genes. It has perfect homology with the β element found in the mouse ribosomal protein genes L32 and L30 promoter (Atchison et al., 1989; Hariharan et al., 1989), and an 8 out of 9 nucleotides homology with both the XrpF1 binding site in the Xenopus Laevis L14 promoter (Carnevali et al., 1989) and the Su2 element in the promoter of mouse Surf-1 and Surf-2 genes (Lennard and Fried, 1991). All six genes lack a TATA box and are ubiquitously expressed. However, the ribosomal protein genes have a single transcriptional start site, whereas the TS and Surf-1 and Surf-2 genes have multiple transcriptional start sites. Therefore, it appears that this consensus sequence does not have any role in specifying the transcriptional initiation pattern. Furthermore, inactivation of this consensus sequence in the ribosomal protein genes only leads to a relative small (2 to 3-fold) reduction in promoter activity whereas inactivation of the CIII element leads to a 12-fold reduction in TS gene expression. It will be interesting to see whether the same factor is responsible for controlling the transcription of these housekeeping genes as well as other ubiquitously expressed genes that lack TATA element.

S1 nuclease protection analysis indicated that inactivation of the GC box at -80 causes a significant increase in utilization of the immediately proximal start sites and a decrease in use of the downstream start sites as discussed
previously. However, inactivation of the CII and CIII binding sites has no effect on either the distribution or the position of initiation sites except for the occurrence of an additional site due to the creation of a TATA box by the mutation within the CIII binding site. The result of CII site mutation is a little surprising since the mutations in this site were very close to the first start site (Fig. 2).

As mentioned earlier, the region containing the transcriptional start sites has been demonstrated to specify the pattern of transcription initiation of several genes with TATA-less promoters (Means and Farnham, 1990; Smale and Baltimore, 1989; Ayer and Dynan, 1988). In addition, Means and Farnham (1990) even proposed a start site consensus sequence for housekeeping genes. However, it does not seem to apply in the case of mouse TS gene. For example, unlike the above genes which have only one or a few tightly clustered start sites, the TS gene has multiple transcriptional start sites across a 60 nucleotide region, and sequences around each start site are very different. This suggests that, in the case of TS gene, there is no consensus sequence which can specify the position of transcriptional initiation as suggested by the observations with mouse DHFR gene. This, however, does not mean that the sequences around the transcriptional start sites of the TS gene are not important for transcription. In fact, comparison of the 5'-flanking sequences of the human and mouse TS genes revealed that there are many significant blocks of sequence homology between the two genes, especially in the 5' untranslated region which covers all of the start sites (Deng et al., 1986; Takeishi et al., 1989). S1 nuclease protection analyses showed that the two major start sites in the human TS gene are
located at the same or almost the same positions as the two major start sites in the mouse TS gene. These observations suggest that sequences around the start sites may be important in transcription of TS genes. A complete understanding of the mechanism for transcriptional initiation of genes with TATA-less promoters awaits a more detailed analysis.

In summary, interaction of the CIII binding protein(s) with the CIII site is critical for TS gene expression. However, the CIII and CII interactions do not have apparent effects on the relative utilization and positions of transcriptional start sites. The role of the protein(s) interacting with the CII site is not clear so far, but CII could play a role in regulating the TS gene expression in response to growth-stimulation. As discussed previously, the GC boxes are required not only for efficient expression of the TS gene but also for proper utilization of the individual transcriptional start sites.

The mouse TS gene expression is stimulated by some (but not all) introns

Previous studies showed that intronless TS minigenes are expressed at a very low level when transfected into ts- hamster V79 cells (DeWille et al., 1988). This low level of expression could be due to a number of factors, including the absence of intron sequences in the minigenes. To determine whether TS introns can stimulate the expression of the intronless minigenes, numerous intron-containing minigenes were constructed and analyzed by transient transfection assays. The results in Table 4 clearly demonstrate that inclusion of some of the introns from the mouse TS gene has significant effects on the expression of the intronless minigene. Inclusion of introns 1
and 2, intron 1, or introns 5 and 6 resulted in significant increases in both TS enzyme and mRNA levels relative to the intronless minigene. However, inclusion of intact intron 4 did not stimulate minigene expression. Instead, it led to a two-fold reduction in enzyme level.

Previous studies have shown that the effects of introns on gene expression vary considerably depending on the gene under consideration. It was demonstrated in early studies that introns are essential for production of mature mRNA from certain viral or cellular genes (Hamer and Leder 1979; Gruss et al., 1979; 1980). Recently, many studies demonstrated that introns are also necessary for high level expression of many recombinant minigenes (Buchman and Berg, 1988; Chung and Perry 1989; Lipson et al., 1989; Ottavio et al., 1990), as in the case of the TS minigene. However, several mammalian genes which lack introns are also expressed at high levels. Examples include histone (Hentschel and Birnstiel, 1981), heat shock (Hunt and Morimoto, 1985) and interferon genes (Lengyel, 1982). In addition, introns are not essential for expression of several recombinant genes (Chee et al., 1986; Ng et al., 1985; Gross et al., 1987). Brinster et al. (1988) found that introns have little effect on the expression of several chimeric genes in transfected cell lines. Curiously, when the expression of these genes was examined in transgenic mice, 10 to 100 times more mRNA was produced from minigenes that contained introns than from those that lacked introns.

The mechanisms by which introns stimulate gene expression have been identified for a number of genes. Although introns may regulate gene expression through splicing, introns have been shown to stimulate transcription of several genes due to the presence of transcriptional enhancer
sequences, usually within the first intron (Gillies et al., 1983; Bornstein and McKay, 1988; Horton et al., 1987; Rossi and de Crombrugghe, 1987; Laherty et al., 1990). If this is the case, intron 1 of the TS gene is the most probable candidate to contain a transcriptional enhancer. Introns 5 and 6, however, do not seem to have an enhancer element since placement of these two introns upstream or downstream of the TS coding sequence failed to show a stimulation effect (Tiling Deng and Dawei Li, unpublished observations).

Introns may affect gene expression at a variety of levels in addition to transcriptional initiation. For example, some introns contain sequences that appear to function as transcriptional terminator (Bentley and Groudine, 1988). Sequences present in certain introns may stimulate or retard the rate of degradation of an hnRNA molecule or the rate of splicing (Pkielny and Rosbash, 1985; Gudas et al., 1988; Ghogawala et al., 1989; Bonadio et al., 1990). More recently, it has been demonstrated that introns can increase efficiency of RNA 3' processing and accumulation of cytoplasmic RNA (Manley et al., 1990; Niwa et al., 1990). The TS gene contains multiple introns, it is possible that different introns stimulate TS gene expression via different mechanisms. It would be interesting to determine whether introns 5 and 6, or rather just 6 alone can stimulate TS gene expression by increasing the efficiency of polyadenylation since it has been suggested that stimulation of polyadenylation is linked with the presence of the last intron (Niwa et al., 1990).

S1 nuclease protection analysis revealed that inclusion of intron 4 into the TS intronless minigene does not have any influence on the level of mRNA. However, TS enzyme assay showed that presence of intron 4 causes
a reduction in TS enzyme level by about 2-fold (Table 4). It appears that the minigenes containing intron 4 can produce the same amount of mRNA in the cytoplasm as that of the intronless minigene, but not all of the mRNA is functional for TS synthesis. The most straightforward explanation for this observation is that the mRNA derived from minigenes containing intron 4 is a mixture of both normal and abnormal splicing products. Sequence analysis revealed that besides the normal 3' splicing site, there is at least one more putative 3' splicing site in intron 4. Thus, it is possible that the RNA splicing machinery may use both sites to splice intron 4 and consequently result in production of both normal and abnormal mRNA species. As a result, the relative mRNA level derived from the intron 4-containing TS minigenes is higher than TS enzyme level.

**Identification of the regulatory elements of the mouse TS gene by chimeric gene analysis**

TS represents only about 0.01% of total cell protein in all types of proliferating cells, which is typical for the product of a housekeeping gene. It has been postulated that the low level expression of many housekeeping genes is due to inefficient promoters (Dynan, 1986), since the promoter structure of many housekeeping genes is very different from that of most specialized genes which are usually expressed at a very high level in selected cell types. To investigate the mechanism for this low level of the expression of mouse TS gene, a series of chimeric TS minigenes (Fig. 11) were constructed and analyzed by transient expression assays.
The low level of expression of the mouse TS gene is probably not due to a weak promoter, it is instead due to an inefficient polyadenylation signal. To determine whether the low level of expression of the TS gene is due to an inefficient promoter, the strength of the mouse TS promoter was compared with that of two well-characterized viral promoters, the HSV TK promoter and the SV40 early promoter. The results in Table 5 suggest that the strength of the TS promoter is approximately the same as that of SV40 early promoter but greater than that of the HSV TK promoter. Therefore, the low level of expression of this housekeeping gene is not due to a weak promoter. This result implies that an inefficient posttranscriptional event may be a major factor responsible for the low level of expression of the mouse TS gene.

To address this question, the efficiency of the TS polyadenylation signal was compared with that of the SV40 early polyadenylation signal. The results in Table 5 clearly indicate that the TS gene has a very inefficient polyadenylation signal. It is about 10 times less efficient than the SV40 early polyadenylation signal regardless which promoter (TS or SV40 early promoter) is used to direct the synthesis of the TS mRNA. This result suggests that the TS promoter can actually direct the synthesis of plenty of TS transcripts, but a large fraction of the transcripts fails to be processed to mature mRNA, which consequently results in the low level of TS mRNA in cytoplasm. This result is consistent with the observation of Chris Harendza who found that when the TS non-consensus hexanucleotide AUUAAA (the upstream polyadenylation signal) was mutated to a consensus hexanucleotide AAUAAA, the level of TS gene expression increased about 2
to 7-fold, indicating that the TS polyadenylation signal is relatively weak (Harendza and Johnson, 1990). Similar observations have been also made with the DHFR gene (Kaufman and Sharp, 1982).

In addition to the inefficient polyadenylation process, several observations suggest that the low level of expression of the TS gene may be in part due to the abnormal splicing of TS primary transcript. It has been found that about 10% of the cytoplasmic TS mRNA is nonfunctional due to aberrant splicing of intron 5 (Li and Johnson, 1989). Moreover, inclusion of intron 4 results in the production of a mixture of TS mRNA species of which only 50% is functional (Table 4).

In summary, the low level of expression of the mouse TS gene is probably not due to a weak promoter, it is instead due to inefficient polyadenylation and possibly other processing steps such as RNA splicing.

The stimulatory effect of introns 5 and 6 is not specific for the TS promoter or TS polyadenylation signal. Table 5 shows that inclusion of introns 5 and 6 into the intronless chimeric TS minigenes can stimulate the expression 4 to 11-fold no matter which promoter (TS, or SV40 early, or HSV TK) is used to direct the synthesis of the TS transcript. This result indicates that the stimulatory effect of introns 5 and 6 is not restricted to minigenes driven by the TS promoter. This is not unexpected since it is unlikely that transcriptional stimulatory signals are present in introns 5 and 6 as discussed in a previous section. Thus, it appears that the stimulatory effect of the introns 5 and 6 is primarily a posttranscriptional phenomenon. In line with this conclusion, Table 5 shows that inclusion of introns 5 and 6 into the
intronless TS minigenes with SV40 early polyadenylation signal also led to a significant increase (2 to 5-fold) in the level of expression. This result suggests that the stimulatory effect of introns 5 and 6 is not specific to TS polyadenylation signal. However, the stimulatory effect of introns 5 and 6 is larger when coupled with TS polyadenylation signal than with SV40 early polyadenylation signal. This observation might be very significant. As mentioned previously, it was demonstrated that introns can stimulate gene expression by increasing the efficiency of polyadenylation (Niwa et al., 1990; Huang and Gorman, 1990). It is conceivable that the stimulatory effect of introns may be more profound on a weak polyadenylation signal than on a strong one. If this is the case, it is possible that the stimulatory effect of intron 5 and 6 is linked with polyadenylation of the TS transcript.

Study of the regulation of the mouse TS gene in response to growth stimulation

It has been demonstrated that the expression of the TS gene is highly regulated in growth-stimulated cells (Conrad, 1971; Navalgund, et al., 1980; Jenh, et al., 1985). Both TS enzyme and mRNA levels are very low when cells are in the G0 (resting) phase but reach a maximum level when cells are in S (DNA synthesis) phase. The mechanism for this regulation, however, is still not clear. One of my primary goals in this Ph.D research was to study the regulatory mechanism for the expression of mouse TS gene in response to growth-stimulation, particularly, to identify the regulatory sequences which are responsible for the growth-regulated expression.
The sequences responsible for growth-regulated TS gene expression are present and functional in the intron-containing TS minigenes. Fig. 17 and Fig. 18 show that minigenes which included 1 kb of the TS 5′ flanking region, the TS coding region containing introns 5 and 6 or introns 1 and 2, and 0.25 kb of the TS 3′ flanking region are regulated as well as the endogenous TS gene in growth-stimulated cells. Therefore, the sequences responsible for growth-regulated expression reside within these intron-containing minigenes.

It has been demonstrated that introns 5 and 6 or introns 1 and 2 can stimulate the expression of TS minigenes about 8 to 20-fold relative to an intronless minigene (Deng et al., 1989b; Li et al., 1991). Inclusion of these introns into the intronless minigene made it possible to study the growth-regulation of mouse TS minigenes (since it was not possible to detect the transcripts of the intronless TS minigenes in stably transfected cells).

It has been shown that introns can stimulate the expression of a number of other S-phase genes, including the Chinese hamster DHFR gene (Venolia et al., 1987), the human TK gene (Lipson et al., 1989), and the human proliferating cell nuclear antigen (PCNA) gene (Ottavio et al., 1990). A question one might ask is that apart from the stimulatory effect, do introns also play a role in growth regulation of these S-phase genes? In regard to this question, Ottavio et al. (1990) found that introns of PCNA gene, in general, increase the expression of the gene in proliferating cells. However, in the absence of intron 4, the PCNA gene cannot be regulated normally in G₀ phase cells. On the other hand, Lipson et al. (1990) have shown that introns of human TK gene can increase the amount of TK mRNA, but have no effect
on the growth-stimulated expression. In the case of the TS gene, it is unlikely that the putative regulatory sequence is located within a specific intron, since minigenes containing either introns 5 and 6 or introns 1 and 2 are regulated normally in response to growth-stimulation. However, the possible involvement of TS introns in growth-regulation of the TS gene cannot be excluded. Unfortunately, the intronless minigene could not be studied due to its extremely low expression in stably transfected cells.

In summary, sequences responsible for growth-regulated expression of the TS gene reside within these intron-containing minigenes. TS introns are necessary for high level expression of TS gene but the sequences responsible for growth-regulated expression of the TS gene do not appear to reside within a specific intron.

The 5'-flanking region of the mouse TS gene is necessary but not sufficient for normal regulation in growth-stimulated cells. To study the role of the 5'-flanking sequences in the growth-regulated expression of mouse TS gene, I constructed a chimeric TS minigene which contains the SV40 early promoter, TS coding region with introns 5 and 6, and TS polyadenylation signal. Fig. 19 shows that substitution of the SV40 early promoter for the TS promoter led to a dramatic reduction in growth-regulated expression of TS minigene. This finding suggests that sequences which are important for the normal regulation of the TS gene in response to growth-stimulation are located, or at least in part located, within the 5'-flanking region of the TS gene.
To further delineate the approximate location of the regulatory sequences in the 5'-flanking region, the regulation of several TS minigenes with shortened 5'-flanking sequences was analyzed. Fig. 20 shows that deletion of sequences upstream of -250 results in a partially growth-regulated expression. When the 5'-flanking region is deleted to -150, the growth-regulated expression is dramatically reduced (Fig. 21). Since all of the elements that are sufficient for normal transcriptional initiation are located downstream of -150, at least some of the regulatory sequences are located upstream of the essential promoter elements.

The promoters of several S-phase genes have been demonstrated to be important for growth regulated expression of these genes. Examples include the TK gene (Roehl and Conrad, 1990; Kim et al., 1988; Knight et al., 1987; Travali et al., 1988; Stewart et al., 1987; Lipson et al., 1990), Histone genes (Ito et al., 1989; Dalton and Wells, 1988; Artishevsky et al., 1987; Seiler-Tuyns and Paterson, 1987; van Wijnen et al., 1988), and the PCNA gene (Chang et al., 1990). Moreover, some sequence elements in the promoter region of several S-phase genes have been identified to function as regulatory elements for growth-regulated gene expression. Dalton and Wells (1988) have clearly demonstrated that the sequence 5′AAACACA3′ (HI-box) in the promoter of chicken histone HI gene can modulate S-phase expression of the gene. Ito et al. (1989) have found that the binding activity of the octamer 5′ATTTGCAT3′ in the promoter of human H2b histone gene varies during the cell cycle (Ito et al., 1989). More recently, gel mobility shift analysis showed a protein factor (designated Yi) that binds to a consensus sequence CCCNCNNNCT in the promoter of the murine TK gene is absent in G₀ and G₁
cell extracts but present in G₁/S cell extract, presumably due to the change in
the stability of Yi protein (Dou et al., 1991; Bradly et al., 1991). Interestingly, as
shown in Fig. 2, the TS 5'-flanking region also contains the above putative
elements but with different extents of homology, 6 out of 7 nucleotides
identical to HI box, 6 out of 8 nucleotides identical to the octamer, and 10 out
of 10 nucleotides identical to the Yi sequence, respectively. In addition, all of
the putative sequences are located within a 35-nucleotide region between
-162 to -197 relative to ATG translational start codon. This finding appears to
correlate well with the deletion analysis which shows that when deleted to -150,
the minigene can no longer be regulated in response to growth stimulation.
The significance of these consensus sequences in mouse TS promotor is presently being established by further deletion and point mutation
analysis.

Although the upstream sequences are necessary for growth-regulation
of TS gene expression, it appears they are not sufficient since a chimeric
minigene with 1-kb of the TS 5'-flanking sequence linked to the CAT coding
region is expressed constitutively in growth-stimulated cells (Li et al., 1991).
This observation suggests that proper regulation of the mouse TS gene
depends on the presence of multiple regulatory elements that are located in
more than one region of the gene. Similar observations have been made
recently by Fridovich et al. (1991). One possibility is that TS gene transcription
is regulated by a downstream enhancer sequence that is necessary (but not
sufficient) for growth phase-specific transcriptional regulation. This
downstream control sequence would presumably function in cooperation with
upstream elements to bring about proper regulation. It is unlikely that the
putative enhancer sequence is located within one of the TS introns, since minigenes containing either introns 5 and 6 or introns 1 and 2 were both regulated normally as discussed previously. It is possible that the putative enhancer is located in the 3'-flanking region or even within the coding region.

Earlier studies on the metabolism of TS mRNA indicated that TS gene expression is controlled to some extent at the transcriptional level but to a much greater extent at the posttranscriptional levels in both mouse and human cells (Ayusawa et al., 1986; Jenh et al., 1985a). Therefore, an alternative explanation for the above results is that sequences upstream of the transcriptional initiation sites are somehow having an effect on a posttranscriptional regulatory process. It is possible that communication between transcription and RNA processing could be mediated by direct interactions between transcriptional initiation complexes and splicing or polyadenylation complexes. The concept of communication between transcriptional and posttranscriptional processes is not entirely novel. For example, it was shown that splicing and polyadenylation occur very inefficiently when the promoter for a rRNA gene is used to drive transcription of a protein-coding region (Sisodia et al., 1987). In addition, proper 3′ end formation of U1 and U2 small nuclear RNAs appears to require a special transcription complex which is specified by sequences in the 5'-flanking region of the small nuclear RNA gene. Substituting other RNA polymerase II promoters results in incorrect 3′ end formation (Hernandez and Lucito, 1988; Hernandez and Weiner, 1986; Neuman de Vegvar et al., 1988).
The 3'-flanking region is not essential for growth-regulated expression of the mouse TS gene. Early studies have demonstrated that the polyadenylation state of TS transcripts in the cells varies during growth stimulation (Jenh et al., 1985a), which suggests that the polyadenylation of the TS transcript may be regulated in growth-stimulated cells. To further investigate the role of the 3'-flanking region, I constructed a chimeric TS minigene which contains the TS promoter, the TS coding region with introns 5 and 6, and the SV40 early polyadenylation signal. As shown in Fig. 22, replacement of TS 3'-flanking region with SV40 early polyadenylation signal has no significant effect on growth regulation of the chimeric minigene.

The most straightforward explanation for this result is that 3'-flanking region is simply not essential for growth-regulation of the mouse TS gene. However, it should be noted that the upstream polyadenylation signal AUUAAA of TS gene is within the TS coding region and therefore is still present in the chimeric gene. It is possible that this TS hexanucleotide sequence is still functional as an upstream polyadenylation signal in the chimeric minigene. This would require that the SV40 sequence provide the downstream polyadenylation signal. However, 3' S1 nuclease protection analysis by John Ash revealed that most, if not all, of the TS mRNA derived from the chimeric gene is polyadenylated at the SV40 early polyadenylation site, indicating the upstream polyadenylation signal of TS gene is not functional. Thus, it appears that the 3'-flanking region of mouse TS gene is not essential for growth-regulation.
In summary, the 5'-flanking region is necessary but not sufficient for normal regulation of mouse TS gene in response to growth stimulation, whereas the 3'-flanking region does not seem to be essential for the growth regulation of the mouse TS gene. A complete understanding of the complexity in growth regulation of gene expression is still to come, although significant progress has been made in this direction.
LIST OF REFERENCES


