THE RAT C-MOS: ITS EXPRESSION AND STRUCTURE

by

KYOKO KOISHI

A thesis submitted for the degree of
Doctor of Philosophy

of the Australian National University

March 1990
DECLARATION

I hereby declare that the work recorded in this thesis has been performed by myself and that it has not been submitted in any previous application for a degree.

[Signature]

31 March, 1990

Kyoko Koishi
ACKNOWLEDGEMENTS

I came to Australia with many purposes. I wanted to study more about Biology, to train my English as a communication tool and wanted to know what I can do in this field. These questions are still in my mind as they were and what I have realized is that I will be asking the same questions in my life.

I would like to thank my supervisor, Dr Hiroto Naora, for giving me an opportunity to study in his unit and his patience.

I am grateful to Antony W. Braithwaite and Ian S. McLennan for their advice and comments on my thesis. Every effort given by them without any obligation has been appreciated. Thank you (to be honest, it was sometimes hard to listen, though). I would also acknowledge Peter J. Matthews for his encouragement. I was not only helped by them but also impressed very much by their personal attitude to seek blunt truth in Nature. I would think their straight-forwardness at work should be a part of scientific inquiry.

The support given by Brett M. Tyler, D. Stuart Letham and Chris D.S. Buller is acknowledged greatly. I thank for help and friendship given by Helen Liszczynsky, David Buckle, Felice Driver, Kaoru Miyahara, Hisakazu Matsushige and Mei-Lin Wong. During my PhD course, I have had lots of help from many people.

This study has been supported by a PhD scholarship given by the ANU, which is greatfully acknowledged.
ABSTRACT

This thesis is concerned with the study of the mechanisms that regulate gene expression. The focus of the study was the c-mos, a cellular homologue of the transforming gene of the Moloney murine sarcoma virus. This gene was of particular interest as several previous studies had failed to detect its transcription in a range of different tissues.

The expression of the c-mos gene was investigated in several tissues using Northern blot hybridization. Transcripts were detected in the brain and testis, but not the liver. The sizes and amounts of the transcripts differed between the brain and testis, with the latter producing the most abundant transcripts. The smallest transcript, 0.6 kb, was detected in the brain and the largest, 4.0 kb, in the testis. Intermediate-sized transcripts were also observed in both tissues. The 5' ends of the testis poly (A)+ RNA were localized at 0.68, 1.0 and 1.2 kb upstream and 3' ends at 0.26 and 0.43 kb downstream from the c-mos coding region. This indicated there is a long leader sequence present in the c-mos transcripts. The leader sequence was analyzed by sequencing and compared to recent reports on the structure of the gene in other species. Nine ATGs were newly found in addition to previously identified 3 ATGs proximal to the major c-mos open reading frame and the presence of multiple ATGs may be involved in translational control of the c-mos expression.
Expression of the c-mos had been suggested to be repressed by the presence of a transcriptional unit in close proximity to it. The possible presence of an open reading frame near c-mos was investigated by hybridizing it with a cDNA probe prepared from rat liver RNA. A positive signal was obtained and localized to 5.6 kb downstream of the coding region. Further analysis indicated the putative transcriptional unit contained two repetitive sequences, a B2 element and the rat ID sequence. These sequences are included in many transcripts and the relationship between the observed transcripts and the c-mos locus is thus uncertain.

Methylation of the c-mos locus had also been suggested to repress its expression and this was examined by investigating the methylation status of the c-mos with methylation-sensitive and -insensitive restriction enzymes (HpaII/MspI). The Hpa II/Msp I sites in the c-mos coding region were highly methylated in tissues which express the c-mos, testis and brain, and also in the liver, that does not express c-mos. There were, however, tissue-specific variations in the sites and density of the methylation. Two Hpa II/Msp I sites at approximately 3.5 kb upstream of the rat c-mos coding region were observed to be unmethylated in all three tissues examined.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>CDTA</td>
<td>trans-1, 2-diamino cyclohexane-N,N,N',N'-tetraacetic acid</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>c-mos</td>
<td>cellular mos</td>
</tr>
<tr>
<td>CsCl</td>
<td>caesium chloride</td>
</tr>
<tr>
<td>dATP</td>
<td>deoxyadenosine triphosphate</td>
</tr>
<tr>
<td>DEPC</td>
<td>diethylpyrocarbonate</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNase</td>
<td>pancreatic deoxyribonuclease I</td>
</tr>
<tr>
<td>d.H2O</td>
<td>distilled water</td>
</tr>
<tr>
<td>dsDNA</td>
<td>double stranded DNA</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>E. coli</td>
<td>Escherichia coli</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediamine tetraacetic acid</td>
</tr>
<tr>
<td>EtOH</td>
<td>ethanol</td>
</tr>
<tr>
<td>G6PD</td>
<td>glucose-6-phosphate dehydrogenase</td>
</tr>
<tr>
<td>GRE</td>
<td>glucocorticoid-responsive element</td>
</tr>
<tr>
<td>HEPES</td>
<td>4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid</td>
</tr>
<tr>
<td>hpert</td>
<td>hypoxanthine phosphoribosyl transferase</td>
</tr>
<tr>
<td>IAP</td>
<td>intercisternal A particle</td>
</tr>
<tr>
<td>IgG</td>
<td>γ-immunoglobulin</td>
</tr>
<tr>
<td>IgHs</td>
<td>immunoglobulin heavy chains</td>
</tr>
<tr>
<td>IgM</td>
<td>μ-immunoglobulin</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl-β-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>LTR</td>
<td>long terminal repeat</td>
</tr>
<tr>
<td>5mC</td>
<td>5-methylcytidine</td>
</tr>
</tbody>
</table>
MMTV  mouse mammary tumour virus
Mo-MLV  Moloney murine leukemia virus
Mo-MSV  Moloney murine sarcoma virus
MOPS  3-[N-morpholino]propanesulphonic acid
mos  mouse sarcoma
mRNA  messenger RNA
nt  nucleotide
PEG  polyethylene glycol
PIPES  piperazine-N,N'-bis[2-ethanesulphonic acid]
RIS  rat inhibitory sequence
RNA  ribonucleic acid
RNase  pancreatic ribonuclease A
r.p.m.  revolution per minute
SDS  sodium dodecylsulphate
snRNA  small nuclear RNA
SV40  simian virus 40
TAT  tyrosine aminotransferase
tk  thymidine kinase
Tris-HCl  tris-(hydroxymethyl)aminomethane hydrochloride
tRNA  transfer RNA
UPEs  upstream elements
UMS  upstream mouse sequence
v-mos  viral mos
X-Gal  5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Acknowledgements</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>4</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>6</td>
</tr>
</tbody>
</table>

## CHAPTER ONE. GENERAL INTRODUCTION:

### BACKGROUND TO THE PROJECT.

1. 1 The experimental material: the *c-mos* gene.  

1. 1. 1 Discovery of the *c-mos*.  

1. 1. 2 Oncogenes.  

1. 1. 3 Transforming activity of *mos*.  

1. 2 Transcriptional regulation of eukaryotic gene expression.

1. 2. 1. 1 DNA signals.

1. 2. 1. 2 Sequence-specific DNA binding proteins.

1. 2. 2 RNA sequence-protein interaction as a regulator.

1. 2. 2. 1 Capping and 5'end formation.

1. 2. 2. 2 Termination and 3'end formation of mRNAs.

1. 2. 2. 3 Splicing: Introns and exons.

1. 2. 3 Production of multiple mRNA species.

1. 3 Factors working on gene expression in a wide context.

1. 3. 1 DNA Methylation.
1.3.1.1 Methylation as a modulator of protein interactions with DNA signals.

1.3.1.2 CpG islands; non-methylated G+C-rich clusters in vertebrate genomes.

1.3.1.3 Methylation as a long term regulator.

1.3.2 Chromatin structure and gene expression.

1.4 A search for regulatory mechanisms of \( c-mos \) expression.

CHAPTER TWO.

MATERIALS AND METHODS.

2.1 Source of materials.

2.1.1 Animals.

2.1.2 DNA clones.

2.2 DNA preparations.

2.2.1 Growth of bacteriophage and preparation of its DNA.

2.2.2 Growth of bacteria and preparation of plasmid DNAs.

2.2.3 Mini-scale preparations of plasmid DNA.

2.2.4 Transformation of \( E. coli \).

2.2.5 Genomic DNA preparation.

2.3 RNA preparation from tissues.

2.3.1 Isolation of total RNAs from frozen tissues.

2.3.2 Preparation of poly (A)+ RNA enriched fractions from total RNAs.
| 2.4 | Electrophoresis of DNA or RNA fragments on agarose gels. | 74 |
| 2.5 | Preparation of radioactively-labelled probes for hybridizations. | 75 |
| 2.5.1 | Randomly-labelled DNA probes. | 75 |
| 2.5.2 | End labelling of DNA fragments. | 76 |
| 2.5.3 | Single-stranded RNA probes. | 77 |
| 2.6 | Hybridization conditions. | 78 |
| 2.7 | S1 nuclease protection assays. | 81 |
| 2.8 | Determination of DNA sequences. | 83 |

**CHAPTER THREE.**

**CHARACTERIZATION OF C-MOS TRANSCRIPTS IN RAT.**

| 3.1  | Introduction. | 86 |
| 3.1.1 | Characteristics of murine c-mos transcripts. | 86 |
| 3.2 | Results. | 88 |
| 3.2.1 | Detection of c-mos transcripts in rat tissues. | 88 |
| 3.2.2 | Localization of the 5' and 3' termini of testis c-mos transcripts on a genomic DNA map. | 90 |
| 3.2.3 | Multiple ATGs are found in the leader sequence of rat c-mos transcripts. | 92 |

**Figures**

| 3.3 | Discussion. | 109 |
3. 3. 1  Rat \textit{c-mos} transcripts.  
3. 3. 2  The \textit{c-mos} transcripts of various species.  
3. 3. 3  The termini of rat testis transcripts.  
3. 3. 4  Characteristics of the leader sequence in the rat \textit{c-mos} transcripts.  
3. 3. 5  Summary.  

\textbf{CHAPTER FOUR.}  
A SEARCH FOR TRANSCRIPTIONAL UNITS IN THE PROXIMAL REGION OF C-MOS.  
4. 1  Introduction.  
4. 1. 1  A search for mechanisms that regulate expression of the \textit{c-mos} gene.  
4. 1. 2  Spatial arrangement of genes in chromosomes.  
4. 2  Results.  
4. 2. 1  cDNA hybridization assays to detect putative neighbouring transcriptional units.  
4. 2. 2  Characterization of a DNA fragment from the 3' \textit{c-mos} flanking region.  
4. 2. 3  Characterization of the sequences in the 2.7 kb \textit{Pst} I fragment.  
4. 2. 4  Does the murine \textit{c-mos} flanking region contain a homologous sequence to the \textit{Sac I-Eco} RI fragment?  
4. 2. 5  Nucleotide sequence of the \textit{Sac I-Eco} RI fragment.
Figures

4. 3 Discussion. 147
4. 3. 1 Transcriptional units in the proximal region of the c-mos. 147
4. 3. 2 Repetitive sequences in the downstream region of the c-mos. 148
4. 3. 3 Possible roles of the B2 elements in the c-mos transcription. 150
4. 3. 4 Summary 153

CHAPTER FIVE.

DNA METHYLATION OF THE RAT C-MOS REGION. 154

5. 1 Introduction. 155
5. 1. 1 Tissue-specific pattern of DNA methylation. 155
5. 1. 2 Methylation of the c-mos sequence. 157
5. 2. Methods: Detection of DNA methylation using restriction endonucleases. 159
5. 3 Results. 161
5. 3. 1 Methylation of Hpa II/Msp I sites in the rat c-mos region. 161
5. 3. 2 Presence of non-methylated Hpa II/Msp I sites in the 5′ flanking region of the rat c-mos. 164

Figures 166
5. 4 Discussion. 173
5. 4. 1 DNA methylation of the c-mos region.
5. 4. 2 Various modifications observed between testis and brain.
5. 4. 3 Unmethylated Hpa II sites in the 5' flanking region of c-mos.
5. 4. 4 Summary.

CHAPTER SIX

GENERAL DISCUSSION.

6. 1 Introduction.
6. 2 C-mos transcription in various species.
6. 3 Transcriptional unit for c-mos expression.
6. 4 Regulation of c-mos expression in rats.
6. 5 Biological properties of the c-mos product.

REFERENCES.
CHAPTER ONE

GENERAL INTRODUCTION: BACKGROUND TO THE PROJECT.
1.1 The experimental material: the \textit{c-mos} gene.

\textit{C-mos} is a cellular homologue of the transforming gene of the Moloney murine sarcoma virus. This gene has the potential to cause drastic phenotypic changes in cells when it is transduced into a viral genome. However, at the start of this project, \textit{c-mos} transcripts had not been detected in normal animals. Given the sensitivities of the techniques that were used, this indicated that the \textit{c-mos} is not transcribed at more than an average of 1-3 copies per cell (Flankel & Fishinger, 1976; Gattoni-Celli et al, 1982; Müller et al, 1982). It was thus suspected that the \textit{c-mos} is a silent gene or possibly a pseudo-gene and hence nonfunctional. The putative \textit{c-mos} coding region is, however, highly conserved between species, which is a feature of genes of a vital importance in cellular functions. Comparison of the nucleotide sequences of \textit{c-mos} clones from rat (van der Hoorn & Firzlaff, 1984), mouse (Oskarsson et al, 1980; Jones et al, 1980; van Beveren et al, 1981b), human (Watson et al, 1982), chicken (Schmidt et al, 1988) and \textit{Xenopus} (Sagata et al, 1988; Freeman et al, 1989) indicates that they contain continuous open reading frames of approximately 1 kb. The similarity of the DNA sequences between species is: mouse-rat, 93\% (van der Hoorn & Firzlaff, 1984); mouse-human, 77\% (Watson et al, 1982); human-chicken, 63\% (Schmidt et al, 1988); mouse, human-\textit{Xenopus}, 50-52\% (Sagata et al, 1988). The amino acid sequences deduced from those nucleotide sequences has an average similarity of 75\% between mammalian species (Watson et al, 1982).
An initial objective of my thesis was, therefore, to examine whether or not the c-mos is transcribed in normal rat tissues. Low levels of c-mos transcripts were detected and these observations are discussed with reference to other recent reports of c-mos expression (Chapter 3). The second question that I sought to address in my thesis was: Does some feature(s) of the flanking or coding region of the c-mos account for the low levels of its expression? Experimental results relating to this question are presented in Chapters 4 and 5.

1.1.1 Discovery of the c-mos.

The Moloney murine sarcoma virus (Mo-MSV), in which the mos was initially found, was unexpectedly discovered through studies of a tumour virus, the Moloney murine leukemia virus (Mo-MLV). The Mo-MLV causes leukemia in mice, rats, and guinea-pigs with a long latent period (Moloney, 1960; Dunn et al, 1961). However, when mice were inoculated with high doses of a murine-derived Mo-MLV stock, they developed rhabdomyosarcomas instead of leukemia (Perk & Moloney, 1966). The virus stock was recovered from the sarcomas and found to reproducibly cause rhabdomyosarcomas in mice, in 3 to 5 days, with a 100% incidence. This virus stock did not induce tumours in rats, hamsters or guinea-pigs, but could be maintained in selected culture cells derived from rodents. Antisera against Mo-MLV reduced its tumorigenic activity but did not change the type of tumour produced or its species.
specificity (Moloney, 1966). Thus, the Mo-MSV was identified as a new virus which has an immunological and morphological inter-relationship with the Mo-MLV.

An initial question was, what was different between the Mo-MLV and Mo-MSV? The biological properties of the Mo-MSV measured by tumorigenicity, were stable after sequential inoculation. Therefore, the difference was assumed to be inheritable. The genetic material, which is RNA in these viruses, was isolated from Mo-MLV and Mo-MSV and compared with each other by hybridization. It was found that the Mo-MSV contains a specific sequence which is not shared by the Mo-MLV (Scolnick et al, 1975). The next question was, where did the unique portion of the Mo-MSV genome come from?

It was also known at this stage of the study that vertebrates possess sequences related to RNA tumour virus genomes in the form of DNA (Lowy et al, 1971; Benveniste et al, 1974). The cellular DNA of the host was therefore taken into account as a possible source of the Mo-MSV specific sequence. To examine this possibility, complementary DNA (cDNA) was prepared from the Mo-MSV and the portion which did not hybridize with the Mo-MLV genome was isolated. This Mo-MSV-specific cDNA was hybridized with DNAs derived from mice and other animals. The Mo-MSV-specific cDNA hybridized with murine cellular DNA at a high efficiency and with the DNAs from other animals to lesser
The results indicated that there is a homologous sequence in the murine genome (Frankel & Fischinger, 1976; 1977). The cellular homologue of the Mo-MSV-specific sequence was isolated from a murine genomic DNA library and was designated as mos (mouse sarcoma) (Oskarsson et al, 1980; Jones et al, 1980; Coffin et al, 1981). The sequence in the original virus genome is called v-mos (viral-mos) and the homologue in the animal genome is called c-mos (cellular-mos).

The mechanism by which the Mo-MSV arose is thought to have been a recombination event that led to replacement of a part of the genome of the Mo-MLV by a piece of host DNA (Frankel & Fischinger, 1976; Jones et al, 1980). The recombination of a viral genome and cellular DNA is not unique to the generation of the Mo-MSV and may be common to many other viruses. The c-mos is one of a group of genes, called proto-oncogene, which were discovered during the investigation of the biological properties of tumour viruses.

1.1.2 Oncogenes.

The concept of oncogenes arose from the study of tumour viruses (Huebner & Todaro, 1969). Some tumour viruses contain RNA as their genetic material and are thus designated as retroviruses as their genetic information is converted from RNA to DNA during their propagation (reviewed by Hughes, 1983). When genomes of the acutely transforming
retroviruses were analyzed, parts of the genomes were found to be closely related, but not identical, to host DNA sequences that occur in uninfected host cells (Benveniste et al, 1974; Stehelin et al, 1976). Such nucleotide sequences have been characterized by physical, biochemical, and recombinant DNA techniques to define the limits of the sequences in the viral genomes. For example, this can be done by comparing the nucleotide sequences of a transforming virus, its non-transforming but replication competent helper, and the related cellular sequences with each other and with the amino acid sequences of the suspected gene products. The specific sequences have been generally referred to as onc genes (Baltimore, 1975). Related sequences found in the cells are designated as c-onc genes (cellular oncogenes) (Coffin et al, 1981) or proto-oncogenes (Bishop, 1981). The viral transforming gene, on the other hand, is referred to as v-onc genes (viral oncogenes) (Coffin et al, 1981).

Do cellular oncogenes cause cancer? The answer may be that cellular oncogenes can cause neoplasia in animals when they are inappropriately expressed in particular cell types at a particular time. In other cell types, or at a different stages of development, expression of the oncogene may be vital for the cell's survival, proliferation or differentiation. For instance, the c-ras gene is expressed in embryonic and foetal tissues (Slamon & Cline, 1984), and its protein product can either promote or inhibit cellular differentiation, depending on the cell types in which the c-ras is expressed (Feramisco et al, 1984; Bar-Sagi &
Feremisco, 1985). *C-ras* is not normally active in a transformation assay using NIH3T3 cell (Stacey & Kung, 1984). However, transformation activity of the *c-ras* can be induced by DNA point mutations which cause a single amino-acid change (Taparowfsky et al, 1982) or a high level of production of the normal transcripts (Pulciani et al, 1985). Abnormal expression of *c-ras* has also been shown to cause unscheduled cell proliferation *in vivo*. When the *c-ras* gene was fused with the elastase I gene, integrated into germ cells and transgenic mice were generated, most of the mice developed pancreatic tumours (Quaife et al, 1987). The transformation is specific to the combination of gene and tissue: transgenic mice bearing a chimeric gene consisting of elastase I and *c-myc* do not show neoplastic change in their pancreases (Quaife et al, 1987), even though *c-myc* is capable of transforming lymphoid cells in a different induction sequence (Stewart et al, 1984; Adams et al, 1985).

Studying oncogenes has contributed not only to understanding the mechanisms of carcinogenesis but also to the study of proteins involved in the growth regulation of cells. A good example is the thyroid hormone receptor (Weinberger et al, 1986; Evans, 1988) which was known to be a nuclear protein and some of its biochemical properties had been characterized. Attempts to isolate and purify the receptor had failed, mainly because it is present only in small amounts in the cell nucleus. When the human glucocorticoid receptor was purified and its amino acid sequence was determined, homologous proteins to it were searched for in
a databank. This search revealed certain homologies between the receptor protein and the \textit{v-erb-A} oncogene product of the avian erythrocytosis virus (Weinberger et al, 1985). The cellular counterpart of the virus gene, \textit{c-erb-A}, was cloned using a part of the \textit{v-erb-A} gene as a probe and its protein product was identified as a thyroid hormone receptor by binding assays with a number of ligands (Weinberger et al, 1986). Having the cDNA of the \textit{c-erb-A} made it possible to discover multiple forms of the receptor protein, which had not previously been known to exist (Izumo & Mahdavi, 1988). This case has shown that results obtained from the study of oncogenes have a direct relationship to phenomena which have vital importance in living cells.

Another example is the human \textit{c-sis} gene that encodes the B chain of platelet derived growth factor (PDGF). While PDGF was being characterized by amino-acid sequencing, it was found that the sequence is more than 90\% identical with the amino acid sequence deduced from the \textit{v-sis} gene (Simian sarcoma virus transforming gene) (Waterfield et al, 1983). These findings led to the identification of the \textit{c-sis} product as the B chain of PDGF (Jonsson et al, 1984). When a \textit{c-sis} cDNA clone was examined in an \textit{in vitro} transformation assay, it transformed NIH3T3 cells (Josephs et al, 1984). This indicates that cellular oncogenes have the potential to cause cancer, but it should be noted that the \textit{c-sis} gene was in an entirely different context to its natural circumstances. The \textit{v-sis} gene is necessary for the virus to initiate and maintain the transformed condition.
and its regulatory region is completely heterologous with the host genome, in contrast to the highly conserved sequence of coding region. This implies that the \textit{v-sis} gene originated from a cellular gene and that its regulatory region has been modified (Jonsson et al, 1984).

More than 40 distinct oncogenes have now been identified in viral or cellular genomes (Weinberg, 1985; Bishop, 1987). The biochemical functions so far detected in the products are: (a) protein phosphorylation, with either tyrosine or serine and threonine as the substrate amino acids (\textit{src}, \textit{mos})(Hunter & Sefton, 1980; Maxwell & Arlinghaus, 1985); (b) metabolic regulation by proteins that bind GTP (the \textit{ras} family)(Hurley et al, 1984); (c) control of gene expression by influencing the biogenesis of mRNA (\textit{jun}, \textit{fos}, \textit{ski}) (Angel et al, 1988; Setoyama et al, 1986; Chiu et al, 1988; Barkas et al, 1986); (d) participation in the replication of DNA (\textit{myc})(reviewed by Cole, 1986). This variety of functions found in the oncogene products implies that there are multiple routes to stimulate cell reproduction. Categorizing these known oncogenes according to the location of the protein products (Weinberg, 1985) or relatedness (Bishop, 1987) reveals the presence of a network of signals used for replication control of vertebrate cells. Studying oncogene expression and their biological properties thus leads to insights into basic and common mechanisms involved in cellular proliferation.
1. 1. 3 Transforming activity of mos.

V-mos is part of an oncogenic virus and thus one of the first questions asked about v- and c-mos is whether they are oncogenes. When plasmids containing either the v-mos or c-mos, along with their flanking sequences, are transfected into murine fibroblasts, only the v-mos-transfected fibroblasts were transformed (Andersson et al, 1979; Oskarsson et al, 1980). However, when c-mos was connected with the 5′ flanking region of v-mos, which was assumed to contain regulatory sequences for the v-mos expression, the recombinant c-mos transformed the transfected cells (Oskarsson et al, 1980). This suggests that mos is an oncogene and that the c-mos and v-mos are able to produce functionally similar proteins. Mos is now normally classified as a member of the src family of oncogenes (see above) as its nucleotide sequence contains a region that is homologous to a protein domain common in kinases (Barker & Dayhoff, 1982) and because the protein product of v-mos has serine/threonine kinase activity (Maxwell & Arlinghaus, 1985).

The control of the expression of an oncogene is of vital importance to an organism due to the severe consequences of its inappropriate expression (see above). How then is the expression of c-mos controlled? The initial transformation experiments discussed in the previous paragraph suggest that some feature of the flanking region of c-mos
represses its expression, whereas the viral regulatory sequences promote transcription of \( v\text{-}mos \), independent of cellular control (Oskarsson et al, 1980; van Beveren et al, 1981b). The notion that \( c\text{-}mos \) expression is repressed due to its genomic circumstances was further supported by observations that melanoma cell lines, which contain rearrangements in their \( c\text{-}mos \) region, express \( c\text{-}mos \) (Rechavi et al, 1982; Cohen et al, 1983; Gattoni-Celli et al, 1983). For instance, XRPC24 melanoma cells contain an intercisternal A particle (IAP) gene inserted into the 5' region of \( c\text{-}mos \) (Rechavi et al, 1982; Canaani et al, 1983). The IAP contains enhancer-like sequences (Kuff et al, 1983) and may be responsible for the activation of \( c\text{-}mos \) expression in XRPC24 cells (Canaani et al, 1983).

As a prelude to considering how \( c\text{-}mos \) expression may be regulated by its normal cellular environment (Section 1.3), I will discuss some of the current ideas about how mRNAs are produced from a gene, with particular emphasis being given to regulatory mechanisms. As discussed in Chapter 3, \( c\text{-}mos \) is now known to be expressed in a tissue-specific manner, but only at low levels. This may be a reflection that the number of cells that produce \( c\text{-}mos \) transcripts is very small, or that \( c\text{-}mos \) is only expressed in rare circumstances or in small amounts or that its mRNA may be very short lived. Thus in the review, I have covered mechanisms which may govern cell-specific gene expression, mechanisms that may alter the efficiency of transcription and subsequent processing of the primary transcripts. The way in which multiple RNAs are produced
from a gene is also discussed as variously-sized *mos* transcripts have been
detected in testis and other tissues (Chapter 3).

1. 2 Transcriptional regulation of eukaryotic gene expression.

Transcription of genes in eukaryotes is performed by three RNA
polymerases, polymerase I, II and III, which have distinctive substrate
specificities. Polymerase I catalyzes the production of large ribosomal
RNAs, whereas the polymerase III produces 5S ribosomal RNA, transfer
RNAs (tRNAs) and some small nuclear RNAs (snRNAs). Protein-coding
genes are transcribed by RNA polymerase II in the nuclei of eukaryotic
cells. This review concentrates particularly on the transcription
performed by polymerase II, as the production of many diverse cellular
proteins depends on this enzyme's activity. Furthermore, *c-mos* has been
assumed to be a protein-coding sequence and the protein product of *v-mos*
has been recognized in cells transfected by a recombinant plasmid clone

Regulatory mechanisms commonly involve an interaction between
a nucleotide sequences and specific proteins leading into the formation of
a catalytic complex. DNA signals in the flanking region of a gene, in
particular, contribute to the activation of transcription of a gene in a cell-
specific and a temporal manner (McKnight & Kingsbury, 1982;
1. 2. 1. 1 DNA signals.

In prokaryotes, there is a sequence, TATAATG, which the RNA polymerase binds directly to and is indispensable for the initiation of transcription (Pribnow, 1975; Ptashne, 1986). A similar short AT-rich region known as a "TATA-box" or Hogness box has detected by comparing the nucleotide sequences in the promoter region of viral and vertebrate genes (the consensus sequence is: TAT/AAT/AA; Benoist & Chambon, 1981). The TATA-box is present in many, but not all, eukaryotic genes at 20-30 base pairs (bp) upstream of the coding region (Gannon et al, 1979; Reynolds et al, 1984; Melton et al, 1984a). Genes which lack the TATA-box are still transcribed but tend to have a variety of 5' ends, suggesting that transcription began at various points (Grosschadl & Bimstiel, 1980).

DNA also contains signals which appear to be involved in the activation or repression of gene expression. Promoters are an example of this type of signals. Promoters are typically approximately 100 bp long and consist of an AT-rich sequence, a TATA-box, and some short elements (Benoist & Chambon, 1980; Mathis & Chambon, 1981; reviewed by Maniatis et al, 1987; Dynan, 1989). The TATA-box is
virtually important in expression of many genes, however, a population of genes produce their transcripts despite lack of this element (Smale & Baltimore, 1989). The short DNA elements, which are designated upstream elements (UPEs) by Maniatis and his colleagues, are usually 8-12 bp sequences and contain one or more recognition sites for transcriptional regulatory proteins (Maniatis et al, 1987). Experimentally induced mutations or substitutions of a UPE in a promoter region decreases the expression of a gene in a transfection assay, indicating that UPEs normally enhance transcription (Dieks et al, 1983; Baty et al, 1984; Graves et al, 1986). Promoter regions often have a combination of UPE elements with differing protein binding specificities or competency to stimulate transcription (Mason et al, 1985). The biological properties and characteristics of UPEs are very similar to a second category of cis-acting elements, enhancers.

Enhancers have been identified in many viral genomes by using their recombinant clones and transfection into cultured cells. Polyoma virus, bovine papilloma virus and Mo-MSV are known to have strong enhancers (de Villierz & Schaffner 1981; Lusky et al, 1983; McClements et al, 1981). Some of the viral enhancers have host specificity and may result in host cells predominantly producing viral specific proteins (Lamins et al, 1982). Among viral enhancers, the SV40 enhancer element is a well studied example (reviewed by McKnight & Tjian, 1986). At a position very close to the viral replication origin, there is a cluster of G+C
rich sequences and two copies of a 72 bp element, which is the enhancer leading to the viral late transcriptional unit. Upstream from the G+C rich cluster, there are the viral replication origin and early viral transcriptional unit. Extensive disruption or removal of the 72 bp repeat impairs viability and reduces the early viral transcription (Benoist & Chambon, 1980). Only a single 72 bp unit appears to be required for transcription as removal of additional repeats of the 72 bp does not prevent transcription (Gruss et al, 1981).

The action of an enhancer is not specific to a particular gene or position relative to a coding region. The SV 40 enhancer, for instance, increases the transcription of a rabbit globin gene at locations up to 1400 bp upstream or 3300 bp downstream of the gene, in either orientation (Banerji et al, 1981). The definition of an enhancer is thus based on its functional properties and not on its position relative to a gene. For instance, the DNA signals located in the 3' flanking regions of the chicken β-globin gene are designated as enhancers (Choi & Engel, 1986).

Enhancers have been identified in several cellular genes (Maniatis et al, 1987), including the immunoglobulin genes (Banerji et al, 1983; Gillies et al, 1983; Picard & Shaffner, 1984; Bergman et al, 1984), glucocorticoid-inducible genes (Klein-Hitpass et al, 1986; Jantzen et al, 1987) and heat-shock protein genes (Bienz & Pelham, 1986). A well characterized example is the enhancers of the immunoglobulin heavy
chain gene, a gene which is expressed in B-lymphocytes at a low level in pre-stimulated cells and at an extremely high level in antigen-stimulated cells. The interaction of the enhancers with B-cell specific trans-acting protein factors has been suggested by protection of the regions from DNA reacting chemical reagents (Ephrussi et al, 1985; Mercola et al, 1985). The B-cell specific enhancer is located within the introns of the μ heavy chain and κ light chain genes (Banerji et al, 1983).

The analysis of the immunoglobulin heavy chain (IgH) enhancer has revealed an interesting feature of cis-acting signals. The enhancer includes several short elements that are collectively required to maximize the transcription rate. Investigation of the effect of each element on IgH transcription, by its removal or nucleotide substitution in the sequence, shows that a few of them have inhibitory influences for the transcription when they are exposed to non-B cell nuclear extract (Scholer & Gruss, 1985). A series of experiments with IgH enhancers have shown that an enhancer can be composed of multiple DNA elements which bind different protein factors and can influence the transcription in a positive or negative way. Negative enhancers, or “silencers”, are thus recognized as regulatory elements whose removal results in a gene having a higher rate of transcription (Kadesch et al, 1986).
1. 2. 1. 2 Sequence-specific DNA binding proteins.

DNA signals are commonly recognized by multiple protein factors and a combination of these factors define the cell- and temporal-specificity of the expression of a gene. Protein factors have been detected in nuclear extracts and identified according to their sequence-specific binding in various *in vitro* assays (Galas & Schmidt, 1978; Fried & Crothers, 1981). cDNA libraries cloned in expression vectors can be directly screened by DNA motifs (Vinson et al, 1988; Singh et al, 1988). Specificity of the DNA binding proteins has been analyzed by construction of the deletion mutants of both target DNA sequences and cDNAs encoding the proteins and their usage for binding assays. Analysis of the protein structures based on their amino-acid sequences, chemical or immunological characteristics provide insights into the binding property of these proteins. Such studies have shown that they are composed of a variety of usually separable DNA binding and transcriptional activation domains (reviewed by Jones et al, 1988; Mitchell & Tjian, 1989; Johnson & McKnight, 1989).

Within the DNA binding domain, a number of structural motifs are recognized and designated according to models of their three-dimensional structures; for instance, helix-turn-helix, zinc finger or leucine zipper (Steitz et al, 1982; Miller et al, 1985; Landschulz et al, 1988a). One of them, the zinc finger motif consists of about 30 amino acids with two pairs
of cysteine-cysteine or cysteine-histidine residues that stabilize the domain by tetrahedrally co-ordinating with a Zn$^{2+}$ ion. This motif was originally found in the RNA polymerase III transcription factor TFIII, which binds to the internal control region of the 5S RNA gene (Brown et al, 1985; Miller et al, 1985). Since then it has been established that the zinc finger motif is involved in many RNA polymerase II transcription factors including the glucocorticoid hormone receptor family (Evans & Hollenberg, 1988).

Another interesting motif found in DNA binding proteins is the "leucine zipper" (Johnson et al, 1987; Landchulz et al, 1988a, b). This motif was first described in the mammalian enhancer binding protein C/EBP (Johnson et al, 1987). It was recognized that a group of DNA binding proteins share a bipartite region of primary sequence that consists of a highly conserved stretch of about 30 amino acids with a substantial net basic charge which is followed immediately by a region containing four leucine residues positioned at intervals of seven amino acids. This latter part of the domain is named the leucine zipper. From the deduced protein structure and the results of deletion mutant analysis, it has been suggested that dimerization of proteins in this group is stabilized by hydrophobic interactions between closely opposed $\alpha$-helical leucine repeat regions of the two subunits. The basic region adjacent to the leucine motifs seems to be necessary for DNA binding but not for dimerization, and both subunits appear to contribute their basic regions to
form the DNA binding domain (Kouzarides & Ziff, 1988).

How do the various protein domains work to promote the initiation of a transcription? Recombination of DNAs encoding the motifs and formation of chimeric proteins have been used to investigate this question. The study of the yeast transcription factors, GAL4 or GCN4 provides a good example (Ma & Ptashne, 1987; Hope & Struhl, 1986; reviewed by Ptashne, 1988). When either of them is used together with heterologous DNA binding domains derived from other transcriptional activator proteins, the GAL4 or GCN4 can activate the transcription of reporter genes in yeast or in cells of a variety of higher organisms (Brent & Ptashne, 1985; Ma & Ptashne, 1987). Deletion analyses of the GAL4 or a GCN4-recombinant revealed that the two activators need stretches of amino acids bearing negative net charges but not common protein structures (Ma & Ptashne, 1987; Hope & Struhl, 1986). This characteristic is shared by domains of other transcription factors such as glucocorticoid receptors and AP-1/Jun (Bohmann et al, 1987; Hollenberg & Evans, 1988). It has been proposed that the acidic domain working for transcriptional activation may facilitate transcriptional initiation by interacting with general components of the transcriptional complex or even with RNA polymerase II (Guarente, 1984).

In summary, DNA signals in the flanking region of a transcriptional unit contribute to initiate production of an RNA in
conjunction with sequence-specific binding proteins. Both promoters and enhancers are composed of multiple DNA elements which interact with various protein factors; enhancers alter the transcription rate from a promoter region quantitatively, whereas promoters are essential for the accurate start of transcription. The enhancers can act from a great distance in an orientation-independent manner and can also exert an influence downstream from a transcription unit (Banerji et al, 1981; reviewed by Maniatis et al, 1987; Dynan, 1989).

1. 2. 2 RNA sequence-protein interaction as a regulator.

Nucleotide sequence signals in the primary transcripts contribute to the regulation of post-transcriptional modification and thus make up a second level of gene regulation before translation (reviewed by Manley, 1988; Humphrey & Proudfoot, 1988; Smith et al, 1989). Some of the mechanisms working in this phase involve sequence-specific protein binding to the signals and thus resemble the DNA signals-protein recognition that are involved in initiation. That is, RNA not only carries codes for the amino acid sequence of a protein(s) but also contains signals that affect its translation through the regulation of its stability, transport to the cytoplasm and cutting and splicing to form various mRNA species.
1. 2. 2. 1 Capping and 5'end formation.

Shortly after the initiation of transcription, the 5' end of the RNA is modified by addition of a guanosine triphosphate and methylation to form a "cap structure" m7G(5')ppp(5')N (Furuichi et al, 1975; Ziff & Evans, 1978). This modification is unique to eukaryotic mRNAs and, unlike other processes of transcription, the mRNAs from some mammalian viruses are not modified in this way (reviewed by Shatkin, 1985). This modification may not to be essential for translation of some mRNAs as uncapped mRNA produced from the poliovirus is translated in the host cells to produce a functional protein (Banerjee, 1980).

The presence of the cap structure stabilizes mRNA in the nucleus and cytoplasm (Furuichi et al, 1977; Green et al, 1983). Chemical analogues like m7G(5')p disturb the translation of capped mRNAs by inhibiting the interaction between 40S ribosome and the mRNAs (Sonenberg et al, 1978). This suggests that the cap structure is recognized by cytoplasmic protein factors which may be important in the initiation of translation. These protein factors have been identified by chemical cross-linking of the capped-mRNA-protein complex in which the cap position has been labelled and they have been designated as cap binding proteins (Sonenberg et al, 1978).
Another possible role of this capping is to increase the efficiency of the subsequent RNA processing by the interaction with nuclear proteins (Ederly & Sonenberg, 1985). Some experiments have indicated that splicing and 3' formation of mRNAs proceeds efficiently when the transcripts contain capped 5' ends (Konarska et al, 1984; Patzelt et al, 1987). However, the greater processing of capped RNA may merely reflect an increased stability of the capped intermediates in the nucleus rather than a direct effect on the way that the RNA interacts with the splicing machinery.

1.2.2 Termination and 3' end formation of mRNAs.

The 3' end of a mRNA is produced through post-transcriptional processes such as polyadenylation (described later), rather than the detachment of the RNA polymerase II from the 3' flanking region of a gene. The detachment sites have only been determined in a few genes due to the instability and scarce amount of the non-polyadenylated primary transcripts (reviewed by Platt, 1986; Friedman & Imperiale, 1987). From the limited information available it seems that the termination of eukaryotic transcription occurs at various points in the 3' flanking region (Citron et al, 1984; Hagenbuchle et al, 1984). For instance, examination of the murine α-amylase transcripts by S1 protection assay revealed that prior to polyadenylation its nuclear transcripts contain a long stretch of sequence which varies in length from
hundreds to thousands of base pairs (Hagenbuchle et al, 1984). Similarly, the chicken ovalbumin gene has a termination region, instead of a single termination site, about 900 bp downstream from the last exon. The ovalbumin transcripts are polyadenylated at various points in this region, which is 200 bp long and AT rich (LeMeur et al, 1984).

Polyadenylation was first observed in the mRNAs of globin, immunoglobulin and ovalbumin as clusters of adenine residues which are not present in the genomic sequences (Lim & Canellaskis, 1970; Edmonds et al, 1971; Birnboim et al, 1973). It appears to be a process separate from termination and involving the addition of a few hundred base pairs of adenine to the 3' end of a primary transcript. Recently, a poly(A) polymerase and protein factors involved in poly-adenylation have been purified from HeLa cells (Takagaki et al, 1988; Christofori & Keller, 1988). The cellular proteins were fractionated by chromatography and centrifugation, and assayed for their biological properties using an in vitro reaction. A combination of experiments revealed that correct poly-adenylation needs at least a poly(A) polymerase, a protein factor which governs precise cleavage at a fixed point and one of the small nuclear RNAs (snRNAs) (Christofori & Keller, 1988). The protein factor is thought to work in this system by recognizing consensus sequences, including AAUAAA, in order to determine the position of the polyadenylation.
Comparison of the 3' non-coding sequences of genes has revealed the presence of nucleotide sequences used as signals for 3' end formation. Two of them, AAUAAA and YGUGUUYY (Y = pyrimidine), are common in many eukaryotic mRNAs (Proudfoot & Brownlee, 1976; McLauchlan et al, 1985). The biological importance of these sequences was examined by creating mRNAs with point mutations or deletions in the consensus sequences. When the AAUAAA is modified to become AAGAAA, the transcripts have 3' sequences of variable length, although they are still properly modified by poly-adenylation, indicating this sequence is unimportant for determining the position of the polyadenylation site and the efficiency of reaction (Montell et al, 1983). Another consensus sequence YGUGUUYY (Y = pyrimidine) is found downstream from the AAUAAA (McLauchlan et al, 1985). Although this sequence is not as fundamentally important as AAUAAA, as the lack of the sequence does not abolish functional RNA production (Danner & Leder, 1985), requirement for downstream sequences have been observed in a number of systems (Gil & Proudfoot, 1984; 1987).

A group of RNAs produced by RNA polymerase II, for instance histone mRNAs and sn RNAs, are not processed by polyadenylation. They seem to have alternative combinations of nucleotide signals for 3' end formation. The histone pre-mRNAs possess two G+C-rich regions and a 15 bp purine-rich domain, which is complementary to U7 snRNA (Luscher et al, 1985). These homologous sequences work together to
produce the proper 3' end of the histone mRNA (Strub & Birnstiel, 1986).

The termination of the RNA polymerase II reaction is not as well understood as the initiation. However, the termination is an important part of transcription as inappropriate termination may lead to the production of a nonfunctional protein. Faulty termination has been postulated to be a part of the aetiology of genetically-based diseases such as thalassemia in man (Higgs et al, 1983; Orkin et al, 1985).

1. 2. 2. 3 Splicing: Introns and exons.

In eukaryotic genomes, many protein coding genes are split into blocks of coding region interrupted by non-coding regions (Brack & Tonegawa, 1977; Tonegawa et al, 1978; Mandel, 1978). The non-coding regions are designated as intervening sequences or "introns", whereas the sequences which are finally converted to matured mRNAs are called "exons" (Tonegawa et al, 1978). Prior to the transport and translation of a mRNA, the introns are deleted and the exons are ligated at precise positions (Breathnach et al, 1978). The splicing of introns from a precursor RNA occur after capping and either after or during polyadenylation of the primary transcript (Ross & Knecht, 1978; Ross, 1978; Gilmore-Hebert & Wall, 1979).
The splicing of nuclear transcripts of higher eukaryotes is performed by a two-step reaction in a multi-component complex, which contains snRNAs and protein-ribonucleoprotein complexes (Maniatis & Reed, 1987), which is referred to as a spliceosome (Padgett et al, 1984; Aebi et al, 1986). The first step of the splicing process is a cleavage of the 5' splice site and the formation of a linear upstream moiety and a downstream lariat structure in which the 5' end of the intron is covalently joined to the original RNA at 30-40 nucleotides upstream of the 3' splice site (Ruskin et al, 1984). In the second step, the upstream moiety is joined to the downstream exon, and the intron is released as a lariat (Zeitlin & Efstratiadis, 1984).

The start and end of an intron are indicated by dinucleotide consequence signals; GT at the start of an intron, splice donor site, and AG at its end, splice acceptor site (Breathnach et al, 1978; Mount, 1982). Mutations which occur at the 3' splice site influence splicing more than mutations at the 5' site (Wieringa et al, 1983; Aebi et al, 1986). Destruction of the consequence signal, AG, at the 3' splice site of an exon in a β-globin gene results in aberrant 3' splicing of the gene in vivo. When the expression of this mutated globin gene was studied in vitro it was observed that cleavage of the 5' end of the intron and lariat formation were inhibited (Aebi et al, 1986). This indicates that the structure of a transcript, which in vivo is presumably organized in conjunction with proteins, plays a critical role in splicing.
The removal of introns from an RNA has been postulated to be necessary to get the RNA through the nuclear membrane (Gruss et al, 1979). When the 16S mRNA transcription unit of SV40 was replaced by a cDNA copy of itself or a cDNA copy of β-globin, transcripts of the inserted cDNAs did not appeared in the cytoplasm as mRNAs (Hamer & Leder, 1979). However, this result was not reproduced when a different combinations of genes were used (Treisman et al, 1981; Gething & Sambrook, 1981). Furthermore, genes which originally have no intron, such as histone genes, are transported to cytoplasm, suggesting that splicing of a gene is not always a prerequisite for its transport to the cytoplasm. Another biological importance of having introns is that it provides an opportunity to have protein variants without a major change of the gene structure.

1. 2. 3 Production of multiple mRNA species.

Many eukaryotic genes are now known to generate multiple mRNA species from a single genetic locus (reviewed by Leff et al, 1986; Breitbart & Nadal-Ginard, 1987; Smith et al, 1989). The production of multiple mRNA molecules can occur at initiation, splicing or polyadenylation but not during elongation of the RNA. The three positions will be discussed separately.
Alternative 5' end: Some genes, such as the chicken and murine myosin light chains, murine α-amylase and *Drosophila* alcohol dehydrogenase genes, have multiple promoters (Nabeshima et al, 1984; Robert et al, 1984; Hagenbuchle et al, 1984; Benyajati et al, 1983). The multiple promoters can be used by RNA polymerase II separately and produce various 5' ends.

Multiple promoter regions do not result in the production of multiple protein isoforms except when subsequent selection of exons occurs. For example, the α-amylase gene has two promoter regions that do not directly affect the amino acid sequence of the proteins produced (Hagenbuchle et al, 1984; Nawa et al, 1984). Another example with an alternative production of isoforms is the myosin light chain. The comparison of the nucleotide sequences of genomic and cDNA clones of chicken myosin light chains indicates that two protein isozymes are formed from a single gene by using the two promoter regions followed by exclusive selection of two translated exons (Nabeshima et al, 1984).

Choice of variable exons: Regulated alternative splicing has been observed in some virus genes (Ziff, 1980) and cellular genes, with the pattern of splicing being dependent on sex (Nagoshi et al, 1988), tissue (Laski et al, 1986; Leff et al, 1987) or developmental stage (Breitbart & Nadal-Ginard, 1987). For instance, the gene for fast skeletal muscle troponin T (TnT), which confers calcium ion sensitivity to the myosin
ATPase, is encoded by a single gene with 18 exons that can produce 64 protein isoforms according to developmental and tissue-specificity (Medford et al, 1984; Breitbart et al, 1985). The isoforms are produced in two different ways: splicing of two exons at the 3' end (exons 16 and 17) in a mutually exclusive manner and selective combinations of five exons (exons 4-8). The mechanism governing the selection of the TnT gene exons was studied by using recombinant plasmids which included a part of the troponin gene (Breitbart & Nadal-Ginard, 1987). The same combination of exons and introns were alternatively spliced when they were transfected to non-muscle and muscle cells, suggesting that cell-type specific protein factors work on the selection. As exons 4-8 are similarly very small (12-18 nucleotide) and relatively homologous in their nucleotide sequence (Breitbart et al, 1985), the conformation of the initial transcript and cis-acting element could also be involved in determining the pattern of splicing.

**Termination variety:** Some genes have multiple polyadenylation sites and produce mRNA species having various 3' ends (Leff et al, 1986; Breitbart et al, 1987). A typical example is the adenovirus 2 major late message, which generates a diverse group of proteins from a large single transcript of 25 kilobases (kb) (Ziff, 1980). An example involving a cellular gene is the production of membrane bound and secreted forms of IgM (Perry & Kelly, 1979; Rogers et al, 1980; Early et al, 1980). The two forms of IgM have different carboxyl termini and only one can bind
to the cellular membrane. The mRNAs coding for the two proteins have the same sequence at the beginning and differ in the final parts of the coding and poly(A) sites. The gene encoding IgM has two sets of "exon-poly(A) sites" in its 3' region and two different forms of mRNAs are generated by selection of one or other of the sites. The mechanism governing the selection is not known.

In addition to the production of multiple transcripts from a single gene, the presence of two genes encoding a major unit of human glucose-6-phosphate dehydrogenase (G6PD) has been recently reported (Kanno et al, 1989). Prior to this discovery there has been a discrepancy that the previously cloned cDNA of the protein did not contain the message for a part of the amino-terminal peptide directly determined by amino acid sequence analysis (Takizawa et al, 1986; Persico et al, 1986). This discrepancy could not be explained by an incomplete message in the cDNA clone as a consensus translation start codon is present in the cDNA sequence (Persico et al, 1986). It was found that the amino-terminal portion of the G6PD major unit is encoded on a separate locus in chromosome 6 (Kanno et al, 1989), even though the gene encoding the carboxy-terminus of the protein is located in the X chromosome (Pai et al, 1980). Both of the two cDNAs isolated independently have a typical structure of a functional mRNA, therefore, it has been speculated that the G6PD mRNAs are generated separately and connected to produce a mature mRNA before the translation. A possibility, if two peptides are
separately produced and connected with an inter-peptide exchange, should be examined as a Northern-blot hybridization detected messages for the two distinct parts of the G6PD major unit (Kanno et al, 1989) but not for the integrated molecule.

1. 3 Factors working on gene expression in a wide context.

Various combinations of specific interactions between nucleotide sequences and proteins are an important part of transcriptional regulation (Jones et al, 1988; Mitchel & Tjian, 1989). However, this type of regulatory system can not operate if nucleotide sequences are, by some manner, masked from the protein factors. The accessibility of protein factors to DNA is another level of transcriptional regulation and the factors that may restrict the binding of regulatory proteins to DNA sequences are discussed in the following sections.

1. 3. 1 DNA Methylation.

DNAs from a broad spectrum of eukaryotes contain 5-methylcytosine, in addition to the four unmodified bases. 5-methyldeoxycytidine (5mC) is predominantly found as part of the dinucleotide, 5mCG, which is a relatively rare dinucleotide. In vertebrate genomes, the dinucleotide CG occurs at only one-fifth of the frequency expected from a
random distribution of nucleotides (Russell et al, 1976) and 60-90% of them are methylated at the C-residue (Grippo et al, 1968; Razin & Riggs, 1980; Bird, 1986). Although eukaryotes with small genome sizes such as yeast (Profitt et al, 1984) and Drosophila (Urieli-Shoval et al, 1982) lack this modification, DNA methylation is a dominant modification.

DNA modification has been postulated to play an important role in transcriptional regulation. This hypothesis is based on three lines of experimental results (reviewed by Razin & Riggs, 1980; Doerfler 1983, 1984). The first is that the enzymatic function of prokaryote restriction endonucleases, which are essentially sequence-specific binding proteins, is easily disturbed by modification of nucleotide residues of the target sequences (Bird & Flavell, 1978). The second is that some genes have tissue-specific methylation patterns which partially correlate with their tissue-specific transcriptional activity, although some genes do not show a particularly good correlation (van der Ploeg & Flavell, 1980). The third is that viruses, particularly the endogenous retroviruses of chicken, which are silent in the host cells can be activated by a potential methylation inhibitor, 5-azacytidine (Groudine et al, 1981). Furthermore genes or viral genomes integrated into host cells undergo gradual inactivation of their transcription which coincides with their methylation (Gautsch & Wilson, 1983). In the next section, I will discuss the current knowledge about DNA methylation in relationship to the regulation of transcription in eukaryotes.
1. 3. 1. 1 Methylation as a modulator of protein interactions with DNA signals.

The change in the configuration of a DNA sequence produced by its methylation can, in some instances, prevent regulatory proteins from binding to it and thus affect transcriptional activity. An example of this is the rat tyrosine aminotransferase (TAT) gene, which is known to be transcribed only in the parenchymal cells of the liver (Glueckson-Waelsch, 1987). The presence of nuclear protein factors bound to a region of the promoter was investigated using dimethylsulphate (DMS). DMF diffuses into cellular nuclei and modifies guanosine residues in DNA sequences which are not masked by proteins or some other factor (Ogata & Gilbert, 1978). The TAT gene promoter was protected from DMF modification in the TAT-expressing hepatoma but not in the TAT-non-expressing hepatoma variants or fibroblasts, indicating the presence of cell-type specific nuclear binding proteins (Becker et al, 1987). Furthermore, when the TAT gene promoter was methylated and assayed for protein binding using the DNase-I foot-printing method, the methylated region did not interact with the nuclear extract prepared from the TAT-expressing cells (Becker et al, 1987). This implies that DNA methylation can abolish the binding of specific nuclear transcription factors.
Does the interference of binding of nuclear factors caused by DNA methylation affect the transcriptional activity of a gene? Experiments done to date have not revealed a universal answer to this question. Methylation appears to have different affects on different genes and even on different regions within a given gene. One of the more extensive studies of the effect of methylation on a gene has been carried out by Doerfler and his colleagues using adenovirus genes (summarized in Dobrzanski et al, 1988). When the Ad 2 E2A gene was methylated with a sequence-specific methylase and introduced into *Xenopus* oocytes, the gene was not transcribed (Langner et al, 1984). In addition, when recombinant constructs containing the promoter region of the E2A gene and chloramphenicol transferase gene were used in a similar system, it was also shown that methylation of the promoter region was sufficient to inactivate or inhibit their transcriptional activity (Langner et al, 1986).

A more refined analysis of the importance of methylation was then undertaken (Toth et al, 1989). Various portions of the E2A late promoter region were methylated, transfected into cultured cells and transcriptional activity of the downstream genes was examined (Müller & Doerfler, 1987). The methylation of the region was directly checked by genomic sequencing, in which the methylated cytidine was detected as a gap in the C lane (Church & Gilbert, 1984). It was revealed that the methylation of selective parts of the E2A promoter region inhibited the transcription of the downstream genes (Toth et al, 1989). It is
noteworthy, however, that the methylation of a few of the DNA elements present in the E2A late promoter region did not inhibit transcription in *Xenopus* oocytes or in cultured cells (Verdimon et al, 1982).

On the other hand, some DNA sequences, such as the Sp1-binding site, appear to be unaffected by the methylation of their cytosine residues. Sp1-binding sites have been detected and shown to affect the transcriptional activity of SV40 early genes (Fradin et al, 1982; Greasmann et al, 1983; Tjian & McKnight, 1986) and many vertebrate genes, including the human and murine metallothionein genes (Briggs et al, 1986; Karin et al, 1984). In *in vitro* studies, the binding of nuclear extracts to the consensus sequence for Sp1 is not influenced by cytosine methylation (Harrington et al, 1988; Holler et al, 1988). Furthermore, methylated Sp1 binding sequences can enhance transcription of downstream genes when transfected into cells (Holler et al, 1988), thus suggesting that methylation does not always disturb or decrease transcriptional activity of genes even though the methylation of a single site can cause reduction of the transcription of some genes.

1. 3. 1. 2 CpG islands; non-methylated G+C-rich clusters in vertebrate genomes.

DNAs prepared from vertebrates are not easily cleaved with methylation-sensitive restriction endonucleases because of their high
density of methylation (Gruenbaum et al, 1981; Kunnath & Locker, 1982; Van der Ploeg & Flavell, 1980). However, about 1% of the genome of a wide range of vertebrates can be cleaved into a distinctive fraction of DNA by either \textit{Hpa} II, which recognizes CCGG and is methylation sensitive, or \textit{Hha} I, which recognizes GCGC and is methylation sensitive (Cooper et al, 1983). The characteristics of the C+G-rich fraction revealed by these enzymes are: (a) they contain short fragments of DNA with sizes ranging from 40 to 400 bp with a mean of 120 bp (Cooper et al, 1983; Bird et al, 1985); (b) the cytosine residues are non-methylated. The G+C-rich sequence clusters have been variously designated as \textit{Hpa} II tiny fragment islands (Bird et al, 1985), CpG islands (Bird, 1986; Gardiner-Garden & Frommer, 1987) or methylation-free islands (Kolsto et al, 1986). I will refer to the G+C clusters as CpG islands.

The numbers of CpG islands in chicken and murine genomic DNAs have been estimated to be 30,000 by Southern blot hybridizations (Cooper et al, 1983; Bird et al, 1985). To date, no extensive sequence homologies have been detected between the CpG islands that have been isolated and cloned (Bird et al, 1985). What they appear to have in common is their location- many CpG islands are located in the flanking regions of either low-copy number genes or ribosomal RNA genes (Cooper et al, 1983; Bird et al, 1985; Bird, 1986; Gardiner-Garden & Frommer, 1987).
Several lines of evidence indicate that the methylation status of a CpG island can influence the transcription of neighbouring genes. For instance, *in vitro* enzymatic methylation of the CpG island in the human *c-Ha-ras* gene flanking region reduces the transcription of a downstream reporter gene by 70-80% (Rachal et al, 1989). It is also known that CpG island-like sequences in LINE elements (transcriptionally active middle-repetitive sequences in humans) are heavily methylated in cells which do not produce detectable transcripts (Loeb et al, 1986; Hattori et al, 1985). Similarly, CpG islands on inactive X chromosomes are heavily methylated (Lock et al, 1986).

The biological properties of the CpG islands have not been clearly defined yet. However, it has been suggested that they contribute to the regulation of the transcription of large areas (Bird, 1986; 1987). This hypothesis has been supported by three lines of experiments, which were discussed above: (a) CpG islands have the potential to inhibit transcription of the downstream gene by methylation (Rachal et al, 1989); (b) the CpG islands frequently occur in the flanking region of genes (Bird, 1986; Gardiner-Garden & Frommer, 1987); (c) inactive X chromosomes of mammals have heavily methylated CpG islands (Yen et al, 1984; Lock et al, 1986). According to this model, large parts of the eukaryotic genomes, which are transcriptionally silent, will have highly methylated CpG islands (Antequera et al, 1989) whereas the CpG islands near transcriptionally active genes, particularly "house-keeping genes", will
be non-methylated and exposed to relatively non-specific transcriptional activator proteins (Bird, 1986; Holler et al, 1988). The genes expressed in a tissue-specific manner will need additional and temporal combinations of transcriptional activator proteins for their expression (Kolsto et al, 1986).

1.3.1.3 Methylation as a long term regulator.

DNA methylation is stable through generations of cells (Stein et al, 1982). When a plasmid or a cloned gene is methylated and introduced into cultured cells, the modification is faithfully inherited through more than 50 doubling generations (Wigler et al, 1981; Jaenish et al, 1983). Methylation is maintained during cell division by endogenous methylases (Adams, 1971; Woodcock et al, 1982).

DNA methylation of some tissue-specific genes is also very stable (Groudine & Conkin, 1985; Bird, 1986). Some tissue-specific genes are almost fully methylated in sperm and probably also in the female germ cells and remain methylated throughout early development and in all non-expressing adult somatic tissues. Transcription occurs when the gene is demethylated in a tissue specific manner (Groudine & Conkin, 1985; Swain et al, 1987; reviewed by Bird, 1986; Cedar, 1988). For instance, the IgG κ gene which is highly methylated and inactive in pre-B cells, can be transcriptionally induced by treatment with
lipopolysaccharide. During the induced differentiation the IgG κ sequences undergo demethylation and then are maintained in the hypomethylated condition after the treatment (Kelly et al, 1988). The demethylated IgG κ gene can be transcribed in cultured cells which do not have the phenotypic characteristics of B cells. This indicates that the initial methylation of the gene has been used to shut off the transcription completely in a developmental stage-specific manner. It has been speculated that groups of genes, such as those that encode the liver specific proteins, follow a similar pattern of methylation and demethylation, and expression, during development (Wilks et al, 1982; Vedel et al, 1983; Benvenisty et al, 1985).

DNA methylation may be involved in parental imprinting (Groudine & Conkin, 1985). An example of this category of genes is the X-chromosome-linked genes (Chapman et al, 1982; reviewed by Gartler & Riggs, 1983). In mammals, females and males express equivalent levels of most X-linked gene products despite the fact that females have twice the number of X-linked genes per diploid as do males. Studies of the expression of X-chromosome-linked genes has provided indirect evidence that DNA methylation plays a role in the maintenance of inactivation of one of the X chromosomes in cells of a female (Lock et al, 1988). The murine hypoxanthine phosphoribosyl-transferase (hprt) gene on an inactive X chromosome is inefficient in transforming hprt-deficient cultured cells to a hprt-positive phenotype, whereas the hprt allele
carried on an active X chromosome can transform these cells efficiently (Liskay & Evans, 1980; Venolia & Gartler, 1983). The DNA isolated from cells treated with the methylase inhibitor, 5-azacytidine, can result in efficient transformation in the type of experiment described in the previous sentence (Lester et al., 1982; Venolia et al., 1982). Furthermore, the DNA methylation pattern of a particular region in the 5' flanking region of the murine hprt gene is correlated with whether the chromosome is active or inactive (Lock et al., 1987). Thus, DNA methylation can be a long term and sequence-non-specific repressor working on sizes of DNA ranging from chromatin domains to chromosomes.

1.3.2 Chromatin structure and gene expression.

Eukaryotic DNA is folded up around a protein core made by histones and makes up a structural unit called a "nucleosome" (Olins & Olins, 1974; McGhee & Felsenfeld, 1980). The bulk of eukaryotic nuclear DNA is thus packaged into a higher order structure which is like beads on a string when extended in a hypotonic solution (Gasser & Laemmli, 1987). The histones supporting this nucleosome interact with DNA by hydrogen bonds and can decrease the transcription of a gene, to a greater or lesser extent, by limiting the access of large proteins to the DNA (McGhee et al., 1981). Thus some of the sequence-specific DNA binding proteins mentioned in previous sections have to negotiate
with histones before they can interact with desirable DNA signals.

The hypothesis that a signal used by a DNA-binding protein needs to be in an exposed area of folded DNA has been experimentally examined in various genes with DNase I, which degrades DNA if it is not protected by proteins (reviewed by Weintraub, 1983; Eissenberg et al, 1985; Gross & Garrard, 1988). For example, SV40 has nucleosome-free regions detected with DNase I and micrococcal nuclease. The SV40 enhancer element for the early transcriptional unit and its flanking region is sensitive to DNase I-digestion (Jongstra et al, 1984). The DNase I sensitivity of the region disappears if the enhancer element of the SV40 is deleted, indicating that the chromatin structure sensitive to the DNase I can be sequence-specific (Gerard et al, 1985; McKnight & Tjian, 1986). These experiments have shown that both multiple sequence motifs and protein factors contribute to make up a specific chromatin structure, which can be used to identify a functional area of the chromatin. It has been speculated that transcriptionally active genes and their flanking region have a particular chromatin structure (Pederson et al, 1986). Although the fundamental characteristics of chromatin structure recognized by DNase I are not well known, DNase I hypersensitive sites of about 70 genes have been described(Gross & Gerrad, 1988).
Some nuclease hypersensitive sites are inducible by trans-acting protein factors. The promoter region of MMTV (mouse mammary tumour virus), for instance, has a stretch of DNA designated LTR (long terminal repeat), which contains a glucocorticoid-responsive element (GRE). This element stimulates transcriptional initiation within a minute after exposure of cells to hormone-containing medium (Chandler et al, 1984). Changes that occur prior to the induction of transcription include sequence-specific organization of nucleosomes in the LTR region (Richard-Foy & Hager, 1987), generation of endonuclease sensitive sites and probably displacement of a single nucleosome from the GRE region. This structural change is accompanied by the binding of two nuclear protein factors (Cordingley et al, 1987). The generated nuclease sensitive site disappears with removal of the hormone (Zaret & Yamamoto, 1984) and the process is blocked by antagonists which bind covalently to the hormone receptor. Therefore, the entire process of generation of the nuclease sensitive structure and transcriptional activation is thought to be controlled by multiple trans-acting protein factors. As the GRE is recognized by the receptor-hormone complex when the sequence is in the non-stimulated state, the necessary region is thought to be a specific configuration of the chromatin (Richard-Foy & Hager, 1987).

Besides the DNA signals and trans-acting protein factors, DNA itself has the potential to modulate the template activity for RNA
polymerases by changing its three dimensional configuration or helicity (Villeponteau et al, 1984; Weintraub, 1983). Non-B form DNA and unpaired sequences are preferentially digested by many nucleases, including Bal31, mung bean, S1 and *Neurospora crassa* nucleases (Lassen & Weintraub, 1982; Kohwi-Shigematsu et al, 1983; Han et al, 1984). These DNA conformations can be induced locally by changes in torsional stress (Ptashne, 1986), binding of protein (Shuey & Parker, 1986) or by a particular nucleotide sequence tract. DNA conformations which do not contribute to a stable interaction with the histone core disfavour formation of the nucleosome structure and thus tend to make exposed area of DNA (Nickol et al, 1982; Luchnick et al, 1985; Pederson et al, 1986).

A characteristic of eukaryotic transcription is that it has various levels of regulation. A functional RNA is often encoded by a number of DNA segments which are separated spatially from each other in the genome, which is supported by histones and wound into a DNA-protein complex. Transcriptional machinery approaches the region preceding a gene and the transcripts have to be modified into a shape which is stable and translatable in a cell. The system is complex, however, it provides a variety of protein products and flexibility of temporal and cell-specific regulation.
1.4 A search for regulatory mechanisms of \textit{c-mos} expression.

Which of the various mechanisms discussed above is applicable to \textit{c-mos}? One of the first ideas proposed to explain the low expression of \textit{c-mos} was that its flanking region contained an inhibitory element (a silencer?) but not an activator. There were two lines of supportive observations for this hypothesis. The first was that \textit{c-mos} only transforms cultured fibroblasts when it is linked with viral enhancers rather than its own flanking region (see Section 1.1.3). Secondly, the flanking region of the murine \textit{c-mos} contains a segment that can inhibit the expression of genes linked to it. This inhibitory segment of the flanking region is located about 1.8 kb upstream of the putative \textit{c-mos} coding region and is designated as the UMS (upstream \textit{mos} sequence) (Wood et al, 1984; McGeady et al, 1986). In transfection assays, the UMS can inhibit the expression of either of a thymidine kinase gene or the \textit{c-mos} gene when it is inserted between the viral LTR and downstream coding regions (Wood et al, 1984; McGeady et al, 1986). The UMS may work as a termination signal for transcription as it is only inhibitory when located upstream of a gene (Wood et al, 1984; McGeady et al, 1986).

The rat \textit{c-mos} 5' flanking region has also been analyzed in a similar manner and a portion of it, the rat inhibitory sequence (RIS), was found to have the potential to prevent the accumulation of transcripts
generated by a downstream gene (van der Hoorn et al, 1985; van der Hoorn & Neupert, 1986). The RIS element has several similar characteristics to the murine UMS; it is located 1.8 kb upstream of the putative c-mos coding region; its nucleotide sequence is highly homologous to a part of the UMS and it contains a repeat of the polyadenylation signal (van der Hoorn et al, 1985; van der Hoorn & Neupert, 1986). However, transcripts which start upstream of the RIS do not terminate in it, indicating that unlike the murine UMS the RIS does not repress transcription by acting as a termination signal (van der Hoorn & Neupert, 1986). Putative protein-binding sequences have recently been suggested to occur in a region close to the RIS (van der Hoom, 1987; van der Hoom & Nordeen, 1989) but their function is unexplored.

The RIS sequence may, or may not, account for the limited expression of c-mos. Given the limited knowledge about the structure of c-mos it is important to consider other putative regulatory influences. In particular, factors that work in a wider context (Section 1.3) have received little attention and two such mechanisms are examined in this thesis. In Chapter 4, the hypothesis that c-mos expression is restricted by the presence of a neighbouring transcriptional unit has been examined. The search for this putative transcriptional unit lead to the discovery of a novel combination of two repetitive sequences in the c-mos 3’ flanking region.
In Chapter 5, the methylation status of the \textit{c-mos} region has been studied in three tissues. DNA methylation is generally inhibitory to transcription (Section 1.3.1) and there are some indications that methylation may affect \textit{c-mos} expression. Genomic DNAs prepared from cultured rodent cell lines and murine tissues, which do not express \textit{c-mos}, are highly methylated (Gattoni-Celli et al, 1982) and when the \textit{Hpa} II sites in \textit{v-mos} are methylated \textit{in vitro} the ability of the \textit{v-mos} to transform cultured murine cells is reduced (McGeady et al, 1983). Both these observations are suggestive that methylation of \textit{c-mos} may repress its expression, however, none of the materials examined express \textit{c-mos}. In order to assess the possibility that DNA methylation in the \textit{c-mos} region may contribute to it transcriptional regulation, this question has been re-examined (Chapter 5) using tissues in which \textit{c-mos} transcription had been observed.
CHAPTER TWO

MATERIALS AND METHODS.
2. 1 Source of materials.

2. 1. 1 Animals.

Adult Donryu rats (10-15 weeks old) were maintained in standard conditions and fed *ad libitum* until they were killed while under ether-anaesthesia. Various tissues were removed by dissection, frozen in a liquid nitrogen bath and stored at -95°C until used.

2. 1. 2 DNA clones.

A λ phage clone, λD3e, containing the rat *c-mos* with its 5' and 3' flanking regions was generously provided by Dr van der Hoorn. A recombinant plasmid pMS2 which harbours the murine *c-mos* and its 3' flanking region was kindly provided by Dr Vande Woude, Bionetics Research Inc., USA.

2. 2 DNA preparations.

2. 2. 1 Growth of bacteriophage and preparation of its DNA.

The procedure used to prepare bacteriophage DNA was a modification of the method described in Molecular Cloning (Maniatis et
al, 1982). The modification was that the caesium chloride centrifugation step was omitted and phage DNA was directly recovered from the pelleted phage particles.

Liquid medium and reagents:
NZCYM medium: 10 g of NZ amine A (casein enzymatic hydrolysates, Sigma), 5 g of NaCl, 5 g of Bacto-yeast extract, 1 g of casamino acids and 2 g of MgSO4-7H2O were dissolved in 1 litre of d.H2O and autoclaved.
SM solution: 5.8 g of NaCl, 2 g of MgSO4-7H2O, 50 ml of 1 M Tris-(hydroxymethyl)aminomethane hydrochloride (Tris-HCl) (pH 7.5) and 5 ml of 2% gelatin were dissolved in d.H2O, made up to 1 litre and autoclaved.

Exponentially growing *Escherichia coli* (*E. coli*) cells of a strain LE392 were harvested by centrifugation in a Sorvall SS34 rotor at 6,000 r.p.m. for 5 min at 4°C. The cells were suspended in 10 mM MgSO4 at a density of 2 x 10^8 cells/ml (OD660 -2.0) to prepare competent cells. A lysate stock of λD3e (8 x 10^7 plaque forming unit (p.f.u.)/ml) was added to the suspension and incubated at 37°C for 20 min. The infected cells were inoculated with 100 ml of pre-warmed medium and grown overnight at 37°C with shaking. The cultured cells spontaneously lysed during the incubation and the lysis was completed by addition of 1 ml of
chloroform. Cell debris was pelleted by centrifugation in a Sorvall GSA rotor at 7,000 r.p.m. for 10 min at 4°C. The supernatant was collected and treated with DNase I and RNase A at concentrations, 1 μg/ml and 5 μg/ml, respectively to digest the nucleic acids liberated from the lysed bacteria. Solid sodium chloride and a solution of 50 w/v% polyethylene glycol (PEG 6000) were added to give final concentrations of 1 M and 10 w/v%, respectively. The suspension was cooled on ice for 2 hours and the bacteriophage particles were recovered by centrifugation in a Sorvall GSA rotor at 8,500 r.p.m. for 20 min at 4°C. The pelleted phage particles were suspended in SM solution and lysed by adding 20 w/v% sodium dodecyl sulfate (SDS) and 0.5 M ethylenediamine tetra-acetic acid (EDTA)(pH 8.0). The lysate was extracted twice with a mixture of aqueous-saturated phenol and chloroform (1:1). Nucleic acids in the aqueous phase were precipitated in 0.3 M sodium acetate (pH 5.2) and 70% ethanol (EtOH). The pellet recovered by centrifugation was washed with 70% EtOH twice, briefly dried and finally dissolved in an appropriate volume of d.H2O. Cellular RNA present in this fraction was degraded by incubation with 1 μg/ml of RNase A at 37°C for 1 h. DNA was recovered by ethanol precipitation and centrifugation after extraction with a mixture of phenol and chloroform.
2.2.2 Growth of bacteria and preparation of plasmid DNAs.

All the plasmid DNAs were prepared by the alkaline-lysis method described by Birnboim & Doly (1979), except that EDTA and acetic acid was used instead of CDTA and formic acid. The plasmid DNAs prepared by this procedure were used to isolate DNA fragments to use them for hybridization probes or for subcloning.

Liquid medium and reagents:
Luria-Bertani (LB) medium: 10 g of Bacto-tryptone, 5 g of Bacto-yeast extract and 10 g of NaCl were dissolved in 1 litre of d.H2O and autoclaved.
1 x STE: A solution containing 0.1 M NaCl, 10 mM Tris-HCl (pH 8.0) and 1 mM EDTA.
1 x GTE: 50 mM glucose, 25 mM Tris-HCl (pH 8.0) and 10 mM EDTA.
1 x TE (pH 8.0): 10 mM Tris-HCl (pH 8.0) and 1 mM EDTA.

*E. coli* transformants containing desirable plasmids were streaked out on selective agar plates containing an appropriate antibiotic and incubated overnight at 37°C. Single colonies were inoculated and grown overnight in 20-50 ml of LB media containing the same antibiotic as used in the plates. The confluent cultures were subsequently added to 500 ml LB media containing the same antibiotic in 2 litre flasks and incubated
overnight at 37°C. Amplification of the plasmids, by addition of chloramphenicol to the culture media, was used only for the recombinants cloned into pBR322 and was performed exactly as described by Maniatis et al (1982). The cells grown were collected by centrifugation in a Sorvall GSA rotor at 6,000 r.p.m. for 10 min at 4°C and washed once with 1/5 of the volume of the culture media of 1 x STE. The cells obtained were resuspended in 10 ml of 1 x GTE containing 5 mg/ml of lysozyme and incubated at room temperature for 10 min. Twenty ml of a solution containing 0.2 M NaOH and 1 % SDS was added to the cell lysates and incubated on ice for 10 min. The host cell debris was aggregated by addition of 10 ml of a solution containing 3 M potassium acetate and 2 M acetic acid and pelleted by centrifugation in a Sorvall SS34 rotor at 10,000 r.p.m. for 10 min at 4°C. The supernatant was collected and 0.6 volumes of iso-propanol was added to precipitate the DNAs and RNAs present. The precipitate was pelleted by centrifugation in a Sorvall HB4 rotor at 10,000 r.p.m. for 15 min at 4°C. The pellets were washed once with 70 % EtOH, briefly dried and dissolved in an appropriate volume of 1 x TE (pH 8.0). One gram of caesium chloride (CsCl) was added per ml of the solutions and thoroughly dissolved. An aqueous solution of ethidium bromide (10 mg/ml) was added to give a final concentration of 600 µg/ml. Plasmid DNAs were separated from cellular RNA and DNA by CsCl density gradient centrifugation in a Beckmann VTi80 rotor at 60,000 r.p.m. for 16 hours
at 20°C. When highly purified plasmid DNAs were needed, the CsCl density gradient centrifugation was performed twice.

The plasmid DNAs were removed from centrifuge tubes by syringes and collected into glass-top tubes. Ethidium bromide in the solution was extracted with water-saturated n-butanol and then the aqueous solutions were diluted with 4 volumes of 1x TE (pH 8.0). The DNAs were precipitated by addition of cold ethanol to give a final concentration of 70% and collected by centrifugation in a Sorvall HB4 rotor at 10,000 r.p.m. for 15 min at 4°C. The pellets were washed twice with 70% EtOH, dried and finally dissolved in d.H2O. The amount of DNA was estimated by an OD260 measurement of the solution and a standard conversion of OD260 of 1.0 equivalent to 50 μg/ml DNA.

2. 2. 3 Mini-scale preparations of plasmid DNA.

The method described in the previous section was used in a small scale when plasmid DNAs were prepared to check the inserted fragment by a restriction endonuclease cleavage site mapping or when they were used for DNA sequencing.

The volume of bacterial cultures used was 2.5 ml and the cells collected from the confluent cultures were suspended in 0.2 ml of 1 x
GTE. The other solutions were added in the same ratio described in the previous section but scaled to the volume of this suspension. All the centrifugations were performed using an Eppendorf centrifuge. Caesium chloride gradient centrifugation was omitted and RNase was used to remove RNA derived from the host cells. RNase A, which was initially boiled for 1 hour to inactivate contaminated DNase activity, was added to the supernatants separated from the host cell debris to give a final concentration of 12 μg/ml and the mixtures were incubated for 30 min at 37°C. The reaction mixtures were then extracted with a mixture of phenol-chloroform and the plasmid DNAs, precipitated by addition of isopropanol and collected by a centrifuge. The recovered DNA pellets were washed twice with 70% ethanol, dried and dissolved in an appropriate volume of d.H2O.

2. 2. 4 Transformation of *E. coli*.

DNA fragments isolated from original clones were subcloned into various plasmid vectors depending on the use required. The procedure was according to the method described by Messing (1983).

Medium and reagents:

2 x YT liquid medium: 16 g of Bacto-tryptone, 10 g of Bacto-yeast extract and 5 g of NaCl were added to 1 litre of d.H2O and autoclaved.
0.1 M IPTG: Approximately 24 mg of isopropyl-β-D-thiogalactopyranoside (Sigma) were dissolved in 1 ml of autoclaved dH2O.

2% X-Gal: Approximately 20 mg of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (Sigma) was dissolved in 1 ml of dimethyl-sulfoxide.

DNAs prepared from phage or plasmids which contained DNA fragments of interest were cleaved with restriction endonucleases to release the DNA fragments from the vector. The reaction conditions of each restriction enzyme were adjusted as suggested by the manufacturers. The DNA fragments were fractionated on agarose gels containing 0.5 μg/ml of ethidium bromide, chosen by their sizes and electroeluted from the gel. The ethidium bromide in the elutates was removed by an extraction with a mixture of phenol and chloroform. The isolated DNA fragments were precipitated in 70% ethanol and 0.3 M sodium acetate, washed with 70% ethanol, dried and dissolved in dH2O.

Ligation of a DNA fragment with an appropriate vector was performed with reaction conditions described in Molecular Cloning (Maniatis et al, 1982). Typically, 100-200 ng of vector DNA was mixed with a 3 molar excess of a fragment and incubated at 8°C overnight.

Appropriate strains of E. coli (JM 107 or DH5α) were grown in 2 x YT or LB media to prepare competent cells. When the cells had grown to
an exponential phase (OD660 = 0.4), they were harvested by centrifugation and suspended in a half the volume of the culture media of an ice-cold solution of 50 mM CaCl2. The cells were incubated on ice for 20 min, collected by centrifugation and resuspended in 1/10 volume of the culture media of 50 mM CaCl2. Part of ligation mixtures containing approximately 10-50 ng of plasmid DNAs was mixed with 150 µl of the competent cell suspension and incubated on ice for 30-40 min. The cell suspensions were incubated at 42°C for 2 min to give a heat shock. The heat-treated mixtures were rapidly chilled on ice, centrifuged to pellet the cells and then suspended in 200 µl of LB medium and incubated for 1 hour. The suspensions were spread with 20 µl of 0.1 M IPTG and 40 µl of 2% X-Gal on LB agar plates containing an appropriate antibiotic. Transformed cells grew as white colonies on the plates after incubation for 16-20 hours at 37°C. The transformation efficiency of this procedure was 5-10 x 10^5 colonies/µg DNA.

### 2. 2. 5 Genomic DNA preparation.

Genomic DNAs were prepared from frozen tissues according to the procedure described in Molecular Cloning (Maniatis et al, 1982).

Reagents:

Lysis solution: 0.5 M EDTA (pH 8.0), 100 µg/ml of proteinase K and
0.5% N-lauroylsarcosine (sodium salt).

Frozen tissues were poured with liquid nitrogen into a stainless-steel cup attached with a blender and ground to a fine powder. Excess liquid nitrogen was evaporated and lysis solution of approximately 10 times the volume of the tissue was added. The suspensions were homogenized in the lysis solution by a hand-driven teflon-potter homogenizer and incubated at 50°C for 3 hours to degrade proteins. The reaction mixtures were extracted three times with an equal volume of phenol and dialysed against 4 litre of a solution containing 50 mM Tris-HCl (pH 8.0), 10 mM EDTA and 10 mM NaCl. The dialysed solutions were incubated with heat-treated RNase A at a concentration of 100 µg/ml for 3 hours at 37°C to remove RNA and extracted twice with a mixture of phenol and chloroform (1:1). DNAs in the aqueous solution were recovered by precipitation by addition of 2 times of the volume of ethanol and centrifugation.

2.3 RNA preparation from tissues.

2.3.1 Isolation of total RNAs from frozen tissues.

RNAs were prepared from frozen tissues by the guanidinium-caesium chloride method described by Chirgwin et al (1979).
Reagents:
Distilled water used for the RNA preparation was treated with diethylpyrocarbonate (DEPC) at a concentration of 0.1% and autoclaved.

5 M guanidinium isothiocyanate solution: 60 g of guanidinium isothiocyanate (Fluka), 2 ml of 0.5 M EDTA, 1 ml of 1 M Tris-HCl (pH 7.0) and 5 ml of 2-mercaptoethanol were dissolved in d.H2O and made up to 100 ml.

20% N-lauroylsarcosine (sodium salt) (20% Sarkosyl solution): 10 g of N-lauroylsarcosine (sodium salt) (Sigma) was dissolved in d.H2O and incubated at 65°C for 1 hour.

5.7 M CsCl in 1 x TE (pH 7.4): 48 g of CsCl was dissolved in a solution containing 10 mM Tris-HCl and 1 mM EDTA (pH 7.4) (1 x TE (pH 7.4))

Two gram of the frozen tissue were homogenized in a Sorvall stainless-steel homogenizer with 15 ml of 5.7 M guanidinium isothiocyanate solution at maximum speed for 2 min. The homogenates were transferred to Corex glass centrifuge tubes and 20% Sarkosyl solution was added to a final concentration of 1%. The mixtures were stirred for 5 min at room temperature and 8.4 g of solid CsCl was added and then stirred for another 30 min. The solutions were centrifuged to remove insoluble components in a Sorvall HB4 rotor at 12,000 r.p.m. for 15 min at 4°C. RNAs in the supernatant were pelleted through 3.5 ml of 5.7 M CsCl by centrifugation in a Beckmann SW41 rotor at 28,000 r.p.m for 24
hours at 20°C. The supernatants were discarded and the remaining RNA pellets were thoroughly dissolved in a solution containing 10 mM Tris-HCl, 1 mM EDTA (pH 7.5) and 0.1% Sarkosyl and extracted with an equal volume of phenol. The RNAs in the aqueous solution were precipitated by addition of 5 M NaCl and EtOH to give final concentrations of 0.2 M and 75%, respectively. The RNAs were recovered by centrifugation in a Sorvall HB4 rotor at 10,000 r.p.m. for 20 min at 4°C, washed twice with 70% EtOH, briefly dried and dissolved in an appropriate volume of 10 mM Tris-HCl (pH 7.0). The amounts of RNA in the solutions were estimated by OD260 measurements, and standard conversion of 1 OD260 equivalent to 33 μg/ml.

2.3.2 Preparation of poly (A)+ RNA enriched fractions from total RNAs.

Poly (A)+ RNA enriched fractions were isolated from total RNA fractions by oligo-(dT) cellulose column chromatography (Aviv & Leder, 1972).

Elution buffers:

Loading buffer; 20 mM Tris-HCl (pH 7.6), 0.5 M NaCl, 1 mM EDTA and 0.1% SDS.

Low-salt buffer; 20 mM Tris-HCl (pH 7.6), 0.1 M NaCl, 1 mM EDTA and 0.1% SDS.
Non-salt buffer; 10 mM Tris-HCl (pH 7.6), 1 mM EDTA and 0.05 % SDS.

Oligo-(dT)-cellulose (Pharmacia) was washed sequentially with d.H₂O, a solution of 0.1 M NaOH and 5 mM EDTA, and d.H₂O. The last d.H₂O washing was continued until the pH of the eluate was lower than 8.0. The cellulose column was then equilibrated with loading buffer. Total RNAs dissolved in 0.5 ml of d.H₂O were incubated at 65°C for 15 min and 0.5 ml of a solution of 2-times concentration of the loading buffer was added to adjust the salt-buffer conditions. One ml of the first eluate was collected, heated and then applied to the column again. The loaded cellulose was washed with 10 ml of the loading buffer and subsequently washed with 4 ml of the low-salt buffer and 10 ml of the non-salt buffer. The eluates from the three different kinds of solutions were collected separately. The non-salt buffer eluates were the poly(A)⁺ enriched fractions. The amount of RNA in each fraction was estimated by a measurement of OD₂₆₀. Typically, when 1 mg of total RNA was applied, 2-10 w/w% of the loaded RNA was eluted by the non-salt buffer. The fractionated RNAs were precipitated in 0.3 M sodium acetate (pH 5.2) and 70% EtOH, recovered by centrifugation and dissolved in d.H₂O.
2. 4 Electrophoresis of DNA or RNA fragments on agarose gels.

DNA fragments were fractionated by their sizes on agarose gels by a method originally described by Loening (1968). The chemical composition of the buffer used was 40 mM Tris-acetate and 2 mM EDTA (pH 7.8). The concentration of agarose used in the gels ranged from 0.7 to 1.2% according to the sizes of the DNA fragments. When ethidium bromide was used to visualize DNA fragments, it was added to both reservoir buffer and gels at a concentration of 0.5 μg/ml. The electrophoreoses were performed with various voltages ranged from 2 to 50 V/cm according to the sizes of equipment used and the DNA fragments to be observed.

Electrophoresis of RNAs was performed using the formaldehyde-3-[N-morpholino]propanesulphonic acid (MOPS) buffer system described by Goldberg (1980). The gels contained 2.2 M formaldehyde and 40 mM MOPS (pH 7.0), and the running buffer contained 40 mM MOPS. The RNA samples were heated in a sample buffer containing formaldehyde and formamide at 55°C for 15 min just before their loading. The electrophoresis was performed at 6-10 V/cm in a cold room at 4°C to avoid uneven-heating of the gels by current.
2. 5 Preparation of radioactively-labelled probes for hybridizations.

2. 5. 1 Randomly-labelled DNA probes.

Radioactively-labelled DNA fragments were synthesized by one of two procedures described below.

The first method used isolated DNA fragments as templates. The DNA probes were prepared by a random priming method described by Feinberg and Vogelstein (1983). The DNA fragments were isolated by electroelution from agarose gels after restriction endonuclease cleavage of the plasmid DNAs purified by CsCl-density gradient centrifugation. The isolated fragments were heat-denatured and the radioactive DNA strands were synthesized with the large fragment of *E. coli* DNA polymerase I (Klenow fragment) and [*α-^32^P]*dATP, using deoxyribonucleotide hexamers generated by DNase degradation of calf thymus DNA (random primers) as primers. Normally, 50-100 ng of DNA fragments were used and 70-90% of [*α-^32^P]*dATP (3,000 Ci/mmol, Amersham) in reaction mixtures was incorporated to obtain a specific activity of 1-2 x 10^8 c.p.m./μg DNA. The labelled DNA fragments were separated from unincorporated [*α-^32^P]*dATP by Sephadex G-50 column chromatography.
The second method used poly(A)+ RNA fractions as the templates. cDNA probes were prepared from poly(A)+ RNA with AMV reverse transcriptase according to a method described in Molecular Cloning (Maniatis et al, 1982). The random primers were used instead of oligo-d(T) nucleotide to avoid synthesizing a population of labelled-DNA fragments predominantly from 3' portions of the template RNAs. After the syntheses of cDNA, the template RNAs were degraded by addition of an alkaline solution and the probes were separated from unincorporated [α-32P]dATP by Sephadex G-50 column chromatography.

2. 5. 2 End-labelling of DNA fragments.

DNA fragments radioactively labelled at single ends were used as probes for S1 nuclease protection assays. Dependent on the labelled position on the DNA fragments, one of the two procedures described below was chosen.

5' end-labelled probes were prepared by phosphorylation with T4 polynucleotide kinase (Maniatis et al, 1982; Arrand, 1985). The linearized plasmid DNAs were dephosphorylated by alkaline phosphatase. The DNAs were extracted twice with a mixture of phenol and chloroform, once with chloroform and precipitated in 70% EtOH. The DNAs were recovered by centrifugation, washed twice with 70%
EtOH, dried and dissolved in 20 μl of a solution containing 15 mM dithiothreitol (DTT), 10 mM MgCl₂, 60 mM Tris-HCl (pH 7.8), 15 mM spermidine tetrahydrochloride and 10-20 μCi of [γ-³²P]ATP (6,000 Ci/mmol). Two to five units of the T4 polynucleotide kinase were added to the reaction mixtures and incubated at 37°C for 60 min. The solutions were extracted twice with a mixture of phenol and chloroform, and once with chloroform. The labelled DNA fragments were cleaved with restriction endonucleases to obtain parts of the fragments which had been labelled at only one end. The fragments were separated by electrophoresis on agarose gels and isolated by electroelution.

"Filling-in" reaction with Klenow fragment was used for radioactive labelling of DNA fragments at their 3' ends (Maniatis et al, 1982). Plasmid DNAs were linearized by restriction endonucleases which produce 3' recessive ends. The 3' ends were labelled with a complementary strand synthesis with the Klenow fragment using either or both of [α-³²P]dATP and [α-³²P]dCTP.

2. 5. 3 Single-stranded RNA probes.

Radioactively-labelled RNAs were prepared with an in vitro transcription system described by Melton et al (1984b). This procedure was used to make orientation-specific probes for Northern blot
hybridizations.

Reagents:

Transcription buffer: 40 mM Tris-HCl (pH 7.5), 6 mM MgCl₂, 2 mM spermidine hydrochloride, 10 mM DTT, 1 unit/ml of RNAsin, 100 μg/ml of bovine serum albumin (BSA) and 500 μM each of ATP, CTP and UTP.

DNA fragments of interest were cloned into a Bluescribe vector (Integrated Sciences) which had a T3 or T7 RNA polymerase promoter sequence at ends of the polylinker integration site. The recombinant plasmid DNAs were linearized by cleavage with restriction endonucleases. 0.5-1 μg of the linearized DNAs was incubated with [α-³²P]GTP and 15 units of T3 or T7 RNA polymerase in the transcription buffer at 40° C for 1 hour. The template DNAs were then degraded by an RNase-free DNase (Promega) and the synthesized RNAs were purified by an elution through Sephadex G-50 or by an EtOH precipitation. [α-³²P]GTP was used to minimize premature termination of the transcription products (Melton et al, 1984b).

2. 6 Hybridization conditions.

Two kinds of nylon membrane, Zeta-probe (Bio-Rad) and Hybond-N (Amersham), were used selectively to achieve the transfer of
DNA or RNA within a short time and to obtain a low background of the hybridization signals on the membrane. DNAs or RNAs were transferred from electrophoresed agarose gels to membranes according to the instruction manual of the membrane manufacturer. Various concentrations of alkaline solutions were used to transfer the polynucleotide fragments to the Zeta-probe, and 20 x SSC solution was used for the transfer to the Hybond-N. The Zeta-probe was used for almost all Southern transfers and hybridizations as it had a high capacity for binding nucleotides and completing the transfer within a short time. The Hybond-N was preferentially used for Northern-blot hybridizations because it gave a low background after a long time exposure.

Reagents:

1 x SSC: 0.15 M NaCl and 15 mM sodium citrate (pH 7.0), therefore, 20 x SSC is 3 M NaCl and 0.3 M sodium citrate (pH 7.0).

1 x phosphate-EDTA solution (pH 6.9)(1 x PE): 1 x PE contains 0.13 M sodium phosphate (pH 6.9) and 1 mM EDTA. 2 x PE contains 0.26 M sodium phosphate (pH 6.9) and 2 mM EDTA.

10 w/v% BLOTTO: 10 w/v% suspension of non-fat milk powder. The BLOTTO is a collective name for commercially available non-fat instant milk powder (Reed & Mann, 1985)

50 w/v% polyethylene glycol 20,000 (PEG 20,000)

10 mg/ml carrier DNA: An approximate amount of salmon sperm DNA sodium salt (Sigma, type III) was dissolved in d.H2O, sheared by
sonication and the concentration was adjusted to 10 mg/ml. The DNA was denatured by boiling in a water bath and rapid cooling in an ice-water bath just before its addition to a hybridization solution.

The compositions of the hybridization solutions, the incubation temperature and the washing conditions for each membrane were changed according to the combination of the type of probe and the sample as follows.

**DNA probes to DNAs immobilized on a membrane.**

Hybridization solution: 2 x PE, 7% SDS, 0.5 w/v% BLOTTO, 1 w/v% PEG 20,000 and 0.5 mg/ml carrier DNA.

The hybridization was performed at 65°C with shaking in a water incubator. The membrane was rinsed in 2 x SSC and washed with a solution of 2 x SSC and 0.1% SDS at room temperature for 15 min, subsequently with a solution containing 0.5 x SSC and 1% SDS at 65°C for 15 min. The membrane was rinsed with a solution of 0.5 x SSC and 0.1% SDS and exposed to an X-ray film at -95°C for an appropriate length of time.
DNA probes to RNAs immobilized on a membrane.

Hybridization solution: 50 v/v% deionized formamide, 2 X PE, 7% SDS,
0.5 w/v% BLOTTO, 1 w/v% PEG 20,000 and 0.5 mg/ml carrier DNA.

The incubation was performed at 50°C. The second wash of the membrane was performed in 0.2 x SSC and 1% SDS at 65°C and the final rinse was done in 0.2 x SSC and 0.1% SDS.

RNA probes to RNAs immobilized on a membrane.

Hybridization solution: 60 v/v% deionized formamide, 2 x PE, 7% SDS,
0.5 w/v% BLOTTO, 1 w/v% PEG 20,000 and 0.5 mg/ml carrier DNA.

The incubation temperature was 55°C and the conditions for rinsing and washing were as described above.

2. 7 S1 nuclease protection assays.

5′- or 3′-end-labelled DNA probes were prepared as described in Section 2. 5. 2. The radioactively-labelled DNAs were co-precipitated
with RNA samples and dissolved in 20 μl of hybridization buffer. The mixtures were incubated at 90°C for 10 min, immediately transferred to a water bath set at 52°C and incubated overnight. Two hundred microlitres of the S1 nuclease digestion buffer were added to each tube and the mixtures were subsequently incubated at 37°C for 30 min. A large amount of RNAs (50~70 μg), a high concentration of S1 nuclease (5,000 units/ml) and a long incubation time of an X-ray film (10-20 days) were used as the Northern blot hybridizations have shown a low level of transcription even in the testis. The digestion was terminated by addition of 20 μl of S1 stop buffer and 2.5 times of the volume of the reaction mixture of cold EtOH were added to precipitate polynucleotides present. The precipitates were pelleted by centrifugation, washed with 70% EtOH and dissolved in 5 μl of d.H2O, and 5 μl of the formamide-dye mixture was added. The mixtures were boiled for 4 min and loaded onto 5% polyacrylamide-7 M urea denaturing gels. The electrophoresed gels were fixed in a solution of 10% methanol-10% acetic acid, rinsed with water, washed with 10% EtOH in water and dried overnight. The dried gels were exposed with X-ray films for an appropriate length of time.

Reagents:
Hybridization buffer: 80% recrystallized formamide, 80 mM piperazine-N,N'-bis[2-ethanesulphonic acid] (PIPES) (pH 6.8), 0.4 M NaCl and 10 mM EDTA.
S1 nuclease digestion buffer: 30 mM sodium acetate (pH 4.5), 0.28 M NaCl, 1 mM ZnSO4 and 5,000 unit/ml of S1 nuclease (Boehringer-Mannheim).

S1 stop buffer: 500 mM Tris-HCl (pH 9.0), 100 mM EDTA and 50 mg/ml of tRNA.

Formamide-dye mixture: A solution containing 98% recrystallized formamide and 0.1 w/w% each of bromophenol blue and xylene cyanol.

2. 8 Determination of DNA sequences.

The nucleotide sequences of DNA were determined by a dideoxy-nucleotide-chain termination method initially described by Sanger et al (1977) using double-stranded plasmid DNA (dsDNA) as templates. To generate DNA templates with readable sizes, subcloning and unidirectional digestion with exonuclease III of the previously cloned plasmid DNAs were performed as described by Henikoff (1987). The plasmid DNAs were prepared by the miniscale-preparation method described in Section 2.2.3. They were denatured in a solution of 0.1 N NaOH at room temperature and hybridized with appropriate oligonucleotides. The DNA synthesis by Klenow fragment was performed at 42°C as described by Hattori and Sakaki (1986). The composition of the reaction buffer used with the Klenow fragment was the same as described by Messing (1983). When T7 DNA polymerase was
used for the DNA synthesis, its reaction was performed according to a protocol provided by the manufacturer. The DNAs synthesized with [α-\textsuperscript{32}P] dATP were analyzed by electrophoresis on 5 or 6% polyacrylamide-7 M urea denaturing gels and exposure to X-ray films overnight at -95°C.

The DNA sequencing data was stored in a VAX computer system and compared with previous data in GENBANK or EMBL databank using a program (Wilbur and Lipman, 1983). When two nucleotide sequences were aligned and compared for similarities, a program initially described by Needleman & Wunsch was used (1970).
CHAPTER THREE

CHARACTERIZATION OF C-MOS TRANSCRIPTS IN RAT
3. 1 Introduction.

Studies of the regulatory mechanisms of gene expression at present rely on the comparison of two cellular systems, one of which expresses the gene and another which does not. In order to investigate the expression of c-mos in normal animals, the identification of c-mos expressing and non-expressing tissues has been needed. At the start of this project, c-mos transcription had not been detected (see Chapter 1). Since then, the presence of the gene and its expression have been detected in several species (Leivobitch et al, 1987; Paules et al, 1988 (rat); Propst & Vande Woude, 1985; Propst et al, 1987 (mouse); Paules et al, 1988 (human & primate); Schmidt et al, 1988 (chicken); Sagata et al, 1988; Freeman et al, 1989 (Xenopus laevis). However, its transcriptional unit is still not well defined (Vande Woude et al, 1987). For instance, mouse is the only species in which the 3' and 5' ends of c-mos transcripts have been localized on a genomic map (Propst et al, 1987).

3. 1. 1 Characteristics of murine c-mos transcripts.

The characteristics of murine c-mos transcripts can be summarized as follows. The sizes of the transcripts vary between tissues: 1.7 kb in the testis, 1.4 kb in the ovary and a range of sizes between 1.3 and 6 kb in the epididymis and embryo (Propst & Vande Woude, 1985; Propst et al, 1987; Vande Woude et al, 1987).
All these variously-sized transcripts appear to have a common 3’ end, which is located 130 bp downstream from the termination codon of the open reading frame. The 5’ ends are located 280 bp (testis) and 70 bp (ovary) upstream from the first ATG of the two translation start codons (van Beveren et al, 1981a; Propst et al, 1987). The extremely large transcripts (6 kb) observed in the embryo and epididymis may have long leader sequences (Vande Woude et al, 1987). This speculation is based on the fact that a single 3’ end was mapped for the all transcripts studied. However, an attempt to map the 5’ ends was unsuccessful (Propst et al, 1987). If this hypothesis is correct, then their 5’ ends must be upstream of the previously identified inhibitory element, UMS (Vande Woude et al, 1987), which is thought to act as a terminator of transcription (Wood et al, 1984; McGeady et al, 1986).

In this Chapter, I have studied the expression of c-mos in the rat testis, brain and liver and detected it in the testis and brain. The transcripts produced in the testis have been further analyzed by S1 nuclease protection assays to localize their 3’ and 5’ ends on a genomic map. The results obtained are discussed in comparison with recent reports of c-mos transcription in rats and other species.
3. 2 Results.

3. 2. 1 Detection of c-mos transcripts in rat tissues.

RNAs were prepared from livers, testes and brains of adult male rats and assayed for c-mos transcripts using hybridization with DNA probes. Several DNA probes were prepared from plasmid subclones of an original phage clone λD3e (van der Hoorn & Firzlaff, 1984). Three types of hybridization probes were prepared, one of which was used for Northern blot hybridization and the other two were used for S1 nuclease protection assay (Figure 3.1).

A radioactively-labelled fragment derived from the coding region of the c-mos (a 0.6 kb Xho I-Hind III fragment; Probe I in Figure 3.1) hybridized to the testis and brain RNAs (Figure 3.2, lanes 1, 2; testis, lanes 5, 6; brain) but no signal was detected with the liver RNAs (Figure 3.2, lanes 3, 4). The intensities of the signals of the Northern blot hybridization were higher in the testis RNA than the brain RNA, indicating that the amount of the transcripts produced in the testis is more than that in the brain. The sizes of the testis transcripts varied from 2 to 4 kb in the poly(A)− fraction and about 1.5 to 2 kb in the poly(A)+ RNA fraction (Figure 3.2, lanes 1 and 2). The sizes of the brain transcripts varied from approximately 1.5 to 3 kb in the poly(A)− RNA fraction and
0.6 to 1.5 kb in the poly(A)+ RNA fraction (Figure 3.2, lanes 5 and 6).

The intensity of the hybridization signals for the murine testis RNA fractions, which were used as a positive control to assess the experimental conditions, were stronger than those for the rat RNA samples (Figure 3.2, lanes 2 (rat) and 7 (mouse)). This indicates that the proportion of the c-mos transcripts per unit weight of RNA is greater in the mouse than the rat testis. The smaller amount of hybridization to the rat RNAs than to the murine RNA is unlikely to be a reflection of species-specific differences in the hybridization efficiency of Probe I under the present hybridization conditions. The c-mos sequences of mouse and rat have a 90% homology in the region from which Probe I was prepared (van Beveren et al., 1981a; van der Hoorn & Firzlaff, 1984).

A Northern blot hybridization of poly (A)+ RNAs with the DNA probe identical to that used in the previous experiment was performed to measure the sizes of the transcripts more accurately. The amounts of RNAs were increased to 20 μg of testis and 10 μg of brain poly(A)+ RNAs. A 1.8% agarose gel was used to obtain good resolution of low molecular weight RNAs (Figure 3.3). Four RNA fragments with sizes of 1.5, 1.7, 1.8 and 2.1 kb were detected in the testis RNA poly(A)+ fraction (Figure 3.3, lane 2). Although the hybridization to the brain RNAs did not show clear fragment sizes on this X-ray film, the size range of the
transcripts was estimated to be between 0.6 and 1.3 kb. The smallest brain transcript was shorter than the size of the \textit{c-mos} coding region, which has been deduced to be 1.0 kb from the DNA sequence (van der Hoorn & Firzlaff, 1984).

3. 2. 2 Localization of the 5' and 3' termini of testis \textit{c-mos} transcripts on a genomic DNA map.

One way that the multiple transcript sizes observed with Northern blot hybridizations (Figure 3. 2 and 3. 3) could have been produced was if the rat \textit{c-mos} transcripts had multiple 5' ends and/or 3' ends. The termini of the testis transcripts were therefore determined by S1 nuclease mapping. The amount of \textit{c-mos} transcripts present in brain RNA were considered to be too low to successfully determine their 3' and 5' termini.

Two separate DNA fragments isolated from \textit{c-mos} subclones were used to localize the 5' ends of the \textit{c-mos} transcripts in the poly(A)+ fraction. Two DNA fragments sized 0.22 and 0.4 kb were protected when the distal probe, which was a 1.1 kb \textit{Xho I-Hind} III fragment labelled at its \textit{Hind} III end, was used for the S1 nuclease protection assay (Figure 3. 4, lane 1). These signals indicate that two termini are located 1.0 and 1.2 kb upstream of the \textit{c-mos} open reading frame, as the labelled \textit{Hind} III site is located 0.8 kb upstream of the reading frame. A 0.8 kb DNA fragment was also detected (Figure 3. 4, lanes 1 and 2), however, this is probably
due to self-folding of the DNA fragment as a signal of the same size was present in the negative control lane (Figure 3.4, lane 2).

A 0.74 kb DNA fragment was detected when the proximal probe, which was the 0.87 kb \textit{Hind} III-\textit{Xho} I fragment labelled at its \textit{Xho} I site, was used for the S1 nuclease protection assay (Figure 3.4, lane 3). The labelled position in the fragment is 63 bp downstream from the ATG in the \textit{c-mos} open reading frame previously determined (van der Hoorn & Firzlaff, 1984). Therefore, the detection of the 0.74 kb fragment indicates that a region of 0.68 kb in front of the \textit{c-mos} coding region is protected from S1 nuclease digestion. Thus the results obtained suggest the presence of multiple leader sequences in the testis \textit{c-mos} transcripts as the 5' termini of the transcripts are located at 0.68, 1.0 and 1.2 kb upstream from the putative \textit{c-mos} coding region.

The 3' termini of the \textit{c-mos} transcripts derived from the rat testis poly(A)+ fraction were localized on a genomic DNA map with 3'-end-labelled probes. A 0.5 kb \textit{Hind} III-\textit{Pvu} II fragment and a 1.4 kb \textit{Hind} III-\textit{Xba} I fragment, both of which were labelled at the \textit{Hind} III ends, were used for hybridization (Figure 3.5). Two DNA fragments with sizes of 0.23 and 0.4 kb were detected with the short probe (Figure 3.5, lane 3), indicating that the 3' termini of the \textit{c-mos} transcripts correspond to 0.26 and 0.43 kb downstream of the \textit{c-mos} coding region. The intensity of the 0.4 kb DNA fragment was much stronger than that of the 0.23 kb
fragment. Therefore, the majority of the \textit{c-mos} transcripts produced in the testis probably have 3' termini located 0.43 kb downstream of the putative open reading frame. The \textit{S}1 mapping with the long probe did not produce similar signal strengths, which may have been because the reaction conditions were not particularly suitable for the long probe.

The results obtained in this section indicate that the \textit{c-mos} transcripts produced in rat testis have multiple 5' and 3' ends in the poly(A)$^+$ fraction. The sizes of the transcripts that could be produced from the various combinations of the 3' and 5' termini are 2.01, 2.33 and 2.53 kb in conjunction with the 0.26 kb 3' terminus and 2.15, 2.50 and 2.70 kb in conjunction with the 0.46 kb 3' terminus (Figure 3. 6). All of these 6 transcripts are not necessarily produced. The sizes of most of these putative transcripts are longer than the sizes of the transcripts observed by Northern blot hybridization, even though the latter transcripts contain a poly(A) tail of unknown length (Figures 3. 2 and 3. 3).

\textbf{3. 2. 3 Multiple ATGs are found in the leader sequence of rat \textit{c-mos} transcripts.}

The results described in the previous sections demonstrate that the rat \textit{c-mos} transcripts produced in the testis are much larger than the open reading frame. Variable 5' ends with long leader sequences were also
detected and they may be generated by either alternative splicing sites of an upstream intron or alternative initiation sites for the transcription. A portion of the leader sequence, immediately adjacent to the putative open reading frame, was characterized by DNA sequencing (Figure 3. 7). The determined sequence was searched for similarity to consensus sequence for intron/exon boundaries.

Eight ATGs were newly identified (boxed in Figure 3. 7) and most of them are followed by termination codons and are thus unlikely to be used for translation. The eight ATG codons were separated from the \( c-mos \) coding region by stretches of A residues. This A-track and its surrounding sequence was similar to some eukaryotic tRNA sequences (underlined in Figure 3. 7).

The sequence was searched for consensus signals for intron/exon boundaries. Four AGGTs were observed and are identified in Figure 3. 7 by double underlines (AG is the critical dinucleotide, see Chapter 1, Section 1. 2. 2. 3; Mount, 1982). However, the sequences surrounding the AGGTs did not match significantly with the favored sequences for intron generation \((T/C)\_nNC/\_TAG/\_G\) (Mount, 1982; Rogers, 1985).

The nucleotide sequence was also compared to the corresponding region of the murine \( c-mos \) (van Beveren et al, 1981a; Wood et al, 1984).
to detect the most recent changes to the nucleotide sequences. The two sequences had an 88% homology, which was slightly lower than the homology observed in the rat and mouse open reading frames previously (93%, van Beveren et al, 1981b; van der Hoorn & Firzlaff, 1984). The rat sequence had an additional A track (22 bp) and the deletion of a T track (9 bp). Most changes were single base pair substitutions (Figure 3.8).
Figure 3.1 A restriction map of the rat c-mos illustrating the location of the hybridization probes.

The coding (shown as an open box) and its proximal flanking region is illustrated. The hybridization probes used in this chapter are divided into three groups.

Probe I was prepared from a 0.6 kb Xho I-Hind III fragment and was used for Northern-blot hybridizations.

Probe II consists of a 1.1 kb Xho I-Hind III fragment labelled at the Hind III site and a 0.87 kb Hind III-Xho I fragment labelled at the Xho I site. Probe III consists of a 1.4 kb Hind III-Xba I fragment and a 0.5 kb Hind III-Pvu II fragment labelled at the Hind III sites. The position of labelled nucleotides on the fragments are shown by asterisks (*) and the restriction endonuclease cleavage sites are shown by the following abbreviations: H, Hind III; K, Kpn I; P, Pvu II; Xb, Xba I; Xh, Xho I.
Figure 3. 2 Detection of the *c-mos* transcripts in the rat tissues.

Five microgram each of the poly(A)$^+$ fraction and 20 μg each of the poly(A)$^-$ fraction were fractionated on a 1% agarose gel. The RNAs immobilized on a nylon membrane were hybridized with Probe I at 55°C. The positions of ribosomal RNAs determined by ethidium bromide staining are shown as 28S and 18S.

Rat RNA samples:

**poly(A)$^-$ fraction**
- lane 1, testis
- lane 3, liver
- lane 5, brain

**poly(A)$^+$ fraction**
- lane 2, testis
- lane 4, liver
- lane 6, brain

lane 7, 20 μg of unfractionated murine testis RNA.
Figure 3. Multiple transcripts are present in the rat testis.

Poly(A)$^+$ RNAs prepared from rat testis and brain were separated on a 1.8% agarose gel, transferred to a nylon membrane and hybridized with Probe I. The conditions of the hybridization, washing of the membrane and the exposure to an X-ray film were as described in Chapter 2.

Rat RNA samples; lane 1, 20 $\mu$g of liver poly(A)$^-$ fraction;
lane 2, 20 $\mu$g of testis poly(A)$^+$ fraction;
lane 3, 10 $\mu$g of brain poly(A)$^+$ fraction;

Murine RNA sample; lane 4, 20 $\mu$g of unfractionated RNA fraction.
Figure 3. 4 Mapping of the 5’ ends of the rat testis c-mos transcripts (Probe II).

A 1.1 kb Xho I-Hind III fragment labelled at the Hind III site was used for a distal probe. Another fragment, a 0.8 kb Hind III-Xho I fragment labelled at the Xho I site, was used for a proximal probe. The distal or proximal probe was hybridized with 50 µg or 70 µg of poly(A)+ RNA fraction prepared from the rat testes, respectively. Fifty microgram of yeast tRNA were processed in the same way and used as a negative control for the hybridization conditions.

In the map below, the arrows and bars indicate the two probes used and the protected fragments, respectively. The asterisks (*) indicate the radioactively-labelled sites on the fragments.
Figure 3. 5 Mapping of the 3' ends of the rat testis transcripts (Probe III).

A short probe (0.5 kb \textit{Hind} III-\textit{Pvu} II DNA fragment) and a long probe (1.4 kb \textit{Hind} III-\textit{Xba} I DNA fragment) were labelled at the \textit{Hind} III ends by a filling-in reaction and hybridized separately with 50 \( \mu \text{g} \) each of poly\((A)^+\) RNA fraction prepared from testis (lane 1; the long probe, 3; the short probe). Fifty microgram of yeast tRNA were used as a negative control as described in the legend to Figure 3. 4 (lane 2 & 4).

In the map below, the arrows and bars present the two probes and the protected fragments, respectively. The asterisks (\( \ast \)) indicate the radioactively-labelled sites on the fragments.
Figure 3. 6 A schematic diagram of the c-mos transcripts produced in the rat testis.

The 5' ends (left hand side) and the 3' ends (right hand side) of the c-mos transcripts produced in the testis are indicated by vertical arrows in a restriction map. A set of possible transcripts deduced from the results of S1 protection assays are indicated by horizontal arrows with their sizes in kb.
Figure 3. The DNA nucleotide sequence of the leading sequence to the coding region in the c-mos transcripts.

The position of the nucleotide sequence is shown in negative numbers, as number 1 is set at the first nucleotide of the putative c-mos coding region determined previously (van der Hoorn & Firzlaff, 1984). Twelve ATGs found in this region are shown in boxes. A long stretch of A residues is underlined and four AGGT sequences are double-underlined.
<table>
<thead>
<tr>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAGCTTGCT</td>
<td>CTGTGACCTA</td>
<td>ACTTGCTTGG</td>
<td>GAAAGGAAAG</td>
<td>GACGGTAGAA</td>
</tr>
<tr>
<td>60</td>
<td>70</td>
<td>80</td>
<td>90</td>
<td>100</td>
</tr>
<tr>
<td>GTCAGTGTTG</td>
<td>TGTTAAGGCG</td>
<td>TCTTAACTTG</td>
<td>CTGTTTAAGATC</td>
<td>AATGATACAGT</td>
</tr>
<tr>
<td>110</td>
<td>120</td>
<td>130</td>
<td>140</td>
<td>150</td>
</tr>
<tr>
<td>GTCAGATTGC</td>
<td>TCCCTCTTGAC</td>
<td>GACTTAGG</td>
<td>AGGTACGGAT</td>
<td>TCTGCCCTCT</td>
</tr>
<tr>
<td>160</td>
<td>170</td>
<td>180</td>
<td>190</td>
<td>200</td>
</tr>
<tr>
<td>ATCCGAGATT</td>
<td>GTATTTGGAT</td>
<td>CAAAGACATT</td>
<td>AAAATTAGAG</td>
<td>ATGTAATTTC</td>
</tr>
<tr>
<td>210</td>
<td>220</td>
<td>230</td>
<td>240</td>
<td>250</td>
</tr>
<tr>
<td>TTTCAGGT</td>
<td>TGCATCTAAT</td>
<td>AACCCTGCCA</td>
<td>AGTAGTCCA</td>
<td>CAAGTGCTCG</td>
</tr>
<tr>
<td>260</td>
<td>270</td>
<td>280</td>
<td>290</td>
<td>300</td>
</tr>
<tr>
<td>CTGTGTGTAC</td>
<td>TGCAACTCTT</td>
<td>GCAGCAATA</td>
<td>GCTGCTGGGA</td>
<td>CTGATTTAT</td>
</tr>
<tr>
<td>310</td>
<td>320</td>
<td>330</td>
<td>340</td>
<td>350</td>
</tr>
<tr>
<td>CAGCTGGGAG</td>
<td>ACTAGCAGCA</td>
<td>ACAATCAATT</td>
<td>ATAGCCTGAA</td>
<td>AAGTGTCAC</td>
</tr>
<tr>
<td>360</td>
<td>370</td>
<td>380</td>
<td>390</td>
<td>400</td>
</tr>
<tr>
<td>CAAGTTCACT</td>
<td>GTACTTTTGC</td>
<td>TTTTCTCTTC</td>
<td>AGCTAAGGA</td>
<td>CCAATTAT</td>
</tr>
<tr>
<td>410</td>
<td>420</td>
<td>430</td>
<td>440</td>
<td>450</td>
</tr>
<tr>
<td>ATCGTCAGAA</td>
<td>GTCCTTGGAA</td>
<td>ATGTTTTTTT</td>
<td>ATTGTATTAT</td>
<td>GATTCAGAA</td>
</tr>
<tr>
<td>460</td>
<td>470</td>
<td>480</td>
<td>490</td>
<td>500</td>
</tr>
<tr>
<td>TTAGTTGGA</td>
<td>GGAGAGGTGA</td>
<td>GCTTTGGGCC</td>
<td>AGGCCAGTAC</td>
<td>TGTTTAACAG</td>
</tr>
<tr>
<td>510</td>
<td>520</td>
<td>530</td>
<td>540</td>
<td>550</td>
</tr>
<tr>
<td>GTGAGAGGTT</td>
<td>GTTTTGCTAA</td>
<td>GGATTCACTT</td>
<td>TAGCTGTGAG</td>
<td>AGAGCAATCG</td>
</tr>
<tr>
<td>560</td>
<td>570</td>
<td>580</td>
<td>590</td>
<td>600</td>
</tr>
<tr>
<td>GTTTCATCTG</td>
<td>AGACTGCGAC</td>
<td>GCTTCATCTG</td>
<td>CACGCCAACA</td>
<td>CCCACCTGAA</td>
</tr>
<tr>
<td>610</td>
<td>620</td>
<td>630</td>
<td>640</td>
<td>650</td>
</tr>
<tr>
<td>AAAAAAAAAAA</td>
<td>CAAAACAAAAA</td>
<td>AAAACAAAAAA</td>
<td>CCAACCTTGG</td>
<td>TGAGCAGCA</td>
</tr>
<tr>
<td>660</td>
<td>670</td>
<td>680</td>
<td>690</td>
<td>700</td>
</tr>
<tr>
<td>GACGTGAGAG</td>
<td>CACAGTTGTTG</td>
<td>GCTGGGTTTG</td>
<td>AGAAATAACAG</td>
<td>AAGAAGGGAA</td>
</tr>
<tr>
<td>710</td>
<td>720</td>
<td>730</td>
<td>740</td>
<td>750</td>
</tr>
<tr>
<td>GAAGAGTTC</td>
<td>AGGCCGGAAC</td>
<td>CTTCAGCAGATC</td>
<td>GTTCCAAACAC</td>
<td>TTTCCCTGGGT</td>
</tr>
<tr>
<td>760</td>
<td>770</td>
<td>780</td>
<td>790</td>
<td>800</td>
</tr>
<tr>
<td>GTTCCGGTCA</td>
<td>TTTCTCTCTTA</td>
<td>GCGTCTCTTG</td>
<td>TGACGTCCCC</td>
<td>ATCTGAGGGT</td>
</tr>
<tr>
<td>810</td>
<td>820</td>
<td>830</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GTAATCCCT</td>
<td>CGCTCTCTCAT</td>
<td>CCTGTGTGTCG</td>
<td>TACCTCCC</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3. A comparison of the rat and mouse sequences immediately before the coding region.

The DNA sequence of the murine c-mos 5' flanking region was that determined by Van Beveren et al (1981a) and Wood et al (1984) and recorded in the GENBANK (release No. 47, 1987). The comparison was carried out using a program initially described by Needleman & Wunsch (1970) driven by a VAX/VMS system.
3. 3 Discussion.

3. 3. 1 Rat \textit{c-mos} transcripts.

Variously-sized \textit{c-mos} transcripts were detected in the rat testis and in very small amounts in the brain. The rat liver did not contain \textit{c-mos} transcripts at a level detectable by Northern blot hybridization with a DNA probe (Figure 3. 2). The sizes of the rat testis \textit{c-mos} transcripts ranged from 2.2 to 4.0 kb in the poly(A)$^-$ RNA and from 1.5 to 2.1 kb in the poly(A)$^+$ RNA. The brain transcripts were barely detectable and ranged in size from 1 to 3 kb in the poly(A)$^-$ and 0.6 to 1.5 kb in the poly(A)$^+$ RNA fractions (Figure 3. 2 & 3. 3). These observations are in broad agreement with recent reports of \textit{c-mos} expression in rat testis (2.4 - 7.0 kb), embryo (ca 1.3 kb) (Propst et al, 1988) and muscle (3.6 kb) (Leivobitch et al, 1987).

The sizes of the poly(A)$^+$ transcripts observed here were shorter than those observed in the poly(A)$^-$ fraction. For instance, the largest transcripts in the testis were 2.1 and 4 kb in the poly(A)$^+$ and poly(A)$^-$ fractions, respectively. However, 7 kb long transcripts have recently been reported in both poly(A)$^+$ and poly(A)$^-$ fractions, indicating that rat \textit{c-mos} may exist as a very long transcript with only a short poly(A) tail.
A similar discrepancy exists in the literature about murine \textit{c-mos} transcripts. Propst & Vande Woude, observed the same sizes of \textit{c-mos} transcripts in both poly(A)$^+$ and poly(A)$^-$ RNA fractions and suggested that \textit{c-mos} transcripts have an extremely short poly(A) tail as an increase in hybridization signal was not observed after passages of the RNA down an oligo-d(T) cellulose column (Propst & Vande Woude, 1985). In contrast, Mutter & Wolgemuth reported that murine \textit{c-mos} transcripts were longer in the poly(A)$^-$ than poly(A)$^+$ fraction and that a great degree of enrichment of murine \textit{c-mos} signals occurred after the use of oligo-d(T) columns (Mutter & Wolgemuth, 1987). An explanation for the observed differences may be obtained from a study of the structure of the long transcripts detected in mice and rats. However, the possibility that the shorter sizes observed in the poly(A)$^+$ fractions is caused by degradation during fractionation of the poly(A)$^+$ RNAs has not yet been eliminated.

3. 3. 2 The \textit{c-mos} transcripts of various species.

A common feature of the \textit{c-mos} transcripts which have been detected in various species is their tissue-specific abundance. In all species examined, the expression of \textit{c-mos} is most abundant in the reproductive
organs (Propst & Vande Woude, 1985; Schmidt et al, 1988; Sagata et al, 1988; Paules et al, 1988). The testis and ovary produce similar amounts of transcripts in mammals (Propst et al, 1987; Paules et al, 1988) but in *Xenopus*, *c-mos* transcripts appear to be more common in the ovary (Sagata et al, 1988; 1989; Freeman et al, 1989). *C-mos* transcripts have also been detected in the murine kidney, brain, placenta and mammary gland at levels which are only detectable by S1 nuclease protection assays (Propst et al, 1987). These transcripts have not been characterized, due to their small amounts, although this is obviously important given the recent speculation that the *c-mos* may have a biological function in somatic cells (Leivobitch et al, 1987).

Some of the rat *c-mos* transcripts detected were very large (2-4 kb, Figure 3.4; up to 7 kb Propst et al, 1988; testis) compared with the size of the open reading frame (1.0 kb) deduced from the nucleotide sequence (van der Hoorn & Firzlaff, 1984). Large *c-mos* transcripts have also been reported for rat muscle (3.6 kb, Leivobitch et al, 1987), murine epididymis and embryo (6 kb, Propst & Vande Woude, 1985) and *Xenopus* oocyte (3.1 kb, Sagata et al, 1988). The *Xenopus c-mos* transcript was analyzed by cDNA cloning and found to contain an extremely long non-translated 3' region (2 kb; Sagata et al, 1988). The reason why long *c-mos* transcripts are produced in other species, such as the murine epididymis (6 kb in poly(A)^+ RNA), has been speculated to be due to the presence of a long 5' untranslated region (Propst et al, 1987;
Vande Woude et al, 1987). However, the extremely long transcripts have not been analyzed further. Another possibility is that the long transcripts result from a "reading through" from a transcript-ional unit located upstream of the c-mos. Characterization of the large transcripts may help to solve this question.

Short transcripts have been observed in the rat brain (Figure 3.2 and 3.3), murine embryo (1.3 kb; Propst & Vande Woude, 1987) and human and monkey testes (0.87 & 0.97 kb, respectively; Paules et al, 1988). The transcripts produced from the primates have a common 3′ end and their 5′ ends are therefore thought to be within the coding region (Paules et al, 1988). The short transcripts from the rat brain were detected as weak signals in the Northern blot hybridization. This does not necessarily imply that higher levels of the transcripts will not be found within discrete areas within the rat brain and murine embryo. In situ hybridization with various portions of a genomic mos fragment will be a valuable procedure to analyze these short transcripts. An important question to be answered is whether the short transcripts are used to generate a protein or are degradation products.

3.3.3 The termini of rat testis transcripts.

The c-mos transcripts produced from a normal rat tissue were analyzed by S1 nuclease mapping, for the first time. Three 5′ ends (0.68,
1.0 and 1.2 kb upstream from the \textit{c-mos} open reading frame) and two 3' ends (0.26 and 0.43 kb downstream) were detected for the testis poly(A)$^+$ RNA (summarized in Figure 3.6).

Multiple 5' ends may be generated by either alternative splicing sites of an upstream intron and/or alternative initiation sites for transcription (see Chapter 1, Sections 1.2.2.3 and 1.2.3). Consensus splice site sequences for an upstream intron were detected, but the nucleotides surrounding them were not those which are generally favourable for splicing. This needs to be directly examined by sequencing of a cDNA clone and comparing it with the genomic sequence.

The presence of long leader sequences raises possibilities about two different aspects of \textit{c-mos} expression. The first is about its transcriptional unit. The transcriptional unit for the rat \textit{c-mos} may start far upstream (at least 1.2 kb) from the open reading frame. In combination with the total transcripts sizes observed, the possible involvement of long 5' leader (and/or 3' tailing) sequences in the transcriptional unit needs to be considered. The second is that the \textit{c-mos} expression may be regulated at the translational level as well as the transcriptional level. This will be discussed in the next section.

The 3' ends for the rat \textit{c-mos} transcripts produced from recombinant DNAs transfected into cultured murine cells have been
previously mapped to 80 bp downstream of a polyadenylation signal (200 bp downstream of the open reading frame) (van der Hoorn & Neupert, 1986). The results obtained in my study indicate that a minor portion of the rat testis transcripts have 3' end close to the site described in the above study (250 bp), however, the dominant 3' site is 230 bp further downstream from that point. This size difference may be because of the different sources of the RNAs (rat testis vs murine cultured cells).

It should be noted that the lengths of putative transcripts produced from the 5' termini located 1.2 kb upstream (2.53 and 2.70 kb, in combination with the observed 3' termini) are longer than the transcripts detected by Northern blot hybridization (2.1 kb in poly(A)+ RNA). Furthermore, small transcripts (1.5 kb) were detected by Northern blotting which are not readily related to the observed termini. As the same poly(A)+ RNA fraction was used for the Northern blotting and S1 mapping, degradation of the RNA samples is not the reason.

A size discrepancy has also been reported in the murine c-mos transcripts. Estimations made by S1 nuclease mapping and Northern blot hybridization for murine testis poly(A)+ RNA are approximately 1.3 kb and 1.7 kb, respectively (Propst & Vande Woude, 1985; Propst et al, 1987). The longer sizes of the murine c-mos measured in the Northern may be due to polyadenylation, although this does not fit in with the
possible presence of a short poly(A) tail (see previous section).

Polyadenylation is not be the cause of the size difference in the rat c-mos, as estimates made by S1 mapping are exclusive of the poly(A) tail but are still longer than the observations made by Northern analysis. One possibility is that the discrepancy may have resulted from technical reasons, such as a non-linear distribution of RNA molecules in the agarose gels (Lehlach et al, 1977). Another possibility is that a part of the coding region (0.4 kb) is spliced out, although I consider this unlikely as an intron has not been detected in the coding region of c-mos (van der Hoorn & Neupert, 1986).

Lastly, the difference in the estimation of transcript length could be produced if a terminus had not been detected, particularly if it was a 3’ termini located within the coding region. The probes used to detect the termini would not have detected a terminus within the coding region.

3. 3. 4 Characteristics of the leader sequence in the rat c-mos transcripts.

The nucleotide sequence determination of the 5’ non-coding region demonstrated 8 additional ATGs in front of the ATGs initially identified for the c-mos open reading frame (Figure 3. 7; van der Hoorn & Firzlaff, 1984). The presence of multiple ATGs may be a common
feature of the \textit{c-mos} gene as it has been reported in mouse, human, primate, and chicken \textit{c-mos} genes (Wood et al, 1982; Watson et al, 1982; Paules et al, 1988; Schmidt et al, 1988). The chicken \textit{c-mos} has 10 ATGs before its coding region and most of the ATGs are thought to be unfunctional as terminator codons follow immediately after them (Schmidt et al, 1988).

Multiple ATGs or small open reading frames (mini-cistrons) in the 5' non-coding sequence of a transcript have been postulated to decrease the translation efficiency of a gene (Kozak, 1986b). The role of ATGs has been examined by correlating the amount of protein production with the number of 5' ATGs in cloned genes (Kozak, 1984b; 1986a). For instance, it has been demonstrated that the human preproinsulin gene produces less than 10% of its normal amount of protein when a second ATG is inserted (Kozak, 1984b).

The influence of multiple ATGs on the expression of the \textit{c-mos} has not been experimentally examined. However, the human and primate \textit{c-mos} genes can only be activated when a heterologous regulatory sequence is introduced very close to the coding region to abolish the first ATG (Blair et al, 1986; Paules et al, 1988). The human and primate \textit{c-mos} is thought to be more inactive than rodent genes as they show low levels of transformation activity even when linked with viral enhancers (Blair et al, 1986; Paules et al, 1988). This low potential could be a consequence of
translational control, although the evidence for this is indirect.

The presence of the long 5' non-coding sequence and mini-cistrons may be a characteristic of proto-oncogene transcripts (Kozak, 1986; 1987). When the structures of the 5' non-coding sequences of a large number of vertebrate mRNAs were surveyed, it was found that nearly all transcripts derived from 33 proto-oncogenes have leader sequences larger than 100 nucleotides (nt) (Kozak, 1987). Of 346 mRNAs in which transcriptional start sites have been mapped, 73% have a 5' non-coding sequence shorter than 100 nt. It was also noted that-two thirds of the oncogene transcripts cited had extra ATGs preceding the major open reading frames, even though extra ATGs were relatively rare in the rest of the cellular genes (Kozak, 1987). It has been speculated, therefore, that translational regulation contributes to the restriction of proto-oncogene expression (Kozak, 1986; 1987).

3.3.5 Summary.

The c-mos is expressed in various sizes of mRNAs in rat testis and brain, but not in the liver. Investigations of the mechanisms that govern the expression of a gene rely on the comparison of tissues which express the gene with those that do not. The discovery of two tissues that express c-mos was necessary for the experiments reported in Chapters 4 and 5 to be undertaken. The variation in the sizes of the testis transcripts appears to
be due to the presence of transcripts with different 3' and 5' termini. The testis c-mos transcripts contained a long leader sequence, the characteristics of which raises the possibility that the expression of c-mos may be subject to translational regulation. A further consequence of the presence of a long leader sequence is that possible regulatory sequences may be present far upstream of the c-mos open reading frame.
CHAPTER FOUR

A SEARCH FOR TRANSCRIPTIONAL UNITS IN THE PROXIMAL REGION OF C-MOS
4. 1 Introduction.

Expression of the \textit{c-mos} has now been observed in the gonadal tissues of several species, although the level of transcription is commonly low. The observation of large transcripts in rodents suggests that the regulation of \textit{c-mos} expression is complex and that regulatory sequences may be relatively far away from the coding region (Chapter 3). In the next two chapters, I will examine and discuss two hypotheses about transcriptional regulation of the \textit{c-mos}.

4. 1. 1 A search for mechanisms that regulate expression of the \textit{c-mos} gene.

Investigations of the mechanisms that regulate \textit{c-mos} transcription had begun, using recombinant DNAs and transfection assays, even before \textit{c-mos} expression had been observed in normal animals (Chapter 1). These early studies demonstrated that the biological properties of \textit{v-mos} and the murine \textit{c-mos} coding regions are very similar but that their flanking regions differ in their nucleotide sequences. The viral flanking regions promote transcription, whereas the 5' flanking regions of both mouse and rat \textit{c-mos} contain elements which repress the expression of transfected \textit{c-mos} (Oskarsson et al, 1980; Wood et al, 1984; McGeady et al, 1986; van der Hoorn & Neupert, 1986).
4. 1. 2 Spatial arrangement of genes in chromosomes.

Another possible explanation for the low level of \textit{c-mos} expression has been that its transcription is repressed due to the presence of a transcriptional unit closely adjacent to it. Statistical analysis of known gene structures indicate that the distance between transcriptional units of clustered genes is positively correlated with the sizes of the units (Naora & Deacon, 1982a). It has been suggested that there are minimum and maximum intergenic distances between transcription units. When the distance between two units falls below the minimum value then inactivation of one or both of the transcriptional units will occur or a substantially single unit will be formed by reading through (Naora & Deacon, 1982a,b).

The \textit{c-mos} may be an interesting gene to study in relationship to this hypothesis (Naora et al, 1985). The expression of \textit{c-mos} is unusually low and the presence of a polyadenylation signal in the murine UMS sequence may indicate that there is a transcriptional unit upstream of this element (McGeady et al, 1986). Furthermore, the murine \textit{c-mos} is activated when its genomic structure is rearranged by insertion of an LTR-like sequence or the entire genome of an intracisternal A particle (Rechavi et al, 1982; Canaani et al, 1983; Gattoni-Celli et al, 1983). The activation of \textit{c-mos} produced by these rearrangements has been suggested to be due to the capacity of the inserts to act as activators but it is also possible that
they activate c-mos expression by increasing the distance between the c-mos and a putative transcriptional unit.

An aim of this chapter was to investigate the rat c-mos 3' and 5' flanking regions to examine whether or not there is an active transcriptional unit in close proximity to the c-mos. The 3' flanking region of the rat c-mos was partially characterized as part of this investigation.
4.2 Results.

4.2.1 cDNA hybridization assays to detect putative neighbouring transcriptional units.

In order to search for the presence of an actively utilized transcriptional unit near the \textit{c-mos} gene, a cDNA probe prepared from rat liver poly(A)$^+$ RNA was hybridized with λD3e DNA digested with restriction endonucleases. Liver poly(A)$^+$ RNA was chosen to make the $^{32}$P-labelled cDNA as it was assumed that the putative neighbouring gene would be most likely to be actively transcribed in a tissue where \textit{c-mos} expression is repressed. Random oligomers were also used for priming the synthesis of the cDNA probe in order to avoid missing out the 5' portion of mRNA.

The λD3e phage DNA was digested with various combinations of restriction endonucleases (Figure 4.1), size-fractionated on a 1% agarose gel (Figure 4.2a), transferred to a membrane filter and subsequently hybridized with radioactively-labelled cDNA prepared from the rat liver poly (A)$^+$ RNA. The probe hybridized with the \textit{Eco RI-Hind III} 4.2 kb fragment (Figure 4.2b, lane 3), which contains the 3' flanking region of the \textit{c-mos} gene. The position of the signal was determined to be between the \textit{Bgl II} and \textit{Pst I} sites which are located, respectively, 4.5 kb and 6.4 kb
downstream from the end of the c-mos coding region (Figure 4.2b, lane 10). Some DNA fragments derived from the phage vector of λD3e were not distinguishable in size from those derived from the genomic DNA insertion containing c-mos. Therefore, both the 5' and 3' flanking regions of c-mos were subcloned and examined with the same hybridization probe to eliminate the slight possibility that the positive hybridization signal is derived from the vector DNA and also to check the reproducibility of the initial observation.

A 4.2 kb Eco RI-Hind III fragment was isolated from the c-mos 3' flanking region, cloned into a pUC18 plasmid vector and designated as pEH42. A 3.2 kb Eco RI-Hind III fragment was isolated from the 5' flanking region, similarly cloned and designated as pEH32 (Figure 4.1). These two DNAs were cleaved with various combinations of restriction endonucleases (Figure 4.3a, for pHE42, Figure 4.3c, for pEH32), transferred onto a membrane and hybridized with the 32P-labelled cDNA probe, prepared from liver poly(A)+ RNA, which had been used for the previous experiment. The cDNA probe hybridized with DNA fragments derived from the pHE42 (Figure 4.3b) but not with those from the pEH32 (Figure 4.3d). This indicates that (a) the DNA fragments which hybridized with the cDNA probe (Figure 4.2b) are not from the vector and are also not caused by a non-specific binding; (b) a part of the 3' c-mos flanking region contains nucleotide sequences contained in the liver...
transcripts as the hybridization probe was prepared from the liver poly(A)+ RNA; (c) the 5' c-mos flanking region does not produce transcripts at a level detectable by the procedure that was used (Figure 4.3d).

4.2.2 Characterization of a DNA fragment from the 3' c-mos flanking region.

RNAs were prepared from rat liver, testis and brain, so that the tissue specificity and amounts of the putative 3' transcripts could be correlated with the level of c-mos expression (Chapter 3). After fractionation and transfer to a membrane, the RNAs were hybridized with a radioactively-labelled probe prepared from a 2.7 kb Pst I fragment isolated from pHE42. The hybridization signals showed a smear-like feature, ranging in size from 0.5 to 5 kb indicating that the fragment is included in various transcripts (Figure 4.4). The pattern of hybridization was similar in the RNAs of all three tissues, although the strengths of the hybridization signals were greater in the RNAs prepared from testis and brain than the RNA from liver (Figure 4.4, lanes 2, 3 and 4). Thus the strength of the signals was greater in the tissues that express c-mos than in the tissue that does not express detectable levels of c-mos.

Repetitive sequences have been known to give rise to transcripts of variable sizes (reviewed by Jelinek & Schmidt, 1982; Weiner et al, 1987). The possibility that the 2.7 kb Pst I fragment contains repetitive sequences
was examined using a Southern-blot-hybridization to genomic DNAs. High-molecular weight genomic DNAs were prepared from mouse and rat livers and human DNA, prepared from a lymphoid cell line, was provided by Dr Hiroto Naora. One microgram of each of the DNAs was cleaved with \textit{Eco} RI, \textit{Hind} III or a mixture of both enzymes and fractionated on an agarose gel (Figure 4.5a). The 2.7 kb \textit{Pst} I fragment used for the Northern-blot-hybridization described in Figure 4.4 was hybridized with the genomic DNAs. In the rat and murine genomic DNAs the hybridization signals appeared as smears, with a size range of approximately 1 to 12 kb, (Figure 4.5b, lanes 1-6). No hybridization was detected with the human DNA (Figure 4.5b, lanes 7-9). The results indicate that sequences homologous with the 2.7 kb \textit{Pst} I-\textit{Pst} I fragment are present in the rodent genomes but are not present at a detectable level in the human genome. The widely spread sizes of the positive signals in the mouse and rat DNAs demonstrate that the probe includes DNA sequences which are located in various positions in the rodent genome.

4.2.3 Characterization of the sequences in the 2.7 kb \textit{Pst} I fragment.

The hybridization described in section 4.2.1 (Figure 4.2) shows that the sequence which hybridized with the cDNA probe is located between \textit{Bgl} II and \textit{Eco} I. A part of the region, a \textit{Sac} I-\textit{Eco} RI fragment was subcloned (BSS 1; Figure 4.1) and examined with Southern-blot-hybridizations to \textit{λD3e} or genomic DNAs and Northern-blot-
hybridizations to RNAs derived from the rat tissues. The most abundant repetitive sequence family, the B2 element, is known to contain orientation-specific transcripts (Ryskov et al, 1983). Therefore, the vector, Bluescribe, was used to produce strand-specific probes to test whether the repetitive sequence derived from the *c-mos* region have strand-specificity in the transcripts.

A radioactively-labelled probe prepared from the 0.8 kb Sac I-Eco RI fragment hybridized with the murine and rat genomic DNAs but not with humans and produced smear-like features. The 0.8 kb Sac I-Eco RI fragment thus contained the region of DNA which produced the previously described hybridization signals. Northern blot hybridization was used to characterize the transcripts hybridized with the Sac I-Eco RI fragment (Figure 4. 6). Poly(A)+ and poly(A)- RNA fractions prepared from the rat testis, liver and brain were hybridized with strand-specific RNA probes prepared from the BSS1 with either T3 or T7 RNA polymerase. The T3 RNA polymerase was used to synthesize a probe to detect anti-sense transcripts (relative to the orientation of the *c-mos* transcripts) (anti-sense probe; Figure 4. 6a) and T7 RNA polymerase for synthesizing a probe to detect sense-transcripts (sense probe; Figure 4. 6b).

The sense and antisense RNA probes were hybridized with the same filter and resulted in similar hybridization patterns (Figure 4. 6). The transcripts containing the Sac I-Eco RI fragment were rich in the testis
and brain, and ranged in size from approximately 0.6 to 10 kb. The brain poly(A)+ RNA fraction had a large proportion of the transcripts in a size range of 0.6 to 6.0 kb. The pattern of hybridization with the sense- and antisense-probes was similar with the exception that three bands (1.4, 1.7 & 2.1 kb) were detected in the testis poly(A)- fraction by the sense but not the antisense probe. These three bands were also present in the murine testis RNA. Thus, rat and murine testis RNAs contain orientation-specific transcripts, although the size distribution of the transcripts that hybridized with the strand-specific probes were very similar.

4. 2. 4 Does the murine c-mos flanking region contain a homologous sequence to the Sac I-Eco RI fragment?

The murine c-mos flanking region contains a cluster of B1 elements, which belong to a murine repetitive sequence family (Propst & Vande Woude, 1984; Krayev et al, 1980). The possibility that the Sac I-Eco RI fragment contains similar repetitive DNA elements was examined by hybridizing the Sac I-Eco RI fragment to the DNA of a murine c-mos clone.

A murine genomic clone, pM2, which contains the c-mos and its 3' flanking region, was cleaved with various restriction endonucleases and the DNA fragments were fractionated on an agarose gel (Figure 4. 7a). The DNA fragments immobilized on a membrane were hybridized with
the radioactively-labelled Sac I-Eco RI fragment (Figure 4.7b). The region which hybridized with the SacI-Eco RI fragment was localized to approximately 6 kb downstream from the 3' end of the murine c-mos coding region, by using a published restriction map (Blair et al, 1981). The B1 containing regions have previously been shown to be located 10 kb upstream and 0.6, 2.7, and 5.4 kb downstream of the murine c-mos coding region (Propst & Vande Woude, 1984). It therefore appears that the Sac I-Eco RI fragment has a similarity with a part of the murine DNA but not with the repetitive sequences previously described in the murine c-mos flanking region (Propst & Vande Woude, 1984).

4.2.5 Nucleotide sequence of the Sac I-Eco RI fragment.

The DNA sequence of the Sac I-Eco RI fragment was determined as described in Chapter 2 (Figure 4.8) and two data-banks, GENBANK (release 48, 1987) and EMBL (release 14, 1988) were searched for homologous sequences. A region of the Sac I-Eco RI fragment, between nucleotide number 236 to 400, (number 1 is the first nucleotide of the Sac I site in the Sac I-Eco RI fragment in Figure 4.8) was highly homologous with two related repetitive sequences, the B2 element (86%; a consensus sequence is from Krayev et al, 1982) and the rodent Alu type 2 repetitive sequence (71%, Haynes & Jelinek, 1981). This portion of the Sac I-Eco RI fragment also contained signal sequences for RNA polymerase III (Galli et al, 1981); a split promoter sequence (TGGCTTCAGTGG; 237-
248 & TGAGTTCAAATCC; 280-292), a termination signal (TTCTTTTT; 419-425) and an over-lapping poly-adenylation signal (AATAAAATAAA; 385-394) (Figure 4.8).

Another region from nucleotide number 431 to 511 was almost identical to the rat ID sequence (94%; Sutcliffe et al, 1982) (Figure 4.8, hatched underline). This region was located just after the B2-like portion and was followed by a polyadenylation signal (506-511) and an additional consensus signal for polyadenylation (TGTGG; 555-559) (Figure 4.8).

The portion of the Sac I-Eco RI fragment which contained the two repetitive sequences was followed by repeats of CA (563-570) and flanked by adenine-rich sequences (117-223; 707-720) (Figure 4.8).
The restriction endonuclease cleavage sites of \( \lambda D_3e \) were re-determined by myself. A 3.2 kb \( Eco \; RI-Hind \; III \) fragment was isolated from the \( \lambda D_3e \), cloned into a pUC18 plasmid vector and designated as pEH32. A 4.2 kb \( Hind \; III-Eco \; RI \) fragment was isolated, cloned into the same plasmid vector and designated as pHE42. A 0.8 kb \( Sac \; I-Eco \; RI \) fragment was isolated from the subclone pHE42 and cloned into a Bluescribe vector, which is designated as BSS1. The abbreviations of restriction endonuclease cleavage sites are as follows:

- B, \( Bgl \; II \)
- E, \( Eco \; RI \)
- H, \( Hind \; III \)
- K, \( Kpn \; I \)
- P, \( Pst \; I \)
- S, \( Sac \; I \)
- Xb, \( Xba \; I \)
- Xh, \( Xho \; I \)

The restriction map of a recombinant phage \( \lambda D_3e \), which contains the rat \( c-mos \) sequence, and its subclones.

Figure 4.1

:\n
Figure 4.1 Restriction map of a recombinant phage \( \lambda D_3e \), which contains the rat \( c-mos \) sequence, and its subclones.
Figure 4.2 Rat liver cDNA probe hybridizes to the downstream region of \textit{c-mos}.

0.6 \mu g of \lambda D3e DNA was cleaved with the following combinations of restriction endonucleases and hybridized with a cDNA probe prepared as described in Chapter 2, Section 2.5.1:

- lane 1, \textit{EcoR I};
- lane 2, \textit{Hind III};
- lane 3, \textit{Eco RI + Hind III};
- lane 4, \textit{Pst I};
- lane 5, \textit{Pst I + Eco RI};
- lane 6, \textit{Pst I + Hind III};
- lane 7, \textit{Bgl II};
- lane 8, \textit{Bgl II + Eco RI};
- lane 9, \textit{Xba I};
- lane 10, \textit{Xba I + Eco RI};
- lane M, \lambda phage DNA digested with \textit{Pst I} and used as a size marker.

(a) The DNA fragments generated by the cleavage were fractionated on a 1\% agarose gel and visualized by ethidium bromide staining.

(b) A photograph of the X-ray film exposed to the membrane after hybridization with the cDNA probe.
Figure 4. 3 Subcloned flanking regions of *c-mos* hybridize with a rat cDNA probe.

Figures 4. 3 a and b. pHE42 was cleaved with the following combinations of restriction endonucleases: lane 1, *Pst* I; lane 2, *Pst* I + *Eco* RI; lane 3, *Bgl* II; lane 4, *Bgl* II + *Eco* RI; lane 5, *Bgl* II + *Eco* RI + *Pst* I; lane M, λ phage DNA digested with *Pst* I and used as a size marker.


(a, c) The DNA fragments generated by the cleavage were fractionated on a 1% agarose gel and visualized by ethidium bromide staining.

(b, d) A photograph of the X-ray film exposed to the membrane after hybridization with the cDNA probe.
Figure 4. A 2.7 kb \textit{Pst} I fragment hybridizes to variously-sized transcripts in rat tissues.

RNA fractions were prepared from the rat liver, testis and brain as described in Chapter 2, separated on a 1% agarose gel and transferred to a nylon membrane. The immobilized RNAs were hybridized with a DNA probe prepared from the 2.7 kb \textit{Pst} I fragment with the random priming method described in Chapter 2, Section 5.1. The RNAs loaded were as follows:

- lane 1, liver poly(A)$^+$ RNA fraction 5 µg;
- lane 2, liver total RNA 20 µg;
- lane 3, testis total RNA 20 µg;
- lane 4, brain total RNA 20 µg.

The positions of 18S and 28S RNA were determined by ethidium bromide staining of the liver total RNA electrophoresed on the same gel.
Figure 4.5 The 2.7 kb \textit{Pst} I DNA fragment hybridizes with rodent genomic DNA.

(A) Ethidium bromide-stained gel.

(B) X-ray film of the hybridization membrane.

The DNA fragments were fractionated on a 1\% agarose gel, transferred on to a nylon membrane and hybridized with a radioactively-labelled probe prepared from the 2.7 kb \textit{Pst} I fragment.

\textbf{Rat} \hspace{1cm} \text{lane 1} \textit{Eco RI};
\hspace{1cm} \text{lane 2} \textit{Hind III};
\hspace{1cm} \text{lane 3} \textit{Eco RI} \& \textit{Hind III};

\textbf{Mouse} \hspace{1cm} \text{lane 4} \textit{Eco RI};
\hspace{1cm} \text{lane 5} \textit{Hind III};
\hspace{1cm} \text{lane 6} \textit{Eco RI} \& \textit{Hind III};

\textbf{Human} \hspace{1cm} \text{lane 7} \textit{Eco RI};
\hspace{1cm} \text{lane 8} \textit{Hind III};
\hspace{1cm} \text{lane 9} \textit{Eco RI} \& \textit{Hind III}. 
Figure 4. Both DNA strands of the *Sac I-Eco RI* fragment hybridize with RNAs.

(A) Anti-sense probe.
(B) Sense probe.

Strand specific RNA probes were prepared as described in Chapter 2, Section 2.5.3.

(a) (anti-sense): An RNA probe synthesized with T3 RNA polymerase detects transcripts made from the opposite strand to which *c-mos* is transcribed.

(b) (sense): T7 RNA polymerase produces the RNA probe which detects transcripts with the same orientation as the *c-mos* transcripts

Rat RNA samples:

poly(A)- fraction: lane 1, testis  poly(A)+ fraction: lane 2, testis  lane 3, liver  lane 4, liver  lane 5, brain  lane 6, brain  lane 7, 20 μg of unfractionated murine testis RNA.
Figure 4.7 The *Sac I-Eco RI* fragment hybridizes with the 3' flanking region of the murine c-mos.

A plasmid clone, pMS2, which contains the murine c-mos coding region and its 3' flanking region, was cleaved with various combinations of restriction endonucleases and hybridized with a radioactively labelled probe prepared from the 0.8 kb *Sac I - Eco RI* fragment.

The restriction endonucleases used were as follows:

lane 1, *Bgl II*;

lane 2, *Bgl II + Hind III*;

lane 3, *Bgl II + Xba I*;

lane 4, *Xba I*;

lane 5, *Xba I + Eco RI*;

lane 6, *Sac I*;

lane 7, *Sac I + Eco RI*;

lane M, λ phage DNA cleaved with *Pst I* and used as a size marker.
Figure 4. 8 DNA nucleotide sequence of a 0.8 kb \textit{SacI-Eco RI} fragment; the \textit{SacI-Eco RI} fragment.

The number of nucleotides starts from the \textit{Sac} I site of the fragment, which is 5.7 kb downstream of the 3' end of \textit{c-mos} coding region. The sequence which is highly homologous with the murine B2 and rodent \textit{Alu} type-2 elements are underlined (236-400). The hatched underlined sequence shows the part of the sequence which is homologous to the rat ID sequence (431-511). The two sequences homologous with a split type of RNA polymerase III initiation consensus sequence are shown in boxes (237-248). The third box is an over-lapped polyadenylation signal (385-394). Another polyadenylation signal and TG-box identified are double-underlined(506-511, 555-559).
GAGCTCTCTG GTATCTCCTT TTTTTGAGGA GTTCCCGTTC ACCTTCCTCA
TCTAATCCCTG TTTCTAAATC CCAGTGGATT TCCCCACTAGC TAAGCAGCCT
GGAGGGGCTGG TTTCTCAGGG TCTTCTGGGT GTGTTAATAG ATCAGTTACC
TTTCATTTGGC CTTGGTGGTC GTGGGTCTGG ATTTGTACTG ATGCTAAAA
AGTGAGAGAT AAAAGTAAG GAACCTGGGCT GGAGAGATGG CTCAGTGGGT
AAGAGCAGCTG ACTGCTCTTC CAGAGGCTCT GGAGGAGE
CATGTTTTT CACAAACCAC TGTAATGAGA CCTAAACCT CTTCTGTGTTG
TTGAAAGACA GCTACAGTGT ACTTATATAT AACAAATAAA TAAAATCCTTT
TTTTATTTTT TTTTTCTGTT CTCTTTTCTG GAGCTGGGGA CCGAACCAG
GGCTTGCCG CTTTCCTAGGT AAGCCCTCTA ACCCGCTGAG TAAATCCTCA
GCCCAAAATA ATTTTTTTTT AAAGTGAAAT GAGGGAATTG AGCTCCATAA
AAGCTGTTGA TACACACACA CCTCTCTTAC CCCACAGTCC ACCCAGGGA
CTTAGTCCCA CAGCAGCTTG TAGGAGTACA ATCCTGCAAC CACTTCCACA
GTTTCACTGG CTGCAGAATT TCTCAAGACA GAACACAATT CTTACAAATG
AGACTCAAAA AGAGGAATAA TTGAATTCC
4. 3  Discussion.

4. 3. 1  Transcriptional units in the proximal region of the \textit{c-mos}.

One of the purposes of the experiments described in this chapter was to examine the hypothesis that the lack of expression of \textit{c-mos} in most tissues is due to the expression of a neighbouring transcriptional unit. This hypothesis predicts that a transcriptional unit should be detectable near to the \textit{c-mos} gene and that this transcriptional unit should be expressed in tissues that do not produce \textit{c-mos} transcripts and have a lower level of expression in cells that express \textit{c-mos}.

Hybridization of a cloned rat \textit{c-mos} genomic sequence with a cDNA probe revealed the presence of a nucleotide sequence that was transcribed into the poly(A)\(^+\) RNA fraction (Figure 4. 2). On the other hand, the \textit{c-mos} upstream region (0.8-4 kb upstream of the open reading frame) did not demonstrate the involvement in the transcripts (Figure 4. 2 & 4. 3). The DNA region hybridized with the cDNA probe was restricted and found to contain rodent specific repetitive sequences (5.8 kb downstream of the coding region) (Figure 4. 5). Nucleotide sequence determination of the region (\textit{Sac I-Eco RI} fragment) confirmed the presence of two repetitive elements which can be an independent transcriptional unit or a part of it (Figure 4. 8; Haynes & Jelinek, 1981). However, as these
sequences are located in various positions in the rat genome (Figure 4.6) it was not possible to determine whether the observed transcripts were generated from the region adjacent to the c-mos gene or from some other part of the genome.

Thus a possible transcriptional unit was found in the c-mos flanking region, however, the position of the unit was not in a proximity which was postulated to cause interference of the c-mos expression (ca. 2 kb; Naora & Deacon, 1982a). It should be noted that the method used to detect transcriptional units does not detect low levels of transcripts and thus intergenic effects have not been unambiguously ruled out as a regulatory influence for c-mos. Furthermore, even though the expression of the rat c-mos may not be repressed by a neighbouring transcriptional unit, this does not discount the possibility that the expression of some other genes is regulated in this manner.

4.3.2 Repetitive sequences in the downstream region of the c-mos.

The Sac I-Eco RI fragment contained repetitive sequences as it had various and multiple locations in the rat and murine genomic DNA (Figure 4.6). A portion of its nucleotide sequence was highly homologous with a B2 element and another portion was almost identical to the rat ID sequence (Figure 4.8). The B2 element is one of the major
repetitive sequences present in rodents, and is a member of a large
group of sequences named the B2 family (Krayev et al, 1982; Page et al,
element was initially found in mice (Krayev et al, 1982) and Chinese
hamster, where it was named the Alu type 2 sequence (Haynes & Jelinek,
1981). The rat ID sequence is a major repetitive sequence in rats and can
also be categorized into the B2 family (Rogers, 1985; Weiner et al, 1986).

The distinguishing characteristics of the B2 family are: (a) they
contain promoter sequences and termination signals for polymerase III
(Haynes & Jelinek, 1981); (b) their copy number in a haploid genome is
approximately 10^5, which is characteristic of middle repetitive
sequences (Kramerov et al, 1982); (c) their nucleotide sequence has
similarity with tRNAs (Rogers, 1985); (d) they have a direct repeat at
the ends of their units and may thus be mobile in the genome (Jelinek &
Schmidt, 1982; Rogers, 1985). The B2 repetitive sequence family is
involved in many transcripts in rodent cells (Figure 4. 6; Kramerov et
al, 1982; Georgiev et al, 1983; Ryskov et al, 1983), with a large
proportion of the B2 elements being observed in hnRNAs (Kramerov et
al, 1982) and some in mRNAs (Ryskov et al, 1983). The 3' region of
the murine major histocompatibility complex (MHC) genes (Kress et al,
1984; Mellor et al, 1984) and the untranslated 5' exon of a transcript of
the murine c-abl (Ben-Neriah et al, 1986) are examples of polymerase II
transcripts that contain B2 elements.
The murine *c-mos* locus hybridizes with a fragment of the rat *c-mos* flanking region which contains a rat ID sequence and a B2 element, suggesting that the murine *c-mos* flanking region probably contains a B2-like repetitive sequence. The murine *c-mos* locus also contains an LTR-like repetitive sequence and a cluster of B1 elements (Propst & Vande Woude, 1984; Krayev et al., 1980). The significance of the repetitive sequences that occur in the murine *c-mos* locus is largely unexplored, although the LTR-like sequence has been reported to have no influence on transcription in a transfection assay using NIH3T3 cells (Propst & Vande Woude, 1984).

### 4. 3. 3 Possible roles of the B2 elements in the *c-mos* transcription.

Several physiological roles have been proposed for B2 elements, although none of them have been unambiguously examined (Jelinek & Schmidt, 1982; Rogers, 1985; Weiner et al., 1986). One putative function of the B2 elements in the flanking region of a gene is induction of transcription (Murphy et al., 1983; Weiner et al., 1986). The B2 elements contain signal sequences required for transcription by RNA polymerase III and may also create a favorable circumstance for the polymerase II by changing the local structure of DNA. This hypothesis has been indirectly supported by observations that B2 elements are highly expressed when the transcriptional activity in a cell is elevated as occurs in embryos,
transformed cells and relatively undifferentiated cultured cells (Scott et al, 1983; Vasseur et al, 1985; Singh et al, 1985; Suzuki et al, 1986). However, it is not known whether the high incidence of the B2 elements, in the examples listed above, is a result or a cause of the stimulated transcription.

In relation to transcriptional regulation of a gene, the rat ID sequence has been the subject of two studies (McKinnon et al, 1986; Glaichenhaus & Cuzin, 1987). In order to assess whether a rat ID sequence can act as an activator, various cultured cells derived from rat or mouse were transfected by recombinant DNAs containing a reporter gene, a promoter and the ID sequence (McKinnon et al, 1986). The presence of the ID sequence in the constructs increased the rate of the transcription independently of the position or orientation of the ID sequence relative to the promoter-reporter gene combination. These results can be interpreted to imply that ID sequences are enhancers. However, when another ID sequence was linked with a globin gene, and studied in a run-on transcription assay, it increased the total amount of transcript generated, but not the rate of transcription (Glaichenhaus & Cuzin, 1987). It was therefore speculated that the ID sequence stabilizes the transcripts when the element was transcribed after the globin gene.

There are distinct differences in the amounts of c-mos transcripts produced in different species, with rats and mice producing more
transcripts than primates (Chapter 3; Propst & Vande Woude, 1984; Paules et al, 1988). The rat and murine c-mos loci both contain rodent-specific repetitive sequences (Figure 4.2 & 4.7b; Propst & Vande Woude, 1984) and it is possible that these sequences are acting as enhancers, resulting in rodents producing more c-mos transcripts than primates. The lack of information about the flanking region of the primate c-mos makes it difficult to assess this speculation. The question of whether or not the rat ID element or B2 element in the flanking region of rat c-mos stimulates the transcription of c-mos needs to be directly examined. It may be possible to do this by examining the expression of c-mos in cells that have been transfected with rat c-mos genomic clones in which one or both of the repetitive elements have been deleted. This experiment would be best done using a cell line that naturally expresses c-mos, but as such a cell line has yet to be identified it may be necessary to artificially induce c-mos transcription by deleting the part of the 5' flanking region which has previously been shown to have an inhibitory effect on the c-mos transcription (van der Hoorn et al, 1985).

B2 elements have also been postulated to be responsible for the generation of multiple transcription sizes from a gene. For instance, a B2 element is variably included in the K antigen mRNA of the MHC gene (Kress et al, 1984). Various transcripts of the c-mos gene are produced, the sizes of which differ between tissues and species (rat, Chapter 3; Propst et al, 1988; Leibovitch et al, 1987; mouse, Propst &
Vande Woude, 1985; Propst et al, 1987). The largest transcripts recorded are 4 kb (Chapter 3) and 7 kb (Propst et al, 1988) in the testis and this raises the possibility that these long transcripts contain long untranslated 3' regions (see Chapter 3, Section 3.3).

4.3.4 Summary.

The initial objective of this chapter was to examine whether \textit{c-mos} transcription may be repressed by the expression of a neighbouring transcriptional unit. Investigation of this question led to the discovery of a novel combination of two repetitive sequences, a B2 element and an ID sequence, in the far downstream of the \textit{c-mos} coding region. These sequences were involved in various RNAs prepared from the liver, testis and brain, regardless whether or not the \textit{c-mos} transcripts were detected.
CHAPTER FIVE

DNA METHYLATION OF THE RAT C-MOS REGION
5. 1 Introduction.

DNA methylation can act as an inheritable repressor of gene expression when the interaction between DNA signals and protein activators is inhibited by methylation of the regulatory sequences (Chapter 1, Section 1.3.1; reviewed by Razin & Riggs, 1980; Doerfler, 1983, 1984; Cedar, 1988). One of the biological roles of DNA methylation is to regulate tissue-specific gene expression and, as discussed below, this may be one of the factors that influences the transcription of c-mos.

5. 1. 1 Tissue-specific pattern of DNA methylation.

A relationship between DNA methylation and tissue-specific gene expression was first observed by Waalwijk and Flavell (1978) using the rabbit β-globin gene. The methylation of vertebrate globin genes has been extensively studied since then. An example is the human globin locus which contains 4 globin genes that are expressed at various stages of the development (van der Ploeg & Flavell, 1980). The locus contains 17 restriction cleavage sites which show tissue-specific patterns of the DNA methylation, when studied by combinations of methylation-sensitive and -insensitive restriction endonucleases (see Section 5.2 in this Chapter for a discussion of the usage of restriction endonucleases to detect DNA methylation). The tissues that produce globins appear to have lower
levels of methylation than those which do not produce globins. However, there are a few exceptions to this correlation. For instance, globin genes are not expressed in a few cultured cell lines and placenta, yet the globin locus of these cells are not methylated (van der Ploeg & Flavell, 1980).

For some genes there is no noticeable correlation between their methylation status and their tissue-specific expression (Doerfler, 1983; Bird, 1986). An example of this type of gene is the type 1 collagen α2 chain gene, which is transcribed in fibroblasts and forms a major extracellular protein in skin, bones and tendons. When the methylation pattern of the gene was analyzed by restriction endonucleases in collagen-productive and non-productive cells, the coding and 3' regions of the gene were observed to be heavily methylated and the 5' region was entirely unmethylated, irrespective of whether the gene was expressed (McKeon et al, 1982). Vitellogenin, albumin and α-fetoprotein genes are also transcribed even though they are heavily methylated (Gerber-Huber et al, 1983; Vedel et al, 1983).

These observations may be a reflection of the property of DNA signals which need to be unmethylated to bind with the protein factors, therefore the methylation in a coding region may only have a minor effect on the transcriptional activity. However, recent reports have revived the question about the importance of DNA methylation on the whole structure of genes. The necessity of forming a particular chromatin structure in
conjunction with DNA methylation is suggested for repression of the *tk* gene (Chapter 1. section 1. 3; Buschaeusen et al, 1987). Another example is a human γ globin gene whose transcriptional activity is not recovered by the demethylation of a DNA region containing the promoter (Murray & Grosveld, 1987).

From the studies discussed here and in Chapter 1, the methylation of particular sites in some genes seems to influence their transcriptional activity. At present, there is insufficient evidence to predict how methylation will affect the expression of a particular gene and each gene must be experimentally examined. To demonstrate that the tissue-specific expression of a gene is regulated by its methylation it is important to firstly show that there is an inverse correlation between methylation of particular sites within the gene and the expression of the gene and subsequently to demonstrate that experimentally altering the methylation status affects expression of the gene.

5. 1. 2 Methylation of the *c-mos* sequence.

The *c-mos* coding region is highly G+C-rich (68%) and is only expressed at low levels in a few tissues (Chapter 3). Therefore, DNA methylation has been postulated to contribute to the rather inactive nature of *c-mos* in transcription. There have been two publications supportive of this hypothesis (Gattoni-Celli et al, 1982; McGeady et al, 1983). The
genomic DNAs prepared from cultured rodent cell lines and murine tissues, which do not express \textit{c-mos}, are highly methylated as would be predicted for a repressed gene (Gattoni-Celli et al, 1982). Furthermore, when the \textit{Hpa} II and/or \textit{Hha} I sites in the cloned \textit{v-mos} are \textit{in vitro} methylated, the transforming activity of the \textit{v-mos} is reduced (McGeady et al, 1983). Inhibitory effects of the DNA methylation on gene expression is thought to be intermediated through (a) perturbation of binding between protein activators and their participant DNA signals (Kruczek & Doerfler, 1983) and/or (b) alteration of the chromatin structure of regulatory regions or the whole transcriptional unit (Buschhausen et al, 1987). Therefore, the DNA methylation in possible regulatory regions has been notified to be more influential in the gene expression than that in the rest of genomic regions. However, the methylation of the \textit{mos} coding region is suggested to be inhibitory to its transcription as the methylation of the \textit{v-mos} coding region containing dense population of the \textit{Hpa} II sites greatly reduced the transformation efficiency (McGeady et al, 1983).

The degree of methylation of the \textit{c-mos} region in cells which express the gene is not known and this information is needed to determine if there is an inverse correlation between the methylation of \textit{c-mos} and its expression. In this chapter, the methylation status of \textit{Hpa} II/\textit{Msp} I sites in the \textit{c-mos} region has been investigated using the DNAs prepared from rat testis and brain, which express \textit{c-mos}, and from the liver, which does not
produce detectable levels of c-mos transcripts (Chapter 3).


The ability of many restriction endonucleases to cleave DNA is inhibited by methylation of the nucleotide residues in their target DNA sequences (Bird & Southern, 1978). However, the binding and biochemical properties of some restriction endonucleases are not influenced in this manner. The difference in the sensitivity to methylation between various restriction endonucleases has been used to detect the methylation of specific sites in bulk DNA preparations. A methylated residue in a restriction enzyme cleavage site can be detected as an absence of cleavage by a methylation-sensitive restriction endonuclease and the presence of cleavage by its methylation-insensitive isoschizomer (Bird & Southern, 1978). A typical combination of restriction endonucleases used for this procedure is Hpa II and Msp I. Both enzymes recognize the sequence 5′-CCGG-3′, but Hpa II can not cleave DNA when the sequence is methylated at either or both of the C residues (CmCGG or mCmCGG). Msp I can cleave the sequence when the outer cytosine residue is not methylated (CmCGG). In combination with another restriction enzyme, the positions of the methylated cytosine residues can be assessed.
A limitation of the procedure using methylation-sensitive and -insensitive restriction endonucleases is that it does not detect all the methylated cytosine residues present in a DNA fraction as the enzymes only recognize particular sequences. This means that a certain portion of methylated residues are not detectable even though they may be functionally important. This problem can be overcome by using the combination of a sequencing method and Southern blot hybridization (Church & Gilbert, 1984). Genomic DNA fragments are cleaved by nucleotide-specific chemical reactions and detected by a hybridization with radioactively-labelled probes (Maxam & Gilbert, 1977; 1980; Church & Gilbert, 1984). Methylated cytidines do not react with the chemical reagents and thus can be detected as missing bands on sequencing gels. The presence of C-residues can be proven by nucleotide sequence determination of the counter DNA strand (Saluz et al, 1986). A technical problem of detecting DNA methylation in tissues is that all of the methods need a relatively large amount of DNA and it is difficult to analyze the DNA modification at the level of cell types.
5. 3 Results.

5. 3. 1 Methylation of $Hpa$ II/$Msp$ I sites in the rat $c-mos$ region.

A restriction map of the rat $c-mos$ with the positions of the 5'-CCGG-3' sequences ($Hpa$ II/$Msp$ I site) is shown in Figure 5.1. The $Hpa$ II/$Msp$ I sites in the proximal 5' flanking region are from the DNA sequence described in Chapter 3, whereas those in the 5' upstream region were determined by restriction enzyme mapping of the cloned DNA. The rat $c-mos$ coding region contains 7 $Hpa$ II/$Msp$ I sites (van der Hoorn & Firzlaff, 1984), many of which are located in the 3' portion of the sequence.

Genomic DNAs from the rat testis, brain and liver were prepared as described in Chapter 2. Twenty microgram each of the DNA samples were cleaved with either $Hpa$ II or $Msp$ I. The DNA fragments were separated on a 1% agarose gel, transferred to a membrane and hybridized with a DNA probe. The DNA probe was prepared from a plasmid clone, pHT22, which contains the $v-mos$ gene and its flanking region (Vande Woude et al, 1979). This probe was chosen to enable the methylation of the whole $c-mos$ coding region to be investigated with one DNA probe. The pHT22 DNA was cleaved by $Xba$ I to give a 2.2 kb fragment, which included the entire $v-mos$ sequence and a few hundred base pairs of the
flanking region.

Most of each of the DNAs prepared from the liver, testis and brain were not cleaved by *Hpa* II (Figure 5.2, lanes 4, 6 and 8). A 3.2 kb fragment was produced by *Hpa* II cleavage in all the three DNA samples and detected as a signal with a low intensity. This suggests that a minor portion of the DNAs have *Hpa* II-sensitive sites around their *c-mos* coding regions as the fragment size is longer than the coding region. A 3.9 kb fragment was also detected in the DNA prepared from brain (Figure 5.2, lane 8) which was not present in the liver or testis DNA (Figure 5.2, lanes 4 and 6). This indicates that there is a *Hpa* II site which is not methylated in a portion of the DNA derived from brain.

When the DNAs were digested by *Msp* I, three major bands and several weaker bands were detected by hybridization with the *v-mos* probe (Figure 5.2, lanes 5, 7 and 9). The sizes of the three major bands were 0.8, 0.9 and 1.1 kb, which were close to the entire length of the *c-mos* coding region (1.0 kb). If all the *Hpa* II/*Msp* I sites in the coding region were unmethylated then the fragment sizes would have been smaller than 0.3 kb (Figure 5.1). Thus some of the *Hpa* II/*Msp* I sites in the coding region are methylated at the first cytosine residue (mCCGG) and/or double modified (mCmCGG) and therefore can not be cleaved by *Msp* I. There was also a group of low intensity signals corresponding to
DNA sizes of 1.4, 1.9, 2.2, 2.6 and 5.6 kb. This group of signals may be generated from a minor portion of DNA with a high density of DNA methylation in the \textit{c-mos} coding region.

In the experiment using a multiple number of restriction endonucleases, 20 \( \mu \)g each of the DNA was cleaved with either \textit{Hpa} II or \textit{Msp} I and then cleaved by both \textit{Eco} RI and \textit{Hind} III. The DNA fragments produced were fractionated, transferred to a nylon membrane and hybridized with a \textit{v-mos} probe -the radioactively-labelled 2.2 kb \textit{Xba I-Xba I} fragment derived from pHT22. The reactions using \textit{Hpa} II produced a common DNA fragment with a size of 1.2 kb in all of the DNA samples prepared from the liver, testis and brain (Figure 5.3, lanes 1, 3 and 5). This DNA fragment is smaller than the \textit{Hind} III 1.4 kb fragment, which is calculated from a restriction map of the \textit{c-mos} region, indicating that one of the \textit{Hpa} II/\textit{Msp} I sites located in the 3' portion of the \textit{c-mos} coding region is not methylated in the three tissues.

When \textit{Msp} I was used instead of \textit{Hpa} II, three major and two minor DNA fragments were detected (Figure 5.3, lanes 2, 4 and 6). The three major fragments were 0.8, 0.9 and 1.1 kb, which are similar sizes to those detected by a single enzyme reaction with \textit{Msp} I. However, the intensities of hybridization signals were changed by usage of \textit{Eco} RI and \textit{Hind} III (Figure 5.2, lanes 6 and 7; Figure 5.3, lanes 4 and 6). The signal intensity of the 1.1 kb fragment decreased and the intensities of a group of
DNA fragments with sizes of approximately 0.8 kb increased (Figure 5.3).

The intensities of two minor fragments, with sizes of 0.55 and 0.65 kb, were stronger in the brain DNA than in the liver or testis samples (Figure 5.3, lane 2, 4 and 6). These fragments are probably derived from the 5' end of the c-mos coding region because of their sizes (Figure 5.1), suggesting that a Hpa II/Msp I site at the beginning of the coding region is non-methylated in a part of the brain DNA. These results indicate that both the number of methylated cytosine at a Hpa II/Msp I site and the sites of modification are various in the DNAs examined.

5.3.2 Presence of non-methylated Hpa II/Msp I sites in the 5' flanking region of the rat c-mos.

The methylation status of the 5' upstream region of the c-mos was examined with a similar procedure used for the coding region. As the 5' end of a poly(A)+ transcript was located about 1.2 kb upstream of the coding region (Chapter 3), the regulatory sequences for transcriptional initiation locate beyond this point. Therefore, the experiment was performed using a hybridization probe derived from a DNA fragment located far upstream of the c-mos coding region. A 0.7 kb Eco RI-Pst I DNA fragment, which is located at approximately 4 kb upstream of the c-mos coding region, was prepared from pEH32. The pEH32 contains a 3.2
kb Eco RI-Hind III fragment of the 5' flanking region of the rat c-mos.

Twenty microgram each of the testis, brain and liver DNAs were cleaved with Hpa II or Msp I, and then treated with a mixture of Eco RI and Hind III. The generated DNA fragments were fractionated, transferred to a membrane and hybridized with the radioactively-labelled Eco RI-Pst I fragment. Two common bands, whose sizes were 0.3 and 0.5 kb, appeared in all the DNA samples, indicating that two Hpa II/Msp I sites at 3.5 kb upstream from the 5' end of the c-mos coding region are non-methylated (Figure 5.1 & 5.4). One additional 1.2 kb fragment was detected in all the DNAs digested with Hpa II, Eco RI and Hind III (Figure 5.4, lanes 1, 3 and 5). This suggests that another Hpa II/Msp I site downstream of the DNA probe used for this experiment is also non-methylated. Although the DNA fragment patterns derived from the liver, testis and brain were similar to each other, the brain DNA also contained two minor fragments, sized 1.5 and 1.6 kb, that were not detected in the liver or testis DNAs (Figure 5.4, lane 5).
Figure 5.1 Position of *Hpa II/ MSP I* sites in the *c-mos* coding and upstream region.

The *Hpa II/ MSP I* sites in the coding region (shown as a box) are from a DNA sequence described by van der Hoom & Firzlaff (1984). A site immediately adjacent to the 5' end of the open reading frame is from the sequence described in Chapter 3 of this thesis. The rest of the sites were determined by restriction enzyme mapping of cloned DNAs. The abbreviations for restriction endonuclease cleavage sites are follows:

- **E**, *Eco RI*;
- **Hp/M**, *Hpa II/ MSP I*;
- **P**, *Pst I*;
- **H**, *Hind III*.

![Diagram of restriction sites in the c-mos region]
Figure 5.2 The *Hpa* II/*Msp* I sites in the testis and brain DNAs were methylated.

DNA samples were prepared from rat liver, testis and brain as described in Chapter 2. Twenty micrograms of each of the DNAs was cleaved by *Hpa* II (lanes 4, 6 & 8) or *Msp* I (lanes, 5, 7 & 9). The untreated (lanes 1, 2 & 3) and treated DNAs (lanes 4-9) were fractionated and hybridized with the radioactively-labelled *v-mos* 2.2 kb *Xba* I fragment.

Lanes 1, 4 & 5, liver;
lanes 2, 6 & 7, testis;
lanes 3, 8 & 9, brain.
Figure 5. 3  Tissue-dependent difference of methylation density.

Twenty micrograms each of the DNAs prepared from liver, testis or brain were cleaved with *Hpa* II (lanes 1, 3 & 5) or *Msp* I (lanes 2, 4 & 6) and then treated with a combination of *Eco* RI and *Hind* III. The DNA fragments generated by these digestions were then fractionated and hybridized with the *v-mos* 2.2 kb *Xba* I fragment. The experimental conditions for the hybridization, washing and exposure are as described in the legend to Figure 5. 2.

Lanes  1 & 2, liver;

lanes  3 & 4, testis;

lanes  5 & 6, brain.
Figure 5. 4 Unmethylated \textit{Hpa II}/\textit{Msp I} sites in the \textit{c-mos} upstream region.

The genomic DNAs prepared from rat liver, testis and brain were cleaved with \textit{Hpa II} (lanes 1, 3 & 5) or \textit{Msp I} (lanes 2, 4 & 6) and then cleaved with a combination of \textit{Eco RI} and \textit{Hind III}. The DNA fragments were separated and hybridized with the radioactively-labelled 0.7 kb \textit{Eco RI-Pst I} DNA fragment. This fragment was isolated from a subclone pEH32 which contained the 5'upstream region of rat \textit{c-mos} as described in the legend to Figure 4. 1.

- lanes 1 & 2, liver;
- lanes 3 & 4, testis;
- lanes 5 & 6, brain.
5. 4 Discussion.

5. 4. 1 DNA methylation of the c-mos region.

A highly methylated status of the c-mos region had been thought to inhibit its transcription. This hypothesis was based on investigations of several rodent cell lines and murine tissues which do not express the c-mos (Gattoni-Celli et al, 1982). A weak point in this hypothesis is that the extent of DNA methylation of the c-mos region in a tissue which expresses the c-mos had not been examined. In this Chapter, the methylation of c-mos in DNAs prepared from the tissues that have various levels of the c-mos transcription were examined and compared. Irrespective of the levels of c-mos transcription, the c-mos was highly methylated in all of the tissues examined (Figure 5. 2), although a tissue-specific variation was observed for the degree of modification (Figure 5. 2 & 5. 3). This lack of a clear correlation between the level of transcription and methylation status, could be used as an argument against c-mos expression being repressed by DNA methylation. Several points need to be considered, however, before such a conclusion could be drawn (see below).

Technical limit.

The procedure used in this Chapter only detects methylation of Hpa II/Msp I sites and not of other sequences such as 5'-CGCG-3' or 5'-GCGC-
This method was chosen so that the results obtained could be compared with previous studies which had shown c-mos to be highly methylated in cells that did not express it (Gattoni-Celli et al, 1983). Furthermore, the Hpa II sites are of particular interest as the in vitro transformation activity of v-mos is reduced when it is methylated by either a Hpa II-site-specific or Hha II-site-specific methylase (McGeady et al, 1983). The relevance of this observation to c-mos is not entirely clear as it is not known whether the reduced activity of the v-mos was due to methylation of sites in its coding and/or flanking region. The coding region of the v-mos is rich in Hpa II-sites, but there are also some sites in its flanking region. The coding regions of the c-mos and v-mos are highly homologous but their flanking regions are dissimilar (van Beveren et al, 1981b) and thus results obtained with v-mos are only applicable to c-mos if the important methylation sites are in the coding region and not in the flanking region.

**Tissues contain various cell types.**

Tissues contain mixtures of cell types, each of which has its own pattern of DNA methylation. When a tissue is examined for DNA methylation, the pattern that emerges is a reflection of cells that contain methylated residues in a DNA region and cells that do not. Under these circumstances, the various intensities of hybridization signals between two tissues may indicate that different proportions of the cells have the
methylated DNA region in one tissue compared with the other. This cell-type variety is also applicable to the detection of c-mos expression. The detection of low levels of transcription does not necessarily mean the same low level throughout cells in a tissue. The concentration of the c-mos transcripts in individual cells may be high in a tissue which gives the detection of a small amount as a whole. The results given by the analysis of a whole tissue may represent the events happening in a large proportion of the cells.

It may be possible to overcome the problem of cell type diversity by isolating various cell types from the testis by gravity centrifugation. The expression of c-mos in such isolated cell fractions could be measured (Goldman et al, 1987) and correlated with the methylation pattern of their c-mos region. This experiment is a good refinement of tissue-specificity analysis, however, it may not entirely eliminate the influence caused by a mixed population of cells as the cells can be physically similar but differ in their gene expression. For instance, liver parenchymal cells can be purified using a low speed centrifugation, but the purified cells have various amounts of albumin production when observed in a section (Brands et al, 1983). To assay cell-type specific DNA methylation, a new detection procedure at the level of in situ hybridization needs to be developed.
Position-dependent effect of the DNA methylation.

The similarity of the DNA methylation in the three tissues does not necessarily mean that methylation is unimportant for the control of \textit{c-mos} expression. The influence of methylation on transcriptional regulation can be dependent on the position of the modification and not just the presence or the density of methylation (Chapter 1, Section 1.3; reviewed by Doerfler, 1984; Cedar, 1988). In some cases methylation of a part of the promoter region inhibits the expression of following transcriptional units (Toth et al, 1989). Therefore, the DNA methylation of regulatory sequences in 5' regions may be particularly important, although this does not preclude the involvement of the methylation of the coding region in the transcriptional regulation.

The detection of methylation sites that effect the \textit{c-mos} expression could be done using cultured cells that produce detectable levels of the \textit{c-mos} transcripts. Methylation of DNA can work as a passive regulator of transcription, therefore, demethylation means release from inhibition and not necessarily activation. The assay system must be able to provide necessary protein activators to promote \textit{c-mos} transcription. The use of cultured cells which produce \textit{c-mos} transcripts has another advantage in that the cells being studied would be a homogeneous population. There is, however, the disadvantage that the results obtained with cultured cells may not entirely reflect the behavior of cells that express \textit{c-mos in vivo}. 
5. 4. 2 Various modifications observed between testis and brain.

The brain DNA contained a wider variety of modifications than that of liver and testis DNA. For instance, additional \( Hpa \) II sensitive sites were detected as minor bands at higher-molecular weight positions in the brain DNA (Figure 5. 2, lane 8) and after \( Msp \) I digestion, there were differences in intensity of restriction fragments between DNAs prepared from testis and brain (Figure 5. 2, lane 6). These observations indicate that there are brain-specific sites for the methylation of the \( c-mos \).

The heterogeneity of the methylation of \( c-mos \) may be caused by various density of DNA methylation within the \( Hpa \) II/\( Msp \) I recognition sequence (5'-CCGG-3') and/or may be derived from cell-type-specific differences in the sites of methylation (see Section 5. 4. 1). Which cell types express \( c-mos \) is not known in the rat brain and testis, the tissues studied here.

At present the significance of the difference in the minor methylation bands observed between the brain and testis is unknown. The sizes of the \( c-mos \) transcripts in these tissues are different (Chapter 3) and one possibility is that some of the methylation sites in the coding region influence the size of the transcripts. It has been recently reported that the transcription of the chicken \( \beta \)-globin gene starts actively from the
methylated and DNase I resistant form, however, the elongation was perturbed to produce truncated RNA molecules (Lois et al, 1990).

5. 4. 3 Unmethylated Hpa II sites in the 5' flanking region of c-mos.

Unmethylated Hpa II sites were detected upstream of the putative functional DNA regions that had been described in the past (Figure 5. 4; van der Hoorn & Neupert, 1986; van der Hoorn & Nordeen, 1989). The sizes of the DNA fragments detected (0.5 and 0.3 kb) and their unmethylated status in the three tissues tested suggest that CpG islands may occur upstream of the c-mos. The observations described in this Chapter are sufficient for their identification, however, to confirm their identity, two experiments can be performed. The first is determination of the DNA sequence of the far upstream region of the c-mos. A high-density of G+C residues is not always the property of a CpG island (Gardiner-Garden & Frommer, 1987), however, the nucleotide sequence should provide a knowledge of this region. The second is characterization of the region by Southern hybridization to the DNAs prepared from more tissues or from various developmental stages. One of the characteristics of the CpG islands is that their methylation status is constant throughout various tissues and developmental stages (Bird et al, 1985; Kolosto et al, 1986).
5. 4. 4 Summary.

The objective of this Chapter was to further the investigation of the possible role that DNA methylation has on \textit{c-mos} expression. Nine methylation sites were examined. Some of these sites were uniformly methylated in tissues with different levels of \textit{c-mos} expression. It is unlikely that these sites are involved in the regulation of \textit{c-mos} expression, although they could possibly have an effect on the elongation of the \textit{c-mos} transcripts. Other sites had tissue-specific differences in the extent of their methylation and the affect of methylation of these sites on \textit{c-mos} expression needs to be directly examined.
CHAPTER SIX

GENERAL DISCUSSION.
6.1 Introduction.

Since the *c-mos* was identified as a potential transforming gene, the major interest has been the regulation of its expression and biological properties in normal animals. Initially, it appeared not to be transcribed in normal animals, even though its coding region is highly conserved (Chapter 1; Frankel & Fishinger, 1976; Wood et al, 1982). The demonstration in this thesis, and by other investigators, that the *c-mos* is expressed in normal tissues changes the context in which the *c-mos* is being studied (Chapter 3; Propst & Vande Woude, 1985; Schmidt et al, 1988; Sagata et al, 1988; Paules et al, 1988). The *c-mos* can now be studied as one of many cellular genes, although it remains a very interesting gene to study as its expression is restricted and in some circumstances it can transform cells.

6.2 *C-mos* transcription in various species.

The species examined for *c-mos* expression include rat (Chapter 3; Leibovitch et al, 1987; Propst et al, 1988), mouse (Propst & Vande Woude, 1985; Propst et al, 1987), human, monkey (Paules et al, 1988) and *Xenopus laevis* (Sagata et al, 1988; Freeman et al, 1988). The amount of transcripts are commonly very low and sometimes only detectable with the most sensitive method. Its expression is tissue-specific; the *c-mos* is expressed at relatively high levels in gonadal tissues. These two
characteristics were the reason why the transcript-ion was not detected in the initial surveys. It should be noted that a mos-like sequence has been reported to be undetectable in the genomic DNA of a fish by a v-mos probe (Xenophorin; Vielkind & Dippel, 1984). However, it needs another survey to conclude that a mos-homologue does not occur in fish as the sequence homology may be low.

The tissue-specificity of the c-mos expression needs further examination. The c-mos expression has usually been examined in the liver, kidney, brain, testis or ovary and embryos of a species. Other tissues have rarely been investigated and it is possible that the gonadal tissues is not the only part of the body where the c-mos is expressed at a relatively high level.

Another characteristics of the c-mos transcripts is the variety of the sizes detected in various species, in contrast to the relatively constant size of its coding region (1.0 kb). This can be summarized with some other characteristics as follows:

(a) Xenopus laevis: A single ATG. The transcription is predominantly detected in oocytes. The only species in which cDNA clones have been prepared. The transcript size observed in the oocytes is 3.1 kb and its length is due to the presence of a long 3' non-coding region-up to 2 kb (Sagata et al, 1988; Freeman et al, 1989).
(b) Rodents: Multiple ATGs. Various 5' ends have been mapped, although the reason for their production is unknown. (Chapter 3; Wood et al, 1983b; van Beveren et al, 1981). The c-mos is expressed in both female and male gonadal tissues (Chapter 3; Propst and Vande Woude, 1985). The sizes of the transcripts have a wide range (less than 1 kb to 7 kb in rats (Chapter 3; Propst et al, 1988); 1.3 kb to 6 kb in mice (Vande Woude, 1985; Propst et al, 1987).

(c) Primates: Overlapping reading frames in the 5' part of the coding region (Watson et al, 1982; Paules et al, 1988). The testis produce transcripts which are shorter than the coding region and are derived from the 3' region of the major open reading frame (Paules et al, 1988).

The Xenopus c-mos has some distinctive characteristics; it has a TATA-box-like sequence in the 5' flanking region, which has not been detected yet in other species and it produces transcripts with a single 5' end and a long 3' non-translated region in contrast to the variable 5' ends detected in the c-mos transcripts of other animals (Sagata et al, 1988). It is also known that the Xenopus c-mos transcripts are produced in oocytes but not at a detectable level in the testis, which produces c-mos transcripts in other species (Sagata et al, 1988; Freeman et al, 1989). The characteristics of the c-mos transcripts produced in the normal tissues of various species may thus be a reflection of their gene structures.
6. 3 Transcriptional unit for c-mos expression.

The c-mos has been regarded as an intron-less gene according to the DNA sequences of the coding region and the proximal regions (Oskarsson et al, 1980; van der Hoorn & Firzlaff, 1984; reviewed by Vande Woude et al, 1987). However, this does not exclude the possible presence of introns outside of the DNA region whose sequence has been determined. Although neither an intron nor splicing of the c-mos has ever been detected (van der Hoorn & Firzlaff, 1984; Propst et al, 1987; Sagata et al, 1988), the possible presence of a non-translated exon beyond 1.2 kb upstream or 0.43 kb downstream of the coding region (Chapter 3) has not been eliminated. The various 5' ends detected in the rodent transcripts suggests that these are generated by multiple promoters or alternative splicing (Chapter 1, Section 1.2.2.3).

The estimation of size of a c-mos transcriptional unit is reasonably important in studying its expression. It can be used to identify a putative area in the genomic DNA region which may contain regulatory sequences for transcriptional initiation. From this point of view, the newly identified non-methylated Hpa II/Msp I sites located 3.5 kb upstream of the open reading frame (Chapter 5) may suggest the start of the rat c-mos transcriptional unit is present around this point. Although this observation can not be decisive in positioning the transcriptional unit, it has been demonstrated that CpG islands are often present around
transcriptional units (Chapter 1, Section 1.3.1.2; Bird, 1986; Gardiner-Garden & Frommer, 1987).

6.4 Regulation of \textit{c-mos} expression in rats.

The work described in this thesis has demonstrated that the rat \textit{c-mos} transcripts produced in the testis have a long leader sequence (Chapter 3). DNA sequence determination of a part of the leader sequence indicates another level of the regulation of the \textit{c-mos}. The \textit{c-mos} expression may be restricted not only at transcriptional level but also translational level (Chapter 3) as it has been demonstrated that multiple ATGs preceding a gene are generally inhibitory to translation (Kozak, 1986b). This feature may be common in the transcripts of protooncogenes and thus has been suggested to play a role in the translational regulation of the expression of proteins that could be hazardous if inappropriately expressed (Kozak, 1986b; 1987).

This finding makes it possible to exclude a part of the proximal DNA region from the survey of regulatory sequences for transcriptional initiation. As the testis transcripts have leader sequences of up to 1.2 kb, the putative regulatory sequences may be present upstream of this point. It has been postulated that the regulatory sequences may be 1-2 kb upstream from the open reading frame (van der Hoorn, 1987), however, this estimation was not based on the observed transcription. With the
transforming activity as measurement of the \textit{c-mos} expression, an inhibitory DNA element has previously been found at 1.8 kb upstream of the \textit{c-mos} (van der Hoorn et al., 1985). The element can be very proximal to or even inside of the \textit{c-mos} transcriptional unit from the results observed recently (Chapter 3; Propst et al., 1988). In conjunction with the detection of large-sized transcripts in the rat testis, the survey should include the far upstream region of the \textit{c-mos} (ca. 5 kb upstream) and even cloning of a DNA region beyond the 5' end of the present genomic clone.

A survey for a putative transcriptional unit revealed the presence of a combination of repetitive elements in the rat \textit{c-mos} region (Chapter 4). The presence of the repetitive elements may not be unusual as the sequences have a high incidence in rodent genomes (Jelinek & Schmidt, 1982; Rogers, 1985). However, the characteristics of the elements are rather interesting; they are both class specific repetitive elements and are known to be transcribed in rapidly growing cells or in the cells with relatively undifferentiated phenotypes (see Chapter 4). The question to be asked and particularly related to the \textit{c-mos} expression in rats is whether or not the repetitive elements are contained in the large-sized transcripts. As the distance between the \textit{c-mos} coding region and the elements is at least 5.6 kb, the short transcripts do not contain these elements. The long transcripts of the \textit{c-mos} need to be analyzed, possibly after amplification to compensate for the small amounts of these transcripts.
Another direction of the search for regulatory mechanisms of the c-mos expression has revealed that the c-mos coding region is heavily methylated regardless of whether the tissue studied produces c-mos transcripts. However, tissue-specific variations of the DNA methylation were observed (Chapter 5). Two lines of the further investigation were suggested as a consequence of the experiments described. The first is to use relatively homogeneous cell populations, such as cultured or fractionated cells, as the observed tissue-specific difference should be a reflection of cell-type specificities. In practice, the rodent testicular cells can be used for this purpose as the various cell types can be isolated and identified morphologically (Bellve et al, 1977). The c-mos transcripts have been detected in the testis (Chapter 3; Propst & Vande Woude, 1985) and the cell types that generate the highest amount of the c-mos transcripts are suggested to be the round spermatids (Goldman et al, 1987). However, DNA methylation status of the c-mos has not been studied.

The second question is to ask whether or not methylation of the c-mos coding region is causal to the various sizes and/or amounts of the c-mos transcripts, particularly the short transcripts. This could be done by using in vitro DNA methylation of the various regions of c-mos genomic DNA and transfection into cultured cells.
6.5  Biological properties of the c-mos product.

My interest, and this thesis, is primarily concerned with the regulation of gene expression, with c-mos being an interesting gene to study in this context. The advances in our knowledge of the transcription of c-mos raises new questions in fields allied to my interest. In particular, it is important to know whether the variously -sized transcripts give rise to different proteins and if they do, whether the different proteins have different biological properties. It is also important to begin to examine what the biological functions of c-mos are, particularly since mos is an oncogene and is thus likely to be involved in the regulation of vital cellular functions.

Homology searches for the mos amino acid sequence deduced from the nucleotide sequence demonstrated the presence of a protein domain common in kinases (Baker & Dayhoff, 1982). The v-mos product was produced in an in vitro system and found to have serine/threonine kinase activity (Maxwell & Arlinghaus, 1985). However, experimental observations directed toward understanding the biological property of the c-mos have only recently become possible, with the identification of cells that express c-mos (Vande Woude et al, 1987).

This question has now begun to be addressed using various approaches; studies of its transcription in genetically defective mouse
strains (Goldman et al, 1987; Propst et al, 1988b), \textit{in situ} hybridization to detect the transcripts (Mutter & Wolgemuth, 1987; Mutter et al, 1988), and \textit{in vitro} culture of amphibian and murine oocytes which have been injected with sense or anti- sense \textit{c-mos} transcripts (Mutter et al, 1988; Sagata et al, 1988; 1989a; O'Keefe et al, 1989). In the male murine germ cells, the \textit{c-mos} is expressed only transiently in postmeiotic spermatid (Goldman et al, 1987). In the female conterpart, particularly \textit{Xenopus} oocytes, where the \textit{c-mos} product appears to be essential for the oocytes to proceed from the G2 to the M phase of the second meiotic cell division (Sagata et al, 1988; 1989a) and also to the meiotic arrest (Sagata et al, 1988; Watanabe et al, 1989).

The experimental results described above suggest an interesting and fundamentally important question to ask; that is whether or not the biological properties of the \textit{c-mos} product are only limited to meiosis. One way to answer this question is to assay various cell types in one species and a similar cell type in various species for its expression. If the \textit{c-mos} expression is important to drive mitosis in somatic cells, it should appear in various cells, possibly in the actively dividing cell population. Studying various species is also important as the survey of the \textit{c-mos} expression is not thorough yet and the structure of the \textit{c-mos} is various between species, thus its products may have species dependent differences in the spectrum of biological effects.
REFERENCES


Lone, Y.-C., Simon, M.-P., Kahn, A. & Marie, J. (1986). Sequences complementary to the Brain-specific "identifier" sequences exist in L-type pyruvate kinase mRNA (a liver-specific messenger) and in transcripts especially abundant in muscle. J. Biol. Chem. 261, 1499-1502.


molecular mechanism for parental imprinting. Cell 50, 719-727.


Vinson, C.R., LaMarco, D.F., Johnson, D.F., Landschultz, W.H. & McKnight, S.L. (1988). In situ detection of sequence-specific DNA-


