108

T ANTIGEN BINDING SEQUENCES

IN CELLULAR DNA

VIESTURS SIMANIS

Department of Biochemistry,

Imperial College of Science and Technology,

London SW7 2AZ.

Submitted for the degree of Ph. D., September 1984.

ABSTRACT

The cloning and analysis of cellular sequences to which SV40 large T antigen is bound *in vivo* has been undertaken by the use of a novel approach, immunoselection cloning. Chromatin prepared from SV40 transformed cells was digested with restriction endonucleases, followed by immunoprecipitation of the fragmented nucleoprotein complexes with anti - T antigen, or control antibodies. The DNA in the immunoprecipitate was then purified and cloned in bacterial plasmid vectors for further study. The clones thus obtained were then analyzed for their ability to bind T antigen *in vitro* by the McKay immunoprecipitation assay, a DNA cellulose competition assay, and by DNAse I footprinting.

The efficiency and selectivity of the process was initially monitored by demonstrating a very low background of colonies containing plasmids with inserts in the absence of anti - T antigen monoclonal antibody, and by showing that the SV40 origin was present among the clones obtained by precipitation with anti - T antigen, but not control antibodies. Subsequent analysis indicated that an enrichment for R. elements had also occurred by this procedure.

A detailed analysis of three clones suggested that they all contain high affinity binding sites for large T antigen. The location of the site in one clone has been

identified by DNAse I footprinting. A comparison of the expression of some of the clones obtained in normal and SV40 transformed cells has shown that their expression in cytoplasmic polyadenylated RNA is altered, though direct regulation by T antigen remains to be demonstrated.

The strategy of immunoselection cloning therefore seems to represent a potentially useful approach to the study of sequences that DNA binding proteins interact with in vivo.

A rapid, high yield immunoaffinity purification procedure for SV40 large T antigen has also been developed which produces >98% pure, biologically active protein, as judged by its DNA binding properties and enzymic activity.

CONTENTS

	Page.
ABSTRACT	2
ACKNOWLEDGEMENTS	7
ABBREVIATIONS	8
LISTING OF FIGURES AND TABLES	10
CHAPTER ONE : INTRODUCTION	14
Section 1.1 : Levels of control of Gene Expression.	15
Section 1.2 : Models for control of Gene Expression.	23
Section 1.3 : Transcription Factors.	29
Section 1.4 : The interaction of steroid hormone receptors with chromatin and DNA.	37
Section 1.5 : Prokaryotic Gene Regulatory Proteins.	43
Section 1.6 : Simian Virus 40.	57
Section 1.7 : Small t Antigen.	60
Section 1.8 : Large T Antigen.	61
Section 1.9 : Interaction of Large T Antigen with the SV40 origin of replication.	63
Section 1.10: Identification of regions of T antigen involved in binding to DNA.	75
Section 1.11: Interaction of T antigen with the host cell.	82
Section 1.12: Interaction of T antigen with the host cell chromatin.	87
Section 1.13: Conclusions.	89
CHAPTER TWO : MATERIALS AND METHODS.	90
Section 2.1 : Eukaryotic Cells and Viruses.	91

Section 2.2 : Bacteriological Techniques. 94

Section 2.3 : General Techniques. 97

Section 2.4 : Nucleic Acids. 101

Section 2.5 : Electrophoretic Techniques. 112

Section 2.6 : Blotting Techniques. 117

Section 2.7 : DNA Sequencing. 122

CHAPTER THREE : CLONING OF CELLULAR SEQUENCES TO WHICH T ANTIGEN IS BOUND. 127

Section 3.1 : Introduction. 128

Section 3.2 : Materials and Methods. 139

Section 3.3 : Results. 142

Section 3.4 : Discussion. 162

CHAPTER FOUR : PURIFICATION OF T ANTIGEN. 167

Section 4.1 : Introduction. 168

Section 4.2 : Materials and Methods. 172

Section 4.3 : Expression of the SV40 Early Region in Saccharomyces cerevisiae. 183

Section 4.4 : Purification of T Antigen. 203

Section 4.5 : Discussion. 232

CHAPTER FIVE : ANALYSIS OF NON - ORIGIN CLONES I: CLONE p27. 237

Section 5.1 : Introduction. 238

Section 5.2 : Materials and Methods. 241

Section 5.3 : Screening of clones by DNA cellulose competition assay. 247

Section 5.4 : Clone p27. 249

Section 5.5 : Discussion. 274

CHAPTER SIX : ANALYSIS OF NON - ORIGIN CLONES II: CLONES p24 AND p29. 281

Section 6.1 : Introduction. 282

Section 6.2 : Materials and Methods. 282

Section 6.3 : Clone p24. 282

Section 6.4 : Clone p29. 301

Section 6.5 : Discussion. 311

CHAPTER SEVEN : GENERAL DISCUSSION 318

Section 7.1 : Regulation of eukaryotic gene expression. 319

Section 7.2 : Eukaryotic DNA binding proteins. 335

Section 7.3 : Immunoselection cloning. 339

Section 7.4 : General summary. 359

REFERENCES 361

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to my supervisor, David Lane, for his help, advice, tolerance and limitless patience during the course of this work and also for constructive criticism of this manuscript in draft form.

I am also indebted to Jean Beggs in whose laboratory I learned many techniques, and where part of the work presented in this thesis was done.

I am very grateful to Melanie Lee for her advice, encouragement and friendship, and for constructive criticism of this manuscript in draft form.

Sincere thanks are also due to Elizabeth Woodruff for her expert typing of this dissertation, and also for her patient and tireless proofreading.

I also wish thank all the members of the CRC group, past and present, for their help, advice and support, and for making my time here an experience I shall not readily forget.

This work was supported by an MRC studentship, and by the Cancer Research Campaign.

ABBREVIATIONS

ATP	Adenosine 5' triphosphate
dATP	2' deoxy adenosine 5' triphosphate
ddATP	2', 3' dideoxy adenosine 5' triphosphate
BSA	Bovine Serum Albumin
bp	base pairs
dCTP	2' deoxy cytidine 5' triphosphate
ddCTP	2', 3' dideoxy cytidine 5' triphosphate
cpm	counts per minute
DMEM	Dulbecco's modification of Eagle's medium
DMS	Dimethyl sulphate
DNA	Deoxyribonucleic acid
DNAse I	Deoxyribonuclease I
DTT	DL-Dithiothreitol
EDTA	Ethylene diamine tetraacetate
FCS	Fetal Calf Serum
dGTP	2' deoxy guanosine 5' triphosphate
ddGTP	2', 3' dideoxy guanosine 5' triphosphate
IPTG	Isopropyl ß-D Galactoside
kb	1000 base pairs
moi	multiplicity of infection
MOPS	3-N-Morpholino propane sulphonic acid
NP40	Nonidet P40
OD	Optical density
PEG	Polyethylene glycol
PIPES	1, 4 Piperazine ethane sulphonic acid

psi	pounds per square inch
RNA	Ribonucleic acid
mRNA	Messenger RNA
rRNA	Ribosomal RNA
tRNA	Transfer RNA
SDS	Sodium dodecyl sulphate
SV40	Simian Virus 40
TTP	2' deoxy thymidine 5' triphosphate
ddTTP	2', 3' dideoxy thymidine 5' triphosphate
TCA	Trichloroacetic acid
TEMED	N, N, N', N', tetramethylethylene diamine
Tris	tris(hydroxymethyl)amino methane
Triton X-100	octyl phenoxy polyethoxyethanol
U.V.	ultra violet
X-gal	5-bromo-4-chloro-3-indolyl-ß-D-galactopyranoside

LIST OF FIGURES. Page.

1.1 : Coordinate control of gene expression in the Britten - Davidson model. 27

1.2 : A schematic diagram of the SV40 genome. 59

1.3 : The interaction of Large T Antigen with the SV40 origin of replication. 67

3.1 : Strategy for Immunoselection cloning of DNA sequences to which T antigen is bound. 135

3.2 : Structure of the known SV40 inserts in the SV3T3 Cl 38 genome. 138

3.3 : Dot blot analysis of immunoselected DNA. 145

3.4 : Colony hybridization of immunoselected clones using an SV40 probe. 150

3.5 : SV40 origin clones, I. 153

3.6 : Southern blot of clones 106, 200 and 207. 155

3.7 : SV40 origin clones, II. 157

3.8 : Southern blot of clones 323 and 336. 159

3.9 : Proposed structure of SV40 homologous clones. 161

4.1 : Structure of SV40 - pJDB207 recombinants used in these studies. 185

4.2 : Analysis of DNA extracted from yeast transformants. 187

4.3 : Analysis of RNA extracted from yeast transformants: I. 190

4.4 : Analysis of RNA extracted from yeast transformants: II. 192

4.5 : Proteins synthesised in yeast transformants. 195

4.6 : Radioimmunoassay for small t antigen. 198

4.7 : Purification of SV40 large T antigen. 208

4.8 : Western blot of T Antigen purification. 210

4.9 : Titration of Rabbit anti D2 serum. 213

4.10 : Standard curve for Large T Radioimmunoassay. 215

4.11 : McKay immunoprecipitation assay using pure T Antigen. 221

4.12 : DNAse I footprint of the SV40 origin. 223

4.13 : Summary of data from fig. 4.12. 225

4.14 : Autophosphorylation of pure T Antigen. 231

5.1 : Restriction map of clone p27. 251

5.2 : McKay immunoprecipitation assay on p27. 253

5.3 : Footprinting of clone p27: I. 256

5.4 : Footprinting of clone p27: II. 258

5.5 : The DNA sequence of clone p27. 261

5.6 : Sequences homologous to p 27 within mouse DNA. 263

5.7 : Homology between p27, R. elements, and pG3-28. 266

5.8 : Colony screen of clones obtained using a p27 probe. 268

5.9 : Analysis of clones homologous to the p27 insert. 271

5.10 : Dot blot analysis of expression of clones homologous to p24 and p27. 273

6.1 : Restriction map and sequencing strategy for clone p24. 284

6.2 : McKay immunoprecipitation assay on clone p24. 287

6.3 : DNA sequence of clone p24. 289

6.4 : Sequences homologous to p24 in mouse DNA. 292

6.5 : Sequences homologous to p24 in cytoplasmic poly A+ RNA. 294

6.6 : Analysis of immunoselected clones homologous to p24. 297

6.7 : Alignments of p24 with sequences in the EMBL and GENBANK databases. 300

6.8 : Restriction map of clone p29. 303

6.9 : Sequences homologous to p29 in mouse DNA. 306

6.10 : McKay immunoprecipitation assay on clone p29: I. 308

6.11 : McKay immunoprecipitation assay on clone p29: II. 310

LIST OF TABLES

2.1 : Solutions used in tissue culture. 93

2.2 : Media used in bacterial growth. 95

2.3 : Some commonly used solutions. 100

2.4 : Restriction enzyme buffers. 102

2.5 : Buffers used in the electrophoresis of nucleic acids. 115

2.6 : Composition of protein gels. 116

2.7 : Solutions used in Maxam - Gilbert sequencing. 125

2.8 : Solutions used in dideoxy sequencing. 126

3.1 : Summary of data from immunoselection cloning experiments. 147

4.1 : Purification of Large T antigen. 218

4.2 : The ATPase activity of purified T Antigen. 227

To my parents

CHAPTER ONE

INTRODUCTION

CHAPTER ONE

INTRODUCTION

Two of the major unanswered questions remaining in molecular biology are: How is eukaryotic gene expression controlled, and What are the molecular interactions by which the control is exercised?

Section 1.1: Levels of control of gene expression.

That control is indeed exercised over gene expression is beyond doubt. Biochemical and genetic consideration (Bishop, 1974) have produced an estimate of 50,000 as the total number of genes expressed during the lifetime of a mammal. However, analysis of the complexity of cellular messenger RNA populations suggests that only a fraction of these genes are expressed in each cell type, typical numbers obtained being of the order of 8,000 - 17,000 with approximately 2% of the genome being transcribed into mRNA (Axel et al., 1976; Wold et al., 1978; see also Lewin, 1980, Chapter 24).

The two stages on which greatest attention has been focussed as possible points of control are the initiation of transcription and processing of RNA. Examples are known of control being applied at the level of initiation of transcription, splicing of precursor RNA molecules, selection of poly A addition sites and turnover of mRNA in

the cytoplasm. For a comprehensive review, see Brown (1980), and Darnell (1982).

There are many well documented cases of control of initiation of transcription, such as the switching of globin gene expression in chicken embryos (Groudine et al., 1982) and the regulation of ovalbumin and conalbumin in chick oviduct in response to steroid hormones (McKnight and Palmiter, 1979). These studies analyzed "runoff" transcription by endogenous RNA polymerase in isolated nuclei to synthesize labelled RNA which was then hybridized to filters bearing immobilized DNA of the gene under study. The studies of globin transcription using embryonic and adult gene probes to study transcripts made in adult and embryonic nuclei showed that little if any transcription from the adult or embryonic globin genes was detectable in the heterologous nuclei, suggesting that control was at the level of transcription. Similarly, administration of steroid hormones was found to result in at least a 20-fold increase in ovalbumin transcription. Similar techniques have been used to analyze temporal changes in transcription during adenovirus infection (for reviews see: Ziff, 1980; Tooze, 1981; Nevins, 1982). The late transcription unit for polypeptide IX lies within an early transcription unit, yet at early times in infection, no detectable expression from the polypeptide IX gene occurs. Later in infection (6 - 8 hours), the protein IX gene is actively transcribed; the expression of the overlapping early transcription unit is unaffected

(Petterson and Matthews, 1977; Wilson et al., 1979).

The studies outlined above have focussed chiefly upon the products of terminal differentiation, products made in response to hormonal stimulation or during viral infection, where one or a few mRNA species constitute the majority of the cellular mRNA. Analysis of bulk mRNA and hnRNA populations has shown that hnRNA has a greater sequence complexity than mRNA (reviewed in Lewin, 1980, Chapter 25). Cross hybridization studies of messenger RNAs from different tissues have revealed both quantitative and qualitative differences in the fraction of genes being expressed (Hastie and Bishop, 1976; Young et al., 1976; Galau et al., 1976).

It has been postulated (Davidson and Britten, 1979) that regulation of gene expression may occur by each cell type processing a qualitatively and quantitatively different subset of nuclear RNA to mature messenger RNA. These ideas are based on the analysis of RNA populations detailed above, and also studies on sea urchin and mouse cells, where low abundance messenger RNA sequences found on the polysomes of one tissue, but absent from those of another, are readily detectable in the nuclear RNA of both tissues. For example (Wold et al., 1978), comparison of RNA extracted from sea urchin blastula and intestine show that the complexity of the two messenger RNA populations changes considerably and that only 15% of blastula mRNA sequences are present in the intestine mRNA. However, it

is found that virtually all of the blastula mRNA sequences are represented in the adult intestine nucleus, suggesting that for at least some genes, control of expression depends on whether or not their transcripts are processed or selected for transport to the cytoplasm. Studies on RNA populations of tobacco tissues (Kamaly and Goldberg, 1980) have produced similar findings.

The apparent dichotomy between the findings that for bulk populations, control is applied at the level of processing or transport, while there are clear examples of control at the level of transcription, may be resolved by postulating that for proteins such as globin or ovalbumin, expressed at very high levels in terminal differentiation or in response to hormonal stimulation, efficient processing of RNA transcripts would be insufficient, requiring additional increases in the rate of transcription above the basal level for the majority of genes in all cells. Thus these genes are believed to represent a special class, which because of high production levels, are regulated in a different manner to the majority of cellular genes. Examples for individual transcription units of control of gene expression at the level of RNA processing are also known.

A comparison of the transcripts initiating at the major late promoter and terminating at the L1 polyadenylation site at early and late times during adenovirus infection (Nevins and Wilson, 1981; Akusjarvi and Persson, 1981) has shown that at early times during

infection the primary transcript from this transcription unit is processed to give a single mRNA of 4kb, whereas at late times after infection, three mRNAs of 4.3kb, 3.8kb, and 2.3kb are produced from this region. S1 mapping has shown that the 4kb early mRNA has the same body as one of the late species, but has an additional 5' leader segment in the tripartite leader, which alters the translation of the RNA. Analysis of pulse labelled cytoplasmic and nuclear RNAs both early and late in infection has demonstrated that differential splicing rather than differential message stability or transport is the reason for production of a single early L1 mRNA.

Detailed analysis of immunoglobin gene transcription has shown that the choice of polyadenylation site can determine whether a primary transcript from a single gene will encode the secreted or membrane bound form of IgM (Rogers et al., 1980; Early et al., 1980). Use of the five polyadenylation sites in the Adenovirus major late transcription unit varies throughout infection (Nevins and Darnell, 1978; Nevins, 1979; Nevins and Wilson, 1981). At early times in infection, the L1 polyadenylation site is used twice as frequently as the L2 site, while late in infection the reverse is true, with the L2 site being utilized more frequently than the L1 site.

During the synthesis of the mRNA to encode either secreted or membrane bound IgD, the entire coding sequence for IgM heavy chains must be excised from the precursor

RNA. This process necessitates selection of one set of splice junctions over another in that those splice junctions which are used to form the mature Cμ gene transcript are not used in the production of the mature C_δ encoding mRNA (reviewed in Wall, 1980; Blattner and Tucker, 1984), though it has not yet been shown whether the control is applied at the level of splicing, or in the choice of 3' end favoring a particular structure in the RNA in turn favoring the use of some splice junctions over others. Premature termination of transcription, analogous to attenuation control of biosynthetic operons in prokaryotes (for review see Yanofsky, 1981), has been proposed as a control mechanism for gene expression during VSV infection (Iverson and Rose, 1981) and SV40 late gene expression (Hay et al., 1982).

mRNA turnover has been implicated as a control mechanism to explain the changes in the relative levels of Adenovirus Elb region transcripts during infection (Wilson and Darnell, 1981), and has also been shown to be the mechanism by which yeast compensates for the presence of extra copies of histone genes (Osley and Hereford, 1981).

Control at the level of translation has been shown to operate during the heat shock response (Storti et al., 1980; Bienz and Gurdon, 1982), in which the translation of the majority of cellular mRNAs is greatly reduced after heat shock, while those coding for certain "heat shock" proteins is unimpaired. Yeast ribosomal protein production also appears to be controlled at the level of

translation (Pearson et al., 1982). In some Adenovirus-SV40 hybrid viruses it has been found that the amounts of SV40 large T antigen produced depend on the position of the SV40 insert within the messenger RNA (Thummel et al., 1983). A construction which fuses the SV40 early region to the third leader segment produces 20-fold more T antigen than similar constructs with only one or two leader segments. Equal amounts of SV40 mRNA were made from all constructs, and all mRNAs were translated in vitro with equal efficiency, again suggesting some form of translational control in vivo. An additional regulatory mechanism, thus far only demonstrated in prokaryotes, controlling translation by production of an antisense RNA transcript off the noncoding strand has recently been discovered by Mizuno et al., (1984).

The chromatin structure of active (or potentially active) genes differs from that of inactive ones (for a review, see Lewin, 1980; Weisbrod, 1982). The degree to which the genes are methylated varies (Razin and Riggs, 1980), and they are found to be more sensitive to digestion with DNAse I in tissues where they are expressed (Weintraub and Groudine, 1976; Lawson et al., 1980; Lawson et al., 1982). These changes are necessary but not sufficient for transcription as hematopoietic cells not yet induced to undergo erythroid maturation have a globin gene in the "active" conformation yet only make high levels of RNA when induced to differentiate (Miller et

al., 1978). In cells not destined to express globin such as adult liver or fibroblasts, no increase in sensitivity to DNAse I is observed. Reconstitution experiments involving mixing plasmids carrying the adult ß-globin gene, histones and nuclear extracts from erythrocyte nuclei result in formation of a DNAse I hypersensitive region in the 5' flanking region of the gene. No such site is conferred by extracts made from cells in which ß-globin is not expressed. The factor responsible has been partially purified, and binds to DNA fragments containing the hypersensitive region (Emerson and Felsenfeld, 1984). Clearly, further studies of this type will lead to identification of those proteins conferring the "active" state upon a chromatin domain. Comparison of chromatin structure of expressed and silent MAT (mating type) genes on plasmids in yeast (Abraham et al., 1983) has indicated that plasmids bearing active genes are less supercoiled than ones bearing silent copies, suggesting that the transcribed gene has a more open structure.

Actively transcribed genes are found associated with the nuclear matrix (Berezney and Coffey, 1974; Ciejek et al., 1983; Hentzen et al., 1984). Thus, the ovalbumin gene is only found at the matrix when the oviduct tissue is stimulated by steroid hormones. Only the transcribed regions of the 100kb DNAse I sensitive ovalbumin gene chromosomal domain are found associated with the nuclear matrix suggesting it is the site of nuclear transcription. This therefore represents yet another potential level at

which regulation of gene expression may occur.

Control of gene expression by transposition of previously unexpressed DNA segments to "expression sites" has been demonstrated for yeast mating type genes (Klar et al., 1982; Nasmyth et al., 1981) and also for trypanosome variant specific glycoprotein genes (reviewed by Borst et al., 1983).

Section 1.2: Models for control of gene expression.

Models for control of eukaryotic gene expression have generally represented a re-formulation of the classical model of the bacterial operon, proposed by Jacob and Monod (1961). Thus, the model of Britten and Davidson (1969), further elaborated by Davidson and Britten (1973, 1979) proposes the existence of four classes of DNA. Producer genes encode proteins, and are considered directly analogous to the structural genes of bacterial operons. These genes have adjacent to them one or more receptor elements, which through interaction with activator molecules permit transcription of the producer gene to occur. A second set of genes, known as integrator genes, encode the regulatory molecules which interact with the receptor sites. Transcription of these loci is controlled by adjacent sensor elements, with the gene being transcribed only when its controlling sensor element is activated. The model proposes that integrator genes

encode activator RNAs which interact with receptor sites, though formally the activator molecule could also be a protein.

The interaction of the four types of eukaryotic DNA sequence therefore constitutes a system of positive control, in which genes are inactive unless specifically activated in contrast to the system of negative control, as seen in the E. coli lactose operon (for review see Miller and Reznikoff, 1978) when genes are active unless specifically repressed. Integrator genes are analogous to regulator genes in bacterial operons. Sensor elements are proposed to constitute the targets which are recognized by agents which alter the pattern of gene expression, such as hormone-protein complexes for instance, and thereby provide a means for the specific control of gene expression in response to changes in the environment of the cell.

To account for co-ordinated control of many genes in different states of differentiation it is proposed that receptor and integrator elements are repeated many times in the genome, thus permitting a small number of controlling elements to activate a large number of genes. Hence, though structural genes whose products serve related functions are not tightly linked together in operons, possession of a common receptor sequence would permit their coordinate regulation.

Redundancy of receptor elements or integrator genes would permit activation of a particular gene in response

to differing stimuli, with either multiple adjacent receptor elements permitting response to more than one activator molecule, or multiple groups of integrator genes under the control of different sensor elements serving to modulate large changes in gene expression (Figure 1.1).

Repetition of receptor elements and integrator genes would immediately suggest a control function for part of the repetitive sequence component of the eukaryotic genome.

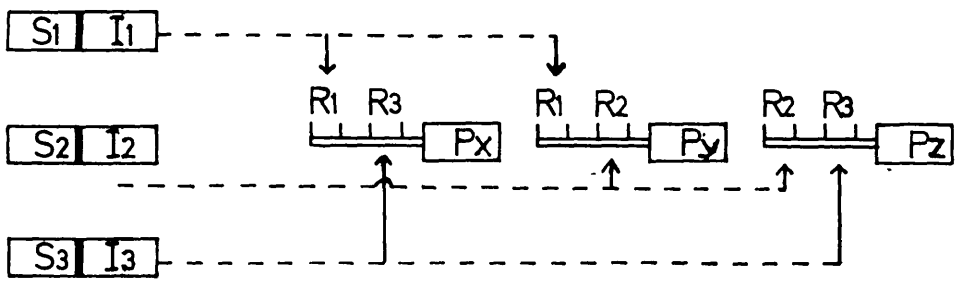
It has thus been postulated (Davidson and Britten, 1979) that the products of some integrator genes, activator RNA molecules, recognize particular hnRNA molecules in the nucleus, and that the formation of the RNA duplex would in some manner provide a signal for RNA processing or transport, thereby accounting for observations such as those of Wold et al., (1978) discussed earlier. Control of some transcription units by activator proteins rather than RNAs cannot of course be ruled out.

There is evidence in support of such a model for eukaryotic gene control. In Aspergillus there is genetic evidence for the existence of an integrator gene (Arst, 1976). Repeated sequences have been found associated with genes expressed only in one tissue. Thus brain-specific genes have so-called identifier sequences in their introns (Sutcliffe et al., 1982, 1984), and a conserved sequence has been identified upstream of immunoglobulin genes,

Figure 1.1: Coordinate control of gene expression in the Britten - Davidson model.

Diagram showing how coordinate control of gene expression might be achieved by redundancy of receptor and / or integrator genes.

Redundancy in Receptor Elements



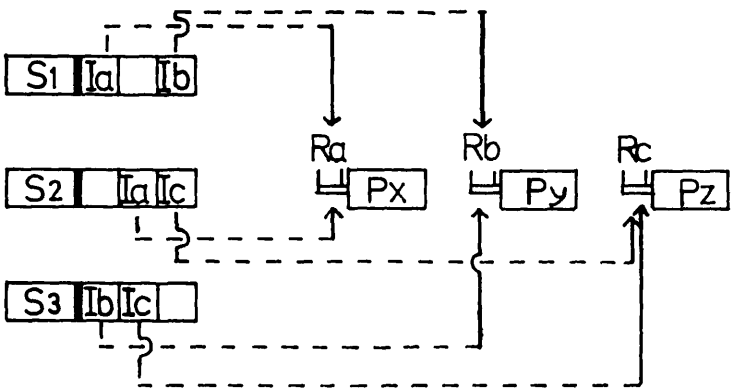
S: Sensor element

I: Integrator gene

R: Receptor element

P: Producer gene

Redundancy in Integrator Genes



necessary for their expression (Parslow et al., 1984; Falkner and Zachau, 1984). The role of short repeated sequences in control of gene expression has been reviewed by Davidson et al., (1983).

RNA polymerase II transcriptional control regions consist of a variety of functional domains (for review see Lewin, 1980). Upstream of some promoters, so called enhancer elements have been identified which potentiate transcriptional activity by an unknown mechanism. They function independently of orientation within a transcription unit, and can act either 5' or 3' to a gene. Initially, enhancers were described in viral systems such as SV40 (Banerji et al., 1981; Benoist and Chambon, 1981; Fromm and Berg, 1982) and Adenovirus (Gaynor and Berk, 1983; Hearing and Shenk, 1983). Similar sequences have since been isolated from human (Conrad and Botchan, 1982) and rat (Fried et al., 1983) DNA.

A tissue specific enhancer element, functional only in mouse B cells but not in fibroblasts has been identified in the major intron of a rearranged immunoglobulin heavy chain gene (Gillies et al., 1983; Banerji et al., 1983; Queen and Baltimore, 1983). It has also been found that the functional activities of different enhancers depend upon the host into which they are introduced (Laimins et al., 1982; De Villiers et al., 1982). This host and tissue dependence could either be a result of primary sequence differences, or it could involve specific cellular factors recognizing the enhancer

elements. Using a co-transfection *in vivo* competition assay, Scholer and Gruss (1984), have obtained evidence for the existence of cellular components which interact with enhancers. A possible candidate for such proteins might be those isolated by Nordheim *et al.*, (1982) which bind selectively to Z DNA, which is a component found in enhancer elements (Nordheim and Rich, 1983). Thus, activation of a particular cellular enhancer element could also provide a means for controlling gene expression. Recent work on the expression of insulin and chymotrypsin genes in different cell types (Walker *et al.*, 1983) has shown that cell specific expression appears to be controlled by elements in the 5' flanking region of the genes though the nature of the elements has not been defined.

Section 1.3: Transcription factors.

Pelham (1982) has described a regulatory upstream promotor element in the *Drosophila* hsp 70 Heat Shock gene. Two recent studies have demonstrated the interaction of proteins with this region. Fractionation of a *Drosophila* RNA polymerase II transcription system (Parker and Topol, 1984a,b) has separated various fractions containing transcription factors necessary for accurate initiation *in vitro*. A heat shock specific transcription factor (HSTF) together with two other factors, one of which binds to the

TATA box and leader region (factor B) have been identified. HSTF seems to be necessary for transcription of the hsp 70 gene, but not other genes such as histone genes, and footprint analysis has shown that it binds to the regulatory region defined by Pelham (1982). Wu (1984) has adopted a different approach, using in situ analysis of chromatin structure by resistance to exonuclease III digestion. It was found that two regions at the 5' ends of heat shock genes were resistant to digestion by the enzyme, suggesting that they have protein bound to them. These two binding sites (I and II) were mapped precisely, and it was found that site I lay over the TATA box region of the promoter, while site II was found to be coincident with the upstream regulatory region. In non-heat shocked cells protein was only bound at site I, while upon heat shock, protein was also found bound at site II. A model therefore emerges for regulation of the heat shock genes, which proposes that the TATA box region binding protein(s) establishes the transcriptional potential of heat shock genes, and that this potential is realized during heat shock activation by the additional binding of a second protein(s) to the upstream regulatory region. Thus HSTF is a positive regulatory molecule whose binding upstream of the hsp 70 gene potentiates its transcription. Taken together, these studies provide excellent evidence for regulation of at least some genes by sequence specific DNA binding proteins. It is of interest to note that another protein fraction with heat-shock locus specificity has

been identified by Jack et al., (1981). By fractionating nuclear extracts, and analyzing them by filter binding and protein blotting (Bowen et al., 1980), a set of three polypeptides have been identified which bind approximately 1 kb upstream of the presumptive transcription start of the hsp 70 gene. The relationship of these proteins to other factors that have been identified, and their role in the heat shock response remain to be determined.

Fractionation of RNA polymerase II in vitro transcription system has identified a number of other factors (Matsui et al., 1980; Tsai et al., 1981; Samuels et al., 1982; Slattery et al., 1983) required for accurate initiation. Some of these act to suppress random or nick induced transcription, the latter being identified as poly (ADP-Ribose) polymerase, while others mediate correct initiation, binding to the template in a sequence specific manner. Davison et al., (1983) have identified a transcription factor from HeLa cells which binds to the TATA box region of the conalbumin and adenovirus major late promoters [possibly analogous to factor B, identified by Parker and Topol (1984a)]. Very excitingly, Dynan and Tjian (1983a) have identified a transcription factor from HeLa cells, Spl, which discriminates between promoters, stimulating transcription from the SV40 early promoters, but having a slight inhibitory effect on the adenovirus major late promoter. This transcription factor has been shown to bind to sequences in the SV40 early promoter,

within the tandem 21 bp repeats located 70 - 110 bp upstream of the transcription initiation sites (Dynan and Tjian, 1983b). These findings suggest that Spl is involved in promoter selection, rather than general stimulation of transcription, and furthermore that there should be a class of cellular promoters with which Spl interacts. In this context, it is interesting to note that sequences isolated from the monkey genome having homology to the SV40 control region (McCutchan and Singer, 1981; Queen et al., 1981) have promoter activity (Saffer and Singer, 1984). These fragments contain sequence motifs found within the SV40 21 bp repeat region; a study of the interaction of Spl with these cellular promoters would clearly be a worthwhile exercise.

RNA polymerase III transcribes small RNAs in the cell, such as 5S RNA, tRNA, and adenovirus VA RNA. A similar fractionation approach to that used for identifying RNA polymerase II transcription factors has been successfully applied to pol III in vitro transcription systems. Three factors, in addition to RNA polymerase III have been identified, called Tf III A, B and C (Segall et al., 1980). Two of these, B and C, are necessary for transcription of tRNA and VA genes, while a third additional component, Tf III A, is required for transcription of 5S rRNA genes. The interaction of these components with their genes in the formation of preinitiation complexes has recently been reviewed (Lasser et al., 1983) and will not be discussed further here.

Much attention has however been focussed on the interaction of Tf III A with 5S RNA and 5S RNA genes. While factors III B and III C have not yet been purified, in the case of Tf III A, a single protein of 38,000 daltons has been purified to homogeneity and shown to interact, in the absence of other factors, with the internal control region of 5S RNA genes, thereby indicating its key role as an initiation factor (Engelke et al., 1980; Sakonju et al., 1981; Sakonju and Brown, 1982).

Binding to the internal control region has been shown to be cooperative, requiring two molecules of Tf III A to saturate the gene (Hanas et al., 1984). This finding has however been disputed by the recent study of Smith et al., (1984), who have convincingly demonstrated that only a single molecule of TF III A binds to each gene. The discrepancy is thought to have arisen due to differences in the methods used by the two groups. This latter study also identified functional domains within the protein. Mild proteolysis has shown that the protein consists of three structural domains: a large 20 kd domain at one end, and two 10 kd domains at the other. The 20 kd domain recognizes the nucleotides at the 3' end of the control region, but will not by itself activate the gene for transcription. An adjacent 10 kd domain, in conjunction with the 20 kd domain, extends binding to the 5' end of the control region, but supports transcription only

weakly. The third domain, which does not bind to DNA is responsible for most of the transcription enhancing activity of the protein. The factor also interacts stoichiometrically with 5S RNA in Xenopus oocytes and this together with the demonstrated potential of 5S RNA to inhibit factor III A function in vitro suggests that in the cell, 5S gene transcription in the presence of limiting amounts of this factor may be subject to autoregulation (Honda and Roeder, 1980; Pelham and Brown, 1980). Protein fractions have also been identified from yeast which will interact with the bipartite promoter of tRNA genes (Klemenz et al., 1980; Ruet et al., 1984; Stillman and Geiduscheck, 1984).

In contrast to the results seen with pol II and pol III in vitro transcription systems, it has been found that RNA polymerase I in vitro transcription system will only transcribe rDNA genes of the species from which the extract was derived, but not genes of a heterologous species. Recent fractionation studies of pol I transcription systems have revealed that three fractions are required for transcription, one of which forms a stable preinitiation complex with rDNA promoters from homologous but not heterologous species (Miesfield and Arnheim, 1984).

Various other proteins or protein fractions with potential regulatory activity, some of which have demonstrable DNA binding activity, have been described from various sources. A study of yeast mutants defective

in mating type interconversion has delineated a sequence within the MAT locus that is essential if switching is to occur, which seems to be the site at which a double stranded cut is made (Weiffenbach et al., 1983). A search for an enzyme that performs this function (Kostriken et al., 1983) has revealed two endonucleases: YZ endo, which cleaves within the MAT locus, and is only present in yeast strains that are undergoing mating type interconversion, and a second endonuclease recognizing a different sequence, present in all yeast strains known as Sce II. A study by Nasmyth (1983) has again demonstrated a site specific cut within MAT and demonstrated that the endonuclease responsible is either encoded or regulated by the HO gene.

The protein product of the nit 2 gene of Neurospora crassa, which controls a number of unlinked genes involved in nitrogen metabolism, has been shown to be a DNA binding protein, which can be eluted from DNA cellulose columns by glutamine, which is the repressor metabolite in whose presence the genes are not expressed (Grove and Marzluf, 1981). This immediately suggests a positive control mechanism of the type discussed earlier, where the nit 2 gene product binds at a regulatory sequence near each nitrogen related structural gene to activate its expression. Glutamine would then displace the positive control factor from the regulator site to turn off transcription of the genes coordinately.

Various proteins of unknown function have been isolated from Drosophila which show sequence specific DNA binding. A protein DB-1 has been purified and shown to bind to nucleolar DNA (Weideli et al., 1978) while a second protein DB-2 binds to a sequence found at a single locus in Drosophila DNA (Weideli et al., 1980). Proteins from both Drosophila (Hseih and Brutlag, 1979) and mouse (Avila et al., 1983) have been described which preferentially bind satellite DNA. Antisera raised against protein fractions released from Drosophila nuclei by DNAse I digestion (Silver and Elgin, 1977; Mayfield et al., 1978) has shown that one protein fraction was found bound at only a few loci, mostly at puffs active during the third larval instar suggesting that the antibodies were raised against some form of regulatory protein(s). The relationship between these proteins and the other Drosophila DNA binding proteins discussed has not to my knowledge been established. However it is noteworthy that the antiserum obtained does not stain heat shock loci until after the cells have been heat shocked.

Filter binding has demonstrated specific protein-DNA interactions around the chicken lysozyme gene (Nowock and and Sippell, 1982). The nature of all the proteins involved is not known, but recent work (Borgmeyer et al., 1984) has shown that extracts of oviduct nuclei contain a protein which recognizes the sequence TGGCANNNTGCCA and has high affinity for sequences in the flanking region of the lysozyme gene. The protein is found in many other

tissues however, and its function is at present unclear.

A protein which will induce differentiation in macrophages and granulocytes has been shown to be a DNA binding protein (Weisinger and Sachs, 1983), and at least one of the retroviral oncogenes encodes a product which is a nuclear DNA binding protein (Donner et al., 1982).

Recently, Nuclear factor I, a 47 kilodalton protein required for initiation of adenovirus DNA replication in vitro has been shown to bind in a sequence specific manner to the origin of replication of adenovirus DNA (Nagata et al., 1983), and a DNA binding protein has been isolated from human cells which will stimulate the activity of DNA polymerase α (Biamonti et al., 1983).

Section 1.4: The interaction of steroid hormone receptors with chromatin and DNA.

Much work has focussed on the mechanisms by which steroid hormones influence gene expression. Steroid hormones enter the cell by diffusion, and bind to a specific class of proteins, the steroid receptors, activating them. These hormone-receptor complexes are found in the nucleus of the cell leading to the suggestion that they were binding to DNA and thereby altering the pattern of gene expression in the cell (Yamamoto and Alberts, 1976).

The original model for the mode of action (reviewed by Clark, 1984) suggested that the hormone enters the cell and complexes with a cytoplasmic, unfilled receptor molecule. This causes a conformational change in the receptor, which activates it, and permits it to bind to and enter isolated nuclei. Thus, steroid filled receptors can exist only in the nucleus. Subsequent studies demonstrated the existence of unfilled receptors in the nucleus, leading to the suggestion that an equilibrium exists between unfilled nuclear and cytoplasmic receptors. Recent work has suggested that the concept of the cytoplasmic receptor is incorrect and due to solubilization of receptors during purification of nuclei. The first approach used (Welshons et al., 1984) was cytochalasin B induced enucleation of cells followed by assay for receptor in the cytoplast and nucleoplast fractions. This indicated very little hormone binding activity within the cytoplast fraction, with most of the unfilled receptor residing within the nuclear fraction. The alternative approach (King and Green, 1984) employed immunocytochemical staining rather than hormone binding to detect receptor in the cell, and again found that specific staining was localized to the nucleus only in all tissue types tested.

This therefore leads to a reformulation of the classical model, in which all receptor molecules are found in the nucleus, and hormone enters the cell, migrates to the nucleus, where it binds to unfilled receptor,

activating it. Thus the activation process does not, as in previous models, permit transport to the nucleus, but alters the affinity of the receptor for some nuclear component, most likely a specific binding site on chromatin. Which of these models applies remains to be determined.

Most work on interaction of receptor-hormone complexes with DNA has focussed on the progesterone receptor of chick oviduct, and the glucocorticoid receptor. The progesterone receptor is a heterodimer of two polypeptides of 110,000 daltons, A component, and 117,000 daltons, B component (Kuhn et al., 1975). DNA binding activity resides in the A component though both subunits will bind hormone. The purified receptor will bind to nuclei and chromatin.

The receptor has a preference for A - T rich DNA, and causes increased sensitivity of double stranded DNA to S1 nuclease suggesting that it destabilizes the DNA helix. The protein also shows a marked preference for binding to single stranded DNA (Hughes et al., 1981).

In vitro studies of the interaction of the progesterone receptor with hormonally regulated genes such as ovalbumin by a variety of techniques such as filter binding (Compton et al., 1982, 1983) and DNA cellulose competition assays (Mulvihill et al., 1982) have identified a fragment at the 5' end of the ovalbumin gene to which the receptor binds. There seem to be a number of

receptor interaction sites in this region, located 100-300 bp upstream from the site of transcription initiation. At least one of these sites contains an A - T rich region of DNA. Comparison of the sequences at the interaction sites with those found upstream of other hormonally regulated genes has lead to the derivation of a consensus sequence for the binding of the progesterone-receptor comples (Mulvihill et al., 1982). Deletion analysis of sequences at the 5' end of the ovalbumin gene required for hormonal regulation in vivo (Dean et al., 1983) is in broad agreement with these studies, defining a region 90-220 bp upstream of the transcription initiation site as necessary. A conserved nine residue sequence that is found upstream of six steroid hormone regulatable genes lies within this region, and the progesterone receptor has been shown by footprint analysis to interact with sequences 150-195 bp upstream of the transcription start site (cited in Dean et al., 1983). Recent studies of receptor DNA interaction (Compton et al., 1984) have shown that binding to DNA is cooperative, and also that it causes a change in DNA conformation, consistent with the earlier result suggesting helix destabilization (Hughes et al., 1981). Thus, it has been postulated that the receptor affects gene expression by causing local unwinding of the helix, to create a single stranded entry site for RNA polymerase thereby permitting, perhaps together with other factors, transcription of the gene. The proof of such a mechanism will clearly await

development of a suitable *in vitro* transcription system.

The glucocorticoid receptor, a 90,000 dalton protein, has also been shown to bind to DNA, chromatin and isolated nuclei (Simons *et al.*, 1976). Studies on mutant receptors isolated from hormone resistant cell lines have indicated that the receptor hormone complex has a reduced affinity for both specific and non-specific DNA sequences (Pfahl, 1982). A similar effect is observed if the DNA binding of receptor complexed to antagonists is studied (Bourgeouis *et al.*, 1984). In both cases however, the receptor retains the ability to discriminate between specific and non-specific DNA sequences. The biological effects of these mutations and antagonists may therefore be mediated by the reduced overall DNA binding affinity.

Transcription of the Mouse Mammary Tumor Virus (MMTV) genome is influenced by glucocorticosteroid hormones, and it has been shown that the sequences required for hormonal control reside within the viral genome (Ucker *et al.*, 1981). Thus, many studies have been undertaken to define the nature of the interaction between the viral DNA and the hormone receptor complex. Analysis by filter binding (Payvor *et al.*, 1981) and DNA cellulose competition assays (Pfahl, 1982) demonstrated that there were high affinity binding sites for the receptor-hormone complex in or near the MMTV promoter.

In vivo studies using deletion mutants (and subfragments of the promoter region) to identify the

element necessary for responsiveness to glucocorticoids (Buetti and Diggelmann, 1983; Chandler et al., 1983; Hynes et al., 1983) have delineated a region 100-200 bp upstream of the transcription start site that is required, in agreement with the in vitro studies mentioned above.

Very excitingly, it was found that the glucocorticoid response element was capable of acting at a distance from a promoter, affecting both proximal and distal promoters (Hynes et al., 1983) and in an orientation independent manner (Chandler et al., 1983) suggesting that it has enhancer like activity. Sequences related to the viral enhancer core consensus sequence (Weiher et al., 1983) are found within this region.

Footprint analysis (Payvar et al., 1983) has defined five binding sites for receptor in MMTV DNA, of variable size, both upstream of and within the transcribed regions. The physiological significance of the sites in the coding region is unclear, but the one upstream of the initiation site coincides with the glucocorticoid response element. Comparison of the various sequences involved lead to the definition of a degenerate octanucleotide sequence recognized by receptor. Footprinting studies of the promoter region (Scheidreit et al., 1983) coupled with methylation interference experiments (Scheidreit and Beato, 1984) have defined a similar protected region and a consensus hexanucleotide sequence TGTTCT which is also part of the octanucleotide defined by Payvar et al., 1983. A similar element has been found in the glucocorticoid

response element in the metallothionine IIA gene (Kanin et al., 1984).

The binding of the hormone-receptor complex to the glucocorticoid response element region together with the finding that this region has enhancer like activity raises the exciting possibility that steroid hormones modulate gene expression by activating cellular enhancer sequences, thereby potentiating transcription. Whether or not such enhancer elements are found upstream of other hormonally regulated genes is not known. Studies on the sequences necessary for hormonal control of the chicken lysozyme gene (Renkawitz et al., 1984) have shown that a region 220 to 140 bp upstream confers hormonal control by both dexamethasone and progesterone. Whether the regulatory sequences recognized by the two hormones overlap, or if the region has hormone dependent enhancer activity was not determined. Comparison of upstream regions from hormonally regulated genes (Davidson et al., 1983) has identified a short repeated element in the sequence required for induction. Its significance remains to be determined.

Section 1.5: Prokaryotic gene regulatory proteins.

From the foregoing discussion, it is clear that the study of eukaryotic sequence specific DNA binding proteins is still in its infancy. Whilst many cellular

subfractions containing protein components with regulatory properties, such as transcription factors, have been identified, often the proteins involved have not yet been identified; in addition, biological regulatory molecules are also likely to be present only in small quantities within the cell, making purification and biochemical analysis difficult. Only a few molecules regulating eukaryotic gene expression have been purified to near physical homogeneity, notable examples being Tf III A, the steroid hormone receptors, and the large T antigen of SV40, which will be discussed in greater detail later. None of these proteins has yet been crystallized so a complete molecular description of their interaction with DNA is not possible.

Analysis of the interaction of prokaryotic gene regulatory proteins with their DNA sequences have progressed much further. Studies have been principally focussed on the lactose operon repressor molecule, and cyclic AMP receptor protein of *E. coli*, and the *cI* and *cro* repressor proteins encoded by bacteriophage λ, although the interactions of many other proteins such as RNA polymerase (Siebenliest and Gilbert, 1979; Siebenliest *et al.*, 1980) and restriction endonucleases with DNA have also been analyzed.

The lactose operon repressor molecule binds as a tetramer to the operator of the *lac* operon and in doing so sterically inhibits RNA polymerase from binding to the promoter and forming an open complex (Reznikoff *et al.*,

1974; Majors, 1975; for review of lac promoter region, see Reznikoff and Abelson, 1978; for general review of lac operon, see Miller and Reznikoff, 1978). A similar mechanism has been proposed for the action of the trp operon repressor (Squires et al., 1975). The physical interaction of lac repressor with lac operator DNA has been analyzed in great detail (for review, see Barkley and Bourgeois, 1978). Filter binding studies (Riggs et al., 1970) have determined that the repressor has a very high affinity for operator DNA, the affinity constant K_d being of the order of 10^{12} M. Inducers of the lac operon such as allolactose (the natural inducer, or IPTG, the inducer most often used in these studies) decrease this affinity by a factor of 10^3, but do not affect the affinity of lac repressor for non-operator DNA. Binding of repressor protects 27 bp of operator DNA from digestion by DNAse I (Gilbert and Maxam, 1973). The cumulative results obtained from studies of the base changes in operator constitutive mutants (repressor binds the mutant operator with decreased affinity), the effects on binding of substituting nucleotide analogues in the operator, and methylation protection and phosphate alkylation interference studies have led to a description of which nucleotides in the operator interact with repressor (Goeddel et al., 1978). The protein makes contacts with seven phosphate residues, and ten purine residues, decreasing methylation at seven, increasing it at three,

over a span of 17 residues in the central region of the operator. Both the N-3 of adenine and N-7 of guanine residues are protected from methylation, indicating both minor and major groove contacts, respectively, are made by the repressor. Analysis of the phosphate contacts suggests that the protein binds to the operator along only one face of the helix. On binding to the operator, it has been shown that lac repressor unwinds the DNA by 90° (Wang et al., 1974).

A variety of experimental approaches have been adopted to try to define a mechanism by which lac repressor finds its target in the E. coli chromosome. Studies have been undertaken using the filter binding method of Riggs et al., (1970), coupled with DNA cellulose chromatography (de Haseth et al., 1977) and gel electrophoresis of nucleoprotein complexes (Fried and Crothers, 1981).

The problem facing the lac repressor is that of rapidly locating a single copy of the lac operator, approximately 20 base pairs in length, in the E. coli chromosome and distinguishing it from the remaining approximately 10^{7} base pairs of the genome. Theoretical considerations for the diffusion controlled, one step association of lac repressor and operator (Berg et al., 1981; Winter et al., 1981b) give a maximum value for the association rate constant as 10^{7} $M^{-1}s^{-1}$, while measured values in vitro exceed this by a factor of about 10^{3} (Riggs et al., 1970; Barkley, 1981; Winter et al., 1981b). Thus

some mechanism must exist by which lac repressor reduces the dimensionality or volume of the search process.

Lac repressor has a measurable affinity for non-operator DNA (Lin and Riggs, 1975; de Haseth et al., 1977) so it has been proposed (Barkley, 1981; Berg et al., 1981; Winter et al., 1981a,b) that the actual kinetic mechanism must involve a two step process, whereby the initial encounter between protein and DNA is the diffusion controlled formation of a complex between repressor and a non-specific DNA site, followed by a facilitated search process involving transfer to other non-specific sites until the operator is located. Two mechanisms have been proposed for this facilitated transfer of non-specifically bound repressor from site to site. The first possibility is sliding, and is defined as the diffusion of non-specifically bound repressor along the DNA in a one dimensional random walk, while the alternative is an intersegment transfer mechanism involving rapid and direct inter- (or intra-) molecular transfer of repressor from one DNA molecule to another, involving a transient intermediate where the repressor is bound to two DNA segments.

Experiments have been devised and undertaken to distinguish these possibilities. There is strong evidence (Winter et al., 1981a,b; reviewed by Berg et al., 1982) that for lac repressor in vitro in dilute solution, the sliding model of searching is applicable. Analysis of the

interaction of lac repressor with non-operator DNA molecules has indicated that the binding is purely electrostatic in nature, with approximately eleven charge-charge interactions per repressor tetramer (Rezvin and von Hippel , 1977). Thus the repressor is viewed as sliding along the DNA helix, displacing the cation in front of it, which is then replaced by one binding to the DNA behind the repressor as it moves on. Hence no net change in the number of ions involved in the interaction occurs, so there is no thermodynamic barrier to the repressor sliding along the DNA if the replacement of ions is rapid compared to the sliding rate. The repressor therefore slides over the DNA on what may be considered an isopotential surface. This raises the question of how the repressor recognizes the operator sequence when it encounters it. A study of the specific interaction (Record et al., 1977) indicates that in contrast to binding to non-operator DNA, binding to operator DNA involves only seven, plus or minus one, charge-charge interactions (a figure in excellent agreement with the study of alkylation of phosphate residues by ethylnitrosourea, which indicates that binding of repressor to operator protects seven phosphate groups from alkylation). In addition to electrostatic interactions the repressor forms a specific matrix of hydrogen bonds with the operator leading to much tighter binding. It is therefore proposed that the repressor can exist in two conformations, a non-specific DNA binding conformation in which the interaction is purely

electrostatic, and a specific binding form in which fewer electrostatic interactions are coupled to other forces to give much tighter binding. Thus, on encountering the operator, it is proposed that the repressor remains in the now more favored specific binding conformation. Such a model requires that interconversion of the two forms is very rapid. Kinetic measurements indicate that such interconversions do indeed occur (Berg *et al.*, 1982).

While it is clear that *in vitro* the sliding search mechanism predominates, it is worth considering what mechanisms might operate *in vivo*. This problem has been addressed by Lin and Riggs (1975) and Winter *et al.*, (1981b), who calculate that under the conditions found in the *E. coli* nucleoid body where the DNA concentration is sufficiently high that the affinity for non-operator DNA becomes a significant factor, the vast majority of the repressor in the cell is bound to non-operator DNA, with very little free repressor. Also, since stretches of naked DNA, that is to say DNA not complexed with proteins or involved in higher order structure, are likely to be short *in vivo*, it seems reasonable to propose that *in vivo* both sliding and intersegment transfer are involved in searching for the operator sequence, since the distance a protein is able to slide under these conditions is limited.

The *lac* repressor has thus far not been crystallized, so a description of which part of the protein interacts

with DNA *via* this route has not been possible. However, analysis of mutant repressor proteins which are defective in specific DNA binding, or have been chemically modified, for instance by alkylation or iodination, has indicated that the N-terminal sixty amino acids form the operator binding domain (Miller, 1978).

λ repressor and *cro* bind to the same tripartite operator in the λ genome, controlling the expression of two promoters (for review of λ repressor and *cro* function see Ptashne *et al.*, 1980; Johnson *et al.*, 1981). Both proteins bind as dimers; however it is found that while *cro* binds non-cooperatively, the binding of λ repressor is highly cooperative, the binding of the first dimer greatly enhancing the binding of a second. Methylation protection studies have indicated that on binding, the N-terminal arms of the repressor wrap around the operator DNA, completely encircling it (Pabo *et al.*, 1981).

In addition to repressing transcription from the P_R promoter when bound at operator sites 1 and 2, λ repressor activates transcription from the P_{RM} promoter, which is that for the C_I gene. The repressor therefore acts as a positive regulator of its own transcription. The suggestion was made that it mediated this effect through a direct interaction with RNA polymerase, stimulating open promoter complex formation (Johnson *et al.*, 1981). Subsequent work (Hochschild *et al.*, 1983; reviewed by Ptashne, 1984) has confirmed this by the isolation of mutants defective in positive control which are found to

be localized on the surface of the protein just where contact with RNA polymerase would be expected. Second site mutations in RNA polymerase have been found which will complement this defect.

The λCII protein, which is a positive regulator required for establishing transcription of repressor following infection, probably acts by facilitating RNA polymerase binding to a promoter which lacks the normal RNA polymerase recognition sequences. Analysis of its binding to the promoters in λ whose transcription it stimulates shows that in keeping with a model of direct interaction with RNA polymerase, it binds to the opposite face of the helix from that bound by RNA polymerase (Ho et al., 1983).

The affinity of λ repressor for its principal binding site is approximately 300 fold lower than that of lac repressor for lac operator. Despite this lower binding constant, the λ repressor is just as efficient as lac repressor in regulating expression from the promoter it controls (Johnson et al., 1981). Firstly, it is present at 10-fold higher concentration in the cell than lac repressor, and its much lower affinity for non-operator DNA ensures that in vivo, non-specific DNA is not an effective competitor for operator binding. Therefore, in contrast to the situation with lac repressor where an appreciable fraction of the repressor is sequestered on non-specific DNA (Lin and Riggs, 1975), little or none of

the λ repressor is bound to non-operator DNA. Thirdly, as a consequence of the cooperative nature of the binding to O_R1 and O_R2, the repression is more efficient. Thus, a repressor concentration only twelve fold above K_d is required for almost 100% repression, while a similar degree of repression by lac repressor needs a concentration 300 fold above K_d (Johnson et al., 1981).

Unlike the lac repressor, the λ represor, λ cro protein and the catabolite activator protein of E. coli have all been crystallized and their structures determined (Pabo and Lewis, 1982; Anderson et al., 1981; Ohlendorf et al., 1982; McKay and Steitz, 1981). All three of these proteins seem to interact with their target sequence on DNA in a similar manner (Sauer et al., 1982; Matthews et al., 1983; Takeda et al., 1983). Each protein contains a pair of α - helices, one of which fits into the major groove of the DNA helix resulting in the formation of contacts between exposed functional groups on the bases and amino acid side chains. A comparison with a variety of other DNA binding proteins indicated that the pair of helices involved in λ repressor and cro binding are also found at analogous positions in many other phage regulatory proteins, and also the lac repressor, suggesting that this structural motif is conserved and funtionally significant.

These models for DNA binding have been based on model building and fitting the protein into the helix of the DNA. Only recently has a co-crystal of a regulatory

protein bound to its target sequence been grown and analyzed (Anderson et al., 1984). Although the structure has not been solved to high resolution, preliminary findings indicate no significant distortion of the helix occurs, and that binding is to B form DNA, as predicted from the model studies.

The interaction of catabolite activator protein with DNA has been much studied, since the protein is a positive and negative regulator of many catabolic operons (see Miller and Reznikoff, 1978; de Crombugghe and Pastan, 1978) and also autoregulates its own synthesis (Aiba, 1983). The protein regulates its own synthesis not by blocking the binding of RNA polymerase (as found with other repressor proteins), but by binding within the leader sequence of the gene and either preventing a bound RNA polymerase from initiating transcription, or by blocking elongation of the transcript. Equilibrium and kinetic studies of the binding to DNA (Fried and Crothers, 1984a,b) using a gel electrophoretic technique have shown that in the absence of cyclic AMP, the protein binds non-specifically, and cooperatively to DNA. In the presence of cAMP, the CAP dimer binds non-cooperatively to the specific sequences it recognizes. Only a single molecule of cAMP binding per dimer is required for this to occur, indicating that the two subunits are non-equivalent, suggesting that there are significant deviations from the twofold symmetry that are seen with λ repressor and cro

bound to DNA. A coupled sliding and intersegment transfer model for binding site location similar to that for the lac repressor has been proposed to operate in vivo, though in vitro, it appears that intersegment transfer plays a significant role (Fried and Crothers, 1984b).

It has been shown that CAP binding to DNA introduces a bend into the DNA at the binding site (Wu and Crothers, 1984). Several mechanisms have been considered for the regulatory effects of CAP on transcription (reviewed by Adhya and Garges, 1982). It is possible that by inducing a conformational change in the DNA the protein creates a preferential entry point for RNA polymerase or that the bend in the DNA allows the CAP protein to interact directly with the RNA polymerase in some way. However, the precise mode of action of CAP still remains to be determined.

The sequence specific bacterial restriction endonucleases are another family of DNA binding proteins that have been much studied (for review see Yuan et al., 1981). The crystallization of EcoRI enzyme with a synthetic oligonucleotide containing its recognition sequence and solving of the structure (Frederick et al., 1984) has shown that the sequence specificity is determined by a tight complementary interface between the major groove of the DNA and α-helices of the two protein subunits. The DNA is found to contain a number of torsional kinks, introduced by the protein binding to the major groove.

A genetic approach to the study of which residues in a protein are important in recognition has been adopted by Youderian et al., (1983), who showed a single amino acid change in the mnt repressor of Salmonella decreases its affinity for wild type operator 1000 fold. This amino acid change also altered the binding specificity allowing the mutant protein to recognize a mutant operator with 1000 fold greater affinity than wild type repressor.

Studies on the int protein of λ (Ross et al., 1979; Ross and Landy, 1982; Ross and Landy, 1983) have shown that it can interact specifically with two different types of sequence on λ DNA, in contrast to most other sequence specific DNA binding proteins which recognize only one consensus sequence.

Thus, it is clear that prokaryotic regulatory proteins bind to defined DNA sequences, generally upstream of the genes they are controlling, and that the sequences are recognized by a combination of electrostatic interactions and hydrogen bond network formation. The proteins probably locate their target sequences in vivo by a facilitated search mechanism involving one dimensional sliding or scanning along the DNA coupled with intersegment transfer between regions of DNA molecules, which greatly increases the rate at which such sequences are located compared with simple diffusion mediated encounters.

Once bound to their recognition sequence, the

proteins may control gene expression in a variety of ways, preventing transcription either by preventing RNA polymerase from binding to promoter or elongating transcripts, as seen with the operon repressor molecules, and CAP in its negative regulatory mode, or potentiating transcription by direct interaction with RNA polymerase, as found in the case of the CI and CII gene products of λ, and probably CAP as well. The proteins when bound frequently alter the local structure of the DNA, unwinding it partially, as seen with lac repressor, or "bending" it, as recently observed with CAP. Clearly, further studies will be required to determine the physiological significance of in vitro phenomena such as bending, and of course a complete molecular description of the interaction of a regulatory protein awaits the solving to high resolution of protein - DNA crystals. The extent to which such conclusions may be applied to eukaryotic regulatory proteins will be discussed later.

Section 1.6: Simian Virus 40.

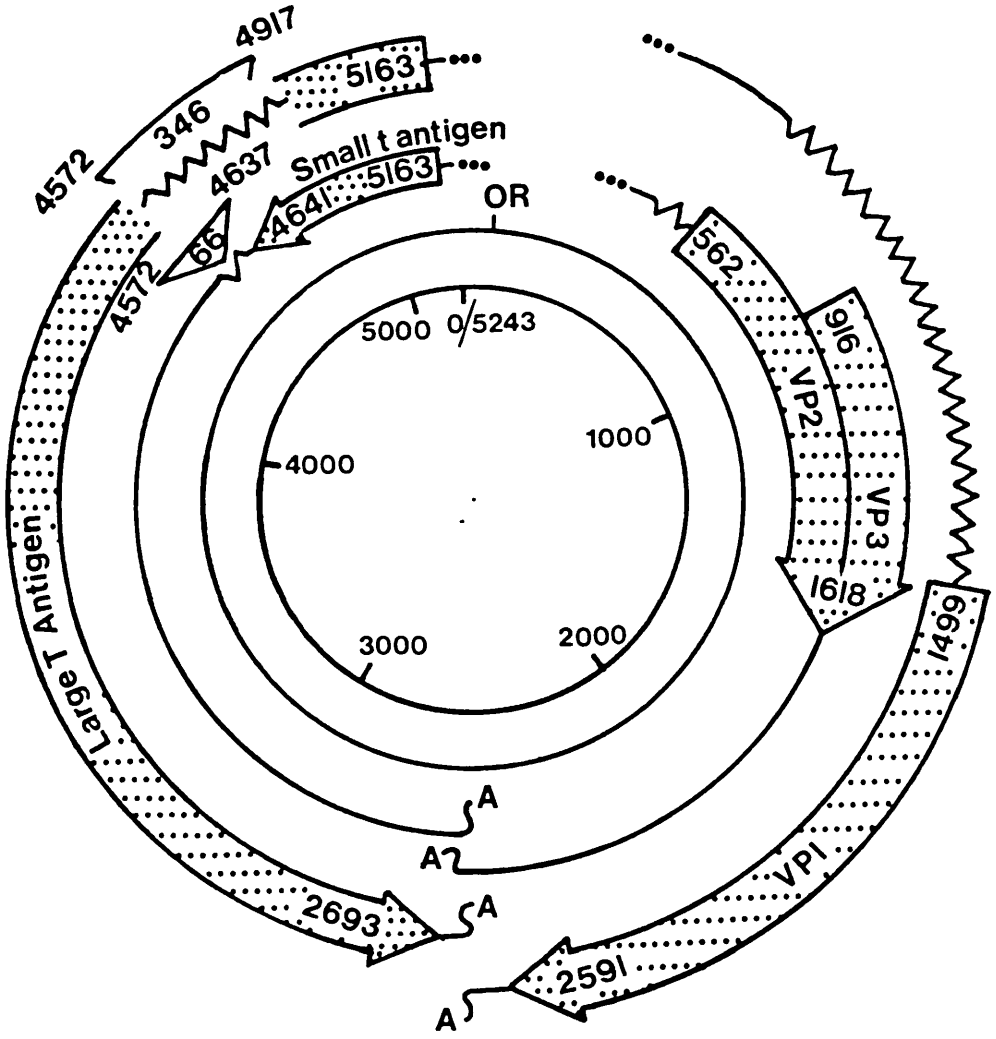
Arguably the best characterized eukaryotic sequence specific DNA binding protein with gene regulatory function is the large T antigen of SV40. The life cycle and molecular biology of SV40 have recently been reviewed by Tooze, (1981).

The SV40 genome encodes six known proteins, the two tumor antigens, large T and small t, the virion structural proteins VP1, VP2 and VP3 (Tooze, 1981), and the agnoprotein (Jackson and Chalkley, 1981; Jay et al., 1981 b; Nomura et al., 1983). Of these, only large T and small t synthesis is normally detected during non-productive infection or in transformed cells, while in lytic infection, all six proteins are detected, with large T and small t synthesis being detectable before the onset of viral DNA synthesis, and continuing throughout the course of the infection (Crawford et al., 1980; Tooze, 1981). The interrelationships of the SV40 mRNAs, the proteins they encode and the structure of the viral genome are shown in Figure 1.2.

The early region of SV40 encodes two proteins: the large T and small t antigens (Prives et al., 1977) which share the same 82 amino acids at their N-termini (Paucha et al., 1978), and are encoded by two distinct early region mRNAs, produced by differential splicing of the

Figure 1.2: Schematic diagram of the SV40 genome.

The SV40 genome is represented as a circle, with the origin of replication at the top. Numbers refer to the SV40 DNA sequence (Tooze, 1981). Arcs with arrowheads denote the position of mRNAs, boxed segments represent coding regions. Numbers within the dotted areas represent initiation and termination points for translation, those outside these regions denote the end points and sizes of the early region introns.


4917
5163
346
4572
4637
Small t antigen
5163
4641
OR
66
4572
5000
0/5243
1000
4000
3000
2000
562
916
VP2
VP3
1618
1499
Large T Antigen
2693
A
A
A
A
VP1
2591

same precursor molecule (Berk and Sharp, 1978; Reddy et al., 1979).

Section 1.7: Small t antigen.

The functions and properties of small t antigen are not very well defined. The △.54 - .59 mutants of SV40 have deletions within the intron of large T antigen which render them unable to produce full length small t while synthesis of large T remains unaffected (Khoury et al., 1979; Shenk et al., 1976). These mutants are viable in tissue culture though they produce a slightly smaller burst size during low multiplicity infections (Topp, 1980). The role of small t in vivo remains to be determined.

Small t antigen seems to be required for in vitro transformation of non-permissive cells, under conditions where soon after infection growth of the cell is restricted by imposing conditions where only a transformed cell will grow. Such conditions might be plating in semi-solid media, or a requirement to overgrow a monolayer of cells. It is found that actively growing cells can be transformed with equal efficiency by wild type and small t deletion mutant viruses (Martin et al., 1979; Rubin et al., 1982; Sugano et al., 1982). It has been suggested by Martin (1981) that small t has growth factor like activity.

The biochemical properties of small t are poorly

defined. It is not a DNA binding protein (Prives et al., 1980). It has been reported that small t causes dissociation of actin cable networks (Graessman et al., 1980) upon microinjection of DNA into cells, and is involved in the stimulation of a host centriolar antigen (Shyamala et al., 1982). It is also found complexed to two host proteins (Rundell et al., 1979). The relationship of these properties to its function is unclear. Small t is found both in the nucleus and cytoplasm of infected and transformed cells (X. Montano and D. P. Lane, personal communication).

Section 1.8: Large T antigen.

Large T antigen is a phosphoprotein with an apparent molecular weight of approximately 90,000 daltons (Tegtmeyer et al., 1975), which occurs in different quaternary structures in the nuclei of infected and transformed cells (Fanning et al., 1981) forming functionally discrete subsets (reviewed by Rigby and Lane, 1983). The protein is subject to a number of post translational modifications, including N-terminal acetylation (Paucha et al., 1978) and ADP ribosylation (Goldman et al., 1981). A fraction of the T antigen in the cell is found on the cell surface (reviewed by Rigby and Lane, 1983), and it has been suggested that it is anchored in the plasma membrane by tightly associated

lipids (Henning and Lange-Mutschler, 1983; Klockmann and Deppert, 1983).

T antigen has intrinsic ATPase activity (Tjian and Robbins, 1979; Griffin et al., 1980), and associated kinase activity (Griffin et al., 1980; Bradley et al., 1982; Van Roy et al., 1984) though whether the kinase activity is intrinsic to the protein has been disputed (Tjian et al., 1980). The pure protein does not unwind DNA (Myers et al., 1981a).

That large T antigen is required both for initiation and maintenance of transformation seems to be beyond doubt. The evidence has been comprehensively reviewed by Tooze (1981) and Martin (1981) and can be summarized as follows: first, all SV40 transformants and tumor lines examined thus far express large T antigen. Transfection experiments using fragments of the SV40 genome have shown that the early region is both necessary and sufficient for transformation (Graham et al., 1974). It has been found that some cell lines transformed by tsA mutant viruses are temperature sensitive for the transformed phenotype, and also that at the non-permissive termperature, tsA viral infection of non-permissive cells fails to induce transformation (Fluck and Benjamin, 1979). The tsA mutations map genetically to the T antigen coding sequence suggesting that T antigen functions are required both for initiation and maintenance of the transformed state. Strong additional evidence is provided by the fact that a high proportion of revertants selected for loss of the

transformed phenotype are found to have lost expression of the SV40 early functions (Martin, 1981). Small t antigen is not required for transformation under certain conditions, as shown by studies of deletion mutants unable to synthesize small t, which are capable of transformation *in vitro*, and tumor induction *in vivo* (Lewis and Martin, 1979), demonstrating that large T alone can cause transformation *in vitro*.

Section 1.9: Interaction of T antigen with the SV40 origin of replication.

Large T antigen is also involved in the lytic growth of SV40. Studies of the replication of *ts*A mutants at the non-permissive temperature showed that functional T antigen is required for the initiation of each round of viral DNA replication (Tegtmeyer, 1972). Analysis of replicating viral nucleoprotein complexes extracted from infected cells has shown that the protein is bound to the viral origin of replication (Mann and Hunter, 1979, 1980; Reiser *et al.*, 1980; Segawa *et al.*, 1980; Rudolph and Mann, 1983; Stahl *et al.*, 1983; Stahl and Knippers, 1983; Tack and de Pamphelis, 1983), a finding confirmed by electron microscopic observations (Harper *et al.*, 1984).

Much work has been expended on the study of the interaction of SV40 T antigen, or related proteins, with the viral origin of replication. Electron microscopic

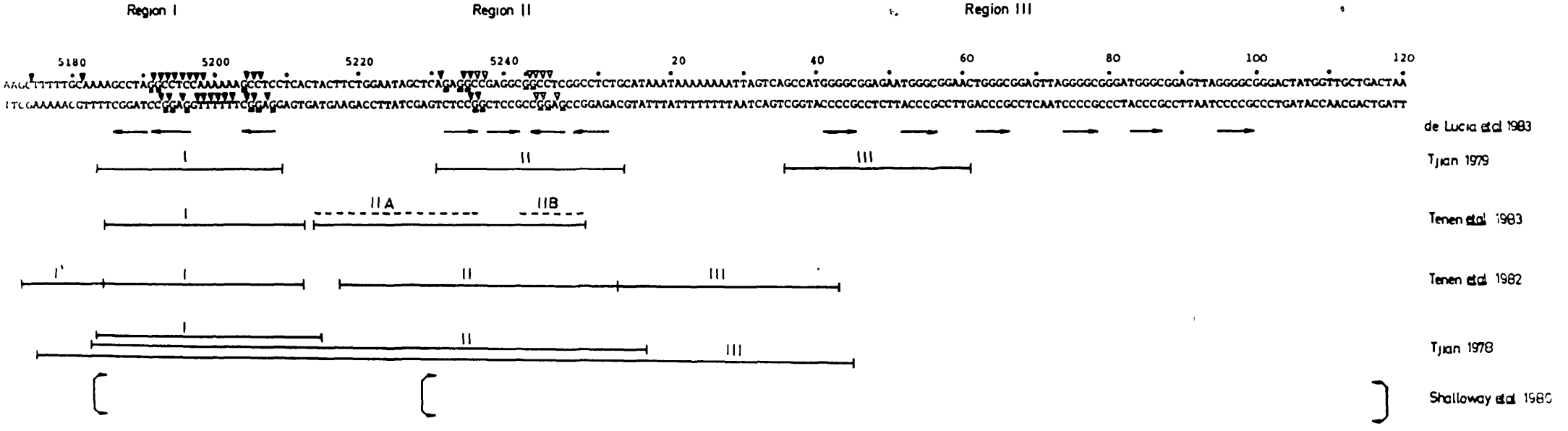
obvservations (Reed et al., 1975) and filter binding studies (Jessel et al., 1976) indicated preferential binding to the origin region. In a pioneering study Tjian (1978) described the purification of a T antigen related protein (encoded by an adenovirus-SV40 hybrid virus) comprised of all of the sequence unique to large T antigen fused to an adenovirus protein which replaces those sequences common to both large T and small T antigens. This protein, known as D2T, was shown to bind to three adjacent regions at the origin (known as sites I, II and III) in a cooperative manner, as judged by protection from DNAse I digestion. Tjian (1979) subsequently extended this analysis using protection of purine residues from DMS methylation by protein using the techniques developed for DNA sequencing (Maxam and Gilbert, 1977) to show that most contacts were in the major groove of the DNA. An alternative approach was used by Shalloway et al., (1980), who analyzed the sequences protected from digestion by exonuclease III when T antigen purified from the SV40 transformed human cell line SV80 was bound to the origin in vitro. A summary of the interaction of T antigen with the origin of replication is shown in Figure 1.3.
Subsequently, Tenen et al., (1982) have shown, using the DNAse I footprinting technology of Galas and Schmitz (1978), that as the D2T concentration is increased, site I is occupied first, confirming the data of Tjian, (1978), followed by sites II and III. They have also detected an additional site, on the early side of site I, known as

site I', which becomes occupied at the same concentration as site II. Studies of the binding of D2T to mutant origins indicated that binding to site II was sequence specific, in that when site II was deleted, normal protection of site I was observed, but no protection of site II equivalent sequences was observed. Thus the binding to site II is not a result of non-specific polymerization of D2T along the DNA following binding at site I. Additionally, it was observed that deletion of site I decreased the binding of D2T to site II, indicating that though binding to site II by D2T is sequence specific, prior binding to site I facilitates binding to site II (Tenen et al., 1982), a situation strikingly reminiscent of that observed with λ repressor binding to its operator (Johnson et al., 1981). Electron microscopic observations (Myers et al., 1981c) have shown that D2T is bound to the origin as a tetramer.

A potential objection which may be raised to these studies concerns the functional integrity of the proteins used. Gish and Botchan (unpublished, cited in Clark et al., 1981) have found that SV80 T antigen is defective for DNA replication, and it has recently been demonstrated using filter binding and the McKay immunoprecipitation assay that it is also defective in its sequence specific DNA binding capability (Myers et al., 1981a; Fanning et al., 1982; Burger and Fanning, 1983; Gruss et al., 1984). Thus experiments with this protein are seriously

Figure 1.3: The interaction of large T antigen with the SV40 origin of replication.

A summary of the accumulated findings concerning the interaction of T antigen with the SV40 origin of replication. Open triangles denote weak or moderate phosphate contacts, closed ones strong phosphate contacts with the strand below the triangle. Double underlined G residues indicate a strong contact point, single underlining a weak or moderate contact with that base (Jones and Tjian, 1984). Arrows indicate the location of pentanucleotide recognition elements (de Lucia et al., 1983). The numbering is according to Tooze (1981). Dots denote 10 bp intervals.


Region I
Region II
Region III
5180
5200
5220
5240
20
40
60
80
100
120
de Lucia et al 1983
I
II
III
Tjian 1979
I
IIA
IIB
Tenen et al 1983
I'
I
II
III
Tenen et al 1982
I
II
III
Tjian 1978
Shalloway et al 1980

compromised. D2T has also been shown to be impaired in its ability to bind to DNA. Comparison of the amounts of protein required to give protection of the binding sites at the origin from DNAse I digestion and DMS methylation showed that higher protein - DNA ratios are required for binding to site II when D2T was used compared to a "wild-type" full length T antigen isolated from lytically infected monkey cells, or T antigen produced by an SV40 - Adenovirus hybrid virus (Myers et al., 1981b). The binding of D2T to the origin of replication was also shown to be more labile to increasing salt concentration than that of lytic T antigen (deLucia et al., 1983). More D2T is also required to repress transcription from the SV40 early promoter in vitro than when authentic T antigen is used (Myers et al., 1981).

Therefore, more recent studies have used either T antigen extracted from lytically infected monkey cells or the T antigens produced by Adenovirus-SV40 hybrid viruses constructed to overproduce full length T antigen. These studies have demonstrated that there are three T antigen interaction sites at the origin of replication though the binding to site III is weak, and its physiological significance is unclear (Wilson et al., 1982; deLucia et al., 1983). deLucia et al., (1983) identified a related family of pentanucleotides with consensus sequence (G>T)(A>G)GGC present three times in region I, four times in region II and six times in region III, in different arrangements, with which the protein appears to make

contact as judged by DMS methylation protection. It has been proposed that a single pentanucleotide will direct and orientate the binding of a single molecule of T antigen (Tegtmeyer et al., 1983). That the pentanucleotides are at least a component of the recognition element has been shown by the subsequent studies which analyzed the binding of lytic T antigen to the origin of replication of BK virus (Ryder et al., 1983), and to mutant origins containing deletions of one or more pentanucleotides in site II (Lewton et al., 1984), where the presence of a single pentanucleotide was shown to be capable of directing the binding of T antigen, though the interaction was not very salt stable.

Tenen et al., (1983) studied the binding of lytic T antigen to wild type and mutant origins. They found that binding to site II was sequence specific, rather than due to processive polymerization of the protein along the DNA following binding at site I, confirming the previous studies using D2T. Additionally, using a restriction endonuclease protection assay, it was found that binding to site II involves sequential interactions with specific subsites, with T antigen binding to the early side of site II before the late side. In contrast to the findings of Myers et al., (1981b) and Tenen et al., (1982), it was found that prior binding of T antigen to site I did not facilitate binding to site II, suggesting that the binding of T antigen extracted from lytically infected cells to

the SV40 origin is not cooperative (DiMiao and Nathans, 1982; Tenen et al., 1983; Jones and Tjian, 1984). Whether the observed differences in binding behaviour between D2T and lytic T antigen results from differences in protein-protein, or protein-DNA contacts, or both, remains to be resolved.

Jones and Tjian (1984) have extended the analysis of the interaction of T antigen with the origin of replication using alkylation interference to define those bases and phosphate moieties involved in contacts with T antigen (Figure 1.3). Confirming earlier work, (Tjian, 1979), most contacts were found to be limited to the major groove of the helix. Sites I and II both contain two clusters of guanine contacts, accesible from one face of the helix only. Consistent with the finding of two subsites of different affinity within site II, it was observed that methylation of residues on the early side of site II interfered with binding more strongly than methylation of residues on the late side of site II. It is also clear from these studies that not all of the pentanucleotide blocks defined by deLucia et al., (1983) contain essential contact sites, suggesting that this putative consensus sequence is not the only structure recognized by T antigen, and that additional structural and / or sequence features may play an important role in directing the binding of T antigen to DNA. This view is further supported by the finding of Tenen et al. (1982), that D2T will bind to the origin of replication of a

mutant virus cs1088, (DiMiao and Nathans, 1980) which has multiple point mutations within site I resulting in the loss of all three pentanucleotides from this site. Despite this however, the affinity of the protein for this template is reduced only four-fold compared to wild type. In contrast, mutants in which site I has been deleted, the affinity is reduced 20 fold compared with wild type template. Clearly, further work is required to precisely define those elements that constitute a high affinity binding site for T antigen. T antigen binds to site I with 20-fold greater affinity than site II (Jones and Tjian, 1984). There is no clear specific feature of site II which would account for this, though it is possible that the spacing between the clusters of guanine residues, which differs between the two sites favours tighter binding at site I due to constraints imposed by the protomeric strucutre of T antigen.

An interesting analogy may be drawn between the λ repressor, discussed earlier, and T antigen in their interaction with DNA, in that they bind with similar affinities to the specific sequences they recognize, making essential contacts with both strands in the major groove of the helix. Like λ repressor the majority of the contact points are found along one face of the helix, with additional phosphate contacts reaching around the helix to embrace a greater surface area of the DNA duplex. Whether T antigen binding to DNA involves a similar arrangement of

α helix - major groove contacts to that of λ repressor awaits the elucidation of the crystal structure of T antigen. Clearly though, there are differences between the two proteins, the most notable being that in contrast to the λ repressor, the binding of T antigen to its specific sites is not cooperative.

At this point it is worth considering the DNA binding properties of polyoma large T antigen. Like SV40 T antigen, it is a nuclear protein (Takemoto et al., 1966). It is required for the initiation of each round of viral DNA replication (Francke and Eckhart, 1973), negatively regulates early transcription (Cogen, 1978), and is found associated with viral chromatin (Clertant et al., 1984). The purified protein binds with high affinity to three sites at the viral origin of replication (Gaudray et al., 1981; Pomerantz et al., 1983; Dilworth et al., 1984), though in contrast to SV40 T antigen, it binds all three sites with equal affinity. A similar pentanucleotide to that identified at the SV40 origin of replication is found repeated in each of these three binding sites, three times in site A, twice in site B, and four times in site C, and has been proposed as a component of the binding site structure that is recognised by Polyoma large T.

What then is the regulatory function of T antigen binding to these sites? The relationships between T antigen binding and viral DNA replication and early transcription have been analyzed by a study of the transcription and replication of a series of mutants with

deletions in the origin region, both *in vivo* and *in vitro*.

Mutants with deletions in site III are fully viable, suggesting that the physiological significance of the weak binding to this region observed *in vitro* is questionable (DiMiao and Nathans, 1980).

Site II lies within the origin of replication (Gutai and Nathans, 1978; Subramanian and Shenk, 1978). Deletions within this region abolish replication, but T antigen can still bind to the residual site I and II sequences *in vitro* (DiMaio and Nathans, 1980; Myers and Tjian, 1980; McKay and DiMaio, 1981; Tenen *et al.*, 1983; Lewton *et al.*, 1984). In contrast, deletion of site I causes only a slight, cold sensitive defect in DNA synthesis (DiMaio and Nathans, 1982). Deletion of site I and most of the sequences between sites I and II abolishes replication, and severely impairs binding *in vitro* (Myers and Tjian, 1980). Thus the data do not provide a clear correlation between *in vitro* binding and DNA replication *in vivo*.

Large T antigen regulates the transcription of its own mRNA in infected permissive cells (Tegtmeyer *et al.*, 1975; Reed *et al.*, 1976; Alwine *et al.*, 1977). This autoregulation has been extensively studied both *in vivo* and *in vitro*. Rio *et al.*, (1980), showed that addition of D2T antigen to an *in vitro* transcription system specifically repressed synthesis of RNA from the SV40 early promoter, without affecting transcription from the

SV40 late promoter, or the major late promoter of adenovirus 2. Subsequently, Myers *et al.*, (1981b) showed that pure T antigen could mediate a similar effect, and that its mode of action was to prevent RNA polymerase from binding to the early promoter. It was found that the transcription of templates lacking binding sites I and II was not repressed *in vitro* by adding T antigen suggesting that binding to these sites was important in autoregulation. It has been shown that at late times in infection, the 5' ends of early region mRNAs shift approximately 40 nucleotides upstream from the initiation sites used early in infection, which lie close to the origin of replication. Studies of the transcription of *ts*A mutants late in infection has demonstrated that functional T antigen is required to maintain the upstream shift of 5' termini (Ghosh and Lebowitz, 1981). Hansen *et al.*, (1981), have shown that both the early and late early mRNA 5' ends are used *in vitro*, and that as increasing amounts of D2T are added, first early early 5' ends and then late early 5' ends are repressed. The sequential repression was found to correlate precisely with the binding of D2T to sites I and II respectively. The precise mechanism of production of late-early transcripts *in vivo* is however, unclear, since a mutant lacking site I sequences undergoes a normal shift during infection (Hansen *et al.*, 1981). Further work will clearly be required to address the question of what role T antigen plays in mediating the shift of 5' ends during infection.

Studies of autoregulation in vivo (DiMaio and Nathans, 1982; Rio et al., 1983) have shown that deletion of either site I or site II sequences causes overproduction of early mRNA and T antigen. However deletion of both site I and site II, or the presence of a temperature sensitive T antigen results in an even greater derepression of the early promoter indicating that binding of T antigen to both sites I and II plays a role in autoregulation.

Section 1.10: Identification of regions of T antigen involved in binding to DNA.

Attempts have also been made to define the region of T antigen which is responsible for binding to DNA, using a variety of different approaches. Studies of the in vitro translation products of SV40 cRNA have shown that only polypeptides which contain the information encoded between coordinates 4400 - 4600 will bind to SV40 DNA, suggesting this region is involved in origin binding (Prives et al., 1982). Analysis of the DNA binding properties of the truncated T antigens encoded by the non - defective SV40 - Adenovirus hybrid viruses (Prives et al., 1982) has again implicated this region in origin binding, and has also shown that proteins carrying coding information from the region 3800 - 4000 are able to interact non-specifically with calf thymus DNA, but do not bind specifically to the

origin. A third approach has been to study the DNA binding properties of mutant T antigens. Wilson *et al.*, (1982) have analyzed the T antigens synthesised by *ts*A mutant viruses during lytic infection of monkey cells. It was found that mutants with lesions between coordinates 3476 - 3733 were defective for DNA binding only after heat shock, while those with lesions between coordinates 3733 and 4002 were severly impaired in their ability to bind the origin of replication both before and after heat shock. Prives *et al.*, (1983) studied the DNA binding properties of mutant T antigens unable to support DNA replication, synthesized by some transformed cells. They found that the T antigen encoded by the transformed cell line C6 carries three point mutations. These were separated *in vitro* and recombinants carrying the separated lesions were used to transform non-permissive cells. The T antigens encoded by these lines were studied for their DNA binding properties. It was found that a T antigen bearing a mutation in amino acid 153 (around coordinate 4350) was unable to bind to the origin of replication. A caveat must be held against the latter report however, since the template encoding the mutant T antigen was not recloned from the transformed rat cells to prove that no other lesions had been introduced during the transformation event. Analysis of truncated T antigens found in transformed cells has shown that proteins extending from the N - terminus to approximately nucleotide 4000 will bind to DNA cellulose, while shorter

ones will not (Chaudry et al., 1982).

In a very elegant study (Shortle et al., 1979; Margolskee and Nathans, 1984), a series of mutants were constructed in the viral origin of replication which impaired the ability to replicate. Second site revertants of these were selected by random local mutagenesis of mutant DNA followed by transfection of cultured cells and selection of large (wild type) plaques. It was found that the second site revertants mapped by marker rescue and in vitro recombination to the gene for T antigen, map in the region betwen coordinates 4000 - 4400. DNA sequence analysis of two of these mutants has shown that the changes are missense mutations in amino acids 157 and 166. The mutants display a relaxed specificity, being able to function with variously modified origin sequences including some with deletions which are inactive for replication when coupled with wild type T antigen. Clearly, a study of the in vitro interaction of these T antigens with mutant and wild type origins would be of great interest.

The DNA binding properties of a series of in vitro constructed deletion mutants have also implied, in agreement with the studies above, that sequences between 3620 - 4610 are important in origin recognition (Clark et al., 1983; Pipas et al., 1983), but also serve to emphasize a major potential flaw in most of these studies, which is the assumption that T antigen functions as a

linear array of amino acids. Thus, a deletion which abolishes DNA binding does not permit one to conclude that the deleted region contained the DNA binding domain, for it is intrinsically just as likely that the deletion alters the tertiary structure of the protein, or precludes some postranslational modification, thereby preventing DNA binding. An excellent example of such an effect is provided by deletion mutant dl 1135, which has a small deletion in the N terminal large T/small t common region, not previously implicated in binding to DNA, and yet is defective in origin binding. It seems highly likely that this defect is mediated via structural effects but serves to emphasize the caution with which the results of this type of analysis must be interpreted. In this respect, the selection of pseudorevertants described above is less compromised, since it measures a positive change, that is to say the ability of the mutant T antigen to interact with a template with which wild type T antigen cannot, rather than a negative one such as loss of DNA binding ability. It is therefore much more likely that these pseudoreversion mutations will define those regions intimately involved in DNA binding, though of course, other effects cannot be ruled out without further biochemical analysis.

The studies are however broadly in agreement, defining a region encoded by nucleotides 4600 - 4200 as being important in origin binding. They are also in good agreement with the study of Morrison et al., (1983) who

identified a 17 kd N terminal fragment of T antigen, produced by partial cleavage with trypsin, which exhibits specific origin binding in the McKay immunoprecipitation assay. This fragment extended from the N terminus to approximately coordinate 4450. However, since neither the affinity of the fragment for the origin, nor its interaction with it were analyzed, it cannot be said that this is the only region involved in DNA binding.

The precise molecular form of T antigen which binds to the origin of replication remains to be determined, and will require the use of protein - nucleic acid and protein - protein cross linking agents. Studies have however been undertaken on the DNA binding properties of the various subsets of T antigen found within the cell. Zonal sedimentation of SV40 infected or transformed cells in sucrose or glycerol gradients resolves a minimum of two forms of large T antigen, one sedimenting at 5 - 7 S, the other at 15 - 16 S. A third species sedimenting at 23 S is also seen, which represents the complex of T antigen with a host protein of 53,000 molecular weight (Lane and Crawford, 1979). The 5 - 7 S peak may be resolved into two separate species, one at 5.5 S, representing monomeric T antigen, the other at 7 S, a dimer (Fanning et al., 1981; Bradley et al., 1982).

The 23 S complex of T antigen and 53K retains origin specific DNA binding (Bradley et al., 1982; Dorn et al., 1982) immunoprecipitable with both anti-T and anti-p53

antibodies (Reich and Levine, 1982; Burger and Fanning, 1983). p53 alone will not bind specifically to the origin of replication of SV40.

The 5 - 7 S form and 16 S form of T antigen differ in their DNA binding properties with respect to both pH optima and salt resistance of binding. The 16 S form has a higher affinity for non-origin DNA, and greater salt resistance reaching an optimum at 150 mM to 200 mM NaCl rather than at 60 mM to 80mM for the lighter species. 5 - 7 S T antigen has a higher affinity for origin DNA than 16 S T antigen (Dorn et al., 1982; Fanning et al., 1982; Gidoni et al., 1982).

Pulse chase experiments (Fanning et al., 1981) have shown that newly synthesized T antigen first appears in the 5 - 7 S form, later aggregating into the 16 S and 23 S forms. The newly synthesized T antigen is less highly phosphorylated than the 16 S form (Fanning et al., 1981; Scheidtmann et al., 1984). It has been reported that in vitro phosphorylation of D2T causes it to aggregate (Baumann and Hand, 1982), and Montenarrh and Henning, (1983) have reported that oligomers can be dissociated to monomers and reassembled in vitro, whereas newly synthesized monomers do not form oligomers in vitro, suggesting that phosphorylation and oligomerization are linked. Recent data (Scheidtmann et al., 1984) suggest that the form most active in origin binding is a 5 S species, probably a monomer. This fraction was found to represent newly synthesized large T, and in agreement with

previous studies was phosphorylated to a low degree compared to "aged" 16 S T antigen. These data are however, at variance with the results of Bradley et al., (1982), who found that the active DNA binding species was the 7 S dimer, and that the monomer did not bind DNA.

This apparent discrepancy might be due to a variety of reasons, such as differences in the binding assay employed, but it should also be noted that the study of Bradley et al, was conducted using T antigen extracted from the SV80 cell line, subsequently shown to be severely defective in DNA binding (Myers et al., 1981a; Fanning et al., 1982; Gruss et al., 1984), so it could be argued that the failure of monomer to bind SV40 DNA is a consequence of this defect. The work of Scheidtmann et al., (1984) used wild type T antigen extracted from lytically infected cells.

Scheller et al., (1982) have reported that a monoclonal antibody PAb100 which recognizes only a subfraction of the T antigen in cells immunoprecipitates the majority of the DNA binding activity. What identifies the fraction that is recognized by PAb100 is not clear; however, PAb100 is a low affinity antibody and it will therefore preferentially precipitate oligomeric forms of T antigen, so an alternative interpretation of the data is that T antigen is oligomerized when bound to DNA.

It has been shown that treatment of T antigen with alkaline phosphatase (Shaw and Tegtmeyer, 1981) does not

alter its binding to the origin. However, since only phosphorserine residues were dephosphorylated, the only conclusion which could be drawn was that the vast majority of phosphoserine phosphorylation sites are not involved in regulating the site specific recognition and affinity of T antigen for the SV40 origin. No conclusion was drawn about the importance of phosphothreonine phosphorylation in binding. Further work is clearly necessary to clarify this. It has been reported that T antigen will also bind single stranded DNA in addition to double stranded DNA (Spillman et al., 1979; Montenarrh and Henning, 1982).

SV40 also supplies a "helper" function, complementing a block to late gene expression to permit human adenoviruses to replicate efficiently in monkey cells. This function has been genetically mapped to the extreme C terminus of T antigen (Cole et al., 1979; Polvino-Bodner and Cole, 1979). Purified T antigen will also provide this function if microinjected into adenovirus infected monkey cells (Tjian et al., 1978) formally proving that it is T antigen that provides the helper function.

Section 1.11: Interaction of T antigen and the host cell.

T antigen also affects DNA replication and transcription in the host cell (reviewed by Rigby and Lane, 1983). Host DNA synthesis is stimulated following infection of both permissive and non-permissive cells (reviewed by Martin, 1981). Temperature shift infection

experiments following infection or transformation by wild type, tsA mutant virus, small t deletion mutant virus or tsA, small t deletion double mutant virus indicated that functional large T antigen is required for the stimulation of cell division (Chou and Martin, 1975; Hiscott and Defendi, 1979; Hiscott and Defendi, 1981; Christensen and Brockman, 1982). These experiments also indicated that small t antigen is not required, at least for the first cycle of division following infection. Fibre autoradiography has shown that frequency of cellular origins of replication is increased, and the origin to origin distance is decreased in SV40 transformed cells relative to normal cells, an effect dependent on the presence of functional T antigen (Martin and Oppenheim, 1977). There is also evidence that large T antigen can alter the time during S phase at which particular regions of the genome are replicated (Marchionni and Roufa, 1981). It was found that in tsA transformed cell lines integrated SV40 sequences were replicated at different times at the permissive and non-permissive temperatures, suggesting that T antigen altered the time at which the SV40 containing replicon was replicated, possibly using the SV40 origin, though this was not proven. Finally, it has also been demonstrated that microinjection of purified D2T into the nuclei of quiescent cells is capable of inducing a round of cellular DNA replication (Tjian et al., 1978). Deletion mutant analysis (Soprano et al., 1983) has

suggested that the region encoded between coordinates 4000 and 4300 is important. This region is part of the domain implicated in DNA binding, and some of the second site revertants isolated by Shortle et al., (1979) also map to this region. Whether T antigen stimulates DNA replication by directly binding cellular origins of replication, or by interaction with some additional protein is unclear, though an obvious candidate for such a factor would be the 53K protein with which T antigen associates in both lytic infection and transformed cells (Lane and Crawford, 1979; reviewed by Lane et al., 1982). That 53K might in some way be involved in control of cell proliferation has been suggested by studies in which microinjection of anti-p53 monoclonal antibodies into quiescent cells was able to inhibit DNA synthesis in response to serum stimulation. Inhibition was only effected if the antibody was injected within two hours of serum stimulation (Mercer et al., 1983). It has been shown that there is an increase in synthesis and steady state levels of p53 protein and mRNA in late G1 prior to the onset of DNA synthesis in response to serum stimulation (Reich and Levine, 1984) suggesting a role for p53 in the progression of cells from a growth arrested state to an actively dividing state. In non-transformed cells, the protein has a very short half life, being stabilized by association with T antigen (Oren et al., 1981). It has also been found that the Adenovirus E1b protein complexes with p53 in adenovirus transformed cells (Sarnow et al., 1982).

It has been found that when several SV40 transformed cell lines were ordered with respect to the amount of T antigen, there was a close correlation between growth rate in subconfluent culture and the amount of T antigen per cell (D. P. Lane and G. Winchester, personal communication). Those cell lines expressing large amounts of T antigen were found to divide more rapidly than those containing low levels of the protein. One may therefore speculate that by binding to p53 and thereby stabilizing it, T antigen might cause rapid cell proliferation by increasing the amount of p53 available. Since p53 is a DNA binding protein (D. P. Lane, personal communication) recognition of cellular origins could be through p53, or alternatively p53 could act as a specificity factor enabling T antigen to recognize cellular replication origins.

The presence of large T antigen also affects transcription of a number of different templates in the host cell. In some human - mouse hybrid cell lines, only the human rDNA complement is normally expressed. Viral infection or transfection of SV40 DNA into such cells leads to transcriptional activation of the quiescent mouse rDNA genes. The reactivation is dependent on functional large T antigen, and is independent of small t antigen (Soprano et al., 1979, 1980). Deletion analysis (Soprano et al., 1983) and microinjection of cloned DNA fragments (Soprano et al., 1981) have localized the transcription

induction function to the region encoded between nucleotides 3500 - 3800 (albeit with the reservations expressed earlier). This activation has been mimiced *in vitro*. It has been found that addition of pure T antigen to a RNA polymerase I *in vitro* transcription system stimulates rDNA transcription (Learned *et al.*, 1983). The mechanism by which this activation is achieved is unclear however since rDNA genes contain no T antigen binding site. Thus, it is possible that T antigen can mediate this effect by directly interacting with the transcriptional machinery. Further work is required to clarify this point.

Effects on the expression of cellular protein coding mRNA species have also been established. RNA/cDNA cross hybridization studies have shown that activation of a small number of poly A^+ RNAs occurs on SV40 transformation of human fibroblasts, whereas expression of the vast majority of poly A^+ RNA was unchanged. It was found that transcription of approximately 30 average sized mRNA molecules was activated (Williams *et al.*, 1977). Subsequently, genes which are differetially expressed in transformed cells have been cloned (Schutzbank *et al.*, 1982; Scott *et al.*, 1983). However, direct regulation of these transcripts by T antigen, rather than *via* indirect effects, remains to be demonstrated by transfection experiments and *in vitro* transcription of such regulated genes. Some indication that T antigen does play a role in the regulation of these genes is provided by analysis of

the transcription of the genes in tsA transformed cell lines, where increased expression is seen only at the permissive temperature.

T antigen has also been shown to directly activate transcription from the SV40 late promoter in the absence of DNA replication (Brady et al., 1984; Keller and Alwine, 1984). It is interesting to speculate that there may be a class of cellular promoters like the SV40 late promoter with which T antigen could interact to alter transcription in the transformed cell. In this context, it is interesting to note that the Adenovirus Ela gene product, which is known to activate both viral and cellular genes (Nevins, 1981, 1982) can substitute for T antigen in activating the SV40 late promoter (Keller and Alwine, 1984). Recent work has demonstrated the presence of a cellular Ela - like activity (Imperiale et al., 1984), raising the possibility that T antigen could act on genes controlled by such an element, activating them in its absence.

Section 1.12: Interaction of T antigen with host cell chromatin.

There is evidence available which demonstrates that a proportion of the T antigen in the nucleus associates with the host cell chromatin. It was found that the chromosomes of SV40 infected cells could be stained with

anti - T antisera (d'Alisa and Gershey, 1978). Subsequent subcellular fractionation studies (Staufenbiel and Deppert, 1983) showed that approximately 13% of the T antigen in the nucleus was tightly associated with chromatin, only being released upon DNAse I digestion and high salt extraction, while 20% was associated with the nuclear matrix. The remainder was defined as nucleoplasmic, and was readily extracted with detergent and low salt. A tightly bound form of T antigen has also been found by Scheidtmann et al., (1984) who found that further high salt extraction of nuclei which had already been extracted with moderate salt concentration and detergents released more T antigen. This fraction bound very tightly to DNA in vitro, and in its properties resembled the newly synthesized T antigen monomer. Further evidence is provided by the finding that tsA transformed cell lines are temperature sensitive for autoregulation of SV40 early genes, suggesting that T antigen binds to the SV40 origin of replication even in a transformed cell (Alwine et al., 1977). It has also been demonstrated that fusion of transformed non-permissive cells with permissive CV-1 cells allows rescue of virus (Botchan et al., 1978), and that this process requires both a functional T antigen, and a functional origin of replication in the transformed cell (Gluzman, 1981), suggesting that the origin of replication is accessible to T antigen in the chromatin of the transformed cell.

Section 1.13: Conclusions.

It is therefore apparent that control of eukaryotic gene expression is a complex process, with many potential levels of control. There is clear evidence that sequence specific DNA binding proteins exert a regulatory role over the process of gene expression. One of the best characterized eukaryotic regulatory proteins is the large T antigen of SV40. Its interaction with the SV40 origin and early promoter have been well characterized, and a consensus recognition sequence at the origin defined. There also exists evidence suggesting that T antigen can affect the expression of mRNA encoding genes, and perturb DNA replication in the host cell, possibly by interaction with specific sequences in the cellular chromatin. Thus it is clearly of interest to study sequences with which T antigen interacts in the cell.

The aim of this study is therefore to attempt to develop a strategy by which such DNA fragments might be selectively isolated from the rest of the genome using SV40 large T antigen as a model system, and then to analyze the interaction of T antigen with these sequences *in vitro*, and the effects, if any, in the host cell of the binding *in vivo*.

CHAPTER TWO

MATERIALS AND METHODS

CHAPTER TWO

MATERIAL AND METHODS

Section 2.1: Eukaryotic cells and viruses.

Balb/c 3T3 Clone A31 cells were obtained from Dr. W. A. Nelson-Rees, Naval Biomedical Research Laboratory, Oakland, CA.

SV3T3 Cl 38 cells, derived by infection of Balb/c 3T3 A31 cells with WT 830 strain of SV40, were obtained from Dr. P. Rigby, Imperial College (Rigby et al., 1980).

Cos-1 cells (Gluzman, 1981) were obtained from Dr. Y. Gluzman, Cold Spring Harbor Laboratory.

293 cells (Graham et al., 1977) were obtained from Dr. P. H. Gallimore, University of Birmingham.

CV-1 cells were obtained from Dr. L. Crawford, I.C.R.F., London.

Ad5 SVR111 virus stock (Gluzman et al., 1982) was obtained from Dr. Y. Gluzman, Cold Spring Harbor Laboratory.

SV40 virus, strain 776 was obtained from Dr. L. Crawford, I.C.R.F., London.

Virus stocks were maintained and propagated by Mrs. E. Woodruff, Imperial College.

All cells were grown at 37°C in humidified incubators gassed with 10% ($^{v}/v$) CO_2 in air, on plastic dishes (Nunc), in Dulbecco's modification of Eagle's Medium [DMEM] (Flow), supplemented with 10% ($^{v}/v$) Fetal Calf

Serum [FCS] (Flow). SV3T3 Cl 38 cells were grown in DMEM supplemented with 2% (^{v}/v) FCS. All media contained 500 µg/ml penicillin and 100 µg/ml Streptomycin.

Infection of cells with virus

The medium was removed from subconfluent 9 cm dishes of cells and replaced with 0.6 ml of TS (Table 2.1) containing 2% (^{v}/v) FCS and sufficient virus to give the desired multiplicity of infection (m.o.i.). The cells were then incubated at 37°C for 90 minutes for SV40 infections or 2 ^{1}/2 hours for adenovirus infections to allow the virus to adsorb to the cells, rocking the dishes every 20 minutes to distribute the virus evenly over the cells. The cells were then overlaid with DMEM containing 2% (^{v}/v) FCS and incubated for the required length of time.

In vivo labelling of cells

The medium was removed from the cells, which were then washed twice in TD (Table 2.1). They were then incubated for 20 minutes at 37°C in DMEM lacking methionine and glutamine (Flow), supplemented with glutamine (2mM). This medium was then replaced with a mixture of fully supplemented DMEM and DMEM lacking methionine and glutamine [1:9 (^{v}/v)] containing 25 µCi/ml ^{35}S-Methionine (1200 Ci/mmol, Amersham International) and 5% (^{v}/v) methionine deficient FCS, supplemented with glutamine, and the cells labelled for the desired time.

TABLE 2.1

Solutions used in tissue culture

TD:	25 mM Tris-HCl, pH 7.4 136 mM NaCl 5.7 mM KCl 0.7 mM Na_2HPO_4
Ca/Mg solution:	21 mM $MgCl_2$ 18 mM $CaCl_2$
TS:	TD + 5% ($^v/_v$) Ca/Mg solution.

Section 2.2: Bacteriological techniques.

E. coli strains

DH-1 (Hanahan, 1983): F^-, recA1, endA1, gyrA96, thi-1, hsdR17 ($r_k^- m_k^+$), supE44.

JM101 (Vieira and Messing, 1983): Δ (lac, pro), supE, thi: F'traD36, proAB, $lacI^q Z\Delta M15$.

HB101 (Boyer and Roulland-Dussoix, 1969): F^- hsd S20 ($r_b^- m_b^-$), recA, ara, proA, lacY, galK, rpsL20 (Sm^R), xyl5, supE.

Strains were stored as stab cultures, or at -70°C in L-broth containing 15% (v/v) glycerol, and propagated on L-plates or Ψ plates (Table 2.2). JM101 was propagated on minimal M9 medium plates, supplemented with thiamine.

Media

Recipies for media are given in Table 2.2. Media were sterilized by autoclaving at 15 p.s.i. for 15 minutes at 121°C. Antibiotics and other supplements were added from filter sterilized stocks when the agar had cooled to approximately 60°C. After mixing, the plates were poured, allowed to set, and dried before use.

Supplements and antibiotics

Ampicillin (Sigma) and Tetracycline (Sigma) were added to plates and liquid media to final concentrations of 100 µg/ml and 25 µg/ml respectively. Amino acids and vitamins were added to 20 µg/ml. All stock solutions were filter sterilized using 0.22 µm filters (Nalgene).

TABLE 2.2

Media used in bacterial growth

L-broth: 1% (^{w}/v) Difco Bacto Tryptone, 0.5% (^{w}/v) Difco Bacto Yeast Extract, 0.5% (^{w}/v) NaCl, 0.1% (^{w}/v) glucose, adjusted to pH 7.2 with NaOH.

L-agar: L-broth containing 1.5% (^{w}/v) Difco Bacto Agar.

Ψ- broth: 2% (^{w}/v) Difco Bacto Tryptone, 0.5% (^{w}/v) Difco Bacto Yeast Extract, 20 mM $MgSO_4$, 10 mM NaCl, 5 mM KCl adjusted to pH 7.5 with KOH.

Ψ- agar: Ψ- broth containing 1.5% (^{w}/v) Difco Bacto Agar.

BBL Top agar: 1% (^{w}/v) BBL Trypticase, 0.5% (^{w}/v) NaCl, 0.65% (^{w}/v) Difco Bacto Agar.

Water Agar: 2% (^{w}/v) Difco Bacto Agar.

M9 Minimal salts: 28 g/liter Na_2HPO_4, 12 g/liter KH_2PO_4, 2 g/liter NaCl, 4 g/liter NH_4Cl.

Minimal M9 Agar plates: M9 salts containing 1.5% (^{w}/v) Difco Bacto Agar, 0.2% (^{w}/v) glucose and 1 mM $MgSO_4$.

TfB I: 100 mM RbCl, 50 mM $MnCl_2$, 10 mM $CaCl_2$, 35 mM CH_3COONa, 15% (^{v}/v) glycerol, adjusted to pH 5.8 with acetic acid, and filter sterilized.

TfB II: 10 mM RbCl, 75 mM $CaCl_2$, 10 mM MOPS, 15% (^{v}/v) glycerol, pH 6.8 adjusted with KOH, and filter sterilized.

Bacterial Transformation

The following protocol represents a modification of that developed by D. Hanahan (cited in Maniatis et al., 1982) for χ^{1776}, and was routinely used for all strains.

A single colony of 2 - 3 mm diameter, taken from a fresh Y plate, was inoculated into 7 mls of Y broth prewarmed to 37°C, and grown with vigorous shaking at 37°C until an OD_{550} of 0.3 was reached.

This culture was then used to provide a 5% (v/v) innoculum into 100 mls prewarmed Y broth which was then grown at 37°C with vigorous shaking to an OD_{550} of 0.48 - 0.5 (approximately 2 hours). The cells were then transferred to sterile Oak - Ridge centrifuge tubes and chilled on ice for 15 minutes. They were then collected by centrifugation at 4°C, 3000 rpm, for 5 minutes and washed in 30 mls of ice cold TfB I (Table 2.2). The cells were pooled, spun down as before and resuspended in 30 mls of TfB I and incubated on ice for 30 minutes. The cells were then pelleted and gently resuspended in 4 mls of ice cold TfB II (Table 2.2). After 20 minutes on ice the cells were snap - frozen in dry - ice ethanol in 200 µl aliquots in prechilled Eppendorf tubes and stored at -70°C until required.

For transformation a tube of cells was thawed at room temperature until just liquid, and placed on ice for 15 minutes. DNA was added to the cells and after 30 - 60 minutes on ice the cells were heat shocked by incubation

in a 42°C water bath for 90 seconds with constant agitation, and then placed back on ice for 2 minutes. 800 µl of Y broth was then added and the cells were incubated at 37°C with gentle agitation for 60 minutes before plating on the appropriate selective media.

DNA to be added to cells was in either TE or ligase buffer (Table 2.4). Transformation was routinely assayed using 10 ng of pAT153 (Twigg and Sherratt, 1980) or 10 ng of M13 mp8 (Messing and Vieira, 1982). After heat shock, JM101 was plated directly on L-plates together with 2.5 mls of BBL top agar to which IPTG and X-gal had been added to 1 mM and 0.02% (w/v) respectively. Plates were incubated inverted overnight at 37°C. Typical efficiencies were 1 - 6 x 10^8 transformants per µg DNA for DH-1 and 0.5 - 5 x 10^7 transformants per µg DNA for HB101 and JM101.

Section 2.3: General Techniques.

Phenol - Chloroform extraction

1 volume of phenol and 1 volume of chloroform were added to the sample and the phases mixed either by gentle vortexing or repeated inversion of the tube. The phases were separated by centrifugation and the upper, aqueous phase transferred to a fresh tube. Residual phenol and chloroform were removed by adding 1 volume of water saturated di-ethyl ether and mixing. After brief

centrifugation, the upper phase was removed and discarded. Residual ether was removed by heating the solution in a water bath at 60°C.

The phenol used for extraction was redistilled and equilibrated with 0.1 M Tris-HCl pH 8.0 until the pH of the aqueous phase was greater than 7.5. 8-Hydroxyquinoline was added to 0.1% (^{w}/v) to act as antioxidant.

Ethanol precipitation

Nucleic acids were precipitated by addition of NaCl to 0.1 M or sodium acetate (adjusted to pH 5.5 with acetic acid) to 0.3 M followed by 2.5 volumes of 96% ethanol. The sample was mixed and incubated for 30 minutes or longer at -20°C or for 10 minutes in dry ice - ethanol. The precipitate was collected by centrifugation for 5 - 10 minutes in a microfuge or 10 minutes at 10,000 rpm in the Sorvall HB-4 swing out rotor at 0°C. The pellet was then washed with 70% (^{v}/v) ethanol, drained and after briefly drying under vacuum, resuspended in TE. When only small amounts of DNA were present, *E. coli* tRNA or nuclease - free BSA were added to 10 µg/ml as carrier.

Preparation of DNAse - free RNAse

RNAse A (Sigma) was dissolved in 0.1 x SSC at a concentration of 10 mg/ml and heated to 100°C for 15 minutes. The enzyme was then stored frozen in small aliquots at -20°C. The presence of residual DNAse was assayed by incubating the enzyme with a restriction digest to check that no smearing of the band pattern occurred.

TCA precipitation

The sample to be assayed was diluted to 100 µl with TE and 20 µg of fragmented salmon sperm DNA added, followed by ice - cold TCA to a final concentration of 10% ($^{w}/v$). After 5 minutes on ice, the precipitate was collected by filtration onto Whatman GF/C filters, washed with cold 10% TCA containing 0.1 M tetra-sodium pyrophosphate followed by ethanol, dried and counted using Aquasol 2 scintillation fluid (NEN).

Subcloning

200 ng of appropriately digested vector was mixed with a 3 fold molar excess of the target DNA and ligated overnight in a 10 µl volume in Ligase buffer using 1000 U/ml T4 Ligase at 6°C for EcoRI cohesive ends and at 15°C for all other ends. After transformation of competent bacteria (Section 2.2) colonies containing the desired insert were identified either by plaque color, DNA sequencing (Section 2.7), colony screening (Section 2.6), inactivation of antibiotic resistance or analysis of mini preps (Section 2.4).

TABLE 2.3

Some commonly used solutions

TE: 10 mM Tris-HCl pH 8.0, 1 mM EDTA.

TNES: 10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.1% (^{w}/v) SDS.

Denhardt's solution: 0.02% (^{w}/v) BSA, 0.02% (^{w}/v) Polyvinyl pyrollidone, MW 40,000, (Sigma), 0.02% (^{w}/v) Ficoll, MW 400,000, (Sigma).

Standard Saline Citrate (SSC): 1 x SSC is 0.15 M sodium chloride, 0.015 M tri-sodium citrate.

Section 2.4: Nucleic acids.

Restriction endonucleases

AluI, BglII, BstNI, HhaI, EcoRl, KpnI, PstI, PvuII, RsaI, TaqYI, XhoI, T4 DNA ligase and E. coli DNA polymerase I were purchased from New England Biolabs.

AvaII, ClaI, HindII, HindIII, SalI, Sau3A, SmaI, XhoII, Calf Intestinal Alkaline Phosphatase and T4 polynucleotide kinase were purchased from Boehringer Mannheim.

HaeIII, HinfI, MboI and E. coli DNA polymerase I, Klenow fragment, were purchased from Bethesda Research Laboratories.

BstI was a gift from T. Anderton, Imperial College.

A.M.V. Reverse Transcriptase was obtained from J. Beard, Life Sciences, Inc., St. Petersberg, Florida.

Terminal transferase was obtained from Ratcliff Biochemicals, Los Alamos, New Mexico.

Restriction digests

Analytical reactions were generally performed in a volume of 20 µl containing 0.5 to 2 µg DNA, the appropriate buffer (Table 2.4) and 2 units of enzyme per µg DNA. Digestion was for 1 hour at 37°C (65°C for TaqYI and BstNI). Reactions were terminated by addition of 0.1 volume of gel buffer I (Table 2.5). For double digests, the enzyme requiring lowest ionic strength was used first, and then, if required, the ionic conditions were adjusted prior to the second digest.

TABLE 2.4

Restriction enzyme buffers

EcoRI buffer: 100 mM Tris-HCl pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$, 10 mM DTT, 0.2% (^{v}/v) Triton X-100.

used for: EcoRI, HindII, HindIII, BstI, SalI, BglII, XhoI, ClaI.

HinfI buffer: 6 mM Tris-HCl pH7.5, 6 mM $MgCl_2$, 6 mM ß-mercaptoethanol, 100 mM NaCl, 100 µg/ml BSA.

used for: AvaII, RsaI, HinfI, AluI, HaeIII, Sau3A, MboI, BstNI, TaqYI, PvuII, BamHI.

KpnI buffer: 6 mM Tris-HCl pH 7.5, 6 mM $MgCl_2$, 6 mM ß-mercaptoethanol, 6 mM NaCl.

SmaI buffer: 10 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 20 mM KCl, 1 mM DTT.

PstI buffer: 20 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM $(NH_4)_2SO_4$, 100 µg/ml BSA.

HhaI buffer: 50 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 0.5 mM DTT.

Ligase buffer: 66 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 50 mM NaCl, 1 mM ATP.

Preparative digests contained 100 - 500 µg/ml DNA and were terminated by phenol - chloroform extraction, ether extraction and occasionally ethanol precipitation prior to electrophoresis.

Preparation of DNA

Extraction of plasmid DNA from E. coli (Clewell and Helinski, 1969; Radloff et al., 1967). 100 - 500 ml cultures of bacteria were grown in L-broth shaking at 37°C under appropriate antibiotic selection for 16 hours. The cells were collected by centrifugation in a Sorvall GSA rotor, 5,000 rpm for 10 minutes at 4°C. The pellets were thoroughly drained and resuspended in 10 mls of 50 mM Tris-HCl pH 8.0, 25% (w/v) sucrose containing 10 mg/ml lysozyme (Sigma). After 10 minutes at room temperature, 2.5 ml of 0.5 M EDTA pH 8.0 were added, followed 10 minutes later by 20 ml of ice - cold 50 mM Tris-HCl pH 8.0, 50 mM EDTA containing 0.3% (v/v) Triton X-100. After 90 minutes on ice, the samples were transferred to Oak-Ridge centrifuge tubes and centrifuged at 16,000 rpm for 60 minutes at 4°C in the Sorvall SS-34 rotor. The supernatants were carefully decanted, and adjusted to 31.5 ml with TE. 30 g of CsCl was added, followed by 1.5 ml of 10 mg/ml ethidium bromide. The samples were then centrifuged in Beckman Quick - Seal tubes in a VTi50 rotor at 45,000 rpm for 16 hours at 20°C. The plasmid band was removed by side puncture, extracted twice with TE saturated butan-1-ol to remove ethidium bromide, and then dialyzed for 1 hour against 200 volumes of TE to reduce

the CsCl concentration. The DNA was then twice extracted with phenol and chloroform, twice with butan-1-ol and then ethanol precipitated. The DNA was collected by centrifugation, washed once with 70% ethanol, briefly vacuum dried, and resuspended in TE at a concentration of 100 - 500 µg/ml.

Typical yields were 2 - 4 mg plasmid DNA per liter of culture. Occasionally, the DNA was rebanded to further purify it by transferring the DNA extracted from the CsCl gradient to a fresh 5 ml Quick - Seal tube and centrifuging for 5 hours at 50,000 rpm in a Beckman VTi65 rotor.

"Mini-scale" plasmid preparation (Birnboim and Doly, 1979). 1.5 ml of a saturated culture of E. coli grown in L-broth containing the appropriate antibiotic was spun down in an Eppendorf tube and the medium aspirated off. The cells were resuspended in 100 µl of 25 mM Tris-HCl pH 8.0, 10 mM EDTA, 10% (^{w}/v) glucose, 5 mg/ml lysozyme and incubated for 10 minutes at room temperature. 200 µl of 0.2 M NaOH, 1% (^{w}/v) SDS was then added and the tubes mixed by gentle inversion. After incubating for 5 minutes on ice, 150 µl of 3 M sodium acetate (adjusted to pH 4.8 with acetic acid) was added and after gentle mixing the tubes were incubated on ice for 30 minutes. The bacterial chromosome and cell debris were pelleted by centrifuging for 10 minutes in a microfuge. 400 µl of supernatant was carefully removed and added to 1 ml of ethanol in a fresh

tube, mixed and placed in dry ice - ethanol for 10 minutes. After centrifuging for 5 minutes the pellets were washed once with 70% (^{v}/v) ethanol, briefly vacuum dried and resuspended in 50 µl of TE. 5 - 10 µl of DNA were used per restriction digest. Occasionally, the DNA was phenol - chloroform extracted prior to digestion. When small fragments (less than 1 kb) were expected, DNAse - free RNAse was added to restriction digests at a concentration of 100 µg/ml.

Preparation of high molecular weight mammalian DNA (Gross-Bellard et al., 1973). The medium was aspirated from five 9 cm dishes of cells, which were then washed three times with 10 mls of TD and then drained well. They were lysed in situ with 5 ml of 10 mM Tris-HCl pH 8, 10 mM EDTA, 10 mM NaCl, 1% (^{w}/v) SDS and 100 µg/ml Proteinase K, and incubated overnight at 37°C, and then extracted three times with phenol - chloroform and twice with ether. RNAse A was added to 100 µg/ml and the sample incubated at 37°C for 6 hours. Proteinase K and SDS were added to 100 µg/ml and 0.5% (^{w}/v) respectively, and incubation continued for a further 16 hours. The DNA was then extracted twice with phenol - chloroform, twice with ether and then dialyzed overnight against several changes of TE at 4°C.

Great care was taken at all stages in the procedure never to vigorously mix the preparation, or to pass it through a narrow opening, in order to minimize shearing of the DNA.

Extraction of DNA from gels (Smith, 1980)

Agarose gels: A gel of appropriate percentage to give maximal resolution of the desired fragment was run (Section 2.5), and after staining with ethidium bromide the bands were visualized on a long - wave U.V. (365 nm) transilluminator. The desired bands were excised from the gel and the DNA purified by hot phenol extraction or electroelution.

For hot phenol extraction, the volume of the gel slice was estimated, and it was then transferred to a 1.5 ml Eppendorf tube. One volume of TE followed by one volume of neutral phenol was added. The gel slice was then melted by incubation at 65 - 70°C with occasional mixing, and the aqueous and organic phases separated by centrifugation for 5 minutes. The aqueous phase was transferred to a fresh tube, and the organic phase re-extracted with 1 volume of TE. The combined aqueous phases were then extracted twice with phenol - chloroform, once with ether and the DNA ethanol precipitated. Generally, gels were made using Rose Chemicals Agarose, gelling temperature 36°C.

For purification by electroelution, the gel slice was sealed in a dialysis bag with 1 - 2 ml of electrophoresis buffer and placed into a gel tank containing sufficient buffer to make electrical contact with the bag, but not cover it. Electrophoresis was at 100 volts for 1 - 5 hours. The polarity of the power pack was then reversed

for 45 seconds and the buffer containing the DNA removed from the dialysis bag. The bag was washed out with fresh buffer. The pooled eluate and wash were concentrated by repeated extraction with butan-1-ol, and then extracted twice with phenol - chloroform, once with ether and ethanol precipitated.

Acrylamide gels (See section 2.5): The desired band was located either by autoradiography or staining with ethidium bromide, excised from the gel and eluted as described by Maxam and Gilbert (1980). The gel slice was crushed in an Eppendorf tube and 5 volumes of 0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, 0.1% ($^{w}/v$) SDS were added. After incubation overnight at 37°C, gel fragments were removed by centrifugation, and the DNA extracted twice with phenol and chloroform, once with ether and ethanol precipitated.

Preparation of cytoplasmic, polyadenylated RNA (Favaloro et al., 1980)

Cells were harvested at 25 - 50% confluence. The monolayer was washed twice with ice - cold TD and the cells scraped off with a rubber policeman. The cells were collected by centrifugation at 2,000 rpm for 4 minutes at 4°C in the Sorvall HB-4 rotor, and washed twice in TD. The cells were lysed by vortexing for 30 seconds in 10 mM Tris-HCl pH 8.5, 1.5 mM $MgCl_2$, 140 mM NaCl, 0.5 mM spermidine, 0.15 mM spermine (buffer A) containing 0.5% ($^{v}/v$) NP-40 (5 ml per 10 dishes). This was then underlayered with an equal volume of buffer A containing

30% (^{w}/v) sucrose and 1% (^{v}/v) NP-40, and nuclei removed by centrifugation at 8,000 rpm for 15 minutes at 4°C. The upper phase was removed and added to an equal volume of 200 mM Tris-HCl pH 7.5, 25 mM EDTA, 300 mM NaCl, 2% (^{w}/v) SDS and 200 µg/ml Proteinase K. After 30 minutes at 37°C the sample was extracted twice with neutral phenol mix (50:48:1 :: neutral phenol:chloroform:isoamyl alcohol), precipitated with ethanol and stored at -20°C until required.

Oligo-dT cellulose fractionation of RNA (Aviv and Leder, 1972).

The RNA pellet was collected by centrifugation, washed once in 70% ethanol and resuspended in a small volume of water. The RNA solution was then adjusted to 10 mM Tris-HCl pH 7.6, 10 mM EDTA, 500 mM NaCl, 0.1% (^{w}/v) SDS (binding buffer) and passed four times over oligo-dT cellulose. The column was then washed with 10 column volumes of binding buffer, 10 column volumes of 10 mM Tris-HCl pH 7.6, 10 mM EDTA, 200 mM NaCl, 0.1% (^{w}/v) SDS, and finally poly A^{+} RNA was eluted with five column volumes of 10 mM Tris-HCl pH 7.6, 10 mM EDTA, 0.1% (^{w}/v) SDS. 0.2 column volume fractions were collected and 1 µl of each was spotted and dried onto 1% agarose gels containing 1 µg/ml ethidium bromide. Peak fractions were located by viewing on a 365 nm UV transilluminator, pooled and ethanol precipitated. Columns were packed in 2 ml disposable syringes. Oligo-dT cellulose was washed in

binding buffer, and after settling and decantation to remove fines it was poured into a column. Columns were prepared for re-use by washing with 10 column volumes of 0.1 M NaOH followed by equilibration with binding buffer. Columns were stored at 4°C in 10 mM Tris-HCl pH 7.6, 10 mM EDTA, 0.1% (w/v) SDS.

Sometimes, peak poly A^+ fractions were adjusted to 0.5 M NaCl and re-chromatographed on the column prior to precipitation to reduce the amount of non-polyadenylated RNA contamination.

^{3}H poly U assay (Bishop et al., 1974).

An aliquot of each RNA to be assayed was diluted into 0.2 ml of 2 x SSC. A saturating amount of ^{3}H poly U (sp. act. 1.53 mCi/mg, Amersham) was added and the sample incubated at 42°C for 30 minutes. The samples were chilled in ice water, then 25 µg RNAse A was added and the reactions left in ice water for 30 minutes. 20 µg of sonicated salmon sperm DNA was added and the sample TCA precipitated (Section 2.3). RNA concentrations were calculated assuming a 1:1 correspondence between poly A and poly U and a 7.5% representation of poly A in mRNA.

In general, the following precautions were taken to avoid ribonuclease contamination. All glassware was rinsed with 0.1% (v/v) diethylpyrocarbonate, and baked overnight at 180°C. All buffers and solutions used were autoclaved prior to use, and gloves were worn at all times when handling RNA.

Labelling of nucleic acids

Nick translation (Rigby et al., 1977). A typical reaction contained, in a total of 20 µl, 200 ng DNA, 20 pg DNAse I, 1 unit E. coli DNA polymerase I, 50 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 0.5 µg/ml BSA. Unlabelled triphosphates were present at a concentration of 50 µM, while ^{32}P-labelled triphosphates were present at 3 µM (spec. act. 410 Ci/mmol, Amersham International).

The reaction was allowed to proceed at 15°C for 90 - 120 minutes and was terminated by addition of EDTA to 10 mM. Following addition of 25 µg of salmon sperm DNA, the labelled DNA was separated from unincorporated triphosphates by passage through Sephadex G-50 in TNES. Probe containing fractions were pooled, ethanol precipitated, redissolved in TE and stored at -20°C. Specific activities of 0.5 - 5 x 10^8 cpm per µg DNA were obtained. Prior to use DNA polymerase I was diluted in 12.5 mM Tris-HCl pH 7.5, 2.5 mM $MgCl_2$, 50 µg/ml BSA, 14 mM ß-mercaptoethanol, 50% (^{v}/v) glycerol.

5' end labelling with polynucleotide kinase (Maxam and Gilbert, 1980). DNA fragments were dephosphorylated by incubation with calf intestinal alkaline phosphatase in a buffer containing 50 mM Tris-HCl pH 8.0, 0.1 mM EDTA for 30 minutes at 37°C. 0.01 units of enzyme were used per pmol of 5' ends. For blunt ends, 20 fold more enzyme was used, incubating at 50°C. EDTA was then added to 10 mM and SDS to 0.1% (^{w}/v) and the reaction heated at 65°C for

1 hour, followed by three phenol - chloroform extractions and ethanol precipitation.

The dephosphorylated DNA was resuspended at a concentration of 5 - 50 µg/ml in a buffer containing 70 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT, 100 mM KCl and 2 mCi/ml of γ -^{32}P ATP (sp. act. >3000 Ci/mmol, Amersham). T4 polynucleotide kinase was added to 200 units/ml and the reaction incubated for 1 hour. EDTA was then added to 25 mM and the sample extracted with phenol - chloroform and ether.

3' labelling with terminal transferase (Tu and Cohen, 1980). Final reaction conditions were 50 - 100 µg/ml DNA in a buffer containing 30 mM Tris, 140 mM Cacodylic acid, 1 mM DTT adjusted to pH 7.4 with KOH. $CoCl_2$ was added to 1 mM just prior to adding enzyme to 1000 U/ml. Cordycepin α-^{32}P-5'-triphosphate (5000 Ci/mmol, NEN) was included at 2 mCi/ml. The reaction was incubated at 37°C for 60 minutes and was terminated by addition of EDTA to 10 mM followed by extraction with phenol - chloroform and ether.

3' labelling and end filling using AMV Reverse Transcriptase. Final reaction conditions were 50 mM Tris-HCl pH 8.3, 6 mM $MgCl_2$, 5 mM DTT, 50 mM KCl, 100 µg/ml BSA. Unlabelled triphosphates were included at 500 µM, labelled triphosphates at 2 mCi/ml (Spec. act. 3000 Ci/mmol, Amersham). Enzyme was added to 1000 units/ml and incubation was at 42°C for 1 hour, followed by addition of EDTA to 10 mM and extraction with phenol - chloroform, and ether.

Kinase labelling of RNA (Milstein et al., 1974). RNA was ethanol precipitated and redissolved at 100 - 200 µg/ml in 50 mM Na_2CO_3. After 30 minutes at 42°C, sodium acetate (pH5.0), was added to 0.25 M and the RNA was ethanol precipitated. Final reaction conditions used for labelling were the same as for DNA.

Section 2.5: Electrophoretic techniques.

Electrophoresis of DNA

Neutral agarose gels (Favaloro et al., 1980). Gels were cast and run in TAE buffer (Table 2.5). Gels contained 0.5% - 2% ($^{w}/v$) agarose (Rose or Seakem) and sometimes 1 µg/ml ethidium bromide. For electrophoresis gels were run submerged at 1 - 3 V/cm for 6 - 16 hours. If required, gels were stained in 1 µg/ml ethidium bromide (Sharp et al., 1973) prior to viewing on a U.V. transilluminator.

Polyacrylamide gels. Gels [3% - 10% ($^{w}/v$) acrylamide; 30:1 ($^{w}/w$) :: acrylamide:bis-acrylamide] were cast and run in TBE buffer (Table 2.5) at 3 V/cm for 16 - 24 hours, fixed in 10% ($^{v}/v$) acetic acid and dried on a Bio-Rad gel dryer. Samples were mixed with $^{1}/10$ volume of gel buffer I (Table 2.5) prior to loading.

Denaturing acrylamide gels (Sanger and Coulson, 1978). Gels, [6% - 15% ($^{w}/v$) acrylamide; 25:1 ($^{w}/w$) :: acrylamide:bis-acrylamide; 40 cm x 20 cm x 0.3 mm]

containing 7 M urea (Aristar Grade, BDH) were cast and run in TBE buffer (Table 2.5) at 45 W constant power until the tracking dyes had migrated a suitable distance. Prior to loading, samples were heated to 100°C in gel buffer II (Table 2.5) for three minutes. After electrophoresis, gels were fixed in 10% (v/v) acetic acid, transferred to Whatman 3 MM paper and dried.

Acrylamide was deionized by stirring with mixed - bed resin AG 501 - X8(D) (Bio-Rad) for 1 hour, filtered through Whatman 3 MM paper and stored at 4°C in the dark.

Electrophoresis of RNA (Lehrach et al., 1977).

Gels were cast and run in 20 mM MOPS (pH 7.0), 5 mM sodium acetate, 1 mM EDTA containing 2.2 M formaldehyde (Koch - Light). RNA samples were precipitated with ethanol and resuspended in running buffer containing 2.2 M formaldehyde and 50% (v/v) recrystallized formamide (Rose). After heating to 60°C for 5 minutes to denature the RNA, the sample was loaded onto the gel, and subjected to electrophoresis at 3 - 5 V/cm for 4 - 8 hours. If desired, bands were visualized by staining in 50 mM NaOH containing 1 µg/ml ethidium bromide, and destained for 30 minutes in 0.2 M acetic acid before viewing on a U.V. transilluminator.

Alternatively, RNA was denatured using glyoxal prior to electrophoresis (McMaster and Carmichael, 1977). RNA samples were precipitated with ethanol and resuspended in 5 µl of water. 20 µl of 1 M glyoxal, 50% (v/v)

recrystallized formamide in 10 mM sodium phosphate buffer, pH 7 was added, and the sample incubated at 50°C for 1 hour. The sample was then loaded onto 1 - 1.5% (W/v) agarose gels cast and run in 10 mM sodium phosphate buffer, pH 7.0. Gels were run at 4 - 5 V/cm and the buffer was mixed every 10 minutes. Before use crude glyoxal was purified by repeated passage over mixed bed resin (Bio-Rad AG 501-X8D) until the pH was > 5.5. This treatment serves to remove glyoxylic acid contaminating the crude glyoxal solution.

Electrophoresis of proteins (Laemlli, 1970).

Gels were cast as vertical slabs according to the recipes given in Table 2.6. After pouring, the separating gel was overlaid with butan-1-ol. When the gels had polymerized, the butanol was washed off, the stacking gel poured and the comb inserted. When the gel had set, the comb was removed, and after flushing the wells out with running buffer (Table 2.6) the samples were loaded.

Prior to loading, samples were mixed with an equal volume of 2 x sample buffer (Table 2.6) and boiled for 3 minutes. Electrophoresis was at 80 V for 16 hours. Gels were stained for 1 hour in 25% (V/v) propan-2-ol, 10% (V/v) acetic acid containing 0.05% Coomassie Brilliant Blue (Bio-Rad), destained in 25% (V/v) methanol, 10% (V/v) acetic acid and dried. Occasionally gels were silver stained by the method of Sammons et al., (1981). Gels were sequentially incubated in 50% (V/v) ethanol,

TABLE 2.5.

Buffers used in electrophoresis of nucleic acids

TAE buffer: 40 mM Tris base, 5 mM sodium acetate, 1 mM EDTA, adjusted to pH 7.8 with acetic acid.

TBE buffer: 89 mM Tris base, 89 mM boric acid, 2.8 mM EDTA.

Loading buffer I: 50% ($^{w}/v$) sucrose, 100 mM EDTA (pH 8.0), 0.05% ($^{w}/v$) Xylene cyanol, 0.05% ($^{w}/v$) Bromophenol blue.

Loading buffer II: 99% ($^{v}/v$) formamide, 25 mM EDTA, 0.05% ($^{w}/v$) Xylene cyanol, 0.05% ($^{w}/v$) Bromophenol blue.

TABLE 2.6

Composition of protein gels.

	% gel		
	15%	8.5%	5% stacking gel
30% ($^{w}/v$) Acrylamide (ml)	15	8.5	1.67
2.5% ($^{w}/v$) bis-Acrylamide (ml)	1.04	2.0	0.65
1 M Tris-HCl pH 8.8 (ml)	11.2	11.2	--
1 M Tris-HCl pH 6.8 (ml)	--	--	1.25
Water (ml)	2.6	8.2	6.43
20% ($^{w}/v$) SDS (ml)	0.15	0.15	0.05

Polymerization was initiated by adding 100 µl of 10% ($^{w}/v$) ammonium persulfate and 10 µl of TEMED.

Running buffer: 25 mM Tris base, 192 mM glycine, 0.1% ($^{w}/v$) SDS.

Sample buffer (1 x): 62.5 mM Tris-HCl pH 6.8, 2% ($^{w}/v$) SDS, 100 mM DTT, 10% ($^{v}/v$) glycerol, 0.01% ($^{w}/v$) Bromophenol blue.

10% (^{v}/v) acetic acid (overnight), 25%(^{v}/v) ethanol, 10% (^{v}/v) acetic acid, (2 x 2 hours), 10%(^{v}/v) ethanol, 0.5%(^{v}/v) acetic acid (2 x 2 hours), and 11 mM $AgNO_3$ (2 hours), followed by a brief rinse in water. The gels were then reduced by incubation in 0.75 M NaOH, 92mM formaldehyde, 2.3 mM $NaBH_4$ for 10 minutes, and transferred to color enhancing solution (80 mM Na_2CO_3) to develop the colored protein bands.

Section 2.6: Southern, Northern and Western blotting.

Western blotting (Towbin et al., 1979; Yewdell and Lane, unpublished).

Protein gels were prepared and run as described in Section 2.5. Proteins were transferred from 15% or 8.5% acrylamide gels to nitrocellulose paper (Schleicher and Schuell) for 2 - 6 hours at 0.8 amp. using an EC Electroblot apparatus. Transfer buffer was 25 mM Tris-HCl pH 8.3, 192 mM glycine, 20% (^{v}/v) methanol.

The papers were then incubated overnight in a solution of 3% (^{w}/v) BSA in PBS (Phosphate buffered saline: 8.1 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 2.7 mM KCL, 137 mM NaCl). The blot was then incubated overnight at 4°C with a 1:5 dilution of tissue culture supernatant from hybridoma cells in 1% (^{w}/v) BSA, 5% (^{v}/v) FCS in PBS, giving a final antibody concentration of 1 - 2 µg/ml. The blot was then washed in PBS (3 x 10 minutes), PBS containing 1% (^{v}/v) NP-40 (4 x 15 minutes) and PBS (3 x 10 minutes) at room temperature, and incubated overnight with

a 1/100 dilution of Horse - Radish Peroxidase conjugated rabbit anti - mouse Immunoglobin (DAKO), in 1% ($^{w}/v$) BSA, 5% ($^{v}/v$) FCS in PBS. The blot was washed in PBS (3 x 10 minutes), PBS containing 1% ($^{v}/v$) NP-40 (3 x 10 minutes), 0.1 x SSC containing 0.1% ($^{v}/v$) NP-40 (3 x 10 minutes) and finally PBS (3 x 10 minutes). The blot was developed using 4-chloro-1-napthol (Aldrich) [5% ($^{w}/v$) in ethanol, diluted 1/100 in PBS and filtered through Whatman 3 MM paper] to which H_2O_2 had been added to 0.006% ($^{v}/v$).

Iodinated BSA and Phosphorylase B (a gift from J. Gannon, Imperial College) and pre-stained protein gel molecular weight standards (BRL) were run on all gels to be transferred.

Northern blotting (Thomas, 1980; Wahl et al., 1979).

RNA samples were denatured and subjected to electrophoresis as described in Section 2.5. The gel was then soaked for 30 minutes in 50 mM NaOH, and neutralized in 0.1 M Tris-HCl pH 7.4, 1 x SSC for 30 minutes. It was then laid on top of two pieces of Whatman 3 MM paper on an elevated glass plate arranged such that the ends of the papers formed wicks into a reservoir of 1 x SSC.

A piece of Gene Screen (NEN) cut to the same size as the gel, and equilibrated with 1 x SSC was then laid onto the gel, taking care not to trap any air bubbles. A stack of paper cut to the same size as the gel was then laid on top of the membrane and a weight placed on top of the paper. Transfer was overnight. The membrane was then

rinsed in 2 x SSC to remove residual agarose, air dried and baked at 80°C for 2 hours.

It was then hybridized according to the manufacturer's instructions. The membrane was prehybridized overnight in 10 - 20 ml of 50% ($^{v}/v$) recrystallized formamide, 10 x Denhardt's solution, 50 mM Tris-HCl pH 7.5, 1 M NaCl, 0.1% ($^{w}/v$) sodium pyrophosphate, 0.1% ($^{w}/v$) SDS, 10% ($^{w}/v$) Dextran sulfate, 200 µg/ml denatured salmon sperm DNA at 42°C.

The bag was then opened, and a radioactive probe added. The bag was resealed and incubation continued for 24 hours at 42°C. The filter was then washed three times for 10 minutes each in 2 x SSC, 0.1% ($^{w}/v$) SDS, four times for 30 minutes each in 2 x SSC, 0.1% ($^{w}/v$) SDS at 60°C and four times for 30 minutes each in 0.1 x SSC, 0.1% ($^{w}/v$) SDS at 60°C.

The filters were then blotted dry and autoradiographed at -70°C with preflashed (Fuji X-ray) film using Ilford Tungstate Intensifying Screens, as described by Laskey and Mills (1977).

Probes were denatured by heating to 100°C for 10 minutes followed by cooling on ice. Probe concentrations of 10 ng/ml or less were used to minimize background binding.

Southern blotting (Southern, 1975; Thomas, 1980).

Agarose gels were cast and run as described in Section 2.5. The DNA was denatured in situ by incubation of the gel in 0.5 M NaOH, 1.5 M NaCl for 30 - 60 minutes

at room temperature with gentle agitation. The gel was then incubated in 1 M Tris-HCl pH 7, 3 M NaCl for 30 - 60 minutes and then laid on top of two pieces of Whatman 3 MM paper on an elevated glass plate arranged such that the ends of the papers formed wicks into a reservoir of 20 x SSC.

A piece of nitrocellulose paper, cut to the same size as the gel, and equilibrated sequentially with water, 2 x SSC and 20 x SSC was then laid on top of the gel, taking care not to trap any air bubbles. A stack of paper the same size as the gel was laid on top, and a weight placed on top of the paper. Transfer was allowed to continue for 4 - 16 hours. The membrane was then rinsed in 10 x SSC to remove residual agarose, air dried and baked at 80°C for 2 hours.

For transfer to Gene Screen (NEN), essentially the same protocol was followed except that the gel was incubated in 1 x SSC following denaturation. 1 x SSC was also used for transfer.

Filters were prehybridized in polythene bags in 10 - 20 ml of 2 x SSC, 5 x Denhardt's solution, 1% (^{w}/v) SDS containing 200 µg/ml denatured salmon sperm DNA for 4 - 16 hours at 68°C. The bag was then opened and denatured radioactive probe added. The bag was resealed and incubated at 68°C for 16 - 24 hours. Typical probe concentrations were 10 - 25 ng/ml.

Filters were washed in 1 x SSC, 0.1% (^{w}/v) SDS (3 x 5 minutes at room temperature), 1 x SSC, 0.1% (^{w}/v) SDS (4 x

30 minutes at 65°C) and finally 0.1 x SSC, 0.1% (w/v) SDS (2 x 30 minutes at 65°C), dried and autoradiographed in the same way as Northern blots.

Colony screening (Grunstein and Hogness, 1975; Hanahan and Meselson, 1980).

Colonies to be screened were plated onto sterile nitrocellulose filters (Schleicher and Schuell, BA-85) laid on L-plates containing appropriate antibiotic selection. When numbers to be screened were small, the colonies were plated out in an ordered array. When 0.1 mm diameter colonies had appeared, they were replicated onto another filter and growth continued until 0.5 mm colonies could be seen. One filter was kept at 4°C, and the other transferred to fresh plates containing 100 µg/ml chloramphenicol for 2 hours at 37°C.

When ordered arrays were to be screened the colonies were allowed to grow up overnight at 37°C.

The filters were then placed on a puddle of 0.5 M NaOH on a piece of Saran Wrap. After 5 - 10 minutes, they were transferred to a puddle of 1 M Tris-HCl pH 7.4 for 2 minutes. They were then transferred to a puddle of 0.5 M Tris-HCl pH 7.0, 1.5 M NaCl for 10 minutes, air dried and baked for 2 hours at 80°C.

Hybridization and washing of filters was performed as described for Southern transfer. Occasionally, filters were washed in 3 mM Tris base for 2 x 15 minutes at room temperature in addition to the other washes following

hybridization.

Section 2.7: Sequencing of DNA.

Chemical degradation (Maxam and Gilbert, 1977, 1980).

DNA fragments were end labelled and gel - purified as described in Section 2.5. The DNA was resuspended in 18 µl of distilled water and 2 µl of 10 mg/ml salmon sperm DNA added.

For T and C cleavage 5 µl of DNA was added to 20 µl of water and 25 µl of Hydrazine (Pierce). After 15 minutes at room temperature, 200 µl of CT stop (Table 2.7) was added together with 10 µg of E. coli tRNA and 750 µl of ethanol.

For cleavage at C alone, the same procedure as for T and C cleavage was followed except that the 20 µl of water were replaced with 20 µl of saturated NaCl.

For G cleavage, 5 µl of DNA was added to 200 µl of DMS buffer (Table 2.7) and 1 µl of DMS (Sigma) added. After 3 minutes at room temperature 110 µl was removed and added to 50 µl of G. stop (Table 2.7). After a further 12 minutes the remaining solution was transferred to the tube containing G. stop. 10 µg of E. coli tRNA and 750 µl of ethanol were then added.

For A cleavage 5 µl of DNA was added to 20 µl of 1.5 M NaOH in 1 mM EDTA (carbonate - free, freshly prepared). The sample was divided into two equal parts and sealed in glass capillary tubes. These were then heated at 90°C,

for 3 minutes or 15 minutes and pooled together in a tube containing 225 µl 0.3 M sodium acetate pH 5.5, 20 µg tRNA and 750 µl ethanol.

After incubating all samples at -70°C for 10 minutes the tubes were spun for 5 minutes. The DNA pellets were reprecipitated twice by redissolving in 200 µl of 0.3 M sodium acetate pH 5.5 followed by addition of 600 µl of ethanol. The pellets were then washed twice in 70% (^{v}/v) ethanol and dried under vacuum. They were then redissolved in 20 µl of 10% (^{v}/v) redistilled piperidine in water, sealed in glass capillary tubes, and heated to 90°C for 30 minutes. The capillaries were then opened and the contents expelled into 250 µl of water. The samples were then dried overnight under vacuum and resuspended in gel buffer II (Table 2.5) prior to electrophoresis on denaturing acrylamide gels (Section 2.5).

Dideoxy - sequencing (Sanger et al., 1977; Sanger et al., 1980).

Recombinant plaques were picked into 1.5 ml of L-broth and grown with vigorous shaking for 7 hours at 37°C. Bacteria were pelleted by spinning for 5 minutes in a microfuge, and 1 ml of the supernatant added to 0.25 ml of 20% (^{w}/v) PEG 6000, 2.5 M NaCl in a fresh tube. After 15 minutes at room temperature, phage were pelleted by centrifugation for 10 minutes, the supernatant aspirated off and the samples recentrifuged for a further 30 seconds. Residual liquid was removed and the phage pellet

resuspended in 100 µl of TE. After extraction with phenol and chloroform the single stranded DNA was precipitated with ethanol, washed once in 70% (^{v}/v) ethanol, resuspended in 25 µl TE and stored at -20°C until required.

For sequencing 4 µl of template was combined with 4 µl of primer mix (Table 2.8), heated to 70°C and slowly cooled to room temperature. 2 µl of 10 mM Tris-HCl pH 8, 5 mM $MgCl_2$ containing 0.8 units of Klenow fragment and 10 µCi of labelled nucleotide (α -^{35}S-dATP, Amersham, spec. act. 410 Ci/mmol) were added, and after mixing, 2.5 µl dispensed to each of four tubes containing 2 µl of A, C, G or T sequencing mix (Table 2.8). The samples were mixed by centrifugation and incubated at 25°C for 20 minutes. 2 µl of Chase mix (Table 2.8) was added and the incubation continued for a further 15 minutes.

4 µl of formamide dyes were then added (Buffer II, Table 2.5), and the sample boiled for 3 minutes prior to loading onto 6% (^{w}/v) denaturing acrylamide gels (Section 2.5) or stored at -20°C until subjected to electrophoresis.

Gels were cast with an additional 0.3 mm thickness of spacer covering the bottom 2 cm of the gel. The resulting "wedge" gel improves the number of bases which can be read off each run by decreasing the separation between bands at the bottom of the gel.

TABLE 2.7

Solutions used in Maxam - Gilbert sequencing

CT stop: 0.3 M sodium acetate, 0.01 M magnesium acetate, 0.1 mM EDTA adjusted to pH 5.5 with acetic acid.

G. stop: 1 M Tris-acetate pH 7.5, 1 M ß-mercaptoethanol, 1.5 M sodium acetate, 0.05 M magnesium acetate.

DMS buffer: 50 mM sodium cacodylate (pH 8.0), 10 mM magnesium chloride, 0.1 mM EDTA.

TABLE 2.8

Solutions used in dideoxy sequencing

Primer mix: 1 µl Universal Primer, 1 µg/ml (New England Biolabs)
1.5 µl 100 mM Tris-HCl pH 8.0, 100 mM $MgCl_2$
3.5 µl H_2O

Sequencing mixes

A mix: 62.5 µM dCTP, 62.5 µM dGTP, 62.5 µM TTP, 25 µM ddATP in TE.

C mix: 4.15 µM dCTP, 83.3 µM dGTP, 83.3 µM TTP, 33 µM ddCTP in TE.

G mix: 83.3 µM dCTP, 4.15 µM dGTP, 83.3 µM TTP, 100 µM ddGTP in TE.

T mix: 83.3 µM dCTP, 83.3 µM dGTP, 4.15 µM TTP, 200 µM ddTTP in TE.

Chase solution: 1 mM dATP, 1 mM dCTP, 1 mM dGTP, 1 mM TTP in TE.

CHAPTER THREE

CLONING OF CELLULAR SEQUENCES TO WHICH T ANTIGEN IS BOUND

CHAPTER THREE

APPROACHES TO CLONING CELLULAR DNA SEQUENCES WITH WHICH T ANTIGEN INTERACTS

Section 3.1: Introduction.

The experiments described in this chapter are designed to address the question posed at the conclusion of Chapter One, that is to say, how might one specifically isolate segments of DNA to which T antigen is bound in a transformed cell.

There are a variety of potential approaches which could be adopted to solve this problem. One might for instance, consider that a cellular DNA sequence with which T antigen would interact, assuming that it did so by directly binding to the DNA, would bear some of the characteristics of the sequence around the SV40 origin of replication, which T antigen is known to bind to. Thus, by probing a genomic library with a nick translated origin probe, one might hope to isolate origin - like sequences. These could then be tested for biological function and T antigen binding *in vitro*. This approach has been adopted by two groups. A screen of human DNA indicated the presence of some 200 origin - like sequences. Of two analyzed, one bound T antigen *in vitro*, while the other did not, but was shown to have enhancer - like function. Neither sequence acted as an origin of replication when re-introduced into cells (Conrad and

Botchan, 1982). A similar screen of Simian cell - DNA (McCutchan and Singer, 1981) identified only three or four such sequences, none of which bound T antigen *in vitro*. However, these sequences, and those of the previous study were found to be very G-C rich, and to be homologous to the region III (Queen *et al.*, 1981) identified by de Lucia *et al.*, (1983). This region of SV40 has been shown to contain a binding site for the promoter - specific transcription factor Sp1 (Dynan and Tjian, 1983a,b). The origin - like monkey sequences have been shown to function as a promoter *in vivo* (Saffer and Singer, 1984). No sequence analogous to the putative high affinity binding site at the origin (site I) was isolated. Both studies showed that the sequences isolated were members of a larger family of sequences, dispersed throughout the genome.

An alternative strategy would be to make use of the technique developed by McKay (1981) to study the interaction of D2T with SV40 DNA to isolate high affinity T antigen binding sequences from cellular DNA. Thus, the DNA would be fragmented with appropriate restriction enzymes, ligated to vector, and the ligation mix incubated with purified T antigen. The T antigen would then be immunoprecipitated, and the co-precipitated DNA used to transform bacteria. The alternative would be to digest the DNA, immunoprecipitate as before and then clone the co-precipitated fragments. The clones obtained by either

method would be pooled and the process repeated several times, analyzing the complexity of the pool at each stage. This type of protocol should enrich for high affinity binding sites, provided that stringent binding conditions (>150 mM NaCl; pH $\geq$7.4) and limiting amounts of T antigen are used. An excellent test system for this kind of protocol would be to try to select the SV40 origin from transformed cell DNA.

Neither of these methods addresses the question of which DNA sequences T antigen binds to in the cell nucleus, where they are assembled into chromatin. The region to which T antigen binds on the SV40 genome is the viral origin of replication. One might therefore consider that some of the sequences with which T antigen interacts *in vivo* might act as cellular origins of replication. Thus, by using a line of transformed Simian cells called cos-1, such sequences could potentially be amplified. Cos-1 cells (Gluzman, 1981) are permissive CV-1 monkey kidney cells transformed by cloned SV40 DNA with a 6 base pair deletion at the origin of replication. Therefore, the viral DNA integrated into the host cell cannot be rescued, but the cells contain functional T antigen and all the cellular factors required for viral DNA replication. If these cells are transfected with plasmid vector DNA which cannot replicate then the DNA is rapidly degraded. However, plasmids containing a functional SV40 origin can replicate by virtue of the T antigen supplied in *trans*, so the transfected DNA increases in abundance.

The strategy then would be to transfect cos-1 cells with a mixture of DNA fragments, and then analyze the replicating DNA by the procedure of Hirt (1967). Multiple cycles of transfection and extraction might be expected to enrich for sequences capable of autonomous replication in cos-1 cells. The interaction of T antigen with these sequences, and whether their replication was T antigen dependent could then be studied.

An alternative method is to search for T antigen DNA complexes in the chromatin of SV40 transformed cells. The assumption made is that such complexes will be stable at high ionic strength, as has been found for viral chromatin, where a fraction of the T antigen remains associated with the DNA after washing with salt concentrations of greater than 1 M, or treatment with dextran sulfate and heparin, both of which remove most chromatin proteins, including histones (Reiser et al., 1980; Rudolph and Mann, 1983; Stahl and Knippers, 1983; Stahl et al., 1983). That such a tightly bound fraction exists in the cell has been shown by differential extraction procedures (Staufenbiel and Deppert, 1983; Scheidtmann et al., 1984). The approach would therefore be to prepare a chromatin fraction from transformed cells, to fragment it using nucleases, and then following immunoprecipitation of nucleoprotein complexes bearing T antigen, to elute the DNA from the precipitates and clone it for further analysis. This has the merit that it

selects those sequences that T antigen is bound to *in vivo*, without imposing any pre-selection for function (e.g., replication).

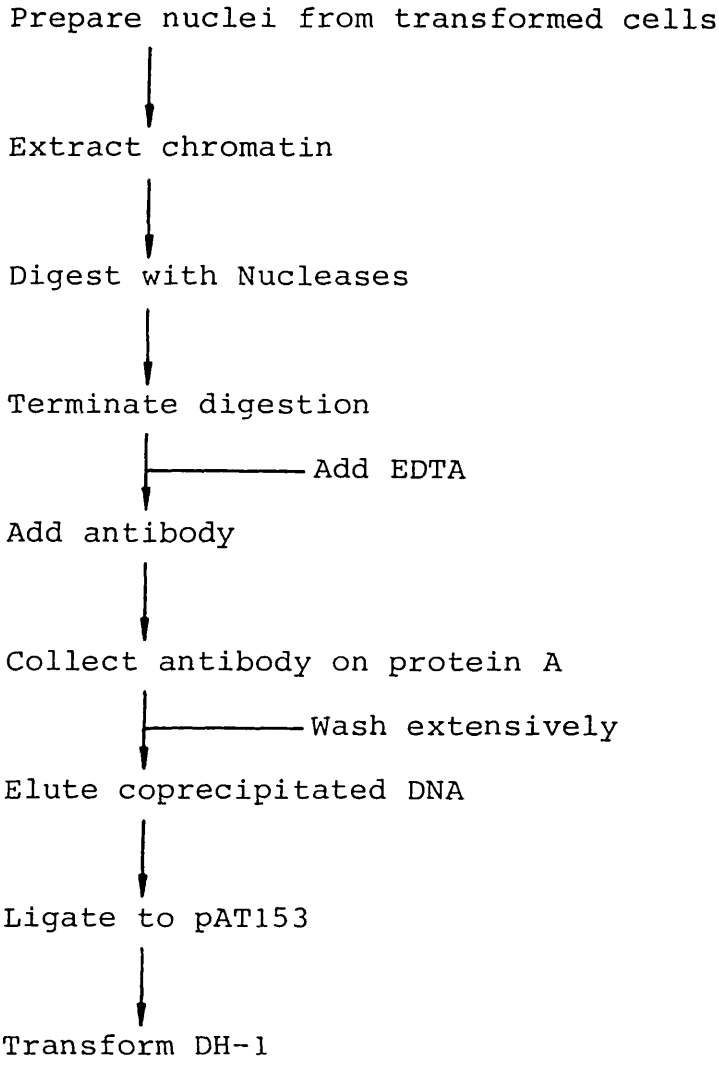
The method for fractionation of chromatin must now be considered. A variety of different techniques have been used, including sonication, digestion with DNAse I, DNAse II or micrococcal nuclease, or digestion with restriction endonucleases (reviewed by Axel, 1978; Gottesfeld, 1978). DNAse II and micrococcal nuclease digest DNA to leave a final product with 5' hydroxyl and 3' phosphate groups. The precipitated DNA would therefore require treatment with alkaline phosphatase, followed by polynucleotide kinase prior to cloning. While this problem does not arise with DNAse I, like the other two enzymes, it does not digest the DNA to leave a blunt end, therefore necessitating either end repair prior to cloning or cloning by homopolymer tailing. The use of a restriction enzyme to digest chromatin therefore seems appealing, since if an enzyme that generates a sticky end is used, the resultant fragments can be cloned directly in a bacterial vector following elution from the immunoprecipitate (blunt end ligation would not be as efficient as high concentrations of ends are required). Conditions for digestion of chromatin with restriction enzymes have been established in studies which have used them to probe chromatin structure (reviewed by Axel, 1978; Gottesfeld, 1978). It has for instance, been demonstrated that the region around the SV40 origin of replication is

more sensitive to restriction enzyme digestion than the rest of the genome, suggesting this region is not assembled into nucleosomes (Varshavsky et al., 1978; Scott and Wigmore, 1978; Varshavsky et al., 1979).

The strategy adopted to try to clone cellular DNA sequences with which T antigen interacts is diagrammed in Figure 3. Nuclei were prepared from transformed cells by detergent lysis. After washing, they were lysed by resuspension in low ionic strength hypotonic buffer to release the chromatin. The chromatin suspension was then adjusted to a buffer suitable for restriction enzyme digestion, and digested for several hours at 30°C using a mixture of Sau3A and BamHI. Digestion was then terminated by adding EDTA, and the NaCl concentration was adjusted to 0.5 M. Monoclonal antibody to SV40 large T antigen, or control antibody, was then added to precipitate the T antigen bearing nucleoprotein complexes. After elution from the immunoprecipitate by digestion with proteinase K in the presence of SDS the co-precipitated DNA was precipitated with ethanol, ligated to BamHI digested pAT153, and used to transform DH-1 to ampicillin resistance. Colonies containing inserts were identified by replica plating onto plates containing tetracycline. In one experiment, nuclei were lysed by incubation with dextran sulfate and heparin, which serves to remove most chromatin associated protein (Adolph et al., 1977). After washing to remove histones, the conditions were adjusted

Figure 3.1: Strategy for immunoselection cloning of cellular sequences to which T antigen is bound.

Strategy for Immunoselection Cloning



to those for restriction enzyme digestion and the sample immunoprecipitated as before.

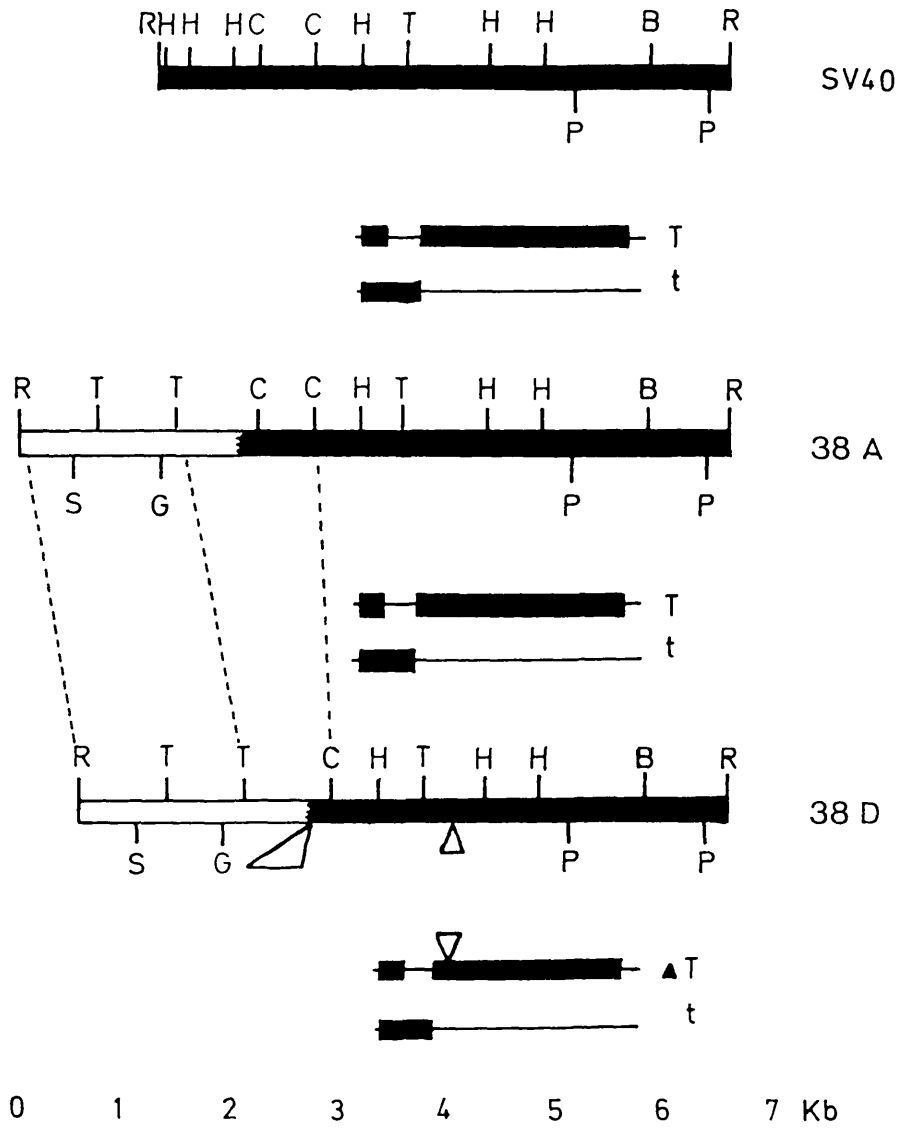
The transformed cell line chosen for studies was the SV40 transformed Balb/c 3T3 line SV3T3 Cl 38. The SV40 DNA inserts in the genome of this line have been well characterized (Rigby, 1979; Clayton and Rigby, 1981). It is known that there are three independent insertions of SV40 DNA, one of which is at least a tandem insertion of SV40. Two of the templates encode full length, replication competent T antigen, SV40 and 38A, while the third, 38D, encodes a replication defective, transformation competent T antigen (Rigby and Lane, 1983). 22% of the T antigen in this cell line is complexed with p53 (Lane et al., 1982). The structure of the integrated SV40 DNA is shown in Figure 3.2. All three templates possess a functional origin of replication which should therefore be capable of interacting with T antigen.

Genes activated in Cl 38 cells upon transformation by SV40 have recently been cloned (Scott et al., 1983). It will therefore also be of interest to determine whether any of the immunoselected DNA corresponds to the clones obtained by Scott et al., (1983).

A similar approach has been adopted by Cuzin and his colleagues to clone binding sites for polyoma large T antigen in chromatin (personal communication), using immunoselection cloning following digestion with DNAse I. They have identified a clone homologous to the MSV LTR

Figure 3.2: Structure of the known SV40 DNA inserts in the SV3T3 Cl 38 genome.

This figure is modified from Clayton and Rigby, (1981). The structures of the 38A, 38D and SV40 templates are shown. 38D is derived from 38A by two deletions, one at the virus host junction, the other in the viral early region, indicated by triangles below the template. SV40 sequences are shaded black. Restriction endonuclease abbreviations are: B, BamHI; C, CfoI; G, BglII; H, HindIII; P, PstI; R, EcoRI; S, SstI; and T, TaqYI.


R H H H C C H T H H B R
P P
SV40
T
t
R T T C C H T H H B R
S G P P
38 A
T
t
R T T C H T H H B R
S G P P
38 D
▲T
t
0 1 2 3 4 5 6 7 Kb

region which will rebind polyoma large T antigen as judged by the DNA cellulose competition assay (Gaudray et al., 1981) and McKay immunoprecipitation assay (McKay, 1981).

Section 3.2: Materials and methods.

General materials and methods used have been described in Chapter Two.

Immunoselection cloning

SV3T3 Cl 38 cells (Rigby et al., 1980) were grown to 25 - 50% confluence in DMEM supplemented with 2% (^{v}/v) FCS. The medium was aspirated from the cells, and after washing twice with ice cold TD, the cells were removed from the dish using TD supplemented with 5 mM EDTA. They were then washed in TD twice by centrifugation at 2,000 rpm, 4°C, for 4 minutes in the Sorvall HB-4 rotor, and lysed by resuspension in 2 ml/dish of Solution I [10 mM Tris-HCl pH 7.4, 120 mM NaCl, 1 mM $MgCl_2$, 1 mM EGTA, 0.1 mM DTT, 0.25% (^{v}/v) Triton X-100, 1 mM PMSF]. After agitation for 20 seconds on a vortex mixer, the nuclei were collected by centrifugation at 2,500 rpm, 4°C, for 5 minutes in the swing out rotor, as before. The nuclei were washed twice in Solution I and then examined by phase contrast microscopy to ensure that whole cells and cytoplasmic material were not present. The nuclei were then washed twice in Solution I without Triton to remove detergent. After the last wash, care was taken to remove all the supernatant. The nuclei were lysed by gentle

resuspension in 0.5 ml/dish of 0.25 mM EDTA (pH 8.0). After 20 minutes in an ice - water bath, the chromatin was adjusted to digestion conditions appropriate for Sau3A, final concentrations being 10 mM Tris-HCl pH 7.4, 50 mM NaCl, 5 mM $MgCl_2$, 0.1 mM DTT, 1 mM EGTA, 1 mM PMSF and 20 units Sau3A and 20 units BamH1 added per OD_{260} unit (OD_{260} was read in 0.1 M NaOH). EGTA is added to chelate Ca^{2+} ions thereby inhibiting endogenous nuclease activity during the incubation. Digestion was for 2 - 3 hours at 30°C. At the end of this time, digestion was stopped by adding EDTA to 10 mM. The concentration of NaCl was adjusted to 0.5 M, NP-40 was added to 0.05% (^{v}/v) and 3 µg of monoclonal antibody was added. For experiments one and two, the antibody used was added as hybridoma cell supernatant (100 µl) which had been heated to 56°C for 60 minutes to inactivate nucleases. The second antibody (rabbit anti-mouse immunoglobulin) necessary with PAb1005, which is of the IgG 1 subclass and so binds protein A poorly, was treated similarly.

After incubation overnight at 4°C, 10 µl of a 1:1 slurry of protein A Sepharose beads in Solution I was added. Three hours later the beads were collected by centrifugation for 2 seconds in a microfuge, and washed four times with Solution I. They were then resuspended in 50 µl of Solution II [50 mM Tris-HCl pH 7.4, 100 mM NaCl, 5 mM EDTA, 1% (^{w}/v) SDS, 100 µg/ml Proteinase K (Merck)] and incubated at 37°C for 1 hour to release the DNA from

the protein complexes. The sample was then extracted three times with phenol and chloroform, twice with ether, and following addition of E. coli tRNA (Sigma) to 10 µg/ml as carrier and $MgSO_4$ to 10 mM, precipitated with ethanol overnight at -70°C. The DNA was collected by centrifugation for 15 minutes at 4°C in an Eppendorf microcentrifuge, washed once, dried and ligated to 20 ng BstI digested pAT153, treated with alkaline phosphatase to prevent self - ligation, in a volume of 10 µl overnight at 14°C. The DNA was then used to transform DH-1 to ampicillin resistance.

This method is based on the methods developed for chromatin preparation and digestion with restriction endonucleases described in Gottesfeld (1978) and Stein and Stein (1978). For Experiment Three, the observation of Reiser et al., (1980) that T antigen forms a heparin - resistant complex with the origin of replication in the SV40 minichromosome was exploited. Nuclei were prepared as described for the standard method and lysed by resuspension in 1 ml/dish of Solution III [10 mM Tris-HCl pH 7.4, 10 mM EDTA, 1 mg/ml heparin (Sigma), 100 µg/ml dextran sulfate (Sigma), 1 mM PMSF]. After 5 minutes on ice, the DNA was pelleted by centrifugation at 12,000 rpm for 20 minutes at 4°C in the Sorvall HB-4 rotor. The DNA was then washed twice with Sau3A buffer and resuspended in 0.5 ml Sau3A buffer and digested for 90 minutes at 30°C using 20 units of BamHI and 20 units of BglII per OD_{260}, and the sample processed as before. The use of dextran

sulfate and heparin to remove chromatin proteins is based on the method of Adolph et al., (1977).

In general, 1 - 2 OD_{260} units were used per precipitation. Purified monoclonal antibodies were used except in Experiment One. Antisera and monoclonal antibodies were a gift from J. Gannon, Imperial College, London.

Section 3.3.

Nomenclature

Clones derived from Experiment One have numbers 0 - 199, those from Experiment Two 200 - 299 and those from Experiment Three 301 - 399. Experiments One and Two used a combined BamHI and Sau3A digestion of chromatin, while in Experiment Three BamHI and BglII were used to digest chromatin depleted by washing with dextran sulfate and heparin.

Results

There is some evidence, discussed in Chapter One, that T antigen can interact with the SV40 origin of replication in a transformed cell. One might therefore expect that if the immunoselection protocol is successful, the DNA co-precipitated using antibodies specific for T antigen should contain viral origin sequences, while control antibodies which do not react specifically with any protein in mouse cells should not.

To test this an experiment was performed as described in Section 3.2 and the co-precipitated DNA was immobilized on nitrocellulose and probed with the HindIII - PvuII fragment spanning SV40 coordinates 5171 - 270, which encompasses the origin of replication. The result is shown in Figure 3.3. From lanes 8 - 10, it is apparent that origin homologous sequences are present only when an antibody reacting with T antigen [PAb419, (Harlow et al., 1982)] was used to immunoselect DNA. PAb419 was chosen as the immunoprecipitating antibody since it has been shown not to interfere with the DNA binding (Clark et al., 1983) activity of large T antigen. Neither second antibody alone, nor a control antibody PAb1005 (Thomas et al., 1982) which reacts with human but not murine p53 precipitated any origin sequences. Second antibody is necessary if PAb1005 is to be collected on protein A Sepharose beads, since being of the IgG 1 subclass, it binds to protein A weakly. This finding demonstrates that the immunoselection protocol is showing the required selectivity for sequences to which T antigen is bound.

The experiment was therefore repeated, and the immunoselected DNA was ligated into the BamHI site of pAT153. Colonies obtained following transformation of DH-1 were screened by replica plating onto ampicillin and tetracycline containing plates. Colonies which had an ampicillin resistant, tetracycline sensitive phenotype were selected for further analysis. Colony numbers obtained are shown in Table 3.1. A small number of

Figure 3.3: Dot blot analysis of immunoselected DNA.

DNA was immunoselected as described in Section 3.2, boiled for three minutes, quenched on ice, adjusted to 10x SSC and filtered through nitrocellulose using a BRL hybridot manifold. The sample wells were then washed with 3 x 0.2ml of 10x SSC, and the filter processed and hybridized as described in Section 2.6. Samples are: 1 - 6, 50, 1000, 200, 50, 20 and 5 pg of SV40 DNA, respectively; 7, 5µg of C1 38 DNA; 8 - 10, DNA immunoselected by precipitation with Rabbit antibody to mouse immunoglobulin (8), PAb1005 + Rabbit antibody to mouse immunoglobulin (9), and Rabbit antibody to mouse immunoglobulin + PAb419 (10), respectively.

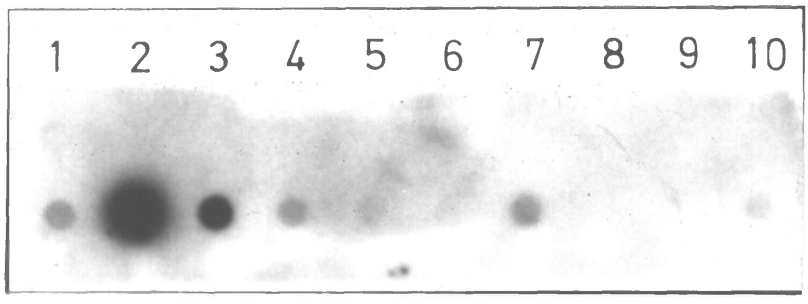

1
2
3
4
5
6
7
8
9
10

Table 3.1: Summary of data from immunoselection experiments.

The number of tetracycline sensitive colonies obtained following immunoselection with the indicated antibodies is shown. The number in brackets is the number of colonies judged to have inserts, where determined. N. D., not determined, Rb α MIg, Rabbit antibody to mouse immunoglobulins.

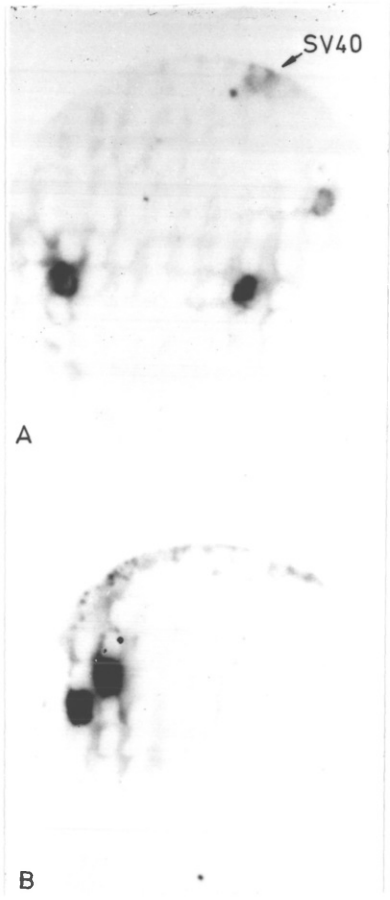
	Number of tet^S colonies(with inserts)		
	Antibody		
Experiment	PAb419	PAb1005	Rb×MIg
1	87 (64)	8 (0)	3 (0)
2	55 (53)	2 (0)	3 (0)
3	98 (ND)	0	ND

tetracycline sensitive colonies were obtained with control antibodies. On analyzing the DNA from these transformants, however, no insert could be detected in the plasmid. The probable reason for this observation is that the BstI enzyme (an isoschizomer of BamHI) used to digest pAT153 to prepare it for cloning contained a trace amount of some contaminating nuclease activity. The alternative explanation is that the calf intestinal alkaline phosphatase used contained nuclease activity. Analysis of the tetracycline sensitive colonies obtained using PAb419 selected DNA also showed that a small number had no detectable insert.

The numbers of tetracycline sensitive colonies with inserts obtained with PAb419 immunoselected DNA was consistent, and small. The earlier dot blot analysis (Figure 3.3) had suggested that the SV40 origin was being specifically immunoselected with PAb419 but not with control antibodies. To investigate whether the origin was amongst those clones obtained, tetracycline sensitive colonies with inserts were plated out in an ordered array, and grown on nitrocellulose filters. The colonies were then lysed according to the procedure of Hanahan and Meselson (1980) and the DNA on the filters hybridized with nick translated SV40 DNA, as described in Section 2.6. The result obtained is shown in Figure 3.4. Filter A has the colonies from Experiments One and Two while filter B has those from Experiment Three. There are clear positive signals from all three experiments. Experiment One has

Figure 3.4: Colony blot of immunoselected clones using an SV40 probe.

Colony hybridization of clones derived by immunoselection with PAb419. The probe was nick translated SV40 DNA. The arrow indicates the position of a colony bearing pTSV3, a clone of SV40 in the EcoRI site of pAT153.


SV40
A
B

given one positive, Experiments Two and Three, two each. These colonies were grown up and the DNA from 2 ml cultures extracted and analyzed by digestion with various restriction enzymes and Southern blotting. Figures 3.5 to 3.9 show the results obtained. For clones 106, 200 and 207, Sau3A digestion gives only one band in addition to those derived from vector, at 1350 bp (Figure 3.5, lanes 3 - 5), comigrating with the largest Sau3A fragment of SV40 (lane 6, Figure 3.5). The HindIII and EcoRI digests are consistent with these clones being an insertion of the Sau3A A fragment of SV40 at the BamHI site of pAT153. The homology of this inserted fragment with SV40 was confirmed by hybridizing a Sau3A digest of clones 106, 200 and 207 with nick translated SV40 (Figure 3.6). A single hybridizing band of 1350 bp length is seen, in excellent agreement with the size of the Sau3A A fragment of SV40, 1347 bp (Tooze, 1981). The HindIII digests demonstrate that clones 106 and 207 have the insert in the same orientation, while clone 200 has the opposite orientation (Figure 3.9).

Clones 323 and 336 would seem to represent full length clones of SV40 at the BamHI site of pAT153 as judged by restriction enzyme analysis and Southern blotting . Both clones give bands of 5.2 and 3.65 kb when digested with BamHI, of which only the 5.2 kb band is SV40 homologous (Figure 3.7, lanes 10 and 11; Figure 3.8, lanes 3 and 4).

Figure 3.5: SV40 origin clones, I.

Samples of clones 106, 200, 207, 323 and 336 were digested with Sau3A and subjected to electrophoresis through a 1% ($^{w}/v$) agarose gel. Lanes 1 - 5, clones 336, 323, 207, 200 and 106 respectively; Lane 6, pTSV3 digested with Sau3A; Lane 7, marker.

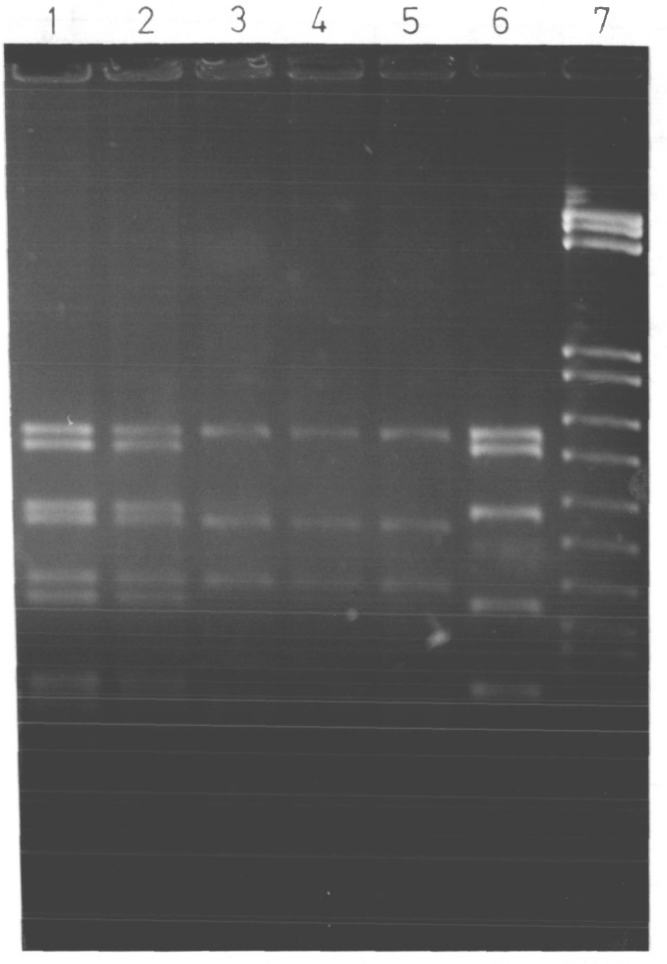

1
2
3
4
5
6
7

Figure 3.6: Southern blot of clones 106, 200 and 207.

Samples of clones 106, 200 and 207 were digested with Sau3A, subjected to electophoresis, transferred to nitrocellulose and hybridized with nick translated SV40 DNA. M, markers; Lanes 1 - 3, clones 106, 200 and 207, respectively.

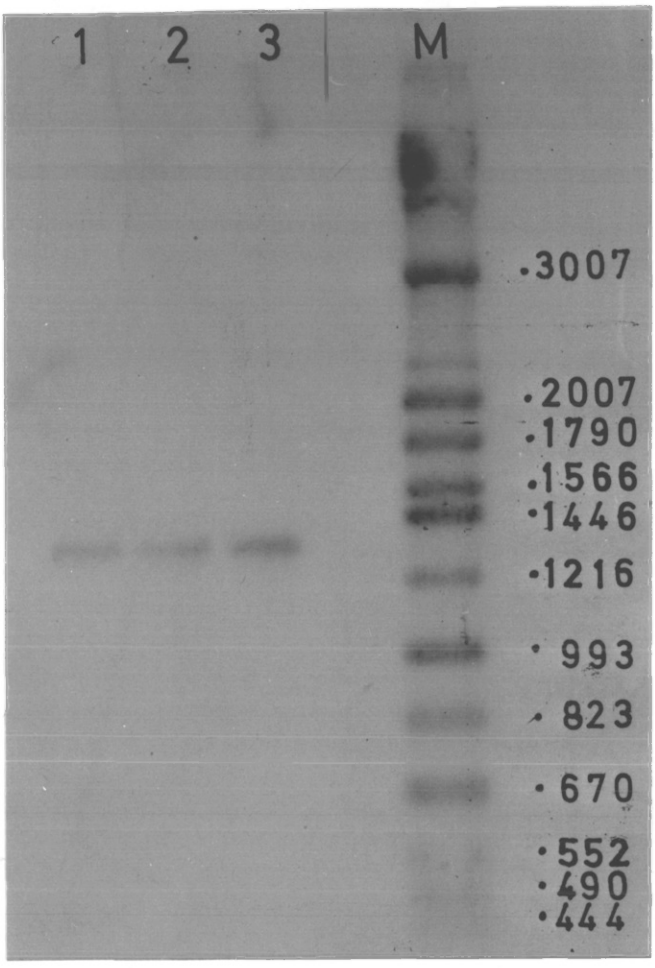

1
2
3
M
·3007
·2007
·1790
·1566
·1446
·1216
· 993
· 823
· 670
·552
·490
·444

Figure 3.7: Agarose gel of SV40 origin clones, II.

Samples of the SV40 origin clones were digested either with EcoRI, BamHI or HindIII and subjected to electrophoresis as described in the legend to Figure 3.5.

Lanes 1, 2 and 9, markers; lanes 3, 4, and 5, HindIII digests of clones 106, 200 and 207, respectively; lanes 6, 7, and 8, EcoRI digests of clones 106, 200, and 207, respectively. Lanes 10 and 11, BamHI digests of clones 323 and 336; Lanes 12 and 13, EcoRI digests of clones 323 and 336.

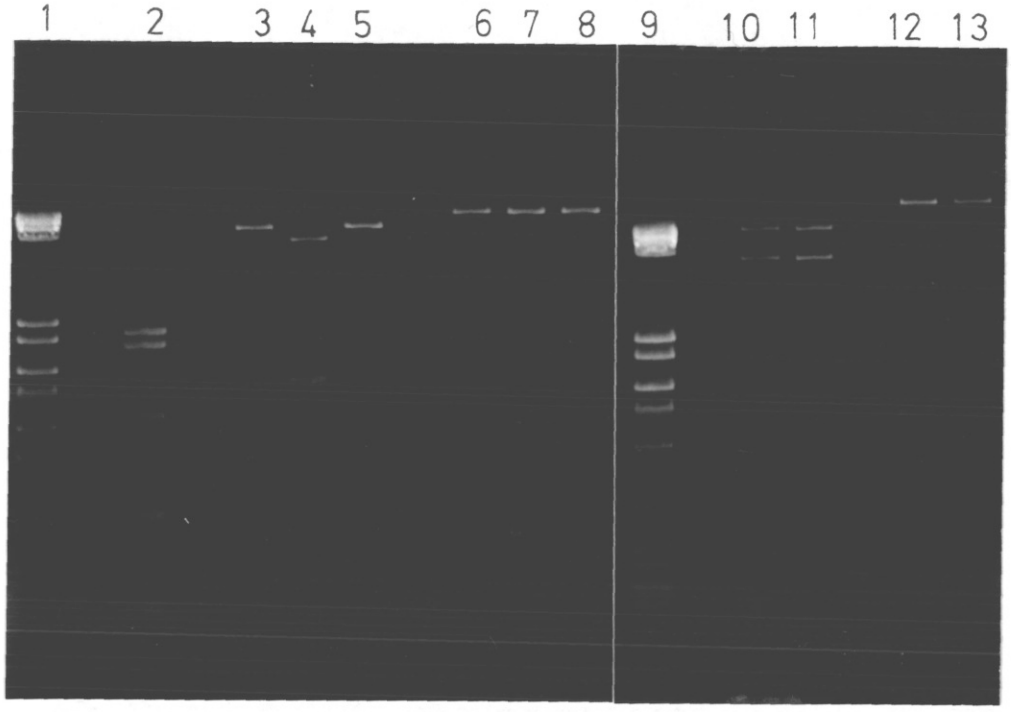

1
2
3
4
5
6
7
8
9
10
11
12
13

Figure 3.8: Southern blot of clones 323 and 336.

Samples of clones 323 and 336 were digested with the indicated enzymes, subjected to electrophoresis, transferred to nitrocellulose and hybridized with nick translated SV40 DNA.

Lane M, marker; Lanes 1 and 2, Sau3A digests of clones 323 and 336; Lanes 3 and 4, BamHI digests of clones 323 and 336; Lanes 5 and 6, EcoRI digests of clones 323 and 336.

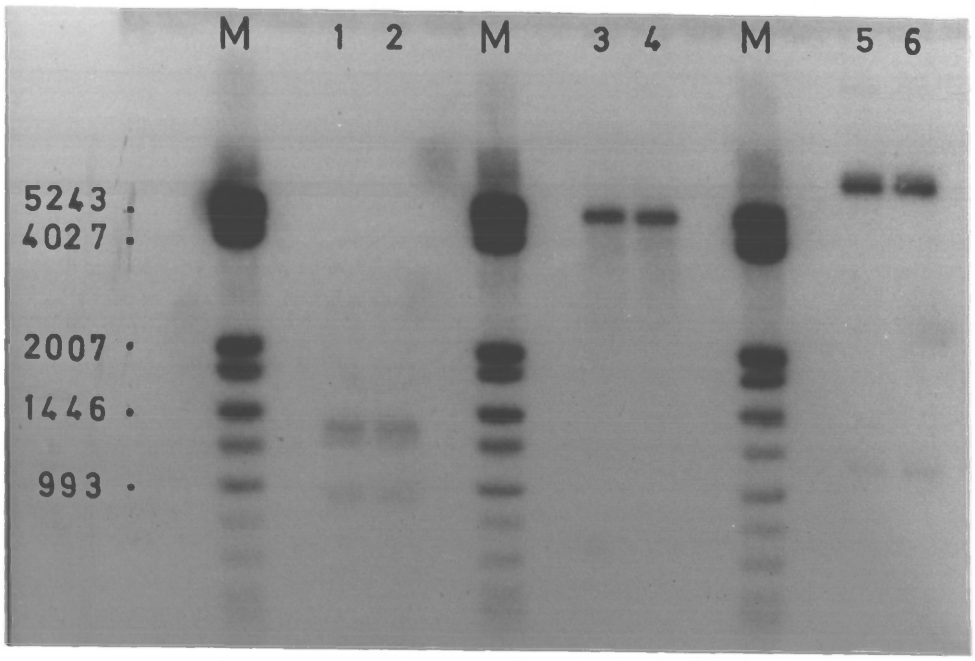
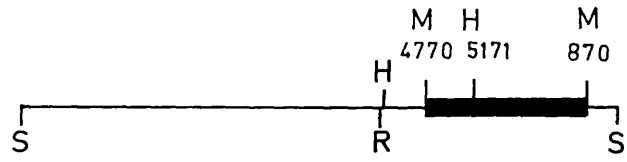
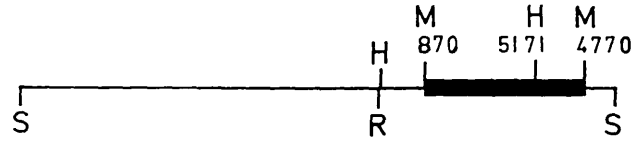

M
1 2
M
3 4
M
5 6
5243
4027
2007
1446
993

Figure 3.9: Proposed structure of SV40 homologous clones.

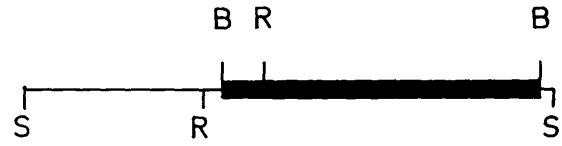
The clones are shown in linear form, opened up at the SalI site of pAT153. Abbreviations for restriction enzymes are: B, BamHI; H, HindIII; M, Sau3A; R, EcoRI; S, SalI. The Sau3A sites of pAT153 are not shown. Thin lines represent vector sequence, thick lines inserted DNA. Coordinates on SV40 are from Tooze, (1981).



Clone 106,207



Clone 200



Clone 323,336

EcoRI digestion (Figure 3.7, lanes 12 and 13; Figure 3.8, lanes 5 and 6) shows that both clones have the same orientation, as shown in Figure 3.9. The weak 8.8 kb band observed upon digestion with EcoRI, lanes 1 and 2, Figure 3.8, is a partial digestion product, and was not observed in other experiments, see for instance Figure 3.7, lanes 12 and 13. Sau3A digestion (Figure 3.5, lanes 1 and 2; Figure 3.8, lanes 1 and 2) at least confirms the presence of the expected large Sau3A fragments of SV40.

A search for origin - like sequences was also undertaken, hybridizing an ordered array of colonies at lower stringency [6 x SSC, 5 x Denhardt's, 1% (^{w}/v) SDS at 65°C, followed by washes in 2 x SSC, 0.1% (^{w}/v) SDS at 55°C]. The probe used was a nick translated fragment of SV40, coordinates 5171 - 270, spanning the origin of replication. A number of apparently positive colonies were identified, but upon analyzing the DNA from them, no significant SV40 homology was found. The reason for this discrepancy remains unclear.

Section 3.4: Discussion.

The cloning of genomic sequences from transformed cells other than the SV40 origin of replication with which T antigen interacts requires a strategy for the selective isolation of such fragments of DNA. A variety of approaches were possible, as discussed in Section 3.1. The method ultimately adopted utilizes restriction enzyme

digestion to fragment chromatin from transformed cells, followed by immunoselection of nucleoprotein complexes to which T antigen is bound using a monoclonal antibody directed against large T. Restriction enzyme digestion was chosen to facilitate the subsequent cloning of the immunoselected DNA. A variant on this method was also tested, removing the majority of chromatin associated proteins by treatment with dextran sulfate and heparin prior to digestion with restriction enzymes. The work of Reiser et al., (1980) had previously demonstrated that T antigen can form a complex at the viral origin of replication resistant to such a washing procedure. The reasoning behind testing this variation of the standard method was that the partial deproteinization of the chromatin would facilitate subsequent enzyme digestion.

Based on studies of autoregulation in tsA transformed cells (Alwine et al., 1977) and rescue of viral DNA following fusion with permissive cells (Botchan et al., 1978) it is clear that the SV40 origin is available to T antigen in the transformed cell. The expectation that the SV40 origin of replication should be cloned therefore constitutes a powerful internal control for the validity of the immunoselection cloning protocol. Hybridization of the clones obtained by this protocol has shown that in all three experiments, SV40 origin sequences were cloned. The distribution of insert sizes has not been determined for Experiment Three, but for Experiments One and Two, the

average size of insert is approximately 1.3 kb. It may be calculated that with an insert of this size, approximately 10^7 colonies would be required for a 99% complete library of mouse DNA (assuming a genome size of 3×10^9 base pairs). There are four or five copies of SV40 DNA in the SV3T3 Cl 38 genome (Clayton and Rigby, 1981). The finding that screening approximately 50 colonies (Experiments One and Two) gives one or two SV40 hybridizing colonies, representing clones of the viral origin of replication suggests an enrichment of $>10^4$ fold for origin sequences in the immunoselected DNA, compared to the chance of finding such a clone by selection of random clones.

Several points concerning these data must be considered however. Two of the SV40 templates in the Cl 38 genome are "marked", that is to say, they are uniquely identifiable as originating from the Cl 38 cells (Figure 3.2). All of the clones thus far obtained which are homologous to SV40 constitute either full length viral DNA or the Sau3A fragment encompassing the origin of replication. The 38-D template should produce a Sau3A fragment containing the origin of replication that is different in size from that obtained from viral DNA. The 38-A and the 38-D templates will both produce "marked" SV40 hybridizing fragments upon BamHl - BglII digestion. Neither the 38-A nor 38-D templates have been recovered in the immunoselected clones. It is conceivable that this is purely by chance, but other explanations must be considered. It is possible that T antigen will not bind

to the origins of replication of these templates *in vivo*, though since both of these origins are competent to support T antigen dependent replication (P. Rigby, personal communication), this seems unlikely.

Plasmids corresponding to clones 106, 200 and 207, namely the *Sau*3A A fragment inserted at the *Bam*HI site of pAT153, have to my knowledge never been constructed before in our laboratory, thus eliminating the possibility of contamination by pre-constructed plasmids. Additionally, the finding that clones 200 and 207 have the insert in opposite orientations would have required simultaneous contamination by two different plasmids if this were the explanation.

In view of the routine use of SV40 DNA in our laboratory, the possibility of contamination with SV40 DNA must be considered. Clones 106, 200 and 207 comprise a subfragment of the SV40 genome. Any contaminating DNA would have had to be present during the restriction enzyme digestion step, and co-precipitated as a result of T antigen binding to it during the digestion step. The same possibility could be used to explain the genesis of clones 323 and 336, and unfortunately cannot be formally eliminated based on the available data.

An experiment which would answer this latter objection would be to repeat the immunoselection cloning experiment using a starting cell line with a uniquely marked origin of replication. Such a cell line is cos-1

(Gluzman, 1981). These cells contain only a single copy of the viral early region and replication origin, with the latter containing a defined six bp deletion at the BglI site. Thus, only mutant origins could be immunoseledted, and any origin - homologous clones could be sequenced to confirm that they did indeed originate from the cos-1 cells. Studies have shown that T antigen will bind to clones of the cos-1 cell SV40 origin of replication, at least *in vitro* (Lewton *et al.*, 1984).

If one assumes that these clones have not arisen as a result of adventitious contamination of whatever kind, then it is apparent that immunoselection cloning provides a very powerful tool for analyzing cellular sequences with which DNA binding proteins interact. Clearly, this approach is not only applicable to large T antigen, but could be used for any DNA binding protein for which an antibody was available. Obvious candidate proteins are the steroid hormone receptors, and the *myc* gene product.

The very low background of non-specifically co-precipitated DNA when control antibodies are used gives one some confidence that the clones obtained might represent a significant subfraction of the cellular DNA with which T antigen interacts *in vivo*, either directly, or through another protein. The experiments described in the subsequent chapters are directed towards analysis of the interaction of T antigen with a few of the clones obtained by this procedure, both in crude extracts, and as a purified protein.

CHAPTER FOUR

PURIFICATION OF LARGE T ANTIGEN

CHAPTER FOUR

APPROACHES TO THE PURIFICATION OF T ANTIGEN

Section 4.1: Introduction.

A rigorous analysis of the biochemical properties of T antigen requires a supply of highly purified, biologically active protein. While procedures have been developed for analyzing interactions of T antigen with DNA (McKay, 1981; Gaudray et al., 1981), and its enzymic activity (Clark et al., 1981), using crude extracts of cells, it is more satisfactory to undertake such experiments using pure preparations of protein.

Much effort has therefore been expended to develop procedures for the purification of biologically active large T antigen from various sources (Tenen et al., 1977; Tjian, 1978; Spillman et al., 1979; Tegtmeyer and Anderson, 1981; Tenen et al., 1983). Problems encountered during purification of T antigen have included proteolysis, dependent upon the pH of the extraction (Schwyzer et al., 1980), and heterogeneous fractionation of protein during column chromatography steps (Griffin et al., 1979a; Tjian et al., 1979; Dorn et al., 1982), resulting in a low final yield of protein.

Because of the small amounts of T antigen produced in lytic infection [4×10^9 cells yield only 22 µg of T antigen after purification (Tjian et al., 1979)], other potential sources of T antigen have been investigated. The SV40 transformed human cell line, SV80, produces large

amounts of T antigen and was therefore exploited to provide starting material for T antigen purification in early studies (Tenen et al., 1977; Spillman et al., 1979). Similarly, Tjian (1978) used HeLa cells infected with $Ad2^{+}D2$ as a source from which to purify D2T. However, as discussed earlier, it is now clear that both of these proteins are in some way defective, in at least their interaction with DNA, when compared with "authentic" T antigen from lytically infected cells (see Chapter One). Therefore Adenovirus - SV40 hybrids have been constructed which express full - length T antigen (Solnick, 1981; Thummel et al., 1981b; Gluzman et al., 1982; Thummel et al., 1983) enabling large amounts of the protein to be purified from infected HeLa cells. There are also caveats concerning the use of T antigen from these cells. It is known that T antigen exists in a variety of different forms in the lytically infected cell, differing for instance, in their oligomeric state, degree of phosphorylation and ability to interact with DNA. Since these T antigens are made in the environment of a human cell, late in adenovirus infection, it is conceivable that the protein will be incorrectly modified, perhaps in such a way as to alter its function. That this could be so in at least one case is suggested by the interaction with DNA of the T antigen produced by Ad-SVR6 (Thummel et al., 1981b), which in contrast to that of T antigen extracted from lytically infected monkey cells is cooperative.

However, other T antigens such as that encoded by Ad-SVR284 (Thummel et al., 1983) do not bind in a cooperative manner (Jones and Tjian, 1984). The reason for this discrepancy is unknown, but serves to illustrate that until all of the biological properties of T antigen and the parameters which govern them have been defined, all work using T antigens from sources other than lytically infected monkey cells is potentially compromised, unless it can be demonstrated that the T antigens from the two different sources are identical with respect to the parameter under study.

Despite this, the use of these hybrid viruses remains an attractive proposition, since the procedures developed for purification of T antigen from lytically infected cells require large quantities of starting material, and produce protein only 20 - 60% pure by the methods of Tjian et al., (1979), Tenen et al., (1983) or Tegtmeyer and Anderson (1981). More seriously, as shown in the study of Tegtmeyer et al., (1981), while these contaminating proteins may not interact with the SV40 origin of replication, they can bind to specific cellular DNA sequences, thus potentially compromising any experiments using such preparations to study the interaction of T antigen with cellular non-origin DNA sequences.

During the course of this study, a number of different approaches aimed at obtaining high level expression and purification of large T antigen were investigated, and these form the subject matter of this

chapter. Initial experiments explored the possibility of using a heterologous system, the yeast Saccharomyces cerevisiae, as a host for expression. Following limited success, an alternative approach using an adenovirus - SV40 hybrid virus designed to overproduce large T antigen was used. An immunoaffinity purification procedure was subsequently developed to obtain >95% pure, full length biologically active T antigen from cells infected with this virus.

Section 4.2: Materials and methods.

For general methods, see Chapter Two.

Yeast strains: MC16 α, *leu* 2-3, *his* 4-712, su F2, *lys* 2-1, *ade* 2-1.

Yeast media: The following media were routinely used for growth of yeast:

YPDA:
2% (^{w}/v) Difco Bacto Peptone
1% (^{w}/v) Difco Bacto Yeast Extract
2% (^{w}/v) D-glucose
20 mg/liter adenine sulfate

YPDA agar: YPDA + 1.5% (^{w}/v) Difco Bacto Agar

Yeast Minimal Medium:
0.67% (^{w}/v) Difco Bacto Yeast Nitrogen Base without amino acids.
2% (^{w}/v) D-glucose
Amino acids and other supplements were added to 20 mg/liter.

Yeast Minimal Agar:
Yeast Minimal Medium + 1.5% (^{w}/v) Difco Bacto Agar

Yeast transformation

This was performed essentially according to the method of Beggs (1978). A 100 ml culture of MC16 grown at 30°C with shaking to mid-log phase was collected by centrifugation, washed once in sterile, distilled water, resuspended in 25 ml of sterile 1 M Sorbitol, 25 mM EDTA (pH 8.0) and solid DTT added to a concentration of 25 mM. After 30 minutes at 30°C, the cells were washed once with 1.2 M Sorbitol, and resuspended in 25 ml of sterile, freshly made 1.2 M Sorbitol, 100 mM citric acid, 10 mM EDTA (pH 5.8, adjusted with NaOH) containing 1% ($^{v}/v$) ß-glucuronidase (Sigma). The cells were incubated at 30°C for 60 - 90 minutes to effect spheroplasting, and then washed three times in sterile 1.2 M Sorbitol, 10 mM $CaCl_2$, and finally resuspended in 1 ml of 1.2 M Sorbitol, 10 mM $CaCl_2$. 1 - 5 µg of DNA were then added to 50 µl of this spheroplast suspension, and incubated at room temperature for 30 minutes. The volume of the DNA solution added did not exceed 15 µl. 500 µl of freshly made, sterile 20% ($^{w}/v$) PEG 4000, 10 mM Tris-HCl pH 7.4, 10 mM $CaCl_2$ was then added to facilitate DNA uptake. After 3 minutes, the spheroplasts were collected by brief centrifugation, resuspended in 200 µl of YPDA containing 1.2 M Sorbitol, and incubated at 30°C for 1 hour. The spheroplasts were then pelleted, resuspended in 1.2 M Sorbitol and plated onto selective media together with 8 ml of 3% ($^{w}/v$) top agar. For transformation, all plates and agar contained

1.2 M Sorbitol as osmotic stabilizer. Plates were incubated inverted at 30°C for 4 - 5 days. Surplus spheroplasts were resuspended in 10 mM MOPS pH 7, 75 mM $CaCl_2$, 10 mM RbCl, 1.2 M Sorbitol containing 15% (v/v) glycerol, incubated on ice for 30 minutes and then snap - frozen (in 50 µl aliquots) in dry ice - ethanol prior to storage at -70°C. For transformation, an aliquot was thawed, and DNA added directly. Typical transformation efficiencies were 1 - 5 x 10^4 colonies per µg DNA for fresh cells, and 10 - 25% of this value for frozen cells.

Extraction of DNA from yeast

This is essentially the method of Beggs (1978) with minor modifications. An aliquot of the culture to be analyzed (typically 0.5%) was removed, and the cells collected by centrifugation in an Eppendorf microcentrifuge. The cells were resuspended in 0.5 ml of 1 M Sorbitol, 25 mM EDTA (pH 8.0), 5.0% (v/v) ß-mercaptoethanol. After 15 minutes at room temperature, the cells were pelleted and then resuspended in 0.5 ml of 100 mM citric acid, 10 mM EDTA, 1.2 M Sorbitol (pH 5.8, adjusted with NaOH) containing 1% (v/v) ß-glucuronidase. After 30 minutes at 37° C the cells were washed once in 1.2 M Sorbitol, resuspended in 400 µl of lysis solution [50 mM Tris-HCl pH 8.0, 20 mM EDTA, 1% (v/v) SDS] and incubated at 60°C for 30 minutes. The sample was then extracted three times with phenol and chloroform, twice with water saturated ether and precipitated with ethanol.

Extraction of RNA from yeast

This was performed according to the procedure of Beggs et al., (1980). 200 ml cultures were grown to mid-log phase at 30°C with shaking, in minimal medium lacking leucine (to maintain selection for the plasmid) and collected by centrifugation. The cells were washed once in cold TNEH (50 mM Tris-HCl pH 7.4, 100 mM NaCl, 5 mM EDTA containing 200 µg/ml Heparin) and resuspended in 1 ml of cold TSEH [50 mM Tris-HCl pH 7.4, 5 mM EDTA, 1 mg/ml Heparin, 10% (^{w}/v) sucrose]. 0.75 volume of baked, acid washed glass beads (BDH; 100 mesh, for gas chromatography) was added, and the tube vortexed for 90 seconds at maximum speed. The supernatant was removed, after allowing the beads to settle, and the beads were washed with 4 ml, then 3 ml ice - cold TNEH. SDS was then added to 0.5% (^{w}/v) followed by extraction with 10 ml of water saturated phenol. The aqueous phase was then twice extracted with phenol and chloroform, and precipitated with ethanol. Typically, 10 - 20 µg of RNA were obtained per ml of culture. All solutions were sterilized by autoclaving, and all glassware was baked. Gloves were worn throughout to avoid contamination with ribonuclease. Heparin was added where required as a solid following autoclaving. Subsequently, RNA was fractionated by affinity chromatography on oligo-dT cellulose as described in Chapter Two.

Labelling and extraction of proteins from yeast

Cultures of transformed yeast cells were grown in

selective minimal medium with shaking at 30°C to mid-log phase. Labelling was initiated by addition of ^{35}S-L-methionine to 100 µCi/ml and was at 30°C with shaking. The cells were then chilled on ice for 10 minutes, and diluted 10 fold in zymolyase buffer (50 mM sodium phosphate pH 7.0, 14 mM ß-mercaptoethanol, 1 M Sorbitol). Zymolyase 60,000 (Kirin brewery) was added to 100 µg/ml and the cells incubated at 37°C for 15 minutes. The cells were then washed once in zymolyase buffer by centrifugation, and lysed by resuspension in 1 - 2% of original volume of 10 mM sodium phosphate pH 7, 50 mM ß-mercaptoethanol, 7 M guanidine HCl containing 1 mM PMSF added just prior to use. The samples were then centrifuged to remove cell debris, and an aliquot of each counted by TCA precipitation. The remainder was used for immunoprecipitation or stored at -70°C until required.

Immunoprecipitation

2 - 10 x 10^6 acid insoluble counts were incubated with 10 µl of Normal Rabbit Serum in NET buffer [50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM EDTA, 0.5% (^{v}/v) NP-40] on ice for 1 hour. Sufficient NET buffer was added to dilute the guanidine HCl concentration to 0.5 M or less, to avoid denaturing the antibody. 25 µl of Protein A Sepharose beads (slurry in NET buffer) were added, and after 1 hour, removed by centrifugation. The supernatant was then divided into two equal portions and incubated with 100 µl of either PAb419 or PAb423 (Harlow et al.,

1981) overnight at 4°C. Antibodies were used as tissue culture supernatant from hybridoma cells. 100 µl of supernatant contain 1 - 2 µg of antibody. 10 µl of protein A Sepharose beads were then added, and after 1 hour at 4°C, collected by centrifugation and washed twice with NET buffer containing 0.1% (w/v) SDS, and three times with NET buffer, prior to resuspension and boiling in sample buffer. Samples were then subjected to electrophoresis on 15% acrylamide gels, enhanced, dried and autoradiographed as described in Chapter Two.

Labelling and extraction of protein from E. coli

Cultures of E. coli strain 294 bearing plasmid HP1 (Thummel et al., 1981) were grown at 37°C with shaking to mid-log phase in M9 medium supplemented with all amino acids except methionine. IPTG was added to 1 mM to induce the lac promoter and 30 minutes later ^{35}S-L-methionine was added to 100 µCi/ml. The cells were labelled for 10 minutes, collected by centrifugation, washed once and extracted by resuspension in 5% of culture volume of 50 mM Tris. HCl pH 8.0, 10 mM EDTA, 10% (w/v) sucrose containing 100 µg/ml lysozyme. After 15 minutes on ice the cells were pelleted and lysed by resuspension in 10 mM sodium phosphate pH7.0, 50 mM ß-mercaptoethanol, 7 M guanidine HCl to which PMSF had been added to 1 mM just prior to use. After centrifugation for 10 minutes at 4°C in a microfuge, to remove debris, the supernatant was used for precipitation or stored at -70°C.

Radioimmunoassay for small t

^{35}S-Methionine labelled small t was purified from HPl bearing E. coli strain 294, following growth and labelling as described above except that following induction with IPTG, the culture was concentrated from 100 ml to 5 ml for labelling. Following labelling, the cells were washed, resuspended in 0.5 ml of cold NET buffer and sonicated on ice (6 x 30 second bursts at maximum amplitude, at 1 minute intervals). PMSF was added to the NET buffer just prior to use to a concentration of 1 mM. The cell debris were removed by centrifugation, and 1 ml of PAb419 tissue culture supernatant was added. After an overnight incubation at 4°C, 75 µl of formaldehyde fixed Staphylococcus aureus (Staphs) bearing A protein [10% (^{v}/v) suspension in NET buffer] were added. Two hours later the bacteria were pelleted, washed five times with NET buffer, resuspended in sample buffer and boiled for three minutes. After centrifugation the supernatant was subjected to electrophoresis on a 15% acrylamide gel. The small t band was located by autoradiography, excised from the wet gel, and the protein eluted from the gel slice by electrophoresis. Approximately 10^6 cpm of small t protein were recovered, 50% of which could be reprecipitated by PAb419. The protein was stored in running buffer at -20°C.

For the assay, protein extracts were prepared as described above, except that proteins were not labelled prior to extraction. 5,000 cpm of ^{35}S-methionine labelled

t-antigen was mixed with varying amounts of yeast or *E. coli* extract in a total volume of 300 µl of NET buffer. 50 µl of PAb419 was then added. Following incubation overnight at 4°C, 20 µl of Staphs were added, and two hours later, the bacteria were pelleted, washed once in NET buffer and counted in 1 ml of Aquasol 2 Scintillation fluid (NEN).

Protein determination

The modification (Peterson, 1977) of the method of Lowry *et al.*, (1951) was used. BSA was used as the protein standard. Alternatively, the optical density of the solution at 280 nm was read.

Preparation of immunoaffinity column (Schneider *et al.*, 1982)

One gram of protein A Sepharose beads were swollen in 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM EDTA (Buffer A) containing 0.1% (^{v}/v) Nonidet P-40 (NP-40) and packed into a 1 cm diameter Bio-Rad Econo column. 300 ml of tissue culture supernatant from PAb419 hybridoma cells (Harlow *et al.*, 1981) was circulated through the column overnight at a flow rate of 100 ml/hr. The column was then extensively washed with Buffer A containing 0.5% (^{v}/v) NP-40, followed by 0.1 M borate buffer pH 9.0. The beads were then resuspended in 10 ml of 0.1 M borate buffer pH 9.0, and dimethylpimelimidate dihydrochloride (Sigma) added to a final concentration of 40 mM. The mixture was agitated at room temperature for 1 hour. The beads were then washed

with 10 ml of 40 mM ethanolamine-HCl pH 8.0 followed by 0.1 M borate buffer pH 8.0 in which they were stored until required. Approximately 2 mg of PAb419 were coupled to 1.5 ml (packed volume) of protein A Sepharose beads.

Gel electrophoresis and Western blotting

See Chapter Two.

Radioimmunoassay for T antigen

Purified T antigen was iodinated by J. Gannon (Imperial College), using the iodogen method. A titration of the polyclonal rabbit serum raised against D2 protein (Lane and Crawford, 1980) to determine the optimal dilution for maximal sensitivity in the assay was performed as follows. 10^4 cpm (50 µl) of ^{125}I-labelled T antigen was mixed with increasing dilutions of antiserum (100 µl) in a total volume of 250 µl of Buffer A containing 1% ($^{w}/v$) BSA, 5% ($^{v}/v$) FCS and 0.5% ($^{v}/v$) NP-40. After incubation overnight at room temperature, 20 µl of 10% ($^{v}/v$) suspension of protein A bearing Staphylococcus aureus was added. After a further three hours the bacteria were collected by centrifugation and washed twice in Buffer A containing 0.5% ($^{v}/v$) NP-40 before counting. The quantitative assay was basically performed in a similar manner. 50 µl (10^4 cpm) of iodinated T antigen were mixed with various amounts of unlabelled T antigen, or T antigen containing extracts in a total volume of 150 µl as described above. 100 µl of a 1/4000 dilution of the rabbit anti-D2 antiserum was then added. Incubation and further processing of the samples

was as described earlier.

ATPase assay

The standard ATPase assay was performed as described by Clark et al., (1981), in 20 - 100 µl of buffer containing 25 mM PIPES (pH 7.0), 100 mM NaCl, 5 mM $MgCl_2$, 0.01% (^{v}/v) NP-40, 5 - 500 µM unlabelled ATP and 0.5 - 2 µCi of [γ -^{32}P] ATP (3000 Ci/mmol, Amersham). After incubation with 50 - 200 ng of T antigen for 10 minutes at 20 - 24°C unreacted ATP was precipitated by the addition of 100 µl of acid washed Norit charcoal [7.5% (^{w}/v) in 50 mM HCl, 5 mM H_3PO_4). After pelleting the charcoal, half the supernatant was analyzed for free ^{32}Pi by Cerenkov counting. For inhibition studies 100 ng of T antigen was incubated with 600 ng of purified PAb419 (Harlow et al., 1981) or PAb204 (Lane and Hoeffler, 1980) for 15 minutes at 0°C before commencing the assay. Monoclonal antibodies were purified by affinity chromatography on protein A Sepharose.

McKay assay and DNAse I footprinting

See Chapter Five, Section 5.2.

Immunoaffinity purification of large T antigen

Twenty 9 cm dishes of subconfluent 293 cells (ca. 10^8 cells) were infected with Ad5 SVR111 at a multiplicity of 5 pfu per cell. 24 hours later the cells were harvested and washed 4 times with ice cold TD. All subsequent steps were performed at 4°C. The cells were then lysed in 10 ml of cold Buffer B [50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1

mM EDTA, 1 mM DTT, 10% (^{v}/v) glycerol, 1 mM PMSF] containing 0.5% (^{v}/v) NP-40. After 30 minutes on ice, cell debris and nuclei were removed by centrifugation at 10,000 rpm at 4°C in the Sorvall HB-4 swing out rotor.

The lysate was then passed 4 times through a 1 ml (packed volume) protein A Sepharose column equilibrated in Buffer B to deplete proteins which bind to protein A Sepharose non-specifically. The lysate was then circulated through the 1 ml volume PAb419 immunoaffinity column, which had previously been equilibrated in Buffer B, at a flow rate of 100 ml/hr. The column was then washed sequentially with 150 volumes of Buffer C [50 mM Tris-HCl pH 8.0, 500 mM LiCl, 1 mM DTT, 1 mM EDTA, 10% (^{v}/v) glycerol, 1 mM PMSF], and 100 volumes of Buffer D [10 mM PIPES pH 7.4, 5 mM NaCl, 1 mM DTT, 1 mM EDTA, 10% (^{v}/v) glycerol, 1 mM PMSF]. Bound proteins were eluted with Buffer E [20 mM triethylamine pH 10.8, 10% (^{v}/v) glycerol]. 200 µl fractions were collected and immediately neutralized by addition of 0.1 volume of 0.5 M PIPES pH 7.0. Protein containing fractions were detected by spotting 1 µl of each fraction onto Whatman 541 paper, allowing it to dry, and then staining with Coomassie Brilliant Blue [0.1% (^{w}/v) in 25% (^{v}/v) propan-2-ol, 10% (^{v}/v) acetic acid] for 10 minutes, followed by destaining in water. Peak fractions were pooled, dialyzed against 500 volumes of Buffer F [10 mM PIPES pH 7.0, 5 mM NaCl, 1 mM DTT, 0.1 mM EDTA, 10% (^{v}/v) glycerol, 1 mM PMSF], rapidly frozen in small aliquots and stored at -70°C. All

quoted pH values were adjusted at 25°C; PMSF was added to the concentrations indicated just prior to use.

Section 4.3: Expression of the early region of SV40 in Saccharomyces cerevisiae.

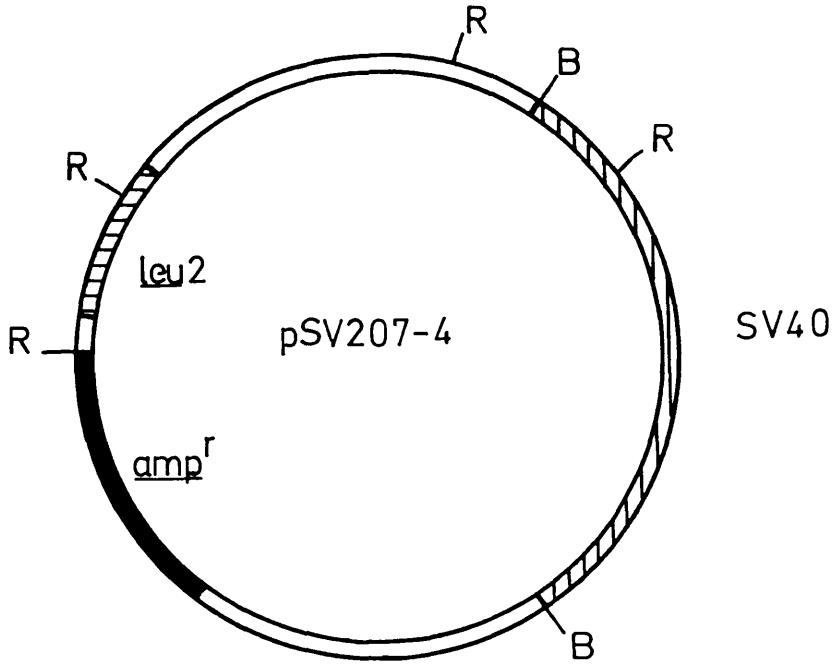
SV40 was digested to completion with BamHI, and ligated to BamHI cut pJDB207 (Beggs, 1981). E. coli strain HB101 was transformed to ampicillin resistance, and the colonies thus obtained were toothpicked onto plates containing ampicillin or tetracycline. Ampicillin resistant, tetracycline sensitive colonies were grown up and mini-scale plasmid preparations were analyzed for the presence of insert DNA by digestion with BamHI. The orientation of the insert was determined by digestion with EcoRI. Two recombinant plasmids, each containing a single insert of SV40 DNA, in opposite orientations were selected for further study. The structure of these plasmids is shown in Figure 4.1. Following a large scale plasmid preparation the DNA of pSV207-4 and pSV207-12 (hereafter abbreviated to p4 and p12) was used to transform S. cerevisiae strain MC16, selecting for leucine independent growth, the marker carried by pJDB207. Individual transformants were streaked out on minimal agar plates lacking leucine and colony purified.

Total DNA from the transformed yeast clones was isolated, digested with BamHI, subjected to

Figure 4.1: Structure of the SV40 - pJDB207 recombinants used in this work.

The size of the plasmids is 12.1 kb.

Abbreviations: B, BamHI; R, EcoRI.


R
B
R
R
leu2
R
pSV207-4
SV40
amp^r
B

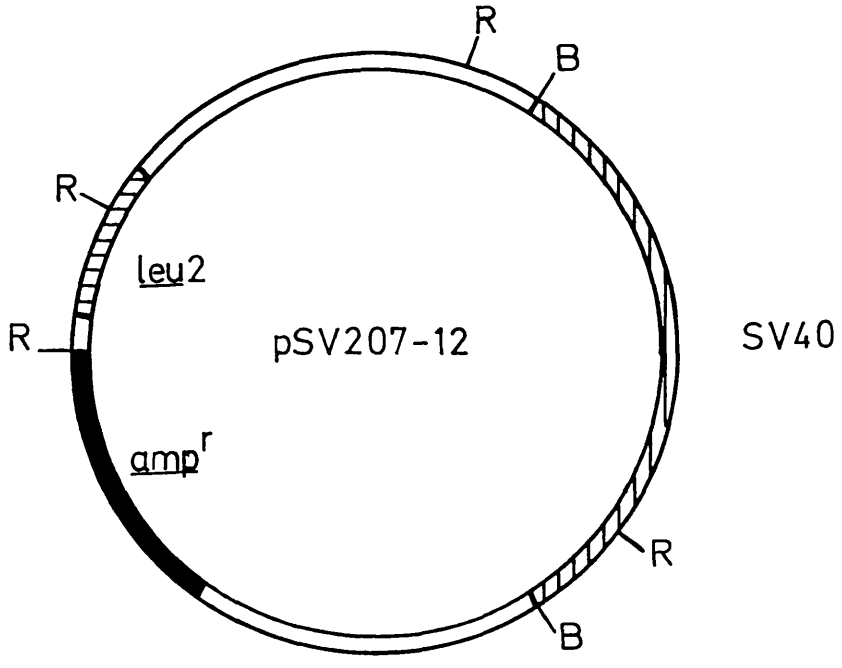

R
B
R
leu2
R
pSV207-12
SV40
amp^r
R
B

Figure 4.2: Analysis of DNA extracted from yeast transformants.

DNA was extracted as described in Section 4.2, subjected to electrophoresis through a 0.7% (w/v) agarose gel, transferred to nitrocellulose and probed with nick translated SV40 DNA.

Lane 1: Undigested p4 DNA.

Lane 2: p4 transformant, undigested, $2\mu^{+}$.

Lane 3: p4 transformant, undigested, $2\mu^{o}$.

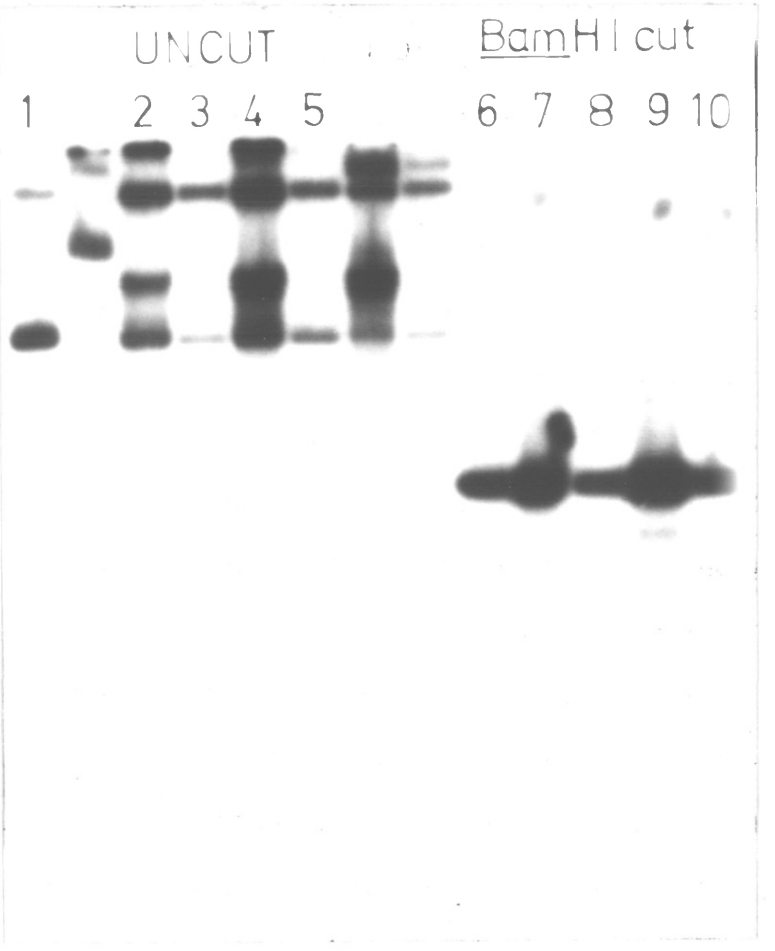
Lane 4: p12 transformant, undigested, $2\mu^{+}$.

Lane 5: p12 transformant, undigested, $2\mu^{o}$.

Lanes 6 - 10, samples as lanes 1 - 5, except that the samples were digested with BamHI.

Unlabelled tracks pertain to other plasmids, not discussed in this thesis.

$2\mu^{+/o}$ indicates whether the strain harbours the endogenous yeast plasmid, two micron.


UNCUT
BamHI cut
1 2 3 4 5 6 7 8 9 10

electrophoresis and transferred to nitrocellulose. The filter was then probed with nick translated SV40 DNA. As shown in Figure 4.2, uncleaved DNA shows several bands in addition to those of supercoiled and open circular DNA. These are most probably recombinants of the plasmid either with itself, or more likely, endogenous 2 μ DNA (since they are not seen in 2 μ^{o} transformants). BamHI cleavage however demonstrates that the majority of the SV40 inserts in the plasmids have not been grossly rearranged, since a 5.2 kb band may be excised, representing full length linear SV40.

Since the constructions described have no known strong yeast promoter positioned such that it would direct transcription of SV40 sequences it was necessary to determine whether any transcription of the inserted DNA was occurring. Therefore, RNA was extracted, denatured with formaldehyde, and transferred to nitrocellulose after electrophoresis. The filters were then hybridized with nick translated pJDB207, SV40, or the HindIII B fragment of SV40, which was used as a probe for early region specific transcripts. The result is shown in Figures 4.3 and 4.4. As shown in Figure 4.3, the RNA from yeast transformed with the vector pJDB207 contains no SV40 homologous transcripts, as expected. However, it is clear that transcripts with SV40 homology are far more numerous and diverse in size than those observed in lytically infected cells. Figure 4.4 shows that some of the transcripts produced are clearly SV40 specific, in that

Figure 4.3: Analysis of RNA extracted from yeast transformants: I.

RNA was extracted from yeast transformants and SV40 infected CV1 cells. The blot was probed with nick translated SV40 DNA. Headings above each track indicate the source of the RNA. 2µg of poly A^+ RNA was loaded per track.

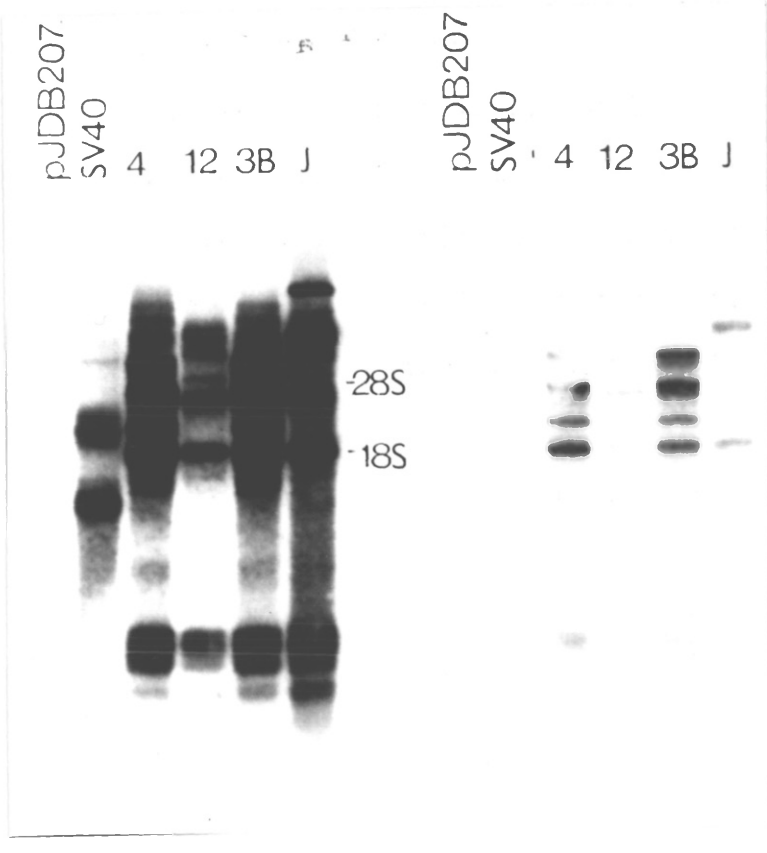
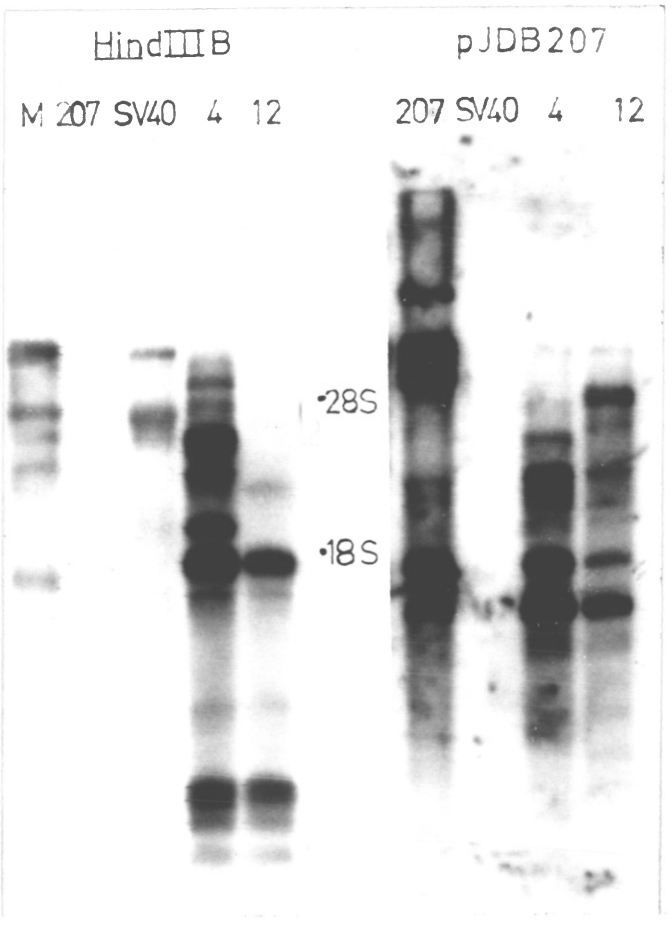

pJDB207
SV40
4
12
3B
J
28S
18S
pJDB207
SV40
4
12
3B
J

Figure 4.4: Analysis of RNA extracted from yeast transformants: II.

Legend is as for Figure 4.3. The filter was first probed with nick translated HindIII B fragment of SV40, then, after removal of probe, with nick translated pJDB207.

M, markers; 207, RNA extracted from pJDB207 transformed MC16; 4 and 12, RNA extracted from MC16 transformed with p4 and p12, respectively; SV40, 10 μg of unfractionated RNA extracted 48 hours after infection by SV40.

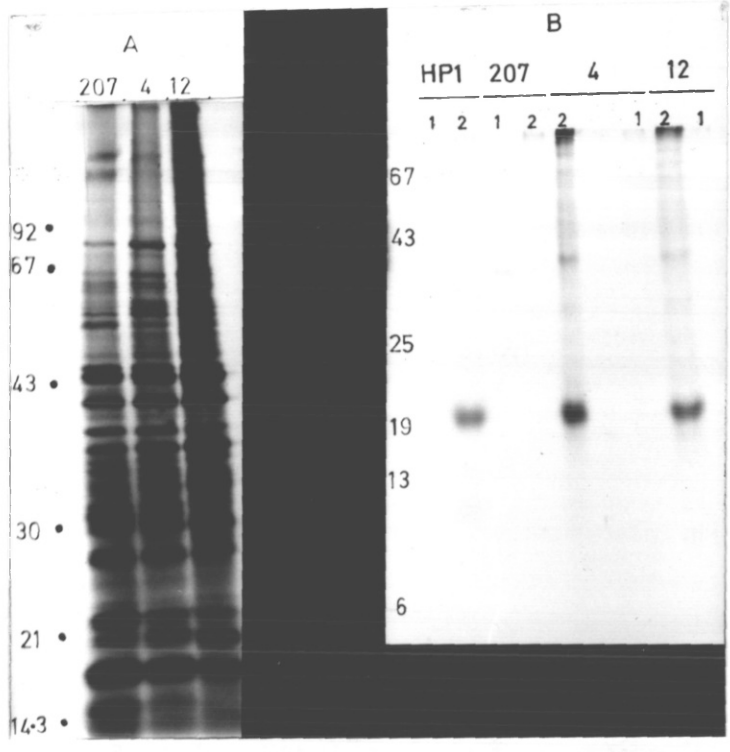

HindIII B
pJDB207
M 207 SV40 4 12
207 SV40 4 12
28S
18S

they have no detectable vector homology, and are present in both orientations, suggesting they are encoded within SV40 itself. It is also intriguing to note that the complexity of the SV40 specific transcripts differs in the two orientations. The reason for the difference is not known. The transcripts that have only early region homology and are produced in both orientations have sizes of 1.6 kb, 0.95 kb, 0.65 kb (doublet), 0.58 kb and 0.5 kb. The 1.8 kb transcript is also produced in both orientations, but co-migrates with a transcript synthesized from pJDB207; therefore, while it is likely that this transcript has no vector homology, this has not been formally proven. None of these transcripts is the same size as authentic early region mRNA (2.3 kb and 2.6 kb), observed during lytic infection. The proteins synthesized by yeast bearing p4 and p12 were then investigated for the presence of T antigen related proteins. Thus proteins from ^{35}S-methionine labelled cultures were extracted as described in Section 4.2 and immunoprecipitations were performed using the anti - large T monoclonal antibodies PAb419 and PAb423 (Harlow et al., 1981). Samples of total protein and the immunoprecipitated proteins were electrophoresed on 15% (^{w}/v) acrylamide gels, and after staining to reveal the positions of marker proteins, treated with En^{3}hance (NEN), dried and fluorographed. The result is shown in Figure 4.5. The immunoprecipitation data indicate that in yeast transformed by p4 and p12, a novel protein is produced,

Figure 4.5: Proteins synthesised in yeast transformants.

Panel A: Total protein extracted from yeast transformants. 10^5 cpm was loaded in each track. 207, 4 and 12 are protein extracts from MC16 transformed by pJDB207, p4 and p12, respectively. The positions of unlabelled marker proteins is indicated. Numbers denote their Mr in kilodaltons.

Panel B: Immunoprecipitation of protein from yeast transformants. 10^6 cpm were added to each precipitation. The nomenclature is as for Panel A. HP1 is protein extracted from *E. coli* strain 294 bearing plasmid HP1, after induction with IPTG.


A
207
4
12
92
67
43
30
21
14·3
B
HP1
207
4
12
1 2 1 2 2 1 2 1
67
43
25
19
13
6

which is recognized by PAb419, but not by PAb423. Since PAb419 reacts with small t antigen, and the protein precipitated co-migrates with authentic small t antigen overproduced in E. coli (Thummel et al., 1981a) it is tempting to speculate that it is indeed small t antigen. To investigate whether this was the case, a radioimmunoassay was devised for measuring levels of small t. The source of small t antigen used was a bacterial extract made from strain 294 bearing plasmid HP1 (Thummel et al., 1981a), which upon induction overproduces small t antigen. Thus the bacteria were labelled with ^{35}S-methionine after induction with IPTG, proteins were extracted and immunoprecipitated with PAb419 and the small t band excised from the gel and the protein electroeluted. This was used as a source of small t for a competition radioimmunoassay when a fixed amount of labelled small t is incubated with a limiting amount of antibody in the presence of increasing amounts of unlabelled protein extracts from the yeast. Small t antigen containing extracts will compete with the labelled small t for the limiting amount of antibody thus decreasing the number of counts precipitated. The results of such an assay are shown in Figure 4.6, which clearly shows that the extracts made from cells transformed with p4 and p12 cause inhibition in the assay while extracts from cells transformed with pJDB207 do not. Approximately 2 - 3 fold more protein is required for 50% inhibitiion of the assay with yeast extracts than with extracts from E. coli

Figure 4.6: Radioimmunoassay for small t antigen.

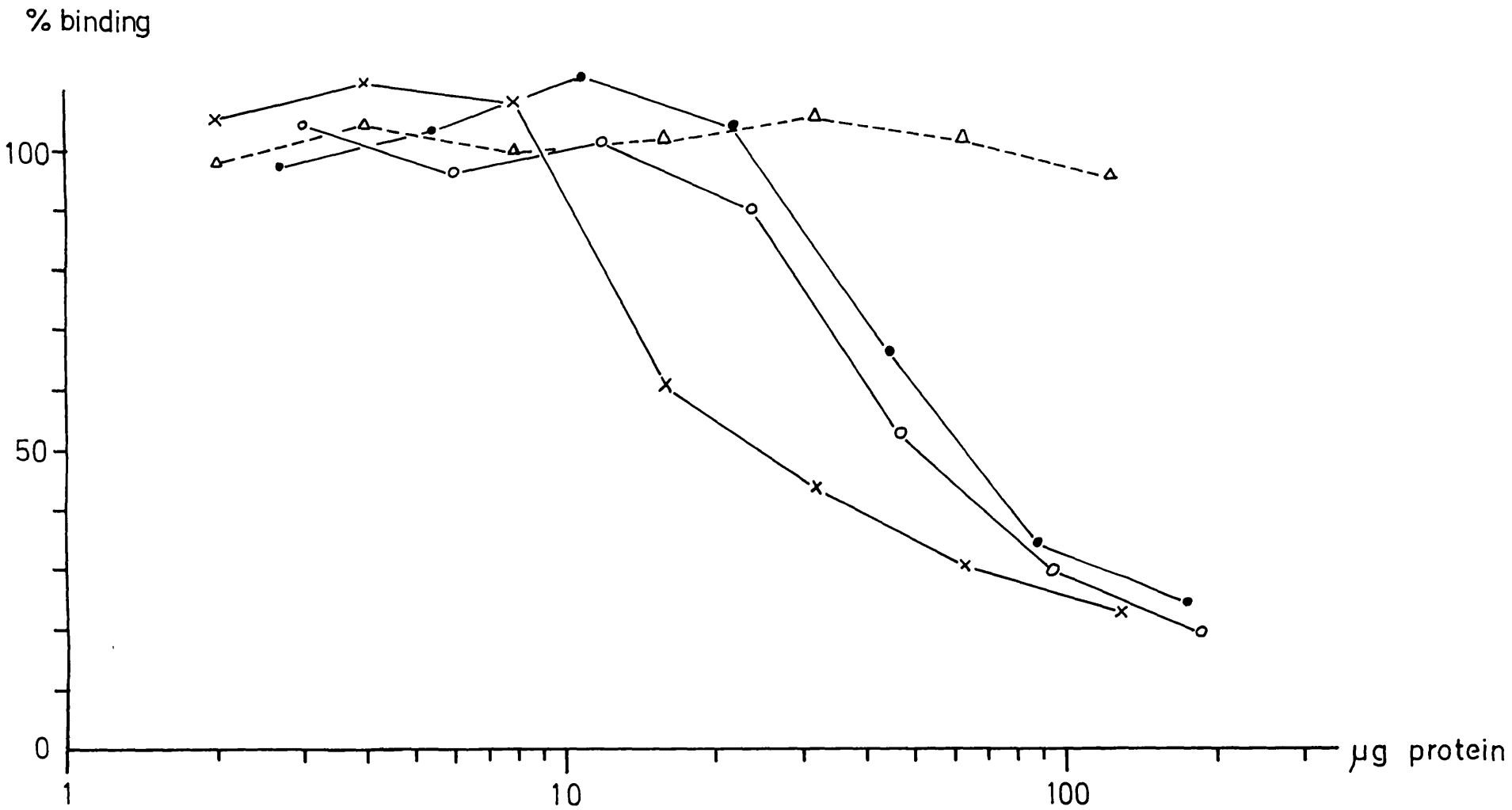
Samples were processed as described in the text. % binding is plotted against the amount of input protein. All points are the average of duplicate determinations. Symbols used:

--Δ--Δ-- : pJDB207 transformant protein extract.

--o--o-- : p4 transformant protein extract.

--●--●-- : pl2 transformant protein extract.

--x--x-- : HPl transformed E. coli extract.


% binding
100
50
0
1
10
100
µg protein

bearing HP1, suggesting that if the latter has a steady state level of 1% of cell protein (Thummel et al., 1981a) then the steady state level in yeast represents 0.2 - 0.3% of cell protein. A similar assay designed to detect large T antigen, described later in this chapter (Section 4.4) failed to detect any T antigen related peptides. Two additional points are worthy of note. Firstly, yeast does not seem to contain detectable levels of the 35 kd protein with which PAb419 reacts in higher eukaryotic cells (Crawford et al., 1982), and secondly no evidence has been found by immunoprecipitation for the presence of the truncated small t antigen produced by HP1 bearing cells. However, since this is thought to consist of the C terminal portion of t antigen, resulting from initiation of translation at an internal AUG codon, it is possible that it does not bear the determinant recognized by PAb419. The state of oligomerization and subcellular localization of small t antigen were not investigated. Thus, it seems that while small t antigen is synthesized in yeast, large T antigen is not. Clearly, a formal proof that the 19k protein precipitated by PAb419 is small t antigen will require analysis by two - dimensional gel electrophoresis and peptide mapping.

Saccharomyces cerevisiae has recently been successfully used as a heterologous system for the introduction, maintainance and expression of foreign genes from higher eukaryotes, with the cloning of a Drosophila gene complementing the ade 8 mutation in yeast

(Henikoff et al., 1981), the expression of interferon (Hitzemann et al., 1981; Tuite et al., 1982), Hepatitis B virus surface antigen (Valenzuela et al., 1982; Miyanohara et al., 1983), calf prochymosin (Goff et al., 1984) and thaumatin (Edens et al., 1984). The cDNA for rat growth hormone is apparently efficiently transcribed, but not translated (cited in Hitzemann et al., 1981). The bacterial antibotic resistance genes ß-lactamase (Roggenkamp et al., 1981) and chloramphenicol acetyl transferase (Cohen et al., 1980) have also been expressed in yeast. It is therefore necessary to consider why large T antigen was not synthesized in yeast. The block to expression could in principle be either at the level of transcription or translation.

The SV40 DNA inserted into the high copy number pJDB207 is clearly transcribed, though whether all, or indeed any of the transcripts are synthesized by RNA polymerase II is not known. The transcripts produced however are not of the size or number expected for the known early region products of SV40. Possible reasons for this include aberrant initiation of transcription, incorrect termination of transcription, or aberrant splicing of the primary transcript. It is known that both early region mRNAs of SV40 are spliced, so an analysis of the RNA synthesized in yeast was undertaken using the S1 nuclease mapping method (Favaloro et al., 1980). No correct use of acceptor splice site was observed, and only

very low level use of correct donor sites. The normal SV40 early polyadenylation site was not used, with at least some transcripts terminating prematurely. Evidence was also obtained that suggested that at least a fraction of the transcripts produced were spliced, though aberrantly. The data pertaining to these points however remain equivocal and the analysis was not pursued further. Such aberrant transcription and RNA processing has been observed in other studies of expression of genes from higher eukaryotes in S. cerevisiae (Beggs et al., 1980; Langford and Gallwitz, 1983; Watts et al., 1983). Yeast nuclear genes for RNA polymerase II do have introns in them (Ng and Abelson, 1980). A conserved sequence, TACTAAC, found within the introns of genes transcribed by RNA polymerase II has been shown to be essential for splicing to occur (Langford and Gallwitz, 1983; Langford et al., 1984). Such a sequence is not found within the intron of the SV40 early region mRNA, and thus provides a probable explanation for the observed lack of correct mRNA processing, though other possibilities, such as the prematurely terminated transcripts being unable to adopt the correct conformation to permit splicing to occur cannot a priori be ruled out. It should of course be noted that synthesis of small t antigen does not require splicing to occur, since the splice junction lies outside the coding sequence (Ready et al., 1979). Thus even an aberrantly spliced mRNA could potentially encode small t, provided that the splice junctions lie outside the coding

region.

Since small t antigen production was observed it suggests that the SV40 early promoter, or some sequence around that region is directing early region transcription. Preliminary primer extension analysis indicated 5' ends in the same region as SV40 early mRNAs initiate at during lytic infection, a suprising finding, in view of the greater distance (50 nucleotides or more) between the TATA box and mRNA 5' ends observed in S. cerevisiae though shorter distances have been found for some yeast genes (reviewed by Sentenac and Hall, 1982). Further work would clearly be necessary to verify this result.

High levels of expression would also require efficient translation, had correctly processed mRNA been synthesized. Following the abandonment of the use of yeast to express large T antigen, a study of codon usage in yeast was published (Bennetzen and Hall, 1982). Comparison of codon usage in T antigen (Tooze, 1981) with the preferred yeast codon bias indicates strong incompatibility in the codons for Phe, His, Pro, Asn, Lys, Gly and Leu, suggesting that even if correctly processed mRNA had been synthesized, translation would probably not have been very efficient.

Thus the use of other vector - host systems for the overproduction and ultimate purification of large T antigen were investigated.

Section 4.4: Purification of biologically active large T antigen.

The purification of large T antigen from lytically infected monkey cells is fraught with problems (see Section 4.1). Thus, attention has focused upon the use of adenovirus-SV40 hybrid viruses, constructed to overproduce large T antigen, to provide a plentiful source of starting material for purification. All of these hybrid viruses have had part of the viral genome replaced by SV40 sequences, thereby rendering them defective for lytic growth. Thus, the resulting defective virus must be grown with a complementing helper under selective conditions which maintain the virus in the population (Solnick, 1981; Thummel et al., 1981b, 1983). Under these conditions, the defective constitutes only a minor part of the population. Thus for ease of propagation, and stability of the virus, helper free adenovirus type 5 vectors have been developed (Gluzman et al., 1982). These have a deletion in the E1 region into which sequences may be inserted. This renders the virus defective for lytic growth, but if the virus is grown on 293 cells (Graham et al., 1977) which provide E1 region functions in trans then the recombinant virus may be propagated without the need for helper.

The virus used in these studies, Ad5 SVR111, has the SV40 early region, expressed under the control of the adenovirus major late promoter, cloned into the E1 region

of virus Δ E1/dl309 (Gluzman et al., 1982).

In developing a strategy for purification to retain biological activity of T antigen, speed is a prime consideration, since its DNA binding activity is known to be labile (Tenen et al., 1983). Also, an excessive number of column chromatography steps is undesirable if high yields are to be obtained, yet it must be ensured that the highest possible degree of purity is achieved to prevent interference in the analysis of T antigen function by contaminating proteins.

A very attractive potential solution to this problem would be to use the technique of immunoaffinity chromatography using an anti-T monoclonal antibody coupled to a solid support. This has the merit of being extremely specific, since only those proteins recognized by the antibody (and those complexed with them) will be retained on the column following extensive washing. The protein can then be eluted from the column by a mild change in conditions, such as a shift in pH. Immunoaffinity purifications have been developed for the middle T and large T antigens of polyoma virus (Walter et al., 1982; Dilworth et al., 1984) and the myc gene product (Donner et al., 1982).

Initial experiments in this laboratory using anti-53k monoclonal antibodies to purify the T antigen - 53k complex from transformed cells proved encouraging (D. P. Lane and J. Gannon, unpublished), so this approach was

pursued. An immunoaffinity matrix was constructed by coupling PAb419 (Harlow et al., 1981) to protein A Sepharose by the method of Schneider et al., (1982). The use of protein A Sepharose as a solid matrix rather than CNBr - activated Sepharose has the advantage that it permits optimal spatial orientation of the antibody, thereby potentially maximizing antigen binding capacity. The use of a monoclonal antibody rather than a polyclonal serum, means that only a single type of antibody - antigen interaction must be disrupted to displace the antigen from the column, leading to more efficient elution and permitting the use of less drastic conditions than might otherwise be necessary.

Thus, a high pH detergent extract of Ad5 SVR111 infected cells was made and formed the starting material for the column chromatography steps. A high pH was chosen for extraction in view of the results of Schwyzer et al., (1981), who showed that this minimizes proteolytic degradation of T antigen during extraction. As a further aid to preventing proteolysis, the protease inhibitor PMSF was added to all buffers. The first step is to pass the extract through a column of protein A Sepharose beads (2 ml volume) to remove proteins which stick non-specifically to protein A Sepharose. The flowthrough from this column, which contains 95% of the T antigen in the starting material, is then circulated through the PAb419 affinity column, which is then washed with high salt buffer (to remove non-specifically bound proteins), followed by a

buffer of low ionic strength so that when the eluting solution is applied, the pH shift is more rapid, thus displacing the T antigen from the column in a minimal volume. Samples from each stage in the purification were analyzed by gel electrophoresis and Western blotting. The results are shown in Figure 4.7 and Figure 4.8. It can be seen that the T antigen purified by this method migrates as a homogeneous, single band upon gel electrophoresis, with an apparent molecular weight of 90,000 daltons (Figure 4.7, lane h, lane T). The Western blot analysis (Figure 4.8, lane h) shows that the purified material reacts with a mixture of PAb204 (Lane and Hoeffler, 1980) and PAb423 (Harlow et al., 1981), different antibodies from those used for the purification. It is also clear that not all of the T antigen was extracted from these cells (Figure 4.8, lane c). This fraction is probably the chromatin and nuclear matrix bound fraction (Staufenbiel and Deppert, 1983) and could be extracted using a buffer containing high salt, e.g. 0.6 M NaCl.

This blot also demonstrates that the immunoaffinity column has very effectively depleted the extract of large T antigen, in that the flowthrough contains very little material that reacts with antibody at the molecular weight of T antigen (Figure 4.8, lane e). The lower band, which is not retained on the PAb419 immunoaffinity column, is a cellular protein with which PAb423 cross reacts (Crawford et al., 1892); also visible in the residual nuclear

Figure 4.7: Purification of T antigen.

Samples were subjected to electrophoresis through a 15% ($^{w}/v$) polyacrylamide gel, and visualized by staining with coomassie blue.

a: 293 cells infected with Ad5 SVR111.

b: Detergent extract.

c: Cell pellet after detergent extraction.

d: Extract after passage through protein A Sepharose column.

e: Extract after passage through PAb419 immunoaffinity column.

f: High salt wash of immunoaffinity column.

g: Low salt wash of immunoaffinity column.

h: High pH eluate from immunoaffinity column.

T: Purified T antigen from a different preparation to that shown in a - h.

M: Marker proteins.

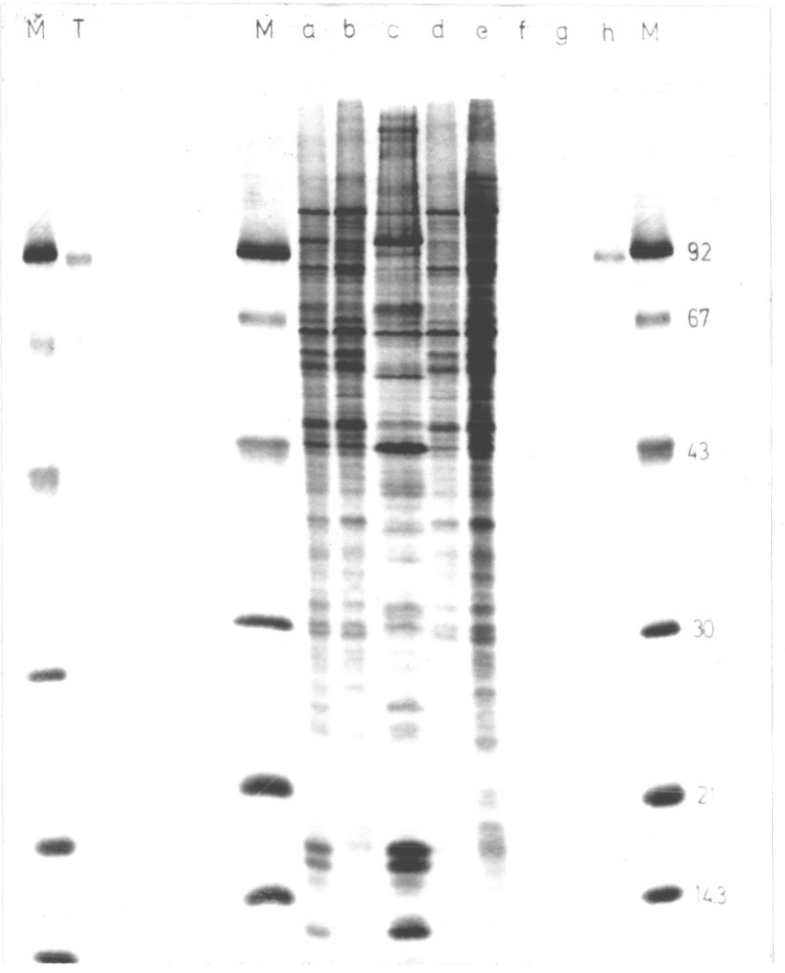
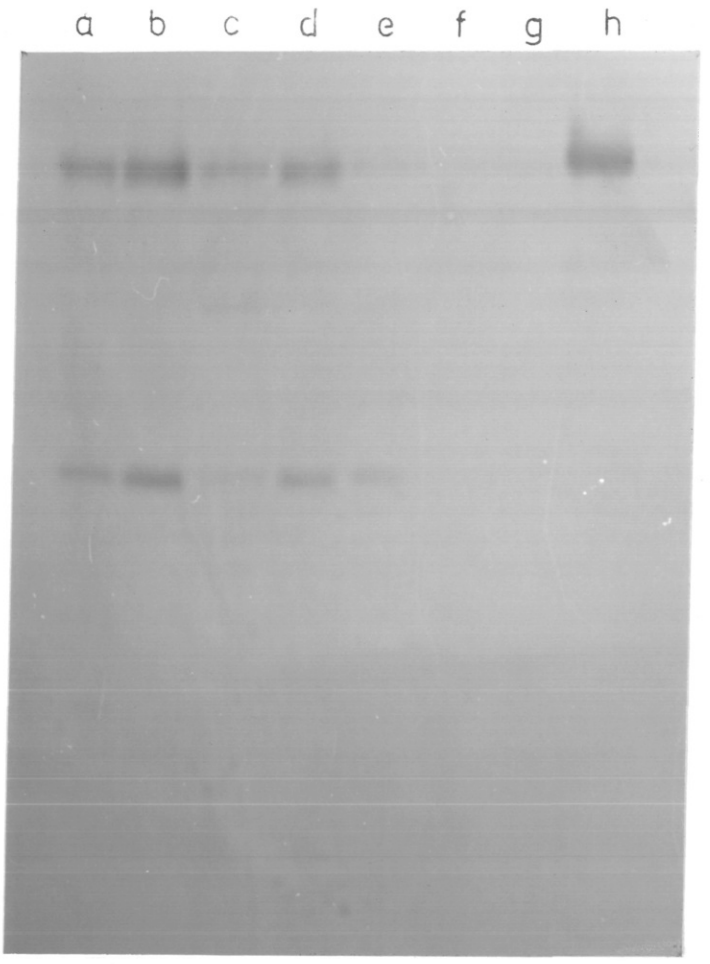

M T
M a b c d e f g h M
92
67
43
30
21
14.3

Figure 4.8: Western blot of T antigen purification.

Samples were subjected to electrophoresis as described in Figure 4.7 legend, transferred to nitrocellulose, and incubated with a mixture of PAb204 and PAb423. The filter was processed as described in Section 2.6. Lanes a - h are as in the legend to Figure 4.7.


a b c d e f g h

protein fraction is the 68k protein identified by Lane and Hoeffler, (1980) (Figure 4.8, lane c).

The availability of pure material permits development of a quantitative radioimmunoassay. Thus, as shown in Figure 4.9, a fixed amount of purified T antigen, labelled with ^{125}I by the iodogen method (the labelling was performed by J. Gannon, Imperial College) was incubated overnight with increasing dilutions of a rabbit serum raised against purified D2T protein (Lane and Crawford, 1980), followed by collection of the immunoglobulin bound T antigen by binding to S. aureus bacteria bearing protein A. Based on these results, 100 µl of a 1/4000 dilution of the serum was used in the subsequent quantitative assay. Thus, the same quantity of labelled T antigen was mixed with increasing amounts of unlabelled T antigen, whose concentration had been determined both by the method of Peterson (1977) and by reading the optical density at 280 nm. Serum was then added, and the samples incubated overnight, and processed as before. The unlabelled large T antigen competes with the labelled T antigen for binding to the limiting amount of antibody, decreasing the number of precipitated counts. This can then be used to quantitate the amount of T antigen in a protein extract, as shown in Figure 4.10. The assay as described is sensitive to at least 2 ng of T antigen. This assay was then used for a quantitative analysis of the purification procedure (Table 1).

The results obtained are in excellent agreement with

Figure 4.9: Titration of Rabbit anti D2 serum.

This figure shows the result of an experiment to determine the optimal dilution of antiserum for use in the large T radioimmunoassay. Samples were processed as described in Section 4.2. A background of 151 cpm was subtracted from all values.

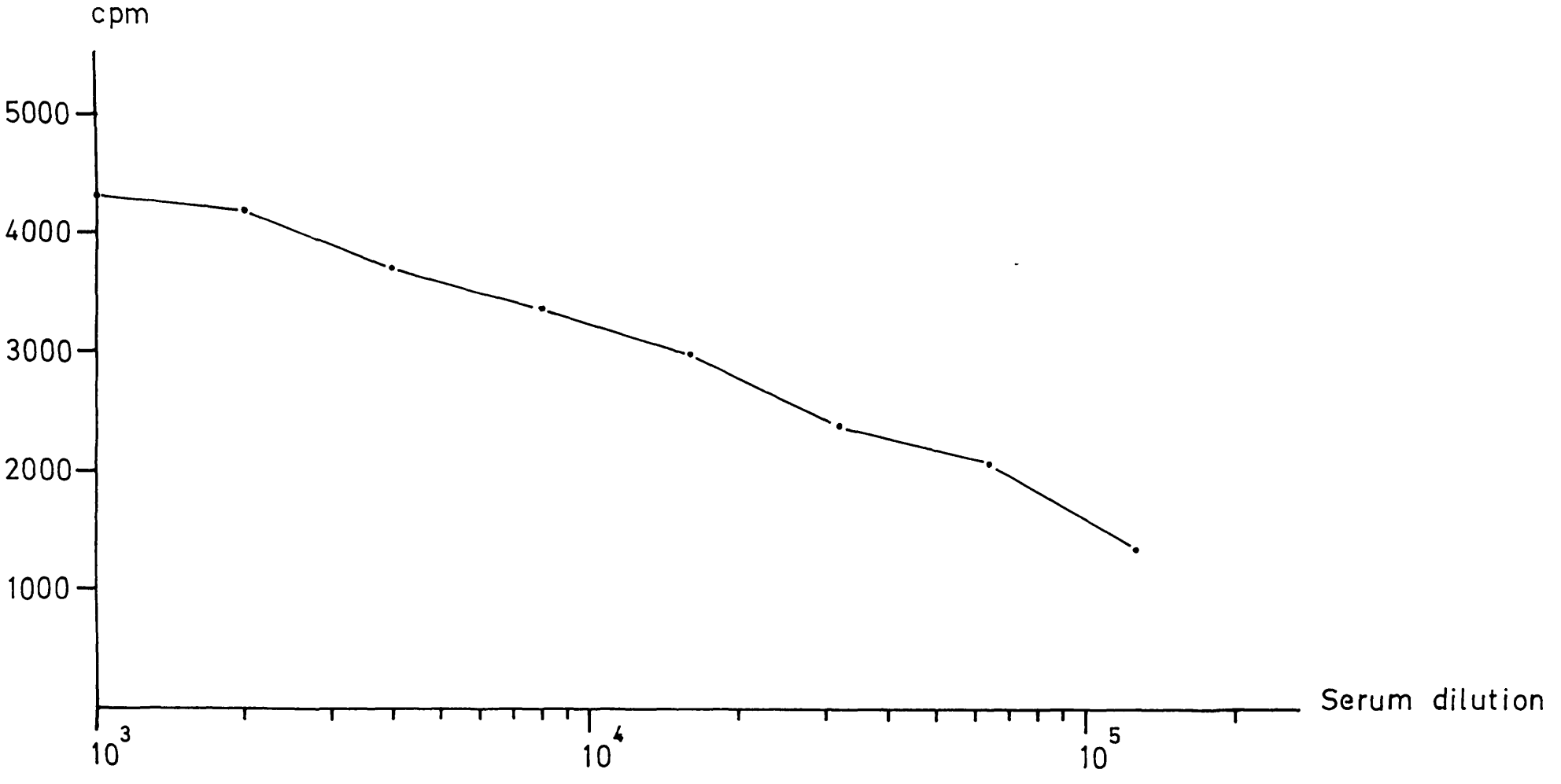
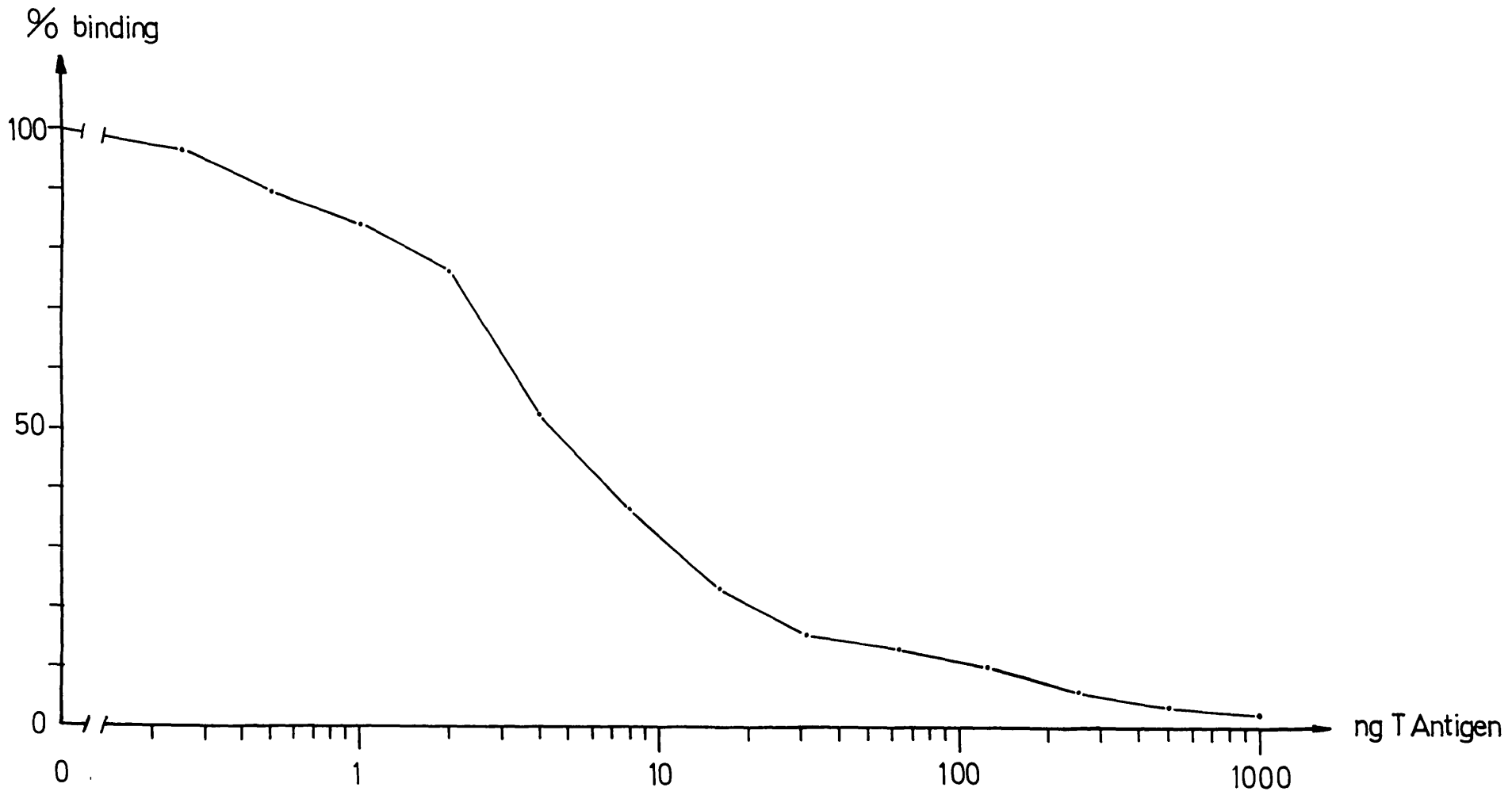

cpm
5000
4000
3000
2000
1000
Serum dilution
10^3
10^4
10^5

Figure 4.10: Standard curve for the large T radioimmunoassay.

Samples were processed as described in Section 4.2. A background of 141 cpm was subtracted from all values.


% binding
100
50
0
0
1
10
100
1000
ng T Antigen

the gel electrophoresis and Western blotting analysis, showing that only 1.4% of the T antigen in the starting material remains in the flowthrough after the PAb419 column. 63% of the starting material was recovered in the pooled eluate, a very high yield, and better than any previously published procedure. Thus, the T antigen produced by this method is undegraded, and at least 98% pure. It is physically undenatured as judged by the fact that all monoclonal antibodies bind to ^{125}I labelled protein in radioimmunoassays, including those which recognize conformationally sensitive determinants, such as PAb203 and PAb205 (Clark et al., 1981) (J. Gannon and D. P. Lane, personal communication). It therefore remains to be determined whether the protein possesses its known biological activities. Two such activities which may readily be analyzed are the ATPase activity of the protein and its interaction with the origin of SV40.

The interaction with SV40 DNA was first investigated using the McKay immunoprecipitation assay (McKay, 1981). SV40 DNA was digested with a combination of TaqYI and HindIII, end labelled with reverse transcriptase and α-^{32}P dGTP, and the mixture of fragments was incubated with purified T antigen. The T antigen was then immunoprecipitated with PAb419, a monoclonal antibody known not to interfere with DNA binding (Clark et al., 1983), and the co-precipitated DNA fragments eluted and analyzed by gel electrophoresis (Figure 4.11). In the absence of T antigen, no fragments are precipitated (lane

Table 4.1: Purification of T antigen.

This table presents the results of a quantitative analysis of the purification shown in Figures 4.7 and 4.8. The amounts of protein were determined by the protein assay described in Section 4.2, and by absorbtion at 280 nm. The amount of T antigen was determined by radioimmunoassay.

Purification of SV40 Large T-Antigen

	Amount of protein (µg)	Amount of T-Antigen (µg)	Recovery (%)
Detergent extract	8.4×10^4	320	(100)
Protein A Sepharose column flowthrough	8.29×10^4	304.5	95
PAb419 Immunoaffinity column flowthrough	7.87×10^4	4.4	1.4
High salt wash	1.5×10^3	none detected	-
Eluate from PAb419 Immunoaffinity column	206	202	63

b), while in the presence of T antigen, the B fragment, which contains the origin of replication, is precipitated, confirming previous studies (McKay, 1981) and demonstrating that the T antigen is functional, at least in this crude assay. It therefore remains to show that the protein will interact correctly with the binding regions at the origin of replication, using the footprinting technique of Galas and Schmitz (1978). The result is shown in Figure 4.12. The footprint obtained shows sequential binding to two regions at the origin. At low concentrations, binding occurs at site I, while at higher concentrations the protein occupies site II. As shown in Figure 4.13, the limits of the two interaction sites are in excellent agreement with earlier studies. No protection of region III was observed.

The second readily testable biological activity of T antigen which was analyzed is the ATPase activity of the protein. The results obtained are shown in Table 4.2. At a substrate concentration of 500 μM ATP, the rate of hydrolysis was 29.3 nmol ATP/min/mg protein, comparing very favorably with the data obtained by Tjian et al., (1979). Another readily testable parameter of the T antigen ATPase activity is its inhibition by monoclonal antibodies. PAb204 (Lane and Hoeffler, 1980) has been shown to strongly and specifically inhibit T antigen dependent ATP hydrolysis (Clark et al., 1981). Incubation of the purified T antigen with excess PAb204 prior to commencing the assay reduced the ATPase activity to less

Figure 4.11: McKay immunopreciptation assay using pure T antigen.

The assay was performed as described in Section 5.2.

Lane 1: Aliqout of input DNA fragments.

Lane 2: Immunoselected fragments in the presence of 50 ng of BSA.

Lane 3: Immunoselected fragments in the presence of 50 ng of pure T antigen.

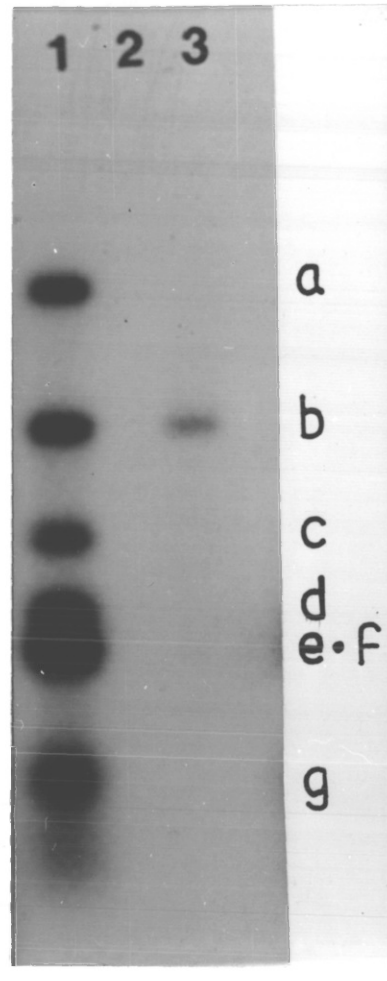

1
2
3
a
b
c
d
e
f
g

Figure 4.12: DNAse I footprint of the SV40 origin.

The footprinting was performed as described in Section 5.2, on an M13 clone of the HindIII - PvuII fragment spanning the origin of replication (coordinates 5171 - 270). The DNA concentration was 5.5×10^{-10} M, and the reaction volume was 25 µl.

Lane a, sample of probe; Lane b, 500 ng T antigen, no DNAse I added; Lanes c - h, 0, 5, 20, 100, 300 and 1000 ng of input T antigen, respectively, digested with DNAse I.

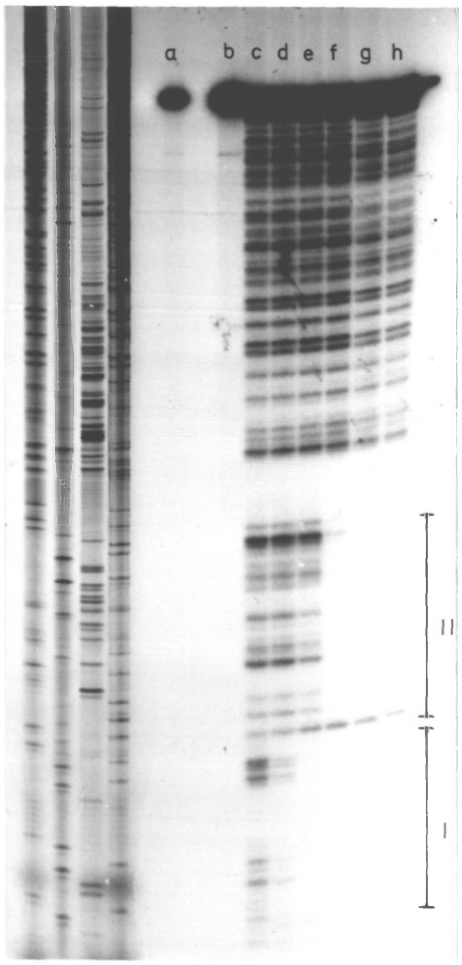

a
b c d e f g h
II
I

Figure 4.13: Summary of data from Figure 4.12.

Numbering is according to Tooze (1981). Data obtained by previous investigators is shown for comparison.

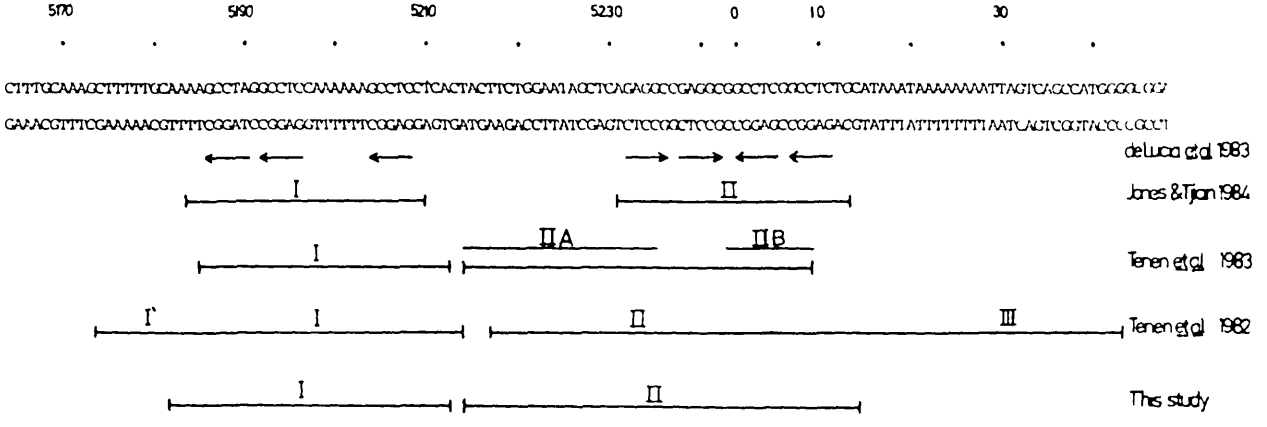

5170
5190
5210
5230
0
10
30
de Lucia et al 1983
I
II
Jones & Tjian 1984
IIA
IIB
I
Tenen et al 1983
I'
I
II
III
Tenen et al 1982
I
II
This study

Table 4.2: The ATPase activity of pure T antigen.

ATPase activity was assayed as described in Section 4.2.

ATPase activity of Immunoaffinity purified Large T-Antigen

Amount of T Antigen	Substrate concentration	Rate of hydrolysis (nmol/min/mg protein)	% activity
50ng	500 µM	28	
600ng	500 µM	29.3	
100ng	5 µM	8.25	100
100ng + 500ng PAb419	5 µM	7.8	95
100ng + 500ng PAb204	5 µM	0.37	4.5

than 5% of control values, while preincubation with PAb419 had only a very slight effect.

Therefore in assays for at least two of the biochemical activities of T antigen, the protein purified by this procedure behaves in a manner very similar to authentic T antigen. Other experiments using this purified material have indicated that it will reassociate with murine p53 *in vitro* as efficiently as T antigen in crude extracts (J. Gannon and D. P. Lane, personal communication).

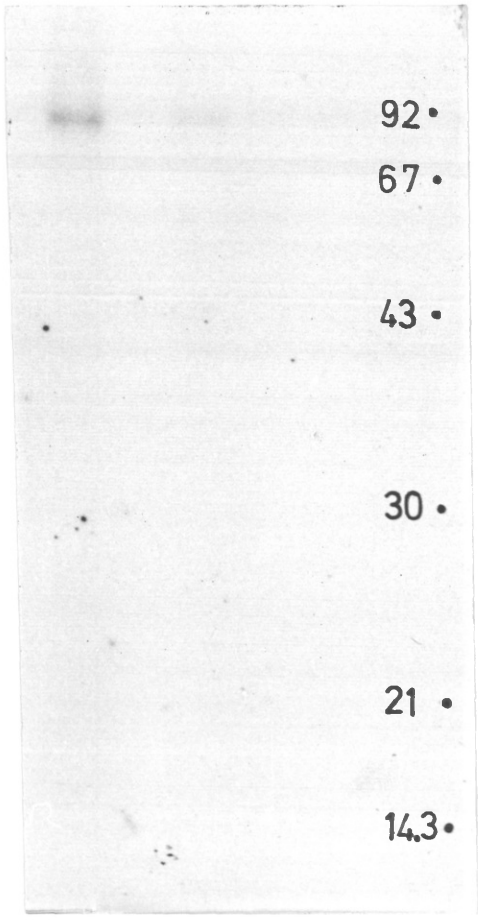
Another activity claimed for T antigen is that it has kinase activity, though whether this is intrinsic to the protein, or merely tightly associated with it remains a matter for debate. It has been demonstrated that the kinase activity is physically separable from T antigen by column chromatography or glycerol gradient centrifugation (Giacherio *et al.*, 1979; Tjian *et al.*, 1979) though Griffin *et al.*, (1979) have claimed that the two activities co-migrate. Since these three studies use different column chromatography matrices to effect separation of T antigen from contaminating proteins, their data are not directly comparable. It is clear however even in the studies of Griffin *et al*, (1979) that at least part of the T antigen can be separated from the kinase activity.

Immune complex kinase assays (Griffin *et al.*, 1979; Jay *et al.*, 1981a; Van Roy *et al.*, 1984) have shown that when the T antigen - 53k complex is precipitated from

transformed cells, the principal acceptors of phosphate *in vitro* are T antigen itself and 53k, rather than immunoglobulin. Thus, it was decided to investigate whether on incubation with γ -[^{32}P] ATP, the purified T antigen became labelled. The result is shown in Figure 4.14, and indicates that the large T band does indeed become labelled. This suggests, but does not prove that T antigen does possess kinase activity. To confirm this, however, the ability of T antigen to transfer phosphate to a known phosphate acceptor such as casein or phosvitin should be studied. Some ATPases proceed *via* an enzyme - phosphate intermediate (reviewed by Walsh, 1979), so it is possible that the labelled intermediate is what is being observed, though since the intermediate is readily labile to both acid and base, other explanations are also possible. The possibility of low levels of a contaminating kinase cannot be eliminated. However, purified T antigen subjected to further chromatography on hydroxyl-apatite also possesses this activity. Further work will be necessary to resolve this point. It would clearly be of interest to determine which amino acid residues become labelled during autophosphorylation both by total acid hydrolysis, and fingerprinting. The T antigen kinase has a very low specific activity (Tjian *et al.*, 1979). Whether this reflects the absence of the normal phosphate acceptor *in vivo* is unknown. It is tempting to speculate that in the cell either the T

Figure 4.14: Autophosphorylation of pure T antigen.

T antigen (1 µg) was incubated in 30 µl of ATPase assay buffer (Section 4.2), containing 10 µCi of γ -^{32}P ATP (3000 Ci/mmol, Amersham) at room temperature for 20 minutes. 30 µl of 2 x sample buffer (Table 2.6) was added, and after boiling for two minutes the sample was subjected to electrophoresis through a 15% (^{w}/v) acrylamide gel. The gel was stained to reveal the position of marker proteins, dried and autoradiographed.


92
67
43
30
21
14.3

antigen - p53 complex is a kinase, or that the T antigen kinase is activated by the binding of p53. It is of course possible that p53 is itself a kinase, though this point has been disputed. Van Roy et al., (1984) have failed to demonstrate any kinase activity associated with p53 from chemically transformed cells, while Jay et al., (1981) have done so. The latter report however did not demonstrate that the protein phosphorylated was p53, and since the two groups used different antibodies to precipitate p53, it could be argued that the discrepancy is due to inhibition of 53k function by some antibodies but not others, as observed for the T antigen ATPase (Clark et al., 1981). Thus, the question of whether T antigen is a kinase remains unresolved.

Section 4.5: Discussion.

The use of cells infected by adenovirus-SV40 hybrid viruses as a source of starting material for purification of biologically active T antigen has superseded the use of all other sources. The exception to this, as discussed in section 4.1 is purification from lytically infected monkey cells. It is clear now that S. cerevisiae does not represent a good host system for expression of genes from higher eukaryotes which contain introns within their coding sequences. Those genes which have been successfully expressed have either been expressed as cDNA clones, or have lacked introns. Thus, to have obtained

expression of large T antigen, the use of a cDNA clone would have been necessary; until recently, (Y. Gluzman, personal communication) no such clone has been available. As discussed in Section 4.3, it is doubtful whether T antigen mRNA would have been efficiently translated, as its codon bias differs significantly from that of abundant yeast proteins. Then, of course, additional problems that could have been encountered are, instability of the T antigen in yeast, difficulties with extraction and the possibilty that the production of T antigen in the yeast cell would be a lethal event. The protein might also have been incorrectly modified in yeast cells thus altering its function.

A rapid, very simple procedure for the purification of T antigen has been developed which combines the power of immunoaffinity chromatography with the use of an SV40-adenovirus hybrid virus which overproduces large T antigen. The final product obtained is >95% pure, full length T antigen, with no detectable proteolytic breakdown products present, as judged by staining of gels and Western blotting, which retains all of the known biochemical activities of T antigen that have been assayed for.

The purified protein has an ATPase activity, whose specific activity compares very favorably with that determined for T antigen purified by other methods, and which is specifically inhibited by some monclonal

antibodies to T antigen but not others. It retains the ability to specifically bind to the SV40 origin of replication, and sequentially protects two regions from DNAse I digestion in a manner very similar to that of T antigen extracted from lytically infected cells. Other experiments using this purified material (J. Gannon and D. P. Lane, personal communication) have indicated that the protein will reassociate with murine p53 *in vitro* as efficiently as crude extracts containing T antigen.

Several points concerning the purification procedure are also worthy of note. In the cell, T antigen is complexed with the 53k cellular protein (Lane and Crawford, 1979) and apart from large T antigen, PAb419 also recognizes small t antigen, and a cellular 35k protein (Harlow *et al.*, 1981; Crawford *et al.*, 1982), yet none of these proteins is detectable in the purified material. The association of T antigen with human 53k is known to be quite weak and readily disrupted (Lane *et al.*, 1982). The use of relatively stringent washing conditions prior to elution may account for its absence. PAb419 has a 10 fold lower affinity for the 35k cellular protein than for T antigen (Crawford *et al.*, 1982), and as the protein is only present in small amounts relative to T antigen in the infected cell, its absence is not surprising. The absence of small t antigen may be due to the fact that the virus overproduces large T antigen, combined with a lower steady state level of small t than large T, and also the fact that PAb419 recognizes only a subset of small t

molecules present in the lytically infected cell (X. Montano and D. P. Lane, unpublished data). The immunoaffinity column made as described in Section 4.2 has been found to be reusable many times without significant loss of binding capacity. Clearly, other monoclonal antibodies could be applied to the column in place of PAb419; for instance, the use of a lower affinity antibody might permit the use of milder elution conditions than those used at present. Antibodies recognizing only a fraction of the T antigen molecules could be used to separate functional subsets of T antigen. Other agents could also be used to effect elution from the column, such as high concentrations of $MgCl_2$, or if the binding site for PAb419 were to be precisely determined, then a peptide covering it could be synthesized and used for elution. Both of these approaches have been successfully applied in immunoaffinity purifications for polyoma middle and large T antigens (Walter et al., 1982; Dilworth et al., 1984). The entire purification procedure as described may easily be completed within a day and is now routinely used in our laboratory for T antigen purification.

The availability of pure protein thus permits detailed biochemical and physical analyses of wild type and mutant T antigens to be undertaken. A very interesing line of study would be to try to separate functional domains of the protein by proteolysis. Preliminary studies of this kind have already been undertaken

(Morrison et al., 1983) and have identified an N-terminal domain capable of binding to the origin of replication. Clearly, this type of study ought to be extended, isolating functional domains on a preparative scale. If sufficient T antigen could be purified, the determination of the crystal structure by X-ray crystallography could be attempted, perhaps initially on isolated domains. It would also be interesting to grow co-crystals of T antigen or the origin binding domain, bound to the origin or subfragments thereof, to determine if the mode of interaction with the DNA is the same as that of bacterial repressor proteins.

Detailed studies of the kinetics of binding to the origin and also to non-specific DNA could be undertaken, and the precise species of T antigen bound to DNA, and the regions of T antigen which make contact with DNA could be investigated using appropriate cross - linking reagents. The enzymology of T antigen could also be analyzed in greater kinetic detail, and active site labelling could be used to identify the location of the ATPase active site (Walsh, 1979). Clearly, were p53 or the T-p53 complex available in pure form then its interaction with T antigen could be studied, determining which regions of the two proteins make contact in the complex, and how the biochemical properties of the complex differ from the individual pure proteins.

CHAPTER FIVE

ANALYSIS OF NON - ORIGIN CLONES I:

CLONE p27

CHAPTER FIVE

INTERACTION OF T-ANTIGEN WITH NON-ORIGIN CLONES: CLONE p27

Section 5.1: Introduction.

A variety of techniques have been developed for studying the interaction of proteins with DNA. The finding that nitrocellulose filters will not retain naked double stranded DNA has been exploited to identify regions of DNA which contain preferred binding sites for DNA binding proteins. The standard approach has been to mix the DNA binding protein of interest with the DNA thought to contain the binding site, and after a suitable period of time, filtering the mixture through nitrocellulose. Protein - DNA complexes are retained on the filter, while uncomplexed DNA passes through. The retained DNA can then be eluted from the filter and analyzed. In general, the DNA is radioactively labelled, and is added to the reaction either as a mixture of restriction fragments, in which case preferential retention of one or more fragments is expected, or the DNA - protein complexes are extensively digested with DNAse I prior to filtration, in which case, only those regions of the DNA to which protein is bound will be protected from digestion and retained on the filter. The retained DNA is then eluted and its origin from within the DNA under study identified. Both of these approaches have been successfully employed to

study the interaction of eukaryotic DNA binding proteins with their specific sequences. Tjian (1978) used DNAse I digestion to identify those regions of SV40 DNA to which large T antigen bound, while Payvar et al., (1983) have used preferential retention of one fragment over others to define binding sites for the glucocorticoid receptor in the MMTV genome. This approach has also been applied to crude, partially fractionated protein extracts to identify proteins that bind specifically in the vicinity of the heat shock loci of Drosophila (Jack et al., 1981), and also to identify specific sequences bound by DNA binding proteins of unknown function (Weideli et al., 1980).

An alternative method is the DNA cellulose competition assay, in which purified protein, or crude extracts are bound to DNA cellulose, and after extensive washing, the DNA cellulose is incubated with various DNA fragments suspected to contain high affinity binding sites for the protein under study. If they contain such a site, then the protein will bind preferentially to the input fragment, and so a greater proportion will be eluted off the DNA cellulose. Thus after removal of the DNA cellulose, the supernatant is assayed for the amount of the protein of interest released. This approach has been used to study the binding of the progesterone receptor to DNA (Mulvihill et al., 1982) and to identify a high affinity binding site for polyoma virus T antigen at the viral origin of replication (Gaudray et al., 1981).

To study proteins for which an antibody is available, an immunoprecipitation assay such as that of McKay (1981) may be used. The basic method is to incubate a mixture of DNA fragments with crude extracts containing the protein, though purified protein can also be used, followed by addition of antibody. The antibody - antigen complexes are then collected on solid - phase protein A, and after washing, the co- precipitated DNA is eluted and analyzed. This immunoselection step permits a DNA binding protein to be analyzed in the presence of many others in crude extracts, and has been successfully applied to study SV40 large T antigen (McKay, 1981) and steroid hormone receptors (Scheidereit et al., 1983).

Once a binding site has been tentatively identified, a more detailed analysis can be undertaken. The DNAse I footprinting technique of Galas and Schmitz, (1978) has been widely used. The method involves incubating a restriction fragment, labelled at one end only, with the protein under study. The DNA is then digested with DNAse I, such that an average of one nick per molecule is incorporated into the DNA. The sample is then analyzed by electrophoresis on denaturing gels. Those regions of the DNA to which the protein is bound are protected from digestion, so comparison of the "footprint" obtained in the presence and absence of protein identifies binding sites for the protein. A similar rationale has been applied to define those purine residues involved in contacts with the protein by their altered susceptibility

to methylation by dimethyl sulfate (see for instance de Lucia et al., 1983). A complementary technique to this is alkylation interference, when the templates are chemically modified, at an average of one modification per molecule, before the protein is added. After binding, bound and unbound templates are separated, typically by immunoprecipitation and analyzed. Under appropriate conditions, unbound templates will be those with residues modified at points which are important in protein - DNA interaction. This can then be used to define purine residues, by DMS methylation, and phosphate groups, by reaction with ethyl nitrosourea, involved in contacts with the protein (see Sakonju and Brown, 1982; Jones and Tjian, 1984).

The experiments described in this, and the following chapter are addressed towards demonstrating a preferential interaction in vitro of T antigen with the clones generated by the immunoselection procedure defined in Chapter Three, initially using crude T antigen containing extracts, and then by the DNAse I footprinting technique described above.

Section 5.2: Materials and Methods.

DNA cellulose competition assay (Gaudray et al., 1981)

Double stranded DNA cellulose (Sigma, 2 mg DNA/g cellulose) was equilibrated with binding buffer [10 mM

MOPS pH 7.0, 1 mM $MgCl_2$, 1 mM DTT, 100 mM NaCl, 0.05% (^{v}/v) NP-40, 10% (^{v}/v) glycerol], at 4°C for 2 hours, or longer. The T antigen commonly used was extracted from SV40 infected CV1 cells, 48 hours after infection at a multiplicity of 10 pfu per cell. Typically, 5 infected 9 cm dishes were washed twice with ice cold TD, and the cells removed from the dish with 5 mM EDTA. Crude T antigen containing extracts were made by lysing the cells with extraction buffer, 50 mM Tris-HCl pH 8.0, 120 mM KCl, 1 mM DTT, 0.5% (^{v}/v) NP-40 containing 1 mM PMSF. 0.25 ml extraction buffer was used per dish. After 15 minutes on ice, cell debris was removed by centrifugation in a microcentrifuge for 15 minutes at 4°C. The supernatant was removed, diluted 10 fold in binding buffer, and applied to 0.2 g of equilibrated DNA cellulose. After one hour at 4°C with end over end rotation, the DNA cellulose was collected by gentle centrifugation and washed with 100 ml of binding buffer over 2 - 3 hours. The DNA cellulose was then resuspended in 4 ml of binding buffer and 200 µl of suspension was added to each assay, which contained 2 - 5 µg of recombinant plasmid, purified either by CsCl - ethidium bromide gradient centrifugation, or by the rapid alkaline lysis procedure of Birnboim and Doly, as described in Chapter Two. When miniprep DNA was used, it was deproteinized by phenol extraction, and treated with ribonuclease prior to use. Samples were incubated at 4°C for one hour with end over end rotation. The DNA cellulose was then removed by centrifugation for 2 minutes

in a microfuge, and 50 µl of the supernatant assessed for T antigen using the assay described below.

ELISA assay for SV40 large T antigen (J. Gannon and D.P. Lane, unpublished data)

PAb204 (Lane and Hoeffler, 1980) was purified by affinity chromatography on Protein A sepharose in 10 mM sodium phosphate buffer, pH 8.0. Antibody was eluted from the resin by washing with 0.1 M citrate buffer, pH 6.0 and stored frozen at -20°C. 1 µg aliquots were bound to the bottom of the wells of 3911 microtiter dishes (Falcon), in 50 µl of PBS, overnight in a humidified box. The wells were then washed twice with PBS, and filled with 5% (^{w}/v) BSA in PBS, and incubated as before. The BSA was then discarded and the plates wrapped in cling film and stored at -20°C until required. The sample was added to the well, and incubated overnight at 4°C in a humid atmosphere. The wells were then washed three times with PBS containing 0.5% (^{v}/v) NP-40 and three times with PBS. 200 ng of biotin - coupled PAb419 (Harlow et al., 1982), a kind gift from J. Gannon, Imperial College, was then added, in 50 µl of 5% (^{w}/v) BSA in PBS. After three hours, the wells were washed three times in PBS, twice in PBS containing 0.5% (^{v}/v) NP-40, and three times in PBS. 5 µg of horse - radish peroxidase conjugated to avidin (Miles) in 50 µl of 5% (^{w}/v) BSA in PBS was added. Three hours later the plates were washed thrice in PBS, twice in PBS containing 0.5% (^{v}/v) NP-40 and then

incubated for 30 minutes in PBS ontaining 0.1% (^{v}/v) NP-40, followed by washing twice in PBS. The detecting antibody was visualized using 3,3'-5,5' Diaminobenzidine (Miles) substrate prepared as follows: The substrate was dissolved at a concentration of 10 mg/ml in DMSO, and diluted to 100 µg/ml in freshly made 0.1 M citrate - phosphate buffer, pH 6.0. H_2O_2 (30 volumes, BDH) was added to 0.02% (^{v}/v) and 50 µl of substrate added per well. After 60 seconds, the sample was transferred to a second plate containing 50 µl 2.5 M H_2SO_4. The yellow color was read on a Flow Titretek ELISA plate reader, using a 450 nm filter.

McKay immunoprecipitation assay (McKay, 1981)

Subconfluent CV1 cells were infected with SV40 virus at a multiplicity of 10 pfu per cell. After 24 hours the cells were removed from the dish with 5 mM EDTA, washed twice in ice cold TD, and resuspended in 100 µl per 9 cm dish of lysis buffer [50 mM Tris-HCl pH 8.0, 120 mM KCl, 1 mM DTT, 5% (^{v}/v) glycerol, 0.5% (^{v}/v) NP-40, 1 mM PMSF]. The cells were vortexed for 30 seconds, and incubated on ice for 15 minutes. Cell debris was then removed by centrifugation for 15 minutes at 4°C in a microfuge. The supernatant was removed and diluted 10 fold in ice cold binding buffer [10 mM Tris-HCl pH 7.4, 150 mM KCl, 1 mM DTT, 5% (^{v}/v) glycerol, 1 mM PMSF, 0.05% (^{v}/v) NP-40] and used immediately. For some assays, lower salt concentrations were used; when this was the case, this is indicated in the appropriate figure legend.

100 µl of T antigen extract (or uninfected cell extract) was incubated with 0.2 - 0.5 µg of end labelled restriction fragments in 50 µl of binding buffer on ice for 15 minutes. 100 µl of PAb419 was then added. After 45 minutes on ice, 10 µl of 10% (^{v}/v) S. aureus bearing protein A, previously washed five times in NET buffer [50 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM EDTA, 0.5% (^{v}/v) NP-40] were added. 15 minutes later the bacteria were collected by brief centrifugation, and washed three to five times with 1 ml of NET buffer. Co-precipitated DNA was eluted by resuspending the washed pellet in 25 µl of 1% (^{w}/v) SDS, 10 mM Tris-HCl pH 7.4, 10 mM EDTA, 5% (^{w}/v) sucrose and incubating at 65°C for 20 - 30 minutes. The bacteria were removed by brief centrifugation and the supernatant analyzed by agarose gel electrophoresis. After electrophoresis, gels were fixed by immersion in 10% (^{v}/v) acetic acid, and dried under vacuum prior to autoradiography.

DNAse I footprinting (Galas and Schmitz, 1978)

In a typical reaction, 5 - 25 ng of end labelled restriction fragment was incubated in 25 - 100 µl volume of binding buffer [10 mM PIPES pH 7.0, 1 mM DTT, 0.1 mM EDTA, 10% (^{v}/v) glycerol, 1 - 150 mM NaCl containing 100 µg/ml BSA). After 15 minutes on ice, $MgCl_2$ and $CaCl_2$ were added to final concentrations of 5 mM and 1 mM, respectively, together with 0.5 ng DNAse I. Digestion was terminated after 90 - 120 seconds on ice by addition of 1

volume of 1 M ammonium acetate, 0.2 M EDTA pH 8.0, 0.2% ($^{w}/v$) SDS and 10 µg of sonicated salmon sperm DNA. After extraction with phenol and chloroform, the sample was precipitated with ethanol and analyzed by electrophoresis on denaturing acrylamide gels as described in Chapter Two.

For some footprints, a modified protocol for synthesizing the template was used. An M13 clone of the fragment of interest was generated, and prepared in bulk. This was annealed with 5' end labelled primer (pentadecamer, New England Biolabs). A typical labelling reaction contained 50 ng primer, 100 µCi (γ-^{32}P) ATP, (Amersham, 3000 Ci/mmol) and 10 units of polynucleotide kinase in a volume of 50 µl. For synthesis reactions, 20 - 25 ng primer was heated to 75°C with 10 - 20 µg of single stranded DNA in 50 µl 10 mM Tris-HCl pH 8.0, 7 mM $MgCl_2$ and cooled slowly to room temperature. Deoxynucleotide triphosphates were added to a final concentration of 200 µM followed by 5 units DNA polymerase I (Klenow fragment, BRL), and the reaction incubated at 25°C for 5 - 10 minutes. The polymerase was inactivated by heating to 70°C for 15 minutes, and ionic conditions adjusted to those suitable for digestion with a restriction enzyme which cleaves within the M13 polylinker, on the 3' side of the inserted DNA. Following digestion, the sample was adjusted to pH 4.5, 4.5 mM zinc acetate, 300 mM NaCl and S1 nuclease added to 500 units per ml. After 10 minutes at room temperature, the sample was precipitated with ethanol following addition of 0.25

volume of 2.5 M ammonium acetate, 100 mM EDTA and tRNA to 10 µg/ml as carrier. The sample was then resuspended in binding buffer, treated with DNAse - free RNAse for 30 minutes at 37°C, extracted with phenol, chloroform and ether, and used for footprinting studies as described above.

Size markers for conventional footprints were generated using C + T cleavage of the end labelled template, while conventional di-deoxy sequencing techniques were used for the alternative method.

Section 5.3: Screening of clones by DNA cellulose competition assay.

A rapid initial screening of the clones from Experiments One and Two were undertaken using a DNA cellulose competition assay, the principle of which has been described in Section 5.1. For most clones, DNA extracted from 10 ml cultures by the alkaline lysis method was used, after deproteination by extraction with phenol and chloroform, followed by treatment with RNAse. Typically, 10 - 20 µg of plasmid DNA were extracted from such a culture.

A variety of problems were encountered using this technique, such as variable background amounts of T antigen eluted either in the absence of added DNA, or by addition of vector DNA, pAT153, and variation in the

amount of T antigen eluted by SV40 DNA experiments. As a consequence of this, only a qualitative judgement of whether or not a particular construct eluted more T antigen than vector alone was made. Using such an assay, it was found that 50 of the 64 clones derived in Experiment One, and 41 of 53 clones in Experiment Two apparently eluted higher than background levels of T antigen from the DNA cellulose. Thus, approximately 78% of clones may be considered to contain DNA sequences for which T antigen has a higher affinity than calf thymus DNA, a finding which is consistent with the view that the immunoselection cloning procedure has enriched for such sequences. A valuable additional experiment which would support this conclusion would be to screen an equivalent number of random clones with similar insert size, to determine the percentage of such clones which will elute signigicant amounts of T antigen.

Three clones generated in the first experiment which displaced higher than background amounts of T antigen from DNA cellulose were selected for further study. The remainder of this chapter is devoted to the analysis of one of them, p27, while the two other clones, p24 and p29 are discussed in Chapter Six.

Section 5.4: Clone p27

Clone p27 was analyzed using six - base recognition restriction enzymes and the restriction map shown in Figure 5.1 derived. Based on this, an EcoRI, SalI double digest of p27 was end labelled using reverse transcriptase and α-^{32}P dTTP, and used in a McKay immunoprecipitation assay. This digest separates the entire insert from the bulk of pAT153, leaving it flanked by approximately 300 bp of vector either side. Control experiments, see Figure 6.2, demonstrated that these vector sequences contain no high affinity T antigen binding site, as judged by McKay immunoprecipitation assay; thus if the smaller, insert containing fragment is preferentially co-precipitated it is likely that it contains a high affinity T antigen binding site. The result of such an asay, carried out under stringent binding conditions (Figure 5.2) shows that the smallest fragment is immunoprecipitated when extracts from SV40 infected monkey cells are used (Figure 5.2, lane 3), but not by extracts from uninfected cells, (Figure 5.2, lane 2) suggesting that the insert contains a T antigen binding site.

The insert DNA was then sequenced using the dideoxy technique, cloning the 200 bp BamHI fragment in both orientations in M13 mp 8 (Messing and Vieira, 1982) and the BamHI - SalI fragment shown in Figure 5.1 into M13 mp 9. Each clone was sequenced twice, and in the case of the 200

Figure 5.1: Restriction map of clone p27.

The restriction map of p27 is presented. The region spanning the EcoRI and SalI sites of pAT153 is shown in linear form, at a scale of 2cm = 100 bp. Abbreviations for restriction enzymes are: B, BamHI; E, EcoRI; R, RsaI; S, SalI. Thin lines denote vector sequences, thick lines denote insert DNA.

Restriction map of p27

100 b.p.

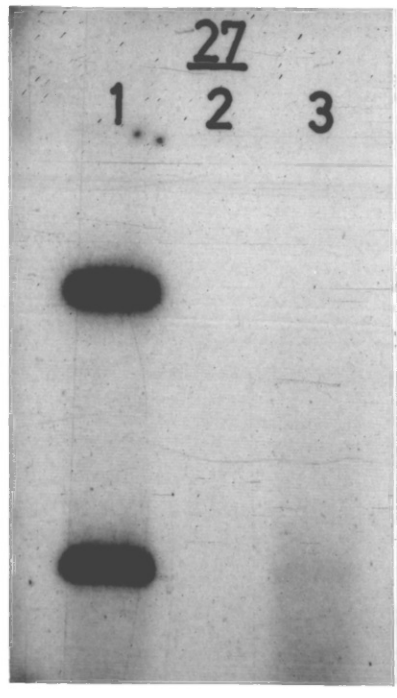
Figure 5.2: McKay immunoprecipitation assay on clone p27.

The assay was performed as described in Section 5.2. The immunoprecipitating antibody was PAb419, and the salt concentration in the binding buffer was 150 mM.

Lane 1, sample of input DNA.

Lane 2, DNA coprecipitated by uninfected CV-1 cell extract.

Lane 3, DNA coprecipitated by SV40 infected CV-1 cell extract.


27
1
2
3

bp BamHI fragment, at least two clones for each strand were sequenced. The sequence obtained is shown in Figure 5.5. The sequence of the 205 bp BamHI fragment was determined on both strands, and that of the remainder of the clone on one strand only. The inserted DNA is 237 bp in length. A small deletion has been introduced in the vector and the inserted DNA, since the sequence of the clone ends at the nucleotide shown, and then reads into the pBR322 sequence beginning at nucleotide 411, deleting the BamHI site at nucleotide 375 in pBR322.

A detailed study of the region of the insert with which T antigen interacts was then undertaken using the footprint technique of Galas and Schmitz. For the experiment shown in Figure 5.3, p27 was digested with BamHI, dephosphorylated, 5' end labelled with polynucleotide kinase, and the small BamHI fragment gel purified. This was then re-cut with RsaI and the large fragment gel purified, to give a fragment labelled on only one strand which was then used for footprinting studies, as described in Section 5.2. From the result, it is clear that one region of the clone is protected from DNAse I digestion by T antigen (Figure 5.3). The protected region is approximately 35 bp in length, in good agreement with the protected fragment size observed by Tjian (1978) when T antigen binds at site I. More precise mapping was undertaken using the M13 footprinting technique described in Section 5.2. The results are shown in Figure 5.4. From these studies it is apparent that both strands are

Figure 5.3: Footprinting of clone p27: I.

10 ng of the 182 bp RsaI - BamHI fragment was mixed with the indicated amounts of T antigen in binding buffer containing 5 mM NaCl, and after 30 minutes on ice, digested with 0.5 ng of DNAse I for 5 minutes at $4^{o}C$. The digest was terminated and the products analysed on denaturing, 15% ($^{w}/v$) acrylamide gels. The wet gel was autoradiographed at $-70^{o}C$. The DNA concentration was $8 \times 10^{-10}M$.
C.T, C + T cleavage of 3' end labelled DNA to provide markers; lane a, input DNA; lanes b - f, 0, 20, 50, 200 and 900 ng of pure T antigen, respectively.

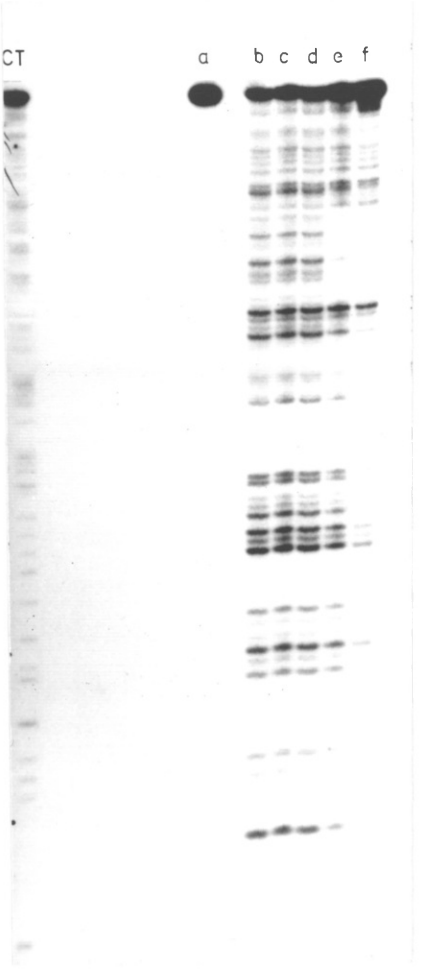
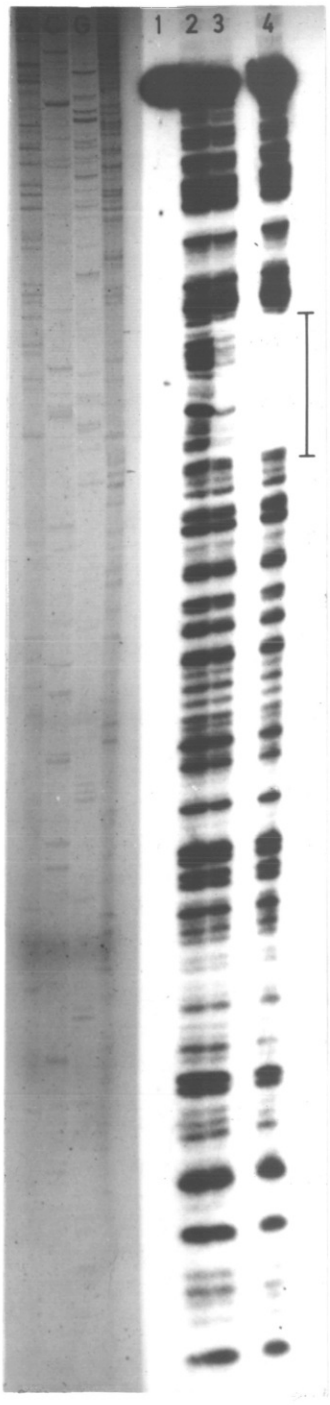

CT
a
b c d e f

Figure 5.4: Footprinting of clone p27: II.

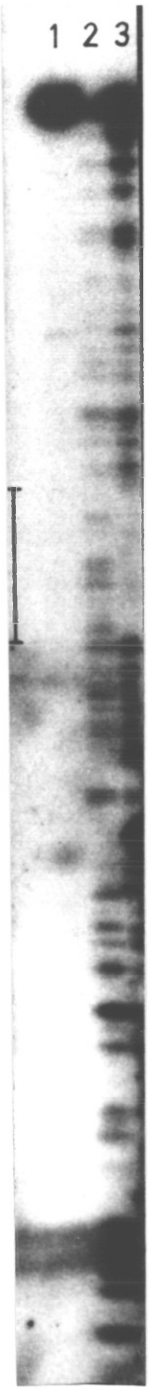
The indicated amounts of T antigen were mixed with p27 template, synthesised on M13 clones of the two strands, and incubated at $4^{o}C$ for 15 minutes. The samples were then digested for 90 seconds with of DNAse I, and further processed as described in Section 5.2.
Panel A: Footprint of the upper strand shown in Figure 5.5: lane 1, input DNA; lane 2, no T antigen added; lane 3, 50 ng T antigen, 5 mM NaCl binding buffer, 0.3 ng DNAse I; lane 4, 400 ng T antigen, 150 mM NaCl binding buffer, 3 ng DNAse I.
Panel B: Footprint of the lower strand shown in Figure 5.5: lane 1, input DNA; lane 2, no T antigen added; lane 3, 400 ng of T antigen, 150 mM binding buffer, 3 ng DNAse I.

In both experiments, the DNA concentration was 6×10^{-10}M, in a 25 µl reaction volume. A, C, G, and T denote bases in the parallel sequencing reaction to provide size markers.

A



B



protected in the same region. Similar results were obtained when the binding was carried out at 150 mM NaCl (Figure 5.4, panels A and B), suggesting that it is the binding to this region that is being observed in the McKay immunoprecipitation assay. The extent of the protected region within p27 is shown in Figure 5.5 within the bracketed region. The T antigen binding site on p27 becomes occupied at protein - DNA ratios intermediate between those required to fill sites I and II at the SV40 origin, suggesting that the affinity of T antigen for this site is intermediate between that of sites I and II. The DNAse I protected region contains three copies of the consensus pentanucleotide thought to comprise part of the structure recognized by T antigen, underlined in Figure 5.5, though whether these residues make contacts with T antigen awaits higher resolution analysis using DMS protection and alkylation interference techniques. It is of interest to note that two of the pentanucleotides are arranged in tandem as found at site I, though separated by six, rather than seven nucleotides, which may account for the lower affinity compared to site I.

p27 was nick translated and used to probe genomic DNA extracted from Balb/c 3T3 cells, digested with BamHI or EcoRI (Figure 5.6). A smear of hybridization was obtained, indicating the presence of a dispersed, repetitive element.

The sequence was then compared with the GENBANK and EMBL sequence databases using the algorithm of Wilbur and

Figure 5.5: The DNA sequence of clone p27.

The DNA sequence was determined using the dideoxy sequencing method. Both strands are shown, starting at the left hand BamHI site shown in Figure 5.1, through to the end of the insert. Pentanucleotides similar to those defined by de Lucia et al., (1983) are indicated by arrows. The brackets indicate the extent of the DNAse I protection.

SEQUENCE OF p27

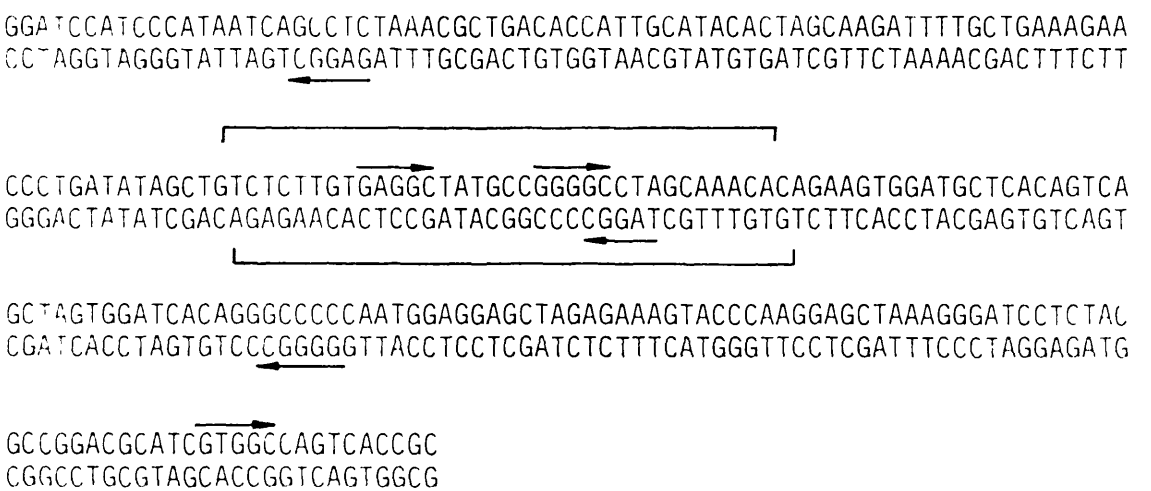
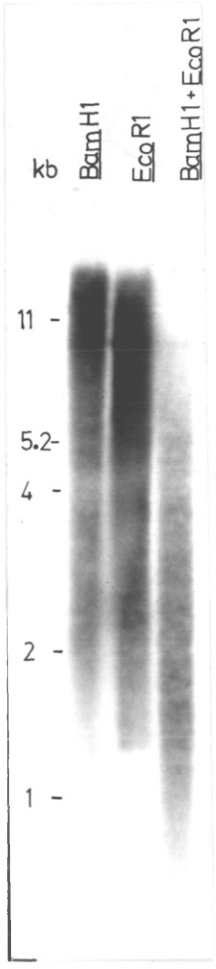


Figure 5.6: Sequences homologous to p27 in mouse DNA.

10 μg of high molecular weight DNA from Balb/c 3T3 cells was digested overnight with 30 units of the indicated enzymes, and transferred to Gene Screen (NEN). The filter was hybridized as described in Section 2.6. The washes were 2 x SSC, 0.1% (^{w}/v) SDS, (2 x 5 minutes, room temperature); 1 x SSC, 0.1% (^{w}/v) SDS, (4 x 15 minutes at 65°C); 0.1 x SSC, 0.1% (^{w}/v) SDS, (2 x 30 minutes at 65°C). The position of DNA size markers is indicated.


kb
BamH1
EcoR1
BamH1+EcoR1
11
5.2
4
2
1

Lipman (1983). A match was found to a short interspersed repeat, related to the R. family of repeats, found adjacent to some mouse structural genes (Leuders and Paterson, 1982). A visual match was also found with members of the R. family of repeated sequences (Gebhard and Zachau, 1983). The matches are shown in Figure 5.7. p27 and R. match at 191 of 205 positions, with one 6 bp gap inserted to maximize homology, (93% match) while pG3-28 matches p27 at 75 of 85 positions (88% match) with the insertion of three gaps in the pG3-28 sequence to maximize the homology. These data suggest that p27 is related to the R. family of repeats in the mouse genome, of which there are approximately 10^5 copies.

It is of interest to establish whether any enrichment for R. family repeats has occurred. To this end, the clones from Experiments One, Two and Three were plated in an ordered array, and colony screened using the BamHI fragment from the p27 insert (Figure 5.1). In parallel, approximately 2,000 randomly plated colonies were screened with the same probe. These colonies were derived by cloning incompletely Sau3A digested mouse DNA into BamHI digested pAT153 which had been treated with phosphatase to prevent self-ligation. Approximately 80% of colonies had inserts, as judged by tetracycline sensitivity. Twenty colonies picked at random had inserts ranging from 500 bp to 2 kb, with an average of 1.3 kb, approximately the same as those of the clones obtained in Experiments One and Two. The results are shown in Figure 5.8.

Figure 5.7: Homology between p27, R. elements and pG3-28.

The homology between p27, the R. element consensus (Gebhard and Zachau, 1982), and pG3-28 (Leuders and Paterson, 1982), is shown. The top row shows the sequence of the BamHI fragment of p27, and underneath, those nucleotides in the R. element consensus and pG3-28 which differ from the p27 sequence. A dot indicates a matched base, a dash indicates a gap inserted to maximize homology. Homology with the pG3-28 sequence was found by a computer search, while the homology with the R. element consensus was found by eye.

Homology between p27, R-elements and pG3-28

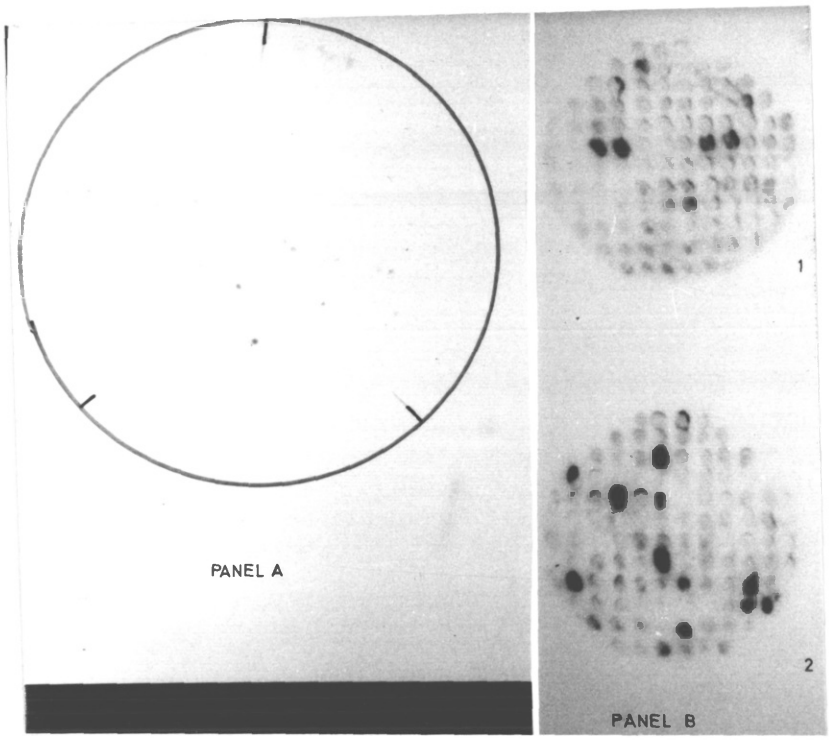
```
p27    :   GGATCCATCCCATAATCAGCCTCTAAACGCTGACACCATTGCATACACTA
R      :   ......................A.C....C.A.....T........T....
pG3-28:

p27    :   GCAAGATTTTGCTGAAAGAACCCTGATATAGCTGTCTCTTGTGAGGCTAT
R      :   A.............................................A....
pG3-28:

p27    :   GCCGGGGCCTAGCAAACACAGAAGTGGATGCTCACAGTCAGCTAGTGGAT
R      :   ...A............T............................T.....
pG3-28:                            .AT....A.T.....A.T...-.....

p27    :   ------CACAGGGCCCCCAATGGAGGAGCTAGAGAAAGTACCCAAGGAGC
R      :   GGGTCA.........................................
pG3-28:    --GGAA.............................C.........-...

p27    :   TAAAGGGGATC
R      :   ........TCT
pG3-28:    .G......GC.
```

Figure 5.8: Colony screen of immunoselected clones using a p27 probe.

Colonies were plated in an ordered array on nitrocellulose filters and grown overnight at 37°C. The filters were then processed as described in Chapter Two, and hybridized with the p27 BamHI insert fragment.

Panel A: Quasi random sample of approximately 2000 colonies.

Panel B: Clones obtained from Experiments One, Two and Three.


PANEL A
1
2
PANEL B

Panel A shows the random colony screen result. Approximately 15 positive signals can be seen, in good agreement with the expected number, given 10^5 copies per genome, and the fact that with an insert size of 1.5 kb, 10^7 colonies are required for a 99% complete library, and assuming only one repeat per insert. The average length of the repeat is approximately 400 bp (Leuders and Paterson, 1982).

Panel B shows the result of screening the ordered array. A number of positive signals can be seen. Filter 1 carries clones from Experiments One and Two, filter 2, those from Experiment Three, and some from Experiment Two. In Experiment One, 7 of 64 clones (11%) contained p27 homologous sequences. The corresponding figures for Experiments Two and Three are 3 of 53 (6%) and 11 of 98 (11%) respectively, suggesting that there has been an approximately 10 fold enrichment of the R. element containing DNA during the immunoselection procedure. Clones with homology to p27 were grown, DNA prepared from 1 ml cultures, and digested with Sau3A to determine whether the characteristic 147 bp Sau3A fragment of the p27 insert was present in the other clones. The digests were transferred to nitrocellulose and probed with nick translated p27 insert BamHI fragment. The result is shown in Figure 5.9. It can be seen that clones 27, 128, 129, 317, 371 and 388 have a fragment of approximately this size.

p27 was used to probe RNA from 3T3 and C138 cells in Northern blots. No evidence for any transcription from

Figure 5.9: Analysis of clones homologous to the p27 insert.

DNA was extracted from 1 ml of saturated bacterial culture, and resuspended in 50 µl of TE. 10 µl was used per digest, using 2 units of Sau3A. Ribonuclease was added to all digests at a concentration of 0.1 mg/ml. The digests were loaded onto a 1.5% ($^{w}/v$) agarose gel, and transferred to nitrocellulose after electrophoresis. Hybridization was performed as described in the legend to Figure 5.6.

Lane S, pTSV3 digested with Sau3A. The sizes of the visible bands in base pairs are: 1347, 1224, 945, 876, 705, 610, 396, 384, 358, 341, 317, 272, 258, 237 and 105; Lane H, pTSV3 digested with HinfI. The sizes of the visible bands in base pairs are: 1847, 1643, 1043, 766, 543, 525, 517, 396, 298, 237, 221, 220, 154 and 145.

Lanes a - u correspond to clones 21, 27, 102, 122, 123, 128, 129, 214, 215, 260, 304, 317, 319, 355, 363, 368, 371, 372, 383 and 394, respectively.

Size markers were end labelled with α-^{32}P TTP and reverse transcriptase.

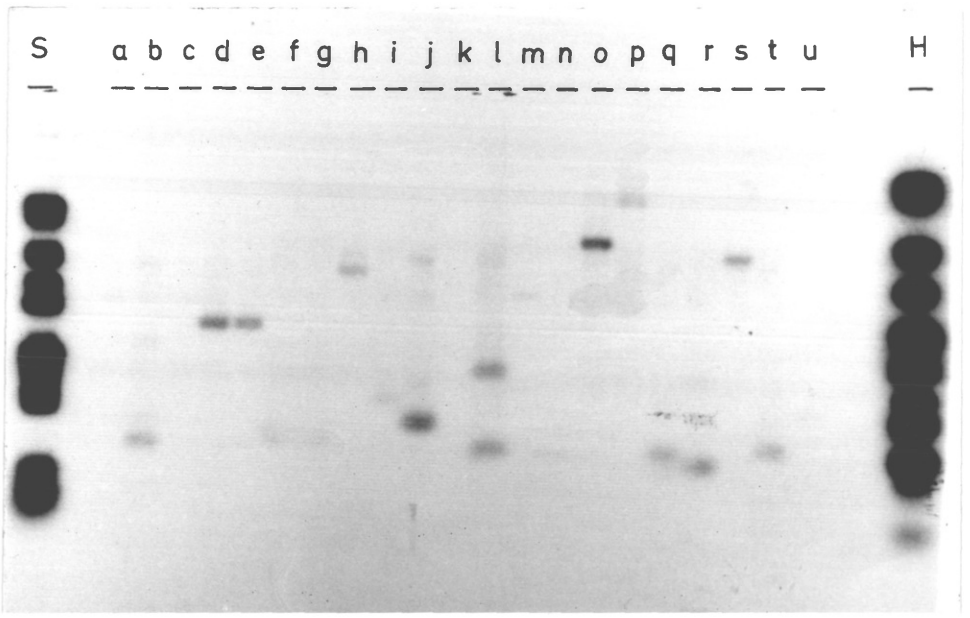
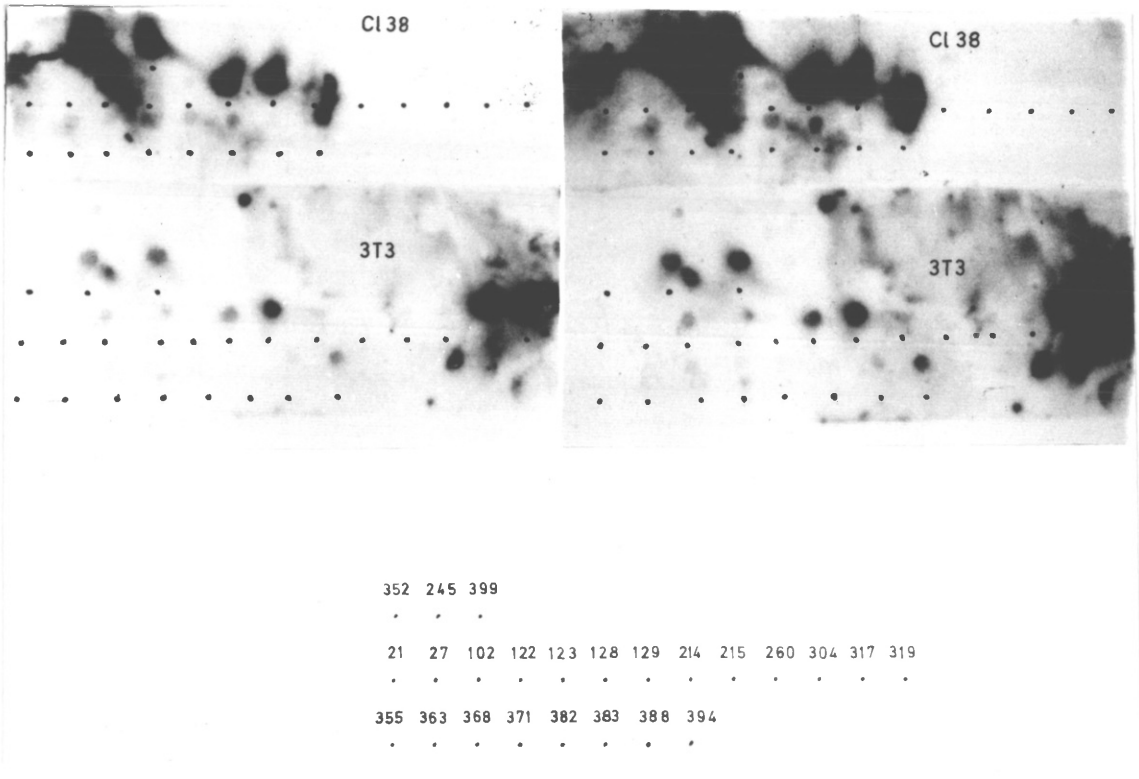

S
a b c d e f g h i j k l m n o p q r s t u
H

Figure 5.10: Dot blot analysis of expression of clones homologous to p24 and p27.

A sample of each DNA miniprep was digested with 0.1 mg/ml ribonuclease at 37°C for 30 minutes and then phenol extracted and denatured. 2 µl of each DNA was spotted onto Gene Screen Plus (NEN), and allowed to dry. Duplicate filters were hybridized with kinase labelled poly A^+ RNA from Cl 38 or 3T3 cells, in 50% (v/v) formamide, 1M NaCl, 1% (w/v) SDS, 10% (w/v) Dextran sulfate, 0.1 mg/ml denatured, fragmented salmon sperm DNA, at 42°C for 48 hours. The filters were then washed with 2 x SSC, 0.1% (w/v) SDS, (2 x 5 minutes at room temperature); 1 x SSC, 0.1% (w/v) SDS (4 x 30 minutes, 65°C); 0.1 x SSC, 0.1% (w/v) SDS, (2 x 15 minutes, 65°C), and autoradiographed. The probe concentration was 10 ng/ml. Two separate exposures are shown. The left hand panel was exposed overnight, the right hand panel for 10 days. The position of the DNA spotted onto the filters is shown by dots. The key below the filters shows the position of the clones on the grid.


Cl 38
3T3
Cl 38
3T3
352 245 399
21 27 102 122 123 128 129 214 215 260 304 317 319
355 363 368 371 382 383 388 394

p27 in poly A^+ RNA was found. The p27 homologous clones were then analyzed for expression in 3T3 and C138 cells by dot blotting. The DNA from each clone was digested with ribonuclease, denatured by treatment with alkali, neutralised and spotted onto filters, which were then probed with alkali fragmented, 5' end labelled poly A^+ RNA from 3T3 and C138 cells. The result is shown in Figure 5.10. Clones homologous to p27 form the bottom two rows of the grid. Although the background, non-specific hybridization is quite high, it can be seen that of the 21 p27 homologous clones spotted onto the filter, two, numbers 317 and 319, are expressed at higher levels in 3T3 cells than C138 cells, whereas seven, numbers 123, 128, 129, 363, 371, 382 and 383 are expressed at higher levels in C138 cells than 3T3 cells. The expression of clones 102, 368 and 394 appears unchanged, though small changes cannot be ruled out based on available data. No RNA complementary to the remaining clones has been detected.

Section 5.5: Discussion.

The experiments described in this chapter have been directed towards an analysis of the non-origin clones obtained using the immunoselection cloning strategy described in Chapter Three. An initial screen of the clones obtained in Experiments One and Two using a DNA cellulose competition assay indicated that approximately 75% of them will elute more T antigen from DNA cellulose

than an equivalent amount of pAT153. These experiments clearly need to be refined and repeated, for instance using varying amounts of each plasmid in the assay to determine the affinity of T antigen for the different plasmids, as described (Gaudray et al., 1981). With the availability of pure T antigen and the development of a radioimmunoassay, the DNA cellulose competition assay could now be repeated to obtain quantitative data of the amount of T antigen eluted using either the radioimmunoassay described in Chapter Four, or standardizing the ELISA assay using known amounts of purified T antigen. Alternatively, the ATPase activity of T antigen could be used to assay for the presence of the protein.

Three clones which were judged to contain binding sites for T antigen by the DNA cellulose competition assay were selected for further study. The work described in this chapter concerns clone p27. This clone has an insert of 237 bp, which has been completely sequenced. Initial studies using the McKay immunoprecipitation assay suggested that the insert contained a strong binding site for T antigen. Further analysis by the footprinting technique demonstrated that in the presence of T antigen a 35 bp region of the clone on both strands was protected from digestion by DNAse I, at protein - DNA ratios intermediate between those required to fill site I and site II at the SV40 origin, suggesting that the affinity

of T antigen for this region of p27 is lower than for site I but higher than for site II. Clearly, it will be necessary to determine the affinity of T antigen for this region using either the filter binding assay of Riggs *et al.*, (1970), used to determine the affinity of *lac* repressor for its operator, or the gel electrophoresis assay of Fried and Crothers (1981).

p27 hybridizes to a very large number of different sized restriction fragments in chromosomal DNA digested with a variety of enzymes, indicating that it contains a dispersed, repetitive sequence element. Comparison of the p27 sequence with the EMBL and GENBANK nucleic acid databases has identified a clone pG3-28 which is identical to p27 at 75 of 85 bases. This clone itself shares a high degree of sequence homology with the moderately repetitive family of DNA sequences known as R. elements, so a visual comparison between p27 and the R. element consensus sequence was undertaken. This showed that p27 was identical with the consensus R. element at 191 of 205 positions, allowing one 6 bp gap in the p27 sequence to maximize homology. This gives 93% identity, well within the average degree of divergence from the consensus, 12.5% (Gebhard *et al.*, 1982), and suggests that part of the p27 insert is related to the R. element family of moderately repeated DNA in the mouse genome. The region of p27 homologous to the R. element family lies wholly within the 205 bp *Bam*HI fragment of the insert. Outside this region, in the remaining 32 bp, no significant homology can be

detected by eye.

Based on the average insert size of the clones from Experiments One and Two, it can be calculated that 10^7 colonies would be required for a 99% complete library of the mouse genome. Assuming one R. element per insert, and a random distribution of R. elements, it would be expected that 1% of clones would have an R. element, given that the estimated copy number in the genome is 10^5. The clones were screened, and it was found that approximately 10% contained p27 homologous sequences, suggesting that the immunoselection protocol has resulted in an enrichment for R. elements. Digestion of the DNA extracted from these clones showed that 6 of 21 had R. homologous sequences on a 145 bp Sau3A fragment similar to p27. This Sau3A fragment would not be predicted from the R. element consensus, or any element yet sequenced (Gebhard et al., 1982; Gebhard and Zachau, 1983), and spans the junction where a 6 bp gap in p27 is inserted to maximize homology, suggesting these R. elements comprise a subset of all R. elements. Similarly none of the sequenced R. elements contains the GAGGC pentanucleotide found within the T antigen binding site, and whether T antigen will bind to this region of a "consensus" R. element remains to be determined. Thus, the possibility exists that a subset of R. elements has been immunoselected; whether this is the case awaits sequencing of the R. elements from these other clones.

What then are the consequences of T antigen binding to this region? p27 itself is not detectably transcribed in cytoplasmic poly A^+ RNA. A dot blot analysis of the p27 homologous clones indicates that 13 of the 21 are transcribed, two at higher levels in untransformed 3T3 cells than Cl 38 cells, while 8 are transcribed at higher levels in Cl 38 cells than 3T3 cells. This finding clearly needs to be verified by running Northern blots and probing with the different plasmids to define the number, and size of transcripts responsible for the observed signal in the dot blot experiment. It is tempting to speculate that within these clones, the R. element comprises the T antigen binding site, and that the binding alters the expression of the surrounding transcription unit. To further investigate this possibility it will be necessary to define the location of the high affinity binding sites for T antigen in these clones by using the McKay assay followed by footprinting of the relevant regions. If the R. element is identified as the binding site, then its position relative to the transcription unit must be determined. If it is consistent, then the repeat unit could be placed in an analogous position relative to some other gene, to investigate its potential regulatory activity. In this context, it is interesting to note that R. elements have within them the dc like sequences, found upstream of immunoglobulin genes, required for their correct expression, (Falkner and Zachau, 1984) and also have homology with the SV40 enhancer region (Leuders and

Paterson, 1982).

Another potential function for p27 might be to act as an origin of replication. Transfection of the plasmid into cos-1 cells, followed by Hirt extraction 48 hours later gave an equivocal answer, in that whereas the parent vector, pAT153 was no longer detectable in low molecular weight DNA at this time, traces of supercoiled p27 were still detectable. It is therefore possible that p27 can replicate, but very inefficiently. The question of whether the DNA seen in Southern blots is persisting in the cell, but not replicating, or is replicating inefficiently, at low copy number could be resolved by digesting the recovered DNA with MboI. This enzyme is an isoschizomer of Sau3A, but will not cleave plasmid DNA methylated at adenine residues by bacteria. On replication in a eukaryotic cell however, the plasmid becomes demethylated, and therefore susceptible to MboI digestion. Thus if the DNA had replicated on transfection into the cos-1 cells it would be cleaved by MboI, whereas if it were just persisting in the cell following transfection, it would not.

In summary therefore it has been shown that T antigen will bind with high affinity to a 35 bp region of the insert of clone p27. This region contains pentanucleotide sequence elements related to those found at the SV40 origin, arranged in a similar though not identical fashion to those found at site I. The sequence of p27 indicates

that it is a member of the R. family of moderately repetitive sequence elements in the mouse genome. R. element containing clones have been isolated at a frequency an order of magnitude higher than expected by chance, suggesting that the immunoselection cloning protocol has enriched for these sequences. It is found that a number of these R. element containing clones are either expressed at higher levels in transformed cells than normal cells or vice versa, which leads to the possibility that T antigen binding to these R elements alters the expression of the adjacent transcription unit. This latter proposal awaits further investigation, but if proven correct would suggest that R. elements might play a role in the regulation of gene expression in the cell.

CHAPTER SIX

ANALYSIS OF NON ORIGIN CLONES II: CLONES p24 AND p29

CHAPTER SIX

ANALYSIS OF NON - ORIGIN CLONES II: CLONES p24 AND p29

Section 6.1: Introduction.

The experiments described in this chapter are concerned with defining whether T antigen will bind *in vitro* to two other clones, p24 and p29. The approach is the same as that for the analysis of p27, described in the previous chapter, using the McKay immunoprecipitation assay to define those regions of the clones which contain high affinity binding sites for T antigen.

Section 6.2: Materials and methods.

All the techniques used in these experiments have been described earlier in Chapter Two, or Section 5.2.

Section 6.3: Clone p24.

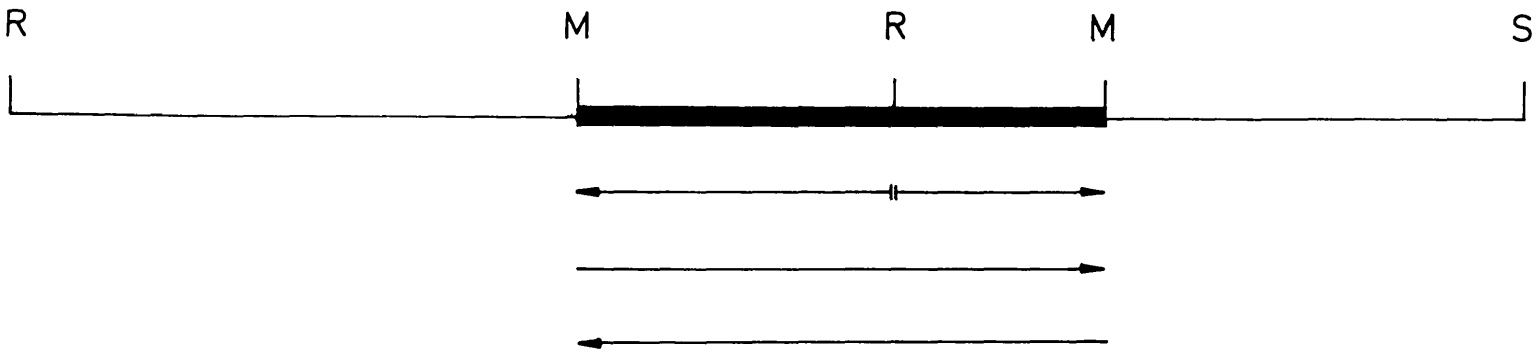
Initial restriction mapping of clone p24 indicated the presence of a single *Sau*3A fragment in addition to those derived from the vector, which contained within it an *Eco*RI site (Figure 6.1). To investigate whether the insert has a high affinity binding site for T antigen within it, the clone was digested with *Eco*RI and *Sal*I, end labelled with [γ -^{32}P]ATP and polynucleotide kinase and

Figure 6.1: Restriction map and sequencing strategy for clone p24.

The restriction map of p24 derived by mapping with BamHI, BglII, ClaI, EcoRI, HindIII, PstI, SalI, and XhoI. The region shown spans from the EcoRI to the SalI sites of pAT153, at a scale of 2cm = 100 bp.

Abbreviations for restriction enzymes are: M, Sau3A; R, EcoRI; S, SalI. Arrows show the sequencing strategy employed, using the dideoxy method. Thin line denotes vector, thick line denotes insert.

Restriction map of clone p 24



used in a McKay immunoprecipitation assay. The result is shown in Figure 6.2. It can be seen that under these conditions, T antigen exhibits specificity for the SV40 origin over pAT153, as shown by the fact that the small EcoRI - SalI fragment of pEP392, containing the T antigen binding region is precipitated preferentially compared to the larger vector fragment (Figure 6.2, pEP392, lane 3). That this precipitation is due to the SV40 sequences is demonstrated by the result obtained using pAT153 DNA alone, when neither fragment is significantly co-precipitated. In the absence of T antigen, no fragments are co-precipitated. It is clear that both insert containing fragments of p24 are preferentially co-precipitated (Figure 6.2, p24, lane 3) suggesting that there are T antigen binding sites on both sides of the EcoRI site in the inserted DNA.

The insert from p24 was then completely sequenced using the dideoxy chain termination technique by the strategy shown in Figure 6.1. Initially, p24 was digested with EcoRI and the fragments cloned into EcoRI digested M13 mp8 and the sequence of the clone determined. Unlike p27, no deletions in the vector were found, and, as expected from the restriction mapping, the insert was found to be a single Sau3A fragment of 339 bp in length. Subsequently, p24 was digested with Sau3A, fragments of this size purified from an agarose gel and cloned in BamHI digested M13 mp8 and mp9, to confirm the sequence obtained, which is shown in Figure 6.3. There are several

Figure 6.2: McKay immunoprecipitation assay on clone p24.

Plasmids pAT153, pEP392, (a clone of the HindIII C fragment of SV40 in pAT153) and p24 were digested with EcoRI and SalI, dephosphorylated with calf intestinal phosphatase, and end labelled with T4 polynucleotide kinase. 0.3 μg of each DNA was used in each assay, performed as described in Section 5.2. For each DNA: lane 1, sample of input DNA; lane 2, DNA fragments coprecipitated by CV-1 cell extract; lane 3, DNA fragments coprecipitated by cos-1 cell extract. The precipitating antibody was PAb419. The salt concentration was 75 mM.

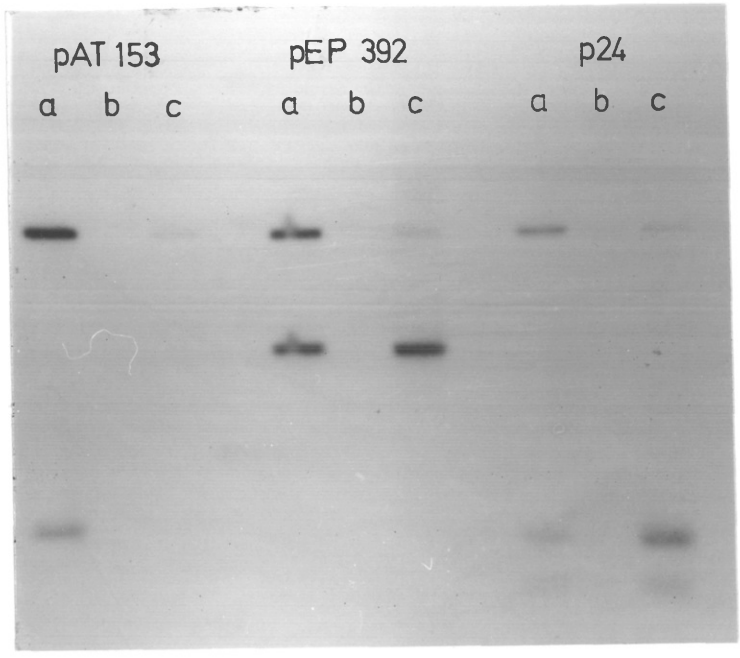

pAT 153
a b c
pEP 392
a b c
p24
a b c

Figure 6.3: DNA sequence of clone p24.

The DNA sequence of clone p24 is shown. The sequence was determined by the dideoxy chain termination technique. Dots denote 10 bp intervals. Pentanucleotides related to those thought to comprise the T antigen recognition element are indicated with an arrow. Putative TATA box elements are underlined with a single solid line, CAT boxes with two lines. A sequence related to the core nucleotide consensus for the viral enhancer is underlined with a single, broken line.

Sequence of Clone p 24

```
                                                  50
GATCACCAATGCACATTTGTAAACCAAGTCTAAGAAGTGGTATATGAGAATAGAACCAAT
CTAGTGGTTACGTGTAAACATTTGGTTCAGATTCTTCACCATATACTCTTATCTTGGTTA
                                          100
ATATTAATATCTAATCAGAGTTCTTAGAGTGCTATTGCTGTGAAGAGACACCTTGACTGT
TATAATTATAGATTAGTCTCAAGAATCTCACGATAACGACACTTCTCTGTGGAACTGACA
                                  150
GACAACTCTTATGAAAGAAAGCATTTCCCTGATAATTAAGGATGTTGAACATTTTTTCAG
CTGTTGAGAATACTTTCTTTCGTAAAGGGACTATTAATTCCTACAACTTGTAAAAAAGTC
                          200
GTGCTTCTCTGCCATTCGGTATTCCTCGGGTGAGAATTCTTTGTTCAATTCTGAGCCCCA
CACGAAGAGACGGTAAGCCATAAGGAGCCCACTCTTAAGAAACAAGTTAAGACTCGGGGT
        250                                                   300
TTTTTTAATGGGTTATTTGATTTTCTGGAGTCCACCCTCTTCAGTTCTTTATATATATTG
AAAAAATTACCCAATAAACTAAAAGACCTCAGGTGGGAGAAGTCAAGAAATATATATAAC

GATATTAGTCCCCTATCTGATTTAGGATAGGTAAAGATC
CTATAATCAGGGGATAGACTAAATCCTATCCATTTCTAG
```

features which are noteworthy. First, the clone is very A - T rich, having an overall G - C content of only 36%. Several TATA and CAT boxes, which form a component of RNA polymerase II promoters can be seen in the sequence, and are underlined. An octanucleotide GTGGTATA, which matches in 7 of 8 positions with the consensus sequence, $GTGG^{A}/T^{A}/T^{A}/TG$, for the viral enhancer element is also present. Only two pentanucleotide elements related to those found at the SV40 origin are found in the sequence, and these are arrowed. No sequences related to RNA polymerase III split promoters were identified.

When p24 is used as a probe onto HindIII digested mouse DNA, at low stringency, a large number of hybridizing fragments are seen. However, when the stringency of washing is raised, a single band of 410 bp is seen (Figure 6.4). This probably reflects the A - T rich nature of p24, permitting short stretches of imperfect homology with many genes to be formed, which are melted off by the higher stringency washes leaving only a single band of hybridization.

To determine whether p24 is transcribed, a Northern blot of cytoplasmic, poly A^{+} RNA extracted from 3T3 and Cl 38 cells was performed. The result is shown in Figure 6.5. It is clear that p24 hybridizes to three transcripts in Cl 38 RNA, not detected in 3T3 cell RNA, of 1.8, 3.3 and 4.05 kb, relative to DNA size markers. Other, higher molecular weight transcripts can also be seen at 4.9 and 5.5 kb, present in both 3T3 and Cl 38 RNA and possibly an

Figure 6.4: Sequences homologous to p24 in mouse DNA.

10 µg of high molecular weight DNA extracted from Balb/c 3T3 A31 cells was digested overnight with 30 units of HindIII, and subjected to electrophoresis through a 0.9% (^{W}/v) agarose gel. Duplicate samples were transferred to Gene Screen Plus (NEN), and hybridized as follows:

Panel A: Hybridization was in 2 x SSC, 1% (^{W}/v) SDS, 5 x Denhardt's solution, containing 0.1 mg/ml denatured, fragmented salmon sperm DNA, overnight at 65°C. The washes were 2 x SSC, 1% (^{W}/v) SDS, (2 x 5 minutes, 68°C); 1 x SSC, 0.1% (^{W}/v) SDS, (2 x 20 minutes, 68°C); 0.1 x SSC, 0.1% (^{W}/v) SDS, (4 x 30 minutes, 68°C).

Panel B: Hybridization was in 1M NaCl, 1% (^{W}/v) SDS containing 0.1 mg/ml denatured, fragmented salmon sperm DNA, overnight at 65°C. Washes were 2 x SSC, 1% (^{W}/v) SDS, (2 x 5 minutes, room temperature); 1 x SSC, 0.1% (^{W}/v) SDS, (5 x 30 minutes, 68°C).

The position of markers is indicated. Exposure was for four days.

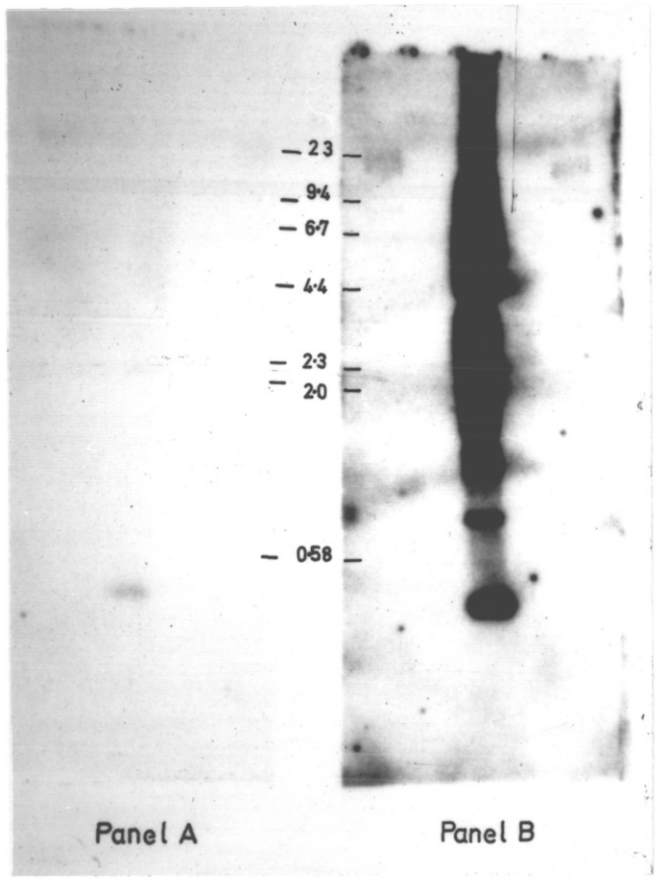
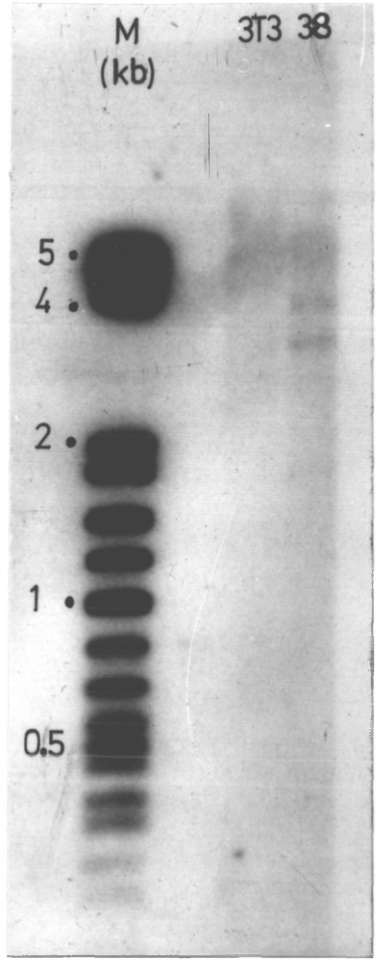

23
9·4
6·7
4·4
2·3
2·0
0·58
Panel A
Panel B

Figure 6.5: Sequences homologous to p24 in cytoplasmic, poly A^+ RNA.

1 µg of cytoplasmic polyadenylated RNA was denatured by heating to 50°C in glyoxal and formamide, as described in Chapter Two. The samples were subjected to electrophoresis through a 1% (w/v) agarose gel, and transferred to nitrocellulose. The markers were M, DNA size markers; 3T3, RNA from Balb/c 3T3 cells; 38, RNA from SV3T3 Cl 38 cells. The filter was washed in 1 x SSC, 0.1% (w/v) SDS, (2 x 5 minutes, then 4 x 30 minutes at 65°C), followed by 2 x 30 minutes in 0.1 x SSC, 0.1% (w/v) SDS at 60°C. Exposure was for 10 days.

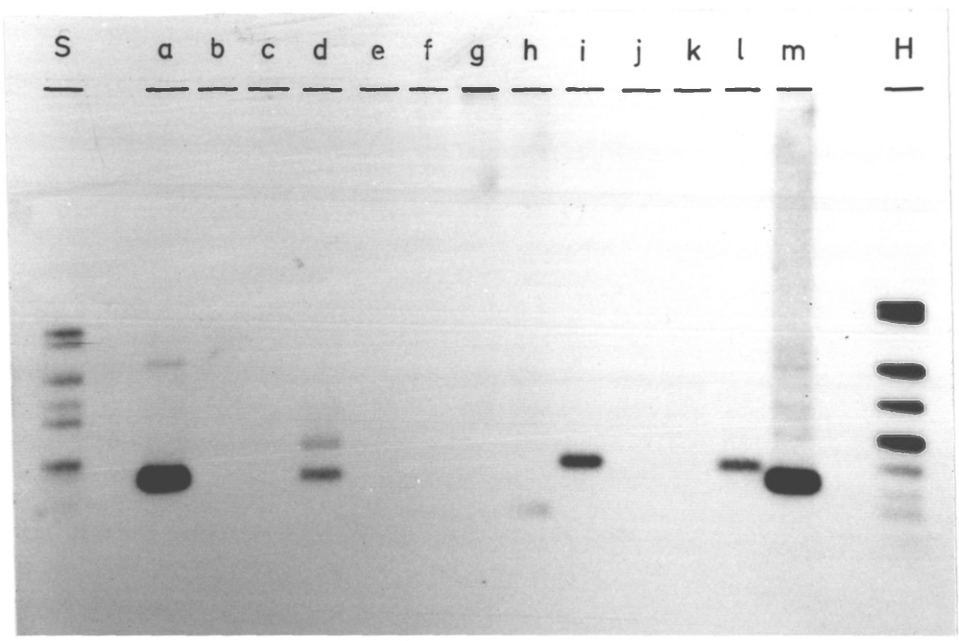

M
(kb)
3T3
38
5
4
2
1
0.5

additional transcript of 4.5 kb present in 3T3 RNA only. Thus it seems that p24 is a member of a class of genes activated upon transformation by SV40.

To determine if p24 or p24 - like clones had been isolated in subsequent immunoselection experiments, the colonies obtained were plated out in an ordered array, and hybridized with the insert from p24. Colonies giving positive signals, together with some others which gave a weaker signal were picked. DNA prepared from these transformants was digested with Sau3A, and probed with the insert from p24. The result is shown in Figure 6.6. Only four of the positive colonies from the colony screen gave a positive signal on digestion with Sau3A. Clones 317 and 352 both have inserts of approximately 415 bp and clone 399 has a single hybridizing fragment of approximately 250 bp. Clone 399 has two Sau3A fragments homologous to p24, one of which is approximately the same size as the p24 insert, 350 bp, while the other, which gives a weaker signal, is 490 bp long. Whether this fragment is a partial digestion product, or a sequence with only partial homology to p24 remains to be determined. These p24 homologous clones were spotted onto nitrocellulose and probed with labelled RNA extracted from 3T3 cells and Cl 38 cells in the manner described earlier for p27 homologous clones. The result is shown in Figure 5.10. It can be seen that of the four clones tested, two appear to be expressed at higher levels in Cl 38 than 3T3 cell cytoplasmic poly A^+ RNA (clones 399 and 245), while one,

Figure 6.6: Analysis of immunoselected clones homologous to p24.

Plasmid DNA was extracted from 1 ml of saturated bacterial culture, digested with Sau3A, subjected to electrophoresis through a 1.4% (w/v) agarose gel, transferred to nitrocellulose, and hybridized with nick translated p24 insert. Hybridization and washing conditions were as described in Panel A of Figure 6.4. S and H are Sau3A and HinfI digests of pTSV3, respectively, and give size markers as listed in the legend to Figure 5.9. Lanes a - m correspond to clones 24, 310, 383, 245, 365, 323, 318, 399, 352, 342, 316, 317 and 24, respectively.


S
a
b
c
d
e
f
g
h
i
j
k
l
m
H

clone 317 is expressed at higher levels in 3T3 cell RNA than Cl 38 RNA. No expression of clone 352 was detected.

A search of the EMBL and GENBANK databases failed to identify any sequence with 100% homology to p24, but produced a variety of alignments of parts of p24 with other genes, shown in Figure 6.7. Panel A shows the alignment between p24, sequences found in Intron A of Kallikrein gene MGK-1 (Mason et al., 1983) and sequences located in Intron D of Prolactin (Cooke and Baxter, 1982). p24 matches with MGK-1 at 33 of 37 positions, and with prolactin at 30 of 37 positions, giving 89% and 81% matches, respectively. The region of p24 which matches with these nucleotides is bases 141 to 177, inclusive. A lower level of homology was also detected with the upstream regions of immunoglobulin heavy chain genes, shown in Figure 6.7, Panel B. Nucleotides 241 - 327 of p24 match at 66 of 86 positions (77%) with these genes, with the insertion of one gap in each sequence to maximize homology. The relevant region of the immunoglobulin genes is found approximately 5 kb upstream of the 5' end of the mRNA (Kim et al., 1981).

The probability of these alignments arising by chance was calculated using the equation given in Davidson et al., (1983). For the immunoglobulin gene alignment the probability of chance occurrence is 10^{-23}, while the matches with Kallikrein gene MGK-1 and prolactin have chance probabilities of 2.8×10^{-16} and 1.1×10^{-12} respectively. The chance of the homology between all

Figure 6.7: Alignments of p24 with sequences in the EMBL and GENBANK databases.

The databases were searched using the algorithm of Wilbur and Lipmann (1983). Panel A shows the homology between prolactin, the MGK-1 gene, and p24. Panel B shows the homology between p24 and the upstream region of an immunoglobulin gene.

Asterisks (*) indicate homologous bases.

A HOMOLOGY BETWEEN p24, KALLIKREIN GENES AND PROLACTIN

```
Kallikrein  :  GCAATTCCCTGATGATTAAGGATGCTGAACATTTTAT
               *** ********* ********** ********** *
p24         :  GCATTTCCCTGATAATTAAGGATGTTGAACATTTTTT
               * ** ****** * * ************ ***** **
Prolactin   :  GTATCTCCCTGGTGACTAAGGATGTTGAGCATTTCTT
```

B Homology between the upstream region of Immunoglobulin Heavy Chain genes and clone p24

```
Ig   :  TTTTTAATTGTGTTATTTGAGGTTTTGGAAGTTAGCTTCTTGTGT
        ******** * *********  ** ****    * * ****  **
p24  :  TTTTTAATGG-GTTATTTGATTTTCTGGAGTCCACCCTCTCCAGT

Ig   :  TCTTTATATTTTTTTGAATATTAGCCCTCTATCAGATGT-GGA
        ********* * *** ******* ** ***** *** * ***
p24  :  TCTTTATATATATTGGATATTAGTCCCCTATCTGATTTAGGA
```

three genes, 27 of 37 positions, arising at random is 10^{-9}, suggesting that all the homologies observed are statistically significant.

Section 6.4: Clone p29.

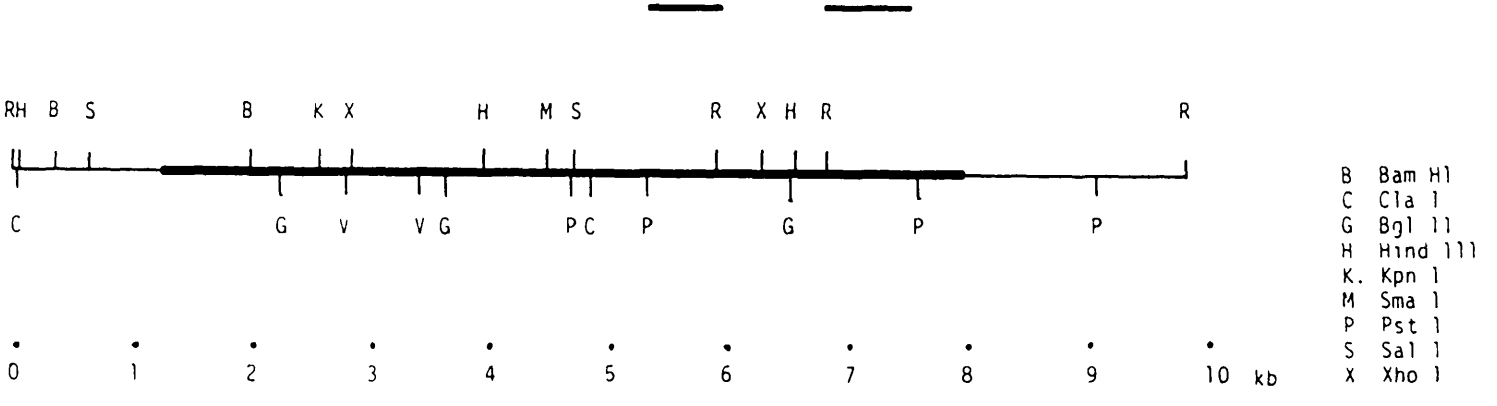
This clone, isolated in Experiment One, has a larger insert than the average size for plasmids obtained in this experiment. The clone was originally isolated as a mixture of two plasmids, one of approximately 10 kb, the other of 3.7 kb. On retransfection into bacteria, the two species were separated. The smaller plasmid had no detectable insert, and by restriction mapping, proved indistinguishable from pAT153, though it had lost the BamHI site possibly accounting for the tetracycline sensitive phenotype of the whole original colony.

The larger plasmid was prepared, and a restriction map derived by a combination of single and double digests (Figure 6.8). Several findings indicated that the clone had undergone some form of rearrangement. It proved impossible to construct a map which could reconcile all of the observed restriction sites with an insert at the BamHI site. Various digests of the clone were therefore hybridized with pAT153, and based on this, insert and vector fragments were identified. Subsequent sequence analysis of the PstI C fragment showed that after 235 bp, the sequence read into pAT153 at nucleotide 1637. At the other end of the clone, the sequence of the BamHI B

Figure 6.8: Restriction map of clone p29.

The restriction map of clone p29 is presented. Solid black bars above the map indicate the regions shown to contain high affinity binding sites for T antigen. The clone is shown in linear form, opened up at the EcoRI site of pAT153. Thin line denotes vector sequence, thick line denotes insert. Abbreviations used for restriction enzymes are: B, BamHI; G, BglII; C, ClaI; R, EcoRI; H, HindIII; K, KpnI; P, PstI; S, SalI; X, XhoI.

Restriction Map of Clone p 29



fragment has been determined for approximately 300 bp, reading towards vector sequence. No vector sequence has been encountered. Thus it cannot be stated whether any deletion of vector sequence has occurred. Precisely how clone p29 arose in its current form cannot readily be explained. The total size of the clone is 9.8 kb.

The question of whether p29 hybridizes to mouse DNA at all therefore needed to be answered. Nick translated p29 was used to probe digests of mouse DNA with EcoRI, BamHI or BglII. The result is shown in Figure 6.9. The clone hybridizes to a 3.7 kb BamHI fragment, a 3.9 kb EcoRI fragment and five BglII fragments of 5.7 kb, 4.7 kb 3.6kb, 3.2kb and 2.3 kb. These findings are clearly incompatible with the observed structure of p29. The approach with p29 has therefore been to define those regions of the clone which contain the high affinity binding sites using the McKay immunoprecipitation assay. The results are shown in Figures 6.10 and 6.11. It is clear that the 2.3 kb PstI B fragment, derived entirely from insert sequences is preferentially co-precipitated by T antigen containing extracts (Figure 6.10, lanes b and c). In a double digest with EcoRI and BglII, the A and C fragments are co-precipitated by T antigen containing extracts, suggesting that the PstI B fragment contains two separate T antigen binding sites, as shown in Figure 6.8. These regions have not yet been completely sequenced, and no sequence resembling a high affinity T antigen binding site has been found. A search of the EMBL and GENBANK

Figure 6.9: Sequences homologous to p29 in mouse DNA.

10 µg of high molecular weight DNA was digested with 30 units of BamHI, EcoRI or BglII, overnight at 37°C, and processed described in the legend to Figure 6.4, Panel A. The filter was hybridized with nick translated p29. The position of DNA size markers is indicated.

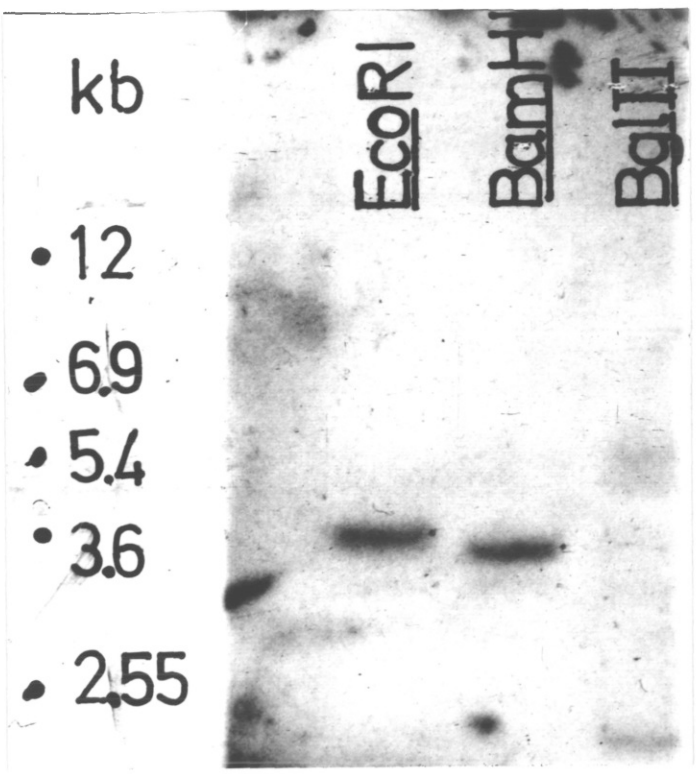

kb
EcoRI
BamHI
BglII
12
6.9
5.4
3.6
2.55

Figure 6.10: McKay immunoprecipitation assay on clone p29: I.

Clone p29 was digested with PstI and 3' end labelled using cordycepin and terminal transferase as described in Chapter Two. 0.3 µg of DNA was used per assay, performed as described in Section 5.2.

Lane a, sample of input DNA.

Lane b, fragments coprecipitated by incubation with uninfected CV-1 cell extract.

Lane c, fragments coprecipitated by incubation with SV40 infected CV-1 cell extract.

The immunoprecipitating antibody was PAb419, and the salt concentration was 150 mM.

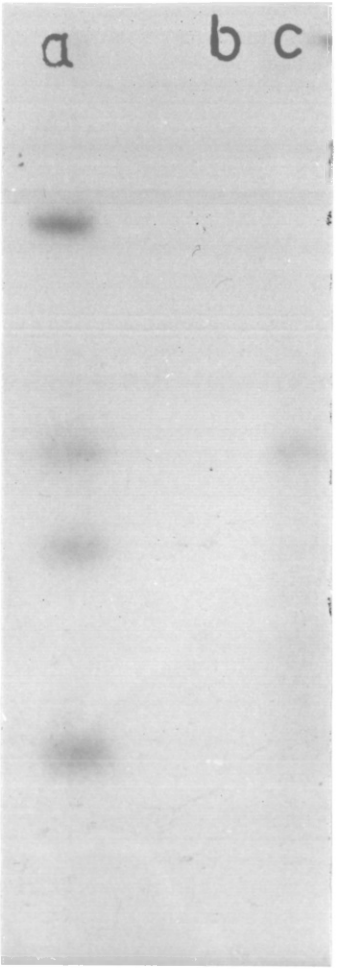
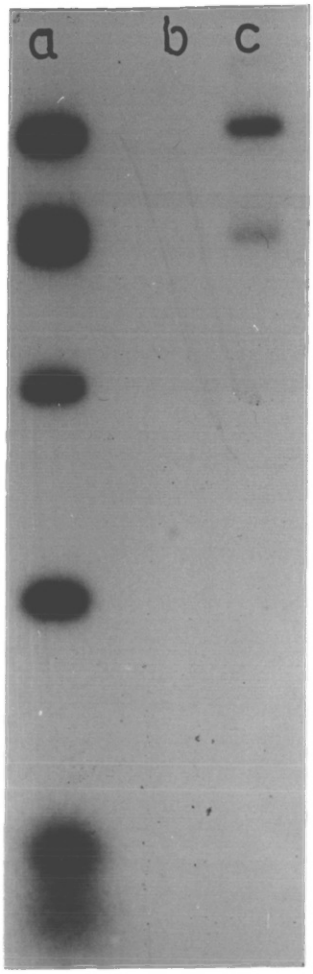

a
b
c

Figure 6.11: McKay immunoprecipitation assay on clone p29: II.

Clone p29 was digested with EcoRI and BglII, and labelled using reverse transcriptase with TTP and dGTP as the ^{32}P labelled nucleotides. The assay was performed as described in the legend to Figure 6.10.


a
b
c

databases with the available sequences has failed to reveal any significant homologies. A colony screen of all the other clones obtained in subsequent experiments showed that p29 homologous sequences had only been cloned once, in Experiment One.

Section 6.5: Discussion

Two clones p24 and p29 have been demonstrated to contain high affinity binding sites for large T antigen *in vitro*, using the McKay immunoprecipitation assay.

p29 contains a large insert of approximately 6.1 kb, and has clearly undergone some form of rearrangement during cloning. It is hard to decide exactly how p29 has been generated, since at least one of the junctions of vector and insert does not lie at the BamHI site of pAT153. Despite this however, two regions of the inserted DNA are preferentially selected in a McKay immunoprecipitation assay. Preliminary sequence data from these regions has not yet identified any arrangement of pentanucleotide elements similar to that found at site I of the SV40 origin. There are therefore several outstanding experiments which need to be performed on p29. First, the sequencing of the regions containing the putative T antigen binding site should be completed, and any potential T antigen recognition sequences identified. This region could then be analyzed in greater detail using the techniques discussed in Section 5.1. One possible

route to rapidly identifying the T antigen binding site would be to randomly fragment the PstI B fragment (which contains the T antigen binding sites), using DNAse I and to clone the fragments in M13. The pool of ligated DNA would then be used in a McKay assay to select those clones which have a T antigen binding site. These could then be sequenced, and also used as probes to select unrearranged copies of the p29 gene(s) from a genomic library for further analysis of expression and structure.

p24 has a single Sau3A fragment, 339 bp in length, inserted at the BamHI site of pAT153. As judged by the McKay immunoprecipitation assay, T antigen must interact with this clone at sites on both sides of the central EcoRI site, as both of these fragments are preferentially immunoselected compared to vector DNA. Inspection of the DNA sequence however, shows that only two pentanucleotide sequence elements related to those found at the SV40 origin of replication are present, one on each side of the EcoRI site. These alone cannot be responsible for the observed precipitation, thus other explanations must be considered. Using a filter binding assay, it has been demonstrated that SV40 T antigen has a 17 - fold higher affinity for single stranded than double stranded DNA *in vitro* (Spillman *et al.*, 1979). It is therefore conceivable, given the A - T rich nature of the clone that localized melting of the DNA strands occurs, and that what T antigen binds to are these single stranded regions, in

addition to the pentanucleotide elements.

Another possibility arises when one considers the cold sensitive mutants generated by Di Maio and Nathans (1980), discussed in the Introduction. It has been shown (Tenen et al., 1982) that mutant origins that have the site I pentanucleotides deleted will no longer bind T antigen in vitro in the McKay immunoprecipitation assay with higher affinity than other fragments. However, a mutant origin which has a number of point mutations within site I, altering all three pentanucleotides is still preferentially precipitated in a McKay assay, albeit with reduced affinity. It is therefore apparent that a T antigen binding site still exists in this mutant, suggesting that the pentanucleotides are not the sole component of whatever constitutes a high affinity T antigen binding site. It is likely that in this mutant, the altered sequence adopts a similar conformation to wild - type site I, allowing it to make at least a fraction of the hydrogen bond and electrostatic contacts that are normally made with T antigen, thus allowing the protein to bind, though with reduced affinity. It is therefore conceivable that p24 contains a site I - like structure which is recognized by T antigen. This possibility could be investigated by footprinting using purified T antigen, as was done for p27, to precisely define those regions which the protein binds to.

A third possibility is that T antigen is binding to the DNA in a complex with another protein, and that it is

the sequence specific recognition of p24 by this second protein or the complex that is responsible for the observed immunoprecipitation. A good candidate for such a protein is p53, the protein to which T antigen is found complexed in both lytically infected and transformed cells (Lane and Crawford, 1979). p53 is known to be a DNA binding protein in its own right (J. Gannon and D. P. Lane, personal communication), so this possibility is not implausible, and may be readily tested by purifying the T antigen - p53 complex and using it in footprinting studies. In addition, p53 could be purified from transformed cell lines which overproduce it, and used similarly either alone, or with T antigen.

In transformed cells, p24 hybridizes to three transcripts not detected in untransformed 3T3 cells. Genes activated in cells upon transformation by SV40 have been cloned in other studies (Schutzbank et al., 1982; Scott et al., 1983), though there is no clear proof in these cases for a direct regulation of these genes by T antigen. In view of the cloning strategy employed in these studies, it is tempting to speculate that the gene defined by the p24 clone is directly regulated by the binding of T antigen and thus is, perhaps, a gene whose aberrant expression is partly responsible for the transformation of the normal 3T3 cell.

Clearly then a high priority is to establish whether purified large T antigen binds to p24 in vitro by

footprint analysis. It will also be necessary to establish whether activation of p24 in cells transformed by SV40 and other agents is a general phenomenon. It would also be of interest to determine whether any normal tissues in the mouse express the p24 gene, or if it is expressed during embryonic development. Whether p24 activation requires functional T antigen in the transformed cell could be analyzed by studying its expression in tsA transformed cells at the permissive and non-permissive temperatures.

p24 contains no long open reading frame, and its DNA sequence, with several potential TATA and CAT boxes, and an octanucleotide homologous in 7 of 8 positions with the consensus sequence for the viral enhancer element, resembles the structure expected of an RNA polymerase II promoter region. Whether this is so could be investigated using the recently developed CAT vectors (Gorman et al., 1982), and by transfection into appropriate cell types, activation by T antigen in vivo could be assessed. If pure T antigen bound to the clone in vitro, then it would be of interest to try to reconstruct the activation using an in vitro transcription system. Precisely how T antigen might activate the p24 gene remains a matter for speculation. It is conceivable that T antigen binds to single stranded regions in the clone, stabilizing them and perhaps creating a preferential "entry site" for RNA polymerase, as suggested for the progesterone receptor (Compton et al., 1984), or it might interact with a

preformed, incomplete transcription complex, activating the gene for transcription.

At high stringency, p24 hybridizes to a single 410 bp HindIII fragment in digested chromosomal DNA. Ultimately therefore murine cDNA and genomic libraries could be screened at high stringency to isolate larger clones from the p24 region to determine the gene structure, identify any potential protein coding regions and subsequently to use this information to raise antibodies to study the protein product.

A computer search of the EMBL and GENBANK databases failed to identify any sequence with extensive homology to p24. However, short regions of homology with sequences found in the intervening sequences of the prolactin and MGK-1 genes were found. One can speculate that this region might have some regulatory function, perhaps comprising the binding site for some transcription factor or other protein potentiating gene expression. Such control sequences within introns are not without precedent, with the finding of an enhancer element within the intron of an immunoglobulin heavy chain gene (see Section 1.2). p24 is also homologous to sequences found approximately 5 kb upstream of the 5' end of the mRNA for an immunoglobulin heavy chain gene. As before, one can speculate that this sequence has some regulatory function, perhaps as the binding site for protein that alters local chromatin structure, or as an anchor point to the nuclear

matrix to permit transcription to occur.

Clearly then, further work is required on both p24 and p29, initially to extend the analysis of the interaction of T antigen with these clones, and subsequently to further study the genes from which they are derived.

CHAPTER SEVEN

GENERAL DISCUSSION

CHAPTER SEVEN

GENERAL DISCUSSION

Section 7.1: Regulation of eukaryotic gene expression.

There are many potential levels at which control of gene expression may be exercised. These have been discussed in detail in Section 1.1. Models for the control of eukaryotic gene expression generally represent a reformulation of the classical model of the bacterial operon, proposed by Jacob and Monod (1961). Thus, the model of Britten and Davidson (1969) proposes the existence of four classes of eukaryotic DNA, defined according to function. Producer genes are those encoding proteins, and may be considered directly analogous to the structural genes of bacterial operons. Adjacent to these genes are one or more receptor elements, which through interaction with activator molecules permit transcription of the producer gene to occur. A second set of genes, known as integrator genes, encode the regulatory molecules which interact with these receptor sites. Transcription of these loci is controlled by adjacent sensor elements, with the gene only being transcribed when its sensor element is activated.

The interaction of the four classes of DNA sequence therefore constitutes a system of positive control, in

which genes are not expressed unless specifically activated. Integrator genes are analogous to regulator genes in bacterial operons. Sensor elements are proposed to constitute the targets which are recognized by agents which alter the pattern of gene expression providing a means for the specific control of gene expression in response to changes in the environment of the cell. This model has subsequently been extended by Davidson and Britten (1979), to propose that the products of some integrator genes, RNA molecules, recognize particular hnRNA molecules in the nucleus and that the formation of this RNA duplex would in some manner provide a signal for RNA processing or transport.

The evidence on which such a model is based is discussed in Chapter One and has been comprehensively reviewed by Britten and Davidson (1979). Briefly, it has been found that polysomal mRNA sequence sets found in a given cell type, but absent from the polysomal mRNA of another cell type, nevertheless appear to be ubiquitously represented in their nuclear RNA. The implication of such data is that each differentiated cell nucleus contains not only all of the genes ever utilized by the organism but also transcripts of all or most of them, and that regulation of gene expression may occur by each cell type processing a qualitatively and quantitatively different subset of nuclear RNA to messenger RNA.

Comparison of the total sequences found in nuclear

RNA of two cell types has shown that while single copy structural gene sequences may be ubiquitously represented, the expression of some non - structural sequences differs between them (Britten and Davidson, 1979; Davidson and Posakony, 1982). Based on such data, it has been proposed that the moderately repeated sequences in the genome are specifically transcribed in different cell types and that these transcripts anneal within the nucleus to a particular subset of hnRNA molecules thereby selecting them for transport to the cytoplasm, or perhaps controlling their processing.

Formally, this model retains all of the essential features of the earlier one. Sensor elements would control the expression of different families of repeat sequences, which thus become the activator molecules. The receptor site thus becomes the complement of the transcribed repeated sequence. It is postulated that producer genes are constitutively transcribed. Such RNA - RNA duplexes have been observed in the electron microscope (Davidson and Posakony, 1982) and as RNAse resistant molecules (Kramerov et al., 1977; 1979). Though attractive, the evidence in favour of such a model remains circumstantial.

Many different families of moderately repeated sequences have been identified in the genome of the mouse (Fanning, 1982; 1983; Gebhard et al., 1982; Gebhard and Zachau, 1983; Heller and Arnheim, 1980; Kominami et al., 1983; Krayev et al., 1980; 1982; Leuders and Paterson,

1982; Wirth et al., 1983). The organization of the repeats and their relationship to one another has recently been reviewed (Jelinek and Schmid, 1982; Rogers, 1983; Singer, 1982). These repeats often have an A - T rich sequence at one end, and are bounded by a short duplication of target sequences flanking them. This has lead to the suggestion that they are mobile within the genome, and are generated by reverse transcription (Sharp, 1983). Their location, within and adjacent to structural gene transcription units has lead to the suggestion that they are involved in some way in the control of gene expression, or replication.

Some transcripts containing B2 elements have been found to be quantitatively and qualitatively regulated during embryonic development in the mouse (Murphy et al., 1983), and to show increased expression in a wide range of transformed mouse fibroblasts (Scott et al., 1983). In Drosophila melanogaster, a repetitive element known as suffix present at the 3' end of many embryonic RNAs has been identified (Tchurikov et al., 1982).

Both B1 and B2 elements have within them sequences homologous to the split promoter recognized by RNA polymerase III (Krayev et al., 1982) and act as templates for transcription in vitro. B1 sequences also have homology to part of the SV40 origin of replication (Krayev et al., 1980). Human Alu sequences also have B1 - like properties (Jelinek et al., 1980). This has lead to the

suggestion that Bl or Alu elements may act as cellular origins of replication. There is however no firm evidence to support this notion. A family of repeated sequences with structural homology to solitary retroviral long terminal repeats has been identified. The members of this family have within them elements implicated in RNA polymerase II transcriptional regulation such as CAT and TATA boxes, and also a polyadenylation signal AATAAA (Wirth et al., 1983). Of particular interest in view of the finding that it contains a binding site for SV40 large T antigen is the structure of an R. element (Gebhard et al., 1982; Leuders and Paterson, 1982). This repeat has been found adjacent to a number of different structural genes in the mouse. Like the other repeats discussed above it bears all the hallmarks of a "retroposon", being flanked by short direct repeats, and having a long, A - T rich region at one end. No RNA polymerase III promoter or viral origin like sequences are present, but it does have TATA and CAT boxes and a polyadenylation signal. It also has a sequence with partial homology to the dc element discovered upstream of immunoglobulin genes (Falkner and Zachau, 1984; Parslow et al., 1984), and homology with the viral enhancer region. It will therefore be of great interest to determine whether the R. elements in the apparently differentially expressed clones isolated during this study have any regulatory function.

It has also been suggested (Doolittle and Sapienza, 1980; Orgel and Crick, 1980), that these repeats have no

function except to replicate themselves; that is to say, they are "selfish" DNA. These authors do not however rule out the possibility that a subset of these repeats might, by mutation have acquired regulatory function, and then spread throughout the genome by retrotransposition events.

Despite a wealth of circumstantial evidence however, no clear function has yet been demonstrated for the moderately repetitive sequence component of the genome. The role of these repeats in controlling gene expression, if any, therefore remains enigmatic.

The results of many studies indicate that in a cell expressing, or committed to the expression of a particular gene, that gene is packed in a different, more open form in chromatin, which renders it sensitive to digestion by DNAse I (Weintraub and Groudine, 1976). This seems to be true for all genes, irrespective of the abundance class of the RNA they encode (Garel et al., 1977).

It is clear that for many genes which encode RNA classed as superabundant in a particular cell type, for instance globin, ovalbumin and alpha-fetoprotein (AFP), control of expression is applied primarily at the level of transcription (Groudine et al., 1982; McKnight and Palmiter, 1979; Latchman et al., 1984). This also seems to be true of moderately abundant, tissue specific mRNA (Derman et al., 1981). This is in contrast to findings such as those of Wold et al., (1978), suggesting that the complex class of messengers (1 - 10 copies per cell) are

ubiquitously represented in the nuclear RNA of different cell types, suggesting they are constitutively transcribed, and that regulation occurs at some post-transcriptional step, with each cell type selecting a subset of these nuclear RNA molecules to become mature mRNAs in the cytoplasm.

This apparent dichotomy may be resolved as follows. For the complex class of RNAs, the basal level of transcription resulting from a gene in the "open" conformation in chromatin, combined with efficient processing of primary transcripts would be sufficient to provide the required level of polysomal RNA. Expression of complex class genes in appropriate cell types would therefore be controlled at the level of RNA processing and/or transport to the cytoplasm. This mechanism alone however is considered insufficient to provide enough mRNA where larger amounts of protein are required. Thus it is postulated that those genes whose mRNA production may be classed as moderately abundant or superabundant are regulated in a different manner, with additional factors causing an increased rate of transcription of the gene above basal levels in the required cell type. In other cell types however, these genes would be held in a "closed" state, and not transcribed. This has been shown to be the case for globin, where transcription in mouse embryonal carcinoma cells (Jacquet et al., 1978) and 293 cells (Treisman et al., 1983) is undetectable. This is in contrast to complex class RNAs, which would be held in an

"open" conformation in all cell types. It should of course be noted that while the observed tissue specific differences in sensitivity to DNAse I do not actually prove that these genes are being differentially expressed as a consequence of their being in an altered configuration in chromatin, they nevertheless define a structural difference at the level of the chromosome whose presence is closely correlated with differential gene expression.

What then is the nature of the "open" state of the expressed genes in chromatin? It has been shown that extraction of chromatin with 0.35 M NaCl eliminates the DNAse I sensitive state. Reconstitution of depleted chromatin with this protein fraction restores the DNAse I sensitivity. Subsequent analysis has shown that addition of two nuclear proteins from this fraction, HMG 14 and HMG 17 will restore the DNAse I sensitivity of the depleted chromatin (Weisbrod et al., 1980). Sensitivity to DNAse I was only restored to those genes which already displayed it before extraction. Other studies (Gottesfeld et al., 1974; Baer and Rhodes, 1983) have shown that a fraction of nucleosomes enriched for transcribed sequences is deficient in H1, H2a and H2b. Reconstitution experiments have demonstrated that H2a and H2b will compete with HMG 14 and 17 for binding to depleted nucleosome cores (Weisbrod et al., 1980). The precise mode of action of HMG 14 and 17 remains unclear, but it is apparent that the

presence of these proteins is correlated with an open state for both abundant and complex class mRNA encoding genes. It may be postulated therefore that they are, perhaps together with other proteins, involved in establishing a gene in an open conformation, thereby permitting a basal level of transcription to occur.

For genes encoding superabundant RNA species in the cell stable so - called transcription complexes have been implicated in controlling gene expression. The regulation of Xenopus 5S RNA gene transcription by RNA polymerase III has been extensively studied (recently reviewed by Brown, 1984). Xenopus has two types of 5S RNA gene, oocyte and somatic, which are differentially regulated during development. Somatic 5S RNA genes, comprising only 2% of the animal's complement of 5S genes encode more than 95% of the 5S RNA synthesized in somatic cells. In oocytes, mainly oocyte - type 5S RNA is synthesized. This control of expression is known to be applied at the level of transcription (Bogenhagen et al., 1982). Transcription of cloned somatic and oocyte 5S DNA templates in vitro with somatic cell extracts fails to reproduce this regulation. However, transcription of native chromatin, extracted from somatic cells, produces only somatic - type 5S RNA (Schlissel and Brown, 1984). If the structure of the chromatin is perturbed by washing with high salt concentrations, or ion exchange chromatography, then the oocyte type 5S genes can also be expressed, but only if a nuclear protein fraction containing RNA polymerase III

transcription factors is added. Thus, transcription of the gene is dependent on the presence of an active, stable transcription complex, which potentiates correct transcription of the 5S RNA genes. It has been shown that neither RNA polymerase II nor III will accurately initiate transcription on a naked DNA template in the absence of transcription factors (Cozzarelli et al, 1983; Lewis and Burgess, 1980). It has also been found that if histones are added to the perturbed chromatin before the nuclear extract, then no activation of oocyte - type 5S genes occurs suggesting that histones and transcription factors are competing for binding to the control region of the gene (Schlissel and Brown, 1984). Nucleosomes may thus be thought of as the general repressor, in that unless a transcription complex has formed on a gene, it is assembled into nucleosomes and will be in a closed or inactive state. It is found that treatments which remove histone H1 will permit reactivation of previously silent genes by permitting the formation of transcription complexes (Schlissel and Brown, 1984). This correlates well with the finding, discussed earlier, that actively transcribed nucleosomes are deficient in H1. Thus, it seems that maintenance of the repressed state requires the presence of histone H1. Such stable transcription complexes have also been described in vitro for genes transcribed by RNA polymerase I (Cizewski and Sollner-Webb, 1983), and RNA polymerase II (Davison et al., 1983).

Recently, it has been shown that it is possible to reconstitute a gene into chromatin in the open state (Emerson and Felsenfeld, 1984). A ß-globin gene was combined with histones and extracts from the nuclei of chick erythrocytes. The resulting complex displayed a DNAse I hypersensitive site at the same region as that found *in vivo*. Nuclear extracts from tissues not expressing ß-globin failed to reconstitute this hypersensitive site. It was found that if histones were added prior to nuclear extract no hypersensitive site was formed. No attempt was made to transcribe this reconstituted gene *in vitro* with RNA polymerase II, so whether the proteins binding to this region formed an active transcription complex awaits further experimentation.

Therefore, for genes encoding abundant RNA molecules, the active state is characterized by the presence of an active transcription complex, manifested in physical terms by an increased sensitivity to DNAse I as the gene is held in an open configuration for transcription. The presence of a transcription complex need not mean that transcription is actually occurring, but may signify that a gene has the potential to be expressed in that cell type. When the gene is due to be expressed, the transcription complex would be completed, either by the binding of an additional factor which allows RNA polymerase to recognize the active complex and initiate transcription, or alternatively, some modification of RNA

polymerase could occur to permit it to initiate transcription under the influence of the preformed complex.

There is some evidence that this might be the case. Globin genes in hematopoetic cells not yet induced to undergo erythroid maturation are in the active conformation, yet high levels of RNA are made only when the cells are induced to differentiate (Miller et al., 1978). Another example is provided by the heat shock genes of Drosophila. Wu (1984) has shown that there are two protein binding sites upstream of the hsp 70 gene. In non-heat shocked cells, only one of these sites is occupied and the gene is not expressed, but is sensitive to DNAse I. Upon heat shock, protein occupies the second site, and the gene is expressed. This data has been discussed in greater detail in Section 1.3. Thus, occupation of the first site can be regarded as establishing the potential of the gene to be transcribed, that is, formation of an incomplete transcription complex, while the binding of the second protein(s) permits the gene to be transcribed.

Some of the protein factors that participate in the formation of a transcription complex would be expected to be sequence specific DNA binding proteins. Such factors which bind to the TATA box region of promoters have been identified (Davison et al., 1983; Parker and Topol, 1984) and a protein TFIII A, which binds to the internal control

sequence of 5S RNA genes, and is required for transcription to occur has been extensively studied (Smith and Brown, 1984, and references therein). Given the inability of RNA polymerase II to initiate transcription correctly in the absence of additional factors to confer specificity, it seems probable that some form of transcription complex will also be present at the promoters of those genes encoding the complex class of mRNAs.

Transcription factors would also enable RNA polymerase to discriminate between promoters. Such a factor, Spl, has been isolated (Dynan and Tjian, 1983a) which permits RNA polymerase to preferentially utilize the SV40 early promoter. It is interesting to speculate that there will be a class of cellular promoters whose activity is dependent upon Spl. How might these transcription factors act? It is possible that such promoter - specific factors such as Spl or HSTF recognize a sequence element within the promoter, and interact by protein - protein contact with either a general promoter - recognition protein, or RNA polymerase, or both. Efficient binding of the polymerase at the promoter site would then require interaction with these protein elements, either simultaneously, or sequentially. Both Spl and HSTF have been shown to recognize the SV40 early promoter and heat shock gene regulatory regions in a sequence specific manner (Parker and Topol, 1984b; Dynan and Tjian, 1983b).

Similarly, the so - called enhancer elements,

discussed in Section 1.2, could be considered as the binding site for a protein which then creates a preferential entry site for RNA polymerase, or some other component of the transcription machinery. Control of the expression of this protein(s) or the ability of RNA polymerase to interact with it would allow an enhancer to be cell type specific in its action.

A potential additional level of control is provided if protein - protein contacts are invoked as the primary mechanism of promoter location by RNA polymerase, in that the interaction could be regulated by covalent modification of RNA polymerase, the transcription complex, or both. It is of interest to note that short (<30 bp) common sequence elements are found upstream of many coordinately regulated genes (Davidson et al., 1983). These sequences may constitute the binding site for regulatory proteins to permit selective activation or repression of genes. They would thus constitute the receptor sites proposed in the original model of Britten and Davidson (1969), while the transcription factors would be the activator molecules, the products of the integrator genes. The nature of the sensor elements controlling the expression of these proteins (or RNA molecules), and the factors binding to them remains to be determined.

While the system of gene control outlined here is primarily one of positive control, it should be noted that examples of negative control are also known, such as the

repression of the silent mating type cassettes in S. cerevisiae by the SIR genes. (Abrahams et al., 1984).

It has been shown that some viral gene products can activate cellular genes, and also functionally circumvent the requirement for an enhancer to potentiate the expression of some genes upon transfection into cells (Green et al., 1983; Treisman et al., 1984). Thus the presence of the adenovirus Ela gene product makes it unnecessary for the SV40 enhancer to be present for ß-globin gene expression to occur in HeLa cells. Both Ela and SV40 T antigen have been shown to activate cellular genes (Nevins, 1982; Schutzbank et al., 1982; Scott et al., 1983), though direct regulation by these proteins, rather than through some secondary agent remains to be demonstrated. Ela is not thought to be a DNA binding protein (Tooze, 1981), so it is possible that it acts through interaction with some other protein to form a complete transcription complex on these genes, thereby permitting RNA polymerase to transcribe the gene. It is noteworthy that Ela only activates the ß-globin gene transfected into the cells, and not the endogenous chromosomal gene in the host cell. An explanation for this, consistent with the model outlined above, is that since the ß-globin gene in the chromosome is not destined to be expressed in this cell type, it is in the closed conformation, lacking any bound transcription factors. The newly transfected, naked DNA can bind the required sequence specific factors as it becomes assembled into

chromatin, and thus respond to Ela, whereas the chromosomal gene cannot. Consistent with this view is the finding that the chromosomal globin gene is insensitive to DNAse I digestion, and so is in the closed, inactive conformation (Green et al., 1983).

SV40 large T antigen might act to alter cellular gene expression, either by binding to another protein, or recognizing a particular DNA sequence and thereby completing a transcription complex. It is of course possible that these viral gene products could act by binding to sensor elements, thereby triggering the synthesis of one or more activator elements. The further analysis of the clones obtained in this study will therefore be of great interest. The fact that these viral gene products can activate cellular genes suggests that in their mode of action they are mimicing some cellular protein, whether it be by binding to a complex to activate it, or by binding in a sequence specific manner to a regulatory element. If the direct binding of T antigen to a particular cellular DNA sequence can be proven to have a biological effect, then it will clearly be of great interest to search for cellular proteins which are capable of binding to the same region, the gene set that they control, and the role, if any, of those genes in establishing the transformed state.

Current evidence therefore suggests that for the majority of cellular genes, which encode the so - called

complex class of RNAs, the prime level of control is post transcriptional, at the level of RNA processing or transport. For a subset of genes, expressed at higher levels in some cell types, the primary point of control appears to be at the level of transcription. Active genes of both types are found in an open configuration in chromatin, reflecting either active transcription, or the potential of the gene to be expressed. For genes encoding abundant RNA species, it has been demonstrated that the formation of a stable transcription complex, thought to mediate binding and promoter recognition by RNA polymerase, is associated with the active state of the gene. Genes of this class are held in a closed, non-transcribed conformation lacking a transcription complex in tissues where they are not expressed. In the closed state the gene is considered to be assembled into nucleosomes. The maintenance of the closed state appears to require the continued presence of histone H1. What factors control the change from the inactive to the active state is unknown.

Section 7.2: Eukaryotic DNA binding proteins.

As discussed in Chapter One, many different eukaryotic DNA binding proteins, or protein fractions have been identified, but none has been as well characterized as prokaryotic DNA binding proteins such as the *lac* and λ

repressors. In this section, the problems facing a eukaryotic regulatory protein will be considered.

The problem facing any sequence specific regulatory protein, prokaryotic or eukaryotic, is to locate its specific recognition sequence, typically 20 - 30 bp in length in the genome, which can range in size from 10^7 bp (*E. coli*) to 3×10^9 bp (mammals) or larger. Thus any single sequence has to be preferentially located in the presence of up to 3×10^9 (or more), overlapping non specific binding sites.

Let us first consider the situation as it applies to the *lac* repressor. Given that there are approximately 10 molecules of *lac* repressor per cell (Miller and Reznikoff, 1978) and 10^7 potential overlapping binding sites, the fractional saturation of any one site at a given moment will be 10^{-6}, assuming most *lac* repressor molecules are DNA bound, and binding is random. Thus the affinity for its specific sequence, the *lac* operator, must be at least 10^6 fold greater than for bulk DNA for there to be a significant probability of it being occupied. Studies of the relative affinities of *lac* repressor for bulk DNA and operator DNA *in vitro* have indicated that the affinity of repressor for operator is 10^8 fold greater than for bulk DNA, satisfying this requirement (Lin and Riggs, 1975). Assuming that the DNA in an *E. coli* cell completely fills it, then the concentration of DNA will be of the order of 10 mg/ml. Under these conditions, more than 98% of the repressor will be bound to DNA at any one time. It has

been shown (Barkley and Bourgeois, 1978) that addition of inducer does not alter the affinity of *lac* repressor for non-specific DNA but reduces the affinity for operator DNA by a factor of 10^3. Thus inducers shift the distribution of repressor in favor of non-operator DNA, but do not remove repressor from DNA.

A eukaryotic nucleus contains 10^2 - 10^3 fold as much DNA as is found in the nucleoid body of *E. coli*, and therefore correspondingly more non-specific binding sites to hamper a DNA binding protein in the search for its operator sequence (to use the bacterial analogy). Lin and Riggs (1975) have considered the hypothetical problem of how well the *lac* repressor would regulate the *lac* operon within a eukaryotic nucleus. Their conclusion, assuming that the number of repressor molecules remains the same as in *E. coli* and that all of the DNA would be available for binding is that the *lac* operon would be constitutively expressed. Thus, one must consider how a sequence specific DNA binding protein can locate its target sequence more efficiently within the eukaryotic nucleus.

One possible route to achieving this would be to modify the protein such that its affinity for its target sequence was increased while its affinity for competing DNA was unaltered, or reduced. Assuming that the same site on the protein is involved in binding to both specific and non-specific DNA, current evidence suggests that it may not be possible to increase specific binding

without also increasing non-specific binding (Lin and Riggs, 1975). In this context, it is noteworthy that the lac repressor mutant i^{x86}, which has 100 fold higher affinity both for lac operator and non-specific DNA than wild type repressor binds non-specific DNA so tightly that its ability to locate the operator is impaired, and the lac operon is expressed constitutively. Another possibility would be to mask off part of the genome, to effectively reduce the amount of competing DNA by assembling it into nucleosomes in a "closed" conformation, that is unavailable to DNA binding proteins. It is known that genes that are being expressed, or have the potential to be expressed (Weintraub and Groudine, 1976; Garel et al, 1979; Wu, 1980) are in an "open" conformation, and that sequence specific binding by proteins occurs in the vicinity of these genes (Wu, 1984; see Section 1.3). A third possibility is to increase the number of molecules of the regulatory protein in the nucleus. It has been calculated that the total DNA concentration is roughly the same in the nucleus of a mammalian cell and the nucleoid body of a bacterial cell. Thus, if an equivalent concentration of regulatory protein in the two cells is present, then they should both locate their specific sequence efficiently. CAP and T antigen both bind to their high affinity recognition sites with approximately equal affinity, approximately 5×10^{-10} M (Fried and Crothers, 1984; Jones and Tjian, 1984). There are approximately 10^3 CAP molecules in the E. coli cell,

giving a concentration of 1.6 μM. In a Cl 38 cell, there are approximately 10^5 molecules of T antigen (my unpublished observations), giving a concentration of 2.3 μM, comparable to that of CAP in E. coli. Thus T antigen should be able to locate its "operator", the SV40 origin, in the cell nucleus. No direct measurement of the affinity of T antigen for non-origin DNA has been made, so an estimate of the fraction of T antigen bound to DNA at any one time cannot be made. Similarly, no studies such as those described for CAP and lac repressor (see Section 1.5), analyzing the method of search used by the proteins to facilitate location of their specific sequences, have been undertaken for T antigen, or any other eukaryotic regulatory protein. It seems reasonable, however, to assume that a facilitated diffusion and intersegment transfer mechanism similar to that proposed to operate in E. coli will occur.

Section 7.3: Immunoselection cloning.

From the foregoing discussion, it is clear that there exists good evidence to suggest that the expression of at least a subset of eukaryotic genes is controlled by DNA binding proteins. However, only a few such proteins have been analyzed or purified to any great extent. One of the best characterized proteins with gene regulatory activity is the large T antigen encoded by Simian Virus

40. As described in Sections 1.8 through 1.11, T antigen acts as both a negative and a positive regulator of SV40 gene expression, and also alters the expression of some cellular genes transcribed by RNA polymerase I and II. T antigen also affects DNA replication in the host cell. The interaction of T antigen with SV40 control sequences has been extensively studied (see Section 1.9), but little is known of the cellular sequences with which T antigen interacts. Clearly, it is of interest to study these sequences. The problem however is to devise an experimental strategy to isolate such segments of DNA to which T antigen is bound *in vivo* from the rest of the genome. As discussed in Section 3.1, a variety of approaches are feasible, but the one ultimately adopted was that of immunoselection cloning, whereby chromatin is prepared from SV40 transformed cells, fragmented with nucleases, and then those nucleoprotein complexes bearing T antigen are separated from the rest by immunoprecipitating with antibodies raised against T antigen. The selected DNA is then purified from the immunoprecipitate and cloned in a bacterial vector for further analysis.

The clear prediction of such a protocol is that one ought to be able to immunoselect the SV40 origin from the transformed cell, since it is the one sequence known to contain a T antigen binding site. Preliminary analysis of the immunoselected DNA by dot - blotting indicated that

the SV40 origin sequences had been immunoprecipitated by an anti - T monoclonal antibody, PAb419, but not by a control antibody, PAb1005. In three subsequent experiments following the procedure outlined above, clones containing DNA fragments encompassing the SV40 origin of replication were obtained. Based on the number of colonies required for a complete library of the SV3T3 Cl 38 genome and the number of colonies screened, an enrichment by a factor of $>10^4$ for the origin has apparently occurred. No clones with detectable inserts were obtained when control antibodies were used, suggesting that the clones derived from the DNA co-precipitated by the anti - T antigen monoclonal antibody were obtained as a result of having T antigen bound to them in chromatin, rather than due to non-specific DNA co-precipitation.

Two of the integrated copies of SV40 in the Cl 38 genome are "marked", that is to say, are uniquely identifiable as originating from Cl 38 cells. All of the clones thus far obtained which are homologous to SV40 constitute either full length viral DNA, or the Sau3A fragment containing the viral origin of replication.

An alternative explanation for the cloning of the SV40 origin must therefore be considered namely that of adventitious contamination by SV40 DNA or plasmids. Since no recombinant plasmid comprising the Sau3A A fragment of SV40, inserted at the BamHI site of pAT153 has ever been constructed in this laboratory prior to this study, the

possibility of contamination by an already existing plasmid can be eliminated. Another explanation, given its routine use in the laboratory, is contamination by SV40 DNA. A possible route by which the clones could be derived is DNA contamination during the restriction endonuclease digestion step, followed by binding of T antigen to the naked origin of replication. This would then be co-precipitated and cloned. To try to circumvent this problem, the immunoprecipitation was performed in the presence of 0.5 M NaCl, conditions which greatly reduce the binding of T antigen to naked SV40 DNA *in vitro* (Dorn *et al.*, 1982; Scheller *et al.*, 1982). This possibility cannot formally be eliminated based on the available data.

An experiment which would answer this latter objection would be to repeat the immunoselection cloning experiment using a starting cell line with a uniquely marked origin of replication. Such a cell line is cos-1 (Gluzman, 1981). These cells contain only a single integrated copy of the viral early region and replication origin, with the latter containing a defined 6 bp deletion at the *Bgl*I site. Thus only mutant origins could be immunoselected, and any origin homologous clones would be sequenced to confirm that they did indeed originate from cos-1 cells. Studies have shown that T antigen will bind to clones of the cos-1 cell SV40 origin of replication, at least *in vitro* (Lewton *et al.*, 1984).

A total of 210 non-origin clones were obtained from

the three immunoselection experiments performed. An initial screen of 117 of these by a DNA cellulose competition assay indicated that approximately 75% would displace more T antigen from DNA cellulose than pAT153 alone, suggesting that the DNA inserted in them contained sequences to which T antigen would bind *in vitro*. Three of these clones were selected for further analysis using the McKay immunoprecipitation assay, and footprinting with immunoaffinity purified T antigen.

Clone p29 has an insert of approximately 6.1 kb, but has undergone some form of rearrangement during cloning, as shown both by sequencing and restriction mapping of the clone, which indicated that the inserted DNA is not at the *Bam*HI site of pAT153, and a comparison of the restriction map with the bands seen on a whole genome blot when p29 is used as a probe. The McKay assay suggests that this clone contains two high affinity binding sites for T antigen. Preliminary sequence data from these regions has not yet identified any sequence with an arrangement of pentanucleotides similar to that found at the SV40 origin of replication. Whether these two regions are found in the same arrangement relative to each other in the genome, or have been placed close together during the rearrangement that generated p29 remains to be determined. Clearly, it is important to isolate a p29 homologous clone from a genomic library to resolve this problem, and also to determine the location and nature of the T binding sites within p29.

Clone p24 contains a 339 bp insert, a single Sau3A fragment inserted at the BamHI site of pAT153, which as judged by the McKay immunoprecipitation assay contains a binding site for T antigen. The precise residues involved in the binding to T antigen remain to be determined by footprinting. Since p24 contains only two pentanucleotide recognition elements several possibilities exist to explain the observed binding. The clone might contain a sequence, which in terms of hydrogen bonds and electrostatic contacts resembles site I, and is therefore recognized with high affinity by T antigen, as discussed in Section 6.5. Another possibility is that T antigen binds to this clone in a complex with another protein, an obvious candidate being p53. It is known that the T antigen - p53 complex will bind to the SV40 origin *in vitro* (Reich and Levine, 1982; Burger and Fanning, 1983), and p53 is itself a DNA binding protein (J. Gannon and D. P. Lane, personal communication). Another possibility given the A - T rich nature of the clone, is that localized melting of the strands occurs, and that T antigen binds to these single stranded regions. It has been found that T antigen has a 17-fold higher affinity for single stranded DNA than double stranded DNA (Spillman *et al.*, 1979). Clearly, footprinting studies and DMS protection analysis will be necessary to define which regions are involved in interacting with T antigen, or the T - p53 complex, using purified T antigen, or purified

complex.

The sequence of p24 has several noteworthy features. A number of CAT box and TATA box elements similar to those found as components of RNA polymerase II promoters can be found on both strands of p24. A region with homology to the core nucleotides found in the viral enhancer elements is also present. It is therefore possible that p24 is a promoter. Whether this is so could be tested using the recently developed CAT vectors (Gorman et al., 1982). The effect, if any, of T antigen could be assessed by transfection into appropriate cell types. In Northern blots, hybridization of p24 against poly A^+ cytoplasmic RNA isolated from Balb/c 3T3 cells and Cl 38 cells shows that p24 hybridizes to a variety of transcripts in Cl 38 cells not detectable in 3T3 cells. The sizes of these are 1.8 kb, 3.3 kb and 4.05 kb. Which strand of p24 is being transcribed could be determined by performing an S1 mapping experiment (Berk and Sharp, 1978), which would also indicate whether transcription was being initiated downstream of any of the potential TATA boxes.

At high stringency, p24 hybridizes to a single band on a whole genome blot. A search of the EMBL and GENBANK databases failed to identify any sequence with homology to the full length of p24, but showed that portions of p24 were homologous to sequences found in the introns of the prolactin and MGK-1 genes over a region of 37 bp, and to sequences found approximately 5 kb upstream of an immunoglobulin heavy chain gene over an 86 bp region.

Whether these sequences have any regulatory significance remains to be determined, but one can speculate that they might constitute the binding site for some transcription factor, or protein regulating local chromatin structure.

At lower stringency, p24 hybridizes to many fragments in digested chromosomal DNA. Whether this is due to short stretches of homology, perhaps some common regulatory sequence found adjacent to many genes, or merely reflects the A - T rich nature of the clone is unclear. At high stringency, a screen of all the clones obtained by the immunoselection showed that p24 homologous sequences had been cloned once in Experiment One, p24 itself, once in Experiment Two and thrice in Experiment Three. Of these additional clones, only one, clone 245, has a Sau3A fragment homologous to p24 which is of similar size to that found in p24. It is therefore possible that this is a larger clone from the p24 region, and as such is worthy of further analysis to determine the structure of the gene(s) encoding the p24 homologous transcripts. A dot blot analysis shows that this clone is, like p24, expressed at higher levels in Cl 38 than 3T3 cells. The three remaining clones all have different sized p24 homologous Sau3A fragments. Of these clones, one is expressed at higher levels in transformed than normal cells, one is expressed at higher levels in normal than transformed cells, and the third is not detectably transcribed. It is tempting to speculate that their

homology with p24 is through some common regulatory element, though proof must await characterization and sequencing of the inserts in these clones, and clone 245 and the elucidation of the relationship between these clones and p24. It is also imperative that these clones be used to probe Northern blots of transformed and normal cell RNA to verify that the data from the dot blot analysis is correct, and also to determine whether the transcripts bear any relationship to those homologous to p24, and the generality of the activation of the p24 gene. An analysis of the binding of T antigen to these clones is also required.

If clones p24 and 245 are indeed derived from the same gene, then it will provide a very strong validation of the immunoselection technique, since the same, low copy number gene will have been isolated from the Cl 38 genome in separate experiments, an event most unlikely to have occurred by chance.

Clone p27 has a 237 bp insert, and contains a high affinity binding site for T antigen, as judged by the McKay immunoprecipitation assay. Footprinting analysis using purified T antigen showed that a 35 bp region of the inserted DNA is protected from DNAse I digestion. This region contains pentanucleotide sequence elements related to those found at the SV40 origin, arranged in a similar, though not identical fashion to those found at site I.

The sequence of p27 has been compared with the nucleotide sequence databases, and has shown that it is a

member of the R. family of moderately repetitive sequence elements in the mouse genome. A screen of all the other clones obtained showed that R. element containing clones have been isolated at a frequency an order of magnitude higher than that expected by chance, suggesting that the immunoselection cloning protocol has enriched for these sequences. A similar screen using a clone of the B2 repeat sequence (Krayev et al., 1982) showed that none of the clones obtained contained a B2 repeat. Since there are approximately 100,000 B2 repeats in the genome (Krayev et al., 1982), one might expect, assuming 10^7 colonies are required for a complete library, that approximately 1 in 100 colonies, at random would have a B2. Since about 200 colonies were screened it is clear that there has been no great enrichment for B2 repeats in the immunoselected DNA. This also suggests that clones such as those isolated by Scott et al., (1983) containing B2 repeated sequence elements, and corresponding to genes expressed at higher levels in Cl 38 cells than 3T3 cells may not be directly regulated by T antigen binding to the B2 repeat, but are activated by some other, perhaps T antigen controlled, mechanism during transformation; it cannot of course be formally ruled out that T antigen is lost from B2 containing nucleoprotein complexes during the immunoselection protocol, because, for instance its binding is not resistant to the high salt concentration used during the immunoprecipitation step.

Dot blot analysis of the p27 clones has indicated that 13 of 21 are detectably transcribed, two at higher levels in untransformed 3T3 than Cl 38 cells, while 8 are expressed at higher levels in Cl 38 than 3T3 cells. The possiblity that the other clones, both p24 and p27 homologous, which show no hybridization signal are transcribed cannot be eliminated, since if the messenger RNA they encode is present at low levels (<10 copies per cell) then it is unlikely that it would be detected by a dot blot assay. Clearly these clones need to be analyzed by Northern blotting. It is tempting to speculate that within these clones, the R. element comprises the T antigen binding site, and that the binding alters the expression of the surrounding transcription unit. This proposal awaits further study, by the kinds of experiments discussed in Section 5.5. If proven correct however, it would suggest that R. elements might play some role in the regulation of gene expression in the cell.

In summary therefore, the problem of selective isolation of those parts of the mouse genome to which SV40 large T antigen is bound in vivo has been approached by fragmentation of chromatin with restriction endonucleases followed by immunoprecipitation using antibodies raised against T antigen. The DNA fragments co-precipitated with the T antigen are then extracted from the immunoprecipitate and cloned in a bacterial vector for further analysis.

The effectiveness of the procedure was initially

assessed by investigating whether the SV40 origin of replication had been specifically immunoselected. It was found that origin sequences were only selected when anti - T antigen antibodies were used for immunoprecipitation. No origin sequences were detected when control antibodies were used. These experiments also showed that in the latter case, no plasmids with detectable inserted DNA were recovered, demonstrating that the level of non-specific background precipitation was low and therefore strongly suggesting that the clones generated by the procedure had arisen by virtue of having been part of a T antigen bearing nucleoprotein complex in the nucleus.

Further evidence supporting the conclusion that an enrichment for a fraction of the genome has occurred is provided by the enrichment for R. elements but not B2 repeats, which have comparable copy numbers in the mouse genome, and tentatively, in Experiments One and Two by the cloning of p24, a low copy number sequence.

The three clones analyzed in detail all contain T antigen binding sites, as judged by the McKay immunoprecipitation assay, and in one case by footprinting under stringent binding conditions. What fraction of all the clones obtained will bind T antigen <u>in vitro</u> remains to be determined. However, as an aid to rapidly identifying those clones which contain high affinity binding sites for T antigen, one of the approaches suggested in Section 3.1 could be adopted. The clones

could be mixed in pools of 10 or 20 and incubated with various amounts of pure T antigen under stringent conditions. Given equal amounts of starting plasmid, the T antigen should, to a first approximation distribute itself between the plasmids according to the relative affinities of the binding sites within them. Following filtration through nitrocellulose, or immunoprecipitation with anti - large T antibodies, the selected DNA would be used to transform bacteria, and the population of plasmids obtained analyzed. By this procedure, enrichment for plasmids with the highest affinity binding sites should occur. Once identified, the binding site within a particular plasmid could be localized by McKay assay. An alternative to this would be to label the clone _in vitro_ and digest with excess DNAse I in the presence of T antigen. The protected fragments could then be isolated and used to identify the region of the clone from which they originated. This approach makes the assumption that T antigen will bind to these clones in an equivalent manner both _in vivo_ and _in vitro_, and does not consider the possibility that T antigen may bind to chromatin in a complex with other proteins, a good candidate being p53, or may not bind directly to the DNA at all, but form part of a transcription complex. The T antigen - p53 complex is amenable to purification by an immunoaffinity procedure similar to that used to purify T antigen (see Chapter Four), and has been purified by chromatography on immobilized anti - p53 antibody columns (J. Gannon and D.

P. Lane, personal communication), but its biological activity remains to be assessed. This could however provide a convenient source of T - p53 complex for binding studies. Procedures are now well established for preparing nuclear protein extracts containing transcription factors (see Section 1.3), and also for reconstituting them into active transcription complexes on a naked DNA template. It might therefore be possible to assess the interaction of T antigen with the clones obtained, in the presence of transcription factors and other nuclear proteins.

Ultimately of course, studies of binding *in vitro* are unsatisfactory, and it is necessary to demonstrate that a similar interaction occurs *in vivo*. Recently, the exonuclease III technique of Wu (1983) and the footprinting with light (Becker and Wang, 1984) and genomic sequencing techniques (Church and Gilbert, 1984) have made such an analysis feasible. Therefore once further *in vitro* characterization of the clones obtained in this study has been performed the location of protein binding sites on them *in vivo* could be determined.

The immunoselection procedure is also clearly open to modification and improvement. The use of restriction enzymes for fractionation of the chromatin was originally for subsequent ease of cloning of the precipitated fragments, and because fragments of defined size would be obtained from the SV40 templates in the Cl 38 genome, thus

easily allowing them to be identified without the need for DNA sequencing. Now that it has been demonstrated that the procedure will produce DNA fragments which contain a high affinity T antigen binding site, alternative methods of fragmentation can be considered.

Restriction enzymes suffer from several drawbacks in a protocol such as this. First, long digestion times are required because under the ionic conditions used for digestion, the chromatin is highly condensed, making it more difficult for the enzyme to gain access to its recognition sequences (Gottesfeld, 1978). The prolonged incubation also provides an opportunity for proteolytic breakdown of chromosomal proteins to occur, and might also permit a "reassortment" of T antigen and other chromosomal proteins to occur to other locations on the chromatin. The possibility that some of the clones obtained might have arisen by this kind of event cannot be excluded.

A viable alternative would be to use DNAse I to fragment the chromatin, performing the digestion in situ in isolated nuclei. After digestion the nuclei could be lysed by altering the osmotic strength and nucleoprotein fragments precipitated as before in high salt. After end repair the fragments could be ligated to linkers to facilitate cloning, or cloned by homopolymer tailing. In the former case, end repair could be facilitated by the use of Mn^{2+} as the divalent cation during digestion, when DNAse I inserts double stranded cuts, rather than nicking only one strand, leaving blunt or nearly blunt ends

(Melgar and Goldthwait, 1968). The use of this enzyme would also reduce both the temperature and length of the digestion period required and hence the possibility of reassortment or proteolysis occuring.

This approach is clearly applicable not only to SV40 large T antigen, but to any DNA binding protein for which an antibody is available. When a protein whose sequence specificity has not previously been defined is under study, the initial approach would be to digest extensively with DNAse I to obtain a "minimal" protected fragment, which could then be cloned. All the fragments obtained would then be sequenced and compared in order to define a consensus recognition sequence, which could then be tested for its ability to rebind the protein in interest in vitro using standard techniques. The experiment could then be repeated with a less extensive digestion to clone adjacent sequences and study their expression and structure. Candidate proteins to which such an analysis could be applied are the steroid hormone receptors, p53, and the retroviral oncogene products myc and myb, though the analysis can clearly be extended. For instance, antibodies against DNA binding proteins from different tissues could be raised and screened to find those that recognize tissue specific proteins. Assuming that such a protein had regulatory function, one could then use immunoselection cloning to isolate those genes whose expression are under its control.

It is worth considering the number of potential high affinity sites for T antigen which one might expect in the genome. A sequence such as that found at site I - GAGGCTTTTTTGGAGGC might be expected to occur $(^1/_4)^{17}$ or once in every 1.7×10^{10} base pairs, thus at most once in the mouse genome. However, if one allows degeneracy in the central nucleotides, and within the pentanucleotides, thus; $G(^A/_G)GGC(N_7)G(^A/_G)GGC$, then the probability of occurrence is once in every 3×10^7 base pairs, or 100 times in the mouse genome. However, since the permitted degree of degeneracy within site I which will retain its capability to act as a high affinity binding site for T antigen has not yet been fully defined, such considerations are perhaps of limited value. A further objection to this kind of analysis is the possibility mentioned earlier that T antigen could bind to DNA in complex with other proteins. It will therefore be of great interest to determine the number of different sequences which will bind T antigen that there are among the clones obtained by the immunoselection procedure.

The relevance of the clones obtained in this study to the process of transformation by SV40 must also be considered. It is clear that the ability of T antigen to bind to the SV40 origin is not an absolute requirement for transformation of non-permissive cells to occur (Prives et al., 1983). It has also been found that many species of T antigen from transformed cells are unable, or impaired in their ability to bind specifically to the SV40 origin in

vitro (Burger and Fanning, 1983). The T antigen from Cl 38 cells, encoded by four of the five SV40 templates is replication competent and therefore presumably able to bind to the SV40 origin specifically. T antigen extracted from Cl 38 cells will discriminate between the SV40 origin and other sequences (pAT153) in vitro (my unpublished data), though the affinity of the protein for the origin has not been measured.

It cannot be excluded that in vivo, the mutant T antigens still show some sequence specific binding, either alone, or in a complex with some other protein(s), and that this contributes towards the establishment of the transformed state. Clearly, since the data presented in this study are as yet only preliminary, much additional work is required, to determine the structure and function (replication origin, transcription unit) of the clones obtained, by the kinds of experiments described in Chapters Four and Five, before any conclusions concerning their role in the transformation event can be drawn.

It is also worth considering what methodology could be applied to studying the converse problem to that addressed in this work, that is, given a defined DNA sequence with suspected or proven regulatory function, how might one identify proteins which interact with it?.

As discussed briefly in Section 1.3, the protein blotting technique of Bowen et al., (1981). has been applied to the study of proteins which interact with the

heat shock genes of Drosophila (Jack et al., 1981). In this method, nuclear protein extracts were prepared and fractionated according to size on a polyacrylamide gel. The proteins were then transferred to nitrocellulose which was incubated with a mixture of labelled restriction fragments from the gene of interest. After suitable washing, the filter was autoradiographed, and the molecular weights of the labelled proteins determined. The DNA was then eluted from the filters and analyzed to determine if any fragments had been selectively retained.

This method, while attractive in principle, suffers from several potential drawbacks. Since the method uses a denaturing SDS polyacrylamide gel only those proteins which retain DNA binding activity after boiling in 2% ($^{w}/v$) SDS will be detected. It is also conceivable that the capability of the protein to recognize DNA in a sequence specific manner may be impaired by such treatments. Denaturation of the proteins prior to electrophoresis will of course also disrupt any protein - protein complexes, so any specific binding which results from the interaction of two or more subunits will not be detected by this technique. Another potential problem may be the presence of highly basic proteins such as histones, which may give false positive signals by binding DNA fragments non-specifically. This will of course be readily detectable when the retained DNA fragments are eluted and analyzed, but serves to increase the amount of work. A way in which this might be overcome would be to

preincubate the filter with a large excess of unlabelled non-specific DNA to saturate these proteins prior to addition of the probe.

Another approach which could be used would be to fractionate nuclear extracts by ion exchange chromatography, gel filtration or sedimentation in sucrose or glycerol gradients, and to assay the fractions thus obtained for the ability to bind to the sequence of interest. Initially this would be done by nitrocellulose filter assays of the type described in Section 5.1, and then by footprinting. In both cases the assay would be performed in the presence of excess, unlabelled non-specific DNA to compete out proteins binding only non-specifically. Such an approach is likely to prove fruitful, and has been used in the search for transcription factors (see Section 1.3).

Another approach would be to use sequential chromatography on two DNA cellulose columns, one to which non-specific DNA had been coupled, and the other to which the sequence of interest had been coupled. A nuclear protein extract would then be prepared and circulated sequentially over the two columns, linked in tandem, initially at high salt, then gradually decreasing salt concentrations. A protein should then distribute itself between the two columns according to its relative affinities for the specific and non-specific DNAs. Thus it should be possible to enrich for proteins binding to

the sequence of interest, providing their affinity for it is significantly higher than for non-specific DNA. Both this approach and the previous one will permit the analysis of binding by complexes of proteins and thus have a distinct advantage over the protein blotting technique in this respect. Clearly, it will be of interest to use techniques such as these to identify cellular proteins which interact with the DNA sequences isolated in this study.

Section 7.4: General Summary.

The cellular DNA sequences with which T antigen interacts have been isolated from the rest of the genome by immunoprecipitation with antibodies against T antigen following nuclease digestion of chromatin prepared from SV40 transformed cells. These sequences have then been cloned and analyzed.

The selectivity of the procedure has been demonstrated by a variety of criteria. Firstly, there is a very low background of colonies containing plasmids with inserts when control antibodies are used for precipitation; second, SV40 origin sequences were selectively immunoprecipitated only with anti - T monoclonal antibodies by this procedure; third, enrichment for a subset of the rest of the genome was also observed.

Detailed analysis of three clones has shown that

all appear to contain binding sites which T antigen will recognize *in vitro* with reasonable affinity. Footprint analysis of one clone has demonstrated that T antigen binds to a region containing pentanucleotide recognition elements of the type found at the SV40 origin of replication. Analysis of the expression of the clones obtained suggests that some are differentially transcribed in normal and transformed cells. However, direct regulation by T antigen remains to be proved.

A novel purification procedure for SV40 T antigen has been developed which combines the use of an Adenovirus SV40 hybrid virus and immunoaffinity chromatography. The procedure is rapid, and gives a very high yield of biologically active T antigen that is greater than 98% pure. An attempt to produce T antigen using *S. cerevisiae* as a heterologous host was unsuccessful.

It is clear that immunoselection cloning potentially affords a very powerful tool for the analysis of the interaction of gene regulatory proteins with the genes they control. The technique described here should therefore, with minor modifications, be applicable to the analysis of proteins other than T antigen and the sequences they interact with.

REFERENCES

REFERENCES

Abraham, J., Feldman, J., Nasmyth, K. A., Strathern, J. N., Klar, A. J. S., Broach, J. R., and Hicks, J. B. (1982) Cold Spring Harbor Symp. Quant. Biol. 47, 989 - 998.

Abraham, J., Nasmyth, K. A., Strathern, J. N., Klar, A. J. S., and Hicks, J. B. (1984). J. Mol. Biol. 176, 307 - 331.

Adhya, S., and Garges, S. (1982). Cell 29, 287 - 289.

Adolph, K. W., Cheng, S. M., and Laemmli, U. K. (1977). Cell 12, 805 - 816.

Aiba, H. (1983). Cell 32, 141 - 149.

Akusjarvi, G., and Persson, H. (1981). Nature 292, 420 - 426.

Alwine, J. C., Reed, S. I., and Stark, G. R. (1977). J. Virol. 24, 22 - 27.

Anderson, J., Ptashne, M., and Harrison, S. C. (1984). Proc. Natn. Acad. Sci. USA. 81, 1307 - 1311.

Anderson, W. F., Ohlendorf, D. H., Takeda, Y., and Matthews, B. W. (1981). Nature 290, 754 - 758.

Arst, H. N. (1976). Nature 262, 231 - 241.

Avila, J., de Garcini, E. M., Wandosell, F., Villasante, A., Sogo, J. M., and Villanueva, N. (1983). EMBO J. 2, 1229 -1234.

Aviv, H., and Leder, P. (1972). Proc. Natn. Acad. Sci. USA. 69, 1408 - 1412.

Axel, R. (1978). Methods in Cell Biology 18, 41 - 53.

Axel, R., Friegelson, P., and Schutz, G. (1976). Cell 7, 247 - 254.

Baer, B. W., and Rhodes, D. (1983). Nature 301, 482 - 488.

Banerji, J., Olsen, L., and Schaffner, W. (1983). Cell 33, 729 - 740.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Cell 27, 299 - 308.

Barkley, M. D. (1981). Biochemistry 20, 3833 - 3842.

Barkley, M. D., and Bourgeois, S. (1978). in The Operon (eds. Miller, J. H., and Reznikoff, W. S.), Cold Spring Harbor Laboratory, New York.

Baumann, E., and Hand, R. (1982). J. Virol. 44, 78 - 87.

Becker, M. M., and Wang, J. C. (1984). Nature 309, 682 - 687.

Beggs, J. D. (1978). Nature 275, 104 - 109.

Beggs, J. D. (1981). Alfred Benzon Symposium 16, 429 - 435.

Beggs, J. D., van den Berg, J., van Ooyen, A., and Weissman, C. (1980). Nature 283, 835 - 840.

Bennetzen, J. L., and Hall, B. D. (1982). J. Biol. Chem. 257, 3026 - 3031.

Benoist, C., and Chambon, P. (1981). Nature 290, 304 - 310.

Berezney, R., and Coffey, D. S. (1974). Biochem. Biophys. Res. Commun. 60, 1410 - 1417.

Berg, O. G., Winter, R. B., and von Hippel, P. H. (1981). Biochemistry 20, 6929 - 6948.

Berg, O. G., Winter, R. B., and von Hippel, P. H. (1982). TIBS 7, 52 - 55.

Berk, A. J., and Sharp, P. A. (1978). Proc. Natn. Acad. Sci. USA. 75, 1274 - 1278.

Biamonti, G., Cobianchi, F., Falaschi, A., and Riva, S. (1983). EMBO J. 2, 161 - 165.

Bienz, M., and Gurdon, J. B. (1982). Cell 29, 811 - 819.

Birnboim, H. C., and Doly, J. (1979). Nucleic Acids Res. 7, 1513 - 1523.

Bishop, J. O. (1974). Cell 3, 81 - 86.

Bishop, J. O., Rosbash, M., and Evans, D. J. (1974). J. Mol. Biol. 85, 75 - 86.

Blattner, F. R., and Tucker, P. W. (1984). Nature 307, 417 - 422.

Bogenhagen, D. F., Wormington, W. M., and Brown, D. D. (1982). Cell 28, 413 - 421.

Borgmeyer, U., Nowock, J., and Sippel, A. E. (1984). Nucleic Acids Res. 12, 4295 - 4311.

Borst, P., Bernards, P., van der Ploeg, L. H. T., Michels, P. A. M., Liu, A. Y. C., de Lange, T., and Kooter, J. M. (1983). Eur. J. Biochem. 137, 383 - 389.

Botchan, M., Topp, W., and Sambrook, J. (1978). Cold Spring Harbor Symp. Quant. Biol. 43, 709 - 719.

Bourgeois, S., Pfahl, M., and Baulieu, E. E. (1984). EMBO J. 3, 751 - 755.

Bowen, B., Steinberg, J., Laemmli, U. K. and Weintraub, H. (1980). Nuclei Acids Res. 8, 1 - 20.

Boyer, H. W., and Roulland - Dussoix, D. (1969). J. Mol. Biol. 41, 459 - 472.

Bradley, M. K., Griffin, J. D., and Livingston, D. M. (1982). Cell 28, 125 - 134.

Brady, J., Bolen, J. B., Radonovich, M., Salzman, N., and Khoury, G. (1984). Proc. Natn. Acad. Sci. USA. 81, 2040 - 2044.

Britten, R. J., and Davidson, E. H. (1969). Science 165, 349 - 357.

Brown, D. D. (1981). Science 211, 667 - 674.

Brown, D. D. (1984). Cell 37, 359 - 365.

Buetti, E., and Diggelmann, H. (1983). EMBO J. 2, 1423 - 1429.

Burger, C., and Fanning, E. (1983). Virology 126, 19 - 31.

Chandler, V. L., Maler, B. A., and Yamamoto, K. R. (1983). Cell 33, 489 - 499.

Chaudry, F., Harvey, R., and Smith, A. E. (1982), J. Virol. 44, 54 - 66.

Church, G. M., and Gilbert, W. (1984). Proc. Natn. Acad. Sci. USA. 81, 1991 - 1995.

Ciejek, E. M., Tsai, M - J., and O'Malley, B. W. (1983). Nature 306, 607 - 609.

Chou, J. Y., and Martin, R. G. (1975). J. Virol. 15, 145 - 150.

Christensen, J. B., and Brockman, W. W. (1982). J. Virol. 44, 574 - 585.

Cizewski, V., and Sollner - Webb, B. (1983). Nucleic Acids Res. 11, 7043 - 7056.

Clark, C. R. (1984). TIBS 9, 207 - 208.

Clark, R., Lane, D. P., and Tjian, R. (1981). J. Biol. Chem. 256, 11854 - 11858.

Clark, R., Peden, K., Pipas, J. M., Nathans, D., and Tjian, R. (1983). Mol. Cell. Biol. 3, 220 - 228.

Clertant, P., Gaudray, P., Cuzin, F. (1984). EMBO J. 3, 303 - 307.

Clewell, D. B., and Helinski, D. R. (1969). Proc. Natn. Acad. Sci. USA. 62, 1159 - 1166.

Cogen, B. (1978). Virology 85, 222 - 230.

Cohen, J. D., Eccleshall, T. R., Needleman, R. B., Federoff, H., Buchferer, B. A., and Marmur, J. (1980). Proc. Natn. Acad. Sci. USA. 77, 1078 - 1082.

Cole, C. N., Crawford, L. V., and Berg, P. (1979). J. Virol. 30, 683 - 691.

Compton, J. G., Schrader, W. T., and O'Malley, B. W. (1982). Biochem. Biophys. Res. Commun. 105, 96 - 104.

Compton, J. G., Schrader, W. T., and O'Malley, B. W. (1983). Proc. Natn. Acad. Sci. USA. 80, 16 - 20.

Compton, J. G., Schrader, W. T., and O'Malley, B. W. (1984). J. Steroid Biochem. 20, 89 - 94.

Conrad,S. E., and Botchan, M. R. (1982). Mol. Cell. Biol. 2, 949 - 965.

Cooke, N. E., and Baxter, J. D. (1982). Nature 297, 603 - 606.

Cozzarelli, N. R., Gerrard, S. P., Schlissel, M., Brown, D. D., and Bogenhagen, D. F. (1983). Cell 34, 829 - 835.

Crawford, L. V., Pim, D. C., and Lane, D. P. (1980). Virology 100, 314 - 325.

Crawford, L. V., Leppard, K., Lane, D. P., and Harlow, E. (1982). J. Virol. 42, 612 - 620.

D'Alisa, R. M., and Gershey, E. L. (1978). Nature 274, 164 - 166.

Danchin, A., and Ullmann, A. (1980). TIBS 5, 51 - 52.

Darnell, J. E. (1982). Nature 297, 365 - 371.

Davidson, E. H., and Britten, R. J. (1973). Quarterly Rev. Biol. 48, 565 - 613.

Davidson, E. H., and Britten, R. J. (1979). Science 204, 1052 - 1059.

Davidson, E.H., Jacobs, H. T., and Britten, R. J. (1983). Nature 301, 468 - 470.

Davidson, E. H., and Posakony, J. W. (1982). Nature 297, 633 - 635.

Davison, B. L., Egly, J. M., Mulvihill, J. R., and Chambon, P. (1983). Nature 301, 680 - 686.

de Crombrugghe, B., and Pastan, I. (1978). in The Operon (eds. Miller, J. H., and Reznikoff, W. S.), Cold Spring Harbor Laboratory, New York.

de Haseth, P. L., Gross, C. A., Burgess, R. R., and Record, M. T. (1977). Biochemistry 16, 4777 - 4790.

De Lucia, A. L., Lewton, B. A., Tjian, R., and Tegtmeyer, P. (1983). J. Virol. 46, 143 - 150.

de Villiers, J., Olson, L., Tyndall, C., and Schaffner, W. (1982). Nucleic Acids Res. 10, 7965 - 7976.

Dean, D. C., Knoll, B. J., Riser, M. E., and O'Malley, B. W. (1984). Nature 305, 551 - 554.

Derman, E., Krauter, K., Walling, L., Weinberger, C., Ray, M., and Darnell, J. E. (1981). Cell 23, 731 - 739.

Dilworth, S. M., Cowie, A., Kamen, R. I., and Griffin, B. E. (1984). Proc. Natn. Acad. Sci. USA. 81, 1941 - 1945.

DiMaio, D., and Nathans, D. (1980). J. Mol. Biol. 140, 129 - 142.

DiMaio, D., and Nathans, D. (1982). J. Mol. Biol. 156, 531 - 548.

Donner, P., Greiser-Wilke, I., and Moelling, K. (1982). Nature 296, 262 - 266.

Doolittle, W. F., and Sapienza, C. (1980). Nature 284, 601 - 603.

Dorn, A., Brauer, D., Otto, B., Fanning, E., and Knippers, R. (1982). Eur. J. Biochem. 128, 53 - 62.

Dynan, W. S., and Tjian, R. (1983 a). Cell 32, 669 - 680.

Dynan, W. S., and Tjian, R. (1983 b). Cell 32, 79 - 87.

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R., and Hood, L. (1980). Cell 20, 313 - 319.

Edens, L., Bom, I., Lederboer, A. M., Maat, J., Toonen, M. Y., Visser, C., and Verrips, C. (1984). Cell 37, 629 - 633.

Engelke, D. R., Ng, S. Y., Shastry, B. S., Roeder, R. G. (1980). Cell 19, 717 - 728.

Falkner, F. G., Zachau, H. G. (1984). Nature 310, 71 - 74.

Fanning, E., Nowak, B., and Burger, C. (1981). J. Virol. 37, 92 - 102.

Fanning, E., Westphal, K - H, Brauer, D., and Corlin, D. (1982). EMBO J. 1, 1023 - 1028.

Fanning, T. G. (1982). Nucleic Acids Res. 10, 5003 - 5013.

Fanning, T. G. (1983). Nucleic Acids Res. 11, 5073 - 5091.

Favaloro, J., Treisman, R., and Kamen, R. (1980). Methods in Enzymology, 65, 718 - 750. (Moldave, L., and Grossman, L. eds.)

Fluck, M. M., and Benjamin, T. L. (1979). Virology 96, 205 - 228.

Francke, B., and Eckhart, W. (1973). Virology 55, 127 - 135.

Frederick, C. A., Grable, J., Melia, M., Samudzi, C., Jen - Jacobson, L., Wang, B. C., Greene, C., Boyer, H. W., Rosenberg, J. M. (1984). Nature 309, 327 - 331.

Fried, M., and Crothers, D. M. (1981). Nucleic Acids Res. 9, 6505 - 6525.

Fried, M., Griffiths, M., Davies, B., Bjursell, G., LaMantia, G., and Lania, L. (1983). Proc. Natn. Acad. Sci. USA. 80, 2117 - 2121.

Fried, M. G., and Crothers, D. M. (1984 a). J. Mol. Biol. 172, 241 - 262.

Fried, M. G., and Crothers, D. M. (1984 b). J. Mol. Biol. 172, 263 - 282.

Fromm, M., and Berg, P. (1982). J. Mol. App. Genet. 1, 457 - 481.

Galas, D. J., and Schmitz, A. (1978). Nucleic Acids Res. 5, 3157 - 3170.

Galau, G. A., Klein, W. H., Davis, M. M., Wold, B. J., Britten, R. J., and Davidson, E. H. (1976). Cell 7, 487 - 505.

Garel, A., Zolan, M., and Axel, R. (1977) Proc. Natn. Acad. Sci. USA. 74, 4867 - 4871.

Gaudray, P., Tyndall, C., Kamen, R., and Cuzin, F. (1981). Nucleic Acids Res. 9, 5697 - 5710.

Gaynor, R. B., and Berk, A. J. (1983). Cell 33, 683 - 693.

Gebhard, W., Meitinger, T., Hochtel, J., Zachau, Z. (1982). J. Mol. Biol. 157, 453 - 471.

Gebhard, W., and Zachau, H. G. (1983). J. Mol. Biol. 170, 255 - 270.

Ghosh, P. K., and Lebowitz, P. (1981). J. Virol. 40, 224 - 240.

Giacherio, D., and Hager, L. P. (1979). J. Biol. Chem. 254, 8113 - 8116.

Gidoni, D., Scheller, A., Barnet, B., Hantzopoulos, P., Oren, M., and Prives, C. (1982). J. Virol. 42, 456 - 466.

Gilbert, W., and Maxam, A. (1973). Proc. Natn. Acad. Sci. USA. 70, 3581 - 3585.

Gilkes, S. D., Morrison, S. L., Oi, V. T., and Tonegawa, S. Cell 33, 717 - 728.

Gluzman, Y. (1981). Cell 23, 175 - 182.

Gluzman, Y., Reichl, H., and Solnick, D. (1982). in Eukaryotic Viral Vectors, Cold Spring Harbor Laboratory, New York. (ed. Gluzman, Y), pp 187 - 192.

Goeddel, D. V., Yansura, D. G., and Caruthers, M. H. (1982). Proc. Natn. Acad. Sci. USA. 75, 3578 - 3582.

Goff, C. G., Mori, D. T., Kohno, T., Gravius, T. C., Smith, R. A., Yamasaki, E., and Taunton - Rigby, A. (1984). Gene 27, 35 - 46.

Goldman, N., Brown, M., and Khoury, G. (1981). Cell 24, 567 - 572.

Gorman, C. M., Moffat, L. F., and Howard, B. H. (1982). Mol. Cell. Biol. 2, 1044 - 1051.

Gottesfeld, J. M. (1978). Methods in Cell Biology 16, 421 - 436.

Gottesfeld, J. M., Garrard, W. T., Bagi, G., Wilson, R. F., and Bonner, J. Proc. Natn. Acad. Sci. USA. 71, 2193 - 2197.

Graessmannn, A., Graessmann, M., Tjian, R., and Topp, W. C. (1980). J. Virol. 33, 1182 - 1191.

Graham, F. L., Abrahams, P. J., Mulder, C., Heinjker, H. L., Warnaar, S. O., deVries, F. A. J., Fiers, W., and van der Eb, A. J. (1974). Cold Spring Harbor Symp. Quant. Biol. 39, 637 - 650.

Graham, F. L., Sunley,J., Russel, W. C., and Nairu, R. (1977). J. Gen. Virol. 36, 59 - 72.

Green, M. R., Treisman, R., and Maniatis, T. (1983). Cell 35, 137 - 148.

Griffin, J. D., Spangler, G., and Livingston, D. M. (1979 a). Cold Spring Harbor Symp. Quant. Biol. 44, 113 - 122.

Griffin, J. D., Spangler, G., and Livingston, D. M. (1979 b). Proc. Natn. Acad. Sci. USA. 76, 2610 - 2614.

Gross-Bellard, M., Oudet, P., and Chambon, P. (1973). Eur. J. Biochem. 36, 32 - 38.

Groudine, M., Peretz, M., and Weintraub, H. (1981). Mol. Cell. Biol. 1, 281 - 288.

Grove, G., and Marzluff, G. A. (1981). J. Biol. Chem. 256, 463 - 470.

Grunstein, M., and Hogness, D. S. (1975). Proc. Natn. Acad. Sci. USA. 72, 3961 - 3965.

Gruss, C., Baumann, E., and Knippers, R. (1984). J. Virol. 50, 943 - 946.

Gutai, M. W., and Nathans, D. (1978). J. Mol. Biol. 126, 259 - 274.

Hanahan, D. (1983). J. Mol. Biol. 166, 557 - 580.

Hanahan, D., and Meselson, M. (1980). Gene 10, 63 - 67.

Hanas, J. S., Bogenhagen, D. F., and Wu, C. W. (1984). Proc. Natn. Acad. Sci. USA. 81, 2142 - 2145.

Hansen, U., Tenen, D. G., Livingston, D. M., and Sharp, P. A. (1981). Cell 27, 603 - 612.

Harlow, E., Crawford, L. V., Pim, D. C., and Williamson, N. M. (1981). J. Virol. 39, 861 - 869.

Harper, F., Florentin, Y., and Puvion, E. (1984). EMBO J. 3, 1235 - 1241.

Hastie, N. D., and Bishop, J. O. (1976). Cell 9, 761 - 774.

Hay, N., Skolnik - David, H., and Aloni, Y. (1982). Cell 29, 183 - 193.

Hearing, P., and Shenk, T. (1983). Cell 33, 695 - 703.

Heller, R., and Arnheim, N. (1980). Nucleic Acids Res. 8, 5031 - 5042.

Henikoff, S., Tatchell, K., Hall, B. D., and Nasmyth, K. A. (1981). Nature 289, 33 - 41.

Henning, R., and Lange - Mutschler, J. (1983). Nature 305, 736 - 738.

Hentzen, P. C., Rho, J. H., and Bekhor, I. (1984). Proc. Natn. Acad. Sci. USA. 81, 304 - 307.

Hirt, B. (1967). J. Mol. Biol. 26, 365 - 369.

Hiscott, J. B., and Defendi, V. (1979). J. Virol. 30, 590 - 599.

Hiscott, J. B., and Defendi, V. (1981). J. Virol. 37, 802 - 812.

Hitzeman, R., Hagie, F. E., Levine, H. L., Goeddel, D. V., Ammerer, G., and Hall, B. D. (1981). Nature 293, 717 - 722.

Hitzemann, R. A., Leung, D. W., Perry, L. J., Kohr, W. J., Levine, H. L., and Goeddel, D. V. (1983). Science 219, 620 - 625.

Ho, Y. S., Wulff, D. L., and Rosenberg, M. (1983). Nature 304, 703 - 709.

Hochschild, A., Irwin, N., and Ptashne, M. (1983). Cell 32, 319 - 325.

Honda, B. M., and Roeder, R. G. (1980). Cell 22, 119 - 126.

Horwich, A., Koop, A. H., and Eckhart, W. (1982). Mol. Cell. Biol. 2, 88 - 92.

Hsieh, T. S., and Brutlag, D. L. (1979). Proc. Natn. Acad. Sci. USA. 76, 726 - 730.

Hughes, M. R., Compton, J. G., Schrader, W. T., and O'Malley, B. W. (1981). Biochemistry 20, 2481 - 2491.

Hynes, N., van Ooyen, A. J. J., Kennedy, N., Herrlich, P., Ponta, H., and Groner, B. (1983). Proc. Natn. Acad. Sci. USA. 80, 3637 - 3641.

Imperiale, M. J., Kao, H. T., Feldmann, L. T., Nevins, J. R., and Strickland, S. (1984). Mol. Cell. Biol. 4, 867 - 874.

Iverson, L. E., and Rose, J. K. (1981). Cell 32, 477 - 484.

Jack, R. S., Gehring, W. J., and Brack, C. (1981). Cell 24, 321 - 331.

Jackson, V., and Chalkley, R. (1981). Proc. Natn. Acad. Sci. USA. 78, 6081 - 6085.

Jacob, F., and Monod, J. (1961). J. Mol. Biol. 3, 318 - 356.

Jacquet, M., Affara, N. A., Robert, B., Jakob, H., Jakob, F., and Gros, F. (1978). Biochemistry 17, 69 - 79.

Jay, G., Khoury, G., DeLeo, A. B., Dippold, W. G., and Old, W. J. (1981 a). Proc. Natn. Acad. Sci. USA. 78, 2932 - 2936.

Jay, G., Nomura, S., Anderson, C. W., and Khoury, G. (1981 b). Nature 291, 346 - 349.

Jelinek, W. R., and Schmid, C. W. (1982). Ann. Rev. Biochem. 51, 813 - 844.

Jelinek, W. R., Toomey, T. P., Leinwand, L., Duncan, C. H., Biro, P. A., Choudary, P. V., Weissman, S. M., Rubin, C. M., Houck, C. M., Deininger, P. L., and Schmid, C. W. (1980). Proc. Natn. Acad. Sci. USA. 77, 1398 - 1402.

Jessel, D., Landau, T., Hudson, J., Lalor, L., Tenen, D., and Livingston, D. M. (1976). Cell 8, 535 - 545.

Johnson, A. D., Poteete, A. R., Lauer, G., Sauer, G., Achers, G. K., and Ptashne, M. (1981). Nature 294, 217 - 223.

Jones, K. A., and Tjian, R. (1984). Cell 36, 155 - 162.

Kamaly, J. C., and Goldberg, R. B. (1980). Cell 19, 935 - 946.

Karin, M., Alaslinger, A., Holtgreve, H., Richards, R. I., Krauter, P., Westphal, H. M., and Beato, M. (1984). Nature 309, 513 - 519.

Keller, J. M., and Alwine, J. C. (1984). Cell 36, 381 - 389.

Khandjian, E. W., Loche, M., Darlix, J. L., Cramer, R., Turler, H., and Weil, R. (1979). Proc. Natn. Acad. Sci. USA. 79, 1139 - 1143.

Khoury, G., Gruss, P., Dhar, R., and Lai, C. J. (1979). Cell 18, 85 - 92.

Kim, S., Davis, M., Sinn, E., Patten, P., and Hood, L. (1981). Cell 27, 573 - 581.

King, W. J., and Greene, G. L. (1984). Nature 307, 745 - 747.

Klar, A. J. S., Strathern, J. N., Broach, J. R., Hicks, J. B. (1981). Nature 289, 239 - 244.

Klemenz, R., Stillman, D. J., and Geiduschek, E. P. (1982). Proc. Natn. Acad. Sci. USA. 79, 6191 - 6195.

Klockmann, U., and Deppert, W. (1983). FEBS Letters 151, 257 - 259.

Kominami, Y., Urano, Y., Mishima, Y., and Muramatsu, M. (1983). J. Mol. Biol. 165, 209 - 228.

Kostriken, R., Strathern, J. N., Klar, A. J. S., Hicks, J. B., and Heffron, F. (1983). Cell 35, 165 - 174.

Kramerov, D. A., Grigoryan, A. A., Ryskov, A. P., and Goergiev, G. P. (1979). Nucleic Acids Res. 6, 697 - 713.

Kramerov, D. A., Ryskov, A. P., and Georgiev, G. P. (1977). Biochem. Biophys. Acta. 475, 461 - 475.

Krayev, A. S., Kramerov, D. S., Skryabin, K. G., Ryskov, A. P., Bayev, A. A., and Georgiev, G. P. (1980). Nucleic Acids Res. 8, 1201 - 1215.

Krayev, A. S., Markusheva, T. V., Kramerov, D. A., Ryskov, A. P., Skryabin, K. G., Bayev, A. A., and Georgiev, G. P. (1982). Nucleic Acids Res. 10, 7461 - 7475.

Kuhn, R. W., Schrader, W. T., Smith, R. G., and O'Malley, B. W. (1975). J. Biol. Chem. 250, 4220 - 4228.

Laemmli, U. K. (1970). Nature 227, 680 - 685.

Laimins, L., Khoury, G., Gorman, C., Howard, B. H., and Gruss, P. (1982). Proc. Natn. Acad. Sci. USA. 79, 6453 - 6457.

Lane, D. P., and Crawford, L. V. (1979). Nature 278, 261 - 263.

Lane, D. P., and Crawford, L. V. (1980). Proc. Roy. Soc. Lond. B. 210, 451 - 453.

Lane, D. P., Gannon, J., and Winchester, G. (1982) in Advances in Viral Oncology (ed. Klein, G), Raven Press, New York. Volume 2, pp 23 - 29.

Lane, D. P., and Hoeffler, W. K. (1980). Nature 288, 167 - 170.

Langford, C. J., and Gallwitz, D. (1983). Cell 33, 519 - 527.

Langford, C. J., Klinz, F - J., Donath, C., and Gallwitz, D. (1984). Cell 36, 645 - 653.

Langford, C. J., Nellen, W., Niessing, J., and Gallwitz, D. (1983). Proc. Natn. Acad. Sci. USA. 80, 1496 - 1500.

Laskey, R. A., and Mills, A. D. (1977). FEBS Letters 82, 314 - 316.

Lasser, A. B., Martin, P. L., and Roeder, R. G. (1983). Science 222, 740 - 748.

Latchman, D. S., Brzeski, H., Lovell - Badge, R., Evans, M. J. (1984). Biochem. Biophys. Acta. (in the press).

Lawson, G. M., Knoll, B. J., March, C. J., Woo, S. L. C., Tsai, M - J., and O'Malley, B. W. (1982). J. Biol. Chem. 257, 1501 - 1507.

Lawson, G. M., Tsai, M - J., and O'Malley, B. W. (1980). Biochemistry 19, 4403 - 4411.

Learned, R. M., Smale, S. T., Haltiner, M. M., and Tjian, R. (1983). Proc. Natn. Acad. Sci. USA. 80, 3558 - 3562.

Lehrach, H., Diamond, D., Wozney, J. M., and Boedtker, H. (1977). Biochemistry 16, 4743 - 4751.

Leuders, K. K., and Paterson, B. M. (1982). Nucleic Acids Res. 10, 7715 - 7729.

Lewin, B. (1980). Gene Expression, Volume 2: Eucaryotic Chromosomes. Wiley Interscience

Lewis, A. M., and Martin, R. G. (1979). Proc. Natn. Acad. Sci. USA 76, 4299 - 4302.

Lewis, M. K., and Burgess, R. R. (1980). J. Biol. Chem. 255, 4928 - 4936.

Lewton, B. A., DeLucia, A. L., and Tegtmeyer, P. (1984). J. Virol. 49, 9 - 13.

Lin, S., Y., and Riggs, A. D. (1975). Cell 4, 107 - 111.

Lowry, O. H., Rosebrough, N. J., Farr, A. L., and Randall, R. J. (1951). J. Biol. Chem. 193, 265 - 275.

McCutchan, T. F., and Singer, M. F. (1981). Proc. Natn Acad. Sci. USA. 78, 95 - 99.

McKay, D. B.. and Steitz, T. A. (1981). Nature 290, 744 - 749.

McKay, R. D. G. (1981). J. Mol. Biol. 145, 471 - 488.

McKay, R., and DiMaio, D. (1981). Nature 289, 810 - 813.

McKnight, G. S., and Palmiter, R. D. (1979). J. Biol. Chem. 254, 9050 - 9058.

Mann, K., and Hunter, T. (1979). J. Viorl 29, 232 - 241.

Mann, K., and Hunter, T. (1980) Virology 107, 526 - 532.

Majors, J. (1975). Proc. Natn. Acad. Sci. USA. 72, 4394 - 4398.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning, A Laboratory Manual., Cold Spring Harbor, New York.

Marchioni, M. A., and Roufa, D. J. (1981). Cell 26, 245 - 258.

Margolskee, R. F., and Nathans, D. (1984). J. Virol. 49, 386 - 393.

Martin, R. G. (1981) Adv. Cancer. Res. 34, 1 - 68.

Martin, R. G., and Oppenheim, A. (1977). Cell 11, 859 - 869.

Martin, R. G., Setlow, V. P., Edwards, C. A., and Vembu, D. (1979). Cell 17, 635 - 643.

Mason, A. J., Evans, B. A., Cox, D. R., Shine, J., and Richards, R. I. (1983). Nature 303, 300 - 307.

Matsui, T., Segall, J., Weil, A. P., and Roeder, R.G. (1980). J. Biol. Chem. 255, 11992 - 11996.

Matthews, B. W., Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., and Takeda, Y. (1983). TIBS 8, 25 - 29.

Maxam, A. M., and Gilbert, W. (1977). Proc. Natn. Acad. Sci. USA. 74, 560 - 564.

Maxam, A. M., and Gilbert, W. (1980). Methods in Enzymology. (eds. Grossman, L., and Moldave, K.), Academic Press. Volume 65, 499 - 580.

Mayfield, J. E., Serunian, J. A., Silver, L. M., and Elgin, S. C. R. (1978). Cell 14, 539 - 544.

Melgar, E., and Goldthwait, D. A. (1968). J. Biol. Chem. 243, 4409 - 4416.

Mercer, W. E., Nelson, D., DeLeo, A. B., Old, L. J., and Baserga, R. (1983). Proc. Natn. Acad. Sci. USA. 80, 6309 - 6312.

Messing, J., and Vieira , J. (1982). Gene 19, 269 - 276.

Miesfield, R., and Arnheim, N. (1984). Mol. Cell. Biol. 4, 221 - 227.

Miller, D. M., Turner, P., Niehaus, A. W., Axelrod, D. E., and Golpakrishnan, T. V. (1978). Cell 14, 511 - 521.

Miller, J. H. (1978). in The Operon. (eds. Miller, J. H., and Reznikoff, W. S.), Cold Spring Harbor Laboratory, New York.

Miller, J. H., and Reznikoff, W. S. (eds.) The Operon. Cold Spring Harbor Laboratory, New York. (1978).

Milstein, C., Brownlee, G. G., Cartwright, E. M., Jarvis, J. M., and Proudfoot, N. J. (1974). Nature 252, 354 - 359.

Miyanohara, A., Toh - e, A., Nozai, C., Hameda, F., Ohtomo, N., and Matsubara, K. (1983). Proc. Natn. Acad. Sci. USA. 80, 1 - 5.

Mizuno, T., Chou, M - Y., and Inouye, M. (1984). Proc. Natn. Acad. Sci. USA. 81, 1966 - 1970.

Montenarh, M., and Henning, R. (1982). Biochim. Biophys. Acta. 697, 322 - 329.

Montenarh, M., and Henning, R. (1983). J. Virol. 45, 531 - 538.

Morrison, B., Kress, M., Khoury, G., and Jay, G. (1983). J. Virol. 47, 106 - 114.

Mulvihill, E. R., LePennec, J. P., and Chambon, P. (1982). Cell 26, 621 - 632.

Murphy, D., Brickell, P. M., Latchman, D. S., Willison, K., and Rigby, P. W. J. (1983). Cell 35, 865 - 871.

Myers, R. M., Kligman, M., and Tjian, R. (1981 a). J. Biol. Chem. 256, 10156 - 10160.

Myers, R. M., Rio, D. C., Robbins, A. K., Tjian, R. (1981 b). Cell 25, 373 - 384.

Myers, R. M., and Tjian, R. (1980). Proc. Natn. Acad. Sci. USA. 77, 6491 - 6495.

Myers, R. M., Williams, R. C., and Tjian, R. (1981 c). J. Mol. Biol. 148, 347 - 353.

Nagata, K., Guggenheimer, R. A., and Hurwitz, G. (1983). Proc. Natn. Acad. Sci. USA. 80, 6177 - 6181.

Nasmyth, K. A. (1983). Nature 302, 670 - 676.

Nasmyth, K. A., Tatchell, K., Hall, B. D., Astell, C., and Smith, M. (1981). Nature 289, 244 - 250.

Nevins, J. R. (1979). J. Mol. Biol. 130, 493 - 506.

Nevins, J. R. (1981). Cell 26, 213 - 220.

Nevins, J. R. (1982). Cell 29, 913 - 919.

Nevins, J. R. (1982). Cell 28, 1 - 2.

Nevins, J. R., and Darnell, J. E. (1978). Cell 15, 1477 - 1493.

Nevins, J. R., and Wilson, M. C. (1981). Nature 290, 113 - 118.

Ng, R., and Abelson, J. (1980). Proc. Natn. Acad. Sci. USA. 77, 3912 - 3916.

Nomura, S., Khoury, G., and Jay, G. (1983). J. Virol. 45, 428 - 433.

Nordheim, A., and Rich, A. (1983). Nature 303, 674 - 679.

Nordheim, A., Tesser, P., Azorin, F., Kwon, Y. H., Moller, A., and Rich, A. (1982). Proc. Natn. Acad. Sci. USA. 79, 7729 - 7733.

Nowock, J., and Sippel, A. E. (1982). Cell 30, 607 - 615.

Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y., and Matthews, B. W. (1982). Nature 298, 718 - 723.

Oren, M., Maltzman , W., and Levine, A. J. (1981). Mol. Cell. Biol. 1, 101 - 110.

Oren, M., Winocour, E., and Prives, C. (1980). Proc. natn. Acad. Sci. USA. 77. 220 - 224.

Orgel, L. E., and Crick, F. H. C. (1980). Nature 284, 604 - 607.

Osley, M. A., and Hereford, L. M. (1981). Cell 34, 377 - 384.

Pabo, C. O., Krovatin, W., Jeffrey, A., and Sauer, R. T. (1982). Nature 298, 441 - 443.

Pabo, C. O., and Lewis, M. (1982). Nature 298, 443 - 447.

Parker, C. S., and Topol, J. (1984 a). Cell 36, 357 - 369.

Parker, C. S., and Topol, J. (1984 b). Cell 37, 273 - 283.

Parslow, T. G., Blair, D. L., Murphy, W. J., and Granner, D. K. (1984). Proc. Natn. Acad. Sci. USA. 81, 2650 - 2654.

Paucha, E., Mellor, A., Harvey, R., Smith, A. E., Hewick, R. M., and Waterfield, M. D. (1978). Proc. Natn. Acad. Sci. USA. 75, 2165 - 2169.

Payvar, F., DeFranco, D., Firestone, G. L., Edgar, B., Wrange, O., Okret, S., Gustaffson, J - A., and Yamamoto, K. R. (1983). Cell 35, 381 - 392.

Payver, F., Wrange, O., Carlstedt - Duke, J., Okret, S., Gustaffson, J. A., and Yamamoto, K. R. (1981). Proc. Natn. Acad. Sci. USA. 78, 6628 - 6632.

Pearson, N. J., Fried, H. M., and Warner, J. R. (1982). Cell 29, 347 - 355.

Pelham, H. R. B. (1982). Cell 30, 517 - 528.

Pelham, H. R. and Brown, D. D. (1980). Proc. Natn. Acad. Sci. USA. 77, 4170 - 4174.

Peterson, G. L. (1977). Anal. Biochem. 83, 346 - 356.

Pettersson, U., and Mathews, M. B. (1977). Cell 12, 741 - 750.

Pfahl, M. (1982). Cell 31, 475 - 482.

Polvino - Bodnar, M., and Cole, C. N. (1982). J. Virol. 43, 489 - 502.

Pomerantz, B. J., Mueller, C. R., and Hassel, J. A. (1983). J. Virol. 47, 600 - 610.

Prives, C., Barnet, B., Scheller, A., Khoury, G., and Jay, G. (1982). J. Virol. 43, 73 - 82.

Prives, C., Beck, Y., Shore, H. (1980). J. Virol. 33, 689 - 696.

Prives, C., Covey, L., Scheller, A., and Gluzman, Y. Mol. Cell. Biol. 3, 1958 - 1966.

Prives, C., Gilboa, E., Revel, M., and Winocour, E. (1977). Proc. Natn. Acad. Sci. USA. 74, 457 - 461.

Ptashne, M. (1984). TIBS 9, 142 - 145.

Ptashne, M., Jeffrey, A., Johnson, A. D., Maurer, R., Meyer, B. J., Pabo, C. O., Roberts, T. M., and Sauer, R. T. (1980). Cell 11, 1 - 11.

Queen, C., and Baltimore, D. (1983). Cell 33, 741 - 748.

Queen, C., Lord, S. T., McCutchan, T. F., and Singer, M. F. (1981). Mol. Cell. Biol. 1, 1061 - 1068.

Radloff, R., Bauer, W., and Vinograd, J. (1967). Proc. Natn. Acad. Sci. USA. 57, 1514 - 1521.

Razin, A., and Riggs, A. D. (1980). Science 210, 604 - 610.

Record, M. T., deHaseth, P. L., and Lohmann, T. M. (1977). Biochemistry 16, 4791 - 4796.

Reddy, V. B., Ghosh, P. K., Lebowitz, P., Piatak, M., and Weissman, S. M. (1979). J. Virol. 30, 279 - 296.

Reed, S. I., Ferguson , J., Davis, R. W., and Stark, G. R. (1975). Proc. Natn. Acad. Sci. USA. 72, 1605 - 1609.

Reed, S. I., Stark, G. R., and Alwine, J. C. (1976). Proc. Natn. Acad. Sci. USA. 73, 3083 - 3087.

Reich, N. C., and Levine, A. J. (1982). Virology 117, 286 - 290.

Reich, N. C., and Levine, A. J. (1984). Nature 308, 199 - 201.

Reiser, J., Renart, J., Crawford, L. V., and Stark, G. (1980). J. Virol. 33, 78 - 87.

Renkawitz, R., Danesch, U., Matthias, P., Schutz, G. (1984). J. Steroid Biochem. 20, 99 - 104.

Reznikoff, W. S., and Abelson, J. N. (1978). in The Operon, (eds. Miller, J. H., and Reznikoff, W. S.) Cold Spring Harbor Laboratory, New York.

Reznikoff, W. S., Winter, R. B., and Hurley, C. K. (1974). Proc. Natn. Acad. Sci. USA. 71, 2314 - 2318.

Rezvin, A., and von Hippel, P. H. (1977). Biochemistry 16, 4767 - 4776.

Rigby, P. W. J., Chia, W., Clayton, C. E., and Lovett, M. (1980). Proc. Roy. Soc. Lond. B. 210, 437 - 450.

Rigby, P. W. J., Dieckmann, M., Rhodes, C., and Berg, P. (1977). J. Mol. Biol. 113, 237 - 251.

Rigby, P. W. J., and Lane, D. P. (1983). in Advances in Viral Oncology (ed. Klein, G), Volume 3, pp 31 - 57.

Riggs, A. D., Suzuki, H., and Bourgeois, S. (1970). J. Mol. Biol. 48, 67 - 83.

Rio, D., Robbins, A., Myers, R., and Tjian, R. (1980). Proc. Natn. Acad. Sci. USA. 77, 5706 - 5710.

Rio, D., and Tjian, R. (1983). Cell 32, 1227 - 1240.

Roberts, T. M., Bikel, I., Yocum, R. R., Livingston, D. M., and Ptashne, M. (1979). Proc. Natn. Acad. Sci. USA. 76, 5596 - 5600.

Rogers, J. (1983). Nature 306, 113 - 114.

Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L., Wall, R. (1980). Cell 20, 303 - 312.

Roggenkamp, R., Kustermann - Kuhn, B., and Hollenberg, C. P. (1981). Proc. Natn. Acad. Sci. USA. 78, 4466 - 4470.

Ross, W., and Landy, A. (1982). Proc. Natn. Acad. Sci. USA. 79, 7724 - 7728.

Ross, W., and Landy, A. (1983). Cell 33, 261 - 272.

Ross, W., Landy, A., Kikuchi, Y., and Nash, H. (1979). Cell 18, 297 - 307.

Rubin, H., Figge, J., Bladon, M. T., Chen, L. B., Ellman, M., Bikel, I., Farrell, M., and Livingston, D. M. (1982). Cell 30, 469 - 480.

Ruet, A., Camier, S., Smagowicz, W., Sentenac, A., and Fromageot, P. (1984). EMBO J. 3, 343 - 350.

Rundell, K., Hearing, P., and Yang, Y. C. (1979). Cold Spring Harbor Symp. Quant. Biol. 44, 211 - 214.

Ryder, K., DeLucia, A., and Tegtmeyer, P. (1983). Virology 129, 239 - 245.

Saffer,J. D., and Singer, M. F. (1984). Nucleic Acids Res. 12, 4769 - 4788.

Sakonju, S., and Brown, D. D. (1982). Cell 31, 395 - 405.

Sakonju, S., Brown, D. D., Engelke, D., Ng, S. Y., Shastry, B., and Roeder, R. G. (1981). Cell 23, 665 - 669.

Sanger, F., and Coulson, A. R. (1978). FEBS Letters 87, 107 - 110.

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., and Roe, B. A. (1980). J. Mol. Biol. 143, 161 - 178.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). Proc. Natn. Acad. Sci. USA. 74, 5463 - 5467.

Sammons, D. W., Adams, L. D., and Nishizawa, E. E. (1981). Electrophoresis 2, 135 - 141.

Samuels, M., Fire, A. Z., and Sharp, P. A. (1982). J. Biol. Chem. 257, 14419 - 14427.

Sarnow, P., Shih, Y. H., Williams, J., and Levine, A. J. (1982). Cell 28, 387 - 394.

Sauer, R. T., Yocum, R. R., Doolittle, R. F., Lewis, M., and Pabo, C. (1982). Nature 298, 447 - 451.

Scheidereit, C., and Beato, M. (1984). Proc. Natn. Acad. Sci. USA. 81, 3029 - 3033.

Scheidtmann, K. H., Hardung, M., Echle, B., and Walter, G. (1984). J. Virol. 1 - 12.

Scheller, A., Covey, L., Barnet, B., and Prives, C. (1982). Cell 29, 375 - 383.

Schlissel, M. S., and Brown, D. D. (1984). Cell 37, 903 - 913.

Schneider, C., Newman, R. A., Sutherland, D. R., Asser, U., and Greaves, M. F. (1982). J. Biol. Chem. 257, 10766 - 10769.

Scholer, H. R., and Gruss, P. (1984). Cell 36, 403 - 411.

Schutzbank, T., Robinson, R., Oren, M., and Levine, A. J. (1982). Cell 30, 481 - 490.

Schwyzer, M., Weil, R., Frank, G., and Zuber, H. (1980). J. Biol. Chem. 255, 5627 - 5634.

Scott, M. R. D., Westphal, K - H., and Rigby, P. W. J. (1983). Cell 34, 557 - 567.

Scott, W. A., and Wigmore, D. J. (1978). Cell 15, 1511 - 1518.

Segall, J., Matsui, T., and Roeder, R. G. (1980). J. Biol. Chem. 255, 11986 - 11991.

Segeawa, M., Sugano, S., and Yamaguchi, N. (1980). J. Virol. 35, 320 - 330.

Sentenac, A., and Hall, B. D. (1982). in The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression. (eds. Strathern, J. N., Jones, E. W., and Broach, J. R.). Cold Spring Harbor Laboratory, New York. pp. 561 - 606.

Shalloway, D., Kleinberger, T., and Livingston, D. M. (1980). Cell 20, 411 - 422.

Sharp, P. A. (1983). Nature 301, 471 - 472.

Sharp, P. A., Sugden, B., and Sambrook, J. (1973). Biochemistry 12, 3055 - 3063.

Shaw, S. B., and Tegtmeyer, P. (1981). Virology 115, 88 - 96.

Shenk, T. E., Carbon, J., and Berg, P. (1976). J. Virol. 18, 664 - 671.

Shortle, D. R., Margolskee, R. F., and Nathans, D. (1979). Proc. Natn. Acad. Sci. USA. 76, 6128 - 6131.

Shyamala, M., Atcheson, C. L., and Kasamatsu, S. (1983). J. Virol. 43, 721 - 729.

Siebenliest, U., and Gilbert, W. (1980). Proc. Natn. Acad. Sci. USA. 77, 122 - 126.

Siebenliest, U., Simpson, R. B., and Gilbert, W. (1980). Cell 20, 269 - 281.
Silver, L. M., and Elgin, S. C. R. (1977). Cell 11, 971 - 983.

Simons, S. S., Martinez, H. M., Garcea, R. L., Baxter, J. D., and Tomkins, J. M. (1976). J. Biol. Chem. 251, 334 - 343.

Singer, M. F. (1982). Cell 28, 433 - 434.

Slattery, E., Dignam, J. D., Matsui, Y., and Roeder, R. G. (1983). J. Biol. Chem. 258, 5955 - 5959.

Smith, D. R., Jackson, I. J., and Brown, D. D. (1984). Cell 37, 645 - 652.

Smith, H. O. (1980). Methods in Enzymology Volume 65, (eds. Grossman, L., and Moldave, K.; Academic Press, New York.) pp.371 - 380.

Solnick, D. (1981). Cell 24, 135 - 144.

Soprano, K. J., Dev, V. G., Croce, C. M., and Baserga, R. (1979). Proc. Natn. Acad. Sci. USA. 76, 3885 - 3889.

Soprano, K. J., Galanti, N., Jonak, G. J., McKercher, S., Pipas, J. M., Peden, K. W., and Baserga, R. (1983). Mol. Cell. Biol. 3, 214 - 219.

Soprano, K. J., Jonak, G. J., Galanti, N., Floros, J., and Baserga, R. (1981). Virology 109, 127 - 136.

Soprano, K. J., Rossini, M., Croce, C., and Baserga, R. (1980). Virology 102, 317 - 326.

Southern, E. M. (1975). J. Mol. Biol. 98, 503 - 515.

Spillman, T., Giacherio, D., and Hager, L. P. (1979). J. Biol. Chem. 254, 3100 - 3104.

Squires, C. L., Lee, F. D., and Yanofsky, C. (1975). J. Mol. Biol. 92, 93 - 111.

Stahl, H., Bauer, M., Knippers, R. (1983). Eur. J. Biochem. 134, 55 - 61.

Stahl, H., and Knippers, R. (1983). J. Virol. 47, 65 - 76.

Staufenbiel, M., and Deppert, W. (1983). Cell 33, 173 - 181.

Stein, G. S., and Stein, J. L. (1978). Methods in Cell Biology 14, 379 - 385.

Stillman, D. J., and Geiduschek, E. P. (1984). EMBO J. 3, 847 - 853.

Storti, R. V., Scott, M. P., Rich, A., and Pardue, M. L. (1980). Cell 22, 825 - 834.

Subramanian, K. N., and Shenk, T. (1978). Nucleic Acids Res. 5, 3635 - 3642.

Sugano, S., Yamaguchi, N., and Simojgo, H. (1982). J. Virol. 41, 1073 - 1075.

Sutcliffe, J. G., Milner, R. J., Bloom, F. E., and Lerner, R. A. (1982). Proc. Natn. Acad. Sci. USA. 79, 4942 - 4946.

Sutcliffe, J. G., Milner, R. J., Gottesfeld, J. M., and Lerner, R. A. (1984). Nature 308, 237 - 241.

Tack, L. C., and De Pamphilis, M. (1983). J. Virol. 48, 281 - 295.

Takeda, Y., Ohlendorf, D. H., Anderson, W. F., and Matthews, B. W. (1983). Science 221, 1021 - 1026.

Takemoto, K. K., Malmgren, R. A., and Habel, K. (1966). Virology 28, 485 - 488.

Tchurikov, N. A., Naumova, A. K., Zelentsova, E. S., and Georgiev. (1982). Cell 28, 365 - 373.

Tegtmeyer, P. (1972). J. Virol. 10, 591 - 598.

Tegtmeyer, P., and Anderson, B. (1981). Virology 115, 67 - 74.

Tegtmeyer, P., Anderson, B., Shaw, S. B., and Wilson, V. G. (1981). Virology 115, 75 - 87.

Tegtmeyer, P., Lewton, B. A., DeLucia, A. L., Wilson, V. G., and Ryder, K. (1983). J. Virol. 46, 151 - 161.

Tegtmeyer, P., Rundell, K., and Collins, J. K. (1977). J. Virol. 21, 647 - 657.

Tegtmeyer, P., Schwartz, M., Collins, J. K., and Rundell, K. (1975). 16, 168 - 178.

Tenen, D. G., Garewal, H., Haines, L. L., Hudson, J., Woodard, V., Light, S., and Livingston, D. M. (1977). Proc. Natn. Acad. Sci. 74, 3645 - 3749.

Tenen, D. G., Haines, L. L., Livingston, D. M. (1982). J. Mol. Biol. 157, 473 - 492.

Tenen, D. G., Taylor, T. S., Haines, L. L., Bradley, M. K., Martin, R. G., and Livingston, D. M. (1983). J. Mol. Biol. 168, 791 - 808.

Thomas, P. S. (1980). Proc. Natn. Acad. Sci. USA. 77, 5201 - 5205.

Thomas, R., Kaplan, L., Reich, N., Lane, D. P., and Levine, A. J. (1982). Virology 131, 502 - 517.

Thummel, C. S., Burgess, T. L., and Tjian, R. (1981 a). J. Virol. 37, 683 - 697.

Thummel, C. S., Tjian, R., and Grodzicker, T. (1981 b). Cell 23, 825 - 836.

Thummel, C., Tjian, R., Shiu - Lok Hu, and Grodzicker, T. (1983). Cell 33, 455 - 464.

Tjian, R. (1978). Cell 13, 165 - 179.

Tjian, R. (1979). Cold Spring Harbor Symp. Quant. Biol. 43, 665 - 662.

Tjian, R., Fey, G., and Graessman, A. (1978). Proc. Natn. Acad. Sci. USA. 75, 1279 - 1283.

Tjian, R., and Robbins, A. (1979). Proc. Natn. Acad. Sci. USA. 76, 610 - 614.

Tjian, R., Robbins, A., and Clark, R. (1979). Cold Spring Harbor Symp. Quant. Biol. 44, 103 - 111.

Tooze, J. (ed.) The Molecular Biology of the Tumor Viruses, 2nd edition (1981). Cold Spring Harbor Laboratory, New York.

Topp, W. C. (1980). J. Virol. 33, 1208 - 1210.

Towbin, H., Staehelin, T., and Gordon, J. (1979). Proc. Natn. Acad. Sci. USA. 76, 4350 - 4354.

Treisman, R., Green, R., and Maniatis, T. (1983). Proc. Natn. Acad. Sci. USA. 80, 7428 - 7432.

Tsai, Y. T., Tsai, M. J., Kops, L. E., Minghetti, P. P., and O'Malley, B. W. (1981). J. Biol. Chem. 256, 13055 - 13059.

Tu, C. P. D., and Cohen, S. (1980). Gene 10, 177 - 183.

Tuite, M. F., Dobson, M. J., Roberts, N. A., King, R. M., Burke, D. C., Kingsman, S. M., and Kingsman, A. J. (1982). EMBO J. 1, 603 - 608.

Twigg, A. J., and Sherrat, D. (1980). Nature 283, 216 - 218.

Ucker, D. S., Ross, S. R., and Yamamoto, K. R. (1981). Cell 27, 257 - 266.

Valenzuela, P., Medina, A., Rutter, W. J., Ammerer, G., and Hall, B. D. (1982). Nature 298, 347 - 350.

Van Roy, F., Fransen, L., and Fiers, W. (1984). Mol. Cell. Biol. 4, 232 - 239.

Varshavsky, A. J., Sundin, O. H., and Bohn, M. J. (1978). Nucleic Acids Res. 5, 3469 - 3478.

Varshavsky, A. J., Sundin, O. H., and Bohn, M. (1979). Cell 16, 453 - 466.

Wahl, G. M., Stern, M., and Stark, G. R. (1979). Proc. Natn. Acad. Sci. USA. 76, 3683 - 3687.

Walker, M. D., Edlund, T., Boulet, M., and Rutter, W. J. (1983). Nature 306, 557 - 561.

Walsh, C. (1979). Enzymatic Reaction Mechanisms. (Freeman and Co.)

Walter, G., Hutchison, M. A., Hunter, T., and Eckhart, W. (1982). Proc. Natn. Acad. Sci. USA. 79, 4025 - 4029.

Wang, J. C., Barkley, M. D., and Bourgeois, S. (1974). Nature 251, 247 - 248.

Wall, R. (1980). TIBS 5, 325 - 327.

Watts, F., Castle, C., and Beggs, J. D. (1983). EMBO J. 2, 2085 - 2091.

Weideli, H., Brack, C. H., and Gehring, W. J. (1980). Proc. Natn. Acad. Sci. USA. 77, 3773 - 3777.

Weideli, H., Schedl, P., Artavanis - Tsakonas, S., Steward, R., Yuan, R., and Gehring, W. J. (1978). Cold Spring Harbor Symp. Quant. Biol. 42, 693 - 700.

Weiffenbach, B., Rogers, D. T., Haber, J. E., Zoller, M., Russell, D. W., and Smith, M. (1983). Proc. Natn. Acad. Sci. USA. 80, 3401 - 3405.

Weiher, H., Konig, M., and Gruss, P. (1983). Science 219, 626 - 631.

Weintraub, H., and Groudine, M. (1976). Science 193, 848 - 856.

Weisbrod, S. (1982). Nature 297, 289 - 295.

Weisbrod, S., Groudine, M., and Weintraub, H. (1980). Cell 19, 289 - 301.

Weisinger, G., and Sachs, L. (1983). EMBO J. 2, 2103 - 2107.

Welshons, W. V., Liebermann, M. E., and Gorski, J. (1984). Nature 307, 747 - 749.

Wilbur, W. J., and Lipman, D. J. (1983). Proc. Natn. Acad. Sci. USA. 80, 726 - 730.

Wilson, M. C., Fraser, N. W., and Darnell, J. E. (1979). Virology 94, 175 - 184.

Wilson, V. G., Tevethia, M. J., Lewton, B. A., and Tegtmeyer, P. A. (1982). J. Virol. 44, 458 - 466.

Williams, J. G., Hoffman, R., and Penman, S. (1977). Cell 11, 901 - 907.

Winter, R. B., Berg, O. G., and von Hippel, P. H. (1981). Biochemistry 20, 6961 - 6977.

Winter, R. B., and von Hippel, P. H. (1981). Biochemistry 20, 6948 - 6960.

Wirth, T., Gloggler, K., Baumruker, T., Schmidt, M., and Horak, I. Proc. Natn. Acad. Sci. USA. 80, 3327 - 3330.

Wold, B. J., Klein, W. H., Hough - Evans, B. R., Britten, R. J., and Davidson, E. H. (1978). Cell 14, 941 - 950.

Wu, C. (1980). Nature 286, 854 - 860.

Wu, C. (1984). Nature 309, 229 - 234.

Wu, H. M., and Crothers, D. M. (1984). Nature 308, 509 - 513.

Yamamoto, K. R., and Alberts, B. M. (1976). Ann. Rev. Biochem. 45, 721 - 746.

Yanofsky, C. (1981). Nature 289, 751 - 758.

Youderian, P., Vershon, A., Bouvier, S., Sauer, R. T., and Susskind, M. M. (1983). Cell 35, 777 - 783.

Young, B. D., Birnie, G. D., and Paul, J. (1976). Biochemistry 15, 2823 - 2829.

Yuan, R. (1981). Ann. Rev. Biochem. 50, 285 - 315.

Ziff, E. B. (1980) Nature, 287, 491 - 499.