Current Topics in Microbiology and Immunology

93

Editors

W. Henle, Philadelphia · P.H. Hofschneider, Martinsried
H. Koprowski, Philadelphia · F. Melchers, Basle · R. Rott,
Gießen · H.G. Schweiger, Ladenburg/Heidelberg
P.K. Vogt, Los Angeles

Initiation Signals in Viral Gene Expression

Edited by Aaron J. Shatkin

With 30 Figures

Springer-Verlag
Berlin Heidelberg New York 1981

Aaron J. Shatkin, Ph.D. Head,
Laboratory of Molecular Virology
Roche Institute of Molecular Biology
Nutley, New Jersey 07110
U.S.A.

ISBN-13: 978-3-642-68125-7 e-ISBN-13: 978-3-642-68123-3
DOI: 10.1007/978-3-642-68123-3

Library of Congress Catalog Card Number 15-12910.

Typesetting: Fotosatz Service Weihrauch, Würzburg
Printing and binding: Universitätsdruckerei H. Stürtz AG, Würzburg
2121/3321-543210

Table of Contents

Indexed in Current Contents

Introduction
Elucidating Mechanisms of Eukaryotic Genetic Expression by Studying Animal Viruses

AARON J. SHATKIN*

Eukaryotic genetic expression is carefully regulated. Normal cell growth and division, tissue differentiation, and organism development all depend on a strictly ordered progression of specific events. Perturbation of the control of these processes, for example by exposure to harmful chemicals or infection with viruses leads to aberrant forms of metabolism, often resulting in malignancies and cell death. One of the most challenging problems in biology is to define at the molecular level the mechanisms that govern gene function in higher organisms, including ultimately man. This goal serves to unify the diverse efforts of many investigators, whether studying the precise patterns of embryogenesis, the loss of control that occurs during neoplastic growth or the redirection of biosynthetic pathways in virus-infected cells.

Recently there has been remarkable and exciting progress toward understanding the molecular biology of eukaryotic expression. Much of this rapidly increasing new information has come from studies of animal virus systems. Just as investigations of the relatively simple, rapidly assayed, and easily manipulated bacteriophages lead to basic discoveries about prokaryotic cells, analyses of animal viruses and their interactions with host cells have provided fundamental information about how eukaryotic nucleic acids are organized for regulated replication, transcription, and translation. For example, the small genome of SV_{40}, like cellular DNA in chromatin, is associated with histones to form nucleosomal arrays (*Griffin* 1975). In SV_{40}-infected cells these viral "minichromosomes" are replicated and transcribed by cellular enzymes, thus providing an excellent experimental system that has been effectively exploited as detailed in the review by *Tjian.*

Other studies on SV_{40} (*Lavi* and *Groner* 1977; *Aloni* et al. 1977) and on adenovirus transcription earlier (*Chow* et al. 1977; *Klessig* 1977; *Berget* et al. 1977) provided the first evidence for splicing of mRNA precursors. With this striking finding came a new appreciation of eukaryotic genes as structures consisting of highly specific functional subdivisions, not simply blocks of contiguous nucleotides. This organization is most readily apparent at the level of pre-mRNA cutting and ligation, processes that result in retention of some RNA sequences (exons) in gene transcripts and removal of others (introns). As described in the paper by *Flint,* reconstitution of stretches of pre-RNA by splicing in different ways can enormously increase coding potential without genome alteration.

Discovery of this fundamental mechanism of expression depended partly on another recently described characteristic feature of eukaryotes, the "cap" structure, $m^7G(5')ppp(5')N$ (Figure 1). It is present on the 5'-termini of most viral and cellular messengers and their precursors (*Shatkin* 1976). The structure, synthesis, and functional consequences of capping have been reviewed in detail recently (*Filipowicz* 1978; *Banerjee*

*Roche Institute of Molecular Biology, Nutley, New Jersey 07110, USA

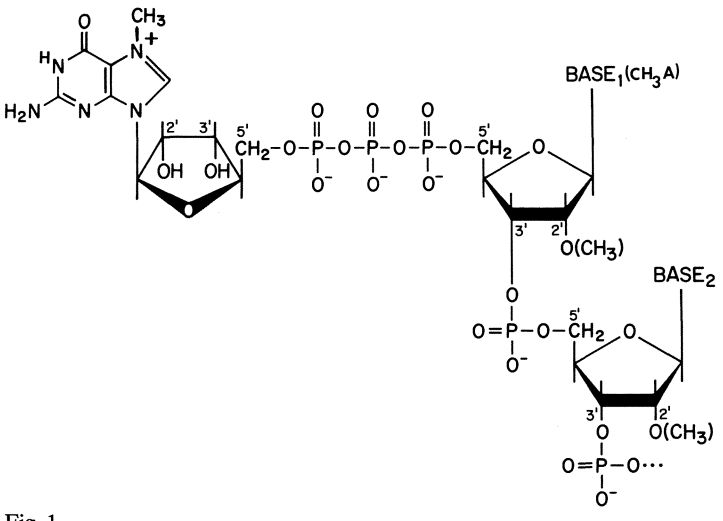

Fig. 1

1980). The steps involved in cellular mRNA capping (*Venkatesan* and *Moss* 1980) are probably the same as those deciphered by studying the enzymatic activities associated with human reovirus type 3 (*Furuichi* et al. 1976). These steps and the enzymes that catalyze them include:

(1) $\overset{*}{p}ppX + pppY \rightarrow \overset{*}{p}ppXpY + PP_i$	RNA polymerase
(2) $\overset{*}{p}ppXpY \rightarrow \overset{*}{p}pXpY + P_i$	Nucleotide phosphohydrolase
(3) $pppG + \overset{*}{p}pXpY \rightarrow G\overset{*}{p}ppXpY + PP_i$	Guanylyl transferase
(4) $G\overset{*}{p}ppXpY$ + S-adenosylmethionine $\rightarrow m^7G\overset{*}{p}pX^mpY$	Methyl transferases

From this scheme it is evident that the cap is added early in transcription, probably during initiation or very soon after nascent chains are started. The cap on nuclear transcripts is retained during conversion of pre-mRNA to mature mRNA. Like the poly(A) detected on the 3′-termini of vaccinia virus mRNA (*Kates*, 1970) and subsequently on most eukaryotic messengers, the cap provides a unique marker for mRNA 5′-termini. With the possibility of defining precisely the 5′-ends of mRNAs came a renewed interest in studying mechanisms of initiation of eukaryotic transcription. The resulting recent findings are reviewed in the chapter by *Shenk*.

Caps also were found to promote the formation of protein synthesis initiation complexes (*Both* et al. 1975), and a cap binding protein that stimulates capped mRNA translation was purified from initiation factor preparations by affinity chromatography on m^7 GDP-Sepharose (*Sonenberg* et al. 1979, 1980). The importance of the 5′ end of mRNA as the site of ribosome attachment is an essential element of the scanning model recently proposed to explain how ribosomes recognize a single initiation site in mRNA (*Kozak* 1978). New evidence concerning monocistronic function of eukaryotic mRNAs is described and the scanning model of protein synthesis initiation is evaluated in the paper by *Kozak*.

Another particularly intriguing finding related to caps is the priming of influenza transcription by capped heterologous mRNAs. Although purified influenza virus contains RNA polymerase associated with its RNA genome, virus replication is dependent on cellular DNA-dependent RNA polymerase II activity (*Lamb* and *Choppin* 1977). Furthermore, influenza mRNAs isolated from infected cells are capped (*Krug* et al. 1976), although capping activity is not associated with the virion transcriptase (*Plotch* et al. 1978). These paradoxes were resolved by the finding that influenza transcriptase is primed, and the cap and approximately 10–15 adjacent nucleotides of the primer mRNA are incorporated into the viral transcripts as preformed 5′-ends. Chimeric influenza messages consisting of viral coding sequences and cellular 5′-termini have also been isolated from infected cells (*Dhar* et al. 1980). While primer-dependence may not be a general property of transcriptases, recent studies of the mRNA synthesizing systems present in two other RNA viruses, human reovirus (*Yamakawa* et al. 1981) and insect cytoplasmic polyhedrosis virus (*Furuichi* 1981), indicate that short 5′-terminal oligonucleotides are made in great molar excess relative to full-length mRNAs. It remains to be determined if these are prematurely terminated, abortive transcripts or if they represent initiator oligonucleotides that function as primers for mRNA synthesis. RNA priming and the mechanism of initiation of influenza mRNA formation is described in the chapter by *Krug;* both the early background work and more recent experiments are summarized.

The last paper in the collection is a comprehensive review of nucleotide sequences involved in the formation and replication of defective interfering particles of animal viruses. Defective particles are generated by members of almost all virus groups. Exactly how they arise is of considerable intrinsic interest. In addition, there is a strong possibility that defective interfering particles are important elements in virus disease, especially in persistent infections. The long-term interactions that may occur between virus- and host-directed pathways of expression are considered by *Perrault.*

It seems likely that future investigations of animal virus systems by such powerful new techniques as recombinant DNA technology and monoclonal antibody production will provide additional insights into eukaryotic gene expression. Hopefully, the following review articles will help stimulate some of these studies.

References

Aloni Y, Dhar R, Laub O, Horowitz M, Khoury G (1977) Novel mechanism for RNA maturation: the leader sequences of simian virus 40 mRNA are not transcribed adjacent to the coding sequences. Proc Natl Acad Sci USA 74:3686–3690

Banerjee AK (1980) 5′-Terminal cap structure in eucaryotic messenger ribonucleic acids. Microbiol Reviews 44:175–205

Berget SM, Moore C, Sharp PA (1977) Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci USA 74:3171–3175

Both GW, Furuichi Y, Muthukrishnan S, Shatkin AJ (1975) Ribosome binding to reovirus mRNA in protein synthesis requires 5′-terminal 7-methylguanosine. Cell 6:185–195

Chow LT, Gelinas RE, Broker TR, Roberts RJ (1977) An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. Cell 12:1–8

Dhar R, Chanock RM, Lai C-J (1980) Non-viral oligonucleotides at the 5′ terminus of cytoplasmic influenza viral mRNA deduced from cloned complete genomic sequences. Cell 21:495–500

Filipowicz W (1978) Function of the 5′-terminal ^{7}mG cap in eukaryotic mRNA. FEBS Lett 96:1–11

Furuichi Y (1981) Allosteric stimulatory effect of S-adenosylmethionine on the RNA polymerase in cytoplasmic polyhedrosis virus. J Biol Chem 256:483–493

Furuichi Y, Muthukrishnan S, Tomasz J, Shatkin AJ (1976) Mechanism of formation of reovirus mRNA 5′-terminal blocked and methylated sequence, ^{7}mGpppGmC. J Biol Chem 251:5043–5053

Griffith JD (1975) Chromatin structure: deduced from a minichromosome. Science 187:1202–1203

Kates J (1970) Transcription of the vaccinia virus genome and the occurrence of polyriboadenylic acid sequences in messenger RNA. Cold Spring Harbor Symp Quant Biol 35:743–752

Klessig DF (1977) Two adenovirus mRNAs have a common 5′ terminal leader sequence encoded at least 10 kb upstream from their main coding regions. Cell 12:9–21

Kozak M (1978) How do eucaryotic ribosomes select initiation regions in messenger RNA? Cell 15:1109–1123

Krug RM, Morgan MA, Shatkin AJ (1976) Influenza viral mRNA contains internal N^{6}-methyladenosine and 5′-terminal 7-methylguanosine in cap structures. J Virol 20:45–53

Lamb RA, Choppin PW (1977) Synthesis of influenza virus polypeptides in cells resistant to α-amanitin: evidence for the involvement of cellular RNA polymerase II in virus replication. J Virol 23:816–819

Lavi S, Groner Y (1977) 5′-Terminal sequences and coding region of late simian virus 40 mRNAs are derived from noncontiguous segments of the viral genome. Proc Natl Acad Sci USA 74: 5323–5327

Plotch SJ, Tomasz J, Krug RM (1978) Absence of detectable capping and methylating enzymes in influenza virions. J Virol 28:75–83

Shatkin AJ (1976) Capping of eucaryotic mRNAs. Cell 9:645–653

Sonenberg N, Rupprecht KM, Hecht SM, Shatkin AJ (1979) Eukaryotic mRNA cap binding protein: purification by affinity chromatography on Sepharose-coupled m^{7}GDP. Proc Natl Acad Sci USA 76:4345–4349

Sonenberg N, Trachsel H, Hecht S, Shatkin AJ (1980) Differential stimulation of capped mRNA translation in vitro by cap binding protein. Nature 285:331–333

Venkatesan S, Moss B (1980) Donor and acceptor specificities of HeLa cell mRNA guanylyltransferase. J Biol Chem 255:2835–2842

Yamakawa M, Furuichi Y, Nakashima K, LaFiandra AJ, Shatkin AJ (1981) Excess synthesis of viral mRNA 5′-terminal oligonucleotides by reovirus transcriptase. J Biol Chem 256:6507–6514

Regulation of Viral Transcription and DNA Replication by the SV40 Large T Antigen

ROBERT TJIAN*

1 Introduction

Simian Virus 40 (SV40) provides a particularly useful model for studying gene regulation at the molecular level because it is one of the simplest and most thoroughly characterized (*Tooze* 1980) viral systems known to undergo a well-defined developmental cycle in mammalian cells. The entire nucleotide sequence of the 5243 base pair genome of SV40 has been determined (*Reddy* et al. 1978; *Fiers* et al. 1978) and important regions such as the origin of replication, transcriptional promoters, structural gene boundaries, splice junctions, and positions of mutations have been located. The small double-stranded circular genome of SV40 encodes only five or six genes and consequently relies in large part on the host cell machinery to carry out complex regulated processes such as transcription and replication. During lytic infection of monkey cells, SV40 undergoes a temporal program of gene expression that is controlled in part by viral coded proteins (*Tegtmeyer* 1972; *Cowan* et al. 1973; *Tegtmeyer* et al. 1975; *Reed* et al. 1976). Immediately after infection, the SV40 A and F genes, encoding the large T (*Black* et al. 1963; *Tegtmeyer* 1975) and small t (*Prives* et al. 1978; *Crawford* et al. 1978; *Sleigh* et al. 1978) antigens respectively, are expressed. Transcription of these early viral genes by the host RNA polymerase II originates from a region of the genome located at approximately 0.67–0.70 fractional units on the conventional genome map (Fig. 1) (*Sambrook* et al. 1973; *Khoury* et al. 1973; *Jackson* and *Sugden* 1972; *Khoury* et al. 1975). During this early phase of the lytic cycle, SV40 DNA replication does not occur (*Tegtmeyer* 1972) and expression of the structural

*Department of Biochemistry, University of California, Berkeley, California 94720, USA

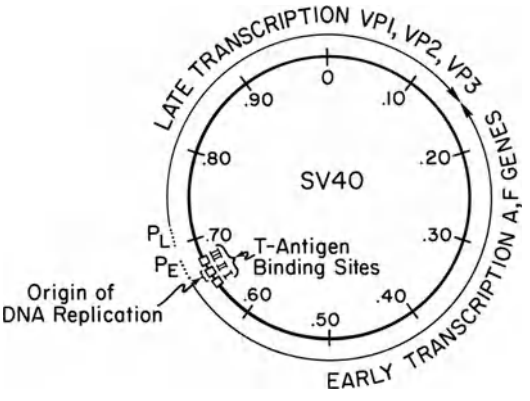

Fig. 1. Map of the SV40 genome depicting the origin of replication, T-antigen binding sites, approximate starts of early and late transcription and coding regions for viral genes

capsid proteins is detectable only at very low levels (*Cowan* et al. 1973; *Alwine* et al. 1977). However, after the intracellular concentration of large T antigen reaches a certain threshold (*Graessmann* et al. 1978), viral DNA replication is initiated. This A-gene-dependent synthesis of SV40 DNA originates from a unique position on the genome (*Danna* and *Nathans* 1972; *Fareed* et al. 1972) at 0.67 map units and proceeds bidirectionally (*Jaenisch* et al. 1971; *Kelly* and *Nathans* 1977). The discontinuous strand synthesis during SV40 DNA replication (*Fareed* and *Salzman* 1972) is thought to be catalyzed by the host DNA polymerase and is primed by short RNA sequences located at intervals along the DNA template (*Su* and *DePamphilis* 1976). After the onset of viral DNA replication, transcription of the early genes is reduced while late gene RNA synthesis becomes activated to a high level. Analysis of temperature sensitive mutants in the A gene revealed that, at the restrictive temperature, early viral mRNAs and proteins are overproduced (*Cowan* et al. 1973; *Tegtmeyer* 1975; *Reed* et al. 1976; *Alwine* et al. 1977). Thus, it appears that the product of the SV40 A gene, large T antigen, plays a pivotal role in viral gene expression by initiating DNA synthesis and autoregulating early transcription. There is also a wealth of evidence that implicates large T antigen in functions involving the initiation and maintenance of virally induced transformation of nonpermissive cells (*Kimura* and *Dulbecco* 1972; *Tegtmeyer* 1975; *Brugge* and *Butel* 1975; *Martin* and *Chou* 1975; *Osborn* and *Weber* 1975b; *Steinberg* et al. 1978). By contrast, mutations in gene F have only a small and poorly defined effect on viral gene expression during lytic infection of permissive cells (*Sleigh* et al. 1978; *Topp* pers. comm.). There is, however, some evidence that implicates small t antigen in functions involving the initiation of virally induced cellular transformation in certain established cell lines but not in primary cultures or rapidly dividing cells (*Sleigh* et al. 1978; *Bouck* et al. 1978; *Martin* et al. 1979; *Martin* 1981).

A comprehensive general review of papovaviruses has recently been published in the second edition of Tooze's *Molecular Biology of Tumor Viruses*. I will therefore not attempt to cover both the lytic and transformation functions of SV40 tumor antigens. Instead, the purpose of this review will be to analyze critically the various lines of investigation pertaining specifically to the molecular mechanisms by which the large T antigen of SV40

regulates viral transcription and replication. In particular, I will focus on the experimental strategies that have been employed to study the biochemical properties of purified SV40 T antigen.

2 Primary Structure of the SV40 Large T Antigen

The SV40 A gene product is a nuclear protein with an apparent molecular weight of 96,000 as determined by sodium dodecyl sulfate gel electrophoresis (*Tegtmeyer* 1975). However, the primary structure of the T antigen polypeptide deduced from the DNA sequence indicates that its actual molecular weight should be closer to 82,000 (*Fiers* et al. 1978; *Paucha* et al. 1978). Chromatography of a denatured T antigen related protein isolated from an SV40 transformed-cell line confirms that its molecular weight is approximately 80,000 (*Griffin* et al. 1978). The discrepancy between the electrophoretic mobility of T antigen and its actual molecular weight may be due in part to postanslational modifications of the T antigen polypeptide. For instance, the amino terminus of T antigen is known to be acetylated in vivo and there is evidence for phosphorylated residues (*Rundell* et al. 1977; *McCormick* et al. 1980; *Mann* and *Hunter* 1980; *Walter* and *Flory* 1979). However, these documented modifications of T antigen do not appear sufficient to account for its aberrant mobility in sodium dodecyl sulfate gels, suggesting that additional modifications altering the primary structure of the polypeptide may have gone undetected.

3 Purification of T Antigen

The isolation of biochemically useful quantities of active SV40 large T antigen presented a formidable problem because most SV40 infected or transformed cells produce relatively low amounts of the viral protein. For instance, large T antigen represents only 0.01–0.03% of the total soluble protein in lytically infected monkey cells. Consequently, a significant amount of effort has gone towards finding or isolating a convenient source from which the SV40 A gene product can be purified.

Early investigators chose to purify T antigen from SV40 transformed cells such as the human cell line, SV80, because these cells were easy to obtain in large quantities and they produced relatively high amounts of viral early proteins (*Livingston* et al. 1974; *Osborn* and *Weber* 1974). It had been assumed that the 96,000 dalton protein purified from transformed cells would be identical in structure and function to authentic SV40 T antigen produced in lytically infected monkey cells. However, a recent study indicates that the commonly used SV80 cells actually produce a T protein that is defective in at least its viral DNA replication functions (*Gish* and *Botchan*, personal communication). Analysis of SV40 sequences rescued from SV80 cells also revealed a lesion in the carboxyl portion of the protein. In retrospect, the finding that SV80 cells contain a mutant A gene is not surprising. Human cells, which are normally semipermissive for SV40 DNA replication, should contain free viral DNA at a low frequency when transformed by SV40. The fact that SV80 cells are stably transformed but do not contain any free viral DNA implies that

excision and replication of viral DNA is somehow blocked. Inhibition of viral DNA synthesis can occur as a result of mutations in the cellular replication machinery or, alternatively, a defect could have arisen either in the viral origin of replication or in the A gene. In the case of SV80 cells, viral DNA can be efficiently rescued and replicated by providing exogenous wildtype T antigen (*Gish* and *Botchan*, personal communication). This finding indicates that a mutation has occurred that affects the replication functions of T antigen. Thus, the SV80 protein may not be appropriate for studies involving the lytic functions of the A gene product. However, the SV40 A gene product is apparently a multifunctional protein with several distinct activities, some involved in lytic functions and others in transformation and induction of cellular gene expression. If so, partially defective products such as SV80 protein could still be useful for investigating properties of the A gene product that are not directly involved with DNA binding and regulation of lytic processes. Moreover, the SV80 protein also provides a source of a mutant protein that can be used to study specific A gene defects affecting viral DNA replication and transcription (see below).

As an alternative source of protein, cells infected with adenovirus-SV40 hybrid viruses have been used for purification of the SV40 T antigen (*Lewis* et al. 1969; *Black* and *White* 1967; *Lewis* and *Rowe* 1970; *Hassell* et al. 1978; *Tjian* 1978b). One naturally occurring virus designated Ad2+D2 directs the synthesis of a T antigen related protein in quantities 50 times greater than cells either lytically infected or transformed by SV40 (*Tjian* 1978a). Moreover, it is easier to isolate viral protein from cells infected with this hybrid virus because Ad2 inhibits host protein synthesis late in infection and extracts of HeLa cells grown in suspension contain less proteolytic activity than either CV1 or SV80 cells. The D2 protein isolated from cells infected with $Ad2^{+}D2$ is a 107,000 dalton hybrid protein comprised of approximately 10% of an Ad2 protein at its amino terminus and about 90% of the SV40 gene A protein carboxy terminus (*Hassell* et al. 1978). Because the D2 protein lacks the amino terminal portion of the SV40 T antigen molecule, it is possible that its biochemical properties would not accurately reflect those of authentic T antigen. To overcome this problem, and at the same time retain the advantages of overproduction in the adenovirus system, a number of novel adenovirus-SV40 recombinants were recently constructed in vitro (*Thummel* et al. 1981b). These hybrid viruses were designed to overproduce the SV40 A gene product under the control of adenovirus promoters. HeLa cells infected with a novel hybrid virus (AdSVR6) synthesize high levels of a 96,000 dalton T antigen that is structurally indistinguishable from the authentic lytic product made in SV40-infected monkey cells. These hybrid viruses have provided a convenient source for the purification of SV40 T antigen.

A typical procedure for purifying T antigen involves isolating nuclei at pH 6.9 in a low salt buffer, extracting the nuclei with a high ionic strength buffer at pH 8.1 (*Tjian* 1978a), ammonium sulfate fractionation followed by three to four chromatographic columns (*Livingston* et al. 1974; *Tjian* 1978a; *Tjian* et al. 1979). In general, some combination of gel filtration, DEAE sephadex, phosphocellulose and heparin or DNA agarose has given the best results. During purification, proteolysis of the T protein is sometimes observed, especially in the case of CV1 or SV80 cells. However, once T antigen has been purified to near homogeneity, it can be stored at -70 °C in 10–20% glycerol without loss of biological activity for up to 2 years. The yields of T antigen purified to greater than 90% homogeneity from various sources of cells are shown in Table 1.

Table 1. Levels of T antigen production and yields of purified protein from different cells

Source of T antigen	Total protein (%)	T antigen in 10 g of cells (μg)	T antigen actually purified from 10 g of cells (μg)	Yield (%)
SV40 infected CV1 cells	0.01-0.04	50- 200	5-10	5
SV80 cells	0.03-0.06	150- 300	30-50	10
Ad-SVR6 infected Hela cells	0.2 -0.3	1000-1500	150	10
Ad2+D2 infected Hela cells	1.0 -1.5	5000-8000	700	15-20

The amount of T antigen related proteins produced by different sources was calculated by quantitating anti-T immunoprecipitates from extracts of labeled as well as unlabeled cells. The values in yields of purified T antigen were determined from an average of many preparations isolated to near homogeneity by published procedures

4 DNA Binding Properties

It was suspected that the SV40 T antigen might be a DNA binding protein even before there was any direct biochemical evidence to support this idea. Its nuclear location in the cell (*Black* et al. 1963) as well as its involvement in viral DNA synthesis (*Tegtmeyer* 1972) and transcription (*Tegtmeyer* 1975; *Reed* et al. 1976) led workers in the field to hypothesize that the SV40 A gene encoded a regulatory protein that interacts directly with DNA to influence gene expression. The remainder of this review will largely be devoted towards documenting the experiments that have confirmed these early predictions.

The general affinity of T antigen for double-stranded DNA was first demonstrated by DNA cellulose chromatography (*Carrol* et al. 1974). Shortly thereafter, the possibility that wild type T antigen from lytically infected cells may bind selectively to the origin region of the SV40 genome was demonstrated by electron microscopy of antibody-T antigen complexes isolated from infected cell lysates (*Reed* et al. 1975). Using a filter binding assay, it was subsequently shown that partially purified SV80 T protein binds with equal affinity to SV40 DNA and a number of phage DNAs (*Jessel* et al. 1975). Moreover, binding of SV40 DNA restriction fragments by the SV80 protein suggested preferential binding to several different regions on the viral genome (*Jessel* et al. 1976). These initial filter binding studies were difficult to interpret because they involved impure preparations of SV80 protein. Moreover, the binding of isolated restriction fragments often correlated with the size of the DNA fragment rather than sequence.

In order to increase the sensitivity and specificity of the filter binding assay, pure D2 protein was used to protect specific sequences of uniformly labelled SV40 DNA from digestion by pancreatic DNAse I (*Tjian* 1978a). These DNAse protection experiments established that specific binding of D2 protein occurred at three closely spaced sites on

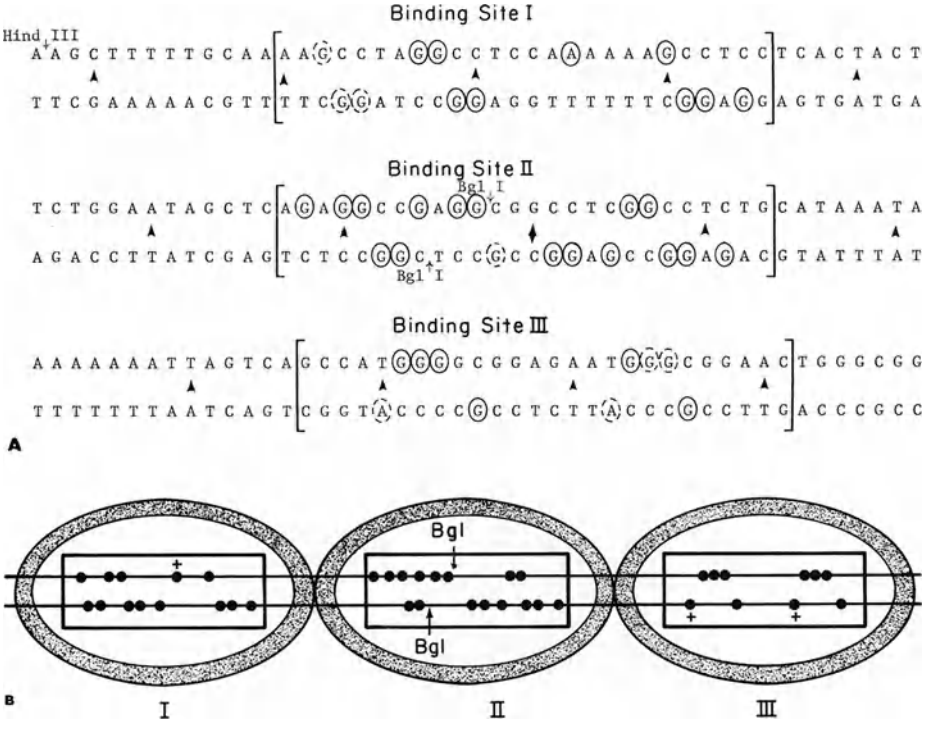

Fig. 2. A,B. SV40 T antigen binding sites. *A* Sequence of nucleotides containing the triple binding sites. *B* A schematic representation of binding sites determined by DNAse protection and dimethyl sulfate protection. The nucleotide sequence in the region overlapping the SV40 origin of DNA replication was reproduced from Fiers, et al. 1978 and Reddy, et al. 1978. The numbered nucleotides are according to the BBB system (*Tooze* 1980). Arrows indicate cleavage sites of restriction enzyme Bgl I. The three tandem binding sites as defined by dimethylsulfate protection experiments are shown in brackets (A) and boxes (B). The ellipses (B) depict the region of DNA sequence protected by T antigen from DNAse digestion. Guanine residues specifically shielded from methylation by T antigen are encircled by solid and broken lines (A) and solid dots (B). Adenine residues enhanced for methylation by T antigen are encircled by broken lines (A) and solid dots with (+) (B). Double arrowhead equals position of first nucleotide and single arrowhead indicates ten-nucleotide spacing

SV40 DNA that extended over a region of 120–140 bp containing the origin of viral DNA replication. Binding of the D2 protein to its three recognition sites occurs in a sequential manner with the strongest affinity at site I and progressively lower affinity to sites II and III (Fig. 2). In addition, dimethylsulfate methylation studies revealed that the D2 protein (*Tjian* 1978b) makes symmetrical contacts with the binding sequences almost exclusively in the major groove of the DNA duplex (Fig. 2).

Papovaviruses such as polyoma and BK contain origin DNA sequences that exhibit a high degree of homology to the SV40 origin (*Soeda* et al. 1979; *Friedmann* et al. 1979; *Fiers* et al. 1978; *Reddy* et al. 1978; *Seif* et al. 1979; *Yang* and *Wu* 1978). Therefore, the ability of the SV40 T antigen to bind origin DNA sequence from these other papovaviruses has also been tested. DNAse protection experiments reveal that the D2 protein binds pre-

ferentially to multiple binding sites in the region of BK and polyoma virus DNA that are analogous to the SV40 origin sequences (*Tjian* 1980). Although the specificity and general features of the interactions at these binding sites are similar, there are some significant differences that probably reflect minor alterations in the nucleotide sequence of the three viruses. For example, SV40 sites I, II, and III bind in a sequential manner while polyoma virus sites I and II bind with equal efficiency and site III binds less well. By contrast, site I of BKV binds most efficiently and sites II and III bind with similar affinities. Presently, it is not clear how these protein-DNA interactions relate to the ability of the SV40 T antigen to carry out replication functions at the different papovavirus origin sequences.

A quantitative filter binding assay recently revealed that purified SV80 protein binds to the specific SV40 origin DNA sequences approximately 20 times less efficiently than D2 protein (*Myers* et al. 1981b). However, the nonspecific DNA binding capabilities of the SV80 protein and D2 protein were very similar. This weakened binding of SV80 protein for origin sequences may help to explain the inability of earlier studies to demonstrate specific DNA binding. Several studies have confirmed that binding of SV80 protein, although weak, occurs predominantly at the origin sequence (*Shalloway* et al. 1980; *Myers*, personal communication). A recent study using an exonuclease III assay reported a fourth T antigen binding site located adjacent to site III (*Shalloway* et al. 1980). However, the SV80 protein was used in these experiments and extremely high concentrations of this defective binding protein were required (500 : 1 molar ratio of protein to DNA) in order to detect binding sites that are protected from exonuclease digestion. Consequently, the assays were performed under conditions that are known to allow nonspecific binding of T antigen to many sites on the SV40 DNA. Moreover, this additional site has not been detected by DNAse footprinting or DMS protection using either SV80 protein or D2 protein under conditions that promote only specific DNA binding (*Tjian* 1978b; *Myers* et al. 1981a).

It is possible that neither the mutant SV80 protein nor the truncated D2 protein bind DNA with the same specificity as wild-type SV40 T antigen. To settle this issue, intact wild-type SV40 T antigen was purified from the newly constructed hybrid viruses (*Thummel* et al. 1981b) and shown by filter binding, DNAse protection, and DMS methylation studies to bind SV40 DNA origin sequences with a specificity identical to that of D2 protein (*Myers* et al. 1981a). Thus, it appears that the DNA binding properties of SV80 protein are aberrant and do not accurately reflect either the strength or the specificity of DNA binding by wild-type SV40 T antigen. Moreover, it is clear that the amino-terminal portion of T antigen is not required for specific DNA binding because both D2 and intact T antigen bind to the triple recognition sites in the same manner.

The size of SV40 DNA fragments protected by T antigen (30–35 bp) from DNAse digestion and the symmetrical contacts made between protein and DNA suggested that the active binding promoter is likely to be either a dimer or tetramer (*Tjian* 1978b; *Myers* et al. 1981a). Moreover, a number of studies reported oligomeric forms of native T antigen by gel filtration and velocity sedimentation (*Osborn* and *Weber* 1974; *Tegtmeyer* pers. comm., *Tjian,* unpublished). Direct electron microscopic examination of purified D2 protein in free form reveals monomers, tetramers, and dodecamers (*Myers* et al. 1981c). However, the predominant species bound to SV40 DNA appears to be a tetramer (Fig. 3).

Quantitative studies of T-antigen binding suggested that interactions at the three sites is not only sequential but may also be cooperative (*Myers* et al. 1981a). This idea was tested

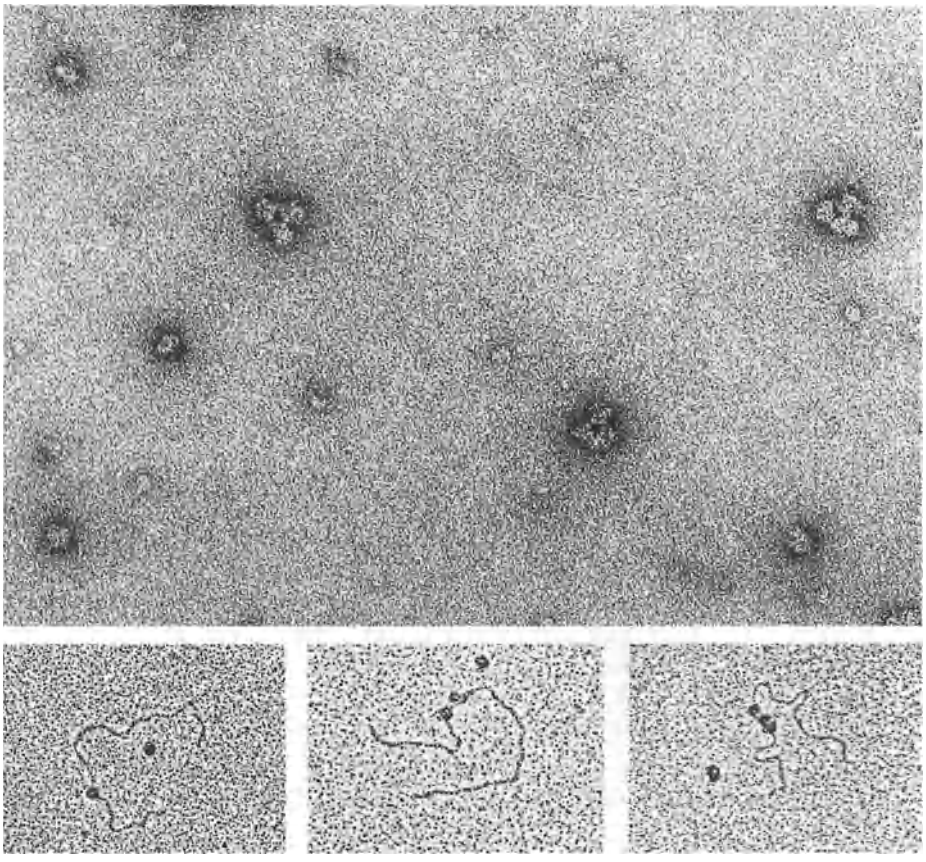

Fig. 3. Electron micrographs of purified D2 protein. *Top Panel:* A preparation of D2 protein negatively stained with uranyl formate. Small particles (monomers), medium size particles (tetramers), and large triangular shaped structures (dodecamers) are seen. *Lower Panels*: Micrographs showing D2 protein bound to the SV40 origin of DNA replication. The origin sequence is approximately 32% from one end of the DNA fragment. Either one, two or three particles of tetramer size are seen bound specifically to the origin region of SV40 DNA. The D2 concentration was 1 µg/ml and protein DNA complexes were visualized after rotary-shadowing

by measuring the quantitative differences between the efficiency of T antigen to bind cloned DNA sequences lacking specific portions of binding site I, II, and III. Removal of sites II and III has no detectable influence on the binding of T antigen to site I, the strongest binding site ($K_d = 6 \times 10^{-10}$ M) (*Myers*, personal communication). By contrast, DNAse protection experiments revealed that T antigen binds site II much less efficiently in the absence of binding to site I. Similarly, binding to site III appears to be influenced by binding to site I and II. Thus, there is direct evidence that a cooperative binding interaction occurs between T antigen at sites I and II and possibly between site II and III. It is interesting to note that potential protein-protein interactions between tetramers of T antigen at the binding sites may be similar to those observed during formation of T-antigen dodecamer structures seen in the electron microscope (*Myers* et al. 1981c).

A number of recent studies have attempted to correlate the DNA binding properties of T antigen with various physical states of the A gene product. For instance, it was found in one study that newly synthesized T-antigen molecules exhibit a greater affinity for SV40 origin sequences than "aged" molecules (*Oren* et al. 1980). Other reports suggested that T-antigen molecules phosphorylated to different extents bind DNA with altered affinities. However, there is presently no clear indication of the biochemical or structural difference between new and old T-antigen molecules nor is there a definitive correlation between the phosphorylated state of the protein and its DNA binding properties. In short, there is no conclusive evidence to indicate the physiological importance of the various T-antigen forms that have been detected in SV40 infected and transformed cells.

Recently, there have been several attempts to determine the binding of T antigen to chromatin rather than naked DNA in vitro. Two such studies report that, in fact, the predominant binding site on chromatin is a region at the origin of viral replication, thus confirming the findings obtained with purified DNA (*Reiser* et al. 1980; *Segawa* et al. 1980). It is worthwhile to note that a region of DNA containing the SV40 origin of replication and transcriptional promoters has been found to be relatively free of nucleosomes when isolated from infected cells (*Varshavsky* et al. 1979). However, the relationship between T-antigen binding and the nucleosome structure of SV40 minichromosomes remains undefined.

5 Regulation of Early Viral Transcription

Having established the specific DNA binding properties of T antigen, it became important to determine how the binding of T antigen to its recognition sites can control viral gene expression. Why does T antigen bind in a cooperative manner to three specific binding sites? What is the spatial arrangement of binding sites to the viral origin of replication and transcriptional promoters? What are the molecular mechanisms by which T antigen is able to initiate viral DNA synthesis and repress early transcription? Although these questions have not all been answered, a number of recent studies have provided some insights into the function of T antigen during the lytic cycle.

Analysis of temperature-sensitive mutants in the A gene of SV40 indicated that both T antigen and its mRNA are overproduced under nonpermissive conditions (*Tegtmeyer* 1975; *Reed* et al. 1976; *Khoury* and *May* 1977). One possibility was for T antigen to act as a regulatory factor that interacts directly with the host RNA polymerase to inhibit early SV40 transcription. Alternatively, T antigen could catalyze a transient modification of some chromatin component in order to inactivate the template for early viral transcription. A third possibility was that T antigen could bind directly to the DNA template and act as a classical repressor to inhibit transcription. The specific DNA binding properties of T antigen led most workers in the field to favor this last possibility. The repressor model seemed particularly attractive because of the similarities between the T-antigen binding sites (*Tjian* 1978b) and features of the triple operator sites found in bacteriophage lambda (*Maniatis* and *Ptashne* 1973; *Meyer* et al. 1975; *Ptashne* et al. 1980). However, there was no direct evidence to indicate that interaction of T antigen to its three binding sites actually led to repression of SV40 early RNA synthesis.

It became clear that the most direct way to determine the mechanism of autoregu-

lation was to devise an in vitro transcription system that would be capable of selectively initiating transcription at the SV40 promoters and would be sensitive to the action of T antigen. Recently two cell-free transcription systems have been shown to initiate RNA from the adenovirus late promoter (*Manley* et al. 1980; *Weil* et al. 1979). Subsequently, these transcription systems have been adapted for synthesis of RNA directed by other viral and cellular promoters (*Rio* et al. 1980; *Proudfoot* et al. 1980; *Corden* et al. 1980; *Yamamoto* et al. 1980). Thus it was shown that SV40 DNA can direct the synthesis of discrete "run off" products corresponding to transcripts initiating from the viral early and late promoters (*Rio* et al. 1980; *Handa* et al. 1981; *Mathis* and *Chambon* 1981). Next it was shown that binding of either D2 protein or T antigen to template DNA containing the binding sites specifically inhibits transcription of SV40 early RNA without affecting late RNA synthesis (*Rio* et al. 1980; *Rio* and *Robbins;* personal communication; *Hansen,* personal communication). In addition, cloned SV40 template DNA lacking sequences for T-antigen binding sites I, or I and II but containing an intact early promoter continue to direct the transcription of early RNA even in the presence of T antigen (*Rio* et al. 1980). These in vitro transcription studies provide direct biochemical evidence that the synthesis of early SV40 RNA is specifically repressed by binding of the SV40 T antigen to its triple recognition sites.

A graphic representation of the interaction between T antigen and SV40 DNA to repress early transcription is shown in Figure 4. The location of the T-antigen binding sites was determined by a combination of DNAse protection, DMS methylation, and direct electron microscopy (*Tjian* 1978a; *Tjian* 1978b; *Myers* et al. 1981c; *Myers* et al. 1981a). Although a bona fide RNA polymerase II binding site has not been defined, the promoter region required for SV40 early transcription can be deduced from a composite of in vivo (*Benoist* and *Chambon* 1980; *Benoist* and *Chambon* 1981; *Gruss* and *Khoury,* pers. comm.) and in vitro (*Rio* et al. 1980; *Rio,* personal communication) deletion mapping data. There is general agreement from in vivo mapping data that a region between 106 and 180 containing one of the 72 base-pair repeats in SV40 DNA is required for A gene expression (see review by *Shenk* in this volume). A similar putative promoter sequence (from 70–160) has also been found to be required for transcription in vitro (*Rio* et al. 1980; *Rio,* personal communication). The small discrepancy in the sequences required for in vivo and in vitro transcription of SV40 early genes is not understood.

Unlike other cases of eukaryotic promoter sequences, deletion of the "TATA" box located 40–45 bases to the left of the putative SV40 early promoter has no apparent affect on the level of synthesis of transcripts in vitro (*Rio* et al. 1980). This result is consistent with the finding that A gene expression is unaltered by "TATA" deletion mutants in vivo (*Benoist* and *Chambon* 1980). Neither the in vivo nor in vitro results rule out the possibility that the "TATA" box has a quantitative affect on the capping and processing of the A gene mRNA (*Mathis* and *Chambon* 1981; *Dynan,* personal communication). The data do, however, suggest that transcription from the SV40 early promoter does not require the "TATA" canonical sequence.

The SV40 early transcriptional "promoter" region lies distal to binding sites I and II but close to, and may actually overlap, sequences of binding site III (*Myers* et al. 1981a). This arrangement of binding sites and promoter suggests that the mechanism of transcriptional repression may involve binding of T antigen at site III to directly inhibit initiation of transcription rather than attenuation of RNA polymerase movement along the template DNA. As a test of this model, it was shown that T-antigen binding sites

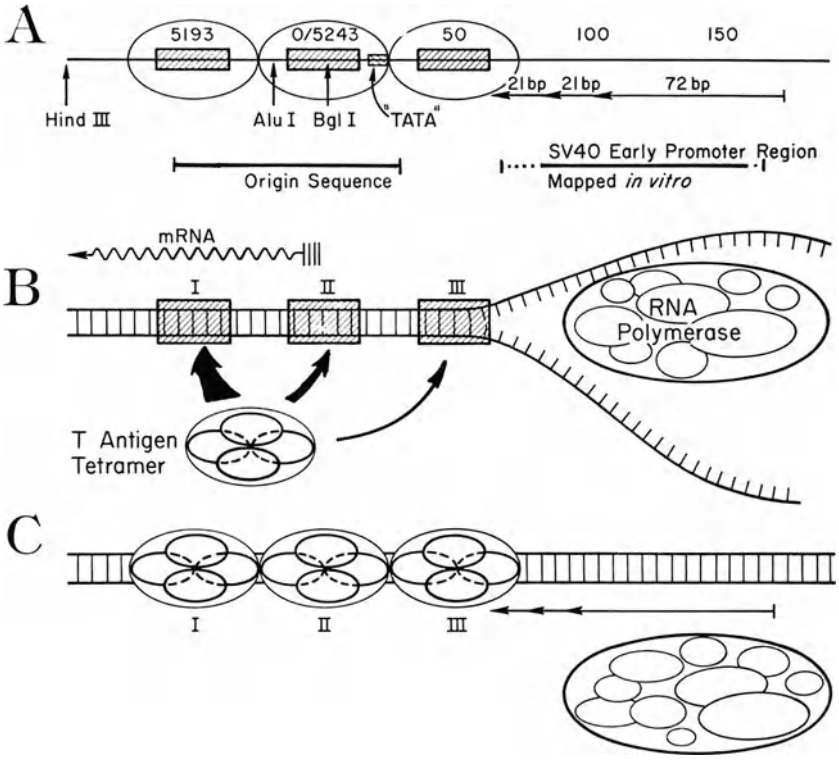

Fig. 4A–C. A schematic representation depicting potential interactions between T antigen, RNA polymerase and SV40 regulatory sequences. *A* Depicts the relative location of T-antigen binding sites (large hatched boxes and ellipses), restriction endonuclease cleavage sites (↑), "TATA" box (small hatched box), repeated sequences (leftward arrows ←); origin of DNA replication and putative early transcriptional promoter region mapped by in vivo and in vitro deletion analysis. The nucleotide numbers at the top are in accordance with the BBB system (*Tooze* 1980). *B* A diagram depicting derepressed SV40 early RNA synthesis is shown as starting from a position to the right of the T-antigen binding sites and the duplex template DNA is shown in an unwound configuration with a molecule of RNA polymerase II holoenzyme (large ellipse with many subunits) positioned to transcribe early SV40 RNA. The small ellipses (containing four isologous subunits) represents T-antigen tetramers and the arrows with different thicknesses represent sequential and cooperative binding of T antigen to its three recognition sites (hatched boxes). The early SV40 mRNA is shown as a wavy arrow with several 5′-capped ends represented by vertical lines. *C* A diagram depicting the repression of SV40 early RNA synthesis by T antigen. All three T-antigen binding sites are filled and RNA polymerase is unable to transcribe SV40 early RNA. The arrows pointing leftward represent the location of the 21 and 72 base-pair repeated sequences in SV40

placed downstream from the adenovirus late promoter are not able to block the transcription of adenovirus RNA synthesis even when all binding sites are saturated with T antigen (*Myers* et al. 1981a). One possibility is for T antigen to repress initiation of transcription by competing for a site that partially overlaps the RNA polymerase binding sequence. Alternatively, binding of T antigen to its sites may stabilize the double stranded

configuration of the template DNA thereby preventing RNA polymerase from unwinding DNA to form a transcriptional initiation complex.

It is worthwhile to note that deletion mutants lacking site I or I and II are no longer able to be repressed by T antigen even though site III is intact (*Rio* et al. 1980). A similar result was recently obtained with repression-defective mutants of early transcription in vivo (*Nathans*, personal communication). This is consistent with the finding that binding to site III may be dependent on a cooperative binding interaction between sites I and II. Thus, it appears that the mechanism of autoregulation may indirectly involve the cooperative binding of T antigen to all three sites. In addition, binding of T antigen to site III may play an active role in repressing transcription by directly preventing RNA polymerase from initiating at the SV40 early promoter.

These studies provide a molecular mechanism for the regulation of early SV40 transcription but do not help explain why late viral RNA synthesis is quiescent at early times and becomes maximally activated only after the onset of SV40 DNA replication. There have been numerous experiments both in vivo (*Cowan* et al. 1973; *Reed* et al. 1976; *Alwine* et al. 1977) and with preinitiated transcriptional complexes in vitro (*Birkenmeier* et al. 1977; *Birkenmeier* et al. 1979; *Ferdinand* et al. 1977) that have been designed to address the problem of late viral transcriptional control. However, the results of these studies have sometimes been controversial and this area of viral gene expression remains ill defined. It is, at present, not even clear whether late SV40 transcription is under negative or positive control. For instance, it was known for some time that mutations mapping in the genes encoding virion proteins exhibited an aberrant late transcriptional pattern (*Avila* et al. 1976; *Cole* et al. 1977). Several recent reports have presented data that supports the idea that late RNA synthesis at early times during infection is suppressed or attenuated by the binding of virion protein to a site on the DNA (*Laub* et al. 1979; *Llopic* and *Stark,* pers. comm.). However, these results are inconsistent with the finding that transcription of late RNA is inhibited at early times even when naked DNA is transfected into the cell (*Parker* and *Stark* 1979). An alternative possibility is for late transcription to be blocked by a cellular DNA binding protein that becomes released after viral DNA replication (*Alwine* and *Khoury* 1980).

A second major model confers the control of late viral RNA synthesis to the positive action of T antigen either directly or indirectly. At present, there is no evidence to implicate T antigen as a direct positive modulator of viral transcription. However, its role as an indirect activator of late transcription via DNA replication and induction of cellular genes has long been a popular model. An alternative simple possibility was recently suggested by the finding that, in vitro, the late promoter becomes active only at high concentrations of template DNA (*Rio* et al. 1980). This apparent differential efficiency of transcription from early and late promoters in vitro suggests that in vivo the early to late switch may also be due to differences in promoter strength. Thus, it is conceivable that the template concentration-dependent activation of late RNA synthesis, working together with the T-antigen mediated repression of early transcription, may contribute to the switch from early to late expression after the onset of DNA replication. However, the ability of cycloheximide to activate late gene transcription is a curious finding (*Handa* and *Sharp* 1980) that does not support the gene dosage model. Thus it is apparent that further investigation will be required to sort out the mechanism of late SV40 transcriptional regulation.