

Structure of Vaccinia Virus Early Promoters

Andrew J. Davison[†] and Bernard Moss[‡]

Laboratory of Viral Diseases
National Institute of Allergy and Infectious Diseases
National Institutes of Health
Bethesda, MD 20892, U.S.A.

(Received 20 April 1989, and in revised form 17 August 1989)

Functional elements of a vaccinia virus early promoter were characterized by making a complete set of single nucleotide substitutions, as well as more complex mutations, and assaying their effects on gene expression. Synthetic oligonucleotides, based primarily on the sequence of the 7.5-kD early promoter, were inserted into a plasmid vector containing the *lacZ* gene of *Escherichia coli* flanked by sequences from the thymidine kinase (TK) gene of vaccinia virus. The *lacZ* gene, under control of the synthetic promoter, was introduced into the vaccinia virus genome at the TK locus by homologous recombination, and each of the 331 different recombinant viruses thus obtained was assayed for β -galactosidase expression. The relative amounts and precise 5' ends of *lacZ* mRNAs specified by a subset of the recombinants were determined by primer extension. Many promoters were tested for their ability to direct specific transcription *in vitro*. A generally good correlation was noted between measurements of promoter strength estimated by β -galactosidase expression, primer extension of *in vivo* mRNA and transcription *in vitro*.

A relatively simple picture emerged from the analysis. The early promoter consists of a 16 base-pair critical region, in which most single nucleotide substitutions have a major effect on expression, separated by 11 base-pairs of a less critical T-rich sequence from a seven base-pair region within which initiation with a purine usually occurs. For the critical region of the 7.5-kD promoter, AAAAGTaGAAATA, any substitution of an upper-case nucleotide reduced expression, usually drastically, whereas certain substitutions of lower-case nucleotides maintained or significantly enhanced expression. On the basis of this analysis, the wide range of activities of natural promoters could be attributed to the presence of one or more non-optimal nucleotides in the critical region. Moreover, single nucleotide substitutions in such promoters had the predicted enhancing effects. Most mutations in the critical region of the 7.5-kD promoter behaved independently, but some nucleotide substitutions compensated for potentially detrimental nucleotides at other positions. Promoters substantially stronger than any natural ones examined were constructed by combining several up-mutations within the critical region of the 7.5-kD promoter and by repeating the critical region sequence. Like the TATA box of eukaryotic RNA polymerase II promoters, the critical region specifies the site of transcriptional initiation.

1. Introduction

Vaccinia virus mRNAs are synthesized in the cytoplasm of the infected cell at both early and late times. Early genes are transcribed and their RNAs processed into capped and polyadenylated species by a complete machinery packaged into virions (Kates & McAuslan, 1967; Munyon *et al.*, 1967; Wei & Moss, 1975). A soluble system for specific tran-

scription of early genes *in vitro* has been prepared from disrupted virions (Golini & Kates, 1985; Rohrmann & Moss, 1985), and many of the components of transcription have been purified and characterized, including RNA polymerase (Baroudy & Moss, 1980; Spencer *et al.*, 1980), poly(A) polymerase (Moss *et al.*, 1975), capping and methylating enzymes (Martin & Moss, 1975; Barbosa & Moss, 1978), as well as factors required for initiation (Broyles *et al.*, 1988; Broyles & Moss, 1988) and termination (Shuman *et al.*, 1987; Shuman & Moss, 1988). The 5' ends of several early mRNAs have been mapped using S₁ nuclease analysis or primer

[†] Present address: MRC Virology Unit, Institute of Virology, Church Street, Glasgow G11 5JR, U.K.

[‡] Author to whom all correspondence should be sent.

extension, but no convincing consensus for early promoters has emerged (Weir & Moss, 1983; Vassef, 1987; Mars & Beaud, 1987).

Late genes are transcribed only after the commencement of viral DNA synthesis. The encoded RNA sequence starts within three consecutive A residues present in a conserved TAAAT element in the non-coding DNA strand (Rosel *et al.*, 1986; Berthelot *et al.*, 1986; Weir & Moss, 1987; Lee-Chen & Niles, 1988), but the capped 5' ends of late mRNAs contain a longer polyadenylate tract (Berthelot *et al.*, 1987; Schwer *et al.*, 1987; Patel & Pickup, 1987; Ahn & Moss, 1989). A soluble system capable of transcribing late genes has been prepared from infected cells (Wright & Moss, 1987; Schwer & Stunnenberg, 1988), and at least one specific late transcription factor has been identified (Wright & Moss, unpublished results).

In contrast to canonical early and late genes, a gene encoding an mRNA that is translated *in vitro* into a protein with a molecular weight of 7.5 kDa is expressed constitutively during infection. (As an abbreviation, we refer to the gene and mRNA encoding the 7.5 kDa polypeptide (Wittekk *et al.*, 1980; Cooper *et al.*, 1981) as the 7.5-kD gene and 7.5-kD mRNA, respectively. Similarly, the promoter is called the 7.5-kD promoter.) In order to understand this apparent exception to regulated transcription, Cochran *et al.* (1985) placed the bacterial chloramphenicol acetyltransferase gene under control of the sequence upstream from the 7.5-kD protein-coding region, and assayed the effects of deleting sequences from the 5' side of the promoter on transcription *in vivo* and *in vitro*. They also characterized the 5' ends of 7.5-kD mRNAs by S₁ nuclease analysis.

	BamHI	SalI	HindIII	XhoI	BglII
TK - GAATT GGATCC GTCGAC AAGCTT				CTCGAG CATGGG AGATCT AATTC - TK	

They found that constitutive expression of the 7.5-kD gene is due to the presence of two distinct, separate promoters, one late and the other early. The former is located about 60 bp† upstream from the latter. They showed that deletion of all but 31 bp upstream from the early mRNA initiation site did not affect early transcription, but that removal

	BamHI	SalI	HindIII	XhoI	Met - Gly - Asp - Pro - . . . β gal
TK - GAATT GGATCC GTCGAC AAGCTT				CTCGAG C ATG GGA GAT CCC . . . lacZ - TK	

of a further 7 bp inactivated the promoter. Similar results were obtained for the thymidine kinase (TK) early gene promoter (Weir & Moss, 1987). In a third study, Coupar *et al.* (1987) deleted sequences from

	BamHI	SalI	HindIII	KpnI	XhoI	Met . . . β gal
TK - GAATT GGATCC GTCGAC AAGCTT				GGTACC	CTCGAG C ATG . . . lacZ - TK	

† Abbreviations used: bp, base-pair(s); TK, thymidine kinase; p.f.u., plaque-forming units; m.o.i., multiplicity of infection; araC, cytosine arabinoside.

the 3' side of another early promoter. The one active mutant of the three tested had a deletion extending 8 bp upstream from the initiation site, and initiation occurred in approximately the same position relative to upstream sequences still present. Taken together, the deletion analyses indicate that the crucial sequence for early promoter activity is located between about 8 and 30 bp upstream from the initiation site, and that the sequence in the vicinity of the initiation site is not critical.

Deletion analysis is not readily applied, however, to detailed study of promoter structure. Also, the gross sequence alterations that occur on deletion may have effects on expression which, being unquantified, are usually ignored. Therefore, in order to identify precisely those elements important in promoter function, we used oligonucleotide synthesis technology to generate a large number of precisely defined mutations in early promoters, and assayed the effect of each mutation on transcription. Reliance was placed on substitution mutations and on sets of sequences differing by single nucleotide insertions.

2. Materials and Methods

(a) Plasmid construction

The plasmid vectors used in this study for testing promoter function were derived from pMM34 (Mackett *et al.*, 1985), which comprises the vaccinia virus TK gene on a HindIII-XbaI fragment linked to a PvuII-AvaI fragment from pBR328, with a small EcoRI fragment of vaccinia virus DNA inserted at the EcoRI site within the TK gene. The small EcoRI fragment was replaced by a synthetic oligonucleotide duplex containing 5 restriction endonuclease sites. The insert had the following sequence in the upper strand:

	XbaI	BglII
TK - GAATT GGATCC GTCGAC AAGCTT	CTCGAG	CATGGG AGATCT AATTC - TK

Two plasmids were obtained; pMJ1, in which the above sequence was in the same orientation as the TK gene, and pMJ2, in which it was in the inverse orientation. Then, a BamHI fragment containing the entire *Escherichia coli* lacZ gene (encoding β -galactosidase), but for the 1st 8 codons, was obtained from pSC8 (Chakrabarti *et al.*, 1985) and ligated into the BglII site of each plasmid. The resulting plasmids, pMJ3 and pMJ4, contained the following sequence:

	BamHI	SalI	HindIII	XhoI	Met - Gly - Asp - Pro - . . . β gal
TK - GAATT GGATCC GTCGAC AAGCTT				CTCGAG C ATG GGA GAT CCC . . . lacZ - TK	

The 4 restriction sites are not present elsewhere in the plasmids. The 1st 8 codons for β -galactosidase were replaced by codons specifying Met-Gly-Asp-Pro. Next, a unique KpnI site was introduced by ligating a synthetic oligonucleotide duplex between the HindIII and XbaI sites, to give pMJ11 and pMJ35.

	BamHI	SalI	HindIII	KpnI	XhoI	Met . . . β gal
TK - GAATT GGATCC GTCGAC AAGCTT				GGTACC	CTCGAG C ATG . . . lacZ - TK	

Candidate promoters were synthesized using an Applied Biosystems 370B DNA synthesizer, and inserted as duplexes usually between the HindIII and XbaI sites in pMJ3, pMJ11, pMJ4 or pMJ35. The former 2 plasmids

have *lacZ* in the same orientation as the TK gene, and the latter 2 have *lacZ* in the opposite orientation. After ligation, remaining circular vector molecules lacking the insert were linearized by cleaving with *Hind*III or *Xba*I (in those instances where 1 or both of these sites was destroyed by inserting oligonucleotides into pMJ3 or pMJ4), or with *Kpn*I (in those instances where oligonucleotides were inserted into pMJ11 and pMJ35), in order to reduce their transforming ability.

The majority of plasmids used for generating vaccinia virus recombinants were made by inserting mixed oligonucleotides based on the 7.5-kD early promoter sequence:



The major sites of initiation of early mRNA from the resident vaccinia virus 7.5-kD gene are at +1(G) and +3(A) (Venkatesan & Moss, 1981; Venkatesan *et al.*, 1981; and this work). The 33 nucleotide promoter segment of each oligonucleotide was synthesized with phosphoramidites that had been contaminated to a level of about 1.5% with each of the other 3 phosphoramidites. Thus, an average of about 1 in 20 nucleotides within the target sequence differed from the sequence above, resulting in about 3 mismatched base-pairs per duplex before replication, and about 1.5 mismatched base-pairs after replication, assuming unbiased repair. The oligonucleotides were inserted between the *Hind*III and *Xba*I sites of pMJ4 and pMJ35, and *E. coli* HB101 cells were transformed to ampicillin resistance. Separate plasmid preparations were made from 1.5-ml cultures of about 800 transformants, and the sequence of each insert was determined as described by Hattori & Sakaki (1986). The parental 7.5-kD promoter (in pMJ21), the majority of the 99 possible mutants with different single substitutions in the 33 bp target, and 125 distinct and more complex mutants were obtained. The remaining single substitution mutants and other early promoter constructions were made using either partially or non-degenerate oligonucleotides. Before generating recombinant viruses, each transformant was colony-purified and the DNA sequence of the promoter was verified.

(b) Generation of recombinant vaccinia viruses

Recombinant viruses were made essentially as described by Mackett *et al.* (1984) and Chakrabarti *et al.* (1985). Plasmid from 1.5-ml bacterial cultures was resuspended in 100 µl of Hepes-buffered saline (0.14 M-NaCl, 5 mM-KCl, 1 mM-Na₂HPO₄·2H₂O, 0.1% (w/v) glucose, 20 mM-Hepes (pH 7.05)) containing 1 µg vaccinia virus (strain WR) DNA/ml and 14 g calf thymus DNA/ml. Then 5 µl of 2 M-CaCl₂ were added, and the mixture was incubated at room temperature for 30 min. The suspensions were then added, 1/well, to a 24-well tray containing drained confluent monolayers of CV-1 cells that had been infected 2 h previously with vaccinia virus (WR strain) at a multiplicity of 0.05. After 30 min at room temperature, each monolayer was overlaid with 1 ml of MEM (minimal essential Eagle's medium supplemented with Earle's salts, 2.5% (v/v) fetal bovine serum, 2 mM-glutamine, 100 units of penicillin/ml and 100 µg streptomycin/ml), and the tray was incubated at 37°C for 3.5 h. The monolayers were aspirated, overlaid with 0.5 ml of MEM and incubated at 37°C for 2 days. Infected cells were scraped into the medium, placed in vials, and disrupted by sonication.

Recombinant viruses, containing *lacZ* inserted into the TK gene, were plaque-purified thrice on 12-well trays containing confluent monolayers of human TK⁻ 143 cells. After absorbing 0.1 ml of appropriate dilutions of virus for 1 h, monolayers were aspirated and overlaid with 1 ml of MEM-PR (MEM lacking phenol red) containing 1% (w/v) low melting point agarose and 25 µg 5-bromo-2'-deoxyuridine/ml. After allowing the agarose to set at 4°C, the trays were incubated at 37°C for 2 days. The wells were then overlaid with 0.5 ml of MEM-PR containing 1% low melting point agarose and 0.3 mg 5-bromo-4-

chloro-3-indolyl β-D-galactopyranoside/ml, and incubated overnight at 37°C. Well-separated blue plaques were picked with Pasteur pipettes and sonicated in 0.5 ml of MEM before the next round of plaque purification. Isolation of recombinants was facilitated by the TK selection, which resulted in a majority of blue plaques at the 1st round of purification. Moreover, even those recombinants lacking an active promoter upstream from *lacZ* produced pale blue plaques. This is presumably due to translation of *lacZ* coding sequences present in long mRNAs (or their degraded products) produced late in infection.

The thrice-picked plaque of each recombinant was used to infect a 25 cm² monolayer of human TK⁻ 143 cells in the presence of 25 µg 5-bromo-2'-deoxyuridine/ml. After incubation at 37°C for 2 days, infected cells were scraped, pelleted and sonicated in 0.5 ml of MEM. A 150 cm² monolayer of CV-1 cells was then infected with a 0.25 ml inoculum and incubated at 37°C for 3 days. The cells were scraped, pelleted and sonicated in 0.5 ml of MEM, and the resulting virus stock was stored at -70°C. Each recombinant was titrated on monolayers of CV-1 cells immediately before assaying β-galactosidase expression. Titres of 2 × 10⁹ to 5 × 10⁹ p.f.u./ml were obtained routinely.

(c) β-Galactosidase assay

Trays (24-well) of confluent CV-1 monolayers were infected with 10⁷ p.f.u. of virus in 0.1 ml of MEM-PR. Representative monolayers were shown by haemocytometric counting to contain approximately 2 × 10⁵ cells. The nominal m.o.i. of 50 was used because it was determined experimentally that β-galactosidase expression from the 7.5-kD early promoter is largely independent of m.o.i. in the range of 20 to 100 (data not shown). After absorbing at 37°C for 1 h, the monolayers were aspirated and overlaid with 0.7 ml of MEM-PR or MEM-PR containing 40 µg cytosine arabinoside (araC)/ml and incubated at 37°C. To assay a group of 24 recombinants, each virus was applied to 1 monolayer on each of 9 trays. Three trays were overlaid with MEM-PR and harvested immediately, 3 were overlaid with MEM-PR and harvested 12 h after infection, and 3 were overlaid with MEM-PR containing araC and harvested 12 h after infection. Thus, each virus was assayed for 3 sample points in triplicate on separate trays. In order to ensure further the statistical significance of the results, all viruses were assayed on a 2nd occasion.

β-Galactosidase assays were carried out essentially as described by Miller (1972). The trays were frozen then thawed, and cells were scraped into the medium and

placed in 1.5-ml Eppendorf tubes. The cells were disrupted by adding 1 drop of 0.05% (w/v) sodium dodecyl sulphate and 1 drop of chloroform. Where appropriate, disrupted extracts were diluted in MEM-PR before repeating the disruption step. After vortexing the tubes twice, 75 µl of 10×Z buffer (0.6 M-Na₂HPO₄, 0.4 M-NaH₂PO₄, 0.1 M-KCl, 0.01 M-MgSO₄, 0.5 M-2-mercaptoethanol (pH 7.0)) and 150 µl of 4 mg *o*-nitrophenyl-β-D-galactopyranoside/ml in Z buffer were added, and the tubes were incubated for 1 h at room temperature. Reactions were terminated by adding 375 µl of 1 M-Na₂CO₃. The tubes were centrifuged at 12,000 g for 2 min, and 200-µl portions were placed in a 96-well microtitre tray. Absorbances were measured simultaneously at 410 nm (released *o*-nitrophenol) and 570 nm (control wavelength) using a Dynatech MR600 microplate reader, with reference to a sample in which MEM-PR had been assayed. Net β-galactosidase synthesis during the period of infection was calculated for each virus, and results from duplicate experiments were normalized with respect to the value obtained with the parental 7.5-kD promoter in order to correct, for example, for variations in monolayer cell density and incubation temperature during the assay. A zero value was assigned to a recombinant lacking a promoter upstream from *lacZ* (vMJ4, made using pMJ4). Final values (mean ± standard deviation) were expressed as a percentage of the mean value obtained using an appropriate control virus included in the same experiments. More accurate values for selected weak promoters were obtained by assaying 6 monolayers for each virus, instead of 3, and allowing the enzyme assay to proceed for 3 h at 37°C.

(d) Analysis of early RNA synthesized in vitro

A DNA fragment suitable for transcribing *in vitro* was prepared by digesting 50 µg of plasmid with *Bam*HI and *Pvu*II. The smallest product fragment (143 bp, not including the 4 nucleotide overhang at the *Bam*HI site, for pMJ21 and its substituted derivatives) was excised from a 5% (w/v) polyacrylamide gel. DNA was eluted overnight at 37°C in 0.5 ml of 10 mM-Tris-HCl (pH 8.0), 1 mM-EDTA, 0.75 M-ammonium acetate, and precipitated with ethanol. Recovery was estimated by comparison with standards on a 5% polyacrylamide gel stained with ethidium bromide. For use in the transcription assays, DNA fragments were diluted to 0.5 µg/ml in water.

Templates were assayed for promoter activity using a modification of the procedure described by Broyles & Moss (1987). The unit reaction mixture contained 0.4 µl of 10×T buffer (0.2 M-Tris-HCl (pH 7.5), 0.1 M-MgCl₂, 20 mM-dithiothreitol), 0.4 µl of 10×NTPs (10 mM each ATP, GTP and CTP, 0.2 mM-UTP), 0.8 µl of [α -³²P]UTP (PB10203; Amersham) and 0.4 µl of vaccinia virus RNA polymerase complex. The RNA polymerase complex was prepared by sucrose gradient-centrifugation of an extract of vaccinia virions that had been subjected to DEAE-cellulose chromatography. In some experiments, as indicated in the Figure legends, [α -³²P]CTP (PB10202; Amersham) replaced [α -³²P]UTP, and the 10×NTPs consisted of 10 mM each ATP, GTP and UTP, 0.2 mM-CTP. Sufficient mixture for the number of transcription assays was constituted on ice immediately before commencing the reactions. For each reaction, one uncapped 1.5-ml Eppendorf tube was placed in a microcentrifuge at room temperature, and 2 µl (approx. 1 ng) of DNA fragment was placed on the inside surface; 2 µl of reaction mix was then placed as a separate drop on the inside surface. The

reactions were initiated by centrifuging briefly, then the caps were closed and the tubes were incubated at 32°C for 15 min. This time was chosen because linear incorporation of isotope continued for about 30 min. The tubes were then placed on ice and the reactions terminated by adding 2 µl of formamide/dyes (0.1% (w/v) bromophenol blue, 0.1% (w/v) xylene cyanol FF, 10 mM-EDTA in 98% (v/v) deionized formamide). The samples were heated in uncapped tubes for 5 min at 100°C, and then electrophoresed on 6% polyacrylamide gels containing 7 M-urea. Gels were fixed in 10% (v/v) acetic acid for 30 min, dried on 3MM chromatography paper and autoradiographed. The approximate size of the major RNA species specified by the parental 7.5-kD promoter was estimated to be 100 nucleotides by comparison with single-stranded DNA markers.

(e) Analysis of early RNA made in vivo

Early RNA was prepared from HeLa suspension cells that had been preincubated at 5 × 10⁵ cells/ml for 30 min, pelleted and resuspended at 5 × 10⁶ cells/ml, infected at a m.o.i. of 20 for 30 min, diluted to 5 × 10⁵ cells/ml, and then incubated with shaking for 4 h. Each of these steps was carried out at 37°C in the presence of 100 µg cycloheximide/ml. Cytoplasmic RNA was isolated and purified by centrifugation through caesium chloride gradients as described by Cooper & Moss (1979). For some experiments, RNA was chemically decapped as reported by Fraenkel-Conrat & Steinschneider (1967). Approximately 5 to 25 µg of RNA was reverse transcribed as described by Wright & Moss (1987), using molar excesses of 2 (5'-³²P)-labelled oligonucleotide primers added simultaneously. One primer complemented nucleotides 30 to 47 of the *lacZ* mRNA under control of the synthetic 7.5-kD promoter, and the other complemented nucleotides 10 to 29 of the 7.5-kD early mRNA specified by the resident vaccinia virus genes. The latter acted as a control for variable amounts of RNA present in the reactions. Primer-extended products were denatured in formamide/dyes and electrophoresed on 6% polyacrylamide gels containing 7 M-urea. Gels were then dried and autoradiographed. Sequencing products obtained using the radiolabelled *lacZ* primer with the plasmid containing the parental 7.5-kD promoter (pMJ21) were used as markers to determine the precise sizes of *in vivo* RNAs. Amounts of *lacZ* mRNA relative to 7.5-kD mRNA produced from the resident 7.5-kD genes were determined by autodensitometry.

3. Results

(a) Effect of single nucleotide substitution mutations on β-galactosidase expression

A complete set of single nucleotide substitutions in the 7.5-kD promoter was produced; 99 mutations in a 33 bp region that includes the normal transcription initiation sites. These sequences, controlling *lacZ*, were recombined into the vaccinia virus genome, so that *lacZ* was in the opposite orientation from the TK gene. The relative amounts of β-galactosidase produced by the recombinants were assayed under conditions that allowed synthesis either of early proteins only or of both early and late proteins. The results are summarized in Figure 1.

Mutations in the region from -28(A) to -13(A) had a marked effect on promoter strength.

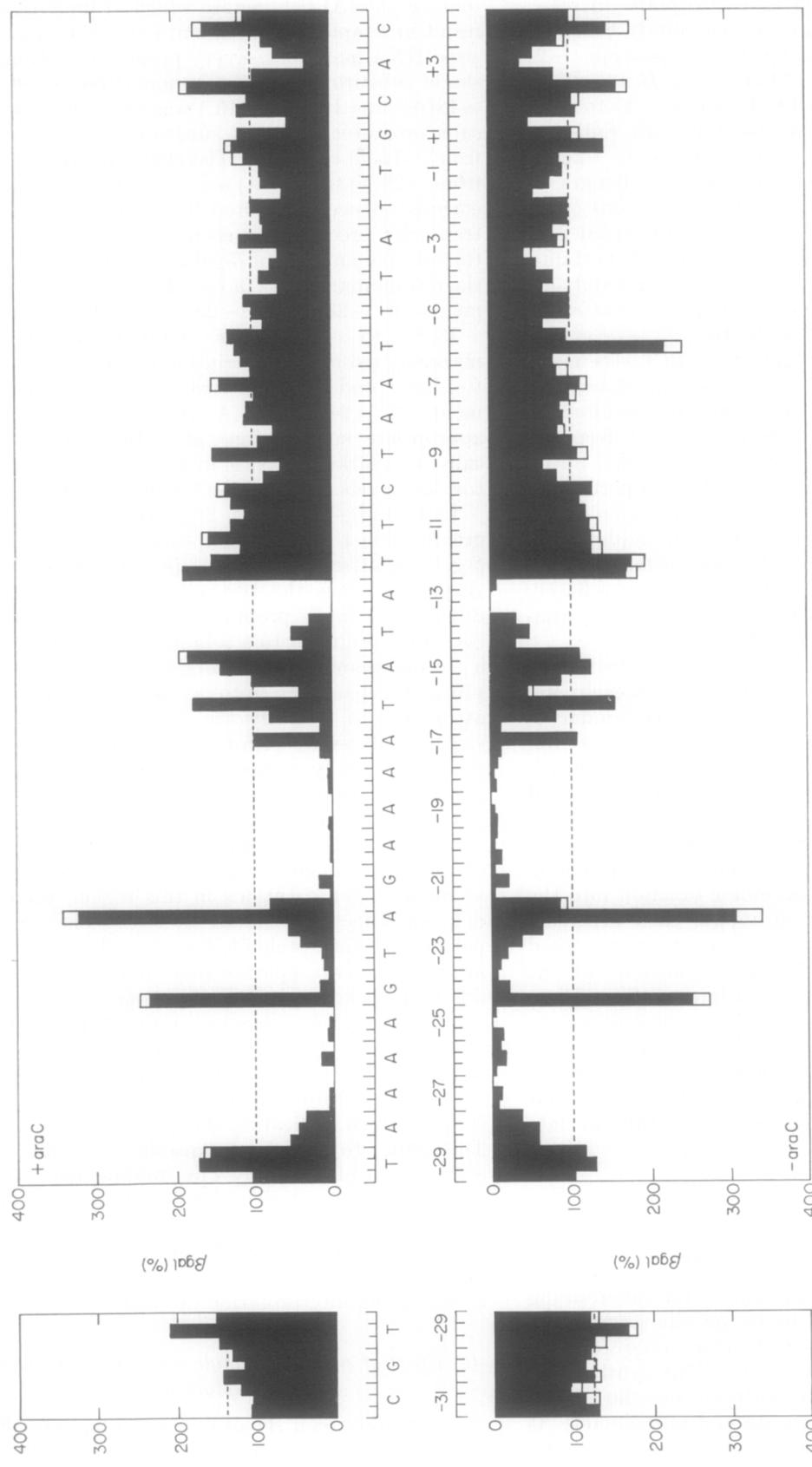


Figure 1. Histogram showing levels of β -galactosidase expressed by cells infected with recombinant viruses containing single nucleotide substitutions in the 7.5-kD early promoter. Results in the larger panel are based on the parental sequence (-29 to +3) present in vMJ21, and those in the smaller panel are based on the same sequence with 2 additional wild-type nucleotides (-31(C) and -30(G)) present in vMJ8. Mean β -galactosidase levels are shown as filled columns and expressed relative to vMJ21 (100%) in both panels. Standard deviations are shown as open columns when they exceed 10% of the mean level of expression by vMJ21. The horizontal broken lines in the larger and smaller panels show levels for the parental sequences present in vMJ21 and vMJ8, respectively. As each parental nucleotide shown may be substituted by 1 of 3 others, 3 values are plotted at each position. The order of substitution, from left to right, is based on G-A-T-C, with the appropriate parental nucleotide missing. Thus, for example, their order at -29(T) is G-A-C. β -Galactosidase expression in the presence of araC is shown above the parental sequence, and that in its absence is shown below.

Mutations at 12 positions in this critical region resulted in total or severe loss of promoter activity, and at four positions led to a substantial increase. The following nucleotides are particularly beneficial for promoter activity; A residues in positions -28 to -24, T residues at -23 and -22, G at -21, A residues from -20 to -18, A or T at -17 to -16, T or C at -15, T at -14 and A at -13. Substitutions further upstream, from -31(C) to -29(T), resulted in measurable, but minor, changes in promoter activity. The latter mutations were obtained using a parental sequence extended 2 bp upstream from that used to obtain the 99 mutants, and so these results are shown separately in Figure 1. The longer promoter sequence was about 30% more active than the shorter.

Substitutions from -12(T) to +4(C) were much less dramatic in their effect on expression than those in the critical region. No mutation resulted in complete loss of activity, and the greatest decreases were caused by substitutions at the normal initiation sites at +1(G) and +3(A). It is important to note, however, that mutations in the region of the initiation sites could have effects in addition to those caused by mutations further upstream. They may directly influence events at, or immediately after, initiation of mRNA synthesis, or the translational efficiency of the mRNA.

The relative activities of single substitution mutants under conditions in which both early and late proteins were produced were very similar to those obtained when late protein synthesis was inhibited by araC. One exception is shown in Figure 1; substitution of -6(T) by an A resulted in increased β -galactosidase expression when late gene expression was allowed. This single nucleotide change created a TAAAT element that probably introduced a weak late promoter function into the early promoter. Similar examples, each involving introduction of a TAAAT element, were noted among the multiple substitution mutants of the 7.5-kD promoter described in a later section. These results indicate that early and late promoters can overlap substantially. It should be noted that, since late promoter function depends on sequences upstream from the TAAAT, most substitutions that introduced this element did not result in late promoter activity.

(b) Effect of single nucleotide substitution mutations on transcription in vitro

The approach described above for determining promoter strength was based on the assumption that the amount of β -galactosidase produced was proportional to the amount of mRNA synthesized. This indirect assay of transcription has the advantage of being simple, accurate and reproducible. It was important, nonetheless, to obtain a direct measure of mRNA production, at least for a representative subset of mutants. As a first step, an *in vitro* transcription system specific for vaccinia

virus early genes was used. RNA polymerase was partially purified from virion lysates in order to remove poly(A) polymerase which, if present, could result in template-independent extension of synthesized RNA molecules. Every promoter containing a single substitution in the regions from -29(T) to -13(A) or -1(T) to +4(C) was assayed, as well as each promoter with a G substitution in the region from -12(T) to -2(T). The effects of substitutions from -29(T) to -13(A) are shown in Figure 2(a). A correspondence was noted between the amount of transcript produced by each promoter *in vitro* and the relative amount of β -galactosidase produced by the recombinant virus *in vivo*. For example, replacements of -28(A) to -25(A), -23(T), -21(G), -18(A) to -16(A) or -13(A) abolished or markedly reduced transcription *in vitro* and β -galactosidase production *in vivo*. Conversely, replacement of -24(G) by A or -22(A) by T, or appropriate substitutions at -16(T) or -15(A), caused a marked increase in transcript and β -galactosidase levels. Some of the G substitutions between -12(T) and -2(T) (Fig. 2(b)) appeared to have a greater effect on transcription *in vitro* than on β -galactosidase production *in vivo*. This is not surprising, since the two assay systems are quite different. A single promoter is presented as a purified DNA fragment to a soluble *in vitro* system, but a recombinant virus contains many competing promoters packaged into a virion. We have not attempted to quantify the difference in promoter strengths obtained with the two systems because of the substantially greater experimental variation of the *in vitro* assay.

Figure 2(a) and (b) show that mutations from -29(T) to -2(T) did not affect the size or pattern of transcripts made *in vitro* but only their overall abundance. Substitutions in this region, therefore, did not affect the initiation sites. The complex pattern of bands probably has several explanations; initiation at two separate sites, +1(G) and +3(A), different phosphorylation and capping states of the synthesized RNA molecules, and perhaps the action of residual poly(A) polymerase present in the RNA polymerase preparation. Mutations from -1(T) to +3(A) (Fig. 2(b)) caused alterations in the transcript pattern, indicating changes in initiation sites. The shift to smaller transcripts produced by replacement of +1(G) by a pyrimidine, for example, is interpreted as abolition of initiation at this position. The complex transcript patterns and unavailability of appropriate RNA molecular weight markers accurate to the nucleotide, however, made a convincing interpretation difficult.

(c) Effect of single nucleotide substitution mutations on transcription *in vivo*

In an additional set of experiments, early RNA was extracted from infected cells, and analysed by primer extension. The results are shown in Figure 2(c). mRNAs specified by the *lacZ* and resident 7.5-kD genes were extended simultaneously in order

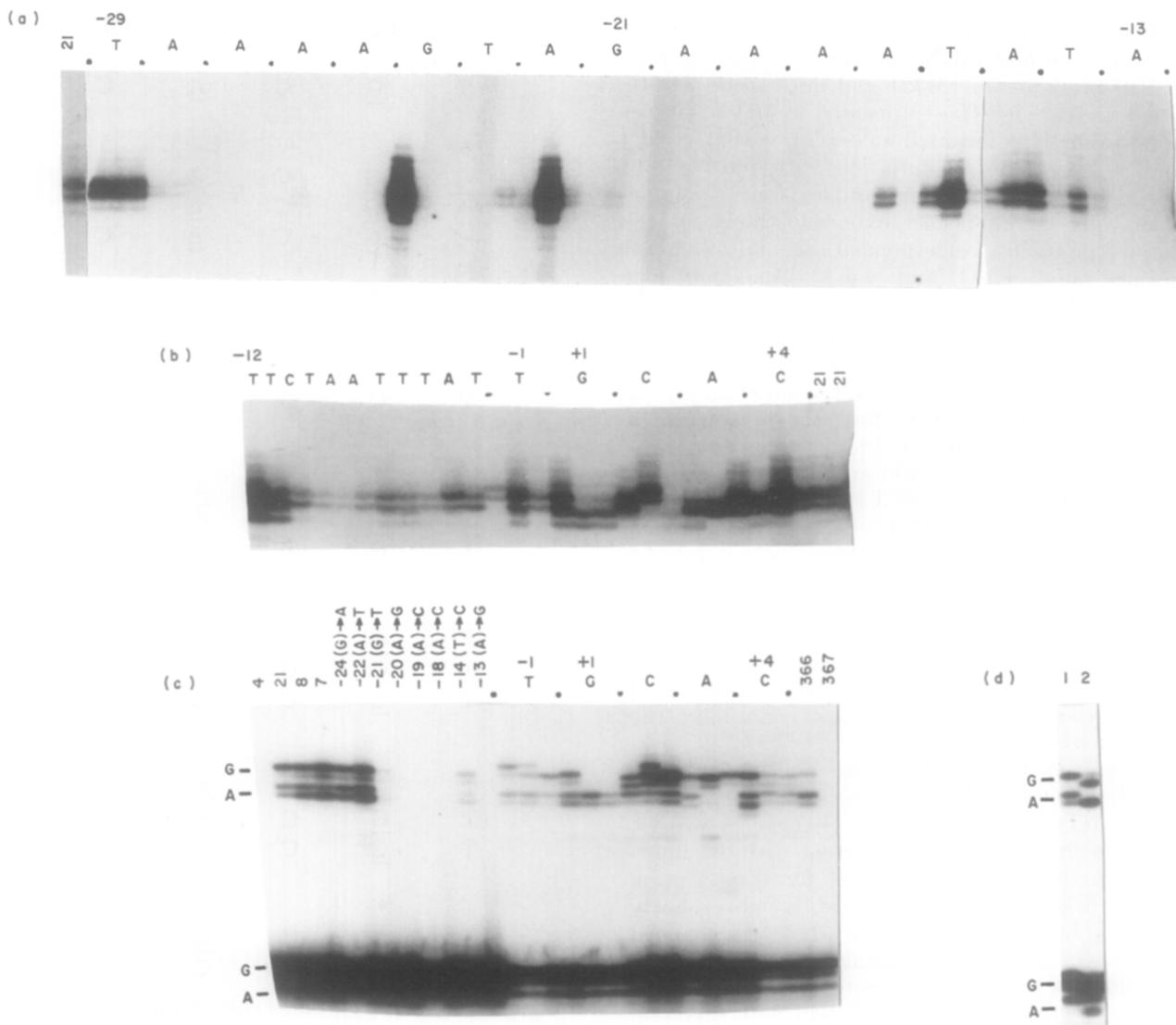


Figure 2. (a) An autoradiograph showing RNA synthesized *in vitro* from templates containing all possible single nucleotide substitutions from -29(T) to -13(A) of the 7.5-kD early promoter. Data are presented in groups of 3 ordered as in Fig. 1. Lane 21 shows RNA (approx. 100 nucleotides) synthesized from the parental promoter present in pMJ21. (b) An autoradiograph showing RNA synthesized *in vitro* from templates containing single nucleotide substitutions from -12(T) to +4(C) in the 7.5-kD early promoter. The ten leftmost lanes show 1 substitution each for nucleotides -12(T) to -2(T). In each case, the substituting nucleotide was G. To the right of these are results for all possible substitutions of -1(T) to +4(C), ordered as in Fig. 1. Lanes 21 show RNA (approx. 100 nucleotides) synthesized from the parental promoter present in pMJ21. (c) An autoradiograph showing primer extended products from mRNA synthesized *in vivo*. The upper group of bands is derived from lacZ mRNA, and the lower from 7.5-kD early mRNA. Marker mobilities corresponding to mRNAs initiated at +1(G) and +3(A) are indicated for each group. Lanes 4, 21, 8, 7, 366 and 367 show results for vMJ4 (no promoter), vMJ21 (parental 7.5-kD promoter), vMJ8 (longer parental 7.5-kD promoter), vMJ7 (longer parental 7.5-kD promoter controlling lacZ in the same orientation as the TK gene), vMJ366 and vMJ367. Sequences of promoters in, and β -galactosidase results for, vMJ8 and vMJ7 are shown in Fig. 4(f)(i) and those for vMJ366 and vMJ367 are shown in Fig. 4(a)(iii). Also shown are selected results for single nucleotide substitutions in the region from -24(G) to -13(A), and results for all single nucleotide substitutions from -1(T) to +4(C) (ordered as in Fig. 1). (d) An autoradiograph showing primer extended products from RNA synthesized *in vivo* by vMJ21 (parental 7.5-kD promoter). The upper group of bands is derived from lacZ mRNA, and the lower from 7.5-kD early mRNA. (1) and (2) show results for untreated and chemically decapped mRNA, respectively.

to control for variable amounts of RNA in the reactions. The two 5'-radiolabelled primers did not, however, have the same specific activity, and so the relative strengths of resident and translocated 7.5-kD promoters was not determined. As expected, little or no lacZ mRNA was detected when substitu-

tions were made at positions -21(G) to -18(A) and -13(A). Good agreement between β -galactosidase levels and the relative amounts of lacZ mRNA produced *in vivo* was taken as further validation of the β -galactosidase assay for measuring promoter strength.

Four major primer-extended products were derived from *lacZ* mRNA specified by the parental 7·5-kD promoter in vMJ21 and most of its substituted derivatives. These appeared at first to correspond to mRNAs initiated at -1(T), +1(G), +2(C) and +3(A); that is, initiation with purines and pyrimidines. When RNA samples were decapped chemically before primer extension, however, the major products corresponded to initiation at purines only; +1(G) and +3(A). An example is shown in the upper group of bands in Figure 2(d). This shift in apparent 5' ends after decapping was noted also for the mRNA specified by the resident 7·5-kD genes, as indicated by the lower group of bands in Figure 2(d). Therefore, the presence of a cap structure apparently induced the reverse transcriptase preparation to add an extra nucleotide of undetermined nature to the 3' end of a proportion of the primer-extended DNA chains. The addition appears to be independent of the sequence near the 5' end of the mRNA, as it was detected using mRNA specified by promoters containing different sequences in the initiation region (Fig. 2(c) and data not shown). The sequence at the 5' end of the mRNA, however, may affect the proportion of primer-extended DNA molecules containing the extra nucleotide.

Considering the results shown in Figure 2(c) in more detail, it is evident that some of the mutations from -1(T) to +3(C) caused changes in mRNA initiation sites. These are consistent with our interpretation of the data obtained by transcription *in vitro* (Fig. 2(b)). For example, a more slowly migrating band appeared when -1(T) was changed to A or G but not to C, indicating a new purine initiation site at -1. Conversely, substitution of T or C for +1(G) eliminated initiation at this site, whereas substitution of A did not. The presence of new intermediate bands when +2(C) was changed to G suggested a new start site at +2. However, the presence of a more slowly migrating band when +2(C) was changed to A was consistent with an additional initiation at -1(T). Although this is the only apparent initiation at a pyrimidine, the conclusion was reinforced using decapped mRNA (data not shown), is apparent in the sizes of *in vitro* transcripts from this promoter (Fig. 2(b)), and was observed when a different promoter containing the same sequence in the initiation region was analysed (data not shown). It is possible that this exceptional result might have another explanation, but the data at their present state of interpretation do not rule out initiation at a pyrimidine under certain circumstances. Substitution of T or C (but not G) for +3(A) eliminated the initiation site at that position. Substitution of G (but not A or T) for +4(C) appeared to allow initiation at +4. Replacement of +1(G) and +3(A) by A and G, respectively, maintained initiation at these two positions, although with a shift of preference towards +3 (Fig. 2(c), vMJ366). Replacement of the purines at +1(G) and +3(A) by pyrimidines abolished initiation at these nucleotides (Fig. 2(c), vMJ367). The β -galactosidase

	-1	+1	+2	+3	+4
	T	<u>G</u>	C	<u>A</u>	C
*					
G	G	C	<u>A</u>	C	
A	<u>G</u>	C	<u>A</u>	C	
C	<u>G</u>	C	<u>A</u>	C	
		*			
T	A	C	<u>A</u>	C	
T	<u>T</u>	C	<u>A</u>	C	
T	C	C	<u>A</u>	C	
			*		
T	G	G	<u>A</u>	C	
T	<u>G</u>	<u>A</u>	<u>A</u>	C	
T	<u>G</u>	<u>T</u>	<u>A</u>	C	
			*		
T	G	C	G	C	
T	<u>G</u>	C	T	C	
T	<u>G</u>	C	C	C	
				*	
T	<u>G</u>	C	<u>A</u>	<u>G</u>	
T	<u>G</u>	C	<u>A</u>	<u>A</u>	
T	<u>G</u>	C	<u>A</u>	T	
		*	*		
T	A	C	G	C	
T	<u>T</u>	C	<u>C</u>	C	

Figure 3. Effects on initiation sites of mutations in the -1(T) to +4(C) region of the 7·5-kD promoter. The natural sequence is shown on the upper line. Mutated nucleotides are indicated by an asterisk above each group, and major initiation sites are underlined. The scheme is derived from the *in vivo* RNA data shown in Fig. 2(c).

results for these double substitution mutants (vMJ366, vMJ367) are shown in Figure 4(a)(iii). In addition to directing initiation within the -1 to +4 region, all active promoters induced a low level of initiation at +7(G) (Fig. 2(c)). Figure 3 summarizes our interpretation of the effects of single and double nucleotide substitutions in the region from -1(T) to +4(C) on initiation sites. Initiation can occur efficiently at least over the 5 bp region from -1 to +4. In addition, initiation can occur inefficiently at +7, at least when initiation is also able to occur in the region from -1 to +4.

(d) *Effect of multiple nucleotide substitution and deletion mutations on promoter strength*

In the process of generating the 99 single substitution mutants, 100 random multiple substitution mutants were obtained. β -Galactosidase expression

(a) DOUBLE SUBSTITUTIONS IN THE 7.5-KD PROMOTER

(i) Both in the critical region

TAAAAAGTAGAAAATATAATTCTAATTATTGCAC

112	T	C		1(1)	[44,1]	:	8(1)	[57,8]
213	G	C		1(1)	[4,1]	:	-2(1)	[6,8]
171	C	G		0(1)	[1,0]	:	1(1)	[-1,8]
301	C	C		1(1)	[1,181]	:	2(1)	[-1,113]
306	C	A		0(0)	[1,51]	:	1(1)	[-1,50]
222	G	G		1(0)	[1,2]	:	-1(2)	[3,6]
242	C	G		0(0)	[0,0]	:	0(0)	[9,8]
165	G	T		1(1)	[6,1]	:	0(1)	[13,3]
187	G	T		0(1)	[6,0]	:	2(1)	[13,-1]
229	T	T		92(9)	[4,321]	:	74(21)	[2,307]
149	T	G		1(1)	[4,3]	:	-3(3)	[2,6]
255	T	A		12(4)	[4,174]	:	9(4)	[2,157]
145	T	A		3(1)	[4,51]	:	6(5)	[2,50]
298	G	G		0(1)	[11,0]	:	-1(2)	[10,8]
207	G	T		21(2)	[11,140]	:	16(2)	[10,127]
249	A	A		0(1)	[14,1]	:	3(2)	[20,4]
246	A	T		0(0)	[14,1]	:	2(2)	[20,3]
279	A	C		6(0)	[14,181]	:	8(2)	[20,113]
170	C	T		1(1)	[41,1]	:	3(1)	[37,3]
223	C	A		20(1)	[41,51]	:	17(3)	[37,50]
235	A	C		0(0)	[1,1]	:	2(1)	[4,8]
241	T	G		0(0)	[16,21]	:	0(1)	[12,6]
194		GC		0(0)	[2,1]	:	0(1)	[6,8]
264		T	G	1(0)	[3,99]	:	1(1)	[2,89]
300		T	A	1(1)	[3,51]	:	6(3)	[2,50]
247			GC	6(1)	[99,26]	:	5(2)	[89,32]

(ii) One in the critical region

TAAAAAGTAGAAAATATAATTCTAATTATTGCAC

317	G		A	48(3)	[55,82]	:	57(7)	[56,93]
232	G		G	1(0)	[4,64]	:	4(2)	[6,66]
240	G	T		0(0)	[1,107]	:	1(1)	[3,93]
259	G	C		1(1)	[1,128]	:	3(1)	[3,96]
184	A	G		312(16)	[232,112]	:	200(12)	[250,79]
204	A		C	206(14)	[232,82]	:	151(9)	[250,79]
283	T		C	6(0)	[17,51]	:	8(1)	[21,49]
231	A	T		7(1)	[14,139]	:	11(2)	[20,113]
172	C	A		35(1)	[41,148]	:	31(4)	[37,109]
221	C		T	10(1)	[41,28]	:	13(2)	[37,38]
148	G	G		81(9)	[57,71]	:	90(12)	[63,84]
198	T		T	388(13)	[321,113]	:	249(12)	[307,85]
152	T	G		21(2)	[16,82]	:	26(4)	[12,66]
273	T	G		0(0)	[1,64]	:	11(2)	[3,66]
135	T	C		1(1)	[1,74]	:	14(1)	[3,59]
121	C	T		0(0)	[0,83]	:	8(2)	[6,83]
102	C	G		1(0)	[0,82]	:	1(1)	[6,66]
286	C	T		0(0)	[0,139]	:	2(1)	[0,113]
272	T	G		1(0)	[3,64]	:	5(2)	[2,42]
299	T	C		3(0)	[3,106]	:	3(1)	[2,88]
143	T	T		4(1)	[3,113]	:	9(2)	[2,105]
226	G	A		20(2)	[14,125]	:	21(1)	[12,124]
250	C	G		13(1)	[13,112]	:	14(3)	[11,79]
155	G	A		81(8)	[79,120]	:	190(9)	[83,219]
195	C	G		31(2)	[40,82]	:	29(4)	[45,66]
168	G	T		67(7)	[99,83]	:	52(7)	[89,83]
127	T	A		144(9)	[140,133]	:	129(2)	[127,129]
289	C	G		57(2)	[181,64]	:	45(5)	[113,42]
139	G	G		87(7)	[36,186]	:	126(13)	[31,171]
263	A	A		111(3)	[51,150]	:	126(11)	[50,177]
270	A	A		51(2)	[51,88]	:	52(3)	[50,81]
120	C	A		30(2)	[26,133]	:	33(1)	[32,129]
237	T	T		0(0)	[0,107]	:	0(1)	[-1,93]

(iii) Neither in the critical region

TAAAAAGTAGAAAATATAATTCTAATTATTGCAC

293	A	T		108(4)	[112,107]	:	95(7)	[127,93]
304	C	T		115(4)	[112,105]	:	104(10)	[127,88]
257	C	C		141(8)	[112,128]	:	142(14)	[127,96]
109	C	C		68(8)	[112,74]	:	56(7)	[127,59]
314	C	C		92(3)	[112,106]	:	72(7)	[127,88]
303	C	C		95(5)	[152,711]	:	100(5)	[126,84]
156	G	G		121(6)	[106,107]	:	102(6)	[120,93]
164	C	T		93(4)	[106,98]	:	104(9)	[120,83]
103	C	C		96(4)	[125,139]	:	94(8)	[111,113]
158	G	T		53(5)	[83,51]	:	39(4)	[83,49]
297	T	C		67(4)	[62,71]	:	61(5)	[63,84]
214	GG			47(3)	[62,105]	:	47(7)	[63,88]
215	GC			53(3)	[62,97]	:	57(7)	[63,99]
154	G	A		121(4)	[90,107]	:	116(13)	[102,93]
281	CT			89(5)	[107,139]	:	94(15)	[93,113]
239	TT			86(4)	[107,74]	:	64(5)	[93,59]
267	T	C		116(8)	[107,113]	:	94(10)	[93,105]
209	T	T		97(5)	[120,85]	:	310(7)	[219,98]
130	A	G		61(3)	[113,57]	:	49(7)	[85,54]
320	TC			155(6)	[113,115]	:	121(11)	[85,110]
256	T	G		113(7)	[106,115]	:	86(9)	[88,110]
196	C	G		158(10)	[122,95]	:	152(14)	[147,81]
366	A	G		2(0)	[113,69]	:	6(3)	[105,55]
367	T	C						

Fig. 4.

(b) TRIPLE SUBSTITUTIONS IN THE 7.5-KD PROMOTER

(i) All in the critical region

TAAAAAGTAGAAAATATATTCTAATTTATTGCAC

244 T C G . . . 0(1) [16,0,0] : -3(1) [17,6,-3]

(ii) Two in the critical region

TAAAAAGTAGAAAATATATTCTAATTTATTGCAC

260	G	C	C	G	4(1) [4,78,105]	: 7(2) [6,85,88]
225	G	G			1(0) [1,2,85]	: -3(1) [3,6,77]
295	C		C	G	0(0) [1,1,85]	: 1(1) [4,8,98]
280	G		T	A	0(0) [11,0,82]	: 3(1) [10,-1,93]
233	T	C		C	5(1) [321,1,74]	: 3(1) [307,8,59]
161	C	G	T		17(1) [78,36,83]	: 13(2) [85,31,83]
269		C G		G	40(2) [40,36,112]	: 34(4) [45,31,79]

(iii) One in the critical region

TAAAAAGTAGAAAATATATTCTAATTTATTGCAC

219	C		T	G	1(0) [0,139,85]	: 0(1) [9,113,98]
236	C		A	G	23(2) [41,148,85]	: 32(2) [37,109,98]
205	C		A	A	61(7) [78,148,88]	: 60(5) [85,109,81]
153		G		C G	51(3) [79,57,115]	: 47(5) [83,54,110]
265		T	G	G	132(4) [140,82,115]	: 82(4) [127,66,110]

(iv) None in the critical region

TAAAAAGTAGAAAATATATTCTAATTTATTGCAC

128	C	C	A		89(2) [112,90,82]	: 101(5) [127,102,93]
278		CC	C		61(3) [105,98,74]	: 53(6) [88,83,59]
253		CCG			44(2) [98,128,82]	: 32(3) [83,96,66]
347		A	T	A	187(8) [120,113,122]	: 204(16) [219,85,147]

(c) QUADRUPLE SUBSTITUTIONS IN THE 7.5-KD PROMOTER

(i) Two in the critical region

TAAAAAGTAGAAAATATATTCTAATTTATTGCAC

294	G	C	T	C	0(0) [6,0,83,108]	: 9(2) [13,0,83,102]
287	G	A	C	T	9(1) [57,51,105,28]	: 15(2) [63,50,88,38]

(ii) One in the critical region

TAAAAAGTAGAAAATATATTCTAATTTATTGCAC

274 T C CC . . . 268(8) [321,106,108,74] : 185(11) [307,120,102,59]

(iii) None in the critical region

TAAAAAGTAGAAAATATATTCTAATTTATTGCAC

339 A T AT . . . 120(7) [120,113,122,176] : 134(22) [219,85,147,162]

(d) DELETIONS IN THE 7.5-KD PROMOTER

(i) Single deletions

TAAAAAGTAGAAAATATATTCTAATTTATTGCAC

140	-				2(0) : 9(2)
271	-				0(0) : 0(2)
266	-				97(6) : 78(12)
288	-				109(7) : 105(7)
321	-				108(7) : 77(10)
310	-				113(7) : 91(9)

(ii) Double deletions

TAAAAAGTAGAAAATATATTCTAATTTATTGCAC

311	-	-			54(5) : 40(5)
329	-	-			61(2) : 49(6)
110		--			23(2) : 18(2)

(iii) Multiple deletions

TAAAAAGTAGAAAATATATTCTAATTTATTGCAC

180	-----				1(1) : 1(1)
230		----			10(1) : 10(3)

Fig. 4.

(e) SUBSTITUTIONS ACCOMPANIED BY DELETIONS IN THE 7·5-KD PROMOTER					
(i) Single substitutions					
<u>TAATAAGTAGAAAATATATTCTAATTATTGCAC</u>					
197	T---			0(0) [321]	: 0(1) [307]
238	A	--		0(0) [1]	: -1(1) [4]
227	-		A	0(1) [88]	: 1(1) [81]
258	G-			0(1) [2]	: 3(3) [6]
218	G	-		6(1) [14]	: 4(0) [12]
(ii) Double substitutions					
<u>TAATAAGTAGAAAATATATTCTAATTATTGCAC</u>					
201	C	-	C	0(0) [0,106]	: -1(1) [6,120]
277	-		T C	0(0) [139,74]	: 2(1) [113,59]
206		-	CG	59(3) [108,64]	: 35(4) [102,66]
(iii) Triple substitutions					
<u>TAATAAGTAGAAAATATATTCTAATTATTGCAC</u>					
132	-	G C		1(0) [14,181,106]	: 9(1) [12,113,88]
178		C C C	-	1(1) [181,1,106]	: 1(1) [113,8,120]
(f) MUTATIONS IN THE 7·5KD PROMOTER (LONGER SEQUENCE)					
(i) Substitution mutations					
<u>CGTAAAGTAGAAAATATATTCTAATTATTGCAC</u>					
8				137(5)	: 124(8)
7+				452(26)	: 440(35)
14			A A	83(3)	: 1229(75)
13+			A A	276(12)	: 1981(83)
(ii) Insertion mutations					
<u>CGTAAAGTAGAAAATATATTCTAATTATTGCAC</u>					
341			<TCGAATTTTAT>	15(2)	: 26(5)
340			<TCGATAAAAATT>	90(6)	: 168(13)
(g) STRONG EARLY PROMOTERS AND OTHER PROMOTERS BASED ON THE 7·5KD PROMOTER					
(i) Substitution mutations					
<u>TAATAAGTAGAAAATATATTCTAATTATTGCAC</u>					
361	AA	A A		127(6)	: 146(7)
362	AA T	A A		54(3)	: 49(3)
343	A T	AT		379(22)	: 329(28)
342	A T	AC		427(24)	: 427(38)
344	A T	AT	GATCT	303(19)	: 303(14)
345	A T	AC	GATCT	275(13)	: 298(28)
429	A T	AT	GATCT	344(11)	: 437(32)
(ii) Insertion mutations based on vMJ344					
<u>TAATAATGAAAAATTAGATCTATTATTGCAC</u>					
355	<A>			322(20)	: 222(14)
356	<AA>			306(16)	: 232(12)
357	<AAA>			381(21)	: 291(41)
358	<AAAA>			419(13)	: 318(16)
359	<TACAATAATTAAATTCTCG>			365(13)	: 285(13)
360	<AAAAATTGAAAAATTAGCT>			854(36)	: 745(42)
(iii) Insertion mutations based on vMJ4					
TCTCGAGCATG					
364	<AAAAATTGAAAAATTAGCT>	..		1(0)	: -1(1)
365	<AAAAATTGAAAAATTAGCTAAAAATTGAAAAATTAGCT>			407(11)	: 373(27)

Fig. 4.

for members of this set, and other independently constructed 7·5-kD promoter mutants, are listed in Figure 4(a), (b) and (c). With few exceptions, the level of β -galactosidase expressed by multiple substitution mutants corresponded approximately to the product of the levels expressed by the appropriate single substitution mutants. This indicates that most nucleotides in the promoter act independently. In addition to mutants containing alter-

ations in the initiation region (which might be expected in some cases to give anomalous results), some mutants deviated from this rule. The most serious deviation occurred for vMJ229, which was about seven times more active than expected (Fig. 4(a)(i)), behaving as if the substitution of -22(A) by T compensated for the very detrimental effect of the replacement of -25(A) by T.

Some of the multiple substitutions that led to

(h)	<u>OTHER EARLY PROMOTERS AND SUBSTITUTIONS THEREIN</u>	
(i)	<u>DNA polymerase gene</u>	
	<u>TAAATGAAAATATA</u> <u>TTCTAAATTCTATAC</u>	
27		23(2) : 47(5)
26+		113(5) : 149(9)
(ii)	<u>TK gene</u>	
	<u>TAAAGTGAACAATA</u> <u>TTAATTCTTATTGTCC</u>	
426		17(2) : 17(2)
32+		82(4) : 111(6)
425	A	189(12) : 193(13)
(iii)	<u>RNA polymerase large subunit gene</u>	
	<u>TAAAGTGTAAATATACTATTATTTTATAGTC</u>	
34		18(2) : 26(7)
33+		73(2) : 95(15)
(iv)	<u>19-kD gene</u>	
	<u>TATATTACTGAATTAATAATATAAAATTCCAATC</u>	
424		67(2) : 64(5)
420	A	153(9) : 137(8)
(v)	<u>RNA polymerase 22-kD subunit gene</u>	
	<u>TAAATTATGAAAAAAAGATGTACTACCTTATTAAAGAC</u>	
422		63(4) : 103(7)
421	T	39(3) : 100(9)

Figure 4. A list showing levels of β -galactosidase expressed by recombinants containing mutated early promoters. The identities of the recombinants (vMJ) are shown on the left. The few recombinants containing *lacZ* in the same orientation as the TK gene are indicated by a plus sign. The parental sequence for each group is shown with the critical region underlined and RNA initiation sites *in vivo* marked by dots. Substituted or deleted (-) nucleotides are indicated for each recombinant. Inserted nucleotides are denoted by angle brackets (< >), the left bracket situated below the nucleotide after which the insertion was made. β -Galactosidase results are given on the right relative to vMJ21 (100%), which contains the parental 7.5-kD promoter, with standard deviations in parentheses. Square brackets contain values for appropriate single substitutions contributing to multiple substitutions, and are given from left to right with respect to the parental sequence. Values to the left of the colon were obtained in the presence of araC, and those to the right were obtained in its absence.

generation of TAAAT elements resulted in late promoter activity. Examples are vMJ155 (Fig. 4(a)(ii)) and vMJ130 (Fig. 4(a)(iii)). Another example was vMJ14 (Fig. 4(f)(i)), in which two substitutions led to the addition of a strong late promoter to the early promoter.

In addition to multiple substitution mutants, 11 deletion mutants were obtained, in addition to ten mutants possessing substitutions as well as deletions. β -Galactosidase results for the former are given in Figure 4(d) and (e). The results highlight the importance of the critical region. Single nucleotide deletions in this region (e.g. -21(G) in vMJ140 and -20(A) in vMJ271), or deletions accompanying substitutions that in themselves did not abolish expression, inactivated the promoter. The effect of the deletion of -20(A) was particularly noteworthy, since it removed only one of a series of A residues. In contrast, single nucleotide deletions from -12(T) to +4(C) had little effect. Deletion of two nucleotides in the region from -12(T) to +4(C) reduced expression in vMJ329 and vMJ110, perhaps as a result of the greater decrease in distance between the critical region and the normal initiation sites. Deletion of eight nucleotides in the region between -12(T) and +4(C) effectively abolished expression (vMJ180). A small amount of RNA

produced *in vivo* was detected by primer extension, initiating at +11(A) (Fig. 5(b)). The mRNA thus contains no untranslated leader, and might be translated very poorly. The results for this mutant show that, although the sequence between -12(T) and +4(C) is not critical for transcription, gross changes can have a substantial effect. They also confirm that the site of initiation is determined primarily by the location of the critical region. Another mutant, vMJ230, lacks the four nucleotides from +5 to +8. RNA of the appropriate size was produced *in vitro* (data not shown), but may have been translated poorly *in vivo* owing to the presence of C three nucleotides upstream from the methionine codon; Kozak (1984) has indicated that a purine at this position results in more efficient translation of eukaryotic mRNAs.

(e) Effect of gene orientation on expression

The majority of recombinants were constructed with the promoter-*lacZ* sequence in the opposite orientation from the TK gene, but the results for several recombinants indicate that β -galactosidase expression was three- to fivefold greater when the promoter-*lacZ* sequence was in the same orientation

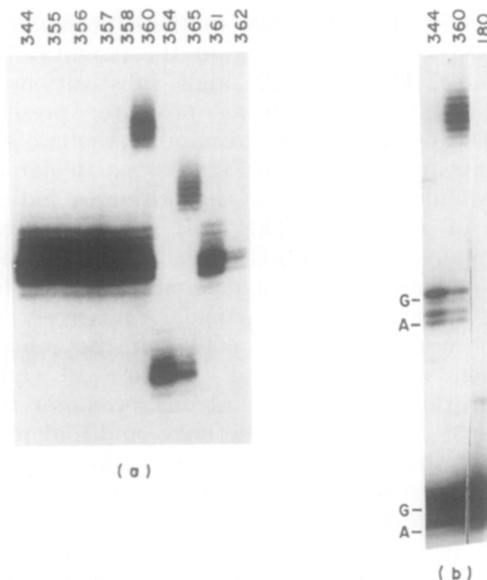


Figure 5. (a) An autoradiograph showing RNA synthesized *in vitro* from templates containing designed early promoters. The sequences of and β -galactosidase results for these promoters, designated at the top of the lanes according to the pMJ nomenclature, are shown in Fig. 4(g). The RNA in the lane on the right has a size of approx. 100 nucleotides. (b) An autoradiograph showing primer extended products from mRNA synthesized *in vivo* from selected early promoters. The sequences of and β -galactosidase results for these promoters, designated at the top of the lanes according to the vMJ nomenclature, are shown in Fig. 4(g) and (d). The upper group of bands is derived from *lacZ* mRNA, and the lower from 7.5-kD early mRNA. Marker mobilities corresponding to mRNAs initiated at +1(G) and +3(A) are indicated for each group.

as the TK gene. These recombinants are indicated by a + in Figure 4, and include the natural 7.5-kD promoter (vMJ8 and vMJ7, Fig. 4(f)(i)), one of its derivatives incorporating a late promoter (vMJ14 and vMJ13, Fig. 4(f)(i)), the DNA polymerase promoter (vMJ27 and vMJ26, Fig. 4(h)(i)), the TK promoter (vMJ426 and vMJ32, Fig. 4(h)(ii)) and the RNA polymerase large subunit promoter (vMJ34 and vMJ33, Fig. 4(h)(iii)). Analysis of mRNA expressed by recombinants containing the two orientations of *lacZ* under control of the 7.5-kD promoter (Fig. 2(c), vMJ7 and vMJ8), however, indicates that the two viruses express approximately equal amounts of *lacZ* mRNA (relative to mRNA expressed from the resident 7.5-kD genes) under conditions of cycloheximide block. The difference in β -galactosidase expression might be manifest at the mRNA level only in the absence of cycloheximide, or might have a post-transcriptional cause. The 5' ends of the mRNAs specified by the two recombinants are identical, but the 3' ends should differ, depending on the locations of early transcriptional terminator signals downstream from *lacZ* (Yuen & Moss, 1987). When *lacZ* is in the same orientation as the TK gene, the shortest predicted

mRNA is about 3350 nucleotides in size, and when it is in the opposite orientation the shortest predicted mRNA is 1000 nucleotides longer. It is possible, then, that the longer mRNA is either preferentially degraded in the absence of cycloheximide, or translated less efficiently. This observation may be relevant to the design of vaccinia virus expression vectors in which foreign genes are expressed under the control of early promoters; inclusion of an early terminator signal immediately downstream from the foreign gene might lead to enhanced antigen expression.

(f) Design of strong early promoters

The analysis of the 7.5-kD promoter allowed the effects of substitutions at each position to be quantified, and thus in principle facilitates the design of early promoters with a desired level of activity. Having identified the optimal nucleotide at each position for maximal expression, two strong early promoters were synthesized, differing from the natural 7.5-kD promoter at four positions in the critical region. The two promoters (in vMJ342 and vMJ343; Fig. 4(g)(i)) differed from each other in one nucleotide, and each was about four times as active as the 7.5-kD promoter in β -galactosidase expression. Replacement of five consecutive nucleotides at -12(T) to -8(A), in order to introduce a convenient unique *Bgl*II site, caused a slight decrease in activity (vMJ344 and vMJ345 in Fig. 4(g)(i)). The *Bgl*II site was utilized in subsequent mutation of the critical region to its left or the non-critical region to its right. *In vitro* and *in vivo* transcription results for vMJ344 are shown in Figure 5(a) and (b). The promoter in pMJ344 produced about 20 times more RNA *in vitro* than did pMJ21, which contains the original 7.5-kD promoter (data not shown), and four times more mRNA and β -galactosidase *in vivo*. This difference is similar to that found with the single substitution mutants, which contribute to the strong early promoters.

The analysis of the 7.5-kD promoter hinted that extension of the tract of A residues at the left end of the critical region might increase expression, so the effects of this in the strong promoter in pMJ344 are shown in Figure 4(g)(ii) and Figure 5(a) (pMJ355 to pMJ358). A minor increase in expression was noted. Similarly, insertion of the 19 bp normally present upstream from -29(T) in the natural 7.5-kD promoter had a minimal effect (vMJ359 in Fig. 4(g)(ii)), confirming that this sequence is unimportant in determining the level of expression.

(g) Compensatory effects of up-mutations and down-mutations

We noted above that one of the double substitution mutants, vMJ229 (Fig. 4(a)(i)) was more active than expected, raising the possibility that substitution of -22(A) by T counteracted the detrimental

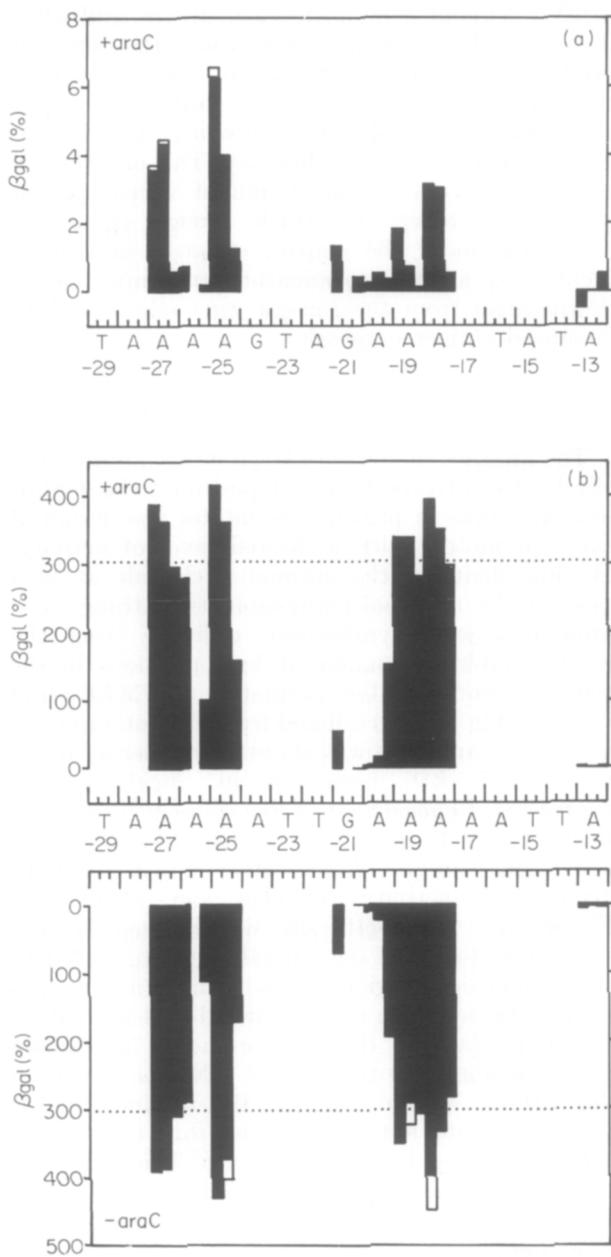


Figure 6. A histogram showing levels of β -galactosidase expressed by cells infected with recombinants containing single substitution mutations in (a) the 7.5-kD early promoter present in vMJ21 and (b) the strong early promoter present in vMJ344. Mean β -galactosidase levels are shown as filled columns relative to vMJ21 (100%). Standard deviations are shown as open columns when they exceeded (a) 0.15% of the mean level of expression by vMJ21 or (b) 10% of the mean level of expression by vMJ344. The horizontal dotted lines in (b) show the levels of expression by vMJ344. Expression in the presence of araC is shown above the sequence, and that in its absence is shown below.

effect of changing $-25(A)$ to T . In a preliminary investigation of this hypothesis, the 22 weakest single substitution mutants of the 7.5-kD promoter were reassayed for β -galactosidase expression under conditions designed to assess weak activity more

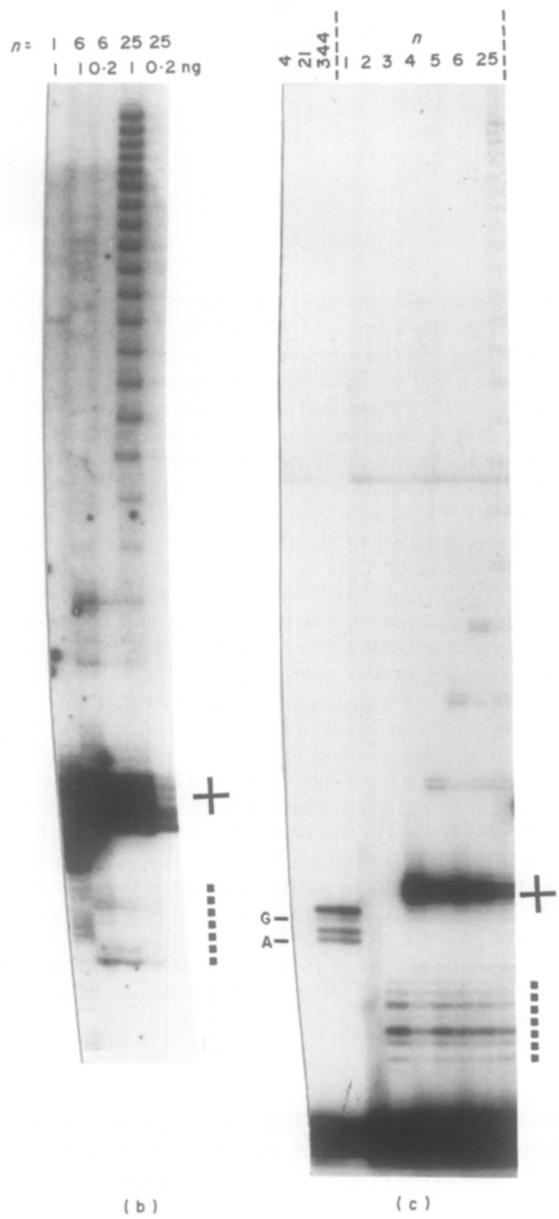
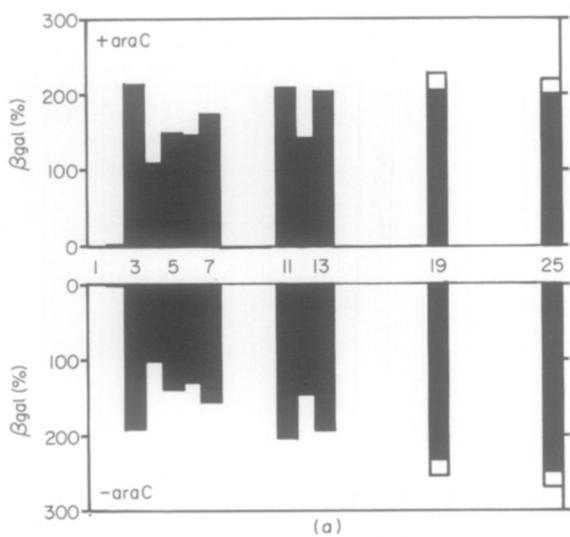
accurately. The strongest of these promoters had only 6.3% of the activity of the parental 7.5-kD promoter. The same 22 single substitutions were then made in the strong promoter present in vMJ344. β -Galactosidase results for the two sets of mutants are compared in Figure 6; particular attention should be given to the ordinates indicating levels of expression. The most striking finding is that many of the substitutions were only moderately detrimental, or not detrimental at all, to the strong promoter. In particular, relatively small effects were caused by many of the mutations between -25 to -27 and -18 to -19 . Only six substitutions reduced the strong promoter to less than 6.3% of its normal activity, and four reduced it to less than 1%. The persistent requirement for A at -13 was most notable. These findings support the hypothesis that one or more nucleotides in the strong promoter partially or totally suppress the effects of potentially detrimental nucleotides at some other positions. The precise locations of the suppressing nucleotides were not determined, as the strong promoter differs from the 7.5-kD promoter at four nucleotides in the critical region. As indicated above, a strong candidate is the presence of T at -22 .

Another illustration of compensatory effects came from the replacement of all nucleotides in the critical region, except $-21(G)$, by A residues. Despite the great reduction of promoter activity expected from changing $-23(T)$ and $-14(T)$, this promoter was slightly more active than the 7.5-kD promoter (vMJ361 in Fig. 4(g)(i); Fig. 5(a)). Even after replacing $-21(G)$ in this promoter with T (vMJ362 in Fig. 4(g)(i); Fig. 5(a)), the promoter remained half as active as the 7.5-kD promoter.

At this stage of the analysis, it appears that there is no single nucleotide, with the possible exception of the A residue at -13 , that is critical for promoter activity.

(h) Promoters with tandemly repeated sequences

Insertion of 19 bp containing an optimal critical region upstream from the optimal critical region already present in vMJ344 more than doubled β -galactosidase expression (vMJ360 in Fig. 4(g)(ii)). Analysis of RNA made *in vitro* (Fig. 5(a)) and *in vivo* (Fig. 5(b)) demonstrated, in addition to the transcripts directed by the original critical region, the synthesis of longer transcripts directed by the inserted sequence. A difference between the balance of RNAs made *in vitro* and *in vivo* was noted; relatively more RNA was specified *in vivo* by the inserted upstream sequence. The functional independence of the critical region from the $-12(T)$ to $+4(C)$ sequence was demonstrated by inserting one or two copies of the optimal critical region into the vMJ4 sequence (vMJ4 contains no promoter at the 5' end of lacZ), to give vMJ364 and vMJ365 (Fig. 4(g)(iii)). The former specified a single set of transcripts *in vitro* and the latter generated two sets



(Fig. 5(a)). The 5' ends of each set of transcripts mapped an appropriate distance from the relevant critical region. Only the longer transcripts specified by vMJ365 expressed β -galactosidase, however, as the shorter transcripts originated within the initiation codon for β -galactosidase.

Having shown that two adjacent critical regions are able to function independently, a set of promoters containing repeats of an 11 bp sequence (GAAAAAAAATT) was constructed. Two adjacent copies of this sequence constitute a single critical region, and each additional copy results in another critical region. The precise construction of this set of promoters is shown in the legend to Figure 7. High levels of β -galactosidase expression did not occur until $n = 3$ (Fig. 7(a)). Expression decreased when $n = 4$, and gradually increased again as more repeats were added. Even when $n = 25$, however, the level of β -galactosidase was not much higher than when $n = 3$.

Transcriptional analyses were carried out in order to determine the sites of initiation of RNA synthesis. Analysis of *in vivo* transcripts by primer extension when $n = 1, 2, 3, 4, 5, 6$ and 25 is shown in

Figure 7. (a) A histogram showing levels of β -galactosidase expressed by cells infected with recombinants containing reiterated, overlapping critical regions. The series was based on the sequence:

BamHI

GGATCCGTCGAAAAAAAATT(GAAAAAAAATT)_n

KpnI XbaI

+
GCTTACCCCTCGAGCATG, $n = 1$ to 25.

The value of n is plotted on the abscissa. Mean β -galactosidase levels are shown as filled columns, relative to vMJ21 (100%). Standard deviations are shown as open columns when they exceeded 10% of the mean level of expression by vMJ21. Expression in the presence of araC is shown above the abscissa, and that in its absence is shown below. (b) An autoradiograph showing RNA synthesized *in vitro* from templates containing reiterated, overlapping critical regions. Values of n , and the amounts of template used in the reaction, are given at the top of the lanes. Transcripts corresponding to initiation at the sites marked in the sequence shown above are indicated to the right. The major transcript (+) has a size of approx. 100 nucleotides. The larger species are due to initiation at the G residue of the reiterated sequence, and differ in size by increments of 11 nucleotides. (c) An autoradiograph showing primer extended products from mRNA synthesized *in vivo* from recombinants containing reiterated critical regions. Values of n are given at the top of the lanes. Lanes marked 4, 21 and 344 show results for vMJ4 (no promoter), vMJ21 (parental 7.5-kD promoter) and vMJ344 (strong early promoter). Transcripts corresponding to initiation at the sites marked in the sequence shown above are indicated to the right. The larger species are due to initiation at the G residue of the reiterated sequence, and differ in size by increments of 11 nucleotides.

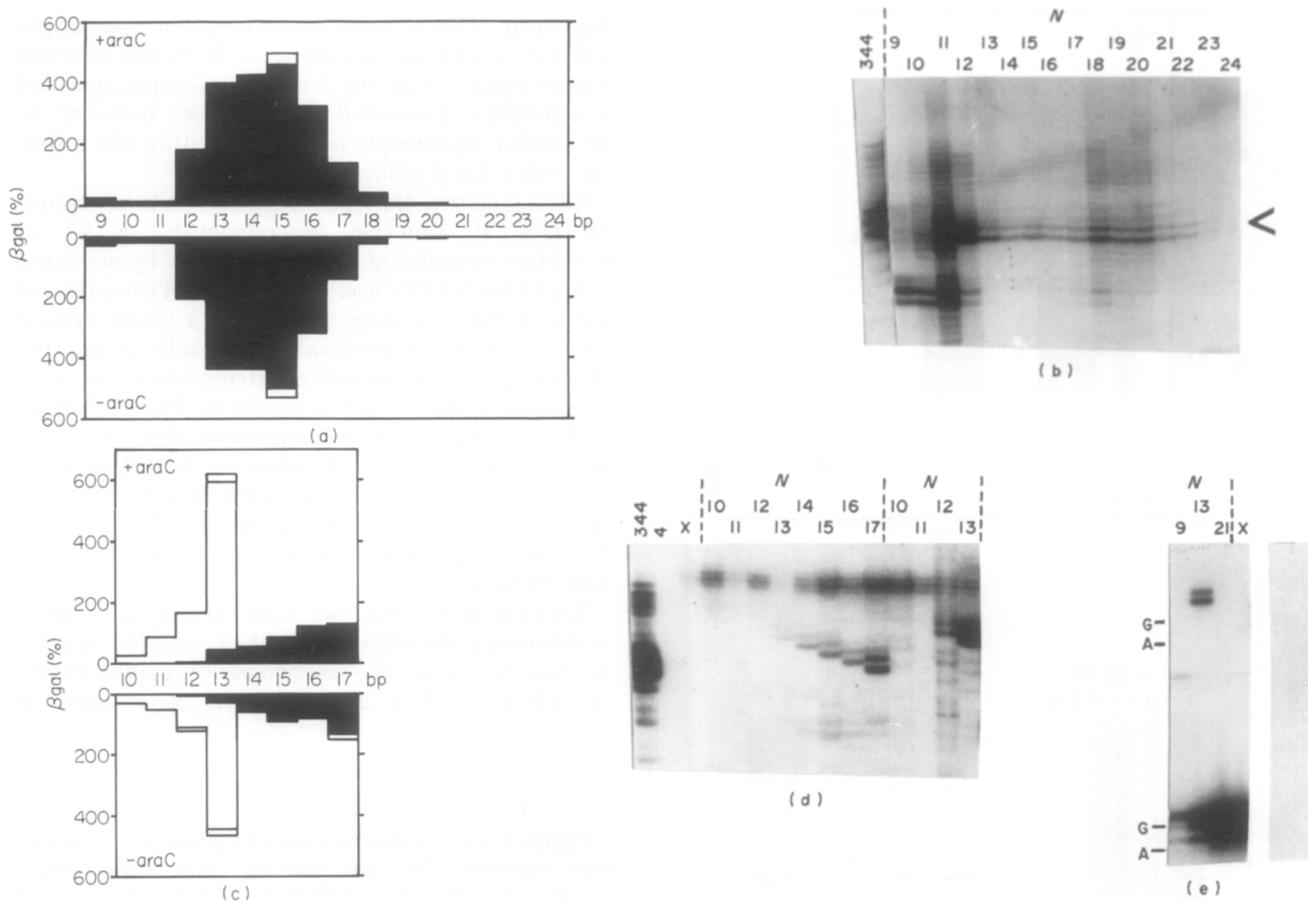


Figure 8. (a) A histogram showing levels of β -galactosidase expressed by cells infected with recombinants containing an isolated purine in the initiation region (1st set). The series was based on the sequence GATCT_nGT₆CGAGCATG ($n = 4$ to 19) immediately downstream from the critical region present in vMJ344. N , the number of nucleotides between the 3' end of the critical region (-13) and the isolated G residue, is plotted on the abscissa, and corresponds to $n+5$. Mean β -galactosidase levels are shown as filled columns relative to vMJ21 (100%). Standard deviations are shown as open columns when they exceeded 10% of the mean level of expression by vMJ21. Expression in the presence of araC is shown above the abscissa and that in its absence is shown below. (b) An autoradiograph showing RNA synthesized *in vitro* from templates containing an isolated purine in the initiation region (1st set). Values of N are given at the top of the lanes. Lane 344 shows results for pMJ344. Radiolabel was added as [α -³²P]CTP. The arrow indicates RNA (approx. 100 nucleotides) initiated at the isolated G residue. (c) A histogram showing levels of β -galactosidase expressed by cells infected with recombinants containing an isolated purine in the initiation region (2nd set). The series was based on the sequence GATCT_nGT_{18-n}CGAGCATG ($n = 5$ to 12) and GATCT_nGT₃GT_{14-n}CGAGCATG ($n = 5$ to 8) immediately downstream from the critical region present in vMJ344. N , the number of nucleotides between the 3' end of the critical region (-13) and the isolated G residue, is plotted on the abscissa, and corresponds to $n+5$. Mean β -galactosidase levels, relative to vMJ21 (100%), are shown as filled columns for the former subset, and as open columns for the latter subset. Standard deviations are shown when they exceeded 10% of the mean level of expression by vMJ21. Expression in the presence of araC is shown above the abscissa, and that in its absence is shown below. (d) An autoradiograph showing RNA synthesized *in vitro* from templates containing an isolated purine in the initiation region (2nd set). Results for the series GATCT_nGT_{18-n}CGAGCATG ($n = 5$ to 12; $N = 10$ to 17) are shown to the left of those for the series GATCT_nGT₃GT_{14-n}CGAGCATG ($n = 5$ to 8; $N = 10$ to 13). Values of N are given at the top of the lanes. Lanes 344, 4 and X show results for pMJ344, pMJ4 (no promoter) and GATCT₁₉CGAGCATG (no isolated purine). Radiolabel was added as [α -³²P]CTP. The major RNA species in lane 344 has a size of approx. 100 nucleotides. The RNA species in adjacent lanes that differ in size by 1 nucleotide increments are due to transcriptional initiation at the isolated G residue. (e) An autoradiograph showing primer extended products from mRNA synthesized *in vivo* from recombinants containing an isolated purine in the initiation region of the 1st set (GATCT_nGT₆CGAGCATG) for 3 values of N . Lane X shows results for GATCT₁₉CGAGCATG (no isolated purine). The upper bands are derived from lacZ mRNA, and the lower bands from 7.5-kD early mRNA. Marker mobilities corresponding to mRNAs initiated at +1(G) and +3(A) are indicated for lacZ under control of the parental 7.5-kD promoter and the resident 7.5-kD genes.

Figure 7(c). When $n = 1$, a complete critical region was absent, and no specific transcripts were detected. When $n = 2$, a single critical region was present, and several transcripts were detected (indi-

cated by the dots in Fig. 7(c)), corresponding to purines within and just upstream from the translation initiation codon for β -galactosidase. This may explain the very poor translation of these mRNAs

(see $n = 2$ in Fig. 7(a)). When $n = 3$, however, the major transcription site occurred at the G residue indicated by a cross in Figure 7(c). As n increased, a series of longer RNAs of approximately equal abundance was revealed by primer extension. Each of these was due to initiation at the G residue of the 11 bp repeat. The major transcription site (denoted by a cross in Fig. 7(c)) was the same for all recombinants for values of n from 3 to 25, but its relative contribution to overall *lacZ* mRNA production gradually decreased as n increased.

Similar results were obtained when the plasmids employed to generate this series of recombinant viruses were used as templates for *in vitro* transcription (Fig. 7(b)), taking into account the fact that *in vitro* transcripts were uniformly labelled internally, whereas *in vivo* transcripts were reverse transcribed using an end-labelled primer.

The results for this set of promoters emphasize the ability of overlapping critical regions to direct transcriptional initiation from independent sites located an appropriate distance downstream (14 bp in the repeated sequence). The preferred use of the major transcription initiation site suggests that complex interactions, possibly of a steric nature, may occur when transcription is directed by overlapping critical regions.

(i) Analysis of other early promoters

Several other vaccinia virus early promoters were tested in order to ascertain the general applicability of findings obtained from the 7.5-kD promoter analysis. β -Galactosidase results are shown in Figure 4(h). The sequences corresponding to the critical region for these promoters are located an appropriate distance upstream from the initiation sites *in vitro*, and show a good fit with the critical region of the 7.5-kD promoter and its active derivatives. Each promoter, however, contains at least one nucleotide that would be expected to be severely detrimental to expression. The DNA polymerase promoter as constructed has C at -28, T at -27 and A at -23, the TK promoter has T at -27, G at -23 and C at -18, the RNA polymerase large subunit promoter has T at -21, the 19-kD promoter has T residues at -26, -25 and -18, and the RNA polymerase 22-kD subunit promoter has T residues at -25 and -24 and G at -13. Certain of these nucleotides are at least in part responsible for the weakness of these promoters compared with the 7.5-kD promoter, since substitution of selected nucleotides in the TK and 19-kD promoters by potentially favourable nucleotides resulted in substantially increased expression (vMJ425 and vMJ420 in Fig. 4(h)(ii) and (iv)). The fact that the parental promoters were considerably more active than anticipated from the single nucleotide substitution analysis of the 7.5-kD promoter may be explained, as in the strong 7.5-kD promoter, by compensatory up-mutations that alleviate the potentially detrimental effects of nucleotides at

other positions. In this connection, it is provocative to note that each promoter except the 7.5-kD promoter has T at -22; this substitution at -22(A) in the 7.5-kD promoter was implicated above in suppression of a detrimental nucleotide at -25. The activity of the 22-kD promoter is an enigma, however (Fig. 4(h)(v)). Substitution of -13(A) by G in both the original and strong 7.5-kD promoter abolished expression, and yet the 22-kD promoter has G at this position. One possible explanation is the presence of A at -12 in the 22-kD promoter, which might substitute for the absence of A at -13. Replacement of A at -14 by T did not have a severe effect on the 22-kD promoter (Fig. 4(h)(v)).

(j) Promoters affecting the site of transcriptional initiation

Two sets of promoters were studied in order to define precisely the region in which transcriptional initiation can occur. Sequences were inserted between the *Bgl*II and *Xho*I sites of pMJ344 (Fig. 4(g)(i)), thus maintaining the critical region from the strong promoter. The first set contained the sequence GATCT_nGT₆CGAGCATG ($n = 4$ to 19) immediately downstream from the highly conserved -13(A). This choice of sequence placed a single purine, with a potential for initiation, in a tract of pyrimidines. We define N as the distance, in nucleotides, between -13 and the G residue; thus, $N = n + 5$. Figure 8(a) shows that efficient β -galactosidase expression occurred when the G initiation site was 12 to 18 nucleotides downstream from -13(A). Expression occurred most efficiently when $N = 13$ to 15. It is evident from Figure 8(a) that a small amount of β -galactosidase was made when $N = 9$ to 11. The *in vitro* and *in vivo* RNA results in Figure 8(b) and (e) show that transcripts directed by these promoters, instead of initiating at the isolated G (as indicated by the arrow) initiated at the next purine downstream, located between 17 and 19 nucleotides from -13, to give shorter RNA species. The lower level of β -galactosidase produced by these three recombinants was probably due, in part, to the initiating purine being rather too far downstream for optimal expression, and possibly to a shorter 5' untranslated leader in the mRNA. The behaviour of this series of promoters differed somewhat *in vitro* and *in vivo*. Figure 8(b) shows that RNA was synthesized *in vitro* from the isolated purine even when $N = 24$, whereas no β -galactosidase was detected *in vivo*. Initiation also occurred at the next purine downstream when it was 20 nucleotides from -13. Moreover, there was evidence of initiation at the isolated purine when it was only 11 nucleotides from -13, whereas very little β -galactosidase was detected. In addition, initiation was detected upstream from the isolated purine *in vitro* but not *in vivo*. These findings indicate that the positional requirements for initiation may be rather less rigid *in vitro* than they are *in vivo*.

The second set of promoters contained the sequence GATCT_nGT_{18-n}CGAGCATG ($n = 5$ to 12) immediately downstream from -13(A). Again, N indicates the distance in nucleotides between 13 and the G, and corresponds to $n+5$. Unlike the first set of promoters, in which the G was located in a pyrimidine tract of variable length, in this set the G was located at a variable position within a pyrimidine tract of constant length. In the first set, transcripts initiated at the G were identical for all promoters, commencing with GU₆, but in the second set they depended on value of n , commencing with GU_{18-n}. The β -galactosidase results for the second set are shown as the filled histogram in Figure 8(c). They differ markedly from those for the first set in Figure 8(a); the level of expression was generally much lower, and increased as N increased to 17 instead of reaching a maximum at $N = 15$. The data from the first and second set of promoters concur, however, regarding the location of the transcription initiation region. The *in vitro* RNA results in Figure 8(d) indicate that initiation occurred on the isolated G, transcripts decreasing in size by single nucleotides as N increased in steps of one. The sequences for $N = 17$ in the two sets of promoters were identical, and levels of expression were very similar, as expected (compare Fig. 8(a) and (c)). When $N = 16$, however, the sequences differed only in that the member of the second set had an additional T in the tract downstream from the initiating G, so that the RNA commenced with GU₇ instead of GU₆. This apparently small difference resulted in a fivefold reduction in β -galactosidase expression. Greater relative reductions were observed as N decreased to 12, and the number of U residues at the 5' ends of the RNAs specified by the second set of promoters increased concomitantly to 11.

The strategy used to obtain the second set of promoters led to the isolation of five additional members. One contained the sequence GATCT₁₉CGAGCATG downstream from -13(A); that is, it lacked an isolated purine. As expected, this promoter, which contains a competent critical region, produced no β -galactosidase *in vivo* or transcript *in vitro* (Fig. 8(d), lane X). However, a detectable level of transcript was generated *in vivo*, commencing at the first G residue downstream from the T tract (i.e. 25 nucleotides from -13) (Fig. 8(e), lane X). This indicates that, although the initiation region is normally located 12 to 18 nucleotides from -13, under certain circumstances initiation *in vivo* may occur inefficiently further downstream. This is illustrated by this unnatural promoter and by the observation that the parental 7.5-kD promoter initiated weakly 19 nucleotides from -13, as described above. The other four promoters generated in the second set contained two potential initiating purines in the pyrimidine tract, rather than one; GATCT_nGT₃GT_{14-n}CGAGCATG, $n = 5$ to 8. The *in vitro* RNA results are shown in Figure 8(d). The β -galactosidase results are shown in Figure 8(c) as the open histogram. Thus,

sequences at a particular value of N giving the activities indicated by the open and filled columns differed in a single nucleotide (the 2nd of the 2 G residues in the T tract). The single nucleotide difference when $N = 13$ resulted in a 20-fold increase in β -galactosidase expression. The *in vitro* RNA data (Fig. 8(d)) indicated that this was a transcriptional effect, and primer extension of *in vivo* RNA (data not shown) showed that the presence of two G residues instead of one resulted in two major mRNA species commencing with GU₃GU₆ and GU₆ instead of GU₁₀, both much more abundant than the GU₁₀ transcript. Thus, a tract of more than six T residues immediately downstream from the initiating purine appeared to be significantly detrimental to mRNA and β -galactosidase expression. The detrimental effect of a tract of T residues appeared to increase with its length. It is notable that such sequences form a subset of the early transcriptional termination signal (T₅NT) identified by Yuen & Moss (1987), and that each addition of a T to a tract of six T residues results in an additional terminator. It is reasonable to hypothesize, therefore, that the behaviour of the second set of promoters was influenced by the presence of transcriptional terminators.

Additional data concerning the effect of a terminator on expression are available from the 7.5-kD promoter mutants. Recombinants vMJ340 and vMJ341, whose promoter sequences and β -galactosidase results are shown in Figure 4(f)(ii), contained an insertion of 12 bp in either orientation close downstream from the initiation sites. The latter contained an early terminator within the inserted sequence, and was sixfold less active than the former. The noting of a similar difference at the mRNA level (data not shown) indicates that the effect occurred at a transcriptional level. Results for two other 7.5-kD promoter mutants suggest that the terminator must be in the mRNA coding region in order to reduce expression significantly. The promoters in vMJ347 and vMJ339 (Fig. 4(b)(iv) and c(iii)) differ in a single nucleotide; the latter possessing a terminator just upstream from the initiation sites. The former promoter is only 50% more active than the latter.

One way in which a terminator might reduce expression is by causing premature termination, rather than by influencing initiation directly. Several lines of evidence failed to support this explanation (data not shown). The results of Yuen & Moss (1987) imply that the prematurely terminated mRNA, excluding the 3' poly(A) tract, would be about 50 nucleotides in size. The 5' end of the 20 nucleotide primer used for primer extension would anneal to nucleotides 28 to 47 of lacZ mRNA, and yet the relative amount of primer extended products were the same when a 20 nucleotide primer annealing to nucleotides 67 to 86 was used. A prematurely terminated *in vivo* or *in vitro* RNA was not detected, and the relative amounts of full-length RNA made *in vitro* corresponded approximately to those made *in vivo*. Therefore, it is possible that a

terminator close to the 5' end of the mRNA acts as an attenuator by reducing the frequency of initiation events or the initial steps in elongation, rather than by causing premature termination. Alternatively, the prematurely terminated RNAs may have been unstable or even shorter than the expected 50 nucleotides.

4. Discussion

This study documents a detailed analysis of the effects of nucleotide sequence on the function of vaccinia virus early promoters. The basic strategy was to place synthetic promoters upstream from a reporter gene encoding β -galactosidase, and recombine this cassette into the vaccinia virus genome. Two types of assay were used to examine expression *in vivo*. The β -galactosidase assay was relied upon most heavily, and was found, except in special cases, to reflect the level of mRNA produced *in vivo*, as determined by a primer extension assay. It should be stressed that these assays measured amounts of β -galactosidase and mRNA present at the time of harvesting infected cells, and thus the results pertain to steady-state levels. Many promoters, however, were assayed for their ability to direct specific transcription *in vitro*, and, in the majority of cases, they behaved similarly *in vitro* and *in vivo*. Some instances where transcription *in vitro* differed from that *in vivo* were noted; particularly for the series of sequences constructed in order to locate the initiation region (Fig. 8).

The vaccinia virus early promoter may be conceived of in terms of three elements; a critical region separated from a 7 bp initiation region by an 11 bp spacer region. There is as yet no indication of the involvement of sequences further upstream or downstream in the regulation of early mRNA synthesis. The critical region is defined as that region upstream from the initiation site within which single substitution mutations may have a drastic effect on transcription; it comprises 16 bp in the 7·5-kD promoter. A parallel may be drawn between it and the TATA element characteristic of eukaryotic genes transcribed by RNA polymerase II. The critical region in the 7·5-kD promoter corresponds to a region recognized by a protein present in vaccinia virions. Yuen *et al.* (1987) demonstrated a specific interaction with the 19-kD early promoter using an agarose gel mobility shift assay, and defined the sequences in the 19-kD and 7·5-kD promoters bound by the protein in DNase protection experiments. The molecular weight of the factor was estimated as 150,000 in glycerol gradients. The inability of two transcriptionally inactive 7·5-kD promoter mutants to bind the factor indicated that it is involved in transcriptional regulation, possibly in initiation. Subsequently, Broyles *et al.* (1988) purified the factor and demonstrated that it is required for transcription and contains an associated DNA-dependent ATPase activity.

Within the critical region of the 7·5-kD promoter,

single nucleotide substitutions at two sites substantially increased expression, at two sites moderately increased expression, and at 12 sites drastically reduced expression as measured by all three assays. Combining several up-mutations appeared to have a greater effect on expression *in vitro* than *in vivo*. Thus, with the "optimized" critical region (-28 to -13) AAAAATTGAAAAACTA, β -galactosidase expression was increased about fourfold, whereas *in vitro* transcription increased more than 20-fold. This plateau effect *in vivo* may mean that other parameters become limiting. It is possible that a second round of base substitutions would further improve expression. For example a T substitution at -23 is beneficial when A is present at -22 in the natural 7·5-kD promoter (Fig. 1). However, T at -23 might not be optimal when T is also substituted for -22(A).

Mutations at most locations in the 7·5-kD promoter behaved independently, but it was clear from mutational analysis of an optimized 7·5-kD promoter that the presence of certain nucleotides may compensate for potentially detrimental nucleotides at other positions within the critical region. Indeed, the full mutational analysis indicates that, in the appropriate contextual sequence, there appears not to be a single essential, unchangeable nucleotide in the early promoter. The flexibility of the early promoter sequence is probably due to stabilization by certain nucleotides of the complex formed between the critical region and the transcription factor. The best candidate for such a nucleotide is a T residue at -22.

The conclusions derived from the mutational analysis of the 7·5-kD promoter were supported by limited studies of other promoters. Thus, as predicted, a single nucleotide change in the weak TK promoter increased expression more than tenfold. Analysis of other promoters also reinforced the idea of sequence flexibility. It was most interesting to compare the sequences of other early promoters with that of the 7·5-kD promoter and its mutations. Alignment of early promoter sequences, however, was hindered by the imprecise locations of mRNA initiation sites in many cases. In deriving the alignment of early promoters shown in Figure 9, therefore, we placed greatest weight on nucleotide elements within the critical region. The most frequent nucleotide at each position within the critical region, when present in at least eight of the 19 promoters, is indicated near the foot of Figure 9. Below this consensus is shown the strong early promoter sequence derived from single nucleotide substitutions of the 7·5-kD promoter. Remarkably, these two sequences differ at only two positions, -15 and -23. The deviation at -15 is not significant, since substantial expression occurred when any nucleotide was substituted for -15(A) in the 7·5-kD promoter (Fig. 1). -23(T) may be beneficial for the 7·5-kD promoter because it lacks a T at -22. Each of the other four promoters in Figure 9 that lacks a T at -22 has a pyrimidine at -23. The best conserved nucleotides in the early promoters shown

	Critical region	Spacer region	Initiation region		
	-28	-13 -12	-2	-1	+6
	:	:	:	:	:
7.5-kD	AAAAGTAGAAAAATA	TTCTAATTAT	T <u>GCA</u> CGG		
DNA pol	TTTAATGAAATA	TTTCTAAATTC	T <u>ATAAAT</u>		
TK	ATAAAGTGACAATA	TTAATTCTT	T <u>TGT</u> CAT		
RNA pol	TTAAAGTGTAAATA	ACTATTATTT	T <u>ATAG</u> TT		
19-kD	TATTA <u>CTGAATTA</u> A	ATATAAAATTC	CC <u>AA</u> CT		
22-kD	AAATTTGAAAAAAG	ATGTA <u>CTACCT</u>	TA <u>ATAA</u> G		
42-kD	AAACACATAAAAATA	GC <u>GT</u> AACTAAT	A <u>AGACAA</u>		
37-kD	GATAACTGAAAAAAA	TTTAT <u>TGTT</u> T	T <u>GTT</u> TATT		
87-kD	TATTA <u>ATGAAAGT</u> A	A <u>ATAA</u> TTTT	T <u>ATTA</u> CA		
H3'	CAGAATTGAAAACGAA	AT <u>GAAGAT</u> C	T <u>AGGC</u> CAG		
H6	AAAAAAATGAAAATAA	TACAA <u>AGG</u> TT	TT <u>GAGG</u>		
D1	AGTA <u>AAATGAA</u> AAA	CT <u>AGTC</u> GT	T <u>AA</u> ATAAA		
D4	GGAA <u>ATGAAAGG</u> TA	CT <u>AGA</u> T <u>AG</u> TA	T <u>AAA</u> AAAG		
D5	CTT <u>TA</u> GTGAA <u>AT</u> TTA	ACT <u>TG</u> T <u>GT</u> CT	A <u>AA</u> ATGGA		
D9	TAAAAA <u>ATGAA</u> ATGTA	ACT <u>GT</u> TTAAA	A <u>TA</u> AGC		
D12	TA <u>ATAA</u> TGAAACAAA	CT <u>ATAGA</u> G <u>TT</u> G	T <u>AA</u> ATG		
I3	AT <u>AA</u> ACTGAAATA	T <u>ATC</u> ATTATAT	T <u>ACA</u> AAAG		
M1	AGAA <u>ATGAA</u> ATAGAA	AT <u>ATTA</u> ATTT	T <u>TAC</u> TA		
N2	AATA <u>ACAT</u> AAAATAA	T <u>ATTA</u> TTTT	AG <u>GA</u> ATTC		
Consensus	AAAAAA <u>ATGAA</u> AAA	T			
Strong promoter	AAAAAA <u>TTGAA</u> ACTA	T			

Figure 9. Alignments of vaccinia virus early promoter sequences with structural elements of the 7.5-kD early promoter. Published initiation sites are underlined. With the exception of the 7.5-kD promoter, they may not be accurate to the nucleotide. A consensus for the critical region was derived using the requirement that each nucleotide should be conserved in at least 8 of the 19 promoters. It is aligned at the foot of the Figure with the sequence of the critical region in the strong early promoter designed using nucleotide substitutions that increase expression of the 7.5-kD promoter. References are as follows: 7.5-kD, Venkatesan *et al.* (1981); DNA pol, Earl *et al.* (1986); TK, Weir & Moss (1983); RNA pol and 22-kD, Broyles & Moss (1986); 19-kD and 42-kD, Venkatesan *et al.* (1982); 37-kD, Slabaugh *et al.* (1988); 87-kD, Tengelsen *et al.* (1988); H3' and H6, Rosel *et al.* (1986); D1, D4, D5, D9 and D12, Lee-Chen *et al.* (1988); I3, Schmitt & Stunnenberg (1988); M1 and N2, Tamin *et al.* (1988).

in Figure 9 are the A residues at -20 (all promoters), -19 (all but the RNA polymerase promoter) and -13 (all but the 22-kD promoter). The 22-kD promoter, however, has A residues at every other position from -20 to -12, some of which may have compensatory effects.

In contrast with the critical region, which tends to be A-rich, the spacer region in natural promoters tends to be T-rich. Moreover, the sequence of the spacer region has only a limited effect on promoter strength. No mutation within this region was found to abolish transcription. Indeed, each copy of the critical region within promoters containing repeated, and even overlapping, copies was able to initiate mRNA synthesis independently at an appropriate distance downstream. These observa-

tions are in support of a relatively simple model for transcriptional initiation, in which the critical region binds a transcription factor, thus allowing entry of the RNA polymerase complex. The latter might associate with the factor before or after the factor has bound to DNA. Initiation then takes place at a certain distance from the binding site, dictated by the stereochemistry of the complex formed between transcription factor, RNA polymerase and template DNA. It is clear that the location of the critical region defines the initiation site; it does not form a site for funnelling of RNA polymerase to an initiation site of specific sequence at a variable distance downstream.

The region in which initiation can occur efficiently has a size of 7 bp *in vivo*, but can be considerably larger *in vitro*. The reasons for this difference are unknown, but may hinge on the higher order of molecular structure present in the virion than is present in the *in vitro* system. In this study, with only one possible exception, initiation was confined to purines. Indeed, initiation did not occur in the initiation region if purines were absent, even though a critical region was present. The sequence at the initiation site has an effect on promoter strength owing to several features; the presence of A and G residues, their locations with respect to the critical region, and their context. Most natural promoters have more than one purine in the initiation region, and the rules governing the relative use of each initiation site appear complex.

The data indicate that there is, in principle, a further level at which transcriptional regulation might occur. The presence of early transcriptional terminator signals close downstream from the initiation site significantly reduced the level of transcription. Additional experiments are necessary to determine whether this is due to a direct effect on the initial stages of mRNA synthesis. The observation that attenuation occurred only when the terminator was capable of being transcribed is in accord with the work of Shuman & Moss (1988), who concluded from the effects of substituted nucleoside triphosphate derivatives on termination *in vitro* that the terminator is recognized in the mRNA. There is as yet, however, no example of a natural promoter regulated by attenuation. Examination of the sequence of the vaccinia virus DNA polymerase gene indicates that there is a potential termination signal approximately 50 nucleotides downstream from the transcription initiation site (Earl *et al.*, 1986). This signal is located considerably further downstream than the terminators present in the promoters above. It is not known whether this terminator serves to attenuate DNA polymerase expression.

Beyond the level of promoter sequence, the possibility remains open that the level of expression of a particular early gene may be influenced by its location with respect to adjacent genes. This aspect was not investigated in this study, but would be facilitated by the availability of dominant selection systems for inserting sequences into any part of the

vaccinia virus genome not essential for growth in tissue culture (Franke *et al.*, 1985; Falkner & Moss, 1988).

The authors thank Norman Cooper for supplying their cell culture needs.

References

- Ahn, B. Y. & Moss, B. (1989). *J. Virol.* **63**, 226–232.
- Barbosa, E. & Moss, B. (1978). *J. Biol. Chem.* **253**, 7692–7697.
- Baroudy, B. M. & Moss, B. (1980). *J. Biol. Chem.* **255**, 4372–4380.
- Bertholet, C., Stocco, P., Van Meir, E. & Wittek, R. (1986). *EMBO J.* **5**, 1951–1957.
- Bertholet, C., Van Meir, E., ten Heggeler-Bordier, B. & Wittek, R. (1987). *Cell*, **50**, 153–162.
- Broyles, S. S. & Moss, B. (1986). *Proc. Nat. Acad. Sci., U.S.A.* **83**, 3141–3145.
- Broyles, S. S. & Moss, B. (1987). *Mol. Cell. Biol.* **7**, 7–14.
- Broyles, S. S. & Moss, B. (1988). *J. Biol. Chem.* **263**, 10761–10765.
- Broyles, S. S., Yuen, L., Shuman, S. & Moss, B. (1988). *J. Biol. Chem.* **263**, 10754–10760.
- Chakrabarti, S., Brechling, K. & Moss, B. (1985). *Mol. Cell. Biol.* **5**, 3403–3409.
- Cochran, M. A., Puckett, C. & Moss, B. (1985). *J. Virol.* **54**, 30–37.
- Cooper, J. A. & Moss, B. (1979). *Virology*, **96**, 368–380.
- Cooper, J. A., Wittek, R. & Moss, B. (1981). *J. Virol.* **37**, 284–294.
- Coupar, B. E. H., Boyle, D. B. & Both, G. W. (1987). *J. Gen. Virol.* **68**, 2299–2309.
- Earl, P. L., Jones, E. V. & Moss, B. (1986). *Proc. Nat. Acad. Sci., U.S.A.* **83**, 3659–3663.
- Falkner, F. G. & Moss, B. (1988). *J. Virol.* **62**, 1849–1854.
- Fraenkel-Conrat, H. & Steinschneider, A. (1967). In *Methods in Enzymology* (Grossman, L. & Moldave, K., eds), vol. 12B, pp. 243–246, Academic Press Inc., New York.
- Franke, C. A., Rice, C. M., Strauss, J. H. & Hruby, D. E. (1985). *Mol. Cell. Biol.* **5**, 1918–1924.
- Golini, F. & Kates, J. R. (1985). *J. Virol.* **53**, 205–213.
- Hattori, M. & Sakaki, Y. (1986). *Anal. Biochem.* **152**, 232–238.
- Kates, J. R. & McAuslan, B. R. (1967). *Proc. Nat. Acad. Sci., U.S.A.* **58**, 134–141.
- Kozak, M. (1984). *Nucl. Acids Res.* **12**, 857–872.
- Lee-Chen, G.-J. & Niles, E. G. (1988). *Virology*, **163**, 80–92.
- Lee-Chen, G.-J., Bourgeois, N., Davidson, K., Condit, R. C. & Niles, E. G. (1988). *Virology*, **163**, 64–79.
- Mackett, M., Smith, G. L. & Moss, B. (1984). *J. Virol.* **49**, 857–864.
- Mackett, M., Yilma, T., Rose, J. K. & Moss, B. (1985). *Science*, **227**, 433–435.
- Mars, M. & Beaud, G. (1987). *J. Mol. Biol.* **198**, 619–631.
- Martin, S. A. & Moss, B. (1975). *J. Biol. Chem.* **250**, 9330–9335.
- Miller, J. H. (1972). In *Experiments in Molecular Genetics* (Miller, J. H., ed.), pp. 352–355, CSHL, Cold Spring Harbor, NY.
- Moss, B., Rosenblum, E. N. & Gershowitz, A. (1975). *J. Biol. Chem.* **250**, 4722–4729.
- Munyon, W., Paoletti, E. & Grace, J. T., Jr (1967). *Proc. Nat. Acad. Sci., U.S.A.* **58**, 2280–2287.
- Patel, D. D. & Pickup, D. J. (1987). *EMBO J.* **6**, 3787–3794.
- Plucienniczak, A., Schroeder, E., Zettlmeissl, G. & Streeck, R. E. (1985). *Nucl. Acids Res.* **13**, 985–998.
- Rohrmann, G. & Moss, B. (1985). *J. Virol.* **56**, 349–355.
- Rosel, J. L., Earl, P. L., Weir, J. P. & Moss, B. (1986). *J. Virol.* **60**, 436–449.
- Schmitt, J. F. C. & Stunnenberg, H. G. (1988). *J. Virol.* **62**, 1889–1897.
- Schwer, B. & Stunnenberg, H. G. (1988). *EMBO J.* **7**, 1183–1190.
- Schwer, B., Visca, P., Vos, J. C. & Stunnenberg, H. G. (1987). *Cell*, **50**, 163–169.
- Shuman, S. & Moss, B. (1988). *J. Biol. Chem.* **263**, 6220–6225.
- Shuman, S., Broyles, S. S. & Moss, B. (1987). *J. Biol. Chem.* **262**, 12372–12380.
- Slabaugh, M., Roseman, N., Davis, R. & Mathews, C. (1988). *J. Virol.* **62**, 519–527.
- Spencer, E., Shuman, S. & Hurwitz, J. (1980). *J. Biol. Chem.* **255**, 5388–5395.
- Tamin, A., Villarreal, E. C., Weinrich, S. L. & Hruby, D. E. (1988). *Virology*, **165**, 141–150.
- Tengelsen, L. A., Slabaugh, M. B., Bibler, J. K. & Hruby, D. E. (1988). *Virology*, **164**, 121–131.
- Vassef, A. (1987). *Nucl. Acids Res.* **15**, 1427–1443.
- Venkatesan, S. & Moss, B. (1981). *J. Virol.* **37**, 738–747.
- Venkatesan, S., Baroudy, B. M. & Moss, B. (1981). *Cell*, **25**, 805–813.
- Venkatesan, S., Gershowitz, A. & Moss, B. (1982). *J. Virol.* **44**, 637–646.
- Wei, C. M. & Moss, B. (1975). *Proc. Nat. Acad. Sci., U.S.A.* **72**, 318–322.
- Weir, J. P. & Moss, B. (1983). *J. Virol.* **46**, 530–537.
- Weir, J. P. & Moss, B. (1987). *Virology*, **158**, 206–210.
- Wittek, R., Cooper, J. A., Barbosa, E. & Moss, B. (1980). *Cell*, **21**, 487–493.
- Wright, C. F. & Moss, B. (1987). *Proc. Nat. Acad. Sci., U.S.A.* **84**, 8883–8887.
- Yuen, L. & Moss, B. (1987). *Proc. Nat. Acad. Sci., U.S.A.* **84**, 6417–6421.
- Yuen, L., Davison, A. J. & Moss, B. (1987). *Proc. Nat. Acad. Sci., U.S.A.* **84**, 6069–6073.