

Structure of Vaccinia Virus Late 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 2 June 1989, and in revised form 17 August 1989)

Functional elements of vaccinia virus late promoters were characterized by mutagenesis. Synthetic oligonucleotides 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 122 recombinants thus obtained was assayed for β -galactosidase expression. The relative amounts and 5' ends of *lacZ* mRNAs specified by a subset of the recombinants were determined by primer extension.

The analysis indicated that late promoters may be considered in terms of three regions; an upstream sequence of about 20 base-pairs, rich in T and A residues, separated by a spacer region of about six base-pairs from a highly conserved $(-1)\text{TAAAT}^{(+4)}$ element within which transcription initiates. All single nucleotide substitutions within the three A residues of the TAAAT, as well as the addition of a fourth A residue, caused drastic reductions in promoter strength. All substitutions of the T residues at -1 and $+4$ were also detrimental to promoter activity, to an extent that depended on the strength of the promoter as determined by the upstream sequence. mRNA synthesis appeared to initiate within the three A residues regardless of promoter strength. The 5'-poly(A) leader, which is a unique feature of poxvirus late mRNAs, was diminished in length when either of the T residues at -1 and $+4$ was mutated, was absent or limited to a few nucleotides when any of the three A residues was substituted, but was unaffected by changes outside the TAAAT sequence. The data are consistent with a model for the generation of the normal 5'-poly(A) leader by an RNA polymerase slippage mechanism requiring three consecutive A residues.

Single nucleotide substitutions within the six base-pairs upstream and three base-pairs downstream from the TAAAT sequence had modest effects on promoter strength. The most and least favourable changes led to a fourfold increase and an eightfold decrease in activity, respectively. Sequences further upstream were essential for late promoter function; tracts of T or A residues enhanced expression up to 20-fold, the former conferring much greater activity. Highest expression was obtained with a tract of 18 or 20 T residues.

Information gained from the mutational analysis allowed the construction of a synthetic late promoter that is 100-fold stronger than the late promoter used initially, and over 50% stronger than copies of the most efficient known vaccinia and cowpox virus natural late promoters. The synthetic promoter was incorporated into new recombination vectors designed for high-level expression of foreign genes in vaccinia virus.

1. Introduction

Transcription of the early and late classes of vaccinia virus genes occurs in the cytoplasm of the infected cell. Mature capped and polyadenylated

early mRNAs are synthesized by a complete transcription system present in the viral core (Kates & McAuslan, 1967; Munyon *et al.*, 1967; Wei & Moss, 1975). Early promoters are relatively small (Cochran *et al.*, 1985; Weir & Moss, 1987a), comprising a critical region of about 16 bp[§], in

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

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

[§] Abbreviations used: bp, base-pair(s); TK, thymidine kinase; m.o.i., multiplicity of infection; araC, cytosine arabinoside.

which the majority of mutations have a profound effect on transcription, separated by an 11 bp spacer region from a 7 bp region in which initiation can occur (Davison & Moss, 1989). The critical region appears to be the site at which a specific early transcription factor binds (Yuen *et al.*, 1987; Broyles *et al.*, 1988; Broyles & Moss, 1988).

Late mRNA synthesis occurs only after the onset of viral DNA replication, in a mode that differs radically from that of early transcription. Most late genes contain a conserved TAAAT element, or a closely related sequence, in the coding DNA strand close to the start of the open reading frame (see Fig. 1). The presence of a G residue immediately following the TAAAT sequence in the majority of late promoters results in overlap between it and the initiation codon for the late protein. Generally, late mRNAs begin with a capped adenylate residue (Boone & Moss, 1977), and some start within the A residues of the TAAAT sequence (Rosel *et al.*, 1986; Weir & Moss, 1987b). Bertholet *et al.* (1987) and Schwer *et al.* (1987), however, provided evidence for a long poly(A) tract at the 5' end. The former workers concluded that the 5'-poly(A) forms a covalent bridge between two non-contiguously coded mRNAs, whereas the latter workers argued that the 5' end of the poly(A) is capped. The latter structure was confirmed by direct analysis of the capped 5' ends of late mRNA (Ahn & Moss, 1989). Promoters that deviate from the consensus in Figure 1 have been described, and may represent other regulatory classes (Lee-Chen & Niles, 1988; Vos & Stunnenberg, 1988).

The mechanism by which late mRNAs are made is likely to be elucidated by fractionation of a recently developed *in vitro* late transcription system that produces 5'-polyadenylated RNAs (Wright & Moss, 1987; Schwer & Stunnenberg, 1988). In two possible mechanisms, *cis*- and *trans*-splicing, the TAAAT would function as part of a processing site rather than a promoter. The ability of linear DNA duplexes containing only a limited sequence upstream from the TAAAT sequence to serve as templates in the *in vitro* system has ruled out *cis*-splicing. Two other possible mechanisms, use of a poly(A) primer and transcriptional slippage within the TAAAT sequence, include the TAAAT sequence as part of a *bona fide* promoter. Schwer & Stunnenberg (1988) presented indirect evidence from the *in vitro* system that supports the latter mechanism. Moreover, both *trans*-splicing and poly(A) priming are unlikely, since the partially purified components of the late transcription system appear to lack endogenous poly(A) and significant independent poly(A) polymerase activity (Wright & Moss, unpublished results).

Bertholet *et al.* (1985) showed that the promoter for the 11-kD late gene (11-kD promoter; Fig. 1) is located within a contiguous sequence containing 100 bp upstream and 6 bp downstream from the TAAAT. Four studies were carried out on separate late promoters (whose sequences are included in Fig. 1), in order to define more precisely the

segment necessary for late gene regulation. Each investigation involved linking a late promoter to the coding region for chloramphenicol acetyltransferase, mutating upstream sequences, and assaying promoter activity in a transient expression system or in recombinant viruses. These studies showed that sequences upstream from the TAAAT element are essential for late promoter function. Cochran *et al.* (1985) defined an essential sequence between 27 and 12 bp upstream from the TAAAT sequence in the 7.5-kD late promoter, and reported that regions further upstream did not modulate the level of transcription. Bertholet *et al.* (1986) showed that an essential sequence is located within the region 16 bp upstream from the TAAAT sequence in the 11-kD promoter, and indicated that additional upstream sequences are required for maximal activity. From their analysis of the 28-kD promoter, Weir & Moss (1987b) concluded that an essential sequence is located between 17 and 7 bp upstream from the TAAAT element, and showed that inclusion of additional sequences up to 60 bp upstream from the TAAAT sequence resulted in maximal expression. Miner *et al.* (1988) analysed two late promoters by transient expression. They found that an essential sequence is located between 24 and one nucleotide upstream from the TAAAT sequence in the A1L promoter, and that sequences further upstream had only slight effects. Their results for the second promoter are difficult to interpret, since the TAAAT element was omitted from the parental expression plasmid. An absolute requirement for at least a part of the TAAAT sequence was reported by Hanggi *et al.* (1986). They showed that gross mutation of the entire element or mutation to TAAAG, albeit with an accompanying deletion close upstream, inactivated that 11-kD promoter.

It is important to note that these analyses were carried out before the discovery of the 5'-poly(A) leader. Levels of mRNA were assayed by S₁ nuclease analysis, which monitors apparent 5' ends mapping in the TAAAT element, and gives no clue regarding the presence or absence of a leader in mRNAs specified by mutated promoters. There has since been one study, involving analysis of a small number of mutations downstream from the TAAAT element in the 11-kD promoter, in which mRNAs were analysed by primer extension, allowing the leader to be detected (de Magistris & Stunnenberg, 1988). An extensive study of sequence elements involved in late gene regulation *in vivo*, however, has not been reported. This paper describes a detailed mutagenic analysis of late promoters, including the TAAAT element and sequences upstream and downstream. The implications of these results for the design of synthetic strong late promoters are discussed.

2. Materials and Methods

(a) Plasmid construction

Plasmids containing sequences for recombination into vaccinia virus were obtained by inserting synthetic oligo-

nucleotide duplexes into the vectors pMJ3, pMJ4, pMJ11 and pMJ35 (Davison & Moss, 1989). These vectors contain 4 (pMJ3, pMJ4) or 5 (pMJ11, pMJ35) unique restriction endonuclease sites immediately upstream from *lacZ*, flanked by vaccinia virus TK coding sequences. The sites are situated so that *lacZ* may be placed under control of synthetic vaccinia promoters. The sequence at the 5' end of *lacZ* in the upper strand is:

<i>Bam</i> HI	<i>Sal</i> I	<i>Hind</i> III	(<i>Kpn</i> I)	<i>Xho</i> I	Met-	Gly-	Asp-	Pro	...	β gal
TK-GGATCC	GTCGAC	AAGCTT	(GGTACC)	CTCGAG	C ATG	GGA	GAT	CCC	...	<i>lacZ</i> -TK

The *Kpn*I site is present only in pMJ11 and pMJ35. Two vectors (pMJ3 and pMJ11) contain the sites and *lacZ* in the same orientation as the TK gene, and 2 (pMJ4 and pMJ35) contain them in the opposite orientation.

Candidate promoters were synthesized using an Applied Biosystems 370B DNA synthesizer, and inserted as partially or fully matched duplexes, usually between the *Hind*III and *Xho*I sites. After ligation, remaining circular vector molecules lacking insert were linearized by cleaving with *Hind*III or *Xho*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 or pMJ35). Separate plasmid preparations were made from 1·5 ml cultures of transformants, and the sequence of each insert was determined as described by Hattori & Sakaki (1986). Required transformants were then colony-purified and the DNA sequence of the promoters verified.

(b) Generation of recombinant vaccinia viruses

Recombinants were isolated using TK⁻ selection and chromogenic detection as described by Davison & Moss (1989).

(c) β -Galactosidase assay

β -Galactosidase assays were carried out as described by Davison & Moss (1989), except that a m.o.i. of 10 was used, and infected monolayers were incubated for 48 h before carrying out the assay procedure. It was determined experimentally that β -galactosidase expression was independent of m.o.i. in the range 1 to 100. Each recombinant was assayed in triplicate on 2 separate occasions. Net β -galactosidase synthesis during the period of infection in the presence or absence of cytosine arabinoside (araC), an inhibitor of DNA replication, was calculated for each virus, and results from duplicate experiments were normalized. A zero value was assigned to a recombinant lacking a promoter upstream from *lacZ* (vMJ4, made using pMJ4). Final values (mean \pm standard deviation) were obtained as a fraction of the mean value obtained using vMJ23, which contains a parental 28-kD promoter, included in the same experiments. Standard deviations are not quoted in the results, but were approximately 10% of the mean values. Except where stated otherwise, negligible β -galactosidase was expressed in the presence of araC, as would be expected of late promoters (that is, less than 0·1% of the level expressed in the absence of araC, even for the strongest promoters).

(d) Analysis of early RNA made in vivo

Late RNA was prepared from CV-1 monolayers that had been infected at a m.o.i. of 10 and incubated for 8 h at 37°C. Total infected cell RNA was isolated by extraction with acid guanidinium thiocyanate/phenol/chloroform

(Chomczynski & Sacchi, 1987). For some experiments, RNA was chemically decapped as described by Fraenkel-Conrat & Steinschneider (1967). Approximately 50 µg of RNA were reverse transcribed as described by Wright & Moss (1987), using a molar excess of (5'-³²P)-labelled oligonucleotide primer. One primer complemented nucleotides +34 to +51 of the vaccinia-specified *lacZ* mRNA,

and the other complemented nucleotides +34 to +51 of the 11-kD mRNA specified by the resident vaccinia virus gene. The latter acted as a control for variable amounts of RNA present in the reactions. Primer-extended products were extracted with phenol, precipitated with ethanol, and treated with RNase A at 20 µg/ml for 1 h at 37°C. The products were precipitated with ethanol, and the dried pellets were denatured in formamide/dyes (0·1% (w/v) bromophenol blue, 0·1% (w/v) xylene cyanol FF, 10 mM-EDTA in 98% (v/v) deionized formamide) by heating at 100°C for 5 min, and electrophoresed on 6% (w/v) 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 28-kD promoter were used as markers to determine the precise sizes of *in vivo* mRNAs.

3. Results

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

The approach to analysing the effect of nucleotide substitutions on late promoter function was similar to that used to study early promoters (Davison & Moss, 1989). Essentially, candidate promoter sequences were prepared as duplexes with cohesive ends using a DNA synthesizer, and ligated to a linearized plasmid containing *lacZ* flanked by vaccinia virus TK gene sequences. After transformation of *Escherichia coli*, the DNA sequence of each promoter was verified and the plasmids were then used to generate recombinant vaccinia viruses. Ultimately, cell monolayers were infected with the recombinants and β -galactosidase expression was measured.

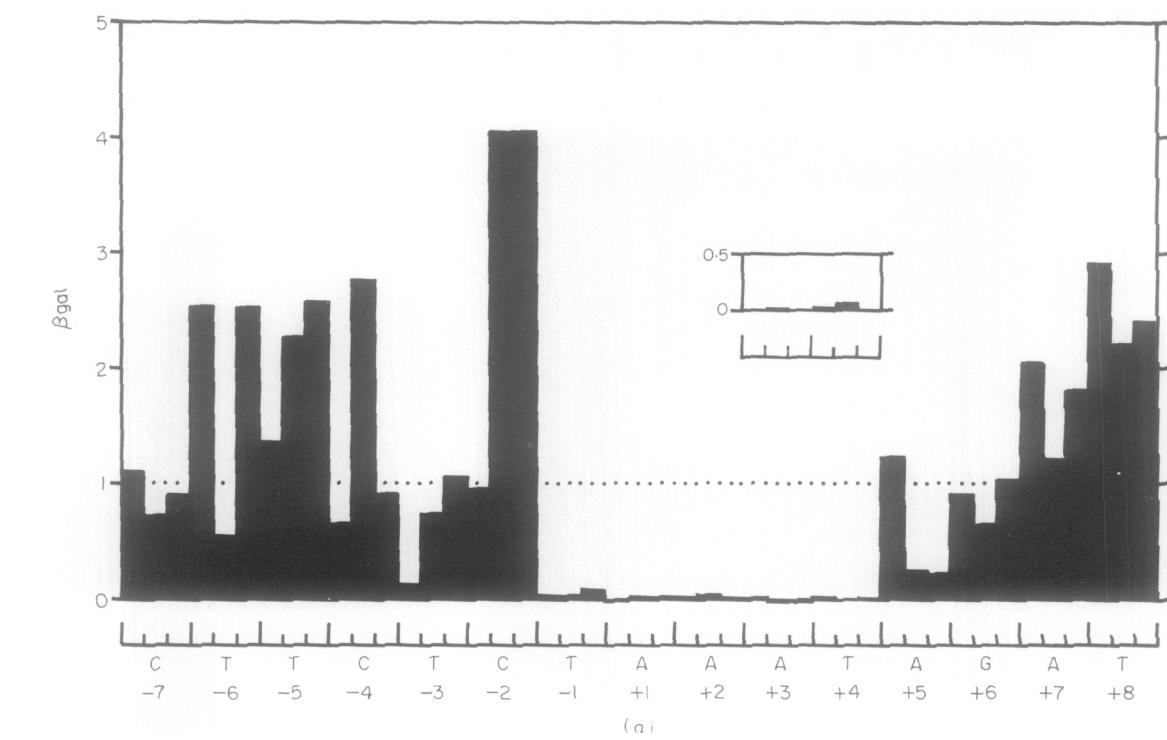
The sequence AAGCTCACAAAAAAACTTCTC-TAAATAGATTGAGC**ATG** was chosen as the initial target for detailed mutagenesis. The first five nucleotides were dictated by the use of a *Hind*III (AAGCTT) cloning site. The next 22 nucleotides, including the TAAAT (underlined), correspond to the natural 28-kD late promoter (Fig. 1). The succeeding nucleotides are not vaccinia-derived. The initiation codon for β -galactosidase is shown in bold type. Since the mRNA begins with a poly(A) sequence, we denote the first T of the TAAAT element as -1 and the adjacent A as +1. The parental promoter (in vMJ23) and all single nucleotide substitutions in the region CTTCTCTAAAT-AGAT, 45 in total, were assayed for activity in recombinant viruses. Levels of β -galactosidase

	-40	-30	-20	-10	-1	+4	+9
	:	:	:	:	:	:	:
7·5-kD	TCCAAACCCACCGCTTTATAGTAAGTTTCACCCA				TAAAT	AATAA	
M1L	GACATAAAAGAACTAAGTTATGATATTGTTATAGAGAG				TAAAT	TGTTG	
37-kD	ATCGTTGATAGAACAGGAATGTATAAGTTTATGTTAAC				TAAAT	GTGGC	
11-kD	AAAATATAGTAGAATTCACTTTGTTTTCTATGCTA				TAAAT	GAATT	
I1L	CTATTGATATATTGATTTAAAAGTTGTTGGTGAAC				TAAAT	GGCGG	
I2L	GATGAAGATAGCGATAAAGAAAAGCCAATATTCAATGTA				TAAAT	GGATA	
I5L	AGGACTTTGTCACATATTCTTGATCTAATTTAGATA				TAAAT	GGTGG	
I7L	TTGGAAAAAAGAAGATATCTGGTAAATTCTTTCCATGA				TAAAT	GGAAA	
28-kD	ATCGGTACGGGTATTCACTTACACAAAAAAACTTCTC				TAAAT	GAGTC	
H1L	GCGTTACCGATAAAGTAGTTTATCCATTGACGTTA				TAAAT	GGATA	
H3L	ACTCGTATTAAGAGTTGTATATGATTAATTCATAAAC				TAAAT	GGCGG	
H5L	TTATACATCATAAACCAATTCTCTAGTTGTTGTAAC				TAAAT	GGACT	
H6R	TCTTGAGGGTTGTGTTAAATTGAAAGCGAGAAATAATCA				TAAAT	TATT	
H8R	GGATGATATAGATCTTACACAAATAATTACAAAACGA				TAAAT	GGAAA	
D11L	TGCTGTGATTTTAAACATAGTTATTACTTATCACTCA				TAAAT	GAGTA	
D12L	GATAATAACTAATAATAATGAAAACAAACTATAGAGTTG				TAAAT	GGATG	
D13L	ATATTCTCTACGGAGTTATTGTAAGCTTTCCATT				TAAAT	AGAAA	
A1L	TACTTGCTCATTAGAAGTATAAAAAAATAGTTCCGTAAT				TAAAT	GGCTA	
A2L	GGATTTCTGGCAACGTCTAGAAATAAAATGTTTTATA				TAAAA	TATTG	
A3L	CGCTAGACATATTACAGAACTATTTAGATTATGATATT				TAAAT	GAGTT	
P4b	GGATATTAAAATCACGCTTCGAGTAAAACATACGAATA				TAAAT	AATGG	
Consensus	GGATTAATAAAAAATTTAAAAAAATTTCAAAATA				TAAAT	GGATA	
	ATC A C T TA T T T A T						G
	TA A						

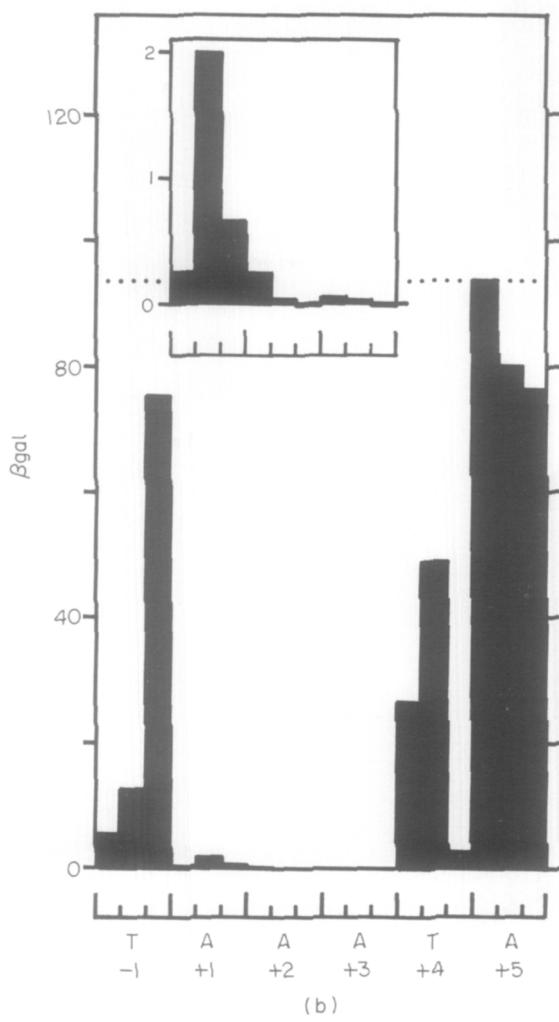
Figure 1. Alignment of vaccinia virus late promoter sequences. Published initiation sites map within, or close to, the TAAAT element. References are as follows: 7·5-kD, Venkatesan *et al.* (1981); M1L, Tamin *et al.* (1988); 37-kD, Hirt *et al.* (1986); 11-kD, Bertholet *et al.* (1985); I1L to I7L, Schmitt & Stunnenberg (1988); 28-kD, Weir & Moss (1984); H1L to H8L, Rosel *et al.* (1986); D11L to A3L, Weinrich & Hruby (1986); p4b, Rosel & Moss (1985). The most common nucleotide(s) at each position is indicated at the foot. Where the next most common nucleotide differed in frequency from the most common by only 1, that nucleotide is listed also.

expressed are shown in Figure 2(a). All recombinants would in principle specify identical β -galactosidase species, with the exception of that containing TAAATG, which would specify a polypeptide with an additional four amino acid residues at the amino terminus. Results for an additional six recombinants, based on a parental sequence in which the A residue immediately downstream from the TAAAT sequence was G, as in the natural 28-kD promoter, are shown in the inset in Figure 2(a). As expected from its high level of conservation in natural late

promoters, mutations in the TAAAT element had a profound effect on expression. All sequence variations except CAAAT (followed by A) and TAAAA (followed by G) were inactive, and even those were reduced to less than 10% of the activity of the parental sequence. Mutations in the six nucleotides upstream (-2 to -7) and four nucleotides downstream (+5 to +9) from the TAAAT element influenced the level of expression, but none abolished it. The greatest changes were an eightfold decrease in activity caused by a G substitution at



(a)



(b)

Figure 2. A histogram showing levels of β -galactosidase expressed by cells infected with recombinants containing single nucleotide substitutions in the TAAAT element and neighbouring nucleotides of (a) the 28-kD late promoter in vMJ23 (modified parental sequence AAGCTCACAAAAA-AAACTTCTCTAAATAGATTCGAGCATG) and (b) the strong late promoter in vMJ480 (parental sequence AAGCT₂₀GGCATATAAAATAGACTCGAGCATG). Mean β -galactosidase levels are shown as filled columns relative to the parental 28-kD promoter in (a). The horizontal dotted lines show expression levels for the parental sequences used in the 2 analyses. As each parental nucleotide 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 the T residue at -1 is G-A-C. The inset in (a) shows results for mutants based on a parental sequence in which A at +5 was replaced by G. The natural 28-kD promoter has G at this position. In the inset in (b), the β -galactosidase results for substitutions of the A residues at +1 to +3 have been replotted on an expanded scale, equivalent to that used in (a).

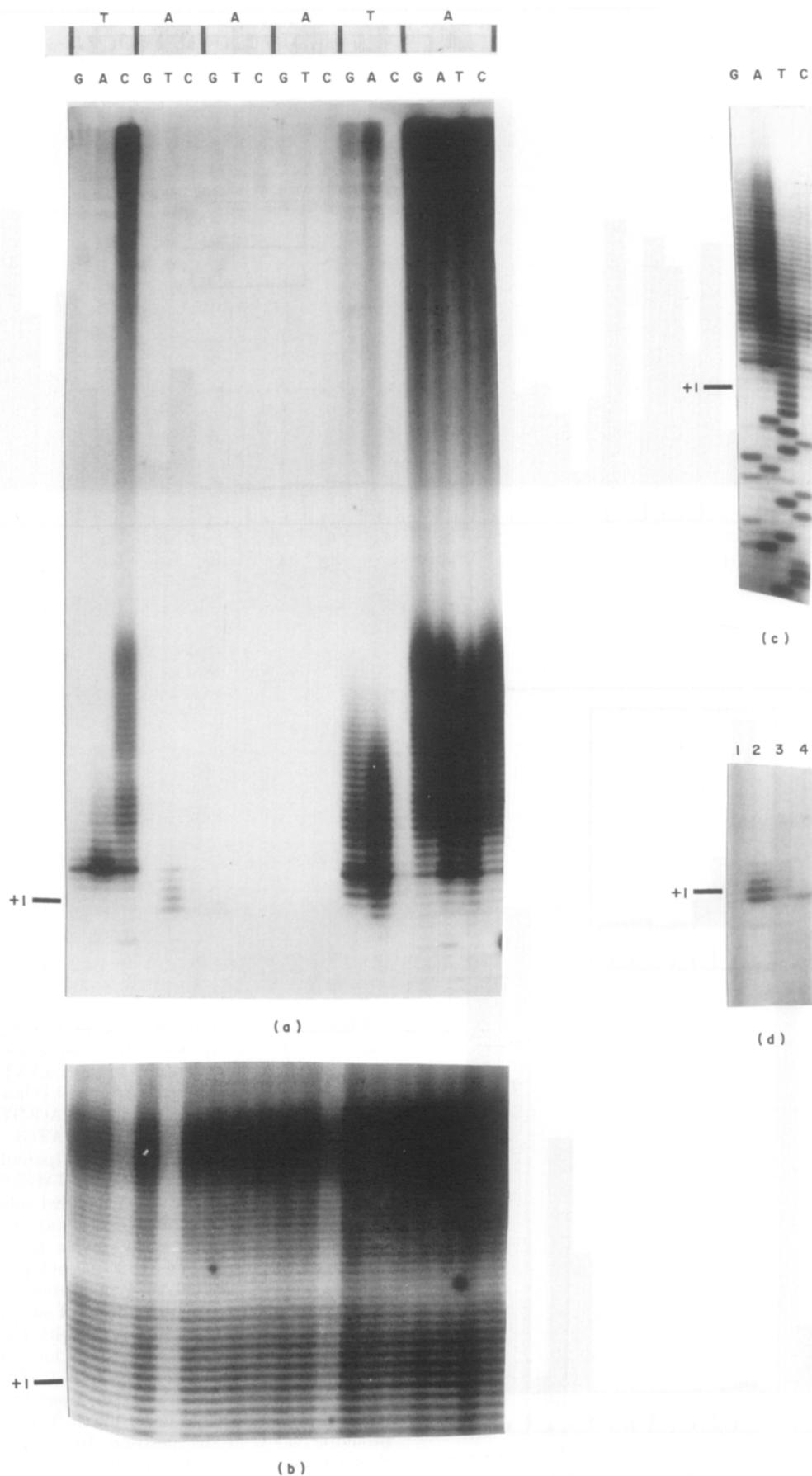


Fig. 3.

-3 and a fourfold increase in activity due to an A or T substitution at -2.

Attempts using primer extension to examine mRNA expressed by the mutants were unsuccessful, since even the parental promoter is rather weak. This left open the possibility that differences in β -galactosidase expression might be due to aberrant RNA processing rather than to differences in transcriptional activity of the promoters. Therefore, a second set of single substitution mutants was made in a much stronger late promoter AAGCT₂₀ GGCATATAAATAGACTCGAGCATG, whose derivation is described in a later section. The parental promoter in vMJ480 is almost 100-fold stronger than the 28-kD promoter in vMJ23. β -Galactosidase results for all single substitutions in the sequence TAAATA are shown in Figure 2(b). As before, all mutants, except TAAATG, would in principle specify identical β -galactosidase polypeptides. In addition to the standard procedure of verifying promoter sequences during the original plasmid construction, several of the mutations were confirmed by cloning from recombinant virus DNA into bacteriophage M13. Most mutations in the TAAAT element of the strong promoter severely reduced β -galactosidase expression, as was observed for the 28-kD promoter. Nevertheless, there were significant differences between the behaviour of the two promoters. All substitutions of the two T residues (-1 and +4) and the first A residue (+1) in the TAAAT element were active, and replacement of the A at +2 by a G residue produced a low level of activity, as shown in the inset in Figure 2(b). Only TATAT, TACAT, TAATT, TAACT and probably TAAGT were inactive. Replacement of T at -1 with C resulted in only a 19% reduction in activity, whereas other substitutions were considerably more detrimental. In contrast, replacement of T at +4 with A, G or C reduced expression by 48%, 72% and 96%, respectively.

(b) Effect of single nucleotide substitutions on RNA synthesis

Primer extension was used to analyse β -galactosidase mRNA synthesized in infected cells by lacZ under control of the strong promoter and its

mutated derivatives. The data shown in Figure 3(a) indicate that recombinants that generated no β -galactosidase failed to produce detectable levels of lacZ mRNA. lacZ mRNA levels (Fig. 3(a)) could not be quantified precisely by comparison with 11-kD mRNA levels (Fig. 3(b)), owing to the non-linearity of autoradiographic response and the representation of transcripts as a ladder of bands, each of a different intensity. Nevertheless, a reasonable correlation was noted between relative amounts of lacZ mRNA and β -galactosidase levels.

Reverse transcriptase can induce the addition of one extra non-complementary nucleotide at the 3' ends of a proportion of synthesized DNA chains, and thus gives the impression that initiation can occur upstream from the actual initiation site (Davison & Moss, 1989). This artefact is dependent upon the presence of a cap structure. Therefore, certain RNA samples were chemically decapped before reverse transcription. Mature and decapped mRNA specified by the parental promoter were indistinguishable in the pattern of their leaders. The ladder-like series of bands obtained with the parental promoter extends more than 40 nucleotides past the TAAAT element, and is indicative of the heterogeneous 5'-poly(A) leader. Sequencing of mRNA produced by the parental promoter that had been decapped chemically showed that at least five A residues are present upstream from the U residue corresponding to T at +4 (Fig. 3(c)). The proximal three of these could be coded conventionally by the DNA template. The nature of the remainder of the leader was ambiguous from the sequence ladder in Figure 3(c), probably as a result of size heterogeneity (Ahn & Moss, 1989). Mutations at +5 (Fig. 3(a)) and other sites outside the TAAAT element (data not shown) did not alter the size of the leader; mRNAs contained leaders of heterogeneous length, ranging from six to over 40 nucleotides.

In contrast to changes outside the TAAAT sequence, most mutations within it resulted in apparent shortening of the leader. This was not an artefact dependent upon the concentrations of RNA in the primer extension reactions (data not shown). Substitution of T at -1 by a purine resulted in a marked decrease in the length of the 5'-poly(A) leader, but the minimum size of six nucleotides was

Figure 3. An autoradiograph showing primer extended products from mRNA expressed by cells infected with recombinant viruses containing single nucleotide substitutions in the TAAATA element of the strong late promoter controlling lacZ (parental sequence AAGCT₂₀GGCATATAAATAGACTCGAGCATG). (a) Results for lacZ mRNA. Each parental nucleotide is shown above the 3 possible substitutions at that position. The results correspond in order to the β -galactosidase results shown for the same mutants in Fig. 2(b), except that data for the unsubstituted parental sequence are included in the last set. The expected mobility of the primer extended product from mRNA initiated at the A residue at +1 in the parental promoter is indicated on the left. (b) Results for 11-kD mRNA. These act as a control for the amount of RNA added to the reactions in (a). The details are as in (a). (c) Direct sequence of the cDNA obtained by primer extension of mRNA expressed by the parental promoter. The RNA sample was chemically decapped before sequencing. The band corresponding to the A residue at +1 is indicated on the left. (d) Results for the TGAATA (lane 1), TTAATA (lane 2), TCAATA (lane 3) and TAGATA mutants (lane 4). Primer extended products were made using chemically decapped mRNA. The expected mobility of the primer extended product originating from an mRNA initiating at the A residue at +1 in the parental promoter is indicated on the left.

(a) Complex mutations in the 28-kD promoter

AAGCTCACAAAAAAACTCTCATAAATAGATTCGAGCATG

23			1.00
6		G	1.26
5+		G	2.83
9	-		0.80
92	-		0.63
16	-	T	0.02
448	<A>		0.06
439	GCCA	T	3.36
440	GGCA	T	2.52
452	GCCA	A	5.93
444	GGGGGGGG		0.00
445	CCCCCCCC		-0.03
443	TTTTTTTT		19.51
12	A A	AAAAAA	0.43
10+	A A	AAAAAA	0.84
19	TTTTTTTTTTTT	T T	6.03
18+	TTTTTTTTTTTT	T T	5.80
20	TTTTTTTTGT	T T	0.70

(b) Reiterated TAAAT elements

X = AGCTAAATAAATAAATAAATAAATAAAT
 X' = AGCTATTATTTATTATTTATTATTT

AAGCTTCTCGAGCATG

69	<X>	0.84
70	<XX>	1.83
71	<XXX>	2.92
72	<X'>	0.02
73	<X'X'>	0.02
74	<X'X'X'>	0.01
75	<X'X>	2.30
77	<X'X'XX>	3.68

(c) Replacement of upstream T tract

AAGCTTTTTTTTTAAATGGACTCGAGCATG

63		4.40
87	GGG	0.47
88	CCCGGG	1.14
438	GCCCCGGG	0.33
451	GGCCCGGG	0.14
89	GGGCCCGGG	0.02
457	AAGCTTTTTTCCCCCCCCGG <u>AAATGGACTCGAGCATG</u>	-0.01

(d) Strong late promoters based on an upstream T tract

54B	AAGCTTTTTTTTTTTTT <u>AAATGGACTCGAGCATG</u>	27.16
377	AAGCTTTTAAGCTTTTT <u>AAATGGACTCGAGCATG</u>	21.81
441	AAGCTTTTTTT <u>AAATGGACTCGAGCATG</u>	105.51
442+	AAGCTTTTTTT <u>AAATGGACTCGAGCATG</u>	109.35
480	AAGCTTTTTTT <u>AAATGGACTCGAGCATG</u>	93.50
479	AAGCTTTTTTT <u>AAATGGACTCGAGCATG</u>	96.27

(e) Strong natural late promoters

455	GGATCCTAAAAATATAGTAGAATT <u>CATTTGTTTTCTATGCTATAAAT</u>	56.89
454+	GGATCCTAAAAATATAGTAGAATT <u>CATTTGTTTTCTATGCTATAAAT</u>	73.02
453	GGATCCTAAAAATATAGTAGAATT <u>CATTTGTTTTCTATGCTATAAAT</u>	41.16
496	AAGCTTATAATTACACGATTGTAAGTT <u>GAATAAAATTTTTATAATAAAT</u>	60.21
	AGACTCGAGCATG	

Fig. 4.

unchanged (Fig. 3(a)). Weakly active promoters with substitutions of A residues at +1 to +3 specified particularly aberrant mRNAs. Mutations of A at +1 resulted in a ladder of five bands corresponding in size to transcripts initiating only two to six nucleotides upstream from the T residue at +4 (Fig. 3(a) and (d)). It is possible, however, that this pattern may indicate that the leader consists of between two and six A residues, two of which could be coded conventionally by the DNA template. The pattern was not altered by decapping the RNA (Fig. 3(d)). The one active mutation of the A residue at +2 (to give TAGAT) resulted in two bands corresponding in size to transcripts initiated at the G and the upstream A residues. Primer extension of decapped RNA showed only one band corresponding to initiation at the G residue (Fig. 3(d)). Again, there is no direct evidence that the mRNA commenced with G; the mRNA might have a leader consisting of two A residues, one of which could be coded directly by the DNA template. The slightly greater mobility of the primer extension products (Fig. 3(a) and (d)) compared with the corresponding bands in the normal ladder is, however, consistent with initiation at the G residue and the absence of a 5'-poly(A) leader. Sequencing of RNAs specified by promoters with substitutions of the A residues at +1 to +3 was prevented by their low abundance. Substitutions of T at +4 also caused a decrease in the length of the 5'-poly(A) leader (Fig. 3(a)).

In addition to the primer extension products discussed above, Figure 3(a) shows that reverse transcripts of high molecular weight were synthesized from mRNAs with normal leaders. They were synthesized in significantly lower molar amounts from mRNAs with shorter leaders, and were not detected using mRNAs with particularly short leaders.

(c) *Effect of more complex mutations on promoter strength*

Several other more complex mutations in the 28-kD promoter were assayed for β -galactosidase activity. Results for these are listed in Figure 4(a). mRNAs produced by certain recombinants were assayed by primer extension, and showed normal 5'-poly(A) leaders (data not shown). Single nucleotide deletions upstream from the TAAAT element (vMJ9, vMJ92) reduced expression, but not severely. Conversion of the sequence TAAATA to CAAATT, by deletion of one nucleotide and substitution of another, abolished expression (vMJ16).

Expansion of TAAAT to TAAAAT reduced activity severely (vMJ448). Stronger derivatives of the 28-kD promoter (vMJ439, vMJ440 and vMJ452) were made by substituting the six nucleotides upstream and two other nucleotides downstream from the TAAAT sequence with optimal nucleotides deduced from single substitutions in the 28-kD promoter (Fig. 2(a)). The maximal effect achieved was almost a sixfold increase in expression.

Replacement of the tract of eight A residues by C or G residues abolished expression (vMJ444 and vMJ445), but substitution by eight T residues caused a 20-fold increase in promoter strength (vMJ443). Similarly, substitution of appropriate nucleotides in the 28-kD promoter to give a tract of 17 A residues upstream from the TAAAT element resulted in a weaker promoter (vMJ12), whereas substitutions to give 18 consecutive T residues strengthened the promoter sixfold (vMJ19). Replacement of the T residue at -9 by G weakened this promoter significantly (vMJ20). It is pertinent to note that Weir & Moss (1987b) showed that replacement of the fourth residue in the A tract (-11) by G was severely detrimental to the 28-kD promoter. These results indicate that tracts of A or, preferably, T residues (or, possibly, a combination) may form part of late gene regulatory sequences, in combination with the downstream TAAAT sequence.

Reiterated overlapping copies of TAAAT were inserted upstream from lacZ, as shown in Figure 4(b). The elements could then serve as initiation sites and as activating upstream sequences. These promoters stimulated lacZ mRNA production in only one orientation, as expected, but were weak (compare vMJ69 with vMJ72, vMJ70 with vMJ73, and vMJ71 with vMJ74). mRNA was not expressed at sufficiently high levels to allow the initiation sites to be located. The weakness of these promoters might be due to the preponderance of A residues, which, when located in the upstream sequences, do not potentiate transcription as markedly as T residues. The presence of inverted repeats in the promoter did not inhibit activity (vMJ75 and vMJ77).

(d) *Effect of the length and location of the upstream T tract on promoter strength*

In order to determine the effect of length of an upstream T tract on promoter activity, a series of recombinants containing promoters based on the sequence AAGCTTAAATGGACTCGAGCATG was

Figure 4. The levels of β -galactosidase expressed by cells infected with recombinants containing mutated late promoters. The identities of recombinants (vMJ) are shown on the left, and β -galactosidase results are given on the right relative to the parental 28-kD promoter in vMJ23 (1·00). Sequences are shown from the HindIII site, except for 3 members of group (e), which are shown from the BamHI site. 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 TAAAT element underlined and the initiation codon for β -galactosidase in bold type. Substituted or deleted (-) nucleotides are shown for each recombinant. Inserted nucleotides are indicated (<>), with the 1st angle bracket below the nucleotide after which the insertion is present.

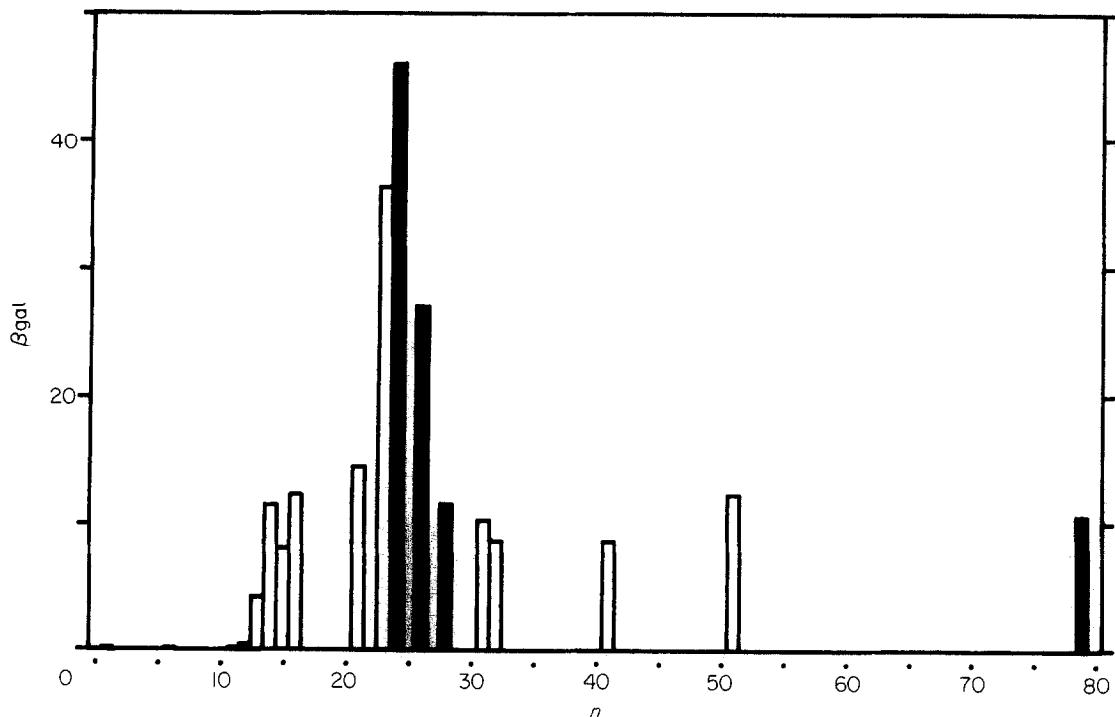


Figure 5. A histogram showing levels of β -galactosidase expressed by cells infected with recombinant viruses containing late promoters with upstream tracts of T residues. The series of promoters was based on the sequence AAGCTTAAATGGACTCGAGCATG, with the value of n ranging from 1 to 79. Results for promoters that were sequenced only in the plasmid used to obtain the recombinant virus are shown as open columns, and results for promoters that were also cloned from the recombinant virus into bacteriophage M13 and resequenced are shown as filled columns ($n=24, 26, 28, 79$). The level of β -galactosidase indicated on the left is relative to that obtained with vMJ23 (1.00), which contains the parental 28-kD promoter.

constructed. The β -galactosidase results are shown in Figure 5. Promoters in which $n=1$ to 11 were inactive, and promoter strength increased as n increased from 12 to 24. A reduction in activity was noted as n increased further to 28, and at greater values no significant change in promoter strength was detected. *lacZ* mRNA produced by selected recombinants corresponded well in amounts with β -galactosidase levels, and active promoters produced RNA with normal leaders (data not shown). Certain of the promoters with longer T tracts were cloned from recombinant virus DNA into bacteriophage M13 and sequenced in order to investigate whether length variation had occurred. Recombinants checked thus are indicated by filled columns in Figure 5. Recombinant $n=79$ (checked using 2 independent M13 clones) was constructed from a plasmid in which $n=60$. At smaller values of n , variation was less. Of the two M13 clones used to check recombinant $n=28$, which was constructed from a plasmid in which $n=28$, one had $n=27$ and the other had $n=28$. Recombinant $n=24$ was constructed from a plasmid in which $n=26$, and one of the two M13 clones had $n=25$ and the other had $n=24$. A separate $n=24$ recombinant, generated using a plasmid in which $n=24$, produced the same level of β -galactosidase. Taking these findings into account, a T tract extending at least 12 nucleotides upstream from the TAAAT element is sufficient for

late promoter activity. Further extension to 24 or 25 nucleotides upstream resulted in maximum strength (nearly 50 times that of the parental 28-kD promoter), and additional T residues were counteractive, giving promoters no stronger than one containing only 14 T residues.

Having determined that the T tract must extend at least 12 nucleotides upstream from the TAAAT sequence, a set of recombinants was made in which nucleotides in the T tract nearest to the TAAAT element were replaced by G or C residues, in order to ascertain how close to the TAAAT sequence the T tract must extend for activity, and to go some way towards defining the minimal sequence requirements for a late promoter. A parental sequence containing 13 T residues was chosen, since, although one containing 12 residues was active, the additional T residue conferred a substantial increase in activity. The β -galactosidase results for these recombinants are shown in Figure 4(c). Replacement of the three T residues adjacent to the TAAAT sequence by G residues (vMJ87) reduced promoter activity tenfold, but substitution of a further three T residues by C residues partially alleviated the reduction. Reference to the effects of single substitutions within the six nucleotides upstream from the TAAAT element in the 28-kD promoter (Fig. 2(a)) indicates that these effects are likely to have been caused by the substitutions rather than by shorten-

ing of the T tract. Further replacement of T by G residues (vMJ438, vMJ451 and vMJ89) caused a progressive weakening of the promoter, and activity was abolished when nine residues had been substituted, leaving only four upstream T residues. An additional promoter containing seven T residues separated from the TAAAT element by 11 C or G residues (vMJ457) was inactive. These results indicate that late promoter regulation may be conferred by a T tract of minimum length situated an appropriate distance from the TAAAT sequence; in this case, the minimum length is five nucleotides located not more than ten and not less than seven nucleotides from the TAAAT element.

Although a short T tract located appropriately with respect to the TAAAT sequence confers late gene regulation, longer tracts add substantially to promoter strength. The degree to which each T residue in the tract contributes towards promoter strength, however, is unknown. Clearly, substitutions of some other nucleotides for T residues had severe effects upon activity (e.g. compare vMJ19 and vMJ20 in Fig. 4(a)). The importance of T residues in at least some of the positions 11 to 13 nucleotides and 21 to 24 or 25 nucleotides upstream from TAAAT is underlined by the data in Figure 5. In contrast, promoters containing 16 or 21 T residues upstream from TAAAT differed little in strength (Fig. 5), suggesting that T residues at this location are unimportant in modulating activity. This was supported by the finding that replacement of the T residues located 20 to 17 nucleotides from the TAAAT sequence by AAGC caused only a slight reduction in activity (compare vMJ54B with vMJ377 in Fig. 4(d)).

(e) Effect of orientation within the vaccinia virus genome on expression

In the majority of recombinant viruses, the promoter-lacZ element was inserted in the opposite orientation from the TK gene, but in some it was inserted in the same orientation. The latter are denoted by a plus sign in Figure 4(a), (d) and (e). The activities of stronger promoters showed little dependence on orientation (compare vMJ19 with vMJ18+ in Fig. 4(a), vMJ441 with vMJ442+ and vMJ455 with vMJ454+ in Fig. 4(d)). Weaker promoters, however, were more active when in the same orientation as the TK gene (compare vMJ6 with vMJ5+ and vMJ12 with vMJ10+ in Fig. 4(a)). The reasons for this are unknown. A similar phenomenon was found for early promoters, but was not restricted to weaker ones (Davison & Moss, 1989).

(f) Construction of strong late promoters

The strongest promoters obtained in this study were based on a tract of 20 T residues separated from the TAAAT element by six optimal nucleotides deduced from the analysis of the 28-kD promoter (Fig. 2(a)). These include vMJ441 and

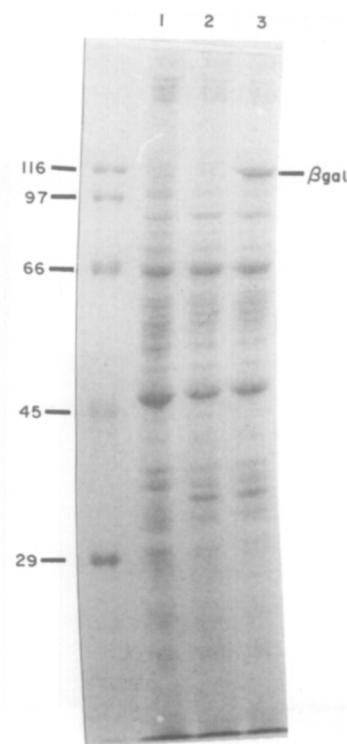


Figure 6. The level of β -galactosidase expressed in cells infected with vMJ480, which contains lacZ under control of the strong late promoter (the parental promoter in Fig. 2(b)). Cell monolayers were infected at an m.o.i. of 10 and incubated for 48 h at 37°C. Extracts representing approximately 10^5 cells were electrophoresed on a denaturing 10% polyacrylamide gel and stained with Coomassie brilliant blue dye. (1) mock-infected cells; (2) cells infected with vaccinia virus strain WR; (3) cells infected with vMJ480. The molecular weights (kDa) of the marker proteins on the left are indicated.

vMJ479 in Figure 4(d), and vMJ480, the parental promoter in Figure 2(b). It was heartening to find that when cells were infected with vMJ480 (containing the parental strong promoter in Fig. 2(b)), β -galactosidase was the major vaccinia-encoded protein, and formed a major component of the total infected cell protein (Fig. 6). Subsequently, reduction of the 20 T residues to 18 was found not to effect promoter strength (data not shown).

We were interested in comparing the strength of the synthetic promoters developed here with the strongest known natural poxvirus promoters. Accordingly, sequences corresponding to -42 to +4 of the promoter for the major 11-kD structural protein gene of vaccinia virus (Bertholet *et al.*, 1985) and to -50 to +4 of the promoter for the A-type inclusion (ATI) protein gene of cowpox virus (Patel & Pickup, 1987) were assayed in recombinant viruses vMJ455 and vMJ496, respectively. The β -galactosidase expression data listed in Figure 4(e) and depicted in Figure 7 show that the cowpox promoter was slightly more active than the vaccinia virus promoter. The synthetic promoter in vMJ480,

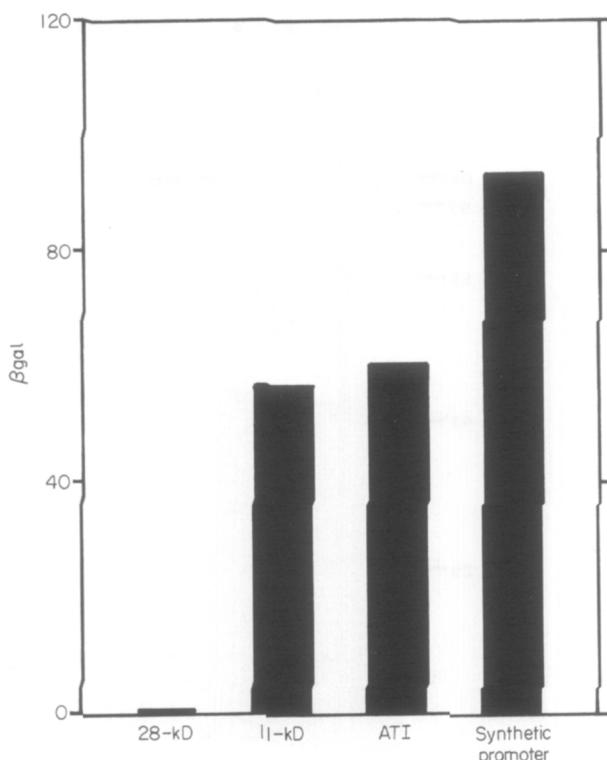


Figure 7. A comparison of the activities of natural and synthetic poxvirus late promoters: the vaccinia virus 28-kD and 11-kD promoters (in vMJ23 and vMJ455, respectively), the cowpox virus A-type inclusion (ATI) protein gene promoter (in vMJ496) and the synthetic promoter (in vMJ480). The sequences of these promoters are shown in Fig. 4. β -Galactosidase activities are expressed relative to vMJ23 (1.00).

however, was considerably more active than either poxvirus promoter.

One additional construction, containing the 11-kD promoter located close upstream from a strong early promoter designed by Davison & Moss (1989), is shown in Figure 4(e) (vMJ453). Early transcription initiated at the underlined G and A (data not shown). When late protein synthesis was prevented by araC, marginally more β -galactosidase was expressed by the early promoter in vMJ453 than by a recombinant containing the early promoter alone (data not shown). Total β -galactosidase expressed by both the early and late promoters was, however, about 72% of that expressed by a recombinant containing the 11-kD promoter alone (vMJ455). This decrease could be a translational effect caused by a longer untranslated leader, rather than an effect on late promoter strength. Nevertheless, these results indicate the feasibility of designing vectors containing adjacent strong early and late promoters, which may be important for a maximal immune response.

4. Discussion

This study has demonstrated the importance of the conserved TAAAT motif of late promoters and

has revealed other elements upstream and downstream from this pentanucleotide that affect expression. Information gleaned from the analysis was used to design strong late promoters. In addition, characterization of the 5' ends of mRNAs produced by mutated promoters provided insights into the mechanism of formation of the 5'-poly(A) leader.

Candidate promoters were synthesized chemically and placed into recombinant vaccinia viruses in order to assess their effect upon expression *in vivo*. We relied primarily on β -galactosidase synthesis for quantification because of the sensitivity and reproducibility of the assay. Nevertheless, there was a good correlation between β -galactosidase expression and the steady-state level of mRNA detected by primer extension. All nucleotide substitutions of the A residues (+1 to +3) in the TAAAT element led to a drastic decrease in promoter activity, whereas certain substitutions of T residues (-1 and +4) had a lesser effect, particularly when placed into an otherwise strong promoter. For example, substitution of C for T at -1 produced greater than a 96% decrease in activity of the weak 28-kD promoter, but only a 19% inhibition of the strong promoter. In the same position, substitution by A or G led to 87% or 94% decreases in the strong promoter, respectively, and complete inhibition of the weak promoter. Similarly, little or no activity was detected with substitutions of T at +4 in the weak promoter, whereas the strong promoter was inhibited by 48%, 72% and 96% with A, G and C changes, respectively, at this position. The difference in the magnitude of the effects of substitutions in the strong and weak promoters was not due to detection limits, but suggests compensating effects in the former. The data indicate that the least detrimental mutations of the TAAAT sequence are CAAAT, TAAAG and TAAAA. A natural example of a functional late promoter with the last of these mutations occurs in the A2L gene (Fig. 1).

The effects of nucleotides surrounding the TAAAT element in the 28-kD promoter were examined. Substitution of G at +5 in the natural 28-kD promoter (Fig. 1) by A decreased activity by only 21%, whereas substitution by T or C led to a reduction of about 80%. In the strong promoter, either purine was approximately equivalent at this position, and a pyrimidine decreased activity by only 15 to 20%. It is notable that 15 of the 21 late promoters listed in Figure 1 have G at +5, three have A, and three have T. Purines also predominate at +6, but positions further downstream have more variable compositions. Not surprisingly, substitutions of +6 to +8 in the 28-kD promoter had minor effects on expression. Some of these changes might have exerted their effect at a translational rather than a transcriptional level. Upstream from the TAAAT motif, A and T predominate in natural promoters at -2, and substitution of either of these nucleotides for the C residue in the 28-kD promoter resulted in a fourfold increase in activity. There is considerable sequence variation between natural

promoters in the region from -3 to -7. Correspondingly, except for a strongly detrimental effect of G at -3, single substitutions in the 28-kD promoter had modest effects on expression. By choosing all optimal nucleotides surrounding the TAAATA sequence, however, we were able to enhance the activity of the 28-kD promoter by a factor of 6.

The region upstream from -7 in natural promoters is essential for late promoter function, and is rich in A and T residues. The presence of tracts of A or T, but not G or C, residues served to activate late promoter function. T tracts had a much greater activating effect than A tracts, indicating that the requirement is not merely for a region with a low energy of denaturation. It is not known whether a particular arrangement of T and A residues would further increase activity. A promoter containing five T residues located between -10 and -14 was active, whereas a tract of seven T residues from -19 to -13 was inactive. Interestingly, T is present at -10 or -11 in 15 of the 21 promoters listed in Figure 1, and at both positions in 13. Tracts of T residues extending to -25 further increased promoter strength, but longer tracts led to a significant loss of activity. T-rich sequences are present upstream from a number of eukaryotic genes, and it has been suggested that they activate RNA polymerase II transcription (Lue *et al.*, 1989).

Primer extension studies indicate that the initiation sites of late mRNAs are dictated by the TAAAT element or its equivalent. Variation of the length of the 5'-poly(A) leader occurred only when modifications were introduced into the TAAAT sequence. Substitutions of T at -4 in the strong promoter by purines were associated with a reduction in the average length of the leader. Substitution of A at +1 by T weakened the promoter considerably, and the 5' ends mapped only a few nucleotides upstream from the TAAAT element. Moreover, when A at +2 was changed to G, initiation appeared to occur only at the G residue. These data are consistent with the formation of the leader by RNA polymerase slippage, and suggest that three consecutive A residues are required at the initiation site for this to occur efficiently. The heterogeneity in length of the leader indicates that after each addition of an A residue by RNA polymerase, there is a probability of proceeding forwards. This probability may be affected by the next nucleotide, since the leader was shorter when other nucleotides were substituted for the T residue at +4. de Magistris & Stunnenberg (1988) also reported an effect of sequence on the length of the leader, and suggested a slippage mechanism. In that report, however, mutations were made downstream from the TAAAT element, at +5, and only a small effect was noted. Using high-resolution sequencing gels, we could detect no effect of mutations at +5 in our template on the length of the leader. Slippage or reiterative copying of a template sequence has been described for bacterial (Chamberlin, 1976) and bacteriophage (Kassavetis *et al.*, 1986; Martin *et al.*, 1988) polymer-

ases, but is usually suppressed completely when a substrate triphosphate complementary to the template base distal to the reiterative sequence is present. In this regard, Schwer & Stunnenberg (1988) reported that the length of the leader is reduced when low concentrations of ATP are used in the *in vitro* late transcription system.

The TAAAT element may be an integral part of the late promoter, or it may be involved only in addition of the 5'-poly(A) leader. These two different rôles might be difficult to distinguish if the leader were required for mRNA stability at late times in infection. In this respect, it is important to emphasize that the analysis described here was done using steady-state mRNA. Pertinent information is available from *in vitro* transcription studies, however. Wright & Moss (1987) reported that extracts of cells infected with vaccinia virus were able to transcribe late genes, and that the resulting mRNAs had 5'-poly(A) leaders. When the TAAAT sequence was mutated, however, specific transcription was not detected. The accompanying observations, that early genes were also transcribed in this system and that resulting RNAs lacked 5'-poly(A) leaders, argue against an RNA stability problem *in vitro*. Thus, the currently available information is consistent with rôles for the TAAAT element in promoter activity, mRNA initiation and formation of the 5'-poly(A) leader.

In addition to cDNAs corresponding to mRNAs with 5'-poly(A) leaders of up to about 40 nucleotides, much longer species were detected in the analysis. Such long cDNAs were described by Bertholet *et al.* (1987) and Schwer *et al.* (1987). The former workers provided evidence that the cDNAs consisted of reverse transcripts of two non-contiguously coded mRNAs physically connected (3' to 5') by a poly(A) sequence, and suggested primer or splicing mechanisms for their formation. The latter workers, however, considered that long cDNAs were formed by reverse transcription of read-through transcripts initiated upstream from the late promoter. We suggest a wholly artefactual origin for at least some long cDNAs. A cDNA formed by reverse transcription of an mRNA with a 5'-poly(A) leader would have a 3'-poly(dT) sequence, which, if separated from the template by RNase H activity, could anneal to the 3'-poly(A) of another physically unconnected mRNA, allowing further extension of the first cDNA. This interpretation is consistent with our finding that greater amounts of long cDNAs correlate with stronger late promoters and longer 5'-poly(A) leaders.

The mutational analysis of late promoters was partly motivated by a desire to enhance the utility of vaccinia virus expression vectors. In this regard, we have constructed late promoters with the following features; a tract of 18 or 20 T residues followed by a six nucleotide sequence containing those substitutions that increased the activity of the 28-kD promoter, followed by TAAATA and then by a three nucleotide sequence that contains those substitutions that increased the activity of the

28-kD promoter. The activity of these synthetic promoters *in vivo* was about 100-fold greater than the original version of the 28-kD promoter, and over 50% more than that of copies of the strong late vaccinia (Falkner & Moss, 1988) and cowpox virus (Patel *et al.*, 1988) promoters used previously for high-level expression. Plasmid vectors for expression of foreign genes are currently being tested that incorporate a synthetic strong late promoter, a range of cloning sites for convenient DNA insertion, and features that allow recombination into the vaccinia virus genome and selection and colour-screening of recombinants.

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

References

- Ahn, B.-Y. & Moss, B. (1989). *J. Virol.* **63**, 226–232.
 Bertholet, C., Drillien, R. & Wittek, R. (1985). *Proc. Nat. Acad. Sci., U.S.A.* **82**, 2096–2100.
 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.
 Boon, R. F. & Moss, B. (1977). *Virology*, **79**, 67–80.
 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.
 Chamberlin, M. J. (1976). In *RNA Polymerase* (Losick, R. & Chamberlin, M., eds), pp. 17–68, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
 Chomeczynski, P. & Sacchi, N. (1987). *Anal. Biochem.* **162**, 156–159.
 Cochran, M. A., Puckett, C. & Moss, B. (1985). *J. Virol.* **54**, 30–37.
 Davison, A. J. & Moss, B. (1989). *J. Mol. Biol.* **210**, 749–769.
 de Magistris, L. & Stunnenberg, H. G. (1988). *Nucl. Acids Res.* **16**, 3141–3156.
 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.
 Hanggi, M., Bannwarth, W. & Stunnenberg, H. G. (1986). *EMBO J.* **5**, 1071–1076.
 Hattori, M. & Sakaki, Y. (1986). *Anal. Biochem.* **152**, 232–238.
 Hirt, P., Hiller, G. & Wittek, R. (1986). *J. Virol.* **58**, 757–764.
 Kassavetis, G. A., Zentner, P. G. & Geiduschek, E. P. (1986). *J. Biol. Chem.* **261**, 14256–14265.
 Kates, J. R. & McAuslan, B. R. (1967). *Proc. Nat. Acad. Sci., U.S.A.* **58**, 134–141.
 Lee-Chen, G.-J. & Niles, E. G. (1988). *Virology*, **163**, 80–92.
 Lue, N. F., Buchman, A. R. & Kornberg, R. D. (1989). *Proc. Nat. Acad. Sci., U.S.A.* **86**, 486–490.
 Martin, C. T., Muller, D. K. & Coleman, J. E. (1988). *Biochemistry*, **27**, 3966–3974.
 Miner, J. N., Weinrich, S. L. & Hruby, D. E. (1988). *J. Virol.* **62**, 297–304.
 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.
 Patel, D., Ray, C. A., Drucker, R. P. & Pickup, D. J. (1988). *Proc. Nat. Acad. Sci., U.S.A.* **85**, 9431–9435.
 Rosel, J. & Moss, B. (1985). *J. Virol.* **56**, 830–838.
 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.
 Tamin, A., Villarreal, E. C., Weinrich, S. L. & Hruby, D. E. (1988). *Virology*, **165**, 141–150.
 Venkatesan, S., Baroudy, B. M. & Moss, B. (1981). *Cell*, **25**, 805–813.
 Vos, J. C. & Stunnenberg, H. G. (1988). *EMBO J.* **7**, 3487–3492.
 Wei, C. M. & Moss, B. (1975). *Proc. Nat. Acad. Sci., U.S.A.* **72**, 318–322.
 Weinrich, S. L. & Hruby, D. E. (1986). *Nucl. Acids Res.* **14**, 3003–3016.
 Weir, J. P. & Moss, B. (1984). *J. Virol.* **51**, 662–669.
 Weir, J. P. & Moss, B. (1987a). *Virology*, **158**, 206–210.
 Weir, J. P. & Moss, B. (1987b). *J. Virol.* **61**, 75–80.
 Wright, C. F. & Moss, B. (1987). *Proc. Nat. Acad. Sci., U.S.A.* **84**, 8883–8887.
 Yuen, L., Davison, A. J. & Moss, B. (1987). *Proc. Nat. Acad. Sci., U.S.A.* **84**, 6069–6073.

Edited by J. Karn