The EMBO Journal vol.3 no.10 pp.2373 – 2382, 1984

Effect of the 21-bp repeat upstream element on *in vitro* transcription from the early and late SV40 promoters

Marc Vigneron, Hugo A.Barrera-Saldana[1], Daniel Baty[2], Roger E.Everett[3] and Pierre Chambon

Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Faculté de Médecine, 11 rue Humann, 67085 Strasbourg Cédex, France
[1]Present address: Universidad Autonoma de Nuevo Leon, Facultad de Medicina, AP 1583, Dtp. Bioquimica, Monterrey, N.L., Mexico 64000,
[2]Present address: Centre de Biochimie et de Biologie Moléculaire du CNRS, 31, Chemin J. Aiguier, BP 13277, Marseille Cédex 09, France,
[3]Present address: Medical Research Council, Virology Unit, Church Street, Glasgow G11 5JR, UK

Communicated by P.Chambon

The role of the 21-bp repeat region [simian virus 40 (SV40) coordinates 40 – 103] on early and late SV40 promoter functions has been investigated *in vitro* using a variety of mutated templates. Using either a HeLa whole cell extract or a S100 extract, we analyzed the transcripts by quantitative S1 nuclease mapping. GC-rich motifs contained in the 21-bp direct repeat constituted an essential element for efficient early transcription *in vitro* in agreement with previous *in vivo* results. These GC-rich motifs act in a non-polar fashion, since inversion of the 21-bp region did not reduce early transcription. Some point mutations in the 22-bp imperfectly repeated sequence, that drastically reduce initiations from the early promoter *in vivo*, had little effect *in vitro*, indicating that all the functions of these GC-rich motifs cannot be reproduced *in vitro* at present. The requirement for the 21-bp repeat region was less stringent when the concentration of the early promoter sequence was increased, which suggests that its function may be to facilitate the recognition of the 'weak' SV40 early TATA box. The multiple late start sites were accurately used *in vitro* and the GC-rich motifs contained in the 21-bp repeat region were an important element for efficient *in vitro* initiation of transcription from the late promoter, irrespective of their orientation. However, the effect of the 21-bp repeat region on late initiations decreased strikingly with increasing distance to the start sites, although it was still detectable over a distance of 220 bp. Under the present *in vitro* conditions, the 72-bp repeat region stimulates weakly both early and late transcription.

Key words: SV40/upstream element/RNA polymerase B(II)/promoter/GC-rich motifs

Introduction

The organization of the promoter regions of the early and late transcription units of the simian virus 40 (SV40) has been extensively studied *in vivo*. The promoter region of the early transcription unit contains two overlapping promoters controlling initiation of transcription at the early-early (EE) and late-early (LE) start sites (see Figure 1). The EE promoter comprises the TATA sequence which ensures efficient initiation of transcription from the EE start sites (EES) (for references, see Wasylyk *et al.*, 1983a), and an upstream sequence, the 21-bp repeat region which contains six GC-rich motifs essential for efficient transcription (Baty *et al.*, 1984 and references therein). The four GC-rich motifs III, IV, V and VI are also important elements of the LE promoter (Baty *et al.*, 1984) but the requirement for the GC-rich motifs I and II appears to be unique to the EE promoter. The enhancer which contains the 72-bp repeat also belongs to both promoters, since it activates transcription from both the EES and LES (Wasylyk *et al.*, 1983a) (a promoter element is broadly defined here as any *cis*-acting sequence important for accurate and efficient initiation of transcription). The late promoter region comprises multiple start sites within a region of almost 200 bp (Figure 1) (Ghosh *et al.*, 1978; see also Hansen and Sharp, 1983 and references therein). The major late start site L1 appears to be positioned by a TATA-like element located ~30 bp upstream (Brady *et al.*, 1982), but to date no such elements have been identified for the other start sites. There are strong indications that some sequences within the 21-bp repeat region are essential for efficient transcription from the late start sites (Fromm and Berg, 1982; Hartzell *et al.*, 1984). Whether the 72-bp repeat enhancer plays a role in the efficiency of late transcription *in vivo* is unknown at present.

To understand the molecular mechanisms by which the initiation of transcription is controlled at the SV40 overlapping early and late promoters, it is essential to identify the protein and DNA sequence components which are involved and to reconstruct *in vitro* an efficient system in which all the requirements for accurate and efficient transcription *in vivo* are reproduced. With soluble cell-free extracts the *in vivo* role of the SV40 early TATA box can be reproduced *in vitro* (Mathis and Chambon, 1981; Wasylyk *et al.*, 1983a; Hansen and Sharp, 1983). Deletion mutants show that some sequences within the 21-bp repeat are essential for efficient initiation of transcription *in vitro* from both the early (Myers *et al.*, 1981; Hansen and Sharp, 1983) and late (Hansen and Sharp, 1983; Brady *et al.*, 1984) start sites. It has been suggested that within the 21-bp repeat region, the DNA sequence requirements for efficient transcription *in vitro* closely resemble those *in vivo* (Hansen and Sharp, 1983). Furthermore a protein factor, Sp1, has been partially purified (Dynan and Tjian, 1983a) and shown to bind *in vitro* to the region containing the GC-rich motifs III, IV, V and VI (Dynan and Tjian, 1983b). The availability of point mutants affecting the six GC-rich motifs contained within the 21-bp repeat region, whose effect on SV40 early transcription *in vivo* has been extensively studied (Everett *et al.*, 1983; Baty *et al.*, 1984), prompted us to study further whether the function played *in vivo* by the 21-bp repeat region is faithfully reproduced *in vitro*. We report here that the role exerted *in vivo* by the four GC-rich motifs III, IV, V and VI on the efficiency of initiation from the EE and LE start sites can be reproduced *in vitro* using a soluble cell-free extract, but that the very strong down effect *in vivo* of a mutation within the GC-rich motif I on initiation for the EES cannot yet be reproduced *in vitro*. We

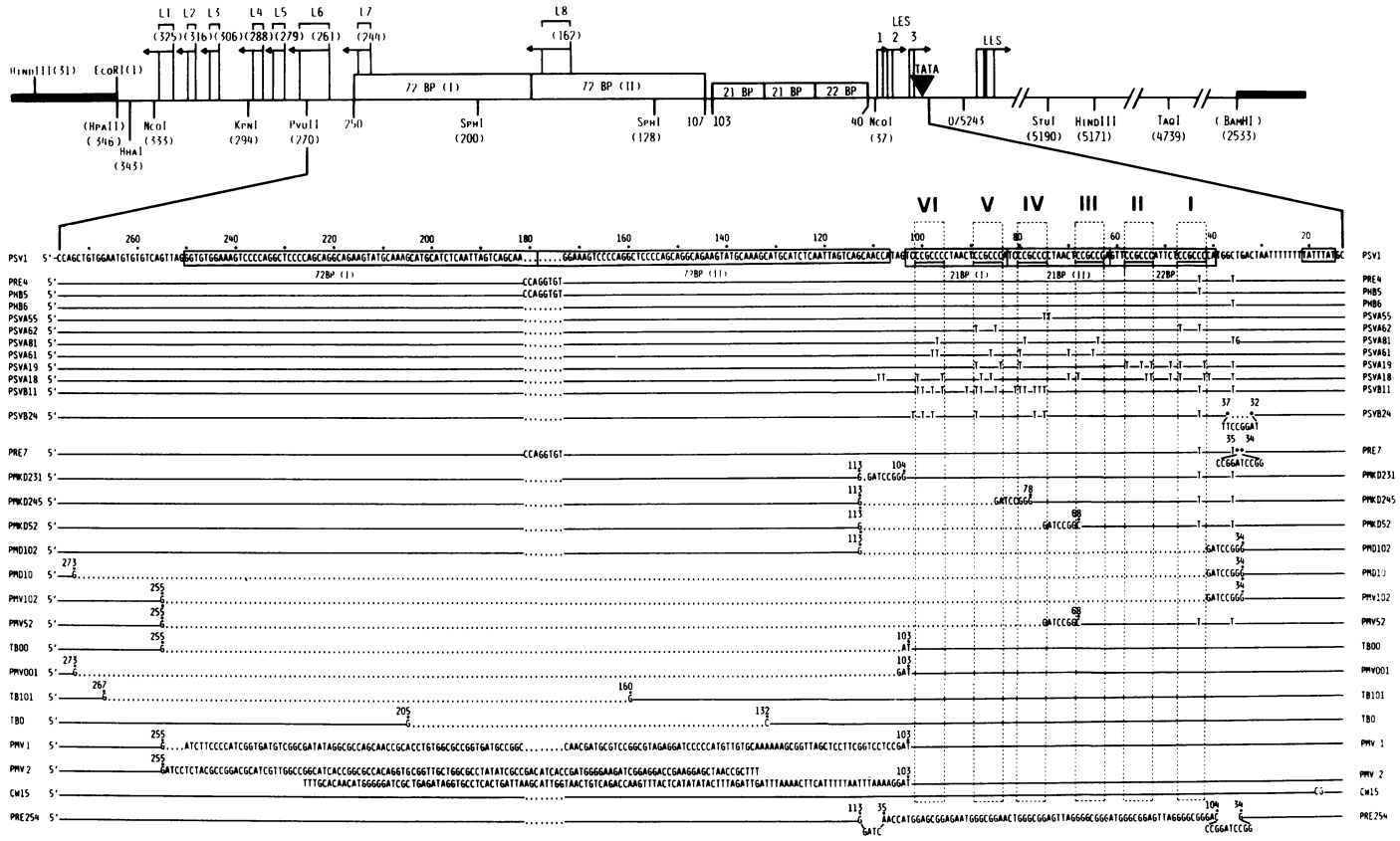

Fig. 1. General organization and sequence of the SV40 early and late promoter region and of its mutant derivatives in the 21-bp and 72-bp repeat regions. The first line represents the wild-type organization of the SV40 origin and early regions in recombinant pSV1 which was derived from SV40 strain 776 (Benoist and Chambon, 1980). Some SV40 (below the line) and pBR322 (above the thick line) key restriction sites are indicated (in parentheses when destroyed during the construction) with their coordinates (numbers in parentheses). Other numbers below the line indicate SV40 nucleotide positions (BBB system, Tooze, 1982). The position of the early start sites EES and LES1 – LES3, and of the late start sites L1 – L8 (their approximate position is indicated in parentheses) are shown above the line. The second line (pSV1) shows the sequence of pSV1 (non-coding early strand) in the region of interest. The six GC-rich motifs (I – VI) are underlined and boxed with dashed lines. The 21- and 72-bp repeat regions are boxed as well as the TATA box. The 72-bp repeat of SV40 strain 776 variant present in pSV1 exhibits an 8-bp deletion (represented by dots, see Benoist and Chambon, 1981) which does not exist in SV40 strain 776 used to construct pRE4, pHB5 and pRE7 (Materials and methods). The sequences of mutants pRE4 to pRE254 are shown below the sequence of pSV1. Deletions are indicated by dots. Dots and numbers above the lines indicate nucleotide positions (see also Materials and methods, and text).

also show that the GC-rich motifs are important *in vitro* for efficient initiation of transcription from the late start sites, especially for those which are most proximal to the 21-bp repeat region. We conclude that at least some of the GC-rich motifs contained within the 21-bp repeat region are functional, irrespective of their orientation (bidirectionality), and that there is no strict distance requirement between these motifs and the 'activated' start sites, although their stimulatory efficiency decreases with increasing distances.

Results

Accurate in vitro initiation of transcription from the SV40 early and late start sites

Transcription from the EE and LE promoters. In vivo studies have identified several start sites for the RNA transcripts from the SV40 early region, the EE start sites (EES and DSB) and the LE start sites (LES) (see Figure 1 and Figure 2B; Wasylyk *et al.*, 1983a; Baty *et al.*, 1984 and references therein). In agreement with previous reports (Mathis and Chambon, 1981; Hansen *et al.*, 1981; Ghosh and Lebowitz, 1981; Myers *et al.*, 1981; Wasylyk *et al.*, 1983a; Hansen and Sharp, 1983) the same start sites are used in an *in vitro* HeLa whole cell extract transcription system (Figure 2A). Under our present stringent S1 nuclease digestion conditions (lanes 5 in Figure 2A and B) the subsets LES0 and LES1 of LE start sites were usually not visible and the subsets LES2 and LES3 were not resolved [the strong bands seen in lanes 2 and 3 of Figure 2A and B are probably due to secondary structure of the origin palindrome (see below); see also Wasylyk *et al.*, 1983a; Baty *et al.*, 1984 for discussion of these points]. In addition to EES, LES and DSB start sites a number of presumptive initiation sites located upstream from LES2 + 3 were detected *in vitro* (see Figure 2A), that do not occur *in vivo*. All these sites disappeared in the presence of 0.1 μg/ml of α-amanitin, indicating that they corresponded to RNA transcribed by RNA polymerase B (see Figure 3, lane 3). Essentially similar results were obtained using an S100 instead of a whole cell extract, although the LES start sites were weaker (data not shown).

Transcription from the late promoter region. The initiation sites for RNA transcribed *in vivo* from the late region of SV40 (Ghosh *et al.*, 1978; Haegeman and Fiers, 1978; Contreras and Fiers, 1981; Hansen and Sharp, 1983) are also used by an *in vitro* transcription system (Hansen and Sharp, 1983). However, because the pattern of the late region initiation site is complex, we decided to compare the location of the *in vivo*

Table I. Effect of mutations on the relative efficiency of transcription from the various SV40 early and late start sites

Mutants	Early transcription start sites				*In vitro* late transcription start sites							
	EES		LES		L1	L2	L3	L4	L5	L6	L7	L8
	In vitro	*In vivo*	*In vitro*	*In vivo*								
pSV1	100	100	100	100	100	100	100	100	100	100	100	100
pRE4	80	9	150	186	115	100	70	75	80	100	190	NA
pHB5	90	10	150	200	ND	ND	ND	ND	ND	ND	ND	NA
pHB6	100	97	100	108	ND	ND	ND	ND	ND	ND	ND	ND
pSVA55	75	95	ND	106	ND	ND	ND	ND	ND	ND	ND	ND
pSVA62	70	4	ND	109	ND	ND	ND	ND	ND	ND	ND	ND
pSVA81	15	17	ND	93	ND	ND	ND	ND	ND	ND	ND	ND
pSVA61	7	8	25	23	75	55	80	50	40	60	50	20
pSVA19	10	4	ND	260	ND	ND	ND	ND	ND	ND	ND	ND
pSVA18	2	1	NA	22	55	40	80	30	20	35	35	5
pSVB11	10	3	ND	49	ND	ND	ND	ND	ND	ND	ND	ND
pSVB24	45	7	ND	390	ND	ND	ND	ND	ND	ND	ND	ND
pRE7	70	20	ND	650	ND	ND	ND	ND	ND	ND	ND	ND
pMKD231	80	10	150	190	ND	ND	ND	ND	ND	ND	ND	ND
pMKD245	15	4	80	60	ND	ND	ND	ND	ND	ND	ND	ND
pMKD52	5	2	60	ND	70	60	70	30	30	35	40	5
pMD102	1	1	NA	10	50	50	70	20	20	25	30	1
pMD10	≤0.5	ND	NA	ND	20	UND	UND	NA	NA	NA	NA	NA
pMV102	ND	ND	ND	ND	20	UND	UND	UND	NA	NA	NA	NA
pMV52	ND	ND	ND	ND	70	UND	UND	UND	NA	NA	NA	NA
TB00	30	ND	30	ND	650	750	500	400	NA	NA	NA	NA
pMV001	25	ND	25	ND	1000	1500	600	NA	NA	NA	NA	NA
TB101	60	0.6	60	ND	250	190	180	90	NA	NA	NA	NA
TB0	60	120	60	ND	150	135	170	140	160	160	NA	NA
CW15	5	3	NA	140	110	80	80	80	85	105	90	100
pRE254	145	22	NA	110	95	80	85	70	60	100	105	70

In vitro transcription data from several independent experiments using different whole cell extracts and DNA preparations, similar to those presented in Figures 2–7, were quantitated by scanning the autoradiograms. The relative intensities of the bands are expressed as the percentage of 'wild-type' pSV1 transcription from the various initiation sites. The *in vivo* data was either reported previously by Baty *et al.* (1984) or unpublished results. ND, not determined; NA, indicates that it is not applicable usually because the start sites have been deleted; UND, below the level of detection.

and *in vitro* start sites as a basis for our future studies with deletion and point mutants. *In vivo* late region RNA was analyzed after transfection of COS cells (Gluzman, 1981) with either SV40 strain 776 DNA or the pSV1 recombinant [which contains the whole early region of strain 776, but lacks the late coding region (Benoist and Chambon, 1980)], or infection with SV40 strain 776 virus. After infection with the virus, several S1 nuclease-resistant bands were revealed (L1–L8 in Figure 2D; see also Figure 1). L1 corresponds to the major late initiation start site previously mapped at position 325, whereas L6 and L8 correspond to the start sites which have been mapped around the *Pvu*II site (position 270) and within the 72-bp repeat (around position 165), respectively (Hansen and Sharp, 1983). The other subsets of bands L2, L3, L4, L5 and L7 which were repeatedly observed most likely correspond to additional start sites of RNA transcribed from the late region and not yet fully characterized (for references, see Hansen and Sharp, 1983). The relative importance of these various late start sites after transfection with SV40 strain 776 DNA or infection with the corresponding virus was estimated by scanning several autoradiograms similar to that shown in Figure 2D (Figure 2E, solid bars).

Two single-stranded DNA probes, the *Stu*I (positions 360–5190 in SV40) and *Ava*II (positions 557–5118 in SV40) probes, were used for quantitative S1 nuclease mapping of the late start sites. No major artefact was introduced for start site L1 when the *Stu*I probe, which results in a short 35–40-bp long hybrid, was used. In fact, the relative importance of the L1 start site rose from ~50% to 65% when the *Ava*II probe was used instead of the *Stu*I probe, whereas no variation was scored for the other start sites. Surprisingly, when pSV1 DNA (which contains pBR322 sequences downstream from SV40 position 346 and lacks all of the late coding sequences, see Figure 1) was transfected into COS cells, the relative importance of the late promoter start sites was markedly different (data not shown). L6 became the major start site of late RNA (~65% of total late RNA), whereas RNA initiated from L1 represented only 7% of this RNA. Whether this puzzling difference is due to variations in RNA stability of the various late RNAs when the late region of SV40 is replaced by pBR322 sequences or reflects an influence of the virion proteins in the relative accumulation of the different RNA initiated from the various late start sites remains to be determined.

However, when SV40 DNA (not shown) or pSV1 (Figure 2C) were used as templates, an identical pattern of late start sites was obtained. Moreover, this pattern (Figure 2E, open bars) was essentially the same as that obtained with SV40 DNA *in vivo* (with possibly some additional minor sites between L7 and L8). It is therefore unlikely that the apparently inefficient use *in vivo* of site L1 of pSV1 (see above) is due to the absence of the virion proteins in cells transfected with pSV1. As in the case of *in vitro* early transcription, essentially the same pattern of late start sites was obtained when a HeLa

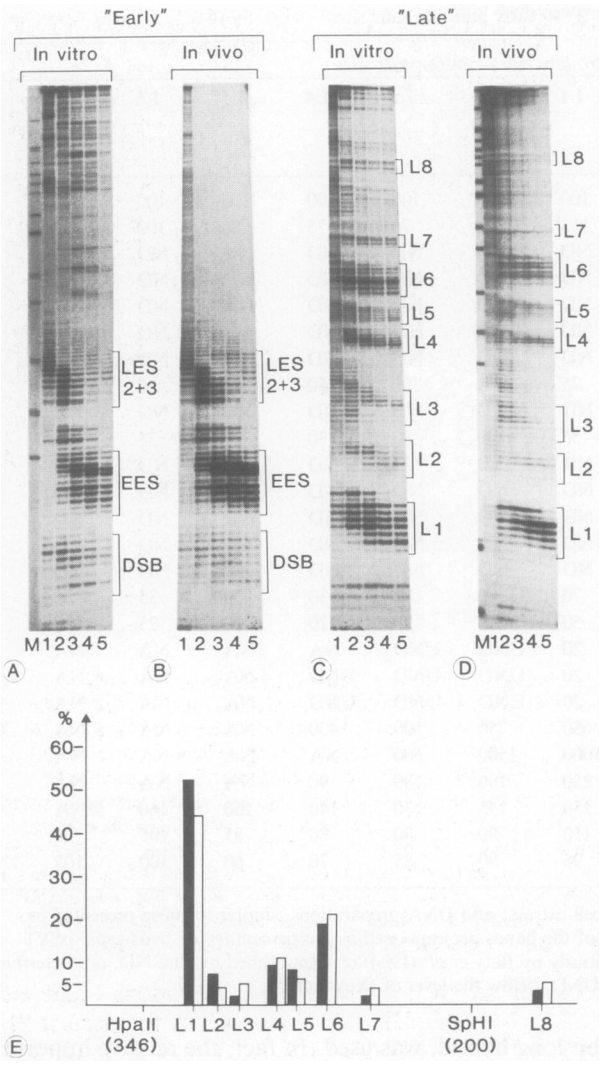

Fig. 2. Comparison of the patterns of transcription from the wild-type early and late promoters *in vivo* and *in vitro* **(A – D): lanes M** correspond to length markers [^{32}P]5′ end-labelled *Msp*I digest of pBR322); **lanes 1 – 5** correspond to digestion with increasing concentrations of S1 nuclease (0.5, 10, 40, 80 and 160 units/450 μl). The S1 nuclease-resistant DNA fragments corresponding to the various early and late start sites are bracketed (see text). The DNA probe for *in vitro* and *in vivo* early and *in vitro* late RNA were the [^{32}P]5′ end-labelled early and late strands of the 443-bp *Hind*III fragment which spans the origin region in pSV1 (Figure 1), respectively. *In vivo* late RNA was probed with the late strand of the 413-bp *Stu*I fragment of SV40 strain 776. *In vitro* early and late RNA were transcribed from superhelical pSV1 DNA (23 μg/ml) in a HeLa whole cell extract (Materials and methods). *In vivo* 'early' RNA was prepared from HeLa cells transfected with pSV1 using calcium phosphate (Wasylyk *et al.*, 1983b). *In vivo* late RNA was obtained from COS cells infected with SV40 strain 776. **(E)** Autoradiograms similar to those shown in panels **C** and **D** were scanned and the relative amount (ordinate) of RNA initiated *in vivo* (solid bars) and *in vitro* (open bars) from the various late start sites (abcissa, L1 – L8) was determined taking total late transcription as 100%. *Hpa*II and *Sph*I sites indicate the relative position of the late start sites (see Figure 1).

cell S100 extract was used in place of a whole cell extract (data not shown) and all L1 to L8 bands disappeared when 0.1 μg/ml α-amanitin was added to the incubation mixture (see Figure 6, lanes 3 and Figure 7, lanes 1). From all the above results we conclude that recombinant pSV1 constitutes a valid template to achieve accurate SV40 late transcription *in vitro* and can, therefore, be used to study *in vitro* the organization and function of the late promoter region.

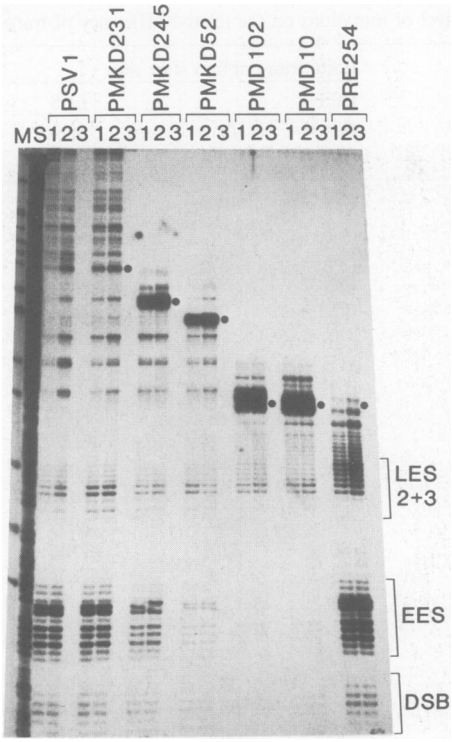

Fig. 3. Effect of deletions within the 21-bp and 72-bp repeat regions or inversion of the 21-bp region on *in vitro* early transcription using a whole cell extract (Materials and methods). DNA template (superhelical DNA) of recombinants as indicated (see Figure 1), concentrations were 19 μg/ml in **lanes 1**, and 27 μg/ml in **lanes 2** and **3**. 0.1 μg/ml α-amanitin was added to the incubation mixtures corresponding to **lanes 3**. The 'early' [^{32}P]DNA probe used for S1 nuclease (160 units) mapping was as indicated in legend to Figure 2. The bands corresponding to the end-point of homology between the probe and the deleted templates are indicated by a black dot. M, length markers as in legend to Figure 2; S, G + A sequence ladder (Maxam and Gilbert, 1980).

Effect of deletions and point mutations within the GC-rich motifs on transcription from the early promoters

A wide range of DNA concentrations were assayed for each SV40 mutant in each experiment to eliminate any possible artefact due to variations in optimum template concentration. The *in vitro* transcription assays were first performed in the presence of a second DNA template containing the adenovirus-2 major late promoter as an internal control. Since the results were identical to those presented here, this internal control was subsequently omitted because minor S1 nuclease-generated bands interfered with the SV40 pattern of start sites. For each mutant the experiments were repeated several times using independent preparations of plasmids and cell extracts. Since very similar results were obtained we are confident of the value of the data.

Initiation from the EE start sites. Superhelical DNA templates were first used to study the effect of mutations within the 21-bp repeat region. The deletion mutants pMKD245, pMKD52 and pMD102 in which two, three and all six of the GC-rich motifs were deleted, respectively, showed a dramatic decrease for EE initiations when compared with pSV1 or pMKD231 (see Figure 1, Figure 3 and Table I), indicating that the 21-bp repeat region constitutes an essential element for efficient *in vitro* initiation from EES and DSB. However, the 21-bp repeat region is not required for the specificity of initiation, since the correct pattern of EES and DSB was still visible for pMD102 on a longer exposure of the gel shown in

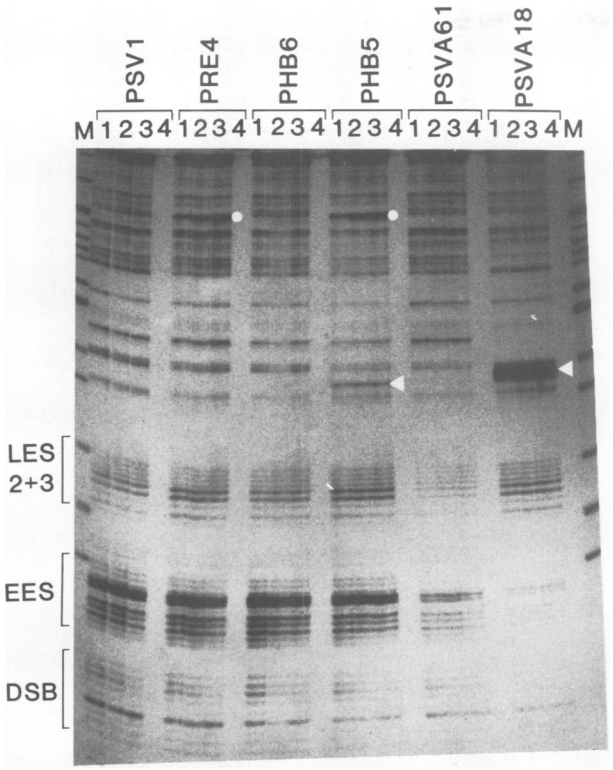

Fig. 4. Effect of point mutations within the 21-bp repeat region on *in vitro* early transcription using a whole cell extract (Materials and methods). DNA template (superhelical DNA of recombinants as indicated, see Figure 1) concentrations were 20 μg/ml in **lanes 1**, 32 μg/ml in **lanes 2** and **4**, 44 μg/ml in **lanes 3**. 0.1 μg/ml α-amanitin was added to the incubation mixture corresponding to **lanes 4**. The DNA probe and the conditions for quantitative S1 nuclease mapping were as indicated in legend to Figure 3. The white arrowheads indicate additional start sites specific to pHB5 and pSVA18. End-points of homology (see legend to Figure 3) are indicated by white dots. M, as indicated in legend to Figure 2.

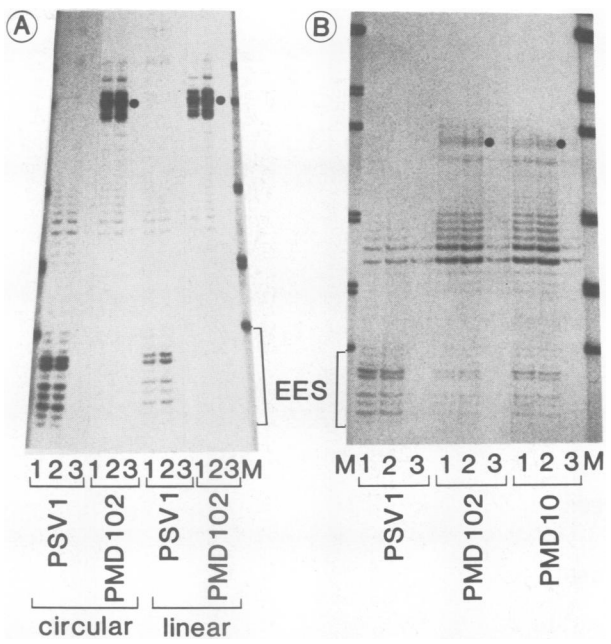

Fig. 5. Effect of the nature of DNA template on *in vitro* early transcription using a whole cell extract (Materials and methods). The origin of the DNA templates was as indicated below the autoradiograms. In all cases the incubation mixtures corresponding to **lanes 3** contained 0.1 μg/ml α-amanitin. The DNA probe and the conditions for quantitative S1 nuclease mapping were as in legend to Figure 3. End-points of homology are indicated by a black dot (see legend to Figure 3). **(A)** Superhelical circular and *Ava*I linearized pSV1 or pMD102 DNA (as indicated) were used as template at concentrations of 19 μg/ml **(lanes 1)** or 27 μg/ml **(lanes 2** and **3)**. **(B)** The purified *Taq*I-*Hha*I fragments (see Figure 1) of pSV1, pMD102 or pMD10 were used at concentrations of 10 μg/ml **(lanes 1)** or 13 μg/ml **(lanes 2** and **3)**. M, as indicated in legend to Figure 2.

Figure 3 (see also below Figure 5B). In mutant pRE254 where the whole 21-bp repeat region has been inverted, initiation from the EES and DSB was increased by ~50% when compared with pSV1 or pMKD231. In agreement with our previous *in vivo* results, the 21-bp repeat region can apparently function bidirectionally *in vitro*, but its insertion in the reverse orientation may be less efficient *in vivo* than *in vitro* (Table I). That the present *in vitro* transcription system may not perfectly reproduce the *in vivo* situation is also apparent from the results obtained with pMKD231, which is strongly down *in vivo*, but not *in vitro* (see below).

We examined the role of the GC-rich motifs themselves by using various single or multiple point mutants (pRE4, pHB5, pHB6, the pSVA and pSVB series, see Figure 1). The results are summarized in Table I (see also Figure 4) in comparison with the *in vivo* data corresponding to the same mutants (Baty *et al.*, 1984 and our unpublished results). With pHB6, pSVA55, pSVA81, pSVA61, pSVA19, pSVA18 and pSVB11 there is a good agreement between the results obtained *in vitro* and *in vivo*, whereas there are marked differences with pRE4, pHB5, pSVA62, pSVB24 and pRE7. The latter all have in common the transition C→T at position 43 which is also present in pMKD231 (see above) [the presence of a complete 72-bp repeat in pRE4 and pHB5 (see legend to Figure 1) does not affect early transcription (our unpublished results)]. Thus the marked down effect *in vivo* of a single C→T transition in the GC-rich motif I was not reproduced *in vitro*. On the other hand, the *in vivo* effect of mutations which affect the GC-rich motifs III, IV, V and VI of the perfect 21-bp repeat (pSVA55, pSVA61, pSVA81) is well mimicked *in vitro*. From these latter results and those obtained with mutants which are heavily mutated in both the 22-bp sequence and the 21-bp repeat (pSVA18 and pSVB11), we conclude that the GC-rich motifs III, IV, V and VI are important for *in vitro* EES initiation, although mutation of one of them alone results in only a moderate down effect (pSVA62). Since the only two mutants which affect the GC-rich motif II (pSVA19 and pSVA18) are also mutated elsewhere in the 21-bp repeat region, we cannot yet say whether this repeat plays any role in the efficiency of early *in vitro* transcription. The effect of mutations on the DSB initiations generally parallels that on EES (Figure 4, and data not shown).

In the above experiments we used superhelical circular templates which are rapidly converted to relaxed covalently closed circles on incubation with a whole cell extract (Hen *et al.*, 1982). Figure 5A presents a comparison of transcription of pSV1 and pMD102 templates, either circular or linearized by digestion with *Ava*I. Although the linear template appears to be used less efficiently, there is still a dramatic decrease in EES transcription when the 21-bp repeat region is deleted. The same result was obtained when *Eco*RI was used instead of *Ava*I, and a S100 extract instead of a whole cell extract.

On the other hand, the length of the linear template has a marked effect on the requirement of the 21-bp repeat region for efficient *in vitro* initiation of transcription from the EES. This is shown in Figure 5B where the template efficiency of the *Hha*I-*Taq*I fragment (positions 343−4739, see Figure 1)

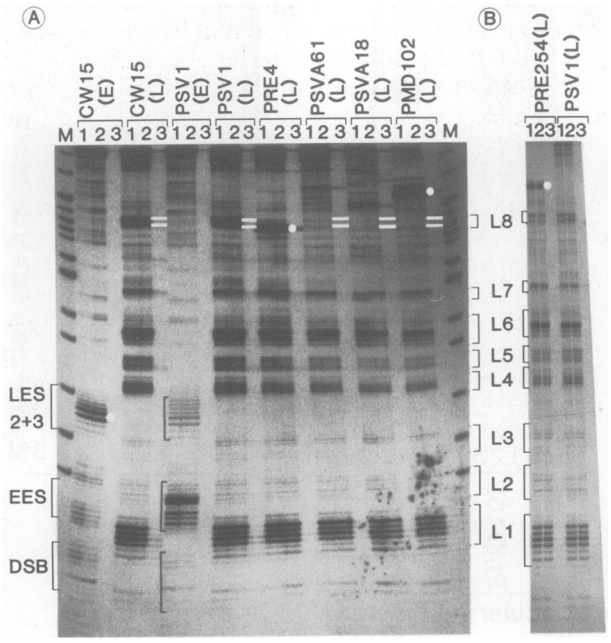

Fig. 6. Effects of mutations and inversion of the 21-bp repeat region on *in vitro* late transcription and of a mutation within the TATA box on *in vivo* early and late transcription, using a whole cell extract (Materials and methods). The origin of the superhelical DNA templates was as indicated above the autoradiograms. In all cases the incubation mixtures corresponding to **lanes 3** contained 0.1 μg/ml α-amanitin and the DNA template concentrations were 19 μg/ml **(lanes 1)** or 29 μg/ml **(lanes 2** and **3)**. The DNA probes and conditions used for quantitative S1 nuclease mapping of early (E) and late (L) RNA transcripts were as in legend to Figure 2. End-points of homology (white dots) are indicated (see legend to Figure 3). **Panels A** and **B** correspond to two different gel electrophoreses. The position of the start site L8 bands is indicated throughout panel **(A)** with white lines. The white dots correspond to end-points of homology between the DNA probe and the DNA template (see legend to Figure 3). M, as in legend to Figure 2.

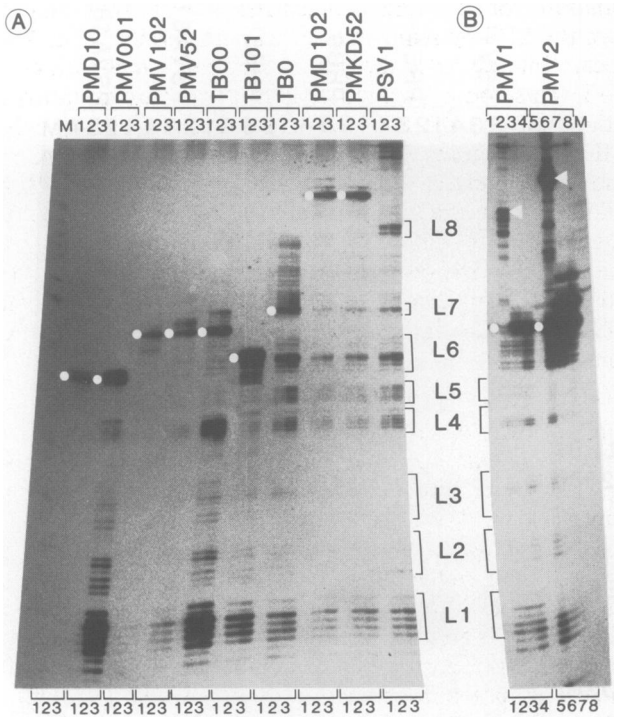

Fig. 7. Effect of distance between the 21-bp repeat and the late promoter regions on efficiency of *in vitro* transcription from the late start sites using a whole cell extract (Materials and methods). The origin of the superhelical templates (see Figure 1) was as indicated above the autoradiograms. In all cases the incubation mixtures corresponding to **lanes 1** contained 0.1 μg/ml α-amanitin. The DNA probe and conditions for quantitative S1 nuclease mapping of *in vitro* late RNA transcripts were as described in legend to Figure 2, with the exceptions noted below. End-points of homology are indicated by a white dot (see legend to Figure 3). **(A)** Superhelical DNA concentrations were 23 μg/ml **(lanes 1** and **3)** and 19 μg/ml **(lanes 2)**. It should be stressed that in the case of TB0 the end-point of homology indicated by a white dot corresponds to the shortest possible hybrid (since TB0 contains only one 72-bp sequence and the DNA probe contains the 72-bp repeat, the possibility exists that the first 72-bp sequence of the probe totally loops out). **(B)** Superhelical DNA concentrations were 23 μg/ml **(lanes 1, 2, 4, 5, 6** and **8)** and 19 μg/ml **(lanes 3** and **7)**. In **lanes 2** and **6**, the homologous probes prepared from pMV1 and pMV2 were used instead of the *Hind*III late probe used in all other lanes. The white arrowheads point to strong start sites located in pBR322 sequences. Note that there are a number of bands located below the L1 region. Since they correspond to an AT-rich region, we interpret them as resulting from artefactual S1 nuclease cuts in the longer hybrid molecules. Therefore we did not take them into account in the determination of the amount of RNA initiated from the L1 start sites in pMV1 and pMV2 which were 35% and 15% of pSV1, respectively. M, as indicated in legend to Figure 2.

of pSV1 (850 bp), pMD102 (780 bp) and pMD10 (620 bp) are compared. In marked contrast with the previous results obtained with the 7.4-kb circular or linear templates (Figure 3 and Figure 5A), the pMD102 and pMD10 deletions cause only ~4-fold and 2-fold decreases in initiation at the EES, respectively. The strong bands present in pMD102 and pMD10 at the level of LES2+3 probably correspond to hybrids between the probe and transcripts initiated upstream from the deletion end-point, which were trimmed down to the AT-rich sequence under the stringent S1 nuclease conditions used. Apparently the 21-bp repeat becomes at least partially dispensable when the concentration of the early promoter sequence increases; this suggests that its function is to facilitate the recognition of the 'weak' SV40 early TATA box, when the latter has to compete with the multiple TATA-like sequences which are present in the rest of the 7.4-kb template (Sassone-Corsi *et al.*, 1981; Mathis *et al.*, 1981).

Initiation from the LE start sites. As discussed above, the LES are used *in vitro* in the presence of a whole cell extract and to a lower extent with an S100 extract. However, there is a major difficulty in quantitating these initiations, because stringent S1 nuclease digestion conditions are required to eliminate all of the partial digests, which are probably caused by the perfect inverted repeat extending from position 5230 to 13 (see Figure 1 and Figure 2A and B). Under these stringent conditions the RNA initiated at the major late LES2 and LES3 start sites located at positions 21−22 and 30−31 (Wasylyk *et al.*, 1983a; Baty *et al.*, 1984) yield bands mapping at around position 15, due to S1 nuclease trimming throughout the AT-rich region. For the same reason any RNA initiated within or close to the AT-rich sequence at positions different from LES2 and LES3 will also map at approximately position 15. Similarly, an end-point of homology between the DNA probe and the RNA, located close to this AT-rich region, will also result in bands mapping at approximately position 15. We have not, therefore, attempted to quantitate initiations from LES2 and LES3 whenever there was a strong band (end-point of homology or putative new start site) immediately upstream from the AT-rich sequence (for example pMD102 and pMD10 in Figure 3 and pSVA18 in Figure 4).

The deletions present in the 21-bp repeat region of pMD245 and pMKD52 have the same effect on transcription *in vitro* and *in vivo* from the LES (Figure 3 and Table I). Apparently

initiation from the LES is much less affected than initiation from the EES by removing the GC-rich motifs of the 21-bp repeat. Interestingly, the point mutation which causes the C→T transition at position 43 is an *in vitro* up mutation for initiation from the LES, as it is *in vivo* (pRE4, pHB5, pMKD231, Figures 3 and 4). In marked contrast this mutation has no effect *in vitro* on initiation from the EES, but strikingly depresses such initiation *in vivo* (see above). The results obtained with the multiple point mutant pSVA61 which exhibits the same decrease *in vitro* as *in vivo* (Figure 4 and Table I), indicates that the GC-rich motifs of the perfect 21-bp repeat are important for efficient *in vitro* initiation from the LES.

Effect of the 21-bp repeat region on late transcription

Effect of deletions and point mutations within the GC-rich motifs on initiation from the late start sites. The 21-bp repeat region plays a definite role in the efficiency of *in vitro* transcription from the late start sites (Figures 6 and 7, Table I). However, it has a much more pronounced effect on the proximal than on the distal start sites. Using as templates the mutants pMKD52 and pMD102 in which three and all of the GC-rich motifs are deleted, respectively, resulted in an almost complete disappearance of initiation from start site L8, whereas initiation from start sites L7−L4 is decreased by ~70% and initiation from start site L1 only by 30% and 50% (Table I, Figures 6 and 7). Similar results were obtained with the multiple point mutants pSVA61 and pSVA18 (Figure 6, Table I), indicating that at least some of the C contained within the GC-rich motifs are involved in this effect of the 21-bp region on *in vitro* late transcription. The C→T transition at position 43 which has a marked effect on EE transcription *in vivo*, but not *in vitro* (see above), does not affect *in vitro* initiation from the late start sites (Figure 6). Using the mutant pRE254, with the 21-bp repeat region inserted in the reverse orientation, as a template did not affect the efficiency of initiation from the late start sites (with the possible exception of L8). This supports the conclusion that the 21-bp repeat region is a bidirectional promoter element.

The distance between the 21-bp repeat and the late promoter regions has a marked effect on the efficiency of in vitro initiation from the late start sites. Since the most distal late start site, L1, was the least affected by mutations within the 21-bp repeat region, we asked if modifying the distance between the 21-bp repeat and the late start sites changed their efficiency as initiators of transcription *in vitro*. Mutants TB0, TB101, TB00 and pMV001 (Figure 1), with deletions of increasing lengths within the 72-bp repeat region between the 21-bp repeat region and the late start sites were used as templates. Initiation from the late start sites increased markedly as the distance between the 21-bp repeat and these sites decreased (Figure 7 and Table I). For example, initiation from start site L1 increased by 1.5, 2.5, 6.5 and 10-fold when the deletion mutants, TB0, TB101, TB00 and pMV001 were used as template, respectively. The results with mutant pMV001 also show that the sequences responsible for initiation from start sites L1, L2 and L3 are located downstream from position 273. A comparison between TB00 and pMV001 (Figure 7) indicates that some sequences necessary for efficient initiation from start site L4 are located between positions 255 and 273. Similarly a comparison between TB0 and TB00 (Figure 7) indicates that sequences important for efficient initiation from start sites L5 and L6 are located upstream from position 255.

Comparison of the results obtained with templates TB00, pMV52 and pMV102, and templates pMV001 and pMD10 (Figure 7 and Table I) shows that it is in fact the 21-bp repeat region which is responsible for the increased efficiency of initiation when brought closer to the late start sites. In both cases, deleting the 21-bp repeat region drastically decreased initiation from the late start sites. We also checked that it is the decrease in distance between the 21-bp repeat region and the late start sites, and not the deletion of some specific 'inhibitory' sequences located within the 72-bp region, which is responsible for the increase in efficiency of late start site initiation. In pMV1 and pMV2 the 72-bp repeat segment from position 104−254 was replaced by pBR322 fragments 138 and 243 bp long, respectively (Figure 1). As shown by using fully homologous (Figure 7B, lanes 2 and 6) and partially non-homologous (lanes 3, 4, 7 and 8) DNA probes for the S1 nuclease analysis, there are numerous start sites within these fragments, whereas initiation of transcription from the SV40 late start sites is much lower than in TB00 (where the 104−254 segment is deleted) and not higher than in pSV1 (where the 72-bp repeat region is present) (see legend to Figure 7).

Effect of the 72-bp repeat region on late in vitro transcription. Since the stimulation of initiation from the late start sites by the 21-bp repeat region is strongly 'distance-dependent', the possible effect of deleting the 72-bp repeat region (the enhancer) on late transcription can be studied only by comparing the recombinants pMV102 and pMD10, in which the 21-bp repeat and the enhancer regions are deleted, with recombinant pMD102 in which only the 21-bp region is deleted. As shown in Figure 7 and Table I, initiation from start sites L1−L4 is lower in pMV102 and pMD10 than in pMD102. It is 2.5-fold lower for start site L1, the only site for which it can be measured with reasonable accuracy. Under the present *in vitro* conditions late transcription is apparently stimulated weakly by the 72-bp repeat, in fact to the same extent as transcription from the EES (Table I).

Effect of point mutations within the early 'TATA' box on early and late transcription

To examine possible competition for the transcription machinery between the early and late promoters, we have used as template recombinant CW15 in which the early TATA box which directs initiation at the EES is mutated from 5′-TATTTAT-3′ to 5′-TATCGAT-3′ (Wasylyk *et al.*, 1983a). The *in vivo* strong down effect of this mutation on initiation from EES can be mimicked using an S100 extract (Wasylyk *et al.*, 1983a). As shown in Figure 6 and Table I, the same result is obtained *in vitro* with a whole cell extract. In addition, as previously noted both *in vivo* and *in vitro* (Wasylyk *et al.*, 1983a), this mutation provokes an increase in initiation from the more distal DSB start sites, when compared with the wild-type recombinant pSV1 (Figure 6). On the other hand, the *in vitro* late transcription pattern is nearly identical to that of pSV1 (Figure 6 and Table I).

Discussion

We have shown that the GC-rich motifs contained in the 21-bp repeat region play a key role in the efficiency of *in vitro* initiation of transcription from the EE and LE start sites of the SV40 early promoter. This conclusion is in keeping with the reports of Myers *et al.* (1981) and Hansen and Sharp (1983) who showed, using deletion mutants, that transcription *in vitro* from the SV40 early promoter was strongly dependent on the 21-bp repeat region. The *in vivo* effect of

mutations which affect the GC-rich motifs III, IV, V and VI (Baty *et al.*, 1984) appears to be accurately reproduced (Table I). On the other hand the C→T transition at nucleotide 43 within the GC-rich motif I, which is a strong down mutation for *in vivo* initiation from the EES, has very little effect on *in vitro* transcription from the same sites. However, as *in vivo*, it appears to increase transcription from the LES. Thus, this latter increase cannot be simply accounted for by a 'decreased competition' from the EE promoter, as it was proposed to explain the stimulatory effect *in vivo* of the C→T transition at position 43 on LES transcription (Baty *et al.*, 1984). It is interesting that the effect of mutations in the GC-rich motifs III, IV, V and VI, on SV40 early transcription can be reproduced *in vitro*, and that these GC-rich motifs correspond to those which have been shown by Dynan and Tjian (1983b) to interact *in vitro* with the SV40-specific initiation factor Sp1 (Dynan and Tjian 1983a). Possibly an additional factor absent or not functional in the HeLa whole cell extract, may be involved *in vivo* in efficient initiation of transcription from the EES. Alternatively, Sp1 could be the only factor interacting with the GC-rich motifs, but the 'correct' *in vivo* organization of the SV40 early promoter region may not be accurately reconstructed *in vivo*.

The almost absolute requirement for the GC-rich motifs III–VI for efficient initiation *in vivo* and *in vitro* from the SV40 EE promoter is unique in comparison with other upstream elements of RNA class B (II) promoters. Dependence on upstream sequences for efficient transcription *in vitro* has been observed for other promoters (Tsuda and Suzuki, 1981; Grosschedl and Birnstiel, 1982; Hen *et al.*, 1982), but the magnitude of this effect was <10-fold. This very strong requirement for the 21-bp repeat is probably related to the 'weakness' of the SV40 early TATA box element which, as pointed out by Hansen and Sharp (1983), is not flanked by GC-rich segments. This assumption is strongly supported by the observation that transcription from the EES becomes visible in the absence of the 21-bp repeat provided smaller DNA fragments are used as templates (Figure 5), presumably under conditions where the SV40 early TATA box has not to compete with the TATA-like sequences present in pBR322 (Sassone-Corsi *et al.*, 1981; Mathis *et al.*, 1981).

The *in vivo* pattern of initiation starts in the SV40 late promoter region is also reproduced *in vitro* with reasonable accuracy under the present conditions of incubation (Figure 2). As noted by Hansen and Sharp (1983) and Brady *et al.* (1984) the major *in vivo* start site L1, at position 325, is also a major site *in vitro*. For the other start sites L2–L8 our results generally agree well with those of Hansen and Sharp (1983), with the exception of L8 which was relatively stronger in their study. The *in vivo* studies of Fromm and Berg (1982) and Hartzell *et al.* (1984) have indicated that the 21-bp repeat region may be an important element of the SV40 late promoter. Using the mutants of Fromm and Berg (1982), Hansen and Sharp (1983) have shown that deletions within the 21-bp repeat region have a marked down effect on the efficiency of initiation of transcription from the late start sites *in vitro*. The present data obtained with C→T transition mutants demonstrate that the GC-rich motifs III, IV, V and VI contained within the 21-bp repeat region belong to a later promoter element indispensable for efficient transcription from the late start sites. In agreement with the results of Hansen and Sharp (1983) obtained with deletion mutants, it is clear from our data that the late start site L8, which is proximal to the 21-bp repeat region, is totally dependent on the GC-rich motifs, whereas transcription from the more distal start sites is only partially dependent on these motifs. Whether or not all four GC-rich motifs III, IV, V and VI are equally important for efficient initiation from the early and late start sites cannot yet be inferred. Also, we do not know whether the present *in vitro* data faithfully mimics the *in vivo* effect of the mutations on late initiation of transcription. *In vivo* and *in vitro* studies, using a series of mutants in which only one of the six GC-rich motifs has been mutated at a time, are in progress to answer these questions and to analyze further how a single region, the 21-bp repeat region, can be shared by the three, early-early, late-early and late overlapping promoters.

Since deletion of the entire 21-bp repeat region (e.g., pMD102) did not lead to a complete disappearance of initiation from start sites L1–L7, the question arises as to which additional DNA sequences are required. The data obtained with the deletion mutation pMV001 showed that sequences responsible for initiation from start sites L1 (position 325), L2 (position 316) and L3 (position 306) are located downstream (in the late direction) from position 273. Similarly, comparison of the results obtained with TB00 and pMV001 indicates that sequences required for efficient initiation from start site L4 (position 288) are located between position 255 and 273. These results, reminiscent of those of Brady *et al.* (1982) who showed that *in vitro* initiation from start site L1 is dependent on a sequence located between positions 296 and 302, suggest very strongly that start sites L2, L3 and L4 are positioned by sequences which are functional counterparts of the canonical TATA box element and are similarly located ~30 bp upstream from these start sites.

In contrast to Brady *et al.* (1984), who concluded from an *in vitro* study that the 21-bp repeat region is an element of the late promoter which may function independently of its position, we found, in agreement with Hansen and Sharp (1983), that deletions which bring the 21-bp repeat region closer to the late start sites gave considerably higher levels of transcription from these sites. Our data also show that this increase is not due to deletion of inhibitory sequences contained within the 72-bp repeat, since replacing the latter with pBR322 sequences did not result in any stimulation of initiation from the remaining late start sites.

A comparison of the results obtained with pMD10 and pMV102 on the one hand, and pMD102 on the other, strongly suggests that the SV40 72-bp repeat enhancer stimulates ~2.5-fold *in vitro* initiation from start site L1. Whether the enhancer sequence which plays a crucial role in *in vivo* initiation from the early promoter (Benoist and Chambon, 1981; Gruss *et al.*, 1981) is also important for efficient initiation of transcription from the late promoter remains to be seen. In this respect, it is noteworthy that, in agreement with our previous report (Sassone-Corsi *et al.*, 1984), the deletion of the 72-bp repeat enhancer results in only a moderate decrease in initiation from the EES under the present *in vitro* conditions (3- to 4-fold for mutants TB00 and pMV001).

As reported by Wasylyk *et al.* (1983a) who used an S100 extract instead of a whole cell extract, a mutation within the early TATA box (recombinant CW15) causes a drastic decrease in *in vitro* initiation from the EES which mimics the *in vivo* situation. This decrease is accompanied by a partially compensatory increase in initiation from 'normally' minor start sites such as DSB (Figure 8). However, no increase was noted for initiation from the late start sites (Figure 6 and Table I), indicating that the early and late promoters are not

competing with each other. Further studies will show whether this lack of competition reflects the *in vivo* situation.

Even though we do not yet know whether all of the GC-rich motifs III – VI have the same effect on the SV40 early and late promoters, it is clear from the data obtained with the recombinant in which the 21-bp region is inverted (pRE254), that at least some of these polar motifs act in both directions to stimulate *in vitro* transcription. That the same bidirectionality operates at least partially *in vivo* has been shown for the early promoter (Everett *et al.*, 1983; Baty *et al.*, 1984). Two other upstream elements, those of the herpes simplex virus type 1 glycoprotein D (Everett, 1984) and thymidine kinase (McKnight *et al.*, 1984) genes, which also contain G- or C-rich motifs, are able to function in both orientations. It is striking that the 21-bp repeat of SV40 stimulates initiation of transcription from both the EES, which is located ~80 bp away, and the late start sites L8 – L1, which are located ~60 – 220 bp away, although the extent of stimulation decreases as the distance between the 21-bp repeat and the potentiated start sites increases. Similarly, the GC-rich upstream elements of the thymidine kinase gene need not be at a fixed distance from the activated start site (McKnight and Kingsbury, 1982). It is also remarkable that the SV40 21-bp repeat region can activate transcription *in vivo* from heterologous TATA box elements (R.Hen and N.Myamoto, personal communication). Thus, the three features, bidirectionality, action at 'distance' and stimulation of heterologous promoter elements, are not unique to enhancer sequences (for references, see Wasylyk *et al.*, 1983b,1984). How general the similarity between the properties of upstream elements and enhancers is, remains to be investigated, but the question of whether they are functionally related and differ mainly by the magnitude of their effect is obviously raised by the above observations. In this respect, it is worth recalling that the ability to generate an altered chromatin structure over its own sequence is a unique property of the SV40 enhancer, not shared by the 21-bp repeat upstream element (Jongstra *et al.*, 1984).

Materials and methods

Bacteria and plasmids

The bacterial host strain and plasmids (Figure 1), pSV1, pRE3, pRE4, pRE7, pRE254, pHB5, pHB6, pSVA18, pSVA19, pSVA55, pSVA61, pSVA62, pSVA81, pSVB11, pSVB24, pMKD52, pMKD231, pMKD245, pMD10 and pMD102, have been described previously (Everett *et al.*, 1983; Baty *et al.*, 1984) as well as TB0 and TB101 (Moreau *et al.*, 1981) and CW15 (Wasylyk *et al.*, 1983a) (see Figure 1).

Constructions

TB00 (Figure 1) which lacks the 72-bp repeat (deletion from positions 104 – 254, see Figure 1) was constructed as follows. The 97-bp *Eco*RI-*Bgl*II fragment of M13mp9TBO*Bgl*II which contains the segment 254 – 346 of SV40, was ligated to plasmid pHB3 digested by *Eco*RI and *Bam*HI to yield pBBD (pHB3, a gift of H.Barrera, was constructed by cloning the 348-bp *Eco*RI-*Hind*III fragment of M13mp9TBO*Bam*HI, which contains the 5171 – 346 segment of SV40, in place of *Eco*RI-*Hind*III fragment of pRE3). The 155-bp *Nco*I fragment of pBBD was inserted in place of the *Nco*I fragment of pSV1 to give TB00.

pMV001 (Figure 1) bearing a deletion from positions 104 to 272 was derived from TB00 and pMD10. The 301-bp *Hind*III fragment of TB00 (Figure 1) was purified, treated with BAP (bacterial alkaline phosphatase), cut by *Sau*3A (position – 102 of TB00, see Figure 1) to yield a 127-bp fragment and a 174-bp fragment. pMD10 (Figure 1) was digested by *Eco*RI, treated with BAP and cut by *Bam*HI to yield an 80-bp *Eco*RI-*Bam*HI fragment that was purified. The 174-bp *Hind*III-*Sau*3A fragment of TB00 and the 80-bp *Eco*RI-*Bam*HI fragment of pMD10 were ligated, digested by *Nco*I and inserted into pSV1 in place of its own *Nco*I fragment to give pMV001.

pMV52 (Figure 1) bearing a deletion from positions 69 – 254 was derived from TB00 and pMKD52. The 413-bp *Hind*III fragment of pMKD52 (Figure 1) was purified, treated with BAP and cut by *Bam*HI to yield a 271-bp fragment and a 142-bp fragment. The 127-bp *Sau*3A-*Hind*III fragment of TB00 and the 142-bp *Hind*III-*Bam*HI fragment of pMKD52 were ligated, digested by *Nco*I and inserted into pSV1 in place of its own *Nco*I fragment to give pMV52.

pMV102 containing a deletion extending from position 35 – 254 (Figure 1) was derived from TB00 and pMD10. The 127-bp *Sau*3A-*Hind*III fragment of TB00 was digested by *Nco*I and cloned into pMD10 digested by *Bam*HI and treated with BAP.

The 127-bp and the 174-bp *Sau*3A-*Hind*III fragments of TB00 were mixed with a total *Sau*3A digest of pBR322 and ligated in the presence of 11% PEG 6000 (Pheiffer and Zimmerman, 1983). The products of ligation were then digested by *Nco*I and cloned into pSV1 in place of its own *Nco*I fragment. pMV1 and pMV2 contain, respectively, a 138-bp and a 243-bp fragment inserted between position 104 and 254 (Figure 1).

In all cases the DNA fragments were purified by polyacrylamide gel electrophoresis and eluted from the slab gel by electroelution. The sequence of the *Hind*III fragment (position – 5171 in SV40 to 31 in pBR322 – see Figure 1), of all of the recombinants was verified according to Maxam and Gilbert (1980).

In vitro transcription and quantitative S1 nuclease mapping analysis of the RNA products

HeLa whole cell extracts prepared according to Manley *et al.* (1980) were dialyzed against a buffer containing 0.1 M KCl, 5 mM $MgCl_2$, 0.5 mM DTT, 50 mM Tris pH 7.9 and 20% glycerol and stored in aliquots at – 180°C. S100 extracts were prepared as described in Weil *et al.* (1979) and similarly dialyzed and stored. The composition of the *in vitro* transcription assay was as described in Wasylyk *et al.* (1980) except that all four nucleotide triphosphates were at 0.5 mM. DNA templates were either superhelical or linearized plasmids or purified DNA fragments as indicated in the legend to figures. The transcripts were analyzed by a quantitative S1 nuclease mapping technique. The nucleic acids extracted from the *in vitro* transcription assay were hybridized for 12 h at 42°C to an excess of ^{32}P end-labelled 5′ single-stranded probe, in the presence of 0.4 M NaCl, 50% formamide and 10 mM Pipes, pH 6.5 (Wasylyk *et al.*, 1980). The hybridization mixture was then diluted 9-fold for S1 nuclease digestion in 0.4 M NaCl, 3 mM $ZnCl_2$ and 30 mM sodium acetate pH 4.5. After incubation for 2 h at 25°C, the products of digestion were phenol-extracted, separated on sequencing gels and autoradiographed as previously described (Wasylyk *et al.*, 1980).

Enzymes

Restriction enzymes and other enzymes were obtained from commercial suppliers and used according to their instructions (S1 nuclease was purchased from BRL).

Acknowledgements

We thank Dr C.Goding for critical reading of this manuscript. We are grateful to Drs T.Reudelhuber for help in *in vivo* experiments, T.Grunstrom, M. Zenke and B.Wasylyk for generous gifts of recombinant plasmids. We thank Mrs B.Augsburger for cell cultures and C.Werlé, B.Boulay, B.Chambon and C.Kutschis for preparing the manuscript. D.Baty was a Boursier de Reconversion (MIR), H.Barrera-Saldana and R.Everett were recipients of EMBO long-term and Royal Society fellowships, respectively. This work was supported by the CNRS (ATP 3582), the Ministère de l'Industrie et de la Recherche (82V1283), the Fondation pour la Recherche Médicale, the Association pour le Développement de la Recherche sur le Cancer and the Fondation Simone et Cino del Duca.

References

Baty,D., Barrera-Saldana,H., Everett,R.D., Vigneron,M. and Chambon,P. (1984) *Nucleic Acids Res.*, **12**, 915-932.

Benoist,C. and Chambon,P. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 3865-3869.

Benoist,C. and Chambon,P. (1981) *Nature*, **290**, 304-310.

Brady,J., Radonovich,M., Vodkin,M., Natarajan,V., Thoren,M., Das,G., Janik,J. and Salzman,N.P. (1982) *Cell*, **31**, 625-633.

Brady,J., Radonovich,M., Thoren,M., Das,G. and Salzman,N.P. (1984) *Mol. Cell. Biol.*, **4**, 133-141,

Contreras,R. and Fiers,W. (1981) *Nucleic Acids Res.*, **9**, 215-236.

Dynan,W.S. and Tjian,R. (1983a) *Cell*, **32**, 669-680.

Dynan,W.S. and Tjian,R. (1983b) *Cell*, **35**, 79-87.

Everett,R.D. (1984) *Nucleic Acids Res.*, **12**, 3037-3056.

Everett,R.D., Baty,D. and Chambon,P. (1983) *Nucleic Acids Res.*, **11**, 2447-2464.

Fromm,M. and Berg,P.J. (1982) *Mol. Appl. Genet.*, **1**, 457-481.

Ghosh,P.K. and Lebowitz,P. (1981) *J. Virol.*, **40**, 224-240.
Ghosh,P.K., Reddy,V.B., Swinscoe,J., Lebowitz,P. and Weissman,S.M. (1978) *J. Mol. Biol.*, **126**, 813-846.
Gluzman,Y. (1981) *Cell*, **23**, 175-182.
Grosschedl,R. and Birnstiel,M.L. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 297-301.
Gruss,P., Dhar,R. and Khoury,G. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 943-947.
Haegeman,G. and Fiers,W. (1978) *Nucleic Acids Res.*, **5**, 2359-2371.
Hansen,U. and Sharp,P.A. (1983) *EMBO J.*, **2**, 2293-2303.
Hansen,U., Tenen,D.G., Livingston,D.M. and Sharp,P.A. (1981) *Cell*, **27**, 603-612.
Hartzell,S.W., Byrne,B.J., Subramanian,K.N. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 23-27.
Hen,R., Sassone-Corsi,P., Corden,J., Gaub,M-P. and Chambon,P. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 7132-7136.
Jongstra,J., Reudelhuber,T., Oudet,P., Benoist,C. Chae,C.B., Jeltsch,J.M., Mathis,D.M. and Chambon,P. (1984) *Nature*, **307**, 708-714.
Manley,J.L., Fire,A., Cano,A., Sharp,P.A. and Gefter,M.L. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 3855-3859.
Mathis,D.J. and Chambon,P. (1981) *Nature*, **290**, 310-315.
Mathis,D.J., Elkaim,R., Kédinger,C., Sassone-Corsi,P. and Chambon,P. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 7383-7387.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-580.
McKnight,S.L. and Kingsbury,R.C. (1982) *Science (Wash.)*, **217**, 316-324.
McKnight,S.L., Kingsbury,R.C., Spence,A. and Smith,M. (1984) *Cell*, **37**, 253-262.
Moreau,P., Hen,R., Wasylyk,B., Everett,R., Gaub,M-P. and Chambon,P. (1981) *Nucleic Acids Res.*, **9**, 6047-6068.
Myers,R.M., Rio,D.C., Robbins,A.K. and Tjian,R. (1981) *Cell*, **25**, 373-384.
Pheiffer,B.H. and Zimmerman,S. (1983) *Nucleic Acids Res.*, **11**, 7853-7871.
Sassone-Corsi,P., Corden,J., Kédinger,C. and Chambon,P. (1981) *Nucleic Acids Res.*, **9**, 3941-3958.
Sassone-Corsi,P., Dougherty,J.P., Wasylyk,B. and Chambon,P. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 308-312.
Tooze,J. ed. (1982) *DNA Tumor Viruses*, published by Cold Spring Harbor Laboratory Press, NY.
Tsuda,M. and Suzuki,Y. (1981) *Cell*, **27**, 175-182.
Wasylyk,B., Kédinger,C., Corden,J., Brison,O. and Chambon,P. (1980) *Nature*, **285**, 388-390.
Wasylyk,B., Wasylyk,C., Matthes,H., Wintzerith,M. and Chambon,P. (1983a) *EMBO J.*, **2**, 1605-1611.
Wasylyk,B., Wasylyk,C., Augereau,P. and Chambon,P. (1983b) *Cell*, **32**, 503-514.
Wasylyk,B., Wasylyk,C. and Chambon,P. (1984) *Nucleic Acids Res.*, **12**, 5589-5608.
Weil,P.A., Luse,D.S., Segall,J. and Roeder,R.G. (1979) *Cell*, **18**, 469-484.

Received on 12 June 1984; revised on 11 July 1984