Volume 12 Number 2 1984 Nucleic Acids Research

Mutational dissection of the 21 bp repeat region of the SV40 early promoter reveals that it contains overlapping elements of the early-early and late-early promoters

D.Baty+, H.A.Barrera-Saldana, R.D.Everett++, M.Vigneron and P.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

Received 4 November 1983; Accepted 28 November 1983

ABSTRACT

Using quantitative S1 nuclease analysis and recombinants which contain the SV40 early promoter region linked to the rabbit β-globin gene coding sequence, we have studied the effect of deletion, inversion and point mutations, located within the 21 bp repeat region, on initiation of transcription from the early-early and late-early startsites in the absence of T-antigen. Our data establish unequivocally that the six GC-rich repeats present in the 21 bp repeat region are essential elements of both the early-early and late-early promoters and that they are not redundant, since mutations, which affect only the GC-rich repeat most proximal to the TATA box, decrease drastically the activity of the early-early promoter, but increase that of the late-early promoter. On the other hand, the four GC-rich repeats most proximal to the 72 bp repeat are common elements of the two overlapping early-early and late-early promoters. Our results, which confirm that the early-early promoter is stronger than the late-early one, also support our previous suggestion that they are in competition for the transcriptional machinery. The general organization of the SV40 early promoter region is discussed.

INTRODUCTION

The organization of the promoter region of the early transcription unit of the Simian Virus 40 (SV40), which encodes the large and small T-antigen and is transcribed immediately after infection, has been extensively studied (1-8 and refs. therein), since it initially appeared to represent a relatively simple experimental model to identify the basic elements involved in the control of transcription by the RNA polymerase class B (II) machinery. In fact, the arrangement of the SV40 early promoter region has been discovered to be more complicated than anticipated. Not only it is partially overlapping with the sequences required for viral replication (see Fig. 2), but two promoters have been identified. The early-early promoter (EEP) controls initiation of transcription from the early-early startsites (EES), which are used chiefly early in viral infection, whereas the late-early promoter (LEP) directs transcription from the further upstream late-early startsites (LES), which are used late in infection when transcription from

© IRL Press Limited, Oxford, England.

the EES is repressed by T-antigen (see Fig. 2) (4, 5, 8-10). It has also been shown that the 72 bp repeat enhancer plays a crucial role in potentiating RNA transcription from both the EEP and LEP (10). The TATA box-like element, 5'-TATTTAT-3', is essential for accurate and efficient initiation of transcription from the EES (10 and refs. therein). In addition, the 21 bp repeat region located further upstream from this sequence (see Fig. 2) plays a major role in determining the efficiency of the whole early promoter region (2,7, 11-16). In particular, we have previously reported that the six GC-rich motifs (Fig. 2) which are repeated in the 21 bp repeat region are important elements of the SV40 early promoter (15). However, in this latter study, the effect of deletions and point-mutations in the 21 bp repeat region on early promoter function was analyzed by measuring T-antigen production, which did not allow us to discriminate between the effects of the mutations on initiation of transcription from the EES and LES. In addition, because T-antigen represses early transcription (9, 11, 17 and refs. therein), the observed effects reflected only indirectly the actual promoter activity of the mutants.

We report here the results of a quantitative analysis of the role of the 21 bp repeat region on initiation of transcription from the EES and LES. Using some of the deletion, inversion and point mutants most of which have been previously described (15), we demonstrate that the 21 bp repeat region contains overlapping promoter sequences which are responsible for efficient transcription from both the EES and LES. However, although all four GC-rich repeats III to VI (see Fig. 2) appear to belong to both the early-early and late-early promoters, the other two repeats I and II are important for transcription from the EES only. Thus, the 21 bp repeat region contains partially overlapping promoter elements of both the EEP and LEP, and the GC-rich repeats are not redundant in their function in agreement with our previous conclusion (15). In addition, we show that the activity of the LEP is increased when that of the EEP is impaired, supporting our previous conclusion that the weaker LEP is in competition for the transcription machinery with the stronger EEP (10).

MATERIALS AND METHODS

1) Bacteria and plasmids.

The bacterial host strains and plasmids pSV1, pRE3, pRE4, pRE7, pRE254, pMKD245, pMD102, the pSVA and pSVB series have been described previously (15). The plasmids pβ(244+)β and pSXβ+, both containing the

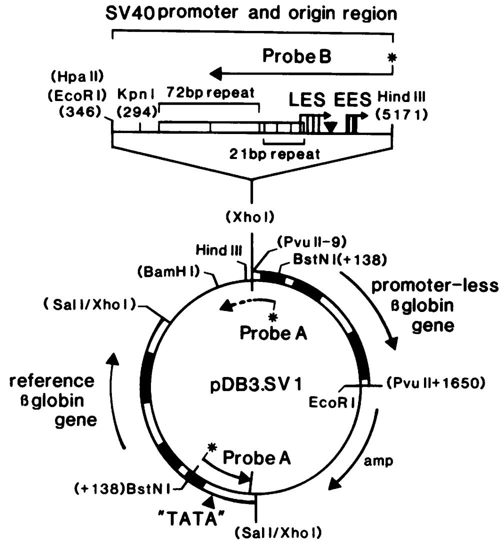


Fig. 1 : The pDB3.SV1 cloning and expression vector. pDB3.SV1 (not at scale) was constructed as described in Materials and Methods. Restriction sites in parentheses were destroyed during the construction of the recombinant. Probe A is the [^{32}P] 5'-end labelled (star) coding strand of the rabbit β-globin gene BstNI fragment (positions +138 to -84). Probe B is the [^{32}P] 5'-end labelled (star) SV40 early coding strand of the small Hind III fragment of pDB3.SV1. LES and EES are the late-early and early-early startsites of the SV40 early promoter region (see text and Ref.10). The position of the SV40 early promoter and rabbit β-globin TATA boxes is indicated by a filled triangle. The SV40 coordinates of the HindIII-HpaII fragment (BBB numbering system, see Ref. 9) are indicated in parentheses.

rabbit β-globin gene, were gifts from W. Schaffner (18, 19).

2) Construction of recombinant plasmids.

Whenever cloning involved non-complementary sticky ends, a treatment with DNA polymerase I (in the presence of deoxyribonucleoside triphosphates) was carried out before ligation to make blunt ends. pDB3 (Fig. 1) was constructed as follows. pRE3 (15) was linearized with EcoRI, repaired with DNA polymerase I, treated with bacterial alkaline phosphatase and ligated to the PvuII fragment of pβ(244+)β which contains the promoter-less rabbit β-globin gene from position -9 to + 1650 to yield pDB1. A XhoI linker was inserted into pDB1 at the EcoRI site located upstream from the β-globin gene (pDB2). pDB3 was derived from pDB2 by inserting at the SalI site the XhoI fragment of pSXβ+, which contains the 2.1 kb BglII fragment of the β-globin gene bearing an intact promoter (see Ref. 18). The EcoRI-HindIII fragments of the SV40

recombinants pSV1, pRE4, pRE7, pRE254, pMKD245, pMD102, pSVA55, pSVA62, pSVA81, pSVA61, pSVA19, pSVA18, pSVB11 and pSVB24, were inserted at the unique XhoI site of pDB3. The resulting pDB3 recombinants were called SV1 (Fig. 1), RE4, RE7, RE254, MKD245, MD102, SVA55, SVA62, SVA81, SVA61, SVA19, SVA18, SVB11 and SVB24 (see Fig. 2). pHB5 and pHB6 (Fig. 2) were obtained by exchanging the NcoI fragments between pSV1 and pRE4, thus separating the two point mutations of pRE4 (see Ref. 15 and Fig. 2). pDHB3.SV1 is a modified version of pDB3.SV1 lacking the HindIII site of pBR322. HB5, HB6 and SAL (Fig. 2) were obtained by cloning the KpnI-HindIII fragments of pHB5, pHB6 and M13mp9TBOSalI (a gift of T. Grundstrom and M. Zenke) in place of that of pDHB3.SV1. All of the pDB3 mutants were sequenced throughout the 21 bp region using the Maxam and Gilbert technique (20).

3) DNA transfection, RNA isolation and S1 nuclease quantitative analysis.

Hela cells at 40% confluency were transfected using the calcium phosphate co-precipitation technique with 10 μg SV40 recombinant per 9 cm plastic petri dishes, as previously described (10, 21). The medium was changed after 24 hrs, cytoplasmic RNA was isolated 24 hrs later and analyzed by quantitative S1 nuclease mapping using 20 μg of cytoplasmic RNA and the [^{32}P] 5'-end labelled single-stranded probes A and B (Fig. 1) [150 units of S1 nuclease (BRL units), 3 hrs at 25°C in 30 mM sodium acetate pH 4.5, 3 mM $ZnSO_4$ and 400 mM NaCl]. Labelling, strand separation and sequencing of the probes were standard techniques as described by Maxam and Gilbert (20).

4) Enzymes.

Restriction and other enzymes were obtained from commercial suppliers and used according to their instructions.

RESULTS

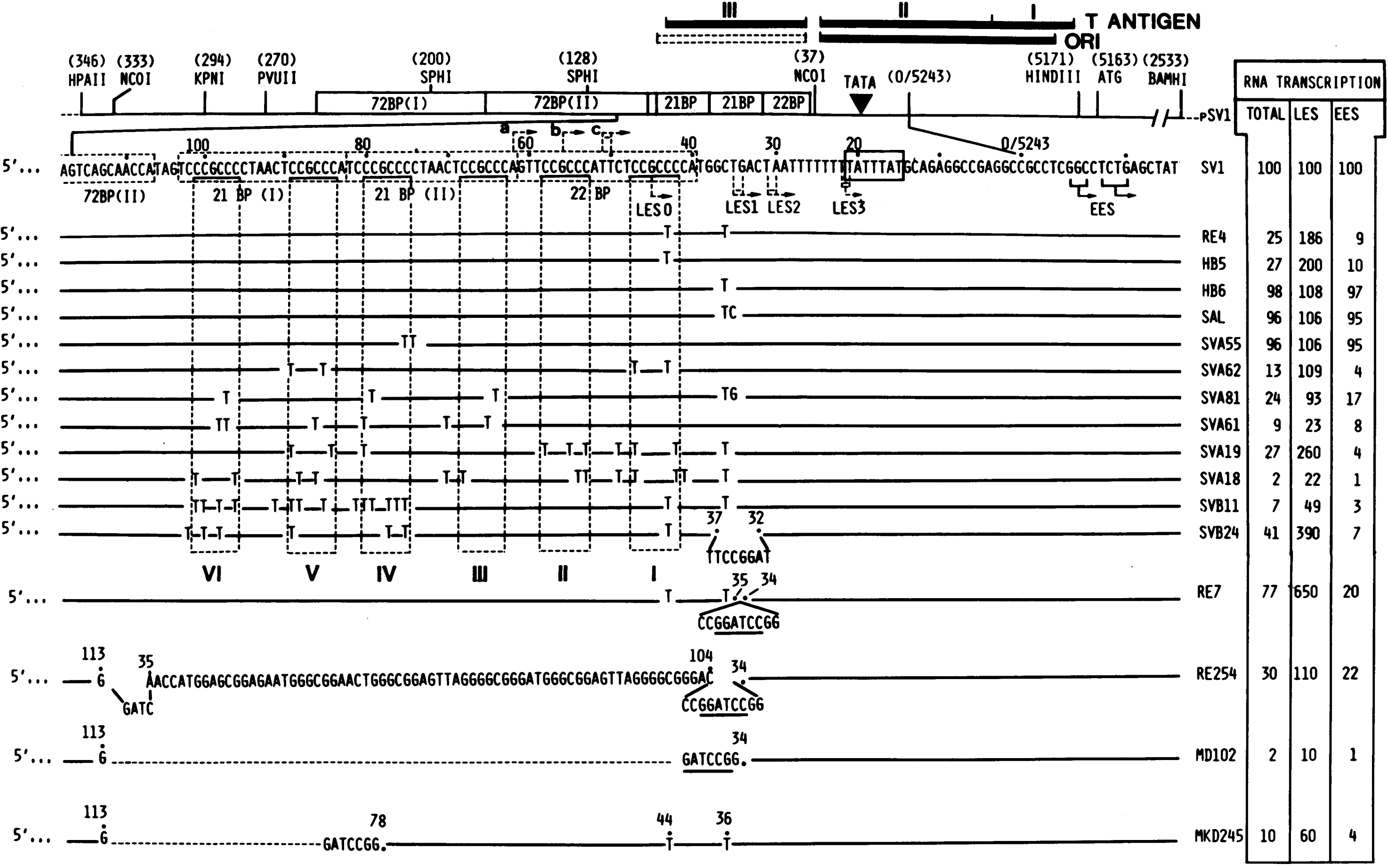
1. Construction of recombinants containing mutations in the 21 bp repeat region of the SV40 early promoter using the pDB3 expression system.

As discussed previously (15) and in the Introduction, the effect of mutations altering the activity of the SV40 early promoter cannot be accurately measured *in vivo* using recombinants producing T-antigen, since the latter represses RNA synthesis from the early transcription unit by binding to sites spanning the promoter region (see Fig. 2). In addition, the same region contains the origin of replication to which the T-antigen binds to trigger replication (see 9 for refs.), thus affecting template abundance when the recombinants are transfected into permissive or semi-permissive cells. To avoid these problems we constructed the cloning and expression vector

pDB3.SV1, in which the coding sequence of the SV40 early transcription unit contained in pSV1 (Fig. 2) have been replaced by the coding sequences of the rabbit β-globin gene, the transcription of which is now under the control of the SV40 early promoter region (Fig. 1 and Materials and Methods). In addition, a complete rabbit β-globin gene, including its promoter region, was inserted in the same recombinant to provide an internal "invariable" gene, the transcription of which can be used as a reference to correct for variations in transfection efficiences in different experiments. A series of mutants was derived from pDB3.SV1 (abbreviated as SV1 in this paper), by replacing the "wild-type" early promoter sequences with the corresponding mutated sequence of mutants altered in the 21 bp repeat region [Fig. 2 and Materials and Methods; the SVA and SVB series, RE4, RE7, RE254, MD102 and MKD245 are derived from some of the previously constructed mutants of the 21 bp repeat region (15), whereas HB5, HB6 and SAL are new mutants].

After transfection into HeLa cells, the amount of RNA initiated from the early promoter region of the "wild-type" recombinant SV1 and the various mutants shown in Fig. 2 was determined by quantitative S1 nuclease analysis. Probe A (Fig. 1 and Material and Methods), was used to measure the amount of RNA initiated for the reference globin gene. It gave a protected fragment of 138 bases after hybridization with RNA initiated from the globin capsite (GLOB band in Figs. 3-5). The same probe A was also used to determine the total amount of RNA initiated from the SV40 early promoter region, since it hybridizes to this RNA up to the end-point of homology which corresponds to the PvuII site located 9 bp upstream from the globin capsite. The corresponding S1 nuclease resistant hybrid yields a fragment (band EP, in Figs. 3-5) of 147 bases which is clearly resolved on shorter exposures from that corresponding to the reference gene. It is important to stress that no RNA was found which could have been initiated from the capsite of the promoter-less β-globin gene located downstream from the SV40 early promoter in the recombinant pDB2.SV1 analogous to pDB3.SV1, but lacking the internal reference globin gene (not shown, see Materials and Methods). This result, which is in agreement with a similar observation made independently by Wasylyk et al. (10) firmly establishes that probe A can be validly used to determine stimultaneously the amount of RNA initiated from the internal reference globin gene and from the SV40 early promoter region.

Probe B was used to accurately map and quantitate RNA initiated from the two early-early (EES) and late-early (LES) sets of startsites of the SV40 early promoter (see Figs. 1, 2 and 3a, lane 2). As previously described by



	TOTAL	LES	EES
SV1	100	100	100
RE4	25	186	9
HB5	27	200	10
HB6	98	108	97
SAL	96	106	95
SVA55	96	106	95
SVA62	13	109	4
SVA81	24	93	17
SVA61	9	23	8
SVA19	27	260	4
SVA18	2	22	1
SVB11	7	49	3
SVB24	41	390	7
RE7	77	650	20
RE254	30	110	22
MD102	2	10	1
MKD245	10	60	4

Wasylyk et al. (10), the LES can be resolved in three subsets, LES1, LES2 and LES3. However, as shown in Fig. 3a (lane 2), the bands corresponding to LES2 and LES3 are not resolved under the present stringent S1 nuclease conditions [see Wasylyk et al. (10), for a discussion of this point]. Early-early RNA initated downstream from the EES (bands DSB, see Ref.10) was not quantitated in the present study. In some of the autoradiograms an additional band, LES0 corresponding most likely to an additional late-early startsite was visible (Fig. 3a, lane 2, and sequence line in Fig. 2). Since the use of probe A generates additional bands in the region which corresponds to LES0, LES1 and LES2 [compare in Fig. 3a lane 1 (Probes A+B) with lane 2 (probe B, alone)], the relative amount of RNA initiated from the LES and EES was determined by scanning autoradiograms corresponding to S1 nuclease analysis with probe B alone.

For any given recombinant, in a given set of transfections, the total amount of RNA initiated from the SV40 early promoter was determined by scanning the EP band of autoradiograms similar to those shown in Figs. 3-5, but exposed for various lengths of time in order to prevent saturation of the film. The figures thus obtained were then corrected for transcription from the internal reference globin gene as estimated from the intensity of the GLOB band, and expressed relative to total transcription from the "wild-type" recombinant SV1 (taken as 100%), which was included in all sets of transfections (Fig. 2, column TOTAL of the RNA TRANSCRIPTION insert). Autoradiograms corresponding to S1 nuclease analysis with probe B alone were similarly

Fig. 2 : General organization, sequence and relative amount of RNA transcribed from the early promoter of pDB3.SV1 (called SV1) and of its mutant derivatives in the 21 bp region. The first line represents the wild-type organization of the SV40 early region in the recombinant pSV1 (1). The location of the three T-antigen binding sites (I, II and III) and of the origin of replication (ORI) is indicated (for refs., see 9, 10). SV40 coordinates (BBB system, see ref. 9) are given in parentheses (see also ref. 10). The second line corresponds to the SV1 wild-type sequence (non-coding strand) between coordinates 5224 and 118. The position of the early-early startsites (EES, arrows) and late-early startsites (dashed arrows, LES0, LES1, LES2 and LES3 are shown, see ref. 10 and text), together with the position of some additional minor startsites (dashed arrows, a, b and c; see Fig. 4 and text). The six GC-rich repeats 5'-CCGCCC-3' are underlined and their position is indicated by the vertical dashed boxes I to VI. The BamHI site present in the insertion mutants is underlined. The sequences characteristic of the various mutants used in this study are indicated below the wild type sequence of SV1 [for some of these mutants, some minor differences were found upon resequencing when compared to the previously published sequence (15)]. The relative amounts of RNA initiated from the early promoter region (TOTAL), the late-early startsites (LES) and the early-early startsites (EES) were determined as described in the Result section, taking SV1 as 100% (see text).

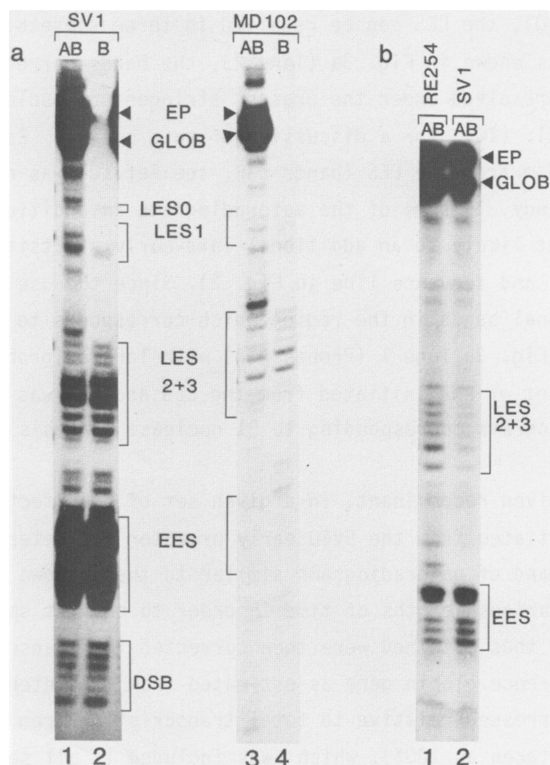


Fig. 3 : Effect of deletion or inversion of the 21 bp repeat region on transcription from the SV40 early-early and late-early promoters. HeLa cells were transfected and total cytoplasmic RNA was analyzed by quantitative S1 nuclease mapping using either a mixture of probes A and B (lanes AB) or probe B alone (lanes B). EP is the S1 nuclease-resistant band which corresponds to the end-point of homology between probe A and the SV40-globin hybrid transcripts (Fig. 1). GLOB is the S1 nuclease resistant band expected for RNA initiated at the startsite of the β-globin reference gene using probe A (Fig. 1). The S1 nuclease-resistant bands LES0, LES1, LES2+3, EES and DSB, which were obtained with probe B (Fig. 1) correspond to RNA initiated from the late-early startsites (LES0, LES1, LES2 and LES3), the early-early startsites (EES) and startsites located downstream from the EES (DSB) (see text, Fig. 2 and ref. 10). The transfected recombinants were as follows : a) lanes 1 and 2, pDB3.SV1 (SV1); lanes 3 and 4, MD102; b) lane 1, RE254; lane 2, SV1. (a) and (b) correspond to separate experiments.

scanned to determine the relative amount (in percent) of RNA initiated from the two sets of LES and EES. These values and the corresponding "corrected" total amount of RNA (see above) were then used to calculate the amount of RNA initiated from the LES and EES of any given recombinant. Again, the figures thus obtained were expressed relative to transcription from the LES and EES

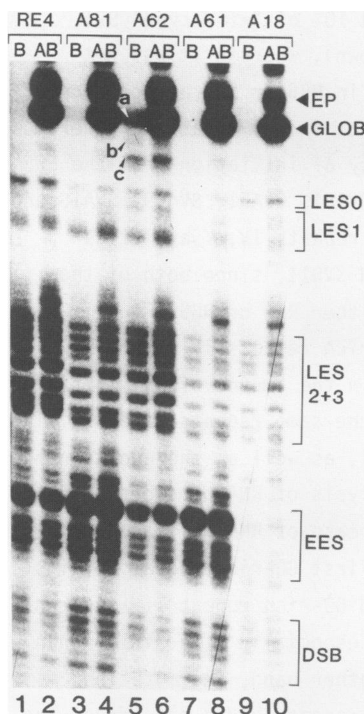


Fig. 4 : Effect of point mutations within the 21 bp repeat region on transcription from the SV40 early-early and late-early promoter. HeLa cells were transfected with the pDB3 mutants RE4, SVA81 (A81), SVA62, (A62), SVA61 (A61) and SVA18 (A18) (Fig. 2) and RNA was analyzed by quantitative S1 nuclease mapping using either probes A and B (AB) or probe B alone (B), as indicated. The arrowheads a, b and c in lane 5 point to S1 nuclease-resistant bands which correspond to minor LES particularly visible in mutant SVA62. All other symbols are as described in legend to Fig. 3.

of the "wild-type" recombinant SV1 (taken as 100%) (Fig. 2, columns LES and EES of the RNA TRANSCRIPTION insert). In agreement with the previous report of Wasylyk et al. (10), we found that the total amount of RNA initiated from the LES and EES represent approximately 10 and 90%, respectively, of the RNA initiated from the whole SV40 promoter early region of the "wild-type" chimeric SV40-globin recombinant SV1. For each recombinant shown in Fig. 2 the values given in the RNA TRANSCRIPTION insert represent the average of at least four independent transfection experiments using different plasmid preparations.

2. Effect of mutations within the GC-rich repeats on RNA initiated from the early-early startsites (EES).

The results obtained by deleting the whole 21 bp region from position 35 to 112 (mutant MD102, Fig. 2 and Fig. 3a) clearly demonstrate that this region plays a key role in initiation of transcription from the EES. The importance of the first GC-rich repeat I contained in the 22 bp segment is revealed by a comparison of recombinants RE4, HB5, HB6 and SAL. Initiation of transcription from the EES was drastically decreased in RE4 and HB5 which

contain a C→T transition at position 43 (9% and 10% of "wild-type" SV1, respectively; Fig. 2, Fig. 4 and results not shown), whereas it was not significantly affected by the mutations present in HB6 or SAL at position 36 or 35 and 36, respectively (Fig. 2 and results not shown). That the other GC-rich repeats do also contribute to the efficiency of initiation from the EES is deduced from a comparison of mutants SVA62, SVA81, SVA61, SVA19, SVA18 and SVB11. The possible contribution of the GC-rich repeats IV, V and VI is suggested by the results obtained with SVA62 and SVB11, since both of them produced less RNA initiated from the EES than either RE4 or HB5 (Figs. 2, 4 and 5a). It is noteworthy that deleting the GC-rich repeats IV, V and VI (mutant MKD245) resulted in a similar decrease in the amount of RNA initiated from the EES as the point mutations present in the same repeats of SVB11. The contribution of the GC-rich repeats IV, V and VI, as well as the additional contribution of repeat III, is supported by analysis of RNA produced by SVA81 and SVA61, both of which exhibited a strong decrease of RNA initiated from the EES in the absence of any mutations in the first GC-rich repeat I (Figs. 2 and 4). The possible involvement of the second GC-rich repeat II is strongly suggested by results obtained with SVA19 and especially with SVA18, when compared with SVB11 (Figs. 2, 4 and 5). On the other hand, the mutation present at position -75 of SVA55 in the GC-rich repeat IV did not affect the amount of RNA initiated from the EES, which indicates that the 3' most distal C of this 5'-CCGCCC-3' repeat may not be important for the function of the GC-rich repeats.

Although not quantitated in the present study, it is clear from the results shown in Figs. 3-5 that the amount of RNA initiated from the early-early startsites which corresponds to the DSB bands was decreased in the above mutants in parallel to that initiated from the EES.

3. Effect of mutations within the GC-rich repeats on RNA initiated from the late-early startsites (LES).

The importance of the 21 bp repeat region for efficient initiation, not only from the EES, but also from the LES was demonstrated by the results obtained with the deletion mutant MD102 (Figs. 2 and 3). However, the same point mutation located within the first GC-rich repeat I, which strikingly decreased RNA initiated from the EES, resulted in a two-fold increase of RNA initiated from the LES (mutants RE4 and HB5 in Figs. 2 and 4 and results not shown). That the integrity of the second GC-rich repeat (II) of the 22 bp segment is also not important for efficient initiation from the LES was indicated by the results obtained with SVA19 (2.6 fold increase in RNA ini-

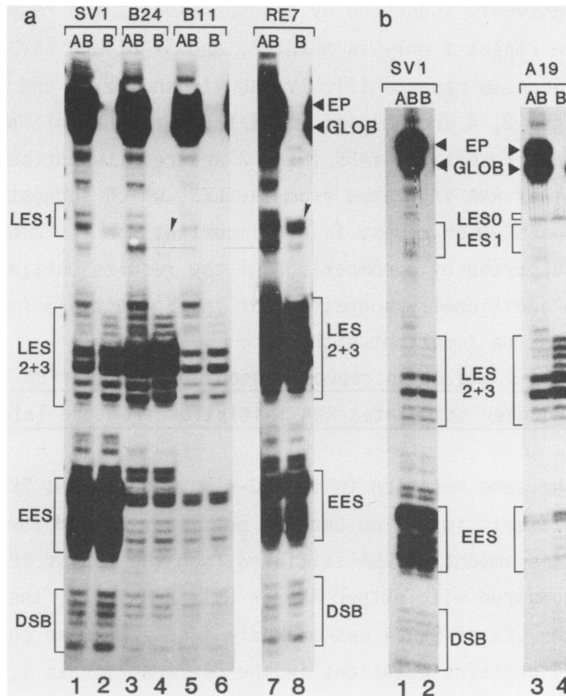


Fig. 5 : Effect of point mutations and insertion within the 21 bp repeat region on transcription from the SV40 early-early and late-early promoters. HeLa cells were transfected with pDB3.SV1 (SV1) and its mutants SVB24 (B24), SVB11 (B11), RE7 (panel a) and SVA19 (A19) (panel b), and RNA was analyzed by quantitative S1-nuclease mapping using either probes A and B (AB) or probe B alone (B), as indicated. (a) and (b) correspond to two separate experiments. Symbols are as described in legend to Fig. 3. The arrowheads in lanes 4 and 8 indicate the position of the S1 nuclease resistant band which corresponds to the end-point of homology of probe B and RNA initiated upstream from the insertion in SVB24 and RE7.

tiated from LES, Figs. 2 and 5b). Clearly, this increase affects all of the LES, since LES0 which was always a minor site in SV1 was easily detected in RE4 and SVA19 (Figs. 3, 4 and 5b). On the other hand, mutations within the GC-rich repeat V affect initiation from the LES, since there was less LES-initiated RNA in SVA62 than in either RE4 or HB5 (Figs. 2 and 4 and results not shown). However, it is interesting to note that there was, in SVA62, an increase in RNA initiated from the normally very minor sites a, b and c. The possible involvement of the GC-rich repeats III, IV, V and VI in efficient initiation from the LES is indicated by the results obtained with SVA61, SVA18 and SVB11 (Figs. 2, 4 and 5A). That the effect of mutations in these

repeats is cumulative is suggested by a comparison of the results obtained with SVA62, where repeat V only is mutated, and SVA61 and SVB11 which bear additional mutations in repeats III, IV and VI, and IV, V and VI, respectively (Figs. 2, 4 and 5A). As it is the case for initiation from the EES, the mutations present in SVA55 (Fig. 2 and results not shown) did not affect the amount of RNA initiated from the LES, which suggests that the 3'-most distal C of a GC-rich repeat is not important for its function. This possibility is supported by a comparison of the results obtained with SVA62 and SVA19, which additionally suggests that the 5'-most proximal C of a GC-rich repeat may not be important either.

4. Insertions between the 21 bp repeat region and the AT-rich region of the early-early promoter stimulates RNA initiation from the late-early startsites.

RE7 has the same mutation in the GC-rich repeat I as RE4, but contains in addition a decameric insertion between positions 34 and 35. As shown in Figs. 2 and 5a, the amount of RNA initiated from the LES of RE7 was markedly increased when compared with either RE4 or SV1. Similarly, the amount of RNA initiated from the LES of SVB24 was markedly increased when compared to SVB11 which, like SVB24 contains mutations in the GC-rich repeats I, IV, and VI, but not the octameric insertion present in SVB24 between positions 32 and 37. Since the precise location of the inserted fragment and its sequence are different in SVB24 and RE7, it is likely that the important parameters responsible for this increase in LES-initiated RNA is not the sequence itself and its exact location, but rather the spacing which is introduced between the 21 bp repeat region and the AT-rich sequence of the early-early promoter. In addition, it is clear from the comparison of the amounts of RNA initiated from the LES of RE7, RE254 and MD102 that the 21 bp repeat region remains indispensable for efficient increased initiation from these late-early sites. In this respect, it is interesting to note that there is much more RNA initiated from the LES of RE254, in which the 21 bp repeat region is present in the inverted orientation, than from those of MD102, in which this region is deleted. This latter result is in agreement with our previous report that the effect of the 21 bp repeat region is to some extent bidirectional (15).

DISCUSSION

Although previous studies using deletion mutants have indicated that the 21 bp repeat region is an essential element of the SV40 early promoter region (2, 7, 11-16), the evaluation of the contribution of the various sequ-

ences contained in this region has led to controversial conclusions. For instance, surmising that the GC-rich repeats could be the important sequences, Fromm and Berg (16) concluded from a study of a number of deletion mutants that at least two of these repeats are needed for maximal promoter function, but that any two of the six are equally effective. In contrast, Everett et al. (15), using a different set of deletion mutants found that the SV40 early promoter activity was decreased by at least 3-fold when only two of the GC-rich repeats were left. In addition, using point mutants, which allowed them to demonstrate unequivocally the involvement of the GC-rich repeats, Everett et al. (15) suggested that a single C→T transition in any of these repeats is sufficient to impair the promoter function of the 21 bp repeat region. Several possibilities could be invoked to explain the origin of these apparently conflicting data. First, the comparison of results obtained with deletion mutants may pose problems, because the new sequence which is substituting for the deleted one is not necesssarily neutral and may affect in one way or the other the activity of the promoter sequences which remain. Second, Everett et al. (15) used an immunofluorescence assay of T-antigen production to determine the promoter activity, whereas Fromm and Berg (16) measured the overall amount of RNA initiated from recombinants containing the SV40 early promoter region joined to the rabbit β-globin coding sequence in order to avoid problems linked to autoregulation of T-antigen synthesis (see Introduction). Third, since the two groups did not use a cotransfected reference gene to standardize their experiments, their results should be considered as semi-quantitative because of the variability in transfection efficiencies from one experiment to another. We have overcome most of these problems in the present study using a chimeric SV40 promoter-β-globin coding sequence recombinant bearing point mutations in the 21 bp repeat region and containing, in addition, a complete rabbit β-globin transcription unit as an internal reference gene. Moreover, the use of appropriate probes has allowed us to determine not only the effect of these point mutations on the overall activity of the SV40 early promoter region, but also to discriminate between their effect on the two overlapping EE and LE promoters which constitute this region. As best shown by the results obtained with mutant RE7, the characterization of the elements which constitute the SV40 early promoter region imperatively requires an independent determination of the amounts of RNA initiated from the EES and LES.

Our present data establish unequivocally that the CG-rich repeats are esential elements of both the EE and LE promoters and that they are not

interchangeable. Table 1 summarizes the effect of point mutations within the GC-rich repeats, assuming that the first and the last C of the repeated sequence 5'-CCGCCC-3' are not as important for promoter activity as the other bases of the repeat, as suggested by the results obtained with mutants SVA55, SVA62, SVA19 (see Result section; note however that the conclusions drawn below are not markedly modified if all of the bases of the GC-rich repeats would be involved in promoter function; see Fig. 2 and Table 1). Quite clearly, the first GC-rich repeat I present in the slightly divergent 22 bp repeat, plays a crucial role in efficient transcription from the EES, but not from the LES (RE4 and HB5 mutants). The results obtained with SVA19, which strongly suggest that the second GC-rich repeat II is similarly important for initiation from the EES, demonstrates that this repeat is not involved in initiation from the LES. That repeat V is important for synthesis of RNA initiated from both EES and LES is borne out from the comparison of the results obtained with mutants RE4, HB5, SVA62 and SVA18. The results obtained using SVA81, SVA61 and SVB11 support the conclusion that either one of repeats III, IV and VI or all of them are important for efficient initiation from both EES and LES. Therefore, it appears that the GC-rich repeats contained in both the 22 bp sequence and the two 21 bp perfect repeats are crucial for efficient transcription from the EES, whereas only those contained in the 21 bp perfect repeats are important for initiation from the LES. Independent mutagenesis of each of the four GC-rich sequences of the 21 bp repeats is in progress to determine whether each of these four repeats are equally involved in RNA initiation from the EES and LES, and whether altering the bases similarly located within each of the six GC-rich repeats has the same relative effect on the promoter function of these repeats. In view of the above discussion concerning the possible difficulty in evaluating the results obtained with deletion mutants, it is interesting to note that the amounts of RNA initiated from the EES and LES were similarly decreased in SVB11 and MKD245, which both contain intact GC-rich repeats II and III, a mutated repeat I, and either a deletion of repeats IV, V and VI (MKD245) or point mutations within these GC-rich repeats.

The role of the 21 bp repeat region in initiation of transcription from the SV40 early promoter has been recently investigated in vitro. Dynan and Tjian (22) have found that activation of in vitro transcription from the EES by the promoter-specific factor Sp1 (23) is critically dependent on the presence of sequences located within the tandem 21 bp perfect repeats. Moreover, these authors have shown that there is a specific binding site for

Table 1 : Effect of mutations within the 5'-CGCC-3' core sequence of the GC-rich repeats on the amount of RNA initiated from the whole SV40 early promoter region (Total) and its early-early (EES) and late-early (LES) start-sites. For each recombinant, + indicates the presence of at least one C→T transition with the core sequence 5'-CGCC-3' of a given GC-rich repeat (see Fig. 2 and text). The RNA transcription results are taken from Fig. 2.

	Mutations in GC-rich repeats						RNA transcription		
RECOMBINANT	VI	V	IV	III	II	I	LES	EES	TOTAL
SV1							100	100	100
SVA55							106	95	96
RE4						+	186	9	25
HB5						+	200	10	27
SVA19					+		260	4	27
SVA62		+				+	109	4	13
SVA81	+		+	+			93	17	24
SVA61	+	+		+			23	8	9
SVA18		+			+		22	1	2
SVB11	+	+	+			+	49	3	7

factor Spl in these 21 bp repeats that can be visualized in DNaseI footprinting experiments. It is very interesting that the DNaseI protected region extends from approximately positions 55 to 97, and therefore includes the totality of the GC-rich repeats III, IV and V, a fraction of repeats II and VI, but not at all the GC-rich repeat I. In vitro transcription studies performed in our laboratory using the deletion and point mutations described here and previously (15) have shown that inititiation of transcription from the EES and LES is strikingly decreased by mutations which affect the GC-rich repeats III, IV, V and VI, but not by mutations which are located within the GC-rich repeat I (RE4, HB5) (M. Vigneron et al., in preparation). Thus, the presently available in vitro transcription systems do not faithfully reproduce the almost absolute in vivo requirement of the GC-rich repeat I for efficient initiation of transcription from the EES. It appears therefore that the 21 bp repeat region may have a double role in initiation of transcription from the EES and LES. One functional unit would be constituted of the four GC-rich repeats contained in the tandem 21 bp perfect repeats and their effect on initiation from both the EES and LES could be mediated through the binding of the Spl factor of Dynan and Tjian (22, 23). The other functional unit would be constituted of most, if not all, of the 22 bp imperfect repeat which is important for efficient initiation of transcription from the EES only. Its effect, which cannot be mimicked in vitro at the present time,

would not seem to be directly related to the binding of factor Sp1. Whether binding of this factor to the 21 bp perfect repeats would nevertheless be a prerequisite for the activation of transcription from the EES, which does require the GC-rich repeats I and II, is unknown. In this respect, it may be significant that the *in vitro* binding of Sp1 to the 21 bp perfect repeat is accompanied by the appearence of two sites of enhanced DNase I cleavage, which map in each of the two GC-rich repeats I and II (22).

Quite clearly the complexity of the organization of the two overlapping promoters which constitute the SV40 early promoter region is increasing as the dissection of their constitutive elements is progressing (the term promoter element is used here to designate any sequence element which is required for accurate and efficient initiation of transcription). Our previous report (10) and the present results indicate that the SV40 promoter region contains the following elements : 1) the TATA box which is required for accurate and efficient initiation from the EES only, both *in vivo* and *in vitro*; 2) the GC-rich repeat I which, most likely together with the repeat II, is very important for efficient *in vivo* initiation of transcription from the EES, but not from the LES (these effects have not yet been reproduced *in vitro*); 3) the four GC-rich repeats contained in the 21 bp perfect repeats which are very important for efficient *in vivo* and *in vitro* initiation from both the EES and LES. Therefore, three elements, the TATA box, the 22 bp imperfect repeat and the 21 bp perfect repeats appear to constitute the EE promoter, whereas the crucial sequences constituting the LE promoter seem to be all contained within the 21 bp perfect repeats. Whether these two repeats contain, in addition to the GC-rich repeats, substitute TATA-like elements directing initiation of transcription from the LES remains to be established. From our previous results (24) and the recent report of Dynan and Tjian (22), it seems very likely that both the TATA box (24) and the 21 bp perfect repeats (22) are binding sites for transcription factors. Whether or not the molecular mechanisms underlying the activation of transcription by the GC-rich repeat contained in the 21 bp perfect repeats are the same for the EE and LE promoters will require a detailed comparison of the effect of individual point mutations in these repeats on initiation from the EES and LES. Further studies are also required to determine whether the two GC-rich repeats of the 22 bp imperfect repeat are indispensable because they also bind an additional, not yet identified, factor, or whether they are important for the generation of an altered DNA structure which may be required for efficient initiation of transcription and may be induced by binding of the

Spl factor to the 21 bp perfect repeats.

In a previous study we have shown that, although both the EE and LE promoters are potentiated by the 72 bp repeat enhancer, about 10 times more RNA is initiated from the EES than from the LES, in the absence of replication and T-antigen production (10). Moreover, we have concluded that the reason for this difference resides in the greater strength of the EE promoter, since altering the TATA box of this promoter resulted in a marked increase of the amount of RNA initiated form the LES (10). Our present data strongly support the conclusion that the two overlapping EE and LE promoters are in competititon for the transcriptional machinery and that the EE promoter is stronger than the LE one. Clearly, point mutations in the EE promoter-specific 22 bp imperfect repeat element, which caused drastic reductions in RNA initiated from the EES, resulted concomittantly in marked increases of RNA initiated from the LES (mutants RE4, HB5, SVA19). In this respect, it is interesting to note that the additional introduction of insertion mutations (RE7 and SVB24) resulted in a marked further increase in the amount of RNA initiated from the LES. Further studies are necessary to determine if this increase reflects a critical spacing requirement between the 21 bp repeat region and the TATA box and capsite elements of the EE promoter or/and the disappearance of steric hindrance problems which would normally prevent simultaneous initiation of transcription from the EES and LES.

ACKNOWLEDGEMENTS

We thank Dr. A. Wildeman for a critical reading of this manuscript. We are grateful to Drs. W. Schaffner, T. Grundstrom and M. Zenke for generous gifts of recombinant plasmids. We thank Mrs. M. Acker and B. Augsburger for excellent technical assistance and C. Werlé, B. Boulay and B. Chambon 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 INSERM (PRC 124026), 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.

* to whom all correspondance should be addressed.
+ Present address : Centre de Biochimie et de Biologie Moléculaire du CNRS, 31, Chemin J. Aiguier, BP 13277, Marseille Cédex 9;
++Present address : Medical Research Council, Virology Unit, Church Street, Glasgow G11 5JR, U.K.

ABBREVIATIONS

EE and LE, early-early and late-early, respectively; EES and LES, EE and

LE startsites, respectively; EEP and LEP, EE and LE promoter, respectively; SV40, Simian virus 40.

REFERENCES

1. Benoist, C. and Chambon, P. (1980) Proc. Natl. Acad. Sci. USA 77, 3865-69.
2. Benoist, C. and Chambon, P. (1981) Nature 290, 304-309.
3. Ghosh, P.K., Lebowitz, P., Frisque, R.J. and Gluzman, Y. (1981) Proc. Natl. Acad. Sci. USA 78, 100-104.
4. Hansen, U., Tenen, D.G., Livingston, D.M. and Sharp, P.A. (1981) Cell 27, 603-612.
5. Ghosh, P.K. and Lebowitz, P. (1981) J. Virol. 40, 224-240.
6. Mathis, D.J. and Chambon, P. (1981) Nature 290, 310-316.
7. Fromm, M. and Berg, P. (1982) J. Mol. Appl. Genet. 1, 457-481.
8. Reddy, V.B., Ghosh, P., Lebowitz, P., Piatak, M. and Weissmann, S.M. (1979) J. Virol. 30, 272-296.
9. Tooze, J. (ed.) (1982) DNA Tumor Viruses, Cold Spring Harbor Laboratory Press, N.Y.
10. Wasylyk, B., Wasylyk, C., Matthes, H., Wintzerith, M. and Chambon, P. 1983) The Embo J. 9, 1605-1611.
11. Myers, R.M., Rio, D.C., Robbins, A.K. and Tjian, R. (1981) Cell 25, 373-384.
12. Lebowitz, P. and Ghosh, P.K. (1982) J. Virol. 41, 449-461.
13. Byrne, B.J., Davis, M.S., Yamaguchi, J., Bergsma, D.J. and Subramanian, K.N. (1983) Proc. Natl. Acad. Sci. USA 80, 721-725.
14. Hartzell, S.W., Yamaguchi, J. and Subramanian, K. (1983) Nucleic Acids Res. 11, 1601-16.
15. Everett, R.D., Baty, D. and Chambon, P. (1983) Nucleic Acids Res. 11, 2447-64.
16. Fromm, M. and Berg, P. (1983) J. Mol. Appl. Genet. 2, 127-135.
17. Rio, D.C. and Tjian, R. (1983) Cell 32, 1227-40.
28. Banerji, J., Rusconi, S. and Schaffner, W. (1981) Cell 27, 299-308.
29. De Villiers, J. and Schaffner, W. (1981) Nucleic Acids Res. 9, 6251-64.
20. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
21. Wasylyk, B., Wasylyk, C., Augerau, P. and Chambon P. (1983) Cell 32, 503-514.
22. Dynan, W.S. and Tjian, R. (1983) Cell, in press.
23. Dynan, W.S. and Tjian, R. (1983) Cell 32, 669-680.
24. Davison, B.L., Egly, J.M., Mulvihill, E.R. and Chambon, P. (1983) Nature, 301, 680-686.