MOLECULAR AND CELLULAR BIOLOGY, Nov. 1986, p. 3667–3676
0270-7306/86/113667-10$02.00/0

Vol. 6, No. 11

Minimal Transcriptional Enhancer of Simian Virus 40 Is a 74-Base-Pair Sequence That Has Interacting Domains

THOMAS A. FIRAK AND KIRANUR N. SUBRAMANIAN*

Department of Microbiology and Immunology, College of Medicine, University of Illinois at Chicago, Chicago, Illinois 60612

Received 14 April 1986/Accepted 21 July 1986

We have assayed the ability of segments of the simian virus 40 (SV40) 72-base-pair (bp) repeat enhancer region to activate gene expression under the control of the SV40 early promoter and to compete for *trans*-acting enhancer-binding factors of limited availability in vivo in monkey CV-1 or human HeLa cells. The bacterial chloramphenicol acetyltransferase and the herpes simplex virus type 1 thymidine kinase genes were used as reporters in these assays. A 94-bp sequence located between SV40 nucleotides 179 and 272, including one copy of the 72-bp repeat, has been termed the minimal enhancer in previous studies. In the present study, we found that the 20-bp origin-proximal region located between nucleotides 179 and 198 was dispensable, since its removal caused only a slight reduction in enhancer activity. However, the deletion of another 4 bp up to nucleotide 202 abolished the enhancer activity. We propose that the minimal enhancer is a 74-bp sequence located between nucleotides 199 and 272, including 52 bp of one copy of the 72-bp repeat and a 22-bp adjacent sequence up to the *Pvu*II site at 272. The nonamer 5′-AAGT/CATGCA-3′, which we term the K core, occurred as a tandem duplication around the *Sph*I site at nucleotide 200, and we found that this duplication was essential for enhancement and factor-binding activities. A heterologous core element (which we term the C core), 5′-GTGGA/TA/TA/TG-3′, identified earlier (G. Khoury and P. Gruss, Cell 33:313–314, 1983; Weiher et al., Science 219:626–631, 1983) also occurred in duplicate, with one of the copies located within the 22-bp sequence near nucleotide 272 present outside the 72-bp repeat. We provide direct evidence that this 22-bp sequence augments enhancer activity considerably. We also found that in addition to the heterologous interaction occurring normally between the K and C cores within the minimal enhancer, certain homologous interactions were also permitted provided there was proper spacing between the elements.

Gene transcription in eucaryotes is increased markedly in the presence of linked, *cis*-acting elements called enhancers (see reference 12 for a review). Enhancers can act in either orientation and over various distances on natural or heterologus promoters. Initially identified in simian virus 40 (SV40) (1, 2, 14, 21), they have been found in numerous virus genomes and near some cellular genes as well (12). Host, tissue, or cell specificity of gene expression is determined by some enhancers (12). Although enhancers work with high efficiency in vivo, they have been found to activate transcription in vitro as well, albeit at a low efficiency (26).

The SV40 enhancer region is located adjacent to the 21-base-pair (bp) repeat promoter element and the core replication origin. In the SV40 numbering system (31), the enhancer spans a 166-bp sequence between its origin-proximal end at nucleotide 107 and its origin-distal end at nucleotide 272 (Fig. 1). This sequence includes the two copies in tandem of the 72-bp repeat and a 22-bp adjacent sequence at the origin-distal end. One copy of the 72-bp repeat can be deleted without affecting the viability of SV40 (14, 28, 32). The remaining 94 bp of the enhancer region, including one copy of the 72-bp repeat and the 22-bp adjacent sequence, located between nucleotides 179 and 272 can provide enhancement outside the context of the viral genome as well and has been taken as the minimal essential region of the enhancer (4, 26, 34, 38).

Considerable research has been carried out with the SV40 enhancer to determine the sequences that are important for its function. Recent evidence indicates the presence of multiple essential domains within the SV40 (16, 26) and polyomavirus (33) enhancers. The system of classification of the SV40 enhancer into 5′ and 3′ domains by Sassone-Corsi et al. (26) will be used in the present study. According to this system, the 5′ domain (called domain B in reference 30) includes the origin-distal portion of the 72-bp repeat and the 22-bp adjacent sequence. The 3′ domain (called domain A in reference 30) occurs towards the origin-proximal portion of the 72-bp repeat around the *Sph*I restriction site (Fig. 1). The 94-bp minimal enhancer region contains one each of the 5′ and 3′ domains. The 166-bp complete enhancer region (Fig. 1) contains two 5′ domains (one of which is truncated) and two 3′ domains, with the 5′ and 3′ domains alternating.

The 5′ domain contains two copies of a so-called core element with the sequence 5′-GTGGA/TA/TA/TG-3′ repeating with a spacing of 15 bp (17, 19, 36). Such a core sequence also occurs in the enhancers of BK, polyoma, and murine leukemia and sarcoma viruses (12) and is believed to be important for enhancer function based on deletion (1, 2, 7, 8, 21) and point mutagenesis (36). A core sequence has not been described for the 3′ domain, but is a possibility. Both the 5′ and 3′ domains also contain short (roughly 8 bp long) stretches of alternating purine and pyrimidine nucleotides (termed Pu/Py elements) that are potentially capable of forming left-handed Z-DNA conformation under certain conditions (23). The Pu/Py elements in the 5′ and 3′ domains have different sequences. The Pu/Py element in the 5′ domain overlaps one of the core sequences, and a similar overlap between core and Pu/Py elements is possible in the 3′ domain as well. Recent work also indicates that the enhancer action is mediated by the binding of *trans*-acting factors of limited availability believed to recognize corelike elements within the enhancer (26, 27).

* Corresponding author.

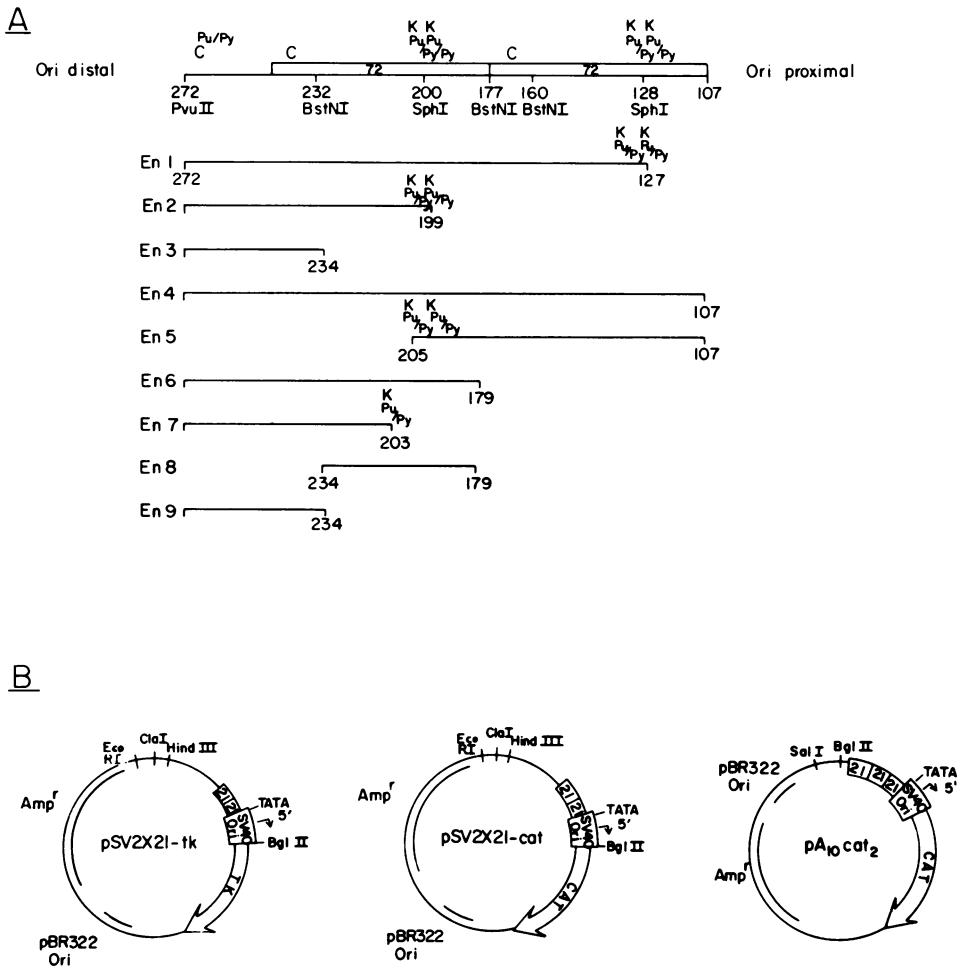

FIG. 1. Construction of CAT and TK gene plasmids containing segments of the SV40 enhancer region. (A) Extent of the enhancer region sequences contained in the segments. The complete SV40 enhancer region, including two copies of the 72-bp tandem repetition, is shown at the top. The nucleotide numbering is according to Tooze (31). The symbol C refers to core element 5'-GTGGA/TA/TA/TG-3' (17, 36); Z refers to a stretch of seven or eight alternating purine and pyrimidine residues (Pu/Py elements) that can assume a left-handed Z-DNA structure under certain conditions (23); K refers to a proposed core element, 5'-AAGT/CATGCA-3', present as a tandem repeat around the two *Sph*I sites overlapping the Z elements. The locations of the repeated C, Z, and K elements are shown at the top. The boundaries of the enhancer segments En1 through En9 are shown at the bottom. The elements K and Z are indicated on some of the segments to show whether they are present in one or two copies each in case these elements happen to be located at the ends of the segments. The elements present internally are as at the top. En3 and En9 refer to the same fragment with *Hin*dIII (En3) or *Bgl*II (En9) ends. The segments were cloned in the vectors shown in panel B. (B) Cloning vectors for the SV40 enhancer segments. The locations of the 21-bp repeats and the TATA elements of the SV40 early promoter, the major early mRNA 5' end (shown by the arrows), and the HSV-1 TK or the bacterial CAT genes that follow are all indicated in the circular plasmids. The enhancer segments were cloned upstream from the early promoter at the unique *Cla*I or *Hin*dIII sites in the pSV2X21-tk or pSV2X21-cat vectors and at the unique *Bgl*II site in $pA_{10}cat_{21}$ as described in the text.

In the present study, using deletion mutagenesis, we have obtained a precise definition of the minimal enhancer sequence required for enhancement function and factor binding. We also show that the loss of enhancer activity resulting from deletion of the 3' or 5' domains can be restored by the duplication of the other domain under certain conditions.

MATERIALS AND METHODS

Plasmid constructions. Plasmid vectors pSV2X21-tk (3), pSV2X21-cat (B. J. Byrne, Ph.D. thesis, University of Illinois at Chicago, 1984), and $pA_{10}cat_2$ (18) were used for cloning the SV40 enhancer region fragments shown in Fig. 1. These vectors contain the herpes simplex virus type 1 (HSV-1) thymidine kinase (TK) gene or the bacterial chloramphenicol acetyltransferase (CAT) gene under the control of the enhancerless SV40 early promoter. Standard cloning techniques (20) were used for plasmid constructions, in accordance with the National Institutes of Health guidelines for recombinant DNA methodology. Enhancer fragments En1 and En2 were cloned at the *Cla*I site and En3 at the *Hin*dIII site 352 and 346 bp, respectively, upstream from the promoter in the pSV2X21-tk and pSV2X21-cat vectors. Fragments En4, En6, En7, En8, and En9 were cloned at the *Bgl*II site immediately upsteam from the promoter in plasmid $pA_{10}cat_2$. Plasmid pEn5-cat was made by cloning the 72-bp permuted fragment released by digestion of the SV40 enhancer region with *Sph*I at the unique *Sph*I site present instead of the *Bgl*II site upstream from the early promoter in a $pA_{10}cat_2$-like vector (18).

Transient expression analysis. The CAT gene plasmid constructs made were studied by transient expression anal-

ysis by transfection of monkey CV-1 or human HeLa cells as described earlier (13, 15, 18). One-day-old subconfluent cells in 60-mm dishes were transfected with 5 μg of each construct by the calcium phosphate coprecipitation technique. Cells were harvested 48 h later. CAT activities of cell extracts were quantitated by measuring the conversion of ^{14}C-labeled chloramphenicol to its acetylated products.

Enhancer competition analysis. CV-1 monkey kidney cells were used for the enhancer competition experiments. CAT plasmid pEn1-cat, having enhancer fragment En1 cloned upstream from the SV40 early promoter, was used as the test gene plasmid. TK gene plasmid constructs having the En1, En2, or En3 fragment cloned upstream from the early promoter were used as competitor plasmids. Cells were transfected with 5 μg of the test gene plasmid and variable amounts of the competitor plasmids, with plasmid pBR322 used as carrier to yield a total of 25 μg of DNA per 60-mm dish of cells. Cells were harvested 48 h later, and CAT activities of cell extracts were determined as outlined above.

Mapping of mRNA 5′ ends by primer extension. COS-1 monkey kidney cells (11) were transfected with 10 μg each of the CAT gene plasmid constructs made in this study containing the various enhancer segments shown in Fig. 1 as described above. The cells were harvested at 48 h post-transfection, and cytoplasmic RNA was isolated by Nonidet P-40 lysis of the cells as described by Favaloro et al. (6). A 21-mer synthetic oligodeoxynucleotide with the sequence 5′-CAACGGTGGTATATCCAGTGA-3′, complementary to the CAT coding sequence located 34 to 14 nucleotides downstream from the AUG start codon, was used as the primer. The primer, 5′ end-labeled with ^{32}P, was annealed with the CAT mRNA present in the cytoplasmic RNA preparations and extended by reverse transcription. The products were resolved by electrophoresis on 8% urea–polyacrylamide denaturing gels and identified by autoradiography.

TK transformation frequency assay. The TK transformation assays were performed by the method of Wigler et al. (37). Subconfluent human 143B TK$^-$ cells in 60-mm dishes were transfected with 5 μg of each plasmid construct by the calcium phosphate coprecipitation technique. The cells were transferred to hypoxyanthine-aminopterin-thymidine selective medium at 16 h after transfection. The medium was replaced once in 3 days. After 14 to 18 days of transfection, the colonies of cells that developed were fixed with methanol, stained with Giemsa dye, and counted.

RESULTS

The SV40 early promoter segments contained in the parent plasmid vectors used for cloning the enhancer segments in this study include the early mRNA initiation sites clustered between SV40 nucleotides 5237 and 5233 (9, 25) and the TATA box element for this initiation located between nucleotides 21 and 15. Additionally, the $pA_{10}cat_2$ plasmid contains all three copies of the 21-bp upstream promoter element (22), having six copies of a GC box binding site for the eucaryotic transcription factor Sp1 (5), while the pSV2X21 plasmids contain four copies of this GC box sequence, which are sufficient for early promoter activity (10). The $pA_{10}cat_2$ plasmid contains, additionally, the origin-proximal 22 bp of the 72-bp repeat enhancer sequence, which is devoid of enhancer activity (18). In monkey CV-1 and human HeLa cells, these enhancerless (but promoter-containing) parent plasmids direct gene expression very inefficiently, but can be made to function well by the linkage in *cis* of an enhancer sequence (3, 13, 18, 27, 36; Byrne, Ph.D. thesis). The expression of the TK and CAT marker genes used can be measured in a convenient, reliable, and sensitive fashion, providing a quantitative measure of enhancer strength.

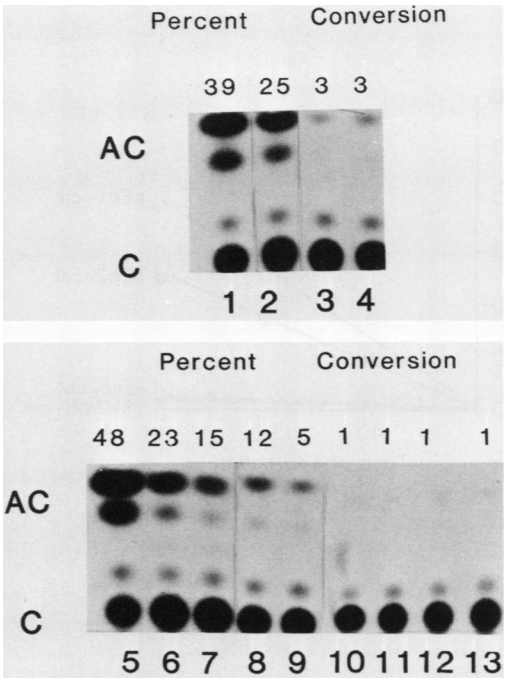

FIG. 2. Thin-layer chromatography (TLC) assay of CAT activities expressed in CV-1 cells from plasmids containing different portions of the SV40 enhancer region. Subconfluent CV-1 cells were transfected with 5 μg of each of the plasmids indicated. The plasmids used in this experiment were: lane 1, pEn1-cat; lane 2, pEn2-cat; lane 3, pEn3-cat; lane 4, pSV2X21-cat; lane 5, pEn4-cat; lane 6, pEn 7(X2)-cat; lane 7, pEn6-cat; lane 8, pEn8(X4)-cat; lane 9, pEn5-cat; lane 10, pEn7-cat; lane 11, pEn8-cat; lane 12, pEn9-cat; and lane 13, $pA_{10}cat_2$. Cells were harvested 48 h later, and CAT activities in cell extracts were determined by measuring the conversion of [^{14}C]chloramphenicol (C) to its acetylated products (AC) by TLC followed by liquid scintillation counting of the spots (13). The cell extract reaction mixture contained 1 μl of [^{14}C]chloramphenicol (0.1 μCi, 47 μCi/mmol), 1 μl of chloamphenicol (10 mM), 5 μl of acetyl coenzyme A (10 mM), 30 μl of cell extract, and 23 μl of Tris chloride, pH 7.8 (0.25 M). Autoradiographs of a TLC assay are shown. The percent conversion of chloramphenicol to acetylchloramphenicol is indicated above each lane.

Origin-proximal 20-bp section of the 72-bp repeat is dispensable for enhancement. Enhancer fragment En6 corresponds to what has been known as the minimal essential enhancer region, including one copy of the 72-bp repeat and a 22-bp adjacent sequence up to the *Pvu*II site at nucleotide 272. Segment En1 contains a repetition of a portion of this region, and segment En4 has two full repetitions of the 72-bp sequence and corresponds to the complete enhancer region. Not surprisingly, constructs pEn4-cat, pEn1-cat, and pEn6-cat expressed the CAT gene quite efficiently under transient expression conditions (Fig. 2 and 3). Similarly, segment En1 functioned as an enhancer efficiently when stably integrated as part of the TK construct pEn1-tk (Fig. 4).

Enhancer fragment En2 was made by truncating the minimal essential enhancer region at the single *Sph*I site at nucleotide 200 to remove the origin-proximal portion of the enhancer. The *Sph*I site in En2 was repaired, however, by

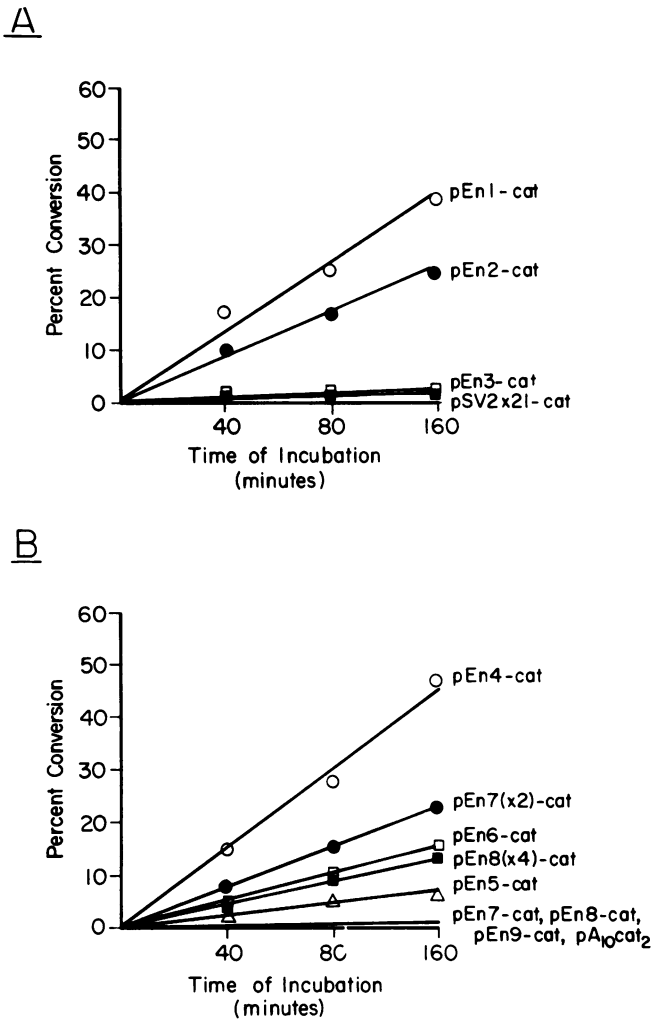

FIG. 3. Analyses of transient expression of the CAT gene from plasmids containing the various enhancer segments. The plasmid constructs were made as described in the legend to Fig. 1 and the text. CV-1 cells were transfected with the plasmids. Cells were harvested 48 h later, and CAT activities in cell extracts were determined by the TLC procedure of Gorman et al. (13) as outlined in the legend to Fig. 2. The percent conversion of ^{14}C-labeled chloramphenicol is plotted against reaction time. Plasmids pSV2X21-cat and $pA_{10}cat_2$ are the enhancerless parent vectors.

joining with plasmid pBR322 cut at the *Sph*I site at its nucleotide 561 position (20). This union with the plasmid sequence also furnished an AT base pair equivalent to that at nucleotide 199 in SV40 DNA. The end result is the deletion in En2 of the origin-proximal 20-bp region of the enhancer located between nucleotides 179 and 198 (Fig. 1). Enhancer fragment En2 functioned quite efficiently in both the CAT (Fig. 2 and 3A) and TK (Fig. 4) assays and was also effective in competition for *trans*-acting factors (Fig. 5). The enhancer activity of fragment En2, determined by these assays, was about 65 to 75% of that of fragment En1, which retains a copy of the origin-proximal 20-bp sequence. The work of Zenke et al. (39) also showed that point mutations in this origin-proximal 20-bp region do not drastically affect the functioning of the enhancer. These results indicate that the origin-proximal 20-bp region of the enhancer is dispensable.

Origin-proximal boundary of the 3′ domain of the enhancer. The 3′ (or A) domains of the enhancer are known to be located around the *Sph*I sites at nucleotides 128 and 200 (26, 30). However, the exact boundaries of this 3′ domain are not known. Enhancer fragments En2 and En7 both have deletions at the origin-proximal portion of the enhancer. Unlike En2, in which the *Sph*I site at nucleotide 200 was restored by fusion with the plasmid *Sph*I site, the *Sph*I end of En7 was trimmed and replaced by a *Bgl*II linker sequence. The result of this alteration was the deletion in En7 of 24 bp of the origin-proximal portion of the enhancer (Fig. 1). Segment En7 failed to function as an enhancer element (Fig. 2 and 3B). Since En2 functioned efficiently as an enhancer, and En7, having a deletion of an additional 4-bp sequence, did not, we place the origin-proximal boundary of the 3′ domain of the enhancer region at nucleotide 199.

5′ domain alone not sufficient for enhancer function even after duplication. Previous work has shown that the 5′ (or B) domain is very important for the activity of the SV40 enhancer (1, 2, 7, 8, 21, 26, 30, 34, 36). The 39-bp fragments En3 and En9 contained the 5′ domain, including both copies of the core sequence identified by Weiher et al. (36), with a 15-bp spacing between them. Segments En3 and En9 did not function as an enhancer sequence under either transient

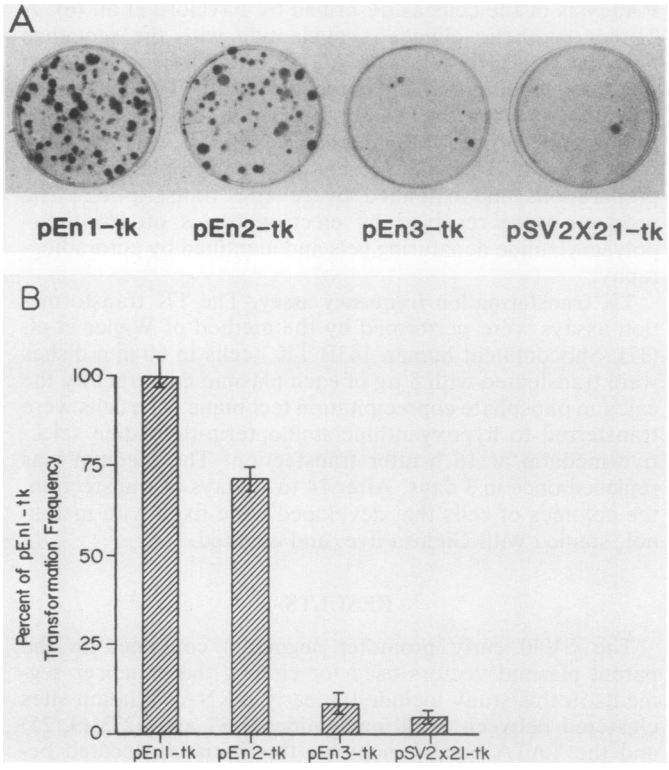

FIG. 4. Enhancer activity of segments of the SV40 enhancer region measured by the TK transformation assay. (A) TK^+ colonies obtained by transformation of human 143B TK^- cells by plasmids pEn1-tk, pEn2-tk, and pEn3-tk, made as described in the legend to Fig. 1. Plasmid pSV2X21-tk is the enhancerless parent vector. The transformation assay was carried out as described in the text. (B) Comparison of TK transformation frequencies. The frequencies of transformation of $143B^+$ TK^- cells to the TK^+ phenotype by the constructs pEn1-tk (104 colonies), pEn2-tk (76 colonies), and pEn3-tk (9 colonies) are presented relative to that (100%) by pEn1-tk. Plasmid pSV2X21-tk (five colonies) is the enhancerless parent vector. The results shown are the means and standard errors of the mean for five different experiments in triplicate.

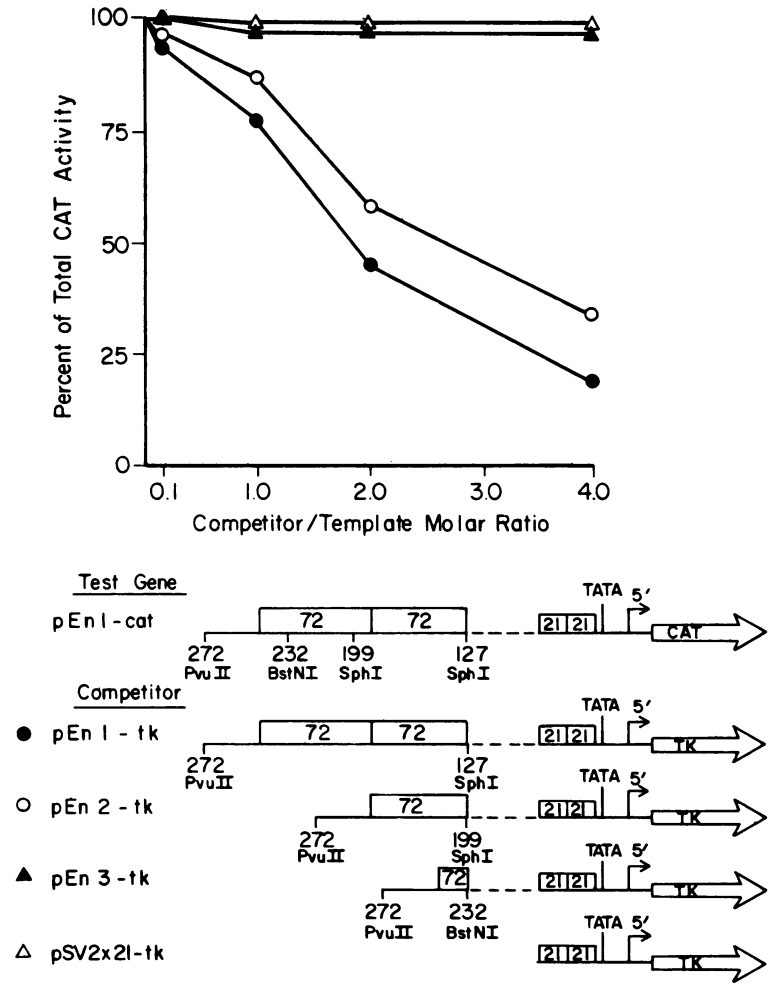

FIG. 5. Competition for *trans*-acting, enhancer-binding factors in vivo in CV-1 cells. The results of the competition experiment are shown at the top, and the plasmids used are shown at the bottom. The test gene plasmid was pEn1-cat. The competitor plasmids were pEn1-tk, pEn2-tk, and pEn3-tk, having decreasing portions of the enhancer region as indicated; pSV2X21-tk is an enhancerless plasmid. All the plasmids contained the 21-bp and TATA elements of the SV40 early promoter and the early mRNA 5′ end as shown. A 5-μg amount of test gene plasmid was used to transfect each 60-mm dish of cells, since this amount was found to be within the linear portion of a dose-response curve in a previous study in our laboratory (17a). Variable amounts of the competitor plasmids were cotransfected with the test gene plasmid, with pBR322 DNA added as carrier, to yield a total of 25 μg of DNA per dish. Cells were harvested 48 h posttransfection, and the CAT activities of cell extracts were determined as outlined in the text. The CAT activity expressed from pEn1-cat in the absence of competitors was taken as 100%, and the expression obtained in the presence of the competitors is plotted as a percentage of that value.

expression (Fig. 2 and 3) or stable transformation (Fig. 4) conditions. Construct pEn9X2-cat contained two copies of the 39-bp En9 fragment cloned in tandem upstream from the SV40 early promoter in $pA_{10}cat_2$. This duplicated segment failed to function efficiently in transient expression assays (Fig. 3B).

The inability of segment En3 to function as an enhancer was also revealed by competition experiments performed under transient expression conditions (Fig. 5). The test gene plasmid in this experiment was pEn1-cat, possessing a good enhancer segment, and CAT gene expression from this plasmid was assayed in the presence and absence of cotransfected competitor plasmids pEn1-tk, pEn2-tk, pEn3-tk, and the enhancerless pSV2X21-tk. These competitor plasmids were of the same background as pEn1-cat except for having the TK gene instead of the CAT gene and the presence (or absence) of different enhancer segments. The results (Fig. 5) showed that pEn1-tk and pEn2-tk competed efficiently with pEn1-cat, causing a five- or threefold decrease, respectively, in CAT gene expression at a 4:1 competitor-template molar ratio. On the other hand, pEn3-tk and the enhancerless pSV2X21-tk competed very poorly, if at all (Fig. 5). These results indicate that the 5′ domain, although necessary, is not sufficient for enhancer activity.

Part of the 5′ domain located outside the 72-bp repeat augments enhancer activity. The complete enhancer region of SV40 customarily includes a 22-bp sequence between nucleotides 251 and 272 located outside the 72-bp repeat. This sequence is also present within the minimal enhancer region and is a part of the 5′ domain of the enhancer. Construct pEn5-cat contained one full copy of the 72-bp repeat plus a portion of the second copy. The major deficiency in this enhancer fragment was the lack of the 22-bp sequence located outside the 72-bp repeat (Fig. 1). The enhancer activity of the En5 segment was approximately 10- and 3-fold less than that of the complete (En4) or minimal (En6) enhancer regions, respectively (Fig. 2 and 3B). Thus, this 22-bp sequence located outside the 72-bp repeat, although

not obligatory for enhancement, can augment enhancer activity significantly.

Certain types of duplications of 5′ and 3′ domains can restore enhancer function. It was shown above that the duplication of the 39-bp En9 segment corresponding to the 5′ domain of the enhancer did not restore enhancer function (Fig. 3B). On the other hand, duplication of the En7 segment (which lacked enhancer activity in its monomer form) restored enhancer activity to a level higher than that of the minimal enhancer region contained in En6 (Fig. 2 and 3B). Both En9 and En7 contained the 5′ domain, but En7 contained in addition a portion of the truncated 3′ domain and an approximately 20-bp-long sequence located between the 5′ and 3′ domains. Similarly, En8 contained the 3′ domain in full and also the origin-proximal 20-bp sequence (shown to be dispensable in this study) and the bulk of the 20-bp sequence located between the 5′ and 3′ domains, but lacked the 5′ domain. En8 lacked enhancer activity in its monomer form, but when present in four tandemly repeated copies possessed a good amount of enhancer activity (Fig. 2 and 3B).

There is some evidence that the procaryotic CAT mRNA sequence is inherently unstable in the eucaryotic environment (8, 22). For this reason, direct quantitations of CAT mRNA levels in animal cells is fairly difficult (22). Since there is no background level of CAT activity in animal cells, the CAT activity produced in transfected cells is an accurate measurement of CAT gene expression. The various CAT constructs used in this study have sequence differences only in the 5′-proximal enhancer region. If differences in CAT activities in transfected cells are to be interpreted as being due to differences in transcriptional enhancement, it is important to prove that the CAT mRNA 5′ ends are the same in all cases. We determined the CAT mRNA 5′ ends in transfections involving various constructs by the primer extension technique. We found that a single mRNA initiation site was used in all these cases (Fig. 6) and that it corresponded to the nucleotide 5230 site used predominantly by the SV40 early mRNA (9, 25). Although different enhancer segments were used, regardless of whether they were immediately upstream or at a distance of about 350 bp upstream from the promoter, the mRNA initiation site was accurately and solely determined by the SV40 early promoter common to all these constructs. The primer extension technique is basically a qualitative one, but as seen from the data in Fig. 6, it appears to be semiquantitative since, for instance, the En7 dimer-containing constructs produced a good amount of the primer extension product, while the monomer lacking enhancer activity did not produce any. Similarly, the transcriptional enhancer activities of the En1, En2, and En3 segments were reflected to some extent in the amounts of the primer extension product that they produced.

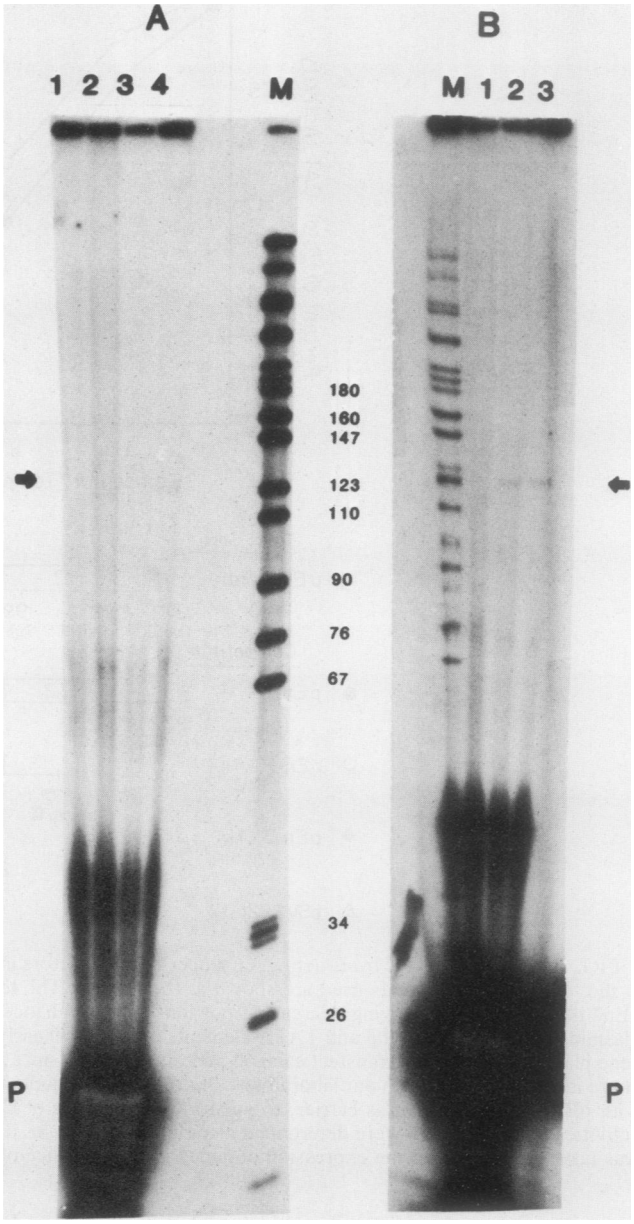

FIG. 6. Mapping the CAT mRNA 5′ ends with the primer extension technique. COS-1 cells were transfected. (A) pSV2X21-cat based constructs: lane 1, pSV2X21-cat; lane 2, pEn2:cat; lane 3, pEn3-cat; lane 4, pEn1-cat. (B) $pA_{10}cat_2$-based constructs: lane 1, pEn7-cat; lane 2, pEn7(X2)-cat; lane 3, pEn8(X4)-cat. Lanes M, *Msp*I digest of plasmid pBR322 for use as size markers. Cytoplasmic RNA was isolated from the cells at 48 h posttransfection. A 21-nucleotide-long synthetic oligonucleotide with the sequence 5′-CAACGGTGGTATATCCAGTGA-3′, complementary to the CAT coding sequence located between 34 and 14 nucleotides downstream from the AUG start codon, was used as the primer. The primer (P), 5′-end labeled with ^{32}P, was annealed with the CAT mRNA present in the RNA preparations and extended by reverse transcription. The products were resolved on 8% urea–polyacrylamide denaturing gels along with the size markers. The main extension product (indicated by arrows in A and B) was approximately 120 nucleotides long, as determined by comparison with the size markers, corresponding to mRNA initiation around nucleotide 5235 in the SV40 sequence. Sizes (in nucleotides) are indicated between the panels.

DISCUSSION

The sequences contained in the SV40 enhancer region required for the enhancement function and for binding *trans*-acting factors mediating the enhancer activity were investigated by using selective deletion mutagenesis in this study. Previous studies of this type have indicated that the minimal essential enhancer element is a 94-bp sequence located between nucleotides 179 and 272 (4, 26, 30, 34). In the present study, we have obtained evidence that the origin-proximal 20 bp of this 94-bp region is dispensable for enhancer function. Since a slightly extended deletion removing another 4 bp of the origin-proximal region abolished

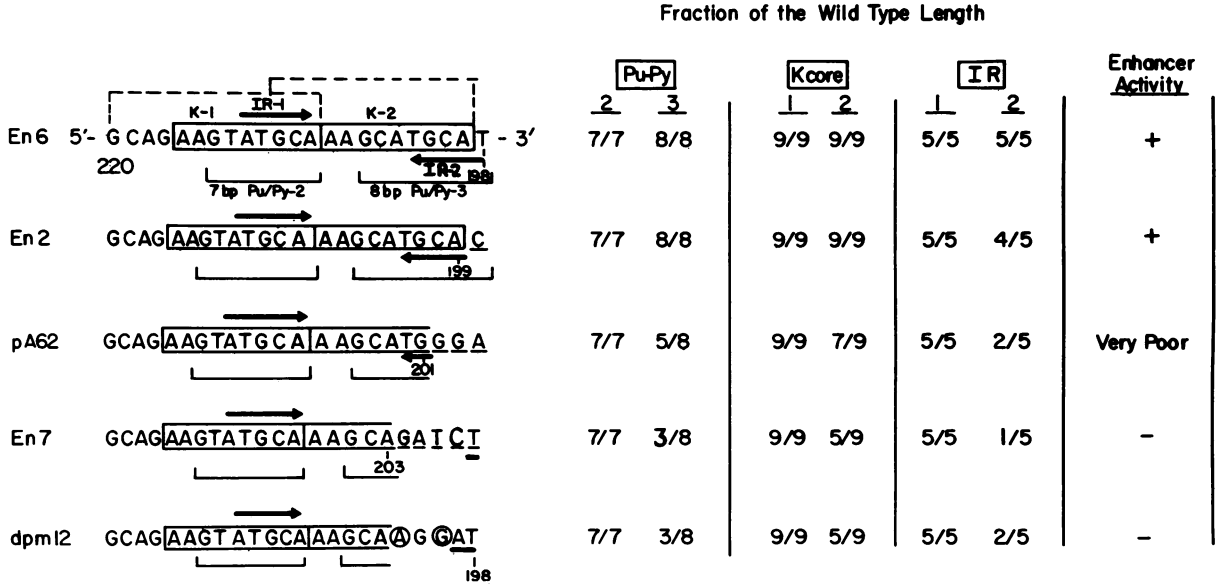

FIG. 7. Sequence requirements at the 3' domain deduced by comparison of various deletion and point mutations in this area of the SV40 enhancer. Segments En6, En2, and En7 were analyzed in this study. Segment A62 was analyzed by Wildeman et al. (38). Segment dpm12 was analyzed by Herr and Gluzman (16). En6 contains the entire 94-bp enhancer region located between nucleotides 272 and 179, including the 5' and 3' domains and the middle repeats (see Fig. 8). The other four fragments contain the 5' domain and the middle repeats but terminate at various points at or near the 3' domain, as indicated by the SV40 nucleotide numbers below the sequences at the right. Only the sequence at the 3' domain is shown. Boxed oligonucleotides are the proposed K core repeats K-1 and K-2 at the 3' domain. The K-2 core overlaps the *Sph*I cleavage site, 5'-GCATGC-3'. The dashed bracket at the top of En6 indicates an overlapping repeat that includes the K core sequences. Brackets below the sequences indicate the Pu/Py elements 2 and 3 that can potentially assume a Z-DNA structure (23). Note the presence of a sequence, 5'-ATGCAAAG-3', overlapping K cores 1 and 2, which is similar to an octanucleotide, 5'-ATGCAAAT-3', occurring in the immunoglobulin enhancer and heavy-chain promoter elements (24). The arrows show the sequence 5'-ATGCA-3' occurring as an inverted repeat (IR) with a spacing of 5 bp. In the K-2 core region, underlined residues indicate neighboring sequences (e.g., linkers) juxtaposed due to the deletion, and circled bases indicate base change mutations. Boxes, arrows, or brackets that are incomplete indicate deletion or point mutations within the K-2 core region. At the right, the correlation between the presence of these sequence features and the enhancer activity of the various fragments is indicated.

enhancer activity, we placed the origin-proximal (or 3') boundary of the enhancer at nucleotide 199 and redefined the minimal essential enhancer as a 74-bp segment located between nucleotides 199 and 272 (Fig. 7).

In other studies of this type performed previously (4, 26, 38), the origin-proximal sequence was deleted by digestion with *Sph*I, but the *Sph*I 3' cohesive end was removed before the fragment was cloned into vectors, resulting in a fragment resembling En7. Such a fragment did not function well in enhancement of transcription (similar to our En7), in competition for enhancer-binding, *trans*-acting factors in vitro (26, 38), or in enhancing the efficiency of replication from the polyomavirus origin as a replacement for the polyomavirus enhancer sequence (4). From these results, it was concluded that the origin-proximal sequence of the SV40 enhancer is not dispensable. In our present studies, we have done further work to restore the *Sph*I site and retain it to produce segment En2, lacking the origin-proximal 20-bp sequence, and found that it had efficient enhancer activity.

Since the 3' domain of the enhancer is believed to be situated around the *Sph*I site, our assignment of nucleotide 199 as the 3' boundary of the minimal enhancer region means that it should also be the 3' boundary of the 3' (or A) domain. As shown in Fig. 7, this 3' domain contains two copies of a nonanucleotide present as a tandem repeat with the sequences 5'-(216)AAGTATGCA(208)-3' and 5'-(207)AAGCAT GCA(199)-3', respectively, the numbers within parentheses being SV40 nucleotide numbers (31). (Note that there is only one mismatch out of nine in this repetition.) This repeated sequence motif (termed the K core sequence) could serve as the recognition signal for a *trans*-acting factor(s) that would bind to the 3' domain of the enhancer region, similar to the role postulated for the core sequence 5'-GTGGA/TA/ TA/TG-3' (termed the C core sequence) found repeated in the 5' domain of the SV40 enhancer and several other enhancer elements (17, 36). The junction between the two K core repeats contains the octanucleotide 5'-ATGCAAAG-3', similar to the consensus octanucleotide in the immunoglobulin promoter and enhancer elements (24). The sequence at the 3' domain can also be written as an inverted repeat with the sequence 5'-(211)ATGCAAAGCATGCAT(199)-3' (the inverted repetition is underlined), which could be the recognition signal for a dimeric protein. Another possibility is an overlapping 12-bp repeat with the sequences 5'-(220)GCA GAAGTATGCA(208)-3' and 5'-(210)GCAAAGCATGCA (199)-3' (note that the first copy has the underlined G as an extra base and that there is only 1 mismatch out of 12 in the two copies).

In the deletion giving rise to the inactive En7 fragment, the origin-distal copy of the K core 5'-AAGTATGCA-3' motif (termed K-1) was retained, but the origin-proximal copy (termed K-2) was deleted except for the first five nucleotides (Fig. 7). The inverted repetition and the overlapping direct repetition pointed out above were also eliminated in the deletion producing En7. All these sequence repetitions are also adversely affected in the double point mutation dpm12 (16) and in the deletion mutant segment A62 (38), lacking enhancement activity, but were retained in our enhancer

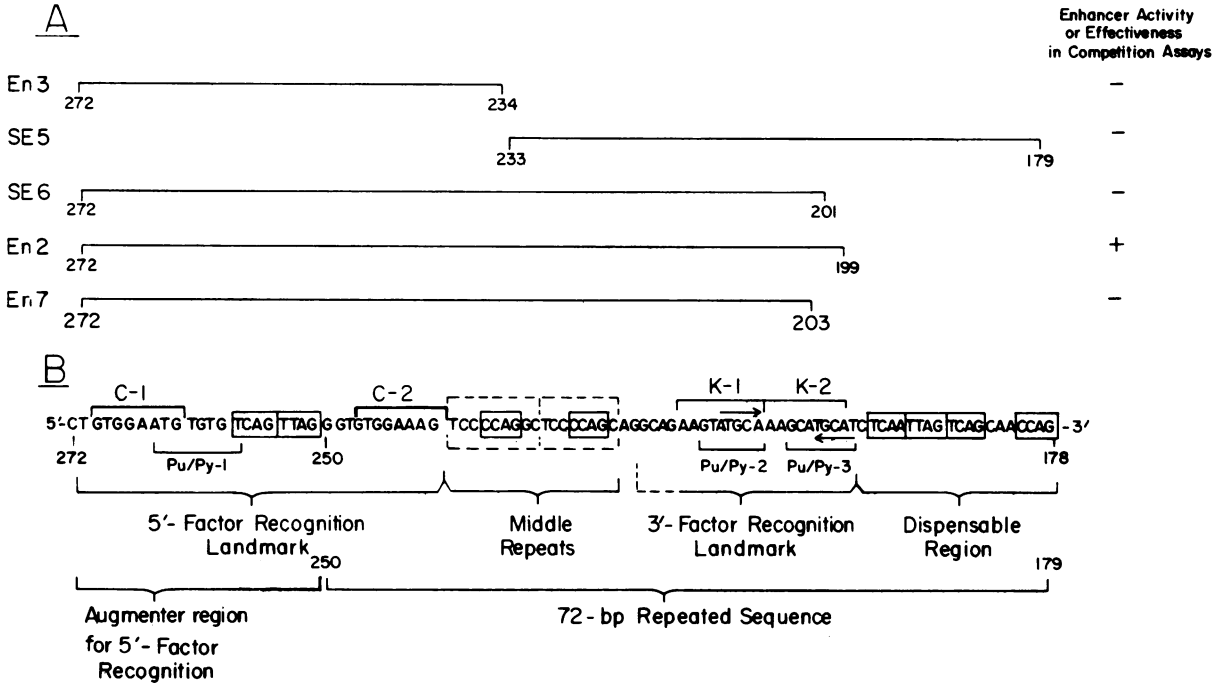

FIG. 8. Structure and function of the SV40 enhancer region. (A) Enhancer region fragments, their enhancer activity, and their effectiveness in competing for enhancer-binding factors. Segments SE5 and SE6 were described by Sassone-Corsi et al. (26). Segments En3, En2, and En7 were analyzed in this study. According to our results, the 74-bp En2 segment located between nucleotides 272 and 199 corresponds to the minimal essential enhancer sequence. (B) Enhancer region sequence and landmarks. The C-1 and C-2 core sequences (17, 19, 36), the K-1 and K-2 core sequences proposed in this study, and the Pu/Py elements (23) are indicated. Boxed tetranucleotides indicate the sequence 5'-PyPyAPu-3' that occurs many times within the enhancer. The inverted repeats occurring in the 3' domain are indicated. Note the presence of the sequence 5'-ATGCAAAG-3', spanning the junction between the K-1 and K-2 cores in the 3' domain, which resembles an octanucleotide that occurs in the immunoglobulin promoter and enhancer sequence (24). Dashed boxes indicate the middle repeats that occur between the 5' and 3' domains. This middle region could be a factor-binding region or serve the purpose of a spacer between the 5' and 3' domains. The 20-bp sequence located between nucleotides 198 and 179 is dispensable for enhancement according to the present study. This study also provides evidence that the sequence between nucleotides 272 and 251 augments factor binding at the 5' domain of the enhancer.

fragment En2, which possessed enhancer activity (Fig. 7). These results indicate that at least one of these different types of sequence repetitions around the *Sph*I site is important for enhancer activity.

The *Sph*I site in the enhancer is also one of the regions that is potentially capable of producing the left-handed Z-DNA conformation under certain conditions (23). The 8-bp Pu/Py element 5'-(205)GCATGCAT(198)-3' (termed Pu/Py-3) located at the *Sph*I site was adversely affected in our enhancer fragment En7 and in mutants dpm12 (16) and A62 (38). However, all these mutants retained a second 7-bp Pu/Py element, 5'-(214)GTATGCA(208)-3' (termed Pu/Py-2), (Fig. 8) located nearby within the 3' domain of the enhancer, which would be expected to be capable of forming Z-DNA conformation (23). The En7 and A62 segments (but not the dpm12 segment) also contain another 8-bp Pu/Py element, 5'-(265)ATGTGTGT(258)-3' (termed Pu/Py-1), located within the 5' domain of the enhancer overlapping one of the 5' domain core sequences (termed C-1) (Fig. 8). The two Pu/Py elements 1 and 2 were of the proper size (seven to nine nucleotides long) and were present at a spacing of 51 bp, similar to the arrangement observed in several enhancer elements (23). If such an arrangement is the only (or major) requirement for enhancer function, then the En7 and pA62 segments should possess enhancer activity which was not found to be the case. Thus, the ability to form the Z-DNA structure may not be the sole criterion for enhancer function.

In the case of certain deletion or point mutations knocking out enhancer activity, duplication of sequences located within or outside the enhancer has been found to restore enhancer activity (16, 29, 33, 35). In our present study, duplication of fragments En7 and En8, but not of En9, restored enhancer activity. En7 has an intact 5' domain and a truncated 3' domain. En8 has an intact 3' domain and lacks the 5' domain. Between the 5' and 3' domains in the wild-type enhancer region are two copies of the sequence motif 5'-(239)TCCCCAGGC(213)-3' and 5'-(230)TCCCCA GC(223)-3', occurring as a tandem repetition (Fig. 8). This motif may not merely have a spacerlike function between the 5' and 3' domains but could potentially be a factor-binding domain in its own right and could be termed the middle repeat domain (Fig. 8). Enhancer segment En7 contains both copies of this motif and En8 contains one of them, but En9 lacks it altogether. The occurrence of this motif at a definite periodicity of repetition may be one of the reasons why En7 and En8 functioned when duplicated while En9 did not.

Another reason for the lack of function of the En9 dimer as an enhancer could be the lack of proper spacing between duplicate copies of the same domain (in this case the 5' domain). Duplication of the 39-bp En9 segment resulted in two 5' domains located side by side within a segment of 78 bp, which could be too frequent a periodicity for proper interaction. Duplications resulting from the En7 and En8 segments contained a spacing of 50 to 70 bp between similar

domains, which might be appropriate for interactions between them. Based on the sequence organization within the wild-type SV40 enhancer, it is optimal for similar domains to be spaced at a distance of approximately 72 bp, although dissimilar domains could be located with much less distance between them.

Though the core sequences of the 5′ and 3′ enhancer domains are apparently quite different, similarities are evident if the core sequences are broken up further. For instance, the tetranucleotide 5′-A/TA/TA/TG-3′ occurs at both the 5′ and 3′ domains and occurs a total of five times within the 74-bp minimal enhancer region. The tetranucleotide 5′-GCAT-3′ or 5′-GCAA-3′ occurs in the 3′ domain, and a slightly variant one, 5′-GCAG-3′, occurs in the potential middle domain. The tetranucleotide 5′-PyPyAPu-3′ (shown boxed in Fig. 8) occurs several times within the enhancer. Thus, it is possible for each domain to be recognized by more than one factor and for a given factor to recognize more than one domain. This could explain why homologous duplications of the same domain as well as heterologous combinations of two different domains could produce enhancer activity.

ACKNOWLEDGMENTS

We thank Julie Yamaguchi for expert technical assistance, S. Chandrasekharappa for synthesizing and providing the 21-nucleotide-long CAT mRNA primer, and Leonor Piedra for expert word-processing of this manuscript.

This research was supported by Public Health Service grant 1-R01-GM-35872-01 from the National Institutes of Health and by grants NP-248F and MV-293G from the American Cancer Society. T.A.F. is partially supported by Public Health Service training grant CA-09318-04 from the National Institutes of Health.

LITERATURE CITED

1. **Banerji, J., S. Rusconi, and W. Schaffner.** 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell **27**:299–308.
2. **Benoist, C., and P. Chambon.** 1981. *In vivo* sequence requirements of the SV40 early promoter region. Nature (London) **290**:304–310.
3. **Byrne, B. J., M. S. Davis, J. Yamaguchi, D. J. Bergsma, and K. N. Subramanian.** 1983. Definition of the simian virus 40 early promoter region and demonstration of a host range bias in the enhancement effect of the simian virus 40 72-base-pair repeat. Proc. Natl. Acad. Sci. USA **80**:721–725.
4. **de Villiers, J., W. Schaffner, C. Tyndall, S. Lupton, and R. Kamen.** 1984. Polyoma virus DNA replication requires an enhancer. Nature (London) **312**:242–246.
5. **Dynan, W. S., and R. Tjian.** 1983. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell **35**:79–87.
6. **Favaloro, J., R. Treisman, and R. Kamen.** 1980. Transcription maps of polyoma virus-specific RNA: analysis by two-dimensional nuclease S1 gel mapping. Methods Enzymol. **65**:718–749.
7. **Fromm, M., and P. Berg.** 1982. Deletion mapping of DNA regions required for SV40 early region promoter function *in vivo*. J. Mol. Appl. Genet. **1**:457–481.
8. **Gheysen, D., A. van de Voorde, R. Contreras, J. van der Heyden, F. Duerinck, and W. Fiers.** 1983. Simian virus 40 mutants carrying extensive deletions in the 72-base-pair repeat region. J. Virol. **47**:1–14.
9. **Ghosh, P. K., and P. Lebowitz.** 1981. Simian virus 40 early mRNA's contain multiple 5′ termini upstream and downstream from a Hogness-Goldberg sequence; a shift in 5′ termini during the lytic cycle is mediated by large T antigen. J. Virol. **40**:224–240.
10. **Gidoni, D., J. T. Kadonaga, H. Barrera-Saldana, K. Takahashi, P. Chambon, and R. Tjian.** 1985. Bidirectional SV40 transcription mediated by tandem Sp1 binding interactions. Science **230**:511–517.
11. **Gluzman, Y.** 1981. SV40-transformed simian cells support the replication of early SV40 mutants. Cell **23**:175–182.
12. **Gluzman, Y., and T. Shenk (ed.).** 1983. Enhancers and eukaryotic gene expression. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
13. **Gorman, C. M., L. F. Moffat, and B. M. Howard.** 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. **2**:1044–1051.
14. **Gruss, P., R. Dhar, and G. Khoury.** 1981. Simian virus 40 tandem repeated sequences as an element of the early promoter. Proc. Natl. Acad. Sci. USA **78**:943–947.
15. **Hartzell, S. W., B. J. Byrne, and K. N. Subramanian.** 1984. The simian virus 40 minimal origin and the 72-base-pair repeat are required simultaneously for efficient induction of late gene expression with large tumor antigen. Proc. Natl. Acad. Sci. USA **81**:6335–6339.
16. **Herr, W., and Y. Gluzman.** 1985. Duplications of a mutated simian virus 40 enhancer restore its activity. Nature (London) **313**:711–714.
17. **Khoury, G., and P. Gruss.** 1983. Review: enhancer elements. Cell **33**:313–314.

17a. **Kumar, R., T.A. Firak, C. T.Schroll, and K. N. Subramanian.** 1986. Activation of gene expression is adversely affected at high multiplicities of linked SV40 enhancer. Proc. Natl. Acad. Sci. USA **83**:3199–3203.

18. **Laimins, L. A., G. Khoury, C. Gorman, B. Howard, and P. Gruss.** 1982. Host-specific activation of transcription by tandem repeats from simian virus 40 and Moloney murine sarcoma virus. Proc. Natl. Acad. Sci. USA **79**:6453–6457.
19. **Lusky, M., L. Berg, H. Weiher, and M. Botchan.** 1983. Bovine papilloma virus contains an activator of gene expression at the distal end of the early transcription unit. Mol. Cell. Biol. **3**:1108–1122.
20. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
21. **Moreau, P., R. Hen, B. Wasylyk, R. Everett, M. P. Gaub, and P. Chambon.** 1981. The SV40 72-base-pair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res. **9**:6047–6068.
22. **Mosthaf, L., M. Pawlita, and P. Gruss.** 1985. A viral enhancer element specifically active in human haematopoietic cells. Nature (London) **315**:597–600.
23. **Nordheim, A., and A. Rich.** 1983. Negatively supercoiled simian virus 40 DNA contains Z-DNA segments within transcriptional enhancer sequences. Nature (London) **303**:674–679.
24. **Parslow, T. G., D. L. Blair, W. L. Murphy, and D. K. Granner.** 1984. Structure of the 5′ ends of immunoglobulin genes: a novel conserved sequence. Proc. Natl. Acad. Sci. USA **81**:2650–2654.
25. **Reddy, V. B., P. Ghosh, P. Lebowitz, M. Piatak, and S. M. Weissman.** 1979. Simian virus 40 early mRNAs: genomic localization of 3′ and 5′ termini and two major splices in mRNA from transformed and lytically infected cells. J. Virol. **30**:272–296.
26. **Sassone-Corsi, P., A. Wildeman, and P. Chambon.** 1985. A *trans*-acting factor is responsible for the simian virus 40 enhancer activity *in vitro*. Nature (London) **313**:458–463.
27. **Scholer, H. R., and P. Gruss.** 1984. Specific interaction between enhancer-containing molecules and cellular components. Cell **36**:403–411.
28. **Shenk, T. E., J. Carbon, and P. Berg.** 1976. Construction and analysis of viable deletion mutants of simian virus 40. J. Virol. **18**:810–817.
29. **Swimmer, C., and T. Shenk.** 1984. A viable simian virus 40 variant that carries a newly generated sequence reiteration in place of the normal duplicated enhancer element. Proc. Natl. Acad. Sci. USA **81**:6652–6656.
30. **Takahashi, K., N. Vigneron, H. Matthes, A. Wildeman, and P. Chambon.** 1986. Requirements of stereospecific alignments for initiation from the simian virus 40 early promoter. Nature (London) **319**:121–126.
31. **Tooze, J. (ed.).** 1981. DNA tumor viruses, 2nd ed. Cold Spring

Harbor Laboratory, Cold Spring Harbor, N.Y.
32. **van Heuverswyn, H., and W. Fiers.** 1979. Nucleotide sequence of the Hind C fragment of simian virus 40 DNA: comparison of the 5'-untranslated region of wild-type virus and of some deletion mutants. Eur. J. Biochem. **100**:51–60.
33. **Veldman, G. M., S. Lupton, and R. Kamen.** 1985. Polyomavirus enhancer contains multiple redundant sequence elements that activate both DNA replication and gene expression. Mol. Cell. Biol. **5**:649–658.
34. **Wasylyk, B., C. Wasylyk, P. Augereau, and P. Chambon.** 1983. The SV40 72-bp repeat preferentially potentiates transcription starting from proximal natural or substitute promoter elements. Cell **32**:503–514.
35. **Weber, F., J. de Villiers, and W. Schaffner.** 1984. An SV40 "enhancer trap" incorporates exogenous enhancers or generates enhancers from its own sequences. Cell **36**:983–992.
36. **Weiher, H., M. Konig, and P. Gruss.** 1983. Multiple point mutations affecting the simian virus 40 enhancer. Science **219**:626–631.
37. **Wigler, M., A. Pellicer, S. Silverstein, and R. Axel.** 1978. Biochemical transfer of single-copy eucaryotic genes using total cellular DNA as donor. Cell **14**:725–731.
38. **Wildeman, A. G., P. Sassone-Corsi, T. Grundstrom, M. Zenke, and P. Chambon.** 1984. Stimulation of *in vitro* transcription from the SV40 early promoter by the enhancer involves a specific *trans*-acting factor. EMBO J. **3**:3129–3133.
39. **Zenke, M., T. Grunstrom, H. Matthes, M. Wintzerith, C. Schatz, A. Wildeman, and P. Chambon.** 1986. Multiple sequence motifs are involved in SV40 enhancer function. EMBO J. **5**:387–397.