Proc. Natl. Acad. Sci. USA
Vol. 75, No. 5, pp. 2160–2164, May 1978
Biochemistry

Nucleotide sequence of the simian virus 40 small-t gene

[oncogenes/early functions/tumor (T) antigen]

G. VOLCKAERT, A. VAN DE VOORDE, AND W. FIERS

Laboratories of Molecular Biology and Physiological Chemistry, State University of Ghent, Ledeganckstraat 35, B 9000 Ghent, Belgium

Communicated by Paul Berg, February 13, 1978

ABSTRACT The nucleotide sequence of the segment of simian virus 40 DNA between standard map positions 0.53 and 0.65, i.e., approximately half of the restriction fragment *Hind* A, is reported. This segment is located near the beginning of the early region and is transcribed counterclockwise. There is a potential initiating ATG signal at 13 nucleotides from the *Hind* C-*Hind* A junction in the strand with the same polarity as the early mRNA. From this signal on, an open reading frame is present which would allow the synthesis of a polypeptide of 174 amino acids until a TAA termination codon is reached at nucleotide 602 (map position 0.547). This polypeptide, revealed by the DNA sequence, corresponds almost certainly to small-t antigen. Correlation of the deduced amino acid sequence with the NH_2-terminal sequences of small-t and large-T (tumor) antigens of simian virus 40, as established by Paucha *et al.* [Paucha, E., Mellor, A., Harvey, R., Smith, A. E., Hewick, R. M. & Waterfield, M. D. (1978) [*Proc. Natl. Acad. Sci. USA* 75, 2165–2169], strongly argues that both proteins are indeed initiated at the ATG triplet. Because the DNA region between 0.547 and 0.534 is blocked for translation in all three reading frames by multiple termination codons, we conclude that the large-T antigen must be coded for by two noncontiguous DNA segments: the segment from 0.65 to around 0.60, which small-t and large-T antigens share, and another segment starting at some point after position 0.534 and continuing counterclockwise until it terminates at map position 0.174. Small-t antigen is methionine-rich and has a remarkably high number of cysteine residues clustered mainly in its COOH-terminal half. It is rich in both basic and acidic residues, the former being slightly in excess.

The genome of simian virus 40 (SV40) contains two transcriptional regions. Early mRNA is synthesized counterclockwise from about standard map position 0.67 (the origin of DNA replication) to 0.16 (see figure 6 in ref. 1). After expression of the early gene(s), DNA replication starts and the late region, extending clockwise from around map position 0.67 to 0.17 (2–4), can be transcribed. Only the early region is transcribed in abortively infected cells and in SV40-transformed cells, and the SV40 early gene(s) play an essential role in the initiation of, and presumably in the maintenance of, the transformed state.

SV40 early mRNA codes for tumor (T) antigen (5–7), a protein of 90,000–100,000 daltons present in the nuclei of productively infected cells and SV40-transformed cells. U antigen and tumor specific transplantation antigen are presumably identical to the T antigen or are derived therefrom (8, 9). T antigen binds specifically to SV40 DNA, is essential for viral DNA replication in permissive cells, induces host cell DNA synthesis and expression of host cell functions, and helps human adenovirus to grow in monkey cells (2–4). Genetic analysis has thus far revealed only one complementation group of temperature-sensitive early mutants, namely tsA. All these mutations map counterclockwise from map position 0.43 [four are in *Hind* H, eight in *Hind* I, and one in *Hind* B (10)]. Deletion mutants have allowed further delimitation of the A gene: while some mutants, like dl-1263 at position 0.21, affect the size of the T antigen, others, in the region 0.54–0.59, do not impair viability and induce an apparently normal T antigen (11, 12). This was somewhat puzzling because the region from position 0.54 to 0.17 encodes only enough information to specify at most a polypeptide of 72,000 daltons, considerably less than the 90,000 daltons observed for T antigen. These A-gene boundaries were confirmed and more precisely mapped by nucleotide sequence analysis: there are several translational termination codons in the three reading frames in the region around 0.53–0.55 (13, 14), and the termination codon TAA for the T-antigen gene has been pinpointed at 0.174 (15).

Another intriguing finding was that serum derived from hamsters bearing SV40-induced tumors not only precipitates the aforementioned large-T antigen, but also reacts with a 15,000- to 20,000-dalton protein called small-t antigen (5, 16, 17). In addition to being immunologically related, large-T and small-t antigens have in common a number of methionine-containing tryptic peptides. Moreover, the deletion mutants in the 0.54–0.59 area lead either to no detectable small-t antigen or to a shortened derivative (M. Sleigh, W. Topp, and J. Sambrook, personal communication; J. Feunteun, personal communication; ref. 17). These observations can all be explained by the following model (ref. 17; see also figure 6 in ref. 1): small-t and large-T antigens would both be initiated at the same position, corresponding to around 0.65 on the standard map. Small-t antigen would be continuously coded for by the region 0.65–0.55 (approximate values). Large-T antigen would be coded for by two noncontiguous segments; one from 0.65 to almost 0.59, and a second from 0.53 to 0.17. The joining of the two segments would presumably occur by splicing at the mRNA level. SV40 mutants with a deletion in the 0.54–0.59 area are in several aspects similar to the hr-t mutants of polyoma virus, which were isolated and extensively characterized by Benjamin and coworkers (18, 19). These polyoma mutants are not only of the host range type, they are also transformation-negative and map in an area (relative to the origin of DNA replication) on the polyoma genome (20) that corresponds to the area on the SV40 genome in which the SV40 mutants map. Also, the SV40 mutants with an impaired small-t antigen have a considerably decreased ability to transform cells to anchorage-independent growth (12).

The nucleotide sequence between map positions 0.649 and 0.533 is presented here. Correlation of this structural information with the data of Paucha *et al.* (1) on the NH_2-terminal amino acid sequence of small-t and large-T antigens provides strong evidence that the aforementioned model for the organization of the small-t and large-T genes is essentially correct.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: SV40, simian virus 40; T antigen, tumor antigen.

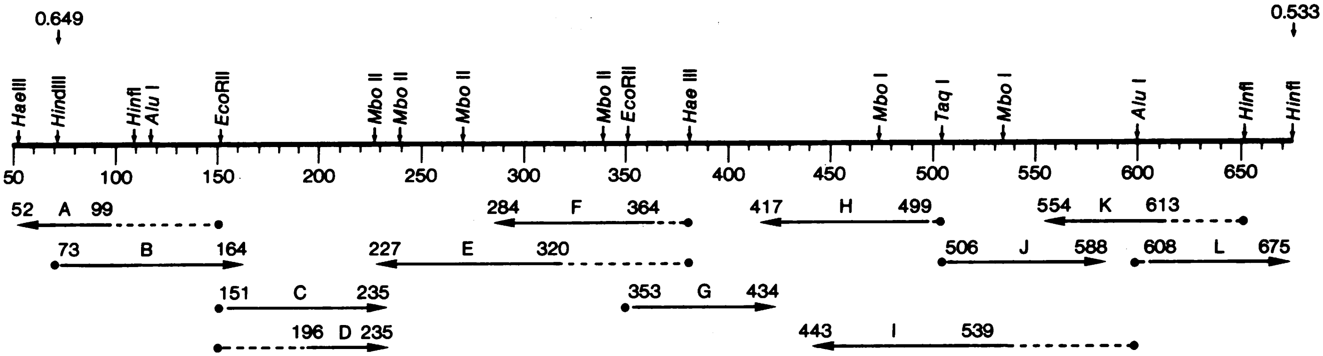

FIG. 1. Distribution of restriction sites and arrangement of the sequences shown in the autoradiograms (Fig. 2). The location of the recognition sites of the restriction enzymes used for sequence determination of this part of the genome is shown above the central line. The numbering is relative to the zero point defined as follows (25): the center of the large palindromic sequence around map position 0.663, where the single *Bgl* I recognition site on SV40 DNA is also located, is taken as the starting point for the early SV40 DNA region. Hence the nucleotide sequence presented here spans the region from *n* = 68 to *n* = 675 as shown in Fig. 3. The lower part of the figure shows the positions of the nucleotide sequences indicated on the autoradiograms in Fig. 2. Filled circles refer to the 5′-labeled end; broken lines represent nucleotide sequences that ran off the gel; solid lines correspond to the nucleotide sequence accompanying each autoradiogram. The numbers refer to the first and last nucleotide indicated alongside the autoradiograms in Fig. 2. The nucleotide sequence to the left of position *n* = 68 will be published elsewhere and clearly shows the overlap of the *Hin*dIII recognition site at the boundary of *Hind* fragments C and A.

MATERIALS AND METHODS

SV40 strain 776, originally provided by D. Nathans, was our main source of viral DNA throughout this work. A few comparative analyses were run with DNA derived from the WT 830 strain obtained from C. Cole. All experimental procedures have been described: cell and virus growth (21), isolation and purification of viral DNA (21), and preparation of the appropriate restriction fragments (22). Nucleotide sequence analysis was essentially done by the method of Maxam and Gilbert (ref. 23; as revised by Maxam and Gilbert, personal communication). The chemical degradation products were separated on 90 × 30 × 0.2 cm polyacrylamide gels containing 20% or, more recently, 10% acrylamide. Exposure to Fuji medical x-ray film was at −70°, with CAWO intensifying screens (Schrobenhausen, Germany) (24).

RESULTS

The SV40 restriction fragment *Hind* A represents 22.38% of the viral genome. The results described here involve the half of this fragment that is closest to the *Hind* C/A junction and the origin of DNA replication (see figure 6 in ref. 1). Fig. 1 shows the distribution of restriction endonuclease sites present in this part of the genome.

The procedure followed for nucleotide sequence analysis was invariably the same for each restriction fragment: DNA fragments with only one ^{32}P-labeled 5′ terminus were prepared and worked up by the procedure of Maxam and Gilbert (23). One polyacrylamide gel was run until the first nucleotide reached the bottom of the gel. Generally, a sequence of about 50 nucleotides on a 20% gel, or about 80 nucleotides on a 10% gel, could be read unambiguously from the autoradiograms. On a second gel, a large number of nucleotides were allowed to run off. This permitted us to read off a following sequence of nucleotides (up to 200 nucleotides from the labeled end). Because the bands migrate closely in the upper part of the gel, care must be taken not to overlook any nucleotide. Sequence results were considered definitive only when the bands were separated on the gel by about 0.5 cm.

Fig. 1 summarizes the positions of the degradation patterns shown in Fig. 2. These patterns are illustrative examples, which together cover the total sequence. All regions, however, have been analyzed at least in duplicate, and every fragment has been analyzed on several gels such that the critical areas were appropriately spread out. Moreover, a large part of the sequence was approached in the same direction from two different restriction sites (see below), and almost 70% has been verified by analysis of the complementary strand. Only some regions close to a restriction site showed anomalous mobility behavior on the gels. A striking example was obtained with a sequence near the *Hind* C/A junction. Our final results indicate that this sequence is pA-G-C-T-T-T-G-C-A-A-A-Gp, but all oligonucleotides smaller than pA-G-C-T-T-T-G-Cp migrated more slowly than expected (Fig. 2*B*). The sequence was confirmed on another gel starting from the *Hae* III site at *n* = 51 in the counterclockwise direction. Under these conditions the same region behaved normally. The sequence was also checked by analysis of the complementary strand (Fig. 2*A*). The cause of retarded fragment migration is unknown, but the formation of secondary structure may be involved. To avoid such difficulties, we also analyzed each sequence around a restriction site on another fragment in which this restriction site was internal.

A partial comparison was made with DNA derived from SV40 WT 830. A *Hinf*I digest was labeled and cleaved by *Taq* I (this produces a single cut in the *Hinf* D fragment). The resulting two subfragments were analyzed and sequences from nucleotide 256 to 350 and from 601 to 507 were derived. They were completely identical to our results obtained on strain 776.

DISCUSSION

The procedure of Maxam and Gilbert allows rapid sequence analysis of DNA. With pure preparations of DNA fragments and high quality reagents, the sequence can be read for over 200 nucleotides, provided several gels are run so that the critical regions can be appropriately spaced out. Only with some smaller fragments did we observe anomalous migrating behavior, but this difficulty could easily be overcome by using other fragments containing the same region. Most of the sequence was confirmed by analysis of the complementary strand. The region reported here contained a sufficient number of restriction enzyme cleavage sites that could be labeled with polynucleotide kinase and that generated fragments suitable for further digestion with another restriction enzyme (see Fig. 1).

The first part of the sequence agrees with the results (as recently revised) of S. M. Weissman and his colleagues (26, 27). The second part, however, shows a number of differences, e.g., in the region between nucleotides 582 and 651, published by

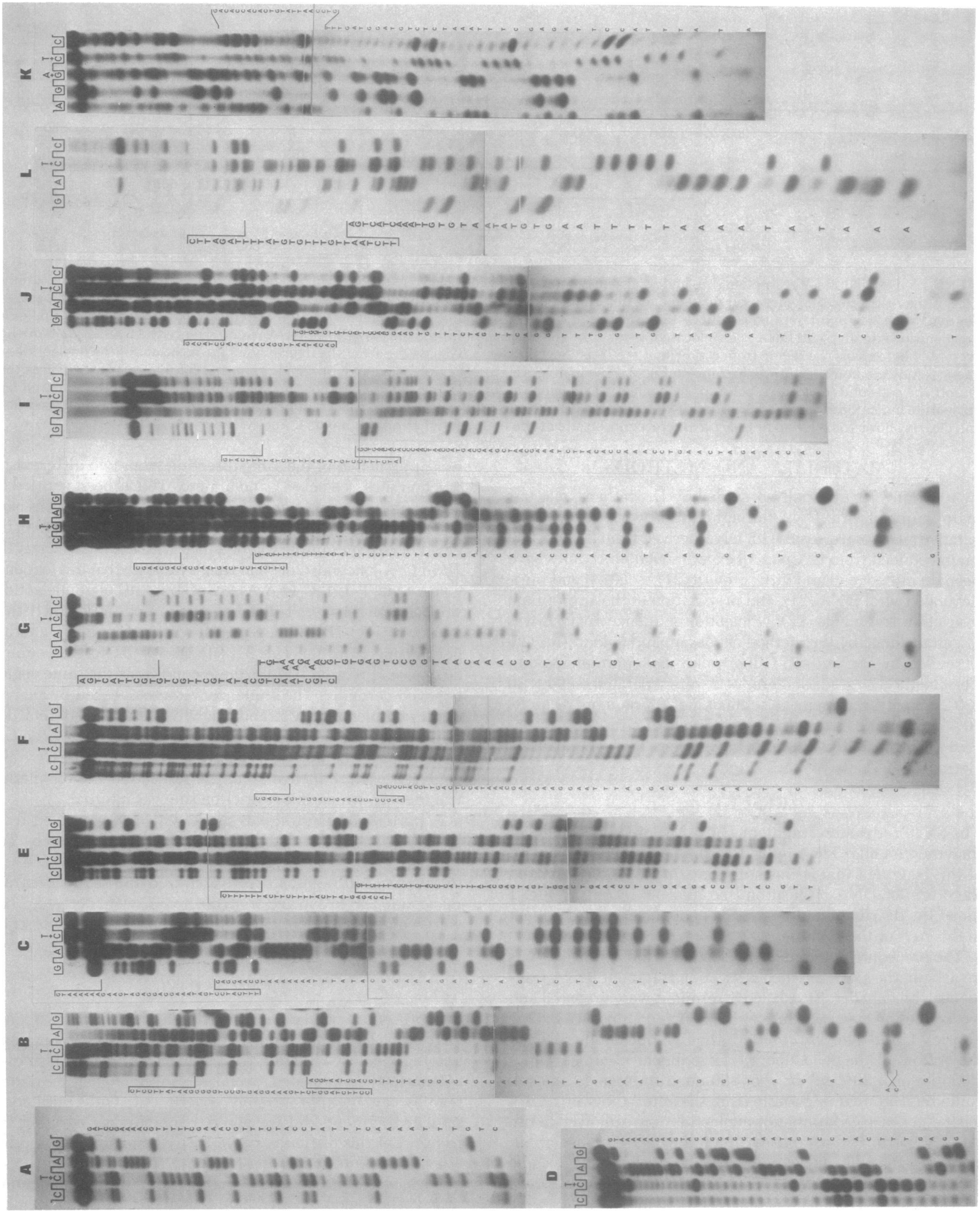

FIG. 2. Chemical degradation patterns covering the total nucleotide sequence. DNA fragments labeled at only one 5′ end were treated by the procedure of Maxam and Gilbert (23). The different digests were run in adjacent lanes. The specificity of each reaction is indicated at the top of the lane. The space between the top of the gel and the position of the undegraded material has been cut off. The degradation patterns are designated *A–L* and correspond to the segments shown in Fig. 1. As discussed in the text, the nucleotide sequences indicated alongside each autoradiogram were not only derived from the pattern represented, but are also based on other, usually longer, runs (not shown). For example, the upper segment of *C* could be spaced out by a longer run, as shown in *D*. Also, many more complementary sequences from the other strand, in addition to those shown here, have been analyzed.

```
                Met-Asp-Lys-Val-Leu-Asn-Arg-Glu-Glu-Ser-Leu-Gln-Leu-Met-Asp-Leu-Leu-Gly-Leu-Glu-Arg-Ser-Ala-Trp-Gly-Asn-

AGC TTTGCAAAG ATG GAT.AAA.GTT.TTA.AAC.AGA.GAG.GAA.TCT.TTG.CAG.CTA.ATG.GAC.CTT.CTA.GGT.CTT.GAA.AGG.AGT.GCC.TGG.GGG.AAT.
TCG AAACGTTTC TAC CTA TTT CAA AAT TTG TCT CTC CTT AGA AAC GTC GAT TAC CTG GAA GAT CCA GAA CTT TCC TCA CGG ACC CCC TTA
 70         80        90        100        110       120        130        140        150

Ile-Pro-Leu-Met-Arg-Lys-Ala-Tyr-Leu-Lys-Lys-Cys-Lys-Glu-Phe-His-Pro-Asp-Lys-Gly-Gly-Asp-Glu-Glu-Lys-Met-Lys-Lys-Met-

ATT.CCT.CTG.ATG.AGA.AAG.GCA.TAT.TTA.AAA.AAA.TGC.AAG.GAG.TTT.CAT.CCT.GAT.AAA.GGA.GGA.GAT.GAA.GAA.AAA.ATG.AAG.AAA.ATG.
TAA GGA GAC TAC TCT TTC CGT ATA AAT TTT TTT ACG TTC CTC AAA GTA GGA CTA TTT CCT CCT CTA CTT CTT TTT TAC TTC TTT TAC
160         170        180        190        200        210        220        230        240

Asn-Thr-Leu-Tyr-Lys-Lys-Met-Glu-Asp-Gly-Val-Lys-Tyr-Ala-His-Gln-Pro-Asp-Phe-Gly-Gly-Phe-Trp-Asp-Ala-Thr-Glu-Val-Phe-

AAT.ACT.CTG.TAC.AAG.AAA.ATG.GAA.GAT.GGA.GTA.AAA.TAT.GCT.CAT.CAA.CCT.GAC.TTT.GGA.GGC.TTC.TGG.GAT.GCA.ACT.GAG.GTA.TTT.
TTA TGA GAC ATG TTC TTT TAC CTT CTA CCT CAT TTT ATA CGA GTA GTT GGA CTG AAA CCT CCG AAG ACC CTA CGT TGA CTC CAT AAA
   250        260        270        280        290        300        310        320        330

Ala-Ser-Ser-Leu-Asn-Pro-Gly-Val-Asp-Ala-Met-Tyr-Cys-Lys-Gln-Trp-Pro-Glu-Cys-Ala-Lys-Lys-Met-Ser-Ala-Asn-Cys-Ile-Cys-

GCT.TCT.TCC.TTA.AAT.CCT.GGT.GTT.GAT.GCA.ATG.TAC.TGC.AAA.CAA.TGG.CCT.GAG.TGT.GCA.AAG.AAA.ATG.TCT.GCT.AAC.TGC.ATA.TGC.
CGA AGA AGG AAT TTA GGA CCA CAA CTA CGT TAC ATG ACG TTT GTT ACC GGA CTC ACA CGT TTC TTT TAC AGA CGA TTG ACG TAT ACG
       340        350        360        370        380        390        400        410

Leu-Leu-Cys-Leu-Leu-Arg-Met-Lys-His-Glu-Asn-Arg-Lys-Leu-Tyr-Arg-Lys-Asp-Pro-Leu-Val-Trp-Val-Asp-Cys-Tyr-Cys-Phe-Asp-

TTG.CTG.TGC.TTA.CTG.AGG.ATG.AAG.CAT.GAA.AAT.AGA.AAA.TTA.TAC.AGG.AAA.GAT.CCA.CTT.GTG.TGG.GTT.GAT.TGC.TAC.TGC.TTC.GAT.
AAC GAC ACG AAT GAC TCC TAC TTC GTA CTT TTA TCT TTT AAT ATG TCC TTT CTA GGT GAA CAC ACC CAA CTA ACG ATG ACG AAG CTA
420        430        440        450        460        470        480        490        500

Cys-Phe-Arg-Met-Trp-Phe-Gly-Leu-Asp-Leu-Cys-Glu-Gly-Thr-Leu-Leu-Leu-Trp-Cys-Asp-Ile-Ile-Gly-Gln-Thr-Thr-Tyr-Arg-Asp-

TGC.TTT.AGA.ATG.TGG.TTT.GGA.CTT.GAT.CTT.TGT.GAA.GGA.ACC.TTA.CTT.CTG.TGG.TGT.GAC.ATA.ATT.GGA.CAA.ACT.ACC.TAC.AGA.GAT.
ACG AAA TCT TAC ACC AAA CCT GAA CTA GAA ACA CTT CCT TGG AAT GAA GAC ACC ACA CTG TAT TAA CCT GTT TGA TGG ATG TCT CTA
   510        520        530        540        550        560        570        580        590

Leu-Lys-Leu

TTA.AAG.CTC TAA GGTAAA TATAAAATTT TTAAGTGTAT AATGTGTTAA ACTACTGATT CTAATTGTTT GTGTATTTTA GATTC
AAT TTC GAG ATT CCATTT ATATTTTAAA AATTCACATA TTACACAATT TGATGACTAA GATTAACAAA CACATAAAAT CTAAG
       600       610        620        630        640        650        660        670
```

FIG. 3. Nucleotide and amino acid sequence corresponding to the small-t gene and adjacent nucleotide sequences. The upper strand of the DNA sequence has the same polarity as SV40 early mRNA. The nucleotide sequence represents the genome segment from map position 0.649 to 0.533. The standard numbering system is explained in Fig. 1 (25). The choice of the ATG triplet at position $n = 80$ as initiating codon is discussed in the text. From this point on, an amino acid sequence of 174 residues can be deduced. The termination codon at the end of the small-t gene is boxed; termination codons in other reading frames are underlined. Beyond the TAA triplet at $n = 602$, a remarkable A·T region of 18 base pairs is present.

Thimmappaya and Weissman (13). These differences have important implications in that they may affect the potential reading frames in which translational termination codons appear and hence they may change the amino acid sequence and the total length of the small-t polypeptide. In order to check whether this part of the early region is more prone to variation by mutational drift, we checked parts of the sequence (a total of about 200 nucleotides) on the WT 830 strain used in P. Berg's laboratory. Not a single residue was different.

The nucleotide sequence for the region 0.649–0.533 is shown in Fig. 3. We know that the sequence reported here must contain (at least part of) the small-t gene because deletions in the region 0.54–0.59 result in truncated small-t polypeptides (see introduction). The strand with the same polarity as early mRNA contains an ATG signal at 13 nucleotides from the *Hin*d C/A junction [this ATG starts at nucleotide 80 in our standard numbering system for the whole genome (25)]. Continuous translation from this potential initiating codon would allow the synthesis of a 174-residue-long polypeptide, while the two other reading frames are blocked by multiple termination codons. This ATG at position 80 is not preceded upstream in phase by another ATG codon, only by termination codons; this is further evidence that codon 80–82 functions in initiation and does not signify an internal methionine residue. Small-t and large-T antigens are normally modified at the NH_2 terminus, presumably by acetylation. Paucha *et al.* (1), however, could avoid this blocking, thus opening the way for sequential chemical analysis of the *in vitro* synthesized polypeptides. Their data are in complete agreement with the amino acid sequence deduced from our nucleotide sequence and are not compatible with any other sequence on the SV40 genome. This proves that the gene for small-t antigen as well as for large-T antigen starts at nucleotide 80 (Fig. 3). The gene for small-t antigen is presumably not interrupted since small-t antigen can be synthesized using as messenger cRNA transcribed *in vitro* from SV40 DNA by means of *Escherichia coli* RNA polymerase (16). It follows that the amino acid sequence shown in Fig. 3 corresponds to the primary structure of the small-t antigen. We do not know whether the primary product is further processed *in vivo*, e.g., at the NH_2 terminus before blocking.

The polypeptide of 174 amino acids has a mass of 20,503 daltons, which is more than the estimate of 17,000 daltons based on mobility in sodium dodecyl sulfate gels (5, 16). The unusually high content of cysteine (11 residues) could possibly account for this discrepancy. Indeed, after reduction and alkylation, small-t antigen moves in a denaturing gel like a 20,000-dalton polypeptide (L. V. Crawford and P. Z. O'Farrell, personal communication). The cysteine residues are clustered particularly in the second half of the protein. Also, the methionine content is rather high, and six of nine internal residues are preceded by a basic amino acid. Small-t antigen is slightly basic (27 Lys + Arg, 3 His, as opposed to 25 Asp + Glu residues), and about half of the basic residues occur in doublets. It is rather low in the polar amino acids Ser and Thr. The relative use of code words is similar to that in other parts of the SV40 genome (14, 25). The dinucleotide CpG, which is very rare in vertebrate

DNA in general and in SV40 DNA in particular, occurs once in this region (as part of the *Taq* recognition site). This CpG is found in a position spanning two codons, the C being part of the codon UUC for phenylalanine. This is somewhat surprising because phenylalanine is predominately coded for by UUU (25), and it seems, therefore, as if the conservation of this particular CpG may have a special, but as yet unknown, significance.

Paucha *et al.* (1) have shown that the large-T antigen starts at the same position as small-t. This explains the occurrence of common tryptic peptides in both proteins and the immunological crossreactivity (5, 16). As mentioned in the introduction, the termination codon TAA for the large-T antigen has been located at map position 0.174 (15). It is preceded by a translatable stretch of DNA (map position 0.534–0.174) that can code for at most 72,000 daltons of protein (25), compared to an estimate of 90,000–100,000 daltons for large-T antigen. The region 0.54–0.59 cannot be part of the structural gene for large-T antigen because deletions that do not affect the size of this protein have been mapped in this area (11) and because multiple termination codons are present in all three reading frames between nucleotides 602 and 671 (Fig. 3). These results firmly establish that large-T antigen is coded for by two noncontiguous regions on the genome. In analogy to the now well-documented splicing events in SV40 late mRNA (28–31) and adenovirus early and late mRNA (32–36), it is very likely that the two structural parts of the genetic information are joined at the mRNA level. The question then remains as to exactly what segment is excised. Paucha *et al.* (1) present evidence that at least the first 25 amino acids are common to small-t and large-T antigens. On the other hand, some of the deletion mutants that affect small-t but not large-T antigen have lost the *Hae* III site at n = 378–381 (e.g., the mutant dl-2112 has a deletion which includes this *Hae* III site as well as the *Alu* I site at n = 597–600; J. Feunteun, personal communication; this corresponds to a deletion of at least 217 base pairs). Another piece of information is provided by analysis of the deletion mutant dl-1001, which lacks the *Hin*dII + III restriction fragments H and I (7). This mutant induces the synthesis of a 33,000-dalton protein which is immunologically related to large-T antigen. The fragment protein must almost entirely be coded for by information present in *Hin*d A [this follows from an inspection of the total DNA sequence (27)]. The second half of *Hin*d A (counterclockwise from position 0.535) has a maximal coding capacity of 20,500 daltons (unpublished data). This leaves 12,500 daltons, i.e., about 95 amino acids, to be coded for by the segment common to small-t and large-T genes. On this basis we would predict that the first boundary of the splicing will be found between nucleotides 350 and 380 (Fig. 3), and that the second boundary will occur very soon beyond the earliest possible site at map position 0.534.

Little is thus far known about the biological functions of the two viral proteins involved in oncogenicity. Benjamin and coworkers (18, 37) postulated that the polyoma small-t protein might play a role in the induction of cellular factors (permissivity factors), perhaps mediated by histone modification. Our hope is that the known primary structure of these SV40 oncogenic proteins may contribute to an understanding of their mode of action.

We are grateful to Dr. A. E. Smith and his colleagues and to Dr. J. Feunteun for keeping us informed of their results. J. Van der Heyden contributed excellent technical assistance. This research was supported by a grant from the Kankerfonds of the Algemene Spaar- en Lijfrentekas (A.S.L.K.) of Belgium.

1. Paucha, E., Mellor, A., Harvey, R. Smith, A. E., Hewick, R. M. & Waterfield, M. D. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 0000–0000.
2. Fareed, G. C. & Davoli, D. (1977) *Annu. Rev. Biochem.* **46**, 471–522.
3. Fried, M. & Griffin, B. E. (1977) *Adv. Cancer Res.* **24**, 67–113.
4. Kelly, T. J. & Nathans, D. (1977) *Adv. Virus Res.* **21**, 81–173.
5. Prives, C., Gilboa, E., Revel, M. & Winocour, E. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 457–461.
6. Carroll, R. B. & Smith, A. E. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 2254–2258.
7. Rundell, K., Collins, J. K., Tegtmeyer, P., Ozer, H. L., Lai, C.-J. & Nathans, D. (1977) *J. Virol.* **21**, 636–646.
8. Robb, J. A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 447–451.
9. Anderson, J. L., Martin, R. G., Chang, C., Mora, P. T. & Livingston, D. M. (1977) *Virology* **76**, 420–425.
10. Lai, C.-J. & Nathans, D. (1975) *Virology* **66**, 70–81.
11. Shenk, T. E., Carbon, J. & Berg, P. (1976) *J. Virol.* **18**, 664–671.
12. Cole, C. N., Landers, T., Goff, S. P., Manteuil-Brutlag, S. & Berg, P. (1977) *J. Virol.* **24**, 277–294.
13. Thimmappaya, B. & Weissman, S. M. (1977) *Cell* **11**, 837–843.
14. Van de Voorde, A., Contreras, R., Haegeman, G., Rogiers, R., Van Heuverswyn, H., Van Herreweghe, J., Volckaert, G., Ysebaert, M. & Fiers, W. (1977) in *Early Proteins of Oncogenic DNA Viruses*, eds. May, P., Monier, R. & Weil, R. (INSERM, Paris), Vol. 69, pp. 17–30.
15. Van Heuverswyn, H., Van de Voorde, A. & Fiers, W. (1978) *Eur. J. Biochem.*, in press.
16. Paucha, E., Harvey, R., Smith, R. & Smith, A. E. (1977) in *Early Proteins of Oncogenic DNA Viruses*, eds. May, P., Monier, R. & Weil, R. (INSERM, Paris), Vol. 69, pp. 189–198.
17. Crawford, L. V., Cole, C. N., Smith, A. E., Paucha, E., Tegtmeyer, P., Rundell, K. & Berg, P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 117–121.
18. Staneloni, R. J., Fluck, M. M. & Benjamin, T. L. (1977) *Virology* **77**, 598–609.
19. Fluck, M. M., Staneloni, R. J. & Benjamin, T. L. (1977) *Virology* **77**, 610–624.
20. Feunteun, J., Sompayrac, L., Fluck, M. M. & Benjamin, T. L. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 4169–4172.
21. Yang, R., Van de Voorde, A. & Fiers, W. (1976) *Eur. J. Biochem.* **61**, 101–117.
22. Volckaert, G., Contreras, R., Soeda, E., Van de Voorde, A. & Fiers, W. (1977) *J. Mol. Biol.* **110**, 467–510.
23. Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
24. Laskey, R. A. & Mills, A. D. (1977) *FEBS Lett.* **82**, 314–316.
25. Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van de Voorde, A., Van Herreweghe, J. Van Heuverswyn, H. Volckaert, G. & Ysebaert, M. (1978) *Nature*, in press.
26. Dhar, R., Subramanian, K. N., Pan, J. & Weissman, S. M. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 827–831.
27. Reddy, V. B., Thimmappaya, B., Dhar, R., Subramanian, K. N., Zain, B. S., Pan, J., Celma, M. L. & Weissman, S. M. (1978) *Science*, in press.
28. Aloni, Y., Dhar, R., Laub, O., Horowitz, M. & Khoury, G. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 3686–3690.
29. Celma, L. Dhar, R., Pan, J. & Weissman, S. M. (1977) *Nucleic Acids Res.* **4**, 2549–2559.
30. Hsu, M. & Ford, J. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4982–4985.
31. Haegeman, G. & Fiers, W. (1978) *Nature*, in press.
32. Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. (1977) *Cell* **12**, 1–8.
33. Klessig, D. F. (1977) *Cell* **12**, 9–21.
34. Dunn, A. R. & Hassell, J. A. (1977) *Cell* **12**, 23–36.
35. Berget, S. M., Moore, C. & Sharp, P. A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 3171–3175.
36. Kitchingman, G. R., Lai, S.-P. & Westphal, H. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4392–4395.
37. Schaffhausen, B. S. & Benjamin, T. L. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 1092–1096.