Volume 8 Number 10 1980 Nucleic Acids Research

A rapid enzymatic DNA sequencing technique: determination of sequence alterations in early simian virus 40 temperature sensitive and deletion mutants

I.Seif, G.Khoury and R.Dhar

Laboratory of Molecular Virology, National Cancer Institute, National Institutes of Health, Bethesda, MD 20205, USA

Received 3 March 1980

ABSTRACT

We have adapted a rapid sequencing technique from the enzymatic nick-translation method of Maat and Smith. The Forward-Backward procedure employs both synthetic and 3' to 5' exonucleolytic activities of E. coli DNA polymerase I to achieve greater reliability, especially in reading stretches of the same nucleotide. The technique has been employed to determine sequence alterations in four early SV40 temperature-sensitive (tsA) point mutants and five early SV40 viable deletion mutants. The nucleotide sequence of these mutants provides an insight into their biological properties.

INTRODUCTION

A major scientific accomplishment in the past few years, has been the development of several rapid nucleic acid sequencing techniques. These methods have permitted the assembling of large banks of sequence data which promise to play a central role in unraveling many of the nucleotide signals which regulate gene expression. For example the complete DNA sequence of three Papovaviruses, simian virus (SV) 40 (1, 2), mouse polyoma virus (3, 4, 5) and human BK virus (6, 7), have recently been obtained. The sequences of these genomes are of particular interest since their regulatory signals, which interact with host cell proteins, are in most cases certain to be analogous to the signals used by eukaryotic cells. The 5153 nucleotide pairs of the BKV (Dun) genome were established in our laboratory (7), using a rapid enzymatic procedure that we now report. In this study, the methodology has been applied to the analysis of a familiar sequence, the early region of SV40 (Fig. 1), to determine nucleotide sequence alterations in temperature-sensitive (tsA) or deletion (dl) mutants. The biological properties of these mutants are analyzed in the context of the specific mutations.

MATERIALS AND METHODS

Enzymes and Nucleotides. Restriction endonucleases were obtained from New England Bio Labs (Beverly, MA), E. coli DNA polymerase I and DNA

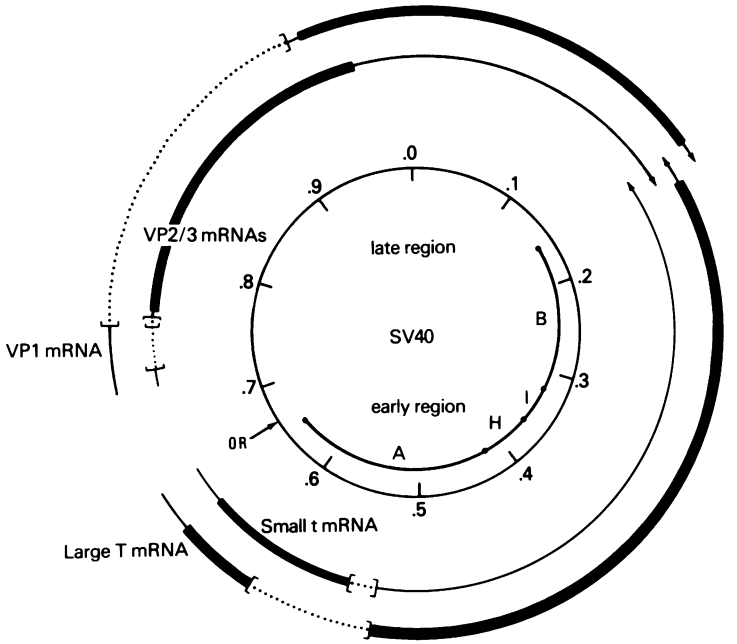


FIGURE 1. Circular map of the simian virus 40 genome.
The genome is a circular double-stranded DNA molecule of about 5200 nucleotide pairs (1, 2) which replicates bidirectionally starting from a unique origin (OR). The map of SV40 DNA is oriented with the single Eco RI site at .0. Two coding regions are defined by the gene functions expressed primarily before (early) and after (late) viral DNA replication. The spliced early and late mRNAs are positioned outside the circular genome with intervening sequences indicated by dotted lines. Protein coding regions are depicted by thick lines on each mRNA. Within the early region two polypeptides are expressed: large T and small t antigens which share the same amino-terminus. The structural proteins VPI, VP2, and VP3 are translated from the late mRNAs. The Hind II+III restriction fragments (A, H, I, B) which cover the early coding region are positioned on the inner circle.

polymerase "large fragment" from Boehringer-Mannheim (Indianapolis, IN), T4 polynucleotide kinase and ddNTPs from P-L Biochemicals (Milwaukee, WI), dNTPs from Sigma (St, Louis, MO). γ^{32}P-ATP (spec.act. 4000 Ci/mmole) was purchased from ICN Pharmaceuticals (Irvine, CA).

Virus Strains. SV40 early viable deletion mutants (8, 9) were provided by T. Shenk (dl-885, dl-886 and dl-888) and M. Sleigh, B. Topp and J. Sambrook (dl-2000 and dl-2007). SV40 early temperature-sensitive mutants (10) were provided by R.G. Martin, C.-J. Lai and J. Alwine (tsA209, tsA239, tsA241 and tsA255). Viral stocks were grown in monkey kidney (AGMK) cells infected at a

multiplicity of 0.1 pfu per cell.

DNA Preparation. SV40 mutants were grown in AGMK cells infected at a multiplicity of 1 pfu per cell. Infection was carried out for 3 days at 37°C for deletion mutants and for 5 days at 32°C for tsA mutants. SV40 DNA was extracted by Hirt's procedure (11) and form I DNA separated by equilibrium centrifugation (12). Form I DNA of tsA mutants was digested with appropriate restriction endonucleases (Mbo I or Hinf I, Hind III, Ava II). Fragments covering the region of the mutated Hind II+III fragment I, were purified by polyacrylamide gel electrophoresis. We found that the DNA of tsA255 has an additional Hinf I site in the wild-type fragment A (in the late region) and that the DNA of mutant tsA209 lacks an Ava II site in the Hind II+III fragment I. For the viable deletion mutants, Alu I and Hinf I DNA fragments were selected which had been shown to cover the deletion sites.

The mutated DNA fragments were labelled at the 5' ends, using γ^{32}P-ATP and T4 polynucleotide kinase (13) and were digested further with appropriate restriction endonucleases (Hinf I or Hae III, Pvu II, BstN I, Hinc II, Hind III) to generate fragments with only one labelled 5' end, which were resolved on polyacrylamide gels.

For DNA elution, the gel slice was ground in 3 ml of a 0.2 M NaCl solution, incubated 12 hours at room temperature, then centrifuged 20 min in a Sorvall HB4 rotor at 2000 rpm (to remove acrylamide particles). The DNA was precipitated from the supernatant by the addition of 2.2 volumes of ethanol.

Sequencing procedure. All the reactions were conducted with either E. coli DNA polymerase I or the "large fragment" of this enzyme. Addition of bovine pancreatic DNAase was not required since the polymerase preparation (see Materials) contained sufficient contaminating nucleases to produce random nicks. Each of the four Forward (FG,FA,FT,FC), four Backward (BG,BA,BT,BC) and four "plus" reactions (+G,+A,+T,+C) was performed in a volume of 5 microliters containing 2 µl of DNA + 1 µl of 5X buffer + 1 µl of enzyme + 1 µl of nucleotide(s). These volumes (2,1,1,1 µl) were taken from the following stock solutions:

→ ^{32}P-DNA ≅ 0.2 pmoles/µl in distilled water.

→ 5X buffer = 33 mM Tris-HCl (pH 7.5), 33 mM $MgCl_2$, 10 mM DTT, 10 mM NaCl.

→ Pol I = 5 units/µl, or "large fragment" = 1 unit/µl.

→ Backward, nucleotide = for BA: solution of 1mM ddA in distilled water.
BG: " 1mM ddG "
BT: " 1mM ddT "
BC: " 1mM ddC "

→ Forward, nucleotides = e.g. FA: (1mM ddA, 200 μM dG, 200 μM dT, 200 μM dC).
→ "plus", nucleotide = e.g. +A: 1mM dA; (fresh solutions for +C and +G).

Each reaction mixture (5 μl) was mixed on plastic, drawn into a 5 μl capillary (unsealed) and incubated at 37°C for 30 min. The reaction mixture was then added, in an Eppendorf tube, to 5 μl of "stop and loading" solution = 99% formamide (deionized) containing 10 mM EDTA, 0.2 mM NaOH and 0.1% xylene cyanol dye. Alternatively, the reaction mixture was added to 5 μl of 10 mM EDTA, lyophilized, and the DNA resuspended in 10 μl of formamide/dye.

Gel Electrophoresis. The 10 μl samples were heat denaturated at 100°C for 30 seconds (3 minutes if NaOH was omitted) and small aliquots were loaded onto the gel wells. To obtain a pattern of bands from which an extensive sequence can be read, the recommended order of loading is: FG,FA,FT,FC,BG,BA,BT,BC. To read nucleotides 20 to 200, 12% and 8% acrylamide gels with 1:30 bisacrylamide were utilized; for nucleotides 200 to 300, 7% acrylamide gels with 1:40 or 1:50 bisacrylamide were employed. Gels (40 x 20 x 0.0004 cm) were run at 20 mA (1000 to 2000 volts) for varying lengths of time (2 to 7 hrs) and autoradiographed at - 80°C with Kodak SB 5 films and DuPont Lightning-plus intensifying screens. Gel bands not visible on the reproduction in Fig.3 were nevertheless distinct on the original films.

RESULTS AND DISCUSSION

Sequencing Procedure. By virtue of their experimental simplicity, the enzymatic methods for DNA sequencing (14, 15, 16) offer certain advantages over the known chemical techniques (13). The sequencing gel patterns obtained by these enzymatic techniques, however, frequently contain ambiguities. We tried to circumvent many of the problems by combining certain aspects of two enzymatic methods in our Forward-Backward (F-B) procedure: the nick-translation method of Maat and Smith (16) and the "plus and minus" method of Sanger and Coulson (14). Maat and Smith have described a procedure in which a double-stranded DNA fragment, ^{32}P labelled at one 5' end, is pretreated with DNAase to create random single-stranded nicks. The 3' terminus created by a nick can bind *E. coli* DNA polymerase I (Pol I) and prime the polymerization of provided deoxynucleoside triphosphates (dG, dA, dT, dC) or prime the addition of a dideoxynucleoside triphosphate chain terminator (ddG, ddA, ddT, ddC). If a mixture contains, for example, the four dNTPs (dG, dA, dT, dC) and an adjusted concentration of ddA, polymerization stops randomly at A residues. Fragments ending in ddA can then be sized on denaturing polyacrylamide gels. This reaction is repeated separately for each

of the four ddNTPs. One difficulty in this procedure is to read the gel band pattern for a row of repeated nucleotides (e.g. 5'...A-A-A-A...3'). The intensity of the bands often decreases drastically in the 5' to 3' direction, a characteristic which is accentuated when the dNTP homologous to the ddNTP is omitted (e.g. + ddA, - dA; see Forward reactions in Fig.2 below). We attempted to solve this problem by conducting, separately, an additional reaction in which the intensity of the bands, in a row of repeated nucleotides, increases in the 5' to 3' direction (see Backward reactions in Fig.2). In Backward reactions, one ddNTP is added, but all the dNTPs are

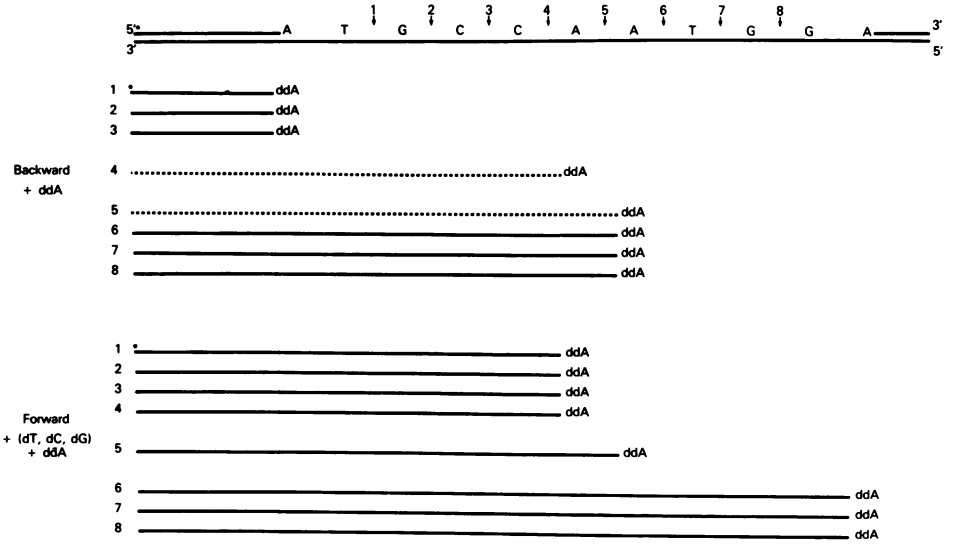
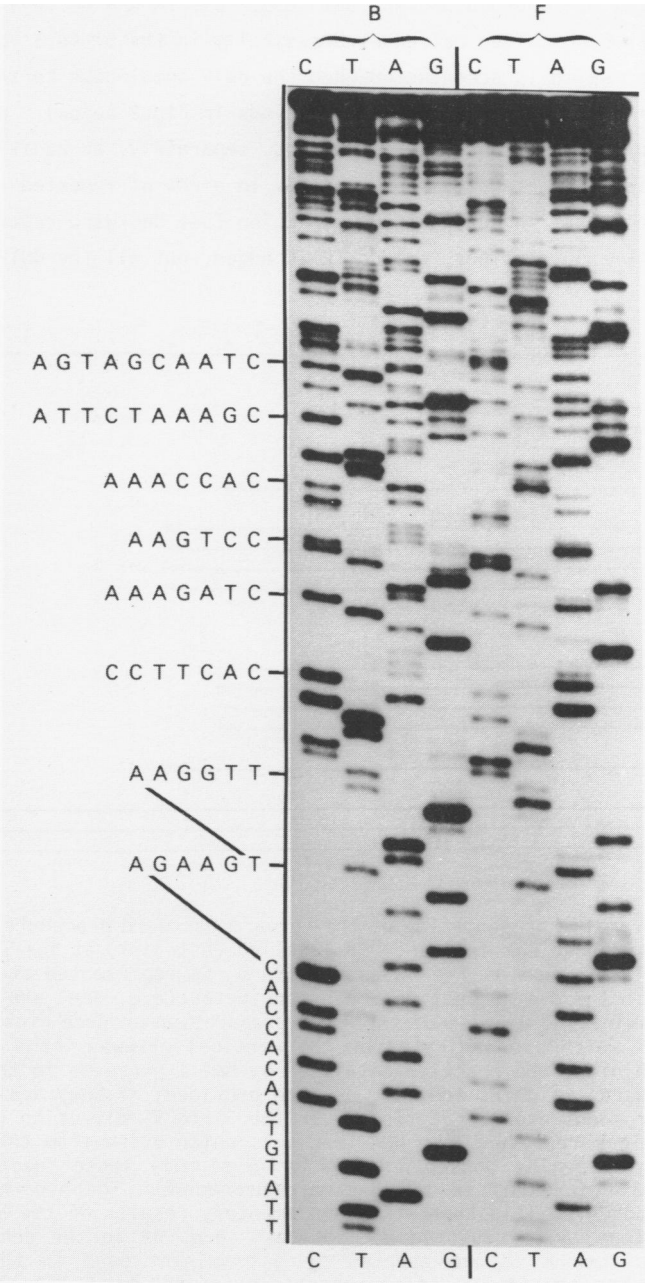


FIGURE 2. Diagramatic presentation of the Forward-Backward procedure.
A double-stranded DNA fragment ^{32}P labelled (asterisk) at one 5' end, with an hypothetical sequence on the labelled strand, is represented at the top of the figure. Pol I and a nucleotide chain terminator (e.g. ddA) are added (see Methods). Contaminating DNAases of the Pol I preparation produce nicks at random sites, some of which are indicated by the vertical arrows. From the 3' end created by each nick, the reaction catalyzed by Pol I proceeds in the 5' to 3' direction (Forward) if dNTPs (dG, dT, dC) are provided; if they are not added, the reaction proceeds exonucleolytically in the 3' to 5' direction (Backward). The numbered lines represent the DNA fragments which arise from the similarly numbered DNA nicks. The dotted lines 4 and 5 signify those reactions which proceed 5' to 3' (forward) in the Backward procedure. The hypothetical DNA sequence was chosen to illustrate the complementary results of the Forward and Backward reactions, with repeated nucleotides: e.g. AA in the center of the figure. The distal A is represented by a prominent band in the Backward reaction while the proximal A is represented by a weak band. The converse is true in the Forward reaction. Non-random nicking, for example, could modify the pattern presented in this figure.


B
F
C T A G C T A G
AGTAGCAATC
ATTCTAAAGC
AAACCAC
AAGTCC
AAAGATC
CCTTCAC
AAGGTT
AGAAGT
CACCACACTGTATT
C T A G C T A G

omitted (see also 17). Pol I binds to the 3' terminus created by a nick, catalyzes one addition of ddNTP (e.g. ddA) if appropriate, or degrades the DNA chain in the 3' to 5' direction until a nucleotide is released which can be replaced by the chain terminator (ddA or ddG, ddT, ddC, in four parallel reactions). The four Backward reactions are then subjected to electrophoresis on the same polyacrylamide gel as the Forward reactions (Fig. 3). A comparison of the two sets of reactions in parallel is directed as follows:

1) In the Backward reaction, the distal (or 3') nucleotide in a row, is the one preferentially replaced by the ddNTP, and therefore gives rises to a pattern opposite in band intensity to that of the Forward reaction. Rare exceptions to the complementary pattern of the two reactions may result from non-random nicking. For example, the distal nucleotide in a row can occasionally be the darkest in a Forward reaction (see Fig.3).

2) Authentic bands which are faint in one type of reaction, are generally much stronger in the other type of reaction, and thus can be confirmed.

3) Artefact bands are rarely present at the same position in the Backward and Forward reactions, and thus can be dismissed.

4) The Backward system more clearly indicates the number of nucleotides in a sequence than does the Forward system, presumably because of a competition between the synthetic and the 3' to 5' exonucleolytic activities of Pol I, under these Backward conditions. In a row of the same nucleotide, for example, this competition slightly reduces the frequency of ddNTP addition at the distal position, in favor of the proximal positions (not represented in Fig. 2). To accurately assess the number of nucleotides in the Forward reaction, Maat and Smith used the dNTP homologous to the ddNTP, ensuring that chain extension continued through several residues of the same nucleotide from the site of the nick. Ideal ratios of dNTP to ddNTP (about 1/500) are

FIGURE 3. Autoradiograph of a Forward-Backward DNA sequencing gel.
The sequence is from the region of the deletion of an early viable SV40 mutant, dl-2000. The 5' end labelled DNA fragment Alu I-B (Hae III)-2, extending from nucleotides 4563 to 4781 (5' to 3', Fig.4), in the system of Reddy et al. (1) was subjected to the Backward-Forward procedure followed by gel electrophoresis (see Methods). The first band at the bottom, band 1 (for a nucleotide T), is the 30th from the labelled 5' end. Interesting points of complementarity between the two patterns (B and F) are located at the following positions: position 2 (FT faint but BT dark), positions 4 and 15 (artifact in FG but not in BG), positions 11-12 (FC-11 unexpectedly lighter than FC-12, but BC pattern normal), positions 17-18 and 23-24 (FA with BA as well as FG with BG show classic complementarity), positions 34-35-36 and 47-48-49 (number of As clearer in BA than in FA), positions 41-42-43-44 (AAGT, best read by comparing F and B).

difficult to obtain. Coupling the Backward reaction to the Forward reaction bypasses the need of this dNTP addition.

Sanger and Coulson have previously described a similar dual sequencing system in which samples electrophoresed in parallel serve as a check on each other: the "plus and minus" method (14). This method, however, appears to be less reliable than is the Forward-Backward procedure. It uses single-stranded DNA in presence of randomly extended primers, dNTPs and T4 DNA polymerase (in the "plus" reaction) or the "large fragment" of E. coli DNA polymerase I (in the "minus" reaction). The "plus" reaction (similar to the Backward reaction) employs only one dNTP (e.g. dA) and the DNA fragments are terminated by this nucleotide. The "minus" reaction (similar to the Forward reaction) uses 3 of the 4 dNTPs (e.g. dG, dT, dC, - dA) and the DNA fragments are terminated by the nucleotide proximal (5') to the one which is deleted in the reaction mixture.

To clarify rare ambiguities in the F-B procedure, the following suggestions are made:

1) Occasionally, in a row of repeated nucleotides, the distal nucleotide (e.g. A) is represented by the faintest (rather than the darkest) band in the Backward reaction just as it is in the Forward reaction. For such situations we have adapted the "plus" method to double-stranded DNA (see Methods). The +A reaction produces a confirming dark band for the distal A, since the presence of dA leads to polymerization, extending to the end of the row. Either Pol I or preferably its "large fragment" can be used to locate nucleotides A or T, under the same experimental conditions as are used in the Backward reactions. Reliable results were less easily obtained with dCTP, partly because of the instability of this dNTP. The 1 mM dCTP solution was prepared immediately prior to the reaction. In spite of this precaution an artifact band was frequently apparent just distal to an authentic band. Furthermore several expected single C-bands were absent, but fortunately never the distal C-band in a row of repeated C's. Similar problems were found with dGTP; since this nucleotide is even less stable than dCTP, the purity of the commercial preparations must be regularly checked. The inherent problems in the "plus" method preclude its routine use, except to clarify specific ambiguities in the F-B procedure.

2) The correspondance between gel bands in the two complementary gel tracks (e.g. FA and BA) is occasionally obscured due to minor inequalities in sample migration. In such cases, there is an advantage in combining equal aliquots from the Forward and Backward samples (e.g. FA + BA) in the same gel track.

Ideally, 12 samples could be electrophoresed in parallel: FG,FA.FT,FC,BG,BA,BT,BC,FG+BG,FA+BA,FT+BT,FC+BC. This also facilitates reading in the upper region of the gel where bands are so close to each other (positions 200 to 350) that two repeated bands (e.g. A-A) one of which is light in the F-track (A-.) and the other, light in the B-track (.-A) are sometimes difficult to distinguish from a single band (A). In this case, the mixed (F + B) sample facilitates reading by presenting two consecutive bands of similar (dark) intensity (A-A).

Other advantages as well as disadvantages of such nick-translation methods have been already reported (e.g. the first 20 residues from the labelled end cannot be determined) and in some cases, solutions have been proposed (16, 17). We describe below sequences of SV40 mutants determined by the F-B technique.

FIGURE 4. SV40 early viable mutants with deletions in the IVS for T mRNA.
The nucleotide sequence (2) of the late DNA strand (the same sense as early mRNAs) is shown with the nucleotides numbered according to Reddy et al. (1). Arrows indicate the donor and acceptor splice sites for the large T mRNA (arrows 1 and 3) and the donor and acceptor splice sites for the small t mRNA (arrows 2 and 3). The nucleotide triplets separated by dots encode the amino-acids for the carboxy terminal end of small t antigen (boxed TAA = the termination codon). None of the 13 nucleotides underlined at the 5' and 3' ends of the large T IVS are removed from known viable deletion mutants. Deletions in mutants dl-886 and dl-885 are illustrated by gray boxes. Deletions in mutants dl-888, dl-2000 and dl-2007 are presented in brackets. The precise position at which the deletions in dl-886, dl-885 and dl-2000, occured, cannot be ascertained since there is a redundancy at their extremities (TG, TA, and TGCTT respectively).

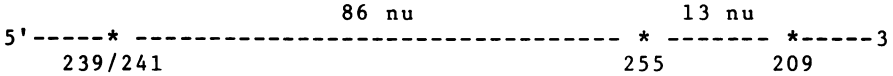
Nucleotide Sequence of SV40 Early Viable Deletion Mutants. The early region of SV40 encodes at least two spliced overlapping mRNAs which appear to be processed from a single colinear transcript (18, 19, 20). These mRNAs differ in the choice of the donor splice site (Fig. 1). The larger mRNA, from which a 65 nucleotide intervening sequence (IVS) has been removed, encodes the small t antigen, a polypeptide of 176 amino-acids which is terminated at a UAA codon prior to the splice junction (21). The smaller mRNA, which deletes a 346 nucleotide IVS extending from an upstream donor splice site to the same acceptor splice site used for small t mRNA, encodes the large T antigen (708 amino-acids; 21) a polypeptide essential for viability. Large T antigen shares 82 N-terminal amino-acids with the small t antigen (21).

Viable deletion mutants have been characterized which have deletions within the IVS of large T mRNA (8, 9, 22). The deletions vary considerably in size (23, 24). There is no portion of the large T IVS which is conserved in all the mutants, except for 13 nucleotides at the donor site and 13 nucleotides at the acceptor site (see the sequences of preferred donor and acceptor sites, 25). The viable deletions also affect the coding region of small t antigen, a protein non-essential for viral replication (8, 9, 22). While the mutations do not appear to disturb the "accumulation" of large T mRNA they affect to a variable extent the accumulation of the "truncated" small t mRNA (26). In this study we have obtained, by the Forward-Backward method, the sequence of five SV40 viable deletion mutants. These mutants were of particular interest because of their levels of small t mRNA production (26).

Monkey cells infected by dl-888 contain no detectable truncated t mRNA or t antigen. From the sequence (Fig.5) it is clear that this results from a removal of the small t mRNA donor splice site (..AAG↓GTAAAT..).

The sequence of the deletions which have little quantitative affect on the production of truncated t mRNA and t antigen (dl-2000 and dl-2007, Fig.5) are, as expected, well separated from the small t mRNA splice sites. The deletion of 9 nucleotides in dl-2000 is a subset of the 76 nucleotides removed from dl-2007.

Infection of cells by dl-885 or dl-886 (as well as by a number of other similar mutants, see 9, 20, 22) gives rise to significantly depressed amounts of t mRNA and t antigen (26). The deletions in this group are, at least in part, upstream from dl-2007 and dl-2000 (Fig.5). However, of the 17 nucleotides missing from dl-885, only two extend beyond the proximal end of dl-2007.



ng of DNA/dish	Temperature	Heteroduplex	pfu/dish
20	40	239-255	79 , 72
		241-255	83 , 81
		239-209	74 , 58
		241-209	69 , 64
		255-209	35 , 34
0.5	32	255-209	24,19,18
	40	255-209	9,2,2

FIGURE 5. Infectivity of ts//ts heteroduplexes between SV40 tsA DNAs.
At the top of the figure is illustrated the physical distance (in nucleotides) between the mutation sites (*) of four SV40 temperature-sensitive mutants in the large T antigen (tsA239 identical to tsA241). The data of Lai and Nathans (30, 34) on the infectivity of the ts//ts heteroduplexes (pfu in each dish) are presented below, for mutants tsA239, tsA241, tsA255 and tsA209. TsA 241//239 heteroduplexes gave no plaques.

Reasons for the variability in the production of small t mRNA by mutants with deletions well separated from the splice sites, are presently unclear. A number of possibilities have been considered: 1: deletions may cause alterations in the "splicing intermediate" which decrease the efficiency of the t mRNA splicing and/or enhance the efficiency of the competing T mRNA splicing (e.g. RNA conformation may play a role in splicing, 26), 2: the stability of the altered t mRNA may be decreased (e.g. during transport), 3: small, unexpected second site mutations beyond the region analyzed by nucleotide sequencing may contribute to the various phenotypes. Other possibilities have appeared less likely: 4: the wild-type small t antigen normally stimulates the small t splicing reaction (22), 5: the IVS of wild-type large T mRNA has a positive regulatory role, in trans, on the splicing or stability of the small t mRNA (22), 6: the mutant small t mRNA is quickly converted to the large T mRNA (25), 7: an alteration in translation (e.g. frame shift) due to the deletion, provokes premature termination of translation and release of the protecting translation proteins. At least some of these possibilities can be further tested experimentally.

SV40 temperature-sensitive mutants of the large T antigen. SV40 large T antigen is a viral DNA binding protein synthesized during the lytic cycle in

permissive (monkey) cells, during abortive transformation of non-permissive cells (e.g. rodent cells), and in transformants. A number of functions have been ascribed to the large T antigen (see 27 for review). These functions include the initiation of viral DNA synthesis and the regulation of the balance between early and late viral transcription in permissive cells, the induction of cellular DNA synthesis in permissive and non-permissive cells, and the initiation of stable cell transformation and the maintenance of the transformed state in certain transformants. Elucidation of most of the large T dependent functions, other than the induction of cellular DNA synthesis, has relied on the use of early SV40 temperature-sensitive (tsA) mutants (28, 29, 10). Using the sequencing technology described above, we have located in four tsA mutants a single base change which is presumably responsible for their temperature-sensitive phenotype.

SV40 tsA mutants have been mapped by the marker rescue technique (30) within a segment of the early gene region encompassed by the Hind II+III restriction endonuclease fragments H, I and B (Fig.1). All but one of the tsA mutants studied (12 of 13 mutants) are clustered within fragment H (269 nucleotides; 4 mutants) and fragment I (257 nucleotides; 8 mutants). We have sequenced the Hind II+III fragment I of four tsA mutants (tsA209, tsA239, tsA241 and tsA255) which have been localized in this segment by marker rescue. In each case, a single nucleotide change was found: C:G → T:A for tsA209, C:G → G:C for tsA239, tsA241 and tsA255. The base substitutions for tsA239 and tsA241 occur at the same position. It is not known whether these two mutants arose independently. Since hydroxylamine, the agent used to generate the mutants (10), converts C:G to T:A base pairs (see 31 for review), the origin of tsA239/241 and tsA255 is not clear.

Using the ts//ts heteroduplex infectivity assay, Lai and Nathans (30) have determined the relative proximity of the mutation sites in the four tsA mutants cited above (Fig.5). After infection of permissive cells with heteroduplexes formed between the DNA strands of two tsA mutants, they obtained a significant number of wild-type plaques at non-permissive temperature ($40^{o}C$). These results were interpreted as a consequence of a cellular excision-repair mechanism acting on mismatched base pairs, in the absence of viral DNA replication. According to this model, when a mismatch is recognized by cellular enzymes, a strand scission is made, the "defect" is excised along with adjacent nucleotides, and the deleted stretch is reconstituted utilizing the non-degraded complementary strand as a template. Several studies have suggested that two pathways for excision-repair of

damaged DNA, operate in mammalian cells (see 32 for review). One process covers sequences of small size (3 to 4 nucleotides), and the other covers sequences of larger size (20 to 100 nucleotides). They have been termed "short patch" and "long patch" pathways. However, little information is avalaible concerning correction of base mismatches in heteroduplexes. In a recent study, mouse cells were infected with heteroduplexes between two variants of polyoma virus (33). No segregation was observed between the mutations whose sites are separated by approximately 90 nucleotides. From these data, it was suggested that the two mutation sites were affected by the same excision-repair events and thus that the patch size exceeded 90 nucleotides.

Having determined the sites of base change in the mutants tsA209, tsA239/241 and tsA255, we can now compare the physical distance between two mutants with the infectivity of the corresponding heteroduplexes (see Fig.5). TsA239 and tsA241 cannot give rise to repaired wild-type progeny since the

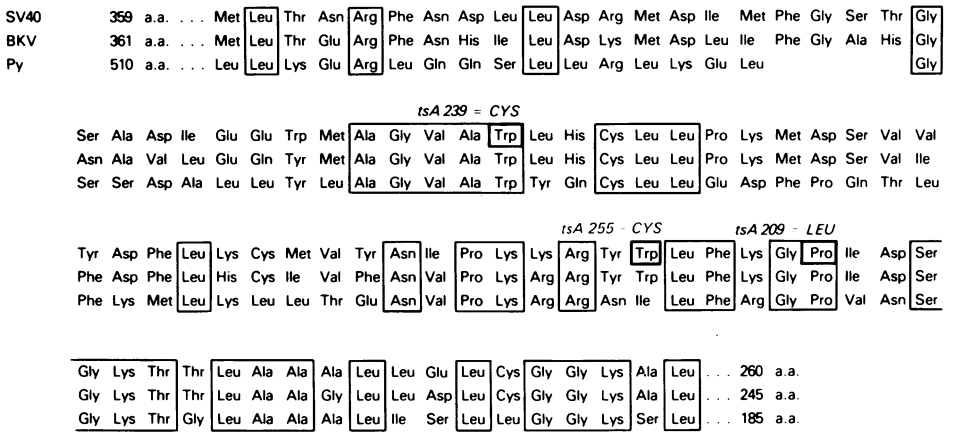


FIGURE 6. Amino-acid substitution in the T antigen of four SV40 tsA mutants.
The Hind II+III DNA fragment I of four SV40 tsA mutants (tsA209, tsA239, tsA241, tsA255) has been sequenced. The amino-acids exchanged in the mutants tsA239, tsA255 and tsA209, are presented above the wild type amino-acid (dark box). TsA241 has the same substitution as tsA239. The corresponding amino-acid sequences in the BKV (human virus), the polyoma (mouse virus) and the SV40 large T antigens are compared (identical amino-acids in light boxes). The positions of these amino-acid sequences in the large T antigens are indicated at the extremities of the sequences. The proximity test (30) suggests that two other mutants map in this region: tsA207 very close to tsA239/241 (possibly at an identical position) and tsA276 affecting another amino-acid in the sequence between tsA239 and tsA209. Thus, four tsA mutants might be clustered within 35 amino-acids (tsA239, tsA276, tsA255, tsA209).

mutants are identical. The mutations in tsA255 and tsA209 are located 87 and 101 nucleotides downstream from that of tsA239/241; infections with heteroduplexes between either tsA255 or tsA209 with tsA239/241 are approximately equally efficient. The sites of tsA255 and tsA209 are separated from each other by only 13 nucleotides; infections with heteroduplexes between tsA255 and tsA209 seem nevertheless efficient, especially when the numbers of plaques obtained at 40°C are compared to the number of plaques obtained at 32°C, the permissive temperature (34). While this suggests that some "patches" may be shorter than 15 nucleotides, a number of factors must be taken in consideration. If, for example, a single-stranded nick can be introduced between two close mismatches (e.g. separated by 13 base pairs), unidirectional long patch repair (e.g. 40 nucleotides) could give rise to a wild-type molecule (however WT molecules would be obtained from only one of the two types of heteroduplexes formed between the DNA strands of the two close mutants). The greater numbers of WT plaques obtained after infection with the heteroduplexes 209//239 or 255//239 (distant mismatches) than after infection with the heteroduplexes 209//255 (close mismatches), indicate that at least some patches exceed 15 nucleotides.

It is not clear that all patches arise from a specific nuclear process of excision-repair. It is possible, for example, that long excisions occur during the transit of the heteroduplexes from the cell surface to the nucleus.

A single amino-acid change results from the mutations in the four SV40 tsA mutants and is responsible for the large T antigen temperature sensitivity: tryptophan (TGG) to cysteine (TGC) both in tsA255 and tsA239/241, and proline (CCA) to leucine (CTA) in tsA209 (Fig. 6). In each case the substitution results in an amino-acid with a greater preference for an hydrophobic environment (35) and in each case this amino-acid appears next to a highly hydrophobic residue (leu or ile). A proline to leucine exchange has also been reported for a ts mutant of the coat protein of the tobacco mosaic virus (36). This exchange resembles, in its amino-acid environment (PRO-ile-glu → LEU-ile-glu) the change in tsA209 (PRO-ile-asp → LEU-ile-asp). A different substitution, at a proline position in a similar amino-acid environment (leu-PRO-leu-asp → leu-SER-leu-asp) has been found for a new ts mutant of the large T antigen (D.J. Cosman and M.J. Tevethia, personal comm.). In tsA255 and tsA239/241 the amino-acid alteration is tryptophan to cysteine (the tryptophan is followed in both cases by a leucine). Two ts mutants of the E. coli lac repressor, 33 amino-acids apart, have the opposite substitution, cysteine to tryptophan (37).

We presume that the region of the three clustered tsA mutations plays an important role in the T antigen functions. The fact that the substitution, in two of three cases (tsA209 and tsA239/241) occurs in an amino-acid which is conserved in both polyoma virus and BK virus, evolutionary relatives of SV40, might further substantiate this contention (see Fig.6).

ACKNOLEDGEMENTS

We thank C.-J. Lai, C. Yasumoto and M. Dobson for helpful discussions, R.G. Martin, T.E. Shenk, M.J. Sleigh and J. Alwine for providing us with the SV40 mutants, D.J. Cosman and M.J. Tevethia for communication of unpublished results, and J. Duvall, M.J. Walling and M. Konig for expert assistance.

REFERENCES

1. Reddy, V.B., Thimmappaya, B., Dhar, R., Subramanian, K.N., Zain, B.S., Pan, J., Ghosh, P.K., Celma, M.L. and Weissman, S.M. (1978) Science 200, 494-502.
2 Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herreweghe, J., Volckaert,G. and Ysebaert, M. (1978) Nature 273, 113-120.
3. Friedmann, T., Esty, A., La Porte, P. and Deiniger, P. (1979) Cell 17, 715-724.
4. Deininger,P., Esty, A., La Porte, P. and Friedmann, T. (1979) Cell 18, 771-779.
5. Soeda, E., Arrand, J.R., Smolar, N., Walsh, J.E. and Griffin, B.E. (1980) Nature 283, 445-453.
6. Yang, R.C. and Wu, R. (1979) Science 206, 456-461.
7. Seif, I., Khoury, G. and Dhar,R. (1979) Cell 18, 963-977.
8. Shenk, T.E., Carbon, J. and Berg, P. (1978) J. Virol. 18, 664-671.
9. Sleigh, M.J., Topp, W.C., Hanich, R. and Sambrook, J.F. (1978) Cell 14, 79-88.
10. Chou, J.Y. and Martin, R.G. (1974) J. Virol. 13, 1101-1109.
11. Hirt, B. (1967) J. Mol. Biol. 26, 365-369.
12. Radloff, R., Bauer, W. and Vinograd, J. (1967) Proc. Nat. Acad. Sci. USA 57, 1514-1521.
13. Maxam, A.H. and Gilbert,W. (1977) Proc. Nat. Acad. Sci. USA 74, 560-564.
14. Sanger, F. and Coulson, A.R. (1975) J. Mol. Biol. 94, 441-448.
15. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Nat. Acad. Sci. USA 74, 5463-5467.
16. Maat, J. and Smith, A.J.H. (1978) Nucl. Acids Res. 5, 45374546.
17. Hindley, J. and Phear, G.A. (1979) Nucl. Acids Res. 7, 361-375.
18. Berk, A.J. and Sharp, P.A. (1978) Proc. Nat. Acad. Sci. USA 75. 1275-1278.
19. Lai, C.-J., Dhar, R. and Khoury, G. (1978) Cell 14, 971-982.
20. May, E., Kress, M. and May, P. (1978) Nucl. Acids Res. 5, 3083-3099.
21. Reddy, V.B., Ghosh, P.K., Lebowitz, P., Piatak, M. and Weissman, S.M. (1979) J. Virol. 30, 279-296.
22. Feunteun, J., Kress, M., Gardes, M. and Monier, R. (1978) Pro. Nat. Acad. Sci. USA 75, 4455-4459.
23. Thimmapaya, B. and Shenk, T. (1979) J. Virol. 30. 668-673.
24. Volckaert, G., Feunteun, J., Crawford, L., Berg, P. and Fiers, W (1979) J. Virol. 30,674-682.

25. Seif, I., Khoury, G. and Dhar, R. (1979) Nucl. Acids Res. 6, 3387-3398.
26. Khoury, G., Gruss, P., Dhar, R. and Lai, C.-J. (1979) Cell 18, 85-92.
27. Kelly, T.J. and Nathans, D. (1977) in Advances in Virus Research, Lauffer, M.A., Bang, F.B., Moramorosch, K. and Smith, K.M., Eds., Vol. 21 pp. 86-173 Academic Press, New York.
28. Tegtmeyer, P. and Ozer, H.L. (1971) J. Virol. 8, 516-524.
29. Tegtmeyer, P. (1972) J. Virol. 10, 591-598.
30. Lai, C.-J. and Nathans, D.(1975) Virology 66, 70-81.
31. Banks, G.R. (1971) Sci.Prog.Oxf. 59, 495-497.
32. Hanawalt, P.C., Cooper, P.K., Ganesan, A.K. and Smith, C.A. (1979) Ann. Rev. Biochem. 48, 783-836.
33. Miller, L.K., Cooke, B.E. and Fried, M. (1976) 73, 3073-3077.
34. Nathans, D. and Lai, C.-J. (1977) in The Molecular Biology of the Mammalian Genetic Apparatus, Ts'o, P., Ed., pp. 269-275 Elvisier North-Holland Biomedical Press.
35. Manavalan, P. and Ponnuswamy, P.K. (1978) Nature 275, 673-674. 36.
36. Wittmann-Liebold, B., Jauregui-Adell, J. and Wittmann, H.G. (1965) Zeitschr. Naturf. Teil II 20b, 1235-1249.
37. Miller, J.H., Coulondre, C., Hofer, H., Schmeissner, U., Sommer, H., Schmitz, A. and Lu. P. (1979) J. Mol. Biol. 131, 192-222.