Cell, Vol. 11, 845-857, August 1977, Copyright © 1977 by MIT

The Nucleotide Sequence of Repetitive Monkey DNA Found in Defective Simian Virus 40

Martin Rosenberg
Laboratory of Molecular Biology

Shoshana Segal, Edward L. Kuff and Maxine F. Singer
Laboratory of Biochemistry
National Cancer Institute
Bethesda, Maryland 20014

Summary

DNA fragments containing monkey DNA sequences have been isolated from defective SV40 genomes that carry host sequences in place of portions of the SV40 genome. The fragments were isolated by restriction endonuclease cleavage and contain segments homologous to sequences in both the highly repetitive and unique (or less repetitive) classes of monkey DNA. The complete nucleotide sequence of one such fragment [151 base pairs (bp)] predominantly homologous to the highly reiterated class of monkey DNA was determined using both RNA and DNA sequencing methods. The nucleotide sequence of this homogeneous DNA segment does not contain discernible multiple internal repeating units but only a few short oligonucleotide repeats. The reiteration frequency of the sequence in the monkey genome is $>10^6$. Digestion of total monkey DNA (from uninfected cells) with endonuclease R·Hind III produces relatively large amounts of discrete DNA fragments that contain extensive regions homologous to the fragment isolated from the defective SV40 DNA.

A second fragment, also containing monkey sequences, was isolated from the same defective substituted SV40 genome. The nucleotide sequence of the 33 bp of this second fragment that are contiguous to the 151 bp fragment has also been determined.

The sequences in both fragments are also present in other, independently derived, defective substituted SV40 genomes.

Introduction

Several defective variants of simian virus 40 (SV40) with genomes that contain DNA sequences derived from the host (monkey) genome have been reported in recent years (Lavi and Winocour, 1972; Martin et al., 1973; Lavi et al., 1973; Rozenblatt et al., 1973; Brockman and Nathans, 1974; Frenkel, Lavi and Winocour, 1974; Lee, Brockman and Nathans, 1975; Oren, Kuff and Winocour, 1976; Davoli et al., 1977; Rao and Singer, 1977b). The genomes of the defective variants are closed circular duplex molecules (DNA I), and typically consist of several tandem repeats of a DNA segment containing the monkey sequences and a portion of the wild-type SV40 DNA sequences including the origin of replication. The monkey sequences in a given defective may be derived from either the highly repetitive portion of the monkey genome or from the less repetitive or unique portion or from both. Furthermore, there is evidence indicating that the monkey sequences present in independently derived defective variants are not a random selection from the monkey genome. Rather, certain sequences, both of the repetitive and unique classes, are more likely to occur than are others (Frenkel et al., 1974; Oren et al., 1976). It is presumed that recombinational events between monkey DNA and infecting or replicating viral DNA give rise to these "substituted" variants, but the mechanism of the recombination is unknown. It may involve chromosomal DNA or extrachromosomal DNA known to exist in monkey kidney cells in tissue culture (Smith and Vinograd, 1972). The recombination may reflect site-specific integration into the cellular DNA and subsequent excision, but it is also possible that nonspecific recombination is involved.

Segments containing the monkey sequences and few, if any, SV40 sequences can be isolated from purified, substituted defective genomes by cleavage with restriction endonucleases (Rozenblatt et al., 1973; Lee et al., 1975; Segal et al., 1976; Rao and Singer, 1977b). The substituted defective DNA called CVP8/1/P2 (Eco RI res) DNA I (Figure 1) yields two fragments of interest (Rao and Singer, 1977b). [Restriction endonucleases and fragments of DNA derived from cleavage with restriction endonucleases are abbreviated according to the recommendations of Smith and Nathans (1973).] The (Hind II/Hind III)-E fragment is approximately 3.1% of a wild-type SV40 genome in length and contains monkey DNA of high reiteration frequency; the (Hind II/Hind III)-C fragment is approximately 4.3% of a wild-type genome in length and does not hybridize to filters containing either SV40 or monkey DNA, and may therefore consist predominantly of monkey DNA from the low repetitive or single-copy class (Segal et al., 1976; Rao and Singer, 1977b). Since the CVP8/1/P2 (Eco RI res) DNA I consists of a 4 fold tandeom repeat, each fragment occurs 4 times within the defective genome. Fragments (Hind II/Hind III)-C and (Hind II/Hind III)-E are contiguous within each tandem repeat, and each has one end generated by endonuclease R·Hind II and one generated by endonuclease R·Hind III; the Hind III cleavage site separates fragments C and E. No SV40 sequences are detectable in either fragment by filter hybridization. We have undertaken experiments designed to characterize the se-

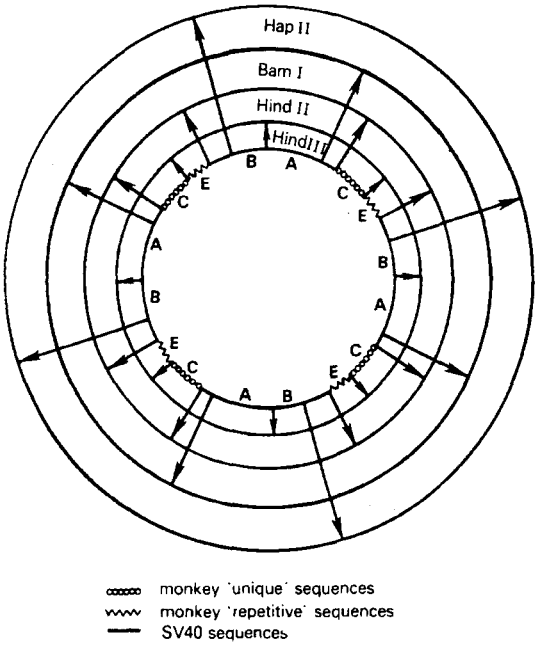

Figure 1. A Schematic Drawing of the Structure of the Substituted Defective Genome CVP8/1/P2 (Eco RI res) DNA I

The figures shows the sites sensitive to endonucleases R·Hind II, R·Hind III, R·Bam I and R·Hap II, as well as (inner circle) the fragments generated by combined cleavage with endonuclease R·Hind II and R·Hind III (Rao and Singer, 1977b). Fragments (Hind II/Hind III)-C (0000) and (Hind II/Hind III)-E (∿∿∿) are indicated.

quences in (Hind II/Hind III)-C and -E in a variety of ways.

Radioactively labeled cRNA copies of (Hind II/Hind III)-E have been prepared and used to study the location of the sequences in the fragment in African green monkey chromosomes by in situ hybridization (Segal et al., 1976). The results indicated many possible chromosomal origins for the (Hind II/Hind III)-E sequences, both centromeric and noncentromeric. The sequence was detected in the arms of between 9 and 11 chromosomes; this location is consistent with an interspersion of repetitive and unique sequences in the monkey genome as has been observed in a variety of other eucaryotic DNAs (Davidson and Britten, 1973).

In the present paper, we report the complete nucleotide sequence of (Hind II/Hind III)-E fragment (151 residues), as well as the sequence of the first 33 nucleotide residues of (Hind II/Hind III)-C from the endonuclease R·Hind III site that separates it from (Hind II/Hind III)-E. The monkey sequences in fragment E are shown to have a reiteration frequency of approximately 1.6×10^6 in the monkey genome. It was recently reported that cleavage of total monkey DNA with endonuclease R·Hind III results in a series of discrete DNA fragments resolvable on polyacrylamide or agarose gels (Grüss and Sauer, 1975). Our results indicate that the monkey sequence in the (Hind II/Hind III)-E fragment hybridizes to each of the fragments obtained directly from monkey DNA. Finally, we present hybridization data indicating that sequences in both (Hind II/Hind III)-E and (Hind II/Hind III)-C are found within the genomes of other independently derived, substituted SV40 defectives.

Results

The Complete Nucleotide Sequence of (Hind II/Hind III)-E Fragment

Both RNA and DNA sequencing techniques were used to determine an unambiguous sequence for the 151 residues of the (Hind II/Hind III)-E fragment (Figure 2). The RNA sequencing methods involved preparing and characterizing a cRNA transcript of the DNA fragment, and analyzing the oligonucleotide products generated from both complete and partial digests of the cRNA with T1 and pancreatic RNAase. As previously described (Segal et al., 1976), the major transcription product was an essentially full-length copy of one strand of (Hind II/Hind III)-E fragment. A second (minor) transcript, readily separable from the major product on polyacrylamide gels, yielded oligonucleotides consistent with its being a covalently linked RNA copy of both strands of (Hind II/Hind III)-E. Analysis and comparison of the T1 and pancreatic oligonucleotide products obtained from these two transcripts designated two separate sets of oligonucleotides that derived from each of the two complementary strands of the DNA fragment (Figure 3). As shown in Figure 2, the complementary products were used to help specify the relative order of certain of the oligonucleotides from the major transcript. These data, in conjunction with the data obtained from the partial T1 ribonuclease digestions (see Figure 2, RNA), allowed deduction of a unique sequence for most of the fragment, although some ambiguities remained.

Direct DNA sequence analysis was also carried out on (Hind II/Hind III)-E, using both the techniques of partial snake venom analysis (Sanger et al., 1973) and the new chemical procedures introduced by Maxam and Gilbert (1977). For both procedures, (Hind II/Hind III)-E was labeled with ^{32}P at one or the other of its 5′-hydroxy termini, as described in Experimental Procedures.

The dimethyl sulfate-hydrazine (DMS-Hz) (Maxam and Gilbert, 1977) technique for direct DNA sequencing was applied to (Hind II/Hind III)-E labeled separately at either its Hind II or Hind III end, and the nucleotide sequences were deduced from a comparison of the various polyacrylamide gels obtained. Representative examples of these

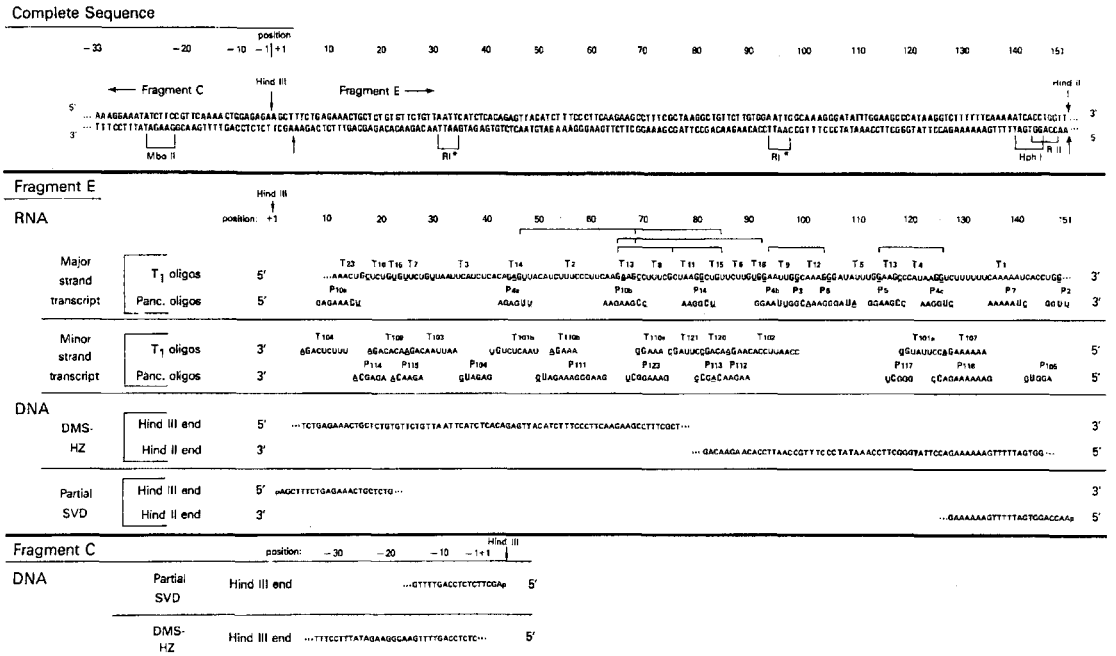

Figure 2. The Derivation of the Nucleotide Sequence of (Hind II/Hind III)-E Fragment and Part of (Hind II/Hind III)-C Fragment

All nucleotide positions are numbered from the R·Hind III site separating the C and E fragments (−1/+1). RNA (E fragment only): numerical designation of T1 (RNAase (T) and pancreatic RNAase (P) oligonucleotides associated with either the major strand or minor strand transcript are those indicated in Figure 3. Only oligonucleotides which were actually used to deduce the relative order of the T1 RNAase products from the major strand transcript are depicted. Partial T1 RNAase digestion products (see Experimental Procedures) obtained from the major strand transcript are designated by brackets (⌐). Nearest-neighbor nucleotides to each oligonucleotide product are underlined (for example, T23, AAACUGC). The nucleotide sequence of each oligonucleotide product was determined by standard RNA sequencing techniques (see Experimental Procedures). These sequences and their relative order within the overall sequence were confirmed by direct DNA sequencing methods. DNA: the sequence of nucleotides depicted for each of the DNA sequencing methods (that is, DMS-Hz; partial SVD) denote only those residues which could be unambiguously determined from that technique at the level of single-nucleotide resolution.

gel patterns are shown in Figure 4, and the derived sequences are indicated in Figure 2 (DNA:DMS-Hz). Partial snake venom analysis of the same end-labeled fragments allowed correlation of the first identifiable nucleotide residue in the DMS-Hz gel patterns with a particular nucleotide position within the fragment (Figure 2, DNA:SVD). Analysis of the products of the partial snake venom digestions by both two-dimensional "homochromatography" (Brownlee and Sanger, 1969) (Figure 5) and one- and two-dimensional paper electrophoresis (Rf values given in Table 1) gave nucleotide sequences for 20–25 nucleotide residues at each end of the fragment. These sequences overlapped the sequences obtained by both the RNA and DNA sequencing techniques, and also confirmed the sequences predicted from the known specificity of the restriction endonuclease cleavage sites at the termini of the fragment (Old, Murray and Roizes, 1975; Smith, 1974).

The results summarized in Figure 2 indicate that the RNA and DNA sequencing methods gave completely consistent data. Although each approach generated certain sequence ambiguities, in conjunction these methods allowed deduction of an unambiguous sequence of 151 nucleotide residues for (Hind II/Hind III)-E. It is important to point out, for assessing the accuracy of the sequencing data, that in addition to using the RNA and DNA sequence information separately, as confirmatory of one another, the techniques were also used to complement one another. Given a relatively easily obtainable catalogue of all the T1 and pancreatic RNAase products obtained from the RNA analysis, they may be ordered within the sequence by reference to the gels obtained from the DNA sequencing method. Knowing the sequence and the nearest-neighbor nucleotide of a particular T1 or pancreatic RNAase product predicts a very specific pattern on the DNA sequencing gels. These patterns can readily be picked out and the oligonucleotides ordered far beyond the single residue resolution of the gel. Thus when used in conjunction, the RNA and DNA sequencing methods allow for greater flexibility and accuracy than does either one alone.

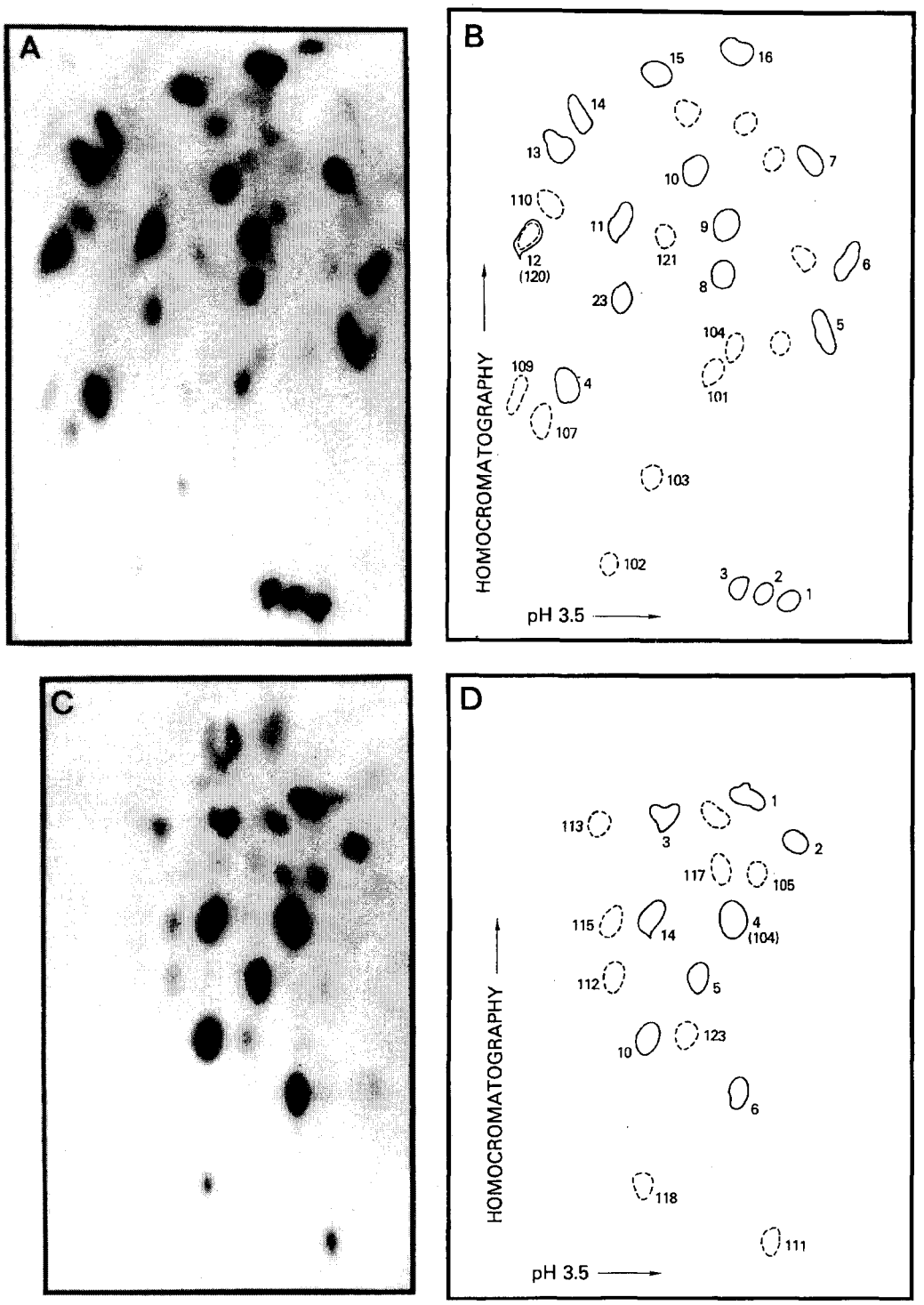

Figure 3. T1 RNAase Fingerprint (A, Autoradiograph; B, Schematic Sketch) and Pancreatic RNAase Fingerprint (C, Autoradiograph; D, Schematic Sketch) of Total cRNA Transcription Product Labeled with α-^{32}P-GTP

First dimension separation: electrophoresis on Cellogel in 8.0 M urea at pH 3.5 (Sanger et al., 1965). Second dimension: ascending thin-layer homochromatography on DEAE-cellulose (9/1, cellulose/DEAE-cellulose, 40 × 20 cm) (Brownlee and Sanger, 1969). Closed circles (○) depict oligonucleotides derived from the major strand transcript; stippled circles (·:·) depict oligonucleotides derived from the minor strand transcript (see Figure 2).

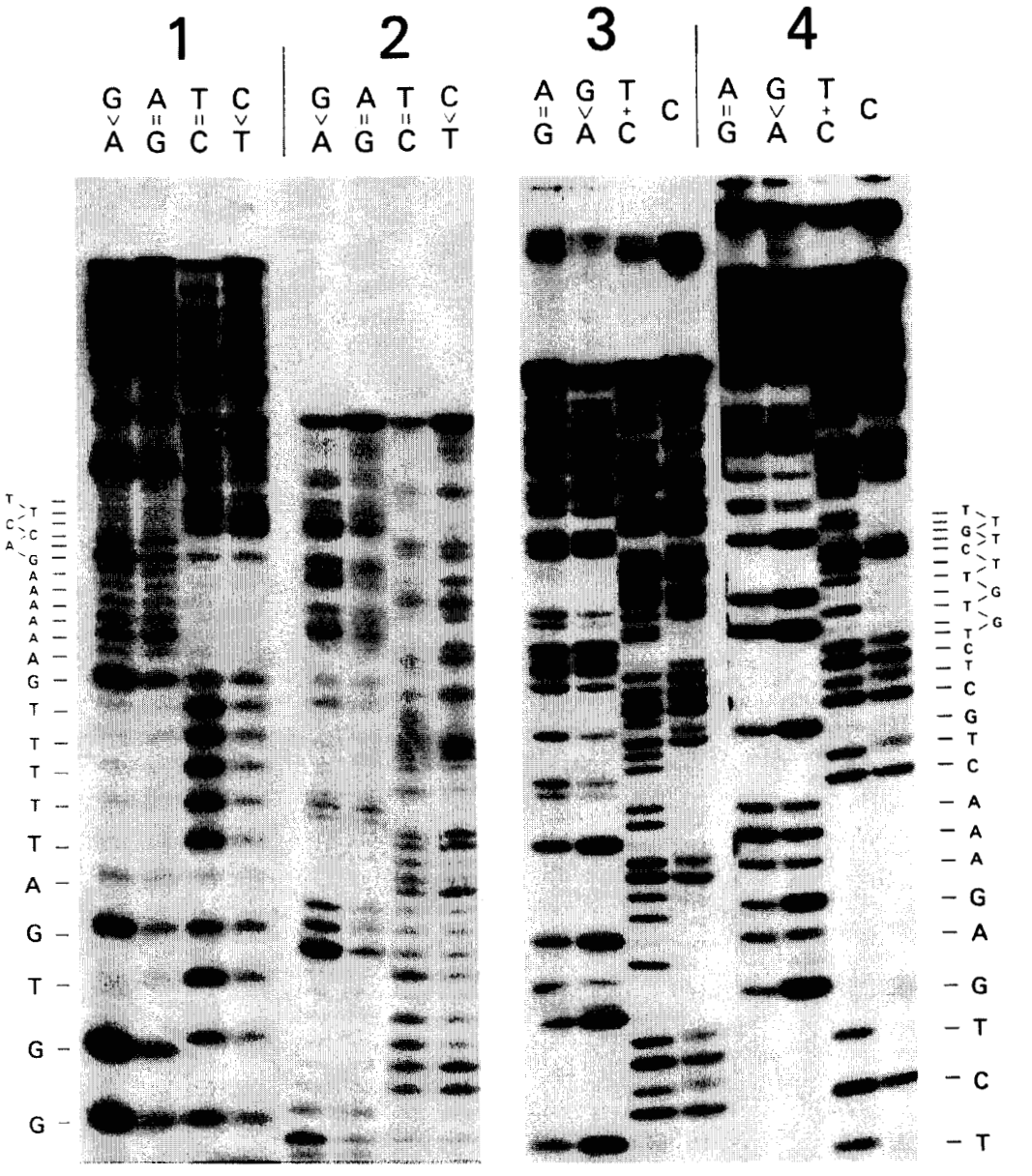

Figure 4. Representative Autoradiographs of DNA Sequencing Gels Obtained from the DMS-Hz Chemical Sequencing Methods (Maxam and Gilbert, 1977) as Applied to (Hind II/Hind III)-E Fragment

(1) ^{32}P-labeled at the Hind II end; first discernible residue (at the bottom) is the G 6 nucleotides from the end at position 146 in the sequence shown in Figure 2; (2) identical to (1), except for longer electrophoresis such that first discernible residue is 22 nucleotides from the end, at sequence position 130; (3) ^{32}P-labeled at the Hind III end; first discernible residue (at the bottom) is at position 6 (Figure 2); (4) identical to (3), except that first discernible residue is at position 17.

Partial Nucleotide Sequence of (Hind II/Hind III)-C Fragment

A sequence of 33 nucleotide residues at the Hind III end of (Hind II/Hind III)-C fragment [contiguous with (Hind II/Hind III)-E] was determined using only the direct DNA sequencing techniques (Figures 2 and 5B). The fragment was labeled with ^{32}P at the Hind III terminus (see Experimental Procedures)

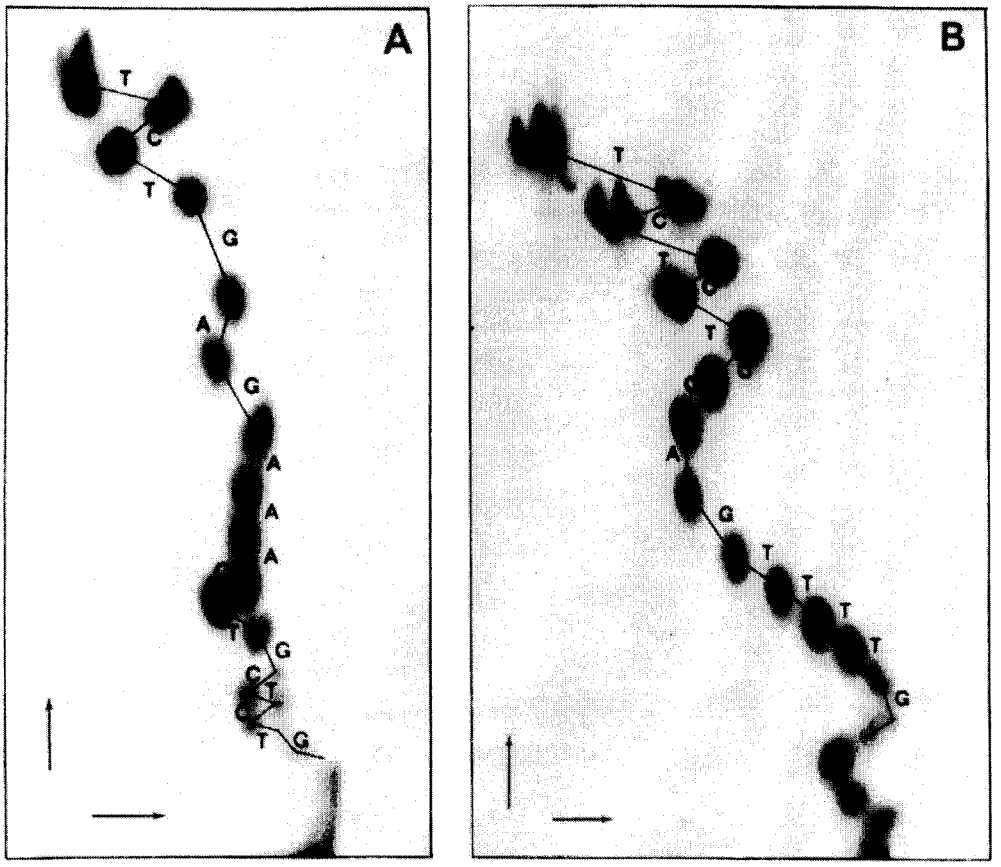

Figure 5. Autoradiographs of Two-Dimensional Fractionation (as Described in Experimental Procedures and Identical to That Used in Figure 3) of the Products Resulting from Partial Digestion with Snake Venom Exonuclease of (Hind II/Hind III)-E Fragment (A) and (Hind II/Hind III)-C Fragment (B) Labeled with ^{32}P-at Their Hind III 5′ Termini

and analyzed by partial snake venom digestion (Figure 2, SVD, and Figure 5B) and the DMS-Hz sequencing technique (Figure 2, DMS-Hz).

The Occurrence of Sequences in (Hind II/Hind III)-E in the Monkey Genome – Cot Analysis

^{3}H-labeled (Hind II/Hind III)-E fragment was allowed to reanneal in the presence of heterologous DNA (E. coli) or DNA prepared from the BSC-1 line of monkey kidney cells, and the percentage of (Hind II/Hind)-E in double-stranded form was determined by chromatography on hydroxylapatite (Figure 6). (Hind II/Hind III)-E alone reannealed with a $Cot_{1/2}$ of 1.7×10^{-4} mole-sec/l; the reaction was accelerated about 7 times in the presence of a 126 fold weight excess about 7 times in the presence of a 126 fold weight excess of BSC-1 DNA. Reactants harvested in the plateau regions of the two reannealing curves were almost equally resistant to digestion with single-strand-specific nuclease S1 (82% in the case of fragment E alone, and 72% for the mixture with BSC-1 DNA). In addition, the melting of reannealed (Hind II/Hind III)-E itself was compared with the melting of duplexes between (Hind II/Hind III)-E and BSC-1 DNA (Figure 6, insert). Melting was sharp in both cases. The Tm value for the duplex between (Hind II/Hind III)-E and BSC-1 DNA was about 1°C lower than was that of the reanneled fragment itself. Both the S1 nuclease data and the melting curves indicate extensive matching between (Hind II/Hind III)-E and the homologous sequences in the cellular DNA, but suggest the possibility of a small percentage of mismatching. The observed Tm values, 83–84°C, are consistent with the fact that (Hind II/Hind III)-E contains 59.6% A·T bp (Figure 2).

From the acceleration in reannealing of (Hind II/Hind III)-E by a known quantity of BSC-1 DNA (see above), it can be calculated that fragment sequences may represent as much as 4.8% of the cell genome. Since monkey cells contain about 5×10^{9} bp of DNA, (Hind II/Hind III)-E sequences may

Table 1. Rf Values of Oligonucleotide Products of Partial Snake Venom Phosphodiesterase Digestion of (Hind II/Hind III)-E

Oligonucleotide Product	Rf Values [a]	
	pH 3.5	pH 1.7
*pdA	1.90	2.25
*pdApdG	0.78	2.00
*pdApdGpdC	0.42	1.85
*pdApdGpdCpT	0.20	0.72
*pdApdGpdCpTpT	0.09	0.25
*pdApdGpdCpTpTpT	0	0.08
*pdApdA	1.06	2.05
*pdApdApdC	0.72	2.00
*pdApdApdCpdC	0.47	1.95
*pdApdApdCpdCpdA	0.15	1.80
*pdApdApdCpdCpdApdG	0.05	1.22
*pdApdApdCpdCpdApdGpdG	0	0.56
*pdApdApdCpdCpdApdGpdGpT	0	0.21

[a] Denotes Rf values = mobility relative to blue marker dye, xylene cyanol FF, determined by electrophoresis on DEAE paper at either pH 3.5 (2.5 hr at 250 ma) or pH 1.7 (4.5 hr at 250 ma) (Brownlee, 1972); appropriate markers were obtained from other DNA fragments of known sequence or purchased from Collaborative Research.

The first six oligonucleotides listed were obtained from the digestion of (Hind II/Hind III)-E labeled with ^{32}P at the 5'-hydroxyl of the Hind III terminus. The last seven oligonucleotides listed were obtained from fragment labeled at the 5'-hydroxyl of the Hind II terminus. * Denotes a radioactive phosphate; dA, dC, dG and T are the four deoxyribonucleosides.

comprise as much as 2.4×10^8 bp per cell. This value corresponds to a reiteration frequency of about 1.6×10^6 in the monkey genome.

The Occurrence of Sequences in (Hind II/Hind III)-E in the Monkey Genome—Hybridization to Fragments of Monkey DNA on Filters

As previously observed (Grüss and Sauer, 1975; F. L. Brown, P. R. Musich and J. J. Maio, personal communication; M. Israel and M. Martin, personal communication), cleavage of total DNA from African green monkey kidney cells with endonuclease R·Hind III and analysis of the products on agarose or polyacrylamide gels yielded several prominent fragments of defined size, as well as a mass of unseparable fragments of relatively high molecular weight (Figure 7). Additional cleavage of the monkey DNA fragments with Hind II [which is necessary to generate (Hind II/Hind III)-E from the defective SV40] did not alter the Hind III digestion pattern seen in Figure 7 in a detectable manner (not shown).

At least five discrete bands were usually visible. The smallest discrete class of fragments [called AGMr (Hind III)-1] is also the predominant class and is slightly larger than the 151 bp (Hind II/Hind III)-E marker (Figure 7). Other experiments using the endonuclease R·Hind III fragments of wild-type SV40 as markers indicated that AGMr (Hind III)-1 is

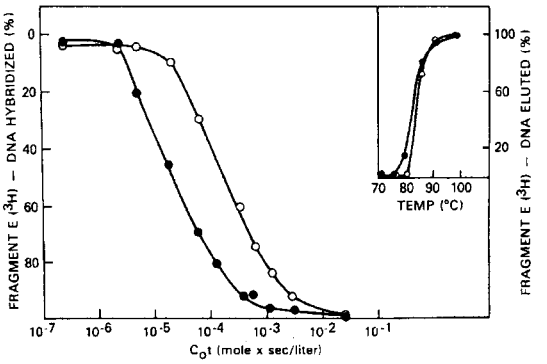

Figure 6. Reassociation of (Hind II/Hind III)-E

^{3}H-labeled (Hind II/Hind III)-E fragment (349,600 cpm/μg, 4.8 ng/ml) was allowed to reanneal at 68°C in 0.12 M sodium phosphate (pH 6.8) in the presence of 605 ng/ml of denatured sonicated BSC-1 DNA (●) or E. coli DNA (○). The extent of hybridization at each time was analyzed by hydroxyapatite chromatography. The abscissa is the molar concentration of nucleotide residues of (Hind II/Hind III)-E fragment times seconds. The insert shows a thermal denaturation of samples first reannealed under the same conditions as above to a Cot value of 10^{-2}.

approximately 170 bp in length (preliminary sequence analysis indicates a chain length of 172 bp; unpublished experiments). The next largest class AGMr (Hind III)-2 is about 350 bp in length, and the remaining fragments appear to be consecutively larger multiples of the size of AGMr (Hind III)-1.

The products of the digestion of BSC-1 DNA with endonuclease R·Hind III were transferred from the agarose slab gel (see Figure 7) directly to nitrocellulose strips and hybridized against ^{32}P-labeled cRNA from (Hind II/Hind III)-E fragment as described in Experimental Procedures. As shown in Figure 7, each of the five discrete bands detected by staining with ethidium bromide hybridized with the cRNA; no other hybrid was detected. Thus all or part of the sequences contained in the (Hind II/Hind III)-E fragment of the defective SV40 genome also occur in each size class of discrete monkey Hind III fragments. Although all the classes of discrete monkey fragments contain some sequences homologous to at least a portion of (Hind II/Hind III)-E, the nature of the additional sequences in the fragments longer than AGMr (Hind III)-1 is not known.

Sequences Present in (Hind II/Hind III)-E and -C Fragments Are Present in Other Substituted Defective SV40 Variants

Previous work (Rozenblatt et al., 1973; Frenkel et al., 1974; Oren et al., 1976) indicated that the DNA associated with the defective SV40 variant CVB/1/P4 contains segments homologous to both highly reiterated and infrequently reiterated or unique monkey sequences. Furthermore, using cRNA

Figure 7. DNA Fragments Obtained from the Digestion of BSC-1 DNA with Endonuclease R·Hind III and Hybridization of the Fragments with (Hind II/Hind III)-E

BSC-1 DNA (3 μg, total volume 50 μl) was digested with endonuclease R·Hind III, and the material was electrophoresed on a 1.4% agarose gel (slot B). A mixture of defective SV40 DNAs (CVP8/1/P2) was digested with endonuclease R·(Hind II/Hind III) to provide a marker of (hind II/Hind III)-E fragment which is the shortest fragment in the digest (slot C) (Rao and Singer, 1977b). The gel was stained with ethidium bromide and photographed (slots B and C). The BSC-1 digest was transferred to nitrocellulose sheets and hybridized with ^{32}P-labeled cRNA to fragment E (spec. radioact. about 1 × 10^8 dpm/μg), and an autoradiogram of the nitrocellulose strip was made (slot A).

probes copied from the monkey sequences present in CVB/1/P4 and other independently derived defective viral DNAs, Oren and his co-workers (1976) demonstrated that the same sequences occur in several of the independently isolated populations of substituted SV40 variants. It seemed probable that the same sequences might be present in CVP8/1/P2 (Eco RI res) DNA 1, since it shares a common ancestry with CVB/1/P4 (Rao and Singer, 1977a). The data in Table 2 indicate that sequences in (Hind II/Hind III)-E fragment hybridize significantly to several of the independent SV40 variants previously described. We consider that hybridization of >5% of the radioactive probe represents a clear indication of positive homology. It is especially noteworthy that the two defectives 776-P11 and 1103 contain sequences homologous to those of (Hind II/Hind III)-E, since these were derived from passage of wild-type strain 776 and not wild-type strain 777, as was CVP8/1/P2 (Eco RI res). No hybridization of the cRNA to (Hind II/Hind III)-E was detected with the defective variants 1101 or DAR-d5, which only contain monkey sequences that are infrequently reiterated (Lee et al., 1975; Davoli et al., 1977). We conclude that the (Hind II/Hind III)-E sequences are frequently incorporated into substituted SV40 defectives.

Table 2. Hybridization of (Hind II/Hind III)-E Fragment Sequences to Independently Derived, Substituted Variants of SV40

Experiment	DNA on Filter		% Input Hybridized
	Type	μg	
1	CVP8/1/P2	0.5	16
	CVG/1/P10	1.25	29
	777-T8	1.25	13
	CVB/4/P8	0.5	12
	CVB/2/P5	1.25	4
	RH911	0.5	3
	CV1/2/P7	1.25	1
	CV1/1/P5	1.25	0.4
2	BSC-1	18	65
	777(CVB)	1.8	0.8
	776-P11	1.2	25
3	777(CVB)	6	2
	CVP8/1/P2	1.25	15
	DAR-d5	1.5	0.8
4	BSC-1	12.5	24
	1101	1.25	23
	1103	1.25	10

The filter hybridization procedure is described in Experimental Procedures. In experiments 1, 2 and 3, the input was ^{32}P-cRNA (7.5 × 10^4, 4.5 × 10^4 and 3.0 × 10^4 cpm, respectively). In experiment 4, the input was (Hind II/Hind III)-E fragment (DNA) labeled with ^{32}P at the Hind II 5′-hydroxyl terminus (9.5 × 10^3 cpm). When first prepared, the cRNA had a specific radioactivity of about 1.1 × 10^8 cpm/μg. The inputs of cRNA correspond, in each instance, to approximately 0.7 ng of RNA. The defective CVP8/1/P2 is the source of (Hind II/Hind III)-E fragment. The other defectives have been described previously by Oren et al. (1976), except for DAR-d5, which was described by Davoli et al. (1977), and 1011 and 1103 which were described by Lee et al. (1975). The viral stocks 777-T8, RH 911 and 776-P11 were derived from the indicated strains of SV40, and have been carried at high multiplicity of infection over a period of years in the laboratory of E. Winocour (Oren et al., 1976). 777(CVB), a plaque-purified strain 777 SV40, and DNA isolated from BSC-1 cells served, respectively, as negative and positive controls for hybridization. The percentage of the input radioactivity bound to blank filters (no DNA) was subtracted from the values shown. In experiments 1, 2 and 3, the blank values were <0.4% of the input; in experiment 4, the blank values were <1.5%. It should be noted that the cRNA-DNA hybridization procedure (experiments 1, 2 and 3) affords lower blank values and greater efficiency of hybridization than the DNA-DNA procedure (experiment 4).

Discussion

The mechanisms responsible for the formation of defective SV40 genomes that are substituted with host (monkey) DNA are not known. Furthermore, those properties of the substituted genomes that

account for their perpetuation during multiple serial passages of defective viral stocks are also not known. It is clear from the available data (reviewed in the Introduction) that such molecules are generated during permissive infection of monkey kidney cells in tissue culture by SV40 and, after high multiplicity passaging, may come to represent a substantial portion of the viral genomes synthesized during infection. The formation of these genomes and their subsequent evolution during serial high multiplicity passage (Brockman, Lee and Nathans, 1973; Martin et al., 1973; Rozenblatt et al., 1973) must involve recombinational events within the SV40 genome, as well as between SV40 and monkey genome DNA. These events may include integration of infecting wild-type SV40 genomes into the chromosomal DNA of the cells and subsequent faulty excision, or exchanges of DNA segments between the DNA of the two genomes in the absence of true integration. It is also possible that they arise by recombination between the wild-type SV40 genomes and nonchromosomal host DNA molecules such as the small polydisperse circular DNA molecules known to occur in uninfected cells of the BSC-1 monkey kidney line (Smith and Vinograd, 1972).

Both low and high reiteration frequency classes of monkey DNA occur in the substituted SV40 defective genomes. Winocour and his co-workers (Frenkel et al., 1974; Oren et al., 1976) have demonstrated that in a variety of independently isolated defectives, a nonrandom selection from the total monkey genome is present. A set of highly reiterated monkey sequences and a set of sequences of low reiteration frequency were detected in several independently isolated defective variants. Other substituted variants did not appear to contain the same monkey sequences. With regard to the reiterated sequences, the data in this paper confirm the earlier observation. Sequences contained in fragment (Hind II/Hind III)-E of the defective SV40 genome CVP8/1/P2 (Eco RI res) DNA I are present in several of those defectives previously characterized as sharing common monkey reiterated sequences. In addition, a completely separate isolate, variant 1103 (Lee et al., 1975), also has (Hind II/Hind III)-E sequences. While it might appear that even random recombination would frequently result in the incorporation of a sequence reiterated 1.6×10^6 times in the monkey genome, in fact, the (Hind II/Hind III)-E sequences represent 10% of the genome and yet are found in about half the substituted variants tested. Thus particular monkey sequences containing at least portions of (Hind II/Hind III)-E fragment are commonly found in substituted variants. Selection of these sequences may occur during the initial recombination events or during the subsequent replication of DNA.

The availability of well characterized, substituted SV40 variants offers an opportunity to elucidate the structures involved in recombination and, perhaps, thereby to increase understanding of the corresponding mechanisms. The incorporation of the monkey sequences present in (Hind II/Hind III)-E fragment into the defective SV40 variant has also provided a convenient tool for the isolation and characterization of a portion of the monkey genome. We describe here the entire nucleotide sequence of the (Hind II/Hind III)-E fragment isolated from the defective CVP8/1/P2 (Eco RI res) DNA I, as well as that of 33 residues in (Hind II/Hind III)-C that are contiguous with (Hind II/Hind III)-E in the defective genome. The direct DNA sequencing data indicate that (Hind II/Hind III)-E and (Hind II/Hind III)-C are each single homogeneous sequences, and not mixtures of sequences as might have been expected if the molecules of CVP8/P2 (Eco RI res) DNA I contained a variety of host sequences.

(Hind II/Hind III)-E fragment is 151 residues long, has an A–T content of 59.6%, and is bounded by an endonuclease R·Hind III restriction site at one end and an endonuclease R·Hind II restriction site at the other. Although the (Hind II/Hind III)-E sequence has a high reiteration frequency in the monkey genome, the nucleotide sequence is markedly different from the sequences reported previously for reiterated satellite and spacer regions of eucaryote genomes (Brownlee, Cartwright and Brown, 1974; Biro et al., 1975; Endow, Polan and Gall, 1975; Salser et al., 1976). The latter are characterized by extensive internal repetitions of relatively short segments (6–15 residues). On the other hand, inspection of the sequence of (Hind II/Hind III)-E (Figure 2) does not indicate extensive internal repetition, although there are a few short repeats. Recent data (Roizes, 1976) suggest that the highly repetitive bovine satellite DNA I (Botchan, 1974) may also contain rather long basic internal repeat units.

There are several short regions of 2 fold rotational symmetry within the fragment (Hind II/Hind III)-E sequence. The longest contain a total of 10 and 12 symmetrically positioned residues, and occur at positions 4–14 and 68–84, respectively (Figure 2). The longer region of symmetry (positions 68–84) contains an unsymmetrical core sequence of five residues (positions 74–78), and potentially gives rise to a stem and loop structure in the single-stranded cRNA copy of this region. In fact, under the conditions used for partial endonuclease digestion of the cRNA transcript, this region was found to be most resistant to ribonuclease attack (Figure 2).

Most strikingly, the sequence contains relatively long stretches of asymmetrically distributed, alternating arrays of purine- and pyrimidine-rich seg-

ments. Examination of either strand indicates that most of these stretches vary from between six and twelve consecutive purine or pyrimidine residues, and have little, if any, exact sequence homology among them.

There is also an interesting array of sequences at the R·Hind II end of (Hind II/Hind III)-E. The G at position 149 is in the recognition site GTTPuAC for endonuclease R·Hind II (Smith, 1974). The same G residue is contained within a recognition site for endonuclease R·Eco RII – that is, CCTGG, residues 145–149 (Bigger, Murray and Murray, 1973; Boyer et al., 1973). Similarly, the two C residues at the 5′ end of the endonuclease R·Eco RII recognition site are the last two residues of a recognition site for the endonuclease R·Hph, TCACC (residues 142–146) (Kleid et al., 1976). There is no apparent reason to expect that sequences for three different bacterial restriction endonucleases should occur in a row in this fragment. (Hind II/Hind III)-E also contains two sites susceptible to cleavage by endonuclease R·Eco RI under the relatively nonspecific conditions termed RI* (Polisky et al., 1975). These sites are the sequences AATT at positions 32–35 and 95–98.

Digestion of total monkey DNA with endonuclease R·Hind III yields a series of well defined DNA fragments of various molecular weights (Grüss and Sauer, 1975). As shown here, all size classes of these fragments contain sequences that hybridize with (Hind II/Hind III)-E. The monkey fragments are of various molecular weights, the smallest and most abundant [AGMr (Hind III)-1] being about 172 bp in length. The other fragments are of increasing size and appear to be multimers of AGMr (Hind III)-1 in length. It is not known whether these longer fragments contain multiple copies of all or of only a part of the (Hind II/Hind III)-E sequence. However, studies on the nucleotide sequence of AGMr (Hind III)-1 (M. Rosenberg, H. Rosenberg and M. F. Singer, manuscript in preparation) indicate that a sequence of 124 residues from one end of AGMr (Hind III)-1 is identical to the 124 residue sequence from the endonuclease R·Hind III terminus of (Hind II/Hind III)-E. Similarly, the 33 determined residues of (Hind II/Hind III)-C fragment are identical, in 32 positions, to the 33 residues at the end of AGMr (Hind III)-1 distal to the (Hind II/Hind III)-E sequences. These data provide direct sequence confirmation of the origin of portions of both (Hind II/Hind III)-C and -E in the monkey genome, and allow us to define the exact positions at which the highly reiterated monkey DNA contained in the SV40-defective genome diverges from the AGMr (Hind III)-1 sequence as it is found in the monkey genome. Subsequent to residue 124, (Hind II/Hind III)-E is different from AGMr (Hind III)-1. This sequence might be an SV40 sequence or, alternatively, a monkey sequence not present in AGMr (Hind III)-1. As reported earlier by Grüss and Sauer (1975), no Hind II sites are detectable in the series of fragments generated from the BSC-1 DNA by cleavage with endonuclease R·Hind III. Thus if residues 125–151 of (Hind II/Hind III)-E were derived from the monkey genome, they must occur in an infrequent segment containing a portion of the highly reiterated AGMr (Hind III)-1 sequence. Furthermore, while it is improbable that the sequence is derived from wild-type SV40 DNA, since nowhere within the wild-type SV40 genome are three contiguous sites specific for endonucleases R·Eco RII and R·Hind II, as there are in (Hind II/Hind III)-E, it could represent an altered wild-type sequence. Sequence determination at the Hind II end of the (Hind II/Hind III)-B fragment known to contain SV40 sequences and to be contiguous to (Hind II/Hind III)-E in the substituted defective CVP8/1/P2 (Eco RI res) DNA I (Rao and Singer, 1977b) should allow a clearer identification of the origin of the sequences at the Hind II end of (Hind II/Hind III)-E. It is possible, of course, that recombinational events themselves generated the nucleotide sequence of residues 125 through 151, and that the observed sequence cannot be characterized either as monkey or SV40 in origin.

About 20–25% of the chromosomal DNA of African green monkey cells is contained in the highly repetitive fraction designated as component α (Maio, 1971; Kurnit, Shafit and Maio, 1973). Grüss and Sauer (1975) reported that component α DNA gives the same series of well defined low molecular weight DNA fragments as does total monkey DNA upon digestion with endonuclease R·Hind III. The reiteration frequency of component α has been estimated to be about 7×10^6, but it has not yet been shown that α is a homogeneous sequence element. It appears probable that the sequence described in this report is a sequence that is included in component α.

Experimental Procedures

Materials

The following materials were prepared as previously described: defective substituted SV40 variant CVP8/1/P2 (Eco RI res) DNA I (Rao and Singer, 1977a); (Hind II/Hind III)-E fragment from CVP8/1/P2 (Eco RI res) DNA I (both unlabeled and ^{3}H-labeled), purified by electrophoresis on polyacrylamide gels after digestion with the endonuclease R·(Hind II/Hind III) (Segal et al., 1976); Rao and Singer, 1977b; ^{32}P-labeled cRNA to fragment E (Segal et al., 1976). Monkey DNA was isolated from BSC-1, an established line of monkey kidney cells, by the procedure of Aloni et al. (1969). E. coli DNA (type VIII) was obtained from Sigma. Endonucleases R·Hind III (Smith, 1974) and R·Hind II, an isochizomer of R·Hind II (Landy et al., 1974), were obtained from New England Biolabs. NE-260 scintillation fluid was obtained from Nuclear Enterprises, Inc. (San Carlos, California).

RNA Sequencing Methods

^{32}P-labeled cRNA to fragment E was prepared by labeling the RNA separately with each of the four α-^{32}P-ribonucleoside triphosphates (New England Nuclear). The transcription products were analyzed either directly or subsequent to electrophoresis on 5% polyacrylamide gels in 8 M urea (Segal et al., 1976). The RNA was digested with either T1 or pancreatic RNAase, and the resulting oligonucleotides were fractionated by two-dimensional homochromatography (Figure 3) (Brownlee and Sanger, 1969). These products were further characterized by a variety of standard sequencing techniques (Brownlee, 1972) identical to those used previously (Rosenberg, Weissman and de Crombrugghe, 1975; Kramer and Rosenberg, 1976). These methods include subsequent digestion of oligonucleotides with the appropriate enzymes (T1, pancreatic and U2 RNAases) and fractionation of the products in one dimension on DEAE paper (Whatman DE 81) at pH 1.7 and 3.5 (Brownlee, 1972); determination of base composition and nearest-neighbor analyses by alkaline hydrolysis; and analysis of certain oligonucleotide products subsequent to their modification with a carbodiimide reagent (Barrell, 1971).

In addition, partial digestion of the cRNA with T1 RNAase was carried out as previously described (Rosenberg and Kramer, 1977). The products were fractionated two-dimensionally as described above using homochromatography solutions which effectively separated oligomers up to 45 residues in chain length. All partial products were further characterized by complete digestion with the appropriate enzymes and fractionation in either one or two dimensions by standard electrophoretic and/or chromatographic techniques (Sanger, Brownlee and Barrell, 1965; Brownlee and Sanger, 1969; Brownlee, 1972). The results of these partial enzymatic digestions are summarized in Figure 2.

Preparation of Fragments Labeled with ^{32}P at a Single 5′-Hydroxy Terminus

The 5′ ends of the DNA fragments were labeled, after dephosphorylation with bacterial alkaline phosphatase (Sigma), by phosphorylation with γ-^{32}P-ATP (ICN; spec. act. >1 Ci/μmole) using T4 polynucleotide kinase as described by Maniatis, Jeffrey and Kleid (1975). The strategy for obtaining the appropriate fragment [(Hind II/Hind III)-C or -E] labeled selectively at only one end will be understood by reference to Figure 1 and as follows: CVP8/1/P2 (Eco RI res) DNA I was cleaved with restriction endonuclease R·Hind III, and the 5′ ends of the mixed DNA fragments were labeled with ^{32}P as described above. These fragments were then redigested with endonuclease R·Hind II, and the products were resolved by electrophoresis on polyacrylamide gels and eluted from the gels as previously described (Rosenberg et al., 1975). Hind II/Hind III-C and -E fragments, labeled uniquely at their Hind III termini, were well resolved (Rao and Singer, 1977a) and can be obtained in pure form directly from these gels.

(Hind II/Hind III)-E fragment, labeled selectively at its Hind II end, was obtained by simply reversing the order of the restriction endonuclease digestions. CVP8/1/P2 (Eco RI res) DNA I was first cleaved with endonuclease R·Hinc II, end-labeled and then cleaved with endonuclease R·Hind III. Again, gel electrophoresis yielded the purified fragments.

DNA Sequencing by Partial Venom Phosphodiesterase Digestion

DNA fragments labeled selectively at one end were partially digested with snake venom exonuclease (Sanger et al., 1973; Maniatis et al., 1975). The resulting products were fractionated by electrophoresis on Cellogel (Kalex) at pH 3.5 followed by either homochromatography on DEAE-cellulose thin-layer plates (see Figure 5) or electrophoresis on DEAE paper at pH 1.7 and/or 3.5 (Sanger et al., 1965; Brownlee, 1972).

DNA Sequencing by Dimethyl Sulfate and Hydrazine Analysis

DNA fragments labeled at one 5′-hydroxy terminus were subjected to direct DNA sequencing according to the methods described by Maxam and Gilbert (1977).

Only representative autoradiograms of the many gel electropherograms obtained in the course of this work are presented here. The autoradiograms of all additional gels used to determine the reported sequences are available to interested investigators upon request to the authors.

Determination of Reassociation Kinetics

BSC-1 and E. coli DNA were sonicated for 30 min at full strength in a 10 kc Raytheon sonifier; the average length of the resulting fragments was about 150 bp as determined on alkaline sucrose gradients. The unlabeled BSC-1 or E. coli DNA was mixed with ^{3}H-labeled (Hind II/Hind III)-E fragment in 8 mM NaOH and denatured by heating at 100°C for 5 min. Thereafter the samples were cooled to 68°C and neutralized with 8 mM HCl. The salt concentration was adjusted to 0.12 M sodium phosphate (pH 6.8), and incubation was carried out at 68°C. At various time intervals, samples were removed and diluted with cold distilled H_2O to give a final concentration of 0.04 M sodium phosphate. The extent of hybridization was determined by hydroxyapatite chromatography at 60°C as described (Britten, Graham and Neufeld, 1974). In brief, unhybridized (Hind II/Hind III)-E was eluted from the columns at 0.12 M sodium phosphate, and hybridized molecules were removed at 0.4 M sodium phosphate. Radioactivity was determined in NE-260 scintillation fluid.

Thermal Denaturation of Reannealed ^{3}H-Labeled (Hind II/Hind III)-E

^{3}H-labeled (Hind II/Hind III)-E fragment was allowed to reassociate as described above in the presence of sonicated denatured BSC-1 or E. coli DNA to a Cot value of 10^{-2} (based on the concentration of the fragment as in Figure 6). The samples were diluted to a final concentration 0.04 M sodium phosphate and applied to hydroxyapatite columns at 60°C. The columns were washed at 60°C with 0.12 M sodium phosphate until no additional radioactivity emerged. The temperature was then increased at 5°C intervals (up to 98°C), and the total radioactivity eluted by 0.12 M sodium phosphate at each temperature was determined in NE-260 scintillation fluid.

Digestion of BSC-1 DNA with Endonuclease R·Hind III (and Hybridization of Resulting Fragments with cRNA to (Hind II/Hind III)-E

BSC-1 DNA was digested with 2 units of endonuclease R·Hind III per μg for 44 hr at 37°C in 0.06 M NaCl, 7 mM $MgCl_2$, 7 mM Tris-HCl (pH 7.4). The reaction mixture was analyzed by electrophoresis through 1.4% agarose slab gels for 5 hr at 40 V: the electrophoresis buffer was 0.04 M Tris-HCl, 0.02 M sodium acetate, 2 mM EDTA (pH 7.8). The gel was then stained with a 0.5 μg/ml solution of ethidium bromide (Sharp, Sugden and Sambrook, 1973), and the DNA was visualized with ultraviolet light. Photographs of gels were taken with a Polaroid camera and high speed type 57 Kodak film. The fragments were then transferred to nitrocellulose sheets by the method of Southern (1975) as modified by Botchan, Topp and Sambrook (1976), and hybridization was carried out with ^{32}P-labeled cRNA to (Hind II/Hind III)-E. Hybridization was in the presence of 50% formamide (Fluka) and 0.75 M NaCl, 0.5% sodium dodecylsulfate, 0.05 M Tris-HCl (pH 7.8) for 24 hr at 37°C. After appropriate washing, the nitrocellulose sheet was placed in contact with Kodak No-Screen film for 48 hr.

Filter Hybridization

Hybridization of ^{32}P-labeled cRNA transcribed from (Hind II/Hind III)-E fragment or of ^{32}P-labeled (Hind II/Hind III)-E itself to DNA immobilized on nitrocellulose filters was carried out by the procedures described previously (Lavi and Winocour, 1972; Segal et al., 1976), except that the samples were treated at 68°C for 30 min in 99% formamide prior to dilution to 50% formamide and addition to

the filter suspended in hybridization buffer. DNA from a variety of independently derived, substituted SV40 variants (Oren et al., 1976) was prepared in the laboratory of Dr. Ernest Winocour. The DNA of the substituted variant DAR-d5 (Davoli et al., 1977) was supplied by Dr. George Fareed. The DNA of variants 1103 and 1101 (Lee et al., 1975) was prepared in this laboratory using stocks provided by Dr. Daniel Nathans.

Acknowledgments

We are grateful to Edward Bieber and Cathy Brady for excellent technical assistance. We thank Drs. Ernest Winocour, Daniel Nathans and George Fareed for generously supplying substituted SV40 defectives obtained in their laboratories. We also thank A. Maxam and W. Gilbert for providing us with their DNA sequencing protocols prior to publication.

Received April 8, 1977; revised May 16, 1977

References

Aloni, Y., Winocour, E., Sachs, L. and Torten, J. (1969). Hybridization between SV40 DNA and cellular DNAs. J. Mol. Biol. *44*, 333–345.

Barrell, B. G. (1971). Fractionation and sequence analysis of radioactive nucleotides. In Procedures in Nucleic Acid Research, 2, G. Cantoni and D. Davies, eds. (New York: Harper and Row), pp. 751–779.

Bigger, C. H., Murray, K. and Murray, N. E. (1973). Recognition sequence of a restriction enzyme. Nature New Biol. *244*, 7–10.

Biro, P. A., Carr-Brown, A., Southern, E. M. and Walker, P. M. B. (1975) Partial sequence analysis of mouse satellite DNA: evidence for short range periodicities. J. Mol. Biol. *94*, 71–86.

Botchan, M. (1974) Bovine satellite I DNA consists of repetitive units 1400 base pairs in length. Nature *251*, 288–292.

Botchan, M., Topp, W. and Sambrook, J. (1976). The arrangement of simian virus 40 sequences in the DNA of transformed cells. Cell *9*, 269–287.

Boyer, H. W., Chow, L. T., Dugaiczyk, A., Hedgpeth, P. and Goodman, H. M. (1973). DNA substrate site for the EcoRII restriction endonuclease and modification methylase. Nature New Biol. *244*, 40–43.

Britten, R. J., Graham, D. E. and Neufeld, B. R. (1974). Analysis of repeating DNA sequences by reassociation methods. In Methods in Enzymology, *29E* (New York: Academic Press), pp. 363–390.

Brockman, W. W. and Nathans, D. (1974). Isolation of simian virus 40 variants with specifically altered genomes. Proc. Nat. Acad. Sci. USA *71*, 942–946.

Brockman, W. W., Lee, T. N. H. and Nathans, D. (1973). The evolution of new species of viral DNA during serial passage of simian virus 40 at high multiplicity. Virology **54**, 384–397.

Brownlee, G. G. (1972). In Laboratory Techniques in Biochemistry: Determination of Sequences in RNA, T. S. Work and E. Work, eds. (New York: American Elsevier).

Brownlee, G. G. and Sanger, F. (1969). Chromatography of ^{32}P-labeled oligonucleotides on thin layers of DEAE cellulose. Eur. J. Biochem. *11*, 395–399.

Brownlee, G. G., Cartwright, E. M. and Brown. D. D. (1974). Sequence studies of the 5S DNA of Xenopus laevis. J. Mol. Biol. *89*, 703–718.

Davidson, E. H. and Britten, R. J. (1973). Organization, transcription and regulation in the animal genome. Quart. Rev. Biol. *48*, 565–613.

Davoli, D., Ganem, D., Nussbaum, A. L., Fareed, G. C., Howley, P. M., Khoury, G and Martin, M. A. (1977). Genome structures of reiteration mutants of SV40. Virology *77*, 110–124.

Endow, S. A., Polan, M. L. and Gall, J. G. (1975). Satellite DNA sequences of *Drosophila melanogaster*. J. Mol. Biol. *96*, 665–692.

Frenkel, N., Lavi, S. and Winocour, E. (1974). The host DNA sequences in different populations of serially passaged SV40. Virology *60*, 9–20.

Grüss, P. and Sauer, G. (1975). Repetitive primate DNA containing recognition sequences for two restriction endonucleases which generate cohesive ends. FEBS letters *60*, 85–88.

Kleid, D., Humayun, Z., Jeffrey, A. and Ptashne, M. (1976). Novel properties of a restriction endonuclease from *Haemophilus parahaemolyticus*. Proc. Nat. Acad. Sci. USA *73*, 293–297.

Kramer, R. A. and Rosenberg, M. (1976). The isolation and characterization of bacteriophage T7 messenger RNA fragments containing an RNase III cleavage site. Nucl. Acids Res. *3*, 2411–2426.

Kurnit, D. M., Shafit, B. R. and Maio, J. J. (1973). Multiple satellite deoxyribonucleic acids in the calf and their relation to the sex chromosomes. J. Mol. Biol. *81*, 273–284.

Landy, A., Ruedisueli, E., Robinson, L., Foeller, C. and Ross, W. (1974). Digestion of DNAs from bacteriophage T7, λ, and ϕ80h with site-specific nucleases from *Hemophilus influenzae* strain Rc and strain Rd. Biochemistry *13*, 2134–2142.

Lavi, S. and Winocour, E. (1972). Acquisition of sequences homologous to host DNA by closed circular simian virus 40 DNA. J. Virol. *9*, 309–316.

Lavi, S., Rozenblatt, S., Singer, M. F. and Winocour, E. (1973). Acquisition of sequences homologous to host DNA by closed circular simian virus DNA: II. Further studies on the serial passage of virus clones. J. Virol. *12*, 492–500.

Lee, T. N. H., Brockman, W. W. and Nathans, D. (1975). Evolutionary variants of SV40; cloned substituted variants containing multiple initiation sites for DNA replications. Virology *66*, 53–69.

Maio, J. J. (1971). DNA strand reassociation and polyribonucleotide binding in the African green monkey, *C. aethiops*. J. Mol. Biol. *56*, 579–595.

Maniatis, T., Jeffrey, A. and Kleid, D. G. (1975). Nucleotide sequence of the rightward operator of phage λ. Proc. Nat. Acad. Sci. USA *72*, 1184–1188.

Martin, M. A., Gelb, L. D., Fareed, G. C. and Milstein, J. B. (1973). Reassortment of SV40 DNA during serial undiluted passage. J. Virol. *12*, 748–757.

Maxam, A. M. and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Nat. Acad. Sci. USA *74*, 560–564.

Old, R. W., Murray, K. and Roizes, G. (1975). Recognition sequence of restriction endonuclease III from *Hemophillus influenzae*. J. Mol. Biol. *92*, 331–339.

Oren, M., Kuff, E. L. and Winocour, E. (1976). The presence of common host sequences in different populations of substituted SV40. Virology *73*, 419–430.

Polisky, B., Green, P., Garfin, D. E., McCarthy, B. J., Goodman, H. M. and Boyer, H. W. (1975). Specificity of substrate recognition by EcoRI restriction endonuclease. Proc. Nat. Acad. Sci. USA *72*, 3310–3314.

Rao, G. R. K. and Singer, M. F. (1977a). Studies on a defective variant of simian virus 40 that is substituted with DNA sequences derived from monkey DNA: I. Origin, properties and purification. J. Biol. Chem., in press.

Rao, G. R. K. and Singer, M. F. (1977b). Studies on a defective variant of simian virus 40 that is substituted with DNA sequences derived from monkey DNA: II. Structure of the DNA. J. Biol. Chem., in press.

Roizes, G. (1976). A possible structure for calf satellite DNA I, Nucl. Acids Res. *3*, 2677–2695.

Rosenberg, M. and Kramer, R. (1977). Nucleotide sequence surrounding a ribonuclease III processing site in bacteriophage T7 RNA. Proc. Nat. Acad. Sci. USA *74*, in press.

Rosenberg, M., Weissman, S. and de Crombrugghe, B. (1975). Termination of transcription in bacteriophage λ. J. Biol. Chem. *250*, 4755-4764.

Rozenblatt, S., Lavi, S., Singer, M. F. and Winocour, E. (1973). Acquisition of sequences homologous to host DNA by closed circular simian virus 40 DNA: III. Host sequences. J. Virol. *12*, 501-510.

Salser, W., Bowen, S., Browne, D., Eladli, F., Federoff, N., Fry, K., Heindell, H., Paddock, G., Poon, R., Wallace, B. and Whitcome, P. (1976). Investigation of the organization of mammalian chromosomes at the DNA sequence level. Fed. Proc. *35*, 23-35.

Sanger, F., Brownlee, G. G. and Barrell, B. G. (1965). A two-dimensional fractionation procedure for radioactive nucleotides. J. Mol. Biol. *13*, 373-398.

Sanger, F., Donelson, J. E., Coulson, A. R. Kössel, H. and Fischer, D. (1973). Use of DNA polymerase I primed by a synthetic oligonucleotide to determine a nucleotide sequence in phage f1 DNA. Proc. Nat. Acad. Sci. USA *70*, 1209-1213.

Segal, S., Garner, M., Singer M. F. and Rosenberg, M. (1976). In situ hybridization of repetitive monkey genome sequences isolated from defective simian virus 40 DNA. Cell *9*, 247-257.

Sharp, P. A., Sugden, B. and Sambrook, J. (1973). Detection of 2 restriction endonuclease activities in *H. parainfluenzae* using analytical agarose-ethidium bromide electrophoresis. Biochemistry *12*, 3055-3064.

Smith, C. A. and Vinograd, J. (1972). Small polydisperse circular DNA of HeLa cells. J. Mol. Biol. *69*, 163-178.

Smith, H. O. (1974). Restriction endonuclease from Hemophilus influenzae RD. In Methods in Molecular Biology, *7*, R. Wickner, ed. (New York: Marcel Dekker), pp. 71-85.

Smith, H. O. and Nathans, D. (1973). A suggested nomenclature for bacterial host modification and restriction systems and their enzymes. J. Mol. Biol. *81*, 419-423.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98*, 503-517.