The EMBO Journal vol.4 no.5 pp.1279-1286, 1985

A new member of the polyomavirus family: the hamster papovavirus. Complete nucleotide sequence and transformation properties

Véronique Delmas, Catherine Bastien, Siegfried Scherneck[1] and Jean Feunteun

Laboratoire d'Oncologie Moléculaire, Institut Gustave Roussy, 94805 Villejuif Cedex, France and [1] Central Institute of Molecular Biology, Academy of the Sciences of GDR, 1115 Berlin-Buch, GDR

Communicated by R.Monier

The hamster papovavirus (HapV) is associated with multiple skin epitheliomas of the Syrian hamster. We have sequenced its genome. It is a double-stranded circular DNA of 5366 bp. The hypothetical genomic organization deduced from this nucleotide sequence is clearly of the polyoma type with the two strands coding in the opposite directions from a non-coding region that shows some of the features of a replication origin and a transcription control region. The amino acid sequences predicted from the open reading frames show an average of 50% homology with polyoma-coded polypeptides. The HapV is, after polyoma, the second example of a papovavirus coding for a middle T antigen. The cloned DNA can immortalize primary rat embryo cells and transform an established rat cell line. The viral DNA is stably integrated into the host genome.

Key words: HapV/sequence/organization/transformation

Introduction

The hamster papovavirus (HapV) has been isolated by Graffi *et al.* from skin epitheliomas arising spontaneously in Syrian hamsters (Graffi *et al.*, 1968). The tumors which are disseminated originate from the hair follicle epithelial cells. In a crystal-like arrangement, virus particles fill the nuclei of cells in the keratinized layer; they are never found in the basal layer of proliferating cells. The tumors are transplantable but no virus can be detected in the transplanted tumors. The virus is morphologically indistinguishable from polyoma virus or SV40. Circular DNA molecules with a mol. wt. of $\sim 3.1 \times 10^6$ have been isolated from homogeneous preparations of virus obtained from the primary skin epithelioma (Graffi *et al.*, 1970). A preliminary characterization of this DNA has established that distinct homologies exist with polyoma and to a lesser extent with SV40 genomes (Scherneck *et al.*, 1984), despite totally divergent restriction maps. Interestingly, both the virus and its DNA are capable of inducing lymphomas and leukemias when inoculated subcutaneously into newborn hamsters; the incidence is high (30 – 80%) and the latency short (4 – 8 weeks). Thymectomized animals are poorly susceptible to these induced diseases suggesting that T lymphocytes may be the target of infection (Graffi *et al.*, 1970). Virions have not been detected in the lymphoma or leukemia cells. Although HapV interaction with keratinized cells may be reminiscent of the papillomavirus pathology, the preliminary characterization of its genome tends to classify it as a polyomavirus. However, its leukemogenic property is unique in this family.

We report here the complete nucleotide sequence of the HapV DNA. This sequence confirms the homologies of the genetic organization with the polyomavirus family. In addition we show that the cloned HapV DNA can both immortalize primary rodent cells and transform established rodent cells. By contrast with papillomavirus DNAs, which can replicate as episomes, the HapV DNA is stably integrated in the host genome.

Results

The general strategy for sequencing is illustrated in Figure 1. The selected restriction fragments were cloned in the appropriate M13 vectors to sequence both strands, including all of the fragment junctions. We have arbitrarily designated as the beginning of the sequence a C on one side of a hairpin structure, which may be an origin of replication. The sequence is presented in Figure 2. HapV is a 5366 bp circular DNA, slightly larger than polyomavirus (5292 bp) and SV40 (5243 bp). The sequence is consistent with the preliminary restriction enzyme mapping (Scherneck *et al.*, 1979). Figure 3A shows the distribution of termination codons in both strands. These diagrams demonstrate that both strands have coding capacities, an observation which immediately establishes HapV as a member of the polyomavirus family. In fact, its overall genetic organization resembles very much that of polyoma virus itself. This similarity includes the existence of three open reading frames in the first half of one strand, suggesting an early region coding for small T, middle T and large T antigens. On the second half of the molecule, the opposite strand shows two consecutive open reading frames that are likely candidates for the VP1, VP2 and VP3 late gene coding sequences. In between these two regions, a non-coding sequence with potential features of an origin of replication and a transcription control region is apparent. A tentative genetic map is presented in Figure 3B. Because the assignments of function are based on comparisons with polymavirus genome, they should be considered as speculative until mRNA mapping and protein identification have been done.

The putative early region

Two possible initiation triplets are present in the putative early reading frame at positions 74 and 192. However, the second AUG provides an N terminus for the early polypeptides homologous to the T antigens of polyoma, and therefore the sequence from position 192 to 2853 represents a likely early coding region. On the basis of the open reading frames and the polyoma model, three mRNAs could be spliced from a putative pre-mRNA. Acceptable consensus splice sites are found at positions compatible with such a model. A polyadenylation signal beginning at position 2881 could direct the maturation of the 3′ end of pre-mRNAs (Table I). Two characteristics of this genetic organization must be emphasized which place the HapV close to polyoma within the polyomavirus family. (i) The existence of a coding capacity for a middle T antigen in the polyoma genome is the major difference between the murine polyoma virus and the primate members of this family; the HapV seems to have this capacity. Moreover, as in polyoma (Treisman *et al.*, 1981), a polyadenylation site present at the 3′ end of the middle T coding sequence (at position 1432 in HapV) could be used to mature truncated

© IRL Press Limited, Oxford, England.

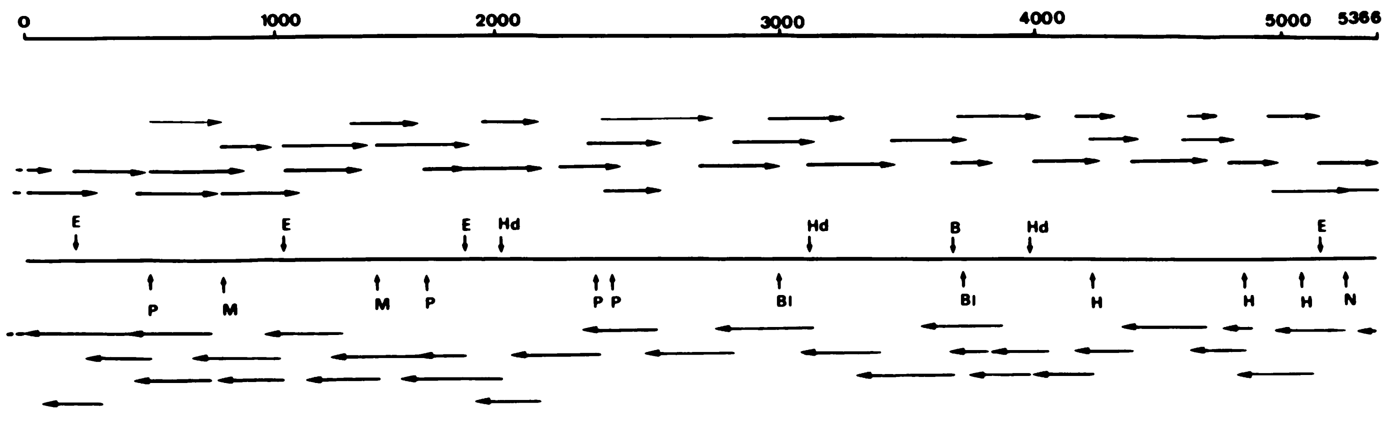


Fig. 1. Sequence strategy: each arrow represents the reading and the strandness of individual sequencing gel. E: *Eco*RI; Hd: *Hind*III; B: *Bam*HI; P:*Pst*I; M: *Mst*II; Bl: *Bal*I; H: *Hinc*III; N; *Nco*I.

```
   1 CCCCTTGCCTCCTTAGCTCTCAAGTAGAAAAGGAAGAGAGGCTTTTGGGGCTTTTTGGCT
  61 TTAAGCCTCATTTTATGAGCAGGAGGAGCTTGTTGCAACTTGAGAGGCGTTTTGAGGCTT
 121 CCAGGCAGAGAATACTCACAGACCCCACACAGTCTAGACGCTCAGAAGCATCTCTAGCTG
 181 CAACAAGCAAGATGGATAGAATTCTTACTAAAGAAGAAAAGCAAGCCTTAATAAGTTTAC
 241 TAGATTTGGAGCCACAATATTGGGGAGACTATGGACGAATGCAGAAATGCTACAAGAAAA
 301 AGTGTCTTCAACTGCATCCTGATAAAGGTGGCAATGAAGAGCTCATGCAACAGCTTAATA
 361 CCCTGTGGACCAAACTAAAAGATGGTCTTTACAGAGTTAGGCTGTTACTTGGGCCTAGTC
 421 AGGTAAGAAGACTTGGAAAAGATCAGTGGAATTTATCTTTACAGCAAACATTTTCTGGTA
 481 CCTACTTTAGGAGGCTCTGCAGACTCCCCATTACCTGCCTAAGAAACAAGGGAATTAGTA
 541 CCTGCAATTGCATACTTTGTTTGCTCAGAAAACAGCATTTTCTGCTAAAGAAGTCCTGGA
 601 GAGTACCTTGCCTGGTGTTAGGAGAATGCTACTGCATAGACTGCTTTGCCTTATGGTTTG
 661 GCCTGCCAGTTACCAATATGCTGGTTCCATTATATGCACAATTTCTTGCTCCAATACCTG
 721 TGGATTGGCTTGATCTGAATGTTCATGAGGTCTACAATCCGGCCTCAGGTATGTATGAAT
 781 ATGGGGGGCTTATAGTTGTAACTGTACAAGTTTAAAATGTGCTTTTTTCAGGACCCTAAT
 841 GCTTCCACCTCCACCAGCAGACCCGGAGAGTTCTACAATCCTGACACAGGAGGATACTGG
 901 TCCTACTCTTATGGGTCAGCAGGATACTCTGACCAGCAGAAGAAATACTGGGAAGAGTTT
 961 TTCTCTAAGTGGGATGTTAATGAGGACCTCACCTGCCAAGAAGAGTTATCATCATCAGAA
1021 GATGAATTCACCCCCTGGCATCCCAATCCCCCCCCCTCCCCTGTTTCTATTTCCAGTGAC
1081 AGCTCCAGTTCCTCCTGTGACGAGGAATACCCAAGAAACTCAAGCAGAAAGAGAAAACGA
1141 GTACATGCCAATGGCTCCCCAAATACACCTATACAGCCAAATAAGAGAGCCCACACACCA
1201 GGAGGAGGAAGAACCACAATACGAGGAGATACCGATATACCTAGAACTCCTGCCAGAGAA
1261 TCCCAATCAACATTTGGCTCTTACTTCAACAGCACGGAGGAGCTTGAGGAGGAAATATCA
1321 CAAACACAACAGTCACATCATAACACAACGCCAAAGAAACCGCCTCCGACGGTTAGTCCT
1381 GATGATTTTCCTACTATCCTTAGGGGGTTTCTTTCTCACGCTATTTTTTCTAATAAAACG
1441 CAAAATGCATTTATAATCTACAGTACTAAGGAAAAATGTGAAGTACTTTATGAACAAATA
1501 GACAAATATAATCCAGACTATAAAGGTATCTTCATTATGAAACAAACAGAAGCATTTGTA
1561 ATGTTTATGACTCCTGGAAAACATAGAGTAGCTGCAGTTAAAAGTTACTGTTGTAAATTT
1621 TGTACCGTTAGCTTCCTGCTATGCAAAGCTGTTACAAAACCGTTAGAGTTGTATAACTGT
1681 GTGGCTAAATGTGATGACTTTCAAATTTTAAAAGAAAATAAGCCTGGTCTATATCATTTT
1741 GAATTCTGTGATGAAAAAAAAGAGGTGAAGCAAATAGACTGGAATTTCCTAACATCTTTT
1801 GCAGTTGAAAATGAGTTAGATGATCCTCTTGTAATTATGGGACATTATCTAGAATTTAGT
1861 CAGTGTGAAAGCTCTTGCAAAAAGTGTGCAGAAGCTTTACCAAGGATGAAAGTCCACTGG
1921 GCTAACCACAGTCAGCACTTAGAGAATGCTGAGCTTTTCTTACACTGCAAACAACAGAAA
1981 AGTATCTGTCAGCAAGCAGCAGATAATGTTCTGGCAAGGAGAAGATTAAAGGTCCTTGAA
2041 TCAACAAGACAAGAATTGTTGGCAGAGAGACTGAACAAACTGTTAGACCAATTAAAAGAT
2101 TTATCTCCTGTAGATAAGCATTTATATCTTGCTGGAGTAGCCTGGTACCAATGTATGTTT
2161 CCTGATTTTGAGATGATGTTATTAGATATTTTAAAATTGTTTACTGAAAATGTTCCAAAA
2221 AAAAGAAATGTACTTTTTAGAGGTCCTGTAAATTCAGGGAAAACTAGCCTTGCTGCAGCT
2281 ATCATGAATCTTGTAGGAGGAGTTGCCCTCAATGTTAATTGTCCTGCAGATAAGCTCAAC
2341 TTTGAACTTGGTGTTGCTATAGATAAATTTGCAGTAGTCTTTGAAGATGTCAAAGGACAA
2401 ACCGGAGATAAGAGACACCTACAGTCTGGACTTGGAATTAATAACCTTGATAACCTGAGA
2461 GATTACCTTGATGGAAGTGTAAAGGTTAATTTAGAAAAGAAGCATGTAAATAAGAGGTCC
2521 CAGATATTTCCTCCTTGTATTGTTACTGCTAATGAATATTTTTTCCTCAAACACTCTAT
2581 GCCAGATTCCATAAAGTTTATAACTTTGAAGTGAAGGATTTTCTTGCCAAGAGCCTTGAG
2641 GAAAACAGTTACATGGGGAGACATAGAGTCTGTCAAAGTCCACTTACAATGCTGATAGCA
2701 TTGCTTTGGAATGTACCCACTGAAAATTTTGATAAGTCTCTCAAAGAGAAGGTGGAAACA
2761 GAAAAGAAGGTTTTGTCTGATATGTGTAACTTTACTACATTTGCAGAAATGTGTCTCAAT
2821 ATTCAGAGGGGTGCTGATCCCCTTGAGGCATTGTAATTGAGGAGGAAACAATAATTGATG
2881 AATAAAGCATTTATTAGAAGCTCTGTGTACAGTCATTTTTCAAGCATTAGTTTGCTGGTT
2941 TTGCAGGGGGTTTAGTATGCTGTTGGCCATACTTGTCAATGAACCTATTCACATCTGGGT
3001 CACCAGGAACAGCCTCTGTACCCTCATAAATCCTGACTTCTTCTACCTGAGCAGCTTCTC
3061 CTTCCATGGGCTGGCCTTCAATTGTTGGAAGCATATTGTTGTACAAAGAAGCTAGCAAGC
3121 TTGTAACTGGGTAAGGATTTTTCACCCATCTTTTTCTCAAGGTCACATTAAAATATCTAG
3181 GCAGCCCCCTCCAATGCCAGCCTGCACTGTTGTATTCTATGTACCAGCCCATAACATCTG
3241 CTGCACTGAGATAAAGCCCATCTCCTTTGCAAAGAGGCCCAACCCCATTTTCATCCAGAA
3301 GCACAGTAGTCAAGGTATTAGTAAACTGCATCACTGGTGGAGTACCAGTACCACCTGTGA
3421 GGTACCTACCATCCTTGTCCAATTTTGCTTTTGCAGTAGGGTCCAGCACCTGGTTTGTGG
3481 AAGTCATTGCTTTGCCAGTAACAGTTTTGATACTAACAATAGCTGCCTCATAATTTGCAT
3541 TATAGTTCTGCACTAGGCCTTGCAAATCTAATGGTTCTCCTCCCACTGCAAACATGTGGT
3601 AAGTTGTACCCTCAACTGGTTTGGAAATTCCAATATCCTTTGTCTCACTTCTGGAGCCAT
3661 ATCCATGCACATTTAGAAGGGATCCCACTCCAACAACTTCAGTTTTTACAGATACAGCCT
3721 CCCACATTTGAAGGGTATCACAGGTCAAATCTTCATTCAGTGTTGGAAGCTGTATTTTAG
3781 CCATACTGTAATATGGCAGTTGATTAGCCTTCACTTCATCAGCAGTAAGGGAGCTATTTA
3841 CTTTAATACTCTGGGAGAACCCATAATACTGGCCATCAGTTCCTGTGCCAGGCTTGTTCT
3901 GACCCATTCTAGGATTAAGGTAGGCCTCAATTTGTGTGATACTGTCTTCTCCTGTTACAA
3961 GATCAAGCACACCCACACCACCCCGCATAATAAGCTTGGGAACATTAGCAGGCTTTGGAC
4021 AGGGTTTCCACAGGGGTTTGCACATCTACTGGAAGCGCCGCTTTTTCTTTTTGGGGCCAT
4081 ACTCAACCTCATCAATGTATGTCTGCCAAGTAGGACTAATGTCTCCGTACAATCCTAGAA
4141 TTAAAGGAAGCATCCAATCAGGTGTCACTCTTTGGTGGGCTCCACCAGGAGCAAAATACC
4201 TCATGATATTTGCCCCTGATTCAAACCAACTAGAAGAGTCTTCCTGCTGTTGACTTCTTC
4261 TCTGGACATCAGGTCTCCCTAGTTCAGCTTCTAATACTTGTCTGCTATTGGCATCTTCAA
4321 TAGAAGGTCTATTACTGTATTCTAAAGCTCTTTCTATTTGTCTTCTTTGAGCTGGATTAA
4381 TTCCTGGAAGTTCTGCATAGTAGTTTTGTAGGCCACCATATATTCTACTATAGGCCTCTC
4441 TAGGTAAATTAGTAACAACCCATCTACTATTTTCCATCATTCTGGCAATGGCATCTAAAA
4501 ATTGATGGGTGGTCTGCAAACTTAAGTCCCTCACAGCAGATTCTACAGCTCCTTGAGTTT
4561 CCCTTCTCAAAGTATCCCAGATATACTCTCCTACAGATTGAAATAATGAGTGGCCCCAGC
4621 CATGAATAACATCCAAGGCATGGGTAAATGACTGTACACCAGGGAATAATATATCATAGT
4681 AGTCAGCTGGTCTCCTGGGAATAAGTGCCATGTTTCTATTCACAATCGGTACTTCGTGAG
4741 CAAGATAGCCGTGTAGACTTCCCAAAGAGAAGGCTGCTGAACCTGCAACAGTTTGAAAAA
4801 TAAATGCTGTCTGGACTGATTCTCTCACAAACTCAGTCATTACTGTTGATGTTAACTCAG
4861 GGGCAGCTTGCATAAATATAAACATGTCTTCACTTAGGCCAATTGAAGACAAAGCTGTCT
4921 CAGCTCCTAAAAACCCCTCCATAGTTATTAATGAAGTAACTTGGGCATCTATCGCGGCAA
4981 AGGCTTCTCCACTAAGTATGGCCTCTACTGAAATTCCAGTAACTGATGAAATTTCGGAGA
5041 GGTAGCTGATCATCTCAATAATCACTGAAATGGCAGATCCCATGTTGACTTACTTGAACA
5101 GTTTGAAAATCTTCTGAACTGTTTCAGGCAGGTTTTTAGGCCGAATTCTAAAGAAACAGA
5161 AAGCAAACACTCAGCGCCGAAGAGCAGGAAATGGCTGACCACTGCACTTGGGCGACACGA
5221 CACGCCTAGCGATAAGGAAGTCACCATGGCAACATAACCGCAGCACTGCTGTTGTCACAG
5281 TTGCCTAGCAAATGACAGACTCAGCAACCACAGGAGAGGAAATGATAGGGCTAGCATTTT
5341 TTCAAATGTAAACCAGAGGCTAGGGG
```

Fig. 2. Nucleotide sequence of HapV DNA. The sequence is 5366 nucleotides long and only the early strand is shown from 5′ to 3′. We have arbitrarily designated as the beginning of the sequence a C on one side of a hairpin structure which may be a possible origin of replication.

forms of early mRNAs. (ii) The splicing of the small T antigen mRNA occurs within the coding sequence as in polyoma and in contrast to SV40 and BK, where it occurs after the termination codon. Comparison of HapV putative protein sequences with polyoma and SV40 early gene products shows homologies ranging from 42% to 61% for polyoma and 32% to 49% for SV40. When the diagonal method (Maizel and Lenk, 1981), seeking at least 40% homology over 30 residue stretches, is used to compare polyoma and HapV large T and middle T antigens, a striking heterogeneity becomes apparent (Figure 4). Next to the region common to the three polypeptides that falls on the diagonal, the sequences strongly diverge between the two viruses both for large T antigen and middle T antigen (<17% homology). In the C-terminal part of both polypeptides, however, the homology is restored to the average value. This is particularly interesting to consider in the case of middle T antigen, which carries a conserved hydrophobic C terminus essential for membrane anchorage as demonstrated in the case of polyomavirus (Carmichael *et al.*, 1982). The remarkable homologies with polyoma early antigens include the existence in the HapV middle T antigen of a Glu_4 ProGlnTyr sequence (amino acid positions 312–318) homologous to the Glu_6Tyr sequence (amino acid positions 309–315) present in polyoma middle T antigen that has been proposed as a consensus sequence for tyrosine phosphorylation (Patschinsky *et al.*, 1982). A second tyrosine present at position 324 may be the equivalent of polyoma tyrosine 322, which is also phosphorylated *in vitro* (Carmichael *et al.*, 1984; Harvey *et al.*, 1984). These possible phosphorylation sites are among the few conserved stretches within the highly divergent region of middle T antigen.

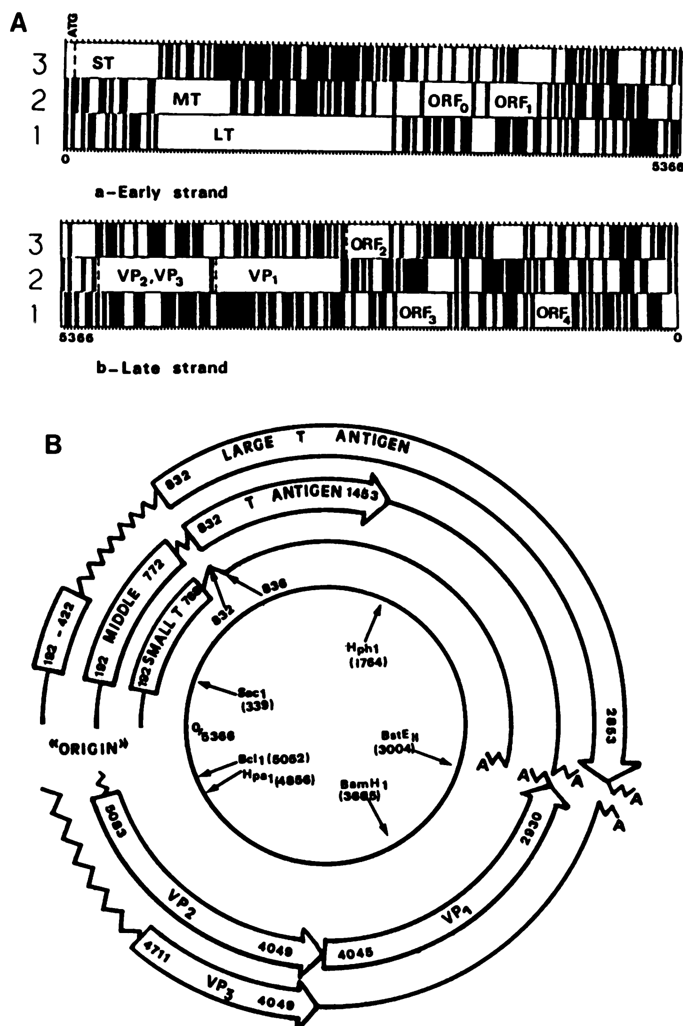


Fig. 3. (A) Distribution of termination codons in the HapV genome. The two diagrams show the distribution of termination codons on the early and late strands. From left to right, the sequence is 5′ to 3′. In each of the three reading frames a vertical bar is drawn when at least one termination codon is present in a group of 10 codons. Dotted vertical bars represent the first ATG present in the corresponding open reading frame. Open reading frames encoding polypeptides homologous with the polyomavirus-coded proteins are designated ST, MT, LT, VP1, VP2 and VP3. Other open reading frames potentially encoding peptides >120 amino acids in size have been designated ORF0 to ORF4. (B) Postulated landmarks on HapV DNA: the postulated physical location of the putative coded protein. The three early proteins: small T, middle T, large T and the three late proteins VP1, VP2 and VP3 are indicated in the appropriate curved bars. Location of non-coding parts of RNAs are indicated as single curved lines. DNA sequences absent in mRNA, which are probably removed by splicing are indicated by jagged lines (⌇). The unique restriction endonuclease sites are shown to facilitate the orientation of the proteins relative to the physical map.

The putative late region

The putative late coding region extends from positions 5083 to 2930 counterclockwise on the HapV physical map. The first open reading frame codes for the capsid proteins VP2 and VP3, the VP3 sequence is totally included in VP2. This open frame is closed by a single termination codon TAG at 4048, and then re-opened between the AUG at 4045 and the TAA at 2930. This region codes for the major capsid protein VP1. This situation distinguishes the HapV from the other polyomaviruses, since polyoma, SV40 and BK late proteins are coded by two different open reading frames in their respective late coding strand. Consequently in HapV there is no overlap between the sequence coding for the C terminus of VP2/VP3 and the N terminus of VP as it is found in the other polyomaviruses. HapV and polyoma differ significantly in the size of their capsid polypeptides. The HapV VP1 protein is shorter by 12 residues than polyoma VP1 protein, whereas HapV VP2 and VP3 proteins are larger by 27 and 17 residues, respectively. In the case of VP2 and VP3, the difference corresponds to an insertion of 25 residues at position 240 in the polyoma gene. The overall homology between HapV and polyoma is 61% for VP1, and 42% and 43% for VP2 and VP3, respectively. These homologies are reasonably constant along VP1 whereas they are much scattered in the VP2 and VP3 sequences.

The putative non-coding region

A sequence of 475 nucleotides bounded by the putative initiation codons for late and early translation at positions 5083 and 192 respectively, appears to be equivalent to the control region of the polyomavirus, containing at least an origin of DNA replication and the signals for transcription control (Figure 5). No HapV DNA replication has been demonstrated *in vitro* and no precise data exist about its transcription; our speculations on the characteristics of this sequence will be based on consensus sequences and/or homologies with other polyomaviruses. This non-coding region presents several interesting features that might be relevant to DNA replication. Two consecutive near-perfect palindromic structures are located between bases 5320 and 5339 and between 5356 and 11. The second palindrome is highly homologous to the polyoma structure considered to be the origin of replication (nucleotides 5281 – 20) and it is tempting to postulate that it plays the same role in HapV. Partially included in this palindrome, and extending upstream to the region between the two palindromes (bases 5338 – 5364), another sequence is recognizable by its homology with the PMS1 sequences required for autonomous replication of the bovine papillomavirus type 1 (BPV1) genome (Lusky and Botchan, 1984). Two PMS sequences have been characterized in BPV1: PMS1 is located in the non-coding region (positions 6945 – 7476) upstream to the 5′ end of early mRNAs and PMS2 lies in the E1 coding region (positions 1515 – 1655). PMS1 and PMS2 contain segments of extensive homology (73%). The HapV non-coding sequence contains a stretch of 20 bp displaying 80% homology with PMS1 (see Figure 5). Another characteristic feature of the polyomavirus replication origin is the presence of a GPuGGC sequence repeated several times; these sequences belong to the T antigen-binding site on polyoma DNA (Gaudray *et al.*, 1981). They are also present in HapV DNA. Finally, the presence of a 6-bp A-T tract upstream of the polyoma-like palindrome is also characteristic of polyomavirus replication origin.

Unlike the transcription control regions from SV40 or several murine retroviruses, there is no tandem repeated sequence in the enhancer of the wild-type polyoma strain A2. The HapV resembles polyoma in this respect; no extensive repeats are apparent in the non-coding region, no TATA box is recognizable upstream of the early or late genes initiation codons. However, enhancer elements are located on the late side of the putative origin of replication. The sequence (5304 – 5312) fits perfectly with the consensus sequence proposed for the SV40 core enhancer. The sequences (5183 – 5195) and (5232 – 5244) are also in good agreement with the consensus adenovirus E1A enhancer (Yaniv, 1984). Again, this situation is reminiscent of the polyoma enhancer sequences which are a mosaic of two elements: one SV40-like and one adenovirus E1A-like (Yaniv, 1984).

Table I. Landmarks on hamster papovavirus genome as suggested by the DNA sequence

Nucleotide No.	Sequence	Significance
Non-coding region		
5183–5195	AGCAGGAAATGGC	'Adenovirus-like' enhancer
5232–5244	ATAAGGAAGTCAC	
5304–5312	GCAACCACA	'SV40-like' enhancer
5320–5339	see Figure 5	Palindrome:hairpin loop
5338–5364	see Figure 5	Region of high homology with BPV PMS1 sequence
5356–11	see Figure 5	Region around putative viral origin of replication, predicted from Py homologies (hairpin loop structure)
5356–5360	(A)GAGGC	Possible T antigen binding sites
37–42		
103–108		
114–118		
118–122	CTTCC	Possible ribosome binding site
Putative early region		
192–194	ATG	Probable initiation codon for early proteins (T antigens)
419–425	TCAG/GTA	Possible splicing signals for maturation of a putative large T antigen mRNA
827–834	TTCAG/GAC	
765–771	TCAG/GTA	Possible splicing signals for maturation of a putative small T antigen mRNA
827–834	TTCAG/GAC	
769–776	GTAT/GT	Possible splicing signals for maturation of a putative middle T antigen mRNA
827–834	TTCAG/GAC	
837–839	TAA	Proposed termination codon for small T antigen
1432–1438	AATAAA	Possible processing signal for polyadenylation of a truncated early mRNA
1454–1456	TAA	Proposed termination codon for middle T antigen
2854–2856	TAA	Proposed termination codon for large T antigen
2881–2886	AATAAA	Possible processing signal for polyadenylation of early mRNAs
3194–3553		Open reading frame beginning with an ATG (ORF0). Coding capacity:119 amino acids
3677–4135		Open reading frame without ATG (ORF1). Coding capacity:153 amino acids
Putative late region		
5097–5091	TCAA/GTA	Possible splicing signals for a putative VP2 mRNA
5093–5088	GTAAG/T	
5097–5091	TCAA/GTA	Possible splicing signals for a putative VP3 mRNA
4780–4785	CAG/GTT	
5097–5091	TCAA/GTA	Possible splicing signals for a putative VP1 mRNA
4095–4080	GAG/GTT	
5083–5081	ATG	Putative initiation codon for VP2
4711–4709	ATG	Putative initiation codon for VP3
4045–4043	ATG	Putative initiation codon for VP1
4048–4046	TAG	Putative termination codon for VP2–VP3
2929–2927	TAA	Putative termination codon for VP1
2895–2890	AATAAA	Possible processing signal for polyadenylation of late mRNAs
2916–2493		Open reading frame beginning with an ATG (ORF2). Coding capacity:142 amino acids
2465–1994		Open reading frame without ATG (ORF3). Coding capacity:158 amino acids
2536–2527	AAG/GAG GAG/GAA	Possible splicing signals for a putative 'ORF2 + ORF3' RNA
2456–2451	AG/GTTA	
1282–917		Open reading frame without ATG (ORF4). Coding capacity:120 amino acids

Transformation

Transformation by polyomavirus is a multistep process involving the cooperation of at least two gene products, the large T and the middle T antigen (Cuzin, 1984). The functions carried out by these two genes can be demonstrated in two different assays. The first one is the immortalization assay which measures the ability of a given gene to promote the unlimited *in vitro* proliferation of primary cells. The polyoma large T antigen gene is active in this assay, as well as other viral genes such as SV40 large T and adenovirus E1A genes. These genes are referred to as class I transformation genes. The second assay is the malignant conversion assay which measures the capacity of a gene to induce a wide spectrum of phenotypic alterations in the cells, ultimately including tumorigenicity. The middle T antigen gene plays this role in polyoma transformation and hence is a class I transformation gene. We have tested the HapV DNA in these two assays. The immortalization capacity was measured on secondary rat or mouse embryo cultures according to the protocol described in Materials and methods. Colonies of immortal cells were obtained with HapV DNA at a frequency five times lower than with polyoma DNA (1–2 colonies/plate). This result establishes that the HapV carries a class I transforming gene but one which is clearly less efficient than the polyoma large T antigen gene. The presence of an active class II gene was investigated by transfection of the established rat cell line F111. Foci of transformed cells could be observed 3 weeks after transfection with a frequency four to five times lower than that obtained with polyoma DNA (Figure 6). In addition, the HapV foci appeared later and consequently were smaller at the time

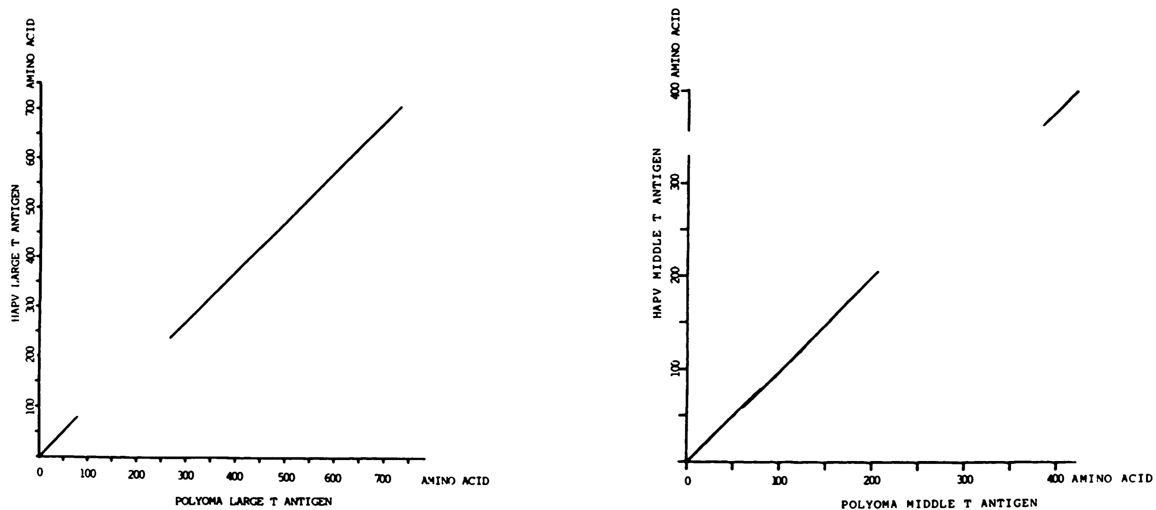


Fig. 4. Graphic matrix of **(A)** polyoma large T and HapV putative large T amino acid sequences, **(B)** polyoma middle T and HapV putative middle T amino acid sequences. Polyoma large T **(A)** or polyoma middle T **(B)** are on the vertical axis, and HapV large T **(A)** or HapV middle T **(B)** are on the horizontal axis. Groups of 30 amino acids are compared between the two amino acid sequences, and when at least 40% homology is detected, a line is drawn on the graph.

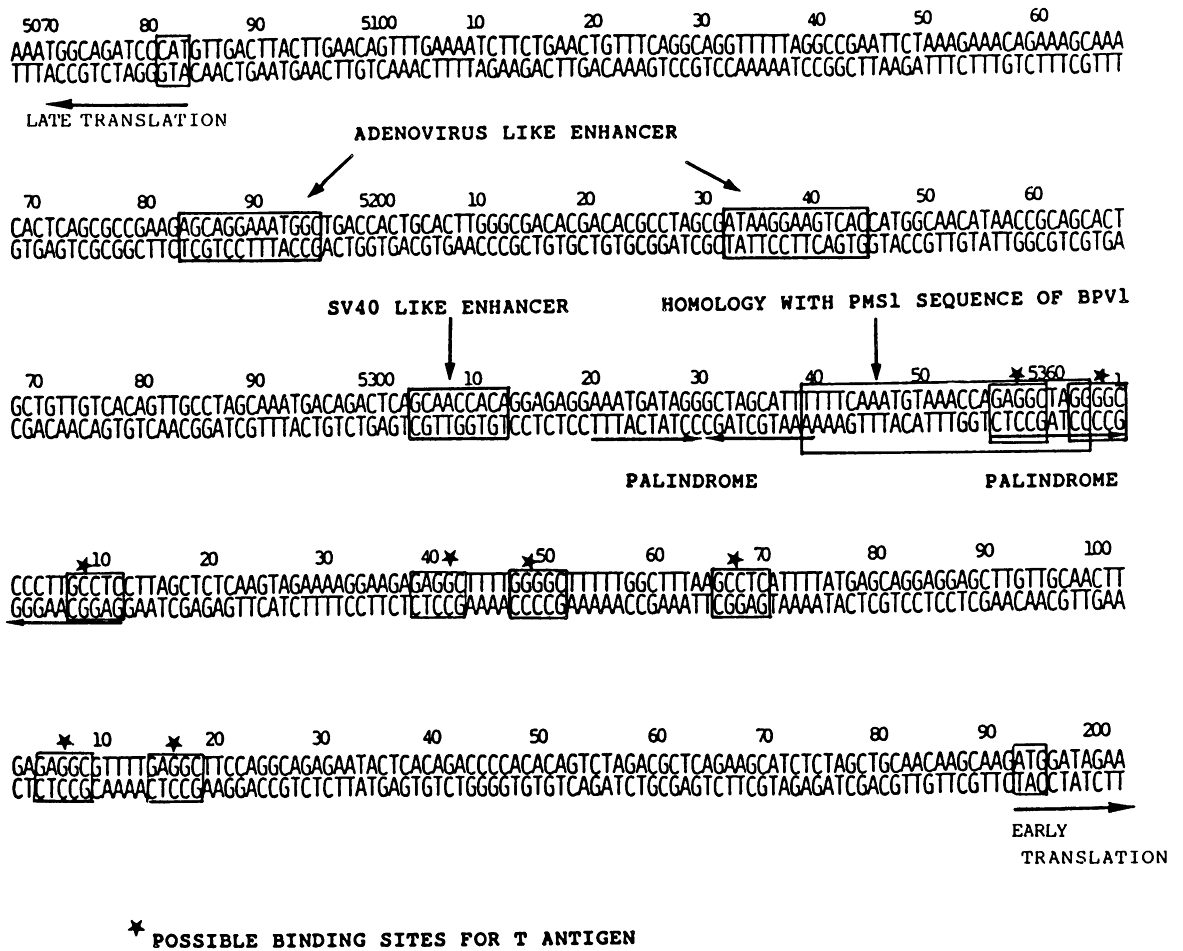
the plates were photographed. The HapV expresses a somewhat weak class II transformation gene able to stimulate the capacity for the transformed cell to overgrow the monolayer.

A major biological difference between the polyomavirus and the papillomavirus concerns the physical status of the viral genome in the transformed cells. It is a property of papillomavirus DNA to replicate autonomously as an episome in the transformed cells whereas polyomavirus DNA is usually integrated in the host genome. It was therefore interesting to study the status of the HapV genome in the transformed cells. The results of such a study performed on one mouse embryo line transformed by HapV is presented in Figure 7. The DNA from this cell line (ME HapV) was analysed by Southern blotting (Southern, 1975) and hybridization of an agarose electrophoresis gel. The undigested DNA stayed in the high mol. wt. region, and no free supercoiled form could be detected. When digested with *Bgl*II, which does not cut HapV, the HapV DNA remains as a single band in the high mol. wt. range, suggesting an integration at a single site within the host genome. The digestions with the single site enzymes *Bam*HI and *Hpa*I show an identical pattern of two heavily and one weakly labeled bands. One of the heavily labeled bands has the mobility of full length linear HapV DNA and represents a complete copy of the viral genome; the others presumably are the junction fragments. The weakly labeled bands represent the second junction containing only a short segment of HapV as the results of an integration between the *Hap*I and the *Bam*HI sites. This experiment suggests that the HapV DNA is probably integrated into the host genome as a tandem repeat (Figure 7). A similar conclusion has been reached on the status of the HapV DNA in an immortalized rat line (V. Delmas, unpublished).

Discussion

The papovaviruses represent a homogeneous family in most of their structural characteristics (icosahedric capsid, double-stranded DNA genome, etc.). However major biological differences divide this family into two distinct groups: the polyomaviruses and the papillomaviruses. The complete nucleotide sequence established for viral genomes representing several members of these two groups has provided a strong basis for this subdivision by showing that polyomaviruses and papillomaviruses use divergent strategies to express their genetic information. At the same time, the careful comparison of the respective coding sequences has demonstrated the existence of short but statistically significant homologies. For example, the SV40 large T antigen and the BPV1 E1 protein share striking sequence homologies which suggest that at least a domain of their sequence has derived from a common ancestor (Clertant and Seif, 1984; Seif, 1984).

We report here the nucleotide sequence of a viral genome (HapV) isolated from Syrian hamster skin epitheliomas. The analysis of the open reading frames clearly demonstrates that the HapV belongs to the polyomavirus group since both DNA strands have coding capacities in a configuration similar to the polyoma genetic map. In addition, five open frames with coding capacity of at least 199 amino acids have been detected both on the early and the late strands (see Table I). The significance of these potentially coding sequences which have no counterparts in the polyomavirus remains to be established. The HapV sequence raises interesting questions about evolution of the polyomaviruses. So far, the coding capacity for a middle T antigen had been demonstrated exclusively in the murine polyoma genome by contrast with the primate polyomaviruses like SV40 or BKV. HapV is a second example of a middle T antigen coding virus. This situation might be specific of the rodent viruses and it would be interesting to extend the observation to another murine papovavirus, the K virus. The amino acid sequence comparison between the polyoma and HapV early proteins reveals a highly divergent region covering ~190 residues. This regions, which is unique in both genomes by its double coding capacity in two different frames for middle T and large T antigen, seems to be totally absent in SV40 and BKV genomes. In addition, polyoma can afford some deletions in this part of the genome without consequence to its biological properties (Mes-Masson *et al.*, 1984). This observation challenges the concept of a greater genetic stability associated with a double coding capacity. Another aspect of papovavirus evolution has been illustrated by Soeda *et al.* (1980), who have suggested a co-evolution of the polyoma viruses and their host species. Our analysis of the homologies between HapV and polyoma or SV40-coded proteins (Table II) supports such a model by showing convincingly that HapV is closer to



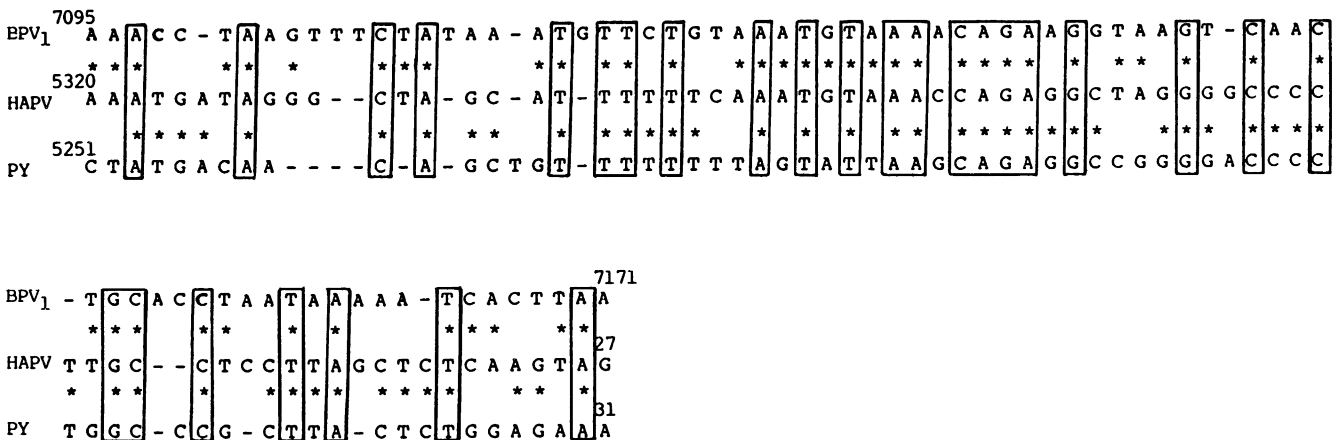


Fig. 5. **(A)** Non-coding region of HapV. The sequence of HapV between nucleotides 5070 and 200 is shown. Horizontal arrows indicate the initiation codons for late (5083) and the early (192) translation. The two palindromic structures at position 5390–9 and 5334–11 are indicated. The sequences homologous with the SV40 core enhancer (5304–5312), the consensus adenovirus E1A enhancer (5183–5195 and 5232–5244) and the PMS1 sequence of BPV1 (5339–5364) are boxed. The possible binding sites for T antigen are boxed and indicated by a star. **(B)** Comparison of sequences around the putative origin of replication of HapV, polyomavirus and plasmid maintenance sequence (PMS1) of BPV1. The HapV sequence between nucleotides 5320 and 27 is compared with parts of polyoma sequence between nucleotides 5251 and 31 and PMS1 sequence of BPV1 between nucleotides 7095 and 7171. Dots above each sequence indicate the respective homology between HapV and BPV1 PMS1 sequence and between HapV and polyoma (Py). The conserved nucleotides between all three sequences are boxed. Gaps are introduced to maximize homologies.

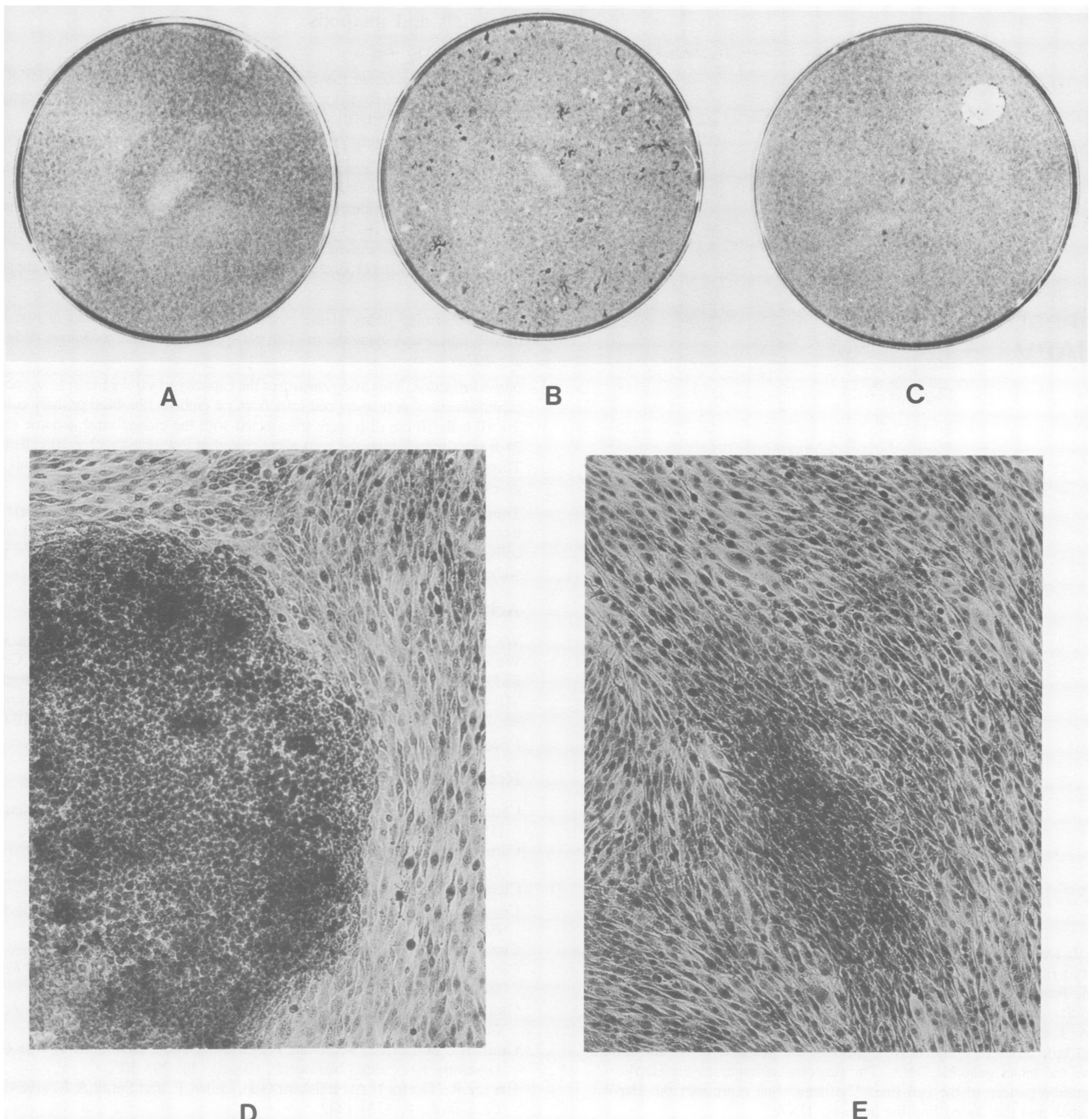


Fig. 6. Transformation assayed by focus formation on the established rat fibroblast cell line (F111). F111 cells (1 x 10^6 in 55 mm dishes) were transfected with 2 μg of PUC13 **(A)**, Py DNA **(B)** and HapV DNA **(C)**. Plates were stained with Giemsa after 3 weeks. Microscopic morphology of typical foci induced by polyoma **(D)** and HapV **(E)**.

the rodent (polyoma) than the primate (SV40) polyomavirus. The HapV is a fully transforming papovavirus carrying both class I and class II transformation genes, by contrast with the papillomaviruses which seem to express only the class II gene (J. Feunteun, unpublished results). The efficiency of the HapV genes is somewhat lower than the efficiency of the polyoma genes, this might reflect differences in the activity of the respective viral early promotors in the recipient rodent fibroblasts. The viral DNA is stably integrated in the genome of the *in vitro* transformed cells; this is the general situation in polyomavirus transformation. However, preliminary data suggest that in the lymphomas induced *in vivo*, the viral DNA is present as free concatenated molecules, a situation similar to that described for the papillomaviruses (Law *et al.*, 1981). It should be emphasized that the HapV may not be a unique isolate. Coggin *et al.* (1983) in Alabama (USA) described a very similar virus associated with skin epithelioma of the Syrian hamster. A detailed comparison of these two independently isolated viruses would be of great interest. In conclusion, the nucleotide sequence data definitively classify the HapV as a polyomavirus; mRNA and protein identification experiments are in progress which should confirm this classification. The biological properties of this virus isolate must be

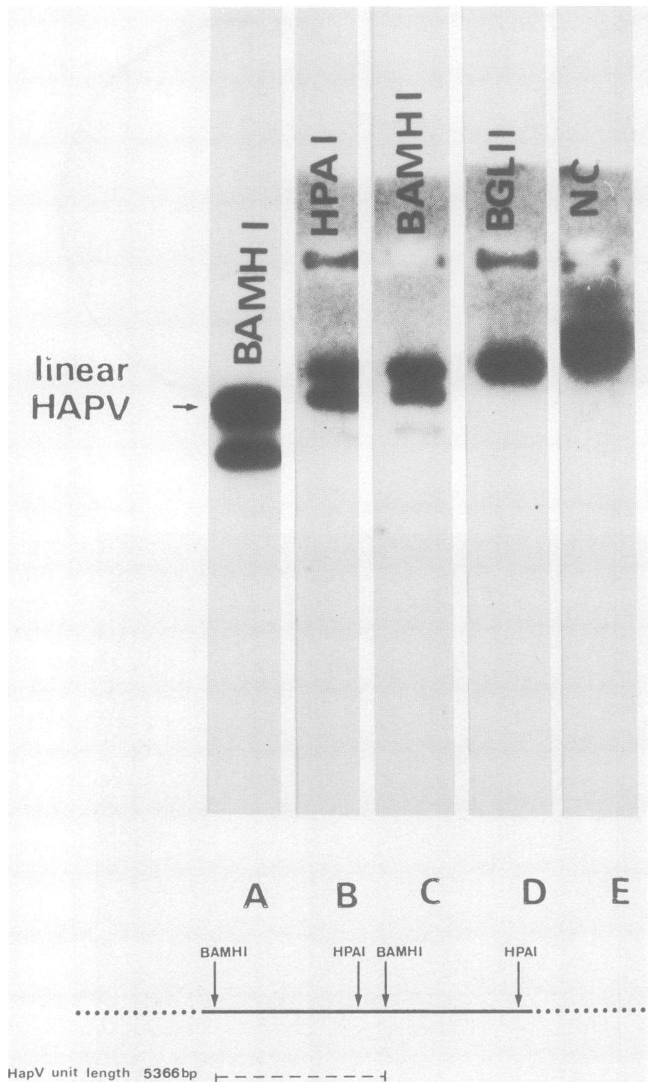


Fig. 7. Detection of the viral sequences into the HapV immortalized mouse embryo fibroblasts. High mol. wt. DNA (10 μg/slot) were cleaved with *HpaI*, *BamHI* or *BglII* and the digests analysed by Southern blot hybridization. Filters were probed with nick-translated ^{32}P-labelled HapV DNA. **Lane a** gives the positions of the full length linear HapV DNA (5366 bp); **lanes b, c** and **d** are digests of the genomic DNA and **lane e** shows the intact genomic DNA. The diagram summarizes the HapV integration pattern of the cell line. The heavy line represents the HapV sequence and the dotted line the genomic DNA.

Table II. Homologies between HapV, polyoma and SV40 coded proteins

	Protein theoretical sizes			% Homology	
	HapV	Polyoma	SV40	HapV/Polyoma	HapV/SV40
Large T	751	785	708	49.5%	34.7%
Middle T	401	432	–	42.4%	–
Small T	194	195	174	55.6%	33.5%
VP1	373	385	364	61.6%	49.3%
VP2	346	319	352	42.5%	32.5%
VP3	221	204	234	43.5%	33.8%

carefully compared with the properties of other polyomaviruses and papillomaviruses in order to extend the characteristics of the affiliation with these two families.

Materials and methods

Source of DNA

The DNA was extracted from a virus preparation purified from a pool of skin epitheliomas. It was cloned at the unique *BamHI* site in pUC13 and propagated in *Escherichia coli* HB101.

DNA sequencing

Restriction fragments were cloned in M13 phages Mp8, Mp9 or tg130 and tg131 and propagated in *E. coli* JM101. The preparation of single-stranded DNAs and the dideoxy chain termination sequencing technique were performed as described by Sanger *et al.* (1977). [^{35}S]Deoxyadenosine triphosphate (Amersham) was used in most of the sequencing reactions. The restriction enzymes were from New England Biolabs or Boehringer and used as recommended by the suppliers.

Transformation

The transforming genes of the cloned HapV DNA were tested in two DNA transfection assays using the calcium phosphate precipitate technique (Wigler *et al.*, 1978).

Immortalization. This assay measures the capacity of viral gene(s) to support the establishment of permanent cell lines from rat embryo fibroblast primary cultures. 3 x 10 cells/10 cm dish were transfected with the cloned viral genome (5 μg). 24 h after transfection the cells were plated at low density (2 x 10 cells/dish). Three weeks later, colonies were counted and picked for further study. This assay gives rise to truly immortal colonies (Petit *et al.*, 1983).

Transformation. 10 F111 cells (an established Fisher rat 3T3 cell line), were transfected with 2 μg of HapV cloned DNA. No carrier DNA was used and the cells were exposed to the precipitate for 4 h. After 3 weeks at 37°C, the plates were stained with Giemsa and dense foci were counted.

Acknowledgements

We are grateful to Dr. R. Monier for useful discussions throughout this work. We are indebted to I. Giri and M. Yaniv (Institut Pasteur, Paris) and J. Cebrian and P. Sheldrick (IRSC, Villejuif) for making available to us computer programs. The help of F. Royer in the preparation of the manuscript is gratefully acknowledged. This work was supported in part by the CNRS grant (No. 508109) and l'Association pour la Recherche sur le Cancer, Villejuif.

References

Carmichael,G.G., Dorsky,D., Oliver,D., Schaffhausen,B. and Benjamin,T.L. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 3579-3583.
Carmichael,G., Schaffhausen,B., Mandel,G., Liang,T. and Benjamin,T. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 679-683.
Clertant,P. and Seif,I. (1984) *Nature*, **311**, 276-279.
Coggin,J.M.,Jr., Bellomy,B.B., Thomas,K.V. and Pollock,W.J. (1983) *Am. J. Pathol.*, **110**, 254-266.
Cuzin,F. (1984) *Biochim. Biophys. Acta*, **781**, 193-204.
Gaudray,P., Tyndall,C., Kamen,R. and Cuzin,F. (1981) *Nucleic Acids Res.*, **9**, 5697-5710.
Graffi,A., Schramm,T., Graffi,I., Bierwolf,D. and Bender,E. (1968) *J. Natl. Cancer Inst.*, **40**, 867-873.
Graffi,A., Bender,E., Schramm,T., Graffi,I. and Bierwolf,D. (1970) *Comp. Leukemia Res., Bibl. Haematol.*, (Basel), **36**, 293-303.
Harvey,R., Oostra,B.A., Belsham,G.J., Gillett,P. and Smith,A.E. (1984) *Mol. Cell. Biol.*, **4**, 1334-1342.
Law,M.F., Lowy,D.R., Dvoretski,I. and Howley,P. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 2727-2731.
Lusky,M. and Botchan,M.R. (1984) *Cell*, **36**, 391-401.
Maizel,J.V.,Jr. and Lenk,R.P. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 7665-7669.
Mes-Masson,A.M., Schaffhausen,B. and Hassell,J.A. (1984) *J. Virol.*, **52**, 457-464.
Patschinsky,T., Hunter,T., Esch,F.S., Cooper,J.A. and Sefton,B. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 973-977.
Petit,C.A., Gardes,M. and Feunteun,J. (1983) *Virology*, **127**, 74-82.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
Scherneck,S., Böttger,M. and Feunteun,J. (1979) *Virology*, **96**, 100-107.
Scherneck,S., Vogel,F., Nguyen,H.L. and Feunteun,J. (1984) *Virology*, **137**, 41-48.
Seif,I. (1984) *Virology*, **138**, 347-352.
Soeda,E., Maruyama,T., Arrand,J.R. and Griffin,B.E. (1980) *Nature*, **285**, 165-167.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503-517.
Treisman,R., Novak,U., Favaloro,J. and Kamen,R. (1981) *Nature*, **292**, 595-600.
Wigler,M., Pellicer,A., Silverstein,S. and Axel,R. (1978) *Cell*, **14**, 725-731.
Yaniv,M. (1984) *Biol. Cell*, **50**, 203-216.

Received on 25 February 1985