The obtention of simian virus 40 recombinants carrying \( d(CG \cdot GC)_n \), \( d(CA \cdot GT)_n \) and \( d(CT \cdot GA)_n \) sequences

Stability of the inserted simple repeating sequences

Jose M. CASASNOVAS 1, Michael J. ELLISON 2, Antonio RODRIGUEZ-CAMPOS 1 and Fernando AZORIN 1

1 Grupo de Quimica Macromolecular, Centro de Investigacion y Desarrollo. Consejo Superior de Investigaciones Cientificas, Universidad Politecnica de Catalunya, Barcelona
2 Department of Biology, Massachusetts Institute of Technology, Cambridge

(Received March 17, 1987) — EJB 87 0367

DNA sequences with a simple repeating motif represent a significant proportion of the repetitive elements found in eukaryotic genomic DNA. In particular, \( d(CA \cdot GT)_n \) and \( d(CT \cdot GA)_n \) sequences are widely found in eukaryotic DNA [1, 2] while \( d(CG \cdot GC)_n \) sequences are also present in eukaryotic genomes but to a lower frequency [1]. Furthermore, these simple repeating sequences are known to adopt non-B DNA conformations \( \text{in vitro} \). Several groups have now established that the alternating purine-pyrimidine sequences \( d(CG \cdot GC)_n \) and \( d(CA \cdot GT)_n \) adopt the left-handed \( Z \) conformation readily in response to negative supercoiling (see [3] for review). More recently the repeated sequence \( d(CT \cdot GA)_n \) has been demonstrated to undergo a structural transition to an underworld form in response to low pH and negative supercoiling [2].

In order to study the factors affecting the stability of these simple repeating elements in eukaryotic chromatin as well as their structural and functional properties, we have developed a general strategy for the introduction of repeated DNA sequences into the simian virus 40 (SV40) genome at two different positions: at the \( TaqI \) site (position 4739) localized in the intron of the T-antigen gene or at the \( HpaII \) site (position 346) found in the leader region of the late SV40. These sequences have been cloned into the SV40 genome and two different positions: at the \( TaqI \) site (position 4739) localized in the intron of the T-antigen gene, or at the \( HpaII \) site (position 346) found in the leader region of the late mRNAs situated at the border of the nucleosome-free gap

In this paper we describe the obtention of SV40 recombinants carrying the following simple repeating DNA sequences: \( d(CG \cdot GC)_5 \) (SV40/CG5), \( d(CA \cdot GT)_{30} \) (SV40/CA30) and \( d(CT \cdot GA)_{22} \) (SV40/CTZ2). We have found that long \( d(CG \cdot GC)_n \) stretches are more unstable than either \( d(CA \cdot GT)_n \) or \( d(CT \cdot GA)_n \), so that \( d(CG \cdot GC)_5 \) is the longest sequence which appears to be stably maintained in SV40. These sequences have been cloned into the SV40 genome at two different positions: at the \( TaqI \) site (position 4739) localized in the intron of the T-antigen gene, or at the \( HpaII \) site (position 346) found in the leader region of the late mRNAs situated at the border of the nucleosome-free gap which spans the SV40 control region [5–7].

MATERIALS AND METHODS

Plasmids

Plasmids were propagated in \textit{Escherichia coli} strain HB101. Plasmid DNAs were prepared by standard procedures.

Cells and viruses

All viruses were propagated in CV1 cells. Viral stocks were kept at \(-80^\circ\text{C}\).

DNA transfection

DNAs were transfected into CV1 cells by the DEAE-dextran method [7]. Individual plaques were picked and the viruses liberated by freeze thawing and sonication. These low-titer viral stocks were used to infect fresh plates of CV1 cells.

DNA sequencing

For sequencing, viral DNAs were obtained according to Hirt [8] and purified by centrifugation through CsCl
gradients. Purified DNAs were digested with NdeI restriction endonuclease, end-labeled with [γ-32P]ATP and digested with MboII. The appropriate DNA fragments containing the inserted sequence were purified by gel electrophoresis and sequenced according to Maxam and Gilbert [9].

Enzymes

All enzymes were obtained from New England Biolabs.

RESULTS

Fig. 1 outlines the construction strategy followed for the obtainment of SV40/ recombinants. They derive originally from plasmid pDPL6 [10, 11] which is a 2.2-kb pBR322 derivative that contains a polylinker with the following restriction sites: HindIII, XbaI, SmaI, BamHI, SalI. Several simple repeating DNA sequences were cloned at the SmaI site of pDPL6: d(CA·GT)30 (plasmid pDHf14 [11]), d(CG·GC)12 (plasmid pDHg16 [10]) and d(CT·GA)22 (plasmid pTC45 [2]). Plasmids pE/ (Fig. 1) were constructed by fusing the large SalI/PvuI fragment of pUC12 [12] with the small SalI/PvuI fragment of either one of the plasmids described above. Therefore plasmids pE/ contain the following polylinker: EcoRI, SstI, SmaI, BamHI, XbaI, HindIII, ClaI. The asterisk indicates the SmaI site derived from plasmid pDPL6, which corresponds to the site of insertion of the different repeating DNA sequences. Plasmids pE/ were then inserted into the SV40 genome at either the Tag1 site or the HpaII site, by linearization of the pE/ plasmid DNAs with AccI which cleaves at their unique SalI site and leaves the same sticky ends as HpaII and TaqI. Those pE/SV constructs containing plasmid sequences in the orientation shown in Fig. 1 were then selected.

Recombinant viruses SV40/ were obtained from pE/SV/ shuttle vectors by cleavage with XbaI which releases the plasmidic sequences but leaves part of the polylinker described above, from SalI to XbaI, attached to the SV40 genome (including the SmaI* site where the different repeating DNA sequences were cloned). Restricted DNAs were then ligated at low DNA concentration (0.01 µg/ml) in order to maximize formation of single-copy circles.

SV40/ DNAs were band-purified in a 1% low-melting-point agarose/Tris/borate/EDTA gel containing 0.5 µg/ml ethidium bromide and the purified closed circular SV40/ DNAs were then transfected to just confluent CV1 cells by the DEAE-dextran method [7]. Individual plaques of lysis were picked 21 days after transfection and used to inoculate fresh plates of CV1 cells. When marked cytopathology was observed, about 15 days after infection, Hirt extracts [8] were prepared and the DNAs analyzed by restriction endonuclease mapping and DNA sequencing.

Conservation of the inserted repeating DNA sequence in SV40/ viruses was determined by cleavage of the corresponding Hirt DNAs with XbaI and BamHI. The sequences inserted in SV40/ recombinants are flanked by recognition sites for these restriction endonucleases (Fig. 1). Therefore cleavage at these sites should result in a small XbaI/BamHI fragment harboring the repeating sequence. Fig. 2 shows a 10% polyacrylamide gel electrophoresis of the XbaI/BamHI digests of the Hirt DNAs obtained from five individual plaques picked for each one of the SV40/ recombinants carrying the insertion at the Tag1 site. From Fig. 2 it is evident that all five SV40/CA30 or SV40/CT22 recombinants maintained the inserted repeating sequence since they contain a small XbaI/BamHI fragment of identical size to the same fragment of the corresponding pE/SV/ shuttle vector (Fig. 2, lanes *).

Similar results were obtained when the repeating DNA sequences were inserted at the HpaII site (data not shown). The precise DNA sequence of the region surrounding the insertion was determined according to Maxam and Gilbert [9]. They were found to correspond exactly to the expected DNA sequences (Fig. 3).

On the other hand, the d(CG·GC)12 sequence was found to be more unstable in the SV40 genome than either d(CA·GT)30 or d(CT·GA)22. In this case the inserted re-
Fig. 2. Conservation of the repeating DNA sequences inserted at the TaqI site in SV40 recombinants. A 10% polyacrylamide-TBE gel is shown for the XbaI/BamHI digests of the DNA obtained from five individual plaques of lysis (lanes 1 – 5) picked for each one of the SV40 recombinants analyzed. Lanes * correspond to a XbaI/BamHI digest of the corresponding pE·SV/ shuttle vector. The arrows indicate the restriction fragment containing the inserted repeating sequence.

Fig. 3. Determination of the DNA sequences of the region surrounding the insertion of the different repeating DNA sequences in SV40 recombinants. The asterisks (*), correspond to the SmaI site derived from pDPL6 where the different repeating DNA sequences were cloned. SV40/0 correspond to the recombinant viruses derived directly from plasmid pDPL6, therefore they do not contain any inserted repeating sequence. The gel shown for SV40/CA30 corresponds to a recombinant virus carrying the insertion at the HpaII site. The DNA sequence of SV40/CA30 recombinants, carrying the insertion at the TaqI site, was also found to correspond exactly to the expected DNA sequence (data not shown).
peating sequence was present only in one out of five SV40/ recombinants analyzed (Fig. 2). Moreover, the insertion was found to be considerably shorter, only 10 bp in length (Fig. 3), than the original 24-bp d(CG·GC)12 insert of plasmid pDHg16 [10]. The d(CG·GC)12 sequence is probably lost early during transfection. This conclusion is based on our observation that, once packaged into virus, the d(CG·GC)5 insert does not seriously impair viral growth and is stably maintained through several passages.

DISCUSSION

DNA sequences with a simple repeating motive are widely found in eukaryotic DNA but their biological significance has not yet been established. In this paper several of these simple repeated sequences have been introduced into the SV40 genome and their stability in the viral minichromosome investigated.

Long d(CG·GC)n sequences are markedly more unstable in SV40 than either d(CA·GT)n or d(CT·GA)n. This higher instability might explain the relatively low frequency of appearance that these sequences show in eukaryotic DNA. It is known that d(CG·GC)n sequences can adopt the left-handed Z-DNA conformation quite readily [2]. However, the high instability of these sequences does not appear to be related to their conformational properties since d(CA·GT)n sequences, which can also form Z-DNA, are stable in SV40. More likely this instability is related to their palindromic character. During replication, palindromic sequences can form loops or other aberrant structures stabilized by intrastrand pairing, which would be eliminated through DNA repair. Consistent with this interpretation it has been reported that the palindromic sequence (C-T-C-T-G-A-G-A)n is also very unstable when inserted in SV40 [14]. The shorter the length of the d(CG·GC)n stretch the more difficult would be the formation of such aberrant structures, so that d(CG·GC)5 might represent the limit beyond which intrastrand pairing is no longer favoured.

Alternatively the d(CG·GC)n insert might be lost via homologous recombination with the host genome or between viral particles. However, this possibility does not appear to be very likely since: (a) repeated sequences that, like d(CA·GT)n or d(CT·GA)n, are highly represented in the host genome are stable when inserted in SV40; (b) sequences that, like (CT)4·(GA)4, do not share any similarity other than the palindromic character with d(CG·GC)n are also unstable in SV40.

The method for the introduction of repeating DNA sequences into the SV40 genome that we have described in this paper is general since any sequence which can be cloned at the Smal site of pDPL6 can be transferred, via the protocol described in Fig. 1, to SV40 provided that it is tolerated by the virus.

SV40/ recombinants should also be useful for studying the conformational state that the inserted DNA sequences adopt in SV40 minichromosomes. It is interesting that the protocol in Fig. 1 allows cloning of any given sequence into two regions of the SV40 minichromosome showing a different structural organization. The region surrounding the TagI site is organized as nucleosomes while the HpaII site is localized at the border of the SV40 control region which is known to be nucleosome-free in a subset of SV40 minichromosomes [5–7]. These investigations should provide important insights on the determination of the degree of conformational variability that DNA might display in eukaryotic chromatin and their potential biological function.

We are grateful to Dr. D. E. Pulleyblank for plasmids and, to Drs. J. A. Subirana and L. Cornudella for advice and support. We are also grateful to Dr. C. Escarmis and S. Prat for experimental assistance. The technical assistance of the Taller de la Catedra de Tecnologia Mecanica is also acknowledged. This work was supported by grants from the Comision Asesora de Investigacion Cientifica y Tecnica, the Consejo Superior de Investigaciones Cientificas (CSIC) and the Comissio Interdepartamental de Recerca i Innovacio Tecnoligica. J. M. C. was a recipient of a doctoral fellowship from the Ministerio de Educacion y Ciencia. M. J. E. was supported by a fellowship from the Medical Research Council of Canada. A. R. C. was a postdoctoral fellow from the C. S. I. C.

REFERENCES