MOLECULAR AND CELLULAR BIOLOGY, Dec. 1986, p. 4570–4577
0270-7306/86/124570-08$02.00/0

Vol. 6, No. 12

# Functional Analysis of the Role of the A+T-Rich Region and Upstream Flanking Sequences in Simian Virus 40 DNA Replication

ROBERT GERARD†* AND YAKOV GLUZMAN

*Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724*

Received 30 June 1986/Accepted 3 September 1986

**One boundary of the minimal origin of replication of simian virus 40 DNA lies within the A+T-rich region. Deletion of only a few bases into the adenine-thymine (AT) stretch results in a DNA template which is defective for replication both in vivo and in vitro (B. Stillman, R. D. Gerard, R. A. Guggenheimer, and Y. Gluzman, EMBO J. 4:2933–2939, 1985). In the present study, such deletion mutations have been reconstructed into a simian virus 40 genome containing an intact early promoter-enhancer region. The resulting mutants synthesized wild-type levels of T antigen, but were defective for replication and would not form plaques on CV-1 monkey cells. Replication-competent phenotypic revertants were selected after transfection of large quantities of the replication-defective viral DNAs into CV-1 cells. DNA sequence analysis showed that most of these revertants contained insertions or point mutations which partially regenerate the length of the AT stretch. These genotypic alterations were shown to be responsible for the revertant phenotype by replication analysis in vivo of subcloned revertant origin fragments. In general, our results emphasize the importance of the AT region to simian virus 40 origin function. However, one revertant retained the altered AT region but deleted six nucleotides upstream. Experiments using this mutant indicate that the 21-base-pair repeats identified as part of the early transcriptional promoter may compensate for defects in simian virus 40 DNA replication in vivo caused by mutations in the A+T-rich region when positioned at an appropriate distance from the core origin.**

The boundaries of the simian virus 40 (SV40) origin of replication have been delineated in previous studies (3, 6, 9, 21, 25, 28, 36). A 65-base-pair (bp) minimal core sequence (nucleotide positions [np] 5209 through 30) has been defined which contains all necessary *cis*-acting sequences required for replication when introduced into a permissive cell environment containing SV40 T antigen. The core origin encompasses the sequences between the transition point for leading to lagging strand synthesis (17) and the 17-bp contiguous stretch of adenines and thymines (AT stretch) and includes the second binding site for T antigen (Fig. 1). Small deletions (16, 28) or base substitutions (32) within this binding site completely abolish replication activity. Using deletion mutants, we previously determined that the integrity of the 17-bp AT stretch in the core origin was crucial to *ori* function both in vivo and in vitro (36). Deletion of even a few base pairs of this region was highly deleterious to origin function. These results are in agreement with those in which point mutations (43) and small deletions (5, 9) within the AT region were shown to reduce the efficiency of replication.

The AT region also serves as the TATA box for the SV40 early promoter (Fig. 1) which also contains the 21-bp (promoter) and 72-bp (enhancer) repeats (1, 9, 14, 43). These upstream repeated sequences are not necessary for origin function (36), although they enhance the extent of replication by 5- to 10-fold (3, 6, 9, 19, 25, 33). The extent of enhancement depends on the experimental parameters of the replication assays, such as the amount of DNA transfected (33), the duration of replication (6), or the presence of competing origins (25). The 21- and 72-bp repeats are not required for SV40 DNA replication in vitro (25, 33, 36).

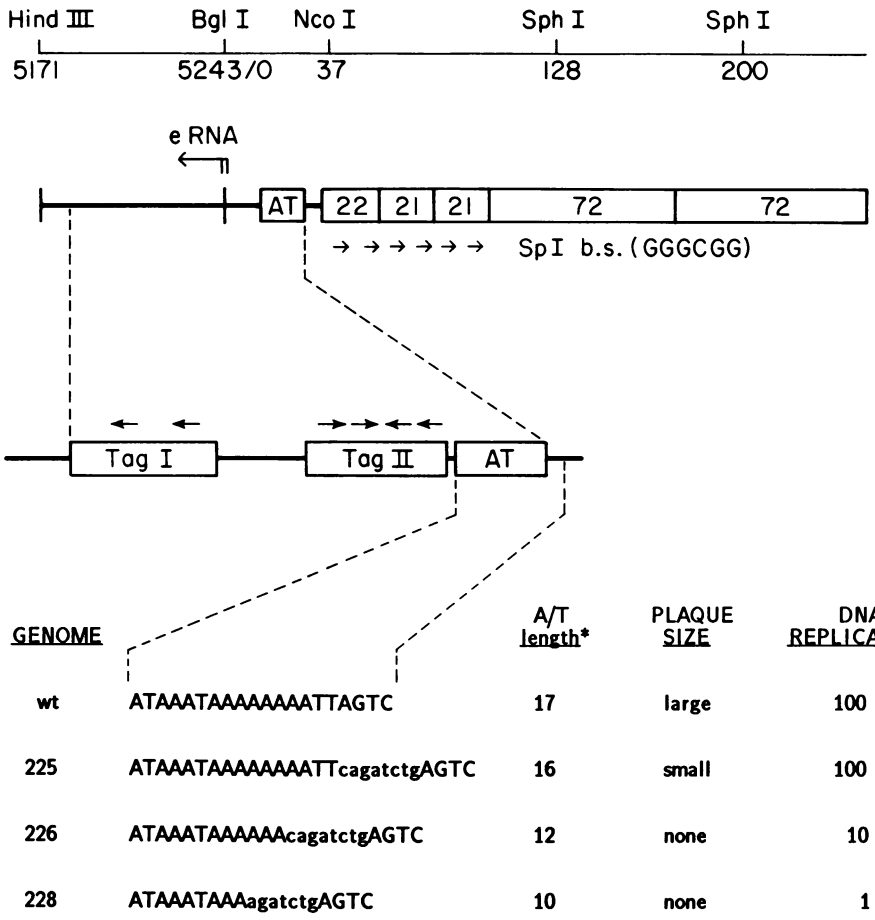
* Corresponding author.
† Present Address: Departments of Internal Medicine and Biochemistry, University of Texas Health Science Center, Dallas, TX 75235.

The experiments described in the present study were initiated to test the effect of upstream sequences on DNA replication potential in vivo of the AT-region deletion mutations. To this end viral genomes were reconstructed which contained wild-type sequences with the exception of the deletions at the late boundary of the A+T-rich region. We found that the upstream sequences did not restore the ability of the mutant origins to replicate, even though efficient transcription of the T-antigen gene occurred. Phenotypic revertants of the AT-region deletion mutations, however, could be selected which replicated more efficiently and formed virus plaques. The structure of these revertants confirmed that the A+T-rich region is critical for *ori* function. Surprisingly, alterations to the upstream sequences comprising the Sp1 binding sites I through III within the 21-bp repeats affected the efficiency of replication of mutant origins containing a shortened AT stretch.

## MATERIALS AND METHODS

**DNAs. (i) Viral genomes containing deletions in the 17-bp AT stretch.** Deletions in the origin of the cloned SV40 genome in pK1 have been described previously (11). pK1*dl*22 contains a *Bgl*II linker in place of the deletion of np 5187 through 30 in the SV40 sequence. The large *Bgl*II-*Bst*XI fragment of pK1*dl*22 missing the SV40 origin region and the first exon of T antigen was ligated to the small *Bgl*II-*Bst*XI *ori* fragment of pS1*dl*5, pS1*dl*6, and pS1*dl*8 DNAs (36) to generate the plasmids pK1*dl*225, pK1*dl*226, and pK1*dl*228, respectively. The SV40 DNA sequences cloned in these plasmids were indistinguishable from the wild-type genome, with the exception of the sequence around the A+T-rich region (Fig. 1).

**(ii) Isolation of revertant viruses.** Plasmids pK1*dl*225, pK1*dl*226, and pK1*dl*228 were digested with *Eco*RI to release the linear SV40 genomes and ligated at a total DNA




| GENOME | | A/T length* | PLAQUE SIZE | DNA REPLICATION** |
|---|---|---|---|---|
| wt | ATAAATAAAAAAAATTAGTC | 17 | large | 100 |
| 225 | ATAAATAAAAAAAATTcagatctgAGTC | 16 | small | 100 |
| 226 | ATAAATAAAAAAcagatctgAGTC | 12 | none | 10 |
| 228 | ATAAATAAAagatctgAGTC | 10 | none | 1 |

FIG. 1. Map of the SV40 origin region depicting the relative positions of the 72- and 21-bp repeated sequences, the A+T-rich region, and the early mRNA starts. Locations of restriction enzyme cleavage sites used in this study are shown above the map. The small arrowheads beneath the 21-bp boxes denote the six binding sites (GGGCGG) for the Sp1 protein (15). The boxes labeled TagI and TagII represent two regions which bind SV40 T antigen, and the arrows above these boxes represent the individual pentanucleotide binding elements (GAGGC) for T antigen (7, 22, 38). The AT box representing the 17-nucleotide-long AT stretch is expanded to show the sequence of the wild-type DNA from np 15 through 34 (39). The sequences of mutants used in this study are shown below. *, The length of the AT stretch is measured as the amount of uninterrupted run of A · T pairs. **, The percentage of DNA replication in vivo of the mutant templates is given in comparison with the wild-type template.

concentration of 5 μg/ml to promote recircularization. Transfection of the recircularized viral genomes into CV-1 monkey cells was as previously described (11). Monolayers were overlaid with 0.9% Noble agar containing Dulbecco minimal essential medium and 5% fetal bovine serum on the day after transfection and stained with neutral red after 2 weeks to visualize plaques. Individual plaques were picked and used to prepare virus stocks. Viral DNA was prepared by the Hirt (18) procedure.

The *Hin*dIII-*Kpn*I origin fragments from the viral DNAs were cloned into m13mp18. Sequencing of individual m13 clones was performed by the dideoxy chain termination method (30).

**(iii) Subcloning of revertant origin fragments from m13mp18.** Revertant origin fragments were isolated from sequenced m13 clones. After annealing of the sequencing primer to a 0.5-μg single-strand template, second strand synthesis was performed. Klenow fragment (2 U) extension of the primed template was performed at 30°C for 30 min in the presence of all four deoxynucleoside triphosphates (50 μM). The duplex DNA was digested with *Hin*dIII and either *Nco*I or *Sph*I and ligated into a similarly digested vector. The *Hin*dIII-*Nco*I origin fragments (np 5171 through 37) from the revertant DNAs were cloned into the vector pS1*dl*20. pS1*dl*20 was constructed by partial digestion of pS1 with *Nco*I as previously described (36). It contains the entire early region of SV40 from *Hpa*II to *Bam*HI with a *Bgl*II linker at np 333 of the SV40 sequence in lieu of the *Nco*I site. In the case of revertant viral DNAs lacking the *Nco*I site at np 37, *Hin*dIII-*Sph*I origin fragments were cloned into pAT153. Recombinant plasmids were identified by appropriate restriction enzyme screening procedures. Two sets of plasmids were obtained. The pSVO-derived series contains SV40 sequences between the *Hin*dIII and *Sph*I sites and includes the upstream promoter 21-bp repeats. The pS1*dl*20-derived series contains origin sequences between the *Hin*dIII and *Hpa*II sites (np 5171 through 346). Only those origin sequences from *Hin*dIII to *Nco*I were derived from the revertant DNAs.

**(iv) BAL 31 deletion series of pSVO$^+$ and pSVO 226r12.** After linearization of pSVO$^+$ and the pSVO 226r12 plasmid with *Sph*I, the DNAs were digested with BAL 31 at 37°C for various periods of time to remove between 0 and 125 bp from each end. The DNAs were flush ended with Klenow frag-

ment and all four deoxynucleoside triphosphates before ligation. Plasmids containing appropriate deletions were sequenced after cloning of *Hin*dIII-*Sal*I origin fragments into M13mp18. For deletion plasmids lacking an *Sal*I site, the entire genome was cloned into m13mp18 via the *Hin*dIII site.

**(v) Insertion mutations of 4 bp.** Plasmids containing a *Bgl*II site at the deletion endpoint were digested with *Bgl*II, flush ended with Klenow fragment and all four deoxynucleoside triphosphates, and ligated in dilute solution to promote recircularization. Insertion mutants were identified by their resistance to *Bgl*II.

**Cells and transfection.** CMT3 monkey cells containing multiple copies of the metallothionein promoter-SV40 T-antigen structural gene have been described (11). They support the high-level replication of SV40 origin-containing molecules, particularly when T-antigen synthesis is induced by the addition of heavy metals to the culture medium. Plasmid DNAs were transfected into CMT3 monkey cells, and the extent of replication was assessed after 50 h of induction by heavy metals as previously described (11). Quantitation of the extent of DNA replication was performed by one of two methods. Visible DNA bands on ethidium bromide-stained agarose gels were photographed with Polaroid type 55 positive-negative film. Negatives were scanned with a densitometer, and the amount of DNA in the bands was determined by comparison of the film density to that of a standard curve of known amounts of DNA. Alternatively, gels were blotted to nitrocellulose (34) and probed with an appropriate nick-translated plasmid. Radioactive bands were located by autoradiography, excised from the blot, and counted in liquid scintillation cocktail. The amount of DNA was determined by comparison of the counts with a standard curve.

## RESULTS

**Construction of mutant viral genomes defective for replication.** Deletion mutations in the 17-bp contiguous AT stretch of the core origin element of SV40 have deleterious effects on DNA replication. These results were observed in plasmids lacking the upstream viral sequences required for the early promoter function (9, 36). We were curious as to whether restoration of the upstream sequences to the plasmids containing AT deletions would improve replication of the origins containing *dl*6 and *dl*8 mutations. Restoration of the upstream sequences was accomplished by inserting the origin fragments from the pS1*dl* series into the Ori$^-$ pK1*dl*22. pK1*dl*22 was chosen because the *Bgl*II linker was fused to np 31, which is adjacent to the AT stretch, and the regenerated plasmids would contain the entire SV40 genome, each harboring the specific AT-region deletion. The sequences around the late boundary of the AT stretch in the viral genomes reconstructed from the *dl*5, *dl*6, and *dl*8 mutants are shown in Fig. 1. The pK1*dl*225 genome, henceforth designated simply as 225, contains a clean insertion of a *Bgl*II linker between np 30 and np 31 of the SV40 sequence and preserves a contiguous stretch of 16 A · T base pairs. Like the parental mutation in pS1*dl*5, the 225 genome replicates at wild-type levels in CV-1 cells (Fig. 1). By comparison, the 226 genome derived from the *dl*6 mutation has lost 5 of the 17 contiguous A · T base pairs (np 27 through 30) and replicates at only 10% the level of the wild type, which is about the same as that of the *dl*6 parent. Like the genome of the *dl*8 parent, the 228 genome is missing a C · G base pair from the linker sequence and therefore contains 10 of the 17 contiguous A · T base pairs (np 25 through 30 deleted). It replicates at a level comparable to that of *dl*8 and is less than 1% the efficiency of the wild type.

**Expression of T antigen by mutant genomes.** To determine whether the effects on DNA replication were a consequence of defects in early gene transcription and synthesis of T antigen, the rate of synthesis of T antigen was measured in cultures of CV-1 cells transfected with the recircularized mutant genomes. [$^{35}$S]methionine-labeled cell extracts were prepared 50 h posttransfection and immunoprecipitated by using monoclonal antibody to SV40 T antigen. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis revealed that approximately equal amounts of T antigen were synthesized by the wild-type and mutant viral genomes (data not shown). Duplicate cultures stained for T antigen by indirect immunofluorescence indicated that both the efficiency of transfection and the staining intensity of individual cells were equivalent in cultures transfected with the different viral genomes.

**Isolation and characterization of phenotypic revertants of mutant viral genomes.** The efficiency with which the mutant viral genomes formed plaques on CV-1 monkey cells was measured after excision and recircularation from the mutant pK1 plasmid series. The 225 genome formed plaques with an efficiency similar to that of the wild-type genome (225 gave $1.2 \times 10^4$ plaques per μg compared with $3.5 \times 10^4$ plaques per μg for the wild type), although the plaques formed were small, indicating an effect of the linker insertion on virus growth without significantly affecting DNA replication. DNA sequencing showed that all viruses recovered after transfection of the 225 DNA had the same origin structure as the original 225 plasmid genome.

Both the 226 and the 228 genomes formed plaques at low efficiency (approximately 10 to 20 plaques per μg for 226 and <10 plaques per μg for 228) and indicate the severe effect that the A+T deletions have on virus growth. Restriction enzyme patterns of viral DNAs obtained after infection of CV-1 cells with virus stocks prepared from the individual plaques were indistinguishable from that of wild-type virus DNA. However, DNA sequencing of the origin regions of these viruses revealed that in no instance was a virus with the original 226 or 228 genome isolated. Instead, the virus plaques isolated represent phenotypic revertants of the original input viral genomes. The predominant genotypic alteration was an increase in the length of the contiguous stretch of AT residues adjacent to the 27-bp palindrome in the core *ori* sequence. The sequences of these origin regions around the AT stretch are shown in Table 1.

The most frequently observed alterations were point mutations in the *Bgl*II linker sequence that convert G · C base pairs to A · T base pairs. These changes increase the overall length of the contiguous AT stretch without changing the genome length. Although most revertants with a specific sequence change in the A+T-rich region were isolated only once, one particular mutation affecting the AT stretch was independently isolated 14 times. This revertant is exemplified by the double point mutant 228r1, which has converted the *Bgl*II linker sequence AGATCTG to AAATCTA.

Only once was a single point mutation observed that restored replication function to a mutant viral genome. This was the revertant 228r23 in which a C · G base pair was converted to a T · A base pair in the linker sequence. This mutation did not increase the overall length of the contiguous 10-bp AT stretch. However, the virus structure was stable as shown by DNA sequencing of six independent plaque isolates from the 228r23 virus stock.

Mutants containing insertions of additional base pairs

TABLE 1. DNA structure of the revertant origins

| Virus | Sequence | Reversion[a] | AT sequence length (bp) | Plaque size | Other mutations |
|---|---|---|---|---|---|
| 226 | ATAAATAAAAAAcagatctgAGTCAGCC | | 12 | None[b] | |
| 226r13 | ATAAATAAAAAAAaAatctgAGTC | (dpm) | 17 | Small | Has T at np 53, A at np 57 |
| 226r3 | ATAAATAAAAAAAAAcagatctgAGTC | (+AAA) | 15 | Medium | |
| 226r2 | ATAAATAAAAAAAagatctAAGTC | (dpm) | 14 | Medium | |
| 226r12 | ATAAATAAAAAAcagatctgAGTCAGGG | (−CCATGG) | 12 | Medium[c] | |
| 228 | ATAAATAAAagatctgAGTC | | 10 | None[b] | |
| 228r13 | ATAAATAAAAaAaATtgAGTC | (+A, tpm) | 16 | Small | |
| 228r17 | ATAAATAAAaAatTtGAGTC | (dpm) | 16 | Small | |
| 228r28 | ATAAATAAAaAatctTAGTC | (dpm) | 13 | Medium | Has A at np 41 |
| 228r1, 228r3, 228r7, 228r8, 228r10, 228r11, 228r12, 228r15, 228r16, 228r18, 228r22, 228r24, 228r26, 228r30 | ATAAATAAAaAatctAAGTC | (dpm) | 13 | Medium | 228r8 has A at np 57 and 63, 228r10 has A at np 5221 and 5222, 228r11 has G at np 48 and A at np 63, 74, 84, and 88, 228r12 has A at np 41, 228r18 has A at np 58 and 64, 228r24 has A at np 5221 |
| 228r4 | ATAAATAAAaAaActgAGTC | (dpm) | 13 | Small | |
| 228r23 | ATAAATAAAagatTtgAGTC | (spm) | 10 | Small | |

[a] spm, dpm, and tpm refer to single, double, and triple point mutations, respectively.
[b] No plaques with this DNA structure were recovered.
[c] Virus stocks of this mutant were unstable.

were observed at a much lower frequency than those with simple point mutations. One of the revertants (228r13) includes the net addition of a single A · T base pair in conjunction with a triple point mutation to increase the length of the AT stretch. 226r3 is an example of a virus which has reverted by the insertion of three A · T base pairs to lengthen the AT stretch without mutating the linker sequence.

Additional point mutations were sometimes observed in revertant viruses and are indicated in Table 1. These mutations clustered predominantly within two regions of the genome. The first region is at np 5221 and 5222 within the core origin region. Examples of revertants with these mutations are 228r10 and 228r24. Apparently, both positions can be either an adenine or a guanine residue without adverse effects on virus growth and replication. The second region is between np 41 and 88 within the 21-bp tandemly repeated sequences. Examples of this class of revertants are relatively common; 4 out of 14 independent isolates of the revertant origin region exemplified by 228r1 contain different point mutations over the 21-bp elements. Obviously, none of these mutations appears to have a detrimental effect on virus growth as judged by the ability of the virus to form plaques.

**Phenotypic reversion is due to increased origin function.** To determine unequivocally that the mutations observed were responsible for the revertant phenotype, subcloning of the revertant origins and analysis of their replicative potential in vivo were performed. The smaller *Hin*dIII-*Nco*I fragment was subcloned from the m13 vector when possible. Use of the *Nco*I site makes it possible to remove any additional mutations within the 21-bp repeats (for example, in 228r8) from the plasmid construction used in the in vivo analysis (Fig. 1). *Hin*dIII-*Sph*I fragments from viral genomes lacking the *Nco*I site at np 37 were cloned out of the m13 genome used to sequence the mutations into a plasmid vector.

In vivo replication of the subcloned *ori* fragments from revertant viruses was tested in CMT3 monkey cells. Hirt supernatant DNAs were digested with both *Hin*dIII and *Dpn*I to convert all replicated DNA to a single band and to degrade unreplicated input plasmid. Figure 2 shows a Southern blot of a gel demonstrating both the replication of plasmids containing the *Hin*dIII-*Nco*I revertant origin fragments (Fig. 2A) and that of plasmids containing the *Hin*dIII-*Sph*I origin fragments (Fig. 2B). In all cases, the revertant origins replicated to a greater extent than did their parental origins. This increase ranged from 5- to 20-fold for the revertants of 226 and 7- to 50-fold for the revertants of 228. The revertant phenotype of the viruses therefore correlates with an increased ability of the origin region to support DNA replication in vivo.

**The relative alignment of the core origin with upstream sequences can effect replication.** A totally different type of revertant was 226r12, which has deleted the 6 bp comprising the *Nco*I restriction site upstream from the AT stretch. The *Bgl*II linker sequence and the original length of the AT stretch in the 226 genome were preserved. However, this revertant virus proved to be unstable; replaquing of the 226r12 virus stock and DNA sequencing of three independent plaque isolates revealed that these viruses had undergone further mutation. Specifically, the C · G base pair in the linker sequence adjacent to the AT stretch had been changed to an A · T base pair in all three cases. This additional mutation was also present in the 226r13 and 226r2 revertant viruses. No other changes in the genomic structure of the three viral origins were observed.

We considered the possibility that the revertant phenotype in 226r12 might be due to the deletion of specific sequences which acted to suppress replication of a genome containing a shortened AT stretch such as that in 226. However, the 228r23 genome retains these sequences and it contains an even shorter AT stretch. Although the subcloned origin from 228r23 replicates only as well as the 226 origin, 228r23 can be

TABLE 2. DNA structure of the parental and *in4* origins

| Mutant | Sequence[a] | Length (bp) | Replication (%) |
|---|---|---|---|
| Wild type | ATAAATAAAAAAAATTAGTCAGCCATG*GGGCGG* | +0 | 100 |
| 225 | ATAAATAAAAAAAATTcagatctgAGTCAGCCATG*GGGCGG* | +8 | 97 |
| 225*in*4 | ATAAATAAAAAAAATTcagatc**GATC**tgAGTCAGCCATG*GGGCGG* | +12 | 114 |
| 226 | ATAAATAAAAAAcagatctgAGTCAGCCATG*GGGCGG* | +4 | 14 |
| 226*in*4 | ATAAATAAAAAAcagatc**GATC**tgAGTCAGCCATG*GGGCGG* | +8 | 42 |
| 226r12 | ATAAATAAAAAAcagatctgAGTCA*GGGCGG* | −2 | 44 |
| 226r12*in*4 | ATAAATAAAAAAcagatc**GATC**tgAGTCA*GGGCGG* | +2 | 32 |
| 228 | ATAAATAAAagatctgAGTCAGCCATG*GGGCGG* | +0 | 0.5 |
| 228*in*4 | ATAAATAAAagatc**GATC**tgAGTCAGCCATG*GGGCGG* | +4 | 3 |

[a] The sequences of the four inserted nucleotides are shown in boldface type, and those of G·C box I are italicized.

propagated as a stable virus stock, whereas 226 cannot. We therefore consider it unlikely that the increase in replication of 226r12 is due to the deletion of suppressor sequences.

An alternative explanation was that the revertant phenotype of 226r12 was due to an alteration in the spacing between the core origin containing the AT stretch and the upstream sequence elements. The deletion of 6 bp results in a change of about one-half of a helical turn of the DNA in terms of the relative alignment of sequences upstream from the deletion with those located downstream. Effects of flanking sequences on SV40 replication have been demonstrated previously (3, 6, 9, 19, 25), although the effects of the relative spacing of these sequences and the core origin have not been examined. To test this hypothesis, 4 bp was inserted at the *Bgl*II restriction site in the mutant origin regions subcloned into plasmids to alter the relative spacing of the core origin and upstream sequences by about one-half of a helical turn (see Materials and Methods). The sequences of the *in*4 mutants are shown in Table 2. The insertion of 4 bp changes the net length of the genomes in the manner shown. These values represent the deviation in base pairs from the relative alignment present in the wild-type genome. The effects of these insertions on DNA replication in vivo are shown in Fig. 3, and quantitation from the gel is shown in Table 2. The ability of the 225 genome to replicate at wild-type levels was not affected by the insertion of 4 bp. This result was not unexpected, as the original *dl*5 mutation replicates at wild-type levels even in the absence of the upstream sequences (36). The revertant phenotype of the 226r12 deletion mutant (−2 bp) was also unchanged by the

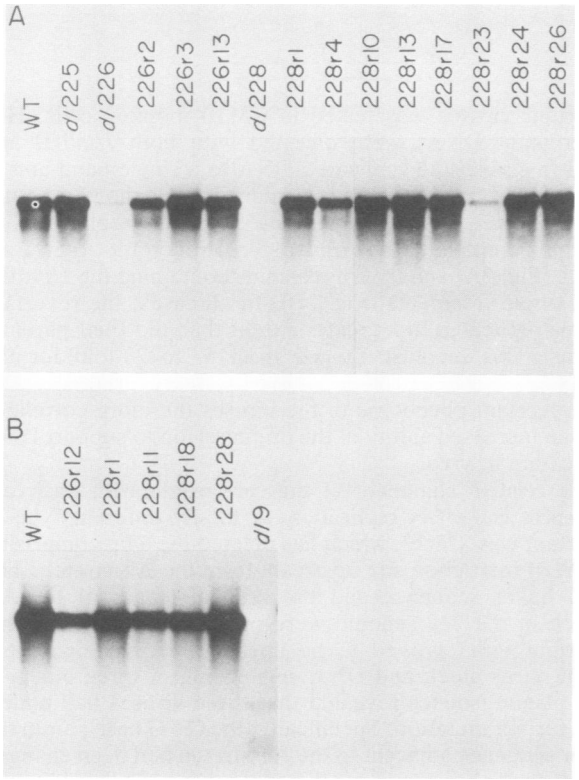



FIG. 2. Replication of subcloned *ori* fragments of revertant viruses. *Hin*dIII-*Nco*I (A) or *Hin*dIII-*Sph*I (B) *ori* fragments were subcloned from the indicated genomes and tested for their ability to replicate after transfection into CMT3 monkey cells. PSVO*dl*9 was used as a negative control. After 2 days of incubation, low-molecular-weight DNA was isolated (18), and the amount of DNA from one-fifth of a 60-mm culture dish was digested with *Hin*dIII and *Dpn*I. After electrophoresis on a 1.4% agarose gel, the DNA fragments were transferred to nitrocellulose and probed with nick-translated pSVO$^+$ DNA. The blot was exposed directly to Kodak XAR-5 film.

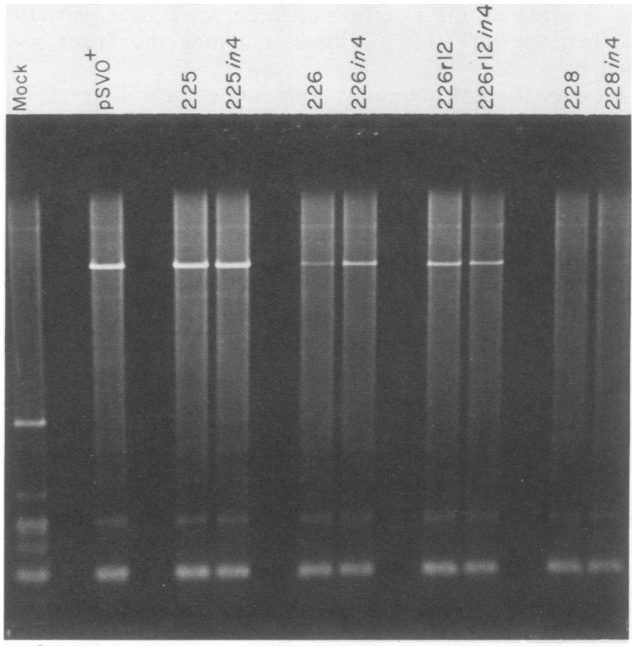



FIG. 3. Replication of 4-bp insertion mutants. Mutant plasmids (100 ng) were transfected into CMT3 cells and extracted by the Hirt procedure after 2 days. The mock transfection was spiked with 0.5 μg of plasmid DNA to assess the efficiency of *Dpn*I digestion. After digestion with *Hin*dIII and *Dpn*I, the DNA from one-third of a 60-mm dish was run on a 1.4% agarose gel, stained with ethidium bromide, and photographed under UV light.

TABLE 3. Deletion analysis of pSVO$^+$ and pSVO 226r12 origins

| Plasmid | Endpoint (np) | Replication (%) |
|---|---|---|
| pSVO$^+$ | 133 | 100 |
| pSVO*dl*22L | 85 | 68 |
| pSVO*dl*16L | 79 | 86 |
| pSVO*dl*18L | 59 | 81 |
| pSVO*dl*14L | 47 | 68 |
| pSVO*dl*23L | 47 | 70 |
| pSVO*dl*19L | 46 | 73 |
| pSVO*dl*28L | 37 | 52 |
| pSVO*dl*39L | 11 | <1 |
| 226r12 | 133 | 38 |
| 226r12*dl*34L | 113 | 53 |
| 226r12*dl*27L | 86 | 56 |
| 226r12*dl*15L | 82 | 59 |
| 226r12*dl*31L | 80 | 62 |
| 226r12*dl*18L | 72 | 38 |
| 226r12*dl*47L | 67 | 9 |
| 226r12*dl*41L | 66 | 9 |
| 226r12*dl*44L | 58 | 5 |
| 226r12*dl*32L | 58 | 5 |
| 226r12*dl*45L | 56 | 2 |
| 226r12*dl*29L | 16 | <1 |

insertion of 4 bp. In this case, the net insertion of 4 bp produces a DNA template which is only slightly different (+2 bp) from the wild-type genome in terms of the relative spacing between the core origin and the upstream sequences. In contrast to these results, both 226 and 228 genomes demonstrate a three- to sixfold increase in the ability to replicate in vivo when 4 bp is inserted at the *Bgl*II site. Even though the 226*in*4 mutant (+8 bp) differs from the wild type by almost a full turn of the DNA helix, its ability to replicate has dramatically increased by comparison with 226 (+4 bp). It is apparent that the 226*in*4 mutation replicates at a level indistinguishable from that of 226r12 (−2 bp), suggesting that either the insertion or deletion of approximately one-half of a helical turn into the replication-defective 226 genome will increase the efficiency of replication. Although not rigorously tested, these results indicate that a change in the relative spacing between the core origin and a sequence element located upstream from the site of the 6-bp deletion in 226r12 was responsible for the observed increase in the level of replication. The upstream elements include the three copies of the 21-bp repeated sequence containing the six G · C boxes which bind transcription factor Sp1 (15) and a small segment (26 bp) of the 72-bp repeat.

**The critical upstream sequences are located with G · C boxes I through III.** To delineate the upstream sequence element involved in the replication of pSVO 226r12, a series of BAL 31 deletions was constructed which progressively removed increasing amounts of the upstream sequences. These deletions started at the *Sph*I site and progressed toward the core origin. A similar series of deletions was constructed in the wild-type origin. The ability of the deleted plasmids to serve as templates for replication in vivo was assayed by transfection into CMT3 monkey cells and analysis of Hirt DNA prepared 50 h posttransfection. The deletion endpoints within the SV40 sequence are shown in Table 3 together with the results of the replication experiments. Deletion mutants in the wild-type sequence seemed to show a slow, progressive decrease in the ability to replicate as the upstream elements were removed. However, even the removal of all upstream SV40 sequences in pSVO*dl*28L resulted in only a twofold decrease in the ability to replicate in vivo. These data are in agreement with our previously published results (36), where we used the same transfection protocol (100 ng of DNA per 6-cm dish of CMT3 cells, chloroquine boost, and the collection of DNA 50 h posttransfection [11]). Further deletion into the AT stretch essentially abolished replication of the wild-type template (5, 36).

In the case of the pSVO 226r12 genome, deletion of the far upstream sequences had essentially no effect on replication of the revertant; removal of the remaining segment of the 72-bp repeat and of G · C boxes IV, V, and VI in deletions ending upstream from np 72 did not result in substantial loss of the DNA replication activity. However, further deletion of sequences resulted in a reduction of the ability of the template to replicate. Replication levels first dropped to the level of the original 226 mutation in *dl*47L and *dl*41L, which removed G · C box III. Loss of additional nucleotides corresponding to sequences over G · C boxes I and II in deletions 44L, 32L, and 45L resulted in an even further loss of the ability to replicate. Finally, deletion of the AT stretch in *dl*29L essentially abolished replication. The analysis of the pSVO 226r12 deletion series indicates that the sequences within the 21-bp repeats, presumably G · C boxes I through III, interact with the core *ori* element in 226r12.

## DISCUSSION

In this study, the sequences essential to the function of the SV40 origin of replication have been analyzed in greater detail. In particular, the effects on DNA replication of sequences on the late side of the core origin were assessed. Sequences upstream from np 30 were previously determined to be unnecessary for efficient *ori* function (36). The present experiments show that restoration of these sequences did not improve the ability of A+T-rich region deletion mutants to replicate, even though the mutants synthesize T antigen at wild-type levels. This allowed us to select for viral revertants which had regained a functional *cis*-acting *ori* element. Characterization of the genomic structure of these revertant viruses revealed alterations in the origin sequence which at least partially restore the length of the AT stretch to that of the wild type in most of the genomes. The restoration of the length of the AT stretch, however, did not reproduce the original sequence, indicating that certain sequence flexibility is tolerated at this region in spite of a strong local homology between origins of SV40, BK, JC, and SA12 viruses (5).

The variant 226r12 (isolated only once) represented a different class of revertants, which led us to observe effects of the upstream sequences on DNA replication in the absence of an intact AT stretch. The increased efficiency of DNA replication in 226r12 was specific for sequences located over G · C boxes I through III and was dependent on the relative alignment of these sequences with the core origin and the A+T-rich region. Specifically, deletion of G · C box III had the largest effect on DNA replication of 226r12 in vivo.

The upstream elements which interact with the core origin coincide with those previously demonstrated to comprise the early promoter sequences and to bind the transcription factor Sp1 (2, 15). The effect on DNA replication of altering the relative spacing between the mutated core origin and the 21-bp repeated sequences containing the G · C boxes correlates precisely with the effect observed previously on the efficiency of initiation of early transcription in SV40 (37). The relative alignment of the 21-bp repeated sequences with either of the two TATA boxes was shown to be critical;

misalignment of these sequences by one-half of a helical turn dramatically reduced the frequency of initiation at early start sites. In our experiments, the proper alignment of the 21-bp repeats with the shortened AT stretch within the core origin (deletion of six nucleotides in 226r12, and insertion of four nucleotides in 226 and 228 mutant origins) was crucial for DNA replication. Two possible mechanisms can be suggested to explain this enhancement in DNA replication. First, there may be direct interaction either between the DNA sequences of the 21-bp repeats and the A+T-rich region or between protein factors which recognize these sequences. Second, the observed effect of the 21-bp repeat could be mediated by nuclease-sensitive chromatin structure caused by the repeats (12, 13, 19, 23). Insertion of one-half of a helical turn of DNA between the wild-type core origin and the 21-bp repeats, however, has no effect on either replication of the template or the nuclease-sensitive structure (19).

Many procaryotic replication origins contain direct repeats located immediately adjacent to AT stretches (35, 40, 42), which are required for origin function (42). In addition to these structural similarities to the SV40 origin of replication, the involvement of promoter sequences in procaryotic DNA replication has been well documented (10, 20). In these instances, the oligoribonucleotide priming process used to initiate DNA synthesis is directed by specific promoter elements which have been localized to specific sequences. Although the initiation of SV40 replication also occurs by a ribopriming process (17), the sequences responsible for this promoter activity have not heretofore been localized on the genome. SV40 replication in vitro is α-amanitin insensitive (24), implying that RNA polymerase II-directed transcription from the early promoter is not involved in the initiation of replication. Either some other enzyme such as DNA primase is utilized to begin the priming process, or the in vitro system does not precisely mimic the in vivo situation. Relevant to this latter possibility are our preliminary observations on the replication in vitro of the mutant DNA templates reported in this study. In collaboration with Bruce Stillman, we have not been able to observe an increase in the replication activity of 226r12 or 226in4 by comparison with 226 in vitro.

The similarities between polyomavirus and the SV40 mutants is particularly striking. The polyomavirus origin contains an AT stretch only 9 bp long adjacent to a G+C-rich 32-bp palindrome and the binding sites for polyomavirus large T antigen. The presence of upstream sequences is absolutely required for *ori* function (27, 41). The upstream sequences which activate polyomavirus replication are localized to the same region as the enhancer element for early gene transcription. The *cis* requirement for a functional enhancer element has been demonstrated by replacing this region with enhancers from either SV40 or the immunoglobulin gene (8). Deletion of the AT stretch abolishes replication, even when upstream sequences are present (J. Hassell, personal communication). Apparently, some process related to early transcription is relevant to polyomavirus DNA replication. The SV40 mutants 226 and 228 are similar to polyomavirus in this regard; they contain a shortened AT stretch, and their ability to replicate is strongly dependent on the presence of upstream sequences normally associated with early gene transcription.

We currently favor the idea that specific protein factors bound to the core origin of SV40 are necessary and sufficient for the initiation of DNA replication on a wild-type template. Certainly T-antigen binding to site II and perhaps other proteins which bind to the AT stretch are involved. Specific protein factors which bind to TATA boxes have been described (4, 29). One may speculate that the AT stretch in SV40 is long enough to bind two such factors, since it contains two TATA boxes located on opposite sides of the DNA helix. These factors may interact with one another and activate the replication process. In the absence of a sufficiently long AT stretch capable of binding two factors, the effect of upstream flanking sequences on replication can be observed. The G · C boxes within the flanking sequences are known to bind the Sp1 protein to form a transcription complex (15), and it may be that Sp1 or some other protein facilitates the initiation of the replication process by interacting with the factor bound to the AT stretch. The specific interaction of a TATA binding protein with an upstream binding factor which facilitates DNA binding and the transcription process has been observed (31). As in procaryotic replication, the local unwinding (26) or bending (45) of the DNA helix by specific replication proteins may be involved in the initiation process. Recent evidence from in vitro competition experiments suggests that SV40 sequences between the core origin and the 21-bp repeats can bind factors other than T antigen which are required for DNA replication (44). However, the nature of such protein factors and the precise mechanism by which initiation occurs remain to be elucidated.

## ACKNOWLEDGMENTS

Our special thanks go to Adam Kansler and Joan Cossetti for technical assistance and to Bruce Stillman, Joe Sambrook, and Winship Herr for critical readings of the manuscript.

This work was made possible by Public Health Service grant CA 13106 from the National Cancer Institute to Cold Spring Harbor Laboratory and by National Research Service Award CA 07493 to R.G.

## LITERATURE CITED

1. **Banerji, J., S. Rusconi, and W. Schaffner.** 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell **27**:299–308.
2. **Barrera-Saldana, H., K. Takahashi, M. Vigneron, A. Wildeman, I. Davidson, and P. Chambon.** 1985. All six GC-motifs of the SV40 early upstream element contribute to promoter activity *in vivo* and *in vitro*. EMBO J. **4**:3839–3849.
3. **Bergsma, D. J., D. M. Olive, S. W. Hartzell, and K. N. Subramanian.** 1982. Territorial limits and functional anatomy of the simian virus 40 replication origin. Proc. Natl. Acad. Sci. USA **79**:381–385.
4. **Davison, B. L., J. M. Egly, E. R. Mulvihill, and P. Chambon.** 1983. Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences. Nature (London) **301**:680–686.
5. **Deb, S., A. L. DeLucia, C.-P. Baur, A. Koff, and P. Tegtmeyer.** 1986. Domain structure of the simian virus 40 core origin of replication. Mol. Cell. Biol. **6**:1663–1670.
6. **DeLucia, A. L., S. Deb, K. Partin, and P. Tegtmeyer.** 1986. Functional interactions of the simian virus 40 core origin of replication with flanking regulatory sequences. J. Virol. **57**:138–144.
7. **DeLucia, A. L., B. A. Lewton, R. Tjian, and P. Tegtmeyer.** 1983. Topography of simian virus 40 A protein-DNA complexes: arrangement of pentanucleotide interaction sites at the origin of replication. J. Virol. **46**:143–150.
8. **deVilliers, J., W. Schaffner, C. Tyndall, S. Lupton, and R. Kamen.** 1984. Polyoma virus DNA replication requires an enhancer. Nature (London) **312**:242–246.
9. **Fromm, M., and P. Berg.** 1982. Deletion mapping of DNA regions required for SV40 early region promoter function *in vivo*. J. Mol. Appl. Genet. **1**:457–481.
10. **Fuller, C., B. Beauchamp, M. Engler, R. Lechner, S. Tabor, J.**

White, and C. Richardson. 1983. Mechanisms for the initiation of T7 DNA replication. Cold Spring Harbor Symp. Quant. Biol. 47:669–679.
11. Gerard, R. D., and Y. Gluzman. 1985. New host cell system for regulated simian virus 40 DNA replication. Mol. Cell. Biol. 5:3231–3240.
12. Gerard, R. D., B. A. Montelone, C. F. Walter, J. W. Innis, and W. A. Scott. 1985. Role of specific simian virus 40 sequences in the nuclease-sensitive structure in viral chromatin. Mol. Cell. Biol. 5:52–58.
13. Gerard, R. D., M. Woodworth-Gutai, and W. A. Scott. 1982. Deletion mutants which affect the nuclease-sensitive site in simian virus 40 chromatin. Mol. Cell. Biol. 2:782–788.
14. Ghosh, P. K., P. Lebowitz, R. J. Frisque, and Y. Gluzman. 1981. Identification of a promoter component involved in positioning the 5′ termini of simian virus 40 early mRNAs. Proc. Natl. Acad. Sci. USA 78:100–104.
15. Gidoni, D., J. T. Kadonaga, H. Barrera-Saldana, T. Takahashi, P. Chambon, and R. Tjian. 1985. Bidirectional SV40 transcription mediated by tandem Sp1 binding interactions. Science 230:511–517.
16. Gluzman, Y., R. Frisque, and J. Sambrook. 1980. Origin-defective mutants of SV40. Cold Spring Harbor Symp. Quant. Biol. 44:293–298.
17. Hay, R. T., and M. L. DePamphilis. 1982. Initiation of SV40 DNA replication *in vivo*: location and structure of 5′ ends of DNA synthesized in the *ori* region. Cell 28:767–779.
18. Hirt, B. 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. 26:365–369.
19. Innis, J. W., and W. A. Scott. 1984. DNA replication and chromatin structure of simian virus 40 insertion mutants. Mol. Cell. Biol. 4:1499–1507.
20. Itoh, T., and J. Tomizawa. 1980. Formation of an RNA primer for initiation of replication if ColE1 DNA by ribonuclease H. Proc. Natl. Acad. Sci. USA 77:2450–2454.
21. Jones, K. A., R. M. Myers, and R. Tjian. 1984. Mutational analysis of simian virus 40 large T antigen DNA binding sites. EMBO J. 3:3247–3255.
22. Jones, K. A., and R. Tjian. 1984. Essential contact residues within SV40 large T antigen binding sites I and II identified by alkylation-interference. Cell 36:155–162.
23. Jongstra, J., T. L. Rendelhuber, P. Oudet, C. Benoist, C.-B. Chal, J.-M. Jeltsch, D. J. Mathis, and P. Chambon. 1984. Induction of altered chromatin structures by simian virus 40 enhancer and promoter elements. Nature (London) 307:708–714.
24. Li, J. J., and T. Kelly. 1984. Simian virus 40 DNA replication *in vitro*. Proc. Natl. Acad. Sci. USA 81:6973–6977.
25. Li, J. J., K. W. C. Peden, R. A. F. Dixon, and T. Kelly. 1986. Functional organization of the simian virus 40 origin of DNA replication. Mol. Cell. Biol. 6:1117–1128.
26. Mukherjee, S., I. Patel, and D. Bastia. Conformational changes in a replication origin induced by an initiator protein. Cell 43:189–197.
27. Muller, W. J., C. R. Mueller, A.-M. Mes, and J. A. Hassell. 1983. Polyomavirus origin for DNA replication comprises multiple genetic elements. J. Virol. 47:586–599.
28. Myers, R. M., and R. Tjian. 1980. Construction and analysis of simian virus 40 origins defective in tumor antigen binding and DNA replication. Proc. Natl. Acad. Sci. USA 77:6491–6495.
29. Parker, C. S., and J. Topol. 1984. A Drosophila RNA polymerase II transcription factor binds to the regulatory site of an hsp 70 gene. Cell 37:273–283.
30. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain- termination inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467.
31. Sawadogo, M., and R. G. Roeder. 1985. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell 43:165–175.
32. Shortle, D., and D. Nathans. 1979. Regulatory mutants of simian virus 40: constructed mutants with base substitutions at the origin of DNA replication. J. Mol. Biol. 131:801–817.
33. Smale, S., and R. Tjian. 1986. T-antigen-simian virus 40 DNA polmyerase α complex implicated in SV40 DNA replication. Mol. Cell. Biol. 6:4077–4087.
34. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by agarose gel electrophoresis. J. Mol. Biol. 98:503–515.
35. Stalker, D. M., R. Kolter, and D. Helinski. 1982. Plasmid R6K replication: complete nucleotide sequence of an autonomously replicating segment. J. Mol. Biol. 161:33–43.
36. Stillman, B., R. D. Gerard, R. A. Guggenheimer, and Y. Gluzman. 1985. T antigen and template requirements for SV40 DNA replication *in vitro*. EMBO J. 4:2933–2939.
37. Takahashi, K., M. Vigneron, H. Matthes, A. Wildemann, M. Zenke, and P. Chambon. 1986. Requirement of stereospecific alignments for initiation from the simian virus early promoter. Nature (London) 319:121–126.
38. Tjian, R. 1978. The binding site on SV40 DNA for a T antigen-related protein. Cell 13:165–179.
39. Tooze, J. (ed.). 1981. DNA tumor viruses. Molecular biology of tumor viruses, 2nd ed., part 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
40. Tsurimoto, T., and K. Matsubara. 1981. Purified O protein binds to four repeating sequences at the replication origin. Nucleic Acids Res. 9:1789–1799.
41. Tyndall, C., G. LaMantia, C. M. Thacker, J. Favaloro, and R. Kamen. 1981. A region of the polyoma virus genome between the replication origin and late protein coding sequences is required in *cis* for both early gene expression and viral DNA replication. Nucleic Acids Res. 9:6231–6250.
42. Vocke, C., and D. Bastia. 1983. Primary structure of the essential replicon of the plasmid pSC101. Proc. Natl. Acad. Sci. USA 80:6557–6561.
43. Wasylyk, B., C. Wasylyk, H. Matthes, M. Winzerith, and P. Chambon. 1983. Transcription from the SV40 early-early and late-early overlapping promoters in the absence of DNA replication. EMBO J. 2:1605–1611.
44. Yamaguchi, M., and M. L. DePamphilis. 1986. DNA binding site for a factor(s) required to initiate simian virus 40 DNA replication. Proc. Natl. Acad. Sci. USA 83:1646–1650.
45. Zahn, K., and F. R. Blattner. 1985. Binding and bending of the λ replication origin by the phage O protein. EMBO J. 4:3605–3616.