Xrep, a plasmid-stimulating X chromosomal sequence bearing similarities to the BK virus replication origin and viral enhancers

Donald E. Riley, Raymond Reeves⁺ and Stanley M. Gartler

Departments of Genetics and Medicine and the Center for Inherited Disease, University of Washington, Seattle, WA 98195 and ⁺Programs in Biochemistry, Biophysics, Genetics and Cell Biology, Washington State University, Pullman, WA 99164, USA

Received 8 August 1986; Revised and Accepted 27 October 1986

ABSTRACT

The human X chromosome-linked fragment, "Xrep," was sequenced because it exerts a positive effect on plasmid growth in both *E. coli* and *Saccharomyces cerevisiae*. The sequence revealed three features similar to the human BK virus replication origin: (1) Xrep has a true palindrome, CCTCC(T)₃CCTCC, which is similar to "true" palindrome-like sequences found at the replication origins of polyoma [CCTC(T/C)₁₀CTCC], BK [CCTC(A/G)₈CCTCC] and SV40 [CCTCC(A)₆GCCTCC] viruses. (2) Twenty nucleotides away from the true palindrome, Xrep has the sequence GAATCCTATTCACCTTTT while BK virus, the human analogue of SV40, has GAAATCCCTATTCTTTT in exactly the same position relative to the true palindrome. These two 17-mers differ only in the positions of two nucleotides comparing Xrep and BK virus. (3) Also similar to the replication origins of DNA viruses, Xrep appears to have a cluster of enhancers adjacent to the origin-like sequences. Potent enhancer-like activity was detected in pSV1-CAT/Xrep constructs. Xrep may originate from an endogenous virus, or from an X chromosomal replication origin.

INTRODUCTION

Because of their relative simplicity, small DNA viruses like SV40 have been used as model systems for the chromosomes of higher eukaryotes. Soeda et al. compared three DNA viruses, BK, SV40 and polyoma, and found that all three had similar "true" palindromes of the general configuration: CCTC(N)_nCTCC at their replication origins (4). Another study in which this same sequence is referred to as an "inverted repeat" showed that deletion of this region results in a drastic reduction in the replicative activity of SV40 (5). The so-called "true palindrome" or inverted repeat also is identical to the T antigen binding region I of SV40 in which mutations lead to defects in DNA replication (6-9). The pentanucleotide contacts that T antigen makes in its binding region I are the most conserved feature of the true palindromes comparing SV40, BK and polyoma (4,6). There is evidence that region I may have a bent DNA conformation (10). Another common feature of polyoma, BK and SV40 viruses is that they all have enhancer sequences near their replication origins (2,11,12).

This report describes the sequence of a cloned, human X chromosomal DNA fragment which has marked similarities with the replication origins of the DNA viruses described above, especially BK. BK is a human virus with 65% to 85% homology with the coding sequences of SV40 (2, 13).

MATERIALS and METHODS

Cell lines and transformations.

The human lymphoblast cell line, GM1416, was propagated in Eagle's minimal medium with 10% fetal calf serum and 40 µg/ml gentamycin under a 5% CO₂ atmosphere. GM1416 contains four X chromosomes per cell as was verified by cytological Barr body analysis. Southern blots with Hind III-digested DNA from various mouse-human hybrid cells were a gift from Dr. Thomas Shows. Procedures for agarose gel electrophoresis and Southern blotting were performed as previously reported (14).

Yeast transformations were performed by the spheroplast technique (15). Spheroplast formation was monitored by phase-contrast microscopy of aliquots from a zymolase-treated cell suspension maintained at 30°C. Saccharomyces cerevisiae 127 has the following genotype: adel,2; leu2; try1; ura3; can1; cyh2; mat a. Yeast cells were grown in either the non-selective, yeast extract-peptone-dextrose (D) medium or supplemented, synthetic minimal medium (17).

Construction of recombinants and deletion mutants.

Bacterial cells were made competent and transformed with recombinant DNA essentially according to Norgard et al. (18). The plasmid, pUC18, and host strain, JM107, were generous gifts from J. Vieira and J. Messing (19). Deletion mutants were constructed by double digestion of the pUC18 polylinker region (eg. Pst I/Xba I) such that exonuclease III could be used to attack the Xrep insert without affecting plasmid sequences (20). Exonuclease III digestion was followed by S1 nuclease treatment and ligation of the resulting blunt-ended molecules. End filling-in with Klenow fragment prior to ligation was found to be unnecessary.

Assays of plasmid recovery.

Plasmid recovery was assayed by a number of methods. Statements made about relative plasmid yield are supported by independent determinations by two or more of these methods. In the first, plasmid-harboring E. coli strains, RRI or JM107, were grown in ampicillin-containing L broth to an OD₆₀₀ of 0.6 followed by chloramphenicol amplification and plasmid purification by the method of Mukhopadhyay and Mandal (21). The resulting plasmid prepara-

tions contained approximately 40% RNA which was subsequently removed either by RNase treatment or by CsCl gradient purification of the plasmids. Subsequent A₂₆₀ readings determined the yield of plasmid.

The second method involved growing plasmid-harboring RRI or JM107 cells to saturation followed by a scaled up (100 ml) Ish-Horowicz and Burke (22) method of purification and equilibrium sedimentation in CsCl gradients. After dialysis against 0.2 mM EDTA (pH 7.0), the yield of plasmid was determined by A₂₆₀ readings.

A third method is less quantitative but its simplicity allowed its application to essentially all of the constructs. Small scale (2 ml) cultures of plasmid-containing JM107 cells were "mini-prepped" (22) and a 5 microliter aliquot of each miniprep was digested with an appropriate restriction enzyme, run on agarose gels containing ethidium bromide and photographed. Visual inspection of the staining intensity, comparing 12 to 14 minipreps per gel, yielded a fairly reliable impression of relative yield.

Nucleotide sequencing.

DNA sequences were determined by the method of Sanger et al. (23) employing ³⁵S-dATP and buffer gradient gels (24). GC-rich regions were sequenced using AMV reverse transcriptase at 55°C with 10-fold elevated free nucleotide concentrations. The reverse transcriptase is approximately 10-fold more sensitive to dideoxynucleotides than is Klenow fragment (25). Sequence ambiguities were resolved by sequencing of both DNA strands.

Transfection and assay of CAT activity.

For the detection of enhancer-like activity, 4-5 X 10⁵ COS-1 cells in logarithmic-phase growth were transfected with 10 ug of plasmid DNA (supercoiled, carrier-free) by an optimized calcium phosphate procedure (37). Thirty-six hours after transfection, the cells were harvested, then extracts were made and assayed for chloramphenicol acetyltransferase (CAT) activity by thin layer chromatography (30).

RESULTS

The initial interest in Xrep was due to unanticipated observations regarding plasmid production in E. coli. Total genomic DNA from a human lymphoblast cell line (GM1416, Methods) containing an abnormal number of X chromosomes (four per cell) was digested with BamH I and electrophoresed on a 0.75% agarose gel. DNA fragments migrating in the 1.9 to 2.3 kb region were cut out, electro-eluted and cloned into the BamH I site of pBR322. It was noticed that one of the clones (now termed "pBXrep") was obtained at a signif-

FIGURE 1. Characteristic increase in recovery of plasmids containing Xrep. Shown are ethidium bromide stained 0.75% agarose gels of minipreps with (+) and without (-) portions of Xrep.

a. The two lanes on the left are pBR322 and Xrep subclone 3C15 (Fig. 4) respectively both doubly digested with HindIII and KpnI. Other lanes labeled with "+" are independent, BamHI digested clones of Xrep inserted into the BamHI site of pUC18. The + lanes are vastly overloaded, but two BamHI fragments (Xrep and pUC18) of the correct size are present. The lanes labeled "-" represent uncut pUC18 (forms I, II and III, as well as minor concatenated forms) with the BamHI site removed. Minipreps were treated and loaded equivalently (ie. identical volumes) except that Xrep-containing clones obviously result in more plasmid DNA. The plasmid in the lane labeled + is about twice as large as the other plasmids, but still shows some increase in intensity due to an Xrep fragment.

icantly higher yield (10-to 20-fold) than other plasmids similarly isolated. For example, plasmids pBD-31 and pBD-41 had anonymous, genomic inserts identical in size to that of pBXrep and were recovered at 160 ug and 100 ug DNA, respectively, per 100 ml culture while pBXrep was recovered at 1.8 mg per 100 ml. Bacterial cell densities were identical in each case. The Xrep fragment was subsequently cloned into pUC18 to determine if the positive effect on plasmid yield was due to some accidental combination of an X-linked DNA fragment with vector sequences. The Xrep fragment was recovered in both orientations, pUCXrep1 and pUCXrep3, in the BamHI site of pUC18 (Methods).

Several independent clones identical with either pUCXrep1 or pUCXrep3 consistently showed a positive effect on the production of plasmid (Fig. 1). Typically, Xrep-containing recombinants (denoted by "+") appear vastly overloaded while empty pBR322 or pUC18 vectors are much reduced in intensity. Quantitative determinations (Methods) revealed a 10- to 20-fold increase in

FIGURE 2. Construction and analysis of yeast vectors with and without Xrep.

- a. Shown is pXrYIP2 in which Xrep (crosshatched) is inserted into the BamHI site of YIP5. Open regions indicate pBR322 DNA. Abbreviations are: R, EcoRI; B, BamHI; S, SalI; Ba, BalI; K, KpnI.
- b. Shown is pXrYrp which has Xrep inserted into the BamHI site of YRp7' (27). The TRP1 selectable marker and the ARS are previously described (27).

recovery in numerous isolates of Xrep compared with controls. Chloramphenicol amplification of cultures increased the yield of plasmids, both with and without Xrep, so that the 10- to 20-fold increase in yield of Xrep-containing plasmids persists independently of chloramphenicol stimulation. Bacterial colony size is not affected by Xrep, and liquid cultures typically grew to the same cell density whether or not Xrep was present in the plasmid. Thus, the effect of Xrep is likely to be either at the level of plasmid replicative activity or stability. Since pUC18 and pBR322 have quite different sequences flanking the Xrep insert, we conclude that the positive effect resides in the Xrep fragment itself.

Plasmid stimulation in yeast.

In order to determine the effects of Xrep on plasmid replication in a eukaryote, the constructions shown in Figure 2 were performed. In figure 2a, Xrep has been inserted into the BamHI site of YIP5, a test vector for autonomously replicating sequences (ARS) in yeast (26). The resulting recombinant, pXrYIP2 was tested for ARS activity in *Saccharomyces cerevisiae* 127 under URA^+ selection (Methods). As a control, strain 127 was transfected with YIP5 which resulted in no transformants on -uracil plates after 4 days at 30°C. Plasmid pXrYIP2, on the other hand, after 4 days, produced very small (1-2 mm) slowly growing colonies suggesting very weak replicative activity. However, the latter transformants were abortive and did not grow upon

Table I

	No. of colonies counted		Colony forming units per ml (10^{-4})		% of cells with plasmid
	D	-Trp	D	-Trp	
pXrYrp-1	137	45	1,370	21	1.5%
pXrYrp-2	334	174	1,670	179	7.7%
pXrYrp-3	335	271	3,350	100	3.0%
pXrYrp-4	558	379	5,580	379	6.8%
pXrYrp-7	378	27	378	7	2.0%
YRp7'-1	153	23	1,530	4	0.26%
YRp7'-2	111	10	1,110	5	0.45%
YRp7'-3	307	44	3,070	16	0.52%
YRp7'-4	220	19	2,200	5	0.23%

Determination of plasmid stability in *Saccharomyces cerevisiae*. Independent colonies of YRp7' and pXrYrp (Fig. 2) grown on plates lacking tryptophan were suspended and plated on selective (-trp) and nonselective medium (D). Cells lacking plasmid are maintained by cross-feeding and by residual gene product.

restreaking to -uracil plates. Since Xrep reproducibly appeared to have weak replicative activity in yeast, we determined whether Xrep might further stabilize a plasmid which already has ARS activity.

For this purpose, Xrep was inserted into the BamH I site of YRp7' to form pXrYrp (Fig. 2b). YRp7', by itself, is an unstable plasmid under TRP⁺ selection. Only a relatively small percentage of cells in a colony derived from a transformed cell actually contain plasmid (Table I). By plating dispersed cells from individual, transformant colonies on selective (-tryptophan) and nonselective media (D) we determined the percentage of cells containing plasmid in the case of both YRp7' and pXrYrp. On the average, YRp7' was maintained in 0.36% of cells while pXrYrp was maintained in 4.2% of cells, an average 11-fold increase (Table I). This increase was reflected in a similar increase in total plasmid DNA per culture as evidenced by Southern blot hybridization of total yeast DNA using pXrYrp as a probe (not shown).

We conclude that, while Xrep does not have full ARS activity by itself in yeast, it does increase significantly the replication or stability of a plasmid that already contains an ARS element.

FIGURE 3. X chromosome localization of the Xrep fragment.

- a. Human fibroblast DNAs were digested with EcoR I. Equivalent amounts were Southern blotted and probed with Xrep. XY, cell line 456, a human male fibroblast line; 3X, 73-126, an abnormal human female fibroblast cell line with 3 X chromosomes per cell; 4X, GM1416, an abnormal human cell line with 4 X chromosomes per cell.
- b. Various mouse, human, and mouse-human hybrid DNAs. "no X," JWR-22H, mouse-human hybrid with no X chromosome. The two lanes marked "X" are NSL-9 and XTR-22 hybrids, respectively, with the human X chromosome. HU, human DNA; MO, mouse (RAG) DNA.

Localization to the X chromosome.

The data of Figure 3 show that Xrep is present in approximately two distinct copies on the human X chromosome. Briefly, Southern blot hybridization intensity of Xrep increases with the number of X chromosomes per cell. Similar analysis of mouse-human hybrid DNAs confirmed the X chromosome localization (Fig. 3). The Xrep copies are distinct because HindIII digestion of genomic DNA results in two bands (approximately, 14 and 19 kb) while the probe itself contains no HindIII sites. Hybridization intensity of Xrep is comparable with that of human PGK (phosphoglycerate kinase) a gene present in one copy, plus a pseudogene on the X chromosome (14). The possibility that some small portion of Xrep is highly repeated cannot be excluded since slight, background smearing within lanes is more evident with Xrep than with other probes used on the same blots (not shown). The possible significance of the

FIGURE 4. Restriction map, G+C content and deletion mutants of Xrep.

- Restriction map of Xrep. The large open box indicates the cluster of enhancer-like sequences (ELS) found within Xrep α . The black box corresponds to a 19 base pair true palindrome described in the text. B, BamH I; P, Pvu II; X, Xba I; K, Kpn I; H, Hinf I; R, EcoR I; M, Msp I.
- G+C content. The numbers between tick-marks indicate the % G+C content for each 120 base segment as determined from the nucleotide sequence.
- Shown are the positions of Xrep α and Xrep β , two regions found to be important in the positive effect on plasmid growth in *E. coli*.
- Deletion mutants of Xrep. The names of these subclones are found over the common end of a series. That is, Os9, Os2, etc. have various deletions of the more GC-rich end (right). Note: X20 is a different subclone from E21 (labeled on right); similarly, X19 is distinct from E16. The deletion mutants are drawn this way to conserve space. Each mutant contains only sequences indicated by the presence of a black line.

latter is that a genomic replication origin might have a short consensus sequence which is highly repeated and dispersed (Discussion).

Construction of deletion mutants for sequencing.

During the construction of deletion mutants, recovery of plasmids from *E. coli* was monitored by several different assays (Methods). All of the assays revealed an occasional 2- to 3-fold variation in recovery of the same plasmid construct isolated as different clones. However, two regions of Xrep (α and β , Fig. 4) were consistently found to be important in the positive effect originally observed in *E. coli*.

The upper bands of Figure 5 illustrate some of the findings of these comparative studies. Firstly, plasmids lacking the α region but containing the β region (eg. the E(n) series plasmids, Fig. 5b) retain a residual positive effect. This effect decreases when sequences near the left end of subclone E29 (Figs. 4 and 5b) are deleted, thus defining the left, or upstream, border of region β . Also, region α has little or no positive effect in the absence of region β (e.g. subclones X19, X20 and X27, Fig. 5a). The X(n) plasmids shown contain only portions of the α region, but plasmids Os4 and Os50 contain the entire α region (Fig. 4) and were also produced at unstimulated levels. In contrast, quantitative assays consistently showed that the α region in undeleted Xrep increases the effectiveness of the β region by two- to three-fold. The right hand borders of the α and β regions were defined by increased production of subclone 3C15 over both E21 and OS2 (Fig. 4). Thus far, studies of internal deletions of Xrep are minimal.

Nucleotide sequence analysis.

The 2,356 bp Xrep sequence was determined (Methods). Computer assisted homology searches revealed an approximate 500 base region of Xrep with numerous close homologies with known enhancers (ELS, Fig. 6). In most cases, the homologies shown are 100%. For example, within a 270 base pair region of sequence Xrep α were found 4 copies of the so-called Ela enhancer core sequence (11, 28, 29) ~~CTTGGT~~
GAACCA which in one case is repeated in tandem and, in another case, the sequence is inverted. Statistically, a given hexameric nucleotide sequence is expected to occur about once every 4,100 base pairs (31); so the appearance of the Ela enhancer core sequence four times in 270 base pairs appears significant. Additionally, the same 270 base pair sequence contains one copy of a fifth sequence ~~CTATTG~~
CATAAACGCTCT which is distinct from the Ela sequence but closely resembles the SV40 enhancer core sequence (11). Two separate 11-mers of the ELS are identical to 11-mers repeated in the SV40 (34) and BK virus enhancers (2) respectively. Xrep/pSV1-CAT constructs described below indeed revealed potent enhancer-like activity.

Within the ELS sequence of Xrep α , a 19 base pair true palindrome was found which is quite similar to true palindromes located near the replication origins of SV40, BK and polyoma viruses (4). These are the region I sequences described in the Introduction. In Figure 7a is shown a comparison of viral region I sequences and the Xrep true palindrome. According to Day and Blake (39) true or "mirror" palindromes 16 or more bases long are statistically expected to be quite rare. Using their equations, we estimate that the probability of finding a 19 base true palindrome anywhere within Xrep is 0.01.

and within the ELS, 0.002. The similarities with viral true palindromes diminishes the probability of chance occurrence still further.

Finally, within the ELS region of Xrep α and 20 bases away from the true palindrome of Figure 7a is the sequence 5'GAATCCTATTCACITTTT3'. We note that 20 bases away from the "true" palindrome of the human virus, BK, there is a nearly identical (except for the position of two bases) sequence 5'GAAATCCCTATTCITTTT3' (13). The function of this sequence is not known, but it is within the non-coding, replication origin and transcriptional control region of BK. A schematic comparison of Xrep sequences with the BK virus replication origin region is shown in Figure 7b. It is difficult to estimate the probability of finding the arrangement shown for Xrep in Figure 7b. However, the probability of finding a true palindrome of this size in the exact position shown is less than 4×10^{-6} not taking into account the similarities with viral true palindromes.

The distances between the true palindrome and the first enhancer core sequence is 100 bases for BK virus and 120 bases for Xrep. The open reading frame (orf) shown (Fig. 7b) for Xrep has little or no significant homology with viral peptide coding sequences but may encode a cellular message as indicated by RNA dot blots (not shown). Inversion of the bracketed DNA fragment (Fig. 7b) would lead to very similar positioning of the enhancers, true palindromes, the 17-mers shown as well as the open reading frames (Discussion).

FIGURE 5. Miniprep comparisons of Xrep deletion mutants. Identical size aliquots of minipreps were cut with PvuII which in all cases results in a 2.4 kb band (top bands) derived entirely from the vector. Since different size DNA fragments bind proportionately, different amounts of ethidium bromide, the identical 2.4 kb bands in each case were used for comparative purposes. The variable size, middle bands of the E(n) series represent various inserts.

Subclone X27 and other X(n) series plasmids are consistently produced at about the level of empty pUC18 and other non-Xrep-containing plasmids. The two dark arrows in panel b show the sub-clones where a decline of the residual positive effect is consistently observed leading to the designation of the β region (Fig. 4). There is a gradual increase in recovery (see the upper bands) as plasmid size decreases down to the size of subclone E7. This plasmid size effect is well known and actually mitigates in favor of our conclusions.

Residual, uncut, supercoiled molecules, not to be confused with PvuII fragments, are indicated by arrows in panel a. The fastest migrating bands derive from both vector and insert (exceptions: X27 and pUC18).

Sequences homologous with viral protein coding sequences such as those of BK or SV40 were not found within Xrep. Also not found were sequences homologous either with known prokaryotic replication origins, promoters or with yeast ARS sequences.

Tests for enhancer activity.

In order to test for enhancer activity, Xrep was inserted 5' to the chloramphenicol acetyltransferase gene (CAT) of plasmid pSV1-CAT (30) forming pRS23. Xrep was inserted between the two HaeII sites of pSV1-CAT. Similarly, a 3' construct (pSCR5) was made with Xrep in the BamHI site 3' to the CAT gene

GATCCAGTGC	CAGACAGCTG	TATGAACACG	CTCTCCTGGC	TACCTGGGCC	CCATCTGTTT	60
GCTTCGGGCT	GACAGTGAGA	TAAGGGACAA	TGACTTTTCC	CTCCTGAGCC	ACATGTTAGT	120
ELS						
GAAGCGGTGG	AGATAATTCT	GCTCTGAGTC	CTGACGGGAA	GTTCCTGGGG	AGCAACTTTA	180
CAGTGCTCCT	GGCTTGGGGC	ACTACCCACT	CGGGGCTTTG	TGGCAAGCTA	TTCCCCCTCA	240
true palindrome						
CGGCCAAGGA	ACCTGGCTTC	CCATACCAAC	TCCTTTCTTC	CACCCACCGT	CTCCCTATCT	300
(GAATCCTATT	CACCTTTT	CAG	GTCTTGGCTC	AAATCCTGAC	TCCCTGTGAA	360
CCCTCCCGAGC	CTAGAGCCAC	TACTCCCTCC	TGGAATCTCA	GGATCTCCAA	CTGTGACATG	420
1						
GCCATACTAT	TTGCCACATC	CCCACAGACG	GATTGTGATG	AGGAAGCAAA	GAGAGAGTCC	480
2						
ATGTAGCATG	CTAGATGCCG	TGTCTGGTGC	AGGCCTGTGG	TGAGTTCCCT	GCCTCTTCCT	540
3						
CTTCCTCCCC	TCTGACCCCC	AGCATGCAGG	CATGGCGGCC	TTGATCACGT	GCTGCCTTCT	600
CTAGATAATC	CTTGGGGGATC	ATTCCCTGTG	TAGTTGTCCC	CTCAGTTGCA	ATGTCACAGA	660
4						
5						
CCATGGGGCC	ACATCTCATC	CTTCCTCACCT	TTCCAGCATC	CAGCATACAA	TAAACACCAT	720
CGTAACCAT	TTTCATTGGG	CATCTTAGGG	TACCATGTGG	CAGGCACTCG	GCTGTGTAGG	780
TAATGCACAC	AAGGTAGTGC	CTCACTCATT	CCCCTCCTAG	CTCTCTTCCA	GGATGTGCAG	840

FIGURE 6. Nucleotide sequence of the portion of Xrep bearing similarities with viral replication origins. The 19 base pair "inverted repeat" or true palindrome is boxed and labeled. As shown in Figure 7a, the CCTCC repeats within this true palindrome are similar or identical to repeats found in three DNA viruses. Twenty nucleotides from the second CCTCC repeat, the sequence in parentheses is similar to a BK viral sequence (Fig. 7b). Enhancer core sequences are boxed and numbered in order of appearance. Core sequence #1 is nearly identical to the core sequence first described in SV40 (35, 36). Core sequences 2-5 are identical to the Ela homology or core sequence (11, 12, 28) except that #2 is on the complementary strand. The 11 base sequences at nucleotides 164-174 and 461-471 are identical with repeats found in the SV40 (34) and BK (2) virus enhancers, respectively. Reversed arrows indicate that the homology is on the strand complementary to those published (2, 34). Abbreviations: ELS, enhancer like sequence.

of pSV1·CAT. Additionally, the 753 bp BamHI to KpnI Xrep fragment containing the ELS region was inserted 5' to the CAT gene forming pRE9.

The Xrep/pSV1·CAT constructs were transfected into monkey COS-1 cells (Methods). Thirty six hours after transfection, cell extracts were assayed for CAT activity (30). Also transfected and assayed were pSV1·CAT which has only 22 base pairs of one SV40 enhancer repeat and pSV2·CAT which has both 72 bp SV40 enhancers intact. Although its activity is reduced, the control pSV1·CAT still has three times the transcriptional enhancement of background (30). Both pRS23 (Xrep 5' to the CAT gene) and pRE9 (ELS 5' to CAT) consistently yielded 4 to 6-fold stimulated CAT activities compared to the control pSV1·CAT (Fig. 8). The enhancement level observed for pSV2·CAT (complete SV40 enhancers) was nearly identical to the level observed for pRS23 and pRE9. Thus, in our hands, the Xrep ELS region appears close to the potency of the complete enhancers of SV40. The 3' Xrep construct, pSCR5 also resulted in significant activation of the CAT gene although the effect was less (1.5 to 2

FIGURE 7. Structural features of Xrep and small DNA virus sequences.

- a. True palindromes in small DNA viruses and Xrep. Unlike an ordinary palindrome, a "true palindrome" is not self-complementary along one strand. Rather, it reads the same in forward and reverse directions on one strand. Soeda et al. (4) first compared the imperfect, true palindromes of Polyoma (Py), BK virus (BK) and SV40. Here, these are redrawn emphasizing that the pentanucleotide T antigen contact sites (i.e. the CCTCC repeats, except for a one base lateral shift) of SV40 (6-9) are mostly conserved among the three viruses. Xrep is seen to have a similar true palindrome with six additional, symmetrically added nucleotides. Justifying this comparison is the sequence environment of the Xrep true palindrome (below).
- b. Comparison of the BK virus replication origin region with Xrep. Top, BK virus; Bottom, Xrep. Twenty nucleotides away from the true palindromes (compared in a.) a nearly identical (except for nucleotides) 17 base sequence is found. Although not drawn to scale, the distance between true palindromes and enhancer core sequences is comparable (100 to 120 bases) in both cases.

fold) than the 5' constructs. Possibly, this is due to a "shielding" effect of an extremely GC-rich region (Fig. 4) situated between the ELS region and the CAT gene of pSCR5.

Since the control plasmid, pSV1·CAT, has a low level of CAT activity, it is possible that the apparent enhancement might, at least in part, be due to increased plasmid replication due to Xrep. All of the constructs in this section, including pSV1·CAT, already contain the SV40 origin of replication. The fact that Xrep inserted 5' to the CAT gene consistently results in 2 to 3 times the stimulation of the 3' construct suggests an effect on transcription. However, in order to clearly distinguish enhanced replication from enhanced transcription, the enhancer-like region of Xrep will be inserted into pSV0·CAT (30) which has no background replication or transcription in COS-1 cells. Assays for quantitating replication are also being devised for this system.

Figure 8. Enhancer-like activity of pSV1-CAT/Xrep constructs. Shown is an autoradiograph of CAT assays of COS-1 cell extracts 36 hrs after transfection with pSV1-CAT/Xrep constructs. Lanes: (1) Control COS-1 cell extracts (no DNA); (2) Cells transfected with pSV2-CAT (complete SV40 enhancers); (3) Cells transfected with pSV1-CAT (control); (4) Cells transfected with pSCR5 (Xrep at 3' end of CAT gene); (5) Cells transfected with pRS23 (Xrep at 5' end of CAT gene); (6) Cells transfected with pRE9 (the 750 bp enhancer-like region of Xrep at the 5' end of the CAT gene). Key: Ori-, origin; Un-, unacetylated chloramphenicol; Mono-, monoacetylated forms of chloramphenicol (two species); Di-, diacetylated chloramphenicol.

DISCUSSION

Positive effect(s) on plasmid yield were observed when Xrep-containing plasmids were transfected into the heterologous systems, *E. coli* and *Saccharomyces cerevisiae*. While these heterologous system effects can be said to parallel the effects of region I sequences (below and Introduction), we will first discuss Xrep in the framework of the homologous system, human cells.

Comparing X rep with viral replication origin enhancer regions, there were remarkably similar features including similar true palindromes, similar 17 base sequences (BK) near the true palindromes, and the presence of enhancer-like sequences. Further, while we think Xrep is more closely

related to BK virus, other similarities were found comparing Xrep with the more thoroughly studied SV40 core replication origin (38). For example, 4 bases downstream of the true palindrome of SV40 is an imperfect, ordinary palindrome, CTACTTCTGGAATAG, known as the "early palindrome." Similarly, 4 bases from the Xrep true palindrome, on the side opposite the 17 base BK homology, is an imperfect, ordinary palindrome, GGAACCTGGCTTCC, of similar size and position to the SV40 early palindrome, although the actual sequences bear little homology. For a comparison of similar structures at a variety of prokaryotic and viral replication origins see ref. 39.

Despite the similarities, however, it should be considered that there is at least one substantial difference between Xrep and well-characterized viral replication origins. That is, an A/T-rich region found within replication origins of BK, polyoma and SV40 viruses (4) is not found near the true palindrome of Xrep. In the case of SV40, deletions into the A/T-rich region greatly reduce viral replication (5, 9) so that the A/T-rich region seems an important part of the replication origin. BK virus also has an A/T-rich segment close to the true palindrome. One interpretation is that Xrep contains an endogenous, incomplete remnant of a viral replication origin related to BK. We cannot exclude this; however, for a number of reasons, a less trivial interpretation seems more likely.

For example, a chromosomal replication origin which replicates once per cell cycle would be expected to differ from a viral replication origin due to the former's more stringent requirements in tissue-specific replication timing (32). Possibly, an A/T-rich region located near such a chromosomal replication origin would compromise some aspect of replication control.

Also, viral DNA probes detect relatively few sequences (1 to 3) homologous with DNA viruses in monkey or human genomes (1, 2). These viral-like sequences may or may not have cellular functions. Rosenthal et al. reported a human DNA sequence homologous with enhancers of the BK virus (2). The sequence of Rosenthal et al. is distinct from sequences reported here, except that their sequence, like Xrep, has multiple repeats of the Ela enhancer core sequence (Results). In contrast with the relative, low number of SV40 or BK viral-like sequences found in primate genomes, there is evidence that the number of replication origins is in the thousands (32). Since viral probes were not employed by us in the isolation of Xrep, it is much more probable, based on the numbers given above, that Xrep represents an X chromosomal replication origin, or related sequence, than a viral remnant. From this perspective, the similarities between Xrep and BK virus are based

entirely on the need of the virus to utilize the host replication and transcription machinery. What is clear from studies which employ viral probes is that the majority of genomic replication origins do not share sufficient homology with viral origins of replication to be detected in large numbers (1, 2). However, under conditions of lowered stringency, increased numbers of genomic sequences are detected (3).

As mentioned previously, the region I sequences of SV40 have been shown to greatly facilitate replication of the virus (6-9). Although region I sequences initially interact with T antigen, there is likely a later interaction with DNA polymerase α . DNA polymerase I from Saccharomyces cerevisiae in many ways resembles DNA polymerase α from higher eukaryotes (33) which might account for a 11-fold increase in replication in yeast (Table I) of a plasmid which carries Xrep, region I-like sequences. This is speculation only and it is unclear whether the positive effect observed in E. coli could be similarly explained. However, it would be surprising if the effects observed in E. coli and Saccharomyces cerevisiae were totally unrelated to one another and also independent of the similarities with BK virus.

Deletion analysis showed that the Xrep β region independently exerts a "residual" positive effect on plasmids in E. coli, while a further positive effect exerted by the α region depends on the presence of the β region. Thus, although the α region contains striking similarities with viral replication origins as well as strong enhancer-like activity, other sequences within Xrep may also be of interest. Mammalian systems will be used to further characterize these sequences.

Acknowledgements

Appreciation is extended to Dr. Walton Fangman and colleagues who lent much assistance with the yeast vector studies. Thanks to Theresa K. Canfield for outstanding technical assistance and to Kenneth Johnson, Ph.D., who performed CAT assays. This work was supported by the National Institute of Child Health and Development Grant 1R01 HD16659 and by NSF Grant PCM 8309322. S.M.G. is the recipient of a Research Career Award from the National Institutes of Health.

REFERENCES

1. McCutchan, T. F. and Singer, M. F. (1981) Proc. Natl. Acad. Sci. USA **78**, 95-99.
2. Rosenthal, N., Kress, M., Gruss, P. and Khoury, G. (1983) Science **222**, 749-754.
3. Conrad, S. E. and Botchan, M. R. (1982) Mol. Cell Biol. **2**, 949-965.

4. Soeda, E., Arrand, J. R., Smolar, N. and Griffin, B. E. (1979) Cell **17**, 357-370.
5. Stillman, B., Gerard, R. D., Guggenheimer, R. A. and Gluzman, Y. (1985) EMBO **4**, 2933-2939.
6. DeLucia, A. L., Lewton, B. A., Tjian, R. and Tegtmeyer, P. (1983) J. Virol. **46**, 143-150.
7. DiMaio, D. and Nathans, D. (1980) J. Mol. Biol. **140**, 129-142.
8. DiMaio, D. and Nathans, D. (1982) J. Mol. Biol. **156**, 531-548.
9. Bergsma, D. J., Olive, D. M., Hartzell, S. W. and Subramanian, K. N. (1982) Proc. Natl. Acad. Sci. USA **79**, 381-385.
10. Ryder, K., Silver, S., DeLucia, A. L., Fanning, E. and Tegtmeyer, P. (1986) Cell **44**, 719-725.
11. Herbolmel, P., Bourachot, B. and Yaniv, M. (1984) Cell **39**, 653-662.
12. Veldman, G., Lupton, S. and Kamen, R. (1985) Mol. Cell Biol. **5**, 649-658.
13. Yang, R. C. A. and Wu, R. (1979) Science **206**, 456-462.
14. Riley, D. E., Canfield, T. K. and Gartler, S. M. (1984) Nucleic Acids Res. **12**, 1829-1845.
15. Hinnen, A. H., Hicks, J. B. and Fink, G. R. (1978) Proc. Natl. Acad. Sci. USA **75**, 1929-1933.
16. Mechali, M. and Kearsey, S. (1984) Cell **38**, 55-64.
17. Sherman, F., Fink, G. R. and Hicks, J. (1979) Methods in Yeast Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
18. Norgard, M. V., Tocci, M. J. and Monahan, J. J. (1980) J. Biol. Chem. **255**, 7665-7672.
19. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene **33**, 103-119.
20. Yoshitake, S., Schach, B. G., Foster, D. C., Davie, E. and Kurachi, K. (1985) Biochemistry **24**, 3736-3750.
21. Mukhopadhyay, M. and Mandal, N. C. (1983) Analyt. Biochem. **133**, 265-270.
22. Ish-Horowicz, D. and Burke, J. F. (1981) Nucleic Acids Res. **9**, 2989-2998.
23. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463-5467.
24. Biggin, M. D., Gibson, T. J. and Hong, G. F. (1983) Proc. Natl. Acad. Sci. USA **80**, 3963-3965.
25. Smith, A. J. (1980) Methods in Enzymology **65**, 560-580.
26. Botstein, D. and Davis, R. W. (1982) In: The Molecular Biology of the Yeast Saccharomyces. Eds. Strathern, J. N., James, E. W. and Broach, J. R. Cold Spring Harbor Laboratory 621.
27. Stinchcomb, D. T., Struhl, K. and Davis, R. W. (1979) Nature **282**, 39-43.
28. Hearing, P. and Shenk, T. (1983) Cell **33**, 695-703.
29. Veldman, G. M., Lupton, S. and Kamen, R. (1985) Mol. Cell Biol. **5**, 649-658.
30. Gorman, C. M., Moffat, L. F. and Howard, B. H. (1982) Mol. Cell Biol. **2**, 1044-1051.
31. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) In: Molecular Cloning a Laboratory Manual, Cold Spring Harbor Laboratory, p. 4.
32. Hand, R. (1978) Cell **15**, 317-325.
33. Chang, L. M. S. (1977) J. Biol. Chem. **252**, 1873-1881.
34. Gruss, P., Dhar, R. and Khoury, G. (1981) Proc. Natl. Acad. Sci. USA **78**, 943.
35. Laimins, L. A., Koury, G., Gorman, C., Howard, B. and Gruss, D. (1982) Proc. Natl. Acad. Sci. USA **79**, 6453-6457.
36. Wether, H., Konig, M. and Gruss, P. (1983) Science **219**, 626-631.
37. Gorman, C., Padmanabhan, R. and Howard, B. H. (1983) Science **221**, 551-553.
38. Deb, S., DeLucia, A. L., Baur, C., Koff, A. and Tegtmeyer, P. (1986) Mol. Cell Biol. **6**, 1663-1670.
39. Day, G. R., and Blake, R. D. (1982) Nucl. Acids Res. **10**, 8323-8339.