

Long-term stability of large insert genomic DNA episomal shuttle vectors in human cells

Richard Wade-Martins, Jon Frampton¹ and Michael R. James*

The Wellcome Trust Centre for Human Genetics, Nuffield Department of Clinical Medicine, University of Oxford, Windmill Road, Oxford OX3 7BN, UK and ¹Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DS, UK

Received November 25, 1998; Revised and Accepted February 12, 1999

ABSTRACT

We have constructed an episomal shuttle vector which can transfer large (>100 kb) human genomic DNA inserts back and forth between bacteria and human cells and which can be tracked in rapidly dividing human cells using a live cell assay. The vector (p5170) is based on the F factor-derived bacterial artificial chromosome cloning vector used in *Escherichia coli*, with the addition of the family of repeats element from the Epstein–Barr virus (EBV) latent origin of replication. This element provides nuclear retention in cells expressing the EBV protein EBNA-1. We have subcloned a series of genomic DNA inserts into p5170 and transfected the constructs into an EBNA-1⁺ human cell line. Episomal mitotic stability was quantitatively analysed using flow cytometry. The episomes were also tracked by time course photography of expanding colonies. A 117 kb episome was retained at ~2 copies/cell and could be shuttled unrearranged from the human cells into bacterial cells after 15 months of continuous cell growth. Furthermore, the episome could still be rescued from human cells cultured in the absence of selection for 198 days. Such a trackable *E.coli*/human cell line shuttle vector system capable of carrying >100 kb of genomic DNA in human cells could prove a valuable tool in gene expression studies.

INTRODUCTION

The recent development of P1-based artificial chromosome (PAC) (1) and bacterial artificial chromosome (BAC) (2,3) human genomic DNA libraries has greatly aided physical mapping and large-scale sequencing projects. However, PAC and BAC systems have yet to be exploited to their full potential for functional expression studies. There are considerable advantages to be gained from using genes delivered as regions of large genomic DNA, including introns and long stretches of flanking sequences, in transgene expression vectors (4–7). Non-coding flanking DNA regions are usually omitted from small cDNA-based expression vectors but may contain controlling elements important for subtle regulation of tissue-specific gene expression (8).

Functional expression studies in murine cells and in transgenic mice have used BACs (9,10) and yeast artificial chromosomes (YACs) (4,6,11,12). Equivalent experiments have not been successfully performed in human cells because it is extremely difficult to obtain stable human cell lines transformed with large fragments of intact DNA (13–16). A systematic means to introduce and retain large inserts of genomic DNA in human cells would be a powerful tool in studies of gene expression.

Mammalian artificial chromosomes may prove suitable vectors for expression studies from genomic DNA (17–20). An alternative approach may be to develop a shuttle vector which could move large genomic DNA fragments freely between human and bacterial cells. Such a system would allow construct manipulation in bacterial cells, simple preparation of large quantities of purified vector for human cell transfection and a method of reintroducing the vector back into bacteria to assess any construct rearrangement. Episomal vectors based on the mechanisms of stable latent replication and extrachromosomal persistence of the 172 kb human herpesvirus Epstein–Barr virus (EBV) represent a promising large capacity extrachromosomal vector system (21–24).

The functional components required for the episomal persistence of EBV-based vectors are the EBV latent origin of replication (*oriP*) present *in cis* (25) and expression of the viral protein EBV nuclear antigen 1 (EBNA-1) (26,27). The EBNA-1 protein binds to the family of repeats (*FR*) element, a tandem array of 20 repeats of a 30 bp motif within *oriP*, and this association promotes nuclear retention (28). Another region of *oriP*, the dyad element, acts as the latent replication origin for EBV (29,30). Incorporating the *oriP* elements into plasmid vectors has been shown to allow constructs to remain stably episomal for long periods of time (26). As an alternative to the whole *oriP* region being used in these plasmids, the *FR* can be used on its own if another replication origin is provided, for example, by including human genomic DNA. Randomly cloned pieces of human genomic DNA of an average size of 12 kb have been shown to support replication in *FR* vectors in human cells expressing EBNA-1 (28), with a positive correlation in short-term replication assays between replication strength and fragment length over the range 1–24 kb (31).

Plasmids carrying *oriP* are maintained as multicopy episomes within human cells and can be easily isolated using techniques to prepare low molecular weight DNA, followed by transfection into competent *Escherichia coli* cells. Such vectors have been

*To whom correspondence should be addressed. Tel: +44 1865 740015; Fax: +44 1865 742187; Email: michael.james@well.ox.ac.uk

used in a number of functional cloning studies based on expression from cDNA (32,33). The ability to propagate the vector in bacteria provides ease of construct manipulation before introduction into human cells and the plasmid rescue from transfected human cells allows constructs to be assayed for any rearrangements. Until the development of the PAC and BAC cloning vectors the capacity of bacterial cloning systems was limited to ≤ 40 kb which prevented the development of a large-insert *E. coli*/human cell shuttle vector. To overcome this the *oriP* system has been incorporated into the YAC system as a method of shuttling a 90 kb YAC between human cells and *Saccharomyces cerevisiae* (34). YAC systems have proved invaluable for physical mapping studies, but are laborious to work with for transfection studies when microgram amounts of purified vector DNA are required.

In contrast to YAC-based systems, we have constructed a shuttle vector system to carry large inserts (>100 kb) of genomic DNA as extrachromosomal plasmids in human and *E. coli* cells. The vector (p5170) is based on the F factor-derived BAC vector used for cloning large human genomic DNA inserts in *E. coli* (2). The vector also includes the *FR* element from EBV *oriP* and the enhanced green fluorescent protein (*GFP*) gene (35) as our reporter system to follow the fate of the plasmid in live human cells. As a reporter gene *GFP* offers the advantage of being able to track the episomes in live cells using time course photography.

For this study we subcloned a size series of well-characterised human genomic DNA inserts from PAC library clones (1) into the unique *NotI* site of p5170. The inserts were in the size range 23–105 kb and were taken from a contig previously constructed within the laboratory (unpublished data). Plasmids carrying the large inserts were shown to behave episomally in a manner comparable with small EBV *oriP* vectors. The large episomes were maintained without rearrangement in most cases as extrachromosomal elements in human cells grown for prolonged periods of time. In the presence of selection a 117 kb episome was present at ~ 2 copies/cell, whereas a small *oriP* vector was maintained at ~ 15 copies/cell in the same clonal cell line. In the absence of selection the episomal constructs carrying large inserts of human genomic DNA were lost from actively dividing cells at $\sim 4\%$ per generation. The 117 kb episome was routinely shuttled unrearranged back into the bacterial host during 15 months of continuous cell culture. The construct could also still be detected in the human cell line after 198 days growing in the absence of selection.

MATERIALS AND METHODS

Plasmid construction

The vector p5170 is based on the *NotI*–*XhoI* fragment of the BAC vector pBAC108L (2), with the addition of three fragments. First, the EBV *FR* region from the vector pHEBo (36); second, the hygromycin-B phosphotransferase (*HPH*) (36) gene under the control of the Rous sarcoma virus promoter (*PRSV*); third, the green fluorescent protein (*GFP*) open reading frame from the plasmid pEGFP-N1 (Clontech) driven by the immediate-early promoter of cytomegalovirus (*PCMV*). The BAC-based vector p5170 is a low copy plasmid, so the pUC link from pCYPAC2 (1) was added at the unique *NotI* site, creating p5171, a high copy vector which can be prepared in large quantities. To prepare the vector for large insert cloning p5171 was digested with *NotI*

releasing the 11.9 kb vector; the digested DNA was dephosphorylated and the 11.9 kb band was purified using the Qiaex II Gel extraction kit (Qiagen). To prepare the 55 and 105 kb inserts, 5 μ g of PAC DNA from individual library clones was digested to completion with *NotI* and loaded onto a 2.2 ml sucrose gradient (20–40%, with a 60% cushion). The gradients were spun at 16 000 r.p.m. for 16 h in a TLS-55 swing-out rotor in a Beckman Optima Ultracentrifuge. Following the spin the gradient was fractionated and fractions with only insert present were identified. The insert DNA was ethanol precipitated in the presence of Dextran 40,000 carrier, resuspended and quantified. The 23 kb insert was purified by electroelution from a 1% pulsed field gel electrophoresis (PFGE) agarose gel. The purified inserts were ligated to the prepared vector using standard techniques. Vector p5173 is based on pHEBo (36) with the addition of the *PCMV*–*GFP* expression cassette from p5170.

Tissue culture

The SV40-immortalized human male fetal lung fibroblast cell line MRC-5V2 (37) (referred to here as MRC5) was cultured in Dulbecco's modified Eagle's medium (Sigma) supplemented with 10% fetal calf serum, 2 mM L-glutamine, 100 U/ml penicillin and 0.1 mg/ml streptomycin. Cells were incubated in a humid incubator at 37°C in 6% (v/v) CO₂.

Transfections and low molecular weight DNA preparation

Plasmids ≤ 15 kb in size were transfected into MRC5 cells by electroporation (260 V, 1050 μ F and an infinite resistance using the Easyject Plus; EquiBio). MRC5 cells transfected with p410⁺ were selected with G418 at 1 mg/ml. For the plasmid rescue assays low molecular weight DNA was prepared by an alkaline lysis method from a confluent 10 cm dish as follows. The cells were scraped into 1.5 ml of phosphate-buffered saline (PBS) without Mg²⁺ or Ca²⁺ and centrifuged at 5000 g for 3 min. The cell pellet was resuspended in 60 μ l of STET buffer (8% sucrose, 5% Triton X-100, 50 mM EDTA, 50 mM Tris, pH 8). The cells were lysed with 130 μ l of alkaline SDS (1% SDS, 0.2 N NaOH), neutralised with 110 μ l of 7.5 M ammonium acetate, incubated on ice for 5 min and spun at 14 000 g for 30 min at 4°C. The cleared lysate was extracted twice with 1:1 phenol:chloroform and once with chloroform and the nucleic acids were precipitated and resuspended in 50 μ l TE (10 mM Tris, 1 mM EDTA) containing 5 μ g/ml RNase A. For episome rescue, 2 μ l of this extract was electroporated into DH-10B ElectroMax *E. coli* cells (Gibco BRL) which were plated onto LB agar containing either ampicillin (100 μ g/ml) or chloramphenicol (12.5 μ g/ml). Plasmid DNA was prepared from bacterial colonies using standard methods and analysed by restriction digests and agarose gel electrophoresis. The plasmids p5170, p5179, p5180 and p5176 were transfected into MRC5(p410⁺) cells using calcium phosphate co-precipitation (38) and selected using hygromycin-B at 200 μ g/ml, in addition to the 1 mg/ml G418 already present. Individual colonies which appeared were picked and amplified into clonal cell lines, grown in hygromycin-B at 125 μ g/ml and 1 mg/ml G418.

Southern blot analysis

DNA extract was prepared from a confluent 10 cm tissue culture dish of MRC5 cells by either the method described by Sun *et al.*

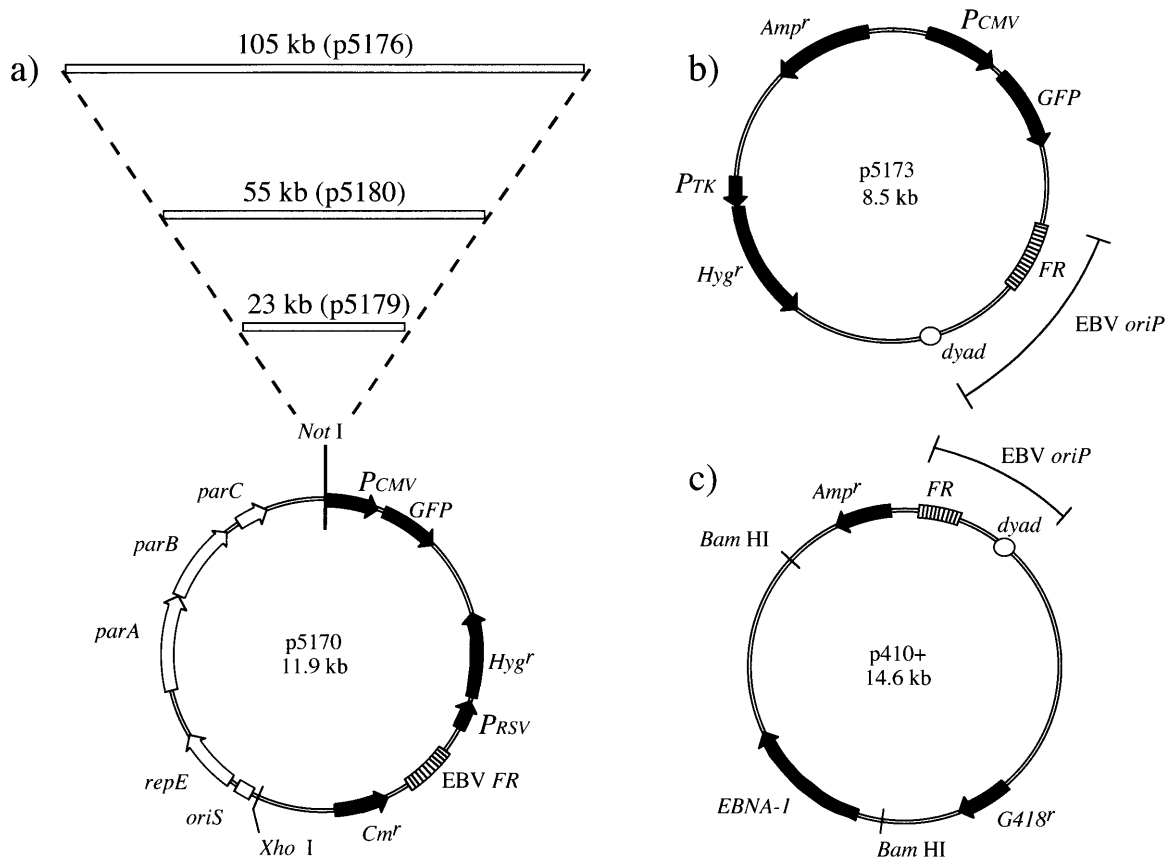


Figure 1. Vector maps. (a) The p5170 vector size series used in this study, with the range of genomic DNA inserts. *parA*, *parB*, *parC*, *repE* and *oriS* are F factor regulatory genes from pBAC108L which control replication and copy number in bacteria. (b) p5173 is an *oriP* vector containing the *PCMV-GFP* expression cassette. (c) p410⁺ is an *oriP* vector expressing EBNA-1 protein.

(22) or by a standard genomic DNA preparation method. The DNA was digested with restriction enzymes (New England Biolabs) and the digests were separated on a 0.8% agarose gel. The gel was blotted and the membrane hybridized by standard methods (39).

Flow cytometric analysis and fluorescence microscopy detection of GFP expression

GFP expression was analysed by flow cytometry at an excitation wavelength of 488 nm using a Beckton Dickinson FACScalibur with associated CellQuest software. 10 000 cells were assayed in each flow cytometric analysis. The episomal rates of loss given in Table 1 are calculated from GFP expression data on days 0 and 24 (p5173), days 0 and 21 (p5179 and p5180) and for days 0 and 22 (p5176, pEGFP-N1, p5170 and p5171), using a first order rate-of-loss model: rate of loss (λ) = $(-1/t)(\ln N_t/N_0)$. N_0 is the percentage green cells at the start of the release experiment and N_t is the percentage green cells after t generations. The fluorescence microscopy was carried out using a Nikon Diaphot 300 inverted microscope fitted with a standard FITC filter set. Photographs were taken at 10 \times magnification with a Nikon 35 mm F801 camera using Fujifilm Provia 400 ASA colour slide film, developed by standard E6 processing.

RESULTS

Vector construction

The vectors used for these experiments are shown in Figure 1. The size series of vectors was built by subcloning human genomic DNA inserts into the unique *NotI* site of p5170 (Fig. 1a) from PAC library clones. Vector p5179 carries a 23 kb insert, p5180 carries a 55 kb insert and p5176 carries a 105 kb insert. The PAC clones were taken from a contig previously constructed within the laboratory covering chromosome 11q22–23 (unpublished data). The 55 and 105 kb inserts lie within a completely sequenced region containing the ataxia telangiectasia (*ATM*) locus (40). Random fragments of human genomic DNA in this size range are expected to provide replication origins for the episomes in human cells (28). The vector p5173 (Fig. 1b) is based on pHEBo (36) with the addition of the *Pcmv-GFP* expression cassette and was included in our study to allow a comparison of the large p5170-based episomes with a regular EBV *oriP* vector within the same cell culture system.

Transfection of vectors into MRC5 cells

The immortalised human fibroblast cell line MRC-5V2 (37) (referred to here as MRC5) was transfected with the EBNA-1-expressing episomal plasmid p410⁺ (Fig. 1c; 30) and G418 selection was applied to create an EBNA-1⁺ cell population. The

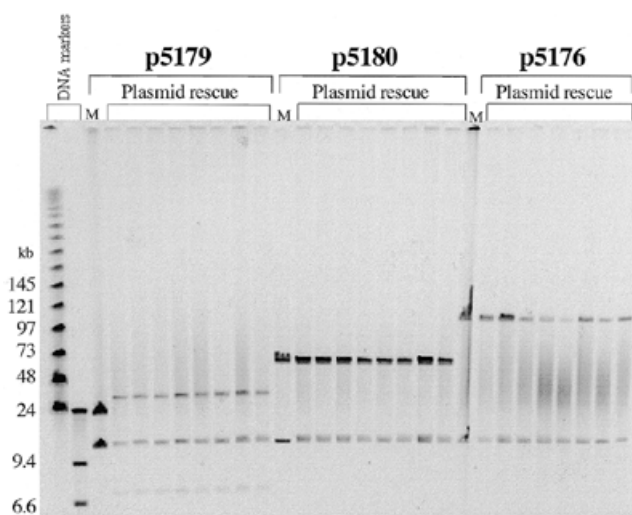


Figure 2. Plasmid rescue from the three clonal cell lines MRC5(p410⁺,p5179), MRC5(p410⁺,p5180) and MRC5(p410⁺,p5176). Plasmid DNA preparations from eight chloramphenicol-resistant bacterial colonies obtained from a plasmid rescue assay on each cell line were analysed by *NotI* digestion and PFGE. M, marker lane containing a *NotI* digest of the plasmid DNA transfected into the human cells.

episomal status of p410⁺ was confirmed by Southern blotting (see below) and by plasmid rescue into *E.coli* (not shown).

The ability of p410⁺ EBNA-1 expression to support *oriP* function *in trans* in a cell line under prolonged selection was confirmed by transfecting p5173 into the MRC5(p410⁺) cells, applying hygromycin selection and successfully rescuing both the p410⁺ and p5173 plasmids back into *E.coli* cells 25 days later (not shown) from a pooled cell population named MRC5(p410⁺,p5173).

The plasmids p5179, p5180 and p5176 were then transfected into the MRC5(p410⁺) cells. Hygromycin-resistant colonies, which also expressed GFP strongly under fluorescence microscopy, were picked and expanded into clonal cell lines. After 5 months of continuous culture, low molecular weight DNA was extracted from these cell lines and transfected into *E.coli* which were plated onto LB agar containing either ampicillin or chloramphenicol. Using restriction digest analysis the plasmid DNA preparations from bacterial colonies from the ampicillin plates were compared with p410⁺ and those from the chloramphenicol plates were compared with either p5179, p5180 or p5176. The p410⁺ results confirmed its episomal status (not shown) and Figure 2 shows the *NotI* restriction digest analysis and PFGE of plasmid DNA prepared from the chloramphenicol-resistant *E.coli* colonies. All three cell lines contained episomes which could be rescued, and were named MRC5(p410⁺,p5179), MRC5(p410⁺,p5180) and MRC5(p410⁺,p5176). In two cases [MRC5(p410⁺,p5180) and MRC5(p410⁺,p5176)] the episomes were intact and unarranged; in one case [MRC5(p410⁺,p5179)] the episome had undergone rearrangement, increasing the insert size from 23 to ~37 kb and creating a new *NotI* site. A stably transfected cell line carrying an integrated form of the p5170 vector not containing any human DNA was also created, and named MRC5(p410⁺,p5170).

Demonstrating autonomous replication of the 117 kb episome

Extrachromosomal DNA was prepared using the method described by Sun *et al.* (22) from MRC5(p410⁺,p5176) grown in the presence of hygromycin. The presence of episomal DNA was confirmed by performing plasmid rescue of intact, unarranged p5176 from the extract (not shown). For the replication assay extract DNA digested with *XbaI* was further digested with either *DpnI*, *Sau3AI* or *MboI*. All three enzymes recognise the same cutting site (GATC), but the activity of *DpnI* and *MboI* depends on the methylation pattern of the DNA. *DpnI* only cuts DNA which retains the bacterial methylation pattern, whereas *MboI* only cuts DNA which has lost the bacterial methylation by replicating twice in human cells. *Sau3AI* cuts regardless of the methylation status and acts as a control. The restriction digestion fragments were separated on a 1% agarose gel which was blotted and probed with a 987 bp fragment containing the chloramphenicol resistance (*Cm^r*) gene from p5170. The appearance of a 1.75 kb band in the *MboI* digest, not present in the p5176 DNA prepared from *E.coli*, demonstrates that the p5176 episome is replicating in human cells (Fig. 3a). The blot was stripped and reprobed with a 2 kb *EBNA-1* DNA probe, which shows that p410⁺ is still present and replicating (Fig. 3b).

Estimating the copy number of the 117 kb episome

Total genomic DNA was prepared from MRC5(p410⁺,p5176) cells growing in hygromycin and the presence of intact episomes was confirmed by plasmid rescue (not shown). Five micrograms of total MRC5(p410⁺,p5176) genomic DNA was digested with *BamHI* and the digestion fragments were resolved on a 1% agarose gel alongside a set of 3-fold serial dilutions of *BamHI* digests of accurately quantified p5171 and p410⁺. Prior to gel loading these dilution series standards were pooled together and mixed with 5 µg of *BamHI*-digested rat genomic DNA. The gel was blotted and probed with the *Cm^r* gene (Fig. 3c) and bands quantified using a Molecular Dynamics PhosphorImager. Vector p5171 was used as the quantitative control in place of p5176 because p5171 DNA is purified twice on caesium chloride density gradients and thus can be quantified much more accurately than can preparations of pBAC DNA. The *Cm^r* gene probe hybridises to a 12 kb band of p5171 and a 14 kb band of p5176. One copy of a 10 kb double-stranded control plasmid is ~6 pg of DNA in 5 µg of genomic DNA prepared from the hyperloid MRC5 line (modal chromosome number of 60–80 per metaphase spread; 37). Based on these data, p5176 was estimated to be present at ~2 copies/cell in MRC5(p410⁺,p5176). The blot was stripped and reprobed with the 2 kb *EBNA-1* DNA probe (Fig. 3d) and PhosphorImager quantitative analysis showed p410⁺ to be present at ~13 copies/cell in MRC5(p410⁺,p5176). The analysis was repeated using an *EcoRV* digest of 5 µg of total genomic DNA prepared from MRC5(p410⁺,p5176) and compared with an appropriate quantitative dilution series of *EcoRV* digests in a separate independent assay (not shown). Values of 2 and 16 copies/cell of p5176 and p410⁺, respectively, were obtained.

Assessing episomal maintenance in the absence of selection

Episomal maintenance in the absence of selection was assessed in two distinct ways. First, the cell lines were analysed using a

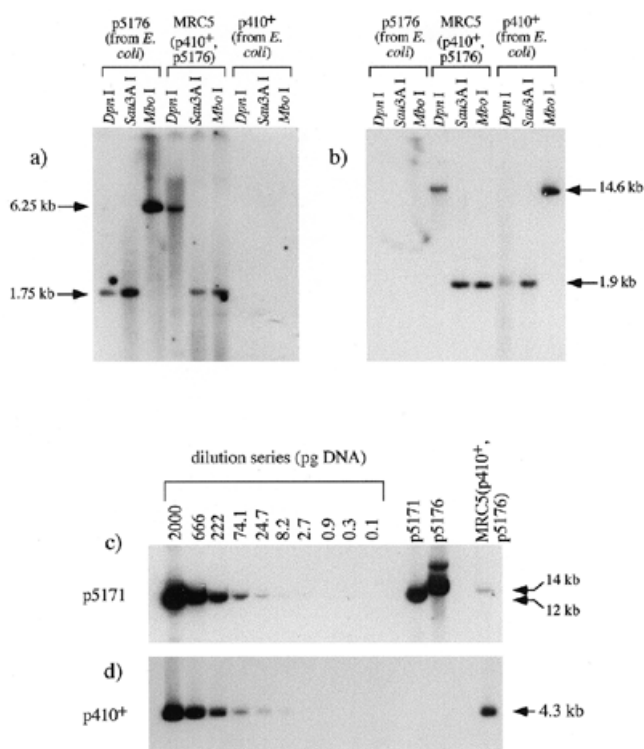


Figure 3. Southern blot analysis of the MRC5(p410⁺,p5176) clonal cell line. (a and b) Extrachromosomal DNA was prepared by the method of Sun *et al.* from MRC5(p410⁺,p5176) and double-digested with either *DpnI*, *Sau3A I* or *MboI* together with *XbaI*. (a) The blot was probed with a 987 bp vector-specific fragment. The appearance of a 1.75 kb band in the *MboI* digest of DNA from human cells, not present in digests of the bacterial DNA preparation, indicates that p5176 has replicated. (b) The same blot rehybridised with a 2 kb *EBNA-1* probe to show the continued presence and replication of p410⁺. (c and d) Five micrograms of MRC5(p410⁺,p5176) genomic DNA was digested with *BamHI*. Serial dilutions of *BamHI*-digested p5171 and p410⁺ were pooled and mixed with 5 µg of *BamHI*-digested rat genomic DNA before loading. (c) The blot was hybridised with a 987 bp probe from the p5170 *Cm^r* gene. (d) The same blot rehybridised with a 2 kb *EBNA-1* probe.

flow cytometer to assess the proportion of cells expressing GFP over time. Second, *GFP* was used to track the fate of the episomes in expanding colonies by time course photography. In both sets of experiments G418 selection was continuous so as to maintain *EBNA-1* expression from p410⁺.

Flow cytometric analysis

MRC5(p410⁺,p5170) was grown for 115 days of continuous culture in the presence and absence of hygromycin selection. The cells were maintained in an actively growing state by regular passaging in the presence and absence of hygromycin selection. At each passage a sample of cells was taken and the proportion of cells expressing GFP was assessed by flow cytometry. The MRC5(p410⁺) cell line was used as a negative control in each assay to determine the background level of fluorescence. In the absence of hygromycin no loss of GFP expression from integrated p5170 was seen (Fig. 4a). These data show that in the clone studied the *PCMV-GFP* expression cassette has not been down-regulated by epigenetic effects, such as DNA methylation, during prolonged periods of cell culture.

The cell lines MRC5(p410⁺,p5173), MRC5(p410⁺,p5179), MRC5(p410⁺,p51780) and MRC5(p410⁺,p5176) were cultured as described above. At the start of the experiment nearly 100% of the cells in each cell line were expressing GFP, causing the peak of GFP fluorescence to shift upwards by three orders of magnitude as compared with the negative control (not shown). The peak of fluorescence shifted downwards over time as the episome was lost from the cell line during continuous growth in the absence of selection. The percentage of cells expressing GFP above background was calculated for each construct at each time point (Fig. 4b–e). For each episomal construct the rate of loss of episome over the first 3 weeks of growth in the absence of selection was calculated, and is given in Table 1. Under continuous selection the expression of GFP remains at, or very near to, 100%. A set of control plasmids was also included in the flow cytometric study to provide a rate of loss value in *EBNA-1*⁺ MRC5 cells for a plasmid lacking the *oriP* dyad symmetry element replication origin or a human genomic DNA insert. The control plasmids were the p5170 vector without a human DNA insert, p5171 [a form of p5170 carrying the pUC insert from pCYPAC2 (1) in the *NotI* cloning site] and pEGFP-N1 (a control plasmid from Clontech lacking *FR*). The rates of loss of these control plasmids in transient transfections of MRC5(p410⁺) cells are shown in Table 1.

Table 1. Rates of plasmid loss calculated from the data in Figure 4 and the rates observed for three control plasmids lacking replication origins (see text for details)

Construct	Rate of loss per cell generation (%)
EBV-based episomes	
p5173	7.8
p5179	2.1
p5180	6.5
p5176	4.1
Control plasmids	
p5170	26
p5171	18
pEGFP-N1	24

After the MRC5(p410⁺,p5173) cell line had been growing in the absence of selection for 45 days, ~7% of the cells still carried the episome according to the flow cytometry data and it remained possible to rescue the plasmid into *E.coli*. At this point selection was reapplied to the cell line. After 21 days of reselection, the expression of GFP returned to >90% (Fig. 4b) and it remained possible to rescue p5173 into *E.coli*. This experiment was repeated on the MRC5(p410⁺,p5176) cell line. In that case after 100 days growth in the absence of selection ~10% of the cells retained the episome and it was still possible to rescue the plasmid into *E.coli*. Hygromycin selection was reapplied and after 14 days of reselection the expression of GFP in the reselected line was back to 100% (Fig. 4e) and it remained possible to rescue p5176 into *E.coli*.

Time course photography of expanding colonies

Cells from the clonal cell lines carrying each episome were plated at low density and isolated colonies were allowed to expand in the

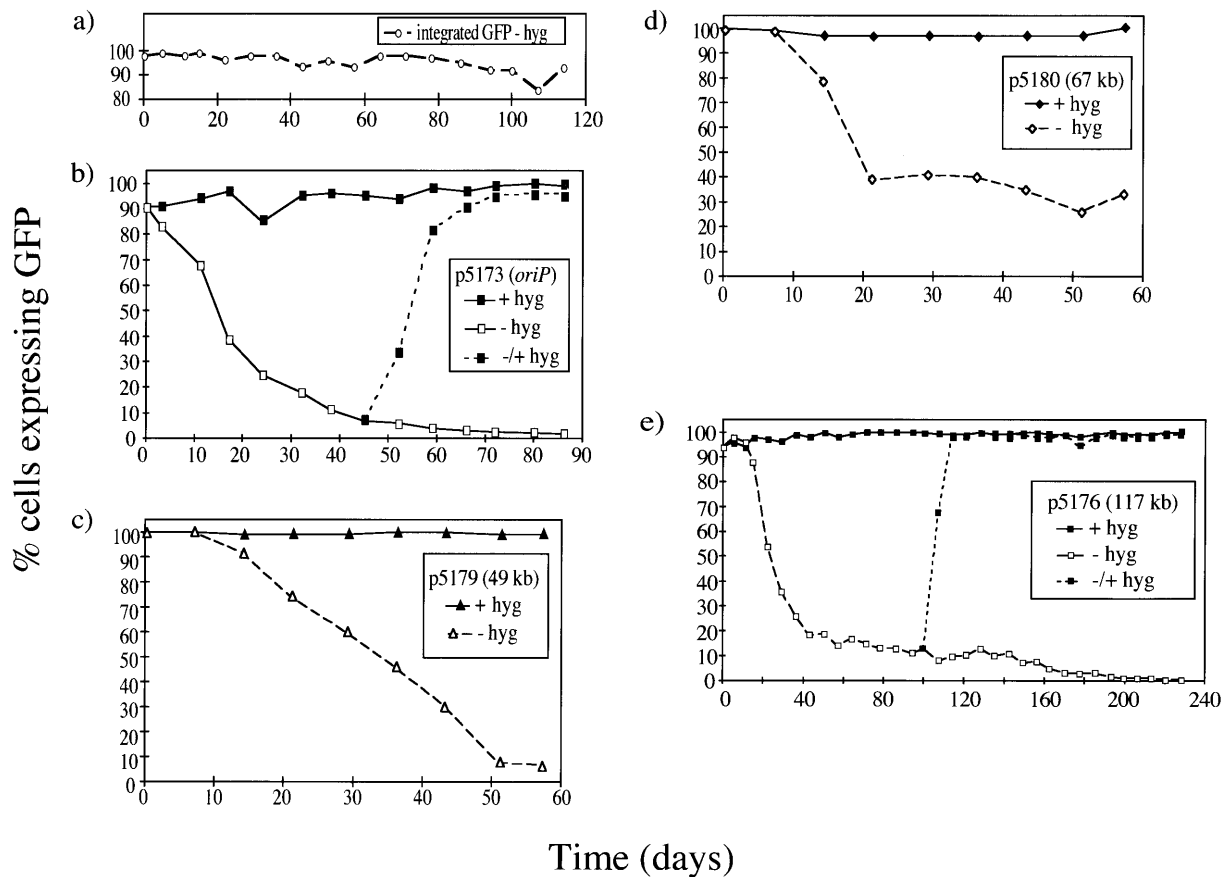


Figure 4. Time course of long-term GFP expression in the presence or absence of hygromycin selection for various cell lines. (a) MRC5(p410⁺,p5170); (b) MRC5(p410⁺,p5173); (c) MRC5(p410⁺,p5179); (d) MRC5(p410⁺,p5180); (e) MRC5(p410⁺,p5176).

absence of hygromycin, but in the continued presence of G418. Individual colonies were followed by photography as they grew over a period of weeks. Figure 5 shows an example of a colony carrying the 117 kb construct expanding in the absence of selection over 38 days. At early times (days 12 and 18) the loss of GFP expression at the edge of the colony can be seen. With time (days 31 and 38), this gives rise to a sectorized colony. Strong expression is maintained at the centre of the colony where the cells are either not dividing or are doing so much more slowly than at the edge. No sectorized colonies were seen when the MRC5(p410⁺,p5170) cell line expressing stably integrated GFP was used for this assay (not shown).

Assessing the retention of the 117 kb episome by plasmid rescue

While the MRC5(p410⁺,p5176) cell line was being grown in the absence of selection for flow cytometric analysis, low molecular weight DNA preparations were made at several time points to assess if it remained possible to rescue the 117 kb plasmid as the fraction of cells carrying the episome diminished. The number of bacterial colonies obtained from chloramphenicol-resistant (p5176) plasmid rescue fell during the duration of the time course, whereas the efficiency of ampicillin-resistant (p410⁺) rescue remained constant (data not shown). Furthermore, it was still possible to rescue p5176 from the cell line growing in the absence of selection after

198 days without hygromycin, when only 1% of the cells were carrying the episome, as assayed by flow cytometric analysis of GFP expression. After this time point, plasmid rescue of the 117 kb proved impossible and the experiment was terminated. In total, 94 chloramphenicol-resistant plasmid rescue bacterial colonies obtained from the MRC5(p410⁺,p5176) cell line grown for up to 15 months in continuous culture were analysed. All colonies examined contained intact, unrearranged forms of p5176 (not shown).

DISCUSSION

This study has shown that the p5170 vector, based on pBAC108L and incorporating the EBV *FR/EBNA-1* system, can act as a shuttle vector carrying large (>100 kb) inserts of human genomic DNA between human and *E.coli* cells. Using a BAC-based system allows routine manipulation in bacterial cells and efficient recovery of episomes from human cells back into bacteria for further study. Although functional expression of a genomic locus remains to be shown, the system we describe offers considerable potential for gene expression studies. Furthermore, *in vitro* modification of BACs by mutagenesis (41) and homologous recombination (10) could allow functional analysis in human cells of targeted mutations in both coding regions and non-coding elements such as locus controlling regions, intronic enhancers or silencers (8,42).

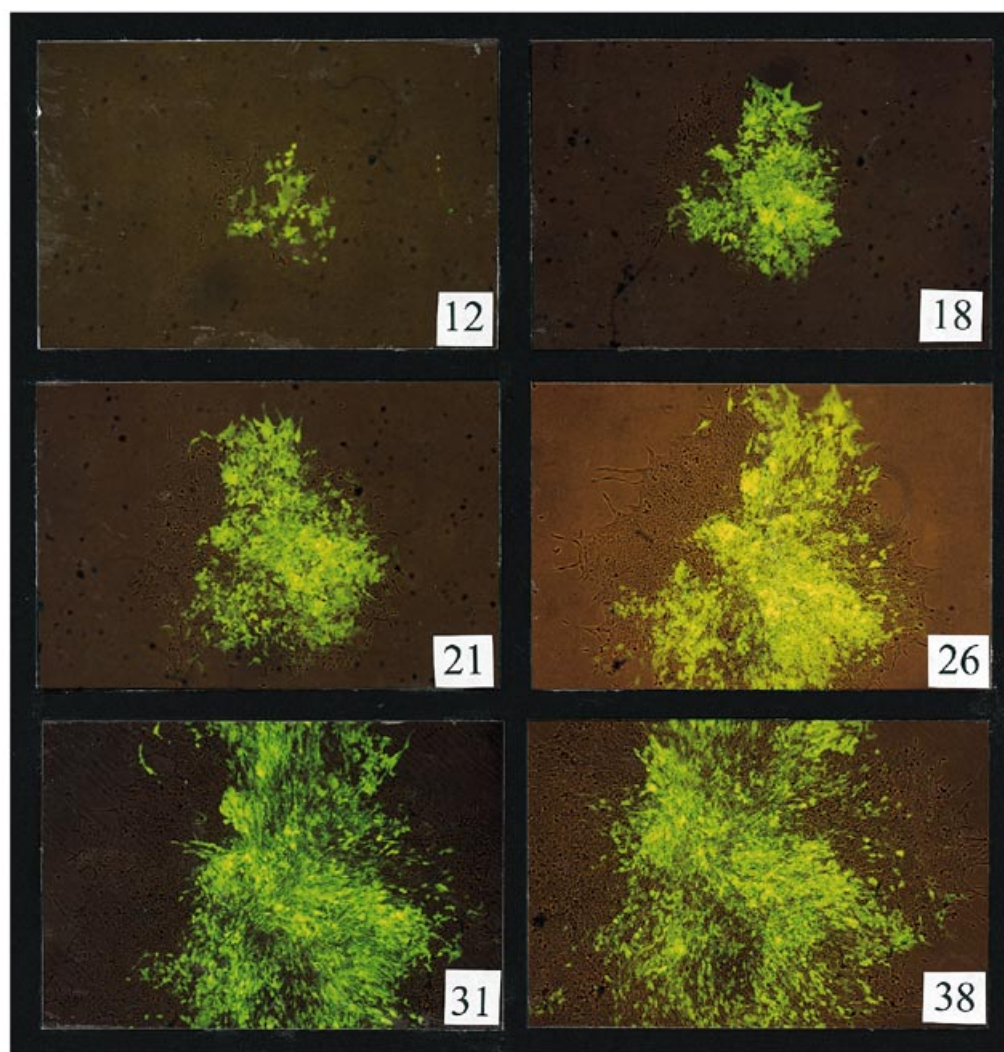


Figure 5. Tracking the fate of episomes in live cells by GFP fluorescence microscopy. The growth of a colony derived from MRC5(p410⁺,p5176) expanding in the absence of hygromycin selection over 38 days is shown; the numbers relate to days of growth since the removal of hygromycin selection.

We have shown that episomal plasmid constructs carrying human genomic DNA inserts >100 kb in size can persist intact in human cells replicating autonomously. An episome carrying a human genomic DNA insert of 105 kb was shown to be perfectly stable, persisting for 15 months in continuously passaged cells with no rearrangement. These observations were based on plasmid rescue analysis of nearly 100 chloramphenicol-resistant plasmid rescue bacterial colonies obtained from the MRC5(p410⁺,p5176) cell line. Other reports have also described the episomal maintenance of large constructs in human cells (21,43), although the present study is the first time to our knowledge that episomal constructs carrying >100 kb in human cells have been rescued in bacterial cells. Plasmid rescue data shows that the system is a true shuttle vector, defined by the ability to move the vector back and forth between two hosts.

Another study has developed an *oriP*/EBNA-1 shuttle vector based on YACs in which a 90 kb YAC was successfully shuttled from yeast cells into human cells and back into yeast (34). The YAC system is based on linear DNA molecules, which have to be converted into the circular form required for EBV episomes.

However, the BAC-derived p5170 vectors are already circular and although they are low copy *E.coli* vectors, they can be prepared simply at high yield using conventional bacterial plasmid preparations. Another feature of the YAC cloning system is the high proportion (40–60%) of chimeric clones (44) which is not seen in BACs. The BAC cloning system can be used to clone human DNA inserts of up to 300 kb (2) allowing p5170-based episomes to carry all but the largest of human genes as genomic DNA.

In the present study the episomal persistence of the p5170 vectors in cells cultured in the absence of selection was measured by following the expression of the reporter gene *GFP* in two ways. First, flow cytometric GFP expression assays over time found the rate of episomal loss of the three constructs carrying genomic DNA to be 2–6% per cell division in the absence of selection. The values are comparable with retention rates reported for the large, circular *oriPYAC* constructs (43), for other *FR*-based constructs carrying smaller inserts of human genomic DNA (45) and for smaller, full *oriP* plasmids (26), including the p5173 construct used in the current study. Second, *GFP* was used as a live cell marker to track the fate of the episomes in expanding cell colonies. The episomes were lost

from the rapidly dividing cells at the edge of colonies, but remained present in the cells at the centre of the colony which had either stopped dividing or were doing so much more slowly (Fig. 5). This study is the first time to our knowledge that time course photography of expanding colonies has been used to assess episome stability in human cells.

Copy number assessment of the episomes carrying genomic DNA was carried out as part of the present study. The 117 kb p5170-based episome (p5176) was found to be present at a low copy number of ~2 copies/cell compared with the smaller, full *oriP* vector p410⁺, present at ~15 copies within the same clonal cell line. The copy number of p5176 is in agreement with the low copy numbers of 90 and 660 kb extrachromosomal *oriPYACs* previously reported (43). The copy number for p410⁺ is in close agreement with previously published data (30). In contrast to the high copy of the *oriP* vectors, the large p5170-based episomes are present at low copy number in the same cells. No attempt was made to assess cell-to-cell variation in copy number of the large episomes, however, it is unlikely to vary significantly given the low average copy number found in the cell line as a whole.

Small overall size, but large cloning capacity, makes the vector a good candidate for gene expression studies and gene therapy protocols. Efficient transfection of the current vector could be achieved through recent developments in mammalian cell transfection, using an inactivated adenovirus carrier (46) or integrin targeting peptides (47,48), for example. The EBV episomes incorporating *FR* and human genomic DNA inserts can function in both human and rodent cell lines (49), allowing the study of gene expression from the genomic inserts to be carried out in a wide variety of tissue culture and animal model systems. A recent study (21) claimed to show gene expression from a vector carrying a 185 kb insert of human genomic DNA from the human globin locus, however, only a low level of illegitimate β -globin transcription was detectable by reverse transcription-based polymerase chain reaction and globin protein was not detectable by immunoblotting.

The p5170 series of vectors represents a first generation of EBV vectors designed to carry genomic DNA inserts in human cells. Currently the vectors require EBNA-1 expression *in trans* for episomal function but we are generating new vectors which express EBNA-1 from an open reading frame *in cis*. Furthermore, it remains to be seen if EBV-based vectors containing large genomic DNA can be packaged efficiently by the *in vitro* culture system that has been used for viral-only or cDNA constructs (50,51). Experiments to this end are currently underway with derivatives of the vector shown in this study. If successful, this would extend the advantages of an infectious viral delivery system to large genomic loci.

ACKNOWLEDGEMENTS

We would like to thank Louise Devenish for help with tissue culture, Dr Martyn Bell for his flow cytometry expertise, Dr Ian Hickson for the use of the phosphorimager and Robert White for critical reading of the manuscript. R.W.-M. is a Wellcome Trust Prize Student, while J.F. and M.R.J. are supported by The Wellcome Trust. This work was partly supported by EC grants BMH4-CT97-2240 and BMH4-CT95-0914.

REFERENCES

- Ioannou,P.A., Amemiya,C.T., Ganes,J., Kroisel,P.M., Shizuya,H., Chen,C., Batzer,M.A. and deJong,P.J. (1994) *Nature Genet.*, **6**, 84–89.
- Shizuya,H., Birren,B., Kim,U.-J., Mancino,V., Slepak,T., Tachiiri,Y. and Simon,M. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 8794–8797.
- Kim,U.G., Birren,B.W., Slepak,T., Mancino,V., Boysen,C., Kang,J.L., Simon,M.I. and Shizuya,H. (1996) *Genomics*, **34**, 213–218.
- Schedl,A., Montoliu,L., Kelsey,G. and Schutz,G. (1993) *Nature*, **362**, 258–261.
- Peterson,K.R., Clegg,C.H., Huxley,C., Josephson,B.M., Haugen,H.S., Furukawa,T. and Stamatoyannopoulos,G. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 7593–7597.
- Manson,A.L., Trezise,A.E.O., MacVinish,L.J., Kasschau,K.D., Birchall,N., Episkopou,V., Vassaux,G., Evans,M.J., Colledge,W.H., Cuthbert,A.W. and Huxley,C. (1997) *EMBO J.*, **16**, 4238–4249.
- Lamb,B.T., Sisodia,S.S., Lawler,A.M., Slunt,H.H., Kitt,C.A., Kearns,W.G., Pearson,P.L., Price,D.L. and Gearhart,J.D. (1993) *Nature Genet.*, **5**, 22–30.
- Blackwood,E.M. and Kadonaga,J.T. (1998) *Science*, **281**, 60–63.
- Antoch,M.P., Song,E.J., Chang,A.M., Vitaterna,M.H., Zhao,Y., Wilsbacher,L.D., Sangoram,A.M., King,D.P., Pinto,L.H. and Takahashi,J.S. (1997) *Cell*, **89**, 655–667.
- Yang,X.W., Model,P. and Heintz,N. (1997) *Nature Biotechnol.*, **15**, 859–865.
- Schedl,A., Ross,A., Lee,M., Engelkamp,D., Rashbass,P., van Heyningen,V. and Hastie,N.D. (1996) *Cell*, **86**, 71–82.
- Perou,C.M., Justice,M.J., Pryor,R.J. and Kaplan,J. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 5905–5909.
- Hoeijmakers,J.H.J., Odijk,H. and Westerfield,A. (1987) *Exp. Cell Res.*, **169**, 111–119.
- Mayne,L.V., Jones,T., Dean,S.W., Harcourt,S.A., Lowe,J.E., Priestley,A., Steingrimsdottir,H., Sykes,H., Green,M.H.L. and Lehmann,A.R. (1988) *Gene*, **66**, 65–76.
- Mayne,L.V., Jones,T., Dean,S.W., Harcourt,S.A., Lowe,J.E., Priestley,A., Steingrimsdottir,H., Sykes,H., Green,M.H.L. and Lehmann,A.R. (1988) *Gene*, **83**, 395.
- Lohrer,H., Blum,M. and Herrlich,P. (1988) *Mol. Gen. Genet.*, **212**, 474–480.
- Ikeno,M., Grimes,B., Okazaki,T., Nakano,M., Saitoh,K., Hoshino,H., McGill,N.L., Cooke,H. and Masumoto,H. (1998) *Nature Biotechnol.*, **16**, 431–439.
- Huxley,C. (1994) *Gene Ther.*, **1**, 7–12.
- Huxley,C., Farr,C.J., Gennaro,M.L. and Haaf,T. (1994) *Biotechnology*, **12**, 586–590.
- Harrington,J., Bokkelen,G.V., Mays,R.W., Gustashaw,K. and Willard,H.F. (1997) *Nature Genet.*, **15**, 345–355.
- Westphal,E.M., Sierakowska,H., Livanos,E., Kole,R. and Vos,J.-M. (1998) *Hum. Gene Ther.*, **9**, 1863–1873.
- Sun,T.-Q., Fenstermacher,D.A. and Vos,J.-M.H. (1994) *Nature Genet.*, **8**, 33–41.
- Kelleher,Z.T., Fu,H., Livanos,E., Wendleburg,B., Gulino,S. and Vos,J.-M. (1998) *Nature Biotechnol.*, **16**, 762–768.
- Calos,M.P. (1996) *Trends Genet.*, **12**, 463–466.
- Yates,J., Warren,N., Reisman,D. and Sugden,B. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 3806–3810.
- Yates,J.L., Warren,N. and Sugden,B. (1985) *Nature*, **313**, 812–815.
- Wendleburg,B.J. and Vos,J.-M.H. (1998) *Gene Ther.*, **5**, 1389–1399.
- Krysan,P.J., Haase,S.B. and Calos,M.P. (1989) *Mol. Cell. Biol.*, **9**, 1026–1033.
- Reisman,D., Yates,J. and Sugden,B. (1985) *Mol. Cell. Biol.*, **5**, 1822–1832.
- Lupton,S. and Levine,A.J. (1985) *Mol. Cell. Biol.*, **5**, 2533–2542.
- Heinzel,S.S., Krysan,P.J., Tran,C.T. and Calos,M.P. (1991) *Mol. Cell. Biol.*, **11**, 2263–2272.
- Legerski,R. and Peterson,C. (1992) *Nature*, **359**, 70–73.
- Strathdee,C.A., Gavish,H., Shannon,W.R. and Buchwald,M. (1992) *Nature*, **356**, 763–767.
- Simpson,K. and Huxley,C. (1996) *Nucleic Acids Res.*, **24**, 4701–4707.
- Cubitt,A.B., Heim,R., Adams,S.R., Boyd,A.E., Gross,L.A. and Tsien,R.Y. (1995) *Trends Biochem.*, **20**, 448–455.
- Sugden,B., Marsh,K. and Yates,J. (1985) *Mol. Cell. Biol.*, **5**, 410–413.
- Huschtscha,L.I. and Holliday,R. (1983) *J. Cell Sci.*, **63**, 77–99.
- Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1996) *Current Protocols in Molecular Biology*. Wiley-Interscience, New York, NY.

- 39 Sambrook,J., Fritsch,E. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- 40 Platzer,M., Rotman,G., Bauer,D., Uziel,T., Savitsky,K., Bar-Shira,A., Gilad,S., Shiloh,Y. and Rosenthal,A. (1997) *Genome Res.*, **7**, 592–605.
- 41 Borén,J., Lee,I., Callow,M.J., Rubin,E.M. and Innerarity,T.L. (1996) *Genome Res.*, **6**, 1123–1130.
- 42 Higgs,D.R. (1998) *Cell*, **95**, 299–302.
- 43 Simpson,K., McGuigan,A. and Huxley,C. (1996) *Mol. Cell. Biol.*, **16**, 5117–5126.
- 44 Monaco,A.P. and Larin,Z. (1994) *Trends Biotechnol.*, **12**, 280–286.
- 45 Wohlgemuth,J.G., Kang,S.H., Bulboaca,G.H., Nawotka,K.A. and Calos,M.P. (1996) *Gene Ther.*, **3**, 503–512.
- 46 Baker,A. and Cotten,M. (1997) *Nucleic Acids Res.*, **25**, 1950–1956.
- 47 Harbottle,R.P., Cooper,R.G., Hart,S.L., Ladhoff,A., McKay,T., Knight,A.M., Wagner,E., Miller,A.D. and Coutelle,C. (1998) *Hum. Gene Ther.*, **9**, 1037–1047.
- 48 Hart,S.L., Arancibia-Carcamo,C.V., Wolfert,M.A., Mailhos,C., O'Reilly,N.J.O., Ali,R.R., Coutelle,C., George,A.J.T., Harbottle,R.P., Knight,A.M., Larkin,D.F.P., Levinsky,R.J., Seymour,L.W., Thrasher,A.J. and Kinnon,C. (1998) *Hum. Gene Ther.*, **9**, 575–585.
- 49 Krysan,P.J. and Calos,M.P. (1993) *Gene*, **136**, 137–143.
- 50 Hammerschmidt,W. and Sugden,B. (1988) *Cell*, **55**, 427–433.
- 51 Banerjee,S., Livanos,E. and Vos,J.-M.H. (1995) *Nature Med.*, **1**, 1303–1308.