

---

**The construction of cosmid libraries which can be used to transform eukaryotic cells**

---

F.G.Grosveld, T.Lund, E.J.Murray, A.L.Mellor, H.H.M.Dahl and R.A.Flavell

---

Laboratory of Gene Structure and Expression, National Institute for Medical Research,  
Mill Hill, London NW7 1AA, UK

---

Received 10 August 1982; Accepted 21 September 1982

---

**ABSTRACT**

Cosmid vectors have been developed which carry selective markers for growth in bacteria ( $\beta$  lactamase gene) and animal cells (the Herpes Simplex virus thymidine kinase gene, the transposon Tn-5 aminoglycosyl 3' phosphotransferase gene and the *E.coli* guanine phosphoribosyltransferase gene). The design of the cosmids allows the exchange of the eukaryotic markers in recombinant cosmids. Human and mouse cosmid libraries containing DNA inserts of about 40kb have been generated by an improved method. Several clones from the human  $\beta$ -globin locus were isolated. These cosmids transform mouse L cells at high efficiency in both circular and linear form. The newly introduced genes are expressed accurately in L cells.

**INTRODUCTION**

A great deal about the structural organization of eukaryotic genes has been learned from detailed studies on cloned DNA (1). In a few cases it has also been possible to analyse certain aspects of the expression of these genes. For example, a number of cloned genes are expressed when injected into the nuclei of *Xenopus* oocytes (2, 3) and extracts of certain cell-types direct the transcription of eukaryotic genes by eukaryotic RNA polymerases (4, 5). Ideally, it is desirable to study the expression of the genes of interest in a variety of cellular backgrounds, in particular, in those cells that actively express the same gene. Partly for this reason, considerable attention has been paid to the study of the expression of mammalian genes after their (re)introduction into cultured animal cells. This expression can be studied by the stable association of the gene in question with the genetic material of the recipient cell (6). Since it is not always possible to select for the desired eukaryotic gene (e.g. a globin gene), that gene is linked to a gene for which such a selection is possible; the most commonly selective marker used to date is the Herpes Simplex virus (HSV-1) thymidine kinase (tk) gene (7, 8, 9), although other selection

systems have become available (10, 11, 12). The DNA is introduced into animal cells usually as a calcium phosphate coprecipitate (13) or by micro-injection (14).

It would clearly be of great use to be able to routinely isolate clones that were directly suitable for gene expression studies of the type described above. Ideally, such clones should contain large segments of eukaryotic DNA, so that all 'regulatory' regions are present on the cloned material, linked to the DNA sequences necessary to study gene expression. In this article, we describe the construction and use of a set of cosmid vectors that can be used to generate libraries of any DNA and that can be subsequently and directly employed for the study of gene expression. These cosmid vectors are designed to allow the exchange of eukaryotic selection markers by a simple procedure. This creates the possibility to generate recombinant cosmids from a single library with different markers to perform multiple transformations and selections on eukaryotic cells. We describe here the use of these vectors to generate DNA libraries by an improved cloning procedure and the isolation and expression properties of one cosmid recombinant from the human  $\beta$ -globin locus.

### MATERIALS AND METHODS

Restriction endonucleases, T4-DNA ligase and calf intestinal phosphatase were from N.E.Biolabs and Boehringer-Mannheim GmbH, nitrocellulose filters (HAT F) from Millipore.

#### The preparation of the cosmid vectors

All the vectors were prepared by standard procedures for restriction enzyme digests, DNA fragment isolations, linker additions and ligations. The various cosmids were constructed by using fragments and linkers from the following sources: pTKM2 (8) pSVgpt (10) pJB8 (15), pAG60 (12), pBR327 (16), pHC79 (17), BamHI and EcoRI linkers (N. E. Biolabs). The resulting plasmids were transformed into E.coli HB101.

#### Construction of cosmid libraries

The recombinant cosmid described in the expression experiments was isolated from a library which was derived from human placental DNA from a family without any known haematologic disease. The method for the isolation of the DNA was as described previously (18). Extreme care was taken in the preparation to avoid shearing of the DNA, resulting in a DNA preparation that has a molecular length above 100kb. This is a crucial

factor to obtain high packaging efficiencies in the construction of a DNA library.

The libraries were constructed and screened as described (19) with a number of modifications, resulting in the following procedure:

Preparation of the packaging extracts

This procedure is exactly as published (19) with one optional modification. The buffer M1 (containing spermidine and putrescine) can be replaced by Q1 (containing 100mM spermidine without putrescine). This results in larger DNA inserts (40 kb) in the recombinant cosmids (20) without any loss of packaging efficiency.

Preparation of chromosomal DNA for cloning

The high molecular weight (M.W.) chromosomal DNA is digested with MboI or Sau3A (0.5 units per  $\mu\text{g}$  of DNA) for a series of increasing times. Three digestion times, a short time (DNA mostly  $>45\text{kb}$ ), a middle time ( $45\text{kb} < \text{DNA} < 35\text{kb}$ ) and a long time (DNA mostly  $<30\text{kb}$ ) are chosen for the preparative digestion (100-300 $\mu\text{g}$  of DNA). After digestion the three digests are phenolized, combined and ethanol precipitated.

The DNA was dissolved in 10mM Tris pH7.5 1mM EDTA (Tris-EDTA) and loaded onto a 13mls 1.25M-5M NaCl (Tris EDTA) gradient. Centrifugation was for 3 hrs at 39K rpm in an SW40.1 Beckman rotor.

Twelve fractions were collected, diluted with one volume Tris EDTA and ethanol precipitated. The DNA of the middle six fractions was pelleted, the DNA dissolved in 40 $\mu\text{l}$  10mM Tris pH7.5 and analyzed by gel electrophoresis.

Three subsequent fractions with a M.W. around 35-45kb were tested for the capacity to self ligate and for packaging efficiency.

Preparation of the vector DNA arms

Two sets of enzymes were used for all the cosmids to prepare the vector arms (15):

ClaI, BstEII and BamHI for the cosmids with the BamHI site at map position 0kb (Fig. 1) and ClaI, HpaI and BamHI for the cosmids with the BamHI site at map position 2.5kb (Fig. 1). 20 $\mu\text{g}$  of cosmid DNA was digested to completion with BstEII (EcaI) or HpaI and another 20 $\mu\text{g}$  with ClaI.

EDTA was added to a final concentration of 10mM, followed by phenol extraction and precipitation.

Both DNA samples were dissolved in 18 $\mu\text{l}$  PME buffer (10mM Tris pH9.5, 1mM spermidine, 0.1mM EDTA) and incubated for 30' at 37°C with 0.05 U (2  $\mu\text{l}$ ) of calf intestinal phosphatase. The samples were phenolized and the DNA ethanol precipitated. The DNA was dissolved in 20 $\mu\text{l}$  10mM Tris

pH7.5 and checked for its inability to ligate. Each sample was degraded with BamHI, followed by a test to establish whether the DNA could be ligated again. The DNA is phenolized, ethanol precipitated and redissolved in 10mM Tris pH7.5.

At this stage the "correct" Bam-Cla and Bam-Bst or Bam HpaI fragment could be isolated by preparative gel electrophoresis (see Fig. 1); usually the packaging efficiency is high enough to omit this step.

### Ligation and packaging of vector and donor DNA

0.5 $\mu$ g fractionated chromosomal DNA and 0.3 $\mu$ g of each of the vector preparations were ligated in a volume of 5 $\mu$ l for 3 hrs. at 15 $^{\circ}$ C. The ligated DNA was packaged without further treatment by mixing in the following order:

7 $\mu$ l of buffer A (19), 1 $\mu$ l of ligated DNA (up to 0.25 $\mu$ g of total DNA), 1 $\mu$ l of Q1 (see above) or M1 (19) buffer, 3.5  $\mu$ l of sonicated extract and 5 $\mu$ l of freeze-thaw lysate. The mixture was incubated for 1 hr at 25 $^{\circ}$ C and diluted by mixing in 230 $\mu$ l of sterile phage buffer (6mM Tris pH8.0, 10mM MgCl<sub>2</sub>, 100mM NaCl 0.5 mg/ml gelatin). Packaged cosmids can be stored over chloroform at 4 $^{\circ}$ C for more than one year.

The transduction was carried out as described (18) using E.coli ED8767 (K. Murray) as the host. The bacteria were plated on nitrocellulose filters on ampicillin containing L-broth agar plates.

The modified procedure yielded normally 100.000-300.000 colonies per  $\mu$ g of fractionated donor DNA. The packaging procedure was scaled up to prepare a complete library.

### Plating and screening

The library was plated and screened as described (18) with two modifications: the bacteria are plated on much larger filters (23 x 23cm) and at a much higher density, 150.000-200.000 colonies per large filter. Both these modifications reduce the amount of work involved in the replica plating and screening of the library.

### Isolation of DNA and RNA

The method for the isolation of DNA from transformed cells was as described previously (18).

RNA was extracted as described by Auffrey and Rougeon (21) with some modifications. The cells were homogenized in 5ml 6M urea, 3M LiCl and precipitated overnight at 0 $^{\circ}$ C. The precipitate was collected by centrifugation for 30 min. at 30,000 rpm at 4 $^{\circ}$ C in a Beckman SW50.1 rotor and taken up in 200 $\mu$ l 15mM NaCl 1mM EDTA, 0.1% SDS containing 2mg per ml

proteinase K. After 30 min. digestion at 37°C the nucleic acids were extracted with an equal volume of phenol, chloroform, isoamylalcohol (25:24:1) and precipitated with ethanol. The precipitated RNA was centrifugated and taken up in 200µl 10mM Tris-HCl pH7.5, 10mM MgCl<sub>2</sub>, 50mM NaCl and digested with 20µg per ml DNaseI for 10 min. at 37°C. The DNase was removed by treatment with 1mg proteinase K per ml in the presence of 0.1% SDS for 10min. at 37°C. The RNA was extracted with phenol, chloroform isoamylalcohol and precipitated with ethanol.

#### Transformation of mouse L-cells

Transformation of mouse L-cells (LMtk<sup>-</sup>) using tk-cosmid DNA and subsequent selection for tk<sup>+</sup> transformants in HAT medium was carried out essentially as described by Wigler *et al.*, (7). Mouse L-cells grown in DME supplemented with 10% newborn calf serum were seeded at a density of about 1-2 x 10<sup>6</sup> cells per 90 mm dish 24 hours prior to transfection. 1ml aliquots of Ca<sub>3</sub>(PO<sub>4</sub>)<sub>2</sub> DNA co-precipitates were prepared using appropriate amounts of cosmid DNA in the presence of 20 or 10µg of salmon sperm carrier DNA per 90 mm dish. Co-precipitates were added directly to 10ml of medium over the L-cells and were left for 8 hours. The precipitate was removed, the cells rinsed once in PBS and fresh medium was added. Selective medium (HAT) was added after 20 hours (on day 1) and was changed at days 3, 6 and 10. tk<sup>+</sup> colonies growing in HAT medium were detected first on about day 6-8 and were scored by microscopic examination on day 14-16 before secondary colonies started growing.

#### Preparation of terminally labelled probes for S<sub>1</sub>-mapping

To map the 3' ends of β-globin mRNA, pHβ<sup>0</sup>G1 (22) was digested with EcoRI and MspI, and the fragments were separated on a 0.7% agarose gel. The 760bp EcoRI/MspI fragment was recovered from the agarose gel by DEAE cellulose chromatography and 3' end labelled using <sup>32</sup>P-dATP and <sup>32</sup>P-TTP and reverse transcriptase in 6mM MgCl<sub>2</sub>, 40mM KCl 50mM Tris pH8.3 for 90 min. at 37°C. The reaction was terminated by phenol extraction and the DNA was purified by gel filtration over Sephadex G50 followed by ethanol precipitation. The specific activity of the 3' β-globin gene fragment was 6 x 10<sup>5</sup> cpm per p mol. of DNA fragment.

To determine the 5' end of the β-globin mRNA produced in L cells, a fragment was prepared by digesting a plasmid containing the human β-globin gene (pAT153-4.4kb Pst-β-globin gene fragment) with CvnI. The DNA was recovered by phenol extraction and ethanol precipitation. The sample was

then dephosphorylated and 5' end labelled using T4 polynucleotide kinase and  $^{32}\text{P}$ -ATP and the entire digest was loaded on an acrylamide gel (23). After electrophoresis the 1205bp fragment, which contains the 5' end of the  $\beta$ -globin gene, was isolated according to Maxam and Gilbert (23). This fragment was further digested with MboII, phenol extracted, ethanol precipitated and loaded on a strand separation gel (23). The 215 NT CvnI-MboII fragment, which contains the 5' end of the  $\beta$ -globin gene, was ethanol precipitated and used for  $S_1$ -mapping. The specific activity of the 5' fragment was  $3.6 \times 10^6$  cpm per p mol.

### $S_1$ -mapping

The  $S_1$ -mapping procedure was as described (24, 25, 26). For detection of the 5' end the RNA was hybridized to the 5' terminal fragment at  $37^\circ\text{C}$ ; the 3' end determination was done by hybridization to the 3' terminal fragment at  $54^\circ\text{C}$  (25). The  $S_1$  nuclease treatment was for 2 hrs. at  $20^\circ$  using 3000 units of  $S_1$  nuclease in each  $300\mu\text{l}$  digest.

## RESULTS

### Construction of expression-cosmid vectors

The cosmid vector pJB8 (5.3kb) has a single BamHI site suitable for cloning (15); it carries the  $\beta$ -lactamase gene and has been used to generate a library of 35-40kb segments of human DNA (19). We used this vector as a starting point in the construction of the expression cosmids pRT1 and pOPF (Fig. 1). First a 1.5kb fragment between a single HpaI and BalI site was deleted to yield pJBF. Into this cosmid we inserted a PvuI-BamHI fragment containing the tk gene from pTKLM16 (Lund and Mellor, unpublished) resulting in the tk-cosmid pRT1 (Fig. 1). Finally, the 1.1 HindIII segment of SV40 which contains the replication origin, was inserted in the HindIII site of pRT1 to generate the SV40-tk cosmid vector pOPF (Fig. 1). To verify that the HSV-1 tk gene was intact in these vectors, we showed that this DNA could transform mouse L tk<sup>-</sup> cells to the tk<sup>+</sup> phenotype at high efficiency (Table I).

A second set of cosmids was constructed on the basis of pOPF and the plasmid pBR 327 (16) which is free of the sequences thought to inhibit replication in eukaryotic cells (27). The pOPF and pRT cosmid sequences from ClaI to PvuI were replaced by the equivalent sequences (cloned in pBr327) resulting in the tk-cosmids pHEP and pSAE (fig. 1). A different eukaryotic marker was introduced by replacement of the central part of the tk-gene by the dominant marker aminoglycosyl 3' phosphotransferase (11) (agpt)



TABLE I

DNA	FORM I		PvuI cut	
	Experiment 1	Experiment 2	Experiment 1	Experiment 2
pLM16	169	332	400	1844
pRT1	26	828	66	184
pOPF	29	479	169	272
TK2	37	392	34	7
TK5	93	972	77	36
SV2	55	110	30	N/D
SV3	88	952	78	N/D
SV4	152	748	46	47
SV6	38	689	76	38

Transformation of L tk<sup>-</sup> cells with tk-cosmids. Form I or PvuI cut tk-cosmid DNA was prepared for Ca<sub>3</sub>(PO<sub>4</sub>)<sub>2</sub> transformation as described in Materials and Methods. In each case, 1 ml of Ca<sub>3</sub>(PO<sub>4</sub>)<sub>2</sub> precipitate containing 100 ng of cosmid DNA and 20µg of salmon sperm DNA was added to 10 ml of H21 medium supplemented with 10% newborn calf serum in an 80 mm dish seeded with 10<sup>6</sup> L tk<sup>-</sup> cells 24 hours prior to the experiment. After 8 hours the precipitate was removed, the cells were rinsed once in PBS and fresh medium added. Selective medium (HAT) was added after a further 20 hours (day 1) and changed subsequently on days 3, 6 and 10. tk<sup>+</sup> colonies were counted on day 16 (experiment 1) or day 14 (experiment 2). Different passages of L tk<sup>-</sup> cells were used in experiments 1 and 2. Cells used in experiment 2 were growing faster than those used in experiment 1 prior to use in the transformation experiment. tk2 and 5 are two human DNA-pRT cosmids, SV2, 3, 4 and 6 are four human DNA-pOPF cosmids.

constructed by positioning the single BamHI site at a different site. The BamHI site in pTM was removed by digestion with BamHI, removal of the "sticky" ends (28) and religation. A new BamHI site was introduced by ligation of a BamHI linker in the single SallI site, which restores the SallI site immediately flanking the BamHI site on both sides. This has the additional advantage that the inserted DNA in a recombinant cosmid can be removed completely by a single SallI digestion. The resulting cosmid pTCF was used to construct three other cosmids, pTBE, pGNC and pNNL (Fig.1) by either removal or replacement of the agpt sequences with the tk gene or the guanine phosphoribosyltransferase (gpt) gene, obtained from the plasmid SVgpt2 (10).

This last set of cosmids (pTBE, pGNC, pTCF and pNNL) can be used to exchange eukaryotic markers in recombinant cosmids. This has two



advantages: 1) the marker can be chosen in accordance with the cell phenotype (as pGNC for tk<sup>-</sup> cells). 2) a eukaryotic cell can be transformed several times by different recombinant cosmids, using a different marker for each transformation. This procedure, (see Fig. 2 ) is based on two principles: 1) the position of the cos site, the cloning site (BamHI) and the eukaryotic marker in relation to each other and three restriction enzyme sites, which each contain two CG base pairs in their recognition sites, i.e. XmaIII, PvuI and NruI (Fig. 1 ). 2) The low occurrence of CG base pairs in eukaryotic DNA's (29).

The frequency of any of these sites occurring in a random piece of 40kb of human DNA is 19%. Consequently, the chance of all three sites occurring in the same stretch of DNA is less than 1%. It follows that >99% of the recombinant cosmids obtained from a library will not contain one of the three sites in the inserted DNA.

The recombinant can, therefore, be cleaved with, for instance, PvuI in the vector only. After treatment with phosphatase to prevent self ligation, a new cosmid with a different marker, also cleaved with PvuI can be ligated to the recombinant. Packaging of this ligated cosmid will result in a new recombinant cosmid in which the eukaryotic marker is

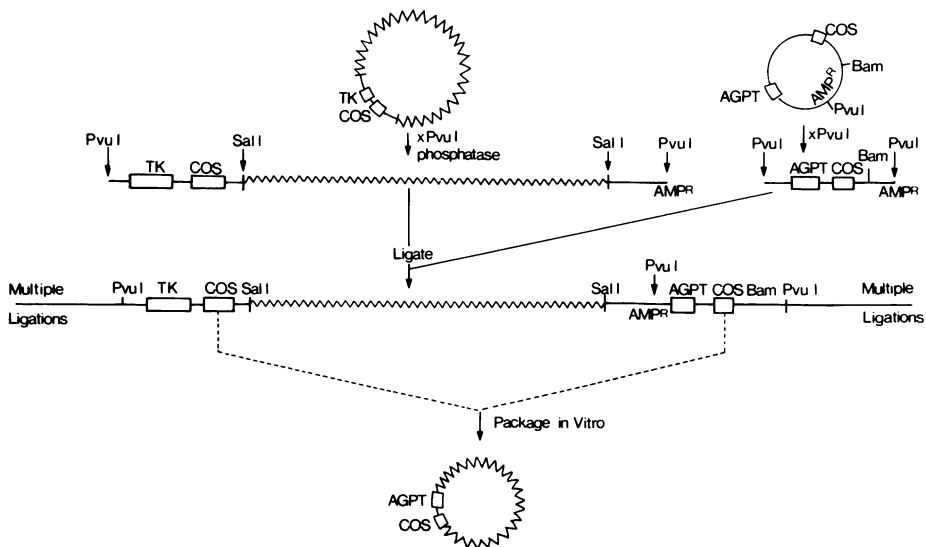


Fig. 2. Exchange of eukaryotic marker in a recombinant cosmid, symbols as in Fig. 1.

exchanged (Fig. 2 ). In the vast majority of the few recombinant cosmids that contain all three of the above enzyme sites, the marker could still be exchanged easily by the use of a combination of the above enzymes.

To date we have constructed human and mouse libraries with the cosmids pOPF, pSAE, pTM and pTCF. Cosmids from all of these libraries have been used to transform a number of different cell types successfully (L cells, MEL cells, K562 cells, rat 1 cells and rat 2 cells). In this paper we present the results from cosmids of the first library (pOPF).

tk cosmids transform tk<sup>-</sup> L cells to tk<sup>+</sup>

Preliminary to the construction of a cosmid library with these vectors, we prepared a number of human DNA-containing cosmids and checked that these had the expected 35kb inserts by isolating the cosmid DNA and digesting this with EcoRI. By this criterion, all of these clones contained inserts of about 35kb.

To establish that our cosmids could indeed confer the tk<sup>+</sup> phenotype on mouse L cells, we chose at random four human DNA cosmids prepared with pOPF1 and two human DNA cosmids prepared with pRT1 (Table I). It can be seen that all cosmids transformed L cells at high efficiency, either in a circular or linear form and that there is no significant effect of the SV40 origin fragment on transformation frequency. The yield of tk-positive L cell clones is broadly proportional to the amount of cosmid DNA applied over a 100 fold range (Table II) and approximately 0.3 positive colonies are obtained per ng cosmid DNA. L cell transformants were cloned, grown up and cellular DNA isolated. The DNA was cleaved with EcoRI, electrophoresed in an agarose gel and blotted onto a nitrocellulose filter. The blots were hybridized with the 2.4kb EcoRI fragment of pTKLM16, which contains the HSV tk gene (data not shown). Most of the clones grown from transformants generated with circular cosmid DNA contained the expected 2.4kb HSV1 tk gene fragment. Some of the clones generated by transformation with PvuI cut cosmids showed more than one band homologous to the tk probe suggesting that more than one cosmid copy is present. We have been unable to detect free circular DNA in the transformed L cells by blotting undigested DNA or by EtBr CsCl fractionation of cellular DNA followed by blotting (data not shown).

### Construction of a human cosmid library in pOPF1

A human library containing pOPF1 was constructed as described in Materials and Methods. A library of approximately 350,000 recombinant (SVtk) cosmids was generated after plating and growth on nitrocellulose

TABLE II

DNA	Amount (ng)	Salmon Sperm DNA (μg)	Colonies
pLM16	50	19	27
pOPF1	50	10	37
SV3	100	-	4
SV3	10	10	3
SV3	100	10	88
SV3	500	10	185
SV3	1000	10	332

Effect of increased amounts of tk-cosmid on transformation efficiency. Form I tk-cosmid was mixed with salmon sperm DNA (where appropriate) and CaPO<sub>4</sub> co-precipitates were prepared, as described in Materials and Methods. 1 ml of precipitate containing the amounts of DNA indicated was added to L tk<sup>-</sup> cells and the experiment was carried out subsequently as described in the legend to Table I and the passage of L cells used was that of experiment I (Table I). Colonies were counted on day 14.

filters. Screening of this library with probes for human ε, γ, δ and β-globin genes led to the isolation of a series of overlapping clones whose map co-ordinates are shown in Fig. 3. Southern blotting experiments,

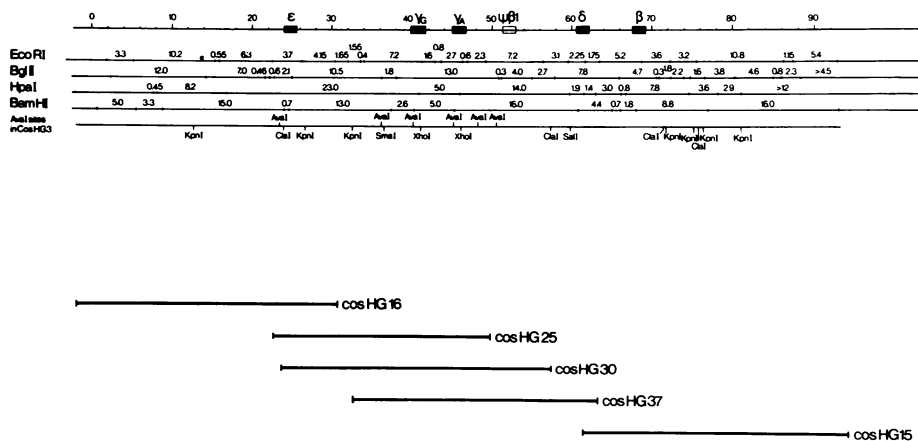


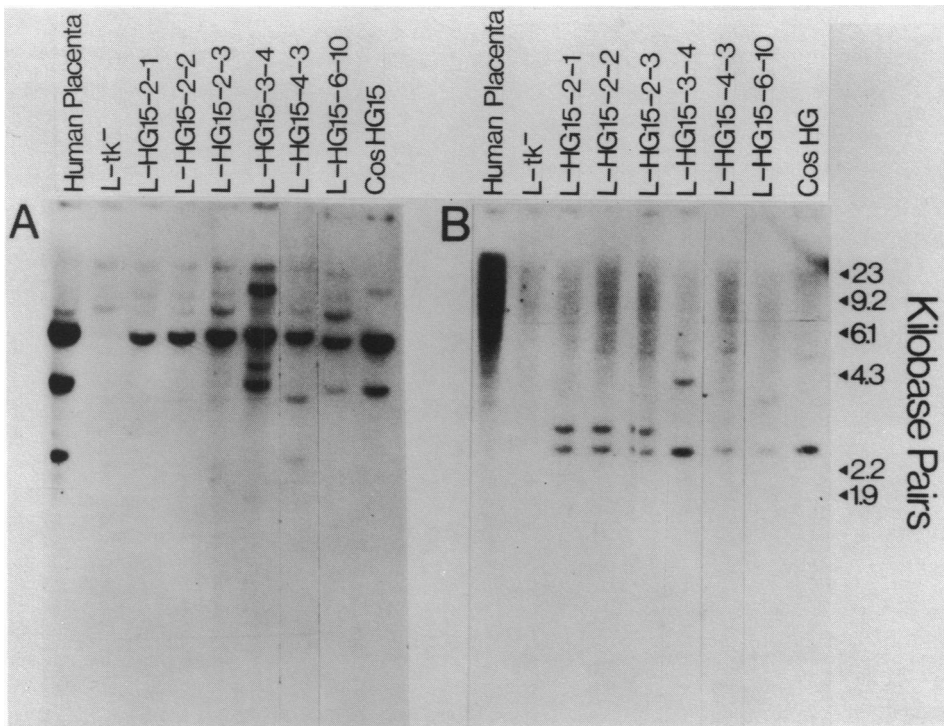
Fig. 3. Physical map of the human β-globin region and location of cosmid clones containing the human β-globin region. The positions of the globin genes were summarized from a large number of publications. For a recent review see Maniatis *et al.* (1).

together with comparison of the restriction maps of the cloned DNA and the known map of the human  $\beta$ -globin gene region establishes that these cosmid clones are colinear with the human DNA in this region.

### Expression of SVtk cosmids in stably transformed cell lines

To test whether the genes carried on SVtk cosmids are expressed specifically in transformed cells, we transformed mouse L tk<sup>-</sup> cells with cos HG15 (see Fig. 3 ) and selected a number of tk<sup>+</sup> transformants. This cosmid contains the  $\beta$ -globin gene flanked by 7kb of human DNA to the 5' side of the gene and about 25kb to the 3' side. DNA was isolated from these clones and hybridized with probes for the human  $\beta$ -globin gene or the HSV-1 tk gene (Fig. 4). As expected, L cells transformed with cos HG15 contained both components. The copy number of the tk and globin components is similar and low in all cases - there appear to be from 1-3 copies per diploid genome when the blots are compared with standards such as human placental DNA. The 5.2kb 5' EcoRI fragment of the  $\beta$ -globin gene in the L cells is intact in all clones examined, but the 3' 3.5kb fragment is only seen in clones L-HG-3-4 and 6-10. This suggests that an integration event has occurred in the latter fragment in clones L-HG 15-2-3 and H-3. Further mapping experiments (data not shown) indicate that the integration event in the first two clones has occurred close to the junction of the inserted and cosmid DNA on the 3' side of the  $\beta$ -gene. Three subclones of an L cell clone (LHG15-2) have been analyzed after 28 generations. All three subclones exhibit the same blotting pattern for both the tk and the globin genes. In fact, two tk-gene containing fragments are observed which probably results from two integrated tk genes, since the smaller of these fragments contains the intact EcoRI fragment of the tk gene which was used as a probe and which is common to all the clones that have been analyzed.

To determine the level of expression of the  $\beta$ -globin gene in these clones, we used S<sub>1</sub> nuclease mapping (24, 26). The 5' end of the  $\beta$ -globin mRNA was determined by S<sub>1</sub> nuclease mapping using as probe a 212NT strand-separated MboII/CvnI fragment labelled at its 5' CvnI end. S<sub>1</sub> nuclease treatment of hybrids formed with both  $\beta^S$ -globin mRNA from a patient with sickle cell anaemia, or from the transformed L cells, gave the same fragment of about 70 nucleotides. This fragment co-migrates with the A+G sequence-marker DNA fragments corresponding to the underlined region of CTT<sup>\*</sup>ACATT (where <sup>\*</sup>A is the capped nucleotide of human  $\beta$ -globin mRNA). Clearly, in all cases the only detectable 5' end of the  $\beta$ -globin mRNA produced in L cells corresponds to the natural cap site of  $\beta$ -globin mRNA. The 3' end of

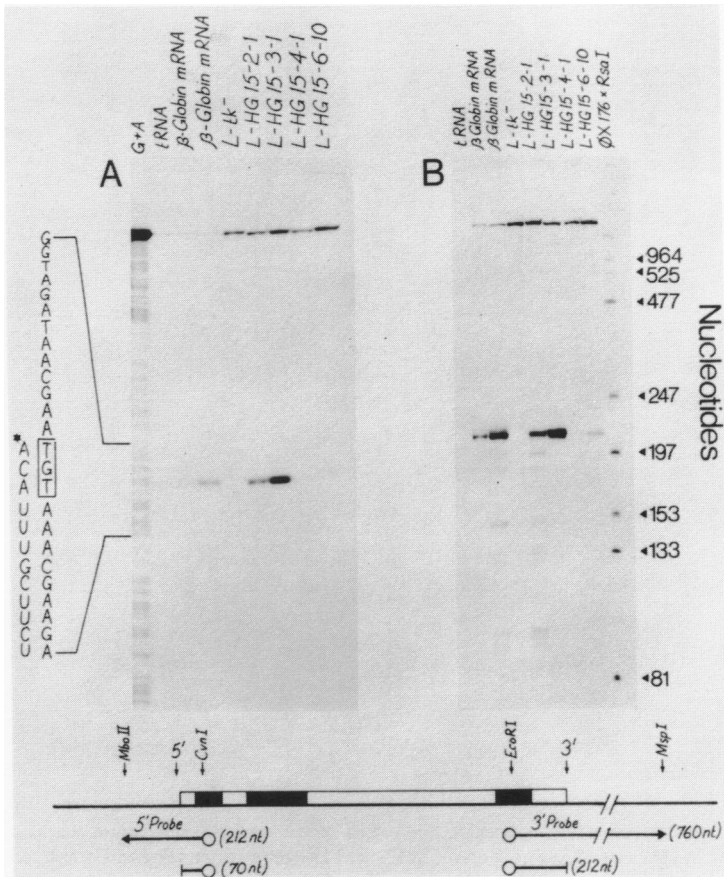


**Fig. 4.** The presence of the human  $\beta$ -globin gene and HSV-tk gene in L cells transformed with cosHG15. Mouse Ltk<sup>-</sup> cells were transformed with undigested cosHG15 and selected in HAT medium. The DNA was digested with EcoRI and approximately 10 $\mu$ g digest was separated on a 0.5% agarose gel, transferred to a nitrocellulose filter and hybridized with <sup>32</sup>P-labelled 4.4kb PstI fragment containing the human  $\beta$ -globin gene (A) or <sup>32</sup>P-labelled 2.4kb EcoRI fragment containing the HSV-tk gene (B).

L-HG15-2-1, L-HG15-2-2 and L-HG15-2-3 are different subclones of the same transformant. In the cosHG15 lane, approximately 10  $\mu$ g of cosHG15 was mixed with 10 $\mu$ g of spleen DNA from B10 mice.

the  $\beta$ -globin mRNA was also determined to be correct by S<sub>1</sub> mapping, using as probe a 760bp EcoRI/MspI fragment 3' labelled at its EcoRI site. Again, the same 3' end was observed for both human  $\beta^s$ -globin mRNA and the human  $\beta$ -globin mRNA from transformed L cells (Fig. 5).

We have estimated the level of  $\beta$ -globin mRNA produced in the L cell clones by comparing the yield of  $\beta$ -globin DNA/RNA hybrids with the yield obtained with a standard which contains a known amount of  $\beta^s$ -globin mRNA. By this criterion, the level of  $\beta$ -globin transcripts varies from 150 to 5000 copies per cell with different L cell clones (Table III).



**Fig. 5.** Detection of  $\beta$ -globin transcripts in L-HG15 transformants.

S<sub>1</sub>-mapping was performed with total RNA from the isolated L-HG15 transformants and single stranded <sup>32</sup>P 5' labelled CvnI MboII coding fragment (A) or double stranded <sup>32</sup>P 3' labelled EcoRI-MspI fragment (B). After S<sub>1</sub> treatment the 5' and 3' fragments were separated on an 8% or 5% polyacrylamide sequencing gel (23) respectively.

(A) Lane 1 fragments from a G+A cleavage of the 5' CvnI-MboII coding strand.

Lanes 2-9: the 5' labelled CvnI-MboII fragment was mixed with 5  $\mu$ g tRNA (lane 2), 0.5  $\mu$ g (lane 3) and 2  $\mu$ g (lane 4) of PolyA enriched RNA from patient with sickle cell anaemia, 20  $\mu$ g RNA from Ltk<sup>-</sup> cells (lane 5), 20  $\mu$ g RNA from L-HG15-2-1 (lane 6), 20  $\mu$ g RNA from L-HG15-3-1 (lane 7), 20  $\mu$ g RNA from L-HG15-4-1 (lane 8) and 20  $\mu$ g RNA from L-HG15-6-10 (lane 9).

(b) In lanes 1 to 8, the 3' EcoRI-MspI fragment was mixed with 5  $\mu$ g tRNA (lane 1), 0.5  $\mu$ g poly(A) RNA from a patient with sickle cell anaemia (lane 2), 2  $\mu$ g of poly(A) RNA from a patient with sickle cell anaemia (lane 3), 20  $\mu$ g RNA from Ltk<sup>-</sup> cells (lane 4), 20  $\mu$ g RNA from L-HG15-2-1 (lane 5), 20  $\mu$ g RNA from L-HG15-3-1 (lane 6), 20  $\mu$ g RNA from L-HG15-4-1 (lane 7) and 20  $\mu$ g RNA from L-HG15-6-10 (lane 8). Lane 9 <sup>32</sup>P-end labelled ØX174 x TaqI DNA fragments.

TABLE III

<u>L-cell clones</u>	<u><math>\beta</math>-globin mRNA molecules per cell</u>
L-HG15-2-1	1700
L-HG15-3-1	5000
L-HG15-4-1	150
L-HG15-6-10	700

Level of human  $\beta$ -mRNA molecules in different L-HG15 tk<sup>+</sup> transformed cells. The numbers of  $\beta$ -globin mRNA molecules per cell were estimated from Fig. 5 using the assumption that 25% of the polyA enriched RNA from the patient with sickle cell anaemia is  $\beta$ -globin RNA and that an L-cell contains 40 pg RNA.

There is no obvious relationship between the apparent site of integration of the cosmid and the level of expression of the  $\beta$ -globin gene. Two L cell clones do not exhibit a normal 3.5kb EcoRI fragment which contains the 3' end of the  $\beta$ -globin gene, yet one (L-HG15-2-1) shows a high level of  $\beta$ -globin mRNA, while the other (L-HG15-4-3) gives a low level.

#### DISCUSSION

In this article we describe the development of a cosmid cloning system that has made possible the construction of libraries of human DNA-cosmid recombinants, each member of which can be selected for growth in bacterial cells, on the basis of ampicillin resistance, or for stable integration into animal cells by selection for the presence of the HSV-1 tk, the E.coli gpt or the Tn5 agpt gene.

We have used these libraries to isolate several overlapping cosmids which contain the human  $\beta$ -globin related genes. These cosmids are colinear with the known map of the chromosomal DNA in this region. In addition, at least a dozen other genes or gene complexes have been isolated from these libraries in collaboration with others and significantly we have never failed to find a gene when the libraries were screened for it. It seems likely, therefore, that such libraries are broadly representative of the genome.

Upon introduction into L cells, the  $\beta$ -globin gene present on one of these cosmids is expressed when the cosmid genomes are stably associated (presumably integrated) with the L cell genome. The level of expression of the  $\beta$ -globin gene in four different transformants varies from 150 to

5000 copies of mRNA per cell. This is a similar value to that seen in proerythroblasts (30).

In L cells, the 5' and 3' ends of the human  $\beta$ -globin mRNA are the same as natural  $\beta$ -globin mRNA. This contrasts with the results obtained by Wold *et al.* (9) who introduced the rabbit  $\beta$ -globin gene into L cells by cotransformation with the marker; in this case, five copies of  $\beta$ -globin mRNA were obtained per cell, but the 5' ends of this RNA were about 50 nucleotides to the 3' side of the natural  $\beta$ -globin cap site. In addition, in experiments where the rabbit  $\beta$ -globin gene was ligated to the gene and introduced into L cells, Mantei *et al.* (8) and Dierks *et al.*, (31) observed a variable level of RNAs with the same 5' ends as rabbit  $\beta$ -globin, together with significant levels of RNAs that originate far upstream from this site and RNAs with 5' ends within the  $\beta$ -globin gene. In this case, from 50 to 2000 copies of rabbit  $\beta$ -globin mRNA were obtained per cell (8). This corresponds to a level of about 50-100 copies of  $\beta$ -globin mRNA per  $\beta$ -globin gene, assuming that all globin genes are expressed. Abnormal 5' ends were also seen for ovalbumin mRNA derived from the chicken ovalbumin gene in L cells (32). It is not clear why this qualitative difference exists between our results and the data in refs. (8, 9, 31). Although it may be thought that the SV40 origin region could influence this, in fact we have recently analyzed L cell clones derived from globin cosmids lacking this viral DNA segment and found the same specificity and level of expression in L cell clones (H. H. M. Dahl, unpublished observations). From these data it is also clear that the efficiency of transformation is not influenced by the presence of the SV40 origin in the cosmids in contrast to the data obtained by DNA injections (14). One possibility is that the correct specificity seen in our case is a consequence of the fact that a large part ( $\pm$  35kb) of the  $\beta$ -globin gene region has been introduced into these cells. Alternatively, it remains possible that this is a genuine difference between the human  $\beta$ -globin gene and the other gene systems used, irrespective of the size and the integration site of the  $\beta$ -globin gene DNA. It may be significant to note that the integration event in all the clones tested seems to have taken place in DNA fragments that contain repetitive DNA sequences. From the data presented and our unpublished data involving multiple selections with different eukaryotic markers (gpt, agpt, tk), different gene inserts (class I and II MHC antigens,  $\alpha$ - and  $\beta$ -globin genes) and different eukaryotic cells (mouse MEL cells, human K562 cells, mouse L cells, rat 1 and rat 2 cells), it seems



likely that any cosmid clone carrying a functional gene(s) from these libraries can be assayed rapidly for phenotypic expressions. Since the cloned segments are between 33 to 43kb, it should be possible to isolate all but the largest genes from these libraries and to study their expression directly without any further manipulation.

#### ACKNOWLEDGEMENTS

T.L. was a grateful recipient of an EMBO long term fellowship. We thank Drs. C. Weissmann, P. Berg, D. Ish Horowicz, A. Garapin, F. Bolivar and J. Collins for their generous gifts of the different plasmids and C. O'Carroll for preparing the manuscript. This work was supported by the British Medical Research Council.

#### REFERENCES

1. Maniatis, T., Fritsch, E. F., Lauer, J. and Lawn, R. M. (1981). *Ann. Rev. Genetics* 14, 145-178.
2. Mertz, J. E. and Gurdon, J. B. (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74, 1502-1506.
3. Kressmann, A., Clarkson, S. G., Telford, J. L. and Birnstiel, M. L. (1977) *Cold Spring Harbor Symp. on Quant. Biol.* 42, 1077-1082.
4. Weil, P. A., Luse, D. S., Segall, J. and Roeder, R. G. (1979). *Cell* 18, 469-484.
5. Manley, J. L., Fire, A., Cano, A., Sharp, P. A. and Gefter, M. L. (1980). *Proc. Natl. Acad. Sci. U.S.A.* 77, 3855-3859.
6. Flavell, R. A. (1980) *Trends in Biochemical Sciences*, 5, 313-315.
7. Wigler, M., Sweet, R., Sim, C. K., Wold, B., Pellicer, A., Lacy, E., Maniatis, T., Silverstein, S. and Axel, R. (1979) *Cell* 16, 777-785.
8. Mantai, N., Boll, W. and Weissmann, C. (1979) *Nature* 281, 40-46.
9. Wold, B., Wigler, M., Lacy, E., Maniatis, T., Silverstein, S. and Axel, R. (1979). *Proc. Natl. Acad. Sci. U.S.A.* 76, 5684-5688.
10. Mulligan, R. and Berg, P. (1980) *Science*, 209, 1422-1427.
11. Jimenez, A. and Davies, J. (1980) *Nature* 287, 869-871.
12. Colbere-Garapin, F., Horodniceanu, F., Kourilsky, P. and Garapin, A.C. (1980) *J. Mol. Biol.* 150, 1-14.
13. Graham, F. L. and van der Eb, A. J. (1973) *Virology* 52, 456-467.
14. Capecchi, M. (1980) *Cell* 22, 479-488.
15. Ish-Horowicz, D. and Burke, J. F. (1981) *Nucl. Acids Res.*, 9, 2989-2998.
16. Soberon, X., Corarrubias, L. and Bolivar, F. (1980) *Gene* 9, 287-305.
17. Hohn, B. and Collins, J. (1980) *Gene* 11, 291-298.
18. Grosveld, F. G., Dahl, H. H. M., deBoer, E. and Flavell, R. A. (1981) *Gene* 13, 227-237.
19. Scalenghe, F., Turco, E., Edström, J. E., Pirrotta, V. and Melli, M. (1981) *Chromosoma* 32, 205-216.
20. Meyerowitz, E. M., Guild, G. M., Preslidge, L. S. and Hogness, D. S. (1980) *Gene* 11, 271-282.
21. Auffray, C. and Rougeon, F. (1980) *Eur. J. Biochem.* 107, 303-314.
22. Moschonas, N., deBoer, E., Grosveld, F. G., Dahl, H. H. M., Wright, S. Shewmaker, C. K. and Flavell, R. A. (1981) *Nucl. Acids Res.* 9, 4391-4401.

23. Maxam, A. M. and Gilbert, W. (1980) In: "Methods in Enzymology"  
Eds. Grossman, C. and Moldave, K., Academic Press, New York.  
65, 499-560.
24. Berk, A. J. and Sharp, P. (1977) Cell 12, 721-732.
25. Grosveld, G. C., Koster, A. and Flavell, R. A. (1981a) Cell 23, 573-584.
26. Weaver, R. F. and Weissmann, C. (1979) Nucl. Acids Res. 6, 1175-1193.
27. Lasky, M. and Botchan, M. (1981) Nature 293, 79-81.
28. Vogt, V. M. (1973) Eur. J. Biochem. 33, 192-200.
29. Subak-Sharpe, H., Blrk, R. R., Crawford, L. V., Morrison, J. R.,  
Hay, J. and Keir, H. M. (1966) Cold Spring Harbor Symp. on  
Quant. Biol. vol. XXXI 737-748.
30. Clissold, P. M., Arnstein, H. R. V. and Chesterton, C. J. (1977)  
Cell 11, 353-361.
31. Dierks, P., van Ooyen, A., Mantei, N. and Weissmann, C. (1981).  
Proc. Natl. Acad. Sci. U.S.A. 78, 1411-1415.
32. Breathnach, R., Mantei, N. and Chambon, P. (1980) Proc. Natl. Acad.  
Sci. U.S.A. 77, 740-744.