

Isolation of β -globin-related genes from a human cosmid library

(DNA cloning; lambda phage packaging; cosmid vectors; placental DNA)

Frank G. Grosveld, Hans-Henrik M. Dahl, Ernie de Boer and Richard A. Flavell

Laboratory of Gene Structure and Expression, National Institute for Medical Research, Mill Hill, London, NW7 1AA (U.K.)

(Received January 12th, 1981)

(Accepted January 26th, 1981)

SUMMARY

A human gene library was constructed using an improved cloning technique for cosmid vectors. Human placental DNA was partially digested with restriction endonuclease *Mbo*I, size-fractionated and ligated to *Bam*HI-cut and phosphatase-treated cosmid vector pJB8. After packaging in λ phage particles, the recombinant DNA was transduced into *Escherichia coli* 1400 or HB101 followed by selection on ampicillin for recombinant *E. coli*. 150 000 recombinant-DNA-containing colonies were screened for the presence of the human β -globin related genes. Five recombinants were isolated containing the human β -globin locus and encompassing approx. 70 kb of human DNA.

INTRODUCTION

Development of recombinant DNA technology has made it possible to isolate specific DNA fragments and to study genes in detail (e.g. Maniatis et al., 1980). It is now clear that a number of problems concerning gene structure and expression can be more easily approached if large DNA fragments are isolated. Some of the advantages of cloning large DNA fragments are: (i) large genes can be isolated on a single recombinant molecule; (ii) gene linkage studies are facilitated, since several linked genes can, in many cases, be isolated on the same recom-

binant molecule; (iii) genes can be isolated together with large areas of their surrounding sequences. This might be particularly important when the control of gene expression is studied, for example, by the reintroduction of recombinant genes in eukaryotic cells. There is some evidence that mutations in the human β -related globin gene locus influence DNA sequences far from the site of the mutation. This in turn might suggest that the unit of developmental-stage-specific expression may well be much larger than a single gene and the sequences immediately surrounding it. To obtain specific expression of a cloned gene after its introduction into animal cells, the cloned DNA might therefore need to carry large amounts of the flanking DNA sequences; (iv) fewer colonies (or plaques) have to be screened in a library to isolate the gene of interest,

Abbreviations: bp, base pairs; DTT, dithiothreitol; kb, kilobase pairs; K, 1000 rev./min; SSC, 0.15 M NaCl + 0.015 M Na-citrate, pH 7.8.

since less recombinant molecules are required to represent the entire genome.

Several vector systems such as plasmids, phage λ , and cosmids, are available for cloning in *E. coli* (Morrow, 1980). The λ cloning systems have been used most frequently; in this case, the upper size limit of inserted DNA is approx. 20–25 kb. The size of inserts using plasmids as vectors is theoretically unlimited, but in practice transformation of plasmids larger than approx. 15 kb is very inefficient (Kushner, 1978). In contrast, the cosmid system allows efficient cloning of DNA fragments of up to approx. 45 kb (Feiss et al., 1977; Collins and Hohn, 1978). Although in a single case several ovalbumin-related genes have been isolated on a recombinant cosmid (Royal et al., 1979), this approach has not yet been applied in anything approaching the frequency of cloning in phage λ .

We report here the development of a cloning procedure in cosmids, the efficiency of which made possible the construction of human-DNA cosmid libraries and isolation of recombinant cosmids containing human β -globin-related genes.

MATERIALS AND METHODS

(a) Strains, enzymes

Bacterial strains *E. coli* BHB 2688 and BHB 2690 (Hohn, 1980) were used for preparation of the λ -packaging mix. Packaged DNA was transfected into *E. coli* 1400 (Cami and Kourilsky, 1978) or HB101 (Boyer and Rouland-Dussoix, 1969). Cosmid pJB-8 was kindly obtained from Drs. D. Ish-Horowicz and J. Burck. Restriction endonucleases were from N.E. BioLabs Inc., and nitrocellulose filters were Millipore HAWP or HATF, Sartorius SM or Schleicher and Schüll BA85. Bacterial alkaline phosphatase was from Worthington (BAPF). Human placental DNA (from a family without any known haematologic disease) was purified as described previously (Flavell et al., 1978).

(b) Partial digestion of chromosomal DNA with restriction enzyme *Mbo*I

400 μ g of human placental DNA was digested with 80 units of *Mbo*I for 30 min at 37°C in a final

volume of 2 ml (10 mM Tris · HCl, pH 7.5, 8 mM $MgCl_2$, 6 mM β -mercaptoethanol, 0.01% gelatin). The reaction was terminated by the addition of 0.5 M EDTA (pH 7.5) to a final concentration of 10 mM; the mixture was extracted with phenol–chloroform–isoamylalcohol (25 : 25 : 1) and DNA in the aqueous phase recovered by ethanol precipitation.

(c) Sucrose gradient fractionation

200 μ g of *Mbo*I digested DNA was layered in 200 μ l onto 30 ml 5–25% sucrose gradients in 10 mM Tris · HCl (pH 7.5), 1 mM EDTA and centrifuged for 14 h at 22.5 K in a Beckman SW27 rotor. 2 ml fractions were collected and 20 μ l of each fraction layered on a 0.5% agarose gel for size analysis.

(d) Digestion and phosphatase treatment of the vector DNA

100 μ g of pJB8 DNA was digested with 30 units of restriction enzyme *Bam*HI for 1 h at 37°C, phenol-extracted and ethanol-precipitated. The DNA was dissolved in 100 μ l 10 mM Tris · HCl (pH 8.0), 1 mM EDTA and treated with 25 units of alkaline phosphatase (BAPF, Worthington) for 30 min at 37°C. The DNA was extracted three times with phenol, ethanol precipitated and dissolved in 10 mM Tris · HCl (pH 8.0) at a concentration of 1 μ g/ μ l.

(e) Ligation

In the standard reaction, 1 μ g of chromosomal DNA and 0.8 μ g of vector DNA were ligated in 5 μ l of 50 mM Tris · HCl (pH 8.0), 7.5 mM $MgCl_2$, 1 mM ATP, 10 mM DTT, 0.02% gelatine with T4 ligase, for 12–16 h at 14°C.

(f) Preparation of packaging extracts and packaging of DNA

Two packaging mixes were prepared by a procedure of V. Pirotta (personal communication) using two λ lysogen *E. coli* strains, BHB 2688 and BHB 2690.

Three 500 ml cultures of *E. coli* BHB 2688 and one 500 ml culture of *E. coli* BHB 2690 were grown

in L-broth to a density of $3 \cdot 10^8$ cells/ml and induced by a 15 min incubation, without shaking, in a 45°C waterbath. The cultures were then vigorously shaken for 1 h at 37°C, followed by rapid cooling on ice. The cells were harvested by centrifugation (9 K for 10 min) and the pellets carefully drained of all supernatant.

The *E. coli* BHB 2688 pellets were resuspended in 3 ml cold 10% sucrose, 50 mM Tris · HCl pH 7.5 and dispensed into two 10 ml Oak Ridge ultracentrifuge tubes. 75 µl fresh lysozyme solution (2 mg/ml in 0.25 M Tris pH 7.5) was added to each tube and mixed in gently. The tubes were then frozen in liquid nitrogen and stored at -70°C. The lysates were thawed at 4°C and 75 µl of buffer M1 (10 mM Tris pH 7.5, 0.03 M spermidine and 0.06 M putrescine, both neutralised with Tris, 18 mM MgCl₂, 15 mM neutralised ATP and 0.2% (v/v) β-mercaptoethanol) was added to each tube and mixed gently. The tubes were centrifuged at 35 K for 35 min at 4°C, then the supernatants were removed and aliquoted into precooled tubes. These were frozen in liquid nitrogen and stored at -70°C.

The *E. coli* BHB 2690 pellets were resuspended in 3.1 ml buffer A (20 mM Tris pH 8.0, 3 mM MgCl₂ 0.05% (v/v) β-mercaptoethanol and 1 mM EDTA) and sonicated, avoiding foaming, until the solution was no longer viscous. Debris was pelleted by centrifugation at 6 K for 6 min at 4°C. The supernatant was aliquoted in precooled tubes, frozen in liquid nitrogen and stored at -70°C. Both extracts are stable at -70°C for at least six months.

In a standard packaging reaction, the two extracts were thawed and mixed with up to 0.5 µg ligated DNA in the following order: 7 µl of buffer A, 1–2 µl of ligated DNA, 1 µl of buffer M1, 7 µl of *E. coli* BHB 2690 extract and 10 µl of *E. coli* BHB 2688 extract. The mixture was incubated at 25°C for 1 h and diluted to 250 µl with 10 mM Tris pH 7.4, 100 mM NaCl, 10 mM MgCl₂, 0.2% gelatin.

(g) Transduction and plating of packaged DNA

In a standard transduction 500 µl of late log phase *E. coli* 1400 or HB101 (grown in L-broth containing 0.4% maltose) were pelleted, the cells resuspended in 250 µl 10 mM MgCl₂ and added to 250 µl of packaged DNA (0.5 µg). Transduction was allowed to take place for 15 min at room temperature

after which 2 ml of L-broth was added. The cells were incubated for an additional 30 min at 30°C (1400) or 37°C (HB101) to allow drug resistance to be expressed. The cells were then pelleted and resuspended in L-broth.

(h) Nitrocellulose filters

After transduction, the cosmid-containing bacteria are grown on nitrocellulose filters on agar plates containing the selective medium (as described by Hanahan and Meselson, 1980). Growing the bacterial colonies on nitrocellulose filters facilitates easy handling, replica-plating, detection and storage of positive colonies. In addition, secondary growth is significantly reduced (which occurs when ampicillin-resistant colonies deplete the medium of ampicillin). Various types and brands of nitrocellulose filters were compared for this type of work. Compared with direct plating on agar, we found that Millipore HAWP, HATF and Sartorius filters gave more than 95% plating efficiency, whereas various batches of Schleicher and Schüll gave from 10 to 95% efficiency. Low plating efficiencies are believed to be due to detergents in the filters. All filters tested could be washed and sterilised, after which the plating efficiency for all was higher than 95%. We routinely use Millipore or Sartorius filters, since they can be used directly without requiring sterilisation or washing.

It is also of importance that the filters do not stretch or shrink during the hybridisation procedure. Changes are minimal in the Millipore or Sartorius filters tested; but Schleicher and Schüll filters shrank, which makes the location of positive colonies at high plating density difficult. Finally, the various filters were tested for their durability in several rounds of hybridisation and exposure. Millipore filters were successfully used for three subsequent hybridisations with various probes. Sartorius filters are more fragile, while Schleicher and Schüll filters start to disintegrate after the second hybridisation.

The wetness of the filters during plating is of importance for a high plating efficiency. If the medium, in which the bacteria are spread on the filter, is adsorbed very fast, a low plating efficiency results; too much medium leads to difficulties in drying the plate. The highest plating efficiency was noted when the medium is adsorbed within

0.5–5 min after plating. The bacteria are usually resuspended in 600 μl medium for a 15 cm petri dish (filter) and the plates are dried for approx. 15 min at 37°C prior to plating. After all the medium is adsorbed, the plates are incubated overnight at 30°C.

(i) Replica-plating, storage and screening of the library

For a genomic library up to 10 000 colonies per filter were plated and replica-plated by the method described by Hanahan and Meselson (1980) with some minor modifications. This method is easy and gives 95–100% transfer of colonies to replica filters.

The blank replica filters are placed on agar in petri dishes. A sterilised Whatman No. 1 filter is placed on a glass plate, followed by the wetted nitrocellulose filter, the template filter (colonies downward), and another Whatman No. 1 filter. A glass plate is placed on top and the glass plates are pressed firmly together. The top glass plate and Whatman filter are removed and the nitrocellulose filters are keyed using a needle dipped in Indian ink. The nitrocellulose filters are then separated and put back on agar plates.

Best results are obtained when the colonies to be replica-plated are fresh and with a diameter of less than 1 mm. It is also important for even transfer that the two nitrocellulose filters during replica-plating are of equal wetness. From one master filter four replicas can usually be made. After replica-plating, both the master filter and replica filters are incubated until the colonies have grown up. Because many bacteria in a colony are transferred to the replica filter during this procedure, the colonies will grow up relatively fast, compared to the initial master plating. We have noted that colonies of unequal size on the master filter tend to give more uniformly sized colonies on the replica filters.

We have successfully stored our colonies on the original nitrocellulose filter, by first transferring it to an agar medium with 25% glycerol, as described by Hanahan and Meselson (1980). After a 2 h incubation at 30°C, the plates containing the filters with the glycerol-soaked colonies are stored at –20°C. Bacteria can be recovered by carefully thawing the plates (upside down, to prevent liquid running onto the nitrocellulose filter).

We have also observed that it is possible to isolate the cosmid DNA from a given bacterial strain and then, after *in vitro* packaging, to transfer this to another strain. The efficiency of packaging and transduction (see below) of this process is 10^5 colonies/ μg of hybrid cosmid DNA.

In preliminary experiments a major problem in screening for positive colonies was high unspecific background hybridisation. The background was reduced significantly by modifying the method described by Hanahan and Meselson (1980). When an *E. coli* strain, containing a λ prophage (e.g., *E. coli* 1400), is used, the plates are first incubated at 37°C for 4–6 h to induce the prophage and thus to cause lysis.

Colonies are lysed by floating the nitrocellulose filters (colonies up) for 4 min in 0.5 M NaOH. They are then immediately floated for another 4 min on 1 M Tris, pH 7.4, and washed by shaking in 0.5 M Tris, pH 7.4, 1.5 M NaCl. To remove all cell debris from the filters they are wiped thoroughly with soft medical tissue soaked in $2 \times \text{SSC}$, 0.1% SDS. Following washing in $2 \times \text{SSC}$, the filters are baked for 2–6 h at 85°C. To avoid a high background, it appears to be important that the filters do not dry before the baking step.

The nitrocellulose filters are then treated according to the procedure used for Southern blots in this laboratory (Flavell et al., 1978). The hybridisation and nick translation procedure has been described previously (Bernards and Flavell, 1980).

(j) Preparation of cosmid DNA

Cosmid DNA was prepared by a modification of the procedure of Birnboim and Doly (1979). Recombinant cosmid containing bacteria were grown in 5 ml L-broth (50 $\mu\text{g}/\text{ml}$ ampicillin) to saturation density. The cells were then pelleted, resuspended in 200 μl 25 mM Tris pH 8.0, 50 mM glucose, 10 mM EDTA and lysed by the addition of 400 μl of 0.2 M NaOH, 1% SDS. After 5 min 200 μl of 3 M Na acetate pH 4.8 was added, the solution cooled on ice, and the precipitate pelleted by centrifugation. 0.6 vols. of isopropanol were added to the supernatant and the resulting precipitate was immediately pelleted by centrifugation. The pellet was first washed with 70% ethanol, 100 mM Tris pH 8.0 and then dissolved in 100 μl 10 mM Tris pH 7.5, 150

mM NaCl, 1 mM EDTA, 0.2 $\mu\text{g}/\mu\text{l}$ of RNase. After incubation for 30 min at 37°C, the sample was phenol-extracted and the DNA precipitated by the addition of 2 vols. of ethanol to the aqueous phase.

The DNA was finally collected by centrifugation and dissolved in 10 mM Tris pH 7.5.

RESULTS

The general procedure as represented in Fig. 1, includes the following steps:

- (i) the preparation, ligation and packaging of chromosomal and vector DNA;
- (ii) transduction of the packaged DNA into *E. coli* and plating on a selective medium to obtain a genomic library, and
- (iii) replica-plating, storage and screening of the library.

(a) Preparation and ligation of chromosomal and vector DNA

To obtain a random cosmid library, human placental DNA was partially restricted with *Mbo*I. Since

*Mbo*I produces DNA fragments with the same 5'-single-stranded ends as *Bam*HI, the *Mbo*I-cut fragments can conveniently be inserted into a cosmid containing a single *Bam*HI site (Roberts, 1980). Obviously the use of a "4 bp" enzyme (like *Mbo*I) rather than a "6 bp" specific endonuclease (like *Bam*HI, *Eco*RI) to restrict the chromosomal DNA minimises the risk of obtaining a non-random population of cloned DNA fragments. The chromosomal DNA was extracted as described (Flavell et al., 1978); the size of the starting DNA should be large enough to obtain 30–50 kb fragments with two *Mbo*I-cut ends per fragment after degradation. After determining the incubation time to produce the optimal 30–50 kb fragment distribution, 400 μg of human placental DNA was digested with *Mbo*I and the partially degraded DNA was size-fractionated by sucrose-gradient centrifugation or preparative agarose gel electrophoresis. Sucrose-gradient centrifugation is faster and usually yields DNA which ligates well, although agarose-gel electrophoresis gives a better size fractionation in a single round of purification. Fractions were analysed by agarose-gel electrophoresis (Fig. 2A) and the appropriate fractions pooled. The DNA was ethanol precipitated and redissolved in 10 mM Tris · HCl (pH 8.0).

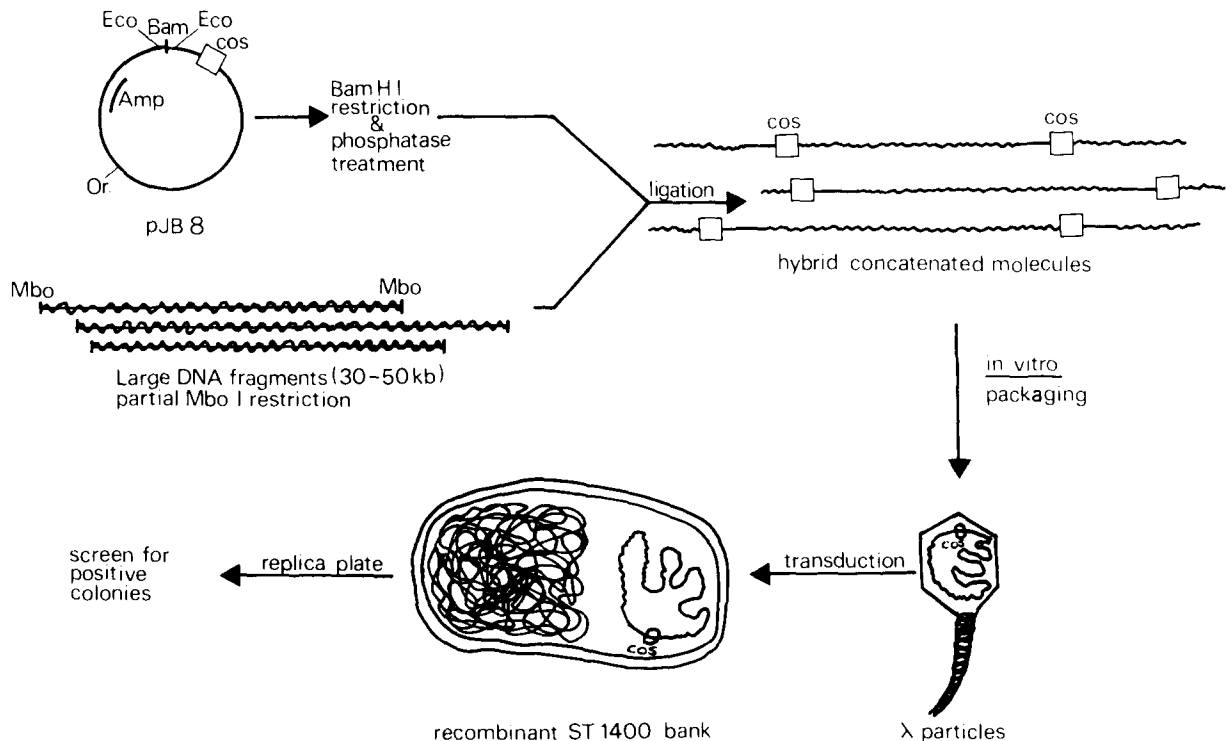


Fig. 1. The DNA cloning procedure using cosmid pJB8.

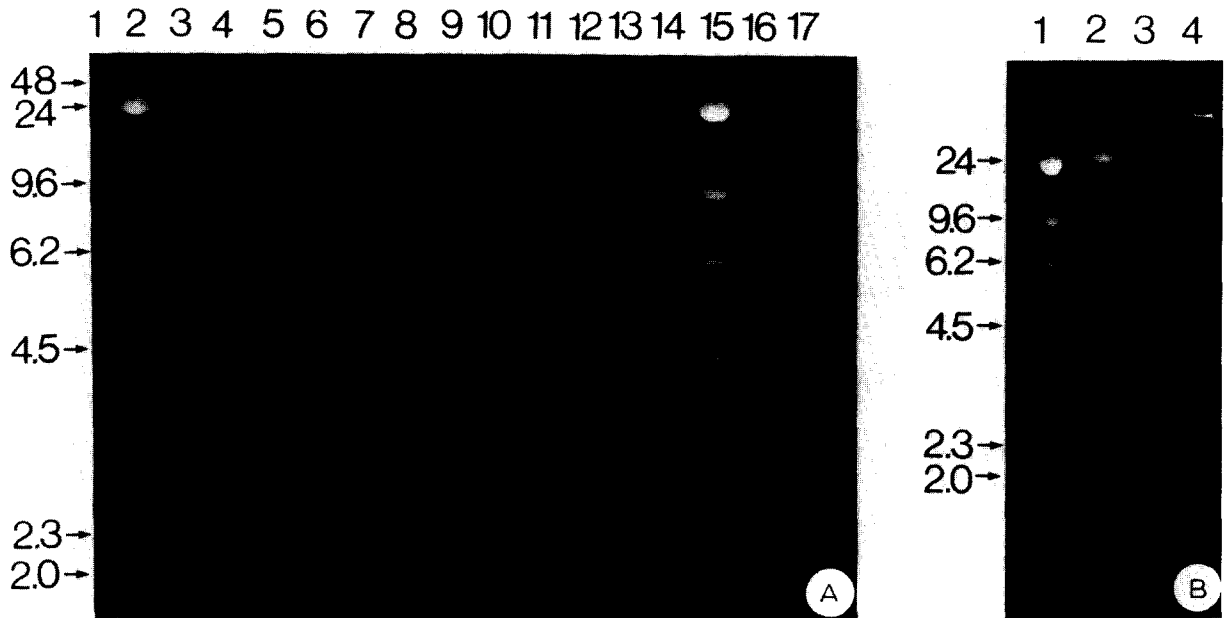


Fig. 2. Analysis of chromosomal and vector DNA. (A) Human DNA was partially digested with restriction endonuclease *Mbo*I and fractionated on 5–20% sucrose gradients. Aliquots of the collected fractions were analysed by agarose gel electrophoresis (see Materials and Methods). Lanes 3–13 and 16, 17 are sucrose gradient fractions, lanes 1 and 14 are wild type bacteriophage λ DNA and lanes 2 and 15 are λ DNA cut with restriction endonuclease *Hind*III (fragment sizes in kb are specified in the left margin). (B) Ligation of chromosomal and vector DNA; lane 2, unligated fractionated chromosomal DNA; lane 3, ligated chromosomal DNA and phosphatase treated cosmid DNA; lane 4, ligated fractionated chromosomal DNA; lane 1, λ DNA *Hind*III marker fragments. The sizes of marker DNA bands in kb are indicated.

In the experiments described in this paper cosmid pJB8 was chosen as vector DNA for the following political and scientific reasons:

(1) pJB8 is classified as a disabled vector in *E. coli* HB101 by the British GMAG;

(2) the single *Bam*HI site is located on a linker between two *Eco*RI sites which enables excision of inserted DNA by *Eco*RI cleavage;

(3) it was the smallest cosmid available (5.2 kb).

100 μ g of pJB8 was digested with *Bam*HI and treated with alkaline phosphatase. The phosphatase treatment of the vector DNA was introduced after testing the various combinations of phosphatase or non-phosphatase treated vector and chromosomal DNA respectively, for cloning efficiency and the size of human DNA insert. Colonies were randomly picked after packaging, transduction and plating of ligated DNA in the following combinations:

(i) Unfractionated, phosphatase-treated chromosomal DNA ligated to vector DNA in a weight ratio of 1 : 2 (it is not possible to give an accurate molar ratio for unfractionated DNA) – the yield was 160 000

colonies per μ g chromosomal DNA.

(ii) Fractionated chromosomal DNA ligated to vector DNA in a molar ratio 1 : 4 – the yield was 60 000 colonies per μ g chromosomal DNA.

(iii) Vector DNA ligated to fractionated, phosphatase-treated chromosomal DNA in a molar ratio 4 : 1 – the yield was 60 000 colonies per μ g chromosomal DNA.

(iv) Fractionated chromosomal DNA ligated to phosphatase-treated vector DNA in a molar ratio of 1 : 4 – the yield was 15 000 colonies per μ g chromosomal DNA.

Ligations of vector DNA alone, with or without phosphatase treatment, gave less than 2000 colonies per μ g DNA. DNA from 5 colonies from each experiment was extracted according to a modification of the method described by Birnboim and Doly (1979), digested with *Eco*RI and analysed by agarose gel electrophoresis. Each DNA preparation contained the 5.2 kb pJB8 vector band. However, only the DNA preparations from combination (iv) contained this band in a 1 : 1 molar ratio relative to the additional

*Eco*RI bands from the inserted DNA (which together comprise approx. 35 kb). Combinations (i)–(iii) contained molecules which had several vectors, in some cases together with more than one human DNA insert. These results imply that the efficiencies obtained for the first three combinations are meaningless in terms of uninterrupted 30–45 kb DNA inserts and that these approaches are impractical to use, since in cases (i) to (iii) a much larger number of colonies would have to be screened to find a colinear insert from a given specific region.

In the standard ligation reaction the vector was present in a 4–5-fold molar excess relative to the chromosomal DNA to minimise self-ligation of the human donor DNA and to optimise the subsequent packaging reaction. Varying the total DNA concentration in the ligation reaction between 200 and 1000 $\mu\text{g}/\text{ml}$ did not affect the efficiency of packaging of the ligated DNA (unpublished results).

Ligated DNA (Fig. 2B) was packaged without further treatment according to the method of V. Pirrotta (pers. commun., see MATERIALS AND METHODS). This yielded up to 50 000 infectious phage particles containing recombinant cosmids per μg chromosomal DNA, depending on the size and quality of the DNAs and the efficiency of the packaging mixes. The packaging mixes usually gave $5 \cdot 10^7$ – $3 \cdot 10^8$ plaques per μg packaged wild-type λ DNA.

(b) Screening of a human cosmid library for the β -globin related genes

Using the procedure described in MATERIALS AND METHODS, a human cosmid library was obtained in *E. coli* 1400, containing approx. 150 000 colonies. The size of the DNA inserts was an average of 35 kb as judged from restriction analysis of 20 isolated clones. The screening of the library was carried out on duplicate sets of filters using probes for the ϵ , γ , δ , and β -globin genes. Two subsequent hybridisations showed in total 12 “positive” spots (one shown in Fig. 3). Of these, 11 spots were present on duplicate filters and one strong spot only on one of the two duplicate filters (this was subsequently shown to be the globin gene cosHG 28). The positive colonies were located on the original plates, using the keyed marking, and then picked up.

Each colony was then resuspended in 0.5 ml L-broth and three serial dilutions were plated on one

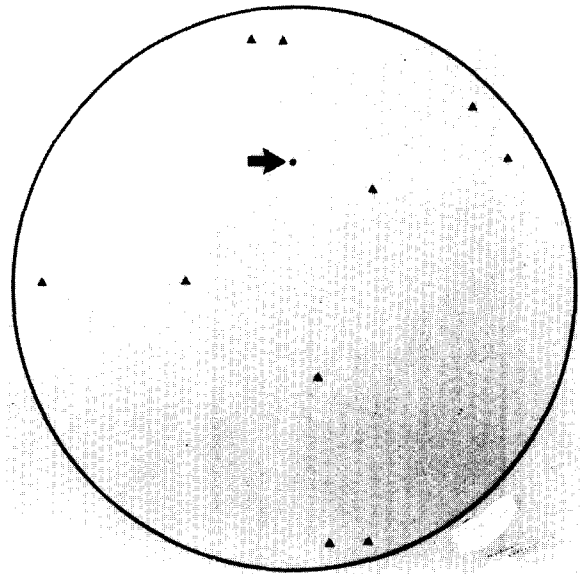


Fig. 3. Screening of the human cosmid library for β -globin-like sequences. Approx. 5000 recombinant DNA colonies on a nitrocellulose filter were lysed, treated and hybridised with a mixture of ϵ and β -globin probes, as described in Materials and Methods. On the filter shown, one positive spot was detected (indicated by the arrow). \blacktriangle indicate keying of the filter.

nitrocellulose filter in three sections before incubations at 30°C. Each filter was replica-plated and duplicate sets of filters incubated, processed and hybridised to single probes as before. Positive clones in the re-screening were then picked from low-colony density areas, grown to mid log phase and stored as glycerol cultures at –70°C. From the original strongly hybridising colonies five were positive in the re-screen. DNA was prepared from five positive clones by a modification of the procedure of Birnboim and Doly (1979; see MATERIALS AND METHODS) and analysed by agarose gel electrophoresis, followed by Southern blotting. To facilitate multiple screening with different hybridisation probes, the gel was soaked in 0.25 M HCl for 40 min followed by alkali treatment and neutralisation and then sequentially blotted onto three nitrocellulose filters (Wahl et al., 1979). This was performed by replacing the first filter with a fresh filter after 15 min of blotting and subsequently replacing filters at 10 min intervals.

The multiple filters were hybridised with probes for the ϵ , γ , and β -globin genes respectively (Fig. 4). In two recombinant DNA clones, cosHG1 and

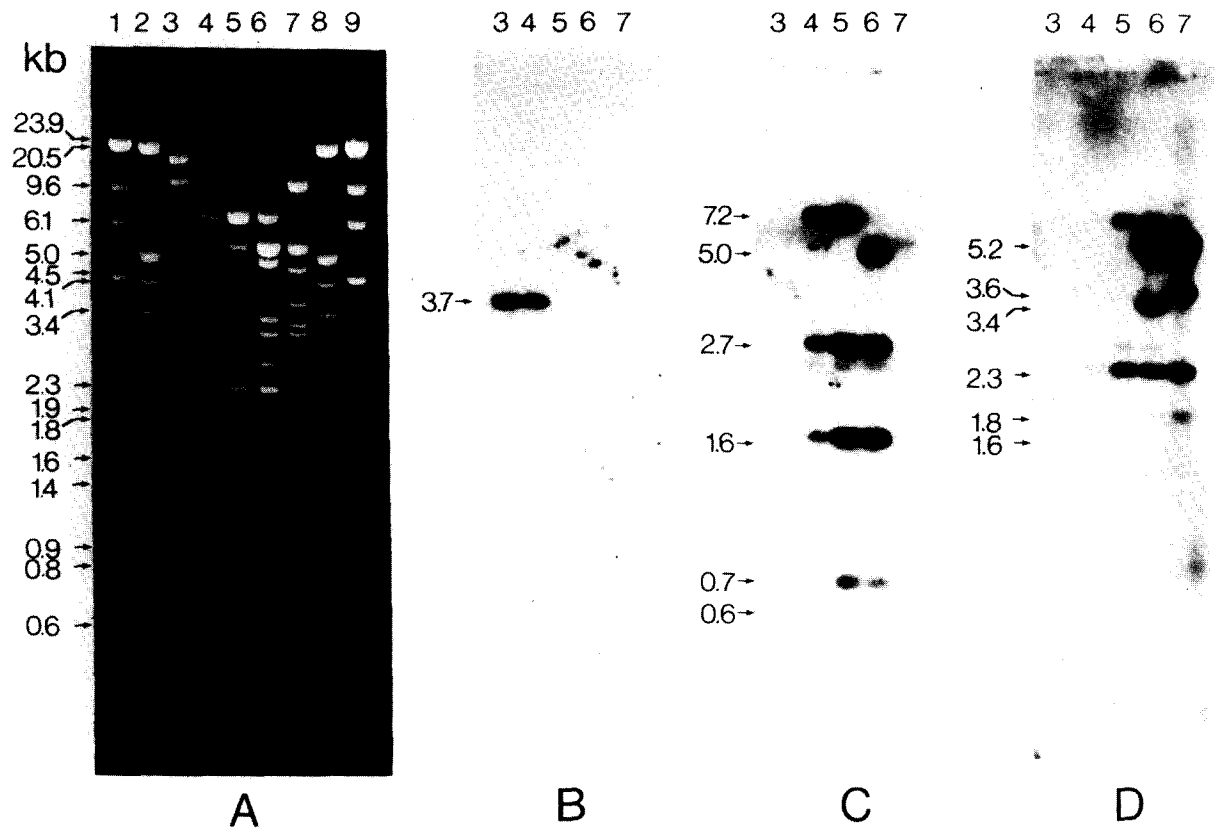


Fig. 4. Analysis of cosmid clones containing human β -globin-related genes. DNA from cosmid clones cosHG1 (lane 3), cosHG3 (lane 4), cosHG30 (lane 5), cosHG28 (lane 6) and cosHG29 (lane 7) were digested to completion with restriction enzyme *EcoRI* and electrophoresed on a 0.6% agarose gel (A). The DNA was transferred to nitrocellulose sheets and hybridised to human ϵ -globin probe (B), γ -globin probe (C), and β -globin probe (D). Clone cosHG30 contains the human γ and δ genes and is contained in cosHG3. Lanes 1 and 9 contain λ DNA digested with *HindIII*, and lanes 2 and 8 λ DNA digested with *HindIII* + *EcoRI*. DNA fragment sizes in kb are indicated.

cosHG3, a 3.7 kb *EcoRI* fragment hybridises with the ϵ -globin specific probe. Two DNA clones contain several fragments which hybridise with the γ -globin specific probe: cosHG3 contains a 7.2 kb, 1.6 kb (both from the γ G-globin gene), 2.7 kb and a 0.6 kb (both from the γ A-globin gene) fragment; cosHG28 shows an identical pattern with the exception that the 7.2 kb 5' γ G fragment is replaced by a 4.4 kb fragment which is the terminal 5'-*EcoRI* fragment of the DNA insert, and which contains the 3' regions of the 7.2 kb fragment. Three recombinants hybridise with the probe for the β -globin gene (which also detects the δ -globin gene (Flavell et al., 1978): cosHG3 contains a 5' (2.3 kb) and 3' (1.6 kb – the 3' end of the human DNA insert) fragments of the δ -globin gene. cosHG29 and cosHG28 both contain the 5' (2.3 kb) and 3' (1.75

kb) fragments of the δ -globin gene, together with the 5' (5.2 kb) and 3' (3.5 and 3.2 kb) fragments of the β -globin gene, respectively, the latter being the 3'-terminal fragment of the clone which terminates in the 3.5 kb, 3'-*EcoRI* fragment of the β -globin gene. This insert found in the fifth recombinant, cosHG30, is contained within cosHG3. Comparison of the *EcoRI* fragments of all the cloned fragments with additional standard restriction enzyme mapping data (obtained by double digestion using all the combinations of the enzymes indicated in Fig. 5; not shown) and the blot-hybridisation data (Fig. 4) result in the map shown in Fig. 5. This map agrees with the data previously obtained by genomic blotting (e.g. Bernards et al., 1979; Fritsch et al., 1979) and cloning in phage λ (Fritsch et al., 1980; Slightom et al., 1980; Kaufman et al., 1980), with the exception of the 5'

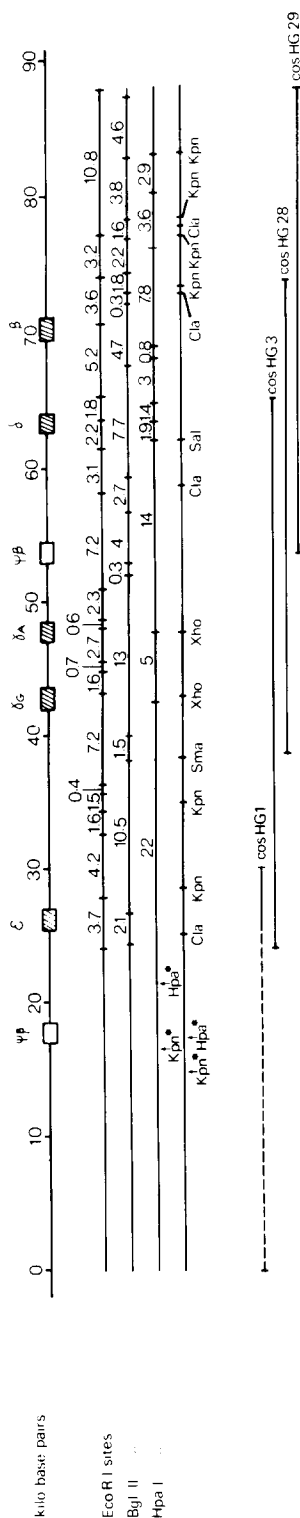


Fig. 5. Physical map of restriction sites in human globin DNA and location of cosmid-cloned DNAs containing the human β -globin-related genes. Hatched boxes indicate the position of the β -globin-related genes, open boxes indicate the position of pseudo genes. The KpnI and HpaI sites marked with an asterisk indicate the difference between genomic DNA (bottom line) and cosHG1 DNA (top line).

part of cosHG1. To find out whether this recombinant actually represented the 5' region of the ϵ gene, it was compared with genomic DNA by blotting procedures. Either human placental DNA (the same DNA as was used for the library), of a mixture of placental and cosmid DNA was digested with *KpnI*, *KpnI* + *HpaI*, *KpnI* + *AvaI*, or *AvaI*, run on a 0.5% agarose gel and hybridised with a nick-translated ϵ gene probe (Fig. 6). The *KpnI* fragment seen with genomic DNA is about 15 kb while the cloned fragment is only about 13 kb. Moreover, cleavage of this fragment with *HpaI* gives a 9.6 kb fragment for the genomic DNA, but only a 6.5 kb fragment for the cosmid. This suggests that the *KpnI* and *HpaI* sites to the 5' side of the ϵ gene in cosHG1 are not located at the same sites as in the genomic DNA. The *AvaI* site to the 5' side of the ϵ gene is, however, localised at the same position, since the *KpnI* + *AvaI* fragment (which extends from the *KpnI* site to the 3' side of the ϵ gene) is seen in both the genomic and cloned DNA. We therefore believe that this recombinant is either rearranged or, alternatively, consists of two segments of human DNA that are not contiguous in the genome and that have been joined during the ligation step.

DISCUSSION

The library described in this article contains approx. 150 000 recombinant colonies. The average size of human DNA inserts in the recombinants is 35 kb, which is 5–10 kb shorter than expected when comparing the size of recombinant cosmids with that of wild-type λ DNA. With a human genome size estimated to be $3 \cdot 10^9$ bp, 85 000 recombinants would be required for a "complete" library. The region which we have screened for specific sequences is about 45 kb (from the 5' end of the human ϵ -globin gene to the 3' end of the β -globin gene (e.g. Maniatis et al., 1980). Therefore, two recombinants containing DNA from inside the 45 kb region and two recombinants containing only either the 3' or 5' part of this region could be expected. Three clones (namely cosHG3, 28 and 30) and two clones (cosHG1 and 29) respectively were found, indicating that the frequency of clones for at least the globin gene region is that expected for a random library

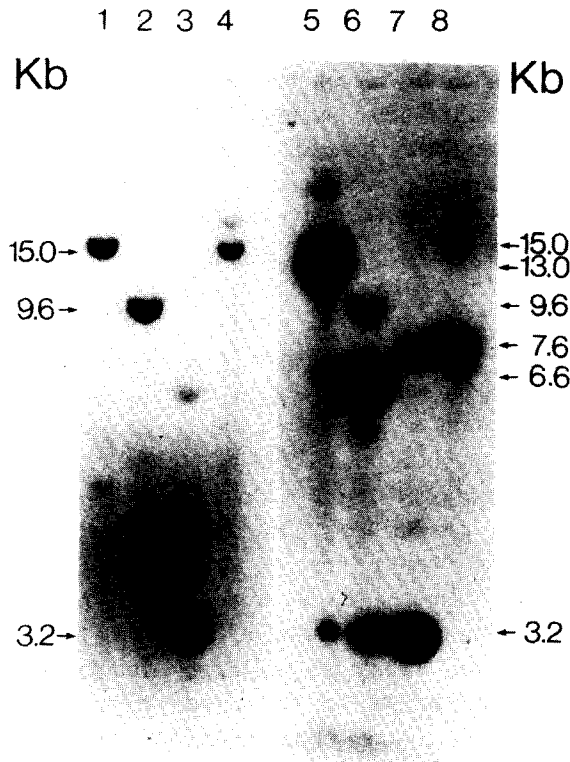


Fig. 6. Comparison of cosHG1 and chromosomal DNA by blotting and hybridisation to an ϵ gene probe. Human placental DNA (lanes 1–4) or human placental DNA plus cosHG1 DNA (lanes 5–8) were digested with *Kpn*I (lanes 1 and 5), *Kpn*I + *Hpa*I (lanes 2 and 6), *Kpn*I + *Ava*I (lanes 3 and 7) and *Ava*I (lanes 4 and 8), blotted and hybridised with a 1.3 kb *Bam*HI-*Eco*RI fragment containing the 3' end of the ϵ -globin gene. Fragment sizes are indicated in kb.

(assuming that all β -globin-related recombinants in the library were found). Moreover, we have also subsequently used the library to isolate human collagen and interferon clones (unpublished data).

The recombinants isolated from this library all proved to be stable on replating and also after growing large volume cultures, with the exception of cosHG3. This recombinant was once lost during propagation due to a deletion in the γ globin region, but could be recovered by repackaging of the original recombinant DNA followed by transduction into *E. coli* 1400. It has subsequently remained stable during propagation. It seems that both the size fractionation and phosphatase treatment are important to obtain a library containing single contiguous 35–40 kb inserts. The physical maps of the

cloned DNAs were compared with that of the human β -globin-related gene locus by Southern blotting and one of the clones, cosHG1, was found not to be collinear with that of the genome. The most likely explanation for this is that some human DNA fragments – one of which contained β -globin related sequences – were ligated into a packageable cosmid. The omission of the size fractionation of the donor DNA increases the risk that non-contiguous DNA fragments will be ligated together with the cosmid vector. This could be minimised by using an even higher vector to chromosomal DNA ratio, however, since the total amount of DNA increases, the disadvantage of such a procedure is that it would require more packaging materials.

When the cosmid DNA is not phosphatase-treated our results suggest that recombinant DNAs containing multiple vector molecules are often formed, packaged and stable in *E. coli*. Recent results (D. Ish-Horowicz, personal communication) indicate that some of these problems can be solved when fragments of cosmid DNA incapable of forming multimers are used in the ligation reaction.

In summary, using the procedures described here, it is possible to construct cosmid libraries and to isolate stable cosmids containing inserts of about 35 kb of genomic DNA. Recombinant cosmid DNA can be obtained in high yield (usually about 600 μ g DNA per liter of culture) using standard propagation methods, and since the cosmid DNA is in itself a substrate for in vitro packaging, it is possible to transfer the DNA to any *E. coli* host that carries phage λ or ϕ 80 receptors.

ACKNOWLEDGEMENTS

We thank Drs. D. Ish-Horowicz and J. Burck for the cosmid vector pJB8 and Dr. T. Baralle for a human ϵ -globin recombinant. We are also grateful to Dr. J. Collins for his advice and hospitality and to EMBO for short-term fellowships to F.G.G. and R.A.F. to permit work in Dr. Collins' laboratory. We thank C. O'Carroll and E. Heather for typing the manuscript. F.G.G. was a SON and Royal Society/ZWO fellowship holder. H.H.M.D. is an EMBO long-term fellowship holder. This work was supported by the British Medical Research Council (M.R.C.).

REFERENCES

- Bernards, R. and Flavell, R.A.: Physical mapping of the globin gene deletion in hereditary persistence of foetal haemoglobin (HPFH). *Nucl. Acids Res.*, 8 (1980) 1521–1534.
- Bernards, R., Kooter, J.M. and Flavell, R.A.: Physical mapping of the globin gene deletion in ($\delta\beta$)^o-thalassaemia. *Gene* 6 (1979) 265–280.
- Birnboim, H.C. and Doly, J.: A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids Res.* 7 (1979) 1513–1523.
- Boyer, H.W. and Rouland-Dussoix, D.: A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *J. Mol. Biol.* 41 (1969) 459–472.
- Cami, B. and Kourilsky, P.: Screening of cloned recombinant DNA in bacteria by in situ colony hybridization. *Nucl. Acids Res.* 5 (1978) 2381–2390.
- Collins, J. and Hohn, B.: Cosmids: A type of plasmid gene-cloning vector that is packageable in vitro in bacteriophage λ heads. *Proc. Natl. Acad. Sci. USA* 75 (1978) 4242–4246.
- Feiss, M., Fisher, R.A., Crayton, M.A. and Egener, C.: Packaging of the bacteriophage λ chromosome: Effect of chromosome length. *Virology* 77 (1977) 281–293.
- Flavell, R.A., Kooter, J.M., De Boer, E., Little, P.F.R. and Williamson, R.: Analysis of the β - δ -globin gene loci in normal and Hb Lepore DNA: Direct determination of gene linkage and intergene distance. *Cell* 15 (1978) 25–41.
- Fritsch, E.F., Lawn, R.M. and Maniatis, T.: Characterization of deletions which affect the expression of fetal globin genes in man. *Nature* 297 (1979) 598–603.
- Fritsch, E.F., Lawn, R.M. and Maniatis, T.: Molecular cloning and characterization of the human β -like globin gene cluster. *Cell* 19 (1980) 959–972.
- Hanahan, D. and Meselson, M.: Plasmid screening at high colony density. *Gene* 10 (1980) 63–67.
- Hohn, B.: In vitro packaging of λ and cosmid DNA, in Wu, R. (Ed.), *Methods in Enzymology*, Vol. 68, Academic Press, New York, 1980, pp. 299–308.
- Kaufman, R.E., Kretschmer, P.J., Adams, J.W., Coon, H.C., Anderson, W.F. and Nienhuis, A.W.: Cloning and characterization of DNA sequences surrounding the human γ , δ and β -globin genes. *Proc. Natl. Acad. Sci. USA* 77 (1980) 4229–4233.
- Kushner, S.R.: An improved method for transformation of *Escherichia coli* with ColEI derived plasmids, in Boyer, H.W. and Nicosia, S. (Eds.), *Genetic Engineering*, Elsevier/North-Holland, Amsterdam, 1978, pp. 17–23.
- Maniatis, T., Fritsch, E.F., Lauer, J. and Lawn, R.M.: The molecular genetics of human hemoglobins. *Annu. Rev. Genet.* 14 (1980) 145–178.
- Morrow, J.F.: Recombinant DNA techniques, in Wu, R. (Ed.), *Methods in Enzymology*, Vol. 68, Academic Press, New York, 1980, pp. 3–24.
- Roberts, R.J.: Restriction and modification enzymes and their recognition sequences. *Gene*, 8 (1980) 329–343.
- Royal, A., Garapin, A., Cami, B., Perrin, F., Mandel, J.L., Le Meur, M., Bregegere, F., Gannon, F., Le Pennec, J.P., Chambon, P. and Kourilsky, P.: The ovalbumin gene region: common features in the organization of three genes expressed in chicken oviduct under hormonal control. *Nature*, 279 (1979) 125–132.
- Slightom, S.L., Blechl, A.E. and Smithies, O.: Human fetal $G\gamma$ - and $A\gamma$ -globin genes: Complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* 21 (1980) 627–638.
- Wahl, G.M., Stern, M. and Stark, G.R.: Efficient transfer of large DNA fragments from agarose gels to diazobenzyl-oxymethyl-paper and rapid hybridization using dextran sulphate. *Proc. Natl. Acad. Sci. USA* 76 (1979) 3683–3687.

Communicated by W. Szybalski.