Volume 10 Number 8 1982 Nucleic Acids Research

The construction of cosmid libraries of eukaryotic DNA using the Homer series of vectors

William Chia[1], Michael R.D.Scott[2] and Peter W.J.Rigby

Cancer Research Campaign, Eukaryotic Molecular Genetics Research Group, Department of Biochemistry, Imperial College of Science and Technology, London SW7 2AZ, UK

Received 9 March 1982; Accepted 6 April 1982

ABSTRACT

We have constructed and characterised a series of approved, disabled cosmid vectors which we call Homer cosmids and have examined the optimal conditions for the construction of libraries of eukaryotic DNA segments using these vectors. Analysis of these libraries shows that most of the sequences we have tested for are present at the expected frequency and that the libraries can be stably propagated. We have also directly tested the stability of cosmid clones carrying tandemly repeated inserts. This work shows that it should be possible to clone most eukaryotic genes using cosmid vectors and that such cloning systems have considerable advantages over those more commonly used.

INTRODUCTION

The use of molecular cloning techniques to isolate segments of eukaryotic chromosomal DNA has been central to recent major advances in our understanding of gene organisation and expression. Most of this work has depended on the use of cloning vectors based on bacterial plasmids or on the bacteriophage λ but such systems suffer from a number of disadvantages. It is desirable to be able to isolate any gene within one recombinant molecule, but the presence of multiple introns makes many eukaryotic genes very large, for example the chicken proα2 collagen gene (1) and the mouse dihydrofolate reductase gene (2) are approximately 40Kb in length. While plasmid vectors can theoretically accomodate inserts of such a size, the transformation efficiency of large plasmids is extremely low (3) and thus plasmids containing large inserts are likely to be under-represented in plasmid libraries. λ vectors have a maximum capacity defined by the size of DNA that can be packaged into a phage particle; this is 15Kb to 23Kb for the commonly used systems (4).

Collins and Hohn have described novel vectors, called cosmids, in which the λ sequences required for the packaging of DNA into phage particles are incorporated into a plasmid (5). λ heads will package up to 53Kb of DNA (4)

0305-1048/82/1008-2503$ 2.00/0

and thus a small cosmid vector should accept fragments 45Kb to 50Kb in length, thereby obviating the problems associated with the cloning of large genes. They should similarly facilitate the isolation of linked genes within one recombinant molecule and of individual genes together with large segments of flanking DNA. These capabilities are likely to be of great importance in studies in which the cloned DNA is to be introduced into cultured cells in order to study the control of gene expression. The use of cosmid vectors also provides other advantages. The large insert size means that the number of clones required to represent a complete eukaryotic genome is significantly reduced. Moreover, certain eukaryotic sequences, particularly those containing tandem duplications, are unstable in λ vectors (6, 7) while other types of sequence have been shown to undergo severe rearrangement during the propagation of λ-derived recombinants (8, 9, 10). Cosmids might overcome such problems, particularly as they can readily be propagated in *recA*$^-$ hosts unlike many of the commonly used λ vectors (4).

Despite these potential advantages cosmid vectors have not been widely used. One major reason for this has been the ease with which λ libraries can be screened but the development of methods for screening large numbers of colonies (11) has obviated this problem. Moreover, many of the earlier cosmid vectors were rather large (12, 13) and thus did not capitalise on the potentially large insert size and none of the previously described cosmids has been certified as being biologically disabled.

We describe here the construction of a small, 5.44Kb, cosmid vector, HomerI, which is based on the plasmid pAT153 (ref. 14). This vector, and derivatives thereof, have been approved by the United Kingdom Genetic Manipulation Advisory Group as a disabled (EK2-equivalent) system when used with any *recA*$^-$ strain of *E. coli* K12. We discuss the optimal conditions for the use of this vector in the construction of libraries of eukaryotic DNA fragments and present data on the sequence representation in these libraries. We also consider the stability of tandemly repeated sequences cloned in cosmids.

MATERIALS AND METHODS

Nucleic Acids. pAT153 DNA was isolated from saturated, unamplified cultures of *E. coli* HB101 using the procedure of Wensink *et al.* (15) as were the DNAs of Homer cosmids and recombinants derived therefrom. For small scale preparations the cells were lysed in the same way; the cleared lysates were extracted with phenol, then with chloroform and the DNA was precipitated and

washed with 70% (v/v) ethanol before being resuspended in an appropriate buffer.

Charon 4 and λC_{I857} Sam_7 DNAs were prepared by proteinase K digestion and phenol extraction of virions purified by caesium chloride density gradient centrifugation. Charon 4 was propagated by the PDS method of Blattner *et al.* (16), λC_{I857} Sam_7 was prepared from cultures of induced lysogens. SV40 DNA was isolated from productively-infected CV1 cells as described by Shenk *et al.* (17). Adenovirus type 2 DNA was the gift of Val Lewis, while *Drosophila melanogaster* (OregonR) embryonic DNA was generously provided by Diane de Cicco.

DNAs were generally labeled *in vitro* using the nick translation procedures of Rigby *et al.* (18). Charon 4 DNA was labeled at its termini using a saturating amount of terminal transferase in 10mM potassium cacodylate, pH7.5, 0.1mM DTT, 20μg/ml BSA, 8mM $CoCl_2$ and 20μM [α-^{32}P]TTP.

Restriction endonuclease digests of genomic DNA were fractionated by sucrose density gradient centrifugation as described by Maniatis *et al.* (19). Adenovirus type 2 (35Kb) and Charon 4 (45Kb) DNAs were run in parallel gradients as size markers. Appropriate fractions of the genomic DNA were then further sized by agarose gel electrophoresis.

The cohesive termini of restriction enzyme fragments were filled in by incubating the DNA with reverse transcriptase (500 units/ml) in 50mM Tris HCl, pH8, 25mM NaCl, 10mM $MgCl_2$, 10mM DTT and 1mM dATP, dCTP, dGTP and TTP at 37°C for 30 min.

D. melanogaster rRNA was the gift of Simon Kidd. It was partially hydrolysed by incubation in 0.1N NaOH for 1h. at 0°C and then labeled using T4 polynucleotide kinase and [γ-^{32}P]ATP as described by Maniatis *et al.* (19).

Enzymes. Restriction endonucleases were purified and given to us by Lesley Woods or were purchased from Uniscience Ltd., Cambridge, England. T4 DNA ligase was purified from *E. coli* lysogenic for the recombinant λ phage NM989 using the published procedure (20) and was the gift of Lesley Woods. Calf intestinal phosphatase was purchased from Boehringer Mannheim and was further purified as described by Efstratiadis *et al.* (21). *Eco*RI methylase was purified from *E. coli* RY13 using the procedure of Greene *et al.* (22). Calf thymus terminal transferase was the generous gift of R.L. Ratliff, Los Alamos, New Mexico, U.S.A. AMV reverse transcriptase was obtained from J. Beard, Life Sciences Inc., St. Petersburg, Florida, U.S.A.

*Eco*RI* digestions were performed in the presence of dimethyl sulphoxide (23). The reaction conditions were: 25mM Tris HCl, pH7.7, 10mM NaCl, 6mM

$MgCl_2$, 2mM 2-mercaptoethanol, 0.04% (v/v) Triton X100, 0.04% (w/v) gelatin and 6% (v/v) dimethyl sulphoxide.

Introduction of Recombinant DNA Molecules into *E. coli*. Recombinant plasmids were introduced into *E. coli* HB101 using a transformation protocol derived from published procedures (24, 25). Cosmid DNA was packaged into phage particles *in vitro* using a procedure originally developed by V. Pirotta and communicated to us by D. Ish-Horowicz (26). After packaging, between 100µl and 500µl of phage buffer (10mM Tris HCl, pH7.4, 100mM NaCl, 10mM $MgCl_2$, 0.1% (w/v) gelatin) together with 2 or 3 drops of $CHCl_3$ were added to the packaging reaction mixtures (25µl to 125µl) which were then shaken vigorously. Debris was removed by centrifugation in a microfuge and the supernatant was used directly for infection. We have not found it necessary to purify the packaged phage by caesium chloride equilibrium density gradient centrifugation. Cells from a saturated, overnight culture of *E. coli* HB101 grown in L broth supplemented with 10mM $MgSO_4$ and 0.4% (w/v) maltose were used for infection. Up to an equal volume of phage buffer containing the *in vitro* packaged transducing particles was added to the cells. After a 10 min. incubation at room temperature between 50µl and 500µl of the infection mixture was spread onto a 82mm nitrocellulose filter laid on top of a plate of L-agar containing 100µg/ml ampicillin. If a fresh overnight culture is used it is not necessary to activate the expression of ampicillin resistance prior to selection.

Gel Electrophoresis and Transfer Hybridisation. Agarose gels, containing 1µg/ml of ethidium bromide, were cast and run in either 89mM Tris, 89mM H_3BO_3, 2.5mM EDTA, pH8.2 or 40mM Tris, 50mM sodium acetate, 1mM EDTA adjusted to pH7.8 with acetic acid. Electrophoresis was normally at 1.5v/cm overnight. DNA was transferred from agarose gels to nitrocellulose filters by the procedure of Southern (27). The filters were preincubated, hybridised and washed as described by Jeffreys and Flavell (28) and then autoradiographed at -70°C using Fuji RX X-ray film and Fuji Mach II intensifying screens (29). The sizes of DNA fragments were determined relative to suitable pBR322, SV40 and λ DNA standards.

The purification of DNA fragments or of recombinant cosmids by gel electrophoresis was performed using low gelling temperature agarose (Miles Laboratories). Gel slices containing the DNA of interest were liquefied by incubation at 67°C for 4 min. This solution was extracted several times with phenol at 37°C and then the DNA was precipitated from the aqueous phase with ethanol, resuspended, re-extracted with phenol and then

reprecipitated with ethanol.

Electron Microscopy. Supercoiled cosmid DNA was nicked by X-irradiation and then spread from aqueous solution using the methods of Davis *et al.* (30). The grids were rotary shadowed with platinum/palladium and examined in a Hitachi H600 electron microscope. 35mm photographic negatives were measured by projecting them onto a Hewlett-Packard 9864A digitiser in line with a 9821A calculator. Open circular SV40 DNA (5.243Kb) was used as length standard.

RESULTS

(i) Construction and Characterisation of Homer Cosmids.

As the basis for our cosmid we chose the plasmid pAT153 because it is small, 3.66Kb, and has been approved by the U.K. G.M.A.G. as a disabled system in conjunction with any *recA*$^-$ strain of *E. coli* K12. pAT153 is derived from pBR322 (ref. 31) by the deletion of two *Hae*II fragments (14; our unpublished data), from nucleotide 1646 to nucleotide 2351 in the pBR322 sequence (32). The λ *cos* site was introduced into pAT153 in the following way. The cohesive ends of Charon 4 DNA were annealed and the circularised DNA was digested with *Bgl*II; because the right end of Charon 4 is derived from Ø80 the *cos* site is located within a 1.78Kb fragment (4). pAT153 was digested with *Bst*I and the resultant linear plasmid molecules were ligated to the *Bgl*II fragments of Charon 4 DNA; the ligation mixture was then digested with *Bst*I and treated with calf intestinal phosphatase. *Bgl*II fragments can be ligated into the *Bst*I site of pAT153 because the recognition sequences of these enzymes have the same central tetranucleotide; however, hybrid sites are refactory to both enzymes. *Bst*I digestion following ligation will thus cleave recircularised vector molecules and those recombinants in which the Charon 4 insert contains a *Bst*I site, which the *cos* fragment does not. The phosphatase treatment prevents recircularisation of *Bst*I cleaved molecules and thus enriches for the desired recombinants. The DNA was then introduced into *E. coli* HB101 by transformation and ampicillin-resistant colonies were selected; all were tetracycline-sensitive, consistent with their having an insert at the *Bst*I site, which is within the tetracycline resistance gene. These transformants were then screened for the presence of λ *cos* sequences using Charon 4 DNA which had been labeled at its termini by tailing with terminal transferase and [α-^{32}P]TTP and then digested with *Bgl*II; this probe should detect only the *Bgl*II fragment containing the *cos* site. This screen was equivocal,

presumably because of the presence of cryptic lambdoid prophages within the *E. coli* genome (33). The colony giving the strongest hybridisation signal was picked, grown and plasmid DNA was isolated. Restriction endonuclease mapping showed the presence of the 1.78Kb fragment of Charon 4 DNA inserted into the *Bst*I site of pAT153. The orientation of the insert was determined by mapping the cleavage sites for *Bgl*I. There are three such sites in pAT153 (ref. 32) and one site located 424 base pairs from the left end of λ DNA (F. Sanger, personal communication). The sizes of the *Bgl*I fragments of this cosmid, HomerI, showed it to have the structure diagrammed in Figure 1.

We have constructed several derivatives of this basic cosmid. HomerII is suitable for cloning *Sst*I fragments. λC_{I857} Sam_7 DNA was digested to completion with *Eco*RI and *Hin*dIII; the 0.95Kb *Eco*RI/*Hin*dIII fragment, which contains an *Sst*I site, was purified by agarose gel electrophoresis. HomerI was digested with *Eco*RI and *Hin*dIII and the larger fragment was similarly purified. The two fragments were ligated and transformed into *E. coli* HB101

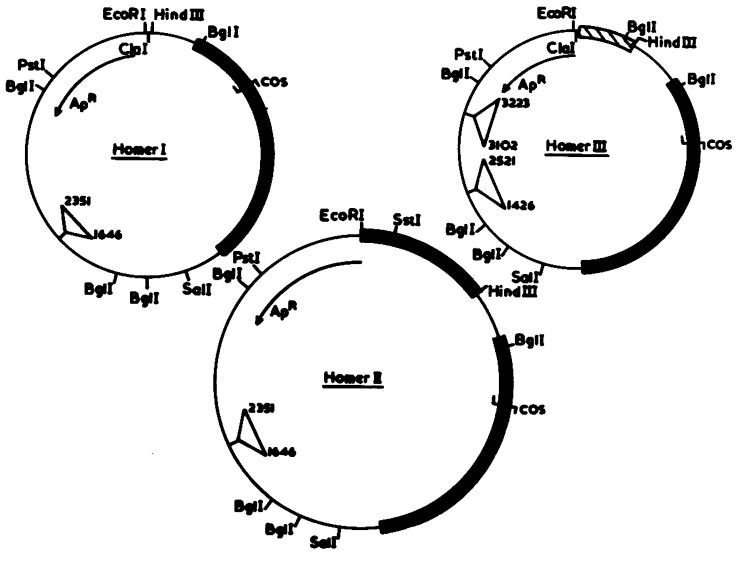

Figure 1. Structures of Homer Cosmids. Thin lines indicate segments derived from pBR322, solid sections segments derived from Charon 4 and the cross-hatched section a segment derived from SV40. Ap^R denotes the location of the ampicillin resistance gene of pBR322, *cos* the location of the λ *cos* site. The numbers show the limits of the deletions in the pBR322 sequence involved in the construction of pAT153 and pXf3. The three maps are to scale. HomerI is 5.44Kb, HomerII is 6.38Kb and HomerIII is 5.34Kb.

selecting for ampicillin resistance. Plasmid DNA isolated from three colonies was examined by restriction endonuclease mapping which showed that a new cosmid, HomerII, had been produced; its structure is shown in Figure 1.

HomerIII is based on the plasmid pXf3 (D. Hanahan, personal communication) rather than on pAT153. pXf3 is also derived from pBR322 but contains two deletions, extending from nucleotide 1426 to nucleotide 2521 and from nucleotide 3102 to nucleotide 3223 (D. Hanahan, personal communication; our unpublished results). The advantage of using pXf3 is that this plasmid lacks the poison sequence which blocks the efficient replication of pBR322 DNA in animal cells (34). HomerIII also contains a 0.415Kb segment of SV40 DNA, from nucleotide 5171 to nucleotide 346 on the BB system of numbering the viral genome (35). This segment contains the viral origin of DNA replication (35) and the 72 base pair repeats which are essential for the *in vivo* transcription of viral early genes (36) and which potentiate the expression of adjacent segments of DNA (37). HomerIII should thus be of great utility for experiments in which the final objective is the re-introduction of the cloned segments into cultured animal cells. It was constructed in the following way which was designed to preserve the single *Cla*I site of pAT153. SV40 DNA was digested to completion with *Msp*I and the cohesive ends of the resultant linear molecules were filled in using reverse transcriptase. This DNA was then digested with *Hin*dIII and the 0.415Kb *Hin*dIII/*Msp*I fragment was purified by agarose gel electrophoresis. pXf3 DNA was digested to completion with *Hin*dIII. 80% of the digested DNA was incubated with reverse transcriptase to fill in the cohesive termini and then mixed with the remaining 20%. The mixture was denatured and reannealed to generate a population of DNA molecules some of which have both *Hin*dIII termini filled in while some have one cohesive terminus and one filled-in terminus. This DNA was ligated to the *Hin*dIII/*Msp*I fragment of SV40 DNA and the resultant hybrid molecules were transformed into *E. coli* HB101. Ampicillin-resistant colonies were screened by hybridisation with ^{32}P-labeled SV40 DNA and positive-scoring clones were examined by restriction endonuclease digestion to determine the orientation of the SV40 insert. In pSO59 the SV40 *Bgl*I site is proximal to the tetracycline resistance gene of pXf3. pSO59 was digested with *Hin*dIII and *Sal*I and the larger, 2.94Kb, fragment containing the vector origin, the ampicillin resistance gene and the SV40 segment was purified by agarose gel electrophoresis. HomerI was also digested with *Hin*dIII and *Sal*I and the smaller, 2.40Kb, fragment containing the λ *cos* site was similarly purified. The two fragments were

then ligated, transformed into *E. coli* HB101 and ampicillin-resistant colonies were selected. Restriction endonuclease digestion of DNA prepared from such clones showed that they contain the desired cosmid, HomerIII, the structure of which is shown in Figure 1.

(ii) The Construction of Recombinant DNA Libraries.

The utility of HomerI as a cloning vector was demonstrated as follows. *D. melanogaster* DNA, partially digested with *Eco*RI to a mean length of 30Kb, was ligated to *Eco*RI-cleaved HomerI DNA at a total DNA concentration of 300μg/ml and a molar vector:target ratio of 8:1. The resultant concatameric DNA molecules were packaged *in vitro*, using the system of Hohn and Murray (38), and the phage particles produced were used to transduce *E. coli* HB101 with selection for ampicillin resistance. Cosmid DNA isolated from individual colonies was analysed by *Eco*RI digestion and agarose gel electrophoresis. Recombinants containing large inserts (average length 25Kb) were obtained but these generally contained more than one vector molecule as evidenced by the disproportionately intense vector band. This result is due to the fact that *cos* sites are not cleaved at 100% efficiency (12). Furthermore, a significant proportion of the clones (10% to 40% in various experiments) contained only monomeric HomerI DNA. These presumably result from the packaging of concatameric vector molecules which are unstable even in a *rec*A$^-$ host.

We next explored the effect of varying the vector:target ratio. In this and all subsequent work *in vitro* packaging was done as described in 'Materials and Methods'. *D. melanogaster* DNA was partially digested with *Eco*RI and fractionated by sucrose density gradient centrifugation to give a preparation with a mean size of 40Kb. This DNA was ligated to *Eco*RI-cleaved HomerI at a total DNA concentration of 300μg/ml and at vector:target ratios varying from 1:3 to 9:1. The concatameric DNA molecules were then packaged and transduced into *E. coli*. The results in Table 1 show that the cloning efficiency increases as the vector:target ratio is increased. However, restriction enzyme analysis of DNA isolated from randomly picked ampicillin-resistant colonies indicated that under conditions which give optimal efficiency the problems of high vector background and of recombinants containing more than one vector molecule still occur.

These problems can be overcome by using dephosphorylated vector DNA. Since such molecules can not concatamerise, vector background can be eliminated even when using the high vector:target ratios required for maximal cloning efficiency. Furthermore, since each target DNA terminus can

Table 1

The Effect of Vector to Target Ratio on Cloning Efficiency

Molar Ratio of Vector:Target	µg/ml HomerI	µg/ml 40Kb *Drosophila* DNA	Colonies per µg HomerI	Colonies per µg *Drosophila* DNA
1:3	12	288	3.0×10^4	1.3×10^3
1:1	33	267	7.5×10^4	9.3×10^3
3:1	82	218	1.2×10^5	4.3×10^4
9:1	159	141	6.1×10^4	6.9×10^4
Vector alone	159	-	2.2×10^3	-

be ligated to only one vector molecule, if carefully size selected target DNA is used each recombinant obtained should contain only one target fragment and one vector molecule. Partial *Eco*RI digestion is not a good way of generating random fragments of target DNA. The use of an enzyme recognising a tetranucleotide sequence is desirable and we have therefore used *Eco*RI* digestion to generate the target fragments. In this way they can be inserted into the *Eco*RI site of HomerI and the experiments are thus directly comparable to those described above.

D. melanogaster DNA was methylated with *Eco*RI methylase to protect *Eco*RI cleavage sites and then partially digested with *Eco*RI* to give a mean size of 40Kb. The methylation step is necessary because under the *Eco*RI* conditions we have used *Eco*RI site cleavage is complete before there is appreciable *Eco*RI* digestion. The digested DNA was fractionated by sucrose density gradient centrifugation and the 35Kb to 50Kb fraction was ligated, at a total DNA concentration of 300µg/ml and a vector:target ratio of 10:1 to *Eco*RI-digested, dephosphorylated HomerI DNA. Following *in vitro* packaging and transduction, ampicillin-resistant colonies were selected and stored on nitrocellulose filters (11).

The cloning efficiency was 4×10^4 cfu/µg target DNA. Cosmid DNA from 18 independent clones was prepared and analysed by *Eco*RI digestion and agarose gel electrophoresis (Figure 2). In 13 cases only a single vector fragment, larger than HomerI, was detected; thus each recombinant molecule contains only a single copy of the vector. The remaining 5 colonies yielded monomeric HomerI, implying that *Eco*RI sites were being reconstructed at a higher than expected efficiency. *Eco*RI* may cleave preferentially at GAATTX and XAATTC, rather than cleaving all XAATTX sequences with equal proficiency.

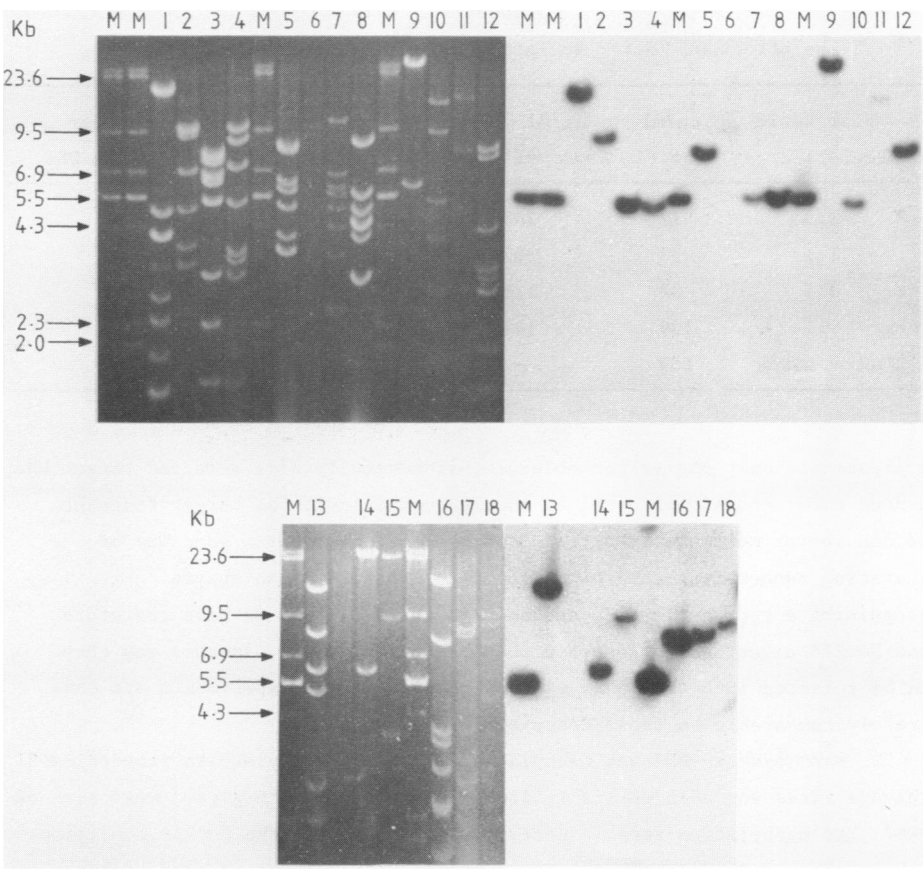

Figure 2. Characterisation of Cosmids from the *Eco*RI* *D. melanogaster* Library. Cosmid DNA was isolated from 18 randomly picked, ampicillin-resistant colonies, digested with *Eco*RI and analysed by agarose gel electrophoresis and transfer hybridization using ^{32}P-labeled HomerI as probe. Both the ethidium bromide stained gel and the autoradiogram are shown. Tracks labeled M contain DNA size markers, the tracks labeled 1 to 18 contain the recombinant cosmids. HomerI is 5.44Kb.

In all 18 clones an insert is present; indeed by examining a pool of plasmid DNA prepared from 15,000 clones we could not detect any monomeric HomerI DNA. Vector background has thus been eliminated.

DNA from the pool of 15,000 recombinants was analysed by electron microscopy. As shown in Figure 3 the average size of the molecules was 41.6Kb (range 35Kb to 52Kb). The insert length is thus 30Kb to 47Kb with an average of 36Kb. The observed size distribution suggests that there is no

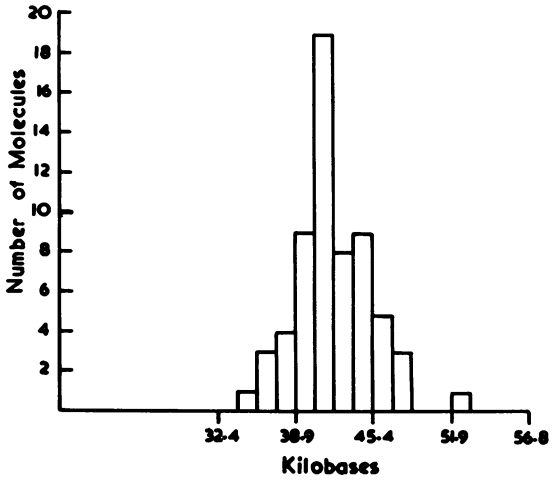

Figure 3. The Size of Cosmids from the *Eco*RI* *D. melanogaster* Library. Open circular DNA derived from a pool of 15,000 recombinant cosmids was spread from aqueous solution. Random fields were photographed and the sizes of individual molecules were measured relative to open circular SV40 DNA as length standard.

gross instability in the library. In agreement with this gel electrophoresis of total DNA from the pool of 15,000 recombinants showed a broad supercoiled band with a minimum size of 32Kb.

(iii) Sequence Representation in the Library.

The data presented above show that large segments of *D. melanogaster* DNA can be cloned in HomerI at high efficiency. However, it remains important to establish that such cosmid libraries are representative of the genome. We prepared six filters each carrying 300 independently derived colonies, in duplicate, and hybridised them with kinase-labeled *D. melanogaster* rRNA. 61 duplicate signals were obtained and we thus calculate that the *D. melanogaster* genome contains 387 rDNA units per haploid DNA complement which agrees well with the published haploid copy number of 255 (39). The library was also screened for zeta elements (40) and the number of clones obtained agreed well with the published haploid copy number of 40 (Diane de Cicco, personal communication).

(v) Stability of Cosmid Clones.

Our original purpose in developing this cosmid system was to clone the primary products of the integrative recombination between SV40 DNA and cellular DNA. We have shown that integration probably involves a tandemly

arrayed polymer of SV40 DNA (41) and thus expect the primary insertion to contain several SV40 genomes in tandem 'head to tail' array. Such tandemly duplicated viral sequences are known to be unstable in λ vectors (7) so we directly tested their stability when cloned in HomerI.

SV40 DNA was modified by *in vitro* methylation with *Eco*RI methylase to protect the single *Eco*RI site and then digested to completion with *Bgl*I. This DNA was then mixed, in a 9:1 ratio, with *Bgl*I-cleaved, unmodified SV40 DNA and the mixture was ligated at a concentration of 300μg/ml. The asymmetry of the *Bgl*I recognition site means that this ligation can produce polymers in which the viral genomes are orientated only in a 'head to tail' fashion. This was verified by digesting a sample of the ligated DNA with *Taq*I, which cleaves the SV40 genome once, and analysing the products by gel electrophoresis.

The ligated DNA was then digested with *Eco*RI; because most of the SV40 genomes have modified *Eco*RI sites this procedure will produce packageable polymers with two *Eco*RI cohesive termini which are, on average, eight SV40 genomes in length, i.e. 42Kb. The *Eco*RI-digested DNA was ligated to *Eco*RI-cleaved, dephosphorylated HomerI DNA and ampicillin-resistant colonies were selected after *in vitro* packaging and transduction. The cloning efficiency was greater than 10^5 cfu/μg target DNA.

Cosmid DNA was prepared by the small-scale procedure from six independently isolated clones and analysed by *Eco*RI digestion and agarose gel electrophoresis. Each cosmid yielded two fragments, one the size of monomeric SV40 DNA. However, the relative intensities of the two bands were not consistent with the cosmid DNAs being of packageable size, i.e. the SV40 DNA is under-represented. DNA was isolated from three clones by caesium chloride-ethidium bromide density gradient centrifugation and analysed by restriction endonuclease digestion and agarose gel electrophoresis. Figure 4 shows that each isolate comprises a heterogeneous mixture of circular plasmid DNAs. *Sal*I cleaves HomerI once but does not cut SV40 DNA and thus linearises each cosmid. The observed bands (Figure 4) correspond to cloned monomers, dimers, trimers and tetramers of SV40 DNA. Cloned tetramers are not of a packageable size and so the simplest interpretation is that cloned 'head to tail' polymers seven to nine SV40 genomes in length are unstable. This instability could be due to reversion of the *rec*A$^-$ mutation in HB101. However, the clones containing the heterogeneous mixtures of cosmids do not allow the growth of Charon 4, a *red*$^-$, *gam*$^-$ phage, but will efficiently propagate λ*vir*, indicating that they retain the *rec*A$^-$ marker.

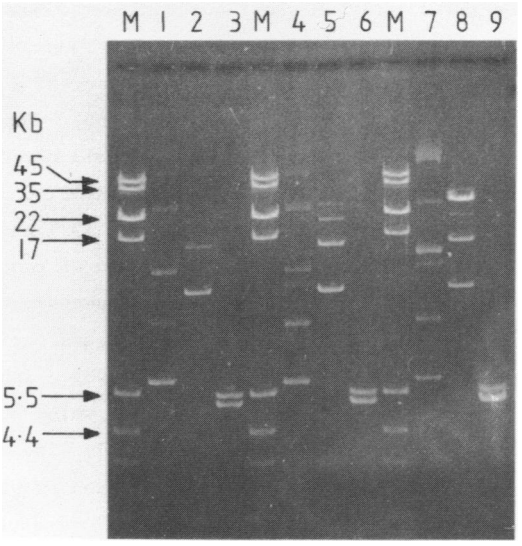

Figure 4. Instability of Cosmid Clones Containing Tandemly Repeated SV40 Genomes. Cosmid DNA was isolated from three independent colonies. Tracks 1, 4 and 7 contain undigested DNA, tracks 2, 5 and 8 contain *Sal*I-digested DNA and tracks 3, 6 and 9 contain *Eco*RI-digested DNA. Tracks labeled M contain DNA size markers. HomerI is 5.44Kb, SV40 DNA is 5.24Kb.

The circular cloned tetramer of SV40 DNA was purified by agarose gel electrophoresis and transformed into *E. coli* HB101. DNA was prepared from eleven ampicillin-resistant colonies and analysed by agarose gel electrophoresis. The cloned tetramer is stably maintained and has been over many generations.

DISCUSSION

HomerI offers some advantages over previously described cosmids. It is smaller than pFF2 and pHC79 (refs. 13, 42), thus allowing better use of the inherent capacity of such vectors, and has the advantage over pFF2, pHC79 and MUA-3 (ref. 43) that it is an approved, disabled (EK2-equivalent) system. Several derivatives of this vector are available. HomerII allows the cloning of *Sst*I fragments while HomerIII is designed for the cloning of segments of DNA that are to be analysed by transfection into cultured eukaryotic cells. It contains all of the *cis*-acting sequences required for the replication of SV40 DNA (44) and lacks the pBR322 sequences which inhibit efficient DNA replication in eukaryotic cells (34). Recombinants derived from HomerIII

will thus replicate efficiently in the SV40-transformed monkey cells (COS cells) described by Gluzman (45) which contain wild-type SV40 large T-antigen. The amplification of the DNA template thus obtained greatly facilitates the analysis of the transcripts of the cloned gene(s) (46). Moreover, the segment of SV40 DNA contained in HomerIII includes all of the sequences required for efficient *in vivo* transcription from the SV40 early promoter, including the 72 base-pair repeats. Banerji *et al.* (37) have shown that these sequences act *in cis* to enhance the expression of genes to which they are linked. Vectors containing the 72 base-pair repeats are thus of great utility in the analysis of gene expression by DNA transfection even in the absence of T-antigen-mediated DNA replication (47, 48). Ish-Horowicz and Burke (26) have described another variant, pJB8, in which a synthetic *Bam*I site has been inserted into the *Eco*RI site of HomerI. The restoration of a *Bam*I site allows the cloning of DNA fragments produced by partial digestion with *Mbo*I or *Sau*3A.

The optimal conditions for using Homer cosmids as cloning vectors have been explored. We conclude that the target DNA should be prepared by partial digestion with a restriction enzyme recognising a tetranucleotide sequence followed by careful size fractionation. Ligation of such DNA to a molar excess of dephosphorylated Homer DNA followed by *in vitro* packaging, transduction and selection for ampicillin-resistant colonies allows the cloning of eukaryotic DNA at an efficiency of 5×10^4 clones per µg of target DNA with no detectable vector background. Such efficiencies are as good as, if not better than, those obtainable with plasmid or λ vectors and, given the large inserts carried by cosmids, the number of clones that must be screened to isolate any sequence is significantly reduced. For example, a complete (>99%) mouse genomic library with an average insert size of 36Kb comprises 4×10^5 colonies, which can be made from 8µg of target DNA. Ten-fold higher efficiencies are obtainable if partial *Eco*RI or *Sst*I digests are cloned, but such procedures are less likely to lead to complete sequence representation and may well result in the cloning of non-contiguous fragments.

The average insert size in our *D. melanogaster* library is 36Kb which is slightly lower than that theoretically achievable. Analysis of the DNA prepared from pools of recombinant clones indicates that this insert size is maintained during propagation and suggests that there is no gross instability in the library. The *D. melanogaster* library constructed by partial *Eco*RI* digestion has been screened for both rDNA units and for zeta elements in

order to test the sequence representation and in both cases the number of clones detected corresponded well with the published copy number for the sequence in question. Other *Drosophila* libraries have been constructed by partial *Eco*RI or partial *Sst*I digestion using HomerI and HomerII, respectively. From such libraries we have isolated the alcohol dehydrogenase gene from seven ADH mutants of *D. melanogaster* and also from *D. mauritiana* and *D. simulans* (W. Chia, R. Karp, M. Bodmer and M. Ashburner, unpublished data). In each case the desired sequence was recovered at close to the expected frequency. rDNA units with and without intervening sequences have also been recovered from the *D. mauritiana* and *D. simulans* libraries (M. Browne and D.M. Glover, personal communication). Moreover, libraries constructed in HomerII from partial *Sst*I digests of SV40-transformed or SV40-infected mouse cell DNA have been used to recover both integrated viral DNA (10) and intact rDNA units (T. Gurney, C.E. Clayton, W. Chia and P.W.J. Rigby, unpublished data).

Cosmid libraries appear to be stable at a gross level but we do not yet have sufficient experience with them to comment in detail on the stability of all types of cloned sequence. We have, however, tested the issue of stability by attempting to construct a 'worst case' cosmid clone containing an octameric 'head to tail' tandem repeat of the SV40 genome. Such clones decayed rapidly to lower oligomers even in a *recA*$^-$ host. However, a cloned tetramer of SV40 DNA is stably maintained. We do not understand this difference between a tetramer and an octamer.

Since this work was completed others have reported on the construction and use of cosmid libraries. Ish-Horowicz and Burke (26) have described an elegant procedure using 'half-cosmids' and dephosphorylated target DNA which should totally exclude the cloning of non-contiguous fragments. Their work involved the insertion of partial *Mbo*I digests into the *Bam*I site of the HomerI derivative pJB8, but the same rationale can be used to insert partial *Taq*I and/or *Msp*I digests into the *Cla*I site of the Homer cosmids described here. They report slightly higher efficiencies than we have achieved but a comparable insert size of 38.8Kb. Meyerowitz *et al.* (43) have described a slightly smaller, 4.8Kb, cosmid called MUA-3 and have constructed *Drosophila* libraries using sheared DNA and synthetic *Eco*RI linkers. Their procedures gave somewhat lower efficiencies than ours but did result in a significantly larger average insert size, 45.5Kb. They report that several unique regions of the *Drosophila* genome can be isolated from their library but discuss stability problems with clones containing histone

genes, which are organised as tandemly arrayed units. Excision of complete repeat units is a likely explanation for this instability and the deleted clones obtained are subsequently stable, results similar to those we have obtained with the SV40 polymers. Grosveld *et al.* (49) have also used the HomerI derivative pJB8 and procedures very similar to ours to construct cosmid libraries of human DNA. They have isolated from such libraries β-globin, collagen and interferon genes and in the case of β-globin the number of clones obtained suggests proper sequence representation. In one case they have evidence for rearrangement of the inserted DNA. HomerI has also been used by Arrand *et al.* (50) to clone very large segments of the Epstein-Barr virus genome and by Lonsdale *et al.* to clone segments of maize mitochondrial DNA (51).

Our work, and that of others discussed above, clearly establishes that cosmid vectors can be efficiently and effectively used for the isolation of large segments of eukaryotic chromosomal DNA. The large insert size obtainable with cosmids is likely to be particularly advantageous in experiments that involve either 'walking' along long stretches of a chromosome or the subsequent analysis of the cloned segment by transfection into cultured cells. Moreover, the large insert size relative to the vector means that the preparation of eukaryotic DNA is much more efficient than with λ vectors and that a smaller number of clones is required to represent an entire eukaryotic genome. The large insert size does have one disadvantage, in that constructing a restriction enzyme map of the cloned segment, which is a necessary prerequisite to almost every type of experiment, can be a considerable undertaking. However, we (M.R.D. Scott and P.W.J. Rigby, manuscript in preparation) have developed a novel two-dimensional restriction mapping procedure which greatly simplifies the analysis of such large cloned segments. We have shown that several different types of sequence are appropriately represented in cosmid libraries and that, in general, such libraries can be stably propagated. Some sequences are unstable in cosmids but it is unlikely that any one vector system will be applicable to all segments of eukaryotic DNA and there is no doubt that cosmid cloning represents a valuable addition to our armoury of techniques.

ACKNOWLEDGEMENTS

We are grateful to Doug Hanahan for making pXf3 available to us and to Sue Hayman for her expert preparation of the manuscript. M.R.D.S. held a Research Studentship from the Medical Research Council while P.W.J.R. holds

a Career Development Award from the Cancer Research Campaign which also paid for this work.

[1]Present Address: Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, England.

[2]Present Address: Department of Microbiology and Immunology, University of California Medical School, San Francisco, California 94143, U.S.A.

REFERENCES

1. Wozney, J., Hanahan, D., Tate, V., Boedtker, H. and Doty, P. (1981) Nature 294, 129-135.
2. Nunberg, J.H., Kaufman, R.J., Chang, A.C.Y., Cohen, S.N. and Schimke, R.T. (1980) Cell 19, 355-364.
3. Kushner, S.R. (1978) in Genetic Engineering, Boyer, H.W. and Nicosia, S. Eds., pp. 17-23, Elsevier, Amsterdam.
4. Williams, B.G. and Blattner, F.R. (1980) in Genetic Engineering: Principles and Methods, Setlow, J.K. and Hollaender, A. Eds., Vol. 2, pp 201-281, Plenum Press, New York.
5. Collins, J. and Hohn, B. (1978) Proc. Natl. Acad. Sci. U.S.A. 75, 4242-4246.
6. Cooke, H.J. and Hindley, J. (1979) Nucl. Acids Res. 6, 3177-3197.
7. Fried, M., Murray, K., Klein, B., Greenaway, P., Tooze, J., Boll, W. and Weissmann, C. (1979) Nature 279, 811-816.
8. Lauer, J., Shen, C-K.J. and Maniatis, T. (1980) Cell 20, 119-130.
9. Villeponteau, B. and Martinson, H. (1981) Nucl. Acids Res. 9, 3731-3746.
10. Clayton, C.E. and Rigby, P.W.J. (1981) Cell 25, 547-559.
11. Hanahan, D. and Meselson, M. (1980) Gene 10, 63-67.
12. Collins, J. and Bruning, H.J. (1978) Gene 4, 85-107.
13. Baldacci, P., Royal, A., Bregegere, F., Abastado, J.P., Carni, B., Daniel, F. and Kourilsky, P. (1981) Nucl. Acids Res. 9, 3575-3588.
14. Twigg, A.J. and Sherratt, D. (1980) Nature 283, 216-218.
15. Wensink, P.C., Finnegan, D.J., Donelson, J.E. and Hogness, D.S. (1974) Cell 3, 315-325.
16. Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L-A., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L. and Smithies, O. (1977) Science 196, 161-169.
17. Shenk, T.E., Carbon, J. and Berg, P. (1976) J. virol. 18, 664-671.
18. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
19. Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978) Cell 15, 687-701.
20. Murray, N.E., Bruce, S.A. and Murray, K. (1979) J. Mol. Biol. 132, 493-505.
21. Efstratiadis, A., Vournakis, J.N., Donis-Keller, H., Chaconas, G., Dougall, D.K. and Kafatos, F.C. (1977) Nucl. Acids Res. 4, 4165-4174.
22. Greene, P.J., Poonian, M.S., Nussbaum, A.L., Tobias, L., Garfin, D.E., Boyer, H.W. and Goodman, H.M. (1975) J. Mol. Biol. 99, 237-261.
23. Kamp, D., Kahmann, R., Zipser, D. and Roberts, R.J. (1977) Molec. Gen. Genet. 154, 231-248.
24. Norgard, M.V., Keem, K. and Monahan, J.J. (1978) Gene 3, 279-292.
25. Morrison, D.A. (1977) J. bacteriol. 132, 349-351.
26. Ish-Horowicz, D. and Burke, J.F. (1981) Nucl. Acids Res. 9, 2989-2998.
27. Southern, E.M. (1975) J. Mol. Biol. 98, 503-515.

28. Jeffreys, A.J. and Flavell, R.A. (1977) Cell 12, 429-439.
29. Laskey, R.A. and Mills, A.D. (1977) FEBS Lett. 82, 314-316.
30. Davis, R.W., Simon, M. and Davidson, N. (1971) in Methods in Enzymology, Grossman, L. and Moldave, K. Eds., Vol. 21D, pp 413-428, Academic Press, New York.
31. Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H. and Falkow, S. (1977) Gene 2, 95-113.
32. Sutcliffe, J.G. (1979) Cold Spring Harbor Symp. Quant. Biol. 43, 77-90.
33. Kaiser, K. and Murray, N.E. (1979) Molec. Gen. Genet. 175, 159-174.
34. Lusky, M. and Botchan, M. (1981) Nature 293, 79-81.
35. Tooze, J. (1980) Molecular Biology of Tumor Viruses. Part 2, DNA Tumor Viruses. 2nd Edition. Cold Spring Harbor Laboratory, New York.
36. Benoist, C. and Chambon, P. (1981) Nature 290, 304-310.
37. Banerji, J., Rusconi, S. and Schaffner, W. (1981) Cell 27, 299-308.
38. Hohn, B. and Murray, K. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 3259-3263.
39. Tartof, K.D. (1973) Genetics 73, 57-71.
40. Kidd, S.J. and Glover, D.M. (1980) Cell 19, 103-119.
41. Chia, W. and Rigby, P.W.J. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 6638-6642.
42. Hohn, B. and Collins, J. (1980) Gene 11, 291-298.
43. Meyerowitz, E.M., Guild, G.M., Prestidge, L.S. and Hogness, D.S. (1980) Gene 11, 271-282.
44. Learned, R.M., Myers, R.M. and Tjian, R. (1981) in Developmental Biology Using Purified Genes, Brown, D. and Fox, C.R. Eds., ICN-UCLA Symposia on Molecular and Cellular Biology, Vol. 23, in press. Academic Press, New York.
45. Gluzman, Y. (1981) Cell 23, 175-182.
46. Mellon, P., Parker, V., Gluzman, Y. and Maniatis, T. (1981) Cell 27, 279-288.
47. Busslinger, M., Moschonas, N. and Flavell, R.A. (1981) Cell 27, 289-298.
48. Grosveld, G.C., de Boer, E., Shewmaker, C.K. and Flavell, R.A. (1982) Nature 295, 120-126.
49. Grosveld, F.G., Dahl, H-H.M., de Boer, E. and Flavell, R.A. (1981) Gene 13, 227-237.
50. Arrand, J.R., Rymo, L., Walsh, J.E., Bjorck, E., Lindahl, T. and Griffin, B.E. (1981) Nucl. Acids Res. 13, 2999-3014.
51. Lonsdale, D.M., Thompson, R.D. and Todge, T.P. (1981) Nucl. Acids Res. 9, 3657-3669.