Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 414–418, January 1984
Biochemistry

Direct isolation of the functional human thymidine kinase gene with a cosmid shuttle vector

(Southern blot analysis/human DNA library/hypoxanthine/aminopterin/thymidine selection)

YUN-FAI LAU AND YUET WAI KAN

Howard Hughes Medical Institute Laboratory and the Division of Medical Genetics and Molecular Hematology of the Department of Medicine, University of California, San Francisco, CA 94143

Communicated by Donald S. Frederickson, October 4, 1983

ABSTRACT **We have developed a new recombinant DNA cloning system to isolate directly the functional unit of the human thymidine kinase (TK) gene. The system utilizes a cosmid vector that can shuttle cloned DNA sequences between bacteria and mammalian cells. A complete human cosmid library was constructed and DNA from the total library was transfected to mouse L cells deficient in TK (LTK^-) by calcium phosphate precipitation. The transfected cells were then selected with hypoxanthine/aminopterin/thymidine (HAT) medium, and one HAT-resistant cell clone was isolated. This cell line became resistant to HAT selection by acquiring the TK gene derived from the human cosmid library. As the cosmid vector contains the cohesive ends of the bacteriophage, we could directly retrieve the human DNA sequences from the transformed mouse L cells. Total DNA from the transformed TK^+ L cells was packaged *in vitro* with lysogenic bacterial extracts and used to infect *Escherichia coli*. One of the two recombinant cosmids isolated contained a 43.8-kilobase human DNA insert and was capable of converting TK^- L cells to the TK^+ phenotype in both acute and stable transformation assays. Thus, we have isolated the functional human TK gene in this recombinant cosmid. The gene was further localized on a 14.5-kilobase *Bam*HI DNA fragment, and it transcribed a mature mRNA of about 1,500 nucleotides. This method of gene isolation has several special features: (*i*) an intact structural gene can be cloned directly based on its function without knowledge of its amino acid or nucleotide sequence; (*ii*) the functional gene sequences can be recovered faster and more efficiently than with the usual DNA transfection method; and (*iii*) in conjunction with cell-sorting techniques, this method can be used to clone genes encoding cell surface markers.**

Shuttle vectors are cloning vehicles that can be propagated in mammalian as well as bacterial cells. These vectors are useful for studying the structure and function of eukaryotic genes because (*i*) a cloned gene can be transformed into mammalian cells for expression studies and (*ii*) the DNA sequence can be rescued in bacterial cells to facilitate identification of essential DNA structures. Several shuttle vectors have been constructed (1–7); however, most of them are limited in the size of the insert DNA they can accommodate (2–4, 7), and some require extensive manipulation before the transforming DNA can be shuttled between mammalian cells and bacteria (1–7). We recently constructed a series of versatile cosmid vectors that can be used as vehicles for genomic DNA cloning and as shuttle vectors (8). The vectors were constructed by inserting the mammalian selectable gene markers, SV2-gpt (9), SV2-DHFR (10), or SV2-neo (11), into the cosmid pJB8. The cloned DNA could be stably transformed in cultured cells by using the appropriate selection system. In addition, because these vectors contain the replication origin and the 72-base-pair repeat enhancer sequences from simian virus 40 (SV40), the cloned genes also could be expressed transiently in COS cells. Because cosmids contain the cohesive ends of bacteriophage, total DNA from the recipient cells could be packaged directly and used to infect bacteria, thereby enabling retrieval of the transformed DNA from the mammalian cells. We used this cosmid vector system to study the human α-globin gene cluster isolated from a human recombinant DNA library (8). The α-globin genes were expressed both transiently in COS cells and stably in mouse L cells, and the α-globin gene cluster could be rescued intact from the total DNA of the transformed cells.

In this study we used these cosmids as shuttle vectors for the direct isolation of a eukaryotic gene based on its function. We isolated the human thymidine kinase (TK; ATP: thymidine 5′ phosphotransferase, EC 2.7.1.21) gene by transforming mouse L cells deficient in TK (LTK^-) with a human cosmid library DNA containing the complete human genome. After selection in HAT medium [Dulbecco-modified Eagle medium (DME medium) containing HAT (100 μM hypoxanthine/0.5 μM aminopterin/20 μM thymidine)] the cells that expressed TK enzyme activity were propagated, and the human TK gene was rescued directly as a functional unit from the transformed mouse L cells by *in vitro* packaging reactions.

MATERIALS AND METHODS

Preparation of Human Cosmid Library DNA. The construction of human recombinant cosmid libraries by using the shuttle cosmid vectors pCV103, pCV107, and pCV108 that contained the mammalian selectable markers SV2-gpt, SV2-DHFR, and SV2-neo, respectively, has been described (8). For this study, about 2×10^8 bacteria containing the human DNA library cloned in pCV108 were inoculated into 1 liter of L broth in the presence of ampicillin (50 μg/ml). Cosmid sequences were amplified with chloramphenicol (200 μg/ml) at an OD_{600} of 0.8 for 16–20 hr. Total cosmid DNA was then isolated by the alkaline lysis procedure (12). The DNA preparations were treated with RNase (50 μg/ml) for 30 min at room temperature and then with proteinase K (100 μg/ml) at 37°C for 1 hr. Then they were extracted with phenol, precipitated with ethanol, resuspended in 10 mM Tris·HCl/1 mM EDTA, pH 7.5, and dialyzed extensively against the same buffer.

Transformation of Mouse TK^- Cells. The mouse LTK^- cell line (obtained from M. Bishop) was maintained in DME medium supplemented with 10% fetal calf serum and penicillin/streptomycin. Cell transformations were performed by calcium phosphate coprecipitation as described (8). About 10 μg of cosmid library DNA/phosphate precipitates per 10^6 LTK^- cells was used. As positive controls we used plasmids pTK (13) and pMK (14), which contain, respectively, the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HAT, hypoxanthine/aminopterin/thymidine; TK, thymidine kinase; SV40, simian virus 40; kb, kilobase.

herpes simplex virus TK gene and the same gene under control of the mouse metallothionein gene promoter. For the transformation experiments with these plasmids and with the isolated DNA fragments purified from the rescued cosmids, 0.25 μg of DNA per equivalent number of LTK$^-$ cells was used. DNA/phosphate precipitates were added to cells for 16 hr. The cells were then treated with 15% glycerol in phosphate-buffered saline for 2 min, washed with phosphate-buffered saline, and grown for 24 hr in nonselective medium before selection in HAT medium. HAT-resistant cells (HAT^R) were grown to mass culture and further selected with the antibiotic G418.

Assay of TK Gene Activity. Cells were labeled with 5 μCi (1 Ci = 37 GBq) of [^{3}H]thymidine per ml in complete DME medium for 48 hr in tissue culture chamber slides and fixed with acetic acid/methanol, 1:3 (vol/vol). Autoradiography was performed with Kodak NTB2 emulsion. Exposure time was 2 days at 4°C.

Rescue of Human TK Sequences. Total DNA was isolated from transformed mouse L cells that were resistant to both HAT and G418 selection. The transforming cosmid sequences were rescued to bacteria by *in vitro* packaging of the total cellular DNA with lysogenic bacterial extracts as described (8).

Mapping the TK Gene. Southern blot and RNA blot hybridization analyses were performed as described (15, 16). Restriction enzyme mapping of DNA fragments was carried out with single and multiple enzyme digestions and partial digestion of DNA fragments end-labeled with ^{32}P (17).

RESULTS

Transformation of LTK$^-$ Cells and Rescue of Transforming Sequences. Primary transformation was performed with a recombinant cosmid library of normal human DNA by using the cosmid pCV108 that contained the selectable gene marker SV2-neo. The cosmid library contained about 5×10^5 individual recombinant cosmids and, hence, represented the human genome six to seven times. Total DNA isolated from this library was digested with the restriction enzyme *Eco*RI to release the inserted DNA from the cosmid. Ethidium bromide staining of DNA size-fractionated by agarose gel electrophoresis showed a smear pattern indistinguishable from that obtained from similarly digested total human DNA. Thus, the DNA isolated from the cosmid library is a fair representation of the human genome, and selected sequences were not preferentially amplified.

Two HAT-resistant colonies were isolated from the primary transformation of about 10^7 cells, yielding a transformation efficiency of approximately 1 per 5×10^6 cells. One of these colonies was grown to mass culture and was resistant to both HAT and G418 selection. Total cellular DNA was isolated from this cell line and packaged *in vitro* with lysogenic bacterial extracts. About 200 individual cosmids were rescued from 6 μg of total cellular DNA, giving a rescue efficiency of about 33 colonies per μg of DNA. DNA was prepared from 84 individual colonies and digested with *Eco*RI. Two cosmid patterns, designated pHTKA and pHTKB, were observed (Fig. 1). They contained a 39- and a 43.8-kilobase (kb) DNA insert, respectively. pHTKA was the major pattern and was present in 80 of the 84 cosmids screened.

DNA from these two cosmids was used for secondary transformation of LTK$^-$ cells. As indicated by [^{3}H]thymidine incorporation and resistance to HAT selection, only pHTKB was capable of transforming the TK$^-$ cells to TK$^+$ (Fig. 2 and Table 1). Therefore, it must contain the functional human TK gene. The major species, pHTKA, was incapable of transforming the TK$^-$ cells. Preliminary studies indicated that a gene encoded in pHTKA was actively transcribed in L and COS cell cultures.

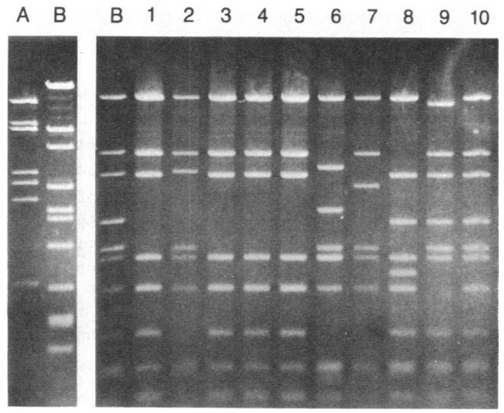

FIG. 1. Ethidium bromide stain of *Eco*RI-digested DNA from cosmids rescued from the primary (*Left*) and secondary (*Right*) transformations. Lanes: A and B, cosmids pHTKA (lane A) and pHTKB (lanes B) rescued from the primary transformants; 1–10, cosmids rescued from the DNA of the secondary transformants. Note that rearrangements have occurred in most cosmids. The cosmid shown in lane 10 is identical to the original pHTKB.

The transforming DNA again was rescued by *in vitro* packaging of DNA isolated from the secondary transformants. Although all the rescued cosmids contained the same basic patterns as pHTKB, most displayed some rearrangement (Fig. 1).

Identification of the Functional Unit of the Human TK Gene. In order to define the functional unit of the human TK gene within the 43.8-kb insert in pHTKB, the cosmid was

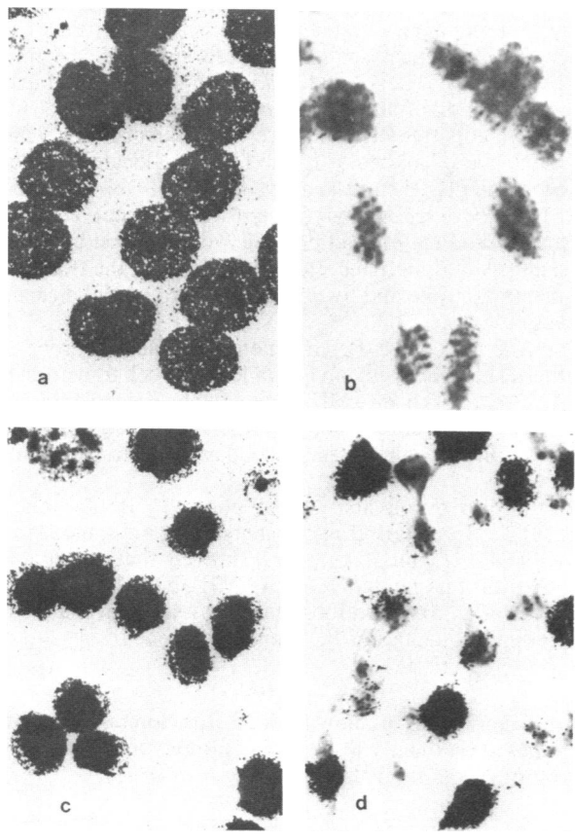

FIG. 2. Detection of TK activity by autoradiography. Cells were incubated with [^{3}H]thymidine for 2 days, and TK activity was detected by the incorporation of [^{3}H]thymidine into the cell nuclei. (*a*) CHO TK$^+$ cells. (*b*) LTK$^-$ cells. (*c*) LTK$^-$ cells after primary transformation with the cosmid library and HAT selection. (*d*) LTK$^-$ cells after acute transfection with the rescued cosmid pHTKB.

Table 1. Transformation efficiency of LTK$^-$ cells

Transforming DNA	Colonies/μg DNA/10^6 cells
pCV108 (vector)	0
pHTKA	0
pHTKB	41
pHTKB-*Bam*HI (14.5 kb)*	284
pHTKB-*Nru* I (42 kb)*	332
pMK†	2,160
pTK†	552

*Linearized DNA fragments.
†pMK and pTK are plasmids containing the herpes simplex TK gene under the control of the metallothionein gene promoter and its own promoter, respectively.

digested with different restriction enzymes, and the larger fragments were purified from agarose gels and used for LTK$^-$ cell transformation. Table 1 compares the transformation efficiency of pHTKB to two linearized fragments derived from it: a 14.5-kb *Bam*HI fragment and a 42-kb *Nru* I fragment. As controls, cosmid vector pCV108 without the human DNA insert and two plasmids containing the herpes TK gene under its own promoter (pTK) or the mouse metallothionein promoter (pMK) were used. Both DNA fragments consistently transformed TK$^-$ cells to TK$^+$. Therefore, the functional unit of the human TK gene is contained in the 14.5-kb *Bam*HI fragment.

The 14.5-kb *Bam*HI DNA fragment was labeled and used as probe in Southern blot analysis of total human, primary, and secondary transformant DNA (Fig. 3). A smeared pattern, characteristic of human *Alu*-like repeat DNA sequences was observed with total human DNA (18). However, defined bands were observed in the DNA from both primary and secondary transformants, suggesting that most other *Alu*-like repeat sequences were absent in the transformed L cells. The 14.5-kb *Bam*HI fragment was further digested with different restriction enzymes to produce small DNA segments free of *Alu*-like repeat sequences. Two such fragments were isolated: a 1.6-kb fragment located between an *Xho* I and *Eco*RI site and a 0.95-kb *Eco*RI DNA fragment. The 14.5-kb *Bam*HI fragment and its adjacent DNA sequences within the cosmid pHTKB were mapped by Southern analysis of doubly digested cosmids using the two DNA fragments as probe and by end labeling and partial digestion studies (Fig. 4).

The *Xho* I/*Eco*RI DNA fragment was used as probe for Southern analysis of human genomic and L cell transformant DNA. The 14.5-kb *Bam*HI fragment that presumably contained the TK functional unit and its flanking sequences was preserved in the primary and secondary TK$^+$ transformants (Fig. 3).

Blot-hybridization analysis of poly(A)$^+$ RNA isolated from LTK$^+$ secondary transformants using the same *Xho* I/*Eco*RI DNA fragment as probe showed that the mature mRNA from this human TK gene was about 1,500 nucleotides long (Fig. 5). Therefore, this DNA segment must contain coding sequences of the human TK gene.

DISCUSSION

Several approaches are now available for cloning eukaryotic genes based on their expression in cultured cells (1, 20–23).

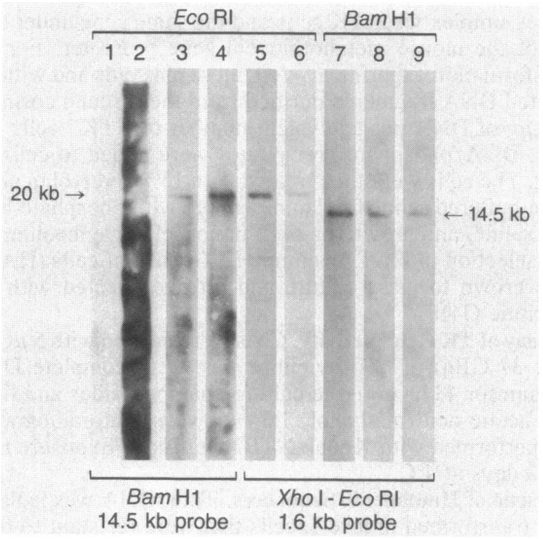

FIG. 3. Southern blot analysis of human, L cell, and cosmid DNA with rescued DNA sequences as probes. Lanes: 1, LTK$^-$ cells; 2, 6, and 8, human DNA; 3, primary LTK$^+$ transformants; 4 and 9, secondary LTK$^+$ transformants; 5 and 7, cosmid pHTKB. Genomic DNA (7 μg) or cosmid DNA (10 μg) was digested with *Eco*RI and *Bam*HI as indicated. The filters in lanes 1–4 were probed with ^{32}P-labeled 14.5-kb *Bam*HI fragment containing the functional unit of the human TK gene. The filters in lanes 5–9 were proved with the ^{32}P-labeled 1.6-kb *Xho* I/*Eco*RI DNA segment within the 14.5-kb *Bam*HI fragment.

In the most commonly used method, total genomic DNA serves as donor sequences and, after a series of secondary transformations, cells that are free of most nonessential transforming DNA sequences are isolated. The target gene can then be rescued by constructing a recombinant DNA library and identified with repeat DNA sequences (21) or linked plasmid probes (20). Such procedures are tedious and time consuming. Occasionally the target genes are separated into different recombinant DNA molecules in the final rescue steps, thereby hampering identification of the functional DNA sequences.

Cloning genes with shuttle vectors by DNA-mediated gene transfer techniques improves the efficiency of the entire cloning operation (1). Unfortunately most shuttle vectors do not meet the basic requirements as general genomic cloning vectors for construction of recombinant DNA libraries that represent the total genome of the donor. The shuttle cosmid vectors described in the present study constitute the only system available that can efficiently fulfill both requirements. Furthermore, inclusion of selectable gene markers in these cosmid vectors reduces the background of revertants during cell selection.

The identification of recombinant cosmids containing the intact human TK gene was straightforward in the HAT selection system. The gene was localized onto a 14.5-kb *Bam*HI DNA fragment, and it transcribed a mature mRNA of about 1,500 nucleotides in length. Based on our analyses with the 1.6-kb *Xho* I/*Eco*RI DNA probe, no rearrangements from the human genome were detected in cosmid pHTKB. In addition, the restriction patterns of all the cosmids rescued

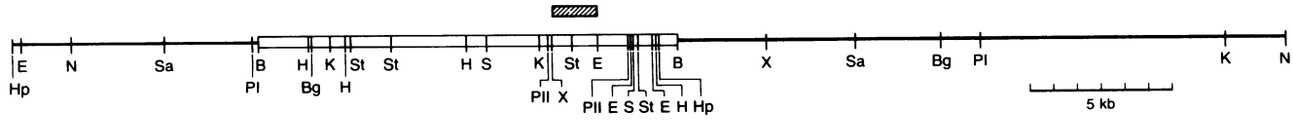

FIG. 4. Partial restriction map of the human TK gene and its flanking sequences in cosmid pHTKB. □, The functional unit as determined by transformation assay; ▨, the 1.6-kb *Xho* I/*Eco*RI DNA segment used as probe in Southern analysis. Restriction enzyme sites: *Bam*HI, B; *Bgl* II, Bg; *Eco*RI, E; *Hin*dIII, H; *Hpa* I, Hp; *Kpn* I, K; *Nru* I, N; *Pvu* I, PI; *Pvu* II, PII; *Sal* I, Sa; *Sma* I, S; *Sst* I, St; *Xho* I, X.

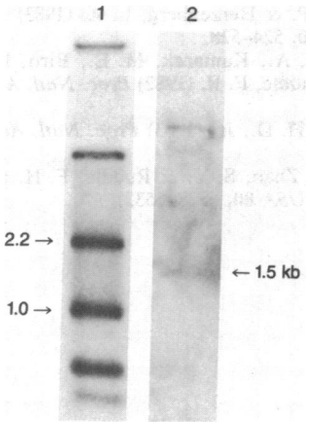

FIG. 5. Blot-hybridization analysis of 20 μg of poly(A)$^+$ RNA isolated from LTK$^-$ cells transformed with pHTKB to TK$^+$. Lanes: 1, ^{32}P-end-labeled *Hin*dIII-digested phage PM2 DNA; 2, the 1.5-kb RNA band corresponding in size to mature human TK gene transcripts (19).

from the primary transformants were identical to either of the two patterns observed. As occurred in our previous studies with the α-globin gene cluster (8), various rearrangements from the original input pattern were found in the majority of the cosmids rescued from the secondary transformants. Restriction enzyme analysis of these cosmids indicated that the intact TK gene was not involved in these changes. Because the secondary transformant remained TK$^+$, the functional unit of the TK gene was preserved.

The physical state of the cosmid DNA in the primary LTK$^+$ transformants is not known. In order for the cosmids to be rescued from the host cells, they must either be arranged in tandem or exist as circular forms within the cells. Several transformation studies performed with cloned DNA sequences showed that donor DNA sequences were tandemly integrated in the host genome (14, 24, 25). However, based on our data, tandem integration of the primary transformants seems unlikely for several reasons. First, because total human cosmid DNA was used as donor sequences, the unique TK gene would be represented once per haploid genome or one TK-containing cosmid per 80,000 recombinant cosmids. When carrier DNA has been used for transformation experiments, the transforming sequences were physically linked to the carrier sequences in the host cells (26). If, in our study, nonessential cosmids within the DNA library served as carrier sequences in the primary LTK$^+$ transformation, we would expect the rescued cosmids to be complex and of multiple species. However, only two species of cosmids were rescued from the DNA of these cells. Hence, if the cosmids were integrated, it must have occurred at a late stage of transformation—i.e., after most nonessential cosmids except pHTKA and pHTKB were depleted from the cells. Second, all rescued cosmids belonging to the same species had identical restriction enzyme patterns, suggesting that no junction sequences between pHTKA and pHTKB from any tandemly integrated forms were represented. Finally, plasmids devoid of pBR322 "poison" sequences have been shown to replicate extrachromosomally in mammalian cells with the SV40 origin of replication (6, 27, 28). Because the cosmid vectors in this study were derived from plasmid pAT153 that had deleted such inhibitory sequences from pBR322 (29) and because the SV40 origin of replication was included in the selectable markers (9–11), it is therefore probable that the transforming cosmids existed as circular extrachromosomal elements rather than integrated forms in the primary transformant cells.

In addition to cloning genes by culturing transformed cells in selective media, this type of cosmid is also useful for isolating genes that are expressed as cell surface markers (30, 31). Cells can be transformed with total DNA from the human cosmid library, and those that retain the human DNA can be selected with antibiotic G418. The cells containing the genes that express the surface marker of interest can then be isolated by fluorescent staining with antibodies against that marker and by cell sorting. The gene for the human Leu-2 antigen has been successfully isolated by using the total DNA transformation system (30). Transforming cells with a total human cosmid library DNA as donor sequences would greatly facilitate this method of gene isolation. One disadvantage of the present system is the limitation on the size of the insert imposed by cosmid cloning. Genes larger than 45 kb would not be represented in their entirety within a recombinant cosmid and could not be cloned with this shuttle vector.

Note Added in Proof. The papers by Bradshaw (32) and Lin *et al.* (33) have now been published.

The authors thank Drs. H. D. Bradshaw and P.-F. Lin for sending us preprints of their manuscripts and J. Gampell for assistance in preparation of the manuscript. This research is supported in part by National Institutes of Health Grant AM16666. Y.W.K. is an Investigator of the Howard Hughes Medical Institute.

1. Perucho, M., Hanahan, D., Lipsick, L. & Wigler, M. (1980) *Nature (London)* **285**, 207–210.
2. Law, M.-F., Lowy, D. R., Dvoretzky, I. & Howley, P. M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2727–2731.
3. DiMaio, D., Treisman, R. & Maniatis, T. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4030–4034.
4. Sarver, N., Byrne, J. C. & Howley, P. M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7147–7151.
5. Breitman, M. L., Tsui, L.-C., Buchwald, M. & Siminovitch, L. (1982) *Mol. Cell. Biol.* **2**, 966–973.
6. Conrad, S. E., Liu, C.-P. & Botchan, M. R. (1982) *Science* **218**, 1223–1225.
7. Kushner, P. J., Levinson, B. B. & Goodman, H. M. (1982) *J. Mol. Appl. Genet.* **1**, 527–538.
8. Lau, Y.-F. & Kan, Y. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5225–5229.
9. Mulligan, R. C. & Berg, P. (1980) *Science* **209**, 1422–1427.
10. Subramani, S., Mulligan, R. & Berg, P. (1981) *Mol. Cell. Biol.* **1**, 854–864.
11. Southern, P. J. & Berg, P. (1982) *J. Mol. Appl. Genet.* **1**, 327–341.
12. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523.
13. Graham, F. L., Bacchetti, S., McKinnon, R., Stanners, C., Cordell, B. & Goodman, H. M. (1980) in *Introduction of Macromolecules into Viable Mammalian Cells*, eds. Croce, C. & Rovera, G. (Liss, New York), pp. 3–25.
14. Mayo, K. E., Warren, R. & Palmiter, R. D. (1982) *Cell* **29**, 99–108.
15. Kan, Y. W. & Dozy, A. M. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5631–5635.
16. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5201–5205.
17. Smith, H. O. & Bernstiel, M. L. (1976) *Nucleic Acids Res.* **3**, 2387–2398.
18. Schmid, C. W. & Jelinek, W. R. (1982) *Science* **216**, 1065–1070.
19. Lin, P.-F., Yamaizumi, M., Murphy, P. D., Egg, A. & Ruddle, F. H. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4290–4294.
20. Lowy, I., Pellicer, A., Jackson, J. F., Sim, G.-K., Silverstein, S. & Axel, R. (1980) *Cell* **22**, 817–823.
21. Jolly, D. J., Esty, A. C., Bernard, H. U. & Friedmann, T. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5038–5041.
22. Pulciani, S., Santos, E., Lauver, A. V., Long, L. K., Robbins, K. C. & Barbacid, M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2845–2849.
23. Shih, C. & Weinberg, R. A. (1982) *Cell* **29**, 161–169.
24. Folger, K. R., Wong, E. A., Wahl, G. & Capecchi, M. (1982) *Mol. Cell. Biol.* **2**, 1372–1387.

25. Kudo, A., Yamamoto, F., Furusawa, M., Kuroiwa, A., Natori, S. & Obinata, M. (1982) *Gene* **19,** 11–19.
26. Perucho, M., Hanahan, D. & Wigler, M. (1980) *Cell* **22,** 309–317.
27. Spandidos, D. A., Harrison, P. R. & Paul, J. (1982) *Exp. Cell Res.* **141,** 149–158.
28. Lusky, M. & Botchan, M. (1981) *Nature (London)* **293,** 79–81.
29. Ish-Horowicz, D. & Burke, J. F. (1981) *Nucleic Acids Res.* **9,** 2989–2998.
30. Kavathas, P. & Herzenberg, L. A. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 524–528.
31. Barbosa, J. A., Kamarck, M. E., Biro, P. A., Weissman, S. M. & Ruddle, F. R. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 6327–6331.
32. Bradshaw, H. D., Jr. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 5588–5591.
33. Lin, P.-F., Zhao, S.-Y. & Ruddle, F. H. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 6528–6532.