Volume 8 Number 22 1980 Nucleic Acids Research

The construction of a recombinant cDNA library representative of the poly(A)$^+$ mRNA population from normal human lymphocytes

Derek Woods, Julian Crampton, Berwyn Clarke and Robert Williamson

Biochemistry Department, St. Mary's Hospital Medical School, University of London, London W2 1PG, UK

Received 10 October 1980

ABSTRACT

A recombinant library has been constructed using the plasmid pAT153 and double stranded cDNA prepared from normal human lymphocyte poly(A)$^+$ RNA. Transformation conditions were optimized to yield approximately 200,000 recombinants per μg of double stranded cDNA. Statistical analysis as well as sequence complexity analysis of the inserted sequences indicates that the cDNA library is representative of >99% of the poly(A)$^+$ RNA present in the normal human lymphocyte.

INTRODUCTION

Recombinant DNA technology has made possible the detailed molecular analysis of a considerable number of specific genes such as those coding for globin, histone and rRNA (1, 2,3). Libraries of total genomic DNA from various organisms (4) have been used to isolate specific genomic DNA sequences. This approach is limited to some extent by the need to obtain a purified probe for the specific gene or RNA molecule being studied. This is simplest when such a sequence is present at high abundance in a particular cell type.

The construction of cDNA libraries from a variety of cell types is a necessary prerequisite to a more detailed analysis of the distribution of mRNA populations. The analysis requires that the cDNA libraries are representative of the mRNA populations of the cell type in question. By comparing libraries derived from different cell types it may be possible to isolate the expressed sequences which are common to both, and those unique to only one cell type.

This paper describes the construction of such a library derived from normal human lymphocyte poly(A)$^+$ RNA. Complementary DNA was synthesised from this poly(A)$^+$ RNA,

double stranded and inserted into a plasmid vector by G-C tailing. The main problem associated with such a cloning procedure is ensuring that the cloned sequences are representative of the poly(A)$^+$ RNA population. Analysis of the cloned sequences indicates that the library closely reflects the original poly(A)$^+$ RNA population both in nucleotide complexity and relative abundance.

MATERIALS AND METHODS

Isolation of poly(A)$^+$ RNA from normal circulating human lymphocytes

Normal circulating human lymphocytes were purified from buffy coat residues using the Ficoll-hypaque separation procedure described by Harris et al. (5). Total RNA was isolated from the lymphocytes by a modification of the method described by Strohman et al. (6) for cells containing high levels of ribonuclease. Pelleted lymphocytes were suspended in 10-20 volumes of a solution containing 19 parts of 6M guanidine hydrochloride (GuHCl), one part 2M potassium acetate, pH5. The cells were sonicated (3 x 15 sec bursts) to shear the DNA, and the RNA precipitated at -20^o overnight by addition of a half volume of ethanol. The RNA was pelleted at 16000g for 10 minutes and dissolved in 1/10th the original volume of 9.5 vol. 6M GuHCl, 0.5 vol. 2M potassium acetate pH5, 0.5 vol. 500mM EDTA pH7.0. The RNA was again precipitated by addition of a half volume of ethanol. The above step was repeated and the pelleted RNA dissolved in 20mM EDTA pH7.0 and the solution extracted with an equal volume of chloroform:butanol (4:1). The RNA was precipitated with ethanol, dissolved in NTS buffer (0.5M NaCl, 10mM Tris pH7.5, 0.1% SDS) and the poly(A)$^+$ RNA purified using oligo(dT) cellulose chromatography (7).

cDNA synthesis

cDNA was synthesised in a reaction mixture containing the following: 50mM Tris/HCl, pH8.2; 50mM KCl; 6mM MgCl; 5mM dithiothrietol; 500μM each of dTTP, dCTP, dGTP, dATP; 25 units/ml rat-liver RNase inhibitor (Searle); 1 unit/ml oligo $(dT)_{12-18}$ (Collaborative Research), 250 units/ml reverse transcriptase (obtained through the Biological Carcinogenesis

Branch, Viral Oncology Program, National Cancer Institute) and 10-50μg/ml poly(A)$^+$ RNA. The reaction mixture was incubated at 42^o for 1 hr. The reaction was stopped by addition of 1/10th volume of 100mM EDTA and 1/15th vol. of 5M NaOH followed by incubation at 60^o for 30 mins. The solution was neutralised with 1N HCl and the cDNA desalted on a G50 Sephadex column. ^{3}H cDNA (10^7 cpm/μg) was synthesised using 1 mCi/ml of ^{3}H dCTP (22 Ci/mMole) and the 500μM unlabelled dCTP was excluded from the incubation. The ^{32}P labelled cDNA (500,000 cpm/μg) used in the cloning experiments was synthesised by adding 100μCi/ml ^{32}P dCTP (400 Ci/mMole) and reducing the unlabelled dCTP concentration to 100μM.

Synthesis of double stranded cDNA

cDNA was used as a template for the double stranded cDNA (ds cDNA) synthesis using the Klenow fragment of E.coli DNA polymerase I (Boehringer) (8). The reaction conditions were essentially those described by Humphries et al. (9): 30mM Tris/HCl, pH7.5; 4mM MgCl; 0.5mM β-mercaptoethanol; 70mM KCl; 500μM each of dCTP, dGTP, dTTP and dATP; 1 - 15μg/ml of ^{32}P labelled cDNA and 75 units of DNA polymerase I (Klenow fragment); incubation was at 18^o for two hours. After incubation the mixture was made 10mM with EDTA and passed over a G50 Sephadex column equilibrated with S1 buffer (70mM Na acetate, 2.8mM Zn_2SO_4, 140mM NaCl pH4.5). To the excluded volume, 40 units of S1 nuclease were added and the solution incubated at 37^o for 1 hr. The incubate was extracted with phenol;chloroform (1:1) and the aqueous phase fractionated on a 15-30% sucrose gradient. Aliquots of the gradient fractions were analysed on a 1% agarose gel and those fractions containing the ds cDNA (9S and greater) were pooled and precipitated with ethanol.

Homopolymer tailing

The DNA tailing reactions were carried out using a modification of the procedure described by Roychoudhury et al. (10). Incubations containined 100mM Tris/HCl pH7.3, 140mM sodium cacodylate, 1mM β-mercaptoethanol, 1mM cobalt chloride, BSA (nuclease free, BRL) 100μg/ml, DNA termini 10-50nM and dCTP 100μM or dGTP 1mM. Analytical tailing reactions were carried out to determine suitable nucleotide concentrations and incubation

times for each batch of DNA used. The poly dG tailing reaction, of the PstI digested plasmid, was incubated for 1 hr at 37° at a dGTP concentration of 1mM and DNA terminus concentration of 80nM. The ds cDNA tailing reaction was at 37° for 5 min at a dCTP concentration of 100μM and ds cDNA concentration of 5μg/ml. After tailing, both DNA preparations were phenol extracted, precipitated with ethanol and dissolved in TEN buffer (100mM NaCl; 10mM Tris/HCl, pH7.5; 1mM EDTA).

Cloning Procedures

The homopolymer tailed plasmid and ds cDNA in TEN buffer were annealed by mixing at a 5:1 (wt/wt) plasmid:cDNA ratio at a plasmid DNA concentration of 2.5μg/ml. The mixture was incubated at 65° for 2 hr and allowed to cool slowly overnight (average rate 2°/hr).

The E.coli K12 strain HB101 was made competent for transformation by the method described by Wensink et al. (11). The standard transformation method used 2 vol. of competent cells and 1 vol. of the transforming DNA (1μg/ml) in TEN buffer. The competent bacteria and DNA were mixed, left on ice for 10 min and then given a -55° (isopropanol:dry ice) cold shock for 100 seconds. The mixture was allowed to thaw at room temperature then placed on ice for a further 5 min. This was followed by incubation at 42° for 1 min and then by 2 min at 0°. 700μl of L broth was added and the final transformation mixture was incubated at 37° for 30 min. This allows the cells to recover from the CaCl treatment (12).

Following transformation the cells were filtered directly onto nitrocellulose filters to give an average density of 5000 transformants/90mm diameter filter. The filters were incubated at 37° for 1 hr on L-agar containing no antibiotics and then transferred to L-agar containing 10μg/ml tetracycline in order to select for transformants. The presence of plasmid pAT153 confers tetracycline resistance.

Plasmid DNA was prepared from a detergent lysate of bacteria (13) by centrifugation in ethidium bromide/caesium chloride density gradients (14).

All genetic manipulations were carried out under CII containment conditions in accordance with G.M.A.G. advice for

recombinant DNA research.

RESULTS AND DISCUSSION

Preparation of poly(A)$^+$ RNA

High levels of endogenous RNase in human lymphocytes resulted in degradation of RNA isolated using standard extraction procedures, even in the presence of RNase inhibitors. For this reason we have used the guanidine hydrochloride isolation procedure described by Strohman et al. 1977 (6). The cells were solubilised in 6M GuHCl and the total cellular RNA separated from the DNA and protein by selective ethanol precipitation. Under these conditions undegraded total RNA (approx. 2mg/10^9 cells), as judged by analysis on 2% agarose SDS gels, was isolated. Poly(A)$^+$ RNA was purified using oligo (dT) cellulose chromatography (7). This poly(A)$^+$ RNA contains, as a result of the isolation procedure, both cytoplasmic and nuclear poly(A)$^+$ RNA.

When added to the rabbit reticulocyte lysate cell free system (15) this RNA directed the synthesis of high MW polypeptides which migrated as discrete bands on polyacrylamide SDS gels. These are similar to those synthesised *in vivo* by human lymphocytes (S. Humphries, unpublished results).

Preparation of Double Stranded cDNA

Single stranded cDNA was synthesised as described in Materials and Methods and a yield of approximately 20% of the input RNA was routinely obtained. The size distribution of the ^{32}P labelled cDNA was analysed on 1% agarose gels (Fig. 1) and shown to range from 400 to 2000 nucleotides in length. This corresponds to the size distribution of the lymphocyte poly(A)$^+$ RNA as analysed by sucrose gradients and hybridisation to tritiated poly(U), by the method of Rosbash and Ford (16).

The nucleotide complexity of the cDNA preparation was determined by its hybridisation to the lymphocyte poly(A)$^+$ RNA and the hybridisation kinetics are shown in Fig. 2. The extent of hybridisation was determined by S1 nuclease assay. From analyses of the components of the $R_o t$ curve it can be calculated, using mouse globin mRNA:cDNA hybridisation as a standard, that there are approximately 8000 discrete poly(A)$^+$ RNAs present in

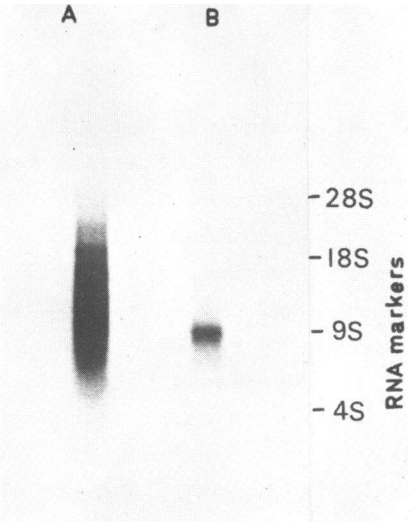

Fig. 1 Analysis of ^{32}P labelled lymphocyte cDNA on a 1% agarose gel. The gel was dried and autoradiographed. A. lymphocyte cDNA B. mouse globin cDNA.

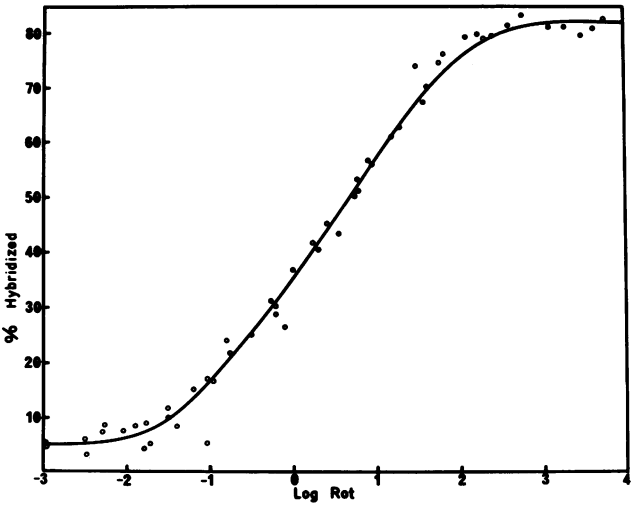

Fig. 2 Lymphocyte poly(A)$^+$ RNA:cDNA hybridisation analysis. ^3H labelled lymphocyte cDNA was hybridised to lymphocyte poly(A)$^+$ RNA in RNA excess and the percentage of the cDNA hybridised was assayed by resistance to S1 nuclease digestion.

normal human lymphocytes (Table I). This figure agrees with that of Ostrow et al. (17) for human lymphoblastic leukaemic cells.

After the second strand synthesis with DNA polymerase I, 50-70% of the ^{32}P labelled first strand cDNA was resistant to S1 nuclease. The S1 treated ds cDNA was fractionated on a 15-30% sucrose gradient in order to remove low MW material. Size analysis of this double stranded, S1 nuclease treated cDNA (Fig. 3) shows that there is a decrease in the mean size of the ds cDNA as compared to the single stranded cDNA, indicating that the double stranding reaction does not give full length ds cDNA.

Cloning the ds cDNA

The cloning of the ds cDNA was performed using the plasmid vector pAT153 in the host bacteria HB101. The single PstI site in pAT153 was used for the insertion of the cDNA using G and C tailing. Under these conditions recombination of the homopolymer tailed plasmid (poly dG) and ds cDNA (poly dC) results in reconstruction of the PstI site at either end of the inserted DNA. This allows the inserted sequences to be excised from the vector for further analysis. Another advantage is that the PstI site of pAT153 is situated in the ampicillin resistance gene, thus screening for ampicillin sensitive clones is a simple test for recombinants.

Conditions were determined which allowed us to tail the

Table I

a Component	b % of cDNA hybridised	c $Rot_{\frac{1}{2}}$	d $Rot_{\frac{1}{2}}$ corrected	e Number of 1000bp seq.	Relative copy number
1	54	0.22	.119	35	260
2	46	56	26	7800	1

a) The Rot curve Fig. 2 was divided into 2 components representing pseudo first order reactions i.e. 2 log units/ component. b) % of cDNA in each component after normalising the curve in Fig. 2. c) Rate at which each component is 50% hybridised. d) $Rot_{\frac{1}{2}}$ corrected for the percentage of the cDNA in each component. e) Calculated using mouse globin cDNA:mRNA hybridisation as a standard and assuming the average size of a mRNA to be 1000 base pairs.

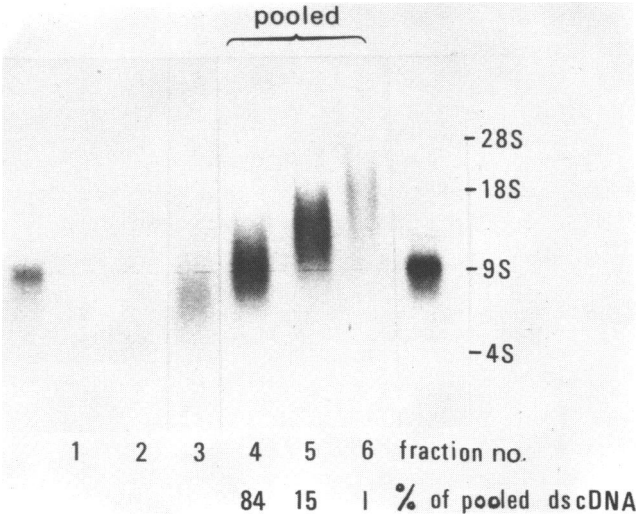

Fig. 3 Autoradiograph of sucrose gradient fractions of ds cDNA analysed on 1% agarose gel. DNA was denatured prior to electrophoresis. The radioactive marker on the gel is 9S mouse globin cDNA.

PstI digested pAT153 to an average tail length of 15 dG residues per terminus. The optimum ds cDNA tailing conditions were difficult to determine as the concentration of the DNA termini could only be estimated, due to the cDNA size heterogeneity. The ds cDNA was incubated under conditions which we estimated would add 10-30 dC residues per DNA terminus. It is important that the DNA used in the tailing reactions is not nicked as this results in homopolymer addition at the nicks, giving overestimates of the tail length, resulting in low cloning efficiencies (i.e. yields of recombinants).

The plasmid and ds cDNA tails were annealed at a weight to weight ratio of 5:1 respectively. Increasing the ratio of plasmid to ds cDNA did not increase the yield of recombinants significantly but resulted in higher backgrounds of non-recombinants. Transformation efficiencies were 2-5 x 10^6 transformants/μg supercoiled plasmid DNA, 5 x 10^3 transformants/μg linear tailed plasmid and 5 x 10^4 transformants/μg linear tailed plasmid after annealing to the tailed ds cDNA. Therefore

the cDNA library of 200,000 recombinant clones was prepared from 1μg of double stranded tailed cDNA annealed to 5μg of tailed plasmid. When analysed by ampicillin sensitivity the cDNA library contained less than 10% non-recombinant plasmids.

The plasmid DNA used in the preparative cloning was not exhaustively digested with PstI, in order to ensure that the PstI cohesive ends were maintained. Thus the linear tailed plasmid probably contained small amounts of uncut plasmid.

Storage and Analysis of the Library

Little is known about the stability of complex populations of recombinant clones. For this reason we decided to store the cDNA library in three ways; as clones on nitrocellulose filters obtained from the original transformation; as purified supercoiled plasmid DNA and as glycerol stocks. The storage of the clones on nitrocellulose filters was performed as described by Hanahan and Meselson (23) and under these conditions the library is stable. Glycerol stocks and plasmid DNA were obtained as follows; cloned recombinants, on the nitrocellulose filter, were shaken into one litre of L broth and small aliquots (10ml) were removed and stored as glycerol stocks at -20°. From the remaining 900ml the recombinant plasmid was isolated by the standard cleared lysate method (13). The bacterial suspension was not grown, nor were the plasmids amplified prior to this isolation. The stability of these cDNA libraries on storage for periods of greater than several months is at present under investigation.

The reconstruction of the PstI site on either end of the inserted sequences and the size of the inserted sequences were analysed by the Southern blot technique (20). PstI digested and undigested library plasmid DNA were electrophoresed on a 1% agarose gel and transferred to nitrocellulose paper. The filter was then hybridised with ^{32}P kinase labelled lymphocyte poly(A)$^+$ RNA (21) and autoradiographed (Fig. 4). The result shows, in the PstI digested DNA, no detectable nicked circular or linear plasmid carrying sequences which hybridise to poly(A)$^+$ RNA showing that the PstI sites have been regenerated. This observation was supported by the individual analysis of 20 recombinant clones picked at random from the library; in all of which the

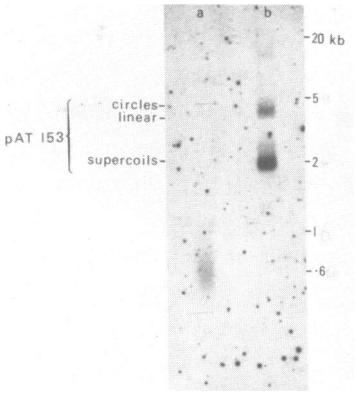

Fig. 4 Southern transfer of (a) PstI digested and (b) undigested lymphocyte library DNA probed with ^{32}P kinase labelled lymphocyte poly(A)$^+$ RNA.

inserted sequence was excised by PstI digestion. The average size of the inserted sequences is approximately 600 base pairs but inserts of up to 1300 base pairs are evident (Fig. 4).

Table I indicates that there are approximately 8000 different poly(A)$^+$ RNA species present in lymphocytes. However, due to the varying abundance of these RNAs, a total of 20,000 RNA molecules from the total population would be required if each of the low abundance RNA species were to be represented at least once. Modifying the formula which Clarke and Carbon (22) applied to genomic cloning, ($P = 1(1-1/f)^N$ where f = number of poly(A)$^+$ RNA sequences present (20,000), N = number of clones, P = probability of any one sequence being cloned) we have calculated that there is a 99% probability that any one cDNA sequence is present in the 200,000 recombinants cloned from the lymphocyte cDNA preparation.

The fraction of the lymphocyte cDNA complexity represented in the cDNA library was determined by comparing the hybridisation kinetics and saturation levels of the cDNA/poly(A)$^+$ RNA reaction shown in Fig. 2 with the hybridisation of the same ^{3}H-labelled cDNA preparation with plasmid DNA isolated from the library (Fig. 5). The kinetics of the two reactions show a similar distribution of abundance classes in the cloned

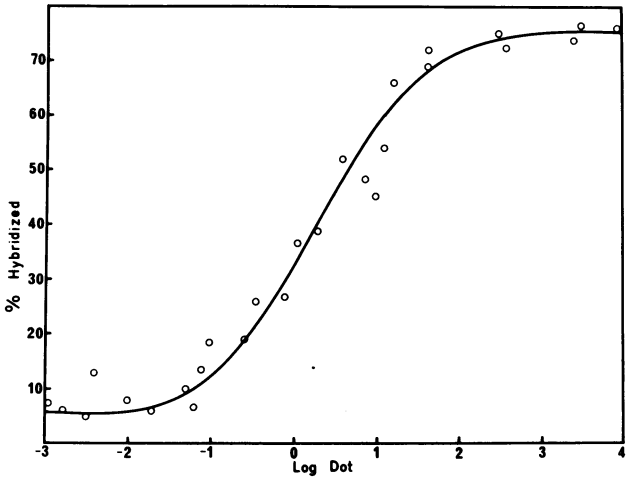

Fig. 5 Kinetic analysis of the hybridisation of ^{3}H labelled lymphocyte cDNA to excess sheared library DNA. The percentage hybridisation was assayed by resistance to S1 nuclease digestions. Dot values were calculated from estimation of the insert DNA concentration.

sequences to those represented in the poly(A)$^{+}$ RNA population. However, it is evident from Figs. 2 and 5 that only 90% of the hybridisable cDNA (that cDNA which hybridises to the poly(A)$^{+}$ RNA at saturation) anneals to the cloned library cDNA sequences (assayed by resistance to S1 nuclease digestion). There are two possible explanations for this. Either the cDNA library only contains 90% of the cDNA sequences, or it contains a complement of all the cDNA sequences but they are only 90% of the cDNA length. As the double-stranded cDNA is smaller than the single-stranded cDNA, this would suggest that the second explanation is correct. This has been confirmed by analysing the cDNA hybridised to the cloned sequences (at saturation) by hydroxyl-apatite chromatography. All the hybridisable cDNA eluted as double-stranded material.

ACKNOWLEDGEMENTS

We wish to thank Teresa Knapp for her technical assistance, the British Blood Transfusion Service for supplying buffy coat residues and Dr. J. Beard (N.I.H.) for a generous gift of RNA

dependent DNA polymerase. This work was supported by grants from the Nuffield Foundation and the Medical Research Council.

References

1. Hardison, R.C., Butler, E.T., Lacy, E., Maniatis, T., Rosenthal, N., Efstratiadis, A. (1979) Cell 18, 1285-1297.
2. Kedes, L.H. (1979) Ann. Rev. of Biochem. 48, 837-870.
3. Sollner-Webb, B., Reeder, R.H. (1979) Cell 18, 485-499.
4. Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K., Efstratiadis, A. (1978) Cell 15, 687-701.
5. Harris, R.T. and Ukaejiofo, E.O. (1970) Brit. J. Haematol. 18, 229-235.
6. Strohman, R.C., Moss, P.S., Micou-Eastwood, J., Spector, D., Przybyla, A. and Paterson, B. (1977) Cell 10, 265-273.
7. Aviv, H. and Leder, P. (1972) P.N.A.S. 69, 1408-1412.
8. Klenow, H., Overjaard-Hansen, K., Palkar, S.A. (1971) Eur. J. Biochem. 22, 371-381.
9. Humphries, P., Cochet, M., Krust, A., Gerlinger, P., Kourilsky, P. and Chambon, P. (1977) Nuc. Acids Res. 4, 2389-2406.
10. Roychoudhury, R., Jay, E. and Wu, R. (1976) Nuc. Acids Res. 3, 101-116.
11. Wensink, P.C., Finnegan, D.J., Donelson, J.E. and Hogness, D.S. (1974) Cell 3, 315-325.
12. Lennox, E.S. (1955) Virology 1, 190- 206.
13. Clewell, D.B. and Helinski, D. (1969) Proc. Natl. Acad. Sci. U.S.A. 62, 1159-1166.
14. Radloff, R., Bauer, W. and Vinograd, J. (1968) Proc. Natl. Acad. Sci. U.S.A. 57, 1514-1521.
15. Pelham, H.R.B. and Jackson, R.J. (1976) Eur. J. Biochem. 67, 247-256.
16. Rosbash, M. and Ford, P.J. (1974) J. Mol. Biol. 85, 87-101.
17. Ostrow, R.S., Woods, W.G., Vosika, G.J. and Faras, A.J. (1979) Biochem. Biophys. Acta 562, 92-102.
18. Twigg, A. and Sherratt, D. (1980) Nature 283, 216-218.
19. Bolivar, F., Rodriguez, R.C., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H. and Falkow, S. (1977) Gene 2, 95-113.
20. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
21. Van de Sande, J.H., Kleppe, K. and Khorana, H.G. (1973) Biochemistry 12, 5050-5055.
22. Clarke, L. and Carbon, J. (1976) Cell 9, 91-99.
23. Hanahan, D. and Meselson, M. (1980) Gene 10, 63-67.