Volume 10 Number 16 1982 Nucleic Acids Research

Isolation of the human HDL apoprotein A1 gene

Carol C.Shoulders and Francisco E.Baralle

Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

Received 9 June 1982; Revised and Accepted 13 July 1982

ABSTRACT

Apo A1 is the major apoprotein of the human plasma high density lipoprotein (HDL). We have isolated apo A1 cDNA and genomic clones and used them to study the gene organisation as defined by its restriction enzyme map. These studies showed that apo A1 is coded by a unique gene. Cross hybridisation was not observed with functionally related apoprotein genes. Increased levels of HDL have been correlated with certain protection against coronary heart disease. If there is a genetic component that contributes to the variable levels of HDL found in the population, it may be possible to correlate these differences with distinct gene organisation patterns.

INTRODUCTION

The function of the lipoproteins in plasma is mainly to transport lipids (primarily cholesterol and triglycerides) from one organ to another. It has recently become evident that they not only solubilize hydrophobic lipids but may also dictate the site in the body to which each lipid class is to be delivered. Defects in the lipoprotein structure or metabolism contribute to the early onset of atherosclerosis as documented by a wide variety of genetic diseases (e.g. familial hyperlipoproteinaemias) (1). The specific genetic defect has not been identified for many of these diseases. Furthermore, many common diseases where lipid transport and metabolism may be altered such as atherosclerosis are very difficult to study from a genetic point of view because they result from complex interactions between several genes or group of genes and the environment. The powerful gene manipulation techniques have not been applied as yet to any significant extent in the study of cardiovascular disease. However, basic work in the study of the structure and function of the genes involved, for example, in lipid transport and metabolism, may provide the framework for the success of future research in the pathogenesis of atherosclerosis. A growing body of evidence (2-8) shows that increased levels of high density lipoprotein-

0305-1048/82/1016-4873$2.00/0

cholesterol (HDL-C) are correlated with certain protection against coronary heart disease. If there is a genetic component that contributes to the variable levels of HDL-C found in the population, it may be possible to correlate these differences with distinct gene organisation patterns as was the case with sickle cell anemia (9-11). We wish to report here the isolation of cDNA and genomic clones for the major apo protein component of HDL (apo A1). The study of the organisation of the human apo A1 gene showed that apo A1 is coded by a unique gene, although polymorphic alleles as defined by variation in restriction enzyme patterns may be present. Once these polymorphisms are precisely defined, they will be used as markers to study families with high incidence of arterial disease lacking any obvious predisposing cause.

MATERIALS AND METHODS

(a) cDNA cloning

The initial cloning of human apo A1 cDNA was carried out following the specific oligonucleotide primer approach (12-14). By selecting a favourable region of the known amino acid sequence of human apo A1 (15), it is possible to predict from the genetic code eight 15 long oligonucleotides, one of which will be exactly complementary to apo A1 mRNA (see Figure 1). The oligonucleotides were synthesised simultaneously (30) using the solid phase phosphotriester method developed by M. Gait (16).

Synthesis of internally or end labelled cDNA for analytical purposes was carried out as previously described (17) using as template 5 μg of a sucrose fraction (smaller than 18S RNA) of human liver RNA and the mixture of synthetic oligonucleotides as primer. The clone, M13A1, was constructed as follows: cDNA synthesis was carried out using 100 μg of RNA template. The cDNA was fractionated on a denaturing gel (see Fig. 1) and band 5 was eluted. This was converted to double strand by the 'loop back' reaction with Klenow DNA polymerase (19). The resulting double strand cDNA was digested with HaeIII and ligated to SmaI restricted and phosphatased M13mp9 vector (20). The ligation mixture was transformed into the bacterial strain JM101. The sequence analysis (21) of four of the recombinant phages allowed us to identify one of them as an apo A1 clone (M13A1). A cDNA library was prepared synthesising cDNA with "smaller than" 18S liver mRNA as template and using oligo dT_{12-18} as primer. The cDNAs produced were converted to double strand as described above and the single strand hairpin loop structure was digested with S1 (23). The products were purified in a

Sephacryl S300 column and the eventual 5' overhanging ends of the cDNA were repaired by "filling in" with the Klenow fragment of DNA polymerase I. The cDNA molecules at this stage were predominantly blunt ended so they were ligated to the PvuII site of pBR322 (previously treated with phosphatase to avoid its circularization) and transformed into MC1061 (24). This procedure avoided the classical tailing of the cDNA and vector ends. About 10,000 recombinant colonies were produced from 10 μg of mRNA. This cDNA library was screened for apo A1 sequences using a probe prepared from M13A1 (see Figure 2).

(b) Genomic blotting

Restriction enzyme digests, blotting and hybridisation were as previously described (29). The molecular weight of fragments was determined using λ HindIII and pBR322 fragments as markers. The probe used was either a copy of the 48 long apo A1 fragment present in M13A1 (see Figure 2B) or nick translated (22) pBA1.2.

(c) Isolation of the apo A1 gene

A human foetal liver gene library obtained from T. Maniatis (31) was screened for apo A1 clones using nick translated pBA1.2 as probe. One positive plaque was observed after screening 5 x 10^5 recombinants. The positive clone (λ A1) was characterised by restriction enzyme mapping. The fragments containing the apo A1 sequences were identified by hybridisation after blotting the λ A1 digests.

RESULTS AND DISCUSSION

(a) Isolation of cDNA clones

The cDNAs synthesised using the specific oligonucleotide mixture as a primer were fractionated in a gel (Figure 1). Bands 1-6 were eluted and analysed in two ways. Firstly, the internally labelled cDNAs were digested with the restriction enzyme HaeIII which under special conditions is able specifically to cut single strand DNA (18). The products of digestion from bands 1-6 were fractionated in a 6% polyacrylamide 7M urea gel (not shown). The sizes of the fragments were compared with the possible HaeIII sites that can be predicted from the apo A1 amino acid sequence. Only the HaeIII fragments of band 5 (about 150, 90 and 50 nucleotides long) were compatible with the apo A1 amino acid sequence.

Secondly, the end labelled cDNAs were eluted and subjected to the Maxam and Gilbert sequencing techniques. The resulting data was of poor quality but tentative sequences could be read from it and again only band 5 was

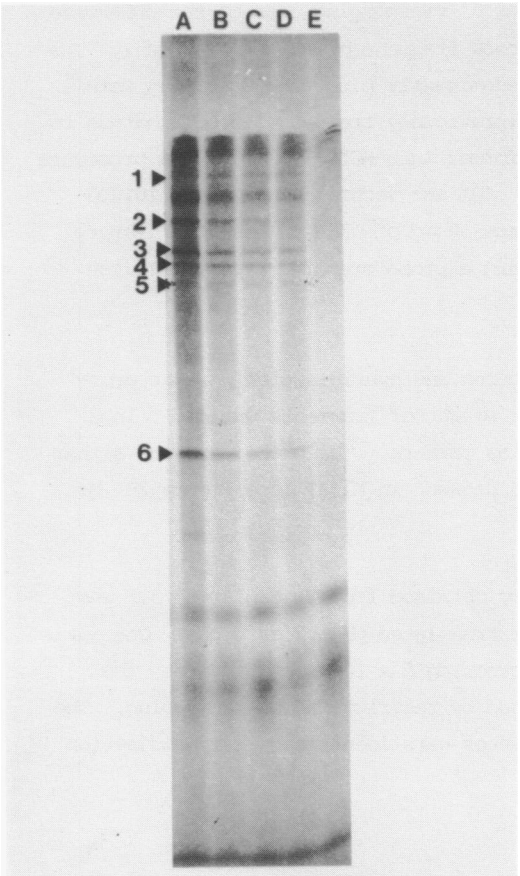

Figure 1 Autoradiograph of a 4% polyacrylamide gel fractionation of cDNA synthesised using an oligonucleotide of defined sequence as primer. The primer sequence was designed to be complementary to the mRNA region coding for amino acids 108 to 112 of human apo A1 (bottom figure). From the genetic code an mRNA sequence can be predicted with 3 ambiguities (R indicates a purine, Y indicates pyrimidine). Hence a mixture of the 8 possible complementary oligonucleotides was simultaneously synthesised (16, 30). Analytical cDNA synthesis was carried out as described in Methods. The pentadecamer primer mixture was added at different concentration as follows: lane A 500ng, lane B 200ng, lane C 100ng, lane D 37ng, lane E no primer added.

compatible with the amino acid sequence. This preliminary characterisation bypassed the classical first step in cDNA cloning that is to establish the presence of the mRNA by *in vitro* translation and allowed us to concentrate directly on the specific apo A1 band. The number of different clones produced by a gel band should be small enough to screen by direct sequencing. The cDNA of band 5 was converted to double strand by the "loop back" reaction with Klenow DNA polymerase (19). The resulting double strand cDNA was not

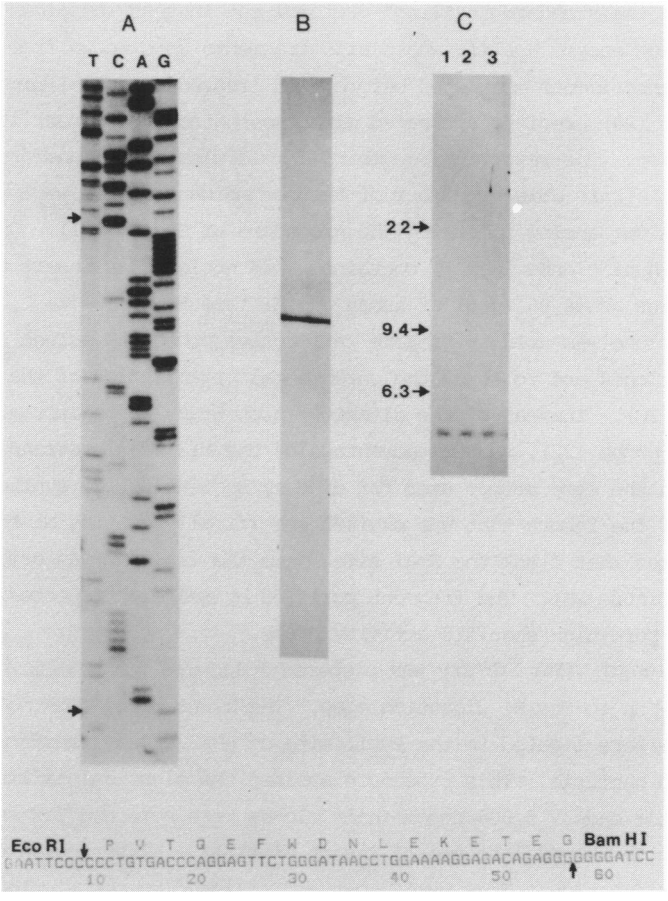

Figure 2

(A) Autoradiograph of a dideoxy sequencing gel of the clone M13A1. The sequence that can be read between the arrows is indicated at the bottom of the figure. The amino acid sequence predicted by the cloned fragment is indicated above in the nomenclature used by Dayhoff (28) and matches completely the apoprotein A1 sequence previously published (15).
(B) Preparation of a radioactive probe from the clone M13A1. The single strand M13A1 DNA was prepared and copied as in the DNA sequencing protocol (21) but in the absence of dideoxynucleotide triphosphates. The apo A1 fragment was then excised by digestion with the restriction enzymes EcoR1 and BamHI, whose sites are flanking the SmaII site where it was originally cloned (see bottom figure). The fragment was then purified on a 12% polyacrylamide gel.
(C) Autoradiograph of a Southern blot of a SacI digest of the human DNA obtained from three unrelated individuals. The blots were carried out as described in reference 29 and the probe used was the M13A1 EcoR1/BamHI fragment prepared as described in (B). Only one band can be seen hybridising in each individual (see text and Figure 3).

taken through the standard "tailing" protocol but instead was digested with the restriction enzyme HaeIII (whose size fragments were known from the preliminary experiments mentioned before) and ligated to a SmaI digested M13mp9 vector (20) previously treated with phosphatase to prevent its circularization. The resulting recombinant molecules were transformed into the bacterial strain JM101 and four of the recombinant clones were analysed by direct DNA sequencing following the procedure of Sanger (21). One of these recombinant clones (M13A1) contained a 48 nucleotides insert that codes for amino acids 66 to 81 of human apo A1 (see Figure 2).

The cDNA clone was used to prepare radioactive probes to screen for full length cDNA clones and to study the chromosomal organisation of the apo A1 gene (Figure 2C). Instead of the classical nick translation approach to prepare the probe (22), we took advantage of the M13 single strand chromosome and with the same primer used for dideoxy sequencing, we synthesised a DNA copy of the insert that was excised specifically cutting at the EcoR1 and BamHI sites that flank the SmaI site where the fragment was originally cloned (see Figure 2B). The fragment purified by gel electrophoresis was an extremely pure high specific activity probe.

A full length cDNA library was prepared following the standard protocol (23) up to the S1 digestion step. The blunt ended cDNA molecules (see Methods) were ligated to the PvuII site of pBR322 and transformed directly into bacteria. This procedure avoided the classical tailing of the cDNA and vector ends. 5,000 recombinant clones were screened for sequences complementary to the 48 long apo A1 insert of M13A1. Eleven colonies gave positive hybridisation. They were grown and analysed by restriction enzyme digest and then the two carrying the larger inserts (pBAI.1 and pBAI.2) were selected for further analysis and DNA sequencing. Figure 3A shows a HinfI digest of pBAI and pBAII. Preliminary sequence data show that pBAI.1 starts around amino acid 44 and lacks part of the 3' non coding region while pBAI.2 starts at amino acid 55 and has all the 3' non coding region.

(b) Gene mapping

The cDNA plasmids M13A1 and pBAI.2 were used as probes for Southern transfers (25) of human genomic DNA digested with various restriction enzymes (see Figure 3B). The human apoprotein A1 gene organisation as defined by preliminary restriction enzyme map is shown in Figure 3. It is clear that human apo A1 is present at a unique position in the haploid genome. There is no cross hybridisation, even at low stringency (2 x SCC final wash), with any other of the related apoprotein genes (28). Further-

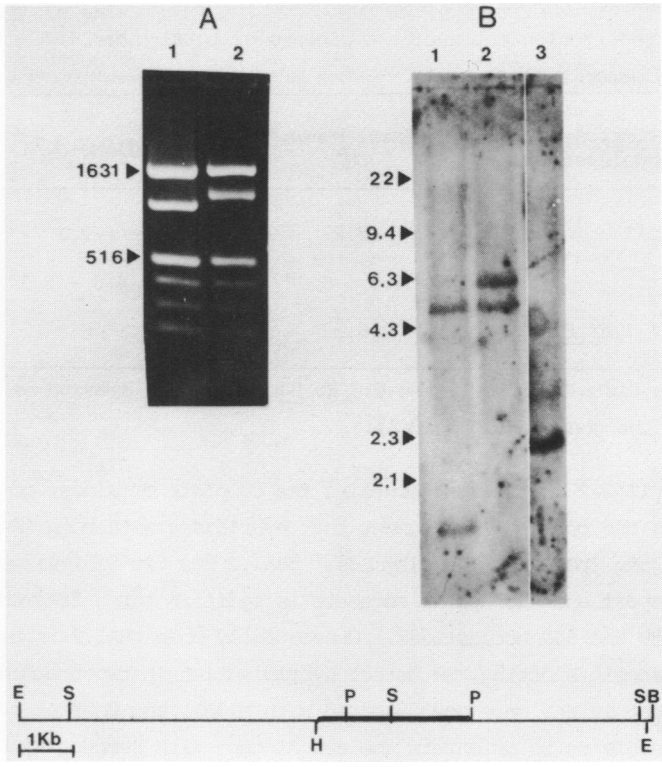

Figure 3

(A) Agarose gel fractionation of a restriction endonuclease HinfI digest of clones pBAI.1 (lane 1) and pBAI.2 (lane 2). The apo A1 sequences do not contain internal HinfI sites. The largest HinfI fragments of pBR322 are 1631 and 516 base pairs (b.p.) long and are indicated in the Figure. The fragments running between them contain the apo A1 sequences and 344 b.p. of pBR322. The size of the inserts was calculated using HpaII and HinfI digests of pBR322 as markers to estimate the respective fragment size and subtracting the pBR322 component of it. pBAI.1 and pBAI.2 have respectively 630 ± 50 and 850 ± 50 b.p. inserts.
(B) Autoradiograph of a Southern blot of human DNA digested with different restriction enzymes (lane 1 HindIII/SacI double digest, lane 2 SacI digest, lane 3 PvuII digest). The blot was carried out as previously described (29). The probe used was nick translated (22) pBAI.2. Similar experiments carried out with different enzymes allowed us to construct a preliminary restriction enzyme map as shown in the bottom of the Figure. E, S, H, P and B denote respectively the restriction enzymes EcoRI, SmaI, HindIII, PvuII and BamHI. The thick line indicates the region that hybridises to the pBAI.2 probe. It should be noted that in the cDNA clones there are two SacI sites 35 b.p. apart. The same arrangement may be present in the genome or the two SacI sites may be farther apart because of the existence of small intervening sequences between them. In either case, the central S indicates two SacI sites relatively close together.

Table 1 Sizes (in kb) of DNA fragments containing the apo A1 gene in restriction endonuclease digests of total human DNA and the genomic clone λ A1

Restriction endonuclease	Total human DNA	Clone λ A1
PvuII	2.2	2.1
SacI	4.8 + 6.0[a]	4.6 + 5.5[a]
EcoR1 + HindIII	6.3	6.3

(a) The distal 0.5 kb of the 6.0 kb SacI fragment seem to be absent in the genomic clone λ A1.

more, the PvuII 2.2 kb fragment contains the complete or almost complete gene as it is the only PvuII fragment that hybridises with both the short (M13A1) and long (pBAI.2) cDNA plasmids. pBAI.2 has two SacI sites 35 nucleotides apart and the apo A1 sequence is split in three fragments of around 35, 250 and 500 nucleotides. Consequently when pBAI.2 is used as a probe in genomic blotting, we detect hybridisation in two fragments, one of 6.0 kb (that is the only one detected with M13A1) and another of 4.8 kb. These data enable us to orientate the apo A1 gene with respect to the restriction sites. The 5' end of the apo A1 gene is on the side of the longer SacI fragment close to the HindIII restriction site, whereas the 3' end is on the shorter SacI fragment on the inside of the PvuII fragment.

From the size of the cDNA synthesised with the specific primer (see Figure 1), it was deduced that the apo A1 mRNA has at its 5' terminus about 110 nucleotides comprising the 5' non coding region and the signal peptide sequence. From the estimated size of the 3' non coding region we can predict the size of the mRNA as about 1,100 nucleotides (not considering the polyA tail). The whole apo A1 gene appears to be contained in the PvuII 2.2 kb fragment, so this leaves a maximum of 1,000 nucleotides of possible non coding sequences in the form of introns. As yet the restriction enzymes used in our mapping experiments have failed to produce any evidence of interruptions in the gene. However, the presence of small introns cannot be ruled out.

(c) Isolation of the apo A1 gene

An apo A1 genomic clone (λA1) was isolated from a human foetal liver

gene library (31) using pBAI.2 as probe. The restriction enzyme map of λ A1 is in agreement with the map deduced from genomic blotting (see Figure 3 and Table 1). The clone contains the PvuII 2.2 kb fragment and this is the only PvuII fragment that hybridises to pBAI.2. The size of the HindIII/EcoRl and SacI fragments is also consistent with the genomic blotting, although it is possible that the distal part of the SacI fragment may be be absent in the clone (Table 1).

A fine mapping and complete sequence analysis of the gene is currently underway. The cDNA clone pBAI.2 and the genomic clone λ A1 are being used to study by blotting the detailed gene organisation in normal individuals. We expect to define in this way restriction enzyme site polymorphisms in the population. Once these polymorphisms are precisely defined, they will be used as markers to study families with high incidence of arterial disease lacking any obvious predisposing cause. In addition to the well established variations in the apo A1 amino acid sequence (26,27), there may exist regulatory alleles (e.g. with a different degree of response to hormone action) that may only be detected at the DNA level. As a given combination of genes involved in lipid transport interacts between themselves and with the environment to provide the fine tuning of lipid transport and metabolism, so it is probable that different combinations of regulatory alleles will result in a marginally different spectrum of plasma lipoprotein levels and consequently in different long term effects in both lipid transport and cell metabolism and thus the state of the vascular system.

ACKNOWLEDGEMENTS

We thank Professor G G Brownlee and K Gould for help and advice in the oligonucleotide synthesis, C E Phillips for excellent technical assistance, Dr E Tuddenham for human liver tissue, Dr M Carroll for liver mRNA and the British Heart Foundation for support. C.C.S. holds an MRC research studentship.

REFERENCES

1. The Metabolic Basis of Inherited Disease, eds. J.B. Stanbury, J.B. Wyngaarden and D.S. Frederickson, McGraw-Hill, New York (1978).
2. A.M. Gotto, N.E. Miller and M.F. Oliver, eds. High Density Lipoproteins and Atherosclerosis, Third Argenteuil Symposium, Elsevier, New York (1978).
3. Gordon, T., Castelli, W.P., Hjortland, M.C., Kannel, W.B. and Dawber, T.R. Am.J.Med. 62, 707 (1977).
4. Rossner, S., Kjellin, K.G., Mettinger, K.L., Siden, A. and Soderström,

C.E. Lancet (1), 577 (1978).
5. Miller, G.J. and Miller, N.E. Lancet (1), 16 (1975).
6. Willett, W. et al, N.Engl.J.Med. 303, 1150-1161 (1980).
7. Stein, Y., Glangeaud, M.C., Fainaru, M. and Stein, O. Biochim.Biophys. Acta 380, 106 (1975).
8. Miller, N.E., Weinstein, D.B., Carew, T.E., Koschinsky, T. and Steinberg, D. J.Clin.Invest. 60, 78 (1977).
9. Kan, Y.W. and Dozy, A.M. Proc.Natl.Acad.Sci.USA 75, 5631-5635 (1978a).
10. Kan, Y.W. and Dozy, A.M. Lancet (2), 910-912 (1978b)
11. Fourth Cooley's Anemia Symposium, W.French Anderson, A. Bank and E.C. Zaino, eds. Ann.N.Y.Acad.Sci., 344, 1-448 (1980).
12. Baralle, F.E. Cell 10, 549-558 (1977).
13. Noyes, B.E., Mevabech, M., Stein, R. and Agarwal, K.L. Proc.Natl.Acad. Sci.USA 76, 1770-1774 (1979).
14. Hudson, P., Haley, J., Cronk, M., Shine, J. and Niall, H. Nature 291, 127-131 (1981).
15. Brewer, H.B.,Jr., Fairwell, T., Larne, A., Ronan, R., Hauser, A. and Bronzert, T.J. Biochem.Biophys.Res.Commun. 80, 623-630 (1978).
16. Gait, M.J., Singh, M., Sheppard, R.C., Edge, M.D., Greene, A.R., Heathcliffe, G.R., Atkinson, T.C., Newton, C.R. and Markham, A.F. Nucl.Acids Res. 8, 1081-1096 (1980).
17. Baralle, F.E. Cell 12, 1085-1095 (1977).
18. Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. Cell 21, 621-626 (1980).
19. Wickens, M.P., Buell, G.N. and Schimke, R.T. J.Biol.Chem. 253, 2483-2495 (1978).
20. Messing, J. and Vieira, J. Analects 9 (8), 1 (1981).
21. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. J.Mol.Biol. 143, 161-178 (1980).
22. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. J.Mol.Biol. 113, 237-251 (1977).
23. Efstratiadis, A., Kafatos, F.C., Maxam, A.M. and Maniatis, T. Cell 7, 279-288 (1976).
24. Casadaban, M.J. and Cohen, S.N. J.Mol.Biol. 138, 179-207 (1980).
25. Southern, E.M. J.Mol.Biol. 98, 503-517 (1975).
26. Franceschini, G., Sirtori, C.R., Capurso, A., Weisgraber, K.H. and Mahley, R.W. J.Clin.Invest. 66, 892-900 (1980).
27. Utermann, G., Feussner, G., Francischini, G., Haas, J. and Steinmetz, A. J.Biol.Chem. 257, 501-507 (1982).
28. Dayhoff, M.O. Atlas of Protein Sequence and Structure Nat.Biomed.Res. Foundn., Washington (1976).
29. Baralle, F.E., Shoulders, C.C., Goodbourne, S., Jeffreys, A. and Proudfoot, N.J. Nucl.Acids Res. 8, 4393-4404 (1980).
30. Wallace, R.B., Johnson, M.J., Hirose, T., Miyake, T., Kawashima, E.H. and Itakura, K. Nucl.Acids Res. 9, 879-894 (1981).
31. Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G. and Maniatis, T. Cell 15, 1157-1174 (1978).