Volume 258, number 1, 71-74 FEB 07839 November 1989

# Direct sequencing of affinity-captured amplified human DNA application to the detection of apolipoprotein E polymorphism

Ann-Christine Syvänen, Katriina Aalto-Setälä*, Kimmo Kontula* and Hans Söderlund

*Orion Corporation, Genetic Engineering Laboratory, and *Institute of Biotechnology, University of Helsinki, Helsinki, Finland*

Received 27 September 1989

We describe a method for the direct sequencing of DNA amplified by the polymerase chain reaction (PCR). Biotin is introduced into one strand of the amplified DNA using a 5'-biotinylated PCR primer. The synthezised fragment is captured on an avidin-matrix and rendered single stranded, whereafter the nucleotide sequence of the immobilized strand is determined by the chain termination method. The method involves few and simple operations and is thus applicable to the analysis of human genes for routine diagnostic purposes. Here we applied the method for determination of the three-allelic polymorphism of the apolipoprotein E (apo E) gene. We were able to correctly identify the alleles in both homozygous and heterozygous samples.

Solid phase nucleotide sequencing; Genetic diagnosis; Biotin-avidin; Polymerase chain reaction; Apolipoprotein E

## 1. INTRODUCTION

The knowledge of the exact molecular defects causing inherited diseases, as well as predisposition to genetic disorders and cancer is increasing rapidly. Diagnostic methods for the direct detection of mutations in the human genes are consequently becoming increasingly important.

The polymerase chain reaction (PCR) technique [1] allows the enrichment of specific DNA fragments from the human genome. Mutations may be detected by analysis of restriction site variation in the amplified fragments [2,3], by hybridization with allele-specific oligonucleotide probes [4] or by nucleotide sequencing. Sequence determination is the most informative approach to genetic analysis, but the methods used for the direct sequencing of amplified DNA [5-8] involve manipulations that are difficult to carry out on a routine basis.

This communication describes a convenient method for the sequencing of PCR-amplified DNA fragments. In this technique biotin is introduced into the 5'-end of the DNA fragments synthesized during the PCR [9]. The fragments are then captured on an avidin-matrix and directly sequenced on this solid support. We applied the method to the analysis of the genetic variation of the apolipoprotein E (apo E). This protein plays an important role in the lipoprotein metabolism: it is both an integral component and a mediator of cellular uptake of several lipoproteins [10]. Three common isoforms of apo E (E2, E3 and E4), encoded by 3 different alleles ($\epsilon 2$, $\epsilon 3$ and $\epsilon 4$), exist in the population. This genetic polymorphism of apo E may explain as much as 10% of the individual variation of serum cholesterol levels [11] and is thus of major clinical significance. The fact that the three-allelic polymorphism of apo E is due to single base substitutions at two DNA loci only 138 base pairs apart from each other, makes it a particularly favorable target for PCR analysis.

*Correspondence address:* A.-C. Syvänen, Orion Corporation, Genetic Engineering Laboratory, Valimiotie 7, SF-00380 Helsinki, Finland

## 2. MATERIALS AND METHODS

### 2.1. *DNA*

Venous blood samples were obtained from patients attending the Lipid Outpatient Clinic of the University Central Hospital of Helsinki. Apo E phenotyping was accomplished by isoelectric focusing [12]. Leukocytic DNA was extracted according to Bell et al. [13].

### 2.2. *Oligonucleotides*

The four primers (P1-P4) were synthesized on an Applied Biosystems 381A DNA synthesizer [14]. A 5'-amino group was added to the PCR primers P2 and P3 with the amino-link II reagent (Applied Biosystems). The amino group was biotinylated using sulfo-NHS-biotin (Pierce Chemical Co.) and purified by reversed-phase HPLC [15]. The sequencing primer (P2 or P3) was labelled with [$\gamma$-$^{32}$P]dATP and T4 polynucleotide kinase [16] to a specific activity of $4 \times 10^6$ cpm/pm.

### 2.3. *Polymerase chain reaction*

The DNA (100 ng per sample) was amplified with the P1 and P4 primers (final concentration 1 $\mu$M) in 100 $\mu$l of a solution of 0.2 mM each of dATP, dCTP, dGTP, dTTP, 20 mM Tris-HCl, pH 8.8, 15 mM $(NH_4)_2SO_4$, 1.5 mM $MgCl_2$, 0.1% Tween 20, 0.1 mg/ml gelatin and 2.5 units of *Thermus aquaticus* DNA-polymerase (Promega or United States Biochemical Corp.) in a DNA thermal cycler

*Published by Elsevier Science Publishers B.V. (Biomedical Division)*
00145793/89/$3.50 

(Perkin-Elmer/Cetus) for 25 cycles of 1 min at 96°C and 2 min at 65°C. A small aliquot (3 μl of a 1:100 dilution) of this first PCR amplification mixture was transferred to a second PCR. This was carried out at the conditions described above and directed by a pair of nested primers (P2 and P3) one of which was 5'-biotinylated.

### 2.4. *Affinity-capture*

A 25 μl aliquot of the second amplification mixture was diluted to 50 μl with 0.15 M NaCl, 20 mM Na-phosphate buffer, pH 7.5 (PBS), after which 5 μl of a 5% (w/v) suspension of avidin-coated polystyrene particles (0.8 μm, Baxter Healthcare Corp.) were added. The samples were kept at 20°C for 1 h. The particles were collected by centrifugation for 2 min in an Eppendorf centrifuge and were washed twice by vortexing with 1 ml of 15 mM NaCl, 1.5 mM Na-citrate, and twice with 1 ml of 0.1% Tween 20 in PBS. The particles were treated twice with 0.15 M NaOH for 15 min, at 37°C, followed by two washes with 0.1% Tween 20 in 50 mM NaCl, 40 mM Tris-HCl, pH 7.5 and a final wash with 0.01% Tween 20 in 50 mM NaCl, 40 mM Tris-HCl, pH 7.5.

### 2.5. *Nucleotide sequencing*

The chain termination method of Sanger [17] was applied. The particles carrying the DNA template were suspended in 10 μl of 50 mM NaCl, 20 mM $MgCl_2$, 40 mM Tris-HCl, pH 7.5, containing 0.5–1 pmol of the sequencing primer P2 or P3. The tubes were heated at 65°C for 2 min and allowed to cool slowly to room temperature. One μl of 0.1 M dithiothreitol, 2 μl of 0.75 μM dCTP, dGTP, dTTP, 0.5 μl of [$\alpha$-$^{35}$S]dATP (>1000 Ci/mmol, Amersham) and 2 μl (3.25 units) of T7 DNA-polymerase (Sequenase, United States Biochemical Corp.) were added and the tubes were kept at 22°C for 3 min. When using an end-labelled primer the labelling step was omitted by substituting the dNTPs with 2.5 μl of $H_2O$. An aliquot (3.5 μl) of the reaction mixture was transferred to 4 tubes containing 2.5 μl of chain termination mixture (80 μM dNTP, 8 μM ddNTP for the A and T reactions; 320 μM dNTP, 8 μM ddNTP for the C and G reactions). The tubes were incubated at 42°C for 6 min and the reaction was stopped by adding 4 μl of a solution containing 95% formamide, 20 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol. The samples were denatured at 80°C for 2 min and the particles were spun down for 2 min. The products of the chain termination reactions were run on 6% sequencing gels for 2.5 h at 2500 V. The sequencing ladders were visualized by autoradiography (Kodak XAR film).

## 3. RESULTS AND DISCUSSION

### 3.1. *Principle of the method*

The method described here is based on PCR amplification of the nucleotide sequence of interest using a primer biotinylated in its 5'-end, which results in the production of 5'-biotinylated DNA fragments [9]. The amplified fragments are captured on a solid matrix taking advantage of the biotin-avidin interaction. The non-biotinylated strand of the captured DNA is released by denaturation and the nucleotide sequence of the remaining immobilized strand is determined by the chain termination method (fig.1). This approach facilitates purification of the amplified DNA from the PCR mixture, eliminating the interference caused by primers and the dNTPs in the sequencing reaction.

### 3.2. *Sequence analysis of the apolipoprotein E gene*

The method was applied for the identification of the genetic polymorphism of apo E. This polymorphism is due to single base substitutions at two loci in the coding area of the apo E gene (fig.2). The allelic frequencies of

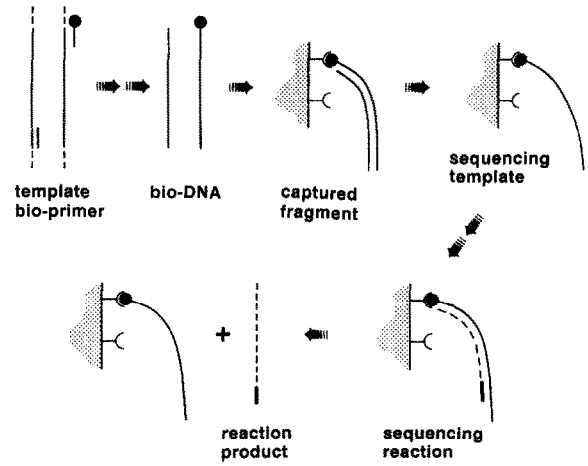
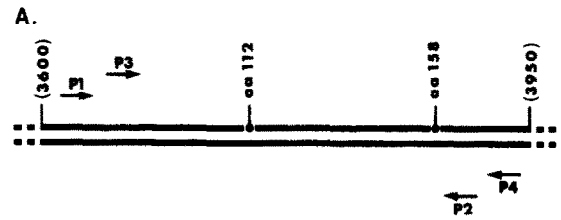



Fig.1. Principle of the method. bio = biotinylated.

$\epsilon$2, $\epsilon$3 and $\epsilon$4 in various populations are approximately 10%, 75% and 15%, respectively [11]. More than 90% of patients with type III hyperlipoproteinemia, a disorder characterized by xanthomatosis, premature




B.

| | | | | |
|---|---|---|---|---|
| ε2 | Cys | Cys | | |
| | TGC | TGC | | |
| ε3 | Cys | Arg | | |
| | TGC | CGC | | |
| ε4 | Arg | Arg | | |
| | CGC | CGC | | |

C.

| PRIMER | SEQUENCE | NUCLEOTIDE NUMBERS |
|---|---|---|
| P1 | 5'-AAGGAGTTGAAGGCCTACAAAT | 3615 - 3637 |
| P2 | 5'-TCGCGGGCCCCGGCCTGGTACA | 3893 - 3914 |
| P3 | 5'-GAACAACTGAGCCCGGTGGCGG | 3649 - 3671 |
| P4 | 5'-GGATGGCGCTGAGGCCGCGCTC | 4012 - 4034 |

Fig.2. (A) The positions of the primers along the fourth exon of the apolipoprotein E gene. The positions of amino acids 112 and 158 are also shown. The numbers in parentheses refer to the nucleotide sequence [18]. The drawing is in scale. (B) Variable amino acid and nucleotide sequences of the alleles $\epsilon$2, $\epsilon$3 and $\epsilon$4 (the corresponding protein phenotypes are designated as E2, E3 and E4, respectively). (C) Nucleotide sequence of the primers. P2 was 5'-biotinylated and P3 was used as the sequencing primer. For sequencing the opposite strand of the DNA P3 was 5'-biotinylated and P2 was the sequencing primer. Nucleotide numbering is according to Paik et al. [18].

atherosclerosis and hyperlipoproteinemia, are homozygous for the $\epsilon 2$ allele [11]. On the other hand the phenotype E4 (encoded by the $\epsilon 4$ allele) seems to be associated with elevated serum cholesterol levels and increased risk of atherosclerosis at the population level [11]. Serum apo E phenotyping can be carried out by two-dimensional electrophoresis or isoelectric focusing combined with immunoblotting [11]. These techniques are laborious and do not permit analysis of rare (other than E2, E3 and E4) apo E variants known to exist [10].

The present procedure was optimized to give clear sequencing ladders. The homogeneity of the DNA fragment to be sequenced is the most crucial factor of the procedure. Low efficiency due to amplification of non-specific DNA is a frequent problem when amplifying fragments of genomic DNA. In our method, this was circumvented by carrying out two consecutive PCR amplifications with a nested set of primers in the second amplification [1]. The primers (fig.2) were chosen from the known sequence of the apo E gene [18]. The primary primers P1 and P4 were 328 base pairs apart on the apo E gene. The nested primers amplified a 265 base pair fragment over the region coding for amino acids 112 and 158 (fig.2).

The sequencing reactions were carried out on the immobilized DNA template essentially as for templates in solution. The amount of DNA produced in the PCR amplification was 0.5–5 pmol per reaction as estimated by the affinity-based hybrid collection method [9]. One-tenth to one-fourth of this amount was optimal for a sequencing reaction with 0.5–1 pmol of the sequencing primer. The label was introduced by primer elongation with $^{35}$S-labelled dATP prior to the termination reactions (fig.3) or alternatively using a terminally $^{32}$P-labelled sequencing primer (fig.4). In the latter case separate labelling and termination reactions in the sequencing protocol were not required. The fact that the template was immobilized on microparticles did not adversely affect the sequencing reactions. The primer annealing efficiency was comparable to that in solution. The same dNTP/ddNTP ratios could be used in both cases.

The sequence adjacent to the sequencing primer was readable as well as the sequence up to 5–7 nucleotides from the 5′-biotin residue attaching the DNA fragment to the microparticle. The codons for the amino acids 112 and 158 in apo E are located 75 and 213 nucleotides, respectively, from the P3 sequencing primer. Fig.3 shows a sequencing ladder over this region. The sequence corresponds to the homozygous phenotype E2E2, in which the amino acid residues 112 and 158 both are cysteines. Fig.4 shows a determination of the base variation at amino acid position 112 in two samples of the phenotypes E3E3 and E3E4. In this case the opposite strand of the DNA was analyzed. In the homozygous sample both alleles code for cysteine residues at position 112 and in the heterozygous sample

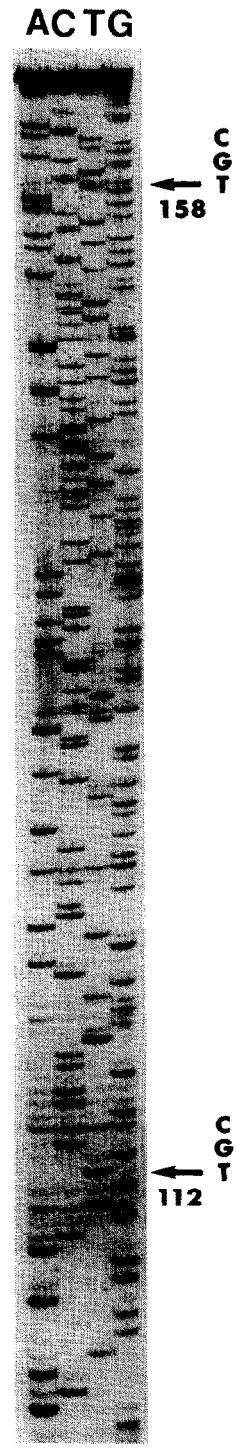



Fig.3. Autoradiograph obtained from sequence analysis of immobilized, amplified apolipoprotein E DNA. The arrows indicate the polymorphic DNA loci. The genotype of the subject is $\epsilon 2\epsilon 2$ (phenotype E2E2).

the sequence corresponding to both an arginine and a cysteine residue at position 112 can be read.

The present technique allows typing of all possible combinations of the three alleles. Variable sites, other than amino acids 112 and 158 of apo E, clustering in the

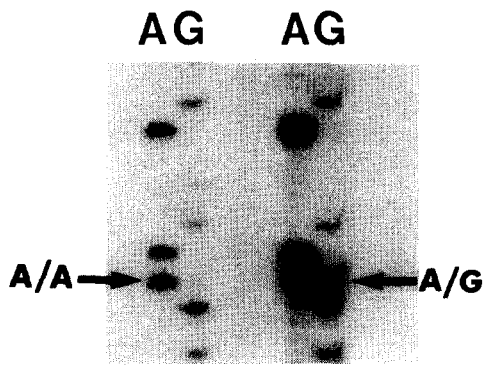



Fig.4. Autoradiographs obtained from sequence analysis of the locus coding for amino acid residue 112 in apolipoprotein E from two samples with the phenotypes E3E3 (left) and E4E4 (right). The relevant lanes (A and G) of the sequencing gel are shown. The arrows indicate the base responsible for the genetic polymorphism at amino acid position 112. A/A corresponds to the $\epsilon 3$ homozygote and A/G corresponds to the $\epsilon 3\epsilon 4$ heterozygote.

same region of the molecule have occasionally been reported [10]. This technique makes feasible the detection of such rare mutations as well as the identification of previously unknown mutations.

## 4. CONCLUSION AND PROSPECTS

Direct nucleotide sequence data allow an unequivocal diagnosis of a genetic defect. The method described in the present communication is suitable for automation. The PCR cycles are carried out automatically. Capturing of the amplified DNA on a solid support comprises a simple and efficient single-step purification of the DNA from the PCR mixture. Solid phase techniques are applicable to automatic sequencing procedures, as shown recently by Hultman et al. and Voss et al. [19,20]. We conclude that nucleotide sequencing of affinity-captured PCR-amplified DNA fragments is a promising approach to the routine diagnosis of genetic disorders.

*Acknowledgements:* We thank Anne Järvinen for excellent technical assistance, Marina Bengtström and Maritta Putkiranta for the synthesis and biotinylation of the oligonucleotides and Jonathan Knowles and Anu Koivula for helpful discussions. This work was supported by grants from the Finnish Academy of Sciences and the University of Helsinki (K.A.-S.). Patent applications relating to the method described here are pending.

## REFERENCES

[1] Mullis, K.B. and Faloona, F.A. (1987) Methods Enzymol. 155, 335–350.
[2] Saiki, R.K., Scharf, S., Falcona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N. (1985) Science 230, 1350–1354.
[3] Kogan, S.C., Doherty, M. and Gitschier, J. (1987) N. Engl. J. Med. 317, 985–990.
[4] Saiki, R.K., Bugawan, T.L., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1986) Nature 324, 163–166.
[5] Wong, C., Dowling, C.E., Saiki, R.K., Higuchi, R.G., Erlich, H.A. and Kazazian, H.H. (1987) Nature 330, 384–386.
[6] Engelke, D.R., Hoener, P.A. and Collins, F.S. (1988) Proc. Natl. Acad. Sci. USA 85, 544–548.
[7] Gyllensten, U.B. and Erlich, H.A. (1988) Proc. Natl. Acad. Sci. USA 85, 7652–7656.
[8] Mitchell, L.G. and Merril, C.R. (1989) Anal. Biochem. 178, 239–242.
[9] Syvänen, A.-C., Bengtström, M., Tenhunen, J. and Söderlund, H. (1988) Nucleic Acids Res. 16, 11327–11338.
[10] Mahley, R.W. (1988) Science 240, 622–630.
[11] Davignon, J., Gregg, R.E. and Sing, C.F. (1988) Arteriosclerosis 8, 1–21.
[12] Ehnholm, C., Lukka, M., Kuusi, T., Nikkilä, E. and Utermann, G. (1986) J. Lipid Res. 27, 227–235.
[13] Bell, G.I., Karam, J.H. and Rutter, W.J. (1981) Proc. Natl. Acad. Sci. USA 78, 5759–5763.
[14] Beaucage, S.L. and Caruthers, M.H. (1982) Tetrahedron Lett. 22, 1859–1862.
[15] Bengtström, M., Jungell-Nortamo, A. and Syvänen, A.-C. (1989) Nucleosides Nucleotides, in press.
[16] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–523.
[17] Sanger, F., Miklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[18] Paik, Y.-K., Chang, D.J., Reardon, C.A., Davies, G.E., Mahley, R.W. and Taylor, J.M. (1985) Proc. Natl. Acad. Sci. USA 82, 3445–3449.
[19] Hultman, T., Ståhl, S., Hornes, E. and Uhlén, M. (1989) Nucleic Acids Res. 17, 4937–4946.
[20] Voss, H., Schwager, C., Wirkner, U., Sproat, B., Zimmermann, J., Rosenthal, A., Erfle, H., Stegemann and Ansorge, W. (1989) Nucleic Acids Res. 9, 2517–2527.