Volume 17 Number 7 1989 Nucleic Acids Research

Direct genomic fluorescent on-line sequencing and analysis using *in vitro* amplification of DNA

Hartmut Voss, Christian Schwager, Ute Wirkner[1], Brian Sproat, Jürgen Zimmermann, Andre Rosenthal[2], Holger Erfle, Josef Stegemann and Wilhelm Ansorge*

European Molecular Biology Laboratory, Postfach 102209, 6900 Heidelberg, [1]German Cancer Research Center, Experimental Pathology, Im Neuenheimer Feld 280, 6900 Heidelberg, FRG and [2]Academy of Science of GDR, Institute of Molecular Biology, GDR-1115 Berlin-Buch, GDR

Received February 15, 1989; Accepted March 1, 1989

ABSTRACT

In vitro amplification of genomic DNA and total RNA, as well as recombinant DNA, using one fluorescently labelled and one unlabelled primer during amplification, together with on-line analysis of the products on the EMBL fluorescent DNA sequencer, is described. Further is reported direct sequencing of fluorescently labelled amplified probes by solid-phase chemical degradation, without subcloning and purification steps involved. At present up to 350 bases in 4 hours are determined with this technique. The fluorescent dye and its bond to the oligonucleotide are stable during the amplification cycles, and do not interfere with the enzymatic polymerization. High sensitivity of the detection device, down to 10^{-18} moles, corresponding to less than 10^6 molecules makes possible analyses of the non-radioactive amplified probes after only 10 amplification cycles, starting with about $5x10^4$ copies of recombinant DNA.

INTRODUCTION

Two of the important recent developments in molecular biology techniques - *in vitro* amplification of DNA fragments via polymerase chain reaction (PCR) (1,2,3) and on-line DNA sequencing using fluorescent labels (4,5,6) - have found an increasing number of applications. Amplification of DNA is used in clinical diagnostics of sickle-cell anemia (7), β-thalassemia (8), hemophilia A (9) and Duchenne muscular dystrophy (10), in forensic analysis (11), DNA fingerprinting (12) and study of DNA-protein interactions (13). Automated fluorescent DNA sequencers have increased the reliability and speed in obtaining sequence data. Many aspects of the sequencing method have been adapted to non-radioactive sequencing, including the enzymatic technique (14), double stranded sequencing (15), chemical degradation (16) and investigation of the kinetics of DNA polymerases (17).

Several strategies to sequence PCR products directly using radioactivity have been reported. They involve the use of a third sequencing primer between the flanking PCR primers (8) or asymetric amounts of PCR primers (18,19) to generate single stranded DNA suitable for dideoxy sequencing with T7- or Taq-Polymerase. In chemical methods end-labelled primers for Maxam-Gilbert degradation (20) and α-thiotriphosphates for PCR (21) have been used. All the reported direct sequencing methods require purification steps after product

amplification to remove impurities, primarily excess oligonucleotide primers and unincorporated deoxynucleotides, before sequencing reactions are carried out.

In this paper we describe *in vitro* amplification of genomic DNA and total RNA, as well as recombinant DNA, using one fluorescently labelled and one unlabelled primer during amplification, together with on-line analysis of the products using our fluorescent DNA sequencer. We further report direct sequencing of fluorescently labelled PCR products by solid-phase chemical degradation, without subcloning and purification steps involved.

EXPERIMENTAL

Materials

Genomic DNA and total RNA were purified from human placenta as described (22,23) in collaboration with W. Pyerin (German Cancer Research Center, Heidelberg). *Taq* Polymerase, T7 DNA polymerase and nucleotides were from Pharmacia PL. Restriction enzymes were from Boehringer Mannheim. Oligonucleotides complementary to human P450arom cDNA (24) together with universal and reverse M13 sequencing primers were synthesized on a multiple DNA synthesizer developed at EMBL (25), allowing simultaneous synthesis of ten or more oligonucleotides. The following primers were used (fluorescent label is indicated by an asteristik *):

ARO1	5'-TTGGTCATGCGCAAAGCCTTAGGA-3'
ARO2	5'-TACTTTCAGCCATTTGGCTTTGGG-3'
ARO3	5'-TATGCTCTCAACACACTGTCCTT-3'
ARO4	5'-GTGTTCCAGACACCTGTCTGAGCTTCTTGGGGTAAAGATCATTTCCAGCATGT-3'
ARO4*	5'-*GTGTTCCAGACACCTGTCTGAGCTTCTTGGGGTAAAGATCATTTCCAGCATGT-3'
UNISEQ	5'-CGTTGTAAAACGACGGCCAG-3'
UNISEQ*	5'-*CGTTGTAAAACGACGGCCAG-3'
REVSEQ	5'-AACAGCTATGACCATG-3'
REVSEQ*	5'-*AACAGCTATGACCATG-3'

Fluorescent oligonucleotides were prepared by labelling with fluorescein as described (5). Primers ARO4* and ARO4, fluorescently labelled respective unlabelled, were used in the same experiments for control. For PCR amplifications we used a thermal cycler constructed at EMBL. The orientation of the PCR primers and an internal *Xho*I restriction site are shown in Fig.1.

Methods

1. PCR with unlabelled oligonucleotides

About 5μg of total RNA from human placenta were used for first strand synthesis with 0.1μM of ARO3 or ARO4, 20 units AMV reverse transcriptase (Life sciences), 10x reaction buffer in a total volume of 20μl. Reaction was for 45 min at 42° C.

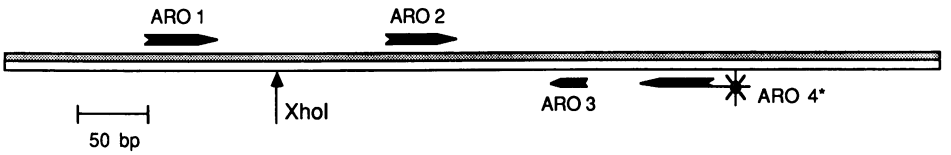


Fig.1 Diagrammatic representation of part of human P450arom gene locus with locations and directions of PCR primers and internal XhoI restriction site used in the experiments (ARO - unlabelled primer, ARO* - fluorescently labelled primer).

For *in vitro* amplification by PCR 10μg RNase A, 10x PCR buffer (670mM Tris-HCl pH 8.8, 67mM $MgCl_2$,166mM $(NH_4)_2SO_4$, 100mM β-mercaptoethanol, 67mM EDTA, 1.7mg/ml BSA), 250μM dNTP, 10% DMSO, 1% gelatine, 0.1μM of the opposite primer ARO1 or ARO2 and 2,5 units Taq polymerase were added to a total volume of 100μl. The solution was overlaid with 100μl mineral oil (Sigma) and subjected to 36 cycles (one cycle consisting of 1 min denaturing at 94° C, 2 min annealing at 40° C and 3 min polymerisation at 72° C). Final polymerisation step after the 36th cycle was for 10 min at 72° C. The products were precipitated, kinased and ligated into *Sma*I cut dephosphorylated Bluescript KS+ vector (Stratagene). Positive clones were isolated and sequenced by double stranded dideoxy method with T7 Polymerase.

In vitro amplification using 1μg human placenta DNA and the same conditions as described above were performed with primer pairs ARO1 & ARO3, ARO1 & ARO4, ARO2 & ARO3, AR02 & ARO4. Amplification products were analysed on ethidium bromide stained agarose gels.

2. PCR with fluorescently labelled oligonucleotides

About 0.1μM of fluorescently labelled oligonucleotides were used for PCR in equimolar amounts to an unlabelled primer in the following combinations, using the same conditions as described above: ARO1 & ARO4*, ARO2 & ARO4*. PCR was performed on genomic DNA as well as on total RNA. The first strand was synthesized with ARO4*. For amplification of recombinant DNA we used 1ng from the clones , as described above under 1, and primers UNISEQ & REVSEQ*. Aliquots of all PCR samples were precipitated and assayed with several suitable restriction enzymes. 10% of the crude amplification products were analyzed on ethidium bromide stained agarose gels and 0.2% on the EMBL automated sequencer (see below).

3. Detection range with fluorescent labelled PCR products

10ng/μl respective 10pg/μl were used for PCR as described previously with labelled universal and reverse primer. 5μl aliquots were taken after 10, 15, 20 and 25 cycles, respectively, and analysed on the automated sequencer.

4. Solid phase chemical degradation of fluorescently labelled PCR products

Chemical degradation of fluorescently labelled PCR samples has been performed essentially as described (16) with slightly modified reaction times, using Hybond M & G paper

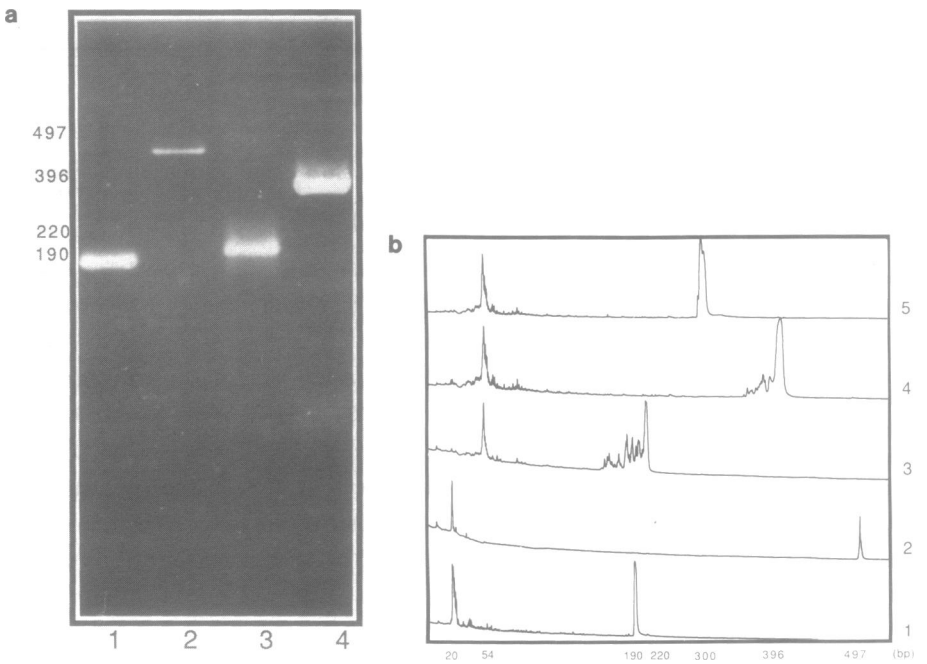


Fig.2 PCR with fluorescently labelled oligonucleotides.
(a): Ethidium bromide stained 2% agarose gel of PCR products after electrophoresis as described in materials and methods. 10% of each assay was loaded. Lane 1: PCR from 1ng plasmid Bluescript KS+ with primers UNISEQ & REVSEQ*. Lane 2: PCR from 1ng plasmid bsARO13R. Lane 3: PCR from 1µg human placenta DNA with primers ARO2 & ARO4*. Lane 4: PCR from 5µg human placenta total RNA with primers ARO1 & ARO4*.
(b): Partial raw data output from a 5% polyacrylamide gel of PCR products analyzed on the automated DNA sequencer. Lane 1 - 4 : 0.2% of PCR products loaded, corresponding to Fig.2a. Lane 5: 2% of PCR shown in lane 4 were restriction-digested with *Xho*I, 10% of this assay were loaded.

(Amersham). About 10% of the unpurified PCR product were applied to the carrier. For degradation of genomic DNA fragments the following reaction conditions were used:

G	:	1% DMS in 50mM ammonium formate buffer pH 3.5 for 2 min
A+G	:	66% formic acid for 20 min
T>Pu	:	0.1mM potassium permanganate for 5 min
C	:	4M hydroxylamine pH 6 for 10 min

After piperidine reaction and lyophilization the samples were dissolved in 10µl 50% of aqueous formamide. This procedure allowed us to resolve up to 350 bases.

5. Sequence determination and detection of amplified DNA fragments

The sequence of the fluorescently labelled PCR products derived from total RNA, genomic DNA and recombinant DNA was determined on-line in the EMBL automated DNA sequencing system

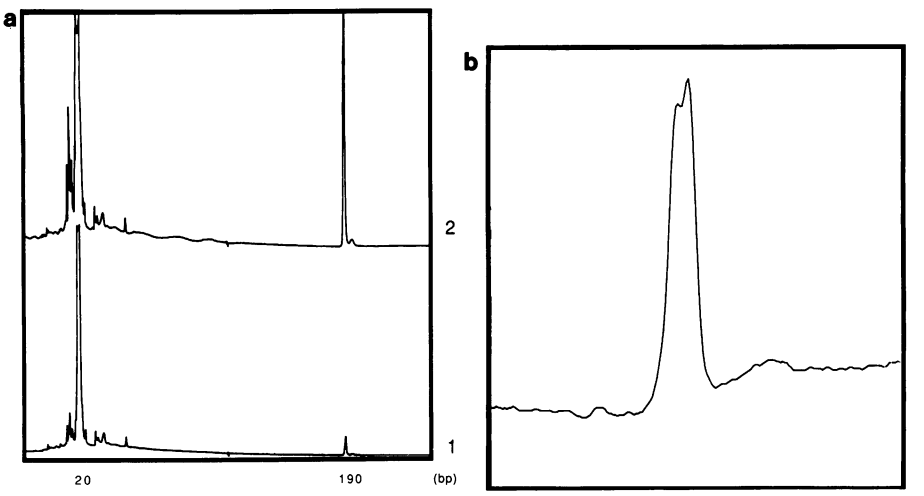
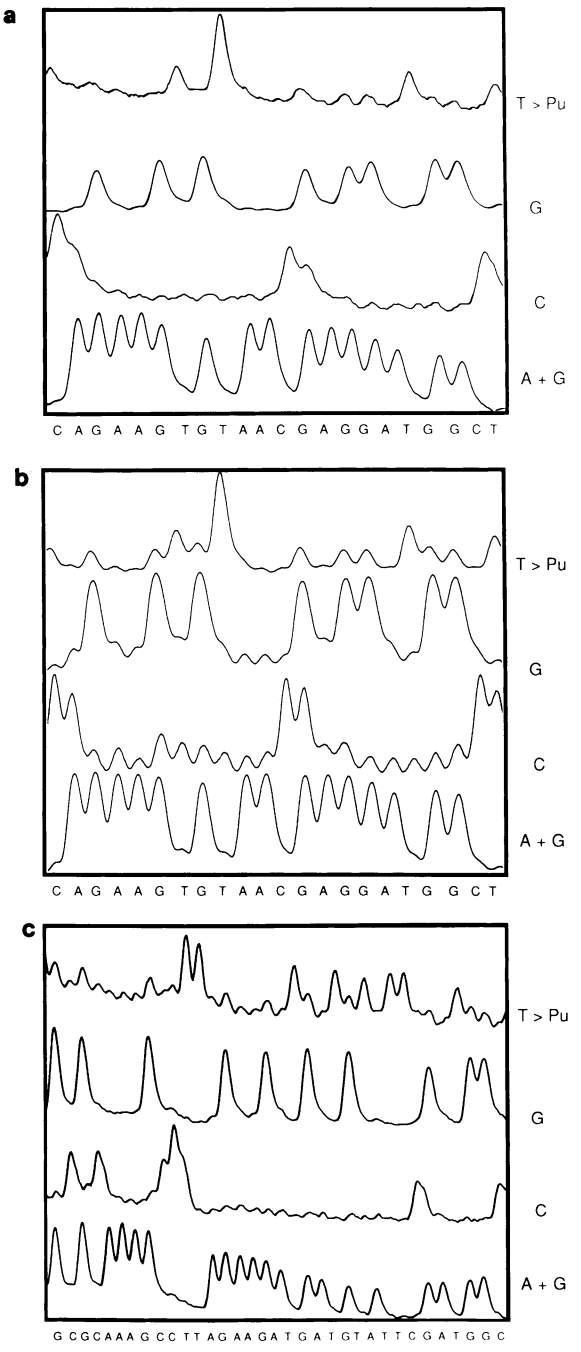


Fig.3 Sensitivity test for PCR, using fluorescently labelled primers for amplification.
(a): Partial raw data output from 5% polyacrylamide gel. Starting material for PCR was $5x10^7$ copies ($=10^{-16}$ moles) of plasmid Bluescript KS+ and primers UNISEQ & REVSEQ*. Lane 1: 5% of PCR assayed after 10 amplification cycles. Lane 2: 5% of PCR assayed after additional 5 cycles (in total 15 cycles).
(b): Starting material $5x10^4$ copies ($=10^{-19}$ moles) of plasmid Bluescript KS+ and primers UNISEQ* & REVSEQ*. 5% of the PCR assay was loaded after 10 amplification cycles, corresponding to approximately $5x10^7$ copies detected on the gel.

described previously (5,14). About 20% of each sequencing reaction product were run on a 7% polyacrylamide gel containing 8M urea. Products of the *in vitro* DNA amplification were detected on-line in the same device.

RESULTS

As shown in Fig. 1, using two unlabelled oligonucleotides for priming we amplified the expected fragments with ARO1 & ARO3 (300 bp), ARO1 & ARO4 (396 bp) and ARO2 & ARO4 (220 bp) from total RNA, with ARO2 & ARO4 (220 bp) from genomic DNA (data not shown.). With the fluorescently labelled primer ARO4*, we obtained products according to the expected length from both the total RNA and genomic DNA (Fig.2a). Amplification of recombinant Bluescript clone bsARO13R with REVSEQ* & UNISEQ primer resulted in the expected 496 bp product. When compared to control experiments with unlabelled oligonucleotides no difference in length or yield of DNA was observed, (data not shown), implying that *in vitro* DNA amplification with fluorescently labelled oligonucleotides is as efficient as amplification with unlabelled primers. Analyzing the PCR products on a 5% polyacrylamide gel the peaks were found in the positions corresponding to the expected molecular weights of the fragments


a
T > Pu
G
C
A + G
C A G A A G T G T A A C G A G G A T G G C T
b
T > Pu
G
C
A + G
C A G A A G T G T A A C G A G G A T G G C T
c
T > Pu
G
C
A + G
G C G C A A A G C C T T A G A A G A T G A T G T A T T C G A T G G C

(Fig.2b). This indicates that the fluorescent dye and its bond to the oligonucleotide were stable during repeated cycles of denaturation, annealing and polymerization, and did not interfere with the extension by polymerization. In no experiments did we observe products longer than the expected full length product.

With primers purified on a gel, lanes 1 and 2 in Fig.2b show two peaks per lane, one corresponding to the unincorporated labelled primer excess and the other to the full length amplification product. Lanes 3 and 4 in Fig.2b display results obtained in amplifications with gel purified labelled primer, while the unlabelled primer was not gel purified. The presence of shorter fragments produced by the amplification is evident from the presence of smaller secondary peaks preceding the main peak corresponding to the full length product. The higher resolution of fluorescent analysis on polyacrylamide gels as compared to the separation of the same sample on agarose gels in Fig.2a, is obvious. The presence of shorter fragments in the final product does not interfere with further analysis, e.g. direct sequencing or restriction digest, because all fragments have in common the labelled 5´ end.

An aliquot of 5% of the ethanol precipitated amplification product from ARO1 & ARO4* was digested with *Xho*I, and 1/10th of the amount was analysed on the automated sequencer. The result was the expected 300 bp fragment (Fig.2b, lane 5), as compared to the 396 bp fragment uncut control (Fig.2b lane 4). The sequence of the restriction site was confirmed by direct sequencing at that position (data not shown).

Fig.3 demonstrates the sensitivity with which the fluorescently labelled PCR products are detected on the automated device. Fig.3a shows amplification products after 10 and 15 cycles, starting with approximately 10^7 copies of Bluescript KS+ plasmid DNA. Numerical integration of both peaks showed, that 5 cycles PCR yielded an amplification of about 30 fold. The high sensitivity of the sequencer (14), down to 10^{-18} moles, enables us to detect at least 10^6 copies. Starting with about $5x10^4$ copies for PCR the amplification results in $5x10^7$ copies after 10 cycles . These amounts are clearly detected as a peak (Fig.3b).

Three different unpurified amplification products were directly sequenced by solid phase chemical degradation, genomic DNA (Fig.4a), total RNA (Fig.4b) and recombinant DNA (Fig.4c). Fig.4a and 4b display partial raw data of the same sequence derived from DNA and

Fig.4 Direct fluorescent sequencing from PCR products by solid phase chemical degradation.
(a): Partial raw data output of direct fluorescent sequencing of unpurified PCR product derived from genomic DNA with primers ARO2 & ARO4*, corresponding to base 90 - 112 after the priming site.
(b): Partial raw data output of the same sequence derived from unpurified PCR of total RNA with primers ARO1 & ARO4*.
(c):Partial raw data output of direct fluorescent sequencing of PCR product derived from recombinant plasmid DNA bsARO13R, corresponding to base 189 - 224 after the priming site.

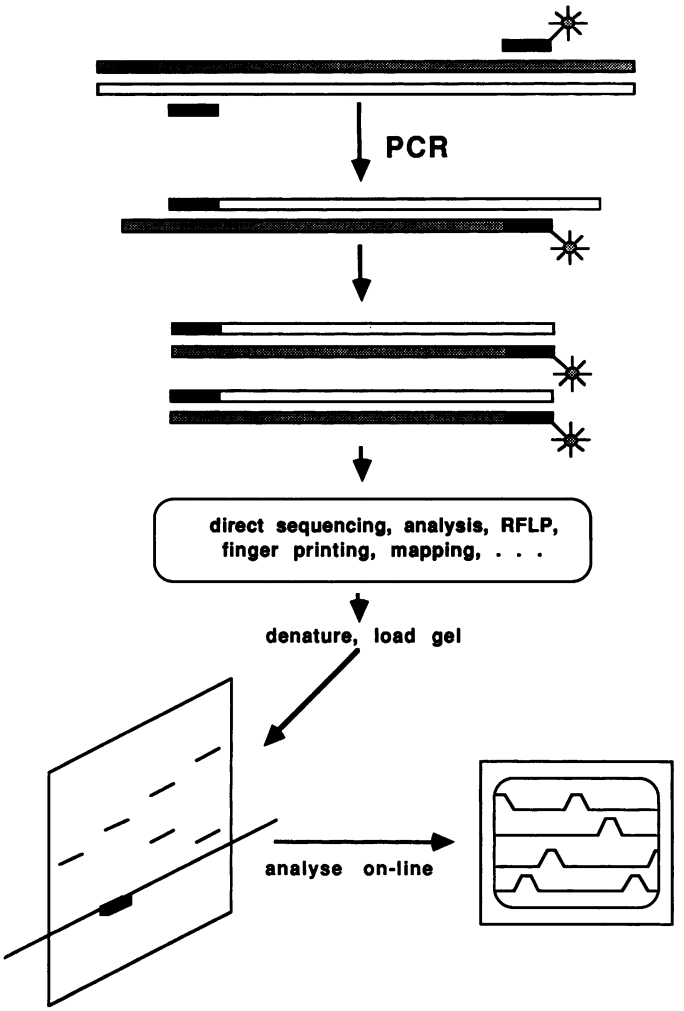


Fig.5 Scheme of the fluorescent DNA fragment analysis.
Via PCR desired DNA or RNA segment is amplified and fluorescently labelled. In the second step analysis of low amounts of PCR products in several applications is carried out on the EMBL automated fluorescent sequencer, functioning as a general on-line detector of fluorescently labelled DNA.

RNA amplification products of different length using identical sequencing reaction conditions. Although there are artefact signals after each real base in the C track, the sequence is readable regarding the rules of chemical degradation sequencing, T > Pu, A+G (16). These artifact signals, which we do not observe in sequencing of recombinant DNA, were found so far in every sequence of DNA fragments amplified by PCR, without influence on the correct sequence

reading. The unincorporated labelled primer and its degradation products usually result in several strong peaks at the beginning of the sequence. The reading starts immediately after that primer. At present up to 350 bases are determined on our automated system in direct genomic fluorescent sequencing with solid phase chemical degradation.

DISCUSSION

Detection of amplification products

In vitro amplification technology has become a powerful tool for fast and precise production of DNA fragments. The products of the amplification are usually detected on ethidium bromide stained agarose gels, but for many applications the sensitivity is not high enough and therefore radioactive probing is necessary to verify the results (10), particularly for allele-specific genetic diseases. Using fluorescently labelled oligonucleotides for PCR and subsequent analysis of the products on the EMBL automated fluorescent sequencer we detect the products on-line with a sensitivity comparable to that of radioactive blots. Using low amounts of uncloned PCR products it is now possible to analyse the products after only 5 to 15 amplification cycles. Because of the high sensitivity of the system, we can detect down to 1/200th part of a restriction enzyme digested PCR product. Hence applications in clinical diagnostics, where only tiny amounts of starting material are available, become possible, e.g. analysis of restriction fragment length polymorphisms (RFLPs).

High resolution of polyacrylamide gels gives more accurate information about length and relative amount of the amplified fragments than separations on agarose gels. This offers the possibility for detailed studies of the *in vitro* amplification technique and for improving the methodology, e.g. to find conditions for precise amplification of longer DNA fragments. With the EMBL automated device is feasible on-line analysis of fragments up to a few kb in length on polyacrylamide gels, and of even longer fragments on agarose gels. Up to 40 samples can be simultaneously analysed.

Direct genomic sequencing

At present the DNA sequencing projects require time and cost intensive subcloning steps. Amplification of inserts in lambda or plasmid vectors up to 2 kb length by PCR has been described (3) avoiding the need for subcloning and purification of recombinant DNA from micro-organisms. We have shown here that fluorescently labelled recombinant DNA obtained by PCR is suitable for direct sequencing without subcloning. Amplification products are labelled by primer extension with the respective fluorescently labelled oligonucleotides and sequenced as described here. Amplification of DNA by PCR with equal amounts of flanking primers is easy to perform and the DNA fragments obtained in high yield can be sequenced by solid phase chemical degradation without any purification steps. Many samples are handled simultaneously and the method can be automated.

The scheme in Fig.5 summarizes the analysis method combining in vitro amplification by PCR and fluorescent detection for several applications, e.g. direct genomic sequencing, restriction analysis and genome mapping. The final products are analysed on the EMBL automated fluorescent sequencer, which is here used as a general on-line detector of fluorescently labelled DNA.

ACKNOWLEDGEMENTS

We thank Professor L. Philippson for support during the course of the project, and Toby Gibson for helpful discussion of the manuscript.

*To whom correspondence should be addressed

REFERENCES

1. Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N. (1985), Science **230**, 1350 - 1354.
2. Mullis K.B.and Faloona, F.A.(1987), Meth.Enzym. **155**, 335 - 350.
3. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B.and Erlich, H.A. (1988), Science **239**, 487 -491.
4. Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Conell, C.R., Heiner, C., Kent, S.B.H. and Hood, L.E. (1986), Nature **321**, 674 - 679.
5. Ansorge, W, Sproat, B.S., Stegemann and Schwager, C. (1986), J. Biochem.Biophys.Meth. **13**, 315 - 323.
6. Prober, J.M., Trainor, G.L., Dam, R.J., Hobbs, F.W., Robertson, C.W., Zagursky, R.J., Cocuzza, A.J., Jensen, M.A. and Baumeister, K. (1987) Science **238**, 336 - 341.
7. Saiki, R.K., Bugawan, T.L., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1986) Nature **324**, 163 -165.
8. Wong, C.,Dowling, C.E., Saiki, R.K.,Higuchi, R.G., Erlich, H.A., Kazazian, H.H. (1987) Nature **330**, 384 - 386.
9. Kogan, S.C., Doherty, M. and Gitschier, J. (1987) N.Eng.J.Med. **317**, 985 - 990.
10. Chamberlain, J.S., Gibbs, R.A., Ranier, J.E., Nguyen, P.N. and Caskey, C.T. (1988), Nucleic Acid Res. **16**, 11141 - 11156.
11. Higuchi, R., v.Beroldingen, C.H., Sensabough, G.F. and Erlich, H.A. (1988) Nature **332**, 543 - 546.
12. Jeffreys, A.J., Wilson, V., Neumann, R. and Keyte, J. (1988), Nucleic Acid Res. **16**, 10953 - 10971.
13. Higuchi, R., Krummel, B. and Saiki, R.K. (1988), Nucleic Acid Res. **16**, 7351 - 7367.
14. Ansorge, W., Sproat, B.S., Stegemann, J., Schwager, C. and Zenke, M. (1987) Nucleic Acid Res. **15**, 4593 - 4602.
15. Kim, H.-R.C., Yew, N.S., Ansorge, W., Voss, H., Schwager, C., Vennström, B., Zenke, M. and Engel, J.D. (1988), Mol.Cell.Biol. **8**, 4416 - 4424.
16. Ansorge, W., Rosenthal, A., Sproat, B.S., Schwager, C., Stegemann, J. and Voss, H. (1988) Nucleic Acid Res. **16**, 2203 - 2206.
17. Kristensen, T., Voss, H., Schwager, C., Stegemann, J., Sproat, B.S. and Ansorge, W. (1988) Nucleic Acid Res. **16**, 3487 - 3496.
18 Gyllenstein, U.B., Erlich, H.A. (1988), Proc.Natl.Acad.Sci.USA **85**, 7652 - 7656.
19. Innis, M.A., Myambo, K.B., Gelfand, D.H. and Brow, M.A.D. (1988), Proc.Natl.Acad.Sci.USA **85**, 9436 - 9440.
20. Cariello, N.F., Scott, J.K., Kat, A.G., Thilly, W.G. and Keohavong, P. (1988) Am.J.Hum.Genet. **42**, 726 - 734.

21. Nakamaye, K.L., Gish, G., Eckstein, F. and Vosberg, H.P. (1988) Nucleic Acid Res. **16**, 9947 - 9959.
22. Blin, N. and Stafford, D.W. (1979) Nucleic Acid Res. **3**, 1799 - 1814.
23. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochem. **18**, 5294 - 5299.
24. Corbin, C.J., Graham-Lorence, S., McPhaul, M., Mason, J.I., Mendelson, C.R. and Simpson, I.R. (1988), Proc.Natl.Acad.Sci.USA **85**, 8948 - 8952.
25. Sproat, B., Schwager, C. and Ansorge, W., A new multiple segmental DNA synthesizer, in preparation.