Pure & Appl. Chem., Vol. 68, No. 10, pp. 1907–1912, 1996.
Printed in Great Britain.

Recent developments in automatic DNA sequencing

A.F.BRADLEY

Perkin-Elmer Hispania S.A., Applied Biosystems Division, Ronda de Poniente 5, Tres Cantos, 28760 Madrid, Spain.

Abstract: Automatic DNA sequencing and fragment analysis applications have increased dramatically during the last few years. We describe here the introduction of a new unattended operation DNA sequencer based on capillary electrophoresis technology and simultaneous fluorescence detection for a variety of applications and the development of a new recombinant thermostable sequencing enzyme (Amplitaq FS[1] DNA polymerase) for fluorescence sequencing.

INTRODUCTION

The demand for highly automated high throughput automatic sequencing is growing very rapidly as a consequence of the different genome sequencing projects currently being performed. To fulfill these requirements as much as possible, new instruments have recently been developed capable of sequencing as many as 36 samples simultaneously at speeds of 200 bases per hour (bph). However, these instruments based on slab gel electrophoresis technology still require manual preparation of gels and manual sample loading, besides their limited flexibility when running small number of samples.

The introduction and extremely fast development of the polymerase chain reaction (PCR[2]) during the last ten years has had an enormous impact in the field of genetic analysis, making it widespread and easily accessible for many different applications. For example, cancer research (1-3), cancer detection (1-6), forensics (1-3, 7-11), mutation detection (1-9, 12-15), prevention of genetic diseases (1,7-9, 16,24), animal husbandry (17, 18), water testing (19-20), pathogen detection (21), plant breeding (21-22), virology (7-9, 23), etc.

Much of this work is presently carried out with traditional manual techniques which still make ample use of radioactive and mutagenic substances and are also very labour intensive. The higher work load, both DNA sequencing and DNA fragment analysis application laboratories are facing has created the need for highly automated new systems, easy to operate, safe, with adequate sample throughput and flexibility and low sample quantity requirements, albeit, at a reasonable cost.

ABI PRISM™ 310 GENETIC ANALYZER

Capillary electrophoresis meets most of these requirements and consequently has been an area of intensive research and developments during the last five years in: DNA sequencing (25-34), DNA fragment analysis (35-41) and PCR product analysis (35-45).

Progress in the application of CE to DNA sequencing had been hampered by the limited stability of gels in capillaries causing bubble formation and steady decline of current (25). These problems and the loss of resolution observed for large DNA fragments when high electric fields are used to speed up the separation process led different researchers to use linear polyacrylamide as the sieving matrix for the separation of the DNA fragments (25, 27, 46).

1 Amplitaq is a registered trademark of Roche Molecular Systems,Inc.

2 The PCR process is covered by patents owned by Hoffmann-La Roche,Ltd.

3 ABI PRISM is a trademark of The Perkin-Elmer Corporation.

4 Sequenase is a registered trade mark of United States Biochemical Corporation.

Recently The Perkin-Elmer Corporation's Applied Biosystems Division has introduced the first commercial CE DNA sequencer, which overcomes most of the above mentioned problems, making automated fluorescence-based CE DNA sequencing practical (47). It provides full automation, no gel pouring, automatic sample loading and automated data analysis under full computer control. The development of two proprietary polymers, one for sequencing purposes and another for DNA fragment analysis has been fundamental for in this step forward.

This instrument takes advantage of the four-dye fluorescent sequencing technology developed by the Applied Biosystem Division. Based on the Sanger sequencing methodology, the use of four different dyes allows distinct labeling of all four extension reactions. Each dye emits light at a different wavelength when excited by laser light and thus all four reactions can be detected and distinguished in a single capillary. This strategy allows therefore simultaneous detection of all four signals in a single run, improving sequencing accuracy, precision and throughput. The fluorescent dye-labels can be incorporated into the DNA extension products using either 5'-dye label primers (dye-primers) or 3'-dye label dideoxynucleotide triphosphates (dye terminators). The most appropriate labeling method to use will depend on sequencing objectives, performance characteristics of each method and on the nature of the sample (48). Dye terminator labeling methodology requires very few pippetting steps as all four sequencing reactions are performed in a single tube, no labeled primers are required and obviously, false stops are not detected. Dye-primer labeling requires four separate extension sequencing reactions and the products of these reactions be combined together before the separation process. The main advantages of this strategy are more even signal intensity of bases - which enables homo and heterozygote determinations - and slightly longer read lengths (48).

The system's main features comprise a single capillary, 15 KV power supply, heat plate for temperature control, fixed laser induced fluorescence simultaneous detection system with CCD camera, syringe pump for low and high viscosity solutions and a 48-96 sample autosampler. The syringe pump is made up of two different plunger positions, one capable of developing pressures up to 1800 psi. for the high viscosity polymer and another capable of generating up to 28 in. Hg vacuum for the low viscosity DNA fragment analysis polymer. The electrophoresis power supply allows control of voltage applied between 100V to 15 Kvolts and the capillary heat plate controls temperature between ambient and 60°C. The argon ion laser with excitation lines at 488 and 514nm is adjustable to 10 miliwatts.

For DNA sequencing purposes a capillary of 75μm ID is used with a length of 36cm from the injection end to the detector window (total length 47cm), externally coated with polyimide to prevent breakage and internally coated with a covalently bonded fluorocarbon to suppress electrosmotic flow. The capillary used for DNA fragment analysis is of the same characteristics except it is not internally coated. Both types of capillaries can be used to run at least 100 samples. The DNA sequencing polymer is pumped fresh for every sequencing reaction analysis overcoming the limited lifetime of gel-filled capillaries and avoiding sample carryover. The polymer is composed of a 5% w/v polymer solution with 6.6M urea (40%) in 0.1M TAPS buffer (pH 8.0) with a proprietary template suppression reagent used to prevent clogging of the capillary with DNA template during injection. Samples are electrokinetically injected at 50V/cm during 20-30 sec. and the separation process is carried out at 42°C, with an applied voltage of 160V/cm during 2.5 hours. The typical length or read is of 425 bases per run (between 25 and 450 bases) with a basecalling accuracy better than 98,5%. Double stranded, single stranded, PCR products or cloned DNAs are adequate templates, and standard sequencing chemistries with amplitaq™, amplitaq FS™ or sequenase™ can be used.

Although DNA fragment analysis is outside the scope of this article, it is important to point out the system can automatically carry out rapid separations (typically 25 minutes) of DNA fragments of around 300 nucleotides in length with a resolution of 2 base pairs for a wide variety of DNA analysis such as : microsatellite studies both of STRs (short tandem repeats) (51) under denaturing conditions and VNTRs (variable number of tandem repeats) (51) of short and long DNA fragments under non-denaturing conditions. Other applications include automatic single strand conformation polymorphism (SSCP) analysis (49), quantitation of RT-PCR products, linkage mapping (52-53), mutation detection (54), etc.

The system's capability of multiple colour detection of laser-induced fluorescence allows simultaneous runs of multiple samples with an internal standard to correct differences in fragment mobilities ensuring accurate and reproducible results.

PCR labeling is easily accomplished either with fluorescent labeled primers or directly during the PCR reaction with fluorescent dNTPs.

Capillaries are refilled with low viscosity sieving polymer (which dynamically coats the capillary wall to prevent electrosmotic flow) before each analysis easily and quickly. Multiple sample injections and variable injection times ensures the analyst to be able to detect weak samples with high sensitivity.

This new automatic laser-induced fluorescence based CE DNA genetic analyzer provides analysts with a highly accurate, reproducible and sensitive system, easy to use yet sufficiently flexible for a variety of applications, for the first time eliminating the need for tedious and labour intensive slab gel pouring and manual sample loading. Multi-colour fluorescence technology simplifies DNA sequencing allowing conventional state of the art Sanger sequencing chemistries to be used unaltered and enhances DNA fragment analysis throughput allowing simultaneous detection of multiplex PCR reactions with an internal standard (50).

AMPLITAQ™ FS THERMOSTABLE SEQUENCING ENZYME

Developments to simplify sample preparation and sequencing chemistries are as important for total sequencing automation as instrumental innovations.

Amplitaq™ FS thermostable sequencing enzyme has been developed and specifically designed by Roche Molecular Systems' scientists for automatic fluorescence sequencing applications. This DNA polymerase is a recombinant form of Amplitaq™ DNA polymerase (55) with two genetically engineered mutations. The first mutation practically eliminates the native enzyme's 5'—>3'nuclease activity generating cleaner data with reduced background noise and virtually no false terminations. The second mutation affects the enzyme's incorporation efficiency of ddNTPs in both dye-primer and

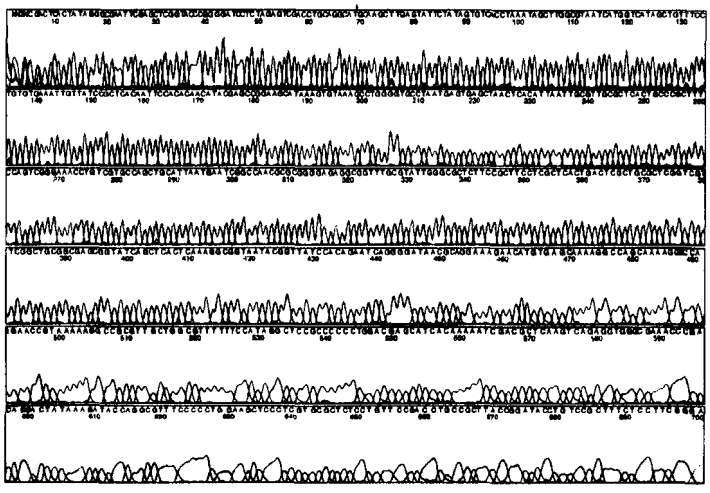

Fig 1. Dye primer kits with AmpliTaq DNA Polymerase, FS Produce data with more uniform peak heights. This results in more accurate automated base calling, especially at longer read distances.

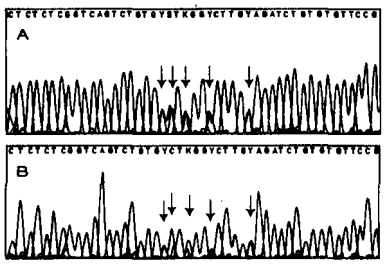

Fig 2. Direct PCR sequencing of the highly polymorphic HLA-B gene using dye primer kits with AmpliTaq DNA Polymerase, FS (A) and AmpliTaq DNA Polymerase, CS+ (B). The more uniform peak heights generated by AmpliTaq DNA Polymerase, FS enable easier identification of heterozygous bases (arrows). Heterozygous bases produce approximately half the amount of signal as homozygous bases.

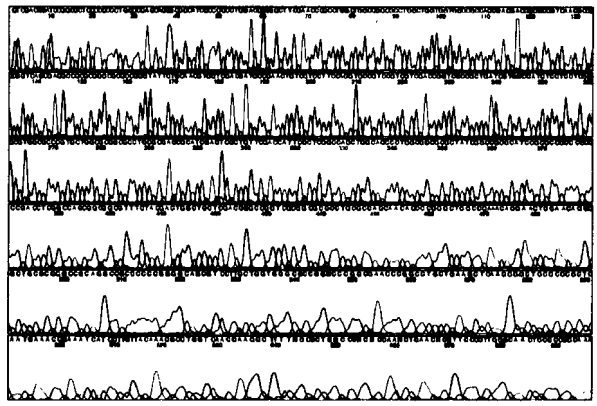

Fig 3. Dye terminator kits with AmpliTaq DNA Polymerase, FS provide up to 650 bases at 98% accuracy. Shown here is sequence data from a GC-rich dsDNA clone.

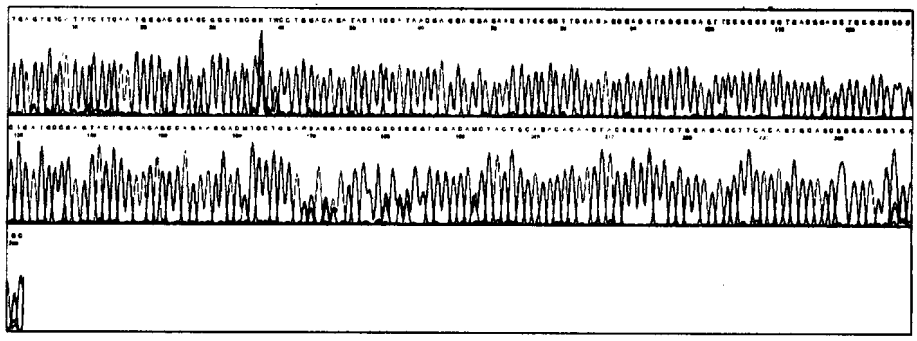

Fig 4. Dye primer sequencing with 310 capillary electrophoresis system of pGEM dsDNA standard.

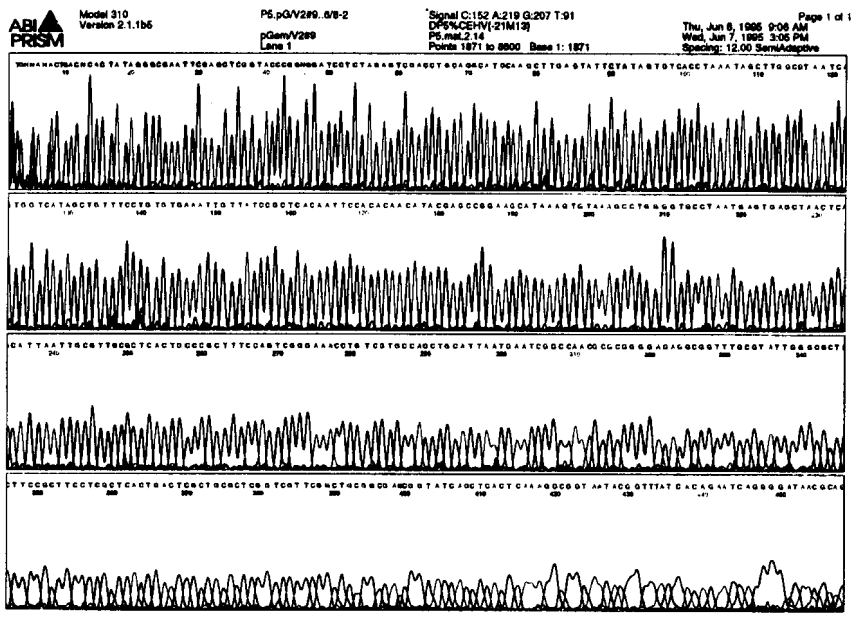

Fig 5. This Dye Primer data is from the HLA DRB1 exon 2 gene of MHC, which is the major histocompatibility complex. The sample was taken from peripheral blood and is a PCR-amplified 300-base product diluted with water. The actual primers were tailed M13 primers incorporated in the PCR primers. This data demonstrates the true power of the AmpliTaq, FS enzyme. Notice the multiple heterozygous base positions at bases 161, 169, 170, 172, 173, 180, 182, 183 and 192; all successfully called with AmpliTaq DNA Polymerase, FS and Sequence Navigator Software. The stop peak at base 37 is a result of primer-dimer formation. It does not adversely affect sequencing results.

dye-terminator reactions, generating higher fluorescent signals. As a result, as little as 0.1- 0.2μg of single stranded DNA or 0.25-0.50μg double stranded DNA starting template is now required with this new enzyme. As a consequence, smaller amounts of dye-labeled ddNTP terminators are needed and a simple ethanol precipitation is all that is necessary for elimination of residual dye-terminator ddNTPs. Additionally, this higher incorporation efficiency of ddNTPs, when using dye-primer chemistry, results in much more uniform peak heights (fig. 1) making this chemistry with the new enzyme an excellent choice for PCR-comparative sequencing aimed at identifying mutations and polymorphisms (fig. 2).

Finally, sequencing templates with high GC content and/or homopolymer regions is easier and more robust with this thermostable enzyme (fig. 3).Following are two examples of sequencing reactions with the new CE DNA sequencer (fig. 4) and an heterozygous sample detection with the above mentioned new enzyme (fig. 5).

These two important new developments in automatic DNA sequencing will make this fundamental technology far more accessible to many laboratories and will also help sequencing of difficult samples and heterozygous/homozygous detection easier.

REFERENCES

1)R.A.Gibbs, Anal. Chem. 62,1202 (1990)
2)W. Bloch, Biochemistry 30 (11), 2735 (1991)
3) H.A.Erlich et al, Science 252, 1643 (1991)
4)E.A. Rose, FASEB J. 5, 48 (1991)
5)C.T. Caskey, Science 236, 1223 (1987)
6)F. McCormick, Cancer Cells 1(2), 56 (1989)
7)H.A.Erlich,PCR Technology,p.1-22,Stockton Press, New York 1989
8)M.A.Innis et al, PCR Protocols,p.3-27, Academic Press, San Diego 1990.
9)N.Arnheim and C.H. Levenson, Chem. & Eng. News, October 1,36 (1990)
10)R. Reynolds et al, Anal. Chem. 63(1), 2 (1991)
11)H.A.Erlich et al, PCR Methods and Applications 1, 241 (1992)
12)C.E. Bronner et al, Nature 368, 258 (1994)
13)P.A. Futreal et al, Science 266, 120 (1994)
14)R.D.Kolodner et al, Cancer Res. 55, 242 (1995)
15) R.D.Kolodner et al, Genomics 24, 516 (1994)
16)E.S.Mansfield, Human Molecular Genetics 2(1), 43 (1993)
17)A. Rodriguez, Virology 189, 000 (1992)
18)Stockmarks for Cattle Paternity- Perkin Elmer Corporation -Applied Biosystems Division.
19) A.K.Bej et al , Biotechniques 15(5), 706 (1992)
20)C.J.Palmer et al, Appl. Environ. Microbiol. 61(2), 407 (1995)
21)G.Nolasco et al, J. Virol. Meth. 45, 201 (1993)
22)B.W.S. Sobral et al, Theor. Appl. Genet. 86, 105 (1993)
23)B.A.Larder et al, Nature 365, 671 (1993)
24)S.H. Embury et al, N. Engl. J. of Med. 316, 656 (1987)
25)M.C.Ruiz-Martinez et al, Anal. Chem. 65, 2851 (1993)
26)J.A.Luckey and L. Smith, Anal. Chem. 65, 2841 (1993)
27)N.Best et al, Anal. Chem. 66, 4063 (1994)
28)H.R. Harke et al, J. Chromatography 608, 143 (1992)
29)D.Y.Chen et al, Nuceic Acids Res. 20,4873 (1992)
30)H. Swerdlow et al, Anal. Chem. 63, 2835 (1991)
31)A.E.Karger et al, Nucleic Acids Res. 19, 4955 (1991)
32)J.A.Luckey et al, Nucleic Acids Res. 18, 4417 (1990)
33)A.S.Cohen et al, J. Chromatography 516, 49 (1990)
34)H.Drossman et al, Anal. Chem. 62, 900 (1990)
35)J.M.Robertson et al, International Society Forensics Haemogenetics 16th Congress, Santiago de Compostela, September 1995
36)H.M.Wenz, Nucleic Acids Res. 22(19), 4002 (1994)
37)J.Cheng et al, Anal. Chem. 66, 4210 (1994)

38)H.Zhu et al, Anal. Chem. 66, 1941 (1994)
39)Y.Baba et al, J. Chromatography 558, 273 (1991)
40)J.P.Landers, Biotechniques 14, 98 (1993)
41)W.Lu et al, Nature 368, 269 (1994)
42)H.E.Schwartz et al, J. Chromatography 559, 267 (1991)
43)K.J.Ulfelder et al, Anal. Biochem. 200, 260 (1992)
44)D.D.Principe et al, J. Chromatography 638, 277 (1993)
45) J.Cheng et al, J. Chromatography 677, 169 (1994)
46)T.Manabe et al, Anal. Chem. 66, 4243 (1994)
47)ABI PRISM 310Genetic Analyzer, The Perkin-Elmer Corp.-Applied Biosystems Div. 850 Lincoln Centre Drive, Foster City CA 94-404-1128 USA.
48)DNA Sequencing Chemistry Guide, The Perkin-Elmer Corp.-Applied Biosystems Div.
49)J.Ellison et al, Biotechniques 15, 684 (1993)
50)P.E.Mayrand et al, Applied and Theoretical Electrophoresis 3, 1 (1992)
51)C.P.Kimpton et al, PCR Methods and Applications 3, 13 (1993)
52)L.E.Pritchard et al, Human Molecular Gen. 4, 197 (1995)
53)J.B.Copeman et al, Nature Genetics 9, 80 (1995)
54)E.Verpy et al, Proc. Natl. Acad. Sci. USA 91, 1873 (1994)
55)F.C.Lawyer et al, PCR Methods and Applications 2, 275 (1993)