Automated DNA sequencing: ultrasensitive detection of fluorescent bands during electrophoresis

Wilhelm Ansorge*, Brian Sproat, Josef Stegemann, Christian Schwager and Martin Zenke

European Molecular Biology Laboratory, Postfach 10.2209, D-6900 Heidelberg, FRG

Received March 31, 1987; Accepted April 24, 1987

ABSTRACT

A simple system has been designed enabling ultrasensitive on-line detection of fluorescently labelled macromolecules, e.g. nucleic acids, proteins and peptides during electrophoretic separations in gels. An important application is the automated DNA sequence determination without radioactivity. Drying of gels, film exposure and handling are not necessary. A sulphhydryl containing M13 sequencing primer has been synthesised and end-labelled in a reaction with fluorescein iodoacetamide. This is then used in the dideoxy reactions. In particular no moving parts or complicated software are required for data collection and analysis. Compared to our first automated device detection sensitivity has been improved by a factor of thirty to about 3×10^{-18} mol per band. The resolution has increased to about 400 bases in 5 hours, with the possibility to read up to about 500 bases when they are properly labelled. Gels shorter than 20 cm may be used for resolution of about 300 bases. The single gel system may be upgraded for simultaneous running and reading of six or ten sequencing samples.

INTRODUCTION

Electrophoretic separations of macromolecules in gel sieving media have become a versatile technique in all laboratories. In a standard procedure, the sample is applied to the gel and macromolecules are separated due to their difference in molecular weight. Modern analytical micro-methods use very small amounts of materials and their fast and reliable detection becomes a limiting factor. Until recently, radioactive labelling has been the most sensitive technique used in DNA sequencing [1,2]. After the separation run, the gel mold is dismantled, gel dried, film is applied and exposed (for about a day or more) to obtain sufficient film darkening allowing detection of bands on the autoradiogram and determination of DNA sequence. Silver staining has been used extensively for detection of DNA and protein bands in gels [3] down to 10^{-13} - 10^{-14} mol per band, but this method fails at still lower concentrations in the range of 10^{-16} - 10^{-17} mol per band, which are required in the DNA sequencing methods.

Time for detection of the separated macromolecules can be significantly shortened when they are fluorescently labelled and detected in the gel already during the electrophoresis, transferring the information directly into a computer, without gel drying,

film exposure, handling and analysis. In one method [4], four fluorescent dyes are used, each to mark one of the bases which are then co-electrophoresed and detected in one gel tube. In spite of the obvious attraction of this method, there seem to be drawbacks associated with it, namely variations in the electrophoretic mobility due to different dyes, spectral overlap of the dyes, additional software needed to correct these problems and a need for moving parts in filtering or scanning.

In our first automated device [5-7], tetramethylrhodamine has been chosen as a single fluorophore for the four bases. DNA sequences are determined in the gel during electrophoresis and read directly into a computer. The chemistry for preparation of the fluorescent primer is reliable and straightforward, and moreover the preparation of the sequencing samples is less expensive compared to the method using radioactive labels. The developed instrument includes several novel features such as detection of laser-induced fluorescence in conjunction with slab gel electrophoresis, excitation laser beam passing inside the gel throughout its entire width, efficient and simple light coupling in and out of the gel, continuous monitoring of the four tracks, no moving parts, and background reduction by adjustment of the plane of polarisation. Another important background reduction is achieved by eliminating the fluorescence of glass, inherent to all techniques where the excitation laser beam traverses glass within the light collecting field. In this paper we describe improvements in sensitivity, resolution and design of the instrument for the automated DNA sequencing method [5-7], with fluorescein as the fluorescent dye for the four bases.

DISCUSSION AND RESULTS

The structure of the fluorescent primer is illustrated in Figure 1. Continuous monitoring of the four gel tracks makes the software required for sequence determination less complicated. In addition, besides the positive peak in one track output, the absence of peaks in the other three track outputs gives further information for the confirmation of a

Figure 1
Structure of the fluorescein labelled M13 sequencing primer. Oligonucleotide = d[GTAAACGACGGCCAGT].

Figure 2.1

Raw data output from a total sequencing run with a 20 cm separating length, 7% polyacrylamide gel, displayed until base number 310 read from the gel. The profile depicts parts of the sequence of the recombinant M13mp8SV containing a rabbit β -globin/Simian Virus 40 (SV40) fusion gene inserted between the *Acc* I and *Eco* RI sites of M13mp8 [8]. The sequence reading starts at position 6287 of M13 sequence [9] and immediately runs through the *Hind* III and *Pst* I sites of the M13mp8 polylinker into the rabbit β -globin sequence from position 310 (TaqI site) to position 19 [10].

sequence decision. A display of raw sequencing data output from the detector is shown in Figure 2 for 20cm separating gel length and in Figure 3 for 30 cm length. The sequence is easily read from the relative peak positions even from the raw data, no complicated software is required. A characteristic feature of the method is that successive peaks exhibit approximately the same width and separation. This observation is explained by the

Figure 2.2

Enlarged part of Figure 2.1

The sequence reads from position 189 to position 153 of rabbit β -globin sequence [10], corresponding to bases 140-176 read from the gel.

Figure 2.3

Enlarged part of Figure 2.1

The sequence reads from position 55 to position 45 of rabbit β -globin sequence [10], corresponding to bases 274-284 read from the gel.

fact that the longer DNA fragments migrate more slowly, but they are concentrated in narrower bands, so that the time for a fragment to pass by the detector remains virtually unchanged. The automated data analysis will make possible a quantification of interband distances and relative band intensities, and will enable the formulation of a set of rules

Figure 3.1

Partial raw data output from a 5% polyacrylamide gel of separating length 30 cm, displayed up to base 440 read from the gel. The DNA sequence represents the rabbit β -globin/SV40 junction where β -globin sequence, at the Pvu II site, (position -9 [10]) has been fused to the Hind III site of SV-40 (position 5171 [11]) via a Hind III linker. The sequence from position 189 to -9 of β -globin and from position 5171 to position 5260 of SV40 is shown.

Figure 3.2

Enlarged part of Figure 3.1

The sequence reads from position +14 to position -9 of β -globin sequence and from position 5171 to position 5182 of SV 40 sequence, corresponding to bases 314 - 350 read from the gel.

appropriate to automated reading. Of particular interest is the analysis of the chemistry involved in the sequence dependent rate of termination of the enzymatic polymerisation. An important aspect of this method is the possibility to read the sequence from the first base (i.e. the first base added by the polymerase to the 3'-end of the primer). Longer DNA fragments up to about 500 bases are resolved as well as the shorter ones providing that

Figure 3.3

Enlarged part of Figure 3.1

The sequence reads from position 5173 to position 5182 of the SV 40 sequence, corresponding to bases 341-350 read from the gel.

the labelling is sufficient. The use of the 5'-terminally labelled fluorescent primer has no effect on hybridisation and does not interfere with the sequence determination. The use of fluorescent dyes absorbing and emitting at wavelengths above 400nm is advantageous, due to very low background caused by Raman emission from the gel, which increases drastically at shorter wavelengths.

We have tested laser power in the range 1-60mW, which is suitable for the fluorescein label. Due to the high optical transparency of the gel only a negligibly small fraction of the incident light energy is absorbed. The low laser power required makes possible the use of smaller air-cooled lasers at a lower price. The present detection limit in the gel, determined at a signal to noise ratio of 2:1, is about 3×10^{-18} mol per band. Photo-bleaching of the fluorescent dye occurs during its passage through the laser beam. To some extent, this may improve the resolution, as the bleached band interferes less with the band following it. The plane of polarisation of the laser beam was rotated from vertical to horizontal in order to minimize the light scattered towards the detectors. Agarose gels were tested in the set-up and could also be used as the separating medium. The length of the separating gel is 20 or 30cm, but even shorter gels may be used for resolution of 300-350 bases. Laser light energy absorbed per band is negligible so that theoretically many hundreds of tracks could be run in parallel. The laser beam traverses smoothly the 20 cm wide gels, making possible simultaneous sequencing of six samples, and a version for ten samples is in preparation.

The accuracy in the sequence reading is limited mainly by occurrence of false bands. To obtain best results with the automated method, the DNA sequencing samples must be of high and reproducible quality. Therefore, automation of the sample preparation techniques will be an important step in the efforts to automate the DNA sequencing methodology.

EXPERIMENTAL

General

5'-(S-Triphenylmethyl-3-mercaptopropyl phospho) d[GTAAAACGACGGCCAGT] was synthesised and purified as described previously [5]. 5-Iodoacetamidofluorescein was purchased from Molecular Probes, Inc. (Eugene, OR, U.S.A.).

Preparation of fluorescein labelled primer

5'-(S-Triphenylmethyl-3-mercaptopropyl phospho) d[GTAAAACGACGGCCAGT] (100nmol) in 0.1M triethylammonium acetate solution (1ml, pH7) was detritylated by treatment with silver nitrate (500nmol) for 1h in an Eppendorf tube. Dithiothreitol (700 nmol) in water (100 μ l) was then added, the tube contents were mixed and left for 30 min. Centrifugation removed the insoluble silver salt. The supernatant was transferred to a fresh Eppendorf

tube, and 1M aqueous sodium bicarbonate was added to give pH8.5. A solution of 5-iodo-acetamidofluorescein (3mg, 6 μ mol) in N,N-dimethylformamide (300 μ l) was added and the contents of the tube were mixed thoroughly and then kept for 6h in the dark. Added 2-mercaptoethanol (10 μ l) to quench the excess iodoacetamide and then dialysed the solution against water (3 x 2l) in the dark.

The solution was concentrated in vacuo and the dye labelled primer was purified by ion-exchange h.p.l.c. on Partisil SAX using a concentration gradient of potassium dihydrogen phosphate pH6.3 in formamide/water (6:4 v/v) as eluant. This procedure removes a large amount of strongly associated but non-covalently bound dye. The product was desalted by dialysis and then further purified by reversed phase h.p.l.c. on a C₈ Aquapore RP-300 column using 0.1M triethylammonium acetate pH7/acetonitrile as eluant. The fully purified fluorescein labelled primer, see Figure 1 for structure, was desalted by dialysis and stored in the dark at -20°C. The final yield of dye labelled primer was 15% based on the S-trityl material. In an identical fashion, eosin labelled primer was obtained in 27% overall yield.

Preparation of Sequencing samples

Single-stranded M13 phage DNA was prepared using standard techniques [12]. Since minute amounts of polyethyleneglycol (PEG) contaminating the single-stranded M13 template DNA lead to unspecific termination of the dideoxy sequencing reaction, efforts have been made to efficiently remove traces of PEG solution after PEG precipitation (e.g. by spinning down the droplets of PEG solution sticking to the wall of the centrifuge tube and removing them by means of a drawn-out capillary). In some instances purification of the M13 phage DNA by CsCl equilibrium gradient centrifugation has proven to be helpful in avoiding the occurrence of unspecific bands. Therefore, standard protocols for the preparation of M13 phage DNA [12] have been scaled up by approximately 200-fold. After PEG precipitation, the M13 phage was lysed in 4ml of 50mM Tris/HCl, pH7.5, 10mM EDTA, 1% SDS, and after proteinase K digestion (100 μ g/ml, 1h, 37°C) extracted with equal volumes of phenol and chloroform and recovered by ethanol precipitation. The phage DNA pellet was dissolved in 4.3ml TE-buffer (20mM Tris/HCl, pH7.5, 1mM EDTA) and after addition of 4.8g of CsCl and 5 μ l of ethidium bromide solution (10mg/ml) was subjected to two consecutive equilibrium density gradient centrifugations (58,000 rpm in a VTi65 rotor, for at least 5 hours at 20°C). The single-stranded phage DNA (lower band in the gradient) was recovered, extracted with an equal volume of 1-butanol (4 times), dialysed against TE buffer and ethanol precipitated.

Sequencing reactions

To ensure proper labelling over more than 300 nucleotides the dideoxytriphosphate to

deoxytriphosphate ratios in the nucleotide specific termination mixes A',C',G',T' have been modified from standard protocols [12] as follows; values given in μl :

		T'	C'	G'	A'
0.5mM	dTTP	25	500	500	500
"	dCTP	500	25	500	500
"	dGTP	500	500	25	500
"	dATP	500	500	500	25
10.0mM	ddTTP	5			
"	ddCTP		1		
"	ddGTP			2	
"	ddATP				5
TE buffer		500	500	500	500

One pmole of single-stranded M13 phage DNA, prepared as described above, was annealed with 1pmole of fluorescein labelled primer (see above) in 10 μl of 10mM Tris/HCl, pH8.0, 10mM MgCl_2 at 65°C for 30-60 minutes. After the addition of 1 unit of *E. coli* DNA polymerase (Klenow fragment, Boehringer Mannheim) 2 μl aliquots of the solution were mixed with 2 μl of the respective termination mix (A',C',G',T', respectively) and incubated at room temperature for 20 minutes. Following the chase reaction (2 μl of 0.25mM dATP, 0.25mM dCTP, 0.25mM dGTP, 0.25mM dTTP, 30-60 minutes at room temperature) 2 μl aliquots were mixed with 2 μl of formamide, heated to 95°C for 2min, chilled on ice, and applied to the polyacrylamide gel for electrophoresis.

Instrumental design

Design of the instrument has been described in [5]. In the present system, the beam of an air-cooled argon ion laser (Lexel, model 65, wavelength 488nm, used in the light regulated mode) traverses a polarisation rotator where its plane of polarisation is rotated through 90°, from vertical to horizontal. The diameter of the laser beam is reduced in a focusing system from 0.6mm to between 0.3 and 0.4mm. The laser output power was in the range 1-60mW. The narrow, focussed beam is then coupled efficiently inside the gel with the aid of light coupling plates, of identical thickness to the gel spacers, and with polished optical surfaces. The beam traverses the entire gel width in a straight line, with no abrupt divergence, and is then coupled out of the gel through a second coupling plate, which faces the first. Inside the gel, fluorescent labelled DNA bands are excited in the order of the DNA sequence, as they enter the laser beam during electrophoresis. This configuration avoids the need for any moving parts and does not lower the counting times per track, as would be the case with a scanning system. This results in increased reliability of the detected sequence.

Taking advantage of the low background inherent in an orthogonal system, the fluorescent light is collected at right angles to the gel plane by light collecting and imaging

optics, of relative aperture 1.2 and focal length 50mm. Fluorescent light from the four DNA tracks is channelled to a filter assembly, used to reduce light scattered by Rayleigh and Raman processes. The filter assembly is composed of a special bandpass interference filter for fluorescein (Schott, FRG), with 45nm full width at half maximum. The detector includes four photomultipliers (type 9954A) and electronic parts from EMI, London. Acquisition and processing of the raw sequencing data is controlled by an IBM PC micro-computer. Raw data from the detector may be displayed on a printer for control purposes during a sequencing run.

Electrophoresis

Electrophoresis was carried out as described in [5]. Since the sequence is deduced from the mutual position of bands in four adjacent tracks thermostated gels without pattern distortion [13] are advantageous for a simple analysis. Generally, 5-10% (total monomer concentration) polyacrylamide-urea gels of dimensions 150-200mm wide, 200-300mm separating length, and 0.5mm thick were used. Even shorter gels (100-150mm) may be used in the system. The gel was mounted securely in a vertical position.

The normal marker dyes used in sequencing gels are strongly absorbing and fluorescing at the wavelengths used, and so were replaced by the fluorescent primer as marker. Volumes loaded on the 0.5mm gels (see under sequencing reactions above) were about twice those loaded on the standard 0.2 - 0.3mm thick sequencing gels [14]. Electrophoretic runs were performed at 1000 - 1500V (about 40V/cm) and 50°C, with the dissipated power limited to 30 watts.

CONCLUSION

The simple system described represents a significant improvement in DNA sequencing methodology. It allows ultrasensitive on-line detection of fluorescently labelled macromolecules, e.g. nucleic acids, proteins and peptides during electrophoretic separations in gels. An important application is the automated DNA sequence determination without radioactivity. A sulphhydryl containing M13 sequencing primer is synthesised and end labelled in a reaction with fluorescein iodoacetamide. The fluorescent bands are excited by a laser and detected in the gel during electrophoresis, and sequence data is transferred directly into a computer. Compared to the first automated device described by us, there are significant improvements in the system performance reported here. Detection sensitivity has been improved by a factor of thirty to about 3×10^{-18} mol per band. The resolution has increased to about 400 bases in 5 hours, with the possibility to read up to about 500 bases when they are properly labelled. Length of the separating gel is 20 or 30cm. The system may be upgraded for simultaneous running and reading of six or ten sequencing samples in one gel. Design of the instrument has been simplified and is less expensive, e.g. by using small air-cooled argon lasers. Both polyacrylamide and agarose may be used as the separating gels.

The chemistry for preparation of the fluorescent primer is reliable and straightforward, and moreover the preparation of the sequencing samples is less expensive compared to the method using radioactive labels. The developed instrument includes several novel features such as detection of laser-induced fluorescence in conjunction with slab gel electrophoresis, excitation laser beam passing inside the gel throughout its entire width, efficient and simple light coupling in and out of the gel, continuous monitoring of the four tracks, no moving parts, and background reduction by adjustment of the plane of polarisation.

The method will be used in applications where speed and high sensitivity are required, like in DNA sequencing or in on-line detection of proteins and peptides labelled by standard techniques with fluorescent dyes. No drying of gels or handling of films are necessary; information about sequence of bands and their profile is stored and recorded directly in the computer.

ACKNOWLEDGEMENTS

We thank Professor L. Philipson for encouragement, and C. Walter-Möller for excellent technical assistance.

*To whom correspondence should be addressed

REFERENCES

1. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
2. Maxam, A.M. and Gilbert, W. (1980) *Methods in Enzymology* 65, 499-559.
3. Ansorge, W. (1985) *J. Biochem. Biophys. Meth.* 11, 13-20.
4. Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B.H., and Hood, L.E. (1986) *Nature* 321, 674-679.
5. Ansorge, W., Sproat, B.S., Stegemann, J. and Schwager, C.H. (1986) *J. Biochem. Biophys. Meth.* 13, 315-322.
6. Ansorge, W., Stegemann, J., and Schwager, C.H. (1986) FRG Patent Application Nr. P36.18.605.8
7. Sproat, B. (1986) FRG Patent Application Nr. P36.42.939.2
8. Zenke, M., Grundström, T., Matthes, H., Wintzerith, M., Schatz, C., Wildeman, A. and Chambon, P. (1986) *The EMBO Journal* 5, 387-397.
9. Yanish-Perron, C., Vieira, J. and Messing, J. (1985) *Gene* 33, 103-119.
10. van Ooyen, A., van den Berg, J., Mantei, N. and Weissmann, C. (1979) *Science* 206, 337-344.
11. Tooze, J. (1982) *Molecular Biology of Tumor Viruses*, 2nd Ed., Part 2, Revised. DNA Tumor Viruses. Cold Spring Harbor Laboratory Press, New York.
12. Bankier, A.T. and Barrell, B.G. in 'Techniques in the Life Sciences', (1983) B5 (Elsevier, Ireland).
13. Ansorge, W. and Garoff, H. (1981) *Anal. Biochem.* 115, 450-457.
14. Ansorge, W. and Labeit, S. (1984) *J. Biochem. Biophys. Meth.* 10, 237-243.