

## DNA Sequencing Directly from a Mixture Using Terminal-Base-Selective Primers

Hiroko FURUYAMA, Kazunori OKANO, and Hideki KAMBARA\*

Central Research Laboratory, Hitachi, Ltd., 1-280, Higashi-koigakubo, Kokubunji-shi, Tokyo 185, Japan

(Received 22 June 1994)

### Abstract

DNA cloning is often used to select and amplify one DNA species from a mixture. However, the cloning process is complex and labor-intensive. We have developed a new two-step method for DNA sequencing directly from a mixture. The first is the introduction of a known oligonucleotide (common part) into the terminus of unknown DNA by ligation. The second is selective DNA sequencing using primers with two additional nucleotides at the 3' terminus in addition to the common part (terminal-base-selective primers). The primers work only for templates on which the primers perfectly hybridized. This method was found to be effective for the *Hind*III digestion products of  $\lambda$  phage.

**Key words:** DNA sequencing; cloning; terminal-base-selective primer; DNA sequencing of mixture

### 1. Introduction

Great progress in the human genome project has been achieved over the last 3 years especially in the mapping field.<sup>1</sup> More recently, however, much more attention is being paid to DNA sequencing. The importance of a large-throughput DNA sequencer as well as a long-base reading technology has been pointed out and a lot of effort has been made to achieve this.<sup>2-7</sup> Various methods for rapid and large throughput DNA sequencers including fluorescent thin-gel electrophoresis and fluorescent gel capillary array DNA sequencers have been reported to improve the overall sequencing throughput.<sup>2-5</sup> Also, experiments using a long-slab as well as long capillary gel electrophoresis have been reported to read 1000 bases in one electrophoresis operation.<sup>6-7</sup> A high throughput of about 1 Mb/day in one laboratory is not unrealistic and may be necessary for the success in the human genome project.<sup>8</sup> In view of this, a new strategic method for sequencing long DNA should be developed because sample preparation in the present DNA sequencing method is very labor-intensive. Although good vectors such as yeast artificial chromosome (YAC) and bacterial artificial chromosome (BAC) can be successfully used to clone long DNA, there are very laborious subcloning steps before sequencing.

Generally, the cloning and the subcloning processes have been considered to be indispensable for sequencing long DNA. Long DNA is first digested into many fragments then the individual DNA fragments are separated

and amplified from the mixed fragments through cloning steps. As we now have a powerful DNA amplification method called the polymerase chain reaction (PCR), the whole sequencing process could be greatly simplified by eliminating the cloning steps if DNA could be sequenced directly from the mixture.

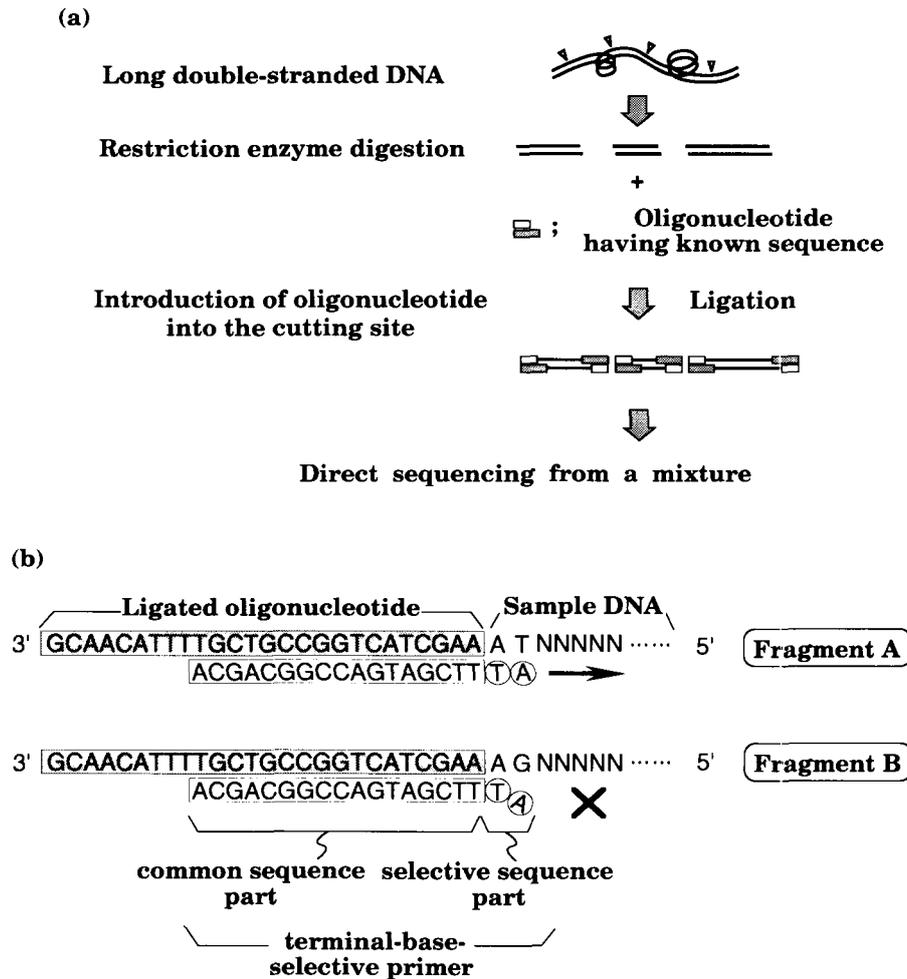
This paper describes a new method of sequencing DNA fragments, produced by enzymatic digestion, directly from the mixture without a separation process such as cloning and colony selection.

### 2. Principle

It has been pointed out that extension of the complementary DNA strand by DNA polymerase is strongly influenced by the matching between the 3' terminus of the primer and the template. We use this matching requirement to distinguish a fragment species from others and to sequence the fragment without any separation procedure such as cloning. The principle of this new method is shown in Fig. 1. Long DNA, which cannot be sequenced by one sequencing process, is enzymatically digested into fragments. As the terminal sequences of these fragments are not known, we introduce a known oligonucleotide into the cutting site terminus by ligation. The ligated oligonucleotides are labeled with a fluorophore at their 5' terminus. The product is analyzed by a fluorescent gel electrophoresis analyzer to determine the number of fragment species, their sizes and the rough amount of each fragment. If the amount is too small for sequencing, PCR can be performed to increase the number of fragment copies because the fragments have a known sequence region at each terminus. Then we

Communicated by Masahiro Hattori

\* To whom correspondence should be addressed. Tel. +81-423-23-1111, Fax. +81-423-27-7784



**Figure 1.** Process of direct sequencing from a mixture. (a) Sample preparation process. (b) Principle of selecting a fragment using terminal-base-selective primers. Polymerase reactions occur only for primers which hybridized completely with sample DNA. In this example, fragment A in the mixture is only sequenced.

can sequence the fragments by the method of the Sanger using the known oligonucleotide terminal sequence for primer binding. However, if the primer hybridized only on the oligonucleotide sequence (common sequence) then every fragment would be extended. Therefore, we add two more nucleotides on the 3' terminus of the primer to select fragments for the polymerase reaction, as shown in Fig. 1(b). We prepare 16 primers which have the common sequence and the two selective bases, [there are 16 possible combinations of these two bases ( $4 \times 4$ )]. As DNA polymerase reactions are very sensitive to the matching of the two-base sequence between the primer and the template, the sequence reactions will occur only for the completely matched primer. This can be used for selective sequencing of mixed DNA fragments. If the number of single-stranded (ss) DNA fragment species in the mix-

ture is large, one primer species will bind to two different fragment species and we would need further selection. Trinucleotides would be used as the selective oligomer part to separate the duplex information. The recommended number of ssDNA fragments in a mixture is 6 or less so there is a small possibility of one primer binding to two fragments. Gel electrophoresis separation and fractionation of the digested fragments might be useful to group the fragments for the selective sequencing, when the number of fragment species is large.

### 3. Materials and Methods

#### 3.1. Oligonucleotides synthesis and labeling

Oligonucleotides were synthesized on an Applied Biosystems 392 DNA synthesizer (Applied Biosystems,

Foster City, CA). The sequences of the terminal-base-selective primers were as follows:

One-base-selective primer set (18-mer, total 4)

ACGACGGCCAGTAGCTTX<sub>1</sub>

Two-base-selective primer set (19-mer, total 16)

ACGACGGCCAGTAGCTTX<sub>1</sub>X<sub>2</sub>

Three-base-selective primer set (20-mer, total 64)

ACGACGGCCAGTAGCTTX<sub>1</sub>X<sub>2</sub>X<sub>3</sub>

where [X<sub>i</sub>, i=1,2,3; and X is any nucleotide]

Texas Red-labeled oligonucleotides were prepared as follows: 5'-terminal Aminolink 2 (Applied Biosystems) binding oligonucleotide ( $2 \times 10^{-3}$  mol) dissolved in 400  $\mu$ l of 0.25 M carbonate buffer (pH 9.0) was conjugated with  $2 \times 10^{-6}$  mol of Texas Red (Sulforhodamine 101 Acid Chloride, Dojindo, Kumamoto, Japan) dissolved in 100  $\mu$ l of acetonitrile. After reaction for 16 hr, the reaction mixture was neutralized by adding 8  $\mu$ l of acetate. The Texas Red-labeled oligonucleotides were collected by adding 33  $\mu$ l of 5 M ammonium acetate, 1300  $\mu$ l of ice-cold ethanol and centrifuge at 12000 rpm for 30 min after incubating for 1 hr at  $-20^\circ\text{C}$ . The precipitation product was resuspended with 40  $\mu$ l of 50% formamide. Further purification of Texas Red-oligonucleotides were carried out by electrophoresis in a 12% polyacrylamide gel containing 7 M urea. The separated bands containing Texas Red-oligonucleotides were cut out and suspended in 0.1 M trimethylamine acetate buffer (pH 7.0) containing 1 mM EDTA. Eluted solutions were added to a single-use type reverse-phase cartridge (Applied Biosystems) and washed with 8% acetonitrile to wash off the unlabeled oligonucleotides. The bound labeled oligonucleotides were dissociated from the column with a 20% acetonitrile aqueous solution and dried. Purified Texas Red-oligonucleotides were dissolved in 5 mM Tris-HCl (pH 8.0) containing 0.5 mM EDTA and stored at  $-20^\circ\text{C}$ .

Fluorescein isothiocyanate (FITC)-labeled primers were prepared by the same method as above but the FITC was dissolved in *N,N*-dimethylformamide before use.

### 3.2. Preparation of samples

Lambda phage DNA was digested by *Hind*III into 8 double-stranded (ds) fragments. The phosphoric acid at the cutting sites was removed by using calf intestine alkaline phosphatase (Takara, Kyoto, Japan) to avoid self ligation during the following ligation process. The oligonucleotide 21 mer (5'-CGTTGTAACGACGGCCAG T-3') and the complementary strand of it, which fit for *Hind*III cutting site (25 mer; 5'-AGCTACTGGCCGTCGTTTACAACG-3'), were used as linkers for ligation. The fragments were ligated by using T4 DNA ligase (Takara) with the linkers followed by centrifugal precipitation in 0.6 volumes of

20% polyethyleneglycol solution including 2.5 M NaCl to remove residual linkers and to purify the modified DNA fragments. The precipitation product was rinsed with 70% ethanol.

### 3.3. Termination reaction for selecting sequencing primers

To determine the number of fragment species and their terminal base species, complementary strand extension reactions were carried out using four non-labeled primers having one selective base and a fluorophore (FITC) labeled dideoxynucleotide (DuPont). Five micrograms of the  $\lambda$  phage fragments ligated with the linker was digested with 6 units of T7 Gene 6 Exonuclease (United States Biochemical, Cleveland, OH) in 5.6  $\mu$ l of exonuclease buffer containing 50 mM Tris-HCl (pH 7.5), 1 mM DTT, and 100  $\mu$ g/ml BSA in a total volume of 28  $\mu$ l after incubating at  $37^\circ\text{C}$  for 15 min to make the ssDNA fragments and the reaction mixture was heated at  $80^\circ$  for 15 min to inactivate the enzyme.

Four reactions corresponding to "A reaction," "C reaction," "G reaction," and "T reaction" were carried out with each of the four one-base-selective primers. After annealing with 7  $\mu$ l of the above DNA solution and dissolving 0.4 pmol of primer in 8.6  $\mu$ l of 130 mM Tris-HCl (pH 7.5), 65 mM MgCl<sub>2</sub>, 163 mM NaCl, 58 mM DTT, 116 mM sodium isocitrate, and 12 mM MnCl<sub>2</sub>, the mixture was distributed for each reaction which contained 25 nM of a ddNTP and 1.5 units of Sequenase Ver 2.0 (United States Biochemical) and was incubated at  $37^\circ\text{C}$  for 10 min. The reaction products were resolved in 80% formamide after ethanol precipitation.

The fragment spectra were obtained with a Hitachi SQ3000 DNA sequencer (Hitachi, Tokyo, Japan). The polyacrylamide gel concentration and migration length were 12% (containing 3% of *bis*-acrylamide) and 30 cm.

### 3.4. Sequence reaction

Linker-ligated  $\lambda$  DNA fragments (0.6  $\mu$ g), 0.05 pmol Texas Red-labeled primer, 0.8 units AmpliTaq (Perkin-Elmer Cetus Instruments; Norwalk, CT), and 3  $\mu$ l of buffer containing 50 mM Tris-HCl (pH 8.5) and 30  $\mu$ M MgCl<sub>2</sub> were mixed in a total volume of 15  $\mu$ l and distributed for each d/ddNTP mix (Takara). The reaction was overlaid with 10  $\mu$ l mineral oil. Denaturation was for 30 sec at  $95^\circ\text{C}$ , annealing was for 30 sec at  $68^\circ\text{C}$ , and polymerization was for 30 sec at  $77^\circ\text{C}$ . The cycle was repeated 30 times in a Perkin-Elmer Cetus thermal cycler.

### 3.5. Separation and detection of DNA fragments

DNA fragment spectra of the digested product as well as the sequencing reaction product were obtained with the highly sensitive fluorescent DNA analyzer, as reported previously.<sup>9</sup> A schematic view of the apparatus

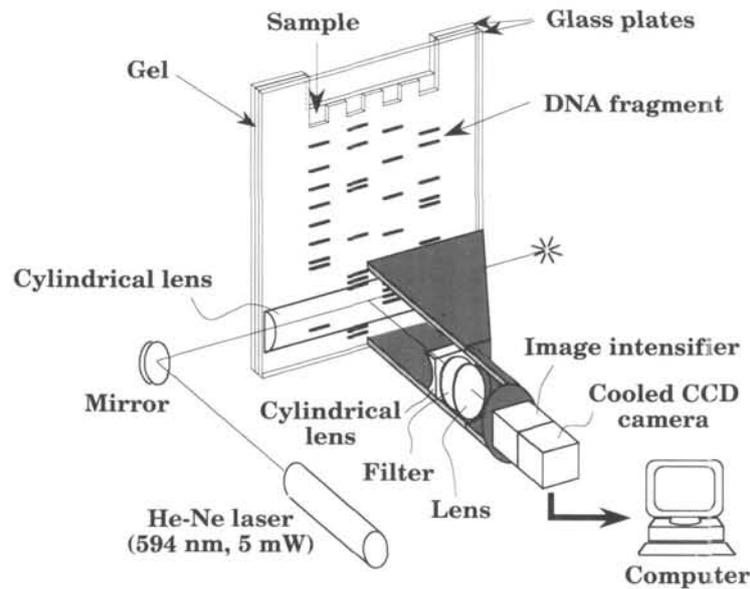


Figure 2. Schematic view of the highly sensitive fluorescent DNA analyzer.

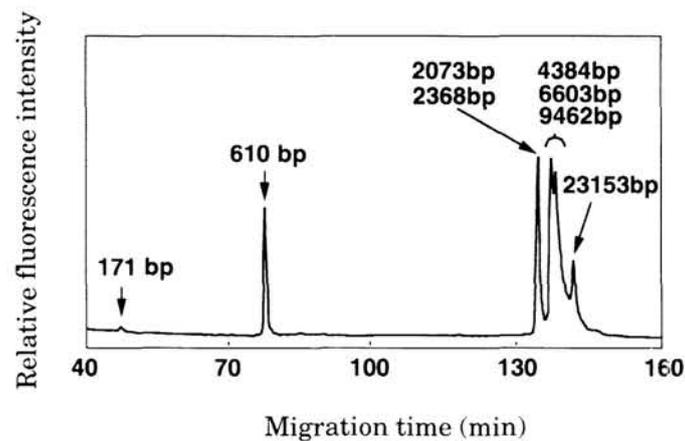


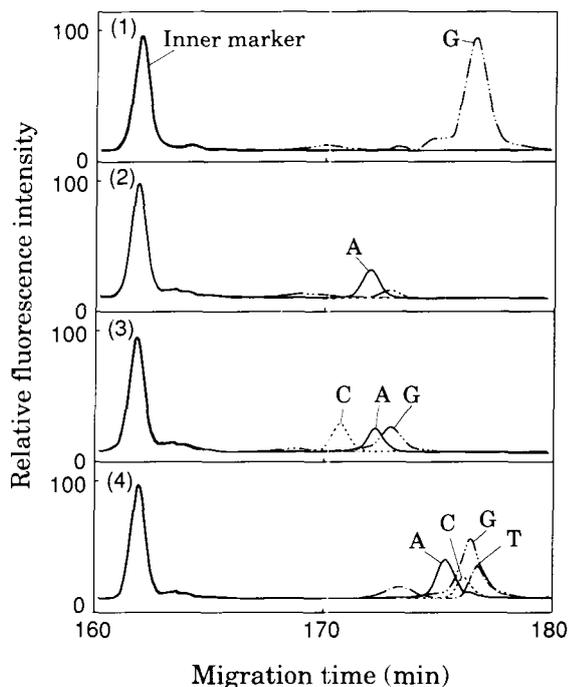
Figure 3. Fragment spectra of  $\lambda$  DNA digested by *Hind*III. Fragments were ligated with Texas Red-labeled oligonucleotides at the termini. The numbers in the figure include the oligonucleotide base number. Although most of the 171-bp fragment was lost during polyethyleneglycol precipitation, the amount of other fragments observed was sufficient for sequencing.

is shown in Fig. 2. It uses gel electrophoresis for the separation and laser-induced fluorescence for the detection. The labeling fluorophore is Texas Red which is efficiently excited by a He-Ne laser (594 nm) and emits a fluorescence maximum at 625 nm. The analyzer can detect labeled DNA at a concentration as small as  $3 \times 10^{-20}$  mol/band. The polyacrylamide gel concentrations and migration length were 3% (containing 5% *bis*-acrylamide)

urea-free and 13 cm for digested DNA fragment separation and 5% (containing 5% *bis*-acrylamide) containing 7 M urea and 30 cm for DNA sequencing.

#### 4. Results and Discussion

Too many DNA fragment species in a mixture are not favorable for direct sequencing from the mixture because



**Figure 4.** DNA fragment spectra obtained by termination reactions with one-base-selective primers. The 3' termini of the selective primers used are A, C, G and T for (1) (2) (3) and (4), respectively. The dideoxynucleotide species are distinguished by lines as —: ddATP, ---: ddCTP, ···: ddGTP, and -·-: ddTTP. The terminal two-base sequences of the mixed sample DNA were determined from the reacted selective primers' 3'-base and introduced ddNTP.

a primer may bind to multiple fragment species. The number of DNA fragment species in a mixture that is convenient for selective DNA sequencing is 6 or less. When the number exceeds 6, it is better to separate the fragments and fractionate them into several groups by gel electrophoresis.

There were five peaks in the fluorescent fragment spectrum of  $\lambda$  phage digested by *Hind*III (Fig. 3). The gel concentration and the migration length are 3% and 13 cm, respectively. This indicates that more than 10 ss-DNA fragments were in the mixture. As the digestion product should contain 8 double-stranded fragments, the resolution of the gel separation seemed to be insufficient for complete separation of the fragments.

Every sequence can be obtained by sequencing with the two selective bases. There are 16 primers and the corresponding 16 sequencing reactions are necessary for sequencing the fragments. However, redundant sequencing reactions can be removed by reading the two-base orders of the termini in advance. The terminal two-base orders were easily obtained as follows.

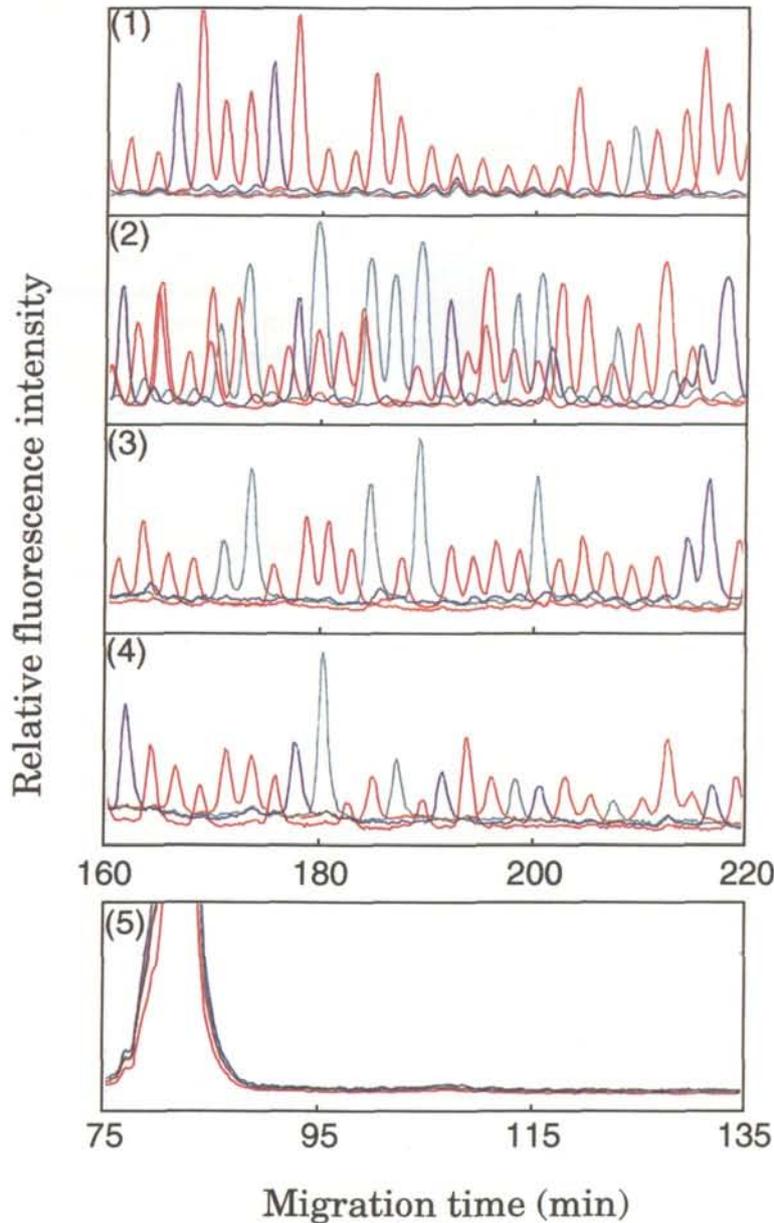
Four reactions corresponding to "A reaction," "C re-

action," "G reaction," and "T reaction" were carried out for each primer. The reactions were the complete termination reactions in which the reaction mixture of A reaction, for example, contained only ddATP. The extended and labeled fragments were observed only when the primers bound and the dideoxynucleotides were taken into their terminus. The DNA fragment spectra of the reaction products showed the approximate number of fragments on which the primers hybridized, and the base species next to the primers (that mean we found all of the terminal two-base sequences included in a mixture). The result is shown in Fig. 4. The spectra were obtained under a high resolution condition using a polyacrilamide gel with 12% (3% *bis*-acrylamide) and a 30-cm migration length.

The peak positions in the spectra are not unique but they depend on the terminal base species because their configuration and migration speed depend on the terminal base sequence. This result shows that there are at least 9 fragment species in the mixture and the sequences of the selective parts AG, CA, GA, GC, GG, TA, TC, TG and TT primers are fit for sequencing. Although several artifacts were observed in the spectra, for example the TT primer gave two peaks in the spectrum, this is not a serious problem because the spectra are only used for selecting primers where a peak is missing. As the peak intensities were normalized by the known amount of oligonucleotide (18-mer) added in the reaction mixture, the relative abundance of the corresponding targets can be deduced. For example, the extended products of GC, GA, and GG gave equal peak heights, which indicates the mixture contains equal amounts of the corresponding complementary strands. The extended product from TG is larger than those of TA and TT, which means that the mixture contains much more template DNA for TG primer than those for other primers.

The sequence reactions were carried out with these primers and the fragment mixture by the *Taq* cycle sequencing method. Parts of the resulting spectra are shown in Fig. 5. All the primers worked and DNA sequencing spectra were obtained except for a few exceptions. The selective primers with AG, CA, GA, GC, GG, TC, and TT termini gave fine spectra for DNA sequencing. The spectrum obtained with the TA primer seemed a mixed one. Therefore, four of the three-base-selective primers (TAA, TAC, TAG and TAT) were used for sequencing and clearly separated spectra were obtained with the TAG and TAT primers, but no DNA fragments were produced from the reactions using TAA and TAC. Similarly, the TG primer gave a three-signal-overlapped spectrum (data not shown). These could also be separated by using primers with more selective bases.

All the sequences obtained for the  $\lambda$  phage fragments agreed well with the reported  $\lambda$  phage sequences. However, the sequences of 125 base fragments (both strands)



**Figure 5.** Sequencing results of a mixture containing 16 DNA fragments. Red, blue, green and orange lines represent A, C, G and T families, respectively. (1) Sequence of a 23130-base fragment determined using the GC primer. (2) Mixed spectra of 564- and 2027-base fragments using the TA primer. (3), (4) Separated sequences of 564- and 2027-base fragments determined using the primers three selective bases (TAG and TAT, respectively). (5) Spectra obtained using the AT primer. This primer did not hybridize perfectly with any fragment in the mixture.

were missing. This is probably because these small base fragments may be lost during the polyethyleneglycol precipitation process. Actually, the amount of 125 base fragments was very small as is shown in Fig. 3.

## 5. Conclusion

DNA sequencing from an unknown mixture of DNA has been successfully demonstrated using  $\lambda$  phage digestion product as an example. Although this selective primer method was successfully used for sequencing DNA in a mixture containing less than 6 ssDNA fragment

species, any mixture containing a large number of DNA fragments will be sequenced by this method coupled with a pre-separation by gel electrophoresis. Size separation and fractionation of DNA fragments with 2–3 base accuracy is possible for up to 1 kb by gel electrophoresis. The DNA size accuracy of DNA fractionated by the gel electrophoresis is several kilobases or larger at long DNA (10–50 kb) regions. However, we can digest them again down to 1 kb or less with other enzymes and fractionate the products to reduce the number and size of the fragments in a mixture for DNA sequencing. The combination of this new method of selective primer sequencing and gel electrophoresis will become a convenient and powerful DNA sequencing method for a long or mixed DNA without cloning. It is also suitable for full automation of the whole sequencing process.

## References

1. Collins, F. and Galas, D. 1993, A new five-year plan for the U.S. human genome project, *Science*, **262**, 43–46.
2. Brumley, R. L., Jr. and Smith, L. M. 1991, Rapid DNA sequencing by horizontal ultrathin gel electrophoresis, *Nucleic Acids Res.*, **19**, 4121–4126.
3. Huang, X. C., Quesada, M. A., and Mathies, R. A. 1992, Capillary array electrophoresis using laser-excited confocal fluorescence detection, *Anal. Chem.*, **64**, 967–972.
4. Kambara, H. and Takahashi, S. 1993, Multiple-sheathflow capillary array DNA analyser, *Nature*, **361**, 565–566.
5. Takahashi, S., Murakami, K., Anazawa, T., and Kambara, H. 1994, Multiple sheath-flow gel capillary-array electrophoresis for multicolor fluorescent DNA detection, *Anal. Chem.*, **66**, 1021–1026.
6. Grothues, D., Voss, H., and Stegemann, J. 1993, Separation of up to 1000 bases on a modified A.L.F. DNA sequencer, *Nucleic Acids Res.*, **21**, 6042–6044.
7. Nishikawa, T. and Kambara, H. 1994, Separation of long DNA fragments by capillary gel electrophoresis with laser-induced fluorescence detection, *Electrophoresis*, **15**, 215–220.
8. Wada, A. 1987, Automated high-speed DNA sequencing, *Nature*, **325**, 771–772.
9. Kambara, H., Nagai, K., and Kawamoto, K. 1992, Photodestruction of fluorophores and optimum conditions for trace DNA detection by automated DNA sequencer, *Electrophoresis*, **13**, 542–546.

