CLIN. CHEM. 35/11, 2196–2201 (1989)

Automated DNA Sequencing Methods Involving Polymerase Chain Reaction

L. J. McBride,[1] S. M. Koepf,[1] R. A. Gibbs,[2] W. Salser,[3] P. E. Mayrand,[1] M. W. Hunkapiller,[1] and M. N. Kronick[1]

Polymerase chain reaction (PCR) as a method for preparing DNA templates has been used for several DNA sequencing applications. An in situ procedure for directly sequencing PCR products by the dideoxy-termination method has been developed by using fluorophore-labeled sequencing primers. Completed sequence reactions were combined and loaded into a single electrophoretic lane of a fluorescence-based DNA sequence analyzer. DNA targets devoid of a universal primer sequence could be sequenced with labeled universal primers by incorporating a universal primer sequence into the PCR product. With this method, the sequence of a 351-bp region in the bacteriophage lambda genome was fully analyzed in a single lane with automatic base identification accuracy of >99%. An unknown sequence, 1.7 kb long, also was sequenced by this procedure, in combination with a "PCR gene walking" strategy. Comparison of the 1110 bases in overlapping sequence data from both strands yielded only two single-base ambiguities. Automated DNA sequence analysis of the highly polymorphic HLA-DQA-1 (alpha) region in the human genome can be performed with this simple methodology. Use of this PCR-sequencing method to analyze DNA extracted from a one-month-old blood sample from an individual who is heterozygous at this locus allowed unambiguous assignment of genotype.

Additional Keyphrases: *dideoxy-termination procedure · fluorometry · "gene walking" · "fingerprinting"*

The advent of the dideoxy-termination method of DNA sequencing (*1*) has encouraged scientists to devise rational strategies to tackle large sequencing projects (*2*). Clearly, development of high-throughput methods suitable for automation is a reasonable focus for the future. Toward this goal, automated DNA sequence analysis has been achieved by combining fluorescence-based detection with dideoxy-termination (*3*). The other aspects of a sequencing project—cloning, preparing templates, and running sequencing reactions—are, however, complex manual operations that require significant labor and skill. Furthermore, the types of manipulations required for these manual methods vary with the type of template and therefore lack the repetitive nature conducive to automation.

Polymerase chain reaction (PCR), an in vitro DNA amplification scheme (*4*), has been proposed as a technology to automate template preparation before DNA sequencing (*5, 6*). The PCR process has been semi-automated largely by virtue of its cyclical nature, consisting of (*a*) denaturation of the DNA template (95 °C), (*b*) hybridization of deoxyoligonucleotide primers to the denatured template (50 °C), and (*c*) template replication (70 °C) catalyzed by *Thermus aquaticus* DNA polymerase (Taq polymerase). We have combined the PCR technology for DNA template preparation with the fluorescence-based method for automated DNA sequence analysis (*3, 7*) to automate more fully the multiple stages of DNA-sequencing projects.

Materials and Methods

Instrumentation and Reagents

For thermal cycling we used a DNA Thermal Cycler (Perkin-Elmer, Norwalk, CT 06859), generally according to the following cycle: 50 °C for 2 min, 70 °C for 2 min + 20 s autoextension, and 94 °C for 1 min. For DNA amplification we used solely the biochemical reagents in the "GeneAmp" DNA Amplification Reagent Kit, which included 10× reaction buffer, deoxynucleotide triphosphates, and Taq polymerase isolated from *Thermus aquaticus* (Perkin-Elmer Cetus, Norwalk, CT 06859). Light mineral oil was purchased from the Sigma Chemical Co. (St. Louis, MO 63178; cat. no. N-3516); sterile 0.5-mL tubes were from Robbins Scientific (Mountain View, CA 94043).

Fluorescence-based DNA sequence analyses were obtained by using a 370A DNA Sequencer [Applied Biosystems Inc. (ABI), Foster City, CA 94404] fitted with a 6% polyacrylamide gel and run with the manufacturer's version 1.3 software. Polyacrylamide gels were prepared on glass plates purchased from ABI. Modified T-7 DNA polymerase ("Sequenase" version 1, cat. no. 70700) and termination nucleotide mixes (cat. nos. 70714, 70716, 70718, 70720) were purchased from United States Biochemical (Cleveland, OH 44122). Sequenase (20 units) was diluted in water containing 50 mmol of dithiothrietol, 3.8 mmol of Tris, and 0.38 mmol of EDTA per liter. Lyophilized sequencing primers (−21 M13 and reverse M13, labeled with the four fluorophores, FAM, ROX, JOE, and TMRA) were purchased from ABI (*8*). Stock solutions of each fluorescent primer were prepared by dissolving a primer into 10 mmol/L Tris buffer (pH 8.0), containing 1.0 mmol of EDTA per liter, to give a final concentration of 0.8 μmol/L, and were stored in the dark at −20 °C.

DNA was extracted from whole blood with an ABI 340A Nucleic Acid Extractor (according to ABI User Bulletin no. 14, October, 1987). A 1.7-kb fragment of human DNA cloned into pGEM was donated by Dr. Craig Venter (National Institutes of Health, Bethesda, MD 20892).

We synthesized deoxyoligonucleotides by the phosphoramidite approach (*9*), using a 381A DNA Synthesizer (ABI) at 0.2 μmol scale (0.2 μmol cycle, version 1.23 software; ABI User Bulletin no. 7, August, 1986) with (2-*O*-cyanoethyl)-phosphoramidites (ABI). Oligonucleotide Purification Cartridges were purchased from ABI.

For ultraviolet absorbance spectroscopy we used a 8451A diode array spectrophotometer (Hewlett-Packard, Palo Alto, CA 94304).

[1] Applied Biosystems, 850 Lincoln Centre Dr., Foster City, CA 94404.

[2] Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030.

[3] Department of Biology, UCLA, Los Angeles, CA 90024.

Presented at the 21st annual Oak Ridge Conference on Advanced Analytical Concepts for the Clinical Laboratory, San Diego, CA, April 1989.

Received April 13, 1989; accepted August 17, 1989.

Primer	Identity	Sequence
101	Forward M13 -21	TGTAAAACGACGGCCAGT
102	Reverse M13	CAGGAAACAGCTATGACC
103	1st step forward	tgtaaaacgacggccagtTCAGAGCGATCGTGTTAA
104	2nd step forward	tgtaaaacgacggccagtGCTATGGCACACAGGTTC
105	3rd step forward	tgtaaaacgacggccagtGAGTAGAAACTACCAGT
106	1st step reverse	caggaaacagctatgaccTAACGGATAGCTTCTGC
107	2nd step reverse	caggaaacagctatgaccGCCTGCCACAAATTTCACT
108	3rd step reverse	caggaaacagctatgaccTGCGCATCGTCCACAACT
109	4th step reverse	caggaaacagctatgaccTCGCTGCTGATAAGTAG
110	lambda excess	CAGCTGCGTCGTTTGAC
111	lambda limiting	tgtaaaacgacggccagtGACTGATAGTGACCT
112	HLA-DQA-1	gtgctgcaGGTGTAAACTTGTACC
113	HLA-DQA-1	cacggatccGGTAGCAGCGGTAGAGT
114	HLA-DQA-1	tgtaaaacgacggccagtTTTACGGTCCCTCTGGC

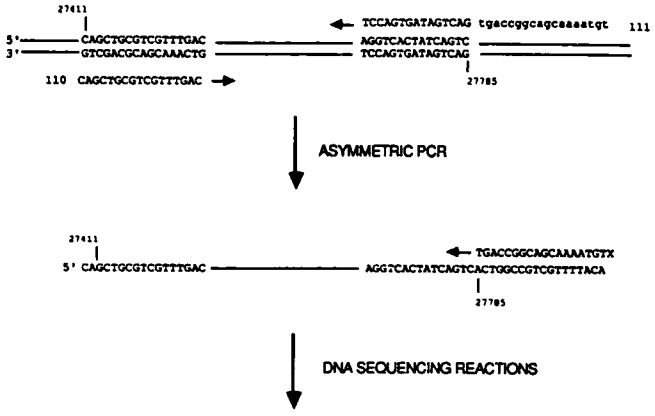
Fig. 1. 2'-Deoxyoligonucleotide primer sequences
Lower-case letters denote regions not complementary to the target sequence

Procedures

Preparation of PCR primers. After incubation in concentrated aqueous NH_3 at 57 °C for 20 h, one half of the crude primer (1.5 mL in NH_3) was enriched on an Oligonucleotide Purification Cartridge as described elsewhere (*10*). Stock solutions of the limiting PCR primers (Figure 1: 101–109, 111, 112, and 114) were prepared by evaporating to dryness an aliquot (approximately 5%) of the enriched product and dissolving the residue in 1 mL of de-ionized water. We determined the concentration spectrophotometrically, assuming that an absorbance of 1.0 *A* at 260 nm indicated the presence of 33 μg of DNA. Final concentrations were between 0.5 and 1.0 μmol/L. Stock solutions of the PCR primer used in excess (Figure 1: 101, 102, 110, and 113) were prepared similarly to give final concentrations of between 20 and 50 μmol/L.

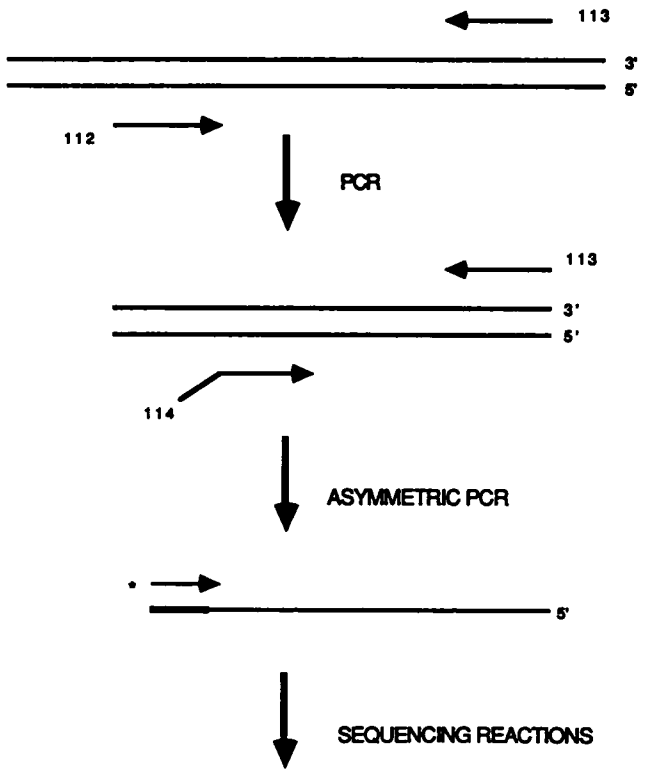
DNA amplification. An example of the general procedure for preparing a single-stranded DNA template (*11*) is as follows: Heat 1 pg to 1 ng of bacteriophage lambda DNA (Figure 2) on a Thermal Cycler at 95 °C for 3 min in the presence of 100 μL of PCR cocktail (1× reaction buffer containing 0.02 nmol of each deoxynucleoside triphosphate, 1.0 pmol of PCR primer 111, and 50 pmol of PCR primer 110), after overlaying with 100 μL of light mineral oil. Then add Taq polymerase (0.5 μL, 2.5 units) to the aqueous phase at 50 °C and perform 35 thermal cycles. Store the aqueous phase at −20 °C until further use.

We modified the above procedure for amplification of human genomic DNA. We amplified the HLA-DQA-1 region (*11*) to single-stranded form with simultaneous incorporation of a universal priming sequence in a two-step procedure (Figure 3) as follows: Dissolve 2 μg of purified genomic DNA in 10 μL of de-ionized water and add this



DNA SEQUENCE ANALYSIS

Fig. 2. PCR-sequencing model system in the bacteriophage lambda genome
Primer 111 (*upper right*) contains the −21 M13 universal sequence denoted by *lower-case letters*. *X* (middle diagram) denotes a 5'-fluorophore-labeled primer



SEQUENCE ANALYSIS

Fig. 3. PCR scheme used to sequence the human HLA-DQA-1 locus (*12*)
The fluorophore-labeled sequencing primer is marked with an *asterisk*

solution to 100 μL of PCR cocktail (1× reaction buffer containing 0.02 nmol of each deoxynucleoside triphosphate, 5.0 pmol of PCR primer 112, and 20 pmol of PCR primer 113). After overlaying the solution with 100 μL of mineral oil, heat for 3 min at 95 °C. After subjecting the sample to 30 thermal cycles (55 °C, 2 min; 72 °C, 1 min; and 94 °C, 1 min), transfer 1.0 μL of the aqueous phase to 100 μL of a similar PCR cocktail containing 3.0 pmol of primer 114 and 50 pmol of primer 113. Subject this solution to 20 thermal cycles (50 °C, 2 min; 72 °C, 1.5 min + 3 s autoextension; 96 °C, 45 s), and store the resulting aqueous phase at −20 °C.

DNA sequencing. For the dideoxycytidine sequencing reaction we added 3 μL of the PCR solution in situ to 1.5 μL of a solution containing 0.4 pmol of fluorescein-labeled (FAM) −21 M13 primer and a final concentration of sequencing buffer (10 mmol/L Tris-HCl buffer, pH 8.5, containing 10 mmol of $MgCl_2$ and 50 mmol of NaCl per liter). After incubation at 90 °C for 5 min, we cooled the solution on ice for 2 min, lightly centrifuged, and added 2 μL of 2′-deoxy/2′,3′-dideoxycytidine-5′-triphosphate mix and 1.5 μL of freshly diluted Sequenase. This solution was incubated at 37 °C for 5 min, then quenched by heating at 65 °C for 10 min. The reaction mixture was manually transferred into heating blocks pre-equilibrated at the appropriate temperatures. The three other sequencing reactions for dideoxyguanosine, dideoxyadenosine, and dideoxythymidine were performed in parallel (with TMRA, JOE, and ROX primers, respectively) in nearly identical fashion (dideoxyguanosine and dideoxythymidine reactions required twice the volume of all ingredients). The four sequencing reactions were cooled to 4 °C, pooled, precipitated with ethanol, resuspended in 6 μL of a solution of formamide/EDTA, 50 mmol/L (5/1 by vol), and loaded into a single electrophoretic lane.

Results and Discussion

The dideoxy-termination method of DNA sequencing is generally accepted as being more robust with single-stranded templates than with double-stranded ones. Therefore, we developed a fluorescence-based sequencing procedure, using single-stranded DNA generated by asymmetric amplification (*12*), a variant of PCR. By using unequal molar amounts of the two amplification primers, it is possible to produce an excess of single-stranded DNA of a chosen strand for direct sequencing. We introduced a universal primer sequence into the PCR product (Figure 2) by synthesizing the limiting PCR primer (111) to contain the −21 M13 universal sequence (*14*). This sequence incorporation method allowed us to use commercially available universal fluorescent primers for sequencing DNA targets originally devoid of universal sequences. Four aliquots of the PCR solution (3 μL each for the sequencing primers labeled with fluorescein derivatives FAM and JOE, and 6 μL each for the primers derivatized with rhodamine deriv-

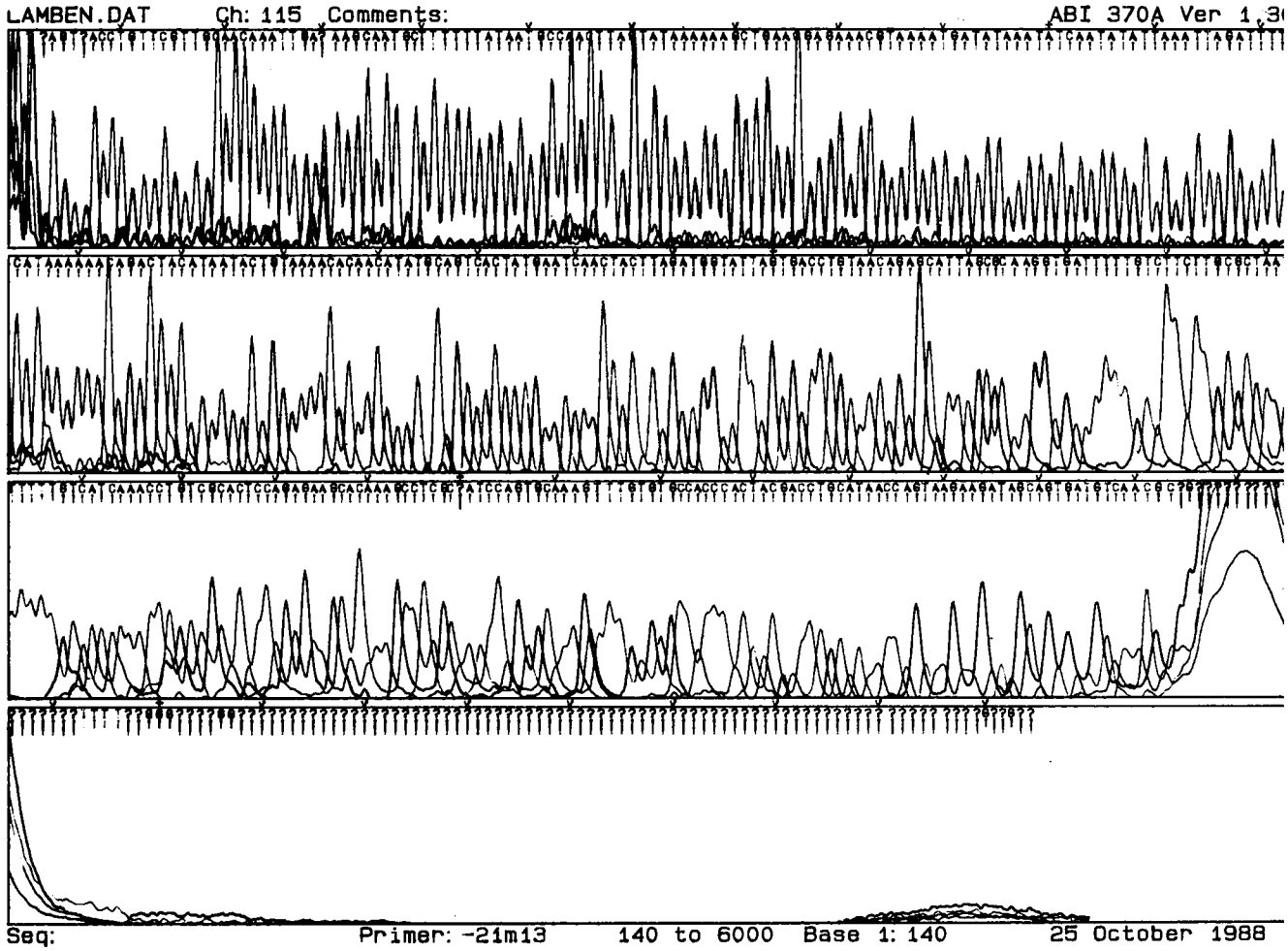


Fig. 4. Example of automated sequence analysis: a 351-bp region in the bacteriophage lambda genome

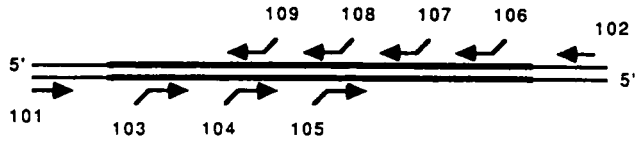


Fig. 5. "PCR gene walking" scheme performed on both strands of a 1.7-kb DNA fragment inserted within pGEM

For the *top* strand, PCR primers 101 and 103–5 were used as limiting primers during amplification, along with the primer in excess, 102. For the *bottom* strand, primers 106–9 were used similarly, along with primer 101. The sequences of these primers are listed in Fig. 1

atives TMRA and ROX) were added directly, without purification, to sequencing reactions. The 351-bp bacteriophage lambda sequence between the PCR primers then was analyzed completely in a single lane with >99% accurate base calling by an automated sequence analyzer with fluorescence detection (Figure 4).

We applied this sequencing procedure without modification (other than the use of different PCR primers) to both strands of an unknown 1.7-kb sequence inserted into a pGEM plasmid (Figure 5). Both strands were amplified asymmetrically. These amplifications used both the forward (101) and reverse (102) universal M13 primers (Figure 1) in either a 50:1 or a 1:50 ratio. After performing the amplifications, we sequenced the PCR products by using the appropriate fluorophore-labeled universal sequencing primers. The sequence information obtained from each strand was used to synthesize two new PCR primers, each approximately 300 bp into the insert from each end. These new primers, 103 and 106, were synthesized to contain one of the two universal primer sequences and were used for a second round of amplification with the universal M13 primers (102 and 101, respectively). These PCR products were then sequenced with the appropriately labeled universal primers. This second round of PCR-sequencing gave an additional 300 bases of sequence information, constituting the second "step" in our "PCR gene walking" method. Overlapping sequence information totaling approximately 1100 bases was obtained after four steps along each strand (Figure 6). Only two single base-pair positions within this 1100-bp region could not be identified with confidence by referring to the analyzed data from both strands.

This gene-walking method required no procedural modification from step to step or synthesis of alternative PCR primers. Significant primer-to-primer variability is encountered during traditional gene-walking methodologies, presumably arising from secondary priming sites. The secondary priming sites of the custom-made PCR primers during this PCR gene-walking method described did not noticeably interfere with the DNA sequencing. However, gene-walking by this in situ strategy likely will be limited to relatively small inserts (<2 kb) in light of the fact that amplification of larger inserts can yield products of variable quantity and purity.

Inverted PCR (*13, 14*) may allow sequencing of unknown regions of DNA much larger than the 1.7-kb region we sequenced. This PCR method requires knowledge of only one primer sequence flanking the unknown DNA sequence of interest; after the linear DNA target is cut, diluted, and then treated with DNA ligase, the circular product is isolated by gel electrophoresis. Formation of nicks in the circular product allows the "sense" and "anti-sense" primer sequences to be used for amplification. Inverted PCR could potentially allow for sequential sequencing of larger sections of DNA, e.g., 10^3 bp to genomic, which would greatly

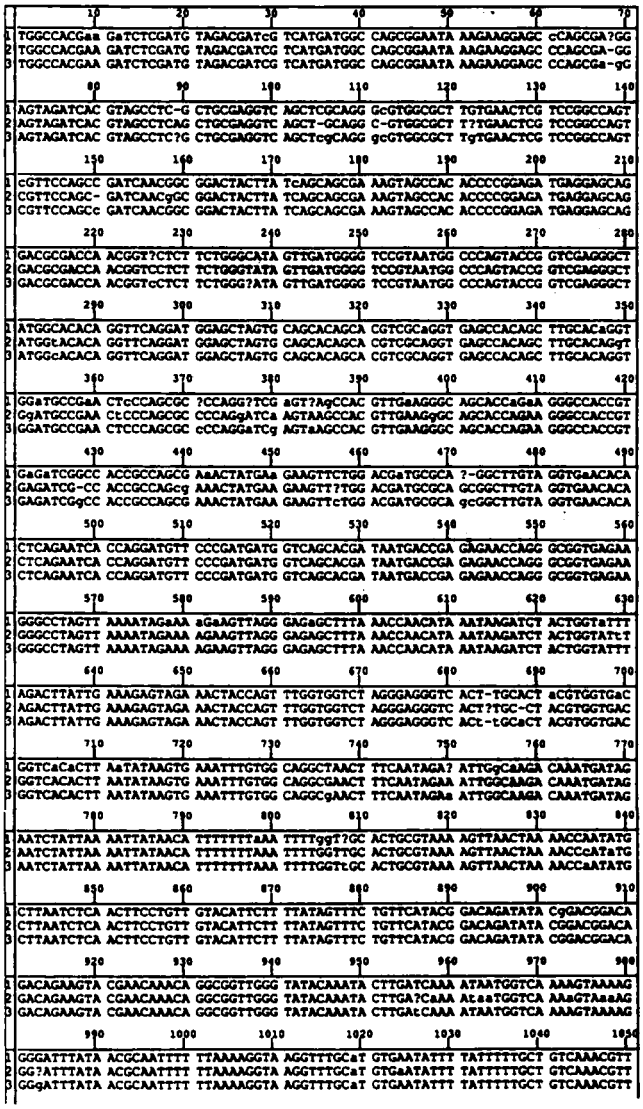


Fig. 6. Sequence information obtained from "PCR gene walking" on both strands of a 1.7-kb insert within pGEM

Sequence 1 is from the M13 −21 sequencing primer. Sequence 2 is the reverse complement of the sequence from the M13 reverse-sequencing primer. Sequence 3 is the consensus sequence derived by manually comparing sequences 1 and 2. For sequences 1 and 2, *upper-case letters* denote bases identified automatically, while *lower-case letters* denote manual identifications that could be made at positions where questions marks were initially assigned in the automated base-identification routine. For sequence 3, *upper-case letters* denote agreement at that position from both strands, while *lower-case letters* indicate that identification could be determined with confidence only upon manual evaluation of the analyzed data from both strands

decrease the overall amount of cloning, clone ordering, and sequence analysis needed for sequencing projects. This methodology might be achieved by use of an array of automated methods—polymerase chain reaction, preparative gel electrophoresis, automated DNA sequence analysis, and a liquid-handling robot to perform, among other things, restriction enzyme digestions and ligations. This automated strategy would greatly minimize or even eliminate the need to clone templates for many sequencing projects.

The in situ nature of the PCR-sequencing procedure (*5, 6*) we developed here (*15*) is an attractive feature for large sequencing projects, by virtue of its potential automatability. Manual steps during PCR amplification (sample prep-

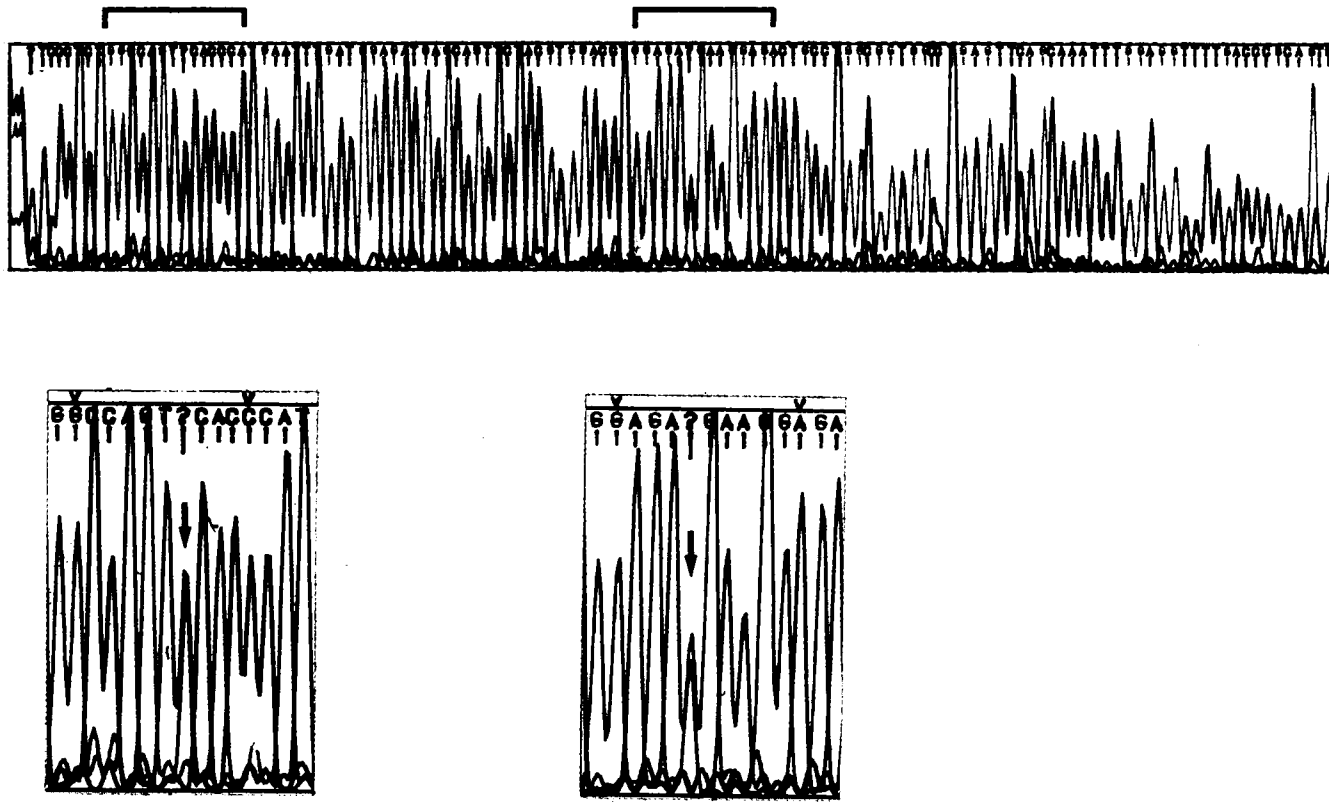


Fig. 7. (*Upper panel*) Automated sequence analysis of a 120-bp region in the human HLA-DQA-1 locus; (*lower panels*) enlargements of the regions in the upper panel around the two question marks, which automatically indicate the presence of polymorphic base positions: A/T (*lower left*), and G/A (*lower right*)

aration) and sequencing reactions essentially have been reduced to handling of small volumes, 1–10 μL, which could be delegated to commercially available liquid-handling robots (*16*).

An automated strategy for rapid DNA sequencing should also facilitate the study of regions within the human genome of medical importance (*17, 18*). These sequencing applications (*18*) necessitate rapid and reliable discrimination of single nucleotide residues relative to normal DNA sequences in both homozygous and heterozygous individuals. Figure 7 illustrates an example of this type of automated sequencing ability on the highly polymorphic HLA-DQA-1 locus within the human genome. For sequence analysis of 120 bases within the HLA-DQA-1 region we started with 2 μg of genomic DNA extracted from blood that had been stored for a month. A two-step PCR amplification method was used (*15*). The HLA-DQA-1 sequence information shown in Figure 7 suggested that this individual was heterozygous at this locus, as indicated by two nucleotide positions marked with question marks. Each of these questions marks, posted from the automatic base-calling software, resulted from the receipt of two overlapping signals of nearly equal intensity. The pairs of overlapping peaks flagged by the two question marks were consistent with nucleotide positions reported to be polymorphic (*12*). Given this sequencing, this individual subsequently was assigned an A1.2/A1.3 genotype (*12*). Our observation that these polymorphic nucleotide positions yielded nearly equivalent peak intensities illustrates our ability to flag automatically a 1:1 mixture of alleles with use of fluorescent-based sequence analysis. This type of automatic analysis may have useful research applications in carrier detection of genetic diseases, as well as in DNA fingerprinting.

We chose this two-step PCR approach for human DNA primarily to minimize the formation of secondary PCR products during amplification of this complex DNA sample. Unwanted PCR products could yield unwanted signal production during in situ DNA sequencing, necessitating gel purification of the single-stranded PCR product before sequencing. This two-step approach also should yield 1000- to 10 000-fold greater amplification, offering the possibility of amplifying genomic sequences directly from <1 μL of unpurified blood. Ultimately, these methodologies should lead to fully automated repetitive DNA sequence analysis of genomic DNA, starting directly from whole blood or tissue.

We thank Curt Becker, Bob Ellis, Dr. Henry Erlich, Dawn Madden, Marilee Shaffer, and Dr. Anne Wan for helpful discussions, and Bill Giusti for oligonucleotide synthesis. Mr. Kevin Clarke donated whole blood for direct sequencing.

References

1. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 1977;74:5463–7.
2. Hood LE, Hunkapiller MW, Smith LM. Automated DNA sequencing and analysis of the human genome. Genomics 1987;1:201–12.
3. Smith LM, Sanders JZ, Kaiser RJ, et al. Fluorescence detection in automated DNA sequence analysis. Nature (London) 1986;321:674–9.
4. Mullis KB, Faloona FA. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. Methods Enzymol 1987;155:335–50.

5. Stoflet ES, Koeberl DD, Sarkar G, Sommer SS. Genomic amplification with transcript sequencing. Science 1988;239:491–4.
6. Innis MA, Myambo KB, Gelfand DH, Brow MAD. DNA sequencing with *Thermus aquaticus* DNA polymerase, and direct sequencing of PCR-amplified DNA. Proc Natl Acad Sci USA 1988;85:9436–40.
7. Connell C, Fung S, Heiner C, et al. Automated DNA sequence analysis. Biotechniques 1987;5:342–8.
8. Fung S, Woo SL, Menchen SM, Connell CR, Heiner C. Method of detecting electrophoretically separated oligonucleotides. Eur. patient appl. no. 0233053, 1986.
9. Caruthers MH, Barone AD, Beaucage SL, et al. Chemical synthesis of deoxyoligonucleotides by the phosphoramidite approach. Methods Enzymol 1987;154:287–313.
10. McBride LJ, McCollum C, Davidson S, Efcavitch JW, Andrus A, Lombardi SJ. A new, reliable cartridge for the rapid purification of synthetic DNA. Biotechniques 1988;6:362–7.
11. Scharf SJ, Horn GT, Erlich HA. Direct cloning and sequence analysis of enzymatically amplified genomic sequences. Science 1986;233:1076–8.
12. Gyllensten UB, Erlich HA. Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. Proc Natl Acad Sci USA 1988;85:7652–6.
13. Ochman H, Gerber AS, Hartl DL. Genetic applications of an inverse polymerase chain reaction. Genetics 1988;120:621–3.
14. Triglia T, Peterson MG, Kemp DJ. A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. Nucl Acids Res 1988;16:8186.
15. McBride LJ, Koepf SM, Gibbs RA, et al. Interfacing in vitro DNA amplification with DNA sequencing strategies. Curr Commun Mol Biol 1989 (in press).
16. Wilson RK, Yuen AS, Clark SM, Spence C, Arakelian P, Hood LE. Automation of dideoxynucleotide DNA sequencing reactions using a robotic workstation. Biotechniques 1988;6:776–87.
17. Caskey CT. Disease diagnosis by recombinant DNA methods. Science 1987;236:1223–9.
18. Gibbs RA, Nguyen P, McBride LJ, Koepf SM, Caskey CT. Identification of mutations leading to the Lesch–Nyhan syndrome by automated direct DNA sequencing of in vitro amplified cDNA. Proc Natl Acad Sci USA 1989;86:1919–23.