
Specific-primer-directed DNA sequencing using automated fluorescence detection

Robert J.Kaiser, Sara L.MacKellar, Ravi S.Vinayak, Jane Z.Sanders, Raul A.Saavedra and Leroy E.Hood

Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

Received April 11, 1989; Revised and Accepted June 22, 1989

ABSTRACT

Automated fluorescence-based DNA sequence analysis offers the possibility to undertake very large scale sequencing projects. Directed strategies, such as the specific-primer-directed sequencing approach ('gene walking'), should prove useful in such projects. Described herein is a study involving the use of this approach in conjunction with automated fluorescence detection on a commercial instrument (ABI 370A DNA sequencer). This includes procedures for the rapid chemical synthesis and purification of labeled primers, the design of primer sequences that are compatible with the commercial analysis software, and automated DNA sequence analysis using such primers. A set of four fluorophore-labeled primers can be reliably synthesized in a twenty-four hour period, and greater than 300 nucleotides of analyzed new sequence obtained using this set in an additional twenty-four hours. Scale-up of these procedures to take advantage of the full capabilities of the sequencer is, at present, too slow and costly to be suitable for routine sequencing, and therefore the use of specific-primers is best suited to the closure of gaps in extended sequence produced using random cloning and sequencing strategies.

INTRODUCTION

Several approaches to the automation of DNA sequence analysis have been described recently (1-6). One approach, initially developed in our group (1), utilizes a set of four discriminable fluorescent dyes covalently attached to the oligonucleotide used as a primer in enzymatic DNA sequencing (7). A different color fluorophore is used to color code each of the four sequencing reactions, A, C, G, and T. The products of the four reactions are co-electrophoresed in a polyacrylamide gel. The DNA sequence is automatically computed from the temporal order in which the different colored bands of DNA pass a high-sensitivity fluorescence detector located near the bottom of the gel during the electrophoretic process. A commercial instrument has been developed based on this general strategy (Model 370A Automated DNA Sequencer; Applied Biosystems, Inc., Foster City, CA).

Automated sequencing technologies such as the one described above offer great potential for the rapid accumulation of DNA sequence information. Large-scale sequencing projects will be typically carried out by first cloning segments of DNA into vectors capable of handling large inserts such as cosmids. These clones will then be sequenced primarily using so-called 'shotgun' techniques. This involves the random fragmentation of the cosmid clones into pieces suitable for subcloning into a standard vector, such as the filamentous bacteriophage M13, and randomly selecting and sequencing these subclones employing the appropriate vector-derived 'universal' primer. These random techniques yield greater than 90% of the desired sequence as overlapping sets of contigs, albeit with much repeated

effort. The sequence information required to fill in the gaps may be obtained by a directed sequencing approach. One such approach is 'specific-primer-directed' sequencing (8). In this procedure, previously obtained sequence information about an unknown insert using a 'universal' primer is employed to generate a new insert-specific oligonucleotide for further enzymatic sequence analysis. Successive cycles of sequencing and new primer generation yield the complete sequence of the DNA fragment of interest. Since this approach demands that a new oligonucleotide be synthesized for every round of sequencing, it necessitates the synthesis and purification of a set of four fluorophore-labeled primers for each cycle of sequence analysis when using the previously described automated sequencer.

The potential utility of specific-primer-directed sequencing prompted us to investigate the requirements for using this approach in conjunction with the ABI 370A DNA sequencer. To this end, we have examined procedures for the rapid synthesis of fluorophore-labeled oligonucleotides, and have tested several such oligonucleotides as specific primers in automated sequencing.

MATERIALS AND METHODS

HPLC-grade solvents (N,N-dimethylformamide, pyridine, dichloromethane, and acetonitrile) for use in chemical synthesis were purchased from Aldrich. Acetonitrile for HPLC was purchased from J.T.Baker. Activated fluorophores for oligonucleotide labeling (Fam-NHS, Joe-NHS, Tamra-NHS, and Rox-NHS) and fluorophore-labeled 'universal' M13 sequencing primer were obtained from Applied Biosystems. All other materials were reagent grade and were used as received unless otherwise indicated.

Synthesis of fluorescent sequencing primers.

Linkers for 5'-fluorophore conjugation. 9-Fluorenylmethyl-oxycarbonyl (Fmoc) protected 2-aminoethanol (C2) and 6-amino-1-hexanol (C6) were synthesized largely as described (9). The crude solids were crystallized twice, N-Fmoc-2-amino-ethanol from 5:1 (v/v) chloroform:hexane (mp. 128–130°C, white fluffy needles), and N-Fmoc-6-amino-1-hexanol from acetone (mp. 142–144°C, fine colorless prisms). Both were homogeneous by thin layer chromatography (TLC) on silica gel (60:40:1 hexane:ethyl acetate:triethylamine; C2: R_f 0.16; C6: R_f 0.23). The corresponding cyanoethyl diisopropyl (CED) phosphoramidites were synthesized by a modification of the standard procedure (10) to minimize untimely removal of the Fmoc group. N-Fmoc-amino alcohol (1.50 mmoles, 0.44 g C2, 0.52 g C6) was suspended in dry dichloromethane (20 ml, HPLC-grade dried by passage over basic alumina) under argon, and dry N,N-diisopropylethylamine (0.78 ml, 4.50 mmoles, dried by refluxing over and distillation from calcium hydride; Aldrich) was added, followed immediately by 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (0.67 ml, 3.00 mmoles; Aldrich). The solid rapidly dissolved to give a clear solution. The reaction was stirred 15 minutes at room temperature, and was diluted with ethyl acetate (100 ml, previously washed with ice cold 5% (w/v) aqueous sodium bicarbonate). The solution was extracted with cold 5% (w/v) aqueous sodium bicarbonate (2×50 ml) and with cold saturated aqueous sodium chloride (50 ml). It was dried over anhydrous sodium sulfate, filtered, and evaporated to a pale yellow oil under reduced pressure (water aspirator) below 35°C. The crude product was dried overnight *in vacuo* over KOH pellets, and used immediately the next day after dissolution to 0.15 M in anhydrous acetonitrile (Applied Biosystems, Inc.). Alternatively, the crude phosphor-amidites were purified by flash chromatography (11) on silica gel using 60:35:5 hexane:ethyl acetate:pyridine as the eluant. Appropriately pure fractions (TLC on silica

gel in 60:40:1 (v/v/v) hexane:ethyl acetate:triethylamine; C2: R_f 0.67; C6: R_f 0.72) were combined, evaporated to dryness, and further dried overnight *in vacuo* over KOH pellets. The clear colorless oils so obtained (in about 90% yield) may be stored desiccated under argon at -70°C . They are stable for at least a month under these conditions.

Oligonucleotide synthesis. Oligonucleotide primers were synthesized on a 1 μmole scale on an Applied Biosystems 380A DNA synthesizer using standard protocols and reagents, and CED phosphoramidites. At the completion of the synthesis, they were retained on the controlled pore glass support using the 'Trityl On' option for further reaction. The support, in its synthesis cartridge, was then dried *in vacuo*, removed from the cartridge, and split into two approximately equal lots for reaction with the two amino alcohol linkers. This was accomplished manually, using a 60-fold excess of N-Fmoc-amino alcohol phosphoramidite (0.2 ml of 0.15 M in anhydrous acetonitrile) and a 200-fold excess of 1H-tetrazole (0.2 ml of 0.5 M in anhydrous acetonitrile), with a 5 minute coupling and a 2 minute oxidation. The 5'-amino oligonucleotides were then cleaved from the support and deprotected using concentrated ammonium hydroxide for 4 hours at room temperature, then overnight at 60°C . The DNA solutions were decanted and evaporated to dryness in a SpeedVac Concentrator (Savant Instruments, Farmingdale, NY), and the resulting dry pellets were dissolved in water (500 μl) and stored frozen.

Alternatively, oligonucleotides were prepared on the automated synthesizer using the more labile base-protecting groups of Schulhof et. al. (12,13). Protected nucleoside phosphoramidites were synthesized essentially as described in the above reference. DMT-dA(phenoxyacetyl) CED phosphoramidite and DMT-dC(isobutyryl) CED phosphoramidite were purified by flash chromatography on silica gel using 9:1 (v/v) ethyl acetate:triethylamine as eluant. No suitable solvent system could be found for the flash chromatographic purification of DMT-dG(phenoxyacetyl) due to its more polar nature. All amidites were obtained as white powders following precipitation into hexane at -70°C , and were stored desiccated under argon at -20°C . These phosphoramidites were also subsequently purchased as a kit under the name 'CED phosphoramidites with labile base protection' from Sigma Chemical Co. (St. Louis, MO). Oligonucleotides synthesized using these phosphoramidites were derivatized with the amino alcohol linkers as described above, then cleaved and deprotected with concentrated ammonium hydroxide for four hours at room temperature. The DNA solutions were decanted, evaporated to dryness and the pellets dissolved in water (500 μl) and stored frozen.

Synthesis of fluorophore-conjugated oligonucleotides. Fluorophore-labeled primers were synthesized using the following general procedure. Crude 5'-amino oligonucleotide (50 μl , about 50 nmoles) and 1 M sodium bicarbonate, pH 9.0 (20 μl) are mixed, and activated fluorophore (6 μl of a solution of 5 mgs in 60 μl of dimethyl sulfoxide, dried by vacuum distillation from calcium hydride and storage over 4\AA molecular sieves) is added. These quantities represent a 10- to 20-fold excess of fluorophore over amino oligonucleotide. The C6 linker is used for the reaction with Fam-NHS and Joe-NHS, while the C2 linker is used for Tamra-NHS and Rox-NHS. The mixture is vortexed briefly to give a clear solution, and the reaction is allowed to proceed at room temperature for the desired period of time (2 hours to overnight) in the dark. The reaction is terminated by the addition of 4 M ammonium acetate, pH 7 (50 μl). Initial purification is achieved by gel filtration chromatography using a NAP 25 column (Pharmacia, Uppsala, Sweden) with 0.1 M ammonium acetate, pH 7, as eluant. The fast moving fluorescent material eluted in the void volume is collected (about 2 ml) and further purified by reverse phase HPLC on

an Axxiom 5 micron ODS (C18) column (4.6 mm×250 mm) using a flow rate of 1.0 ml/minute and a biphasic linear gradient of 10% buffer B:90% buffer A to 30% buffer B:70% buffer A in 30 minutes, then to 70% buffer B:30% buffer A in 15 minutes. Buffer A is 98:2 (v/v) 0.1 M triethylammonium acetate, pH 7:acetonitrile, and buffer B is 50:50 (v/v) 0.1 M triethylammonium acetate, pH 7:acetonitrile. The labeled primer is collected and evaporated to dryness in the dark, and the resulting colored pellet is redissolved in 10 mM Tris, 1 mM EDTA, pH 8.0 (TE) at a concentration of a few pmoles/ μ l. It can be stored at -20°C in the dark for several weeks.

Fluorescence-based DNA sequencing.

Templates. Single stranded M13(mp18) bacteriophage clones containing cDNA inserts of the gene encoding the shark myelin protein PO were obtained as previously described (14). Two such clones, designated PO1 and PO2, containing complementary inserts of about 2.8 Kb were used for DNA sequencing.

Selection of Primer Sequence. Primer sequences were chosen to fit several criteria, generally as recommended by Strauss et. al. (8). A total length of 23–25 nucleotides was judged to be the maximum, primarily for ease of purification by reverse phase HPLC. In order to be compatible with the sequence analysis software, the five bases at the 5'-end of each oligonucleotide were chosen to be identical to the five bases at the 5'-end of the Applied Biosystems universal M13 primer (see Results and Discussion), without regard to complementarity at these positions. The hybridizing portions were thus 18–20 nucleotides, of sufficient length to minimize priming at secondary sites. The sequences were selected to be complementary to a region of previous sequence deemed to be highly reliable, that is, free from compressions, possessing good, uniform signal strength and resolution, and completely analyzed by the instrument without the inclusion of ambiguous base assignments (as represented by '?'s in the computer output). In general this region was about 300 bases from the previous priming site. Sequences ending in G or C at the 3'-terminus were somewhat preferred over those ending in A or T. Sequences possessing a high degree of self-complementarity were avoided. Candidate sequences thus chosen were submitted to computer analysis for complementarity to the M13(mp18) vector sequence. Only sequences having less than 50% complementarity along their entire length were deemed suitable, since this low degree of complementarity should not allow stable hybrids to form under the conditions of the enzymatic sequencing reactions. In no case did this result in a primer having an unacceptably high degree of complementarity to the M13 sequence at its critical 3'-end, where even poor hybridization might cause some non-specific sequence generation and considerably complicate the automated analysis of the fluorescence data (3).

Enzymatic Sequencing Reactions. Fluorescence-based sequencing reactions using modified T7 DNA polymerase ('Sequenase', obtained from United States Biochemical Corporation, Cleveland, OH) were performed essentially according to the Applied Biosystems protocol accompanying the 370A DNA sequencer. The only changes were that 0.2 pmole of template DNA was used for each of the Fam-C and Joe-A reactions and 0.4 pmole was used for each of the Tamra-G and Rox-T reactions, and that the addition of EDTA was used to terminate the reactions rather than elevated temperature. These changes were made in an effort to optimize the conditions which would yield the best sequencing data from the instrument. A 2:1 molar ratio of fluorophore-labeled primer to template was employed. The electrophoresis was performed in 6% polyacrylamide gels at 24 to 29 watts for 14 hours. Data collection began immediately after the primer peak was observed at the detector

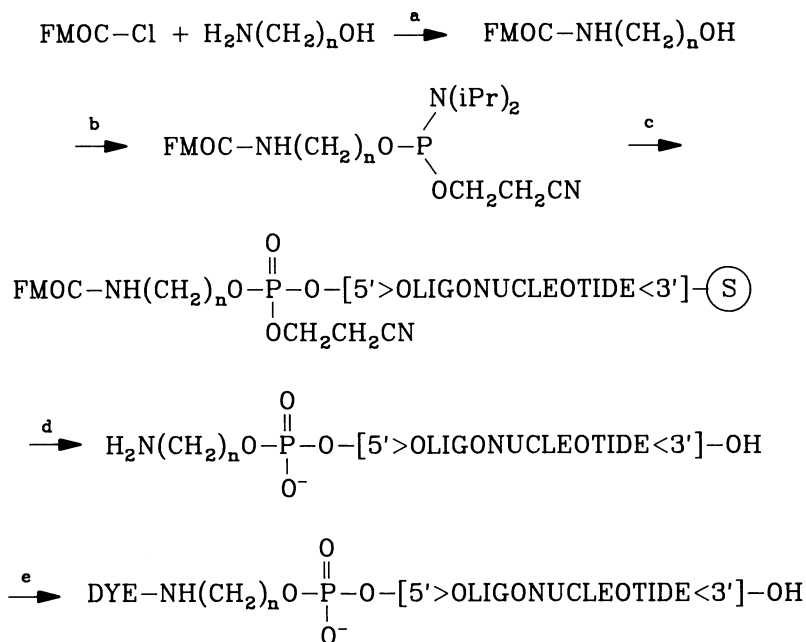


Figure 1. Schematic representation of the synthesis of the fluorophore-labeled primers used in this study. Details are presented in the text. (a) aqueous sodium carbonate; (b) 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite/ N,N-diisopropylethylamine/dichloromethane; (c) 1H-tetrazole/acetonitrile followed by iodine/lutidine/tetrahydrofuran/water; (d) concentrated ammonium hydroxide; (e) N-hydroxysuccinimide ester of the fluorophore (dye) in aqueous sodium bicarbonate, pH 9. FMOC = 9-fluorenylmethoxycarbonyl. n = 6 for the dyes Fam and Joe, and n = 2 for the dyes Tamra and Rox.

(about 90 minutes), and continued for 10 hours (6000 data points collected). At 28 watts, data corresponding to an average of about 400 bases of reliable sequence is obtained during this period. Analysis of the raw data was performed using the commercial ABI software, version 1.2.

RESULTS AND DISCUSSION

Two basic requirements must be fulfilled in order to permit the routine application of the specific-primer strategy to automated DNA sequencing using the ABI 370A. First, sets of purified fluorophore-labeled oligonucleotides must be easily, reliably and rapidly synthesized. Second, these oligonucleotides must not only serve as efficient primers for enzymatic DNA synthesis, but must also yield data that is compatible with the analysis software of the instrument.

Rapid Synthesis of Fluorophore-Labeled Primers.

Fluorophore-labeled primers for automated DNA sequencing using the ABI 370A must meet several criteria. As with any primer, the sequence must be chosen based on known sequence information to hybridize specifically to a single template site. However, the use of fluorescent dyes as detectable moieties imposes additional requirements (1,3). Due to the presence of the four different fluorophores, each set of labeled fragments produced

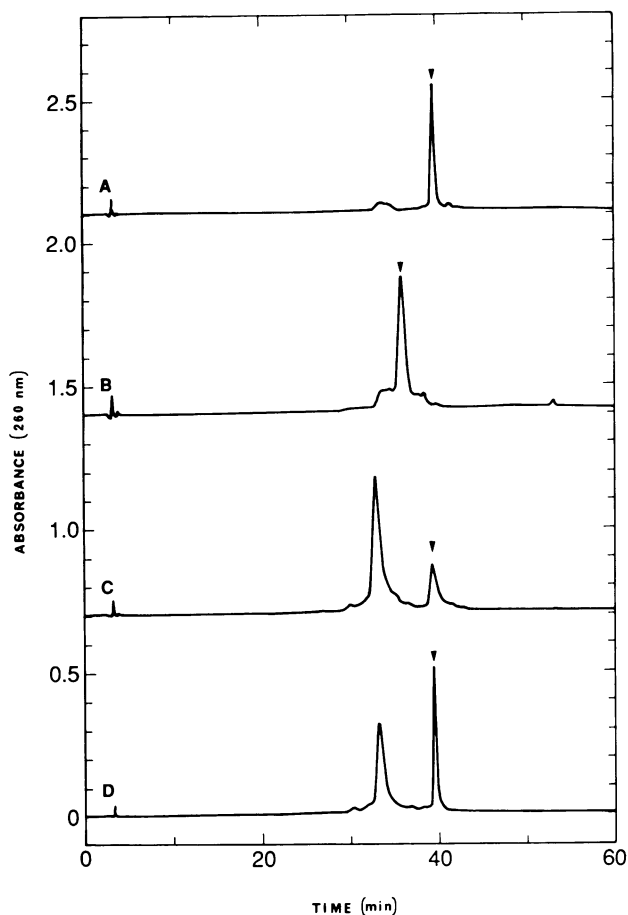


Figure 2. Reverse phase (C18) HPLC traces of a set of crude fluorophore-labeled oligonucleotides corresponding to primer sequence PO1.3 (5' > TGTAAGATCAGCCACCTGCCATAT < 3'). Arrowheads mark the product peaks. Peaks around 34 minutes correspond to unreacted amino-oligonucleotide and failure sequences from the DNA synthesis. (A) Rox-C2-PO1.3; (B) Tamra-C2-PO1.3; (C) Joe-C6-PO1.3; (D) Fam-C6-PO1.3. Chromatographic conditions are given in the text.

in the enzymatic sequencing reactions has a slightly different electrophoretic mobility relative to one another. Thus, the data from the four reactions must be aligned prior to automated sequence analysis. This is accomplished in two ways. A coarse adjustment is made in the labeling chemistry, by using two different straight-chain amino alcohols as the linkers for attaching the fluorophores to the 5'-terminus of the oligonucleotide, as diagrammed in Figure 1. Through the use of a six carbon linker for the two fluorescein derivatives, Fam and Joe, and a two carbon linker for the two rhodamine derivatives, Tamra and Rox, gross differences in molecular weight and charge are ameliorated. This necessitates the use of two different phosphoramidites for introducing the linkers. For the purposes of this study, it was deemed more economical to synthesize the required compounds as needed, since

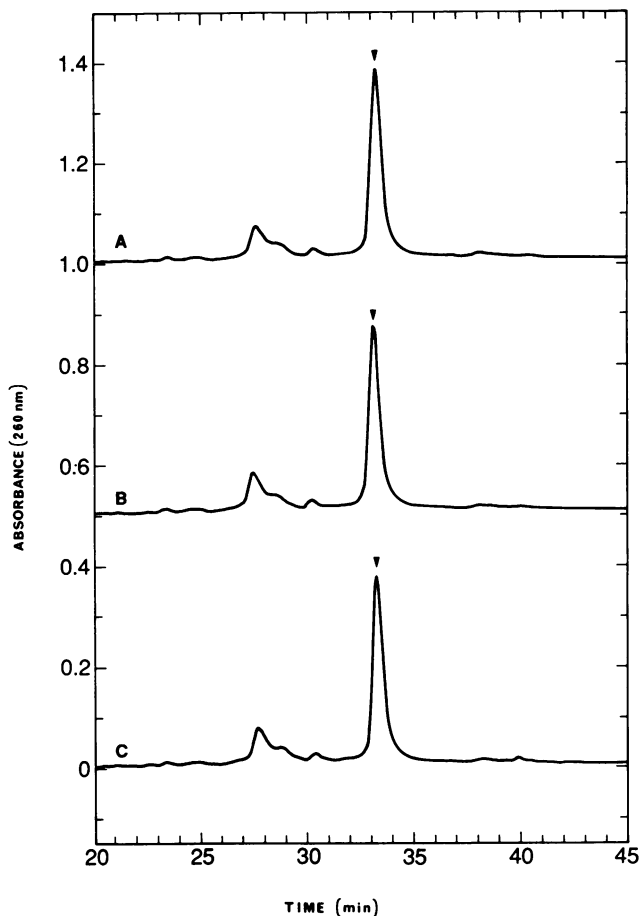


Figure 3. The reaction of Tamra-NHS with $5' > \text{H}_2\text{NCH}_2\text{CH}_2\text{OPO}_3\text{-TGTTAAACGACGGCCAGT} < 3'$ in basic aqueous solution after various times. Arrowheads mark the product peaks. Unreacted amino-oligonucleotide and failure sequences appear around 28 minutes. (A) Overnight (>16 hours); (B) 4 hours; (C) 2 hours. Chromatographic conditions are given in the text.

we could not easily estimate in advance our rate of usage during the development of the methods. However, suitable phosphoramidites are commercially available, and include Aminolink 1 and 2 from Applied Biosystems, and the 5'-amino modifiers C2 and C6 from Glen Research (Herndon, VA) or Clontech Laboratories, Inc. (Palo Alto, CA). A variety of amino-protecting groups have been utilized in such phosphoramidites (3,9,15). We chose to synthesize the appropriate phosphoramidites using the base labile 9-fluorenylmethyloxycarbonyl (Fmoc) group for the following reasons. The Fmoc-amino alcohol precursors are easily obtained in good yield and high purity in an efficient one step reaction and recrystallization, and exhibit excellent long term storage properties. Their phosphoramidite derivatives are also easily prepared by standard reactions and can be readily purified by silica gel chromatography. Alternatively, the crude phosphoramidites may be

Nucleic Acids Research

Table I. Synthesis of Fluorophore-Labeled Primers

Step	Time (hrs)
1. Oligonucleotide synthesis	0.25 per cycle
2. Cleavage/deprotection	8–12 ¹ ; 4 ²
3. Lyophilization	3
4. Fluorophore labeling	2
5. Purification	
a. Gel filtration	1 (4×0.25)
b. Reverse phase HPLC	4 (4×1) ³

¹Using standard base-protecting groups: dA(benzoyl), dG(isobutyryl), and dC(benzoyl).

²Using labile base-protecting groups (see text): dA(phenoxyacetyl), dG(phenoxyacetyl), and dC(isobutyryl).

³May be reduced for shorter sequences that can be purified in pairs.

used within 24 hours of their synthesis without significantly altering their reactivities. They routinely give good yields of both amino-oligonucleotides, and quantitative removal of the Fmoc group is accomplished concomitantly with deprotection of the phosphate groups and the bases. These reagents are thus well suited for the rapid solution synthesis of labeled primers.

Couplings using the Fmoc-amino alcohol amidites were performed manually for convenience, so that a single 1 μ mole automated synthesis might be split in two for derivatization with the amino linkers. Alternatively, it should be possible to carry out two automated syntheses of the same primer sequence at a 0.2 μ mole scale and subsequently perform the linker couplings on the DNA synthesizer.

Labeling reactions were carried out in aqueous/organic solution at pH 9 using crude amino-oligonucleotide. A ten- to twenty-fold excess of activated fluorophore over total DNA was used. This quantity represents a compromise between the yield of labeled oligonucleotide and the cost of the activated dyes. In general, conversion of amino oligonucleotide to fluorophore conjugate were greater than 50% under these conditions, and usually nearly quantitative. The crude reaction mixtures were first purified by gel filtration chromatography on pre-packed NAP 25 columns to remove unreacted dye, and the void volume effluent containing the desired product as well as unreacted DNA species was then further purified by reverse phase HPLC. Representative chromatograms for a typical set of four fluorophore-labeled primers are shown in Figure 2.

Experiments were carried out in order to determine the effect of reaction time on yield of labeled oligonucleotides by performing standard reactions for two hours, four hours, and overnight. Virtually no difference in yield was observed among these three reaction times for any of the four dyes (see Figure 3). Thus, all subsequent labeling reactions were carried out for two hours.

Table I summarizes the steps in the synthesis of fluorophore-labeled primers. The time required for synthesis of the oligonucleotide is essentially set and depends on the length of the desired sequence. The time required for the labeling reactions may possibly be reduced, especially if complete conjugation is not required. The use of pre-packed columns for the gel filtration step dramatically reduces the time and effort required for this step. The time necessary for the HPLC purification of the gel filtered products can occasionally be reduced if sufficiently short primers (18 to 20 nucleotides) are used and two different conjugates can be purified in the same one hour run (data not shown). With longer primers this is generally not possible, and four, one hour runs are necessary. It is doubtful if the

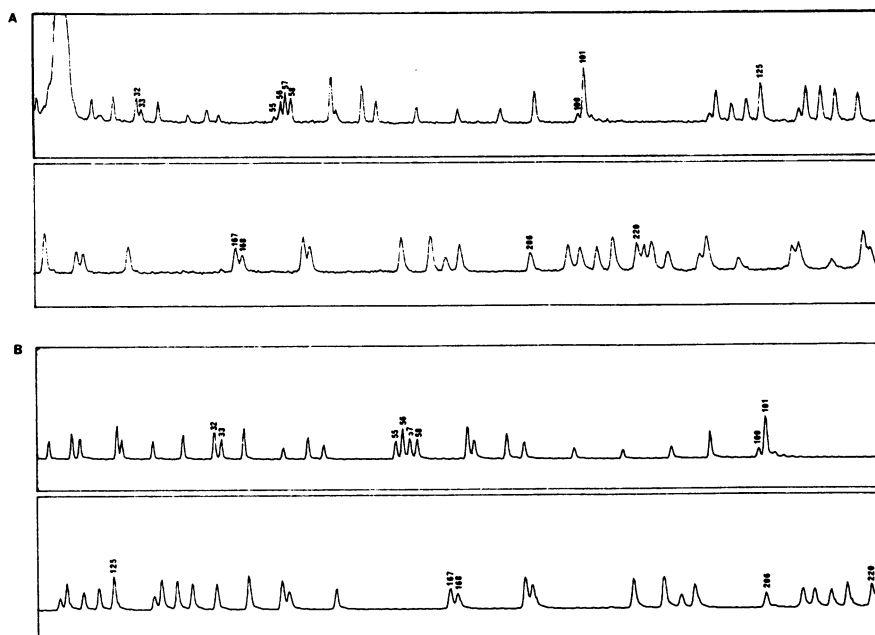


Figure 4. Comparison of automated sequence analysis of M13(mp18) using labeled 'universal' primers synthesized with two types of base-protecting groups. The plots show single-channel raw data for enzymatic dC reactions using 'Sequenase' performed under the conditions given in the text. (A) Tamra-primer synthesized using phosphoramidites with the labile base-protecting groups of Schulhof, et al. (12, 13) [dA(phenoxyacetyl), dG(phenoxyacetyl), dC(iso-butyryl)]; (B) Fam-primer synthesized using phosphoramidites with the standard base-protecting groups [dA(benzoyl), dG(isobutyryl), dC(benzoyl)] (obtained from Applied Biosystems). Numbers above the peaks indicate the relative position in the M13 sequence and are presented only to aid in comparing the two traces. Differences in the rate of electrophoresis between the two traces are due to the use of separate gels run at slightly different powers.

time required for HPLC purification could be much reduced without necessitating the prior purification of the crude amino oligonucleotide, an unsatisfactory alternative.

From Table I, it is clear that a major expenditure of time lies in the usual 8 to 12 hour base deprotection step following DNA synthesis. It would be advantageous to have either a more rapid deprotection reaction for the conventional base protecting groups, or to use alternative protecting groups that may be removed more quickly. We therefore chose to employ the set of base-protecting groups proposed by Schulhof, et. al. (12, 13), which can be removed under mild conditions in four hours. Fluorophore-labeled primers representing the commercially available 'universal' M13 primer sequence (5' > TGT AAA ACG ACG GCC AGT < 3') were synthesized using these phosphoramidites and the procedures described previously. The resulting purified primers were then tested in DNA sequencing to assure that the new chemistry provided oligonucleotides of sufficient quality. As can be seen from Figure 4, little or no difference was observed using these primers as compared to primers synthesized using the conventional protecting groups. The new protecting groups thus represent a viable means for significantly reducing the time required for the synthesis of fluorophore-labeled primers.

Table II. Efficiency of Automated Sequence Assignment Using Specific-Primers

Primer	Sequence	Total Bases ¹	% Assigned ²
PO2.U ³	TGTAAAACGACGGCCAGT	380	95
PO2.1	TGTAATTAATCTACAGTGCATCGCG	360	96
PO2.2	TGTAAAGCAGCCTCCATTGGTGAC	390	94
PO2.3	TGTAACCTGGGATCACGCCACCTGC	400	83
PO1.1a	TGTAACGCGATGCACTGTAGATTAA	—	—
PO1.1b	TGTAATGGCCAATACCGGTGCAC	380	95
PO1.2	TGTAAGTGCACCAATGGAGGCTGCT	400	93
PO1.3	TGTAAGATCACGCCACCTGCCATAT	—	—

¹Total number of bases analyzed by the computer (A,C,G,T, or ? in the output).

²Percent of total bases assigned by the computer (identified as A,C,G, or T in the output).

³U = 'universal' M13 primer (-21), obtained from Applied Biosystems.

The synthesis and purification of fluorescent primers is at present relatively expensive. This is true both because of the cost of the reagents and instrumentation required, and because one generally obtains quantities of labeled oligonucleotides far in excess of those which are needed for use in sequencing. Thus, this approach is generally too costly for routine sequence analysis. However, the use of specific fluorescent primers should prove a valuable option for the occasional filling in of gaps in extended sequence obtained using random cloning and sequencing strategies.

Compatibility of Labeled-Primers with 370A Automated Analysis. As discussed previously, coarse adjustment of the differential mobilities of the fragments formed in the sequencing reactions due to the four fluorophores is accomplished chemically. Fine adjustment is accomplished the data analysis routines (3). A set of mobility shift correction factors must first be empirically determined for a given dye-primer set. These factors describe the time-dependent electrophoretic behavior of three of the sets of dye-labeled fragments relative to the fourth. These correction factors are then applied to the raw data during analysis to align the four data sets. Unfortunately, it appears that each different primer sequence requires a different set of factors. This phenomenon is not yet understood. It is essential, however, to apply the correct set of factors to the raw data in order to obtain maximal accuracy in the subsequent analysis.

The ABI 370A software package includes those sets of correction factors for the 'universal' primers manufactured by ABI. In order to utilize other primer sequences, a strategy employing the extant set of factors was devised and tested. It was suggested (Dr. M. Hunkapiller, Applied Biosystems) to add several bases from the 5'-end of the sequence of the 'universal' primer to the 5'-end of the sequence of a primer for which the factors were unknown. By doing so, fragments generated using the new dye-primer set might exhibit the relative mobility characteristics of fragments generated using the 'universal' dye-primer set, and thus be acceptably analyzed using the existing correction factors. We thus designed and synthesized all labeled specific primers to have an 18–20 nucleotide specific hybridizing sequence and a 5' non-hybridizing tail consisting of the five 5'-most nucleotides of the universal primer sequence (5' > TGTAAC < 3').

The success of this strategy is illustrated by the data in Table II. All of the specific primers synthesized with the five base modification yielded fluorescent sequence data that was, on the average, no less analyzable by the instrument than that obtained using the universal primer with one exception. In the case of PO2.3, the fluorescent signal amplitude

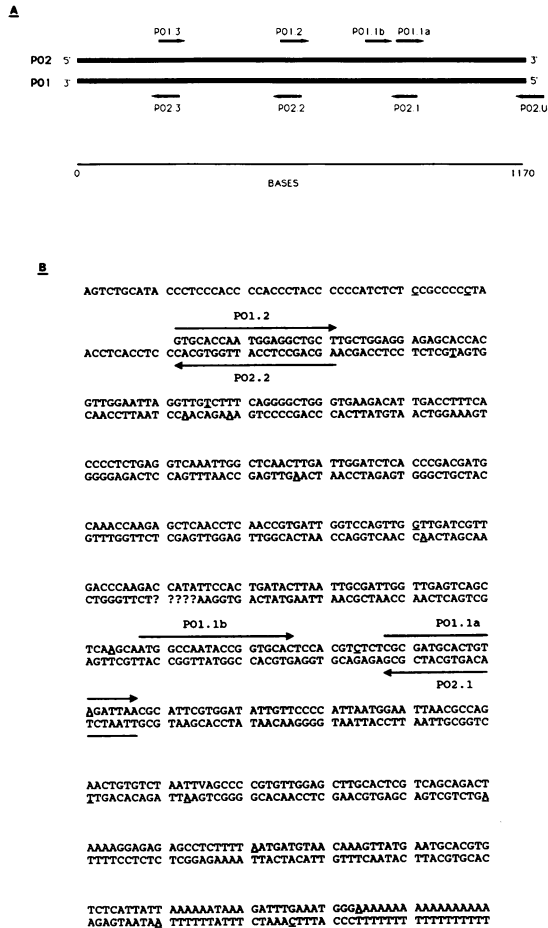


Figure 5. (A) Schematic representation of the location of the primer sequences chosen and the direction of sequencing along the two M13 cDNA clones PO1 and PO2. Sequences are presented in Table II. (B) Region of complementary nucleotide sequence obtained from clones PO1 and PO2 using the automated sequencer. Assignments which are underlined represent ambiguous positions in the automated analysis (represented by '?'s in the output) which were first assigned by manual inspection and then confirmed using sequence information from the opposite strand. The ????? in the figure indicates an aberrant region of the sequence data from one template where analysis could not be made correctly, either automatically or manually; however, the complementary sequence data was unambiguous.

was significantly less than usual, resulting in poorer automated analysis, largely due to interference from excessive baseline noise.

The advantage of the above strategy lies in the simplicity of its implementation, since it requires only an additional five cycles of automated DNA synthesis. Alternatively, an updated version of the analysis software (version 1.3) enables the user to calculate correction factors for primer sequences. However, calculation of a custom set of factors necessitates co-electrophoresing and analyzing the products of a single sequencing reaction using all

four dye-primers to generate the appropriate values. While such a procedure should be useful for the generation of correction factors for use with 'universal' primers for sequencing vectors other than M13 or other commonly used primers, it is too cumbersome for routine use. The strategy used here should be generally preferable.

DNA Sequencing Using Specific Primers. As indicated in Table II and Figure 5, seven specific-primers plus the 'universal' M13 primer were tested in fluorescence-based automated DNA sequencing using the two clones, PO1 and PO2, containing complementary inserts. Of the seven specific-primers, four (PO2.1, PO2.2, PO1.1b, PO1.2) gave excellent results in automated sequencing, one (PO2.3) gave acceptable results, and two (PO1.1a, PO1.3) failed to give sufficient raw fluorescence signal for accurate automated analysis. Representative data for a successful sequence run is shown in Figure 6. In the case of the two primers that failed, steps were taken to ascertain the reason for the failure. Both PO1.1a and PO1.3 were tested in multiple sequencing experiments, each using 'universal' primer as a control in a parallel set of reactions. In each experiment, a normal control result and a specific-primer failure was observed. In order to assure that an error had not been made during the fluorescent primer synthesis in these cases, each primer set was resynthesized at least once. The integrity of the stock solutions of the purified dye-primers was rechecked by HPLC prior to use. We are confident, based on the above considerations, that the observed failures thus represent situations in which the candidate oligonucleotides were unable to act as efficient primers for DNA synthesis by the modified T7 polymerase. Such a result is not uncommon, however, and represents a potential risk of the specific-primer approach at present (17).

Several factors may influence the efficiency with which a homologous DNA sequence can act as a primer for enzymatic sequencing (17). These include poor accessibility of the primer to the template due to template secondary structure, inefficient extension by the polymerase, again probably related to local secondary structure, and multiple site priming. The last possibility can be discounted since it is generally manifested as a multiplicity of overlapping signals (3), and not the near absence of signal, as observed in our experiments. Furthermore, failure was apparently not due to the inability of either oligonucleotide to hybridize to the template, which was tested by allowing equimolar amounts of primer and template to anneal, and then electrophoresing the mixture in a non-denaturing gel using the 370A sequencer. Only a very small amount of fluorescent primer was observed to pass the detector after a time appropriate to its length, while control experiments—either a formamide-denatured equimolar mixture of primer and template that had been previously allowed to anneal, or a non-denatured equimolar mixture that had not been allowed to anneal—gave the expected fluorescent signal at the appropriate time in the same gel (data not shown). It therefore appears that neither oligonucleotide was capable of sufficiently efficient extension by the modified T7 polymerase to provide adequate fluorescent signal for accurate analysis by the automated DNA sequencer.

It is important to note that the signal requirements for automated fluorescence-based sequencing and for conventional radioactive sequencing are somewhat different. Fewer failures due to poor enzymatic extension are expected to be observed using standard radiolabel-based protocols, for two reasons. First, multiple radioactively labeled nucleotides

Figure 6. Representative analyzed sequence data from the Applied Biosystems 370A automated DNA sequencer using a chemically synthesized specific-primer. The data was obtained using primer PO1.2 (5' > TGTAAGTGCACCAATGGAGGCT GCT < 3') under the conditions described in the text. The original output is in four colors: dA, green (Joe); dC, blue (Fam); dG, yellow (Tamra); T, red (Rox).

are generally incorporated in the reaction products during strand extension, thus affording increased detection sensitivity. Second, low signal can be integrated for extended periods by increasing the time of exposure during autoradiography. In contrast, in automated sequencing using fluorescent primers, each extended strand possesses a single label. Furthermore, signal can be accumulated only during the short time in which the electrophoresing fragments are in the read region of the detector. Poorly extended primers may consequently give rise to concentrations of fragments whose fluorescent signal is insufficient for detection and hence accurate automated analysis.

Overall, we were able to obtain about 1100 nucleotides of new sequence from PO2 using four primers, of which about 45% was confirmed by sequence from the complementary clone PO1 using two additional primers. The average amount of reliable sequence obtained per run (based on a conservative estimate of the point at which adequate resolution was lost) was about 380 bases. Of this, an average of 92% of the sequence was assigned by the instrument; a further 98% of the undetermined bases could be assigned by manual inspection of the computer output. In the region where complementary sequence was obtained, the instrument made 28 miscalls out of 980 total assignments (490 base pairs); 20 of these were insertions, most of which arose in a single run in which an abnormally noisy ROX-T signal was observed; 2 were deletions; and 6 were incorrectly identified (eg., G instead of T), 5 of which occurred at an aberrant region of the fluorescence data in one run. The necessity of obtaining complementary sequence information was clearly emphasized in this study. Without exception, areas of doubt observed on one strand could be unambiguously resolved using the information from the opposite strand (see Figure 5).

CONCLUSION

The utility of specific primer-directed automated sequencing using the fluorescent primer approach relies heavily on the ability to produce the necessary sets of four fluorophore-labeled primers sets in a reliable, rapid manner. As described above, a single set of dye-primers can be synthesized and purified in a 24 hour period or less, depending upon the DNA synthesis chemistry used to produce the desired sequence. The majority of the time involved is in the synthesis of the oligonucleotide itself, and is thus a necessary expense for any primer-directed sequencing approach. Once the crude oligomer is obtained, the fluorophore conjugation reactions are rapid. Purification of the conjugates, whether by HPLC or gel electrophoresis, can be a rate-limiting step, especially when multiple primer sets are required.

The use of specific dye-primers in automated DNA sequencing is constrained both by the efficiency with which the candidate sequence can function as a primer in the enzymatic sequencing process and by the need to correct for fluorophore-related mobility differences among the four sets of labeled fragments in the data analysis prior to automated sequence determination. The strategy of using a single set of mobility correction factors by incorporating a short sequence from the 5'-end of a primer whose factors are known at the 5'-end of all other primer sequences appears to be generally successful.

An important result of this study is the observation that not all of the specific-primer sequences tested gave rise to sequencing data that could be analyzed by the 370A DNA sequencer. As discussed previously, this appears to be caused by inefficient extension of the primer by the modified T7 DNA polymerase used in these sequencing experiments due to secondary structure in the primer-template hybrid. It is possible that the use of an alternative enzyme such as *Taq* polymerase from the thermophilic bacterium *Thermus*

aquaticus, which is active at temperatures which generally reduce the amount of template secondary structure, may provide different results. Alternatively, a greater understanding of the relation of primary DNA sequence to secondary structure and its effect on primer extension by DNA polymerases would reduce the incidences of failure through a more informed choice of candidate primer sequences. We are continuing to pursue both of these avenues of investigation.

Our conservative method of choosing candidate primer sequences limited the amount of new sequence information obtained per primer to about 300 bases, out of an average of 380 total bases obtained per run. Future improvements in the automated sequencing technology may be expected to allow for the production of reliable sequence information in the region >380 bases from the primer, thus increasing this number. We also chose a significant overlap (80 bases on the average) between successive rounds of sequencing in order to have an internal check on the accuracy of the sequence data produced from the specific primers. Such a large overlap is generally unnecessary; a small overlap is desirable, however, as the 5–25 bases immediately adjacent to the primer are often poorly resolved.

Based on the information presented herein, we can estimate that a single round of primer-directed automated DNA sequencing can currently be completed in about three days for a single primer. This includes selection of the candidate primer sequence, synthesis and purification of the set of dye-primers, manual performance of the enzymatic sequencing reactions, and electrophoresis and data analysis using the 370A DNA sequencer. From this procedure, one might expect about 400 bases of sequence information, less overlap from the preceding round. However, it would take some 4–5 times longer to produce enough primer sets and sequencing reactions to use the entire 16 lane capability of the instrument. Furthermore, the chemical synthesis and purification of fluorophore-labeled primers is currently quite costly, both in terms of materials and time. Hence, specific-primer-directed sequencing using the ABI 370A is at present better suited to filling in small gaps in extended sequence generated by 'shotgun' sequencing techniques than for producing extensive sequence by 'walking' along large inserts.

ACKNOWLEDGMENTS

We are indebted to Dr. Michael Hunkapiller for his valuable suggestions concerning the automated sequencer during the initial phases of this work. We are also grateful to Dr. J.C. Schulhof for his advice concerning the synthesis and purification of the phosphoramidites with the labile protecting groups. We would like to thank the members of the CIT Biopolymer Synthesis Facility, in particular Dr. Suzanna Horvath and Peter Gaines, for the synthesis of oligonucleotides; Barb Otto for technical help with the DNA sequencer; and Dr. Ulf Landegren for critical reading of this manuscript. This work was supported by National Science Foundation grant DIR8809710. R.A.S. was a postdoctoral fellow of the National Multiple Sclerosis Society.

REFERENCES

1. 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.
2. Ansorge, W., Sproat, B., Stegemann, J., Schwager, C., and Zenke, M. (1987), *Nucleic Acids Res.* **15**, 4593–4602.
3. Connell, C., Fung, S., Heiner, C., Bridgham, J., Chakerian, V., Heron, E., Jones, B., Menchen, S., Mordan, W., Raff, M., Recknor, M. Smith, L., Springer, J., Woo, S., and Hunkapiller, M. (1987), *BioTechniques*

- 5, 342–348.
4. 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.
5. Kambara, H., Nishikawa, T., Katayama, Y., and Yamaguchi, T. (1988), *Biotechnology* **6**, 816–821.
6. Brumbaugh, J.A., Middendorf, L.R., Grone, D.L., and Ruth, J.L. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5610–5614.
7. Sanger, F., Nicklen, S., and Coulson, A.R. (1977), *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
8. Strauss, E.C., Kobori, J.A., Siu, G., and Hood, L.E. (1986) *Anal. Biochem.* **154**, 353–360.
9. Agrawal, S., Christodolou, C., and Gait, M.J. (1986), *Nucleic Acids Res.* **14**, 6227–6245.
10. Sinha, N.D., Biernat, J., McManus, J., and Koster, H. (1984), *Nucleic Acids Res.* **12**, 4539–4557.
11. Still, W.C., Khan, M., and Mitra, A. (1978) *J. Org. Chem.* **43**, 2923–2925.
12. Schulhof, J.C., Molko, D., and Teoule, R. (1987), *Nucleic Acids Res.* **15**, 397–416.
13. Schulhof, J.C., Molko, D., and Teoule, R. (1987), *Tetrahedron Lett.*, 51–54.
14. Saavedra, R.A., Fors, L., Aebersold, R.H., Arden, B., Horvath, S., Sanders, J., and Hood, L. (1989), *J. Mol. Evol.*, in press.
15. Connolly, B.A. (1987), *Nucleic Acids Res.* **15**, 3131–3139.
16. Information contained in the *Biosystems Reporter*, issue 1 (1988) and *ABI 370A Users Bulletin*, issue 8 (1988), available from Applied Biosystems.
17. Barnes, W.M. (1987) In Berger, S.L. and Kimmel, A.R. (eds.), *Methods in Enzymology*, Academic Press, Inc., San Diego, CA, Vol. 152, pp. 538–556.

**This article, submitted on disc, has been automatically
converted into this typeset format by the publisher.**