# A Strategy for Examining Complex Mixtures of Deoxyoligonucleotides Using Ion-Pair–Reverse-Phase High-Performance Liquid Chromatography, Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry, and Informatics

Peter M. Vallone,*,[1] Joseph M. Devaney,† Michael A. Marino,† and John M. Butler*
*Biotechnology Division, National Institute of Standards and Technology, 100 Bureau Drive, Mail Stop 8311, Gaithersburg, Maryland 20899; and †Transgenomic Inc., 11 Firstfield Road, Suite E, Gaithersburg, Maryland 20878

**Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) and ion-pair–reverse-phase high-performance liquid chromatography (IP-RP HPLC) techniques were combined to determine the sequence identity of short single-stranded deoxyoligonucleotides. This methodology is demonstrated using a commercially available multiplex set of eight primer pairs. The primer pairs were separated and collected by IP-RP HPLC. Partial sequence information for IP-RP HPLC fractions was obtained from analyzing exonuclease digestion products by MALDI-TOF MS. IP-RP HPLC, MALDI-TOF MS, exonuclease digests, and a simple computational algorithm provide an integrated strategy for determining the sequence of short nucleic acid oligomers.** © 2002 Elsevier Science (USA)

Nucleic-acid-based assays are prevalent in current high-throughput diagnostic technologies. Applications such as multiplex polymerase chain reaction (PCR)[2] (1), nucleic acid screening through chip and array technologies (2), and aptamer selection (3), along with nucleic acid based hybridization assays rely on short nucleic acid oligomers (4). Development of these assays requires the synthesis of short nucleic acid molecules. Proper quality control of nucleic acid oligomers is essential for success in the development and mass production of commercial assays and kits. The methodology described in this report provides a straightforward means to monitor the quality of a mixture of nucleic acid oligomers by sampling and analyzing a relatively small fraction of the initial batch.

Quality control monitoring is accomplished by applying the established analytical techniques of HPLC and mass spectrometry. Ion-pair–reverse-phase high performance liquid chromatography (IP-RP HPLC) provides a means to separate and concentrate DNA oligomers (varying lengths and sequence contents) from a relatively complex mixture (>20 oligomer components) (5, 6). The emergence of matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) applied to the analysis of DNA in diagnostic assays and minisequencing has been described previously (7–10). MALDI-TOF MS experiments provide the mass of an intact deoxyoligonucleotide sequence while mass spectral analysis of postexonuclease digest products can help elucidate all or portions of an oligomer sequence (7, 11). Characterization by MALDI-TOF MS confirms purity, mass, and sequence identity of the components. Advantages of this approach are the relatively low amount of sample needed to perform this characterization (~10 pmol of each deoxyoligonucleotide) and the ability to directly measure the mass of an oligomer.

[2] Abbreviations used: PCR, polymerase chain reaction; IP-RP HPLC, ion-pair–reverse-phase high-performance liquid chromatography; MALDI-TOF MS, matrix-assisted laser desorption ionization time-of-flight mass spectrometry; STR, short tandem repeat; TMR, carboxytetramethylrhodamine; TEAA, triethylammonium acetate; SVP, snake venom phosphodiesterase; CSP, calf spleen phosphodiesterase.

In demonstrating this DNA oligomer sequencing methodology a set of 8 PCR primer pairs (16 individual primers) is used to illustrate how HPLC, MALDI-TOF MS, and informatics can be used to quality control a set of nucleic acid sequences. In previous work IP-RP HPLC and MALDI-TOF MS techniques were used to generate profiles to provide unique "fingerprint" profiles of various multiplex PCR primer kits. This work takes that method of quality control a step further by probing and confirming the sequence content and context of each nucleic acid component.

## MATERIALS AND METHODS[3]

*Oligonucleotides.* A commercially available multiplex primer kit, PowerPlex 1.1 System (PPlex1.1) was purchased from Promega Corporation (Madison, WI), Lot No. 107616. The PPlex1.1 multiplex kit is routinely used for the PCR amplification of short tandem repeat (STR)-containing regions of human genomic DNA for human identification purposes. The eight STR loci contained in PPlex1.1 with their corresponding fluorescent dye label identity and GenBank accession numbers are as follows: CSF1PO carboxytetramethylrhodamine (TMR) (X14720), TPOX TMR (M68651), TH01 TMR (D00269), vWA TMR (M25858), D16S539 fluorescein (AC024591.3), D7S820 fluorescein (AC004848), D13S317 fluorescein (AL353628.2), and D5S818 fluorescein (AC008512.4).

*IP-RP HPLC.* The HPLC used for separation and collection of the DNA oligomers was a Transgenomic Wave DNA fragment analysis system (San Jose, CA) with a DNASep column (Transgenomic, Inc.) that has a stationary phase consisting of 2-$\mu$m nonporous alkylated poly (styrene-divinylbenzene) particles. UV detection was performed at 260 nm. A fluorescence detector was used to identify primers labeled with a fluorescent dye chromaphore. The IP-RP HPLC separations were run at 70°C with a mobile phase consisting of 0.1 M triethylammonium acetate (TEAA; solvent A) and 0.1 M TEAA–25% acetonitrile (solvent B). The separation conditions were a flow rate of 0.9 mL/min and a gradient of 18 to 50% solvent B over 40 min. A 50-$\mu$L aliquot of the PCR multiplex kit primer solution was injected onto the column for fraction collection purposes. Fractions were collected manually and evaporated to dryness in a Savant SPD111V SpeedVac (Holbrook, NY). Each fraction was reconstituted in 20

$\mu$L of deionized water for further characterization by mass spectrometry.

*Mass spectrometry.* Samples were spotted on a stainless steel MALDI plate using 1 $\mu$L sample and 1 $\mu$L matrix material. The stock matrix solution consisted of 0.7 M 3-hydroxypicolinic acid (Aldrich Chemical, Milwaukee, WI) and 0.07 M ammonium citrate (Sigma Chemical, St. Louis, MO) in 1:1 water and acetonitrile (12). Stock matrix solution was diluted 2.5-fold with water prior to usage. Mass spectra were collected on a Bruker BIFLEX III time-of-flight mass spectrometer (Bruker Daltonics, Bremen, Germany) equipped with a pulsed nitrogen laser (337.1 nm). The following operating voltages were employed for signal collection, IS1 = 19.0 kV, IS2 = 15.0 kV, reflector = 20.0 kV, detector 1.7 kV. All spectra were collected in positive ion mode with between 100 and 300 laser shots collected on each sample for signal averaging purposes. Laser power attenuation was adjusted with each sample to obtain optimal sensitivity and resolution. Prior to data collection the mass spectrometer was calibrated with 15- and 36-base single-strand DNA oligomers, with masses of 4577.1 and 11,091.3 Da respectively, where 1 Da is equivalent to 1 g/mol. The peak masses described throughout this paper are for the singly charged ions and are in daltons.

*Data analysis.* IP-RP HPLC chromatogram baselines were fitted utilizing a loess algorithm (a locally weighted regression smoothing algorithm that performed a full least-squares fit for each data point) and then subtracted from the raw data. Mass spectra were smoothed and enhanced by the application of matrix convolution filters contained in the XMASS 5.0 analysis software package (Bruker Daltonics, Bremen, Germany). Both IP-RP HPLC and mass spectrometry data were normalized to the largest peak in the spectrum as a reference, resulting in relative intensities between 0 and 100%.

*Exonuclease digestion assays.* 5' and 3' acting exonuclease digestion assays were performed as previously described (11). Snake venom phosphodiesterase (SVP) and calf spleen phosphodiesterase (CSP) were purchased from Boehringer Mannheim (Mannheim, Germany), 1 mg SVP in 0.5 mL and 2 mg CSP in 1 mL. The 10× concentrated SVP buffer consisted of 2.5 g of ammonium citrate in 50 mL of water pH 9.4 (~22 mM pH 9.4).

*3' → 5' exonuclease sequencing.* One microliter of stock SVP solution (1 mg in 0.5 mL) was diluted 20-fold in 1× SVP buffer (diluted from the 10× SVP buffer). One microliter of the diluted SVP was added to approximately 6 pmol of oligonucleotide primer in 1× SVP buffer (total volume of 5 $\mu$L). The digestion was incubated at room temperature and 1-$\mu$L aliquots were removed 2, 5, and 10 m after the addition of the SVP.

---

[3] Certain commercial equipment, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

Aliquots were spotted directly onto a 384 position MALDI plate prespotted with $1\mu L$ of the matrix solution and allowed to air dry.

*5′ → 3′ exonuclease sequencing.*   One microliter of the stock CSP solution (2 mg in 1 mL) was added directly to approximately 6 picomoles of DNA in a total volume of 5 $\mu L$. No buffer adjustment was needed for the CSP assay. The digestion was incubated at 37°C and 1 $\mu L$ aliquots were removed 5, 10, and 15 m after the addition of the CSP. Aliquots were spotted directly onto a 384 position MALDI plate prespotted with $1\mu L$ of the matrix solution and allowed to air dry.

## RESULTS

### Strategy for Short Oligonucleotide Sequence Determination

In the simplest case, a single DNA oligonucleotide of known sequence is digested and the resulting partial sequence information is compared to that of the known sequence for confirmation. For practical purposes, if the experimental and theoretical masses are in agreement (to within ~2 Da) then only a few (~5) bases of sequence information are usually needed for an overall sequence confirmation for short nucleic acid oligomers. The strategy described here is related to a more challenging scenario, in which a mixture of nucleic acid oligomers with unknown or limited sequence information is being fully sequenced. The strategy described here is sufficient to be applied to the simplest (single primer with known sequence) and a more complex case (a mixture with unknown sequences).

The flow chart in Fig. 1 depicts the general strategy for determining oligonucleotide sequence(s) from a moderately complex mixture (16 to 32 unique oligomers). When applying this approach to identifying oligomers of unknown sequence content a primary assumption is that a reference sequence is available. The oligomer sequences of interest will be a subset of this reference sequence. In the example illustrated in this report, the reference sequences for the corresponding eight STR loci contained in the PPlex1.1 were obtained from the GenBank public database resource (see Materials and Methods).

Commercially available PCR primer sets are typically available either separately (as a single pair—singleplex) or as a combined set (multiple pairs—multiplex). In the case of a combined set of primers, IP-RP HPLC is essential for the separation and collection of each unique nucleic acid component in sufficient quantity (1 to 10 pmol) for initial characterization and exonuclease digestion reactions with subsequent analysis by MALDI-TOF MS.

Analysis of each separated component (IP-RP HPLC fraction) performed by MALDI-TOF MS provides the mass of each separated deoxyoligonucleotide. The MS
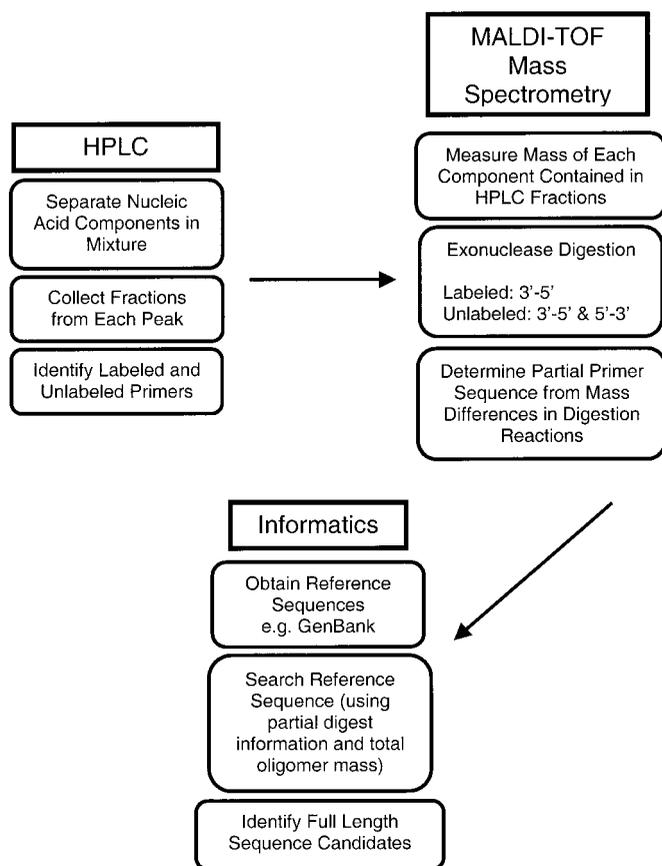


**FIG. 1.**   Flow chart depicting an integrated methodology and stepwise elucidation of unknown oligomer sequences contained within complex DNA oligomer mixtures.

contains the necessary information to determine if any of the fractions contain more than one nucleic acid component, if any failure products ($n - 1$ mers) are present (assuming that the parent mass is known) and if any known modifications (fluorescent dyes, biotin, other labeling) have degraded.

HPLC fractions containing the separated components of the mixture are subjected to digestion by 5′ and 3′ acting exonuclease reactions to determine partial sequence information by MS analysis of the digestion reactions. The resulting partial sequence information and the full-length oligomer mass are combined with a reference DNA sequence to allow for the selection of potential matching oligomer sequences.

The source of the nucleic acid mixture examined in this work was a set of 16 individual DNA primers (8 pairs) in a multiplex PCR kit. These 16 primers were subjected to the described methodology for sequence determination. The 16 oligonucleotides are used specifically for the amplification of STR-containing regions in human DNA for forensic identification purposes. The experimental data used to determine the sequence
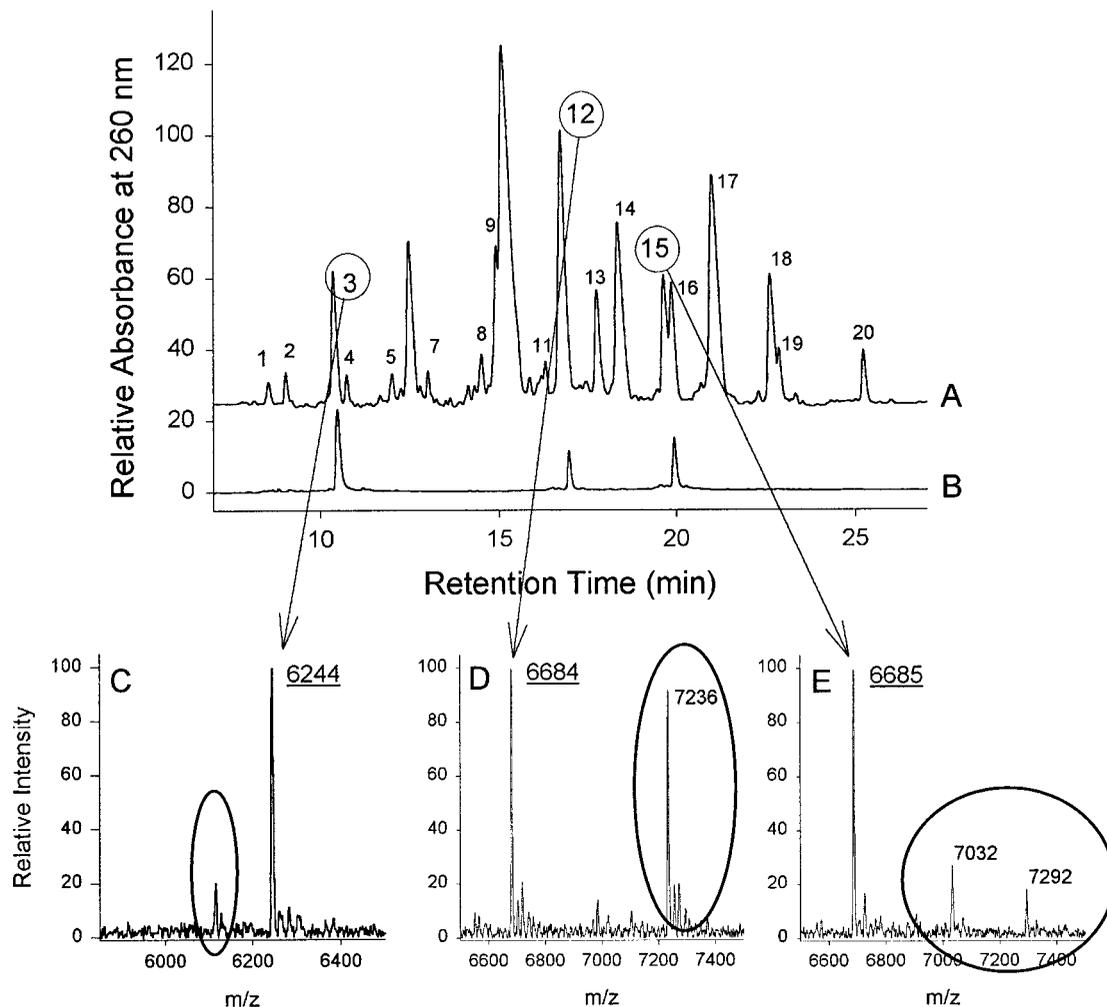
**FIG. 2.** HPLC chromatogram of full primer mixture (A) from a commercial 8-plex PCR amplification kit compared with results from a single primer pair (B). Mass spectra of three of the HPLC fractions are also shown (C–E). Circled peaks are due to carryover in the HPLC fraction collection process. HPLC and MALDI-TOF MS conditions are described under Materials and Methods section (see also Ref. 6).

for 2 of the 16 primers (forward and reverse for a single locus) are shown. The sequences of the primers were identified in a parallel fashion. A parallel approach simplifies the complexity of solving a mixture of oligomers of unknown sequence by reducing the number of possible loci that could be linked to a primer. That is, as a sequence was confirmed to be correct for a specific locus, this information excluded other primers from being connected to the identified locus.

### IP-RP HPLC Chromatogram of a 8plex PCR Multiplex Kit

Figure 2 presents an IP-RP HPLC chromatogram (Fig. 2A) of the 8-plex multiplex PCR kit. There are a total of 16 unique deoxyoligonucleotides in this kit; *at least* 20 significant peaks are present in the chromato-

gram. The literature included with the PPlex1.1 kit indicates that four of the 16 primers are labeled on the 5′ end with the fluorescent dye fluorescein. The four "extra" peaks observed in the HPLC chromatogram are due to oligonucleotides labeled with fluorescein isomers. Evidently the two isomers of fluorescein exhibit different affinities to the IP-RP column, which allows their separation under the described HPLC separation conditions. IP-RP HPLC conditions separated the single-stranded deoxyoligonucleotides based on their sequence content in addition to hydrophobic properties of various fluorescent dyes attached to 8 of the oligomers. The apparent disparity in HPLC peak areas/heights may be due to differing primer concentrations in the kit used to balance amplification yields between the various PCR products.
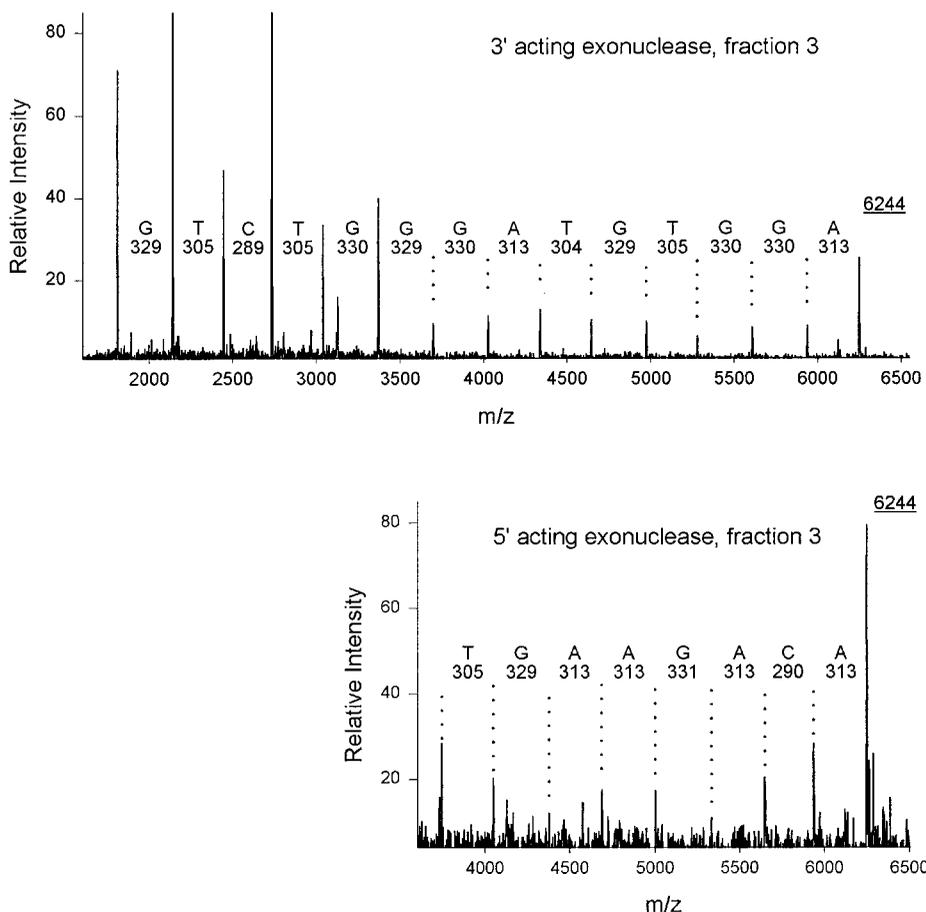
**FIG. 3.** Mass spectra of enzymatic 3′-to-5′ (top) and 5′-to-3′ (bottom) digestions products of HPLC fraction 3. The 3′-to-5′ digestion was performed with SVP and the 5′-to-3′ was performed with CSP.

Figure 2 also presents the IP-RP HPLC chromatogram of the commercially available *singleplex* D13S317 primer set (Fig. 2B). This singleplex chromatogram is depicted solely to illustrate the retention time of the D13S317 primers relative to the multiplex kit. The presence of three distinct peaks in the singleplex primer kit confirms that the isomeric form of fluorescein is responsible for an additional resolved peak. Digest experiments were not performed on the D13S317 singleplex oligomers. Although knowing the exact retention time for the singleplex D13S317 primer set is useful, it is not required to solve the D13S317 sequences contained in a complex mixture.

The mass spectrum of collected IP-RP HPLC fractions 3, 12, and 15 with experimentally measured masses of 6244, 6684, and 6685 Da, respectively, are plotted in the lower inset of Figs. 2C–2E). The extra peak (circled) observed in Fig. 2C is due to depurination of the primer during matrix-assisted ionization/desorption, while the extra peaks (both circled) observed in Figs. 2D and 2E are due to carry over of

primers that were not successfully resolved during IP-RP HPLC fraction collecting.

*Exonuclease digest reactions analyzed by MALDI-TOF MS*

3′ → 5′ exonuclease SVP digests can be performed on all IP-RP HPLC fractions while 5′ → 3′ CSP digests are usually performed only for non-dye-labeled primers. The presence of a dye label on the 5′ end of a DNA oligonucleotide limits the activity of a 5′ → 3′ exonuclease. The use of a fluorescent HPLC detector allowed for the assignment of the dye-labeled oligomers (data not shown). Typically, non-dye-labeled primers exhibited a shorter retention time (less than 15 min with the described HPLC elution conditions). The HPLC retention time along with fluorescent detection information for fraction 3 (Fig. 2C) suggests that it is not dye labeled. Since the oligomer in fraction 3 was not dye labeled it was subjected to both 5′ and 3′ exonuclease digests. The results of exonuclease digestions performed upon fraction 3 are plotted as relative intensity
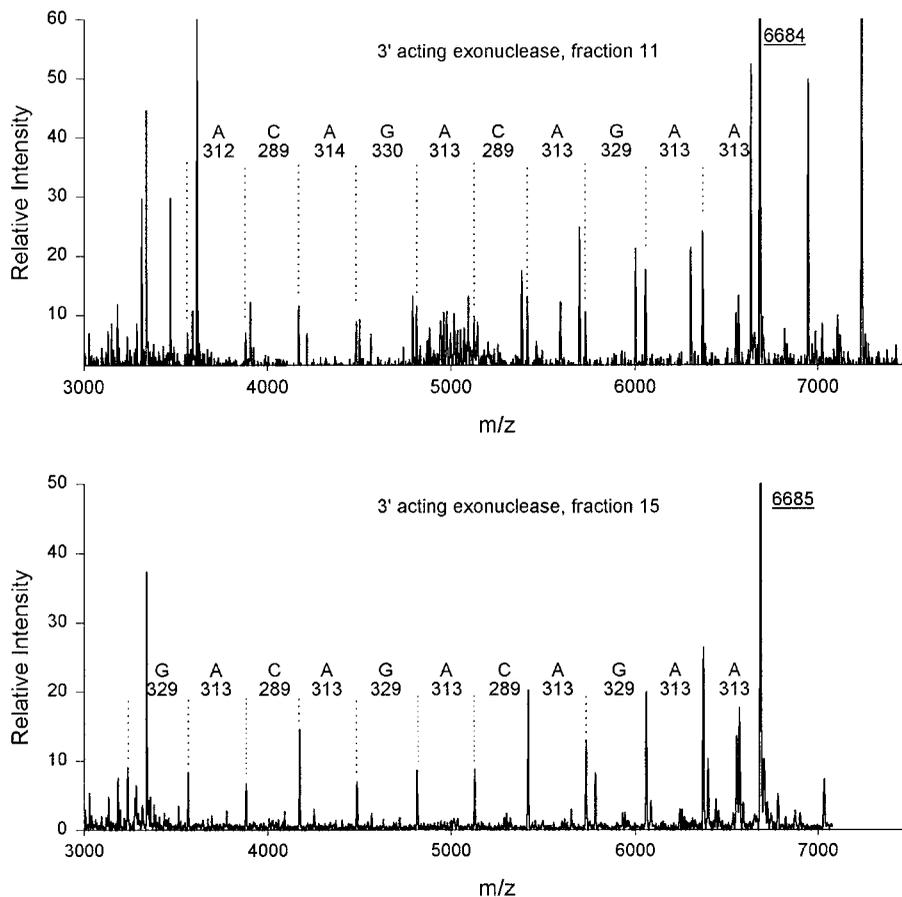
**FIG. 4.** Mass spectra of 3′-to-5′ digests of HPLC fraction 12 (top) and fraction 15 (top). The results from fraction 12 include a "contaminant" (see Fig. 2D) that makes the spectra more challenging to interpret. The "contaminant" is the result of exonuclease digest of the peak at 7236 Da revealing a second partial sequence 5′ . . . ATCCTCATTGAC 3′.

versus $m/z$ in Fig. 3. The mass spectrum of the 3′ → 5′ digestion reaction is shown in the top panel. From mass differences indicated in Fig. 3 measured by MALDI-TOF MS, the sequence of the last 14 bases of the oligomer, 5′. . . GTCTGGGATGTGGA 3′ can be confidently assigned. The bottom panel of Fig. 3 is the result of a 5′ → 3′ digest that allowed the assignment of the first 8 bases, 5′ ACAGAAGT . . . 3′. The intact mass of fraction 3 is 6244 Da corresponding to a DNA oligomer in length ranging from 19 (if sequence content were all Gs; 6244 Da/329 Da = 18.9) to 21 (if sequence content were all Cs; 6244 Da/289 Da = 21.59) bases. Digest experiments performed on fraction 3 resulted in 14 + 8 = 22 bases of sequence information indicating overlapping sequence information (by at least one base). Matching the first 8 and last 14 bases of sequence information into the compiled reference sequence for the 8 STR loci resulted in the 20-base single-strand sequence 5′ ACAGAAGTCTGGGATGTGGA 3′. The matching sequence was found in the forward (relative to the reference sequence) region of the D13S317 locus. The calculated mass of this DNA oligomer is

6246 Da in agreement with the experimentally determined mass of 6244 Da.

Results of 3′ → 5′exonuclease digest experiments on HPLC fractions 12 (Fig. 4, top) and 15 (Fig. 4, bottom) are plotted in Fig. 4. As noted above, the sequence of the DNA oligomer present in fractions 12 and 15 should be the same since the differences in HPLC retention times arises only from the isomeric forms. The mass spectrum of the fraction 12 digestion resulted in the sequence elucidation of the last 10 bases 5′. . . ACAGACAGAA 3′. Interference peaks are present in this spectrum due to the significant carry over of another unresolved oligonucleotide of mass 7234 Da (see Fig. 2D); however, this less intense $m/z$ pattern does not interfere with the sequence assignment for the major component at 6684 Da. The bottom panel of Fig. 4 is the result of a SVP digest on fraction 15 resulting in eleven bases of sequence information 5′. . . GACAGACAGAA 3′. Again, there is a weak signal from the carry over peaks in mass spectrum of fraction 15, which does not interfere with the primary sequence assignment. As expected, the overlapping se-

quence information obtained from fractions 12 and 15 are identical.

## Use of a Sequence Matching Algorithm

Additional $3' \rightarrow 5'$ SVP exonuclease digest experiments could have been performed to further elucidate the sequence of the oligomer contained in fractions 12 and 15, but the use of a computer algorithm described below can assist in solving a complete oligomer sequence without any further digest experiments.

When a nucleic acid oligomer of interest is being used for an application such as PCR (or other diagnostic applications) a reference sequence can be useful for assisting with sequence determination. A reference sequence can be defined as a sequence that contains the sequences of the oligomers of interest as a subset. For example, in the case of eight pairs of multiplex PCR primers the 16 DNA oligomer sequences are merely a subset of this reference sequence.

The reference sequences for each of the eight STR loci that are targets for the PCR amplification were obtained from GenBank for sequence comparisons (as described under Materials and Methods). The expected PCR product size can be estimated from information provided in the multiplex kit. A reduced working size for each GenBank sequence was obtained by including the STR region of interest plus a few hundred bases up and downstream from the STR region. All eight working reference sequences were combined and loaded into Generunner (Hastings Software) sequence analysis software for quick manipulation and matching of partial sequence information. The length of the file containing all eight STR loci was approximately 2575 bp.

Sequence comparisons versus the reference sequences of the 8 loci indicated that the last 10 bases of fraction 12 and the last 11 bases of fraction 15 matched a region of the reference sequence containing the STR site D13S317 in the reverse direction. The reverse orientation of the matched sequence is expected since fraction 3 was matched in the forward orientation in the D13S317 locus.

The oligomer contained in fraction 11 and 15 is dye labeled on the $5'$ end. $5'$ acting CSP exonuclease digest experiments will not assist in elucidating further sequence information. Instead an algorithm described below was employed to determine the complete sequence of the oligomer contained in fractions 11 and 15.

An algorithm for primer searching primarily relying on mass was developed and programmed using Visual Basic 6.0 (Microsoft Corporation, Redmond, WA). The program utilizes the mass of the full length oligomer (as determined by mass spectrometry) with an assumed error range usually 5 to 10 Da, partial sequence information (typically 1 to 10 bases), and the mass of a

specific fluorescent dye to search/"walk" along a reference sequence. The program was employed to help identify the sequence of fractions 12 and 15. In this case the reference sequence consisted of the 8 reference sequences obtained from GenBank. These reference sequences are stored in an Excel (Microsoft Corporation, Redmond, WA) spreadsheet for input purposes. The program output consists of possible oligomer sequences within a designated target molecular weight range.

The program uses the target molecular weight and the assumed error value to define the possible primer size range. The potential oligomer size range is determined by assuming a primer sequence containing exclusively cytosine and then guanosine that provides the upper and lower mass boundaries. Example if no dye label is assumed,

Parent ion molecular weight

$$= 6684 \text{ Da} \pm 5 \text{ Da} \ (6684 \text{ Da} + 5 \text{ Da})/289.19 \text{ Da}$$

$$= 23.1 \text{ bases} \ (6684 \text{ Da} - 5 \text{ Da})/329.21 \text{ Da}$$

$$= 20.3 \text{ bases}.$$

Example if dye label fluorescein $= \sim 536$ Da is assumed,

$$(6684 \text{ Da} - 536 \text{ Da} + 5 \text{ Da})/289.19 \text{ Da}$$

$$= 21.3 \text{ bases} \ (6684 \text{ Da} - 536 \text{ Da} - 5 \text{ Da})/329.21 \text{ Da}$$

$$= 18.6 \text{ bases}.$$

In the case of a dye-labeled primer the program calculates the masses of all 19-, 20-, and 21-base sequences in the template file (and the corresponding complementary reverse sequences) and stores them in an array. Masses of the stored sequences are compared to the input target mass $\pm$ assumed standard deviation and any partial sequence information to narrow down the sequence possibilities.

Masses of commonly used fluorescent dyes were experimentally determined by measuring the mass of a dye-labeled oligonucleotide with known sequence content (data not shown). The experimentally determined mass minus the calculated mass of the sequence provides the mass of the dye/linker attachment. These experimentally evaluated dye-linker masses are listed in Table 1. After calibration of the mass spectrometer the dye mass should be accurate to within *at least* 5 Da.

The following parameters were input into the program for fractions 12/15, a target molecular weight of 6684 Da $\pm$ 5 Da, the sequence of the last 10 bases $5' \ldots$ ACAGACAGAA $3'$, and the mass of the fluorescein dye (536 Da). After completion the algorithm output re-

TABLE 1

Fluorescent Dye Masses Including Oligonucleotide
Attachment Linkers

| Fluorescent dye | Mass (dye + linker) (Da)[a] |
|---|---|
| Fluorescein | 536 |
| 5-FAM or 6-FAM[b] | 540 |
| TMR[c] | 590 |
| JOE[d] | 668 |
| NED[e] | 711 |
| VIC[e] | 719 |
| PET[e] | 902 |

[a] Mass values are estimated to be accurate within *at least* 5 Da.
[b] 5-FAM, 5-carboxy-fluorescein; and 6-FAM, 6-carboxy-fluorescein.
[c] TMR (TAMRA), $N,N,N',N'$-tetramethyl-6-carboxyrhodamine.
[d] JOE, 6-carboxy-2′,7′-dimethoxy-4′,5′-dichlorofluorescein.
[e] NED, VIC, and PET are proprietary to Applied Biosystems (Foster City, CA).

sulted in one sequence that met these criteria. The program returned the sequence 5′ GCCCAAAAAGA-CAGACAGAA 3′. The calculated mass of the selected sequence was 6146 Da. After adding 536 Da for the fluorescein dye label the mass equals 6682 Da. This mass is in agreement with the experimentally determined value of 6684 Da.

The data in Table 2 summarize the results of running the algorithm with the user input of 6684 ± 5 Da and the presence of a fluorescein dye label (536 Da). In Table 2 one can see that as the amount of partial sequence information is decreased the number of potential matches increases. In the example shown when 7–10 bases of sequence information are used only one unique sequence is selected. This can be compared to the extreme case where 163 candidate primer sequences selected when no sequence information is available.

## DISCUSSION

The experimental procedures utilized to determine the sequence of 2 oligonucleotides of unknown sequences from a mixture of 16 total have been described. In the example shown here the primer sequences for amplifying the STR locus D13S317 from an 8-plex PCR primer kit were determined to be forward 5′ ACA-GAAGTCTGGGATGTGGA 3′ and reverse 5′ GC-CCAAAAAGACAGACAGAA 3′ + fluorescein label. The sequence results we obtained for the DNA oligomers contained in HPLC fractions 3, 12, and 15 are an identical match with what was recently published by the Promega Corporation (15) for the D13S317 STR primers contained in PPlex1.1. Sequence information was not available for the PPlex1.1 kit when performing our experiments; thus the exact sequence matches serve as a confirmation for the DNA oligomer sequencing strategy.

IP-RP HPLC was a key technique for the isolation and separation of the 16-component mixture. It is possible to sequence a combination of oligomers simultaneously, but the complexity of the resulting mass spectrum increases rapidly with increasing digest products. An additional benefit of the IP-RP HPLC isolation protocol is that the isolated nucleic acids are essentially salt free. This reduction in salt concentration allows digests to be performed on picomolar amounts of sample.

The reference sequence for each loci obtained from GenBank (containing the PCR product sequence) combined with simple computational methods and information contained in the multiplex kit (dye, loci, etc.) greatly expedites the process of sequence determination. Because of this reference sequence information the complete digestion of each component is not required to determine the full sequence. It is experimentally possible to completely sequence each DNA strand from just the exonuclease digest data. However, some exonucleases exhibit preferential activity to certain sequence content (14). In addition, the dye-labeled PCR primers are blocked at the 5′ end and thus will not usually digest. The use of limited sequence information (5 to 10 bases) combined with a searching algorithm allows the number of digest experiments to be reduced. This results in a more rapid (fewer experiments) and less expensive (less enzyme and HPLC collected sample used) methodology for DNA oligomer sequence determination.

It is important to note that some PCR primers or oligonucleotide probes may be intentionally designed with mismatches, tails, and other modifications. These of course will not be directly selected by the computer algorithm that solely relies on the integrity of the reference sequence from GenBank for inferring the possible sequence. Typically for short oligomers (<30 bases) an agreement between the solved sequence

TABLE 2

Results of Sequence Matching Algorithm

| Sequence input | Output (number of matches) | |
|---|---|---|
| | Forward | Reverse |
| 5′ . . . ACAGACAGAA 3′ | — | 1 |
| 5′ . . . CAGACAGAA 3′ | — | 1 |
| 5′ . . . AGACAGAA 3′ | — | 1 |
| 5′ . . . GACAGAA 3′ | — | 1 |
| 5′ . . . ACAGAA 3′ | 1 | 1 |
| 5′ . . . CAGAA 3′ | 1 | 1 |
| 5′ . . . AGAA 3′ | 1 | 1 |
| 5′ . . . GAA 3′ | 1 | 3 |
| 5′ . . . AA 3′ | 3 | 6 |
| 5′ . . . A 3′ | 29 | 18 |
| No sequence information | 101 | 62 |

mass and the actual sequence mass should be within 2 Da. We have successfully applied the approach described here to deciphering unknown oligonucleotides from a mixture of as many as 34 components. Some components may be solved to a "unique" solution from as little as 5 bases of sequence information plus the overall primer mass and HPLC retention time.

Variations on the methodology described here can easily be applied for the quality control of commercially synthesized nucleic acids. The ability to develop a representative profile for a multiplex PCR amplification kit by establishing a reference mass spectrum and HPLC chromatogram of an intact working kit allows the user or manufacturer to quickly and inexpensively confirm the PCR primers present in the kit. This approach would not necessarily reveal the full sequences of the primers, but rather instill confidence that the correct full-length primers were present in the kit due to the unique masses and retention times observed (6).

The general methodology integrating analytical techniques, molecular biology, and informatics is straightforward and can be modified to quality control differing types of nucleic acid mixtures. This technology has the potential for automated analysis. The high-throughput ability of robotic systems for sample preparation and mass spectrometers for sample analysis makes this methodology reasonable for quality controlling large numbers of batches of synthetic nucleic acids utilized in various DNA applications.

## REFERENCES

1. Edwards, M. C., and Gibbs, R. A. (1994) Multiplex PCR: Advantages, development, and applications. *PCR Methods Appl.* **3,** S65–S75.

2. Wang, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., *et al.* (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280,** 1077–1082.

3. Tuerk, C., and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249,** 505–510.

4. Whitcombe, D., Newton, C. R., and Little, S. (1998) Advances in approaches to DNA-based diagnostics. *Curr. Opin. Biotechnol.* **9,** 602–608.

5. Huber, C. G., Oefner, P. J., and Bonn, G. K. (1993) High-resolution liquid chromatography of oligonucleotides on nonporous alkylated styrene-divinylbenzene copolymers. *Anal. Biochem.* **212,** 351–358.

6. Butler, J. M., Devaney, J. M., Marino, M. A., and Vallone, P. M. (2001) Quality control of PCR primers used in multiplex STR amplification reactions. *Forensic Sci. Int.* **119,** 87–96.

7. Pieles, U., Zurcher, W., Schar, M., and Moser, H. E. (1993) Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: A powerful tool for the mass and sequence analysis of natural and modified oligonucleotides. *Nucleic Acids Res.* **21,** 3191–3196.

8. Nordhoff, E., Ingendoh, A., Cramer, R., Overberg, A., Stahl, B., Karas, M., Hillenkamp, F., and Crain, P. F. (1992) Matrix-assisted laser desorption/ionization mass spectrometry of nucleic acids with wavelengths in the ultraviolet and infrared. *Rapid Commun. Mass Spectrom.* **6,** 771–776.

9. Ross, P., Hall, L., Smirnov, I., and Haff, L. (1998) High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nature Biotechnol.* **16,** 1347–1351.

10. Griffin, T. J., Hall, J. G., Prudent, J. R., and Smith, L. M. (1999) Direct genetic analysis by matrix-assisted laser desorption/ionization mass spectrometry. *Proc. Natl. Acad. Sci. USA* **96,** 6301–6306.

11. Smirnov, I. P., Roskey, M. T., Juhasz, P., Takach, E. J., Martin, S. A., and Haff, L. A. (1996) Sequencing oligonucleotides by exonuclease digestion and delayed extraction matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Anal. Biochem.* **238,** 19–25.

12. Chiu, N. H., Tang, K., Yip, P., Braun, A., Koster, H., and Cantor, C. R. (2000) Mass spectrometry of single-stranded restriction fragments captured by an undigested complementary sequence. *Nucleic Acids Res.* **28,** E31.

13. Lazaruk, K., Walsh, P. S., Oaks, F., Gilbert, D., Rosenblum, B. B., Menchen, S., Scheibler, D., Wenz, H. M., Holt, C., and Wallin, J. (1998) Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument. *Electrophoresis* **19,** 86–93.

14. Bentzley, C. M., Johnston, M. V., and Larsen, B. S. (1998) Base specificity of oligonucleotide digestion by calf spleen phosphodiesterase with matrix-assisted laser desorption ionization analysis. *Anal. Biochem.* **258,** 31–37.

15. Masibay, A., Mozer, T. J., and Sprecher, C. (2000) Promega Corporation reveals primer sequences in its testing kits. *J. Forensic Sci.* **45,** 1360–1362.