STR Analysis by Time-of-Flight Mass Spectrometry

By John M. Butler, Ph.D.
GeneTrace Systems Inc., 1401 Harbor Bay Parkway, Alameda, CA
e-mail: butler@genetrace.com

INTRODUCTION

We live in an age of rapid discovery and development in the fields of DNA information and technology. During the past decade we have seen the development of DNA typing methods for human identification, as well as the development of the polymerase chain reaction (PCR) for amplifying small quantities of DNA. At the same time, technologies for analyzing DNA have become faster, cheaper and more sophisticated. In particular, capillary electrophoresis, microchip hybridization, and mass spectrometry have blossomed into reliable methodologies for DNA analysis. Most important for forensic laboratories, DNA typing results are now widely accepted in courtrooms and will continue to play a significant role with the development of large databases housing DNA profiles of criminal offenders. DNA markers used by forensic and paternity testing laboratories have also been through an evolutionary process from VNTRs, using RFLP methods, to reverse-dot blot systems, to the increasingly popular STR markers that use PCR-based methods. This article will focus on new developments with STR analysis by emerging mass spectrometry techniques.

Mass spectrometry is a versatile analytical technique that involves the detection of ions and the measurement of their mass-to-charge ratio.

MASS SPECTROMETRY AND DNA

Mass spectrometry is a versatile analytical technique that involves the detection of ions and the measurement of their mass-to-charge ratio. Because these ions are separated in a vacuum environment, the analysis times can be extremely rapid, on the order of microseconds. Many advances have been made in the past decade for the analysis of biomolecules such as DNA, proteins and carbohydrates. Two such advances are the introduction of a new ionization technique known as matrix-assisted laser desorption-ionization (MALDI) and the discovery of new matrices that effectively ionize DNA without extensive fragmentation. When coupled with time-of-flight mass spectrometry, this method for measuring biomolecules is commonly referred to as MALDI-TOF-MS.

The MALDI-TOF process involves mixing a DNA sample with an organic matrix and allowing this mixture to co-crystallize in a spatial array on a sample plate. A sample plate may contain hundreds of different samples in a single run. Each sample crystal is then analyzed by moving the sample plate underneath a fixed, pulsed laser beam. A pulse of laser energy liberates and ionizes a small portion of the DNA sample. The matrix protects the DNA molecules from fragmentation and assists in the ionization process. These rapidly generated DNA ions travel through a "flight tube" to the detector in a matter of microseconds with the ions' "time-of-flight" based upon their mass (Figure 1). Small ions have a greater velocity than large ones and reach the detector first. To improve signal processing, the results from multiple laser pulses are averaged, which lengthens the measurement time to a few seconds for each sample. High-speed electronics and computers are required to keep up with the data collection process, which occurs on a nanosecond time scale.

Data points in mass spectrometry are collected in spectral channels that must be converted from a time value to a mass value. This mass calibration is normally performed with two oligonucleotides that span the mass range being examined. The calibration remains consistent over hundreds of samples and is typically performed only once a day. The mass accuracy and precision are such that no sizing standards or allelic ladders need to be run to determine a sample's size or genotype (1). Electrophoretic measurements, whether on a slab gel or in a capillary, require size markers and allelic ladders to adjust for variation in DNA mobilities under slightly different environmental conditions (2). Not only can DNA measurements be obtained more rapidly using mass specrometry, but the measurements are also more accurate.

Not only can DNA measurements be obtained more rapidly using mass spectrometry, but the measurements are also more accurate.

DNA Analysis by Time-of-Flight Mass Spectrometry

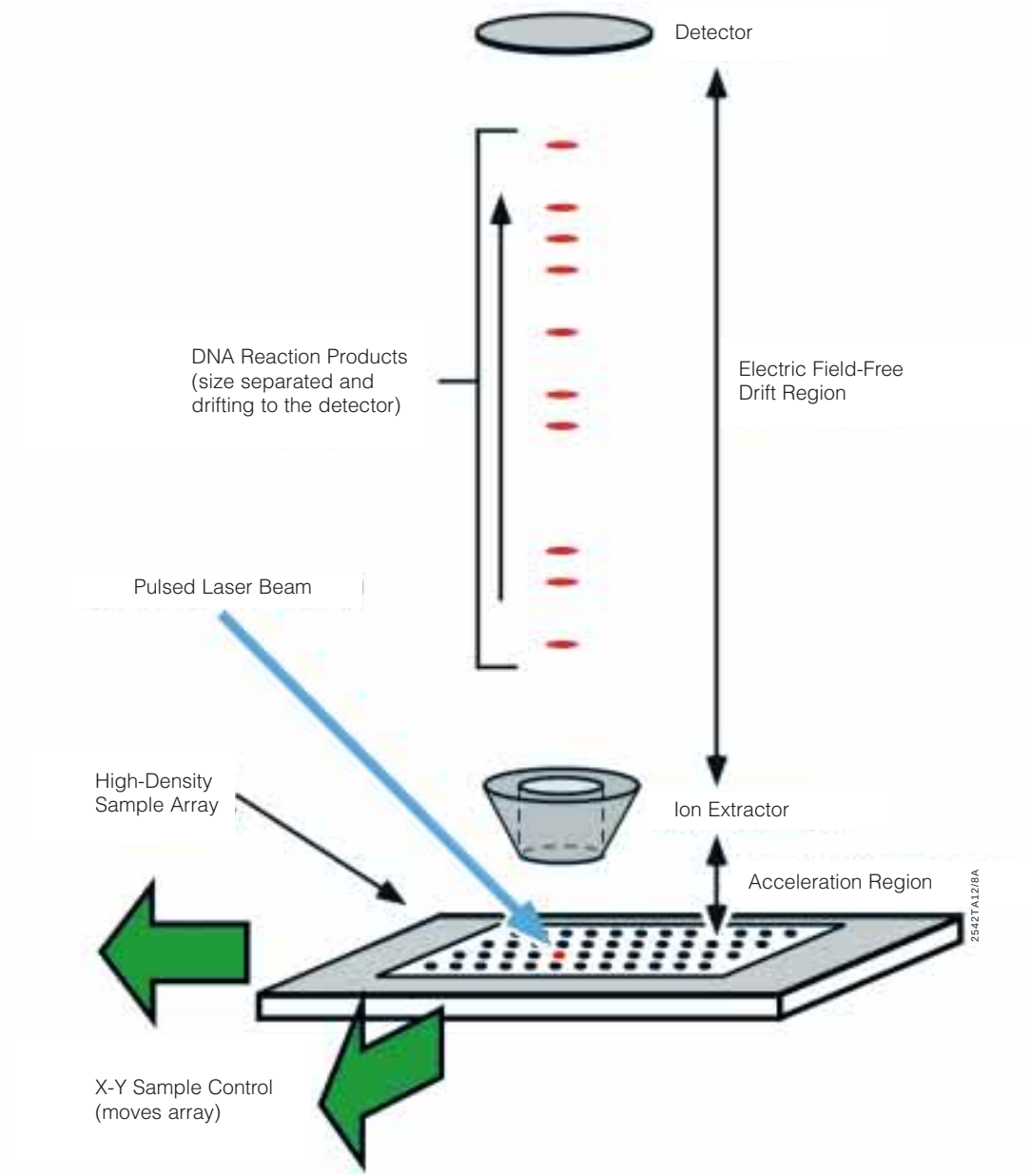
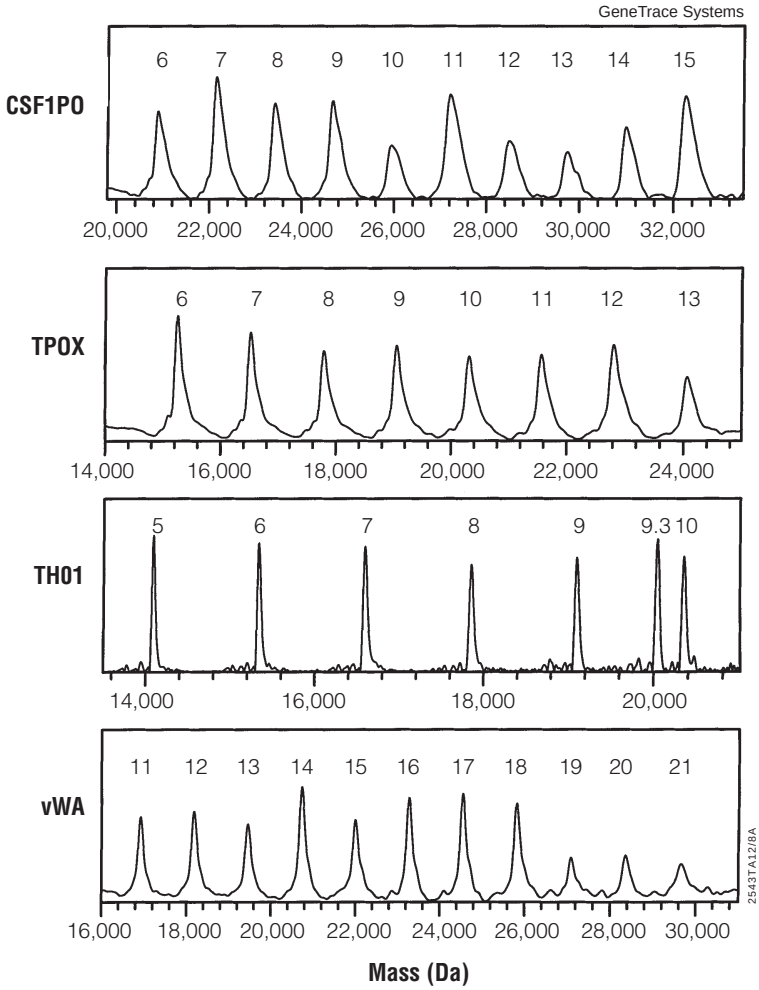


Figure 1. Schematic of GeneTrace automated time-of-flight mass spectrometer. DNA molecules are liberated from a solid-phase matrix environment with a laser pulse. The DNA reaction products are separated by size (mass) in a matter of microseconds. For each run, hundreds of samples are prepared in parallel using a robotic workstation and spotted onto a sample plate that is introduced to the vacuum environment of the mass spectrometer. The sample plate moves under the fixed laser beam to allow sequential sample analysis.

CAVEATS FOR STR ANALYSIS BY MASS SPECTROMETRY

There are several caveats for analysis of PCR products, such as short tandem repeat (STR) markers, using MALDI-TOF-MS. Mass spectrometry resolution and sensitivity are diminished when either the DNA size or the salt content of the sample is too high. By redesigning PCR primers to bind close to the repeat region, the STR allele sizes are reduced and resolution and sensitivity of the PCR products are enhanced. For example, TPOX alleles 6-14 range from 69-101 base pairs in size with GeneTrace-designed primers while with Promega's *GenePrint*™ System primers the same TPOX alleles range in size from 224-256bp. New PCR primers have been designed for numerous STR markers of forensic interest including all of the *GenePrint*™ tetranucleotide STR systems and the 13 CODIS (the FBI's Combined DNA Index System) STR loci. Where possible, we design our primers to produce amplicons that are less than 100bp, although we have been able to resolve neighboring STR alleles that are as large as 140bp in size. GeneTrace scientists are also working on new matrices to extend the size range for high-resolution DNA measurements.

To overcome the sample salt issue, we use a patented solid-phase purification procedure that reduces the concentration of magnesium, potassium and sodium salts in the PCR products (3). Our sample purification procedure, which has been entirely automated on a 96-tip robotic workstation, reduces the PCR buffer salts and yields "clean" DNA for the mass spectrometer. Using our robotic workstation in combination with a single high-throughput mass spectrometer, we have been able to purify and analyze over 2,000 samples in a single day.



THE NEED FOR HIGH-THROUGHPUT DNA ANALYSIS METHODS

Rapid, cost-effective methods for high-throughput DNA analysis are needed to process samples currently being gathered for large criminal DNA databases around the United States. As some states have backlogs of over 50,000 samples and limited funds and manpower to analyze these samples, new approaches are needed. Currently available slab gel or capillary electrophoresis methodologies can handle only a few dozen samples per day. A possible solution for STR analysis with the capability of handling several thousand samples per day is time-of-flight mass spectrometry coupled with robotic sample preparation (4).

The key to high-throughput DNA analysis is automation of every step from PCR reaction assembly to data analysis. Multiple robotic workstations and software packages are necessary to accomplish all of these tasks. GeneTrace uses separate robotic workstations for working with DNA samples and PCR materials prior to PCR, and for manipulating amplicons after PCR, to avoid possible contamination. Hundreds of samples may be prepared and amplified in parallel using, for example, a 384-well sample tray. GeneTrace mass spectrometers have automated data collection capabilities and require only a few seconds to collect and save the mass spectra from STR samples. GeneTrace-designed genotyping software then correlates the observed peak mass back to a genotype based on expected allele masses obtained from a reference sequence, the PCR primer positions and the repeat unit mass.

The data from each sample can be processed and genotyped in approximately one second using a standard desktop personal computer. Accurate genotyping may be performed without the use of an allelic ladder (1). However, we do find allelic ladders to be useful in making resolution and mass accuracy measurements. Figure 2 shows mass spectrometry results with CSF1PO, TPOX, TH01 and vWA allelic ladders. The PCR product size and observed mass of each TH01 allele are shown in Table 1 along with the speed at which the DNA ions are detected with time-of-flight mass spectrometry.

Figure 2. Allelic ladders for CSF1PO, TPOX, TH01 and vWA STR loci. The numbers above each peak designate the allele name (number of repeats). Peak widths vary between samples based on DNA size and salt content. On a mass scale as shown here, each nucleotide is approximately 300 Daltons (Da).

CONCLUSIONS

Time-of-flight mass spectrometry offers a rapid, cost-effective alternative for genotyping large numbers of samples. Each DNA sample can be accurately measured in a few seconds. Due to the increased accuracy with mass spectrometry, STR alleles may be reliably typed without comparison to allelic ladders. Mass spectrometry holds significant promise as a technology for high-throughput DNA processing that will be valuable for large-scale DNA database work.

REFERENCES

1. Butler, J.M. *et al.* (1998) *Int. J. Legal Med.* **112** (1), 45.
2. Schumm, J.W. (1997) *Profiles in DNA* **1**(2), 11.
3. Monforte, J.A. *et al.* (1997) U.S. Pat. No. 5,700,642.
4. Monforte, J.A. and Becker, C.H. (1997) *Nature Med.* **3**, 360.

Table 1. TH01 Allelic Ladder Measurements.

Allele	PCR Size	Measured Mass	Separation Time
5	63bp	13,975Da	170µsec
6	67bp	15,230Da	177µsec
7	71bp	16,492Da	184µsec
8	75bp	17,766Da	191µsec
9	79bp	19,019Da	198µsec
9.3	82bp	19,973Da	203µsec
10	83bp	20,280Da	204µsec

Advantages of Mass Spectrometry for STR Analysis.

- **Rapid results**—STR typing at a rate of seconds per sample
- **Accurate** without allelic ladders
- **Direct DNA measurement**—no fluorescent or radioactive labels
- **Automated** sample preparation and data collection
- **High-throughput**—capable of processing thousands of samples per day