Laser desorption mass spectrometry for high throughput DNA analysis and its applications

C. H. Chen*[a], V. V. Golovlev[b], N. I. Taranenko[b], S. L. Allman[a], N. R. Isola[b], N. T Potter[c], K. J. Matteson[c] and L. Y. Chang[d]

[a]Life Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6378

[b]Oak Ridge Associated Universities, Oak Ridge, TN

[c]University of Tennessee Medical Center, Knoxville, TN

[d]Chinese Academia Sinica, Taipei, Taiwan

ABSTRACT

Laser desorption mass spectrometry (LDMS) has been developed for DNA sequencing, disease diagnosis, and DNA fingerprinting for forensic applications. With LDMS, the speed of DNA analysis can be much faster than conventional gel electrophoresis. No dye or radioactive tagging to DNA segments for detection is needed. LDMS is emerging as a new alternative technology for DNA analysis.

Keywords: DNA sequencing, disease diagnosis, DNA fingerprinting, laser desorption mass spectrometry

1. INTRODUCTION

During the past few decades, the need of rapid and reliable technologies for DNA analysis becomes critical due to the fast progress in molecular biology, bioengineering and gene discovery. In the early 1990's, the Human Genome Project was implemented with the goals to complete the mapping and sequencing of the entire human genome. Many new approaches to analyze and/or sequence DNAs were initiated and pursued. Hybridization on chips and mass spectrometry for DNA analysis are among the most promising new technologies for DNA analysis. However, most routine DNA analysis nowadays still relies on gel electrophoresis. During the past few years, significant progress has been achieved in gel electrophoresis. Automated DNA sequencer has become a major tool for DNA analysis and sequencing. Capillary gel electrophoresis is also expected to be broadly used for sequencing project. Recent development of electrophoresis on chip can also become a valuable tool for DNA analysis in the future. Nevertheless, there are several disadvantages for DNA analysis by gel electrophoresis. They include (1) the time needed to sequence DNA is long. It can take from minutes to hours. (2) all gel electrophoresis approaches need to tag DNA either by radioactive materials or dyes. This process increases time, cost and producing hazardous waste. Furthermore, (3) DNA segments with high guanine (G) component, long tandem repeats and/or secondary structures tend to have band compressions which are often very difficult to resolve. And (4) it is often tedious and need highly experienced technicians to prepare gels. Thus it is very desirable to have a new technology to eliminate the above disadvantages.

It is naturally desirable to consider a time-of-flight mass spectrometry approach for DNA analysis and sequencing since the principle to measure DNAs is similar to gel electrophoresis. However, most mass spectrometric approaches failed to keep the large DNAs intact during the vaporization and ionization processes. In order to overcome this difficulty, mass spectrometry coupled with electrophoresis was pursued for DNA sequencing. In 1987 Hillenkamp and his co-workers[1] discovered that large protein molecular ions can be produced by laser desorption without much fragmentation if these biomolecules are mixed with small organic compounds that serve as matrices for strong absorption of a laser beam. This process is now called matrix-assisted laser desorption/ionization (MALDI). The typical preparation technique for MALDI is to dissolve biomolecular samples in solution, then prepare another solution that contains small organic compounds such as 3-hydroxypicolinic acid. These two

DISCLAIMER

solutions are subsequently mixed and a small amount of solution is placed on a metal plate to dry. After the small organic material crystallizes, the sample plate is placed in the mass spectrometer for analysis. During the MALDI process, matrix materials strongly absorb the laser energy and quickly become vaporized. Large biomolecules are carried along during the fast vaporization process. By this method, large biomolecules can be delivered into space without breakup, which is probably due to the minimal direct absorption of laser energy; thus, "soft" desorption can be achieved. Furthermore, it has also been found that protein ions can be produced during the laser desorption process in addition to the expected neutral molecules. The production of ions is speculated to involve a proton transfer process. With MALDI, the velocity distribution of DNAs is very similar to that of matrix molecule. The mass resolution is often poor due to the large spread of velocity. Optical acoustic desorption was also used to achieve soft desorption of DNAs. Since various laser desorption methods can be used to desorb DNAs, we use the term of laser desorption mass spectrometry (LDMS) to include various laser desorption mechanism for mass spectrometric measurements. In this work, we report our work of using LDMS for sequencing of DNAs, DNA measurements for disease diagnosis and forensic applications.

2. EXPERIMENTAL

The experimental schematic for using a time-of-flight mass spectrometer (TOF-MS) for DNA analysis and/or sequencing is shown in Fig.1. A TOF-MS is a device in which ions of different masses are given the same energy and are allowed to travel in a field free space. Because of their different velocities due to the different masses, the ions of differing mass arrive the detector at different times. An ion detector then produces a time-varying electrical signal proportional to the number of ions impacting it. The amplified current signal is typically displayed on a fast digital oscilloscope, whereupon the molecular weights can be determined by measuring the time interval between the creation and the detection of ions. To achieve high mass resolution, the ions must be produced in a very short time interval and in a very small volume. In our laboratory, either a third harmonic of a Nd-Yag laser beam at 355 nm or a nitrogen laser at 337 nm was used to achieve laser desorption and ionization. A recently developed delayed pulsed ion extraction[2], which is used to compensate the velocity distribution of desorbed ions was installed in the TOF-MS to improve mass resolution. DNA samples for LDMS analysis were placed on a stainless plate with array structure for TOF-MS analysis. One hundred samples can be put in the TOF-MS for analysis at the same time. The time needed for the analysis of each sample is often less than 10 seconds.

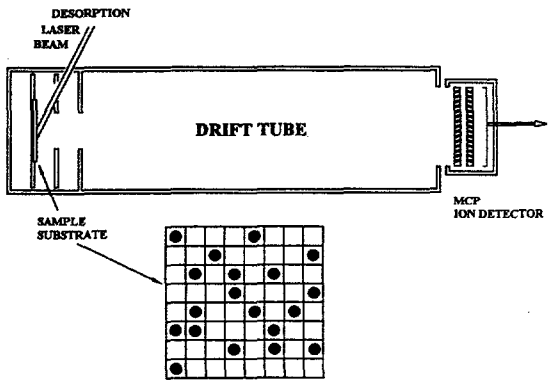


Fig. 1 Schematic of experimental setup for LDMS-TOF-MS.

The typical preparation process for DNA samples for LDMS analysis is to dissolve biomolecular samples in solution, then prepare another solution containing small organic compounds such as picolinic acid. These two solutions are mixed, and a small amount of solution is placed on a metal plate to dry. After the sample has crystalized, the sample plate is placed in the mass spectrometer for analysis. The molar ratio of matrix to analyte is typically more than 100:1. All DNA samples obtained from PCR process or Sanger's process to produce DNA ladders were purified using QIAquick DNA purification kit (QIAGEN, Inc., Chatsworth, GA 91311), precipitated and concentrated to eliminate most impurties.

3. RESULTS AND DISCUSSION

3.1 DNA sequencing

3.1.1 DNA sequencing with Sanger's method to produce DNA ladders

DNA sequencing has been broadly used for biomedical research and clinical applications during the past two decades. With the conventional sequencing approach, different sizes of DNA ladders, which are produced by either Sanger's enzymatic method or Maxam-Gilbert's chemical cleavage method, are separated by gel-electrophoresis to achieve sequencing. As described in the above, LDMS can be used to achieve DNA sequencing at much faster speed. However, the size of DNAs that can be sequenced

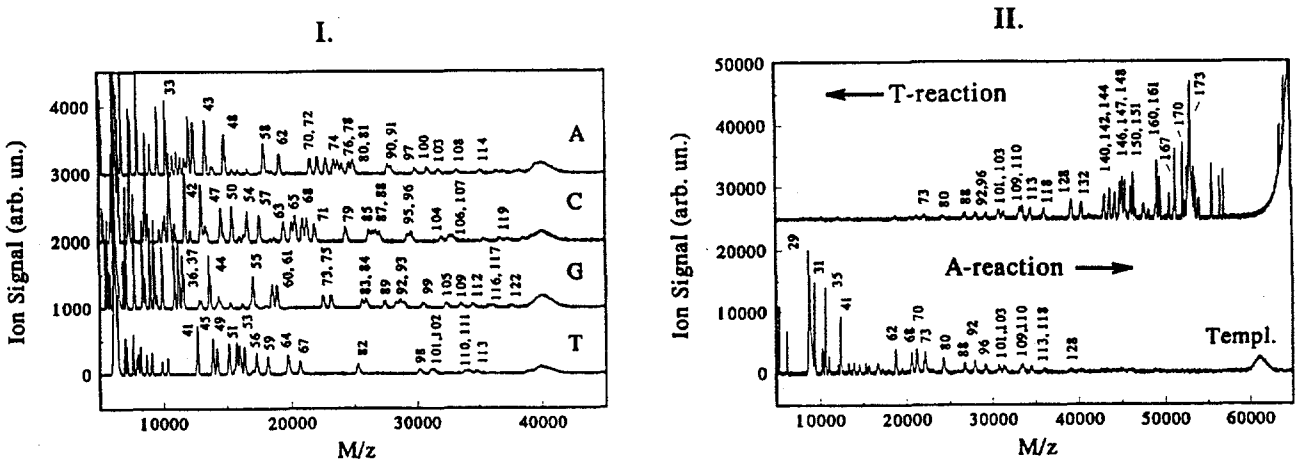


Fig. 2 (I) Negative ion mass spectra of Sanger's sequencing ladders producted by cycle sequencing from PCR products of a double-stranded 130 bp template.

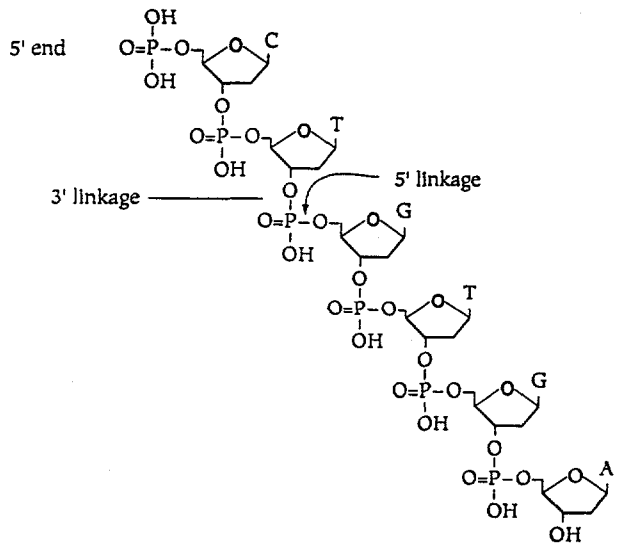
(II) Mass spectra for sequencing with a template of 200 bp. The results for the A reaction with the reverse primer and T reaction with forward primer.

using mass spectrometry is shorter than the gel electrophoresis method because mass resolution and detection sensitivity decrease rapidly for longer DNAs. Recently, we achieved the first LDMS sequencing of DNA with the length longer than 100 nucleotides[3]. Cycle sequencing was used to obtain DNA ladders. Fig. 2(I) shows the negative-ion mass spectra of DNA ladders from A, C, G and T reactions using double-stranded DNA 130 base pairs (b.p.) as the template. Sequencing identification can be obtained up to 120 mer. Sequencing beyond 120 mer can not be identified due to the poor resolution. Since sequencing by using both forward and backward primers can reach 120 mer, complete sequencing of 200 b.p. DNA can be achieved. Figure 2(II) illustrates the idea of combining the spectra from forward and backward primers to sequence ds-DNA with a 200 b.p. template. Due to the measurement of mass to identify the size of DNAs, there is no concern about a missing band which can occur in the gel electrophoresis method. Thus, redundant sequencing can possibly be eliminated with LDMS sequencing of DNAs.

At present, automatic gel sequencers can read up to 1000 b.p., but most routine sequencing ranges from 300-500 bases. We present results utilizing LDMS to determine sequences up to 200 b.p., indicating that mass spectrometry is emerging as a useful tool for DNA sequencing. Mass spectrometry sequence detection utilizes inexpensive unlabelled primers and also has the additional advantages of higher sequencing speed and the ability of obtaining sequence information close to the primer which is often difficult with gel techniques. With rapid discovery of new genes, mass spectrometry should be very valuable for re-sequencing for mutation detection since the size of DNAs to be re-sequenced usually does not need to be long. We expect mass spectrometry will complement the existing sequencing technologies by being a fast, label-free method as well as help in resolving sequence ambiguities.

3.1.2 Direct DNA sequencing by LDMS with selective fragmentation without DNA ladder preparation

In general, the sequencing of DNA required the use of the Sanger or Maxam-Gilbert methods to produce DNA ladder products. Recent developments in laser desorption time-of-flight mass spectrometry has the potential to drastically increase the sequencing speed. However, this approach still needs the time-consuming step of DNA ladder preparation. A more efficient method would be to obtain the DNA sequence information from fragmentation patterns in mass spectroscopy through dissociation. If this approach succeeds, both the time-consuming steps of DNA ladder preparation and gel electrophoresis can be eliminated. Direct DNA sequencing can be easily understood by using the illustration in Figs. 3 and 4. Assuming the selective cleavage of P-O bond at 3'-linkage for an oligonucleotide of 5'-CTGTGA-3', the primary fragmentation of such bond breaking will produce two series of fragments. These are labeled as 5'- and 3'-termini (see Fig. 3). Each series has six members ranging from 1- to 6-mer. The vertical dotted line in Fig. 3 represents the position of cleavage. The fragments in both series can be ionized and resolved in mass spectra. Figure 4 gives the simulated mass spectra for these two series. Trace 'a' in Fig. 4 (top trace) is the spectrum of the series with 3'-termini. The mass difference between two adjacent peaks in the series provides the information for each extra base.



Fragment pattern of breaking 3 linkage

5' -termini

5' -C-T-G-T-G-A-3'
5' -C-T-G-T-G-A-3'
5' -C-T-G-T-G-A-3'
5' -C-T-G-T-G-A-3'
5' -C-T-G-T-G-A-3'

3' -termini

Fig. 3 Illustration of the selective cleavage P-O bond and its fragmentation pattern for a 5'-CTGTGA-3' oligonucleotide.

sequencing information can be obtained by the mass spectra. Short oligonucleotides with the size up to 35 nt were successfully sequenced. Experimental results are shown in Fig. 5. In order to sequence longer DNAs with this approach, mass resolution needs to be very high since the mass difference between adenine (A) and thymine (T) is only 9 Da.

Synthetic oligonucleotide are becoming increasingly important in the research fields of molecular biology and medicine. They can be used as primers in DNA sequencing and in the PCR amplification process; as probes in cDNA hybridization studies; and for assays of target DNA for diagnostic testing. Modified oligonucleotide are increasingly used for the pharmacological control of specific gene expression. The rapid growth in the use of synthetic oligonucleotide has created a demand for a reliable analytical technique which can accurately and quickly determine the integrity of the synthetic product. Up to

For example, the first member of this series is A, and the second member of this series is GA. Due to the selective cleavage of 3'-linkage, the mass difference between these two peaks will be exactly 329.2 Dalton, which represents the mass of dGMP. The sequence information can be obtained by analyzing all values of mass difference in this series together with the total mass of 5'-CTGTGA-3'. The same information can also be obtained from the series with 5'-termini (Fig. 4, trace 'b'). If a complete series (either 5'- or 3'- termini) of such fragmentation for a DNA sample can be experimentally resolved, the full sequence information will be obtained. Trace 'c' of Fig.4 shows a mass spectrum of two series from the cleavage of 5'-linkage P-O bond. This sequencing method is very reliable since the sequence result from one series can be reconfirmed by the other series. However, sequencing longer oligonucleotide will need much better mass resolution since the mass difference between different bases need to be identified. If a 30 mer oligonucleotide is to be sequenced, the accuracy of the ion peaks need to be better than 1 in 1000 since the mass difference between A (adenine) and T (thymine) is only 9 Dalton. With this approach, the sequencing of each probe in average can take less than 1 minute per sample since no wet chemistry processes are needed. By controlling pH values of matrices and laser fluence, we succeeded in producing DNA ladders during the laser desorption process[4]. Thus, the

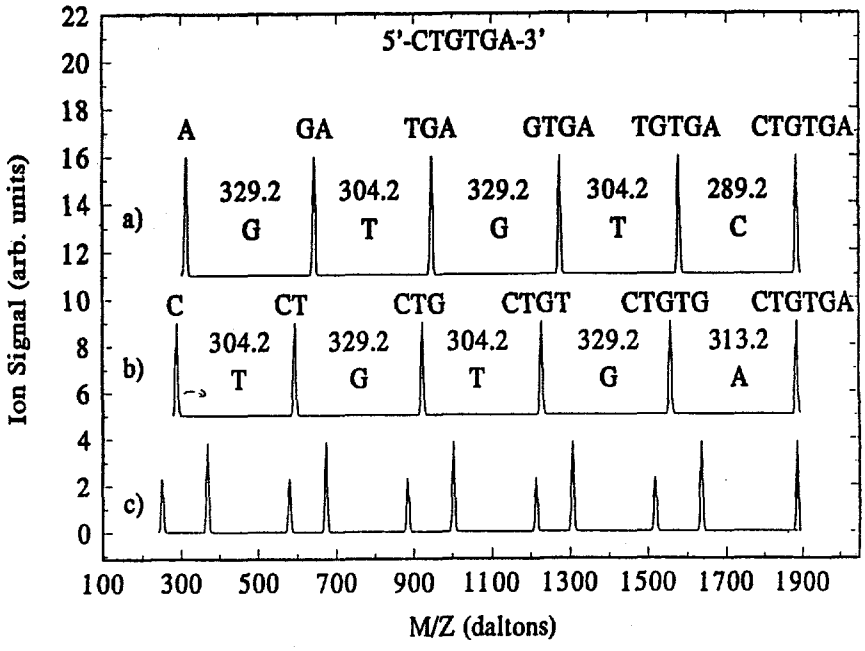


Fig. 4 Simulated spectra of the selective cleavage of P-O 3'-linkage and 5'-linkage for 5'-CTGTGA-3' oligonucleotide. a) spectrum of fragment series with 3'- termini by the 3'-cleavage; b) spectrum with 5'-termini by the 3'-cleavage; and c) combination of both a) and b).

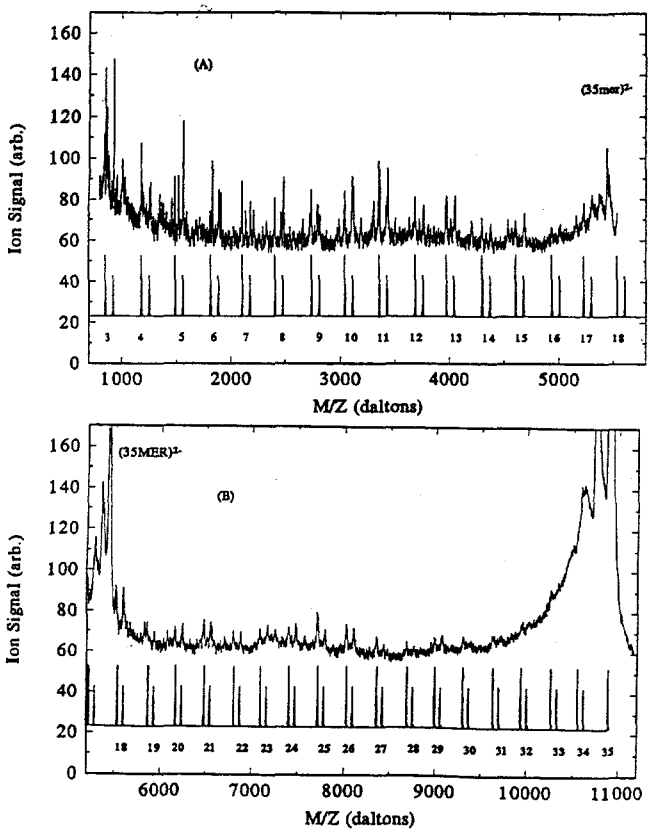


Fig. 5 Direct sequencing of a 35 nt oligonucleotide. A mixture of 2,4,6-trihydroxyacetophenone (1.5 x 10^{-7} mole), citric acid (1 x 10^{-7} mole) and ammonia (2.3 x 10^{-7}) was used as matrix. 100 pmol of oligonucleotide was loaded. The laser fluence was 32 mJ/cm^2. (A) low mass side of spectrum and (B) high side of spectrum.

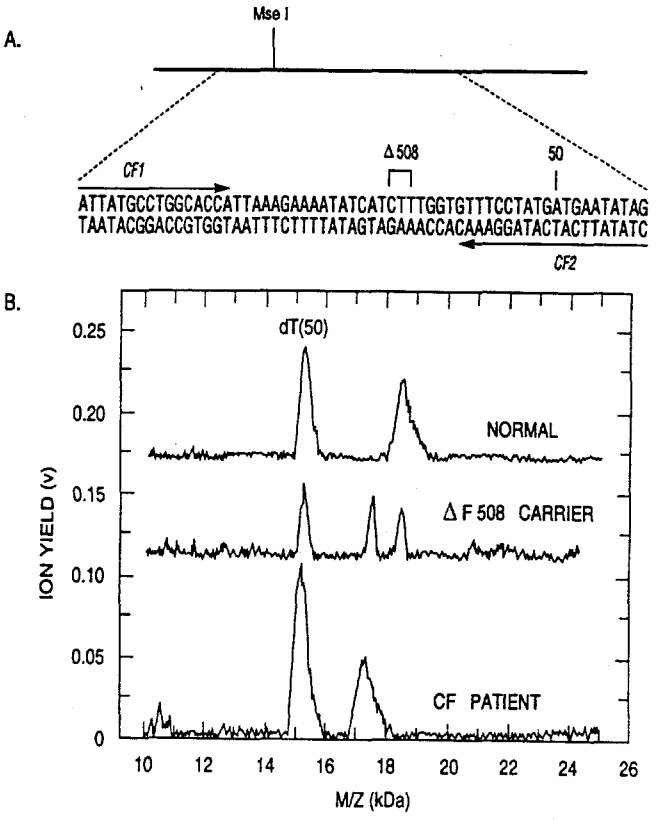


Fig. 6 Cystic fibrosis diagnosis of individuals with the normal and ΔF508 alleles by MALDI-TOF-MS. A) Nucleotide sequence of 59 bp in exon 10 of CFTR gene with primers CF1 and CF2 indicated. B) Negative ion MALDI-TOF-MS of DNA amplified from a healthy volunteer, a carrier, and a cystic fibrosis patient.

now, there have been few methods available for analysis of relatively short synthetic oligonucleotide. The sequencing of oligonucleotide by fragmentation in LDMS developed in this study immediately meets the requirement for the analysis of synthetic oligonucleotide. It can obtain accurate sequence information of synthetic oligonucleotide in short time and low costs.

3.2 Disease diagnosis

Since LDMS provides a fast method to measure DNA size and /or sequence short DNAs, LDMS can be used for the diagnosis of genetic diseases. Recently, we applied LDMS for disease diagnosis due to base deletion or insertion[5], point mutation[6] and dynamic mutation[7] from a high number of tandem repeats.

Cystic Fibrosis (CF) is the most common autosomal recessive genetic disease in Caucasians. The gene responsible for CF is composed of 27 exons spanning 250,000 base pairs (bp) and encodes a protein of 1480 amino acids named the cystic fibrosis transmembrane conductance regulator (CFTR)[8]. In north American population about 70% of CF carriers have a 3 bp deletion in exon 10, resulting in the loss of a phenylalanine residue at codon 508 (ΔF508). The results of the first demonstration of the measurement of ΔF508 mutation in cystic fibrosis carriers and patients by LDMS time-of-flight mass spectrometer are shown in Fig. 6. Two oligonucleotide primers, CF1 and CF2, were designed to do amplification of a DNA segment spanning the deletion and generate a 59-bp or a 56-bp fragment of the normal CF gene or ΔF508 mutation, respectively. (See Fig. 6a) Analysis of these 3-base difference between the normal and the ΔF508 single-stranded alleles in the spectra is shown in Fig. 6b. To validate the utility of LDMS, 30 genomic DNA samples taken from patients were used as templates to amplify DNA segments from exon 10 or other exon/intron regions as negative controls. All samples with ΔF508 mutation were separately verified by size fractionation in polyacrylamide gel. The mass spectrometric and conventional assay were in total agreement.

In addition to the demonstration of the analysis of ΔF508 base deletion for cystic fibrosis, LDMS coupled with allele specific polymerase chain reaction (PCR) was also used successfully for single base mutation for p53 gene and G551D point mutation in CFTR. An example of using LDMS for point mutation simulated for p53 is shown in Fig.7. Primers were extended to the hot spot. With ASPCR, at least one primer is chosen to complement to the segment of the template with the potential hot spot. For point mutation, PCR products will not be produced due to the mismatch of the primer and the template. With LDMS to detect the size of DNAs, the existence of point mutation at a specific site can be determined. In Fig. 7, the chosen template was a simulation of a fragment of p53 gene. The sequence of the template and the primers used for normal template are

5'-CAGCTGTGGGTTGATTCCACACCCCCGCCCGGCACCC
GCGTCCGCGCCATGGCCATCTAC-3' (Template)

5'-CAGCTGTGGGTTGATTCCACACC-3' (Primer 1)

5'-CGGGTGCCGGGCGGGG-3' (Primer 2)

Results are in Fig. 7(I). With mutant template, no PCR products are produced. Results are in Fig. 7(II).

Trinucleotide expansion has been found to associate with a growing number of inherited neurologic disorders[9]. Several human

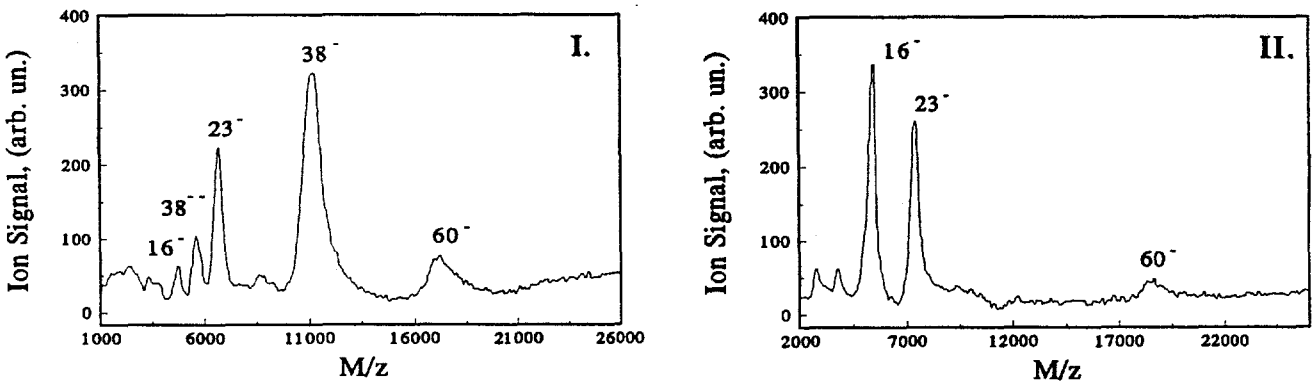


Fig. 7 (I) Negative ion mass spectrum of 38 b.p. PCR product. **(II)** No PCR products were produced with mutant template.

genes have been identified that contain trinucleotide CAG (cytosine-adenine-guanine) repeats within the coding sequences and currently include those genes associated with Huntington's disease (HD), Kennedy's disease, spinocerebellar ataxias and dentatorubral-pallidoluysian atrophy (DRPLA) These repeats are translated as polyglutamine tracts in the protein product. Typically, the stable and nonpathological alleles have between 10-30 repeats, while unstable pathological alleles have modest expansion, often in the range of 40 -100 repeats. Transcription and translation of the genes are not affected by the expansion; while a larger protein is made with the extra glutamines incorporated into the protein. Currently, the quantitative analysis of reiterated trinucleotide repeats is based on gel electrophoretic separation of repeat containing sequences. With LDMS, the analysis can be greatly improved. We recently applied LDMS to achieve successful disease diagnosis of DRPLA and Huntington's disease. An example of a sample from a patient with Huntington's disease is shown in Fig. 8.

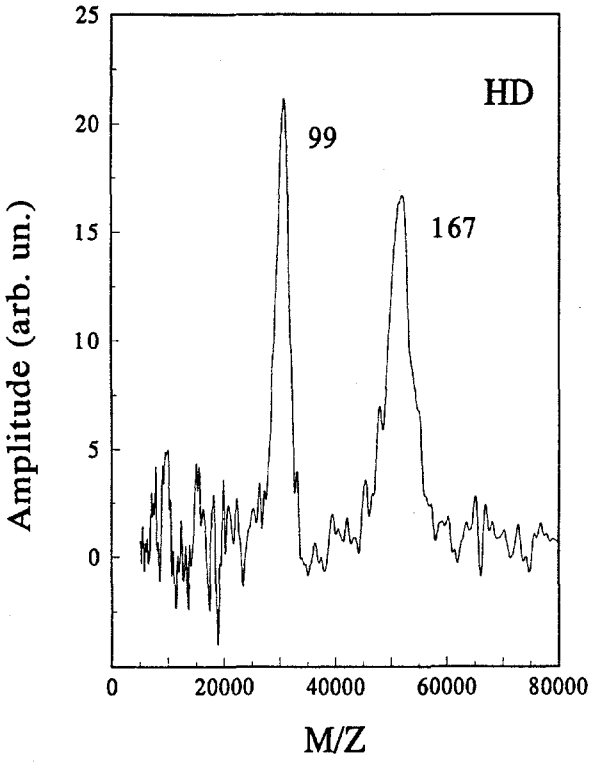


Fig. 8 Mass spectrum from Huntington disease sample. The sequence of primer 1 is 5'-ATGAAGGCCTTCGAGTCCCTCAAGTCCTTC-3'. The sequence of primer 2 is 5'-GGCGGTGGCGGCTGTTGCTGCTGCTGCTGC-3'.

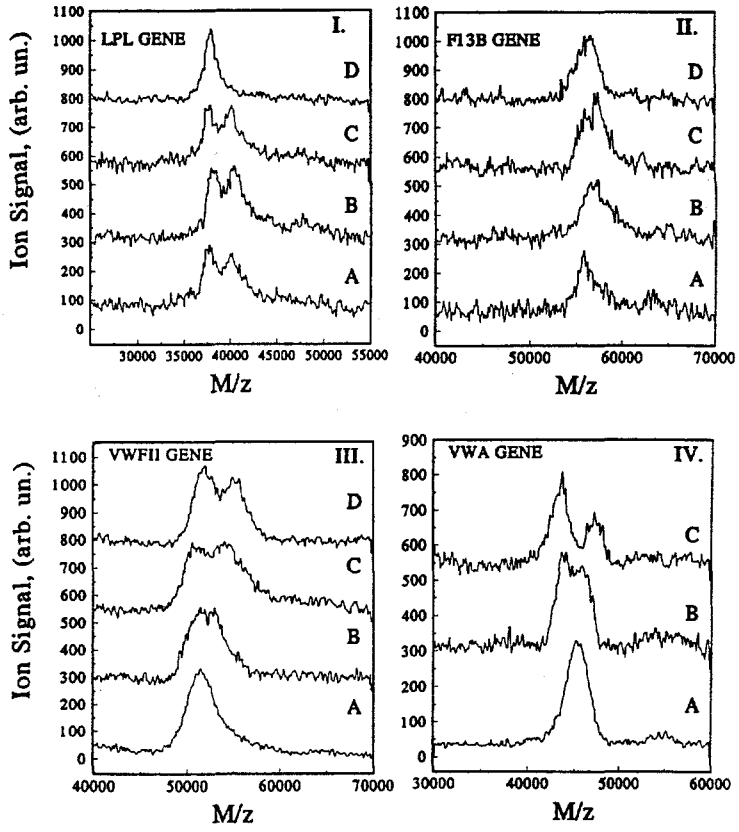


Fig. 9 Negative ion mass spectrum of PCR product from different patients. **I)** lipoprotein lipase gene (LPL), **II)** blood coagulation factor gene (F13B), **III)** VWF gene, and **IV)** VWA gene.

3.3 Measurements of Short tandem repeat (STR) for forensic application

Recently, DNA has become broadly utilized in forensic analysis due to the fact that every individual's DNA structure is nearly identical within all tissues of their body. DNA fingerprinting was initiated by the use of restriction fragment length polymorphism. At the molecular level, various polymorphism can range from a single nucleotide base change to a difference in the number of tandem repeat can be used as a reliable tool for the identification of individuals. Recently, the polymorphic loci contain repeats of 3 to 5 bp in length known as short tandem repeats (STR) has aroused lots attention for forensic application. Since STRs are scattered throughout the chromosomes in enormous numbers, there is a nearly unlimited potential for many loci to be discovered and validated for forensic use. In general, if there are n alleles, there should be n homozygous and n(n-1)/2 heterozygous genotype. If there are m loci with n alleles, there are $(n(n+1)/2)^m$ genotypes overall. Taking n as 10 and m as 4, there are more than 9 million genotypes, If only statistics is applied to two different persons in this example, the accuracy should reach almost one in 10^7. When special ethnic, sex and familial relationship are taken into account, the possibility of accidental matches can increase drastically. Thus, more loci are always preferred for enhancing the reliability of DNA fingerprinting. Most DNA segments which contain STR can be amplified by PCR. Thus, the quantity of the genomic sample needed for analysis can be significantly reduced. An example of using LDMS for STR for person identification is shown in Fig. 9. These spectra indicate that individuals with different alleles can be distinguished. Both homozygous and heterozygous for all 4 loci were observed. For the samples tested in this work, no two individuals have the same 4 base repeats in all 4 loci. The results show that LDMS may become a very valuable tool for STR analysis for forensic application.

4. CONCLUSION

MALDI has been demonstrated for DNA sequencing, disease diagnosis and DNA fingerprinting for forensic applications. However, further improvement in mass resolution is required in order to achieve reliable analysis and/or sequencing of longer DNAs. Automation on sample preparation and DNA purification are also critically needed to achieve high throughput analysis. Nevertheless, LDMS is emerging as a new and valuable tool for DNA analysis. More effort will be put to improve mass resolution and detection sensitivity for large DNA ions.

ACKNOWLEDGEMENTS

Research is sponsored by the National Institute of Justice with the grant number as 97-LB-VX-A047 and by the Office of Biological and Environmental Research, U. S. Department of Energy under contract DE-AC05-96OR22464 with Lockheed Martin Energy Research Corporation.

REFERENCES

1. Karas, M., Bachmann, D. , Bahr, U., and Hillenkamp, F., "Matrix-Assisted Ultraviolet Laser Desorption of Non-Volatile Compounds," Int. J. Mass Spectrom. Ion Process. 78, 53-56 (1987)

2. Vestal, M. L., Juhasz, P., and Martin, S. A., "Delayed extraction matrix-assisted laser desorption time-of-flight mass spectrometer", Rapid Comm. Mass. Spectrom. 9, 1044-1050 (1995)

3. Taranenko, N. I., Allman, S. L., Golovlev, V. V., Taranenko, N. V., Isola, N. R., and Chen, C. H., "Sequencing DNA using mass spectrometry for ladder detection", Nucleic Acids Research, 26, 2488-2490 (1998)

4. Zhu, Y. F., Taranenko, N. I., Allman, S. L., Taranenko, N. V., Martin, S. L., Haff, A., and Chen, C. H. , "Oligonucleotide sequencing by fragmentation in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry", Rapid Comm. Mass. Spectrom. 11, 897-903 (1997)

5. Chang, L. Y., Tang, K., Schell, M., Ringelberg, C., Matteson, K. J., Allman, S. L., and Chen, C. H., " Detection of F508 mutation of the cystic fibrosis gene by matrix-assisted laser desorption/ionization mass spectrometry", Rapid Comm. Mass. Spectrom. 9, 772-774 (1995)

6. Taranenko, N. I., Matteson, K. J., Chung, C. N., Zhu, Y. F., Chang, L. Y., Allman, S. L., Haff, L., Martin, S. A., and Chen, C. H., " Laser desorption mass spectrometry for point mutation detection", Genetic Analysis: Biom.& Eng. 13, 87-94 (1996)

7. Taranenko, N. I., Potter, N. T., Allman, S. L., , Golovlev, V. V., and Chen, C. H., "Detection of trinucleotide expansion in neurodegenerative disease by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry", Genetic Analysis: Biom & Eng. (In press)

8. Drumm, M.L., Smith, C. L., Dean, M., Cole, J. L., Iannnuzzi, M.C., and Collins, F. S., " Physical mapping of the cystic fibrosis region by pulsed-field gel electrophoresis", Genomics 2, 346-354 (1988)

9. Sutherland, G. R., and Richard, R. I., "Dynamic mutation", American Scientist, March-April issue pp 157-163 (1994)

*Author to whom correspondence should be addressed.