DNA sequence analysis by matrix-assisted laser desorption ionization MS

L. A. Haff and I. P. Smirnov
PerSeptive Biosystems, 500 Old Connecticut Path, Framingham, MA 01701, U.S.A.

Introduction
Sequencing oligonucleotides by conventional methods is actually more difficult than sequencing larger DNAs, since the well-established Sanger sequencing method is not easily applied to short non-cloned DNAs. Although methods have been developed for cloning oligonucleotides [1,2], they are laborious. We have developed a facile, reliable and inexpensive chemical method for sequencing oligonucleotides by delayed extraction matrix-assisted laser desorption ionization time-of-flight (DE MALDI-TOF) MS, which is an extension of methods originally described by Schuette et al. [3] and Pieles et al. [4].

The DNA is partially hydrolysed by a series of dilutions of a 3′-to-5′-acting phosphodiesterase and separately with a 5′-to-3′-acting phosphodiesterase (Figure 1). The ladder of partial digestion products is then analysed by a DE-equipped MALDI–TOF mass spectrometer and the sequence inferred from the mass differences between adjacent peaks.

Experimental
MALDI–TOF MS was carried out on a DE linear Voyager® TOF mass spectrometer (PerSeptive Biosystems) equipped with a nitrogen laser operating at 337 nm [5]. The total accelerating voltage was −20 kV. The enzyme digestion employed MALDI-compatible buffers, so the samples were analysed after digestion by simply adding the matrix, 3-hydroxypicolinic acid, and drying an aliquot on a MALDI sample plate. As an example, a 25-mer oligonucleotide (5′-AGG CAT GCA AGC TTG AGT ATT CTA T) was synthesized using an Expedite 8909 Nucleic Acid Synthesis System, cleaved and deprotected. The oligonucleotide was purified by reverse-phase C₁₈ HPLC, desalted with a Quick Spin Sephadex G-25 spin column, and 200 pmol was digested (in the 3′ to 5′ direction) with snake venom 3′-5′ exonuclease.

![Figure 1](image-url)
Method of exonuclease sequencing

Digests from the 3′ end are conducted with a 3′-to-5′-acting phosphodiesterase (snake venom). Separately, digests are carried out from the 5′ to 3′ direction with bovine spleen phosphodiesterase.
Figure 2
DE MALDI–TOF mass spectra obtained from digests of a 25-mer oligonucleotide with six different concentrations of snake venom phosphodiesterase (SVP)

Each sample was digested for 20 min with the indicated concentration of phosphodiesterase in ammonium citrate buffer, pH 9.4. Then 3-hydroxypicolinic acid was added, the sample dried on the MALDI sample plate and spectra were generated.

Figure 3
DE MALDI–TOF mass spectra obtained from digests of a 25-mer oligonucleotide with three different concentrations of bovine spleen phosphodiesterase (BSP)

Each sample was digested for 20 min with the indicated concentration of bovine spleen phosphodiesterase in ammonium citrate buffer, pH 5.1. Then 3-hydroxypicolinic acid was added, the sample dried on the MALDI sample plate and spectra were generated.
phosphodiesterase in 50 g/l ammonium citrate buffer at pH 9.4 for 20 min at 37°C. Typical spectra of six sets of digestion products, each produced by a different dilution of the enzyme (from 0.02 m-units/µl to 0.50 m-units/µl), are shown in Figure 2.

Multiple digestions are carried out in this procedure to ensure an adequate distribution of nested digestion products, regardless of the length, base composition or purity of the oligonucleotide. Snake venom phosphodiesterase generally produces an even distribution of peaks which correspond to every possible size fragment, including the intact (undigested) species and excluding only the three or four smallest 5' residues. The smallest three- or four-base fragments are often not detected because they may fall in an area of chemical noise and because the enzyme appears to digest the last few residues at a higher speed than the rest of the oligonucleotide, leaving little product to be analysed.

The base residue at each position could be unambiguously assigned by correlating the measured mass difference between adjacent peaks with the known masses of the four base residues [3]. Since the smallest mass difference between

![Figure 4](attachment:image.png)

Comparison of spectra from crude oligonucleotide taken in manual and fully automated modes

A 66-mer oligonucleotide, containing failure sequences, was analysed between 1500 and 10,000 Da by DE MALDI-TOF MS, finding the 'sweet spot' in either fully manual (a) or fully automated (b) mode.
the naturally occurring DNA bases is 9 Da, a 4 Da mass difference accuracy is sufficient to determine each base from ladder data.

The 5' terminal base sequence, which could not be fully determined solely from the phosphodiesterase digestion, could be obtained by conducting a series of partial digestions with three different levels of bovine spleen phosphodiesterase in an acidic ammonium citrate buffer. As shown in Figure 3, the peak intensities of the partial digests were not quite as uniform as obtained with the snake venom enzyme. However, nearly all the sequence, starting from the 5' end, could be determined from the three digests alone and the entire sequence could be reconstructed by combining the data from the two sets of enzyme digests.

**Automation**

Recent advances in both hardware and software now permit the full automation of both the MALDI process and sequence assignment. The process of locating a ‘sweet spot’ in the MALDI matrix producing optimized signal-to-noise can be assigned to computer control and sequence ladders produced in the automated mode (Figures 4b) that are of the same quality as those obtained in a fully manual mode (Figure 4a). The length of time required for this process varies with the complexity of the sample, but is typically of the order of 30 s per spectrum. Once these spectra are acquired and semi-manually edited, the computer algorithm can compare the multiple spectra, order their alignment and compute confidence limits for the assignment of the sequence (Figure 5). Typically, only a single spectrum at each of several different enzyme concentrations is required to unambiguously assign an entire sequence to oligonucleotides up to about 50 bases in length. For assignment of sequence to longer oligonucleotides or assignment of sequence to RNA (in which U and C differ by only 1 Da), unambiguous sequence assignment typically requires statistical treatment of several spectra.

**Summary**

The major advantage of this procedure is that, as detailed in Figure 5, the entire sequence of an oligonucleotide is determined in less than 1 h using inexpensive reagents. DE MS, because of its significantly higher resolving power, permits analysis of oligonucleotides as long as 50 bases, about twice as long as that which is possible without DE. Phosphorothioates can be sequenced after oxidation to the phosphodiester state. We have also found that the enzymes will digest through a number of modifying and protecting groups on DNA and RNA, including 2'-amino- and 2'-alkyl-substituted hydroxy groups. The resultant spectra can confirm not only the presence, but also the position, of modifying groups on both DNA and RNA molecules.


Received 18 March 1996