JOURNAL OF MASS SPECTROMETRY, VOL. 31, 1203–1215 (1996)

SPECIAL FEATURE: TUTORIAL

DNA Sequencing by Mass Spectrometry

Kermit K. Murray
Department of Chemistry, Emory University, Atlanta, Georgia 30322, USA

Mass spectrometry has the potential to replace gel electrophoresis for fast DNA sequencing. In this tutorial paper, it is discussed how far mass spectrometry of DNA has come since the advent of electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) and how much remains to be done before mass spectrometry will be a useful tool for DNA sequencing. A brief description of MALDI and ESI mass spectrometric analysis of DNA is presented along with the different strategies for DNA sequencing using mass spectrometry. These include mass spectrometric analysis to replace gel separation in Sanger sequencing, enzymatic ladder sequencing and sequencing by gas-phase fragmentation. The future role of mass spectrometry in DNA sequencing in the Human Genome Project and beyond is also discussed.

KEYWORDS: DNA sequencing; matrix-assisted laser desorption/ionization; electrospray ionization

INTRODUCTION

The Human Genome Project (HGP) was implemented in the late 1980s in an effort to coordinate the huge task of determining the 3 billion base sequence of the human genome and mapping the location of some 100 000 genes.[1] The HGP has the potential to revolutionize the practice of medicine. The human genome map will aid in the understanding and treatment of genetic diseases and will greatly advance both basic and applied research in biology, biochemistry and genetics. In addition to the large-scale mapping and sequencing effort, the broad aims of the HGP are the development of the technology necessary to map and sequence the human genome and that of model organisms, to develop methods to deal with the large quantity of information that will be generated and to investigate the ethical, social and legal implications of the knowledge that will be gained.[2] Mass spectrometry for DNA sequencing falls under the heading of technology development.

Two important concerns from the outset of the HGP have been time and money.[1,2] The time required to sequence the human genome at 1992 sequencing rates was estimated to be 30 000 work-years. The cost, often expressed as the cost of sequencing per base pair, was estimated at \$1–2 per base, i.e. over 3 billion dollars in total. One of the early and continuing project goals was the support of technological advances that might increase the speed and reduce the cost of sequencing. Mass spectrometry has been one of the most promising techniques from the outset and HGP funding has led to many advances in mass spectrometry of DNA during the past few years. Some of the more important of these advances will be discussed in this tutorial paper.

To the delight of molecular biologists (and somewhat to the chagrin of mass spectrometrists), large-scale DNA sequencing is proceeding largely through improvements in conventional technologies and not through new technologies such as mass spectrometry. In 1995, the first complete genome sequence for a free-living organism was obtained when the 1.8 million base pair sequence of *Hemophilus influenzae* was determined.[3] The task required eight scientists operating 14 DNA sequencers for a total of 3 months and the estimated cost for labor and reagents was \$0.48 per base pair. A few months later, the complete sequence of the 580 070 base pair genome of *Mycoplasma genitalium* was reported,[4] and it is expected that more will follow.[5] Advances also continue to be made on the human genome. One recent example is the determination of the complete 685 kilobase (kb) sequence of the human β T cell receptor locus.[6] The success achieved in large-scale sequencing has prompted some researchers to call for a move from DNA mapping to higher resolution sequencing.[7] Currently, the human genome map is complete to a resolution of 300 kb or better. At the current rate of large-scale sequencing, the entire 3 billion base pair sequence of the human genome will be sequenced early in the new millenium.[8]

It is likely that the human genome will be sequenced before mass spectrometry-based sequencing is widely used. However, mass spectrometry-based sequencing will probably be a useful tool for molecular biologists and geneticists of the 21st century. Whereas the genome of a single human can be described by the sequence of 3 billion bases, the database of genetic information for the human species is described by a 10^{16} base sequence catalog that describes the 0.1% of our DNA that differs from person to person.[2] Such a catalog may be necessary to locate rare or complex disease genes and to understand fully the genetic basis of human biology.

This tutorial begins with a brief introduction to the structure of DNA from a mass spectrometrist's point of view. Following this, a chronicle of some recent advances in the mass spectrometry of DNA is provided.

CCC 1076–5174/96/111203–13

Received 3 August 1996
Accepted 13 September 1996

Like other recent progress in mass spectrometry of large bio- and synthetic polymers, advances in the mass spectrometry of oligonucleotides and DNA have been particularly rapid since the introduction of the techniques of electrospray ionization (ESI)[9] and matrix-assisted laser desorption/ionization (MALDI).[10] Several possible approaches to mass spectrometry-based DNA sequencing are described and the relation between what is now possible and what is desired for DNA sequencing is discussed. Lastly, the outlook for mass spectrometry as a sequencing tool is presented.

With thousands of researchers spread out all over the world, the task of managing the information obtained in the HGP is difficult. A fortunate occurrence, although not directly related to the biosciences, is the growth of the Internet and, in particular, the recent growth of the World Wide Web (WWW).[11] Most of the human genome research centers worldwide have WWW pages and many of these have databases of map and sequence information. Many database sites use a WWW browser 'front end' for data management. Anyone in the world with Internet access can obtain the latest human genome map and sequence data. Registered users can enter or change data, or add annotations to augment information in the database. A list of WWW addresses (uniform resource locators, also known as URLs) for some of the HGP centers is given in Table 1. Rather than giving an exhaustive list of sites, it is hoped that the sites in Table 1 will provide a starting point for finding genome resources on the WWW. At the rate that the human genome is being sequenced, the latest data will most likely be found in electronic databases.

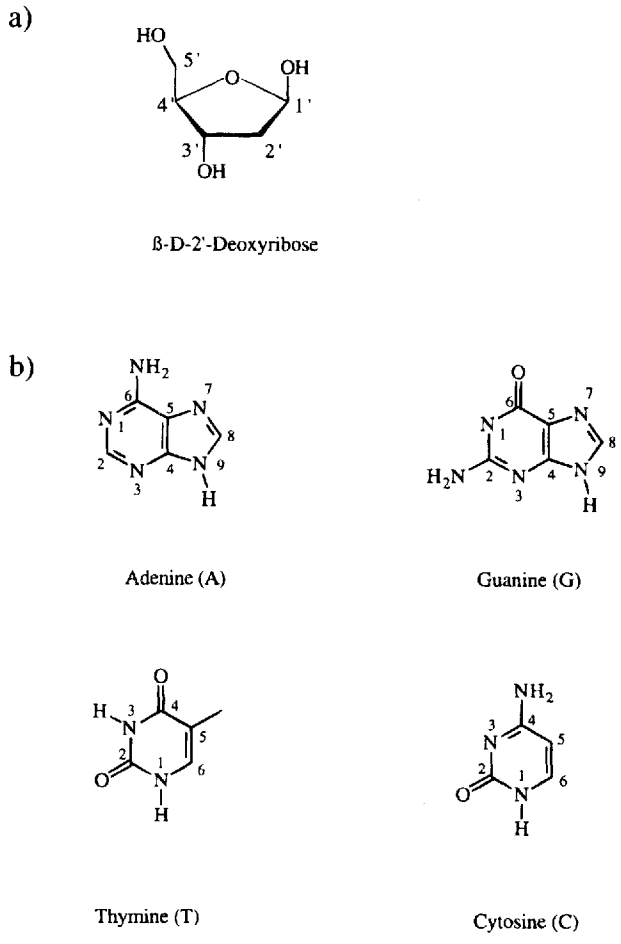

Figure 1. (a) Structure of the sugar β-D-2'-deoxyribose, one of the building blocks of DNA. (b) Structures of the DNA bases adenine (A), guanine (G), thymine (T) and cytosine (C).

DNA STRUCTURE

DNA is a biopolymer consisting of deoxyribonucleotides each containing a sugar, phosphate group and one of four bases.[12] The sugar deoxyribose, depicted in Fig. 1(a), is a derivative of ribose that dehydroxylated at the 2'-position. Of the four bases, depicted in Fig. 1(b), two are purine derivatives, adenine and guanine and two are pyrimidine derivatives, thymine and cytosine. The bases are abbreviated with the letters A, G, C and T for adenine, guanine, thymine and cytosine, respectively. The deoxyribonucleosides are formed by connecting the N-9 of the purine bases or N-1 of the pyrimidine bases to the C-1 position of the deoxyribose. The nucleosides are linked to one another by phosphate groups attached to the 3'-hydroxyl group of one nucleoside and the 5'-hydroxyl group of the adjacent nucleoside. This configuration for single-stranded DNA is shown in Fig. 2(a). A shorthand notation has been developed for depicting the structure of a DNA chain. The sugar is replaced with a vertical line, the base with its letter abbreviation and the phosophodiester linkage with the letter P within a circle. An example of this shorthand notation is shown in Fig. 2(b). Note that the base sequence is written with the 5'-end on the left and the 3'-end on the right. An even more compact notation

Table 1. Human Genome Project Internet resources

Center for Medical Genetics	http://allserv.rug.ac.be/~sdebie/
The Cooperative Human Linkage Center	http://www.chlc.org/
DOE Human Genome Program	http://www.er.doe.gov/production/oher/hug_top.html
Genethon	http://www.genethon.fr/genethon_en.html
Genome Database (GDB)	http://gdbwww.gdb.org/
Human Genome Organization (HUGO)	http://hugo.gdb.org/
Human Genome Project Mapping Resource Centre (UK MRC)	http://www.hgmp.mrc.ac.uk/
The Institute for Genomic Research (TIGR)	http://www.tigr.org/
Los Alamos National Laboratory Center for Human Genome Studies	http://www-ls.lanl.gov/
NIH National Center for Human Genome Research (NCHGR)	http://www.nchgr.nih.gov/
National Center for Genome Resources	http://www.ncgr.org:80/ncgr/ncgr.html
The Sanger Centre	http://www.sanger.ac.uk/

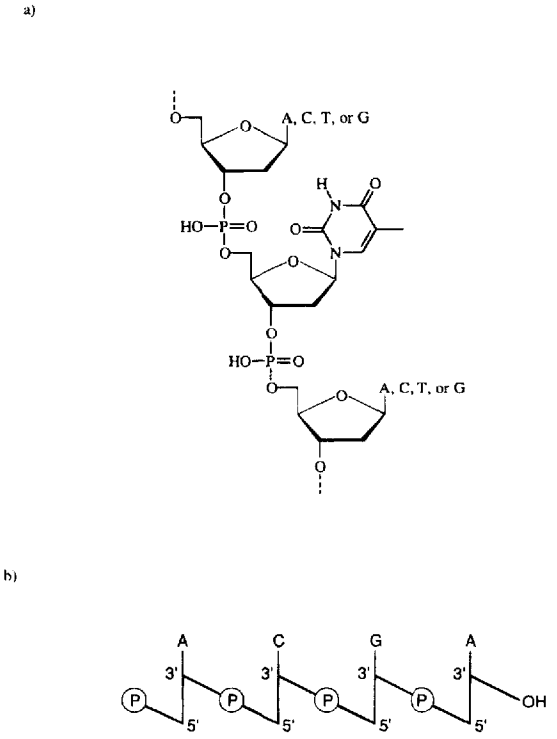

Figure 2. (a) Structure of part of a DNA chain consisting of individual bases distributed on a sugar-phosphate backbone. (b) Shorthand notation for a single-stranded DNA chain.

for the oligonucleotide in Fig. 2(b) is either 5′-ACGA-3′ or d(ACGA).

Double helical DNA is formed when two strands running in opposite directions coil about a common axis. The DNA helix is only 2 nm in diameter, but the fully extended length of even a relatively small chromosomal DNA strand can be several millimeters. The chains are held together by hydrogen bonds which connect the bases on the inside of the coil; the phosphate and deoxyribose units are on the outside. Owing to steric factors, hydrogen bonding occurs only between A–T or G–C base pairs, and thus only complementary DNA strands will bind together. The joining of complementary DNA strands in this manner is called hybridization. It is the specificity of DNA hybridization that allows for DNA replication.

In the past several decades, a variety of tools for the extracellular manipulation of DNA have been developed. These form the basis of what is known as recombinant DNA technology. A variety of restriction enzymes can be used to cleave DNA molecules at sites of specific base sequences and DNA ligase enzymes can be used to join different DNA strands. Oligonucleotides tagged with fluorescent or radioactive molecules can be used to hybridize with and thus tag the complementary site on a DNA molecule. Cloning and polymerase chain reaction (PCR) can be used to amplify a specific DNA strand.[12] In cloning, the DNA of a host cell is cut and the target DNA is inserted. As the cells replicate, the target DNA is also replicated. In the PCR method, many cycles of strand separation and complementary strand synthesis are used to create copies of a particular DNA molecule.

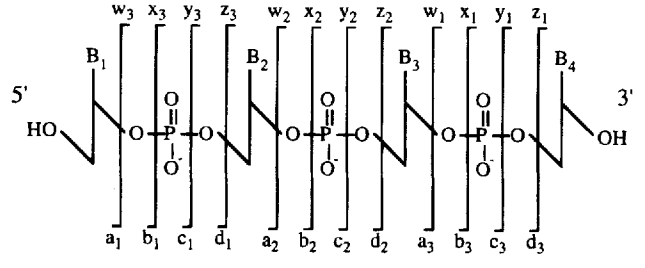

Figure 3. Notation for DNA fragmentation as presented in Ref. 13. The letters a, b, c and d indicate an ion containing the 5′-OH group and the letters w, x, y and z indicate an ion containing the 3′-OH group. The numeral subscript indicates the point of fragmentation in number of bases from the terminal group.

A notation has been developed for designating fragments observed in mass spectrometric analysis of DNA.[13] This nomenclature is similar to that used for peptides and proteins[14] and is depicted in Fig. 3. The four possible phosphodiester bond cleavages are labelled a, b, c and d for fragments containing the 5′-OH group and w, x, y and z for fragments containing the 3′-OH group. The subscript indicates the position of the cleavage, counting the number of bases from the appropriate terminal group. Note that the d- and w-type DNA fragments are *not* analogous to d- and w-type side-chain fragments observed with peptides and proteins.[15] A base is indicated as B_n where the subscript n is the base position with respect to the 5′-end of the molecule. When a fragment ion loses a specific base this is indicated parenthetically, e.g. $a_3 - B_2(A)$ indicates an a-type fragment at the 3-position with additional loss of an adenine base at the 2-position.

GENOME MAPPING AND SEQUENCING

The human genome is the set of genetic instructions that describes a single human being. It consists of 24 distinct chromosomes containing an estimated 10^5 genes that are described by the sequence of 3 billion DNA bases. A genome map consists of the locations of well defined sites in the genome such as the genes themselves or points of restriction enzyme cleavage. The human genome map is being constructed at increasingly fine levels of detail which are shown in Fig. 4. At the highest level are the chromosomes, which can be easily distinguished by light microscopy. Each human chromosome contains approximately 10^8 bases. The proximity of two genes to each other within a chromosome can be inferred by how often two inherited properties are transferred to progeny. Such a determination results in what is called a genetic map, which has a resolution of between 2×10^6 and 10^7 bases. A physical genome map is determined by direct observation rather than inferred by inheritance. For example, a cytogenic map is based on the distinctive banding patterns of stained chromosomes and has a resolution of 10^5–10^7 bases. High-resolution physical mapping is done by randomly cutting the DNA into successively smaller overlapping fragments. Overlapping regions of the different fragments can be identified by hybridizing to tagged complementary oligomers and from this the fragment order is determined. The ordered fragments

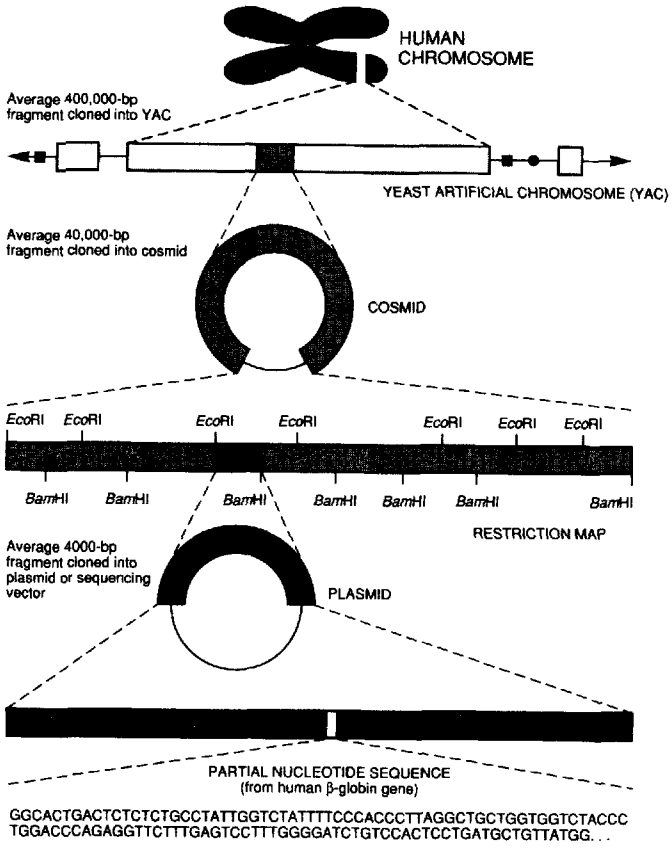

Figure 4. The most common method for DNA mapping and sequencing involves the creation of libraries of successively smaller cloned fragments. Sequencing proceeds with DNA fragments that are a few hundred bases in length. (From Ref. 2).

are then cut again into smaller pieces which undergo the ordering process on a smaller scale. At each stage of division, the fragments are amplified by cloning. Only after the DNA fragments are several hundred bases in length can sequencing of the individual base pairs begin.

The current approach to DNA sequencing involves three stages of DNA fragmentation, starting with the chromosome and finishing with sequencing 400–500 base pair segments.[1,2,16] At each step of fragmentation a new clone library is constructed and the positions of the fragments within the library are determined. Different clones are used at each step, each suited to the particular length of DNA to be propagated. In the first fragmentation step, a low-resolution physical map is constructed by first cutting the DNA into approximately one million base pair segments using restriction enzymes or mechanical shearing. The fragments are inserted into yeast artificial chromosome (YAC) clones for propagation. The set of cloned fragments constitutes a cloned library that can be used repeatedly over long periods and by multiple researchers. The fragmentation is random, thus fragments formed from different DNA molecules may overlap. Shared stretches of the fragments are found by identifying landmarks such as specific hybridization tagged sites or points of restriction enzyme cleavage.

The next fragmentation step leads to high-resolution mapping and involves randomly cutting the DNA that is in the YAC clone library into segments of approximately 40 000 base pairs. The particular clones best suited for this length of DNA are called cosmids, a type of bacterial virus (bacteriophage) that can be used to 'infect' bacteria with the desired DNA strand. A map of the cosmid library is developed by locating landmark overlaps in a manner similar to mapping the YAC library. In the final step, a representative overlapping subset of the cosmid library is selected, randomly cut and propagated in M13 bacteriophage clone libraries. The DNA strands in a M13 phage library are several hundred to several thousand base pairs long. Plasmids can be used in place of M13 phage in some cases to create a library containing DNA strands 5000–10 000 base pairs long. Fragments from the plasmid library of 400–500 base pairs can be sequenced, gel electrophoresis being used to order the fragments by size. The sequence information from the M13 phage library is assembled by computer to form the complete sequence of the cosmid library, the YAC library and finally the chromosome itself.

The Sanger method of dideoxy chain termination is the most common method for DNA sequencing (Fig. 5).[17] In the Sanger method, DNA polymerase is used with deoxyribonucleoside triphosphates to generate DNA strands complementary to the strand to be sequenced. Strands of different lengths are formed when

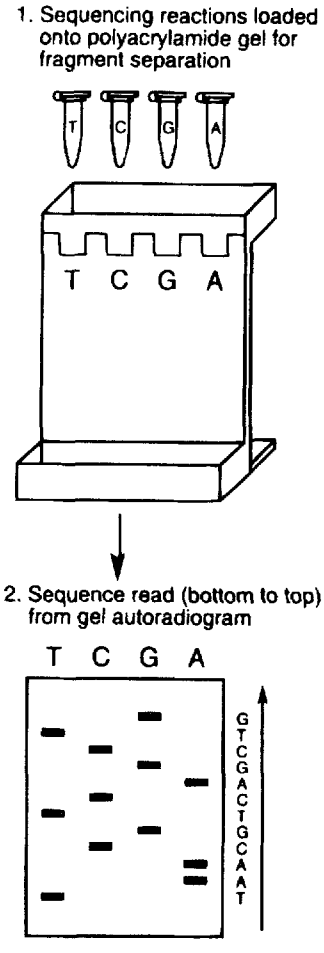

Figure 5. Schematic diagram of Sanger dideoxy chain termination sequencing. Four sets of sequencing products consisting of DNA strands terminated at one of the four bases are separated by gel electrophoresis. The sequence can be read directly from the gel. (From Ref. 2).

a chain termination reagent, the 2',3'-dideoxy analog of one of the four nucleoside triphosphates, is used in the synthesis of the complementary strands. The dideoxy nucleoside terminates the replication of the DNA at the position of the complementary base because there is no 3'-hydroxyl group to form the next phosphodiester bond. With the proper quantity of dideoxy chain terminator, a complete set of fragments can be created with termination at a particular base. Four sets of chain termination fragments are created in all, one for each base. The lengths of the strands are determined for each reaction set and the sequence is read directly from this result. The method of choice for chain length determination at this time is gel electrophoresis. The strands are tagged with either radioactive isotopes or fluorescent dyes.

The time required to perform the gel separation is a significant limitation in the performance of DNA sequencing. Therefore, a major focus of the HGP is to develop methods for sequence readout that are faster than conventional gel electrophoresis. Multiplex capillary gel electrophoresis[18] and miniature capillary electrophoresis chips have been developed[19] which have the potential to reduce the time for a gel separation from hours to minutes. A drawback to these approaches is that the capillaries cannot be reused. Flow cytometry has been used for sizing larger fragments[20] and methods based on microscopic imaging have also been developed.[21,22] Photolithographic techniques similar to those used to fabricate computer chips have been used to create DNA sequencing chips for sequencing by hybridization (SBH).[23] The SBH chip consists of an array of all possible base combinations of a small oligonucleotide bound to a solid support. The tagged target DNA binds to the chip only in regions containing the complementary sequence. Mass spectrometry has been an obvious choice for rapid DNA sequencing from the outset.[1,2] In the most straightforward application, mass spectrometric analysis can be used to replace the gel separation step in Sanger dideoxy chain termination sequencing. Methods based on enzyme ladder sequencing[24] and gas-phase fragmentation[25] have also been proposed. The different approaches to mass spectrometry-based sequencing are discussed in a later section. In the next section, a detailed discussion of recent developments in the mass spectrometry of DNA is presented.

MASS SPECTROMETRY OF DNA

For mass spectrometry to be a useful method for sequencing DNA, the ionization method must be capable of generating ions of DNA containing several hundred bases with mass resolution and mass measurement accuracy adequate to identify peaks corresponding to a difference of one base or better. Before the advent of ESI[9] and MALDI,[10] this was far from being an achievable goal. Mass spectrometry of oligonucleotides was limited to those with fewer than ten bases and was performed with the techniques of fast atom bombardment (FAB) and plasma desorption mass spectrometry (PDMS).[26] It has become possible to envision mass spectrometry-based DNA sequencing only within the past decade as techniques have been developed for the ionization of large biomolecules. Several recent reviews will be of interest for those desiring a more exhaustive treatment of the literature than is given in this section. The recent *Analytical Chemistry* Fundamental Reviews issue contains a brief but informative review of the current status of mass spectrometry of oligonucleotides and nucleic acids.[27] The review of Limbach *et al.*[28] in *Current Opinions in Biotechnology* contains an excellent bibliography, and Fitzgerald and Smith[29] provide a review of MALDI analysis of nucleic acids, including a discussion of fragmentation mechanisms. Other noteworthy reviews cover early work with ESI and MALDI[30] and mass spectrometry of PCR-amplified DNA.[31]

MALDI of oligonucleotides

The first application of MALDI to oligonucleotides was reported in 1990 for several 4–6-base oligomers.[32] It was soon discovered that oligonucleotides were not as easy to analyze with MALDI as were peptides and proteins. By the early part of this decade, it was possible to perform MALDI analysis of proteins with molecular masses over 10^5 Da, yet the limit for oligonucleotides was a few thousand.[10] Over the past few years, the development of new matrix materials, sample preparation methods and a better understanding of the fundamentals of the ionization process have led to steady progress in the mass range accessible for MALDI analysis of oliogonucleotides and DNA.

An early example of a unique and promising sample preparation method is the frozen aqueous matrix method developed by Williams and co-workers.[33] In 1989, the first ice matrix work was reported which indicated that MALDI mass spectra of DNA might be obtained under the proper conditions.[34] It was shown that DNA containing more than 600 bases could be desorbed intact from a frozen aqueous solution on a cooled copper substrate using a visible laser at 581 nm.[34] Although initially detection was done by autoradiography, the DNA was transferred to the gas phase intact and it appeared that at least some of the DNA was ejected from the solid as a free molecule rather than embedded in large ice particles. Subsequent mass spectral studies resulted in excellent mass spectra of oligonucleotides containing up to 60 bases and over 18 000 Da molecular mass.[35,36] A mass spectrum of a mixture of single-stranded oligonucleotides of six different lengths is shown in Fig. 6. It was found that ionization efficiency was enhanced at both 578 and 589 nm, possibly due to resonant absorption by atomic copper at 578 nm or sodium at 589 nm or by a two-photon resonance at 578 nm.[36] It was also found that the best results were obtained from copper surfaces that were deliberately corroded. Unfortunately, these excellent spectra were impossible to generate reproducibly, hampering development of this method.[33]

The next major development in MALDI of DNA was the discovery that exchanging alkali metal cations with ammonium ions significantly decreases adduct ion formation. DNA and oligonucleotide ions formed by

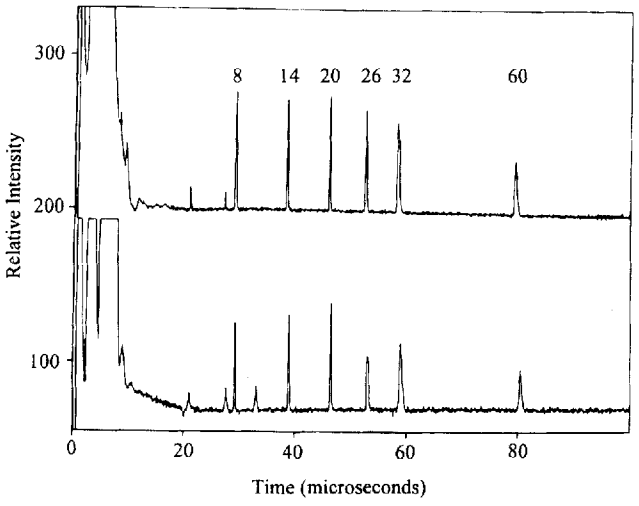

Figure 6. MALDI mass spectra of a mixture of single-stranded oligonucleotides (8, 14, 20, 26, 32 and 60 bases) using a frozen aqueous matrix. The upper trace was obtained from nine laser shots from a permanganate-corroded copper surface and the lower trace from six shots from an EDTA-corroded copper surface. (From Ref. 36).

MALDI have a tendency to form adducts with cation impurities in the sample, primarily sodium and potassium.[37] The positively charged alkali metal ions bind strongly to the oxygen atoms in the phosphodiester groups of the DNA molecule. It was found that the signal from adducts could be reduced by exchanging the metal ions with ammonium ions either by addition of ammonium salts to the sample solution prior to deposition[24] or with a cation-exchange column.[38] It is postulated that cation adduct suppression arises as a result of dissociation of ammonium phosphate ion pairs following desorption. Unlike an atomic cation, the ammonium ion can transfer a proton to the phosphate group to form the free acid and an ammonia molecule:

$$NH_4^+ \cdot {}^-O{-}PO{-}(OR)_2{}^{\pm n} \rightarrow NH_3 + HO{-}PO(OR)_2{}^{\pm n}$$

Here the charge is shown as $\pm n$ to indicate that this mechanism may be in effect for positive or negative ions or for cases of multiple charging as in ESI.

A further development in MALDI of DNA was the introduction of new matrix compounds. It is well known that the MALDI signal for nearly every analyte is highly sensitive both to the matrix compound and to the sample preparation method.[10] Therefore, it was recognized at an early stage that the solution to MALDI analysis of DNA could result from finding the appropriate matrix material. One of the most useful of the matrix compounds used with DNA is 3-hydroxypicolinic acid (HPA), which was first reported by Wu *et al.*[39] 3-HPA is 3-hydroxypyridine-2-carboxylic acid, which is similar in structure to one of the original MALDI matrices, nicotinic acid (3-pyridinecarboxylic acid).[40] It was discovered that out of 45 compounds tested, HPA was a superior matrix for DNA ionization. Mixed-base single-stranded oligonucleotides between 10 and 67 bases in length could be ionized and detected routinely using either 355 or 266 nm radiation for ionization. The reported mass resolution was only 30, but fragmentation was greatly reduced. Spectra were reported in both the positive and

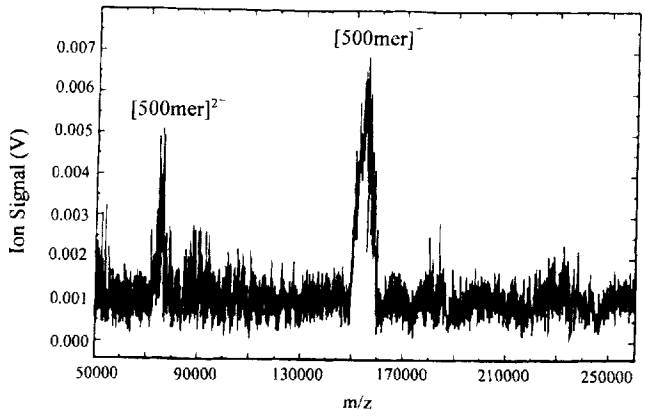

Figure 7. Negative ion MALDI mass spectrum of a 500 base pair double-stranded DNA molecule obtained using a mixture of picolinic acid and hydroxypicolinic acid matrix. Only single-stranded DNA is detected. (From Ref. 41).

negative ion modes, but the negative ion mode gave both better signal and mass resolution.

The largest DNA species yet detected by MALDI mass spectrometry is a 500-base double stranded DNA molecule amplified with PCR from the bacteriophage lambda (λ phage) genome (Fig. 7).[41] Chen and coworkers[41] recorded ions in this mass range using a mixture of picolinic and 3-hydroxypicolinic acid matrices, 266 nm laser desorption and an acceleration voltage of 45 keV. It is possible that the effects of DNA fragmentation are minimized with this large acceleration voltage, which reduces the residence time in the ion acceleration region. For double-stranded DNA, only ions corresponding to the mass of single-stranded DNA are observed, indicating that the DNA is denatured in the sample preparation or fragmented into single strands on desorption and ionization.[42] Although ionizing and detecting a 500-base DNA molecule is an excellent accomplishment, there are still limitations that must be overcome if MALDI is to be a useful tool for mass analysis of Sanger sequencing products. The mass resolution of these large ions is 20–30, more than an order of magnitude worse than necessary for the separation of Sanger sequencing products. Also, the signal-to-noise ratio for the large DNA ions is less than ten.

In the early MALDI studies of oligonucleotides, it was discovered that the mass spectral response was a strong function of base composition.[43–46] Specifically, oligonucleotide homopolymers containing the base thymine were found to yield large MALDI signals, whereas homopolymers containing the other three bases were more difficult to detect. It was also found that RNA gave a larger MALDI signal than DNA and that the length of RNA strands that could be detected also was larger.[38,45,47] Experiments using acridine orange dye to stain the DNA showed that the level of incorporation of oligonucleotide homopolymers dG_{18} and dT_{18} in sinnapic acid matrix was similar, tending to rule out different levels of incorporation of DNA into the matrix as the source of the problem.[46] This evidence pointed to fragmentation as the cause of low MALDI signals for DNA.

A schematic representation of the DNA fragmentation mechanism is shown in Fig. 8. The first step of DNA fragmentation is believed to be protonation of the

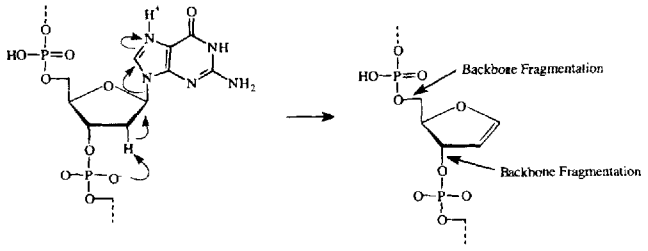

Figure 8. Fragmentation of DNA in the gas phase is believed to occur following base loss. One possible mechanism of base loss is 1,2-*trans* elimination as indicated. Subsequent backbone fragmentation can occur at either the 5′- or 3′-side of the sugar.

base and subsequent *N*-glycosidic bond cleavage.[48–50] Following base loss, phosphodiester backbone cleavage occurs at the 3′-C—O bond. Protonation of the base is necessary for the initial step of base loss, thus the proton affinity of the base is a contributing factor to stability of the phosphodiester backbone. The gas phase proton affinity of deoxythymidine is 224.9 kcal mol^{-1} (1 kcal = 4.184 kJ), between 8 and 10 kcal mol^{-1} lower than that of the other monodeoxynucleotides (dG 234.4, dA 233.6, dC 233.2 kcal mol^{-1}).[51,52] The thymidine base is more difficult to protonate and will, therefore, be more stable with respect to base loss and backbone fragmentation. The stability of RNA with respect to DNA can be explained by the stabilizing effect of the 2′-OH group of the ribose ring. The 2′-OH group prevents 1,2-*trans* elimination that can occur for deoxyribonucleosides.[38,48]

Chemical modifications have been proposed as a means of increasing oligonucleotide stability in MALDI. Oligonucleotides containing the 7-diaza analogs of the purine nucleotides have been found to be more stable than those containing the unmodified nucleotides.[49,53] Replacement of the 7-position nitrogen with a carbon atom decreases the proton affinity of the adenine and guanine bases. As discussed above and depicted in Fig. 8, base loss is believed to be the first step in backbone fragmentation. In the 7-diaza-substituted oligonucleotides, only the cytidine nucleotide is especially susceptible to base loss and fragmentation. It is possible that the enhanced stability of 7-diaza analogs in MALDI is related to the enhanced stability of these analogs in solution.[54]

Delayed ion extraction is a technique for improved mass resolution in MALDI time-of-flight mass spectrometry (TOF-MS).[55] Mass resolution is enhanced with delayed extraction because there is a correlation between the velocity of the laser desorbed ions and their position, the faster ions moving further away from the surface after desorption. If the acceleration voltage is delayed with respect to the firing of the laser, the slower ions that are closer to the surface will get an extra 'kick' to speed them up. Delayed extraction MALDI/TOF-MS has provided unprecedented mass resolution for peptides and proteins; for example, a mass resolution of greater than 10000 has been reported for bovine insulin.[56] As is now familiar, the performance of MALDI with oligonucleotides is not as good as it is with peptides and proteins, and delayed extraction is no exception. Resolution improvements are seen, however. For example, with a linear TOF mass spectrometer, a mass resolution of 860 was achieved for a 27-base oligonucleotide.[57] More recently, a mass resolution of 1000 was reported for oligonucleotides up to 50 bases when analyzed in the linear mode and a mass resolution of 7500 was reported for a 12-base RNA oligomer using delayed extraction with a 1.3 m reflectron apparatus.[58]

Fourier transform (FT) MS has become known as the method of choice for high-resolution MALDI mass analysis since Wilkins and co-workers[59] first demonstrated mass resolution in excess of 11000 for bovine insulin using a unique sugar co-matrix preparation. FTMS provides exceptional mass resolution in the range below *m/z* 15000, but it must be noted that it is difficult to analyze high-mass ions with FTMS mass spectrometers. Trapping ions with *m/z* >50000 is achieved only with great difficulty and with no improvement in mass resolution over that achieved with TOF in this mass range.[60] Recently, Li *et al.*[61] demonstrated mass resolution of 136000 for a singly charged 25-base oligonucleotide using an external ion source FTMS apparatus with a 7 T superconducting magnet (Fig. 9). An important aspect of this experiment is the cooling of ions with a pulse of argon gas prior to excitation and detection. Collisions between the ions and the buffer gas remove translational and internal energy from the ions, reducing fragmentation and allowing high-resolution mass spectra to be acquired.

Oligonucleotides several hundred bases long can be detected at low resolution with TOF and oligonucleotides several tens of bases long can be detected at high resolution with delayed extraction TOF or FTMS. For mass spectrometry to be useful in any DNA sequencing scheme, better mass resolution at higher masses must be achieved.

Electrospray of oligonucleotides

ESI of oligonucleotides was first demonstrated by Covey *et al.*,[62] who detected oligomers up to 14 bases in the negative ion mode. As in MALDI analysis, progress in ESI analysis of oligonucleotides has lagged that of ESI analysis of proteins. This is mainly due to the two problems of alkali metal ion adduction and fragmentation that plague ESI as they do MALDI.

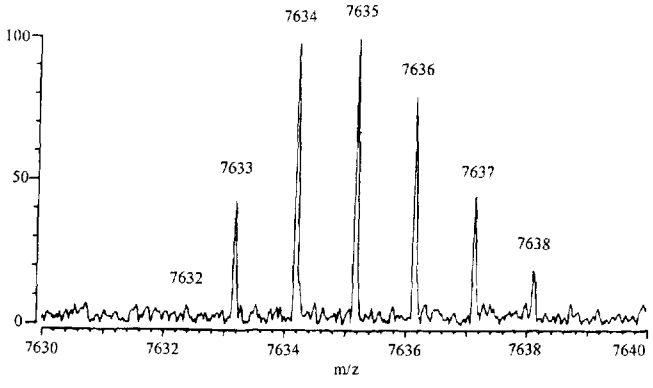

Figure 9. MALDI FTMS of a 25-base oligonucleotide. The mass resolution is 136000 FWHM. Reprinted with permission from *Analytical Chemistry*, **68**, 2090 (1996). ©1996 American Chemical Society.

Stults and Marsters[63] were the first to replace of sodium with ammonium ions in ESI using ammonium salts (and also to suggest their use with MALDI). Oligonucleotide samples were precipitated from an ethanol solution containing ammonium acetate and the ESI sample solutions contained a low millimolar concentration of ammonium acetate or aqueous ammonia. Without addition of the ammonium salt, nearly one sodium counter ion was observed per phosphate group. After treatment with ammonium salts, ions were detected with no sodium adducts for oligomers containing up to 48 bases. An oligomer containing 77 bases was observed with only a single sodium counter ion.

Little *et al.*[64] have used the high resolving power of FTMS to achieve a mass measurement accuracy of <10 ppm for a 50-base oligonucleotide and just over 30 ppm for a 100-mer (Fig. 10). The molecular mass of the oligomer can be used for sequence verification or in conjunction with nozzle–skimmer dissociation for sequencing. However, even the high resolving power and mass accuracy of FTMS are insufficient for sequence determination of unknown oligonucleotides larger than several tens of bases in lengths. For oligomers conta ining more than ~ 15 bases there is at least one possible isobar for every unit mass value and for oligomers larger than 20 bases there are three possible compositions per unit mass.[65]

The high charge states of electrosprayed ions can be a drawback in some cases, even with the mass accuracy and resolving power of FTMS. Space charge effects can limit the number of ions that can be trapped in an ICR cell and a distribution of charges limits the signal level of any one mass spectral peak. Recognizing these limitations, Smith and co-workers[66] have developed methods for ion charge state reduction both by modifications to solution pH and addition of organic bases to the electrospray solution. Organic acids were found useful for charge state reduction, but strong inorganic acids led to suppression of the oligonucleotide ion signal. More recently, a cocktail of imidazole, piperidine and acetic acid in acetonitrile–water (80:20) solution was shown both to suppress alkali adducts and to reduce significantly the charge state of several oligonucleotides.[67]

In other experiments, Smith and co-workers used FT ion cyclotron resonance (ICR) mass spectrometry to trap and detect single DNA ions with masses in excess of 10^8 Da.[68] The ions were created from coliphage T4 DNA using a conventional ESI source and the number of charges was estimated to be in excess of 30 000. The mass to charge ratio was determined from the cyclotron frequency and the charge on the ion was determined by measuring the signal voltage at a known ion radius. The uncertainty in estimating the ion radius is a major contributing factor in the 10% uncertainty of the estimated molecular mass.

An innovative detection method for highly charged megadalton mass DNA was recently reported by Furstenau and Benner.[69] Their mass spectrometer employs a charge detector which is simply an open metal tube connected to a low-noise amplifier. A charged particle entering the tube induces an image charge with an amplitude that is proportional to z, the charge on the ion, similar to the FTMS detection method.[70] An identical signal of opposite sign occurs when the ion exits the tube. The time between the entrance and exit signals can be used to determine the velocity and thus the mass to charge ratio of the ion. With z from the signal amplitude and m/z from the flight time through the tube, the ion mass can be determined. DNA ions between 1×10^6 and 1×10^7 Da were detected at a mass

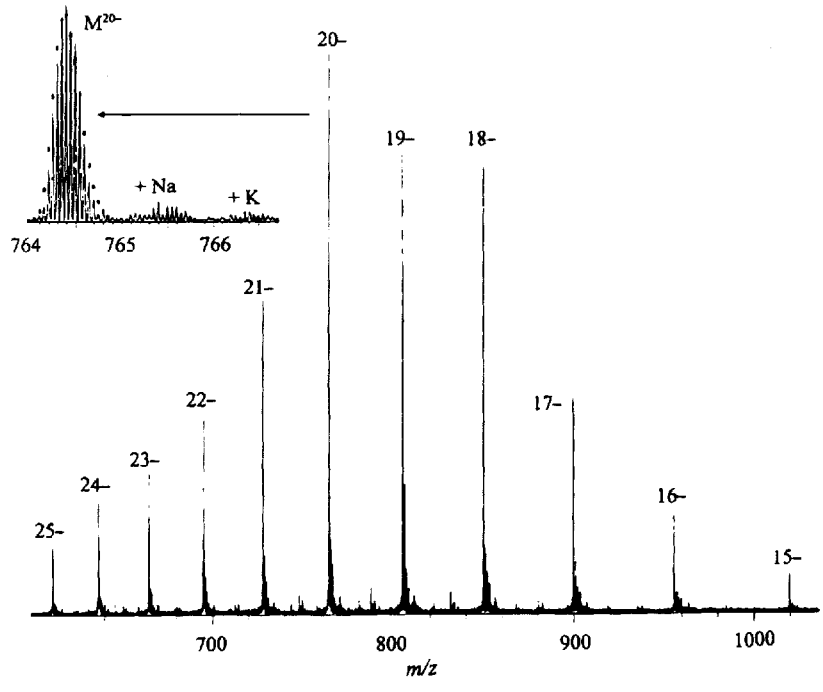

Figure 10. ESI FTMS mass spectrum of a single-stranded DNA 50-mer with a mass measurement accuracy of 7 ppm. The inset is an expanded view of the $[M-20H]^{20-}$ region and the dots indicate the theoretical isotopic distribution. Reprinted with permission from *Proc. Natl. Acad. Sci. USA* **92**, 2318 (1995). ©1995 National Academy of Sciences.

resolution of ten using the charge detection TOF mass spectrometer. Even with no further improvements, charge detection TOF could conceivably compete with pulsed field gel electrophoresis for the analysis of megadalton DNA.

DNA SEQUENCING USING MASS SPECTROMETRY

There are three strategies for DNA sequencing with mass spectrometry: (i) mass analysis of Sanger sequencing reaction products; (ii) ladder sequencing and (iii) gas-phase fragmentation. Mass analysis of Sanger sequencing products is the most straightforward approach and simply involves replacing the gel electrophoretic determination of DNA strand length with the mass spectrometric determination of strand length. Mass spectrometry has potential time and cost advantages over gels. A conventional gel separation may take hours,[71] whereas a mass spectrum can often be obtained in less than 1 min. Massively parallel capillary arrays can greatly reduce the analysis time, but they cannot be reused, thus increasing the cost.[18,19] Mass spectrometric ladder sequencing involves cleaving successive nucleotides from the strand with an enzyme and measuring the resulting change in mass. The mass resolution and mass measurement accuracy requirements are more stringent for ladder sequencing than for Sanger sequencing. For ladder sequencing, a mass difference of 9 Da must be measured to distinguish A from T, whereas for Sanger sequencing the ability to determine a mass difference of ~300 Da is sufficient to define the sequence. In the gas-phase fragmentation method all of the sequencing is done in the mass spectrometer by identifying the fragment ions that result from metastable or collision-induced dissociation (CID). Sequencing by gas-phase fragmentation has the potential of being much faster than methods that rely on solution-phase chemistry, but fragmentation spectra are complex and can be difficult to interpret. Mass resolution and mass measurement accuracy must be as good as or better than that needed for ladder sequencing.

Mass spectrometry and Sanger sequencing

The most straightforward approach to using mass spectrometry for DNA sequencing is to replace the polyacrylamide gel separation of the Sanger sequencing products by mass analysis. This method is particularly attractive when the potential for time savings is considered. The sequence readout from a mass spectrum is similar to the readout from the gel. Instead of one lane for each of the four dideoxy termination reaction solutions, four mass spectra are obtained, one for each of the A, C, T and G bases. Although it is simple to envision, it is less simple to put into operation with current mass spectrometric capabilities. Each sequencing solution contains a mixture of DNA strands several hundred bases in length. The mass spectrometer must be able to generate ions from this mixture at a resolution sufficient to distinguish between two bases in strands up to 500 bases in length. This corresponds to a mass resolution of 500 at a mass range in excess of 100 000 Da. It is obvious from the previous section that mass spectrometry is just starting to approach this goal. Still, some impressive accomplishments have recently been achieved that are highly encouraging and perhaps give a glimpse into the future of DNA sequencing.

An excellent illustration of both the recent progress and remaining barriers for mass spectrometry as a replacement for gel electrophoresis is given in a recent paper by Fitzgerald *et al.*[72] Mock DNA sequencing solutions were analyzed by MALDI mass spectrometry, an example of which is shown in Fig. 11. Four mixtures of oligonucleotides between 17 and 40 bases in length were synthesized to simulate the products of a Sanger sequencing reaction. No impurities such as buffers or Sanger reagents were present, although these could in principle be removed prior to mass analysis in an actual sequence analysis. Calibration was performed using the oligonucleotides as an internal standard. In addition, picomole quantities of each component were used, ~100 times the amount typically used for Sanger sequencing. It is important to note that the authors found that the signal for the high-mass molecules was weaker in the mixture than if the components were analyzed separately. As a result, signals from 36- and 40-base pair oligomers were too weak to be of use. The mass resolution was between 100 and 200, which is more than sufficient for the separation of the mixture components. In a more recent experiment, MALDI was used to analyze the Sanger sequencing products with a 45 nucleotide template and a 12-mer primer.[73] The sequence information could be read to 19 bases past the primer and picomole quantities of the primer and template were used in the reaction. These results indicate that MALDI mass spectrometry must be improved by two orders of magnitude in detection limit and one order of magnitude in mass range to be a legitimate candidate for replacement of gels in the Sanger sequencing protocol.

An important application that is possible using

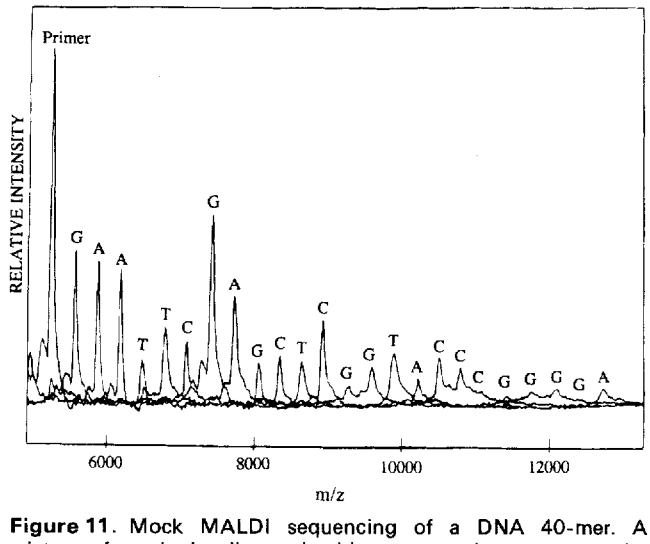

Figure 11. Mock MALDI sequencing of a DNA 40-mer. A mixture of synthetic oligonucleotides was used to represent the products of a Sanger sequencing reaction. The four mass spectra from the sequencing solutions are overlaid (From Ref. 72).

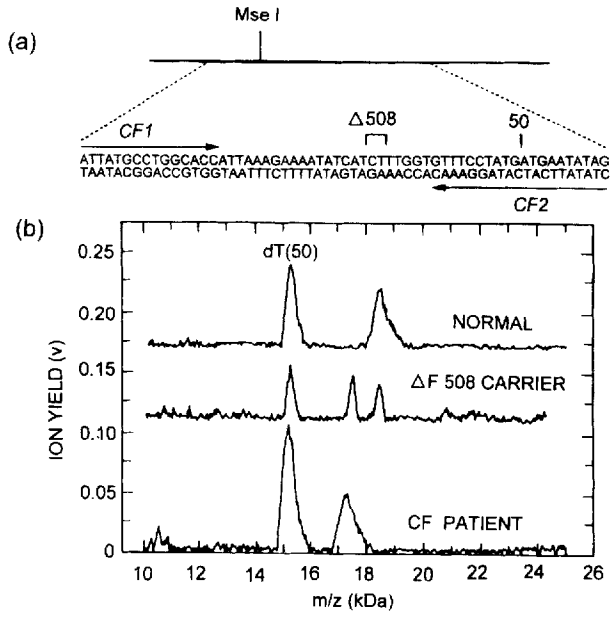

Figure 12. MALDI mass spectra of oligonucleotides obtained from PCR amplification of sections of human DNA from three patients. The top trace was obtained from a normal patient, the center trace was obtained from the DNA of a cystic fibrosis carrier (heterozygous in the cystic fibrosis gene) and the lower trace was obtained from a cystic fibrosis patient (homozygous in the cystic fibrosis gene). (From Ref. 74).

current methods is typified by the detection of the mutant form of the cystic fibrosis gene. Chen and coworkers[74] have devised a method that uses mass spectrometry to identify a DNA mutation that is common to cystic fibrosis carriers and patients. The mutation is a three base pair deletion in one of their chromosomes that results in a loss of a phenylalanine residue in the cystic fibrosis transmembrane conductance regulator protein. A method was devised that produces a 59-base fragment from the normal gene and a 56-base fragment DNA from a gene containing the deletion. In a comparison with the conventional gel electrophoretic method of diagnosis, the mass spectrometric method gave the same results for 30 samples obtained from patients. In Fig. 12, the upper trace is from a normal subject, the middle trace from a cystic fibrosis carrier and the lower trace from a cystic fibrosis patient. The three base pair difference is easily resolved by MALDI-MS.

DNA ladder sequencing

Ladder sequencing oligonucleotides by mass spectrometry is similar to protein ladder sequencing.[75] The oligonucleotide is cleaved using an enzyme that sequentially removes nucleotides from either the 3′- or the 5′-end of the molecule. The sequence is determined by the mass change after each cleavage: 289.2 Da for cytosine loss, 304.2 Da for thymine loss, 313.2 Da for adenine loss and 329.2 Da for guanine loss. Mass measurement accuracy is, therefore, of prime importance to the success of DNA ladder sequencing. Ladder sequencing has been performed with either MALDI or ESI on relatively short oligonucleotide strands: 24-mer for MALDI and 10-mer for ESI.

Pieles *et al.*[24] used the 5′-exonuclease calf spleen phosphodiesterase and the 3′-exonuclease snake venom phosphodiesterase to cleave oligonucleotides from either the 5′- or the 3′-end, respectively. A mixed-base single-stranded 12-mer was used as the test compound and aliquots were taken from the digestion solution every 15 min for analysis. Bentzley *et al.*[76] extended oligonucleotide ladder sequencing with MALDI mass analysis to a 24-base strand using calf spleen phosphodiesterase for enzymatic digestion. A MALDI mass spectrum of a mixed-base 24-mer after 50 min of digestion is shown in Fig. 13. The mass spectral peaks are labelled exo*N*, where *N* is the number of cleavages. As evidenced by the large variation in peak intensity, the rate of digestion is sensitive to the base composition. For example, compare the relatively large exo6 with the relatively small exo10 peak.

Oligonucleotide ladder sequencing has also been demonstrated with ESI. Limbach *et al.*[77] used phosphodiesterase I and II with ESI and a triple quadrupole for mass analysis. Oligonucleotides ten bases in length were digested and mass analyzed after 10 and 20 min and from the data it was possible to obtain partial sequence information. Glover, *et al.*[78] combined enzymatic digestion with off-line liquid chromatography to obtain the full sequence of a single-stranded 10-mer. The long times involved with enzymatic digestion have led many researchers to pursue the avenue of fragmentation in the gas phase. If fragmentation can be induced in the mass spectrometer, the time savings over enzyme digestion-based methods would be considerable.

Gas-phase fragmentation for DNA sequencing

Fragmentation can be regarded as a problem that must be overcome for successful mass spectrometric analysis of DNA, but it can also be used as a means of sequencing itself. With MALDI, fragmentation studies can be

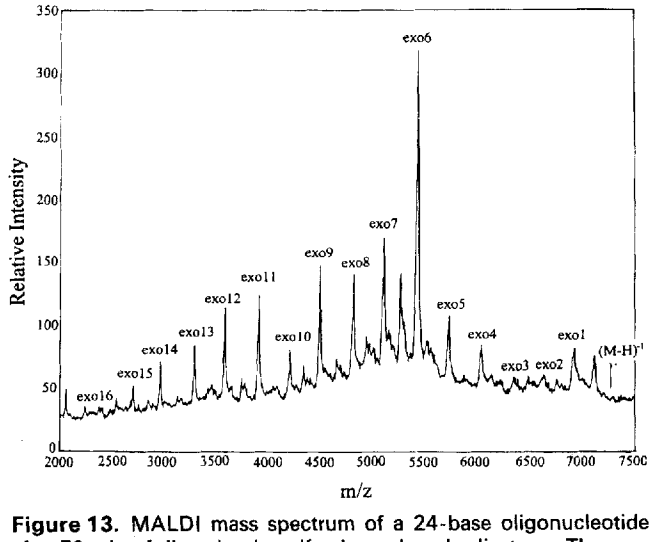

Figure 13. MALDI mass spectrum of a 24-base oligonucleotide after 50 min of digestion in calf spleen phosphodiestrase. The products of the sequencing ladder are labelled exo1 through exo16. Reprinted with permission from *Analytical Chemistry*, **68**, 2141 (1996). ©1996 American Chemical Society.

carried out in a reflectron mass spectrometer using post-source decay.[79] In MALDI/FTMS, ions are trapped for an extended period of time compared with TOF, and more extensive fragmentation can result.[80] Delayed extraction MALDI TOF has also been employed in DNA fragmentation studies.[81] Fragmentation in ESI can be induced during ion formation by changing the nozzle–skimmer potential or by CID.[82] Sequencing by gas-phase fragmentation requires resolution sufficient to separate the often numerous fragments and sufficient mass measurement accuracy to identify the fragments unambiguously. For this reason, DNA fragmentation studies have been carried out either for small oligonucleotides or by high-resolution FTMS.

Much of the initial work on fragmentation of oligonucleotides produced by ESI was done by McLuckey *et al.*[83] Extensive sequence information was obtained for oligonucleotide heximers in an ion trap mass spectrometer using tandem mass spectrometry.[84] Little *et al.*[64] used FTMS to achieve mass accuracy of better than 50 ppm and a mass resolution of 10^5 for oligonucleotides up to 25 bases in length. Sequence information was obtained from fragments created by collisions in the atmospheric pressure region between the electrospray needle and the skimmer orifice (nozzle–skimmer dissociation) or by collisional activation in the ICR cell. Unambiguous sequence information was demonstrated for 14-mer oligonucleotides and sequence verification information was obtained for 25 base oligonucleotides. Ni *et al.*[85] used a triple-quadrupole mass spectrometer to generate CID mass spectra of oligonucleotides up to 15 bases.

Fragmentation of oligonucleotides in MALDI mass analysis has also been used to obtain sequence information. Nordhoff *et al.*[48] used the fragment ions observed in UV and IR MALDI to obtain sequence information for oligonucleotides up to 21-mer in size. IR MALDI was performed at 2.94 μm with an Er:YAG laser and was found to generate more fragments than UV MALDI and was thus more useful for sequencing by gas-phase fragmentation.

The advent of delayed extraction MALDI has led to a new approach for using MALDI fragmentation for sequencing small oligonucleotides. Brown and Lennon[86] demonstrated that delayed extraction is useful not only for improving mass resolution in MALDI/TOF, but also for generating sequence specific fragment ions from proteins. Fragment ions are formed in the ion source a few hundred nanoseconds after laser irradiation, but before ion extraction. Juhasz *et al.*[81] have extended this technique to oligonucleotides. Complete sequence information for an 11-mer was recorded in a single mass spectrum. The abundance of unassigned fragment ion peaks could make the extension of this method to unknowns difficult, however.

OUTLOOK

There is reason for cautious optimism when assessing the outlook for mass spectrometry-based DNA sequencing. While it is likely that the HGP will be completed with minimal contribution from mass spectrometry, there has been significant progress in DNA mass spectrometry in the past several years and there is every reason to believe that in the next few years further significant developments will be forthcoming. Mass spectrometric analysis of 50–100 mixed-base single-stranded DNA is now possible and analysis of single- and double-stranded DNA in excess of 100 bases is possible but difficult. The two techniques of ESI and MALDI continue to evolve together and neither is a clear leader for DNA analysis. ESI leads in both mass resolution and mass measurement accuracy, particularly when combined with FTMS.[64] Owing to the distribution of charges on electrosprayed ions, the ESI method is not well suited to analysis of the multi-component mixtures that are formed in Sanger sequencing reactions. MALDI can be used to ionize DNA of a size that is useful for sequencing, but the resolution at high mass and detection limit are below what are required to compete with gel separations.[29]

Several areas are undergoing rapid development and may soon give rise to new and interesting results. Time-of-flight mass spectrometry of megadalton DNA is one of the more innovative developments in mass spectrometry.[69] If the mass resolution can be improved by an order of magnitude or more, this method could challenge pulsed field gel separation for the analysis of large DNA strands. The commercial availability of optical parametric oscillators as mid-infrared pulsed light sources has led to a revisiting of the frozen aqueous matrix technique.[87] It is possible that infrared absorption of the OH stretch of ice in the 3 μm region could be used as the chromophore for ice matrix MALDI, thereby avoiding the reproducibility problems incurred with visible laser excitation on corroded surfaces. Delayed extraction has rapidly redefined the meaning of high resolution in MALDI/TOF-MS.[55,56] Further innovations in this area may yet provide the necessary mass resolution for DNA molecules several hundred bases in length.

REFERENCES

1. National Research Council, *Mapping and Sequencing the Human Genome*. National Academy Press, Washington, DC (1988).
2. US Department of Energy, Office of Energy Research, Office of Health and Environmental Research, *Human Genome 1991–92 Program Report*, DOE/ER-0544P. US Government Printing Office, Washington, DC (1992).
3. R. D. Fleischmann, M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. Fields, J. D. Gocayne, J. Scott, R. Shirley, L. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. M. Geohagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith, and J. C. Venter, *Science* **269**, 496 (1995).
4. C. M. Fraser, J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. Sutton, J. M. Kelley, J. L. Fritchman, J. F. Weidman, K. V.

Small, M. Sandusky, J. Fuhrman, D. Nguyen, T. R. Utterback, D. M. Saudek, C. A. Phillips, J. M. Merrick, J. Tomb, B. A. Dougherty, K. F. Bott, P. Hu, T. S. Lucier, S. N. Peterson, H. O. Smith, C. A. Hutchison III, and J. C. Venter, *Science* **270**, 397 (1995).
5. R. Nowack, *Science* **269**, 468 (1995).
6. L. Rowen, B. F. Koop and L. Hood, *Science* **272**, 1755 (1996).
7. M. V. Olsen, *Science* **270**, 394 (1995).
8. E. Marshall, *Science* **268**, 1270 (1995).
9. J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong and C. M. Whitehouse, *Mass Spectrom. Rev.* **9**, 37 (1990).
10. F. Hillenkamp, M. Karas, R. C. Beavis and B. T. Chait, *Anal. Chem.* **63**, 1193A (1991).
11. *Nature (London)* **380**, 377 (1996).
12. L. Stryer, *Biochemistry*, 4th edn. Freeman, San Francisco (1995).
13. S. A. McLuckey, G. J. Van Berkel and G. L. Glish, *J. Am. Soc. Mass Spectrom.* **3**, 60 (1992).
14. P. Roepstorff and J. Fohlman, *Biomed. Mass Spectrom.* **11**, 601 (1984).
15. R. S. Johnson, S. A. Martin and K. Biemann, *Int. J. Mass Spectrom. Ion Processes* **86**, 137 (1988).
16. J. C. Venter, H. O. Smith and L. Hood, *Nature (London)*, **381**, 364 (1996).
17. F. Sanger, S. Nicklen and A. R. Couslon, *Proc. Natl. Acad. Sci. USA*, **74**, 5463 (1977).
18. E. N. Fung and E. S. Yeung, *Anal. Chem.* **67**, 1913 (1995).
19. A. T. Wolley and R. A. Mathies, *Anal. Chem.* **67**, 3676 (1995).
20. J. T. Perry, M. E. Johnson, P. M. Goodwin, J. C. Martin, J. H. Jett and R. A. Keller, *Anal. Chem.* **67**, 1755 (1995).
21. J. Frank, *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. Academic Press, San Diego (1996).
22. P. G. Arscott and V. A. Bloomfield, in *STM and SFM in Biology*, edited by O. Marti and M. Amrein, Chapt. 6. Academic Press, San Diego (1993).
23. A. C. Pease, D. Solas, E. J. Sullivan, M. T. Cronin, C. P. Holmes and S. P. A. Fodor, *Proc. Natl. Acad. Sci. USA*, **91**, 5022 (1994).
24. U. Pieles, W. Zürcher, M. Schär and H. E. Moser, *Nucleic Acids Res.* **21**, 3191 (1993).
25. D. P. Little, A. C. Russell, J. P. Spier, M. W. Senko, N. L. Kelleher and F. W. McLafferty, *J. Am. Chem. Soc.* **116**, 4893 (1994).
26. P. F. Crain, *Mass Spectrom. Rev.* **9**, 505 (1990).
27. A. L. Burlingame, R. K. Boyd and S. J. Gaskell, *Anal. Chem.* **68**, 599 (1996).
28. P. A. Limbach, P. F. Crain and J. A. McCloskey, *Curr. Opin. Biotechnol.* **6**, 96 (1995).
29. M. C. Fitzgerald and L. M. Smith, *Annu. Rev. Biophys. Biomol. Struct.* **24**, 117 (1995).
30. J. A. McCloskey and P. F. Crain, *Int. J. Mass Spectrom. Ion Processes* **118/119**, 593, (1992).
31. M. J. Doktycz, G. B. Hurst, S. Habibi-Goudarzi, S. A. McLuckey, K. Tang, C. H. Chen, M. Uziel, K. B. Jacobson, R. P. Woychik and M. V. Buchanan, *Anal. Biochem.* **230**, 205 (1995).
32. B. Spengler, Y. Pan, R. J. Cotter and L.-S. Kan, *Rapid Commun. Mass Spectrom.* **4**, 99 (1990).
33. P. Williams, *Int. J. Mass Spectrom. Ion Processes* **131**, 335 (1994).
34. R. W. Nelson, M. J. Rainbow, D. E. Lohr and P. Williams, *Science* **246**, 1585 (1989).
35. R. W. Nelson, R. M. Thomas and P. Williams, *Rapid Commun. Mass Spectrom.* **4**, 348 (1990).
36. D. M. Schieltz, C. Chou, C. Luo, R. M. Thomas and P. Williams, *Rapid Commun. Mass Spectrom.* **6**, 631 (1992).
37. T. Huth-Fehre, J. N. Gosine, K. J. Wu and C. H. Becker, *Rapid Commun. Mass Spectrom.* **6**, 209 (1992).
38. E. Nordhoff, R. Cramer, M. Karas, F. Hillenkamp, F. Kirpekar, K. Kristiansen and P. Roepstorff, *Nucleic Acids Res.* **21**, 3347 (1993).
39. K. J. Wu, A. Steding and C. H. Becker, *Rapid Commun. Mass Spectrom.* **7**, 142 (1993).
40. M. Karas and F. Hillenkamp, *Anal. Chem.* **60**, 2299 (1988).
41. K. Tang, N. I. Taranenko, S. L. Allman, L. Y. Chàng and C. H. Chen, *Rapid Commun. Mass Spectrom.* **8**, 727 (1994).
42. K. Tang, S. L. Allman, C. H. Chen, L. Y. Chàng and M. Schell, *Rapid Commun. Mass Spectrom.* **8**, 183 (1994).
43. M. Karas and U. Bahr, *Trends Anal. Chem.* **9**, 321 (1990).
44. G. R. Parr, M. C. Fitzgerald and L. M. Smith, *Rapid Commun. Spectrom.* **6**, 369 (1992).
45. E. Nordhoff, A. Ingendoh, R. Cramer, A. Overberg, B. Stahl, M. Karas and F. Hillenkamp, *Rapid Commun. Mass Spectrom.* **6**, 771 (1992).
46. K. Schneider and B. T. Chait, *Org. Mass Spectrom.* **28**, 1353 (1993).
47. F. Kirpekar, E. Nordhof, K. Kristiansen, P. Roepstorff, A. Lezius, S. Hahner, M. Karas and F. Hillenkamp, *Nucleic Acids Res.* **22**, 3866 (1994).
48. E. Nordhoff, M. Karas, R. Cramer, S. Hahner, F. Hillenkamp, F. Kirpekar, A. Lezius, J. Muth, C. Meier and J. W. Engels, *J. Mass Spectrom.* **30**, 99 (1995).
49. F. Kirpekar, E. Nordhoff, K. Kristiansen, P. Roepstorff, S. Hahner and F. Hillenkamp, *Rapid Commun. Mass Spectrom.* **9**, 525 (1995).
50. L. Zhu, G. R. Parr, M. C. Fitzgerald, C. M. Nelson and L. M. Smith, *J. Am. Chem. Soc.* **117**, 6048 (1995).
51. F. Greco, A. Liguori, G. Sindona and N. Uccella, *J. Am. Chem. Soc.* **112**, 9092 (1990).
52. A. Liguori, A. Napoli and G. Sindona, *Rapid Commun. Mass Spectrom.* **8**, 89 (1994).
53. K. Schneider and B. T. Chait, *Nucleic Acids Res.* **23**, 1570 (1995).
54. F. Seela and H. Steker, *J. Chem. Soc., Perkin Trans. 1*, 2573 (1985).
55. S. M. Colby, T. B. King and J. P. Reilly, *Rapid Commun. Mass Spectrom.* **8**, 865 (1994).
56. M. L. Vestal, P. Juhasz and S. A. Martin, *Rapid Commun. Mass Spectrom.* **9**, 1044 (1995).
57. N. P. Christian, S. M. Colby, L. Giver, C. T. Houston, R. J. Arnold, A. D. Ellington and J. P. Reilly, *Rapid Commun. Mass Spectrom.* **9**, 1061 (1995).
58. P. Juhasz, M. T. Roskey, I. P. Smirnov, L. A. Haff, M. L. Vestal and S. A. Martin, *Anal. Chem.* **68**, 941 (1996).
59. C. Köster, J. A. Castoro and C. L. Wilkins, *J. Am. Chem. Soc.* **114**, 7572 (1992).
60. T. Solouki, K. J. Gillig and D. H. Russell, *Anal. Chem.* **66**, 1583 (1994).
61. Y. Li, K. Tang, D. P. Little, H. Köster, R. L. Hunter and R. T. McIver, Jr, *Anal. Chem.* **68**, 2090 (1996).
62. T. R. Covey, R. F. Bonner, B. I. Shushan and J. Henion, *Rapid Commun. Mass Spectrom.* **2**, 249 (1988).
63. J. T. Stults and J. C. Marsters, *Rapid Commun. Mass Spectrom.* **5**, 359 (1991).
64. D. P. Little, T. W. Thannhauser and F. W. McLafferty, *Proc. Natl. Acad. Sci. USA* **92**, 2318 (1995).
65. S. C. Pomerantz, J. A. Kowalak and J. A. McCloskey, *J. Am. Soc. Mass Spectrom.* **4**, 204 (1993).
66. X. Cheng, D. C. Gale, H. R. Udseth and R. D. Smith, *Anal. Chem.* **67**, 586 (1995).
67. D. C. Muddiman, X. Cheng, H. R. Udseth and R. D. Smith, *J. Am. Soc. Mass Spectrom.* **7**, 697 (1996).
68. R. Chen, X. Cheng, D. W. Mitchell, S. A. Hofstadler, Q. Wu, A. L. Rockwood, M. G. Sherman and R. D. Smith, *Anal. Chem.* **67**, 1159 (1995).
69. S. D. Fuerstenau and W. H. Benner, *Rapid Commun. Mass Spectrom.* **9**, 1528 (1995).
70. M. B. Comisarow and A. G. Marshall, *J. Mass Spectrom.* **31**, 581 (1996).
71. P. G. Sealey and E. M. Southern, in *Gel Electrophoresis of Nucleic Acids*, edited by D. Rickwood and B. D. Hames, 2nd edn., p. 51. Oxford University Press, New York (1990).
72. M. C. Fitzgerald, L. Zhu and L. M. Smith, *Rapid Commun. Mass Spectrom.* **7**, 895 (1993).
73. T. A. Shaler, Y. Tan, J. N. Wickham, K. J. Wu and C. H. Becker, *Rapid Commun. Mass Spectrom.* **9**, 942 (1995).
74. L. Ch'ang, K. Tang, M. Schell, C. Ringelberg, K. J. Matteson, S. L. Allman and C. H. Chen, *Rapid Commun. Mass Spectrom.* **9**, 772 (1995).
75. B. T. Chait, R. Wang, R. C. Beavis and S. B. H. Kent, *Science* **262**, 89 (1993).
76. C. M. Bentzley, M. V. Johnston, B. S. Larsen and S. Gutteridge, *Anal. Chem.* **68**, 2141 (1996).
77. P. A. Limbach, J. A. McCloskey and P. F. Crain, *Nucleic Acids Symp. Ser.* **31**, 127 (1994).

78. R. P. Glover, G. M. A. Sweetman, P. B. Farmer and G. C. K. Roberts, *Rapid Commun. Mass Spectrom.* **9**, 897 (1995).
79. R. Kaufmann, D. Kirsch and B. Spengler, *Int. J. Mass Spectrom. Ion Processes* **131**, 355 (1994).
80. E. A. Stemmler, R. L. Hettich, G. B. Hurst and M. V. Buchanan, *Rapid Commun. Mass Spectrom.* **7**, 828 (1993).
81. P. Juhasz, M. T. Roskey, I. P. Smirnov, L. A. Haff, M. L. Vestal and S. A. Martin, *Anal. Chem.* **68**, 941 (1996).
82. M. W. Senko and F. W. McLafferty, *Annu. Rev. Biophys. Biomol. Struct.* **23**, 763 (1994).
83. S. A. McLuckey, G. J. Van Berkel and G. L. Glish, *J. Am. Soc. Mass Spectrom.* **3**, 60 (1992).
84. S. A. McLuckey and S. Habibi-Goudarzi, *J. Am. Soc. Mass Spectrom.* **5**, 740 (1994).
85. J. Ni, S. C. Pomerantz, J. Rozenski, Y. Zhang and J. A. McCloskey, *Anal. Chem.* **68**, 1989 (1996).
86. R. S. Brown and J. J. Lennon, *Anal. Chem.* **67**, 3990 (1995).
87. S. Alimpiev, E. Dratz, P. Kraft and J. Sunner, Proceedings of the 44th ASMS Conference on Mass Spectrometry and Allied Topics, Portland, OR, May 12–16, 1996 p. 645.