

# Analysis and sequencing of nucleic acids by matrix-assisted laser desorption/ionisation mass spectrometry

Marek Šebela

Department of Protein Biochemistry and Proteomics, Centre of the Region Haná for Biotechnological and Agricultural Research, Faculty of Science, Palacký University, Olomouc, Czech Republic. E-mail: [marek.sebela@upol.cz](mailto:marek.sebela@upol.cz)

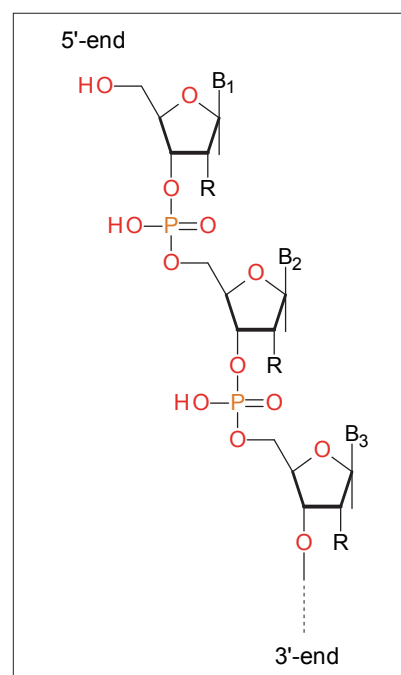
## Introduction

This article describes in brief the use of matrix-assisted laser desorption/ionisation (MALDI) mass spectrometry (MS) for the analysis of nucleic acids. Although the application of this method is limited by the size of the studied DNA or RNA molecules, it allows for determining their molecular mass with a high accuracy and there are also several MALDI-based approaches for reading the nucleotide sequence.

MALDI-MS has become widely popular for the mass analysis of biological molecules including biopolymers and synthetic compounds. The MALDI procedure starts with embedding molecules of the analyte in co-crystals within a suitable matrix compound, which is present in a high excess and absorbs light energy at the given laser wavelength to facilitate the ionisation process. Different lasers in the ultraviolet (UV) and infrared (IR) regions have been explored for this technique, but currently either a nitrogen laser ( $\lambda_{\text{max}}=337\text{ nm}$ ) or a Nd:YAG laser ( $\lambda_{\text{max}}=355\text{ nm}$ ) is typically used in biological research.<sup>1</sup> Since the method became introduced into common use in laboratories in the 1990s, various molecules such as proteins, peptides, oligosaccharides, oligonucleotides or small polynucleotides (e.g. intact tRNA molecules or DNA fragments from the Sanger sequencing) as well as synthetic polymers have proven amenable to routine measurements. The method is currently

applicable also to intact cells and spores of microorganisms and thin tissue samples (in this case via a MALDI imaging arrangement) for diagnostic purposes in science, medicine and technology.<sup>1</sup>

Nucleic acids have attracted the attention of biochemists, biologists, medicinal researchers, physicians, forensic scientists and bioinformaticists as key molecules of life. They are biopolymers composed of deoxyribonucleotides (in DNA) or ribonucleotides (in RNA), each containing a sugar, phosphate group and one of four nitrogenous bases.<sup>2</sup> DNA constituents include 2'-deoxyribose, phosphate, adenine (A), guanine (G), thymine (T) and cytosine (C). RNA contains ribose, phosphate and the same bases with the exception of thymine, which is replaced by uracil (U). As structural units, nucleosides are formed by connecting the N9 atom of the purine bases A and G or N1 atom of the pyrimidine bases C and T (or U in RNA) to the C1' atom of the respective sugar. The nucleosides are then linked to one another by phosphate groups attached to the 3'-hydroxyl group of a nucleoside and the 5'-hydroxyl group of the adjacent nucleoside thereby constituting in this way a phosphodiester bond (see Figure 1). Double helical DNA is formed when two hybridised antiparallel polynucleotide chains coil around a common axis. The chains are held together by hydrogen bonding between each pair of complementary bases: A–T or G–C. There is also a stacking interac-



**Figure 1.** A scheme of the chemical structure of single-stranded DNA and RNA. The symbol R stands for OH in RNA and H in DNA; B stands for any relevant heterocyclic base (A, C, G or T in DNA and A, C, G or U in RNA).

tion of adjacent base pairs involved with a stabilising role.<sup>2</sup>

## MALDI-MS measurements with oligonucleotides

Oligonucleotides are oligomers of ribonucleotides or deoxyribonucleotides with a broad range of applications in molecular biology and genetics, medicine and

forensic science.<sup>3</sup> Natural oligonucleotides are represented, for example, by small RNA molecules with a specific function in the regulation of gene expression (micro RNA) or they are derived from a degradation of large molecules of nucleic acids. Single-stranded synthetic oligonucleotides are prepared as commercial products based on a customer-specified sequence by solid-phase reactions from nucleoside phosphoramidites as building blocks.<sup>3</sup> The chain grows in the 3' to 5' direction, which is backwards relative to biosynthesis. The whole process is automated and the final product is analysed chromatographically to ensure a high level of its purity. Typically, synthetic oligonucleotides find their applications in a polymerase chain reaction (PCR), DNA sequencing by molecular biological methods, artificial gene synthesis and as molecular probes. Most of the applications utilise their property of specific binding to other oligonucleotides with complementary nucleotide sequences. The principle of complementarity is fundamental for various bioanalytical techniques such as DNA/RNA blotting, DNA microarrays or fluorescence *in situ* hybridisation (FISH).

Advances in MS of oligonucleotides and DNA have been particularly rapid since the introduction of electrospray ionisation (ESI) and MALDI. In 1990, Bernhard Spengler and co-workers published an analysis of short oligodeoxyribonucleotides (tetramers to hexamers) performed by acquiring mass spectra in the positive-ion mode with a MALDI instrument equipped with a Nd:YAG laser ( $\lambda_{\text{max}}=266\text{nm}$ ) and time-of-flight (TOF) mass analyser using nicotinic acid as a matrix.<sup>4</sup> The presence of sodium and potassium salts of oligonucleotides in the measured sample and their simultaneous desorption to form multiple peaks in the mass spectrum (reflecting various numbers of metal cations bound to the oxygen atoms of the phosphodiester groups) was shown to be an important reason for both considerably lowered mass resolution and signal intensity values when compared to MS spectra of peptides or proteins of the corresponding molecular masses. Such an obstacle can be over-

come satisfactorily by exchanging the alkali metal ions with ammonium ions. This is achieved either by adding ammonium salts (e.g. diammonium hydrogen citrate or diammonium-L-tartrate) in an excess to the sample solution or using  $\text{NH}_4^+$ -loaded cation exchange beads.<sup>2</sup> The ammonium ion can transfer a proton in the gas phase yielding the free acid of the phosphodiester group and an ammonia molecule. As a result,  $[\text{M}-\text{H}]^-$  or  $[\text{M}+\text{H}]^+$  ions are formed depending on the ionisation mode used. After introducing 3-hydroxypicolinic acid (3-HPA) as an efficient matrix for UV-MALDI of oligonucleotides,<sup>5</sup> mass spectra could be obtained using both ion polarities in a comparable quality. It has been shown that linear TOF mass analysers provide stronger signal intensities and better mass resolution values for molecules larger than 25-mers compared to reflector systems, which is attributed to the loss of ions during transmission through the reflector and a metastable decay.

There are a few other matrices for UV-MALDI, which have been reported suitable for experiments with oligonucleotides (but they are useful rather for molecules smaller than 25-mers): anthranilic acid in a mixture with nicotinic acid; 2',4',6'-trihydroxyacetophenone; 6-aza-2-thiothymine; 5-methoxysalicylic acid; quinaldic acid and 3,4-diaminobenzophenone. Also 2,5-dihydroxybenzoic acid, which is typically used with peptides, is applicable and there are binary matrices available consisting of, for example, 3-HPA and picolinic acid on a nitrocellulose film or 3-HPA and pyrazinecarboxylic acid. The use of these matrices has been described in previous review articles, e.g. in Reference 2. The tetraamine spermine, methylene blue or  $\beta$ -melanocyte-stimulating hormone (a peptide) have been both found to be useful as additives, which stabilise well small duplex DNA structures against their dissociation during MALDI-MS.<sup>6</sup> Synthetic oligonucleotides (i.e. a single-stranded DNA) are typically measured up to 100-mers, which is equivalent to a molecular mass of approximately 35kDa. The reason is that the mass resolution of oligonucleotide peaks decreases (despite sample purifi-

cation and desalting) with their increasing size. In contrast to DNA, larger RNAs (e.g. transcripts) can be analysed efficiently on MALDI-TOF instruments such as a 461-mer because of their higher stability.<sup>2</sup>

## DNA sequencing and MALDI-MS

DNA sequencing is the experimental way of reading the exact order of nucleotides (in other words, the order of the four bases—adenine, guanine, cytosine and thymine) in a DNA molecule. Hence it may be used to determine the sequence of a gene of interest, gene clusters, chromosomes or entire genomes. This kind of information has become indispensable in life sciences, especially in molecular biology, genomics and genetics, and in applied fields such as medicine, forensic science and biotechnology.<sup>7</sup>

The development of the classical DNA sequencing methods, which appeared in the 1970s, was largely stimulated by the discovery in the UK of the double-helix DNA structure by Watson and Crick in 1953 and the genetic code in the USA by the group of Nirenberg in 1965.<sup>3</sup> Because of their size, sequence determination of DNA molecules originally lagged far behind such analyses of RNA. At the beginning of the 1970s, Ray Wu and Frederick Sanger with co-workers independently introduced in their laboratories primer extension strategies with a DNA polymerase and <sup>32</sup>P-labelled nucleotides to produce short radioactive DNA fragments of assorted lengths (i.e. copies of defined regions of the template strand) amenable to a sequence reading. Sanger and Coulson devised the “plus and minus” method allowing the DNA sequence to be read directly from the autoradiograph of a separation electrophoretic gel analysis.<sup>7</sup> This innovative concept resided in the separate use of four pairs of plus and minus reaction mixtures with each reaction limited by a specific deoxynucleoside triphosphate (dNTP). In 1977, Sanger's group in Cambridge (UK) introduced chain-terminating inhibitors (namely 2', 3'-dideoxynucleoside triphosphates; ddNTPs) and applied them to the DNA of bacteriophage

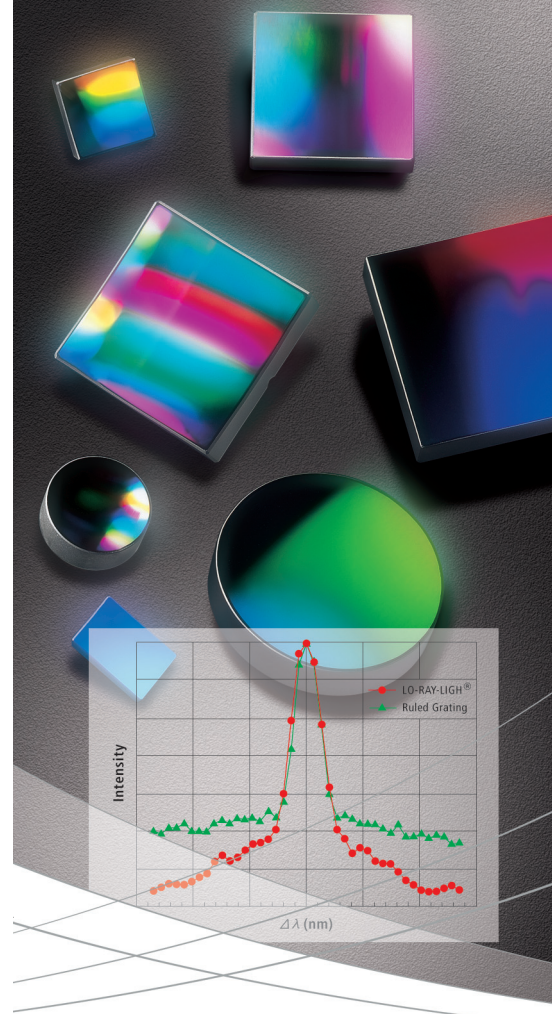
$\phi$ X174.<sup>8</sup> In this arrangement, the replica strand is extended only until a ddNTP is incorporated. The procedure was soon automated and became the method of choice from the 1980s until the middle of 2000s overcoming the contemporary chemical method by Allan Maxam and Walter Gilbert who utilised base-specific chemical cleavage reactions to generate fragments of end-labelled DNA. During that period, important advances were made such as fluorescent labelling and capillary electrophoresis. The first complete genome sequence for a free-living organism was determined for the opportunistic pathogenic bacterium *Hemophilus influenzae* in 1995. Nevertheless, the Sanger method has become largely promoted particularly because of its implementation in the Human Genome Project.<sup>7</sup>

The high demand for low-cost and high-throughput sequencing has driven the development of various "next-generation" technologies, in which numerous sequences are analysed in parallel involving, for example, a fragmentation of large DNA molecules, binding of adapter sequences, solid-phase reactions, miniaturisation, luciferase-based detection or ligation of oligonucleotide probes. This topic has been well reviewed in the literature,<sup>7</sup> thus only pyrosequencing could be mentioned explicitly as a kind of breakthrough.

MALDI-MS analysis of enzymatically synthesised DNA with 3-HPA as a matrix was reported already in the early days of introducing this method into common usage in laboratories.<sup>2</sup> In 1993, Fitzgerald *et al.*, for example, explored four mock sequencing mixtures containing 17–41-mer oligonucleotides terminated either with A, C, G or T. By superposing individual mass spectra of the four samples, the desired sequence could be read from the obtained ladder.<sup>9</sup> Real mixtures of dideoxynucleotide-terminated oligonucleotides from Sanger sequencing reactions became available to analysis by MALDI-MS after including their appropriate purification to remove buffer salts and other impurities. This way of analysis of sequencing products was made even more efficient when all four reactions were performed in one vial prior to MS.<sup>10</sup>

The use of MALDI-MS for DNA sequencing was challenging in the 1990s but the researchers then soon realised that this approach was limited to DNA sequences shorter than 100 nucleotides. This handicap has intensified by the continuous introduction of advantageous "next-generation" sequencing systems over the last decade. For these reasons MS-based sequencing has been limited mainly to genotyping, for the detection of mutations and single nucleotide polymorphisms (SNPs). SNPs are biallelic polymorphisms implicated in DNA variability of individuals. The nucleotide identity at these positions is generally confined to one of two possibilities. The presence of a specific SNP allele can be regarded as a causative factor in human genetic disorders and a screening for its presence might enable the detection of a genetic predisposition to disease.<sup>11</sup> There are numerous benefits that make MALDI-MS an outstanding platform by which to study SNPs: direct detection of the analyte, high accuracy, excellent resolution, simplicity of assay design and short analytical times. The principle resides typically in a primer extension reaction with annealing of the extension primer to a site immediately adjacent to the studied SNP. The key feature is the use of a mixture of dideoxy terminators. This arrangement yields such extension products, which differ in length (in an allele-specific manner) resulting in mass separations between ions reflecting two alleles equal to the mass of a nucleotide (~300 Da) and which are thus very easy for interpretation.<sup>11</sup>

The reverse Sanger sequencing employs  $\alpha$ -thiophosphate-containing dNTPs ( $\alpha$ -S-dNTPs) instead of ddNTPs. The full-length products of the four-aliquot sequencing reactions are subjected to a stringent exonuclease treatment yielding sequence ladders for each different base as the enzyme generates fragments terminated with a thiophosphate linkage. This approach has been combined with MALDI-MS for the sequencing of DNA<sup>12</sup> or RNA (obtained by transcription of a DNA template in the presence of  $\alpha$ -S-dNTPs).<sup>13</sup> Another protocol with DNA transcription involves a base-specific RNase digestion to gener-

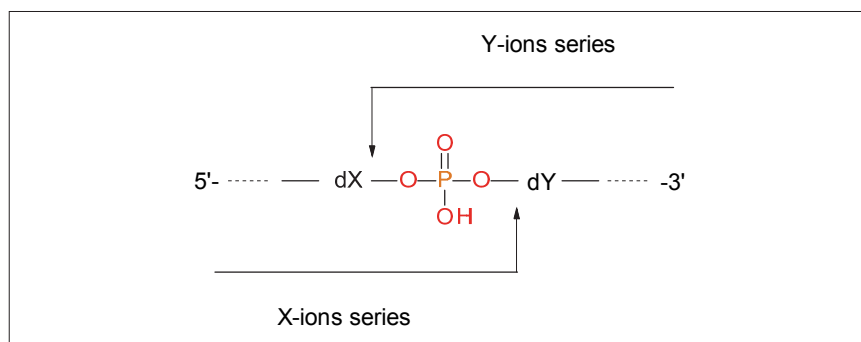


## World class

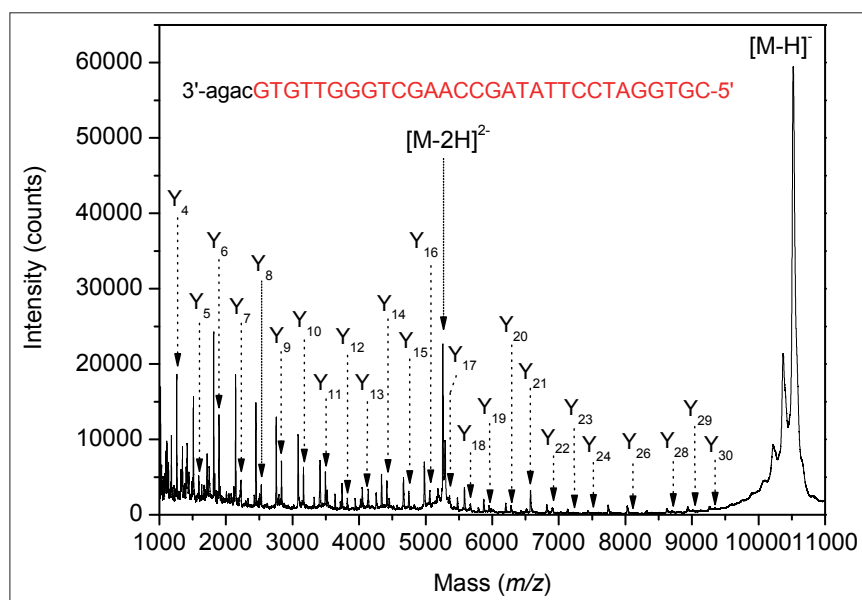
The unique LO-RAY-LIGH® holographic grating technology continues Shimadzu's proven track record on patented and awarded gratings.

- World's highest diffraction efficiency and the world's lowest stray light level
- Higher spectral resolution sharpness rendering better characteristics than conventional gratings
- Custom-made gratings of excellent performance matching the characteristics of different spectroscopic devices
- Widest range of optical components supporting the UV-2600/-2700 series and more

www.shimadzu.eu



**Figure 2.** Formation of the X- and Y-ions during a prompt fragmentation of oligonucleotides in MALDI-MS. This scheme has been adapted from Reference 16. The symbols dX and dY indicate the corresponding fragment chains.



**Figure 3.** MALDI-TOF mass spectrum showing fragmentation of a synthetic oligonucleotide. The spectrum was measured using a 1:1 (v/v) mixture of anthranilic acid (50 mg mL<sup>-1</sup> in 50% v/v acetonitrile) and 3-HPA (50 mg mL<sup>-1</sup> in 50% v/v acetonitrile) containing diammonium hydrogen citrate (1 mg mL<sup>-1</sup>) as a matrix at a relative laser power being higher by 20% than is routinely used on the instrument for molecular mass determinations of intact oligonucleotides. MALDI probes were prepared by depositing a sample drop onto matrix crystals on the target and waiting for a complete drying out. The sequence of the oligonucleotide 5'-CGTGGATCCTTATAGCCAAGCTGGGTTGTCAGA can almost completely be read in the reverse order (highlighted in red capital letters), from the 3'-end, using a series of Y-ions.

ate RNA fragments for MALDI-MS<sup>14</sup> or a chimeric RNA/DNA can be produced by a DNA polymerase incorporating both ribonucleotides and deoxyribonucleotides in the reaction mixture.<sup>15</sup> The product is then subjected to an alkaline hydrolysis, which yields cleavages only at the ribonucleotide sites. For investigating RNA, the mass difference between C and U is only 1 Da. A higher mass accuracy is thus needed.

### Fragmentation of oligonucleotides ions in MALDI-MS

The fragmentation of nucleic acid ions in MALDI-TOF MS has been reviewed comprehensively by Nordhoff *et al.*<sup>2</sup> and its full description is outside the scope of this article. It can be classified into more categories such as prompt, fast and metastable. A high propensity to fragment of oligodeoxyribonucleotide

ions increases strongly with increasing molecular mass. The preferred fragmentation scheme in MALDI-MS involves nucleobase eliminations followed by cleavages of the sugar–phosphate backbone. The stability of oligonucleotides to fragmentation decreases in the order oligothymidylic acids (oligouridylic acids) > RNA > DNA. RNA is more stable due to the effect of the electronegative 2'-hydroxyl group on the N-glycosidic bond, which prevents elimination of the nucleobases and results in a higher sensitivity of MS measurements. The prompt fragment ions of oligodeoxyribonucleotides result from cleavages of the sugar-phosphate backbone (see Figure 2).<sup>16</sup> The major ions series in a negative-mode IR-MALDI-MS with succinic acid as a matrix have been denominated X, X\* (=X+80) and Y. In UV-MALDI-MS, Y-ions are observed with 2,5-dihydroxybenzoic acid or anthranilic acid in a mixture with nicotinic acid rather than using 3-HPA.<sup>2</sup>

In typical UV-MALDI measurements performed to analyse the molecular mass of an oligodeoxyribonucleotide, the laser power is usually held at a threshold level in order to achieve the highest resolution and hence mass accuracy possible. When the laser power is increased by around 20%, a fragmentation is observed (an in-source decay measurement). Fragment peaks form a ladder, from which sequence information can be obtained. Figure 3 shows an experiment using a mixture of 3-HPA and anthranilic acid as a binary matrix. Masses of fragment ions can be identified with the sequence using freely available web software applications (calculators) for oligonucleotides such as OligoCalc (<http://biotools.nubic.northwestern.edu/OligoCalc.html>).

### Conclusions

DNA sequencing has become a routine procedure. There are modern and efficient instruments available for this purpose in research, diagnostic and technological laboratories, or one can take advantage of commercial services. The development in this field of methodology is continuous because of the driving force of broad demands by biologists

for generating a lot of sequencing data in a short time and at the lowest possible price. MALDI-TOF mass spectrometry was in the 1990s considered a big competitor of the traditional molecular biologic approaches to DNA sequencing. The main reasons resided in emphasizing its accuracy and sensitivity. Nowadays the role of MS-based sequencing methods for nucleic acids does not belong in the mainstream. Nevertheless, it has several interesting practical applications including, for example, the verification of synthesis of standard or modified oligonucleotides, characterisation of naturally occurring modified nucleic acids and polymorphisms.

## Acknowledgements

This work was supported by grant no. LO1204 from the Ministry of Education, Youth and Sports, Czech Republic (the National Program of Sustainability I). Štěpán Kouřil is thanked for his help with the fragmentation analysis depicted in Figure 3.

## References

1. R. Cramer (Ed), *Advances in MALDI and Laser-Induced Soft Ionization Mass Spectrometry*. Springer International Publishing Switzerland (2016). doi: <http://dx.doi.org/10.1007/978-3-319-04819-2>
2. E. Nordhoff, F. Kirpekar and P. Roepstorff, "Mass spectrometry of nucleic acids", *Mass Spectrom Rev.* **15**, 67 (1996). doi: [http://dx.doi.org/10.1002/\(SICI\)1098-2787\(1996\)15:2<67::AID-MAS1>3.0.CO;2-8](http://dx.doi.org/10.1002/(SICI)1098-2787(1996)15:2<67::AID-MAS1>3.0.CO;2-8)
3. G.M. Blackburn, M.J. Gait, D. Loakes and D.M. Williams (Eds), *Nucleic Acids in Chemistry and Biology*, 3<sup>rd</sup> Edn. RSC Publishing, Cambridge (2006).
4. B. Spengler, Y. Pan, R.J. Cotter and L.S. Kan, "Molecular weight determination of underivatized oligodeoxyribonucleotides by positive-ion matrix-assisted ultraviolet laser-desorption mass spectrometry", *Rapid Commun. Mass Spectrom.* **4**, 99 (1990). doi: <http://dx.doi.org/10.1002/rcm.1290040402>
5. K.J. Wu, A. Steding and C.H. Becker, "Matrix-assisted laser desorption time-of-flight mass spectrometry of oligonucleotides using 3-hydroxyisobutyric acid as an ultraviolet-sensitive matrix", *Rapid Commun. Mass Spectrom.* **7**, 142 (1993). doi: <http://dx.doi.org/10.1002/rcm.1290070206>
6. A.M. Distler and J. Allison, "Additives for the stabilization of double-stranded DNA in UV-MALDI MS", *J. Am. Soc. Mass Spectrom.* **13**, 1129 (2002). doi: [http://dx.doi.org/10.1016/S1044-0305\(02\)00430-0](http://dx.doi.org/10.1016/S1044-0305(02)00430-0)
7. A. Munshi (Ed), *DNA Sequencing – Methods and Applications*. InTech, Rijeka, Croatia (2012).
8. F. Sanger, S. Nicklen and A.R. Coulson, "DNA sequencing with chain-terminating inhibitors", *Proc. Natl. Acad. Sci. USA* **74**, 5463 (1977). doi: <http://dx.doi.org/10.1073/pnas.74.12.5463>
9. M.C. Fitzgerald, L. Zhu and L.M. Smith, "The analysis of mock DNA sequencing reactions using matrix-assisted laser desorption/ionization mass spectrometry", *Rapid Commun. Mass Spectrom.* **7**, 895 (1995). doi: <http://dx.doi.org/10.1002/rcm.1290071008>
10. E. Nordhoff, C. Luebbert, G. Thiele, V. Heiser and H. Lehrach, "Rapid determination of short DNA sequences by the use of MALDI-MS", *Nucleic Acids Res.* **28**, e86 (2000). doi: <http://dx.doi.org/10.1093/nar/28.20.e86>
11. N. Storm, B. Darnhofer-Patel, D. van den Boom and C.P. Rodi, "MALDI-TOF mass spectrometry-based SNP genotyping", in *Methods in Molecular Biology, Vol. 212-Single Nucleotide Polymorphisms, Methods and Protocols*, Ed by P.-Y. Kwok. Humana Press, Totowa, NJ, p. 243 (2003).
12. T.A. Shaler, C.H. Becker, Y. Tan, J.N. Wickham and K.J. Wu, "Analysis of enzymatic DNA sequencing reactions by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry", *Rapid Commun. Mass Spectrom.* **9**, 942 (1995). doi: <http://dx.doi.org/10.1002/rcm.1290091015>
13. B. Spottke, J. Gross, H.J. Galla and F. Hillenkamp, "Reverse Sanger sequencing of RNA by MALDI-TOF mass spectrometry after solid phase purification", *Nucleic Acids Res.* **32**, e97 (2004). doi: <http://dx.doi.org/10.1093/nar/gnh089>
14. S. Krebs, I. Medugorac, D. Seichter and M. Forster, "RNaseCut: a MALDI mass spectrometry-based method for SNP discovery", *Nucleic Acids Res.* **31**, e37 (2003). doi: <http://dx.doi.org/10.1093/nar/gng037>
15. F. Mauger, K. Bauer, C.D. Calloway, J. Semhoun, T. Nishimoto, T.W. Myers, D.H. Gelfrand and I.G. Gut, "DNA sequencing by MALDI-TOF MS using alkali cleavage of RNA/DNA chimeras", *Nucleic Acids Res.* **35**, e62 (2007). doi: <http://dx.doi.org/10.1093/nar/gkm056>
16. E. Nordhoff, M. Karas, R. Cramer, S. Hahner, F. Hillenkamp, F. Kirpekar, A. Lezius, J. Muth, C. Meier and J.W. Engels, "Direct mass spectrometric sequencing of low-picomole amounts of oligodeoxynucleotides with up to 21 bases by matrix-assisted laser desorption/ionization mass spectrometry", *J. Mass Spectrom.* **30**, 99 (1995). doi: <http://dx.doi.org/10.1002/jms.1190300116>

### EXPERTS in FLUORESCENCE SPECTROSCOPY



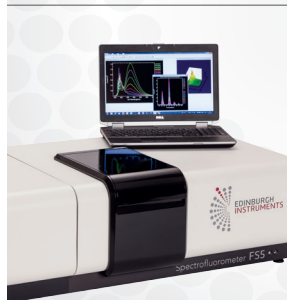
EDINBURGH  
INSTRUMENTS

#### FLS980 Complete Luminescence Laboratory



- > Research-grade fluorescence spectrometer
- > Single photon counting, Fluorescence lifetimes (TCSPC) and Phosphorescence lifetimes (MCS)
- > >25,000:1 Water Raman SNR
- > Multiple sources and detectors
- > Upgrades include - NIR coverage, microscopes, plate reader module, polarisation, integrating sphere and more

#### F55 Photon Counting Spectrofluorometer



- > Fluorescence and phosphorescence measurements capability, including lifetimes, up to 1650 nm
- > >6,000:1 Water Raman SNR, the most sensitive bench-top fluorometer available
- > Transmission detector for UV-VIS absorption spectroscopy
- > UV-VIS-NIR range

sales@edinst.com | www.edinst.com