High-Throughput Analysis of Oligonucleotides Using Automated Electrospray Ionization Mass Spectrometry

by Mark E. Hail, Brian Elliott, and Kathleen Anderson

The power of the oligonucleotide (oligo) microarray has changed the face of biology and brought about terms like "new biology." In his book, *Introduction to Proteomics*, Liebler states that "one array can replace thousands of Northern-blot analyses and can be done in the time it would take to do one Northern."[1] Both the pharmaceutical industry and the FDA are currently evaluating the utility of microarray-based genotyping assays. They are working to determine the accuracy of the data and subsequently how and when to take these data into consideration.[2]

As the genomic era matures, it is becoming increasingly clear that oligonucleotide quality is critical in all applications of genomic analyses. In addition, the increased use of long oligonucleotides (e.g., 70 mers) in printed microarrays has further increased the need for higher-performance analytical systems that have the resolution, speed, and accuracy to detect sequence failures unambiguously. With this in mind, an integrated system was developed with optimized analytical methods that allow for oligonucleotide quality control (QC) or detailed impurity profiling on the same instrument. The system utilizes automated LC-MS and provides a number of advantages relative to traditional methods of analysis in terms of mass accuracy, resolution, and sensitivity for a wide range of oligonucleotide analysis applications.

Accurate molecular weight determination by mass spectrometry has become the method of choice for oligonucleotide QC. Traditionally, these analyses have been done by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS).[3,4] The advantages of MALDI-TOF are that it is simple to operate, offers high sensitivity, and is amenable to high-throughput analysis. However, MALDI-TOF is not particularly effective when analyzing long (>50 bases) and/or fragile oligonucleotides. Furthermore, although MALDI data acquisition is quite fast, the sample must be mixed with matrix and physically deposited onto the MALDI target. In addition, if the oligo samples contain a significant amount of salt, they must be desalted off-line before being applied to the MALDI target. In a high-throughput laboratory, the desalting and spotting are usually accomplished with some type of robotic/liquid handling system, which can add to the cost and complexity of the overall system.

Electrospray ionization-mass spectrometry (ESI-MS) is also a very powerful technique for oligonucleotide analysis. ESI-MS is advantageous in that it can be directly coupled to HPLC, allowing for on-line desalting and chromatographic separation (i.e., LC-MS).[5] The Oligo HTCS (high-throughput characterization system) (Novatia, LLC, Princeton, NJ) utilizes ESI-LC-MS and automated data processing for high-throughput analysis of oligonucleotides. The analytical methods and software were developed initially to address the needs of the company's contract analysis clients. The core technology has been refined over the last three years and has been used to analyze thousands of client samples. The complete oligonucleotide analysis system includes LC-MS hardware, analytical methods, and software.

An LC-MS based approach has many benefits for both oligonucleotide QC and detailed impurity analysis. On the Oligo HTCS, a typical oligonucleotide mass is routinely determined to within 100 ppm or 0.01%. For example, a 60 mer of approx. 18,000 D is measured to within 2 D, which is sufficient mass accuracy to detect an A/T base substitution (a 9-D mass difference). Unlike MALDI-TOF, the mass accuracy, resolution, and sensitivity for the Oligo HTCS are maintained across the entire mass range of oligos analyzed. Oligos exceeding 120 bases in length are routinely analyzed on the system with no degradation in analytical performance versus a standard 20 mer. A typical analysis on the system consumes ~10–100 pmol of oligonucleotide, while absolute detection limits are in the high femtomole range.

A key component of the system is automated data processing capability. ProMass automated biomolecule deconvolution software (Novatia) is used to process raw ESI-MS spectra to determine molecular masses and confirm the masses with those expected from the oligonucleotide sequences. The software also automates the reporting of results and displays them in an easy-to-use Web browser format. Sample throughput is on the order of 1000 samples per 24-hr period. Although throughput is perhaps not as high as can be obtained on some MALDI systems, the advantages of high mass accuracy and resolution and the ability to analyze long and/or fragile oligonucleotides are compelling in many applications of oligonucleotide analysis.

Experimental

MALDI-TOF data were acquired on a Voyager DE system (Applied Biosystems, Framingham, MA) using 3-hydroxypicolinic acid/ammonium citrate matrix and positive ion detection. The Oligo HTCS LC-MS system consisted of a CTC HTS-PAL autosampler (LEAP Technologies, Carrboro, NC), Paradigm MS4 HPLC with UV detector (Michrom BioResources, Auburn, CA), and Finnigan TSQ7000 LC-MS system (Thermo Electron, San Jose, CA). All devices were controlled through the Xcalibur version 1.2 data system (Thermo Electron).

The multiply charged ion ESI-MS data were processed using ProMass version 1.4 software. All data were acquired in negative ion mode. Custom macros written in PAL Cycle Composer were used to optionally dilute and mix samples prior to injection and on-line desalting.

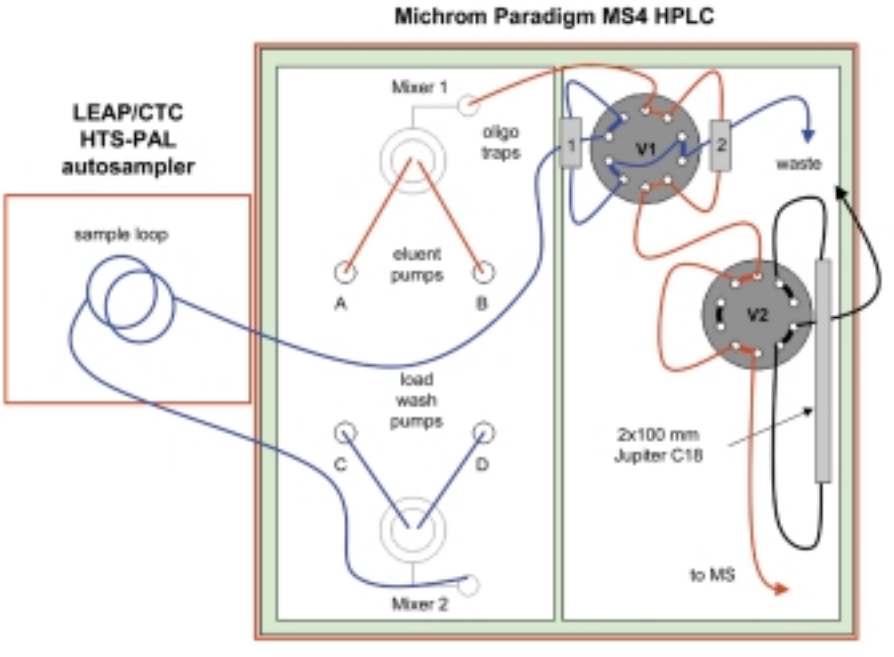

Figure 1 Oligo HTCS LC-MS plumbing diagram. Sample is flushed from the autosampler loop at 1 mL/min using pumps C/D and desalted on trap column #1, while trap column #2 is eluted with pumps A/B. For high-throughput analysis, the column on valve V2 is bypassed. Valve V1 is toggled for each injection.

Maximum sample throughput on the system was obtained with the plumbing arrangement shown in *Figure 1*. With this system, sample was rapidly flushed from the autosampler at 1 mL/min using pumps C/D onto one of two available 1 × 10 mm oligo desalting trap columns (Novatia) installed on valve V1 of the Paradigm HPLC. This arrangement allowed one trap to be eluted (with pumps A/B) while the other was being regenerated for the next run (pumps C/D). Analysis time (inject-to-inject) was ~1.4 min/sample or ~1.7 min/sample with optional predilution and mixing. In high-throughput mode, sample was eluted with 60/40 A/B at 0.2 mL/min, where A contained water and B contained 90% methanol. An ion-pairing mobile phase, adapted from the work of Apffel et al., was used for both high-throughput and detailed LC-MS analyses.[5] Both A and B solvents contained 0.75% hexafluoroisopropanol, 0.0375% triethylamine, and 10 µ*M* ethylenediaminetetraacetic acid (EDTA) (pH 7.3 for solvent A). Mobile phase C was identical to A, while solvent D contained 40/40/20 MeOH/ACN/H_2O. A 20-sec pulse of 100% D was used to clean the injection loop between injections in order to eliminate sample carryover.

The addition of EDTA to the mobile phase was found to be very beneficial in preventing Na, K, and Fe adducts caused by the presence of trace amounts of these cations in the samples or HPLC system. The addition of EDTA was also found to be more effective than simply raising the pH, which could cause degradation of certain oligonucleotides containing base-labile moieties. An advantage of the plumbing arrangement shown in Figure 1 was that the LC-MS system could be automatically switched between high-throughput and detailed LC-UV-MS profiling modes within an analysis sequence. The detailed profiling mode provided additional information about sample purity and the relative abundance of failure sequences, impurities, or degradants. The profiling mode was activated with valve V2 on the Paradigm HPLC by switching an analytical column in-line with one of the desalting traps. Typically, a gradient from 5 to 40% B in 20 min was used with a Jupiter C18 2 × 100 mm 300-Å column (Phenomenex, Torrance, CA).

Results and discussion

A comparison of data from the Oligo HTCS ESI-LC-MS was performed versus data obtained from a high-performance MALDI-TOF system using synthetic oligos ranging from 10 to 120 bases in length (3–37 kD). The oligos and the MALDI data were provided by Integrated DNA Technologies (IDT) (Coralville, IA). The results are summarized in *Table 1*. Both MALDI-TOF and LC-MS provided high-quality data for standard 20 mers, showing comparable mass resolution and accuracy. However, MALDI-TOF displayed marked degradation in resolution and mass accuracy for oligos greater than ~60 bases in length. In contrast, the LC-MS system demonstrated no significant degradation in mass accuracy or resolution with oligo length. Representative data for comparison purposes are shown in *Figure 2*, with overlays of MALDI and ESI-LC-MS data of a 60-mer oligo with a calculated average mass of 18,410.0 D (Figure 2*a*) and a 120-mer with a calculated average mass of 37,031.0 D (Figure 2*b*).

Table 1 Experimental mass accuracy for standard oligonucleotides analyzed by MALDI-TOF and ESI-LC-MS

		MALDI-TOF		ESI-LC-MS	
Oligonucleotide	Calculated mass (D)	Experimental mass (D)	Error (ppm)	Experimental mass (D)	Error (ppm)
10 mer	2993.0	2992.1	–300.7	2993.1	33.4
20 mer	6168.1	6167.4	–113.5	6167.6	–81.1
40 mer	12300.0	12298.6	–113.8	12299.7	–24.4
60 mer	18410.0	18414.6	249.9	18410.1	5.3
80 mer	24661.1	24664.7	146.0	24661.0	–4.1
100 mer	30951.1	30978.3	878.8	30950.6	–16.2
120 mer	37031.0	37121.6	2446.6	37030.0	–27.0

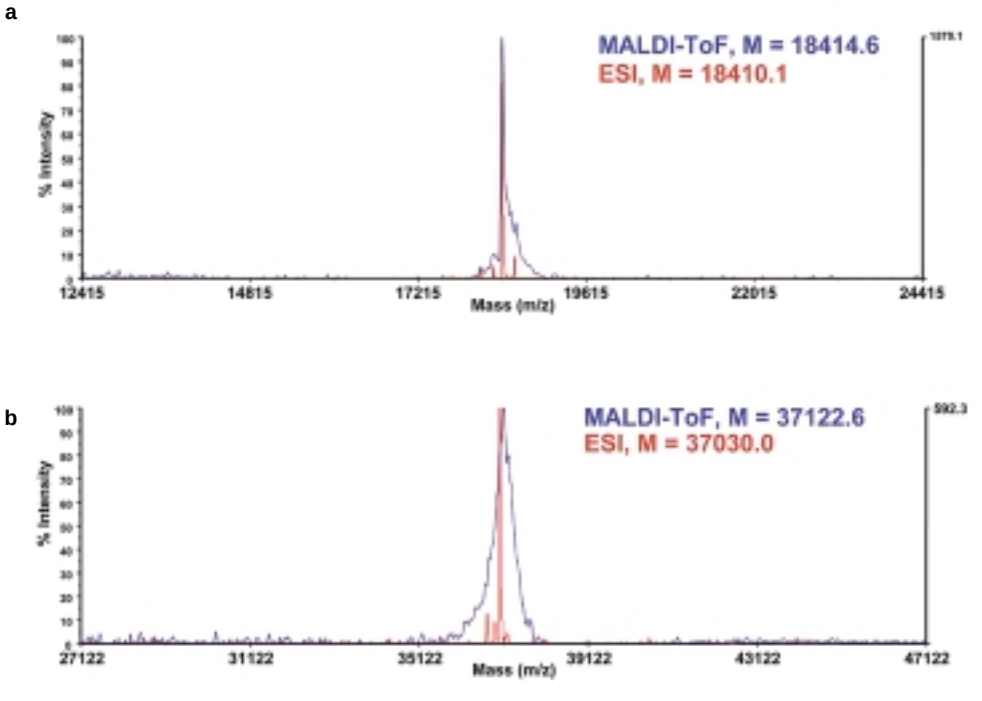

Figure 2 Overlays of MALDI-TOF spectra (blue traces) and deconvoluted ESI-MS spectra (red traces) for oligonucleotides plotted on the same scale: a) 60 mer, 18410.0 D; b) 120 mer, 37031.0 D.

Another advantage of the LC-MS system is that oligos containing labile protecting groups or unusual chemistries are analyzed successfully intact. For example, oligos containing 5′-dimethoxytrityl (DMT) are detected intact with LC-MS, but are frequently detected in MALDI-TOF as a mixture of DMT on/off peaks. Another class of oligos that are difficult to analyze by MALDI but are analyzed successfully by ESI-LC-MS are dual-labeled probes. Some of the quencher moieties strongly absorb at the MALDI UV laser wavelength, causing dissociation of this portion of the oligo during the desorption/ionization process. Data demonstrating this effect are shown in *Figure 3*, with the overlay of MALDI and ESI-LC-MS data of a dual-labeled probe containing a 3′ Black Hole Quencher (Biosearch Technologies, Inc., Novato, CA).

One of the past limitations of using an LC-MS based approach for oligonucleotide analysis is that the ESI mass spectra of oligos are more complex than those produced by MALDI-TOF. For example, most of the ions in a typical MALDI mass spectrum are singly charged (e.g., $[M+H]^+$), allowing for very simple determination of the molecular mass. On the other hand, ESI mass spectra of large biomolecules typically contain many peaks due to the production of multiply charged ions.[6] In the case of oligonucleotides, deprotonation of the phosphate backbone results in the production of a series of negatively charged ions. The charge state distribution, as it is often called, must be transformed mathematically (or deconvoluted) to determine the uncharged masses of all peaks in the original mass spectrum. Even with computer algorithms, this process has traditionally been highly user-interactive and not automated for large numbers of samples.

The approach of Zhang and Marshall was adapted by Novatia to develop an algorithm that automates the ESI spectral deconvolution process for entire LC-MS data sets.[7] The data analysis software, ProMass, produces artifact-free deconvoluted mass spectra from all chromatographic peaks found in an ESI-LC-MS data file.[8,9] In the case of the high-throughput oligo method, a single injection peak elution profile is summed and deconvoluted to determine the masses of oligos present in each sample. The deconvoluted masses are automatically compared with expected target mass(es) from

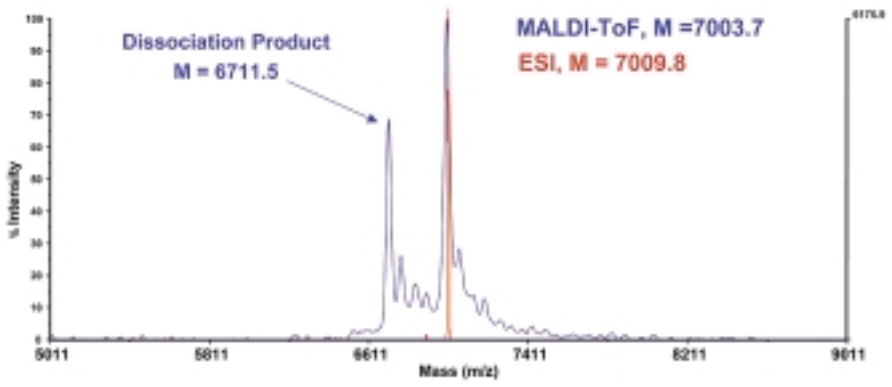

Figure 3 Overlays of MALDI-TOF spectrum (blue trace) and deconvoluted ESI-MS spectrum (red trace) for a dual-labeled probe oligo containing a 3′ Black Hole Quencher. The MALDI spectrum shows a dissociation product induced by strong absorption of the quencher at the UV laser wavelength. No decomposition is observed in the ESI-MS data.

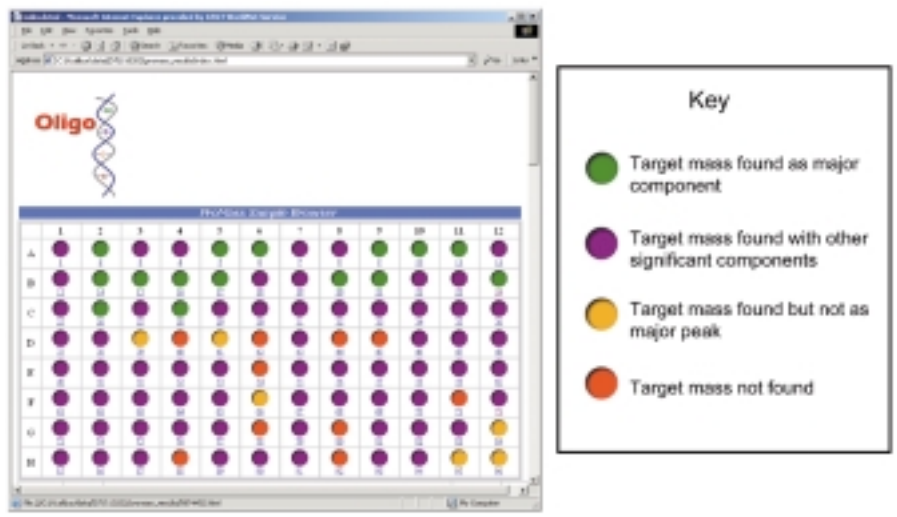

Figure 4 ProMass sample browser Web-based results display from automated ESI-LC-MS analysis of a 96-well plate of oligonucleotide samples. The color coding indicates the presence or absence of target masses as shown in the key. Color-coded wells are hyperlinked to detailed reports that display chromatograms and mass spectra.

each sample. Target masses are automatically calculated from oligo sequences or can be explicitly defined by the user. The software is run in batch mode, permitting an entire sample list to be acquired and processed without user interaction. The results from ProMass processing are Web-based, allowing quick data review from any Web browser. A color-coded sample plate viewer enables the user to quickly review the analysis results from an entire plate and helps to focus analyst attention on those samples that did not produce the expected result. As shown in *Figure 4*, the color codes reflect whether the desired target mass was detected and if any other significant components were present. The user can obtain additional information about any particular sample by clicking on its color-coded sample well to obtain chromatograms as well as raw and deconvoluted mass spectra.

Conclusion

An LC-MS based approach for totally automated analysis of oligonucleotides was presented. The Oligo HTCS integrated system offers significant improvements in mass resolution and accuracy relative to MALDI-TOF for the analysis of long (>50 bases) or fragile oligonucleotides. Sample throughput of 1000 oligos per day on the system is compatible with the output of many high-throughput synthesis laboratories, particularly when it is used to complement existing MALDI-TOF instrumentation. When employed alongside MALDI-TOF, the system can be dedicated for use on those samples that are difficult to analyze by MALDI or other techniques.

References

1. *Liebler DC. Introduction to proteomics: tools for the new biology. Totowa, NJ: Humana Press, 2002.*
2. *McGowan K. The need to make microarrays mainstream. Genome Tech 2002; 22:77–81.*
3. *Ball RW, Packman LC. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry as a rapid quality control method in oligonucleotide synthesis. Anal Biochem 1997; 246(2):185–94.*
4. *Van Ausdall DA, Marshall WS. Automated high-throughput mass spectrometric analysis of synthetic oligonucleotides. Anal Biochem 1998; 256(2):220–8.*
5. *Apffel A, Chakel JA, Fischer S, Lichtenwalter K, Hancock WS. Analysis of oligonucleotides by HPLC-electrospray ionization mass spectrometry. Anal Chem 1997; 69:1320–5.*
6. *Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. Science 1989; 246(4926):64–71.*
7. *Zhang Z, Marshall AG. A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra. J Am Soc Mass Spectrom 1998; 9(3):225–33.*
8. *Hail ME, Whitney JL. Automated LC/MS profiling of proteins using the ProMass protein molecular weight data browser. Proceedings of the 49th ASMS Conference on Mass Spectrometry and Allied Topics, WPH 150, Chicago, IL, 2001.*
9. *Hail ME, Whitney JL, Detlefsen DJ. Improved methods for totally automated analysis, data processing, and results reporting in biomolecule LC/MS. Proceedings of the 50th ASMS Conference on Mass Spectrometry and Allied Topics, MPK 351, Orlando, FL, 2002.*

Dr. Hail is President, and Ms. Anderson is Director of Sales and Marketing, Novatia, LLC, 301A College Rd. E., Princeton, NJ 08540, U.S.A.; tel.: 609-951-0181; fax: 609-951-0185; e-mail: hailm@enovatia.com. Mr. Elliott is Director of Mass Spectrometry, Integrated DNA Technologies, Coralville, IA, U.S.A.