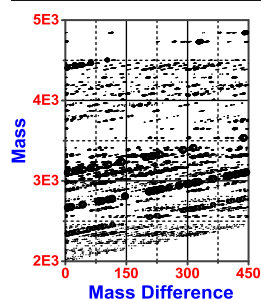


A Novel and Intuitive Method of Displaying and Interacting with Mass Difference Information: Application to Oligonucleotide Drug Impurities

Stilianos G. Roussis

Isis Pharmaceuticals, Carlsbad, CA 92008, USA



Abstract. A new method is presented for determining relationships between components in complex analytical systems. The method uses the mass differences between peaks in high resolution electrospray ionization (ESI) mass spectra. It relates peaks that share common mass differences. The method is based on the fundamental assumption that peaks in the spectra having the same exact mass difference are related by the same chemical moiety/substructure. Moreover, the presence (or absence/loss) of the same chemical moiety from a series of molecules may reflect similarities in the mechanisms of formation of each molecule. The determined mass differences in the spectra are used to automatically differentiate the types of components in the samples. Contour plots and summary plots of the summed total ion signal

as a function of the mass difference are generated, which form powerful tools for the rapid and automated determination of the components in the samples and for comparisons with other samples. For the first time, in this work a unique profile contour plot has been developed that permits the interactive interrogation of the mass range by mass difference data matrix to obtain valuable information about components that share a common mechanism of formation, and all possible mechanisms of formation linked to a selected precursor molecule. The method can be used as an additional and complementary method to the existing analytical methods to determine relationships between components in complex chemical systems.

Keywords: Mass differences, High resolution mass spectra, Sample component relationships, Oligonucleotides, Drug impurities

Received: 22 December 2014/Revised: 26 January 2015/Accepted: 23 February 2015/Published Online: 3 April 2015

Introduction

Significant developments in the understanding of RNAs and their involvement in cellular processes and the regulation of genes have taken place in recent years [1]. RNA is an important potential target for therapeutic modulation, and antisense oligonucleotides can be used as a direct approach to bind to target RNA by well-characterized Watson-Crick base pairing and modulate its function through a variety of post-binding mechanisms [2, 3]. Several such antisense oligonucleotide drugs have been tested in clinical trials and a few of them have been approved by regulatory agencies [3]. However, unmodified oligonucleotides have low specificity and

therapeutic efficacy because of nuclease-mediated degradation of the phosphodiester linkage [3]. In addition, they are unsuitable as systemic therapeutics because they bind weakly to plasma proteins and are rapidly filtered by the kidney. These issues can be overcome by chemical modifications of the phosphodiester backbone, of the purine and pyrimidine heterocyclic bases, and of the sugar moiety [3, 4]. Phosphorothioate (PS)-containing oligonucleotides, where one of the nonbridging phosphate oxygen atoms is replaced by one sulfur atom, were one of the earliest and currently most widely used backbone modifications of oligonucleotide molecules [5]. Some of the desirable phosphorothioate oligonucleotide properties include binding specificity, enzymatic stability, and low toxicity. Owing to their high therapeutic potential, a significant amount of effort is currently being invested in the discovery, development, and manufacture of synthetic oligonucleotide molecules with the most desirable chemical and biological properties.

Electronic supplementary material The online version of this article (doi:10.1007/s13361-015-1115-0) contains supplementary material, which is available to authorized users.

Correspondence to: Stilianos G. Roussis; e-mail: sroussis@isisph.com

One of the most attractive features of antisense oligonucleotide drugs is that their synthesis can be fully automated using solid-phase synthesis approaches with commercially available nucleoside phosphoramidites as starting materials [6]. The development of advanced understanding and control of the manufacturing reactions requires the availability of advanced analytical methods able to characterize the quality of the final drug product and the level and nature of the trace impurities.

Capillary gel electrophoresis (CGE) [7], strong anion-exchange (SAX) chromatography [8], ^{31}P NMR spectroscopy [8, 9], matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (MS) [10, 11], liquid chromatography coupled to electrospray ionization (ESI) mass spectrometry [12, 13], and tandem mass spectrometry (MS/MS) [13–15] are among the analytical methods that have been used to determine the quality of oligonucleotides and characterize their impurities. Ion pair-reversed phase (IP-RP) HPLC coupled with a UV detector and electrospray ionization has been identified as one of the most suitable methods because of its ability to provide increased chromatographic resolution and reliable qualitative and quantitative information [6]. Advanced molecular characterization is possible using high-resolution MS [16, 17] and/or MS/MS instrumentation [13, 18, 19].

Extensive previous work has been done to identify trace impurities in oligonucleotide systems because of their potentially adverse effects on drug quality and their negative impact on manufacturing yields. Even if synthetic yields of 98%–99% are achievable at each oligonucleotide synthesis step, the total amount of impurities tends to increase with the length of the oligonucleotide chain because of the large number of synthesis steps. The nature of the particular deviations from the optimal reaction and purification process conditions controls the amounts and types of impurities formed. Characterization and determination of the source origin of the impurities is the first step in the development of mitigation measures for prevention and elimination of their formation. Phosphorothioate oligonucleotide impurities can be classified in several categories depending on the type of reaction leading to their formation: (1) failure sequences due to coupling inefficiency, incomplete capping, or detritylation (shortmers) [20, 21]; (2) molecules larger than the full length product (longmers) [22, 23]; and (3) other impurity products formed by depurination, deamination, sulfur loss, thermal stress, adduct formation, and other reactions [24–28]. It is possible for some of these impurities to be formed by mixed consecutive or parallel secondary reactions initiated by solvent and raw materials initial impurities, process variables, or the chemical nature of the target oligonucleotide structure.

The conventional approach for the determination of the oligonucleotide impurities involves manual inspection of the liquid chromatograms for the detection of peaks exceeding a minimum amount (e.g., 0.1% by UV detection) in comparison with the amount of the peak of the main drug product. Typically, a list of impurities determined from previous work is manually inspected against the detected peaks in the current chromatograms. Additional computations are required prior to

that step to determine the masses of the expected impurities from their chemical formulas. Also, considerations must be made to account for the expected range of charge multiplicity of the impurities in the mass spectra (i.e., determination of the expected m/z values for a range of possible z charge states). The process, although reliable, is long and tedious and it can take several days to exhaustively inspect all peaks in a series of chromatograms representing replicate analysis or different lots of the same sample. Additionally, the manual inspection of the impurities in the chromatogram can be limited by the resolution of the chromatographic method (i.e., overlapping peaks). Most importantly, relatively weak impurity peaks, previously unidentified, may be missed by the direct comparison with a previously created list of expected impurity peaks. A faster and more reliable approach is needed to overcome these limitations.

In this work, we have explored the use of an approach that is based on the mass differences between peaks in the deconvoluted electrospray ionization mass spectrum obtained by averaging the acquired spectra over a selected retention time region of the IP-RP HPLC MS run. The approach generates a tremendous amount of information in a very short period of time that can be fully automated in tabular reports for rapid computer comparisons or for manual inspections and exploration of the acquired spectra. The method relates peaks in the spectrum that share common mass differences. These mass differences typically represent commonalities in the mechanistic provenance of the impurities and can be used to rapidly test hypotheses of formation and thus develop rapid mitigation measures. Although the method has been developed using accurate masses obtained under high-resolving power experimental conditions using a TOF instrument, it can be used under reduced resolving power conditions (e.g., unit mass resolution) suitable for widely used mass spectrometers (quadrupole and ion trap instruments). Upon rapid initial use of the approach to determine the trace impurities in a particular sample of interest, detailed investigation can be then launched to determine the exact mechanisms and sources of the novel impurities. Subsequently, common routine quantitative approaches can be developed for the targeted measurement of the determined impurities in the given system. The approach and representative application examples are described below.

Principles of Approach

The importance of the characterization of drug impurities and degradation products has been outlined in a recent critical review of the use of modern sophisticated hyphenated analytical methods [29]. Mass spectrometry is one of the main analytical tools used for the characterization of drug impurities and degradation products and is certainly very suitable for nucleic acid studies and the analysis of oligonucleotides [30].

A computer approach (ProMass; Novatia, LLC, Newtown, PA, USA) has been previously developed to facilitate the calculations of the masses in the ESI mass spectra of oligonucleotide drug systems of known sequences [31]. Similar

software capabilities are provided in the packages supplied by most major mass spectrometry instrument manufacturers [16, 32, 33]. In that manner, the existing software tools are used for the confirmation of known or expected sequences of oligonucleotides and not for unknown impurities. They require previous knowledge about the specific oligonucleotide sequence under consideration, and the specific types and chemical formulas of the expected list of potential compounds in the sample. There are many limitations using this approach. The approach developed here does not require prior knowledge about the sequences of oligonucleotides or their impurities.

The current approach is based on the fundamental assumption that peaks in mass spectra having the same exact mass difference are related by the same chemical moiety. This is consistent with the current approaches of determining the most likely formulas of chemical entities from accurately measured mass spectrometric data, within the experimental error windows of the measurements. Further, the presence or absence of the same chemical moiety from a series of molecules may reflect similarities in the mechanisms of formation of each molecule. In general, several different mechanisms can potentially lead to the formation of the same exact molecule. However, due to the complexity of our current oligonucleotide synthesis system and due to the repetitive and additive nature of the synthesis, it is not unreasonable to assume that the same composite chemical mechanism, consisting of a set of individual reactions, controls the formation of chemically related molecules at repeating synthesis cycles. In that manner, oligonucleotides of different length may share common moieties (gains or losses) of the same mass and chemical formula, produced by the same overall composite reaction mechanism. It is possible using the current method to determine the peaks of all molecules sharing the same mass difference to test the hypothesis of a common overall composite mechanism of formation by methodically changing process variable conditions or the various reactants and/or their amounts. This forms an extremely powerful approach capable of interrogating and exploring the reaction space causing the formation of the different impurities.

The current approach is directly applicable to soft ionization mass spectrometric methods (e.g., ESI, MALDI, etc.) as they are generally free of fragment ions and each peak (or set of peaks: isotopic distribution and related peaks due to charge multiplicity) in the spectrum is generally associated with the presence of a single molecule. In that case, the mass difference determined between peaks actually reflects the mass of the same chemical entity gained or lost from a series of related molecules and can be thus used directly to test hypothesis mechanisms of their chemical formation during synthesis. Moreover, the approach is even applicable to ionization methods that lead to the generation of fragment ions (e.g., electron ionization, in-source dissociation, or MS/MS). In that case, considerations generally can be made for the type of ions formed (e.g., even or odd electron ions) and the possible chemical bonding for a more guided chemical formula elucidation. Alternatively, using the current approach, the determined mass differences in the spectra

can be used to automatically differentiate the types of fragment ions formed. The scope of the current work is limited to the formation of precursor ions under soft ionization conditions (ESI) with negligible fragmentation.

Experimental

An Agilent 6224 time-of-flight (TOF) mass spectrometer coupled with an Agilent 1100 series HPLC instrument was used for the experiments (Agilent Technologies Inc., Santa Clara, CA, USA). A detailed description of the HR LC-MS systems, methods, and reagents used can be found in the [supplementary materials](#).

Methods

Data Acquisition

Data acquisition and preliminary data treatment are conducted using the Agilent MassHunter B.04.00 (B4033.1) data system operating on an HP xw4600 Workstation computer. In typical high-resolution experiments, the mass spectrometer is tuned using the Agilent Initial, Standard (~monthly), Quick (daily), and Check (daily) procedures sequentially. The approach allows for a typical resolving power greater than ~10,000 FWHM (full width at half mass). The reported resolving power by the Agilent software ranges from ~10,000 to 30,000 depending on the mass, and it generally increases as the mass increases, although in certain cases it reaches the maximum value at different masses (e.g., at m/z 112.98, resolving power of ~9000; at m/z 1933.93, resolving power of ~33,000, and at m/z 2833.87, resolving power of ~29,000, etc.). To increase the signal-to-noise ratio, a composite mass spectrum is obtained typically by averaging a selected number of scans. Peaks are of a non-symmetrical form characterized by a high mass-end tail, which becomes more pronounced as the mass increases. Despite the non-symmetric peak shape, the software is able to produce acceptable mass accuracies (e.g., <10ppm) across the mass range using the centroids of the peaks. We have found that higher mass accuracy is possible when operating the instrument at the higher resolving power setting (4GHz) even if spectra with higher signal-to-noise values are obtained at the reduced resolving power condition settings. It appears that resolution and peak shape are more important than the number of digitization points across the peak with regard to increased mass accuracy with the current Agilent algorithm. Although it is possible to rigorously determine the mass measurement precision based on peak width and intensity, here it is assumed constant.

Mass Calibration

A fifth order polynomial equation (time-to-mass fit) is used to calibrate the mass scale of the instrument using the Agilent Technologies ESI-L tuning mixture. The compounds in the mixture adequately cover the mass range of interest for our experiments (m/z 80 to 2500). The reported mass measurement

error following tuning and calibration for the compounds in the tuning mixture is ~ 1 ppm or less, across the mass range of interest. A second mass calibration is used by the continuous introduction of reference compounds to correct the initial mass calibration [34]. In principle, more reference compounds would lead to an increased mass accuracy but they could also introduce possibilities of overlapping with our sample peaks of interest. They would certainly increase the background baseline and that could make the detection of sample peaks more challenging. We have found that the use of two or three reference compounds with masses bracketing our masses of interest can provide mass accuracies of less than ~ 3 ppm across the mass range, which is acceptable for the purposes of our current experiments.

Data Treatment Steps

(1) *General Information* Pretreatment of the raw data is done using the existing routines of the MassHunter data system. The pretreated data are subsequently exported to an ASCII file using the existing exporting routines of the system. The exported files are transferred to a different personal computer for further treatment. The file contains the mass-intensity list covering the acquired mass range. Depending on the acquisition digitization rate and the data signal threshold applied, the number of data points in the file can be higher than 250,000. A computer program residing on the PC was written for the calculations. The code was developed in Absoft Pro Fortran 13.0.3, IDE ver. 2.2.

An HPLC chromatographic system coupled with a TOF-HRMS mass spectrometry system has been used to acquire the data in our experiments, but in principle the data from any chromatography system coupled to high-resolving power mass spectrometers can be used by the developed mass difference approach. In fact, the use of chromatography is optional as the averaged mass spectra from flow injection experiments are perfectly acceptable by the current method. The same is true for the use of high-resolution data. Although high-resolution can lead to the determination of the exact mass and formula of the mass difference moiety and can be used to directly rationalize the chemical changes in the structure, low-resolution data can be equally used in a semi-empirical approach to track and eventually eliminate the presence of the undesirable impurity peaks. Similarly, although deconvolution in the current version of the approach is used to determine the “zero-charge” mass of the precursor molecule, the mass differences determined in the raw ESI spectrum can be equally used in a similar semi-empirical approach to track the presence of undesirable impurity peaks. The algorithm in that case will need to account for the charge multiplicity in the peaks.

The steps in the developed approach include the acquisition of the data, the averaging and deconvolution of the mass spectra, and the generation of the mass difference matrix. Averaged spectra are obtained by the summation of several mass spectra in the run to create a composite mass spectrum.

The main benefits of the averaging process are the detection of lower abundance peaks in the spectra and the facilitation of the deconvolution process. For both cases, an increased signal-to-noise ratio is highly beneficial. Although it is possible to develop algorithms to account for the charge multiplicity of the peaks in the spectra, for simplicity we have used Agilent’s existing MassHunter ESI deconvolution routines to generate the “zero-charge” mass spectra. The mass-intensity list in the ASCII file of the averaged deconvoluted spectra is used to generate the mass difference matrix.

(2) *Deconvolution* The resolved isotope deconvolution method of the Agilent MassHunter data system was used for the charge state deconvolution of the averaged mass spectra. The method is related to Marshall’s initial ESI deconvolution algorithm that determines the charge state of the ions using the resolved isotopic peaks in high-resolution mass spectra [35]. Typically, prior to deconvolution the scaling setting of the spectra was set to off and the abundance threshold was lowered to an optimized value for the detection of the lowest abundance sample components. Trial and error was used to determine the optimal deconvolution parameters. Peak location parameters, maximum spike width was set to 2, and required valley to 0.70. Isotope grouping parameters: peak spacing tolerance was set to $0.0025m/z$ and 7.0ppm, isotope model was set to common organic molecules. The maximum charge state was set to -9 . Similar performance was obtained by a range of deconvolution parameters. The most significant parameter affecting the determination of the final number of peaks was the abundance threshold. In this work, prior to deconvolution masses are considered in terms of m/z and after deconvolution in terms of neutral (zero-charge) mass values.

(3) *Generation of Mass Difference Matrix* The deconvolution routine produces mass-intensity lists of centroided peaks reducing thus considerably the size of the original file. Typically, raw mass spectra acquired in the profile mode containing over 250,000 data points are reduced to less than 10,000–15,000 points after deconvolution and conversion to the centroid format. The data array is further reduced by applying an abundance threshold in the mass difference program. The “zero-charge” data array is processed for the determination of the mass differences. A user-defined experimental error window (e.g., 10ppm) is used for the determination of the mass differences. In principle, mass differences between peaks can be determined across the entire mass range acquired (e.g., m/z 80–2500). This would involve a fairly lengthy computation considering the relatively high upper mass of the deconvoluted oligonucleotides (e.g., >7000 Da) and the very narrow 10 (or less) ppm mass difference intervals. Inspection of the mass change during each oligonucleotide synthesis step led to the observation that the maximum change in each step is less than~

$\pm 450\text{Da}$, defined by the mass of the nucleotide with the maximum mass [e.g., mass of P(Moe G) is 419Da ($\text{C}_{13}\text{H}_{18}\text{N}_5\text{O}_7\text{PS}$); Moe: 2-O-methoxyethyl ribose]. It would be very simple to increase this to a higher number but we have found that a mass difference limit of $\pm 450\text{Da}$ accounts for the chemical changes of the molecules in our oligonucleotide systems. There is a tremendous computational benefit to restrict the calculation to a chemically justified finite mass difference range, and this is one of the most important discoveries of the current approach.

In that manner, the intensities of the peaks differing by a constant mass interval are determined by comparison of all masses in the spectra. The sum of all peak intensities is normalized to 1,000,000. Prior to the calculations, the algorithm allows for the intensity weighed combination of the peaks within a user-defined window to reduce the calculation time and eliminate possible redundant peaks produced by the Agilent deconvolution program. This user-defined mass window is based on the resolving power of the instrument used for the experiments. For example, masses produced by the deconvolution program differing by 0.1 or 0.2Da are combined into a single intensity weighed mass to reflect the actual resolving power capabilities of the instrument used for the acquisition. The mass window for the combination of the redundant masses is generally significantly greater than the experimental mass accuracy error window (e.g., 3–10 times).

The program then creates a mass difference grid (in Da or ppm) based on the resolving power of the instrument and the estimated experimental mass accuracy error in the measurement of the peaks to cover the 0 to 450Da mass difference range. In that fashion, this process creates a grid of 15,000 data points for a 0.03Da mass window, which can be reduced to 1500 for a larger window of 0.3Da to cover the 0 to 450Da mass difference range. For unit mass resolution, the total number of points drops to 450. The grid with the highest number of data points provides higher levels of resolution, but it makes the calculations considerably longer to perform. We have found that depending on the experimental needs, a 0.03Da mass window leads to an acceptably rapid computation time (i.e., <1min) for most of our applications. A 3 to 10ppm mass window is the narrowest mass window that could be used and would produce the highest number of data points for the 0 to 450Da mass difference grid, and the highest resolution for the detection of the mass differences between molecules. In that fashion, a matrix $\text{AM}(\text{I},\text{J})$ is created that contains the values of the masses in the measured spectra (I dimension) and the values of the incremental mass differences in the grid (J dimension). The masses in the J dimension of the matrix are obtained by adding the mass difference array (0 to 450Da, in 0.03 or 0.3Da increments, for a total of 15,000 or 1500 data points across, respectively) to the mass of the corresponding initial Ith peak in the spectrum. In that way, for example, the last mass in the J dimension of a peak I with a mass of 5000.00Da will be 5450.00.

Once the $\text{AM}(\text{I},\text{J})$ matrix is created, all peaks in the “zero-charge” mass spectrum are searched to determine those that

differ by a progressively increasing exact mass difference that eventually covers the entire 0 to 450Da range. The absolute mass difference between all peaks in the spectrum is used to determine those that match the exact mass difference under consideration, within the experimental mass error window. The algorithm starts by consideration of very small mass difference intervals, (e.g., 0.3, 0.6Da, etc.) and eventually reaches the highest mass difference intervals (e.g., 449.4, 449.7, 450.00). In that fashion, a new matrix $\text{AY}(\text{I},\text{J})$ is created that contains the intensities of the corresponding masses in the $\text{AM}(\text{I},\text{J})$ grid. Although it is possible to determine two matrices, for the peaks differing by either positive or negative mass difference value, the absolute mass difference has been used for simplicity and rapidity. It has to be noted though, that the plots and results presented here reflect positive or negative mass differences (gains or losses). However, if required, the program can generate both mass difference matrices. The intensities of the masses in the grid are summed across the I or J dimension to generate intensity summaries as a function of the mass difference or the mass range dimensions. These summaries provide very useful information about all impurities in the samples, especially for rapid exploratory comparisons between samples.

Results and Discussion

Purified Drug Substance with and without DMT (4,4'-Dimethoxytrityl) group: Early Eluting Peaks

As a first example of the application of the approach to the determination of trace impurities in oligonucleotide drug systems without prior knowledge of the expected impurities or their sequences, we have chosen a purified drug substance with the DMT group on and the corresponding sample with the DMT group off. The LC-TOF-HRMS total ion chromatograms (TIC) of the two samples are shown in Fig.1. The top panel (Fig.1a) shows the TIC of the purified oligonucleotide drug substance with the DMT group on, and the bottom panel (Fig.1b) the corresponding TIC of the detritylated substance. It is interesting to notice the two partially resolved chromatographic peaks on the right side of the upper panel in Fig.1 corresponding to the two DMT diastereoisomers. These are due to the presence of the chiral phosphorous center, and are only partially resolved. The presence of the aromatic DMT group on the molecules aids in the chromatographic separation by interacting with the hydrophobic C18 group of the stationary phase of the column. We have obtained a better separation using a phenyl hexyl HPLC column because of the stronger interactions with the DMT group of the oligonucleotide, but here we have conducted all HPLC experiments using a C18 HPLC column to allow for direct comparisons between samples. The number of the diastereoisomers is very large (2^{n-1}) [36], where n is the length of the oligonucleotide, which makes it exceedingly difficult [37] and perhaps unnecessary to separate each fully by current analytical techniques. The two broad, unresolved peaks of the diastereoisomers reflect mainly the

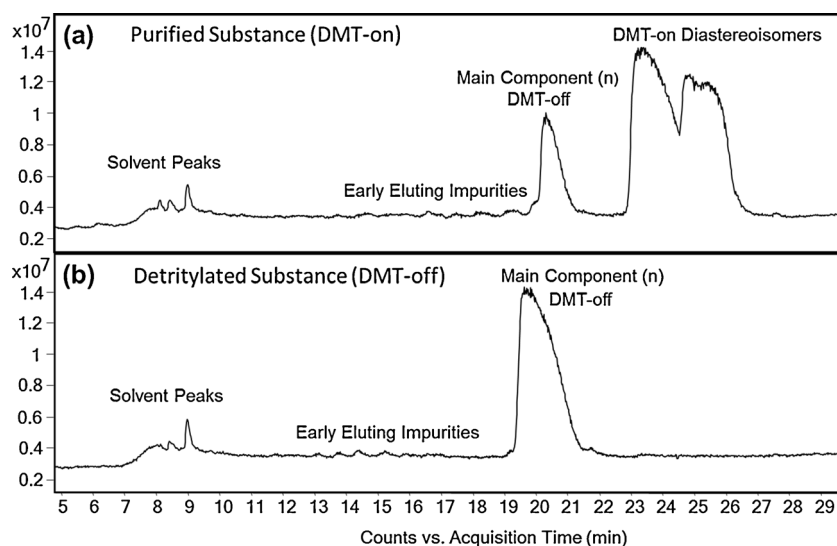


Figure 1. Determination of trace impurities in oligonucleotide drugs. LC-TOF-HRMS total ion chromatograms (TIC) of: (a) purified drug substance with 4,4'-dimethoxytrityl (DMT) group-on; (b) detritylated drug substance (DMT group-off)

separation of the two groups of molecules containing the DMT group on the last nucleotide. A portion of the purified substance sample in Fig. 1a is due to the main (n) component with the DMT group off. This reflects the amount of the material that has undergone detritylation prior to the final acid treatment leading to the complete detritylation of the sample (Fig. 1b). Part of that amount may also be generated during the handling of the analytical sample and the time it remained at room temperature prior to analysis. Generally, the initial amount of the main component (n) in the DMT-on sample is low (e.g., less than ~3%–10%).

It is important to notice the absence of abundant impurity peaks in both Fig. 1a and b. This is primarily due to the low concentration of the impurities as well as the high level of the background baseline signal (chemical noise). The main sources of the background baseline are due to the presence of the second calibration reference compounds continuously infusing into the ionization chamber, and due to the ionization of components from the mobile phase. Some components in the solvents used to prepare the mobile phase are eluting continuously and they contribute to the chemical noise of the system. A few components from the mobile phase elute as distinct peaks. These mobile phase peaks are labeled as “solvent peaks” in the panels of Fig. 1. We have observed that most chromatographic peaks related to the mobile phase are characterized by mass spectrometric peaks with m/z values lower than ~400–500. In that way, it is possible to apply common LCMS algorithms to reduce the contribution of the chemical noise by subtracting the unchanging signal attributable to the continuous elution of the mobile phase components and by excluding mass spectrometric peaks with m/z values lower than ~400–500. Further chemical noise reduction is possible by common subtraction of blank runs. In the present work, the entire mass range acquired has been considered in all experiments and data

treatment steps in order to ensure the detection of all possible trace components.

Average mass spectra were obtained by the summation of all components eluting between approximately 4min and the left base of the main component (n) peak (at ~20min, and ~19min, respectively) for the purified substance with the DMT group on and the corresponding detritylated substance. The most abundant peak in the purified substance spectrum is at m/z 1465.8691. All other peaks in the spectrum are lower than ~50% of the main peak. The spectrum is a composite spectrum of all components eluted within the retention time interval used for the integration of the summed mass spectrum. It also contains the signal attributable to peaks of the same precursor molecule but carrying a different number of charges. Because of this, the spectrum is difficult to interpret directly as it contains a series of different ions and its deconvolution is required to simplify subsequent data treatment. The most abundant peak in the detritylated substance is at m/z 1329.6986. However, there are a few other peaks in the composite mass spectrum of considerable abundance. For example, peaks at m/z 1169.1875, 1465.8711, and 1547.7451 are all at higher than ~50% of the most abundant peak in the spectrum. The two spectra indicate significant differences in the early eluting components present in the two oligonucleotide samples. It is interesting to notice, however, that ions at the same m/z values also, but of different relative abundance, are present in both samples, indicating possible chemical similarities in the samples as well (e.g., m/z 1169.1875, 1465.8711, 1880.2798, 2199.8203, etc.).

The deconvoluted mass spectra of the composite spectra of the early eluting components in the purified substance with the DMT group on and the corresponding detritylated substance with the DMT group off are shown in Fig. 2a and b, respectively. These spectra were obtained by direct deconvolution of

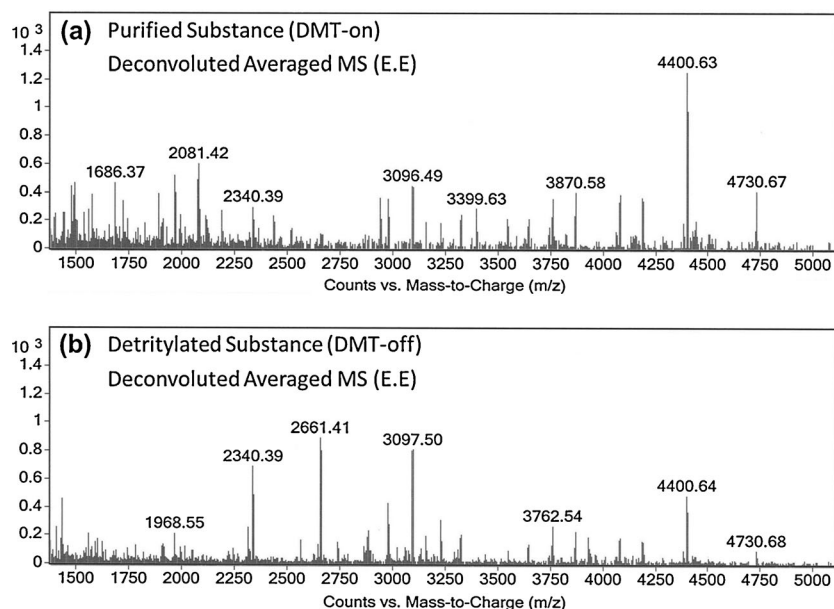


Figure 2. Deconvoluted averaged mass spectra of the early eluting (E.E.) impurities: **(a)** purified drug substance (DMT-on); **(b)** detritylated drug substance (DMT-off)

the averaged mass spectra using the existing MassHunter deconvolution routine. The resolved isotope parameters outlined in the Deconvolution section above were used for the deconvolution. The deconvoluted spectra shown in Fig. 2a and b are the “zero-charge” state spectra obtained by the identification and addition of the signal of all peaks of the same precursor molecules but carrying a different number (z) of negative ion charges (e.g., $z=-2$, -3 , -4 , etc). In that way, the peak at 4400.63Da in the spectrum of Fig. 2a is generated from the peaks at m/z 2199.8 ($z=-2$), and m/z 1465.9 ($z=-3$) in the averaged spectrum. At a first approximation, there is no accounting for the response factors of the peaks obtained at different ion charge states. Very importantly, the deconvolution routine greatly simplifies the composite spectrum by combining the peaks due to the charge multiplicity into a single peak, and at the same time corrects for its mass to produce the “zero-charge” state mass of the precursor molecule. The procedure is successfully applied to the entire spectrum, and we have found it to be able to reliably deconvolute trace components (e.g., at $\sim 1\%$ – 3% or higher of the main peak in the early eluting spectrum) in the spectra of a wide range of samples. We conducted initial evaluation of the MassHunter deconvolution capabilities by manual generation of the “zero-charge” peaks and comparing them with those generated by the Agilent system.

The contour plot of the early eluting impurities in the purified drug substance sample with the DMT group on is shown in Fig. 3a. The abundances of the masses of the peaks detected in the deconvoluted spectrum (Fig. 2a) are plotted against the mass differences between the observed peaks and other peaks present in the spectrum. The mass difference was obtained by considering only the absolute difference between masses. It includes both positive and negative mass differences.

OriginPro 9.1.0 (OriginLab Corp., Northampton, MA, USA) was used to generate the contour plots. The size of the round marker in the plot is associated with the relative abundance of the peaks in the spectra. We found that taking the square root or the logarithm of the peak abundance generates contour plots with more pronounced contrast for better visualization, but we have kept the linear scale in the plots of the current work for simplicity of comparisons. The concept of the mass difference plot can be illustrated by consideration of the most abundant peak in Fig. 2a (at m/z 4400.63). The highest abundance of this peak (m/z 4400.63) is depicted in Fig. 3a at the intercept of the plot at 4400Da (Mass-Scale) and 0Da (Mass Difference-Scale). Peaks detected differing by a given mass difference are assigned the same peak intensity as the initial/precursor peak at 4400Da, reflecting that all detected peaks with a given mass difference are in relation to that precursor peak. In that regard, it is the intensity of the precursor peak (at 0Da mass difference) that is drawn in the plot. The plot represents the fact that there is another peak present in the spectrum that differs by a given mass with regard to the initial/precursor peak. So the size of the markers in the plot represents the intensity of the precursor peak. It is also possible to plot instead the intensity of the peak found to form the detected mass difference, but this representation is not used in this report. The current focus is on the subsequent exploratory capabilities of the approach, which is the detection of the precursor peaks that lead to the observation of a mass difference, as the mass differences are related to characteristic chemical mechanisms leading to the formation of structures with specific masses. In its general case, the approach can relate the mass differences using the intensity of the precursor, the intensity of the actual peak found in the spectra, or a weighed combination of the two. In that way, in the current scheme of presentation, the contour profile plot

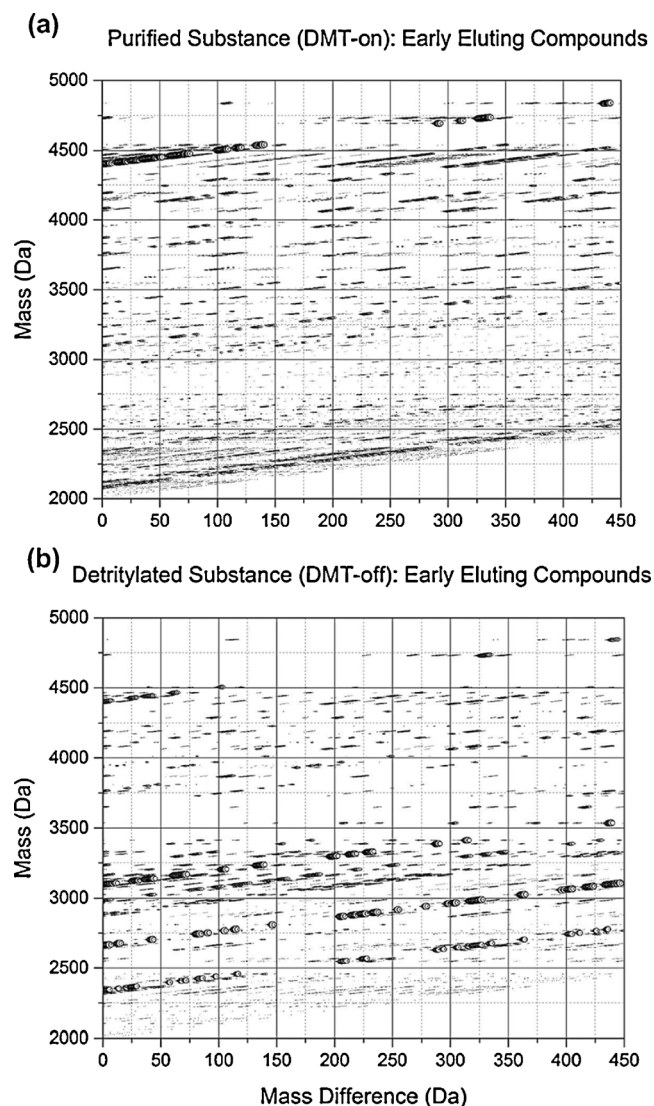


Figure 3. Contour (mass difference vs mass range) plot of the early eluting impurities: (a) purified drug substance sample (DMT-on); (b) detritylated drug substance sample (DMT-off)

consists of sloped parallel lines in a mass range by mass difference plane, intercepting at different positions of the precursor mass range axis. Currently, the size of the markers reflects the overall significance of the precursor mass and the detection of other peaks in the spectrum that forms a pair of masses consistent with a given mass difference. The presence of pronounced signal in the contour plot indicates the presence of peaks at different mass differences. The formation of visible sloped parallel lines in the contour plots indicates the presence of other related peaks in the spectrum and in the presence of intense spots, the presence of peaks with masses of specific mass differences. In that way, an intense spot, shown at around 319Da in Fig.3a, indicates that a peak is present in the spectrum that has gained or lost a 319Da moiety. Although there are many other sloped lines in the plot, only a few show characteristic spots at the 319Da mass difference scale intercept. The

approach allows for the correlation and discovery of such mass difference commonalities between different precursor masses.

The utility of the contour plot is further illustrated by consideration of the plot in Fig.3b, which is produced from the data in the deconvoluted averaged mass spectrum shown in Fig.2b. In similarity to the contour plot shown in Fig.3a, the plot in Fig.3b depicts a series of sloped parallel lines intercepting at different mass range (Da) values, indicating the masses of the precursor masses. Characteristically intense markers reflecting the intensities of the precursor masses are shown in the plot, but this time the lines and spacing of the markers on the parallel lines are considerably different than those present in the lines of the plot in Fig.3a. This is in accordance with the expected differences observed in the deconvoluted spectra of Fig.2, reflecting the actual differences in the types and amounts of the trace components present in the purified oligonucleotide substances with and without the DMT group. Similarly to the contour plot in Fig.3a, the lines and intensity of the markers in plot 3b can be used to detect commonalities in the mass differences and, hence, commonalities in the gains or losses of the same chemical moieties from related precursor structures.

Most interestingly, the data produced by the summation of the total ion current of the peaks detected at each mass difference value as a function of the mass difference (Da) generate characteristic summary plots that can be used for the determination of the most abundant mass differences in the spectra of the samples and for the rapid comparison of the types/families of impurities in the different samples. There are characteristic mass difference values where the signal maximizes for one sample and not for the other. For example, the total ion current detected plotted as a function of the negative mass difference in Fig.4a produces different plots for the two DMT-on and DMT-off samples, respectively (Fig.3a and b). Peak clusters of the mass difference plot in Fig.4a have been labeled with the names of possible moieties consistent with the masses of nucleotides used in the synthesis (Moe: 2'-O-methoxyethyl-2'-deoxyphosphorothioate; d: 2'-deoxyribose; A: adenine; C: cytosine; T: thymine; G: guanine). In addition to the identification of common nucleotide substructures (e.g., P(Moe G), P(Moe A), P(dG), P(Methyl C/T), etc.) the program has identified the presence of structures that gained or lost the DMT substructure. Additionally, there are mass difference peak clusters in the mass difference region of the smaller masses of the plot that are consistent with known mechanisms of depurination and other impurities. Some of these clusters (e.g., depurination peaks at 151, 135, 117Da are more pronounced in the detritylated substance) indicating a higher tendency for this sample to be associated more with this type of reactions.

We have discovered three characteristic mass difference regions in the summary plots of oligonucleotide drug impurities (e.g., Fig.4a). The first region is characteristic of mass differences with values lower than ~160Da, and these are primarily related to gains or losses of relative small chemical moieties. This region contains mass differences that have mainly been historically identified and are related to specific

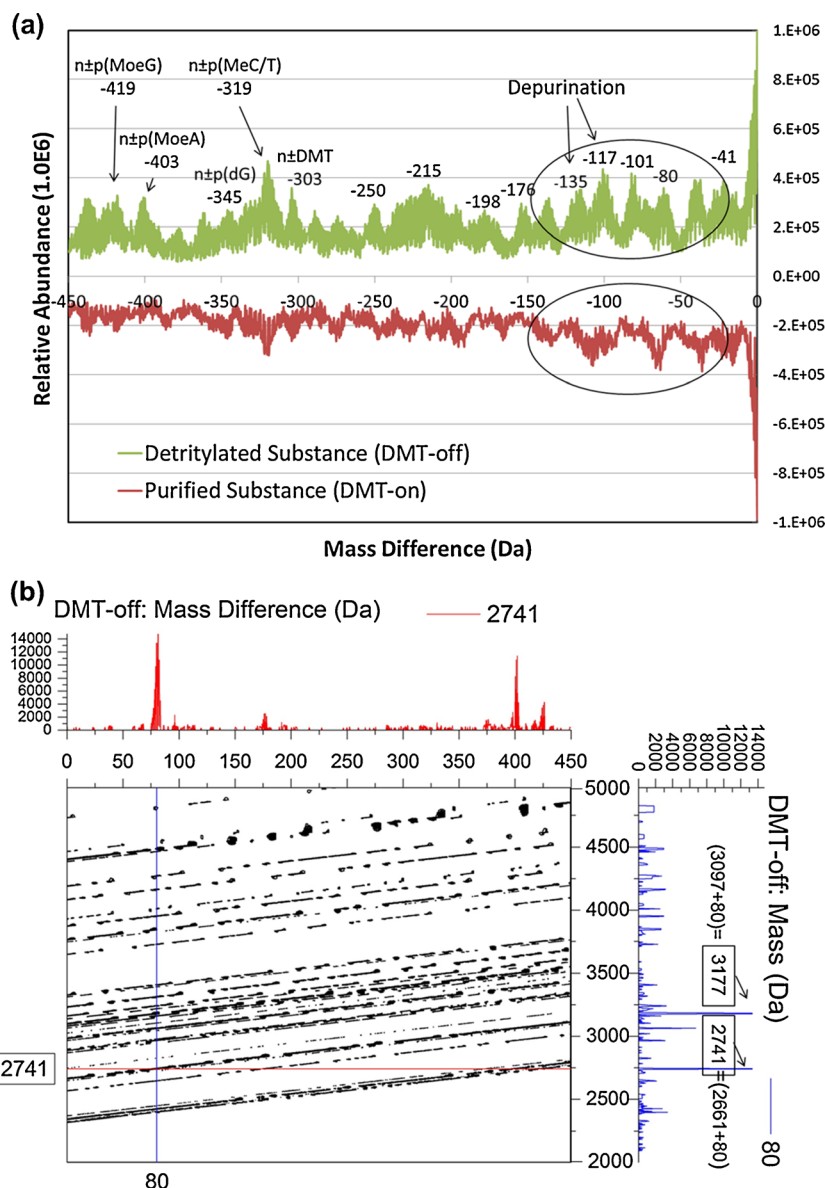


Figure 4. (a) Comparison of total detected impurity signal in the purified (DMT-on) and the detritylated (DMT-off) drug substance samples. The total ion current of the impurities in each sample is plotted against the mass difference range 0 to 450 Da. (b) Profile contour plot of the early eluting impurities in the detritylated drug substance sample (DMT-off). The spectrum on the right side of the plot (blue line) depicts the masses of all impurities in the sample that share a common mass difference of 80 Da. The peaks in this spectrum are expected to reflect information about impurities formed by the same reaction mechanism. The spectrum at the top of the plot (red line) depicts the mass differences of all impurities in the sample related to the precursor mass at 2741 Da. The peaks in this spectrum are expected to reflect information about all reaction mechanisms related to the same precursor mass

chemical moieties and the reactions leading to their formation. The second and most obvious mass difference region covers the range of the nucleotide related masses added or lost at each synthesis cycle, and they range from ~ 290 to ~ 420 Da. A slightly less well understood mass region involves the mass difference range extending between ~ 160 and ~ 290 Da. Although it is not the purpose of the current paper to discuss the nature of the chemical species and reactions involved in the generation of impurities with these characteristic mass differences, this observation indicates the need for future work to fully characterize the sources of the impurities in this middle

mass difference window. Equally, future work needs to be conducted to investigate mass differences extending beyond 450 Da.

The current approach naturally leads to the generation of a profile contour plot, for the first time, for the presentation and interactive exploration of the mass difference data. The profile contour plot generated by the data in the DMT-off deconvoluted mass spectrum is shown in Fig. 4b. The basic 2-D contour plot is the same as the one shown in Fig. 3b. The main difference is that the profile contour plot allows for the interactive interrogation of the data and the extraction of the

spectra in the y-axis (mass range) or x-axis (mass difference) dimensions. In that way, the mass range extracted spectrum for the 80Da mass difference components, on the right side of the plot in Fig.4b (blue line), shows the presence of two peaks at ~3177 and 2741Da in the deconvoluted spectrum, indicating that these peaks share a common chemical gain or loss with a characteristic mass difference of 80Da. Indeed, inspection of the precursor deconvoluted mass spectrum (Fig.2b) reveals the presence of two intense peaks at ~3097 and 2661Da, which differ by the same 80Da mass difference from the detected peaks in the mass range extracted spectrum of the profile contour plot on the right side of Fig.4b (blue line). In that way, the mass range extracted spectrum can rapidly reveal all peaks in the spectrum, which may be formed by the same composite chemical mechanism in the repeating oligonucleotide synthesis cycles, in this case the formation of two impurity structures both containing a moiety of 80Da. The presence of this 80Da peak cluster was initially detected in the summary plot shown in Fig.4a. The formation of such impurities are consistent with previously published work on the composite chemical mechanisms leading to the generation of the 80Da mass difference (i.e., C_5H_4O : 80.0262Da) moiety-containing structures attributable to the reaction of cytosine-containing sequences with depurinated oligonucleotides [26]. The rapid detection of the presence of this type of mechanism clearly demonstrates the very powerful analytical capabilities of the current approach to identify impurities related by the same mechanism of formation.

Additionally, the plot in Fig.4b displays the extracted mass difference spectrum related to a selected impurity peak in the spectrum at 2741Da. The generated mass difference spectrum at the top of the Fig.4b (red line) shows the presence of peaks differing from the peak at 2741 by an absolute mass difference of ~80, ~401, and weaker peaks at ~425, and 175Da. Indeed, there are intense peaks in the spectrum (Fig.2b) at 2661 and 2340Da, which are related to the peak at 2741Da by 80 and 401Da, respectively. The formation of each of these peaks leads to the gain or loss of characteristic mass moieties represented by the observed mass differences in the mass difference extracted spectrum (red line). The presence of each of these mass difference peaks indicates different possible composite mechanisms of formation. For example, the 80Da moiety has already been associated with the reaction of cytosine-containing sequences with depurinated oligonucleotides. The structure of the peak with ~401Da mass difference may be due to the gain/loss of the P(MoeA) moiety (403Da) or a combined structure containing the 80Da and ~321Da (pT) substructures. Further investigation would be required using the exact mass differences to elucidate the substructures and nature of the composite mechanism leading to the formation of the observed ~401Da mass difference. These considerations clearly

illustrate the powerful abilities of the approach to rapidly identify all types of different mechanisms of impurity formation related to specific precursor masses.

Crude Drug Substance Prepared Under Excess and No Capping Experimental Conditions: Early Eluting Peaks

As a second example to illustrate the capabilities of the developed approach, we have analyzed a set of two crude oligonucleotide samples prepared under severe experimental conditions: (a) with excess capping and, (b) with no capping at all. The two samples are characterized by distinctively different chromatographic profiles. Excess capping leads to significantly higher amounts of smaller oligonucleotide impurities in comparison to no capping. Elucidation of the mechanisms is beyond the scope of the current work. Instead, the data are used to further demonstrate the capabilities of the developed approach in the characterization of complex oligonucleotide mixtures.

The deconvoluted mass spectrum corresponding to the crude drug sample prepared under no capping experimental conditions was characterized by a predominantly intense peak (100%) at ~7500Da. The remainder of the peaks in the spectrum displayed peaks with relative intensities lower than ~10%–20% of the main peak. In contrast, the deconvoluted mass spectrum corresponding to the crude drug sample prepared under excess capping conditions was characterized by many intense peaks at the lower mass ranges. The most abundant peak in the spectrum was at 2295Da, followed by other progressively lower intensity peaks at 1975, 2614, 3256, 4250Da, etc. The high mass peak at ~7500Da related to the one in the no capping sample was at ~30%–40% of the most intense peak of the spectrum at 2295Da.

The summary plots for the two samples generated by the summation of the total detected impurity signal as a function the mass difference values are shown in Fig.5a. The clusters of the mass difference peaks for the sample produced under no capping conditions (green line) are of relatively lower abundance with higher abundance peaks observed mainly at the lower mass ranges (i.e., < ~60Da). Considerably higher abundance peak clusters were observed in the trace of the crude sample obtained under excess capping conditions (red line). Although the two samples share peak clusters at common mass difference values, the chart shows that the excess capping sample contains more peak clusters and of higher relative abundance.

It is important to keep in mind that some of the detected impurities may be adducts generated in the ESI process or due to other background impurities in the mobile phase system. The background impurities can be determined by the analysis of blank samples. The presence of possible adducts can be determined from their known mass differences. More complicated systems may involve the reaction of mobile phase reagent impurities with selected oligonucleotide sequences, and more work needs to be done to address these possibilities. The

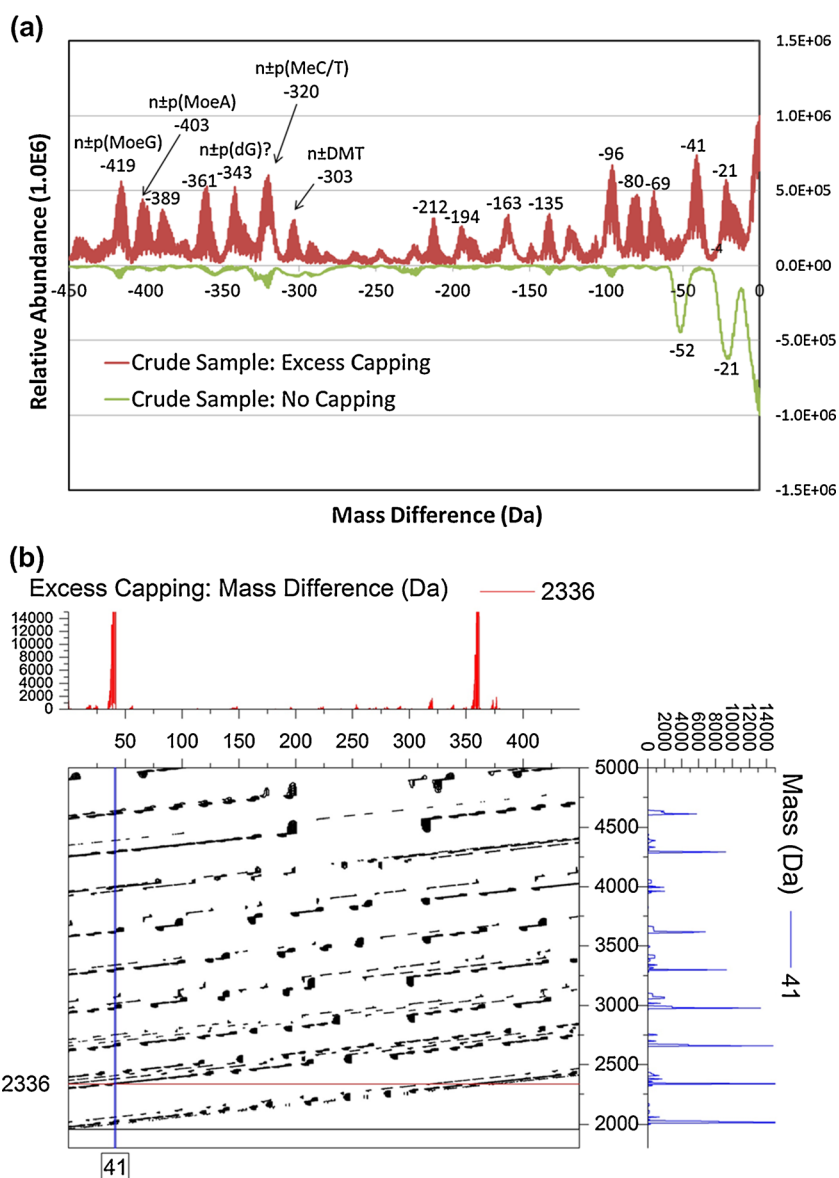


Figure 5. (a) Comparison of total detected impurity signal in crude oligonucleotide drug substance samples prepared under excess and no capping reaction conditions, respectively. (b) Profile contour plot of the early eluting impurities in the crude oligonucleotide drug substance sample prepared under excess capping reaction conditions. The spectrum on the right side of the plot (blue line) depicts the masses of all impurities in the sample that share a common mass difference of 41 Da. The spectrum at the top of the plot (red line) depicts the mass differences of all impurities in the sample related to the precursor mass at 2336 Da

approach developed here can be rapidly applied to determine changes in known systems or evaluate completely unknown systems to determine clusters of potentially novel mass differences.

The profile contour plot of the early eluting impurities in the crude oligonucleotide drug substance prepared under excess capping reaction conditions is shown in Fig. 5b. The figure displays the sloped parallel lines tracking the precursor molecules and the possible presence of related peaks in the spectrum sharing the same common mass differences. As an illustration, we observed a significantly abundant peak cluster of 41 Da in the summary mass difference plot of the excess capping sample in Fig. 5a. It would be interesting to determine the possible

peaks in the spectrum that share the same common mass difference of 41 Da. The mass range extracted spectrum presented on the right side of Fig. 5b (blue line) shows that a significant number of peaks in the initial precursor mass spectrum share the same common 41 Da moiety. It appears that this chemical substructure is formed by the same composite chemical mechanism and is associated with the formation of several impurities in the sample as is shown from the detected peaks at 2016, 2336, 2656, 2976 Da etc. Intense peaks of the corresponding precursor masses are present in the initial deconvoluted spectrum (i.e., 1975, 2295, 2615 Da, etc.). It is quite possible for the same moiety to be added on each oligonucleotide sequence formed at each synthesis cycle, under the

extreme excess capping conditions used. The current method permits the rapid determination of all precursor molecules related to the particular mechanism leading to the formation of the 41Da mass difference series of impurities. Very interestingly, the same 41Da mass difference peak cluster is absent (or of very low abundance) from the spectrum of the sample prepared under no capping conditions, directly indicating that the formation of the 41Da series of impurities is primarily due to capping. The approach can be used in a rapid fashion and without the need of prior knowledge of sequences or mechanisms to guide the determination of experimental conditions leading to the elimination of the 41Da impurity, and in that manner other observed undesirable impurities. Close inspection of the high resolution TOF spectra showed an exact mass difference of 41.0265 and isotopic peak distributions consistent with the C_2H_3N substructure. Studies to elucidate its mechanism of formation are underway.

The spectrum (red line) on the top part of Fig.5b displays the mass difference extracted spectrum for the selected precursor peak with mass 2336Da. As the chart shows, the 2336Da peak is associated with an intense peak in the spectrum at 2295Da, characteristic of the 41Da, C_2H_3N moiety. In addition, a second intense mass difference peak is observed at ~361Da (top spectrum in Fig.5b). The observation of the peak at this mass difference may be associated with the mechanisms of formation of impurities with combined substructures consisting of the 41Da and the 319/320Da PT or P(methyl C) nucleotide moieties. Further work is needed to elucidate the nature and origin of the peak observed at the ~361Da and others at ~381, 212, 194, 163, 96, 69Da, etc. mass difference values.

Assignment of Reaction Types

The mass values of the identified mass difference peak clusters obtained using the current approach (e.g., summary data from Fig.4a, Fig.5a) can be summarized and used as references to assign the types of mechanisms involved in the characterization of future, completely unknown oligonucleotide drug systems. The approach can rapidly determine the novel types of impurities and mechanisms and point to required mechanism elucidation studies and experimental control and mitigation measures for their elimination. The mass differences (Da) of several known oligonucleotide drug impurities from the main oligonucleotide precursor mass (n) at different electrospray ionization charge states (z values) are given in Table1. The table summarizes the types of impurities and the mass differences determined from the chemical formulas of the known nucleotide substructures. The table can be used to associate the observed mass differences between peaks in the spectra to the types of impurities and mechanisms involved in their formation. The table is not complete and more work needs to be done to incorporate the mass differences resulting from other previously unknown mechanisms of impurity formation as they are being determined using the approach developed here.

Further Considerations and Benefits

To account for the lack of calibration standards in the higher mass ranges and the considerable reduction [38] of resolving power as a function of increasing mass using Fourier transform ion cyclotron resonance (FT-ICR) instruments, “formula extension” approaches [39–43] have been developed to aid in the assignment of the most likely candidate chemical formulas from the experimental data. Using these approaches, the determination of the formulas of compounds with masses higher than approximately 500Da is based on finding peaks in the spectra differing from the lower peaks by mass differences consistent with those in a pre-established list of functional groups (e.g., gain or loss of CH_2 , H_2 , CO_2 , etc.). This approach has been particularly useful for applications involving the analysis of complex samples such as crude oil and natural organic matter (NOM), which are characterized by high periodicity. However, the assumed presence of these additional functional groups may not be always valid in the higher masses where new families of compounds may be present, in addition to the possibility of inadequate resolving powers at the higher masses, and thus elaborate procedures must be devised to assess the reliability of the approaches to automatically assign elemental formulas to unknowns [43]. The most important limitation of these approaches lies in the use of non-exhaustive lists of possible functional groups present in the samples. In that way, compounds with unexpected functional groups cannot be determined using these approaches [39]. A significant improvement was made by Kunenkov et al. [40] by the introduction of a statistical approach to determine the high-mass building blocks based on the repetitive mass difference patterns in the spectra. The approach alleviates the pitfalls of the previous methods and is related to the method developed here. However, the probabilistic approach does not consider the ion abundances in the calculations, which can significantly affect the results, especially in the case of determination of low abundance impurities. The calculation also appears to be limited to monoisotopic peaks, which can be problematic for high mass compounds such as oligonucleotides with masses in the range of ~8000Da with hundreds of carbon atoms and wide isotopic distribution envelopes. Although elemental composition assignment was not the purpose for the development of the current approach, it has the benefits of the a priori determination of the mass differences of the components in the spectrum, and it can certainly be used for such applications. Consideration of the ion abundances and the ability to include the isotopic distributions in the calculations are additional significant advantages of the method as illustrated below.

The concept of spectral autocorrelation [44] has been applied to the data of the ESI spectrum obtained under excess capping conditions. Autocorrelation has been found to be useful in describing periodic patterns in high resolution mass spectra of complex samples such as petroleum and synthetic polymers [45]. Figure6 displays the expanded low-mass-region

Table 1. Mass Differences (Da) of Common Oligonucleotide Drug Impurities from the Main Mass (n) at Different Electrospray Ionization Charge States (z)

Impurity type	Mass difference (Da) at charge states (z)			
	z=-4	z=-3	z=-2	z=-1
n-p(MoeG)	-104.77	-139.34	-209.53	-419.07
n-p(MoeA)	-100.77	-134.02	-201.54	-403.07
n-p(Moe methyl U)	-98.52	-131.02	-197.03	-394.06
n-p(Moe methyl C)	-98.27	-130.70	-196.54	-393.08
n-p(dG)	-86.26	-114.72	-172.52	-345.03
n-p(dA)	-82.25	-109.39	-164.50	-329.03
n-pT	-80.01	-106.41	-160.01	-320.02
n-p(methyl C)	-79.76	-106.08	-159.52	-319.04
3'-Terminal phosphorothioate monoester (3' TPT)	-74.28	-98.80	-148.57	-297.13
Depurination species (loss of guanine)	-37.76	-50.22	-75.53	-151.05
Depurination species (loss of adenine)	-33.76	-44.90	-67.53	-135.05
Depurination species (loss of guanine with water added)	-33.26	-44.24	-66.52	-133.04
Depurination species (loss of guanine with MeOH added)	-29.76	-39.57	-59.51	-119.02
Depurination species (loss of adenine with water added)	-29.26	-38.92	-58.52	-117.04
n-108 Da	-27.00	-35.91	-54.01	-108.01
Depurination species (loss of adenine with MeOH added)	-25.76	-34.26	-51.51	-103.03
2' O-methyl species	-11.01	-14.64	-22.01	-44.03
(P=O)2	-7.99	-10.62	-15.97	-31.95
Monophosphate-diester (P=O)1	-3.99	-5.31	-7.99	-15.97
3-(2-Cyanoethyl)thymidine-modified species (CNET)	13.26	17.63	26.51	53.03
Ethylenephosphorothioate diester species (EPD)	34.99	46.54	69.99	139.97
Trichloroethanol-modified species (TCE)	36.98	49.18	73.95	147.91
n+p(methyl C)	79.76	106.08	159.52	319.04
n+pT	80.00	106.41	160.01	320.02
n+p(dA)	82.26	109.41	164.52	329.04
n+p(dG)	86.26	114.72	172.52	345.03
n+p(Moe methyl C)	98.27	130.70	196.54	393.08
n+p(Moe methyl U)	98.51	131.02	197.03	394.06
n+p(MoeA)	100.77	134.02	201.54	403.07
n+p(MoeG)	104.77	139.34	209.54	419.07

Moe=2'-O-methoxyethyl-2'-deoxyphosphorothioate; d=2'-deoxyribose; A=adenine; C=cytosine; T=thymine; G=guanine; U=uracil

of the autocorrelation function for the mass lag range 0 to 450Da. The OriginLab program was used for the autocorrelation calculations with 0.6 sampling interval. The species with mass lag of 320Da was found to have the highest autocorrelation

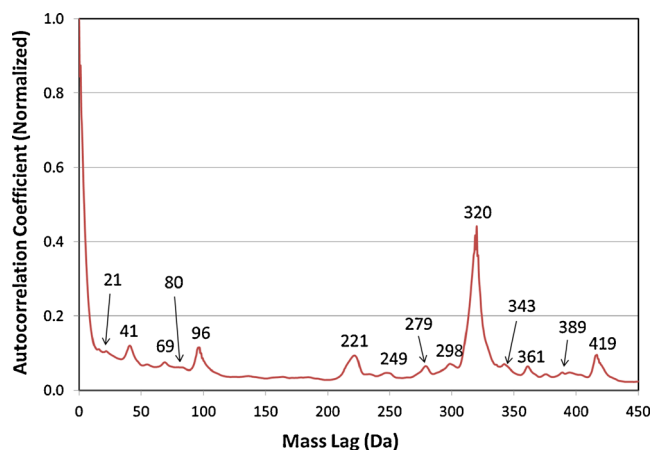


Figure 6. Low-mass-region expanded view of autocorrelation function: autocorrelation coefficient (normalized to 1.0) shown for the 0 to 450 Da mass lag of the crude oligonucleotide drug substance sample prepared under excess capping conditions

coefficient, followed in sequence by those of 221, 419, 96, and 41Da. Autocorrelation was able to determine several of the significant mass differences in the spectrum; however, several of them were barely detected (e.g., 21, 69, 80, 343, 389, etc.). The method developed here produced a significantly more pronounced summary of the mass differences and led to the detection of a much larger data set of mass difference groups (Fig. 5a), indicating the significance of including the ion abundance in the calculations. In addition, the availability of the isotopic distributions allows further verification of the determined exact masses of the differences and possibilities for identification of ion species differing by small mass intervals.

Conclusion

A novel approach has been developed that facilitates the determination of relationships between low concentration components (i.e., oligonucleotide drug impurities) in complex analytical systems. The approach is based on the use of the accurate mass peak differences obtained from high resolution electrospray ionization (ESI) mass spectra.

Different sets of oligonucleotide drug substances were used to illustrate the powerful capabilities of the new approach.

Contour plots were created by plotting the intensities of the mass range by mass difference matrices. The data produced characteristic sloped parallel lines with mass range axis intercepts reflecting the precursor oligonucleotide masses in the deconvoluted initial mass spectrum. The presence of peaks at different mass differences in the contour plot is identified by the presence of intense signal at particular mass difference values. Commonalities in the mass differences identified in the contour plots reflect commonalities in the gains or losses of the same chemical moieties from related precursor structures. Very useful characteristic summary plots were produced by the summation of the total ion current of the peaks detected at each mass difference value as a function of the mass difference (Da). These summary plots can be used for the rapid determination of the most abundant mass differences in the spectra and thus for the rapid comparison of the types/families of impurities in different samples. The mass difference peak clusters generated in the summary plots were found to maximize at values characteristic of known oligonucleotide moieties.

In this work, for the first time, we have developed a unique profile contour plot that allows for the interactive interrogation of the mass range by mass difference data matrix. The profile contour plot permits the identification of all mass difference relationships between peaks and generates extremely valuable information about all sample impurities that share a common mechanism of formation, and all possible mechanisms of formation linked to a selected precursor molecule. In that way, the approach developed in the current work can be used to rapidly classify the types of impurities in oligonucleotide and other sample systems of interest, and directly assist in the elucidation of mechanisms of formation, and, most importantly, to guide the development of manufacturing process measures for impurity elimination.

The approach can be applied to as narrow or as wide a mass range desired by the user and, in principle, can be modified to treat mass lists generated by 2-D (m/z , RT) chromatographic peak detection algorithms instead of raw spectra. This could reduce possibilities of mass interference and obtain information about isomers, provided the components are chromatographically resolved.

The method developed here presents an additional and complementary approach to the conventional methods used for unknown impurity determination in complex systems. Although developed for impurities in oligonucleotide drugs, we believe that is equally applicable to determine trace levels of impurities in any other similarly complex system of interest. Most importantly, impurity determination is exhaustive and only limited by the detection limits of the instrumentation used.

Possible future applications of the method may include analysis of MS-MS data (e.g., determination of related compound classes in complex data from constant neutral loss differences). Also, the advanced fingerprinting capabilities of the contour and summary plots may find applications in many manufacturing areas for quality control and in isolating subtle

differences between complex samples, for example lipid profiles or polymer blends.

References

1. Sharp, P.A.: The centrality of RNA. *Cell* **136**, 577–580 (2009)
2. Crooke, S.T., Vickers, T., Lima, W., Wu, H.: Mechanisms of Antisense Drug Action, an Introduction. In: Crooke, S.T. (ed.) *Antisense Drug Technology: Principles, Strategies, and Applications*, 2nd edn, p. 3. CRC Press, Boca Raton, FL (2008)
3. Bennett, C.F., Swayze, E.E.: RNA Targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annu. Rev. Pharmacol Toxicol* **50**, 259–293 (2010)
4. Verma, S., Eckstein, F.: Modified oligonucleotides: synthesis and strategy for users. *Annu. Rev. Biochem.* **67**, 99–134 (1998)
5. Eckstein, F.: Phosphorothioate oligodeoxynucleotides: What is their origin and what is unique about them? *Antisense Nucleic Acid Drug Dev.* **10**, 117–121 (2000)
6. Capaldi, D.C., Scozzari, A.N.: Manufacturing and Analytical Processes for 2'-O-(2-Methoxyethyl)-Modified Oligonucleotides. In: Crooke, S.T. (ed.) *Antisense Drug Technology: Principles, Strategies, and Applications*, 2nd edn, p. 401. CRC Press, Boca Raton, FL (2008)
7. Gilar, M., Belenky, A., Smisek, D.L., Bourque, A., Cohen, A.S.: Kinetics of phosphorothioate oligonucleotide metabolism in biological fluids. *Nucleic Acids Res.* **25**, 3615–3620 (1997)
8. Srivatsa, G.S., Klopchin, P., Batt, M., Feldman, M., Carlson, R.H., Cole, D.L.: Selectivity of anion exchange chromatography and capillary gel electrophoresis for the analysis of phosphorothioate oligonucleotides. *J. Pharm. Biomed. Anal.* **16**, 619–630 (1997)
9. Bergot, B.J., Egan, W.T.: Separation of synthetic phosphorothioate oligodeoxynucleotides from their oxygenated (phosphodiester) defect species by strong-anion-exchange high-performance liquid chromatography. *J. Chromatogr.* **599**, 35–43 (1992)
10. Cui, Z., Theruvathu, J.A., Farrel, A., Burdzy, A., Sowers, L.C.: Characterization of synthetic oligonucleotides containing biologically important modified bases by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal. Biochem.* **379**, 196–207 (2008)
11. Gao, H., Liu, Y., Rumley, M., Yuan, H., Mao, B.: Sequence confirmation of chemically modified RNAs using exonuclease digestion and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom* **23**, 3423–3430 (2009)
12. Apffel, A., Chakel, J.A., Fischer, S., Lichtenwalter, K., Hancock, W.S.: Analysis of oligonucleotides by HPLC-electrospray ionization mass spectrometry. *Anal. Chem.* **69**, 1320–1325 (1997)
13. Gilar, M., Fountain, K.J., Budman, Y., Holyoke, J.L., Davoudi, H., Gebler, J.C.: Characterization of therapeutic oligonucleotides using liquid chromatography with on-line mass spectrometry detection. *Oligonucleotides* **13**, 229–243 (2003)
14. Ni, J., Pomerantz, S.C., Rozenski, J., Zhang, Y., McCloskey, J.A.: Interpretation of oligonucleotide mass spectra for determination of sequence using electrospray ionization and tandem mass spectrometry. *Anal. Chem.* **68**, 1989–1999 (1996)
15. Wu, J., McLuckey, S.A.: Gas-phase fragmentation of oligonucleotide ions. *Int. J. Mass Spectrom.* **237**, 197–241 (2004)
16. Zou, Y., Tiller, P., Chen, I.-W., Beverly, M., Hochman, J.: Metabolite identification of small interfering RNA duplex by high-resolution accurate mass spectrometry. *Rapid Commun. Mass Spectrom* **22**, 1871–1881 (2008)
17. Nikcevic, I., Wyrzykiewicz, T.K., Limbach, P.A.: Detecting low-level synthesis impurities in modified phosphorothioate oligonucleotides using liquid chromatography-high resolution mass spectrometry. *Int. J. Mass Spectrom.* **304**, 98–104 (2011)
18. Farand, J., Beverly, M.: Sequence confirmation of modified oligonucleotides using chemical degradation, electrospray ionization, time-of-flight, and tandem mass spectrometry. *Anal. Chem.* **80**, 7414–7421 (2008)
19. Gardner, M.W., Li, N., Ellington, A.D., Brodbelt, J.S.: Infrared multiphoton dissociation of small-interfering RNA anions and cations. *J. Am. Soc. Mass Spectrom.* **21**, 580–591 (2010)
20. Krotz, A.H., Klopchin, P., Cole, D.L., Ravikumar, V.T.: Phosphorothioate oligonucleotides: Largely reduced (n-1)-mer and phosphodiester content

- through the use of dimeric phosphoramidite synthons. *Bioorg. Med. Chem. Lett* **7**, 73–78 (1997)
21. Chen, D., Yan, Z., Cole, D.L., Srivatsa, G.S.: Analysis of internal (n-1)-mer deletion sequences in synthetic oligodeoxyribonucleotides by hybridization to an immobilized probe array. *Nucleic Acids Res.* **27**, 389–395 (1999)
 22. Krotz, A.H., Klopchin, P.G., Walker, K.L., Srivatsa, G.L., Cole, D.L., Ravikumar, V.T.: On the formation of longmers in phosphorothioate oligodeoxyribonucleotides synthesis. *Tetrahedron Lett.* **38**, 3875–3878 (1997)
 23. Kurata, C., Bradley, K., Gaus, H., Luu, N., Cedillo, I., Ravikumar, V.T., Sooy, K.V., McArdle, J.V., Capaldi, D.C.: Characterization of high molecular weight impurities in synthetic phosphorothioate oligonucleotides. *Bioorg. Med. Chem. Lett.* **16**, 607–614 (2006)
 24. Suzuki, T., Ohsumi, S., Makino, K.: Mechanistic studies on depurination and apurinic site chain breakage in oligodeoxyribonucleotides. *Nucleic Acids Res.* **22**, 4997–5003 (1994)
 25. Capaldi, D.C., Gaus, H.J., Carty, R.L., Moore, M.N., Turney, B.J., Decottignies, S.D., McArdle, J.V., Scozzari, A.N., Ravikumar, V.T., Krotz, A.H.: Formation of 4,4'-dimethoxytrityl-C-phosphonate oligonucleotides. *Bioorg. Med. Chem. Lett.* **14**, 4683–4690 (2004)
 26. Rentel, C., Wang, X., Batt, M., Kurata, C., Oliver, J., Gaus, H., Krotz, A.H., McArdle, J.V., Capaldi, D.C.: Formation of modified cytosine residues in the presence of depurinated DNA. *J. Org. Chem.* **70**, 7841–7845 (2005)
 27. Capaldi, D.C., Gaus, H., Krotz, A.H., Arnold, J., Carty, R.L., Moore, M.N., Scozzari, A.N., Lowery, K., Cole, D.L., Ravikumar, V.T.: Synthesis of high-quality antisense drugs. Addition of acrylonitrile to phosphorothioate oligonucleotides: adduct characterization and avoidance. *Org. Process Res. Dev.* **7**, 832–838 (2003)
 28. Gaus, H., Olsen, P., Sooy, K.V., Rentel, C., Turney, B., Walker, K.L., McArdle, J.V., Capaldi, D.C.: Trichloroacetaldehyde modified oligonucleotides. *Bioorg. Med. Chem. Lett.* **15**, 4118–4124 (2005)
 29. Singh, S., Handa, T., Narayanam, M., Sahu, A., Junwal, M., Shah, R.P.: A critical review on the use of modern sophisticated hyphenated tools in the characterization of impurities and degradation products. *J. Pharm. Biomed. Anal.* **69**, 148–173 (2012)
 30. Fabris, D.: A role for the MS analysis of nucleic acids in the post-genomic age. *J. Am. Soc. Mass Spectrom.* **21**, 1–13 (2010)
 31. Hail, M.E., Elliott, B., Anderson, K.: High-throughput analysis of oligonucleotides using automated electrospray ionization mass spectrometry. *Am. Biotechnol. Lab.* **22**, 12–14 (2004)
 32. Ivleva, V.B., McCarthy, S.M., Yu, Y.Q., Gilar, M.: ProMass for MassLynx software for oligonucleotide analysis. *Waters Application Note*, 720003577EN AG-PDF, Waters Corp.: June (2010)
 33. Agilent MassHunter BioConfirm Software. Familiarization Guide. G6829-90007, Agilent Technologies Inc.: February (2011)
 34. Fjeldsted, J.: Time-of-Flight Mass Spectrometry. Technical Overview. 5989-0373EN, Agilent Technologies Inc.: December (2003)
 35. Zhang, Z., Marshall, A.G.: A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra. *J. Am. Soc. Mass Spectrom.* **9**, 225–233 (1998)
 36. Wyrzykiewicz, T.K., Cole, D.L.: Stereo-reproducibility of the phosphoramidite method in the synthesis of oligonucleotide phosphorothioates. *Bioorg. Chem.* **23**, 33–41 (1995)
 37. Frederiksen, J.K., Piccirilli, J.A.: Separation of RNA phosphorothioate oligonucleotides by HPLC. *Methods Enzymol.* **468**, 289–309 (2009)
 38. Zubarev, R.A., Makarov, A.: Orbitrap mass spectrometry. *Anal. Chem.* **85**, 5288–5296 (2013)
 39. Kujawinski, E.B., Behn, M.D.: Automated analysis of electrospray ionization Fourier transform ion cyclotron resonance mass spectra of natural organic matter. *Anal. Chem.* **78**, 4363–4373 (2006)
 40. Kunenkov, E.V., Kononikhin, A.S., Perminova, I.V., Hertkorn, N., Gaspar, A., Schmitt-Kopplin, P., Popov, I.A., Garmash, A.V., Nikolaev, E.N.: Total mass difference statistics algorithm: a new approach to identification of high-mass building blocks in electrospray ionization Fourier transform cyclotron mass spectrometry data of natural organic matter. *Anal. Chem.* **81**, 10106–10115 (2009)
 41. Roach, P.J., Laskin, J., Laskin, A.: Higher-order mass defect analysis for mass spectra of complex organic mixtures. *Anal. Chem.* **83**, 4924–4929 (2011)
 42. Tziotis, D., Hertkorn, N., Schmitt-Kopplin, P.: Kendrick-analogous network visualization of ion cyclotron resonance Fourier transform mass spectra: improved options for the assignment of elemental compositions and the classification of organic molecular complexity. *Eur. J. Mass Spectrom.* **17**, 415–421 (2011)
 43. Kilgour, D.P.A., Mackay, C.L., Langridge-Smith, P.R.R., O'Connor, P.B.: Appropriate degree of trust: deriving confidence metrics for automatic peak assignment in high-resolution mass spectrometry. *Anal. Chem.* **84**, 7431–7435 (2012)
 44. Wallace, W.E., Guttman, C.M.: Data analysis methods for synthetic polymer mass spectrometry: autocorrelation. *J. Res. Natl. Inst. Stand. Technol.* **107**, 1–17 (2002)
 45. Apicella, B., Bruno, A., Wang, X., Spinelli, N.: Fast Fourier transform and autocorrelation function for the analysis of complex mass spectra. *Int. J. Mass Spectrom.* **338**, 30–38 (2013)