APPLICATION OF THE ORBITRAP MASS ANALYZER TO TARGETED PROTEOMICS AND GAS CHROMATOGRAPHY/MASS SPECTROMETRY OF SMALL MOLECULES

by

Amelia C. Peterson

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2013

Date of final oral examination: 12/13/2012

The dissertation is approved by the following members of the Final Oral Committee:

Joshua J. Coon, Professor, Chemistry
Timothy T. Donohue, Professor, Bacteriology
Frank N. Keutsch, Professor, Chemistry
Lingjun Li, Professor, Chemistry and Pharmacy
Michael R. Sussman, Professor, Biochemistry

APPLICATION OF THE ORBITRAP MASS ANALYZER TO TARGETED PROTEOMICS AND GAS CHROMATOGRAPHY/MASS SPECTROMETRY OF SMALL MOLECULES

Amelia C. Peterson

under the supervision of Professor Joshua J. Coon

at the University of Wisconsin-Madison

ABSTRACT

The work described herein spans two disparate topics: 1) the use of a high resolution and accurate mass analyzer to facilitate targeted proteomic methods, and 2) the development and applications of novel instrumentation for high resolution gas chromatography/mass spectrometry. While disparate, a common theme of a high resolution and accurate mass analyzer, the Orbitrap, ties the work together into, broadly, a treatise on the benefits of high resolution and accurate mass measurements in mass spectrometric research.

In **Chapter 1**, I present an introduction on the Orbitrap mass analyzer and motivate the need for high resolution and accurate mass measurements in mass spectrometry. I will discuss the theory of Orbitrap operation, as well as the practical factors that may enhance or limit its performance. In **Chapter 2**, I delve into the field of targeted proteomics with the introduction of the "parallel reaction monitoring" (PRM) analysis paradigm using a bench-top quadrupole-Orbitrap LC/MS. While targeted proteomics is usually the domain of low resolution but highly sensitive mass analyzers, I demonstrate the utility of high resolution with a direct comparison to the "gold-standard" triple-quadrupole MS-based selected reaction monitoring (SRM) approach.

This PRM approach has already generated much interest in the proteomics community since its publication.

In **Chapters 3** through **5**, I transition to GC/MS, another field dominated by low resolution and fast scanning mass analyzers. In **Chapter 3**, I detail the construction and characterization of a proof-of-principle instrument for high resolution GC/MS built using an electron transfer dissociation (ETD)-enabled quadrupole linear ion trap (QLT)-Orbitrap MS. While the performance of this instrument was hindered by its design, the benefits of the Orbitrap for unambiguous elemental composition assignment to small molecules were clear, and further development was deemed warranted. In **Chapter 4**, I detail that further development in the design and construction of a dedicated, applications-grade bench-top quadrupole-Orbitrap GC/MS. The application of this instrument to the non-targeted characterization of unknown metabolites is presented in **Chapter 5**. Additionally in **Chapter 5**, I introduce a new method for data-dependent acquisition termed “molecular-ion directed acquisition”, which aims to maximize the information content from non-targeted tandem MS in metabolomic studies.

(350 words)

ACKNOWLEDGEMENTS

I wish to, firstly, express my sincere gratitude to Professor Joshua J. Coon. The environment he has fostered at UW-Madison is truly unparalleled. I cannot be thankful enough for the amazing opportunities he has provided me both inside, and outside, the lab; a graduate student could have no better advocate. Specifically, I greatly appreciate the two internship opportunities he provided me at Thermo Fisher Scientific, in Bremen, Germany, and San Jose, California. These internships have directly set the stage for my post-graduate career, as I will be taking a Scientist position at Thermo Fisher Scientific in Bremen immediately following graduation. Lastly, I greatly value the freedom he has extended to me to follow the paths of my own interest; that, along with his guidance, has been critical to my success.

I would also like to thank my constant companions in this endeavor, my fellow members of Joshua J. Coon research group, both past and present. My special thanks goes to Graeme C. McAlister, whom selflessly devoted much of his time, in my early days in the lab, to instructing me in both the theoretical and practical aspects of mass spectrometry. I could not have asked for a better mentor. I would also like to mention A. J. Bureta, Gloria Sheynkman née Kreitinger, and Molly McDevitt, for not only being fantastic colleagues, but also for their unwavering friendship. Additionally, I thank current group members, Sara Austin, Derek Bailey, Allison Balloon, Dominic Colosi, Emma Coughlin, Alex Hebert, Alan Higbee, Shane Hubler, Nick Kwiecien, Anna Merrill née Larson, Rebeca Pinhancos, Greg Potts, Tim Rhoads, Alicia Richards, Nick Riley, Chris Rose, Arne Ulbrich, and Catie Vincent, and past members, Nicole Ellis née Beauchene, Justin Brumbaugh, David Good, Paul Grimsrud, Aaron Ledvina, Violet Lee, Doug Phanstiel, Nambirajan Rangarajan, Neil Rumachik, Jason Russell, Danielle Swaney, Mark Tervo, Craig Wenger, Qiangwei Xia, and Chenxi Yang, for their inspiration and innumerable

contributions to my success. Lastly, I want to give my sincere thanks to Michael Westphall, the Coon lab's instrumentation innovator, for his tremendous efforts and unfailing support.

I am also grateful for the advice, expertise, and mentorship of my collaborators, and soon-to-be colleagues, at Thermo Fisher Scientific. The opportunity of their mentorship is a unique component of the Coon lab experience, and was invaluable in my development as a scientist. Specifically, I would like to thank Jens Griep-Raming for supervising me during my time in Bremen, as well as Jan-Peter Hauschild, Eduard Denisov, Alexander Makarov, Stevan Horning, Dirk Krumwiede, and Dirk Nolting of Bremen; John Syka, Jae Schwartz, August Specht, and George Stafford (formerly) of San Jose; and Scott Quarmby of Austin.

I will conclude by thanking my friends and family. First, my parents, Paul and René, whom raised me without limitations, inspired my curiosity, and have supported me infallibly throughout my life as a student. It is thanks to them that I have reached this point. I would also like to thank my brother, Brandon, whom has always provided me with someone to look up to. Finally, I thank my friends whom have supported me in my day-to-day endeavors.

TABLE OF CONTENTS

TABLE OF CONTENTS, CONTINUED

TABLE OF CONTENTS, CONTINUED

TABLE OF FIGURES

TABLE OF FIGURES, CONTINUED

TABLE OF FIGURES, CONTINUED

Chapter 6

TABLE OF TABLES

TABLE OF TABLES, CONTINUED

Chapter 7

TABLE OF EQUATIONS

ABBREVIATIONS AND ACRONYMS

-2	penultimate-scan predictive automatic gain control
-1	ultimate-scan predictive automatic gain control
1	pre-scan automatic gain control
%RSD	percent relative standard deviation
$[M+H]^+$	pseudomolecular ion
°C	degree Celsius
11-methyl-12-C18:1	methyl 11-methyl-12-octadecenoate
11-methyl-C18:1	methyl 11-methyl-octadecenoate
1T, 2T, 3T	first, second, third transition
A	solvent A in liquid chromatography
Å	Ångström
α	significance level
ACN	acetonitrile
AGC	automatic gain control
amol	attomole
ANOVA	analysis of variance
AP	atmospheric pressure
ASP	advanced signal processing
AUC	area-under-the-curve
B	solvent B in liquid chromatography
BAME	bacterial acid methyl ester
Br	bromine
C	carbon
C18:0	methyl octadecanoate
C18:1	methyl octadecenoate
C19:0	methyl nonadecanoate
C19:1	methyl nonadecenoate
CAD	collision-activated dissociation
CE	collision energy
CFA	cyclic fatty acid
CH_3CN	acetonitrile
CH_3OH	methanol
$CHCl_3$	chloroform
CI	chemical ionization
Cl	chlorine
cm	centimeter
COM	component object model
Conc	concentration

ABBREVIATIONS AND ACRONYMS, CONTINUED

d	day
D	aspartic acid
Da	Dalton
DC	direct current
DDA	data-dependent acquisition
d_f	film thickness of the stationary phase in gas chromatography
DI	deionized
E	glutamic acid
EI	electron ionization
EL	electron lens
EM	electron multiplier/conversion dynode
EPA	Environmental Protection Agency
ESI	electrospray ionization
ETD	electron transfer dissociation
eV	electron-volt
F	fluorine
FAME	fatty acid methyl ester
FC-43	perfluorotributylamine
FFT	fast Fourier transform
flat	flatapole
fmol	femtomole
FT	Fourier transform
FT-ARM	Fourier transform-all reaction monitoring
FT-ICR	Fourier transform-ion cyclotron resonance
FWHM	full-width-half-max
GC	gas chromatography
GC/MS	gas chromatography/mass spectrometry
h	hour
H	hydrogen
H_2O	water
HCD	higher energy c-trap dissociation
HCl	hydrochloric acid
HCOOH	formic acid
H_o	null hypothesis
HpCDD	heptachlorodibenzo-*p*-dioxin
HpCDF	heptachlorodibenzofuran
HPLC	high performance liquid chromatography
HR/AM	high resolution and accurate mass

ABBREVIATIONS AND ACRONYMS, CONTINUED

HxCDD	hexachlorodibenzo-*p*-dioxin
HxCDF	hexachlorodibenzofuran
Hz	Hertz, per second
IAE	isotopomer abundance error
ICIS	a proprietary peak picking and integration algorithm
ID	inner diameter
inj	injection
int	normalized intensity
IRE	isotopomer ratio error
iso	isolation
iSRM	intelligent selected reaction monitoring
K	lysine
k	rate constant
k′	pseudo-rate constant
kV	kilovolt
L	liter
L	lens
LC	liquid chromatography
LC/MS	liquid chromatography/mass spectrometry
LOD	limit-of-detection
log	logarithm
m	meter
M	monoisotopomer, monoisotopic ion; molar
m	mass
m/z	mass-to-charge
M+1	first isotopomer
M+2	second isotopomer
M+3	third isotopomer
M+4	fourth isotopomer
MALDI	matrix-assisted laser desorption-ionization
mDa	milliDalton
MeOX	methoxime
mg	milligram
MIDA	molecular ion-directed acquisition
MIE	1-methyleneimino-1-ethenylium
min	minute
mL	milliliter
mm	millimeter

ABBREVIATIONS AND ACRONYMS, CONTINUED

mmu	millimass unit
MRM	multiple reaction monitoring
MS	mass spectromet(ry/er), precursor ion mass analysis
ms	millisecond
MS/MS	tandem mass spectrometry, product ion mass analysis
MSTFA	*N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide
MTBSTFA	*N*-(*t*-butyldimethylsilyl)-*N*-methyltrifluoroacetamide
mTorr	milliTorr
MΩ-cm	megohm-centimeter
μ	mean
μA	microampere
μg	microgram
μL	microliter
μm	micrometer
N	asparagine; nitrogen; normal
n	population size
N_2	nitrogen (gas)
NCE	normalized collision energy
NCI	negative chemical ionization
ng	nanogram
NIST	National Institute of Standards and Technology
nL	nanoliter
nM	nanomolar
O	oxygen
OCDD	octachlorodibenzo-*p*-dioxin
OCDF	octachlorodibenzofuran
OD	outer diameter
OFN	octafluoronaphthalene
OT	Orbitrap
P	phosphorus
PCDD	polychlorinated dibenzo-*p*-dioxins
PCDF	polychlorinated dibenzofurans
PCI	positive chemical ionization
PeCDD	pentachlorodibenzo-*p*-dioxin
PeCDF	pentachlorodibenzofuran
pM	picomolar
pmol	picomole
ppm	parts-per-million

ABBREVIATIONS AND ACRONYMS, CONTINUED

PRM	parallel reaction monitoring
PTR	proton transfer reaction
PQD	pulsed-q dissociation
Q	glutamine
q	reduced Mathieu parameter
Q1	first quadrupole for precursor selection
q2	second quadrupole for beam-type collision-activated dissociation
Q3	third quadrupole for product selection
QIT	quadrupole ion trap
QLT	linear quadrupole ion trap
QqOrbi	quadrupole-Orbitrap
QqQ	triple quadrupole
QqTOF	quadrupole-time-of-flight
Q-TOF	quadrupole-time-of-flight
R	arginine
R^2	coefficient of determination of a linear regression
rel	relative
rep	replicate
RF	radio frequency
rms	root-mean-squared
rpm	revolutions per minute
RT	retention time; room temperature
s	second
S	serine; sulfur
σ	standard deviation, sigma
S/N	signal-to-noise
SA	spectral accuracy
sCLIPs	Self-Calibrated Line-shape Isotope Profile Search
Si	silicon
SIL	stable-isotope labeling
SIM	selected ion monitoring
SRM	selected reaction monitoring
SV	semi-volatile
T	threonine
t-	*tert*-
TBDMCS	*t*-butyldimethylchlorosilane
*t*BDMS, TBS	*t*-butyldimethylsilyl
TCDD	tetrachlorodibenzo-*p*-dioxin

ABBREVIATIONS AND ACRONYMS, CONTINUED

TCDF	tetrachlorodibenzofuran
TCP	transmission control protocol
TFS	Thermo Fisher Scientific
Th	Thompson
TMCS	trimethylchlorosilane
TMS	trimethylsilyl
TOF	time-of-flight
TPFHST	2,4,6-tris(pentadecafluoroheptyl)-1,3,5-triazine
TPR	true positive rate
V	volt
v/v	volume per volume
V_{p-p}	volt peak-to-peak
x	times, fold
xfer	transfer
XIC	extracted ion chromatogram
XSC	extracted score chromatogram
z	charge

CHAPTER 1

Introduction

HIGH RESOLUTION AND ACCURATE MASS IN MASS SPECTROMETRY

The benefit of high resolution and accurate mass in mass spectrometry within all fields can be simply summed up as providing a limit on the number of candidate elemental compositions for a given mass spectral feature. In a research context, this simple benefit means significant analytical power. In discovery proteomics, for instance, routine use of high resolution instruments has increased confidence in protein and peptide identifications through resolved and accurate mass measurements and charge state determinations,[1-4] sped up database searching,[5] improved detection of polymorphisms and localization of post-translational modifications,[2,6] and improved quantification through greater resolution of the *m/z*-domain.[2] Accurate mass measurements also make possible the analysis of petroleum,[7-10] coal,[11] oils,[12] and humic/fulvic acids.[13] Likewise, in small molecule GC- and LC-based mass spectrometry, high mass accuracy and resolution are indispensable when the standard method of identification through spectral matching fails.[14] Thus, whether in the context of peptide database searching, compositional analysis of complex mixtures, or manual annotation of unknown metabolites, the goal of the mass measurement is to generate a unique elemental composition which can be subsequently mapped to an analyte.

Regardless of the research question, elemental composition derivation is the first step in identifying a compound. Beynon explained that if mass can be determined with sufficient accuracy, the elemental composition can be deduced.[15] Thus, theoretically, with infinite mass accuracy, and the ability to resolve, completely, all relevant masses,[16] a unique elemental composition can always be determined.[15] In practical terms, if a measurement is known to 0.1 mmu accuracy (not routinely possible), a unique composition can be determined for molecules containing the elements C, H, N, O, and S, and having mass less than 500 Da.[16] If the element space or mass of the compound increase, however, mass accuracy alone cannot point to a single composition; orthogonal filters, such as isotopomer distribution, then become necessary, in conjunction with high mass accuracy, to limit the number of elemental compositions possible.[17-23]

We have performed theoretical calculations to demonstrate the effect of mass accuracy and compound complexity on the number of possible elemental compositions, and motivate the need for isotopic abundance information (**Figure 1.1**). In panels (**A**) – (**D**), we plot the number of even-electron compositions residing between *m/z* 300-301 at mass tolerances from 0.01 mmu (0.033 ppm) to 1 mmu (3.33 ppm). The four panels show the effect of increasing the number of elements considered: panel (**A**) represents peptidic or petroleum-based compounds[16] while panels (**B**) – (**D**) are representative of metabolites, environmental toxins, or industrial byproducts.[17,21,24] In (**A**), 0.1 mmu (0.33 ppm) mass accuracy is sufficient to yield a unique elemental composition. As compound complexity increases from (**B**) to (**D**), however, unique elemental compositions are not always possible. For example, given a mass chosen at random, *m/z* 300.11651, there are 91, 41, 12, 3, and 1 possible elemental compositions at mass tolerances of 1, 0.5, 0.1, 0.05, and 0.01 mmu, respectively, when considering the case illustrated in **(D)**. Using mass accuracy alone, 0.01 mmu mass accuracy, not currently available with any MS, would be required to determine the correct elemental composition ($C_{16}H_{18}N_3OS^+$).

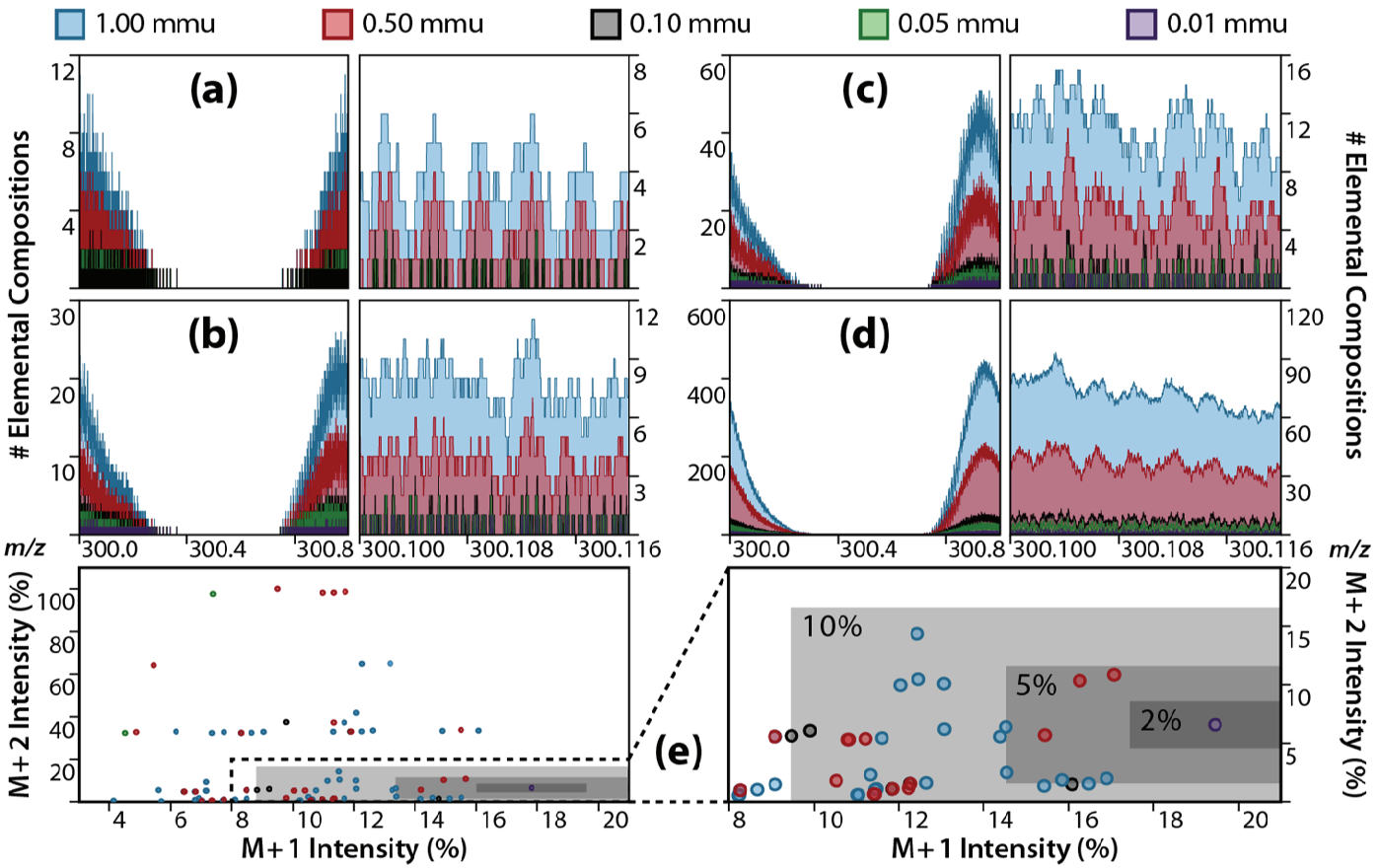


Figure 1.1. (A-D) Count of even-electron elemental compositions residing between *m/z* 300-301 and, at right, *m/z* 300.10-300.12, at mass tolerances of 0.01, 0.05, 0.10, 0.50, and 1.00 mmu. Elements considered increases in each panel: (**A**) $C_xH_xO_xN_xS_x$, (**B**), $C_xH_xO_xN_xS_xP_x$, (**C**) $C_xH_xO_xN_xS_xCl_x$, (**D**) $C_xH_xO_xN_xS_xP_xCl_xF_xBr_xSi_x{}^{13}C_y$, where x = 0-500, y = 0-2. **(E)** Plot of M+1 isotope abundance versus M+2 isotope abundance for all potential elemental compositions at *m/z* 300.11651 within 0.01-1 mmu. Grey boxes bound the 10%, 5%, and 2% isotopic abundance region around the true elemental composition shown in purple ($C_{16}H_{18}N_3OS^+$).

In panel (**E**), we applied an orthogonal isotopic abundance filter on the 91 elemental compositions from panel (**D**) with less than 1 mmu mass error. For a mass spectrometer with 10% isotopic abundance error, the count of possible elemental compositions is reduced on average 4-fold to 19, 12, 4, 1, and 1 at 1, 0.5, 0.1, 0.05, and 0.01 mmu, respectively. With 5% isotopic abundance error, possible compositions are further reduced to 7, 3, 1, 1, and 1; thus, only 0.1 mmu mass accuracy is necessary to make the correct composition assignment. At 2%, the correct assignment can be made even at 1.00 mmu mass accuracy. By inclusion of isotopic distribution accuracy an orthogonal filter, the requirement on mass accuracy to achieve unique elemental compositions is brought into a range more realistic for currently available mass spectrometers. Similarly, in a study by Kind,[17] isotopic distribution information eliminated over 95% of false candidates. The reduction of candidate compositions was greater for measurements with 3 ppm mass accuracy and 2% isotope distribution error than for measurements with 0.1 ppm mass accuracy alone. It should be noted that to achieve high accuracy measurements of mass and isotopic distribution, high resolution measurements are required to remove background and isobaric ion interferences. These studies clearly demonstrate the necessity and utility of high resolution and mass accuracy mass analyzers with low isotopic distribution error in the analyses of complex or unknown samples.

Numerous high-resolution mass spectrometers exist to address this need, including magnetic sector,[25] high-resolution time-of-flight,[26] Orbitrap,[27] and FT-ICR instruments.[28] While the FT-ICR is unparalleled in terms of resolving power,[29] TOFs have the greatest acquisition rates,[26] and isotope ratio measurements are the purview of magnetic sector instrumentation,[25] the Orbitrap mass analyzer possesses a high performance mix of these characteristics that make it the analyzer of choice in this work. The following sections detail how these characteristics of high

resolution and high mass accuracy manifest, and outline the factors which influence their robustness.

A BRIEF HISTORY OF THE ORBITRAP ANALYZER

The Orbitrap mass analyzer, a rare new-comer in the field of mass spectrometry[30], was invented by Alexander Makarov less than 15 years ago.[31,32] Already, its impact on the broad fields of research served by mass spectrometry can arguably be described as nothing short of transformative.[27] Since its initial commercialization by Thermo Fisher Scientific in 2005 as a linear quadrupole ion trap-Orbitrap hybrid MS,[33] thousands of Orbitrap instruments, of various configurations, have been adopted by research labs worldwide. Currently, Orbitrap instruments are available in several different hybrid[34-36] and single-stage bench-top[37,38] configurations that permit multiple ionization types (ESI,[39] MALDI[40]) and numerous ion activation methods (beam-type[41] and resonant excitation CAD, ETD/PTR,[42] PQD) to be used in conjunction with Orbitrap mass analysis. The most recent implementations have made substantial gains in scan speed and resolution through the use of parallel ion injection and mass analysis schema,[37,38] advanced image current transient processing methods,[43-45] improved detection electronics,[43] and a new high-field Orbitrap design.[34,35] At its present stage of development, the Orbitrap can achieve mass resolution exceeding 480,000 (relative to *m/z* 400),[35,46] mass errors less than 1 ppm,[47,48] a *m/z* range of at least 6000 Th, scan rates exceeding 12 Hz at resolution 12,500 (relative to *m/z* 400),[35,37] a dynamic range (and extent of mass accuracy) exceeding 5,000[49] and as high as 25,000,[34] and isotopomer abundance errors typically within 3-10%.[22,50]

Like all inventions, the Orbitrap was not developed in a vacuum (pun intended). Although a wholly new mass spectrometer, the Orbitrap's ancestry can be traced back to the Kingdon trap[51] of the early 1920s and the principles of dynamic orbital trapping in an electrostatic trap (i.e., without magnetic or dynamic electric fields) proposed by Kingdon.[51] In its simplest

configuration, the Kingdon trap is an end-capped cylinder with a wire stretched along its central axis.[51] Application of a DC potential between the wire and cylinder produces a radial logarithmic potential between the electrodes that causes ions with sufficient initial velocity to develop stable orbits radially and axially about the wire,[52] thus achieving orbital trapping (as well as axial confinement by potentials applied to the end-cap electrodes[53]). The radial logarithmic potential distribution (Φ) is given by **Equation 1.1**:

$$\Phi = A\ln(r) + B$$

Equation 1.1

where A and B are constants with value dictated by the applied voltages, and r is the radial coordinate. Ions lacking sufficient initial velocity impinge on the central wire and are lost.[31] While the Kingdon trap found wide-use in ion optical spectroscopy studies,[27,52,54-56] it was not until a variant of the Kingdon trap by Knight[57] in 1981 that the possibility of mass analysis with an electrostatic orbital trap was realized.

Knight tapered the ends of, and split, the outer cylindrical electrode to induce a quadrupolar axial trapping potential in addition to the radial logarithmic potential providing orbital trapping (as in the Kingdon trap).[57] Superimposing the quadrupolar (**Equation 1.2**) and logarithmic fields (**Equation 1.1**) (thereby producing a "quadro-logarithmic" field[31]) caused orbital-trapped ions to harmonically oscillate along the length of the central wire electrode. The quadro-logarithmic potential distribution (**Equation 1.3**) is given simply by the sum of **Equations 1.1** and **1.2**:

$$\Phi = A\left(z^2 - \frac{r^2}{2}\right)$$

Equation 1.2

$$\Phi = A\left(z^2 - \frac{r^2}{2} + B\ln(r)\right)$$

Equation 1.3

where, again, *A* and *B* are constants related to the trap's geometry and applied voltages, and *r* and *z* are the radial and axial cylindrical coordinates, respectively. When Knight applied an RF voltage between the two halves of the outer electrode, he observed resonances proportional to the reciprocal of the *m/z* of the ions when the applied RF frequency was tuned to that of the ion motion.[27,54,55,57] While the observed resonances were broadened, shifted in frequency, and weaker than expected for the population of ions in the trap, this crude mass analysis led Knight to hypothesize that the resonances could be improved by redesigning the central electrode, as it was found distort the intended harmonic axial potential.[57] His proposal was later termed the "ideal Kingdon trap".[58]

Gillig and colleagues[58] confirmed Knight's proposal through theoretical ion simulations in 1996. Their calculations revealed that the equipotential lines of a true quadro-logarithmic field formed a spindle-like central trapping volume (**Figure 1.2A**). Thus, they suggested that an ideal Kingdon trap could be fabricated by either 1) using a spindle-like central electrode with shape given by the quadro-logarithmic equipotential lines, or 2) creating the spindle-like trapping field through fine adjustments in electrode potentials. Gillig, *et al*. followed the latter path.[27,58,59] Using a wire ion guide FT-ICR cell with "tuned" electrode potentials to approximate the ideal

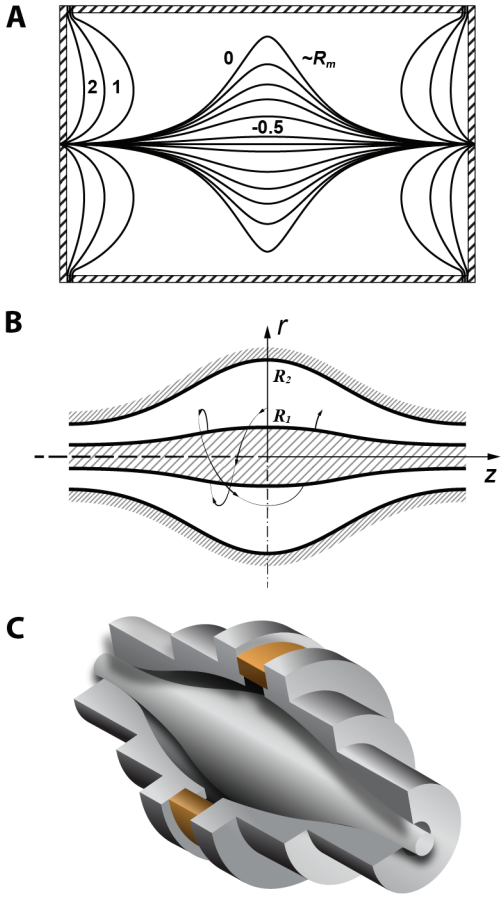


Figure 1.2. (**A**) SIMION plot of the equipotential lines for the "ideal Kingdon trap" with end plates at 14 V and wire at -1 V, adapted from Gillig, *et al*.[58] The equipotential line marked as 0 V roughly corresponds to the trapping stability boundary, R_m, described by Makarov,[31] where ions with radii less than R_m are attracted to the axis and ions with radii greater than R_m are repelled and lost (assuming positively charged ions). (**B**) Equipotential lines of the quadro-logarithmic Orbitrap field with inner and outer electrode radii R_1 and R_2, adapted from Makarov.[31] An example of a stable ion trajectory is shown. (**C**) A three-dimensional representation of the Orbitrap device showing the split outer electrode [not to scale].

field, they demonstrated improved, high-resolution mass analysis in the context of normal FT-ICR experiment.[58] Makarov pursued the former, more technically challenging, path.[31]

In 2000, Makarov coined the term "orbitrap" to describe his device, which consisted of a central spindle-like electrode enclosed in a larger "barrel-like" outer electrode (**Figure 1.2B-C**).[31] In the Orbitrap, ions rotate, and axially oscillate, about a central spindle in "an intricate spiral,"[31] where the frequency of the harmonic ion oscillations along the field's axis is dictated by the mass-to-charge ratio of the ions.[31,32,60] Expectedly, the electrostatic potential distribution (U) for the Orbitrap (**Equation 1.4**), as described by Makarov,[31,34] is quadro-logarithmic like in **Equation 1.3**:

$$U(r,z) = \frac{k}{2}\left(z^2 - \frac{(r^2 - R_1^2)}{2}\right) + \frac{k}{2}R_m^2 \ln\left[\frac{r}{R_1}\right] - Ur$$

Equation 1.4

where k is the field curvature or axial restoring force, R_1 is the central electrode radius, R_m is the characteristic radius (a limit describing the radial boundary of ion trapping), and Ur is the central electrode voltage. $k/2$ and R_m^2 are equivalent to the A and B terms of the prior expression (**Equation 1.3**).

This initial report, using, as proof-of-principle, a pulsed laser ablation or MALDI source with no external ion accumulation or cooling, demonstrated mass resolution up to 150,000, isotopic ratios within 5-10%, and root-mean-squared mass errors of only 5 ppm.[31] Following these promising results, the Orbitrap was coupled to a continuous ion source, ESI, and an RF-only storage device, the c-trap, for accumulation of large ion populations and radial injection into the Orbitrap, was developed.[33,39,49,55] These developments ultimately enabled the Orbitrap's first

commercial implementation as a hybrid instrument with a quadrupole linear ion trap front-end,[33] and guaranteed the instrument's adoption into mainstream mass spectrometry[27,54,55] as the first commercially-viable electrostatic orbital trap.[54] Until the Orbitrap, as Makarov remarked, "dynamic ion trapping in electrostatic fields [had] been regarded [as] *little more than a curiosity of doubtful analytical merit*" [emphasis added].[60]

THEORY AND PRACTICE OF ORBITRAP OPERATION

The geometry of the Orbitrap electrodes (**Figure 1.2B-C**), and trajectories of ion motion within the trap, are derived from the equation for the Orbitrap's quadro-logarithmic potential distribution (**Equation 1.4**).[31] Since several in-depth treatments of the theory of ion capture and motion within the Orbitrap exist in the literature,[27,31,55] only the main relationships governing mass accuracy, resolution, and/or isotopic distributions will be presented here.

Ions are accumulated for Orbitrap mass analysis in a curved RF-only linear quadrupole, the c-trap, external to the Orbitrap, and then injected radially, within a few hundred nanoseconds, through a slot in the electrode at the inside radius of the curved trap.[33,49,54] The ions are accelerated along an S-shaped trajectory and converge as they pass through the entrance to the Orbitrap.[55] As the ions enter the Orbitrap, they experience an increase in the electric field strength as the central electrode voltage rises to its final set-point.[31,54,55] This process is termed "electrodynamic squeezing"[31,32] and does just that; the radius of the ion cloud, and the rotational radius about the central electrode, contract with the increase in field strength. This process prevents collision of the ions with the outer electrode during their initial oscillations, and also makes "room" for later-arriving ions (those of higher *m/z*) to enter.[32,54,55] As a result, the length of the electrodynamic squeezing step (or, rate of the central electrode's voltage ramp, which ultimately is the rate-limiting step for Orbitrap scan rate[54]) dictates the mass range of ions that

make it into the analyzer.[31,55] Following squeezing, voltages in the trap are stabilized and held constant, ions develop stable ion trajectories, and image current detection can begin.[27]

In terms of cylindrical coordinates (i.e., r, z, and φ, the radial, axial, and angular coordinates, respectively), stable ion trajectories within the Orbitrap involve simultaneous r- and φ-motion about the central electrode and simple harmonic oscillations along the z-axis, and meet the requirement that their radii are less than the characteristic radius, R_m (**Figure 1.2A**).[31,32,60] Because there are no cross terms of r/φ and z in the electrostatic quadro-logarithmic field (**Equation 1.4**), an important characteristic emerges: the frequency of oscillatory motion of ions in the z-direction is independent of their orbital motion in the r and φ-directions.[31,54,55] In fact, only the axial frequency is completely independent of the initial ion position and kinetic energy (neither the rotational nor radial frequencies share in this independence, and thus, despite being previously proposed for use in deriving *m/z*,[31,61] would result in poor mass resolution if used for mass analysis).[31] This independence is the crux making the Orbitrap a high mass accuracy and high resolution mass analyzer; the axial frequency depends *only* on an ion's *m/z* and the constant potential between the electrodes and can thus be exploited for mass analysis.[31,32,54,55,60] The relationship between axial frequency (ω_z, in rad/s) and *m/z* is given by **Equation 1.5**.

$$\omega_z = \sqrt{\frac{kq}{m}}$$

Equation 1.5

where m and q are the mass and charge, respectively, and k is related to the field potential.[31]

While not used for mass analysis, orbital motion in the r and φ-directions is nonetheless critical to maintaining the radial stability of the trapped ions.[27,55] Just like in the Kingdon trap,

radial stability is related directly to the initial ion kinetic energy and, in the Orbitrap, is, importantly, independent of *m/z*.[55] Thus, if ion kinetic energy is suitably matched to the radial component of the electric field (in an situation analogous to the selection of ions of a given kinetic energy for transmission in an electric sector with fixed sector radius[62]), ions of the same kinetic energy will form a nearly circular, "thin ring" about the central electrode with radius given by their kinetic energy.[31,55] Thus, because ions of different kinetic energies possess different orbit radii and the Orbitrap can accept a range of initial kinetic energies, the Orbitrap can trap large payloads of ions and possesses high space charge capacity, relative to FT-ICR and Paul-type quadrupole ion traps.[54,55] The end result is a set of ion rings, each with a given *m/z* and kinetic energy, oscillating in the *z*-direction by nature of their *m/z*, and displaced radially from the central electrode by nature of their kinetic energy.

The ensuing axial oscillation of the ions induces an image current on the split outer electrodes, which is detected as a time-domain transient and amplified by a differential amplifier,[31,55] as in FT-ICR.[28] The transient is fast Fourier-transformed (FFT) to generate a frequency-domain spectrum that is translated into *m/z*-space by **Equation 1.5**.[31,54,55] The total detected ion current is the sum of the ion current of each individual species, and has a magnitude directly proportional to the ion population size, charge, and axial amplitude of oscillation.[31,55] Axial oscillations can be detected over 10-100s of thousands of oscillations until ion packets ultimately de-phase and become undetectable.[55] Ion packets might lose coherence due to small imperfections in the Orbitrap field, both intentional (e.g., the injection slot) and unintentional (e.g., machining imperfections, or power supply stability), as well as through collisions with background gas, or space charge effects within or between ion packets.[31,49,54,55,63] Regardless of the cause, when the axial thickness of an initially coherent packet (ring) of ions de-phases to approach the axial amplitude of oscillation, the broadened ring will then induce opposite image

currents on the both outer electrodes simultaneously that ultimately cancel each other out and reduce the signal into the noise.[31,54,55]

The length of time that axial oscillations can continue with coherence ultimately dictates the upper limit of mass resolution.[31,54,55] Since ions of higher *m/z* are more affected by collisions with background gas, by nature of their greater collisional cross-sections (and ion kinetic energy is independent of *m/z*), mass resolution also decreases as *m/z* increases (and the maintenance of high-vacuum is essential).[27,31,33] In contrast to FT-ICR, however, resolution diminishes only as the square root of *m/z*,[31] rather than as m/z^2[28] (i.e., more slowly than FT-ICR, meaning that for a fixed transient length the resolution of the Orbitrap will eventually exceed that of the FT-ICR above some *m/z*[33,34]). The experimental mass resolution is given by the following relationship (**Equation 1.6**):

$$\frac{m}{\Delta m} = 2\frac{\omega}{\Delta\omega} = \frac{1}{2\Delta\omega}\sqrt{\frac{kq}{m}}$$

Equation 1.6

where $m/\Delta m$ and $\omega/\Delta\omega$ are the mass and frequency resolutions at full-width-half max.[31]

While the factors affecting mass resolution also influence mass accuracy,[27] additional factors can cause mass measurement errors. Primarily, mass accuracy is limited by the stability of the voltages applied to the central electrode and deflector electrode (compensating for the ion entrance hole).[54,55] Failure to properly compensate for the fringing fields caused by the ion entrance hole, for instance, causes ions of the same mass to have different axial frequencies, and thus, leads to decreased resolution, peak splitting, and mass shifts in the mass spectrum.[54,55,63,64] Additionally, fluctuations in the ambient temperature result in long-term, rhythmic mass

deviations with a period of about 1 d if externally calibrated.[48,49] At low signal-to-noise, mass errors are primarily the result of electronic and thermal noise;[33] these are unchanged for internal and external calibration.[49,54] However, since increasing the length of the transient (acquiring at higher resolution) enhances signal-to-noise (signal increases proportionally to time, but noise increases with the square-root of time), mass accuracy can be improved in marginal cases by acquiring at higher mass resolution.[49] At a given resolution, the dynamic range over which accurate measurements can be made has been measured to be about 5000 in the Orbitrap.[49]

Just as too few ions decrease mass accuracy, so too will too many ions, while also affecting isotopic abundances. Large cumulative or individual ion populations can lead to Coulombic repulsions, or space charge effects, that cause mass shifts, peak coalescence (mutual phase locking of two ion packets having close *m/z*-values), diffusion (dephasing due to space charge interactions), and synchronization (or "self-bunching,").[54] In self-bunching, Coulombic interactions increase the velocity of slower ions while decreasing the velocity of faster ions to create a "bunch" that oscillates together, i.e., coalesces[54,63,65,66], leading to a shift in measured frequency, and thus *m/z*.[63,65] Furthermore, in some cases, the bunched ion beam may decrease width over time,[65] or split into two or more sub-beams having disparate de-phasing rates, and manifest spectrally as split peaks.[63] High density sub-beams will de-phase at a slower rate than their associated, lower density sub-beams, leading to incorrect isotopic ratios.[63]

For the most part, space charge effects are substantially mitigated by the shielding provided by the central electrode. Especially when rotational radii are small, ions on one side of the central electrode are insulated from the ions other side.[54] Since these effects are relatively small, mass measurement deviations due to space charge (or AGC ion target) can often be calibrated out. Likewise, careful tuning through controlled distortions of the field (to restore the ideal field) can, in principle, mitigate diffusion and synchronization effects.[54,63] However, with

current Orbitrap instrumentation, perturbations made to decrease synchronization effects will affect the device in a *m/z*-dependent manner; thus, optimization for one *m/z* value may result in unchanged or worse performance for other *m/z* values.[63] With the current technology, isotopic abundance accuracy decreases as resolution/transient length increases, with isotopic ratio errors ranging from less than 3% at lower resolutions to greater than 10% at higher settings.[22,50] Although optimization of the injection parameters and field compensation for the injection hole, can improve ratios, at the present time, only limiting ion population size and transient length will universally maintain isotopic abundance ratios at acceptable levels.[54,63,65]

ADVANCES IN ORBITRAP TECHNOLOGY TO IMPROVE MASS ACCURACY, SCAN RATE, AND RESOLUTION

Mass accuracy. One feature of the using an external RF-only device to accumulate and trap ions prior to their injection into the Orbitrap for analysis is that custom ion populations can be formed through multiple fills of the c-trap with different (or the same) ion populations.[37,47-49,67-70] This trait has been exploited in recent implementations of the Orbitrap analyzer to permit multiplexed analysis of fragmented peptides,[37,67,68] improved signal-to-noise in analysis of whole proteins,[69] improved quantification,[70] as well as for the introduction of reference compounds for internal mass calibration.[47,48] The latter strategy, termed "lock-mass", has permitted robust internal calibration of each mass spectrum to result in mass errors typically less than 1 ppm.[47,48]

In one implementation, Olsen and colleagues[47] added, in pre-defined amounts, a known electrospray ionization background ion, polydimethylcyclosiloxane, to each population of analyte ions prior to MS and MS/MS analyses. The calibrant population was prepared by accumulating and isolating the desired ion in the up-stream QLT. This population was then passed to the c-trap, where it remained until joined by precursor- or product-type analytes ions. The populations were

then mixed through mutual storage in the c-trap and simultaneously injected into the Orbitrap for analysis. The authors demonstrate, on average, sub-ppm mass errors, with maximum deviations less than 2 ppm.[47]

A second implementation carried out by Wenger, *et al*.,[48] made use of another key trait of the c-trap, namely that it acts as a t-piece in the otherwise single-direction ion path of hybrid Orbitrap instruments. This means that ions can enter the c-trap from the direction opposite the ESI source, coming from additional sources, such as an NCI source used for generating negative reagent ions for electron transfer dissociation (ETD),[42] or other devices (such as a beam-type CAD collision cell).[41] Wenger, *et al*.[48] used fluoranthene cations generated by chemical ionization in the EI/CI source of an ETD-enabled hybrid Orbitrap,[42] stored a population of these cations in the c-trap, and then proceeded with the normal scan sequence for MS and MS/MS. Like Olsen, *et al.*,[47] this resulted in robust, sub-ppm mass measurement errors for both precursor and product ions, and near elimination of calibration drift over time. This strategy had the benefits, however, that it did not require processing of the calibrant ion population in the QLT, saving a significant amount of time, and that the analyte and calibrant ions were generated by two different ion sources. The latter benefit eliminated concerns of the stability of the calibrant signal over the course of a chromatographic run, and between laboratories.[48]

Resolution and scan rate. On the mass resolution front, two major technological breakthroughs warrant mention: the development of a high-field Orbitrap mass analyzer,[34,35] and advances in signal processing of Orbitrap transients.[43-46] Both of these advances also provide increased instrument duty cycle at the same resolution settings.

Analogous to the quest for larger and larger magnets (with higher field strength) to increase resolution in FT-ICR MS,[28] Orbitrap resolution can also be improved through increases in field strength.[34,54] For an electrostatic trap, increases in field strength come in two forms:

through increasing the voltages applied or through changing the geometry of the trap.[34] If **Equation 1.4** is solved for *k*, the field curvature, and then substituted into **Equation 1.5**, the following relationship is revealed (**Equation 1.7**):

$$\omega_z = \sqrt{\frac{q}{m} \cdot \frac{2Ur}{R_m^2 \ln\left(\frac{R_2}{R_1}\right) - \frac{1}{2}[R_2^2 - R_1^2]}}$$

Equation 1.7

where *Ur* is the central electrode voltage, and R_m, R_1, and R_2 represent the dimensions of the trap (R_m, characteristic radius; R_1, central electrode radius; and R_2, outer electrode radius). **Equation 1.7** shows that the frequency of axial oscillation increases with 1) the magnitude of the central electrode voltage (*Ur*), and with 2) a dimensionally smaller trap or a change in the relative dimensions of the trap (R_2/R_1).[34,44,45,54] Increased axial frequency, in turn, results in increased resolution and dynamic range at the same analysis time, increased scan rate at the same resolution setting, and an increased tolerance to space charge effects.[34,54]

In the bench-top Orbitrap incarnations, the former strategy, of increasing *Ur*, is borne out.[37,38] By increasing the central electrode voltage from 3.5 to 5 kV, resolution per unit time increased approximately 20%.[37] While straightforward, this strategy is increasingly difficult to scale-up because of problems with maintaining and insulating voltages of such magnitude. Furthermore, because the frequency only scales as the square root of *Ur*, it is a rather low "return-on-investment" strategy.[34]

Work on the second strategy of modifying the trap's geometry, led to the development of the high-field Orbitrap by Makarov and colleagues.[34] In the high-field Orbitrap, the central

electrode radius, R_1, is increased from 6 mm to 9 mm, while the outer electrode radius, R_2, remained 15 mm.[34,54] Consulting **Equation 1.7**, we see that this change reduces the dominance of the first term of the denominator such that the axial frequency increases by a factor 1.4. In conjunction with increasing the central electrode voltage as mentioned before, this results in a factor of 1.7 frequency increase over the standard Orbitrap.[34] As a consequence of the increased field strength and decreased trapping volume, several other modifications were necessary to maintain performance: 1) the split between the two outer electrodes was increased to reduce the capacitance in the detector circuit and increase signal-to-noise due to the decreased detected image current on the outer electrodes because of the larger central electrode, 2) the decreased trapping volume required larger pumping apertures at the ends of the outer electrode to ensure proper vacuum in the trapping volume, and 3) the acceleration voltage for injecting ions into the Orbitrap had to be increased 3-3.5-fold such that a total redesign of the c-trap electronics was necessary.[34]

The result of these efforts was an 80% increase in resolution at the same acquisition time, thus making resolutions greater than 600,000 at *m/z* 195 and 380,000 at *m/z* 525 possible at a 3 s acquisition time.[34] In comparison to FT-ICR, the high-field Orbitrap overtakes the resolution of a 7T FT-ICR by *m/z* 280 and of a 15T FT-ICR by *m/z* 1300 at fixed transient length, and is much more tolerant of space charge effects.[34,54] The authors also notably demonstrated a single-shot dynamic range of 25,000 and isotopic resolution of intact proteins exceeding 40 kDa.[34,46,54] However, optical alignment, pressure requirements, and tuning parameters were also found to be more demanding than the standard Orbitrap.[34]

In tandem with the high-field Orbitrap,[34] advanced methods for processing the Orbitrap transient have also been developed that further push the limit of achievable resolution and scan rate in Orbitrap mass spectrometry.[35,43-45] Known commercially as “enhanced FT”,[43] this process

of advanced signal processing (ASP) is based on prior work in the FT-ICR community.[71,72] In general, Fourier-transformation (FT) of a time-domain transient serves as a base for determining the frequency, amplitude (magnitude), and phase information from the damped sinusoidal signals making up the transient.[71] With a phase of zero or some multiple of ±π/2, FT of a damped cosine or sine time-domain signal results in pure absorption (A(ω)) and dispersion spectra (D(ω)) (**Figure 1.3A**).[43,71,72] With any other phase, however, FT results in a linear combination of the absorption and dispersion spectra (mixed mode spectra, **Figure 1.3B**) that can be represented by four types of frequency domain spectra: real (Re), imaginary (Im), magnitude (M) and phase (φ).[71] For a complex frequency-domain spectrum F(ω), these representation are given by the following expressions (**Equation 1.8**):

$$\mathrm{Re}[F(\omega)] = \cos\varphi_0 \mathrm{A}(\omega) + \sin\varphi_0 \mathrm{D}(\omega)$$

$$\mathrm{Im}[F(\omega)] = -\sin\varphi_0 \mathrm{A}(\omega) + \cos\varphi_0 \mathrm{D}(\omega)$$

$$\mathrm{M}(\omega) = \sqrt{(\mathrm{Re}[F(\omega)])^2 + (\mathrm{Im}[F(\omega)])^2} = \sqrt{\mathrm{A}(\omega)^2 + \mathrm{D}(\omega)^2}$$

$$\varphi(\omega) = \tan^{-1}\left(\frac{\mathrm{Im}[F(\omega)]}{\mathrm{Re}[F(\omega)]}\right)$$

Equation 1.8

where φ_0 is the initial phase of all ions, independent of frequency (*m/z*). The magnitude mode spectrum is the almost always used in FT-MS,[28,43,71,72] because the real and imaginary components generally form asymmetric peaks unless the phase of the peak is 0 or a multiple of ±π/2 (**Figure 1.3**).[43,71,72] Since the phase is not generally known for every peak, the magnitude mode spectrum, which is independent of phase variation with frequency (*m/z*), is used. However,

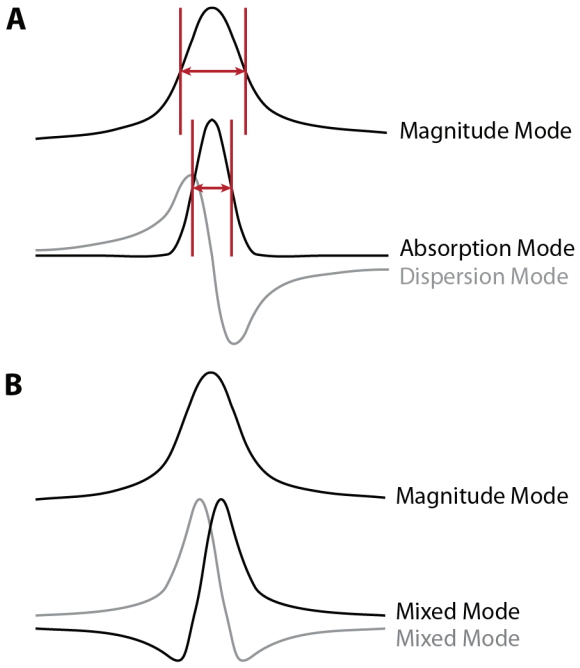


Figure 1.3. (**A**) Pure absorption and dispersion spectra resulting from FFT when signal phase is 0° (i.e., pure cosine) at the start of detection. The corresponding magnitude mode representation is shown above, as well as ranges (in red) denoting the peak full-width-half-maximum in comparison to the absorption mode spectrum. (**B**) Mixed-mode spectra resulting from signal phase of 45° at the start of detection, with corresponding magnitude mode spectrum shown above. Magnitude modes spectra are independent of phase, and are the same in **A** and **B**. Adapted from Beu, *et al*.[5]

neglecting the phase information generated by the FT by using a magnitude spectrum results in reduced resolution compared to the pure absorption mode spectrum by a factor of 2 to 3.5 (as seen in the comparison of magnitude and absorption peak FWHM in **Figure 1.3A**).[43,71] It is obvious then, that if phase information can be used, the higher resolution of real component spectra can be leveraged.

In Orbitrap mass spectrometry, collisionally cooled ions stored in the c-trap all have roughly the same phase prior to their injection (and therefore excitation) into the Orbitrap.[43] This is fundamentally different from FT-ICR where all ions of all *m/z* have the same phase at the *conclusion* of excitation.[71] Thus, if the time it takes for ions to travel from the c-trap to the Orbitrap is known, the phase accumulated during injection ($\Delta\varphi$) can be determined and the absorption spectrum can be recovered from the real and imaginary components (**Equation 1.9**):[43]

$$\mathrm{A}(\omega) = \cos(\varphi(\omega) - \Delta\varphi)\mathrm{Re}[\mathrm{F}(\omega)] - \sin(\varphi(\omega) - \Delta\varphi)\mathrm{Im}[\mathrm{F}(\omega)]$$

Equation 1.9

Fortuitously, a simple phase relationship exists that is independent of *m/z* in the first-order approximation. Since time-of-flight (t_{inj}) from the c-trap to the Orbitrap is proportional to the square-root of *m/z* and the axial frequency of oscillation (ω) is inversely proportional to the square-root of *m/z*, *m/z* cancels out, and the accumulated phase from injection is a constant given by (**Equation 1.10**):

$$\Delta\varphi = \omega \cdot t_{inj} = \frac{L_{eff}}{\sqrt{\frac{2eV}{k}}} = constant$$

Equation 1.10

where, L_{eff} is the effect length from c-trap to Orbitrap. Thus, A(ω) can be calculated in a straightforward manner.[43]

This process has been implemented on the Orbitrap by Lange and colleagues.[43] In practice, two factors are critical to achieving high quality spectra: 1) precise synchronization of the detection event to the exact time injection began, and 2) commencing detection as early as possible after injection. The former is achieved by extrapolating detected signals back in time until their phase differences are minimized and considering this minimum to be the start of the injection/excitation process.[43] The latter was achieved through hardware upgrades to the preamplifier and Orbitrap.[43] By improving the stabilization of the central electrode voltage following electrodynamic squeezing, the 8-9 ms delay usually imposed before the start of detection (to protect the pre-amplifier circuit) was reduced to less than 0.6 ms.[35,43] This advanced processing, implemented in the new bench-top and hybrid configurations of the Orbitrap mass analyzer,[35,37] provides up to a 2-fold increase in resolution at the same transient length, without sacrificing spectral quality or mass accuracy.[43] ASP methods are used in all chapters, except **Chapter 3**, to increase resolution and scan rate.

SUMMARY OF WORK PRESENTED HERE

The concepts of mass accuracy, resolution, and isotopic abundance accuracy will play a major role in this dissertation as I describe the application of the Orbitrap mass analyzer to fields normally dominated by low resolution mass spectrometers: targeted proteomics, and gas chromatography/mass spectrometry. In **Chapter 2**, I report on a new analysis paradigm for targeted proteomics called parallel reaction monitoring, which utilizes mass accuracy, resolution, and selectivity of the quadrupole-Orbitrap LC/MS to improve the confidence of targeted proteomic identifications. **Chapters 3** and **4** deal with the development of Orbitrap

instrumentation for GC/MS; I first describe a proof-of-principle implementation (**Chapter 3**) and then, in **Chapter 4**, introduce a dedicated, applications-grade instrument, the GC/quadrupole-Orbitrap. Finally, in **Chapter 5**, I describe the application of the GC/quadrupole-Orbitrap to the determination of elemental compositions of unknowns in an untargeted metabolomics study of *Arabidopsis thaliana* metabolites. **Chapter 5** also introduces an advanced data-dependent acquisition method, MIDA, or molecular-ion direction acquisition, tailored specifically for derivatized metabolites.

REFERENCES

(1) Haas, W.; Faherty, B. K.; Gerber, S. A.; Elias, J. E.; Beausoleil, S. A.; Bakalarski, C. E.; Li, X.; Villén, J.; Gygi, S. P. Optimization and Use of Peptide Mass Measurement Accuracy in Shotgun Proteomics. *Mol Cell Proteomics* **2006**, *5*, 1326-1337.

(2) Mann, M.; Kelleher, N. L. Precision proteomics: the case for high resolution and high mass accuracy. *Proc Natl Acad Sci USA* **2008**, *105*, 18132-18138.

(3) Spengler, B. De novo sequencing, peptide composition analysis, and composition-based sequencing: A new strategy employing accurate mass determination by Fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass Spectr.* **2004**, *15*, 703-714.

(4) Spengler, B. Composition based de novo Sequencing (CBS) of unknown peptides: A new strategy employing high accuracy ion cyclotron resonance mass Spectrometry. *Mol. Cell. Proteomics* **2004**, *3*, S301-S301.

(5) Bailey, D. J.; Rose, C. M.; McAlister, G. C.; Brumbaugh, J.; Yu, P.; Wenger, C. D.; Westphall, M. S.; Thomson, J. A.; Coon, J. J. Instant Spectral Assignment for Advanced Decision Tree-Driven Mass Spectrometry. *Proc Natl Acad Sci U S A* **2012**, *In press.*

(6) Roth, M. J.; Forbes, A. J.; Boyne, M. T.; Kim, Y.-B.; Robinson, D. E.; Kelleher, N. L. Precise and Parallel Characterization of Coding Polymorphisms, Alternative Splicing, and Modifications in Human Proteins by Mass Spectrometry. *Mol Cell Proteomics* **2005**, *4*, 1002-1008.

(7) Qian, K. N.; Olmstead, W.; Rodgers, R. P.; Marshall, A. G.; Green, L. A. Heavy petroleum composition revealed by high Fourier transform ion cyclotron resonance mass spectroscopy. *Abstr. Pap. Am. Chem. Soc.* **2004**, *227*, U234-U234.

(8) Marshall, A. G.; Kim, S.; Purcell, J. M.; Schaub, T. M.; Smith, D. F.; Rodgers, R. P. Characterization of petroleum by high resolution field desorptionaonization and atmospheric pressure photoionization FT-ICR mass spectrometry. *Abstr. Pap. Am. Chem. Soc.* **2005**, *229*, U865-U865.

(9) Schaub, T. M.; Rodgers, R. P.; Marshall, A. G. Speciation of aromatic compounds in petroleum refinery streams by continuous flow field desorption ionization FT-ICR mass spectrometry. *Energ Fuel* **2005**, *19*, 1566-1573.

(10) Marshall, A. G.; Rodgers, R. P. Petroleomics: the next grand challenge for chemical analysis. *Acc Chem Res* **2004**, *37*, 53-59.

(11) Wu, Z. G.; Jernstrom, S.; Hughey, C. A.; Rodgers, R. P.; Marshall, A. G. Resolution of 10,000 compositionally distinct components in polar coal extracts by negative-ion electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Energ Fuel* **2003**, *17*, 946-953.

(12) Wu, Z.; Rodgers, R. P.; Marshall, A. G. Characterization of vegetable oils: detailed compositional fingerprints derived from electrospray ionization fourier transform ion cyclotron resonance mass spectrometry. *J Agr Food Chem* **2004**, *52*, 5322-5328.

(13) Stenson, A. C.; Marshall, A. G.; Cooper, W. T. Exact masses and chemical formulas of individual Suwannee River fulvic acids from ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectra. *Anal. Chem.* **2003**, *75*, 1275-1284.

(14) Little, J. L.; Williams, A. J.; Pshenichnov, A.; Tkachenko, V. Identification of "Known Unknowns" Utilizing Accurate Mass Data and ChemSpider. *J. Am. Soc. Mass Spectr.* **2012**, *23*, 179-185.

(15) Beynon, J. H. Qualitative Analysis of Organic Compounds by Mass Spectrometry. *Nature* **1954**, *174*, 735-737.

(16) Kim, S.; Rodgers, R. P.; Marshall, A. G. Truly "exact" mass: Elemental composition can be determined uniquely from molecular mass measurement at similar to 0.1 mDa accuracy for molecules up to similar to 500 Da. *Int. J. Mass Spectrom.* **2006**, *251*, 260-265.

(17) Kind, T.; Fiehn, O. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. *Bmc Bioinformatics* **2006**, *7*, -.

(18) Grange, A. H.; Donnelly, J. R.; Sovocool, G. W.; Brumley, W. C. Determination of elemental compositions from mass peak profiles of the molecular ion (M) and the M+1 and M+2 ions. *Anal. Chem.* **1996**, *68*, 553-560.

(19) Roussis, S. G.; Proulx, R. Reduction of chemical formulas from the isotopic peak distributions of high-resolution mass spectra. *Anal. Chem.* **2003**, *75*, 1470-1482.

(20) Stoll, N.; Schmidt, E.; Thurow, K. Isotope pattern evaluation for the reduction of elemental compositions assigned to high-resolution mass spectral data from electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1692-1699.

(21) Kind, T.; Fiehn, O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *Bmc Bioinformatics* **2007**, *8*, -.

(22) Erve, J. C. L.; Gu, M.; Wang, Y.; DeMaio, W.; Talaat, R. E. Spectral Accuracy of Molecular Ions in an LTQ/Orbitrap Mass Spectrometer and Implications for Elemental Composition Determination. *J. Am. Soc. Mass Spectr.* **2009**, *20*, 2058-2069.

(23) Pluskal, T.; Uehara, T.; Yanagida, M. Highly Accurate Chemical Formula Prediction Tool Utilizing High-Resolution Mass Spectra, MS/MS Fragmentation, Heuristic Rules, and Isotope Pattern Matching. *Anal. Chem.* **2012**, *84*, 4396-4403.

(24) Kind, T.; Wohlgemuth, G.; Lee do, Y.; Lu, Y.; Palazoglu, M.; Shahbaz, S.; Fiehn, O. FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal Chem* **2009**, *81*, 10038-10048.

(25) Thomas, R. A beginner's guide to ICP-MS - Part VII: Mass separation devices - Double-focusing magnetic-sector technology. *Spectrosc.* **2001**, *16*, 22-.

(26) Marshall, A. G.; Hendrickson, C. L. High-Resolution Mass Spectrometers. *Annu Rev Anal Chem* **2008**, *1*, 579-599.

(27) Perry, R. H.; Cooks, R. G.; Noll, R. J. Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass Spectrom Rev* **2008**, *27*, 661-699.

(28) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Reviews* **1998**, *17*, 1-35.

(29) Marshall, A. G. Milestones in Fourier transform ion cyclotron resonance mass spectrometry technique development. *Int. J. Mass Spectrom.* **2000**, *200*, 331-356.

(30) Brunnee, C. The Ideal Mass Analyzer - Fact or Fiction. *Int J Mass Spectrom* **1987**, *76*, 125-237.

(31) Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal Chem* **2000**, *72*, 1156-1162.

(32) Makarov, A. Mass spectrometer. **1999**, US Patent 5,886,356.

(33) Makarov, A.; Denisov, E.; Kholomeev, A.; Baischun, W.; Lange, O.; Strupat, K.; Horning, S. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* **2006**, *78*, 2113-2120.

(34) Makarov, A.; Denisov, E.; Lange, O. Performance Evaluation of a High-field Orbitrap Mass Analyzer. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1391-1396.

(35) Michalski, A.; Damoc, E.; Lange, O.; Denisov, E.; Nolting, D.; Mueller, M.; Viner, R.; Schwartz, J.; Remes, P.; Belford, M.; Dunyach, J.-J.; Cox, J.; Horning, S.; Mann, M.; Makarov, A. Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Mol Cell Proteomics* **2011**, *mcp.O111.013698*.

(36) Olsen, J. V.; Schwartz, J. C.; Griep-Raming, J.; Nielsen, M. L.; Damoc, E.; Denisov, E.; Lange, O.; Remes, P.; Taylor, D.; Splendore, M.; Wouters, E. R.; Senko, M.; Makarov, A.; Mann, M.; Horning, S. A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed. *Mol. Cell. Proteomics* **2009**, *8*, 2759-2769.

(37) Michalski, A.; Damoc, E.; Hauschild, J. P.; Lange, O.; Wieghaus, A.; Makarov, A.; Nagaraj, N.; Cox, J.; Mann, M.; Horning, S. Mass spectrometry-based proteomics using

Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol Cell Proteomics* **2011**, *10*, mcp.M111.011015.

(38) Geiger, T.; Cox, J.; Mann, M. Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation. *Mol Cell Proteomics* **2010**, *9*, 2252-2261.

(39) Hardman, M.; Makarov, A. A. Interfacing the orbitrap mass analyzer to an electrospray ion source. *Anal. Chem.* **2003**, *75*, 1699-1705.

(40) Strupat, K.; Kovtoun, V.; Bui, H.; Viner, R.; Stafford, G.; Horning, S. MALDI Produced Ions Inspected with a Linear Ion Trap-Orbitrap Hybrid Mass Analyzer. *J. Am. Soc. Mass Spectr.* **2009**, *20*, 1451-1463.

(41) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* **2007**, *4*, 709-712.

(42) McAlister, G. C.; Berggren, W. T.; Griep-Raming, J.; Horning, S.; Makarov, A.; Phanstiel, D.; Stafford, G.; Swaney, D. L.; Syka, J. E. P.; Zabrouskov, V.; Coon, J. J. A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer. *J. Proteome Res.* **2008**, *7*, 3127-3136.

(43) Lange, O.; Damoc, E.; Wieghaus, A.; Makarov, A. Enhanced FT for Orbitrap mass spectrometry. *Proc. 59th Conf. Amer. Soc. Mass Spectrom* **2011**, Denver, CO, June 5-9.

(44) Lange, O.; Makarov, A.; Denisov, E.; Balschun, W. Accelerating spectral acquistion rate of Orbitrap mass spectrometry. *Proc. 58th Conf. Amer. Soc. Mass Spectrom* **2010**, Salt Lake City, UT, May 23-27.

(45) Makarov, A. Accelerating spectral acquisition rate of Orbitrap mass spectrometry *Proc. 58th Conf. Amer. Soc. Mass Spectrom* **2010**, Salt Lake City, UT, May 23–27.

(46) Denisov, E.; Damoc, E.; Makarov, A.; Lange, O. Orbitrap mass spectrometry with resolving powers above 500,000 and 1,000,000 on a chromatographic time scale. *Proc. 59th Conf. Amer. Soc. Mass Spectrom* **2011**, Denver, CO, June 5-9.

(47) Olsen, J. V.; de Godoy, L. M. F.; Li, G. Q.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **2005**, *4*, 2010-2021.

(48) Wenger, C. D.; McAlister, G. C.; Xia, Q.; Coon, J. J. Sub-part-per-million precursor and product mass accuracy for high-throughput proteomics on an ETD-enabled orbitrap mass spectrometer. *Mol Cell Proteomics* **2010**.

(49) Makarov, A.; Denisov, E.; Lange, O.; Horning, S. Dynamic Range of Mass Accuracy in LTQ Orbitrap Hybrid Mass Spectrometer. *J. Am. Soc. Mass Spectr.* **2006**, *17*, 977-982.

(50) Peterson, A. C.; McAlister, G. C.; Quarmby, S. T.; Griep-Raming, J.; Coon, J. J. Development and characterization of a GC-enabled QLT-Orbitrap for high-resolution and high-mass accuracy GC/MS. *Anal Chem* **2010**, *82*, 8618-8628.

(51) Kingdon, K. H. A method for the neutralization of electron space charge by positive ionization at very low gas pressures. *Phys Rev* **1923**, *21*, 408-418.

(52) Lewis, R. R. Motion of Ions in the Kingdon Trap. *J Appl Phys* **1982**, *53*, 3975-3980.

(53) Sekioka, T.; Terasawa, M.; Awaya, Y. Ion Storage in Kingdon Trap. *Radiat Eff Defect S* **1991**, *117*, 253-259.

(54) Makarov, A. In *Practical Aspects of Trapped Ion Mass Spectrometry, Volume IV*; CRC Press: 2010; pp 251-272.

(55) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Cooks, R. G. The Orbitrap: a new mass spectrometer. *J Mass Spectrom* **2005**, *40*, 430-443.

(56) Yang, L. S.; Church, D. A. Confinement of Injected Beam Ions in a Kingdon Trap. *Nucl Instrum Meth B* **1991**, *56-7*, 1185-1187.

(57) Knight, R. D. Storage of Ions from Laser-Produced Plasmas. *Appl Phys Lett* **1981**, *38*, 221-223.

(58) Gillig, K. J.; Bluhm, B. K.; Russell, D. H. Ion motion in a Fourier transform ion cyclotron resonance wire ion guide cell. *Int J Mass Spectrom* **1996**, *157*, 129-147.

(59) Solouki, T.; Gillig, K. J.; Russell, D. H. Detection of High-Mass Biomolecules in Fourier-Transform Ion-Cyclotron Resonance Mass-Spectrometry - Theoretical and Experimental Investigations. *Anal. Chem.* **1994**, *66*, 1583-1587.

(60) Makarov, A. The Orbitrap: a novel high-performance electrostatic trap. *Proc. 47th Conf. Amer. Soc. Mass Spectrom* **1999**, Dallas, TX, June 13-17.

(61) Oksman, P. A Fourier-Transform Time-of-Flight Mass-Spectrometer - a Simion Calculation Approach. *Int J Mass Spectrom* **1995**, *141*, 67-76.

(62) de Hoffmann, E. Tandem mass spectrometry: a primer. *Journal of Mass Spectrometry* **1996**, *31*, 129-137.

(63) Makarov, A.; Denisov, E.; Jung, G.; Balschun, W.; Horning, S. Improvements in an electrostatic trap. **2006**, Patent WO2006/129109.

(64) Hu, Q.; Makarov, A. A.; Cooks, R. G.; Noll, R. J. Resonant ac dipolar excitation for ion motion control in the orbitrap mass analyzer. *J Phys Chem A* **2006**, *110*, 2682-2689.

(65) Bolotskikh, P. A.; Grinfeld, D. E.; Makarov, A. A.; Monastyrskiy, M. A. Coulomb dynamics of ion bunches in multi-reflection electrostatic traps. *Nuclear Instruments and*

Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **2011**, *645*, 146-152.

(66) Zajfman, D.; Heber, O.; Rappaport, M. L.; Pedersen, H. B.; Strasser, D.; Goldberg, S. Self-bunching effect in an ion trap resonator. *J. Opt. Soc. Am. B* **2003**, *20*, 1028-1032.

(67) Gallien, S.; Duriez, E.; Crone, C.; Kellmann, M.; Moehring, T.; Domon, B. Targeted Proteomic Quantification on Quadrupole-Orbitrap Mass Spectrometer. *Mol. Cell. Proteomics* **2012**, *11*, 1709-1723.

(68) Wang, J.; Bourne, P. E.; Bandeira, N. Peptide identification by database search of mixture tandem mass spectra. *Mol. Cell. Proteomics* **2011**, *10*.

(69) Fornelli, L.; Damoc, E.; Thomas, P. M.; Kelleher, N. L.; Aizikov, K.; Denisov, E.; Makarov, A.; Tsybin, Y. O. Top-down analysis of monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS. *Mol. Cell. Proteomics* **2012**.

(70) Wenger, C. D.; Lee, M. V.; Hebert, A. S.; McAlister, G. C.; Phanstiel, D. H.; Westphall, M. S.; Coon, J. J. Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nature Methods* **2011**, *8*, 933-935.

(71) Vining, B. A.; Bossio, R. E.; Marshall, A. G. Phase correction for collision model analysis and enhanced resolving power of fourier transform ion cyclotron resonance mass spectra. *Anal. Chem.* **1999**, *71*, 460-467.

(72) Beu, S. C.; Blakney, G. T.; Quinn, J. P.; Hendrickson, C. L.; Marshall, A. G. Broadband phase correction of FT-ICR mass spectra via simultaneous excitation and detection. *Anal. Chem.* **2004**, *76*, 5756-5761.

CHAPTER 2

Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics

SUMMARY

Selected reaction monitoring (SRM) on a triple quadrupole (QqQ) mass spectrometer is currently experiencing a renaissance within the proteomics community for its, as yet, unparalleled ability to characterize and quantify a set of proteins reproducibly, completely, and with high sensitivity. Given the immense benefit that high resolution and accurate mass (HR/AM) instruments have brought to the discovery proteomics field, we wondered if highly accurate mass measurement capabilities could be leveraged to provide benefits in the targeted proteomics domain as well. Here, we propose a new targeted proteomics paradigm centered on the use of next generation, quadrupole-equipped HR/AM instruments: parallel reaction monitoring (PRM). In PRM, the third quadrupole of a QqQ is substituted with a HR/AM mass analyzer to permit the parallel detection of all target product ions in one, concerted high resolution mass analysis. We detail the analytical performance of the PRM method, using a quadrupole-equipped bench-top Orbitrap MS, and draw a performance comparison to SRM in terms of run-to-run reproducibility, dynamic range, and measurement accuracy. In addition to requiring minimal upfront method development and facilitating automated data analysis, PRM yielded quantitative data over a wider dynamic range than SRM in the presence of a yeast background matrix due to PRM's high selectivity in the mass-to-charge domain. With achievable linearity over the quantifiable dynamic range found

to be statistically equal between the two methods, our investigation suggests that PRM will be a promising new addition to the quantitative proteomics toolbox.

INTRODUCTION

The most widespread protein sequencing technique is the shotgun method. Proteins are digested into peptides, chromatographically separated, and measured by mass spectrometry (MS).[1-10] Many types of mass spectrometers are used – quadrupole ion traps (QIT), QIT hybrids such as the QLT (quadrupole linear ion trap)-Orbitrap or QLT-FT-ICR, and quadrupole time-of-flight (Q-TOF) hybrids – but, the experiments, from the MS measurement onward, are basically identical: the masses of eluting cationic peptide precursors are measured in a MS scan, and the most abundant precursors are selected in series for successive tandem MS events (MS/MS). This process, called data-dependent acquisition (DDA), continues for the duration of a chromatographic separation, and constant MS operation in this manner can generate hundreds of thousands of spectra in days. These spectra are then mapped to peptide or protein sequence databases using highly-evolved database search algorithms.[11-13] Successful results can be obtained within just a few days and are nothing short of spectacular: tens of thousands of unique peptide spectral matches mapping to several thousand unique protein isoforms have become the norm. While this approach certainly can achieve ultra-high-throughput, it is unfortunately lacking in sensitivity and reproducibility. Specifically, complete coverage of specific biological pathways or functional groups is not typical (i.e., all 500 kinases, 1,400 transcription factors, *etc.*). Likewise, the overlap of identifications in replicate experiments is low (35-60%).[14,15]

The limitations of the shotgun method have propelled a recent fervor in targeted proteomic methods – namely, selected reaction monitoring (SRM, also known as MRM, multiple reaction monitoring).[16-21] SRM achieves the reproducibility and sensitivity that the shotgun approach lacks and even offers a route to determine absolute abundance.[21] By offering superior consistency, completeness, and quantitative accuracy, targeted experiments afford a new avenue to test and generate biological hypotheses.[15] SRM, primarily performed on triple quadrupole

(QqQ) MS, has emerged as the MS "gold-standard" for targeted proteomics[21] and is a broadly accepted alternative to the traditional multiplexed immunoassay.[20] Numerous studies have successfully employed QqQ SRM for both absolute and relative quantitation of analytes in applications ranging from clinical diagnostics[22-27] to whole systems analyses.[28-30] The rising popularity and promise of the SRM technique have also spawned a plethora of new analysis approaches and software tools. For example, many algorithms and software tools have been developed to expedite assay development by aiding in the selection of proteotypic peptides,[31] peptide transitions, and instrument parameters (reviewed in Cham Mead, *et al.*[32]). Additionally, through community effort, several publicly available databases of tandem MS spectra[33-39] and validated SRM assays[40,41] are available to provide empirical guides to transition selection and assay development. Several groups have also presented elegant solutions to maximize instrument duty cycle and improve assay specificity. Picotti and colleagues,[42] for instance, demonstrated use of synthetic peptide libraries to accelerate SRM assay development. Likewise, Kiyonami, *et al*.[43] recently introduced a strategy that boosted SRM bandwidth by restricting the number of transitions per target acquired before and after target elution. By limiting the acquisition of full transition sets to a few occasions during peptide elution, this strategy enabled simultaneous qualitative and quantitative analysis of 6,000 transitions in one hour, a substantial boon to SRM throughput.

In recent years, discovery-based proteomic methods, such as the shotgun method discussed above, have been transformed by significant advancements in instrumentation; key figures of merit, such as sensitivity, duty cycle, mass accuracy, and mass resolution have seen remarkable improvements.[44] While these developments have done little to directly curtail the reproducibility issues that the SRM method so effectively counters, the achievable depth of proteomic sampling (i.e., analytical sensitivity) within the discovery context continues to improve.

Central to this evolution is the increased performance, and availability, of high resolution and accurate mass (HR/AM) instrumentation.[44] Namely, developments in time-of-flight (TOF) technology,[45] and the advent of the Orbitrap mass analyzer in 2005,[46-49] have made fast and sensitive MS/MS scanning[50-53] with <10 parts-per-million (ppm) mass measurement error routine.[54] For discovery experiments, the ability to acquire MS/MS scans with high resolution and low-ppm mass errors offers several advantages, including higher confidence sequence identification,[44,55] post-translational modification site localization,[56] and improved quantitative accuracy. Coming from this perspective, we wondered whether the highly accurate mass measurement capabilities of today's new-generation MS instrumentation could be leveraged to provide benefits within the targeted proteomics domain.

Driving our inquiry was the newly-introduced Q Exactive bench-top quadrupole-Orbitrap MS (QqOrbi),[47] which, along with quadrupole-TOF (QqTOF) instrumentation,[45] possesses a geometry essentially equivalent to a QqQ, except that the third quadrupole of the QqQ is replaced by an Orbitrap (or TOF) analyzer (**Figure 2.1A-B)**. The QqOrbi achieves a 12 Hz scan rate at a resolution of 17,500 for both MS and MS/MS scanning, quadrupole mass filter isolation with mass windows as small as ±0.2 Th, and mass measurement errors typically <1 ppm with internal calibration and <5 ppm with external calibration.[47] With these performance characteristics, we envisioned a targeted proteomics strategy where all products of a target peptide are simultaneously monitored under conditions that offer high resolution and high mass accuracy. Operation would be identical to a SRM scan except that all transitions would be co-detected and distinguished from one another, and from background, by the final mass analysis stage. We call this mode of operation parallel reaction monitoring (PRM).

The PRM technique has several potential advantages over the traditional SRM approach. First, PRM spectra would be highly specific since all potential product ions of a peptide, instead

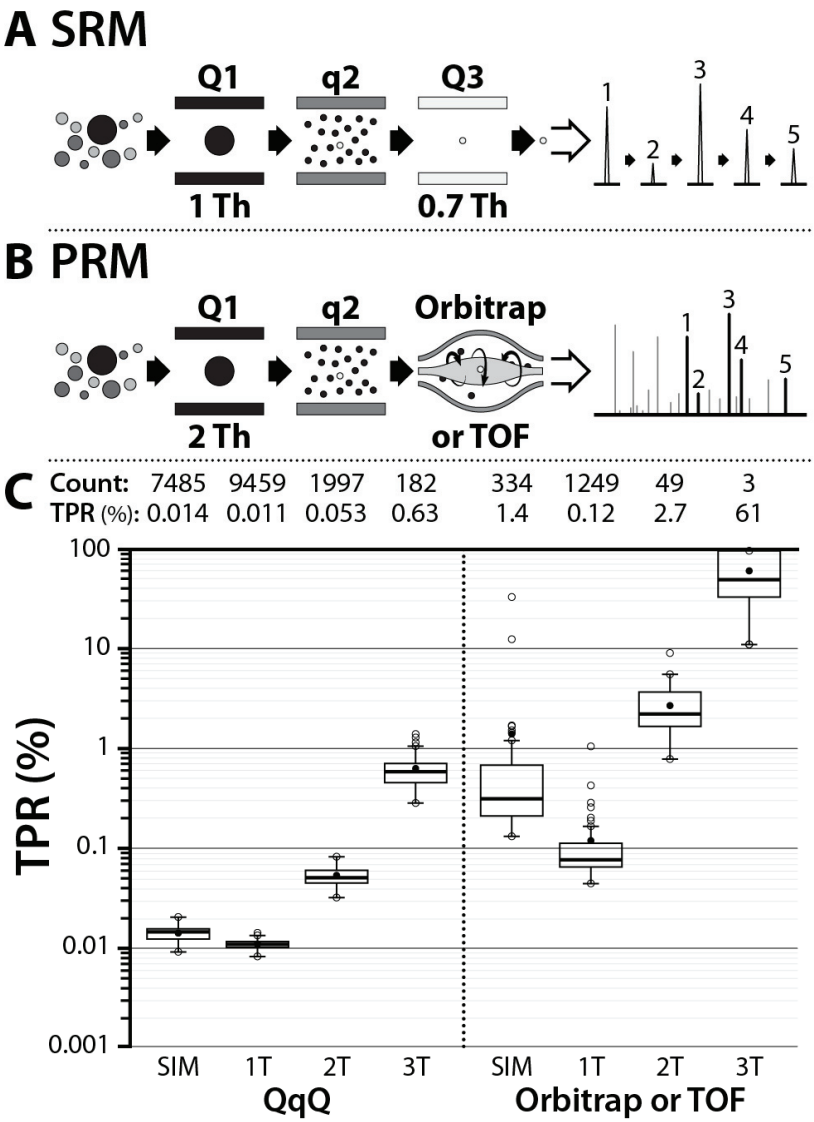


Figure 2.1. Schematic representation of SRM **(A)** and PRM **(B)** as performed on QqQ and QqOrbi instrumentation, respectively. In SRM **(A)**, each product ion transition (white circle) is serially monitored (from 1 to 5) one at a time in distinct scans. In PRM **(B)**, all product ion transitions (1-5, and all possible product ions, shown as black circles) are analyzed/monitored in one concerted, high resolution and high mass accuracy mass analysis. Q1 and Q3 refer to the first and third mass-resolving quadrupoles of the QqOrbi (Q1 only) and QqQ, and q2 to the quadrupole (or cell, in the QqOrbi case) in which beam-type CAD is performed. The isolation widths employed for each device in both experimental and theoretical data are given below each device. **(C)** Theoretical comparison of the rate of correctly identifying a target peptide (as true positive rate in percent, TPR) from all theoretically possible peptides in the human tryptic peptidome in SIM and reaction monitoring experiments in which 1, 2, and 3 *y*-ion transitions (labeled as 1T, 2T, and 3T, respectively) are monitored for Orbitrap instruments (±5 ppm) and QqQ (±250 ppm) for the 25 peptides used in this study in their light and heavy forms (50 total). Count refers to the average number of possible confounding species, including the target peptide, represented by the boxplots. TPR is calculated as 100/count.

of just 3-5 transitions, are available to confirm the identity of the peptide.[57,58] Second, PRM could provide a higher tolerance for co-isolated background peptides/species. Since numerous ions would be available for identification and quantitation purposes, the presence of interfering ions in a full mass spectrum would be less disruptive to overall spectral quality than interference in a narrow mass range, especially since high resolution can often separate these ions from the product of interest. Note that one could extend this concept to a multiplexed PRM scan where the product ions of several target peptides are co-mingled and detected in a single-scan.[47] And third, since PRM monitors all transitions, one need not have prior knowledge of, or pre-select, target transitions before analysis. These points suggest another potential advantage of the PRM approach: elimination of much of the effort required to develop and optimize the traditional SRM assay.

Given that a QqQ possesses a duty cycle approaching 100% and utilizes electron multiplier-based detection, which is inherently more sensitive than image current-based detection (Orbitrap), it is not obvious that the PRM method will afford sensitivity comparable to the current state-of-the-art SRM approach. However, we postulate that what PRM lacks in sensitivity and duty cycle might be effectively countered by the selectivity of HR/AM measurement. Here, we implement PRM on a QqOrbi system and benchmark method performance with triplicate analysis of 25 isotopically heavy-labeled synthetic peptides spanning a concentration range of 10^5 under neat and matrix-containing conditions. We assess key figures of merit, including data quality, run-to-run reproducibility/precision, dynamic range/sensitivity, and measurement accuracy/linearity. Finally, we draw a performance comparison to SRM operating on a common QqQ platform.

EXPERIMENTAL PROCEDURES

Materials and reagents. Unless otherwise specified, all reagents used herein were purchased from Sigma Aldrich (St. Louis, MO). Acetonitrile was purchased from Fisher Scientific (Fair Lawn, NJ) and formic acid (>99%) from Thermo Fisher Scientific (TFS, Rockford, IL). Ultrapure water was supplied from a Barnstead Nanopure Diamond ultrapure water system (resistivity 18.2 MΩ-cm; TFS, Dubuque, Iowa).

Sample preparation. Twenty-five heavy-labeled hypothetical tryptic human peptides (**Table 2.1**) were synthesized by Fmoc solid-phase synthesis, purified by HPLC, and solubilized in 5% v/v CH_3CN/water at a concentration of 5 pmol/µL ± 25% with purity >97% (HeavyPeptides AQUA QuantPro; TFS, Ulm, Germany). All 25 peptides (QqOrbi experiments) or a selection of 14 of the 25 peptides (QqQ experiments) were mixed and used neat with intact bovine serum albumin (BSA) carrier protein (200 fmol/µL in 0.1% HCOOH/water) or spiked into a whole-cell tryptic digest of yeast (1 µg/µL in 0.1% HCOOH/water), prepared as previously reported,[59] at the following concentrations (fmol/µL or nM): 0.002, 0.02, 0.2, 2, 20, and 200. A one microliter aliquot of each sample was subjected to C_{18}-reversed phase liquid chromatography prior to mass spectrometry.

L**iquid chromatography.** For QqOrbi experiments, samples (1 µL) were loaded and desalted on an 8 cm packed pre-column (75 µm ID x 360 µm OD) in 0.2% HCOOH/water (buffer A) at a flow rate of 1 µL/min for 10 min. Analytical separation was performed over a 20 cm packed column (50 µm ID x 360 µm OD) at 300 nL/min with a 60 min gradient of increasing CH_3CN (buffer B, 0.2% HCOOH/CH_3CN) using a nanoAquity HPLC and autosampler (Waters, Milford, MA). Both pre-column and analytical column were packed with Magic C_{18}-reversed phase silica (5 µm diameter, 100 Å pore size; Michrom Bioresources, Auburn, CA) using a pressure bomb. Following sample loading, buffer B was increased rapidly to 5% over 30 s and

Table 2.1. Peptides and instrument parameters used in this study.

#	Peptide*	PRM Parameters			SRM Parameters			
		Mono. Mass	Parent *m/z*	Sched. RT (min)	Parent *m/z*	Transition *m/z*	CE (%)	Sched. RT (min)
1	GVSAFSTWE**k**	1118.5488	560.2817	28.0-33.0	560.584	658.646 805.821 963.977	26	35.0-89.0
2	HFLTLAPI**k**	1046.6368	524.3257	26.5-31.0				
3	ARPA**c**VDA**r**	1024.5112	513.2653	0.0-10.0				
4	SGWT**c**TQPGG**r**	1281.5753	608.7738	13.0-20.0				
5	EGQLAAGT**c**EIVTLD**r**	1741.8544	871.9345	29.0-36.0	872.432	1174.245 1245.323 1316.402	30	37.5-89.0
6	LWSLAEIATSDL**k**	1476.5364	727.9027	40.0-53.0				
7	SEDEDEDEGDAT**r**	1648.8118	739.2755	7.0-14.0				
8	DIQFGSQI**k**	1042.5539	522.2842	23.0-29.0	522.557	540.555 687.731 815.861	25	29.0-37.5
9	HTGTPLGDIPYG**k**	1436.6875	682.3584	22.5-27.0				
10	NSWGTDWGE**k**	1215.5330	594.2640	25.5-30.5	594.580	743.708 800.760 986.971	27	29.0-37.5
11	FSDLTEEEF**r**	1362.7023	641.7949	27.0-31.5	642.134	719.675 820.762 933.920	28	29.0-89.0
12	SFEGTDYG**k**	1010.4436	506.2291	13.0-20.0	506.491	591.556 648.608 777.723	25	22.0-29.0
13	TLNGIQLA**r**	994.5799	498.2972	21.0-26.5	498.544	610.663 667.715 781.818	26	22.0-35.0
14	AFSQNSVLI**k**	1113.6274	557.8210	22.5-27.0				
15	WPGYLNGG**r**	1028.5067	515.2606	26.5-31.0	515.532	689.677 746.729 843.845	25	29.0-89.0
16	GALDGEAP**r**	894.4435	448.2290	10.0-20.0	448.441	539.498 654.586 767.745	23	1.0-29.0
17	AETLVQA**r**	896.4955	449.2550	10.0-20.0	449.471	483.478 596.636 697.741	23	1.0-29.0
18	FLNPEW**k**	940.4898	471.2522	26.5-31.0				
19	LEQNPEESQDI**k**	1453.7908	719.3510	10.0-20.0				
20	SEASSSPPVVTSSSHS**r**	1186.5135	571.2756	10.0-15.0	571.569	583.511 771.693 870.825	26	1.0-22.0
21	AP**c**QAGDL**r**	996.4686	499.2416	10.0-15.0	499.520	541.514 669.644 829.839	24	1.0-22.0
22	TWFQNQ**r**	988.4754	495.2450	20.0-26.0				
23	TVFSSTQL**c**VLND**r**	1710.8048	825.4132	31.0-40.0	825.891	899.973 1129.207 1216.285	29	37.5-89.0
24	FSEVSAD**k**	889.4273	445.7209	10.0-17.0	445.944	527.513 656.628 743.706	22	1.0-29.0
25	GLYEGTG**r**	861.4220	431.7183	10.0-20.0				

c - carbamidomethylcysteine (+ C_2H_3NO) **r** - heavy labeled arginine ($^{13}C_6{}^{15}N_4$) **k** - heavy labeled lysine ($^{13}C_6{}^{15}N_2$)
*All peptides have a charge state (*z*) of +2, with the exception of peptide #20 (*z* = +3).

then shallowly to 25% over 59.5 min, followed by a quick increase to 80% over 1 min. After holding at 80% buffer B for 5 min, buffer B was decreased over 2 min to 0% and the column was re-equilibrated at 100% A for 22 min.

For QqQ experiments, similar LC conditions were used with some notable changes. Pre-column and analytical column were 5 cm and 12 cm in length (75 µm ID x 360 µm OD), respectively, and were packed with 3 µm Magic C_{18}-reversed phase silica. Following sample loading and desalting as above, buffer B was increased to 5% over 30 s and then to 35% over 59.5 min, followed by a quick increase to 75% over 1 min. Following 5 min at 75% buffer B, buffer B was decreased over 1 min to 0% for a 23 min column equilibration.

Mass spectrometry. In all experiments, HPLC eluent was introduced into the mass spectrometer via an integrated electrospray emitter[60] (pulled in-house via laser micropipette puller; Sutter Instrument, Novato, CA) operated at 2.0-2.2 kV and coupled to a custom nano-ESI source. QqOrbi experiments were performed on a quadrupole mass filter-equipped bench-top Orbitrap mass spectrometer (Q Exactive, TFS, Bremen, Germany). Each sample (6 concentrations x 2 background matrix conditions) was analyzed in triplicate under two mass spectrometric conditions: 1) PRM with an isolation width of ±1 Th, and 2) PRM with an isolation width of ±0.2 Th. Samples were analyzed in order of increasing concentration with an extensive column wash between each concentration set to minimize carry-over. In all experiments, a full mass spectrum at 70,000 resolution relative to *m/z* 200 (AGC target $1x10^6$, 250 ms maximum injection time, *m/z* 200-2000) was followed by up to 25 PRM scans at 17,500 resolution (AGC target $2x10^5$, 120 ms maximum injection time) as triggered by a scheduled inclusion list (**Table 2.1**). Ion activation/dissociation was performed by beam-type CAD at a normalized collision energy of 25% in a higher-energy c-trap dissociation (HCD) collision cell. Instrument spectral

mass accuracy was checked and calibrated after each concentration set (approximately every 1.5 days). PRM parameters were optimized and selected from prior experiments (**Appendix 6.1**).

QqQ experiments were performed on a TSQ Quantum Discovery Max (TFS, Austin, TX). Each sample was analyzed in triplicate in order of increasing concentration, targeting a selected set of 14 peptides in scheduled SRM mode (see **Table 2.1** for transitions, collision energies, and scheduling). For each transition, a 35 ms dwell time, Q1 and Q3 selectivities of 1.0 and 0.7 Th (FWHM), respectively, and Q2 collision gas pressure of 1.0 mTorr argon were employed. Collision energies (CE) for each peptide precursor were individually optimized in prior experiments (data not shown) using the following empirically derived formula as a reference point: CE = 0.025 x precursor *m/z* + 12.

Data analysis. PRM data were manually curated within Xcalibur Qual Browser (version 2.2.0.23; TFS, San Jose, CA) and through use of an internally developed script, ElutionProfiler. Elution Profiler developed in C# with Microsoft Visual Studio 2010 and the Microsoft.NET Framework version 3.5 (Redmond, WA) and is available on our website at http://www.chem.wisc.edu/~coon/software.php. Access to data in the proprietary TFS .raw file format was enabled by the XRawfile Component Object Model (COM) library (XRawfile2.dll, installed automatically with Thermo Xcalibur). ElutionProfiler used .raw files as input to generate an extracted score chromatogram (XSC) for each PRM spectrum. The spectral score was calculated based on all present, sequence-specific *b*- and *y*-ions using the following formula (**Equation 2.1**):

$$Score = \sum_{n}^{k} (0.25\delta_b(n) + \delta_y(n))n$$

Equation 2.1

where the result of the Dirac delta functions, δ_b and δ_y, is 1 if the n^{th} ion in the spectrum is a *b*- or *y*-ion with <5 ppm mass error, respectively, and 0 in all other cases. *n* is the product ion number (e.g., 4 for a y_4 or b_4 ion), *k* is the length of the peptide (number of amino acids), and 0.25 is an arbitrarily-chosen scalar to weight *b*-ions (not containing an isotopically heavy-labeled arginine or lysine at the c-terminus) less than equivalently numbered *y*-ions. XICs were generated using the summed intensity of all possible *b*- and *y*-product ions for a particular peptide, extracted at a ±10 ppm mass tolerance. Detection was based on the presence of product ion signals in at least 2 of 3 replicates within ±3σ min of the expected retention time, mass error within ±5 ppm, chromatographic signal-to-noise ≥3, and the presence of a combination of product ions in at least one spectrum with a score meeting or exceeding a peptide-specific threshold equal to the length of the targeted peptide (*k*) in all cases.

For each analysis, individual extracted ion chromatograms (XICs) using a ±10 ppm mass tolerance, unless otherwise noted, were generated for all twenty-five peptides. The ICIS peak detection algorithm was used to calculate peak area-under-the-curve (AUC) and chromatographic signal-to-noise. Expected retention time and product ion abundances were based on the average retention time and product ion abundances observed in the most concentrated (200 nM) analysis within each respective experiment set. In SIM-only analyses, XICs were generated using the monoisotopic precursor mass. Detection of the targeted peptide was based on the presence of the intact precursor signal in at least 2 of 3 replicates within ±1 min of the expected retention time, chromatographic signal-to-noise ≥3, mass error within ±10 ppm, and manual spectral verification.

QqQ data were first processed in Skyline[61] (version 1.1.0.2905) to calculate composite peak AUCs after Savitzky-Golay smoothing. Each transition was then manually inspected within Qual Browser using a ±500 ppm mass tolerance and 7-point boxcar smoothing to generate XICs for each transition. The ICIS peak detection algorithm was employed, as before, for signal-to-

noise determinations. Target peptide detection was based upon meeting at least 8 of 10 acceptance criteria in at least 2 of 3 replicates. The acceptance criteria were as follows: 1) appearance of three co-eluting transitions within ±2.5 min of the expected retention time, 2-4) an average individual transition *m/z* over the elution duration within ±500 ppm of the expected *m/z*, 5-7) an average individual transition relative abundance over the elution duration within ±15% of the expected abundance, and 8-10) an individual transition elution profile with signal-to-noise ≥3. In cases where the ICIS algorithm did not detect an elution peak, the transition was assigned a signal-to-noise of unity.

Raw AUCs were used to calculate run-to-run reproducibility metrics (as percent relative standard deviation, %RSD) for triplicate sets of peptide measurements. Linearity over the quantified dynamic range was calculated as the reproducibility (%RSD) of mass spectral response factors (AUC normalized by concentration). Linearity and reproducibility metrics (%RSDs) were log-transformed to achieve a normal distribution prior to use in significance calculations. Significance calculations (Welch's *t*-test, paired *t*-test, and one-way ANOVA with Bonferroni correction for multiple comparisons) used a threshold (α) of 0.05 and were performed in Origin (version 8.5.1 SR2 b315; OriginLab, Northhampton, MA). For comparison of linearity metrics within each dataset and between the QqOrbi and QqQ datasets, adjusted %RSD values were calculated as given by the following equation (**2.2**) to normalize for the number of concentration orders-of-magnitude quantified in each dataset and individual experiment:

$$\%RSD_{adj} = \frac{\%RSD}{log\left(\frac{[high]}{[low]}\right)}$$

Equation 2.2

where [*low*] and [*high*] correspond to the lowest and highest (always 200 nM) concentration, respectively, at which a peptide was detected. Statistical significance calculations for pairwise comparisons of mean log-transformed %RSD (or adjusted %RSD) values within each dataset and between the QqQ and each QqOrbi dataset were calculated by Welch's *t*-test and pairwise *t*-test, respectively, in Origin.

In silico calculations. A script, available at http://www.chem.wisc.edu/~coon/software.php, was written in C#/.NET using Visual Studio 2010 for all *in silico* calculations referenced in the text. The human tryptic peptidome was modeled using the UniProtKB human protein database (accessed 11 Nov 2011, http://www.uniprot.org/taxonomy/complete-proteomes) with the following parameters: ≥1 enzyme terminus, ≤4 missed cleavages, peptide length of 5-45 amino acids, intact peptide mass of 200-9000 Da, fixed cysteine carbamidomethylation, variable methionine oxidation, and assumed cleavage of initiator methionine residues. Each of the 20,332,717 unique peptide sequences comprising the human peptidome was assessed for its potential to interfere in precursor and product ion measurement of the 25 peptide sequences studied here, in both light and heavy forms (50 total), considering the *y*-ion transitions monitored on the QqQ (**Table 2.1**). For peptides not monitored on the QqQ, equivalent *y*-ions were chosen. Intact confounder peptides were considered in charge states from 1-5 (monoisotopic mass only, isotopes were not considered). Confounder product ions were considered in charge states from 1 to one less the precursor charge state and were of the following ion types: *b*, *y*, *a*, *b*/*y*/*a* – H_2O if containing amino acids D, E, S, or T, *b*/*y*/*a* – NH_3 if containing amino acids K, N, Q, or R, internal fragments, and sequence-specific immonium ions. Interference in all QqOrbi examples met the following requirements: mass measurement tolerance of ±5 ppm and precursor isolation width of ±1 Th. Interference in all QqQ examples met the following requirements: mass

measurement tolerance of ±250 ppm, precursor isolation width (Q1) of ±0.5 Th, and product isolation width (Q3) of ±0.35 Th.

RESULTS AND DISCUSSION

Here, we investigate the performance of parallel reaction monitoring (PRM) on a high resolution and accurate mass (HR/AM) quadrupole-Orbitrap mass spectrometer (QqOrbi) for targeted, quantitative proteomics with respect to the gold-standard selected reaction monitoring (SRM) method performed on a triple quadrupole (QqQ) instrument. In PRM, we have substituted a HR/AM Orbitrap mass analyzer for Q3 within the context of a conventional SRM experiment. Thus, instead of *serially* monitoring target transitions over several ion injections and low resolution mass measurement periods (**Figure 2.1A**), PRM monitors all product ions of a mass-selected peptide target *in parallel* with one ion injection and full mass range Orbitrap mass analysis (**Figure 2.1B**). We queried a set of 25 isotopically heavy-labeled synthetic peptides spanning six orders-of-magnitude in concentration (2 pM to 200 nM, corresponding to 2 amol to 200 fmol on column) with and without peptide background using PRM with isolation widths of ±1 and ±0.2 Th. Three technical replicates were performed for each experiment with one microliter of sample injected on column in all experiments. For comparison, we analyzed 14 of the 25 peptides using a traditional, optimized QqQ SRM assay. Equivalent experiments and comparisons were also performed using selected ion monitoring (SIM) on the QqOrbi. While not described in the body text in detail, interested readers can consult **Appendix 6.4-6.6** for more information.

Theoretical comparison of SRM and PRM. The process of targeting a peptide with SRM involves two stages of quadrupole mass filtering with tight tolerances for both members of a precursor-product ion transition. Since all product ion transitions targeted for a given precursor

peptide (usually 3 to 5) are required to simultaneously elute, the likelihood of mistaking a non-target peptide or background ion for the targeted peptide is a rare occurrence; hence, SRM is considered to be a highly specific assay. The proposed PRM method, however, involves only one stage of quadrupole mass filtering (of the precursor of interest) prior to mass analysis in an Orbitrap. The Orbitrap, however, by nature of its high resolution and high mass accuracy should more effectively separate ions of interest from background ions than the electron multiplier-based detection used in a QqQ. Thus, to motivate our experiments, we asked how PRM compares theoretically to SRM in terms of specificity. In other words, can the selectivity of Orbitrap HR/AM mass analysis make up for use of only one stage of mass filtering?

To answer this question, we digested the human proteome with trypsin *in silico* to yield over 20 million unique peptide sequences (see **Experimental Procedures** for details). For our calculations, we considered these peptides to be potential confounders in the measurement of the precursor and product ions of the 25 isotopically heavy-labeled peptides targeted in this study, as well as their 25 corresponding unlabeled peptides (50 total). For each unique confounder peptide, considered in charge states from 1 to 5, we further generated all possible *b*, *y*, *a*, *b*/*y*/*a* – water, and *b*/*y*/*a* – ammonia product ions, internal fragments, and immonium ions in charges ranging from 1 to one less the precursor charge state. We then asked how often, depending on the amount of evidence required by the assay and the mass analyzer employed, the numerous ions generated by the confounder population resulted in indistinguishable interference in the measurement of one of our target peptides. The results of these queries are summarized in **Figure 2.1C**.

First, we consider a query requiring the least evidence of the targeted peptide, a SIM experiment. In SIM, an intact target peptide ion is isolated in Q1 and mass analyzed without further transformation. We begin our theoretical calculations with SIM to explore the effect of mass accuracy/resolution alone on selectivity. Assuming Q1 isolation widths of ±0.5 and 1 Th,

and mass errors less than ±250 and 5 ppm, for the QqQ and QqOrbi, respectively, highly accurate mass analysis improves the chances of correctly identifying the target peptide (96 ± 3%) by exclusion of a large number of spurious peptides (7151 ± 1447, on average). Still, such an experiment, however, would only produce an unambiguous identification ~1% of the time (**Figure 2.1C**). Note, these calculations assume no upfront chromatographic separations and that all genome-predicted peptides are translated and detectable (i.e., the worst-case scenario).

Though it is obvious that high accuracy mass measurements increase specificity, it is less clear that this benefit persists in reaction monitoring experiments. To test this, we considered the number of intact peptides that could potentially be co-isolated with a given target peptide (assuming ±1 and ±0.5 Th Q1 isolation windows centered on the target peptide *m/z* for the QqOrbi and QqQ, respectively) and then generate at least one product ion (of any type) with a *m/z* within ±5 ppm (QqOrbi) or ±250 ppm (QqQ) of a *y*-ion transition from the target peptide. As shown in **Figure 2.1C**, the specificity of a HR/AM analyzer substantially reduces the number of spurious peptide ions that can interfere in the correct identification of a target peptide by its *y*-ion transitions (shown as the rate of correct target identification, or true positive rate). The inclusion of 1, 2, or 3 *y*-ion transitions with high mass accuracy measurement results in approximately a (10 ± 14)-fold, (53 ± 35)-fold, and (112 ± 81)-fold greater likelihood of correctly identifying a target peptide compared to the inclusion of 1, 2, or 3 *y*-ion transitions at low resolution, respectively. Even if observation of all three *y*-ion transitions were required, as is typically the case in QqQ SRM assays, the likelihood of correctly identifying a target at low resolution and unit mass accuracy is still less than 1% for the 50 peptides considered here (0.6 ± 0.3%). Again, these calculations do not consider the significant benefit that is achieved by chromatographic separation and model a worst-case scenario. Even still, the detection of three *y*-ion transitions at high mass

accuracy (i.e., <5 ppm mass error) provides nearly unambiguous target confirmation (~1.6 ± 1.1 potential confounders) from the background of the entire human peptidome.

These theoretical calculations confirm our guiding supposition that PRM can provide greater routine specificity as compared to conventional SRM. Whether this result implies greater overall performance in targeted, quantitative proteomics studies compared to SRM, however, depends on several additional, but important, factors: whether both analyzers, one image current-based and the other electron multiplier-based, are capable of detecting the targeted species, and whether both analyzers can reproducibly *and* accurately measure the abundance of the targeted species. Given that PRM relies on an analyzer that is fundamentally less sensitive and slower than that used in SRM, investigation of these factors will reveal whether selectivity/specificity can overcome limitations in speed and sensitivity. In the following sections, we empirically investigate this issue through a systematic analysis of reproducibility, sensitivity, and linearity of both methods.

Detection criteria for PRM and SRM. The detection criteria we developed for PRM incorporate the benefits of high selectivity and specificity HR/AM mass analysis, as discussed above. By making use of full mass range MS/MS spectra and the high specificity of product ions when measured with high mass accuracy, we developed an automated detection algorithm that assigned a spectral score to each PRM spectrum based on the presence of *b*- or *y*-ions within ±5 ppm of expected target-specific product ions using **Equation 2.1**, and then generated an "extracted score chromatogram" (XSC) for that peptide. In the design of our spectral score, we chose to weight *b*-ions less than equivalently numbered *y*-ions because all of the peptides targeted in this study were isotopically heavy-labeled at the c-terminus and analyzed in a background of endogenous yeast peptides. Thus, *y*-ions, containing the heavy label, were more specific to our target peptides than *b*-ions and were weighted as such. If the target peptides of interest were not

heavy-labeled, one might consider equally weighting *b*- and *y*-ions for scoring purposes. A positive detection event required that, in at least 2 of 3 replicates, the XSC met or exceeded a peptide specific threshold (equal to the length of the target peptide) within ±3σ min of the expected retention time (**Figure 2.2**).

Figure 2.3 shows the application of this score to exemplary data from the peptide AETLVQAr (#17) under neat and matrix-containing conditions over all 6 peptide concentrations. The XSCs (grey) are overlaid with XICs (blue) generated using the summed intensity of all *b*- and *y*-ions present in each spectrum. Note that XSCs are used solely for establishing a detection event and not for quantification as the score is independent of ion intensity. Following a positive detection event, the target-specific ions that generated the detection event are extracted as an XIC for quantification. This peptide was detected at concentrations spanning 4 orders-of-magnitude (from 20 pM to 200 nM) without matrix, and over 3 orders-of-magnitude (from 200 pM to 200 nM) in the presence of yeast matrix.

In **Figure 2.3**, at high target peptide concentrations (e.g., 20-200 nM), the XSC often increases above threshold before the XIC shows any noticeable intensity change and stays above threshold after the XIC intensity has fallen. This characteristic is ascribable to the intensity-independence of the score, and suggests that high quality Orbitrap spectra, in terms of the presence of product ions with high mass accuracy, can be acquired independently of ion intensity above some intensity threshold. At low concentration, especially in the presence of matrix (see 200 pM in matrix), the XSC easily distinguishes nearby, spurious XIC signals from target-derived signals, demonstrating score specificity. The detection threshold was chosen to ensure that an accepted score (signifying target detection) would, at least, contain one *y*-ion with *m/z* greater than the precursor *m/z*. While this threshold does not definitively ensure that a score above the threshold is entirely specific to only the target peptide, use of the threshold successfully separated

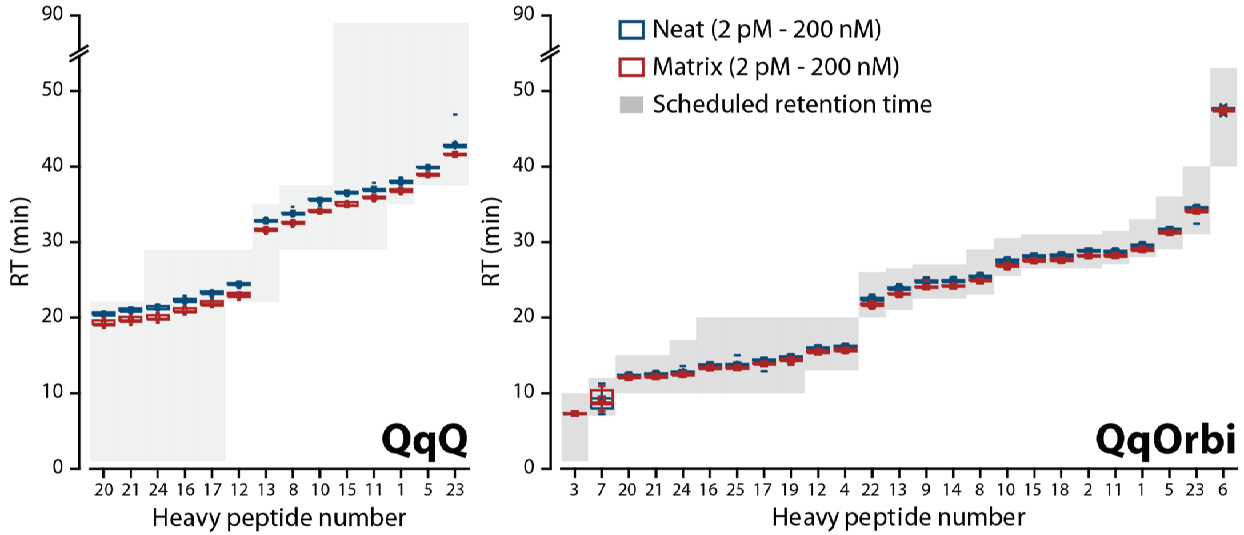


Figure 2.2. Boxplots showing retention time reproducibility across all experiments and concentrations for each peptide targeted in QqQ and QqOrbi experiments (both SIM and PRM) under neat (blue) and matrix-containing conditions (red). Scheduled retention times used for each peptide are shown as grey bars.

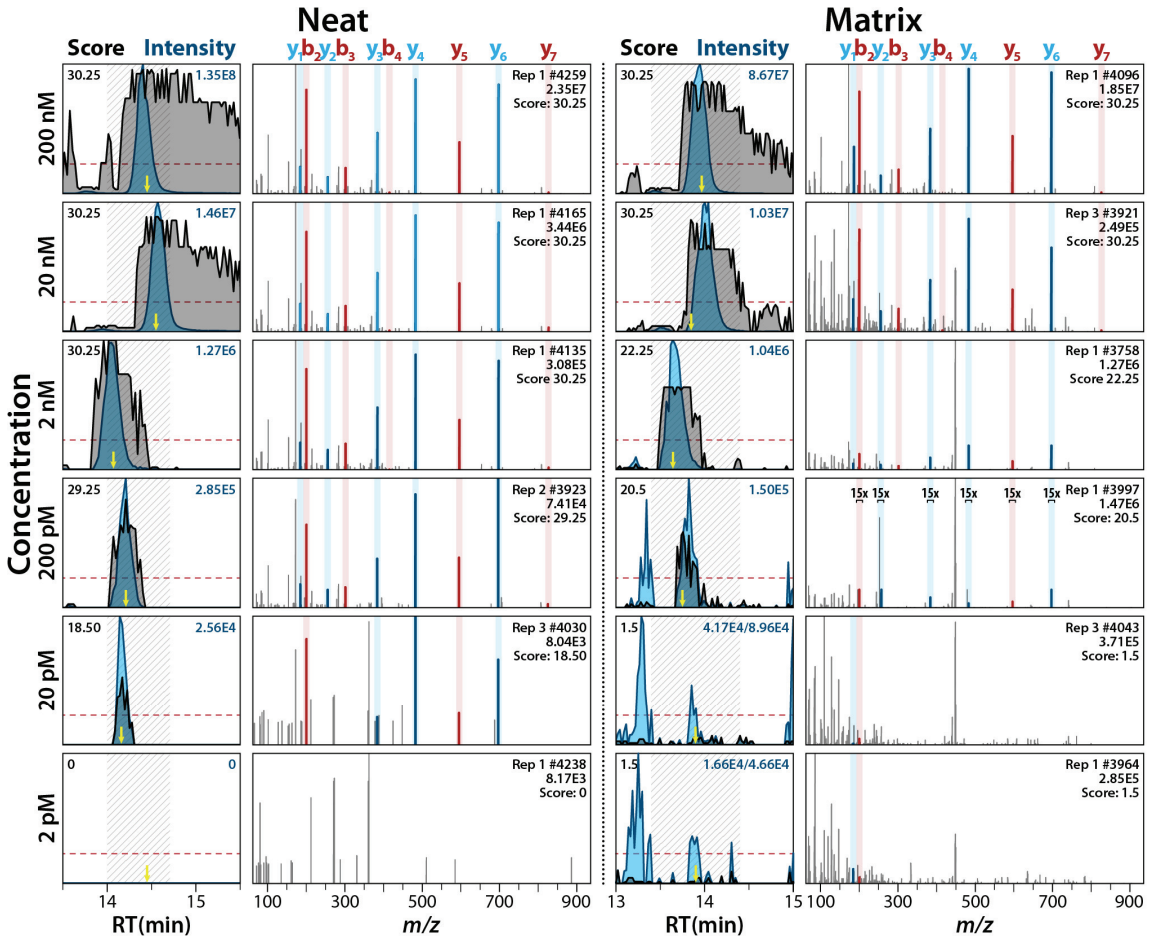


Figure 2.3. Extracted PRM score chromatograms (XSC, ±5 ppm, grey) overlaid with XICs (±5 ppm, blue; 7-point boxcar smoothed) and single-scan PRM spectra for peptide AETLVQAr (#17) isolated at ±1 Th under neat (left) and matrix-containing (right) conditions from 2 pM to 200 nM. Red dotted line in XSC plots represents the score acceptance threshold (8) for this peptide. Product ions detected in each spectrum are highlighted and the spectral score is labeled. Yellow arrows in each XIC/XSC plot indicate the retention time at which the associated single-scan spectrum was acquired. Hashed area in XIC/XSC plots designates the retention time period during which peak elution was expected based on the ±3σ range around the average retention time observed in 200 nM analyses.

spurious signals from target-specific signals, agreeing with or being more conservative than manual interpretation of the data in all cases. **Figure 2.4** plots the distribution of the spectral score for all possible combinations of *b*- and *y*-ions arising from an 8 amino acid peptide, such as AETLVQAr, by the fraction of the score described by each ion. Example ion combinations are given at several scores below and above the threshold. **Appendix 6 Tables 6.1A-D** catalog the maximum XSC scores measured for each replicate, concentration, and PRM experiment type (**Appendix 6.2)**.

QqQ SRM experiments were performed to provide a standard for comparison of our QqOrbi PRM data. QqQ SRM data were analyzed manually based on detection criteria commonly used in the literature for SRM analyses.[42] A positive peptide detection event occurred when 8 of 10 criteria were met in a least 2 of 3 replicates. Originally, we required 10 of 10 criteria to be met in at least 2 of 3 replicates, but relaxed this criterion to improve the completeness of the dataset and thus our ability to compare it to the PRM dataset. The ten criteria were: 1) the appearance of three co-eluting transitions within ±2.5 min of the expected retention time (**Figure 2.2**), 2-4) average transition *m/z* within ±500 ppm of the expected *m/z*, 5-7) average transition abundance within ±15% of the expected abundance, and 8-10) individual transition XICs with signal-to-noise meeting or exceeding 3. Peptides, transitions, and other experimental parameters for SRM are listed in **Table 2.1**.

Figure 2.5A compares PRM with ±1 Th isolation to SRM for the detection of peptide GVSAFSTWEk (#1) at 200 pM and 2 nM (200 amol and 2 fmol, respectively, on column) in the presence of yeast matrix. In PRM, the signals present in the XIC during the entire duration for which this peptide was targeted are exclusively due to the targeted peptide, even at the lowest concentration detected (200 pM; XICs at lower concentrations had zero intensity). Numerous other *y*-ions (shown in grey) are also present at the 2 nM level, adding confidence to the detection

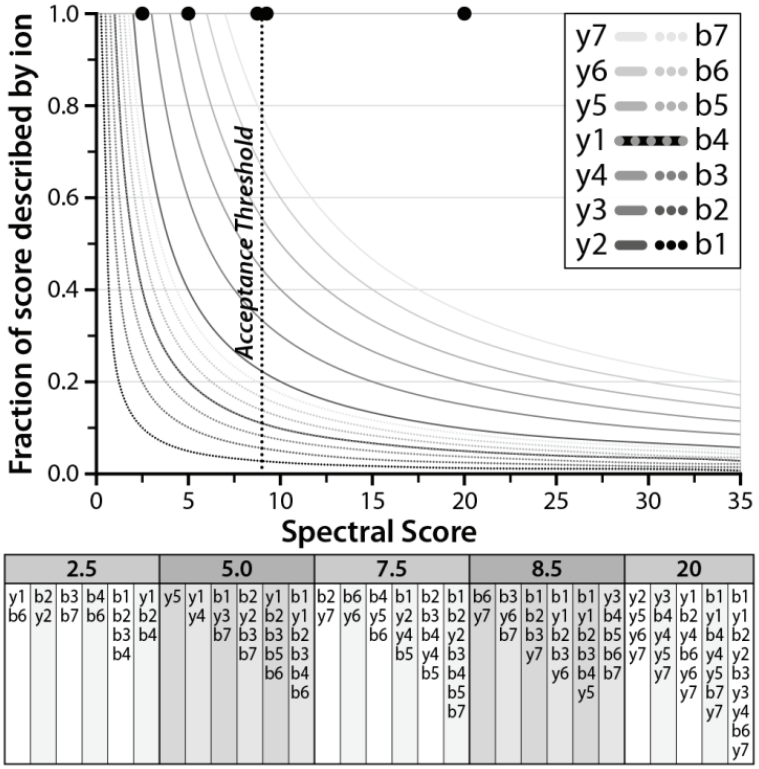


Figure 2.4. Distribution of the spectral score (**Equation 2.1**) for all possible combinations of *b*- and *y*-ions from an 8 amino acid peptide by the fraction of the score described by each ion. Example ion combinations are given at several scores below and above the acceptance threshold (8, vertical dotted line) as designated by dots at the top (y = 1) of the distribution plot.

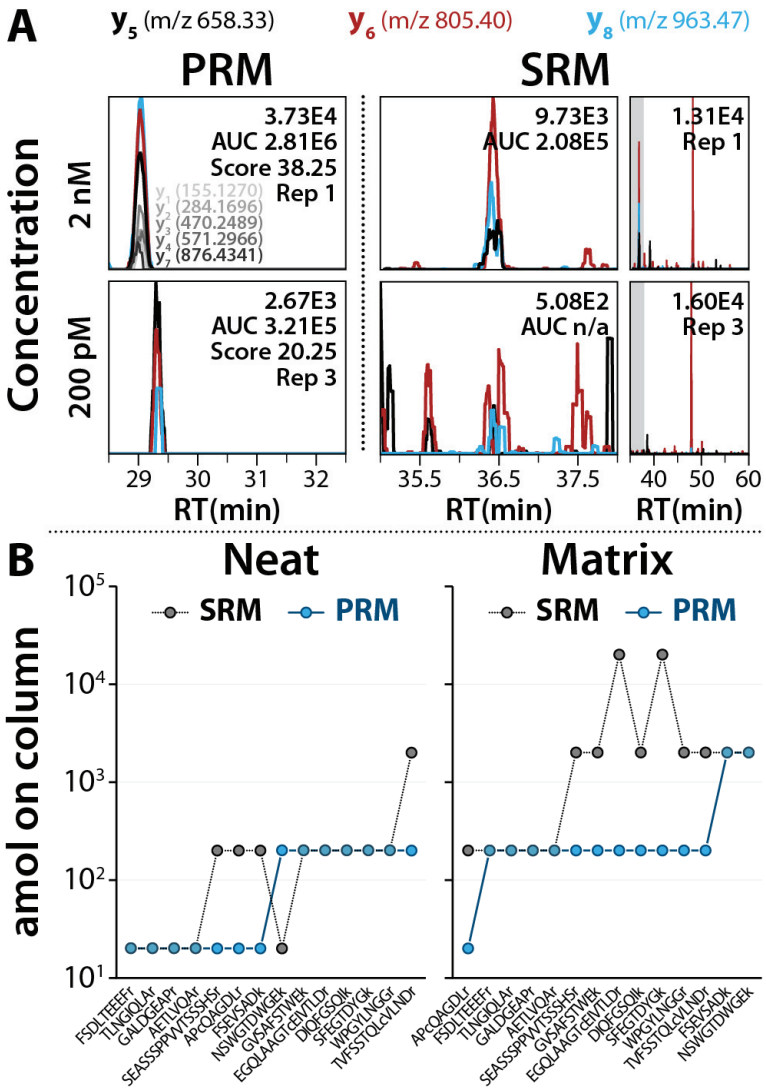


Figure 2.5. (A) Comparison of QqOrbi PRM detection at ±1 Th with QqQ SRM for peptide GVSAFSTWEk (#1) at 200 pM and 2 nM in the presence of matrix. Transition XICs are shown for the entire duration over which the peptide was targeted, with a zoom-in on the relevant time period in the QqQ SRM case (from the region shaded in grey in the chromatograms at the far right). Additional y-type product ions present in the PRM data are shown in grey. XICs were extracted at ±5 and ±250 ppm for PRM and SRM, respectively. The maximum spectral score attained at each concentration in PRM is also labeled. **(B)** Lowest concentration detected (as number of attomoles of peptide on column) for each peptide in neat and matrix-containing experiments for the 14 peptides targeted in both SRM and PRM.

and generating maximum spectral scores well above the threshold of 10 for this peptide. By contrast, transition XICs in the SRM data show numerous background signals throughout the duration over which the peptide was targeted (far right). At the 200 pM level, while numerous signals are present in the XICs for each SRM transition, the high level of chemical noise in all transition channels prevents confident determination of the presence of the target peptide. This example demonstrates the enhancement of sensitivity that can be achieved via PRM due to the high selectivity of HR/AM mass analysis, as well as the ease with which determinations of presence or absence of a peptide target can be made with HR/AM data.

PRM measurement precision. In quantitative studies, high measurement precision is critical to reliably distinguish differences between two analyses or samples. We assessed the degree of measurement precision for the PRM method, defined here as run-to-run area-under-the-curve (AUC) repeatability across technical replicates, for all concentrations and isolation widths.

Overall, PRM exhibited high measurement precision with median percent relative standard deviation (%RSD) less than 10% in most cases (**Appendix 6 Table 6.2A**). The main factors negatively affecting measurement precision (increasing %RSD) were low target peptide concentrations and individual peptide characteristics. Neither isolation width nor the presence of matrix resulted in statistically significant differences in precision ($\alpha = 0.05$; n.b., all discussion of mean %RSDs and statistical significance refer to log-transformed data). The precision of measurements at the lowest concentration typically detected in neat PRM, 200 pM, was significantly lower than at all other (higher) concentrations. In the matrix-containing PRM data, however, no significant differences in the measurement precision of adjacent concentrations were observed (**Appendix 6 Table 6.2B**). When grouped by peptide alone, only very hydrophilic, poorly-retained peptides, #3 and #7, had significantly decreased measurement precision when compared to the other 23 peptides (**Appendix 6 Table 6.2C**).

We find that PRM measurement precision is consistent with studies reporting run-to-run precision data in QqQ SRM experiments. For example, in the 2009 multi-laboratory study of SRM measurement repeatability and reproducibility for peptides spiked into plasma, Addona, *et al.*[20] found both intra- and inter-laboratory precision to be similarly less than 15% RSD across the concentration range studied (1-500 nM). Likewise, precision improved for measurements at higher target concentrations. Kiyonami and colleagues[43] also reported similar precision metrics in their large-scale intelligent-SRM (iSRM) experiments, observing that 80% of the 757 peptides targeted in yeast exhibited less than 10% RSD. Compared to other studies utilizing HR/AM mass spectrometers (quadrupole linear ion trap (QLT)-Orbitrap, QqTOF, and QLT-FT-ICR) for targeted, quantitative measurements,[62-66] QqOrbi PRM achieved, on average, 2-3-fold better run-to-run measurement precision.

PRM dynamic range. On average, PRM experiments yielded quantitative information between 2-4 concentration orders-of-magnitude (93%), with the majority of experiments resulting in quantification across 3 orders-of-magnitude (54%, 0.2 to 200 nM; **Appendix 6 Table 6.3A**, **Figure 2.6**). The wider PRM isolation width, ±1 vs. ±0.2 Th, resulted in improved dynamic range – neat, $10^{3.3}$ vs. $10^{2.6}$, and matrix, $10^{2.7}$ vs. $10^{2.2}$ (**Appendix 6 Table 6.3B**). The presence of matrix in PRM experiments resulted in a modest depression of the quantifiable dynamic range, ~0.5 orders-of-magnitude (matrix vs. neat, $10^{2.4}$ vs. $10^{3.0}$). While tighter isolation widths slightly mitigated the effects of matrix-induced sensitivity depression, at the expense of overall sensitivity, this difference was not significant (**Appendix 6 Table 6.3A**). These results indicate that, although an increase in selectivity due to gas-phase enrichment would be expected by tighter isolation widths, the concomitant decrease in ion transmission at very tight isolation widths (±0.2 Th), in conjunction with Orbitrap detection, results in decreased sensitivity as too few ions are present for the target signal to exceed the Orbitrap's thermal noise band. However, with a

# Orders	PRM							SIM							
	±0.2 Th			±1 Th			Total	±1 Th			±2 Th			Total	Total
	Neat	Matrix	Total	Neat	Matrix	Total		Neat	Matrix	Total	Neat	Matrix	Total		
0	1	1	2	0	0	0	2	0	1	1	0	1	1	2	4
1	0	1	1	2	2	4	5	0	2	2	0	2	2	4	9
2	7	15	22	0	5	5	27	1	19	20	2	20	22	42	69
3	17	8	25	12	17	29	54	1	3	4	0	2	2	6	60
4	0	0	0	11	1	12	12	8	0	8	10	0	10	18	30
5	0	0	0	0	0	0	0	15	0	15	13	0	13	28	28
Total	25	25	50	25	25	50	100	25	25	50	25	25	50	100	200

Figure 2.6. Number of experiments which yielded quantitative data over 0 concentration orders-of-magnitude (200 nM only), 1 order-of-magnitude (20-200 nM), 2 (2-200 nM), 3 (0.2-200 nM), 4 (0.02-200 nM), and 5 orders-of-magnitude (0.02-200 nM). Bars in each cell are normalized to the total in the bottom row.

HR/AM analyzer, the wider isolation width, which provides greater ion transmission, but also higher levels of chemical noise, can be used without a decrease in performance due to the high selectivity of the mass analysis.

Linearity of PRM measurement response. The linearity of measurement response – or, measurement accuracy – was calculated as the percent RSD of response factors (AUC normalized by concentration) over the concentration range for which a peptide was detected. In general, linearity was not significantly influenced by isolation width or the presence of matrix: mean neat %RSDs of 37.1 and 36.4, and mean matrix %RSDs of 35.7 and 34.2, were observed in PRM experiments with ±0.2 and ±1 Th isolation widths, respectively (**Appendix 6 Tables 6.4A-B**). If the lowest detected concentration, associated with lower overall measurement precision (*vide supra*), was excluded from each experiment, however, linearity improved and some significant differences emerged based on the presence of matrix (**Appendix 6 Tables 6.4C-F**). Surprisingly, matrix-containing experiments were found to possess greater linearity than their neat counterparts in all experiments, with %RSDs of 23.2% for neat PRM and 12.0% for matrix PRM (**Appendix 6 Table 6.4D**).

Due to the truncation of the measured dynamic range observed in the presence of matrix, we posited that greater linearity in matrix-containing experiments was an artifact of simply considering a smaller concentration range. Since lower precision on lower concentration measurements make it more challenging to arrive at an accurate average AUC from only three replicates, experiments quantifying over a wide dynamic range (e.g., neat experiments) could have decreased linearity due to low measurement accuracy at the lowest detected concentrations. To address this, we calculated an adjusted %RSD for each experiment and peptide that accounted for the dynamic range represented by a %RSD value. The adjusted %RSD separates out contributions of dynamic range and concentration from the linearity estimate, thereby providing a

more fair comparison between experiments where the dynamic range quantified is drastically different. These adjusted linearity values no longer demonstrate the surprising trend observed above when the lowest concentration was excluded (13.5% Adj. RSD neat vs. 16.6% Adj. RSD matrix). Using the adjusted linearity metric, PRM was not significantly affected by the presence of matrix (or isolation width, as before) (**Appendix 6 Tables 6.4G-H**).

QqQ SRM measurement precision, dynamic range, and linearity. Like the PRM data described above, the QqQ SRM data, which queried a subset of 14 of the 25 peptides analyzed on the QqOrbi, also exhibited a high degree of run-to-run measurement precision, typically less than 15% RSD. Unlike in PRM, however, measurement precision, when all peptides and concentrations were considered together, was significantly greater in the presence of matrix (**Appendix 6 Table 6.5A**). Lower peptide concentrations were again correlated with decreased measurement precision, though only significantly so under neat conditions (**Appendix 6 Table 6.5B**). Peptides were quantified on average over concentration ranges of $10^{3.3}$ (neat) and $10^{2.2}$ (matrix), and the depression of dynamic range under matrix-containing conditions was statistically significant. Linearity under matrix conditions bested that under neat conditions with mean %RSDs of 16.7 and 42.1%, respectively. As in the QqOrbi, the loss of linearity was often the result of peptide detection at the lowest concentrations. Exclusion of these concentrations in %RSD calculations revealed no differences in linearity between matrix and neat conditions. Likewise, if only data over the same dynamic range under both matrix-containing and neat conditions were considered, effectively truncating the neat data to the concentration range quantified in the matrix-containing data, linearity differences due to matrix interferences were not significant. Adjusted %RSDs, however, still exhibited significantly decreased linearity in the neat case. Mean adjusted %RSDs were 13.0 and 8.0% for neat and matrix data, respectively (**Appendix 6 Table 6.5C**).

We postulate that the adjusted %RSD correction did not completely account for differences in linearity, as it did in the PRM data, because the SRM detection criteria were not sensitive enough to detect and exclude data at the lowest detected concentrations in neat experiments that were skewed due to small amounts of chemical interference. In the matrix-containing SRM experiments, on the other hand, more abundant matrix interferences were easily detected and excluded by the detection criteria (**Figure 2.7**). Detection criteria incorporating HR/AM data, however, more sensitively detected and excluded aberrant responses and thus, deviations in linearity were predictable and correctable (as adjusted %RSDs) based on the detected dynamic range. This same rationale can be applied to the discussion of measurement precision above: when all neat data were considered, precision was diminished relative to the matrix-containing dataset due to the inclusion of background-skewed data not excluded by the detection criteria.

Empirical comparison of QqQ SRM and QqOrbi PRM. To compare between QqQ SRM and QqOrbi PRM, we only considered the fourteen peptides analyzed in both datasets. Under neat conditions, run-to-run measurement precision (paired by peptide and concentration) was no different between QqQ SRM and QqOrbi PRM, regardless of the PRM isolation width employed. In the presence of matrix, however, SRM demonstrated significantly better measurement precision compared to both PRM datasets (5.6 versus 11.1% RSD, respectively; **Appendix 6 Tables 6.6A-B**). Since neat SRM measurement precision, as discussed above, is believed to be artificially decreased due to the inclusion of chemical interference, it is likely that SRM exhibits superior measurement precision under neat conditions as well (**Appendix 6.3**, **Appendix 6 Table 6.6B**).

SRM exhibited greater measurement precision likely due to the higher sampling rate of the QqQ (almost twice as many scans were acquired compared to the QqOrbi per 90 min

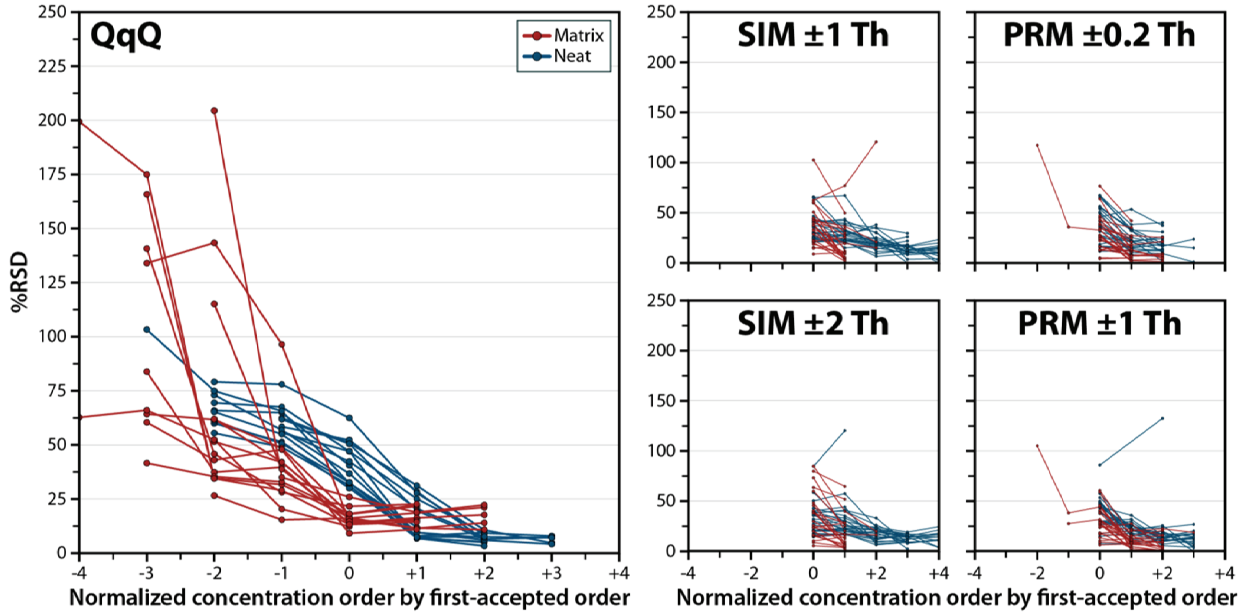


Figure 2.7. Linearity, as %RSD, plotted versus the number of orders-of-magnitude considered in the linearity calculation normalized by the lowest concentration order-of-magnitude accepted as a positive detection for all experiment types. Data greater than or equal to 0 meets the experiment-specific detection criteria. More than half of the measured data in QqQ experiments does not meet detection criteria. The greater %RSD of neat QqQ data in compared to matrix-containing QqQ data at the lowest-accepted concentration (x = 0) suggest that the detection criteria for QqQ are not sensitive enough to detect and exclude low levels of chemical noise in the neat data which inflate the %RSD (decrease linearity). By contrast, HR/AM PRM and SIM data show roughly equivalent %RSDs in matrix-containing and neat data at the lowest-accepted concentration signifying that the detection criteria are sensitive enough to detect and exclude data containing chemical noise.

chromatographic run). The ability of the QqQ to sample more points over a given chromatographic peak provided a more accurate determination of the peak AUC and, in turn, greater run-to-run repeatability. The difference in sampling rate between the two methods is due to the characteristics of the instruments used and, to some extent, necessary aspects of the experimental design. Since the QqQ is a "beam-type" instrument (as opposed to a scanning instrument like the QqOrbi), it operates at a duty cycle nearing 100%; this means that there is very little "down time" where the instrument is not acquiring data. The Orbitrap, conversely, as a scanning instrument, has inefficiencies inherent to its design. The Orbitrap transients employed here (required for MS/MS scans of resolution 17,500) were approximately twice the length of the 35 ms dwell times employed in the SRM method. Additionally, whereas the QqQ cycle time was fixed, the cycle time of the QqOrbi varied based on ion accumulation times, which were dynamically set based on ion flux. At low sample concentrations, QqOrbi injection times could reach as high as 120 ms to result in (at most) ~40 ms of inter-scan "down time" where mass analysis was not occurring. Shorter transients and lower maximum allowable injection times could be used to match the cycle times of the two methods at the expense of overall performance: shorter injection times would decrease spectral quality/usability when ion flux is low, and lower mass analysis resolution would undermine the benefits that high selectivity brings to the PRM method.

Greater QqQ SRM measurement precision did not translate into greater sensitivity or linearity compared to PRM. Under neat conditions, SRM quantified a concentration range spanning $10^{3.3}$ per peptide on average. This range is not statistically different from that achieved with PRM, $10^{3.5}$, at the wider isolation width. In matrix-containing experiments, however, PRM quantified over a broader range at both isolation widths ($10^{2.4}$ and $10^{2.9}$ at ±0.2 and ±1 Th, respectively) than did SRM ($10^{2.2}$), and significantly more at the ±1 Th isolation width (**Table**

2.2). This result is illustrated on an individual peptide basis in **Figure 2.5B**. For both neat and matrix containing experiments, the lowest concentration detected for each peptide studied in the SRM and PRM (±1 Th) experiments is plotted. SRM demonstrates a lower empirical detection limit than PRM in only one case (peptide NSWGTDWGEk, neat). As an aside, under matrix-containing conditions, SRM surprisingly performed no better QqOrbi SIM: SIM quantified on average over a range of $10^{2.0}$ at an isolation width ±1 Th, statistically no different from SRM (see **Appendix 6.5-6.6** and **Appendix 6 Table 6.6C**).

Under neat conditions, PRM demonstrated significantly higher linearity over the quantified dynamic range compared to QqQ SRM (**Table 2.2**). Given that PRM under matrix-containing conditions yielded quantitative data over a wider dynamic range than SRM, we calculated adjusted %RSDs, as before, to normalize the linearity metric for the dynamic range detected and permit a fair comparison of the data. With this consideration, PRM linearity was statistically no different from the linearity exhibited by SRM (**Table 2.2**). **Figure 2.8** plots the SRM linearity data for each of the targeted peptides as a function of PRM linearity and is stratified by the presence of matrix and isolation width. This data presentation demonstrates the effect of using the adjusted linearity metric on mean linearity estimates (shown as vertical and horizontal lines), as well as the relative distributions of the linearity metrics across the datasets.

With these data, we find that a high resolution and accurate mass MS can acquire data of similar quality to that collected on a QqQ in a targeted, quantitative proteomics context with 1) minimal upfront development time, 2) straightforward data analysis, and 3) similar performance metrics. Our PRM methods were designed and optimized on a method-level, rather than on the peptide-level as with SRM methods (i.e., optimization of individual collision energies and transition sets for each target peptide). Thus, the PRM method has fewer parameters that require optimization (namely, only maximum injection time/AGC target, “global” collision energy, and,

Table 2.2. Comparison of QqQ SRM and QqOrbi PRM dynamic range and linearity.

		SRM	PRM ±0.2 Th	PRM ±1 Th
means				
Neat	**# orders**	3.3	2.7	3.5
	%RSD	42.09	29.99	27.36
	%RSD$_{adj}$	12.97	11.80	7.72
Matrix	**# orders**	2.2	2.4	2.9
	%RSD	16.09	29.98	29.08
	%RSD$_{adj}$	8.03	13.07	10.49
p values (SRM vs PRM)				
Neat	**# orders**		**2.61E-02***	1.89E-01
	%RSD		**2.33E-02‡**	**8.74E-04‡**
	%RSD$_{adj}$		2.38E-01	**1.99E-05‡**
Matrix	**# orders**		5.47E-01	**2.77E-03‡**
	%RSD		**2.72E-02***	**3.08E-02***
	%RSD$_{adj}$		1.05E-01	5.11E-01

‡PRM data show significantly greater linearity or dynamic range (α = 5E-02).

*SRM data show significantly greater linearity or dynamic range (α = 5E-02).

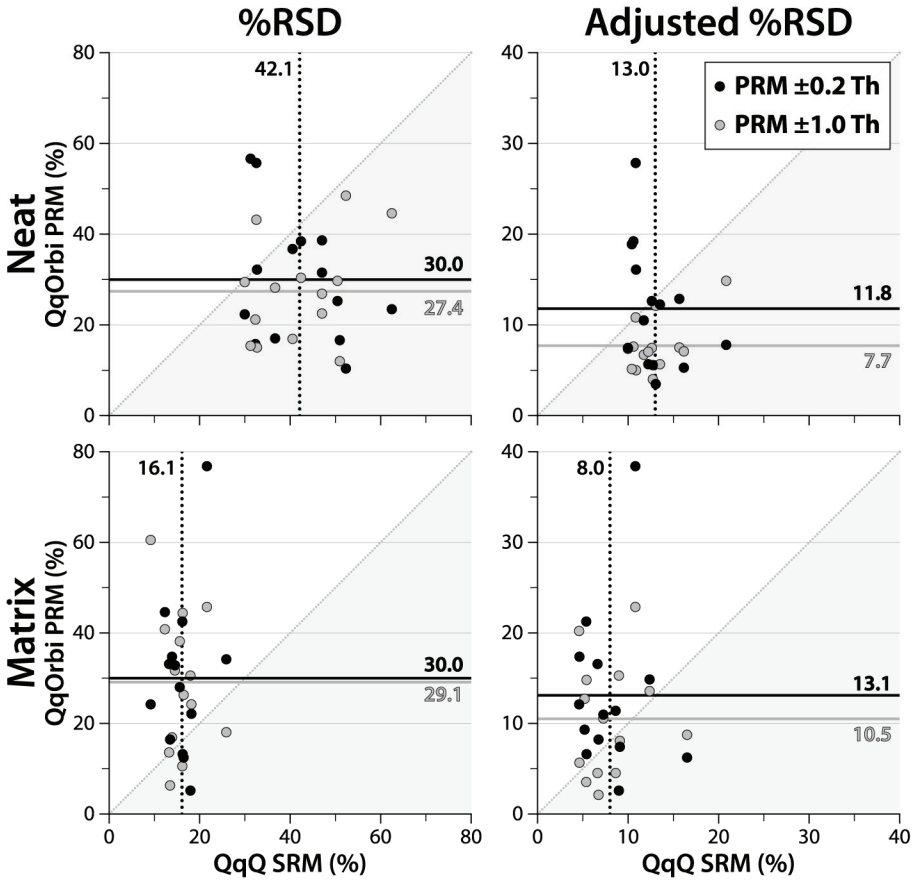


Figure 2.8. Comparison of linearity as %RSD and adjusted %RSD (**Equation 2.2**) for the 14 shared peptides targeted in QqOrbi PRM experiments (y-axis) and QqQ SRM (x-axis) under neat and matrix-containing conditions. Solid horizontal lines and dotted vertical lines represent the mean %RSD value, also labeled in the plot, of the associated dataset. Data falling in the grey region demonstrate greater linearity in PRM experiments. Data in the white region demonstrate greater linearity in SRM experiments.

optionally, target scheduling) and can be performed with solely knowledge of the mass-to-charge ratios of the target peptides (and, optionally, the approximate retention time). These characteristics make PRM amenable to the "walk-up" instrument user who wishes to perform targeted, quantitative assays on a list of targets in a time-efficient manner – collection of MS/MS data and optimization of individual transition sets not required. We believe that the data presented here reflect the analytical performance a typical user might expect for PRM, however the analysis parameters used here may not represent the optimum conditions for all applications of PRM and would require application-specific investigation.

In addition to the minimal upfront method development needed to successfully perform a quantitative PRM assay, a notable aspect of using HR/AM data for targeted proteomics is the ease with which data can be interpreted and data analysis can be automated. Within the course of this study, we developed and automated a novel strategy for scoring and detecting low level analytes in complex, chimeric Orbitrap spectra that allowed sensitive detection of our target peptides without the expense of selectivity. The specificity of accurate mass measurements and the resulting paucity of marginal or ambiguous situations in the data (which are difficult to account for in automated methods) enabled this rather simple algorithm. On the other hand, even using well-established and optimized detection criteria from the literature for SRM, the SRM detection criteria were not sensitive enough to successfully detect and exclude data skewed by small amounts of chemical interference, leading to poor reproducibility and linearity. Furthermore, with the high specificity of a product ion measured at high mass accuracy for a target peptide (as demonstrated by our theoretical calculations) and the benefit of full mass-range mass analysis, the PRM paradigm lends itself well to future development of how targeted, quantitative proteomic data are analyzed. Through use of theoretical calculations, the high specificity of high mass accuracy data lays the groundwork for a generalized and statistically-sound detection algorithm

for reaction monitoring experiments that incorporates probability of correctness measures, as well as obviates the need for manual curation, *ad hoc* detection criteria,[42] "decoy transitions" at the measurement-level,[67] and the potentially high level of human intervention (and error) that comes with such detection strategies.

Lastly, our analysis of the analytical performance characteristics of PRM suggests that targeted HR/AM methods can rival the performance of QqQ SRM in terms of dynamic range, linearity, and, to a lesser extent, precision. While SRM measurement precision was approximately 2-fold better (under matrix-containing experiments) likely due to differences in scan rate between the two mass analyzers, PRM yielded quantitative data over a wider dynamic range than SRM under matrix-containing conditions. High mass accuracy and high resolution underlie this result: when high levels of matrix background are present, a single stage of isolation in combination with highly resolved data is statistically significantly more sensitive than two stages of isolation in combination with low resolution data, the "gold-standard" method for quantitative proteomic analyses. Additionally, achievable linearity over the quantifiable dynamic range was found to be statistically the same between SRM and PRM. Thus, in answer to the query posed in our theoretical comparison of SRM and PRM, our experimental data suggest that high selectivity/specificity can overcome limitations in speed and sensitivity to reliably provide lower detection limits and higher accuracy measurements. We conclude that the proposed PRM analysis paradigm holds promise as a viable, and accessible, alternative and/or complement to SRM for the quantitative proteomics toolbox.

In support of our conclusions, a report by Weisbrod, *et al*.,[62] published recently, described a method for data-independent discovery proteomics on an Orbitrap MS that reinforces the favorable performance of PRM relative to SRM for targeted proteomics. In their "Fourier Transform-All Reaction Monitoring" (FT-ARM) method performed on a QLT-Orbitrap MS, wide

swaths of mass-to-charge space (i.e., 100 Th) were accumulated and isolated in the QLT using broadband waveforms, subjected in bulk to dissociation, and simultaneously mass analyzed to exploit the HR/AM properties of the Orbitrap. The authors noted the ability to quantify peptides with FT-ARM, using the product ions generated from in-bulk dissociation, with modest sensitivity, reproducibility, and precision despite quite significant interferences from other ions present in the 100 Th isolation swath. This further corroborates that HR/AM mass analysis enables real, quantitative information to be extracted from highly interference-riddled measurements with little upfront assay development or optimization. Since PRM involves significantly less co-isolated interference and uses a quadrupole mass filter-equipped platform better suited to accumulation and isolation of large quantities of ions (to improve sensitivity), these findings further support our conclusions that PRM yields high quality quantitative measurements, comparable to QqQ SRM, while simplifying method development.

Looking forward, the PRM paradigm also enables new modes of analysis that are currently unavailable on QqQ platforms. Because of the modular nature of hybrid HR/AM MS, these platforms provide an unprecedented amount of experimental flexibility. Possessing the capabilities of both high performance quantitative and high-throughput discovery proteomics instruments, one can envision mixed mode analysis types in which targeted and discovery experiments are performed simultaneously. For example, when not engaged in profiling an eluting target species, the instrument could be directed to perform conventional data-dependent or data-independent MS/MS, thereby maximizing instrument time, sample usage, and data density. Additionally, because of the high specificity of ions measured at high mass accuracy and consequently the reduced search space for potentially matching peptides, intelligent data acquisition strategies are possible that enable on-the-fly database searching, spectral matching, or spectral scoring (like that used here post-acquisition). Such strategies would enable the mass

spectrometer to make decisions during acquisition, such as assessing whether a target peptide had been adequately identified, changing experimental parameters (CAD energies, isolation width, injection times, etc.) to improve performance for a particular target peptide, or determining its progress in a chromatographic run to make dynamic modifications to target peptide scheduling.[68,69]

REFERENCES

(1) Parker, C. E.; Pearson, T. W.; Anderson, N. L.; Borchers, C. H. Mass-spectrometry-based clinical proteomics - a review and prospective. *Analyst* **2010**, *135*, 1830-1838.

(2) Kirkpatrick, D. S.; Gerber, S. A.; Gygi, S. P. The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods* **2005**, *35*, 265-273.

(3) Phanstiel, D.; Unwin, R.; McAlister, G. C.; Coon, J. J. Peptide Quantification Using 8-Plex Isobaric Tags and Electron Transfer Dissociation Tandem Mass Spectrometry. *Anal Chem* **2009**, *81*, 1693-1698.

(4) Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Matthews, D. E.; Yates, J. R. Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal Chem* **2004**, *76*, 4951-4959.

(5) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **2002**, *1*, 376-386.

(6) Zhu, H. N.; Pan, S. Q.; Gu, S.; Bradbury, E. M.; Chen, X. Amino acid residue specific stable isotope labeling for quantitative proteomics. *Rapid Commun Mass Spectrom* **2002**, *16*, 2115-2123.

(7) Mallick, P.; Kuster, B. Proteomics: a pragmatic perspective. *Nat Biotech* **2010**, *28*, 695-709.

(8) Grossmann, J.; Roschitzki, B.; Panse, C.; Fortes, C.; Barkow-Oesterreicher, S.; Rutishauser, D.; Schlapbach, R. Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods. *J Proteomics* **2010**, *73*, 1740-1746.

(9) Matallana-Surget, S.; Leroy, B.; Wattiez, R. Shotgun proteomics: concept, key points and data mining. *Expert Rev Proteomics* **2010**, *7*, 5-7.

(10) Wu, C. C.; MacCoss, M. J. Shotgun proteomics: Tools for the analysis of complex biological systems. *Curr Opin Mol Ther* **2002**, *4*, 242-250.

(11) Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976-989.

(12) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551-3567.

(13) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X. Y.; Shi, W. Y.; Bryant, S. H. Open mass spectrometry search algorithm. *J Proteome Res* **2004**, *3*, 958-964.

(14) Liu, H. B.; Sadygov, R. G.; Yates, J. R. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **2004**, *76*, 4193-4201.

(15) Tabb, D. L.; Vega-Montoto, L.; Rudnick, P. A.; Variyath, A. M.; Ham, A.-J. L.; Bunk, D. M.; Kilpatrick, L. E.; Billheimer, D. D.; Blackman, R. K.; Cardasis, H. L.; Carr, S. A.; Clauser, K. R.; Jaffe, J. D.; Kowalski, K. A.; Neubert, T. A.; Regnier, F. E.; Schilling, B.; Tegeler, T. J.; Wang, M.; Wang, P.; Whiteaker, J. R.; Zimmerman, L. J.; Fisher, S. J.; Gibson, B. W.; Kinsinger, C. R.; Mesri, M.; Rodriguez, H.; Stein, S. E.; Tempst, P.; Paulovich, A. G.; Liebler, D. C.; Spiegelman, C. Repeatability and Reproducibility in Proteomic Identifications by Liquid Chromatography−Tandem Mass Spectrometry. *J Proteome Res* **2009**, *9*, 761-776.

(16) Gupta, M. K.; Jung, J. W.; Uhm, S. J.; Lee, H.; Lee, H. T.; Kim, K. P. Combining selected reaction monitoring with discovery proteomics in limited biological samples. *Proteomics* **2009**, *9*, 4834-4836.

(17) Yang, X.; Lazar, I. M. MRM screening/biomarker discovery with linear ion trap MS: a library of human cancer-specific peptides. *BMC Cancer* **2009**, *9*.

(18) Schmidt, A.; Gehlenborg, N.; Bodenmiller, B.; Mueller, L. N.; Campbell, D.; Mueller, M.; Aebersold, R.; Domon, B. An Integrated, Directed Mass Spectrometric Approach for In-depth Characterization of Complex Peptide Mixtures. *Mol Cell Proteomics* **2008**, *7*, 2138-2150.

(19) Schmidt, A.; Claassen, M.; Aebersold, R. Directed mass spectrometry: towards hypothesis-driven proteomics. *Curr Opin Chem Biol* **2009**, *13*, 510-517.

(20) Addona, T. A.; Abbatiello, S. E.; Schilling, B.; Skates, S. J.; Mani, D. R.; Bunk, D. M.; Spiegelman, C. H.; Zimmerman, L. J.; Ham, A. J.; Keshishian, H.; Hall, S. C.; Allen, S.; Blackman, R. K.; Borchers, C. H.; Buck, C.; Cardasis, H. L.; Cusack, M. P.; Dodder, N. G.; Gibson, B. W.; Held, J. M.; Hiltke, T.; Jackson, A.; Johansen, E. B.; Kinsinger, C. R.; Li, J.; Mesri, M.; Neubert, T. A.; Niles, R. K.; Pulsipher, T. C.; Ransohoff, D.; Rodriguez, H.; Rudnick, P. A.; Smith, D.; Tabb, D. L.; Tegeler, T. J.; Variyath, A. M.; Vega-Montoto, L. J.; Wahlander, A.; Waldemarson, S.; Wang, M.; Whiteaker, J. R.; Zhao, L.; Anderson, N. L.; Fisher, S. J.; Liebler, D. C.; Paulovich, A. G.; Regnier, F. E.; Tempst, P.; Carr, S. A. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol* **2009**, *27*, 633-641.

(21) Lange, V.; Picotti, P.; Domon, B.; Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol* **2008**, *4*, Msb.2008.2061.

(22) Anderson, N. L.; Anderson, N. G.; Pearson, T. W.; Borchers, C. H.; Paulovich, A. G.; Patterson, S. D.; Gillette, M.; Aebersold, R.; Carr, S. A. A Human Proteome Detection and Quantitation Project. *Mol Cell Proteomics* **2009**, *8*, 883-886.

(23) Keshishian, H.; Addona, T.; Burgess, M.; Kuhn, E.; Carr, S. A. Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol Cell Proteomics* **2007**, *6*, 2212-2229.

(24) Kuzyk, M. A.; Smith, D.; Yang, J.; Cross, T. J.; Jackson, A. M.; Hardie, D. B.; Anderson, N. L.; Borchers, C. H. Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. *Mol Cell Proteomics* **2009**, *8*, 1860-1877.

(25) Whiteaker, J. R.; Lin, C.; Kennedy, J.; Hou, L.; Trute, M.; Sokal, I.; Yan, P.; Schoenherr, R. M.; Zhao, L.; Voytovich, U. J.; Kelly-Spratt, K. S.; Krasnoselsky, A.; Gafken, P. R.; Hogan, J. M.; Jones, L. A.; Wang, P.; Amon, L.; Chodosh, L. A.; Nelson, P. S.; McIntosh, M. W.; Kemp, C. J.; Paulovich, A. G. A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nat Biotechnol* **2011**, *29*, 625-634.

(26) Pan, S.; Chen, R.; Brand, R. E.; Hawley, S.; Tamura, Y.; Gafken, P. R.; Milless, B. P.; Goodlett, D. R.; Rush, J.; Brentnall, T. A. Multiplex Targeted Proteomic Assay for Biomarker Detection in Plasma: A Pancreatic Cancer Biomarker Case Study. *J Proteome Res* **2012**, *11*, 1937-1948.

(27) Anderson, L.; Hunter, C. L. Quantitative Mass Spectrometric Multiple Reaction Monitoring Assays for Major Plasma Proteins. *Mol Cell Proteomics* **2006**, *5*, 573-588.

(28) Costenoble, R.; Picotti, P.; Reiter, L.; Stallmach, R.; Heinemann, M.; Sauer, U.; Aebersold, R. Comprehensive quantitative analysis of central carbon and amino-acid metabolism in Saccharomyces cerevisiae under multiple conditions by targeted proteomics. *Mol Syst Biol* **2011**, *7*, msb.2010.2122.

(29) Picotti, P.; Bodenmiller, B.; Mueller, L. N.; Domon, B.; Aebersold, R. Full Dynamic Range Proteome Analysis of S. cerevisiae by Targeted Proteomics. *Cell* **2009**, *138*, 795-806.

(30) Agard, N. J.; Mahrus, S.; Trinidad, J. C.; Lynn, A.; Burlingame, A. L.; Wells, J. A. Global kinetic analysis of proteolysis via quantitative targeted proteomics. *Proc Natl Acad Sci USA* **2012**, *109*, 1913-1918.

(31) Mallick, P.; Schirle, M.; Chen, S. S.; Flory, M. R.; Lee, H.; Martin, D.; Ranish, J.; Raught, B.; Schmitt, R.; Werner, T.; Kuster, B.; Aebersold, R. Computational prediction of proteotypic peptides for quantitative proteomics. *Nat Biotechnol* **2007**, *25*, 125-131.

(32) Cham Mead, J. A.; Bianco, L.; Bessant, C. Free computational resources for designing selected reaction monitoring transitions. *Proteomics* **2010**, *10*, 1106-1126.

(33) Prakash, A.; Tomazela, D. M.; Frewen, B.; MacLean, B.; Merrihew, G.; Peterman, S.; MacCoss, M. J. Expediting the Development of Targeted SRM Assays: Using Data from

Shotgun Proteomics to Automate Method Development. *J Proteome Res* **2009**, *8*, 2733-2739.

(34) Deutsch, E. W.; Lam, H.; Aebersold, R. PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *Embo Rep* **2008**, *9*, 429-434.

(35) Craig, R.; Cortens, J. P.; Beavis, R. C. Open source system for analyzing, validating, and storing protein identification data. *J Proteome Res* **2004**, *3*, 1234-1242.

(36) Siepen, J. A.; Belhajjame, K.; Selley, J. N.; Embury, S. M.; Paton, N. W.; Goble, C. A.; Oliver, S. G.; Stevens, R.; Zamboulis, L.; Martin, N.; Poulovassillis, A.; Jones, P.; Cote, R.; Hermjakob, H.; Pentony, M. M.; Jones, D. T.; Orengo, C. A.; Hubbard, S. J. ISPIDER Central: an integrated database web-server for proteomics. *Nucleic Acids Res* **2008**, *36*, W485-W490.

(37) Jones, P.; Cote, R. G.; Cho, S. Y.; Klie, S.; Martens, L.; Quinn, A. F.; Thorneycroft, D.; Hermjakob, H. PRIDE: new developments and new datasets. *Nucleic Acids Res* **2008**, *36*, D878-D883.

(38) Desiere, F.; Deutsch, E. W.; King, N. L.; Nesvizhskii, A. I.; Mallick, P.; Eng, J.; Chen, S.; Eddes, J.; Loevenich, S. N.; Aebersold, R. The PeptideAtlas project. *Nucleic Acids Res* **2006**, *34*, D655-D658.

(39) Mathivanan, S.; Ahmed, M.; Ahn, N. G.; Hainard, A.; Amanchy, R.; Andrews, P. C.; Bader, J. S.; Balgley, B. M.; Bantscheff, M.; Bennett, K. L.; Bjorling, E.; Blagoev, B.; Bose, R.; Brahmachari, S. K.; Burlingame, A. S.; Bustelo, X. R.; Cagney, G.; Cantin, G. T.; Cardasis, H. L.; Celis, J. E.; Chaerkady, R.; Chu, F. X.; Cole, P. A.; Costello, C. E.; Cotter, R. J.; Crockett, D.; DeLany, J. P.; De Marzo, A. M.; DeSouza, L. V.; Deutsch, E. W.; Dransfield, E.; Drewes, G.; Droit, A.; Dunn, M. J.; Elenitoba-Johnson, K.; Ewing, R. M.; Van Eyk, J.; Faca, V.; Falkner, J.; Fang, X. M.; Fenselau, C.; Figeys, D.; Gagne, P.; Gelfi, C.; Gevaert, K.; Gimble, J. M.; Gnad, F.; Goel, R.; Gromov, P.; Hanash, S. M.; Hancock, W. S.; Harsha, H. C.; Hart, G.; Hays, F.; He, F. C.; Hebbar, P.; Helsens, K.; Hermeking, H.; Hide, W.; Hjerno, K.; Hochstrasser, D. F.; Hofmann, O.; Horn, D. M.; Hruban, R. H.; Ibarrola, N.; James, P.; Jensen, O. N.; Jensen, P. H.; Jung, P.; Kandasamy, K.; Kheterpal, I.; Kikuno, R. F.; Korf, U.; Korner, R.; Kuster, B.; Kwon, M. S.; Lee, H. J.; Lee, Y. J.; Lefevre, M.; Lehvaslaiho, M.; Lescuyer, P.; Levander, F.; Lim, M. S.; Lobke, C.; Loo, J. A.; Mann, M.; Martens, L.; Martinez-Heredia, J.; McComb, M.; McRedmond, J.; Mehrle, A.; Menon, R.; Miller, C. A.; Mischak, H.; Mohan, S. S.; Mohmood, R.; Molina, H.; Moran, M. F.; Morgan, J. D.; Moritz, R. Human Proteinpedia enables sharing of human protein data. *Nat Biotechnol* **2008**, *26*, 164-167.

(40) Picotti, P.; Lam, H.; Campbell, D.; Deutsch, E. W.; Mirzaei, H.; Ranish, J.; Domon, B.; Aebersold, R. A database of mass spectrometric assays for the yeast proteome. *Nat Methods* **2008**, *5*, 913-914.

(41) Farrah, T.; Deutsch, E. W.; Kreisberg, R.; Sun, Z.; Campbell, D. S.; Mendoza, L.; Kusebauch, U.; Brusniak, M.-Y.; Hüttenhain, R.; Schiess, R.; Selevsek, N.; Aebersold,

R.; Moritz, R. L. PASSEL: The PeptideAtlas SRM Experiment Library. *Proteomics* **2012**, *pmic.201100515*.

(42) Picotti, P.; Rinner, O.; Stallmach, R.; Dautel, F.; Farrah, T.; Domon, B.; Wenschuh, H.; Aebersold, R. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat Methods* **2010**, *7*, 43-46.

(43) Kiyonami, R.; Schoen, A.; Prakash, A.; Peterman, S.; Zabrouskov, V.; Picotti, P.; Aebersold, R.; Huhmer, A.; Domon, B. Increased selectivity, analytical precision, and throughput in targeted proteomics. *Mol Cell Proteomics* **2011**, *10*, mcp.M110.002931.

(44) Mann, M.; Kelleher, N. L. Precision proteomics: the case for high resolution and high mass accuracy. *Proc Natl Acad Sci USA* **2008**, *105*, 18132-18138.

(45) Andrews, G. L.; Simons, B. L.; Young, J. B.; Hawkridge, A. M.; Muddiman, D. C. Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). *Anal Chem* **2011**, *83*, 5442-5446.

(46) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Cooks, R. G. The Orbitrap: a new mass spectrometer. *J Mass Spectrom* **2005**, *40*, 430-443.

(47) Michalski, A.; Damoc, E.; Hauschild, J. P.; Lange, O.; Wieghaus, A.; Makarov, A.; Nagaraj, N.; Cox, J.; Mann, M.; Horning, S. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol Cell Proteomics* **2011**, *10*, mcp.M111.011015.

(48) Michalski, A.; Damoc, E.; Lange, O.; Denisov, E.; Nolting, D.; Mueller, M.; Viner, R.; Schwartz, J.; Remes, P.; Belford, M.; Dunyach, J.-J.; Cox, J.; Horning, S.; Mann, M.; Makarov, A. Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Mol Cell Proteomics* **2011**, *mcp.O111.013698*.

(49) Olsen, J. V.; Schwartz, J. C.; Griep-Raming, J.; Nielsen, M. L.; Damoc, E.; Denisov, E.; Lange, O.; Remes, P.; Taylor, D.; Splendore, M.; Wouters, E. R.; Senko, M.; Makarov, A.; Mann, M.; Horning, S. A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol Cell Proteomics* **2009**, *8*, 2759-2769.

(50) Lange, O.; Makarov, A.; Denisov, E.; Balschun, W. In *58th Conf Amer Soc Mass Spectrom.* Salt Lake City, Utah, 2010.

(51) Zhang, Y.; Hao, Z.; Kellmann, M.; Huhmer, A. HR/AM targeted peptide quantitation on a Q Exactive MS: A unique combination of high selectivity, sensitivity, and throughput. *Thermo Fisher Scientific (San Jose, CA)* **2012**, *Application Note: 554*.

(52) Makarov, A.; Denisov, E.; Lange, O. Performance Evaluation of a High-field Orbitrap Mass Analyzer. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1391-1396.

(53) Nagaraj, N.; Alexander Kulak, N.; Cox, J.; Neuhauser, N.; Mayr, K.; Hoerning, O.; Vorm, O.; Mann, M. System-wide Perturbation Analysis with Nearly Complete Coverage of the Yeast Proteome by Single-shot Ultra HPLC Runs on a Bench Top Orbitrap. *Mol Cell Proteomics* **2012**, *11*, mcp.M111.013722.

(54) Olsen, J. V.; de Godoy, L. M. F.; Li, G. Q.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **2005**, *4*, 2010-2021.

(55) Haas, W.; Faherty, B. K.; Gerber, S. A.; Elias, J. E.; Beausoleil, S. A.; Bakalarski, C. E.; Li, X.; Villén, J.; Gygi, S. P. Optimization and Use of Peptide Mass Measurement Accuracy in Shotgun Proteomics. *Mol Cell Proteomics* **2006**, *5*, 1326-1337.

(56) Roth, M. J.; Forbes, A. J.; Boyne, M. T.; Kim, Y.-B.; Robinson, D. E.; Kelleher, N. L. Precise and Parallel Characterization of Coding Polymorphisms, Alternative Splicing, and Modifications in Human Proteins by Mass Spectrometry. *Mol Cell Proteomics* **2005**, *4*, 1002-1008.

(57) Sherman, J.; McKay, M. J.; Ashman, K.; Molloy, M. P. How specific is my SRM?: The issue of precursor and product ion redundancy. *Proteomics* **2009**, *9*, 1120-1123.

(58) Duncan, M. W.; Yergey, A. L.; Patterson, S. D. Quantifying proteins by mass spectrometry: The selectivity of SRM is only part of the problem. *Proteomics* **2009**, *9*, 1124-1127.

(59) Lee, M. V.; Topper, S. E.; Hubler, S. L.; Hose, J.; Wenger, C. D.; Coon, J. J.; Gasch, A. P. A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. *Mol Syst Biol* **2011**, *7*, msb.2011.2048.

(60) Ficarro, S. B.; Zhang, Y.; Lu, Y.; Moghimi, A. R.; Askenazi, M.; Hyatt, E.; Smith, E. D.; Boyer, L.; Schlaeger, T. M.; Luckey, C. J.; Marto, J. A. Improved electrospray ionization efficiency compensates for diminished chromatographic resolution and enables proteomics analysis of tyrosine signaling in embryonic stem cells. *Anal Chem* **2009**, *81*, 3440-3447.

(61) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26*, 966-968.

(62) Weisbrod, C. R.; Eng, J. K.; Hoopmann, M. R.; Baker, T.; Bruce, J. E. Accurate Peptide fragment mass analysis: multiplexed Peptide identification and quantification. *J Proteome Res* **2012**, *11*, 1621-1632.

(63) Ono, M.; Shitashige, M.; Honda, K.; Isobe, T.; Kuwabara, H.; Matsuzuki, H.; Hirohashi, S.; Yamada, T. Label-free quantitative proteomics using large peptide data sets generated

by nanoflow liquid chromatography and mass spectrometry. *Mol Cell Proteomics* **2006**, *5*, 1338-1347.

(64) Wang, G.; Wu, W. W.; Zeng, W.; Chou, C. L.; Shen, R. F. Label-free protein quantification using LC-coupled ion trap or FT mass spectrometry: Reproducibility, linearity, and application with complex proteomes. *J Proteome Res* **2006**, *5*, 1214-1223.

(65) Wang, W.; Zhou, H.; Lin, H.; Roy, S.; Shaler, T. A.; Hill, L. R.; Norton, S.; Kumar, P.; Anderle, M.; Becker, C. H. Quantification of Proteins and Metabolites by Mass Spectrometry without Isotopic Labeling or Spiked Standards. *Anal Chem* **2003**, *75*, 4818-4826.

(66) Li, X.-j.; Yi, E. C.; Kemp, C. J.; Zhang, H.; Aebersold, R. A Software Suite for the Generation and Comparison of Peptide Arrays from Sets of Data Collected by Liquid Chromatography-Mass Spectrometry. *Mol Cell Proteomics* **2005**, *4*, 1328-1340.

(67) Reiter, L.; Rinner, O.; Picotti, P.; Huttenhain, R.; Beck, M.; Brusniak, M. Y.; Hengartner, M. O.; Aebersold, R. mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat Methods* **2011**, *8*, 430-435.

(68) Graumann, J.; Scheltema, R. A.; Zhang, Y.; Cox, J.; Mann, M. A Framework for Intelligent Data Acquisition and Real-Time Database Searching for Shotgun Proteomics. *Mol Cell Proteomics* **2012**, *11*, mcp.M111.013185.

(69) Bailey, D. J.; Rose, C. M.; McAlister, G. C.; Brumbaugh, J.; Yu, P.; Wenger, C. D.; Westphall, M. S.; Thomson, J. A.; Coon, J. J. Instant Spectral Assignment for Advanced Decision Tree-Driven Mass Spectrometry. *Proc Natl Acad Sci U S A* **2012**, *In press.*

CHAPTER 3

Development and Characterization of a GC-enabled QLT-Orbitrap for

High Resolution and High Mass Accuracy GC/MS

SUMMARY

We detail the development and characterization of a GC/QLT-Orbitrap hybrid mass spectrometer capable of high resolution (up to 100,000 at *m/z* 400) and sub-ppm mass accuracy GC/MS. A high-duty cycle, innovative scan type, the nested scan, was implemented to synchronize the Orbitrap acquisition rate and the timescale of gas chromatography (up to 6.5 Hz at resolution 7,500). We benchmark this instrument's key figures of merit, including resolution, mass accuracy, linear dynamic range, and spectral accuracy, and demonstrate its performance for two challenging applications: the determination of polychlorinated dibenzo-*p*-dioxins (PCDD) and dibenzofurans (PCDF) in environmental samples, and the profiling of primary metabolites in *Arabidopsis thaliana* extracts.

INTRODUCTION

Since the original descriptions of coupling gas chromatography with mass spectrometry by Holmes and Morrell[1] and Gohlke and McLafferty,[2] gas chromatography/mass spectrometry (GC/MS) has developed into a mature and ubiquitous analytical technique. As the mainstay of many industrial quality control facilities and environmental analysis laboratories, the conventional GC/MS instrument employs an ultrafast (often 20-50 Hz) and sensitive, but low resolution and mass accuracy mass analyzer: typically, the single quadrupole, the Paul-type or linear ion trap, or time-of-flight detector.[3] Combined with internal retention time and mass calibration standards, these high-throughput GC/MS instruments are extremely effective tools for efficient determination of the presence and quantity of well-studied, known components in some matrix.[3] Conversely, if the targets of the GC/MS assay are unknown and multitudinous, or require resolution of fine isotopic structure, retention time and unit mass resolution data alone are insufficient to yield unambiguous elemental composition assignments or structural identifications.[4,5]

Several theoretical and empirical treatments[4-9] have addressed the mass accuracy and resolution required to unambiguously assign elemental compositions to unknown compounds. Theoretically, given infinite mass accuracy, the exact mass is sufficient to uniquely identify elemental composition.[6] Practically, however, instrumental limits on mass accuracy, precision, and resolution require additional constraints to filter out spurious potential elemental compositions. For example, by constraining the universe of compounds in question to peptides composed of the 20 common amino acid residues, 1 ppm mass accuracy is adequate to determine amino acid composition.[9] For complex crude oil compounds up to 500 Da that contained only C, H, N, O, and/or S atoms, Kim, *et al.*[7] determined that mass accuracy and resolution of 0.1 mDa was sufficient to yield unique, chemically reasonable elemental compositions. But, for complex

compounds that contain a wide array of elements in any chemically possible configuration, like metabolites, even stricter constraints and new filters are necessary. Even with sub-ppm mass accuracy (i.e., 0.1 mDa for measurements over *m/z* 100), Kind and Fiehn[5] report that assignment of a unique elemental composition is unachievable for small molecule metabolites unless isotopic abundance information is considered as an orthogonal filter to mass accuracy. Isotope distribution error is the deviation of experimentally-acquired isotopic intensities from the theoretically calculated isotopic distribution. In their study, isotopic distribution error, when applied as a filter on candidate elemental compositions, eliminated over 95% of false candidates. Furthermore, the reduction of candidate compositions was greater for measurements with 3 ppm mass accuracy and 2% isotope distribution error than for measurements with 0.1 ppm mass accuracy alone. These studies clearly demonstrate the need for and utility of high resolution and mass accuracy mass analyzers for GC/MS capable of measurements with low isotopic distribution error.

High mass accuracy and resolution analyzers, like the double-focusing magnetic sector and FT-ICR mass spectrometers, are not new to the GC/MS community. The most established of these and one of the first to be coupled to GC,[1] the sector instrument, is currently widely employed for trace detection of toxic tetra- through octa-substituted chlorodibenzo-*p*-dioxin and -furan congeners under EPA method 1613.[10] These instruments typically operate in selective ion monitoring (SIM) mode at resolutions of ~30,000 with ±5 ppm mass measurement accuracy and ~1 s scan speeds.[11] Because the sensitivity of the magnetic sector analyzer falls off precipitously at high resolution,[11,12] this instrument is not well suited for full mass-range profiling or discovery applications. The use of FT-ICR mass spectrometers as GC detectors was first reported in 1980 by Ledford and colleagues[13] and developed extensively in that decade.[14-19] Recently, the Solouki group demonstrated the breadth of potential applications available to this instrument, from

petroleomics to the determination of gas-phase basicities.[20-25] However, due to the cost and complexity of the GC/FT-ICR MS, this instrument is exclusively used in the academic research setting and has yet to be commercialized.

The advent of the Orbitrap and high-field Orbitrap mass analyzers[26,27] has further expanded the range of high mass accuracy and resolution, Fourier-transform instruments available. These instruments, capable of sub-ppm mass accuracy with internal calibration,[28] resolutions up to 150,000[26] and 600,000 (high-field),[27] and isotopic distribution error typically <3-10%,[29] have been extensively developed for use as hybrid[30] and standalone[31] instruments for LC/MS[30-33] and MALDI/MS.[34] Able to collect over a wide mass-range unlike the sector instrument, and not subject to the extensive maintenance requirements of the FT-ICR, the Orbitrap mass spectrometer presents a significant opportunity for the field of high-resolution GC/MS. We report the first adaption of the Orbitrap mass analyzer for GC/MS. Using our laboratory's implementation[33] of electron transfer dissociation (ETD)[35] on a quadrupole linear ion trap (QLT)-Orbitrap hybrid mass spectrometer as a template (where a negative chemical ionization (NCI) source was coupled to the c-trap *via* a long transfer octopole), we modified the NCI source to additionally permit electron ionization (EI) and positive chemical ionization (PCI), and coupled a gas chromatograph, *via* a heated transfer line, directly to the ionization region of the source. Herein, we describe the instrument and software modifications, evaluate the performance of the GC-Orbitrap, and explore two applications germane to high resolution GC-Orbitrap mass spectrometry.

EXPERIMENTAL PROCEDURES

Instrumentation and software modifications. Unless otherwise noted, all experiments were performed on a prototype ETD-enabled hybrid LTQ Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany; described in McAlister, *et al.*[33]) modified to accept a Finnigan MAT (San Jose, CA) GCQ gas chromatograph and to perform EI, PCI, and NCI. Instrumentation modification included removal of the vials, transfer and restrictor lines, and associated heaters and electronics making up the fluoranthene supply system for the NCI source within the ETD module. The NCI source, hereafter referred to as "the rear source," was fitted with a GCQ transfer line and mated to the gas chromatograph. The transfer line temperature was controlled by an external Omega (Stamford, CT) temperature controller. The geometry of the transfer line grounded the normally electrically isolated rear source block. To accommodate the grounded rear source block, a DC voltage was applied to the normally grounded c-trap *via* an external power supply. The rear source was configured for EI or CI by exchange of the appropriate ionization volumes through the vacuum interlock. For CI, methane (99.99%, Airgas Specialty Gases, Port Allen, LA) was introduced through the make-up gas port of the transfer line.

In addition to hardware modifications, the instrument firmware was modified to enable a innovative scan function, the nested scan, to synchronize the Orbitrap acquisition rate and the timescale of gas chromatography separations. The scan sequence optionally allows for internal calibration provided by ions generated by corona discharge (2.8-3 kV) at the front atmospheric pressure (AP) ionization source. The AP-generated ions are injected into the QLT where they are processed by an injection waveform (isolation q, 0.579; waveform amplitude, 0.041; isolation width, 1.7) to result in a mostly homogenous population of a polysiloxane (*m/z* 371.10124).[36] The duration of calibrant transfer from the QLT to the c-trap was optimized on a per-experiment basis to result in a calibrant S/N ~20.

A typical nested scan with internal calibration, illustrated in **Figure 3.1** (top), takes place in five steps (considering scan B, blue): accumulation of internal calibrant ions from the atmospheric inlet in the QLT [a], transfer of calibrant ions to the empty c-trap [b], accumulation of analytes from the GC in the c-trap with the calibrant ions [c], compression and injection of the analyte/calibrant mixture into the Orbitrap [d], and electrodynamic squeezing and mass analysis in the Orbitrap [e]. Due to the modular nature of this dual-ion source QLT-Orbitrap instrument, accumulation of separate ion populations in distinct parts of the instrument (e.g., in the QLT and the c-trap), and Orbitrap mass analysis can take place simultaneously. The nested scan uses this fact to layer or "nest" portions of adjacent scans such that very little time elapses between mass analyses. The "nesting" of three scans (A, B, and C) is illustrated in the bottom portion of **Figure 3.1**. For example, at line [a], while calibrant ions are accumulating in the QLT for scan B [segment 1], scan A (green) analytes from the GC are also accumulating in the c-trap. After the scan A analytes are injected into the Orbitrap just prior to line [b], scan B calibrant ions from the QLT are transferred to the now empty c-trap [segment 2] and scan B GC analytes begin to accumulate in the c-trap with the calibrant ions [segment 3] simultaneous with scan A mass analysis. Thus at line [c], sections of three different scans are occurring: scan A mass analysis, scan B analyte accumulation in the c-trap, and scan C calibrant accumulation in the QLT. When scan A mass analysis ends (line [d]), scan B ions are immediately injected [segment 4] into the Orbitrap for mass analysis [segment 5] and scan C continues on to GC analyte accumulation in the c-trap (line [e]).

In a second implementation, we adapted a commercial LTQ Orbitrap XL system to perform similar experiments. Instrumentation and system operation were identical with the exception of the following improvements: to make use of the high capacity of the XL system's HCD cell and improved HCD cell-to-c-trap ion transfer efficiency, GC analytes were stored in

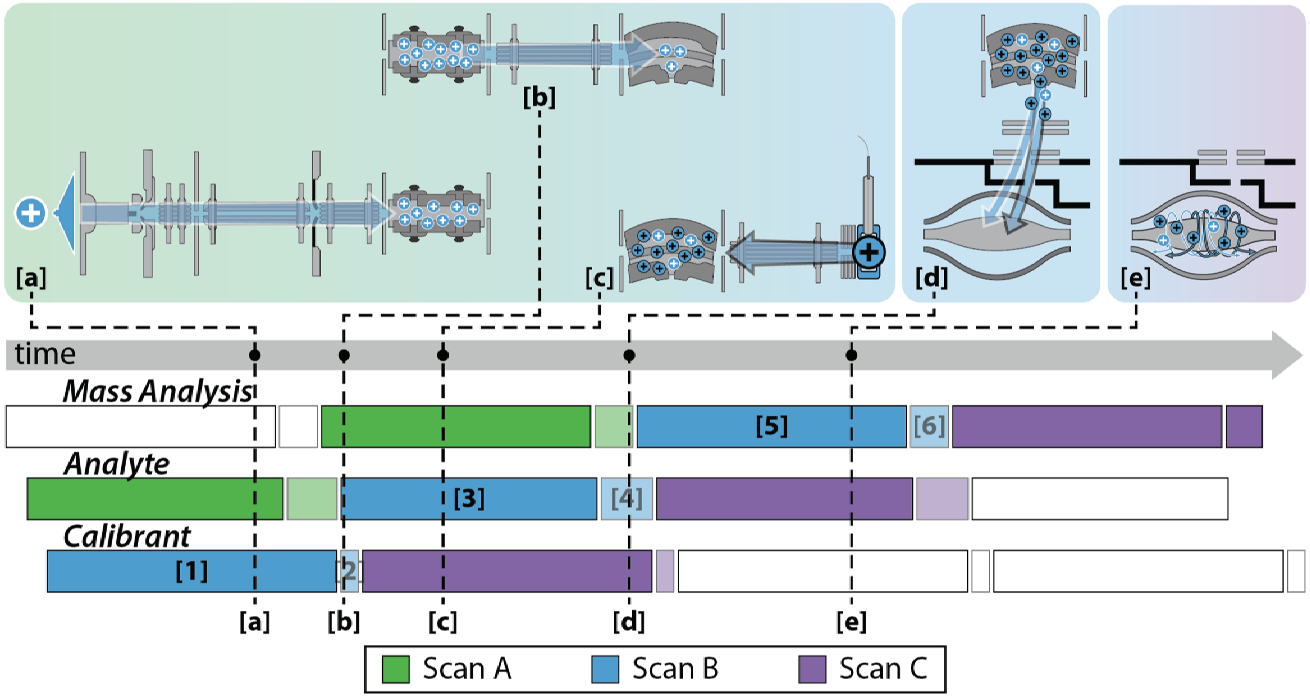


Figure 3.1. A typical nested scan cycle (scan B, blue) with internal calibration proceeds in five steps (top): [**a**] accumulation of internal calibrant ions in the QLT [segment 1], [**b**] transfer of calibrant ion to c-trap [segment 2], [**c**] accumulation of GC analytes in the c-trap with the calibrant ions from step [**b**] [segment 3], [**d**] compression and injection of ions into the Orbitrap [segment 4], and [**e**] electrodynamic squeezing and Orbitrap mass analysis [segment 5]. Segment 6 is the time between mass analyses of adjacent scans. The nested scan layers the steps of scan B with the steps of adjacent scans A (green) and C (purple) as temporally indicated at bottom. For example, at time [**c**], scan C calibrant ions accumulate in the QLT, while scan B GC analytes accumulate in the c-trap and scan A ions undergo mass analysis in the Orbitrap.

the HCD cell during accumulation, and internal calibrant ions from the ESI source were injected after the accumulation of GC analytes rather than before. These changes resulted in a negligible decrease in the overall scan cycle time, and improved both internal calibrant stability and ion transfer efficiencies. With the exception of a discussion on improved internal calibrant stability, no distinction between the two instruments will be drawn below.

Standards. Unless otherwise noted, all chemicals were purchased from Sigma-Aldrich, St. Louis, MO. Preliminary experiments and tuning using calibrant compound perfluorotributylamine (FC-43; Scientific Instrument Services, Ringoes, NJ) were carried out with a custom-constructed calibration gas probe interfaced to the rear source through the vacuum interlock. FC-43 was introduced into the source housing by a passive leak controlled with a fine needle valve and solenoid valve in series. Experiments with high-mass calibrant, 2,4,6-tris(pentadecafluoroheptyl)-1,3,5-triazine (TPFHST), were carried out *via* the heated vial inlet of the unmodified instrument.

Sample preparation and gas chromatography-mass spectrometry. All chromatography was performed using a GCQ gas chromatograph with a 30 m x 0.25 mm (ID) x 0.25 μm (d_f) Crossbond® 5% diphenyl/95% dimethyl polysiloxane column (Restek Rtx-5MS, Bellefonte, PA) and helium carrier gas at 40 cm/s. Unless otherwise specified, all injection volumes were 1 μL, splitless for 1 min, and thereafter, split 50:1. The septum purge flow was 2.5 mL/min. The injector, transfer line, and source temperatures were 250 ºC, 300 ºC, and 220 ºC, respectively. A three minute solvent delay was imposed for all runs. Prior to all experiments, the rear source and instrument ion optics were tuned using FC-43.

Octafluoronaphthalene. Serial dilutions of octafluoronaphthalene (OFN) in isooctane (Fisher Scientific, Fairlawn, NJ) were made to the following concentrations: 100, 50, 10, 5, 1, 0.1, 0.05, 0.01, 0.005, and 0.001 pg/uL. Triplicate injections, starting with the lowest concentration,

were performed by hand for each dilution with a blank solvent injection after every two OFN runs. The following oven program was employed: 1 min isothermal at 40 ºC, 30 ºC/min to 250 ºC, and 8 min isothermal at 250 ºC before rapid cooling to 40 ºC. The rear source was configured for EI (emission current 200 μA, electron energy -70 eV). Mass spectra were acquired from *m/z* 200-300 using the nested scan function without internal mass calibration at a 15,000 resolution.

Polychlorinated dibenzo-p-dioxins and dibenzofurans. A mixture of 17 polychlorinated dibenzo-*p*-dioxin and dibenzofuran congeners in *n*-nonane was purchased from Cerilliant (Round Rock, TX) and used as received. Six microliters were separated using the following oven program: 2 min isothermal at 160 ºC, 25 ºC/min refocusing gradient to 220 ºC, 2 min isothermal at 220 ºC, 5 ºC/min to 310 ºC, and 5 min isothermal before rapid cooling to 160 ºC. EI was employed with an emission current of 300 μA and a reduced electron energy (-35 eV). Mass spectra were acquired from *m/z* 300-500 at 30,000 resolution using the nested scan function with internal calibrant supplied by corona discharge (2.8 kV) at the atmospheric pressure inlet.

Skinner List. A semivolatiles mixture containing 33 Skinner List polycyclic aromatic hydrocarbons was purchased from Restek and used as a 1:100 dilution in dichloromethane. One microliter (10-20 ng per compound) was separated using the following oven program: 2 min isothermal at 35 ºC, 15 ºC/min to 100 ºC, 4 min isothermal at 100 ºC, 3 ºC/min to 310 ºC, and 10 min isothermal at 310 ºC prior to cooling to 35 ºC. PCI was carried out with source housing methane pressure of 3×10^{-4} torr and emission current of 215 μA. Mass spectra were acquired using the nested scan function from *m/z* 60-600 at 30,000 resolution with internal calibration.

Arabidopsis thaliana. Whole *A. thaliana* plants were flash frozen in liquid nitrogen and macerated to a fine powder by mortar and pestle. Aliquots of 33-54 mg homogenized powder were extracted as detailed by Fiehn.[37] Briefly, each aliquot was combined with cold, degassed extraction solution ($CH_3OH/CHCl_3$/DI-H_2O, 2.5:1:1 v/v/v) and incubated at 4 ºC, solids were

pelleted by centrifugation, and the supernatant was extracted with DI-H_2O. The polar and lipophilic phases were separated by centrifugation and polar phase dried under reduced pressure. The dried samples were then methoxyaminated, and silylated with *N*-(*t*-butyldimethylsilyl)-*N*-methyltrifluoroacetamide (MTBSTFA + 1% TBDMCS, Pierce). Samples sat for 2 h at room temperature before analysis. The oven program was as follows: 2 min isothermal at 80 ºC for, 5 ºC/min to 310 ºC, and 5 min isothermal at 310 ºC before cooling. The rear source was configured for methane PCI (emission current 130 µA, electron energy -103 eV, source temperature 200 ºC, and source housing methane pressure of 3.5 x 10^{-4} Torr). Mass spectra were acquired from *m/z* 83-1000 using the nested scan function at resolution 30,000 with internal calibration.

Data analysis and spectral accuracy calculations. In all experiments, spectra were queried by hand in Qual Browser (Thermo Fisher Scientific, version 2.0.7). OFN peak areas were integrated automatically using the ICIS peak detection algorithm within Qual Browser (±1 ppm mass tolerance). Post-acquisition data recalibration using the corona discharge internal calibrant was performed on a per spectrum basis with a 10 ppm mass tolerance in Recalibrate Offline (Thermo Fisher Scientific). Spectral accuracy calculations were performed using the MassWorks software suite (Cerno Biosciences, Danbury, CT).

RESULTS AND DISCUSSION

Instrument modifications and nested scan development. The ultimate goal of this work was to implement the high resolution and high mass accuracy Orbitrap as a viable mass analyzer for gas chromatography applications. To meet the demands of the gas chromatographic time-scale, we set a goal of at least a 4 Hz scan rate at resolution 15,000. In addition to adapting the instrument to operate in a rear source-centric manner, achieving this goal required a significant rearrangement of the typical scan sequence employed by this hybrid instrument.

We first modified the instrument software to permit: (1) transfer of rear source anions or cations (produced via EI, PCI, or NCI) into the Orbitrap after injection into the QLT and, (2) injection of rear source ions directly into the Orbitrap. In both cases, injection times were either determined by a short pre-scan in the QLT immediately prior to the full scan, or fixed by the user. As during conventional operation of this instrument, the pre-scan, injection time, and additional instrument-associated times must all occur during the scan matrix in which ions are analyzed. In addition to Orbitrap mass analysis times of 123, 223, 423, 823, and 1625 ms for resolutions 7,500, 15,000, 30,000, 60,000, and 100,000, respectively, this requirement adds ~5-500 ms for injection, depending on source brightness, and some 200 ms of pre-scan and additional instrument-associated times. Even at the lowest resolution (7,500), the goal of a 4 Hz scan cycle is not possible with these scan types. Attempts to minimize the scan duration to ensure acquisition of multiple spectra over a chromatographic peak require a significant trade-off in sensitivity (available at higher resolutions and longer injection times).

To address these problems, we developed the nested scan, which minimizes scan time and maximizes scan duty cycle without a trade-off in sensitivity. Here, as described in the experimental section and illustrated in **Figure 3.1**, the nested scan layers the steps of one scan—calibrant accumulation and transfer (line [a], [b]), analyte accumulation (line [c]), injection into the Orbitrap (line [d]), and Orbitrap mass analysis (line [e])—to perform in parallel with two other scan cycles. Thus, instead of leaving parts of the instrument idle while others are in use, as in conventional operation (i.e., postponing accumulation of new ion populations until mass analysis ends), injection of analyte ions from the GC/rear source into the c-trap, as well as the accumulation of internal calibrant ions in the QLT, occurs during almost the entire duration of the previous scan's mass analysis. Because the Orbitrap control code prevents the start of the next scan until mass analysis is complete, the length of analyte/calibrant injection/accumulation is tied

to the length of the mass analysis. The mass analysis length, in turn, is determined by the desired resolution of Orbitrap acquisition; thus, analyte and internal calibrant accumulation times are determined directly by the resolution of Orbitrap mass analysis. This link is clearly illustrated by **Table 3.1**, where the duration of each scan segment is given at each resolution (all times determined externally by oscilloscopic measurements). The scan time and rate for the nested scan with internal calibration ranges from 153 ms or 6.5 Hz at 7,500 resolution to 1644 ms or 0.6 Hz at 100,000 resolution. Resolutions of 15,000 and 30,000 mark the most appropriate settings for GC analyses as these settings give acceptable scan rates (4.0 and 2.2 Hz, respectively) and resolution, as well as calibrant and analyte accumulation times sufficient for reasonable signal. The nested scan duty cycle, in terms of the proportion of total scan time dedicated to mass analysis, approaches 100% at 100,000 resolution. This figure is significantly greater than that of conventional scan types because the time required for steps other than mass analysis are effectively reduced to zero.

Instrument and scan figures of merit. As our aim in developing this instrument was to achieve high-mass accuracy and high-resolution mass analysis on a gas-chromatographic time-scale, several figures of merit warrant mention, namely, the internally and externally calibrated mass accuracy, linear dynamic range, sensitivity, limit of detection, and spectral, or isotopic distribution, accuracy achievable with this instrument.

Prior to assessing the mass accuracy achievable with the nested scan and polysiloxane internal calibrant, we first established the baseline optimal mass accuracy achievable using conventional scan sequences and fluoranthene as internal calibrant.[38] **Figure 3.2A** displays a single scan EI spectrum of FC-43 collected at 30,000 resolution with fluoranthene internal calibrant (*m/z* 202.07770). We chose to use fluoranthene as the internal calibrant here to rule out mass accuracy deviations that may have arisen from the implementation of the nested scan or the

Table 3.1. Times (ms) and relevant metrics for nested scan sequence components at all available resolution settings (measured relative to *m/z* 400).

[#]	Description	Resolution Setting				
		7,500	15,000	30,000	60,000	100,000
[1]	Internal calibrant accumulation in linear trap (ms)	138	240	440	840	1640
[2]	Internal calibrant transfer to c-trap; accumulation downtime (ms)	13	13	13	13	13
[3]	GC analyte accumulation in c-trap (containing internal calibrant) (ms)	112	212	412	812	1625
[4]	Compression and injection of analyte/calibrant into orbitrap; accumulation downtime (ms)	41	41	41	41	41
[5]	Orbitrap mass analysis (ms)	123	223	423	823	1625
[6]	Mass analysis downtime (ms)	30	30	30	30	30
	Scan time (ms)	**153**	**252**	**453**	**852**	**1644**
	Scan rate (Hz)	**6.5**	**4.0**	**2.2**	**1.2**	**0.6**
	Duty Cycle (%)	**80.4**	**88.5**	**93.4**	**96.6**	**98.8**

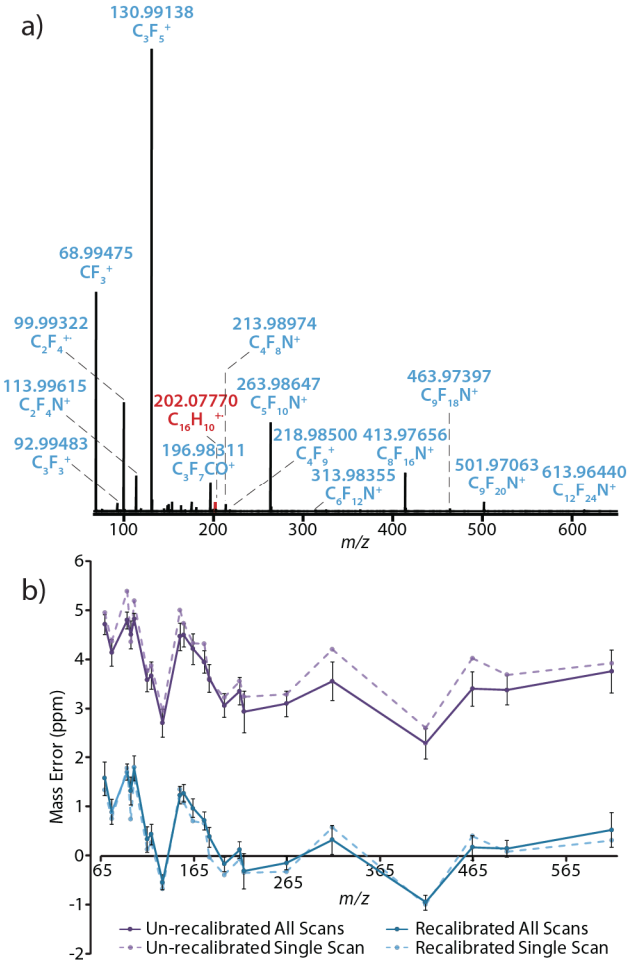


Figure 3.2. (**A**) Single-scan electron ionization spectrum of FC-43 at resolution 30,000 with internal calibrant peak at *m/z* 202.07770. (**B**) Effect of mass recalibration based on *m/z* 202.07770 for each labeled peak in (**A**) plotted for the single scan in (**A**) (dashed lines) and one hundred averaged scans (solid lines) prior to recalibration (purple) and after recalibration (blue).

corona discharge source. This scan was collected several days after external mass calibration and represents the normal extent of instrumental mass accuracy drift. Without internal calibration, all major peaks can be confidently identified with masses known to four decimal places. **Figure 3.2B** shows the effect of mass recalibration based on *m/z* 202.07770. The mass error (ppm) of each labeled peak in **Figure 3.2A** is plotted for the single scan shown in **Figure 3.2A** and 100 averaged scans prior to recalibration (purple) and after recalibration (blue). Recalibration corrects for systemic instrumental mass error (~4 ppm prior to recalibration), as well as minimizes random mass errors, as evidenced, in the recalibrated data, by narrower error bars (±1 standard deviation) and greater adherence of the single scan mass errors to the mean mass errors. The recalibrated mass errors for the remaining data presented herein (following external mass calibration) routinely range within ±1 ppm of the theoretical mass. Our per scan mass accuracy results meet or exceed the results of other such studies employing "lock-mass"-type internal calibration.[28,38,39]

To address concerns that the calibrant ion population delivered by the nested scan cycle would not be of sufficient robustness to provide effective internal calibration, we evaluated the nested scan's internal calibration scheme by tracking the stability of polysiloxane (*m/z* 371.10124) over the duration of an analysis. **Figure 3.3** shows the abundance of the internal calibrant relative to the analyte elution profile. The average internal calibrant intensity was 1.0 x 10^3 while the GC analyte intensity ranged up to 3.5 x 10^6. On closer inspection (inset, **Figure 3.3**), however, the population of internal calibrant dropped to zero when high abundance analytes eluted from the GC, likely due to a loss of ions caused by overfilling the c-trap. While the internal calibrant mass in neighboring scans can be used for recalibration, recalibration from within the same scan is more accurate as scan-to-scan fluctuations (e.g. space-charge effects) are properly accounted for, and is preferred given the constraints of our existing post-acquisition

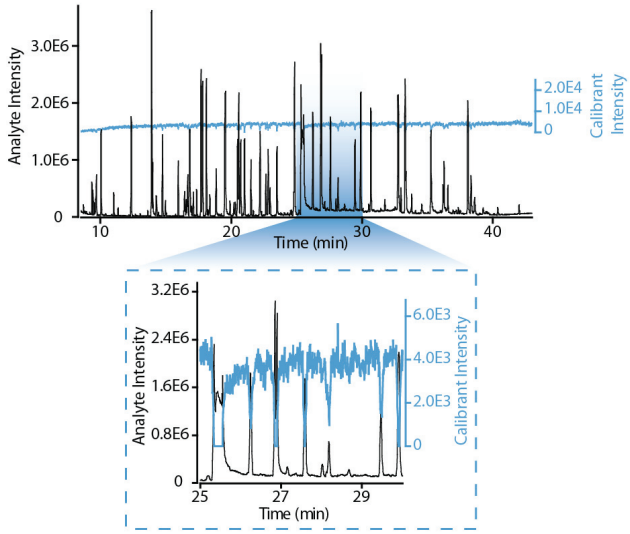


Figure 3.3. Internal calibrant stability plotted over the analyte elution profile. The inset indicates that the population of internal calibrant drops to zero when high abundance analytes elute from the GC.

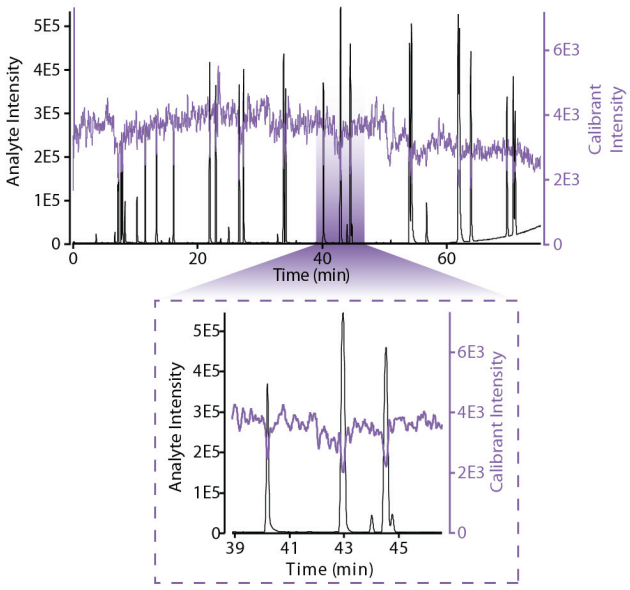


Figure 3.4. Internal calibrant stability plotted over the analyte elution profile in the second implementation of the GC-Orbitrap on an LTQ Orbitrap XL system. The inset shows that the internal calibrant population is less susceptible to fluctuation when high-abundance analytes elute from the GC.

recalibration software. In our second implementation, we modified the nested scan cycle to improve internal calibrant stability by transferring internal calibrant ions from the QLT *after* the payload of ions from the GC had been accumulated in the c-trap. Calibrant stability under this second configuration was substantially improved, with each scan having a robust population of calibrant ions irrespective of analyte abundance (**Figure 3.4**).

In addition to mass accuracy, spectral, or isotopic distribution, accuracy is an important orthogonal filter for determining the elemental composition of unknown compounds. Spectral accuracy is a measure of the degree of congruence between an experimentally-acquired isotopic distribution and a theoretically calculated isotopic distribution. The spectral accuracy obtainable in a mass spectrum is a characteristic of the mass analyzer used and, to date, only one report addresses this figure of merit for the Orbitrap analyzer.[29] In that study, Erve, *et al.* determined spectral accuracy characteristics for ten large natural products (> 600 Da) at all available resolution settings. Higher resolution settings were correlated with decreased spectral accuracy, mirroring a trend more extensively studied in the context of FT-ICR MS. Like in FT-ICR, isotopic distribution errors are believed to arise from destructive and constructive interferences between oscillators having very small frequency differences (e.g., isotopic peaks),[29,40,41] or due to disparate dephasing rates and loss of cloud coherence among ion clouds of different densities.[42,43] Spectral effects of the former phenomenon, called the "isotope beating" effect, worsen at high mass, high charge, and overly-short transient times (relative to beat period), while effects of the latter phenomenon are more apparent at long transient times (>5 s). These phenomena remain relatively unstudied in the context of Orbitrap mass analysis.

To benchmark the effect of these phenomena on small molecule Orbitrap spectra, we analyzed a known mixture of semivolatile polycyclic aromatic hydrocarbons (**Table 3.2**). We determined the spectral accuracy for each compound using the molecular ion or, if the molecular

Table 3.2. Effect of spectral accuracy on accurate elemental composition identifications in a known semivolatile polycyclic aromatic hydrocarbon mixture.

Name	Mass Error (ppm)	Rank by Mass Only[1]	S. A. (%)	Rank by S. A. & Mass[1]	Percent over(+)/under(-) next best molecular formula
pyridine	-0.57	1 (1)	94.2	1 (1)	--
phenol	0.20	1 (1)	93.1	1 (1)	--
1,3-dichlorobenzene	0.67	1 (5)	82.3	1 (5)	19.5
1,4-dichlorobenzene	0.12	1 (4)	81.3	1 (4)	17.8
1,2-dichlorobenzene	0.60	1 (5)	82.3	1 (5)	19.3
indene	0.62	1 (1)	98.4	1 (1)	--
2-methylphenol	-0.01	1 (1)	97.7	1 (1)	--
3-,4-methylphenol	0.08	1 (1)	97.8	1 (1)	--
2,4-dimethylphenol	0.31	1 (1)	98.3	1 (1)	--
naphthalene	-0.83	1 (1)	98.1	1 (1)	--
quinoline	0.26	1 (1)	97.8	1 (1)	--
1-methylnaphthalene	-0.19	1 (1)	97.6	1 (1)	--
dimethylphthalate [1]	0.28	2 (6)	58.8	5 (6)	-0.8
dimethylphthalate [2]	1.22	2 (4)	97.7	1 (4)	2.6
acenaphthene	-0.17	1 (2)	98.1	1 (2)	8.0
2,4-dinitrophenol	-0.15	2 (14)	85.1	1 (14)	0.1
4-nitrophenol	-0.35	1 (3)	98.7	1 (3)	3.2
fluorene	0.38	2 (2)	98.1	1 (2)	7.9
diethylphthalate[1]	0.20	2 (6)	96.2	1 (6)	1.6
diethylphthalate[2]	0.11	1 (3)	97.8	1 (3)	1.8
diethylphthalate[3]	0.20	1 (3)	97.5	1 (3)	3.2
phenanthrene	-0.15	1 (3)	98.0	1 (3)	7.7
anthracene	-1.49	3 (3)	97.2	1 (3)	7.9
di-n-butylphthalate[1]	0.05	1 (6)	94.2	2 (6)	-0.7
di-n-butylphthalate[2]	0.73	2 (2)	97.6	1 (2)	3.0
fluoranthene	-0.13	1 (6)	97.9	1 (6)	6.1
pyrene	0.26	1 (5)	97.6	1 (5)	4.7
benz(a)anthracene	-0.12	2 (12)	89.1	1 (12)	1.4
chrysene	-0.25	2 (12)	97.5	1 (12)	4.8
bis(2-ethylhexyl)phthalate[1]	0.01	1 (15)	91.9	6 (15)	-0.7
bis(2-ethylhexyl)phthalate[2]	1.61	4 (4)	89.9	1 (4)	0.6
benzo(b)fluoranthene	-0.15	3 (19)	97.2	1 (19)	5.0
benzo(k)fluoranthene	0.25	3 (16)	97.4	1 (16)	5.1
benzo(a)pyrene	0.09	2 (13)	96.1	1 (13)	2.4
dibenz(a,h)acridine	0.26	3 (24)	92.0	1 (24)	3.3
indeno(123cd)pyrene	-0.82	17 (43)	94.0	1 (43)	3.4
dibenz(a,h)anthracene	-0.42	6 (26)	86.1	1 (26)	1.2

[1]the value encapsulated by parenthesis is the total number of elemental compositions with that particular mass tolerance

ion was of low abundance, 2-3 fragment ions with the Cerno MassWorks sCLIPS algorithm.[29,44] sCLIPs is a proprietary algorithm that redefines the experimental spectral line-shape as a mathematical function and then matches the recalibrated spectrum to a theoretical distribution within a specified mass tolerance. The output of sCLIPS is a percentage describing the proportion of the experimental line-shape described by the theoretical isotopic profile, where candidates with higher spectral accuracy are considered to be more likely elemental compositions for the ion of interest. At the resolution employed (30,000), the spectral accuracies obtained for ions of sufficient abundance were comparable to those reported by Erve, *et al.,* who employed an unmodified QLT-Orbitrap and electrospray ionization.[29] Using mass accuracy alone (after mass recalibration with the internal calibrant), unique elemental composition assignments were possible for most lower molecular weight compounds. For compounds that were not uniquely assigned, the correct elemental composition was typically ranked within the top four hits. Using both mass and spectral accuracy filters, correct elemental compositions were ranked as the top hit for all compounds except three phthalate compounds (dimethyl-, di-*n*-butyl-, bis(2-ethylhexyl)-). Due to the low abundance of the phthalate molecular ions, these isotopic distributions were susceptible to interference from the Orbitrap's thermal noise band, making the inclusion of spectral accuracy information deleterious to elemental composition assignment of these compounds. In such cases, the phthalates were identified by top ranking hits for their diagnostic fragment ions (*m/z* 163, $C_9H_7O_3$; and, *m/z* 149, $C_8H_5O_3$).[45] For all but two compounds, 2,4-dinitrophenol and a fragment ion of bis(2-ethylhexyl)phthalate, $C_8H_5O_3$, the spectral accuracy of the correct elemental composition bested the second-ranked elemental composition by greater than 1%. The utility of spectral accuracy information is most apparent for the higher molecular weight compounds: for example, by mass alone, indeno(1,2,3-*cd*)pyrene's elemental composition was ranked 17 of 43 candidates within 2 ppm. When spectral accuracy information was included,

this composition was ranked first, surpassing the next ranked elemental composition by 3.4%. Our results suggest that the spectral accuracy obtainable with this instrument, in tandem with low mass errors (typically <1 ppm), will help facilitate assignment of small molecule elemental compositions. Further studies are needed to determine what spectral accuracy is necessary to consider the top-ranked candidate to be statistically a better candidate than other candidates in the list.

Next, to assess the limit of detection, sensitivity, and linearity of response for our instrument, we measured the area under the curve for a triplicate series of 1 μL injections of octafluoronaphthalene (OFN) in iso-octane, scanning from *m/z* 200-300. The Orbitrap response was linear from 10 fg (37 amol) to 100 pg (37 fmol) with 24% RSD of response factors (R^2 = 0.998) and empirical limit-of-detection (LOD) of 10 fg (S/N 5) (**Figure 3.5**). The resolution (20,000) and mass accuracy (<1-3 ppm) of the diagnostic peaks for OFN were unchanged over the entire concentration range.

In the GC/MS field, the use of OFN to define instrument sensitivity is industry standard;[46,47] current GC-enabled mass-selective detectors and triple quadrupoles report S/N 300-500:1 for a 1 μL injection of 1 pg/μL OFN.[47] In our analysis, we achieved a depressed average S/N of 125:1 at 1 pg/μL while maintaining a competitive LOD of 10 fg. The modest sensitivity of the GC-Orbitrap is primarily attributable to inefficiencies in ion transfer and trapping and, to some extent, the inherent diminished sensitivity of image current detection. The competitive LOD, however, is a testament to the enhanced selectivity afforded by high-resolution mass analysis.[46] Even at very low analyte concentrations, matrix interferences are easily resolved. With instrument and firmware optimizations to improve ion transmission and decrease ion loss, the sensitivity of this instrumental is expected to match that of commercial GC-mass spectrometers. Additionally, the development of the more sensitive, compact, high-field Orbitrap

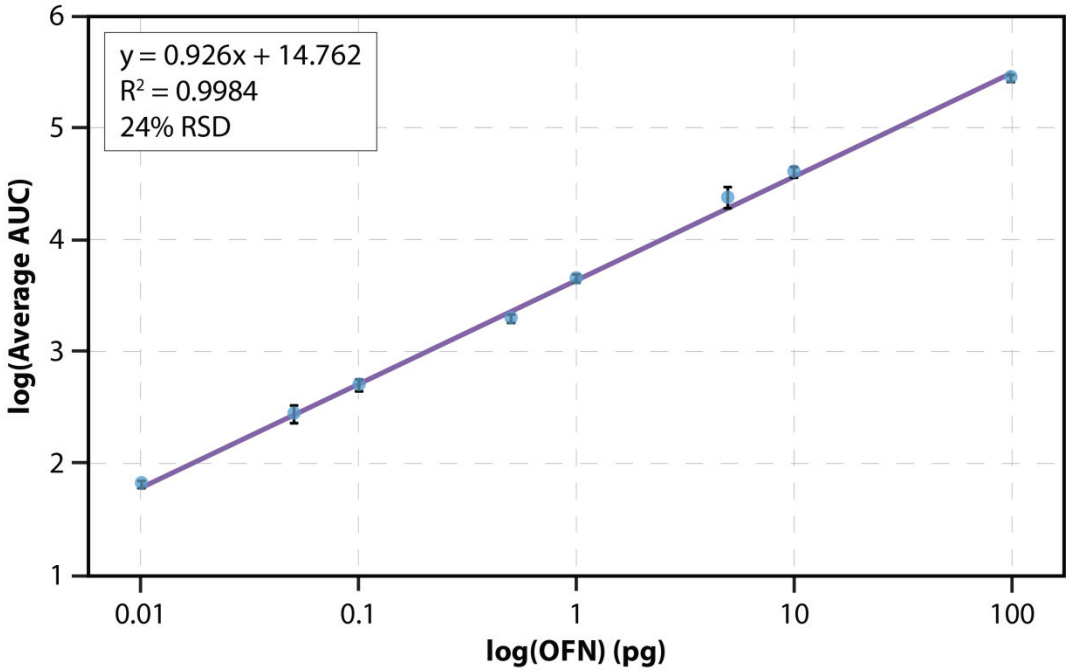


Figure 3.5. Log-log plot of the Orbitrap response, as the average area-under-the-curve (AUC) over triplicate analyses at each concentration, versus octafluoronaphthalene (OFN) concentration from 10 fg/µL to 100 pg/µL. Response was linear over the entire range with a 24% RSD and LOD of 10 fg.

analyzer,[27,48] capable of single ion detection and accelerated spectral acquisition rates, image current detection may cease to be a limiting factor for high sensitivity analyses.

Applications. To demonstrate the utility of our GC-enabled QLT-Orbitrap, we chose two applications for which high resolution GC/MS analysis is critical: the determination of polychlorinated dibenzo-*p*-dioxins (PCDD) and dibenzofurans (PCDF), and the profiling of primary metabolites in *Arabidopsis thaliana* extracts without isotopic dilution.

PCDDs and PCDFs are persistent environmental toxins, released as the byproducts of many industrial processes, which contaminate food supplies (e.g., fish) and lead to adverse health effects in humans and animals.[49] Seventeen of these compounds are considered particularly toxic and are monitored routinely as part of EPA method 1613 revision B[10] and European Standard EN1948-1/2/3.[45] Method 1613 requires the use of a high resolution double-focusing sector mass spectrometer capable of 10k resolution at ±5 ppm mass accuracy, and of monitoring of 12 SIM windows in 1 s with a S/N ≥ 10. In addition, for each compound, ion abundance ratios must fall within ±15% of expected ratios.[10] While no attempt was made to meet the full gas chromatographic requirements of the method, we analyzed seventeen of the most toxic PCDD and PCDF congeners to benchmark our mass spectral capabilities in the context of the EPA method. **Figure 3.6** shows a GC chromatogram of this mixture and a mass spectrum corresponding to overlapping elution profiles of the 2,3,4,6,7,8-hexaCDF (HxCDF) and 1,2,3,4,7,8-hexaCDD (HxCDD) congeners. The mass spectrum demonstrates the ability of the Orbitrap analyzer to fully resolve the isotopic envelopes of both congeners simultaneously (in full scan mode rather than *via* SIM) while maintaining mass and spectral accuracy (see **Table 3.3**). In a single full-scan analysis, all congeners were identified within 1 ppm mass error with internal calibration by the exact masses required by the EPA method, >31,000 resolution for all ions, and well within the ±15% QC limits for theoretical ion abundance ratios. These results suggest that,

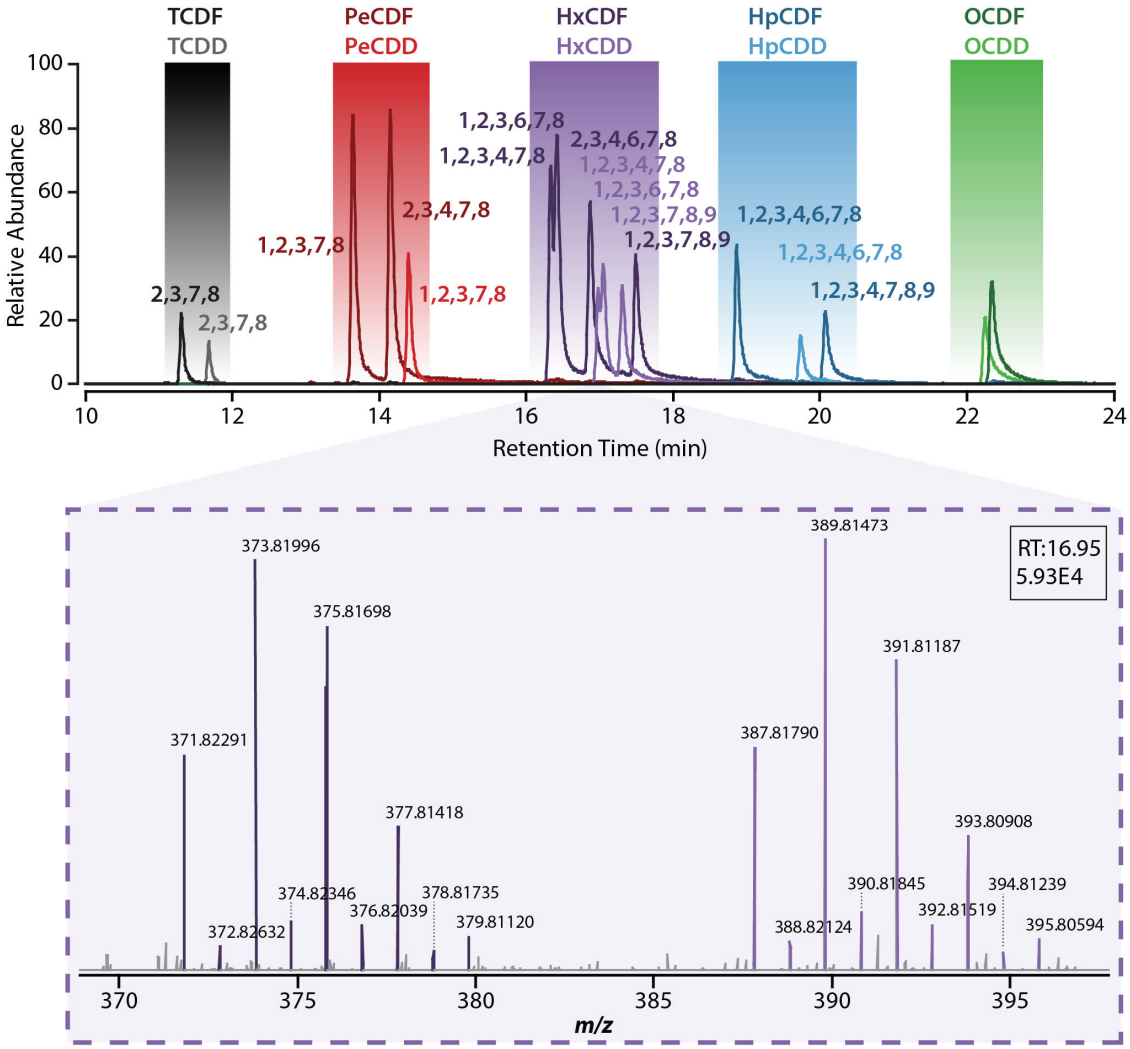


Figure 3.6. GC chromatogram demonstrating resolution of the seventeen toxic polychlorinated dibenzo-*p*-dioxin (PCDD) and dibenzofuran (PCDF) congeners in a standard mixture (T, tetra; Pe, penta; Hx, hexa; Hp, hepta; O, octa). The inset is a subsection of a full-scan mass spectrum corresponding to the overlapping elution profiles of the 2,3,4,6,7,8-HxCDF and 1,2,3,4,7,8-HxCDD congeners. See **Table 3.3** for mass and spectral accuracy characteristics corresponding to this analysis.

Table 3.3. Mass accuracy, resolution, and isotopic abundance ratio errors for PCDF and PCDD congeners.

Congener	Substance	Ion Type	Elemental Composition	Mass Error (ppm)	Resolution	Abundance Ratio	Ratio Error (%)
2,3,7,8	TCDF	M M+2	$C_{12}H_4O^{35}Cl_4$ $C_{12}H_4O^{35}Cl_3{}^{37}Cl$	-0.780 -0.873	38800 38600	M/(M+2)	1.63
2,3,7,8	TCDD	M M+2	$C_{12}H_4O_2{}^{35}Cl_4$ $C_{12}H_4O_2{}^{35}Cl_3{}^{37}Cl$	-0.600 -0.410	37800 37700	M/(M+2)	-4.96
1,2,3,7,8	PeCDF	M+2 M+4	$C_{12}H_3O^{35}Cl_4{}^{37}Cl$ $C_{12}H_3O^{35}Cl_3{}^{37}Cl_2$	-0.927 -0.919	36400 36304	(M+2)/(M+4)	0.94
2,3,4,7,8	PeCDF	M+2 M+4	$C_{12}H_3O^{35}Cl_4{}^{37}Cl$ $C_{12}H_3O^{35}Cl_3{}^{37}Cl_2$	-0.074 0.047	36500 36404	(M+2)/(M+4)	0.80
1,2,3,7,8	PeCDD	M+2 M+4	$C_{12}H_3O_2{}^{35}Cl_4{}^{37}Cl$ $C_{12}H_3O_2{}^{35}Cl_3{}^{37}Cl_2$	-0.756 -0.668	35800 35304	(M+2)/(M+4)	0.07
1,2,3,4,7,8	HxCDF	M+2 M+4	$C_{12}H_2O^{35}Cl_5{}^{37}Cl$ $C_{12}H_2O^{35}Cl_4{}^{37}Cl_2$	-0.086 -0.245	34600 34904	(M+2)/(M+4)	-1.81
1,2,3,6,7,8	HxCDF	M+2 M+4	$C_{12}H_2O^{35}Cl_5{}^{37}Cl$ $C_{12}H_2O^{35}Cl_4{}^{37}Cl_2$	-0.086 -0.085	34600 35004	(M+2)/(M+4)	2.30
2,3,4,6,7,8	HxCDF	M+2 M+4	$C_{12}H_2O^{35}Cl_5{}^{37}Cl$ $C_{12}H_2O^{35}Cl_4{}^{37}Cl_2$	-0.086 -0.005	34600 34804	(M+2)/(M+4)	0.28
1,2,3,7,8,9	HxCDF	M+2 M+4	$C_{12}H_2O^{35}Cl_5{}^{37}Cl$ $C_{12}H_2O^{35}Cl_4{}^{37}Cl_2$	-0.005 -0.005	34600 34904	(M+2)/(M+4)	1.68
1,2,3,4,7,8	HxCDD	M+2 M+4	$C_{12}H_2O_2{}^{35}Cl_5{}^{37}Cl$ $C_{12}H_2O_2{}^{35}Cl_4{}^{37}Cl_2$	-0.121 -0.043	34000 33804	(M+2)/(M+4)	1.97
1,2,3,6,7,8	HxCDD	M+2 M+4	$C_{12}H_2O_2{}^{35}Cl_5{}^{37}Cl$ $C_{12}H_2O_2{}^{35}Cl_4{}^{37}Cl_2$	-0.865 -0.860	34100 33704	(M+2)/(M+4)	0.42
1,2,3,7,8,9	HxCDD	M+2 M+4	$C_{12}H_2O_2{}^{35}Cl_5{}^{37}Cl$ $C_{12}H_2O_2{}^{35}Cl_4{}^{37}Cl_2$	-0.428 -0.273	34000 33804	(M+2)/(M+4)	6.27
1,2,3,4,6,7,8	HpCDF	M+2 M+4	$C_{12}HO^{35}Cl_6{}^{37}Cl$ $C_{12}HO^{35}Cl_5{}^{37}Cl_2$	-0.098 -0.098	33500 33204	(M+2)/(M+4)	0.98
1,2,3,4,7,8,9	HpCDF	M+2 M+4	$C_{12}HO^{35}Cl_6{}^{37}Cl$ $C_{12}HO^{35}Cl_5{}^{37}Cl_2$	-0.613 -0.464	33400 33404	(M+2)/(M+4)	2.03
1,2,3,4,6,7,8	HpCDD	M+2 M+4	$C_{12}HO_2{}^{35}Cl_6{}^{37}Cl$ $C_{12}HO_2{}^{35}Cl_5{}^{37}Cl_2$	-0.484 -0.573	32600 32604	(M+2)/(M+4)	3.20
--	OCDF	M+2 M+4	$C_{12}O^{35}Cl_7{}^{37}Cl$ $C_{12}O^{35}Cl_6{}^{37}Cl_2$	-0.516 -0.311	32200 31900	(M+2)/(M+4)	2.97
--	OCDD	M+2 M+4	$C_{12}O_2{}^{35}Cl_7{}^{37}Cl$ $C_{12}O_2{}^{35}Cl_6{}^{37}Cl_2$	0.192 0.191	31500 31400	(M+2)/(M+4)	-2.62

as a high-resolution GC detector, the Orbitrap is capable of meeting the mass spectral requirements of this EPA method. An advantage of Orbitrap full-scan analysis over sector analysis in this method is the flexibility afforded to the operator to choose any isotopic peaks in a given ion cluster to fulfill the ion abundance ratio requirement, rather than just the particular isotopes monitored. This is particularly empowering in cases where matrix ions interfere at a monitored target mass and render the sector data at that mass useless.

As mentioned in the introduction, identifying unknown compounds in complex samples is significantly more complicated than determining the presence of a known compound with known chromatographic and mass spectral characteristics. So, in contrast to the previous example, we analyzed a complex *Arabidopsis thaliana* whole-plant metabolite extract to gauge the "discovery" capabilities of this instrument. Using standard protocols,[37] primary metabolites were extracted from a whole plant macerate by methanol-chloroform extraction, methoxyaminated and silylated, and analyzed over an hour-long temperature gradient.

As no extensive high-resolution and high-mass accuracy EI spectral databases of *Arabidopsis thaliana* metabolites exist in the public domain to our knowledge, elemental composition determinations were made using a *de novo* approach, based on the molecular ion, rather than spectral matching with unit- spectral accuracy of the pseudomolecular ion (and associated diagnostic ions, *vide infra*) to determine the resolution libraries (as spectral appearances can differ significantly, *vide infra*). To minimize fragmentation and maximize the abundance of the pseudomolecular ion ($[M + H]^+$), we employed positive chemical ionization (PCI) and *t*-butyldimethylsilylation. We then used the accurate mass and elemental composition of the eluting species. In addition to leaving more of the pseudomolecular ion intact, methane PCI provides diagnostic molecular ion adduct peaks ($[M + C_2H_5]^+$ and $[M + C_3H_5]^+$) that help pinpoint the pseudomolecular ion. Similarly, *t*-butyldimethylsilylation of acidic protons directs

fragmentation toward diagnostic losses of *t*-butyl- and methyl-groups from the attached silyl groups, leaving the parent compound intact, and further aiding the localization of the pseudomolecular ion in the spectrum.[50]

All four diagnostic peaks and the pseudomolecular ion were used to make manual assignments of elemental composition. **Figure 3.7** illustrates the process by which assignments were made. For each diagnostic ion, all elemental compositions having the formula $C_xH_xO_xN_xS_xP_xF_xCl_xBr_xSi_x$ (x = 0-1000) within 1 ppm of the observed mass were determined and considered regardless of chemical possibility. Theoretical isotope distributions were calculated for each possible elemental composition (shown in color) at the resolution of data collection and plotted against the experimental isotopic distribution (shown in black). Each distribution was qualitatively compared to the experimental distribution and assigned a '–' or '+' if the distribution could or could not be ruled out, respectively, as potentially 'matching' the experimental data. The results of this comparison are given in **Table 3.4**. Including the [M – *t*-butyl]$^+$ and [M – methyl]$^+$ peaks in **Figure 3.7** allows all but the purple-based ([M – *t*-butyl]$^+$: $C_{11}H_{24}O_2NSSi$) and red-based ([M – *t*-butyl]$^+$: $C_{10}H_{24}O_3NSi_2$) elemental compositions to be eliminated. The pseudomolecular ion and [M + C_2H_5]$^+$ adduct ion provide the discrimination power to eliminate the purple-based elemental composition (not present due to mass accuracy constraints) using the second and third isotopic peaks. Removing the contribution of the two silylations (each composed of $C(CH_3)_3Si(CH_3)_2$ or $C_6H_{15}Si$) in the remaining red-based elemental composition ([M + H]$^+$: $C_{14}H_{34}O_3NSi_2$) and adding the two acidic protons originally displaced by silylation, we can identify this spectrum as arising from the elemental composition, $C_2H_5O_3N$, possibly hydroxyglycine, a hydroxylated amino acid intermediate in many biosynthetic pathways,[51] or a structural isomer. Several other expected metabolites of *Arabidopsis thaliana* were similarly identified; however, given the very high complexity of metabolite samples and the

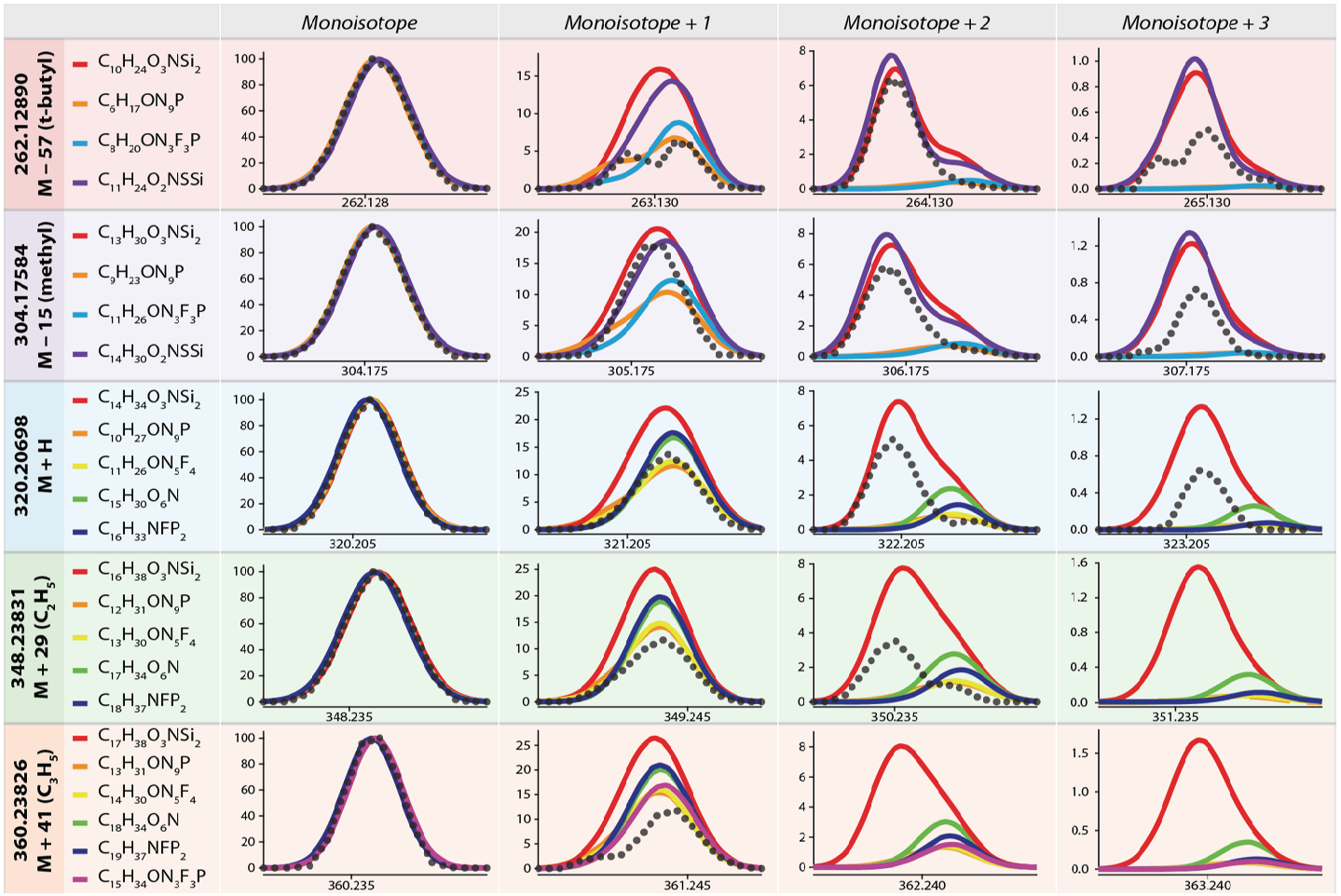


Figure 3.7. Illustrates the process of elemental composition assignment for *Arabidopsis thaliana* metabolites. All elemental compositions having the formula $C_{0\text{-}1000}H_{0\text{-}1000}O_{0\text{-}1000}N_{0\text{-}1000}S_{0\text{-}1000}P_{0\text{-}1000}F_{0\text{-}1000}Cl_{0\text{-}1000}Br_{0\text{-}1000}Si_{0\text{-}1000}$ within 1 ppm (colored traces) of the four diagnostic peaks and the pseudomolecular ion (dotted black trace) were compared by the first three isotopic peaks of their theoretical isotopic distributions. Compounds of the same series maintain their color assignment from row-to-row.

Table 3.4. Results of a qualitative comparison of experimental data with theoretical distributions of potential elemental compositions within ± 1 ppm from an *Arabidopsis thaliana* metabolite sample.

Monoisotopic Peak Mass	Expected Peak Relationships	Possible Elemental Compositions	Mass Error (ppm)	Mono[2]	Mono+1	Mono+2	Mono+3
262.12890	M - 57 (t-butyl)	$C_{10}H_{24}O_3NSi_2$	-0.088	+	+	+	+
		$C_6H_{17}ON_9P$	0.308	+	+	–	–
		$C_8H_{20}ON_3F_3P$	-0.610	+	–	–	–
		$C_{11}H_{24}O_2NSSi$	-0.964	+	+	+	+
304.17584	M - 15 (methyl)	$C_{13}H_{30}O_3NSi_2$	-0.109	+	+	+	+
		$C_9H_{23}ON_9P$	0.232	+	+	–	–
		$C_{11}H_{26}ON_3F_3P$	-0.559	+	+	–	–
		$C_{14}H_{30}O_2NSSi$	-0.864	+	+	+	+
320.20698	M + H	$C_{10}H_{27}ON_9P$	-0.280	+	+	–	–
		$C_{11}H_{26}ON_5F_4$	0.563	+	+	–	–
		$C_{14}H_{34}O_3NSi_2$	-0.604	+	+	+	+
		$C_{15}H_{30}O_6N$	0.674	+	+	–	–
		$C_{16}H_{33}NFP_2$	0.944	+	+	–	–
348.23831	M + 29 (C_2H_5)	$C_{12}H_{31}ON_9P$	-0.171	+	+	–	n/a[1]
		$C_{16}H_{38}O_3NSi_2$	-0.470	+	+	+	n/a
		$C_{13}H_{30}ON_5F_4$	0.603	+	+	–	n/a
		$C_{17}H_{34}O_6N$	0.706	+	+	–	n/a
		$C_{18}H_{37}NFP_2$	0.954	+	+	–	n/a
360.23826	M + 41 (C_3H_5)	$C_{13}H_{31}ON_9P$	-0.304	+	+	n/a	n/a
		$C_{14}H_{30}ON_5F_4$	0.444	+	+	n/a	n/a
		$C_{18}H_{34}O_6N$	0.543	+	+	n/a	n/a
		$C_{17}H_{38}O_3NSi_2$	-0.593	+	+	n/a	n/a
		$C_{19}H_{37}NFP_2$	0.783	+	+	n/a	n/a
		$C_{15}H_{34}ON_3F_3P$	-0.972	+	+	n/a	n/a

[1] Mass spectral data not present/below background noise.
[2] '+' indicates a matching distribution, '-' indicates a non-matching distribution

capability of this instrument to acquire tens of thousands of high quality spectra per chromatographic separation, there exists a clear need for informatics tools that take high-resolution, mass accuracy, and spectral accuracy into account for automated, high-throughput spectral analysis and elemental composition assignment.

Future developments. Two potential shortcomings of the current implementation of the GC-Orbitrap mass spectrometer merit mention. First, the inability to dynamically control analyte injection times is perhaps the most apparent weakness of the nested scan. In normal front source-centric operation of this instrument, automatic gain control (AGC)[52] standardizes each mass analysis to the same target number of charges by adjusting ion injection times to meet the target. Without this feature, under high ion flux conditions, the nested scan can be subject to overfilling of and/or space-charge distortions within the c-trap, particularly at the 60,000 and 100,000 resolution settings. Overfilling and space-charge effects would manifest in the mass spectrum as a loss of high mass ions and systemic loss of mass accuracy. However, at the resolutions with scan rates suitable for a GC-timescale (7,500-30,000), significant problems due to overfilling of the c-trap are uncommon. We tested this by tracking the relative intensities of EI fragment ions of high mass calibrant TPFHST as a function of analyte accumulation time (1-2000 ms) in the c-trap using a modified version of the nested scan. A constant resolution of 7,500 for Orbitrap mass analysis was used at all accumulation times to standardize the level of the thermal noise band. For this mass range, which encompasses the mass range queried in GC experiments, there was no discernible bias against high mass ions at long accumulation times (data not shown).

Similar experiments with FC-43, however, revealed that selected fragment ion intensities decreased with prolonged storage in the c-trap or QLT. Here, EI fragment ions of FC-43 were accumulated in the c-trap for 112 ms (7,500 resolution nested scan) and then held for 0-2000 ms in the c-trap prior to injection into the Orbitrap or transfer to the QLT for analysis. In both

analyzers, the abundance of fragment ion $C_4F_9^+$ (*m/z* 219) decreased as the abundance of an oxygen-containing species $C_3F_7CO^+$ (*m/z* 197) increased. Most other fragment ions were relatively unchanged (**Figure 3.8**; top panel). This phenomenon is characteristic of quadrupole ion traps and was described by Creaser, *et al.*[53] as the following reaction: $C_4F_9^+ + H_2O \rightarrow C_3F_7CO^+ + 2HF$. The background water vapor facilitating this reaction in this system is likely due to the atmospheric pressure source inlet and low purity nitrogen used in the c-trap. Corroborating this hypothesis, this reaction was noticeably slowed (**Figure 3.8**; bottom panel) when the c-trap was supplied and flushed overnight with ultra-high purity nitrogen. The reaction of *m/z* 219 to *m/z* 197 was more pronounced for ions analyzed in the Orbitrap than the QLT (data not shown). In addition to its susceptibility to water reactions, *m/z* 219 is considered a fragile ion.[54,55] The heightened decrease in *m/z* 219 abundance in the Orbitrap is possibly attributable to a breakdown of this ion during the rather "violent" steps of c-trap compression, injection, and electrodynamic squeezing within the Orbitrap. The effects of these and other processes result in different overall spectral appearances for GC-Orbitrap and single-quadrupole EI library spectra, as illustrated in the comparison of FC-43 spectra in **Figure 3.9**.

These two issues highlight potential future improvements to the GC-Orbitrap design and implementation. First, the development of a "predictive" algorithm for proportional adjustment of analyte accumulation times based on the ion source flux within the context of the nested scan type is needed to control ion populations and decrease ion loss. Second, the development of a closed system, lacking an atmospheric pressure inlet, with optimized ion optics is crucial to reduce background water vapor and improve spectral quality.

We have demonstrated the first-ever modification of a QLT-Orbitrap hybrid instrument for high-mass accuracy and high-resolution gas chromatography/mass spectrometry. With the development of a new scan sequence, the nested scan, we have minimized scan times and

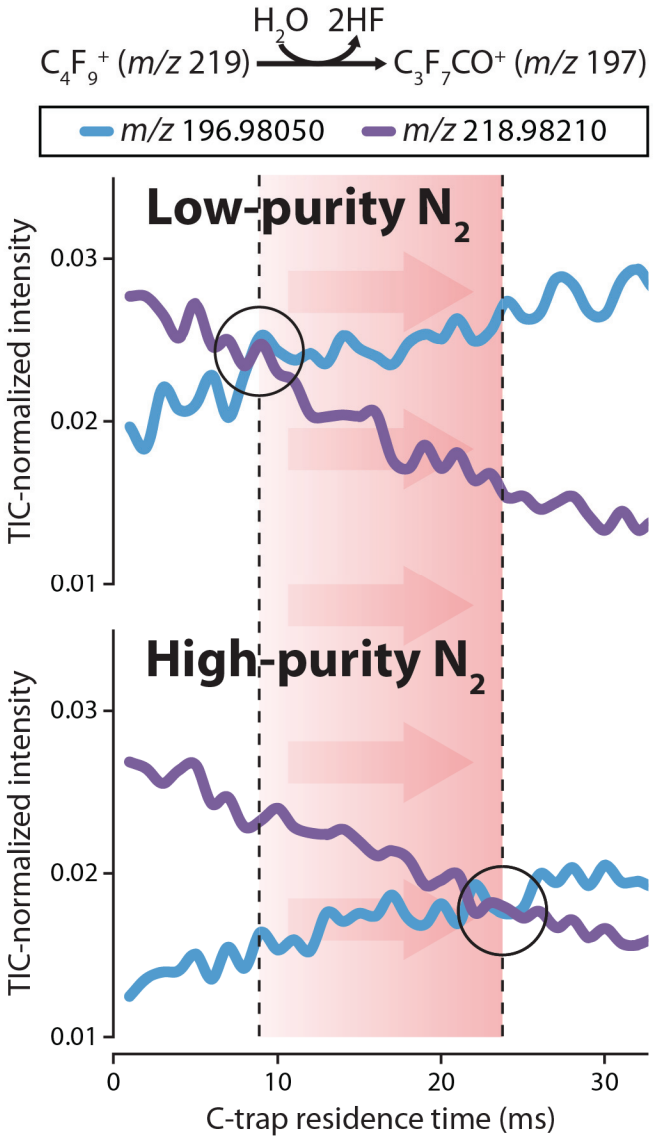


Figure 3.8. Effect of low-purity (top) versus high-purity (bottom) nitrogen in the reaction of FC-43 ion m/z 219 with background water vapor to form m/z 197 as a function of ion residence time in the c-trap.

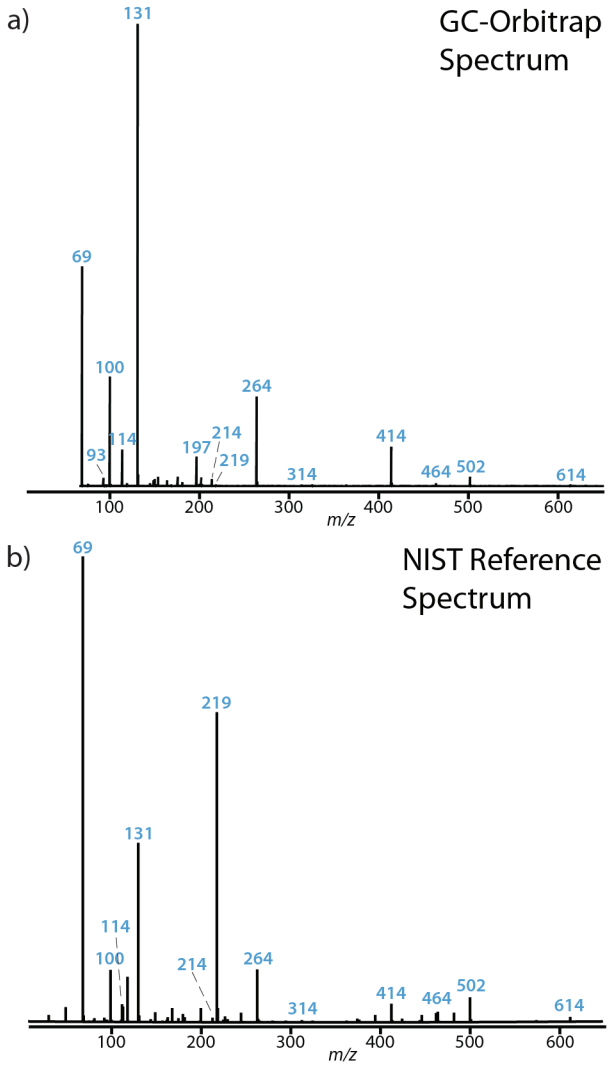


Figure 3.9. Comparison of spectral appearance of a (**A**) GC-Orbitrap spectrum of FC-43 and the (**B**) NIST library spectrum of FC-43 [from http://webbook.nist.gov/cgi/cbook.cgi?ID=C311897&Mask=200].

maximized duty cycle without a trade-off in sensitivity to achieve scan rates of 4 and 2.2 Hz at 15,000 and 30,000 resolution, respectively. This marks a significant improvement over conventional scan cycles and brings the Orbitrap acquisition rate in synch with the timescale of GC separations for the first time. At these scan rates, mass accuracy less than 1 ppm with internal calibration and resolution exceeding 100,000 are routinely achievable. In combination with isotopic distribution information, the GC-Orbitrap facilitates the unambiguous determination of small molecule elemental compositions, as demonstrated in our analysis of primary metabolites from *Arabidopsis thaliana*, as well as meets the mass spectral requirements of challenging EPA methods. We believe the GC-Orbitrap represents a significant step forward in the field of high-resolution, high-mass accuracy GC/MS.

REFERENCES

(1) Holmes, J. C.; Morrell, F. A. Oscillographic mass spectrometric monitoring of gas chromatography. *Appl. Spectrosc.* **1957**, *11*, 86-87.

(2) Gohlke, R. S.; Mclafferty, F. W. Early Gas-Chromatography Mass-Spectrometry. *J. Am. Soc. Mass Spectr.* **1993**, *4*, 367-371.

(3) Masucci, J. A.; Caldwell, G. W. In *Modern practice of gas chromatography*; 4th ed.; Grob, R. L., Barry, E. F., Eds.; John Wiley & Sons: Hoboken, NJ, 2004; pp 339-401.

(4) Kind, T.; Fiehn, O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *Bmc Bioinformatics* **2007**, *8*, -.

(5) Kind, T.; Fiehn, O. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. *Bmc Bioinformatics* **2006**, *7*, -.

(6) Beynon, J. H. Qualitative Analysis of Organic Compounds by Mass Spectrometry. *Nature* **1954**, *174*, 735-737.

(7) Kim, S.; Rodgers, R. P.; Marshall, A. G. Truly "exact" mass: Elemental composition can be determined uniquely from molecular mass measurement at similar to 0.1 mDa accuracy for molecules up to similar to 500 Da. *Int. J. Mass Spectrom.* **2006**, *251*, 260-265.

(8) Stoll, N.; Schmidt, E.; Thurow, K. Isotope pattern evaluation for the reduction of elemental compositions assigned to high-resolution mass spectral data from electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1692-1699.

(9) Spengler, B. De novo sequencing, peptide composition analysis, and composition-based sequencing: A new strategy employing accurate mass determination by Fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass Spectr.* **2004**, *15*, 703-714.

(10) EPA. *Guidelines establishing test procedures for the analysis of pollutants: EPA method 1613*; Federal Register, 1997; pp 48394-48442.

(11) Siuzdak, G. *The expanding role of mass spectrometry in biotechnology*; 2nd ed.; MCC Press: San Diego, CA, 2006.

(12) Thomas, R. A beginner's guide to ICP-MS - Part VII: Mass separation devices - Double-focusing magnetic-sector technology. *Spectrosc.* **2001**, *16*, 22-.

(13) Ledford, E. B.; White, R. L.; Ghaderi, S.; Wilkins, C. L.; Gross, M. L. Coupling of Capillary Gas-Chromatograph and Fourier-Transform Mass-Spectrometer. *Anal Chem* **1980**, *52*, 2450-2451.

(14) Wilkins, C. L.; Giss, G. N.; Brissey, G. M.; Steiner, S. Direct-Linked Gas Chromatography-Fourier Transform Infrared-Mass Spectrometer System. *Anal Chem* **1981**, *53*, 113-117.

(15) Nguyen, M. T.; Wronka, J.; Starry, S.; Ridge, D. P. An Ion-Cyclotron Resonance Spectrometer as a Gas-Chromatographic Detector - Measurement of Momentum-Transfer Collision Frequencies. *Int J Mass Spectrom* **1981**, *40*, 195-210.

(16) Sack, T. M.; Gross, M. L. Pulsed Valve Interface for Gas-Chromatography Fourier-Transform Mass-Spectrometry. *Anal Chem* **1983**, *55*, 2419-2421.

(17) Sack, T. M.; Mccrery, D. A.; Gross, M. L. Gas-Chromatography Multiphoton Ionization Fourier-Transform Mass-Spectrometry. *Anal Chem* **1985**, *57*, 1290-1295.

(18) Cody, R. B.; Kinsinger, J. A.; Ghaderi, S.; Amster, I. J.; McLafferty, F. W.; Brown, C. E. Developments in analytical fourier-transform mass spectrometry. *Anal. Chim. Acta* **1985**, *178*, 43-66.

(19) Larsen, B. S.; Wronka, J.; Ridge, D. P. An Ion-Cyclotron Resonance Spectrometer as a Gas-Chromatographic Detector - the Effect of Continuous Trapping on Performance. *Int. J. Mass Spectrom.* **1986**, *72*, 73-84.

(20) Heffner, C.; Silwal, I.; Peckenham, J. M.; Solouki, T. Emerging technologies for identification of disinfection byproducts: GC/FT-ICR MS characterization of solvent artifacts. *Environ. Sci. Technol.* **2007**, *41*, 5419-5425.

(21) Luo, Z. H.; Heffner, C.; Solouki, T. Multidimensional GC-Fourier Transform Ion Cyclotron Resonance MS Analyses: Utilizing Gas-Phase Basicities to Characterize Multicomponent Gasoline Samples. *J. Chromatogr. Sci.* **2009**, *47*, 75-82.

(22) Solouki, T.; Szulejko, J. E.; Bennett, J. B.; Graham, L. B. A preconcentrator coupled to a GC/FTMS: Advantages of self-chemical ionization, mass measurement accuracy, and high mass resolving power for GC applications. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1191-1200.

(23) Szulejko, J. E.; Luo, Z. H.; Solouki, T. Simultaneous determination of analyte concentrations, gas-phase basicities, and proton transfer kinetics using gas chromatography/Fourier transform ion cyclotron resonance mass spectrometry (GC/FT-ICR MS). *Int J Mass Spectrom* **2006**, *257*, 16-26.

(24) Szulejko, J. E.; Solouki, T. Potential analytical applications of interfacing a GC to an FT-ICR MS: Fingerprinting complex sample matrixes. *Anal Chem* **2002**, *74*, 3434-3442.

(25) Houtman, N. A breath of ... *Chem Brit* **2003**, *39*, 38-40.

(26) Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal Chem* **2000**, *72*, 1156-1162.

(27) Makarov, A.; Denisov, E.; Lange, O. Performance Evaluation of a High-field Orbitrap Mass Analyzer. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1391-1396.

(28) Olsen, J. V.; de Godoy, L. M. F.; Li, G. Q.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **2005**, *4*, 2010-2021.

(29) Erve, J. C. L.; Gu, M.; Wang, Y.; DeMaio, W.; Talaat, R. E. Spectral Accuracy of Molecular Ions in an LTQ/Orbitrap Mass Spectrometer and Implications for Elemental Composition Determination. *J. Am. Soc. Mass Spectr.* **2009**, *20*, 2058-2069.

(30) Makarov, A.; Denisov, E.; Kholomeev, A.; Baischun, W.; Lange, O.; Strupat, K.; Horning, S. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* **2006**, *78*, 2113-2120.

(31) Bateman, K. P.; Kellmann, M.; Muenster, H.; Papp, R.; Taylor, L. Quantitative-Qualitative Data Acquisition Using a Benchtop Orbitrap Mass Spectrometer. *J. Am. Soc. Mass Spectr.* **2009**, *20*, 1441-1450.

(32) Koulman, A.; Woffendin, G.; Narayana, V. K.; Welchman, H.; Crone, C.; Volmer, D. A. High-resolution extracted ion chromatography, a new tool for metabolomics and lipidomics using a second-generation Orbitrap mass spectrometer. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 1411-1418.

(33) McAlister, G. C.; Berggren, W. T.; Griep-Raming, J.; Horning, S.; Makarov, A.; Phanstiel, D.; Stafford, G.; Swaney, D. L.; Syka, J. E. P.; Zabrouskov, V.; Coon, J. J. A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer. *J. Proteome Res.* **2008**, *7*, 3127-3136.

(34) Strupat, K.; Kovtoun, V.; Bui, H.; Viner, R.; Stafford, G.; Horning, S. MALDI Produced Ions Inspected with a Linear Ion Trap-Orbitrap Hybrid Mass Analyzer. *J. Am. Soc. Mass Spectr.* **2009**, *20*, 1451-1463.

(35) Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9528-9533.

(36) Keller, B. O.; Sui, J.; Young, A. B.; Whittal, R. M. Interferences and contaminants encountered in modern mass spectrometry. *Anal. Chim. Acta* **2008**, *627*, 71-81.

(37) Fiehn, O. In *Arabidopsis Protocols*; 2nd ed.; Salinas, J., Sanchez-Serrano, J. J., Eds.; Humana Press: Totowa, NJ, 2006; Vol. 323; pp 439-447.

(38) Wenger, C. D.; McAlister, G. C.; Xia, Q.; Coon, J. J. Sub-part-per-million precursor and product mass accuracy for high-throughput proteomics on an ETD-enabled Orbitrap mass spectrometer. *Mol Cell Proteomics* **2010**.

(39) Erve, J. C.; Demaio, W.; Talaat, R. E. Rapid metabolite identification with sub parts-per-million mass accuracy from biological matrices by direct infusion nanoelectrospray ionization after clean-up on a ZipTip and LTQ/Orbitrap mass spectrometry. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 3015-3026.

(40) Hofstadler, S. A.; Bruce, J. E.; Rockwood, A. L.; Anderson, G. A.; Winger, B. E.; Smith, R. D. Isotopic Beat Patterns in Fourier-Transform Ion-Cyclotron Resonance Mass-Spectrometry - Implications for High-Resolution Mass Measurements of Large Biopolymers. *Int. J. Mass Spectrom.* **1994**, *132*, 109-127.

(41) Easterling, M. L.; Amster, I. J.; van Rooij, G. J.; Heeren, R. M. A. Isotope beating effects in the analysis of polymer distributions by Fourier transform mass spectrometry. *J. Am. Soc. Mass Spectr.* **1999**, *10*, 1074-1082.

(42) Gordon, E. F.; Muddiman, D. C. Impact of ion cloud densities on the measurement of relative ion abundances in Fourier transform ion cyclotron resonance mass spectrometry: experimental observations of coulombically induced cyclotron radius perturbations and ion cloud dephasing rates. *J. Mass Spectrom.* **2001**, *36*, 195-203.

(43) Bresson, J. A.; Anderson, G. A.; Bruce, J. E.; Smith, R. D. Improved isotopic abundance measurements for high resolution Fourier transform ion cyclotron resonance mass spectra via time-domain data extraction. *J. Am. Soc. Mass Spectr.* **1998**, *9*, 799-804.

(44) Gu, M.; Wang, Y. D.; Zhao, X. G.; Gu, Z. M. Accurate mass filtering of ion chromatograms for metabolite identification using a unit mass resolution liquid chromatography/mass spectrometry system. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 764-770.

(45) Santos, F. J.; Galceran, M. T. Modern developments in gas chromatography-mass spectrometry-based environmental analysis. *J. Chromatogr. A* **2003**, *1000*, 125-151.

(46) Fialkov, A. B.; Steiner, U.; Lehotay, S. J.; Amirav, A. Sensitivity and noise in GC-MS: Achieving low limits of detection for difficult analytes. *Int J Mass Spectrom* **2007**, *260*, 31-48.

(47) Sparkman, O. D. Mass Spectrometry PittCon® 2008. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, R1-R6.

(48) Makarov, A. In *58th ASMS Conference on Mass Spectrometry and Allied Topics* Salt Lake City, UT, 2010.

(49) van Bavel, B.; Abad, E. Long-Term Worldwide QA/QC of Dioxins and Dioxin-like PCBs in Environmental Samples. *Anal. Chem.* **2008**, *80*, 3956-3964.

(50) Mawhinney, T. P.; Robinett, R. S.; Atalay, A.; Madson, M. A. Analysis of amino acids as their tert.-butyldimethylsilyl derivatives by gas-liquid chromatography and mass spectrometry. *J. Chromatogr.* **1986**, *358*, 231-242.

(51) Seigler, D. S. *Plant secondary metabolsm*; Kluwer Academic Publishers: Norwell, MA, 1998.

(52) Syka, J. E. P.; Marto, J. A.; Bai, D. L.; Horning, S.; Senko, M. W.; Schwartz, J. C.; Ueberheide, B.; Garcia, B.; Busby, S.; Muratore, T.; Shabanowitz, J.; Hunt, D. F. Novel Linear Quadrupole Ion Trap/FT Mass Spectrometer: Performance Characterization and Use in the Comparative Analysis of Histone H3 Post-translational Modifications. *J. Proteome Res.* **2004**, *3*, 621-626.

(53) Creaser, C. S.; West, S. K.; Wilkins, J. P. Reactions of perfluorotri-n-butylamine fragment ions in the quadrupole ion trap: the origin of artefacts in the perfluorotri-n-butylamine calibration spectrum. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 538-540.

(54) Mosi, A. A.; Cullen, W. R.; Eigendorf, G. K. Ion/molecule reactions of perfluorotributylamine (FC43) cations with polycyclic aromatic hydrocarbons in a quadrupole ion trap. *Int. J. Mass Spectrom.* **1999**, *191*, 195-207.

(55) Vaden, T.; Ardhal, B.; Lynn, B. C. Evaluation of axial DC offsets during scanning of a quadrupole ion trap for sensitivity improvements. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 679-684.

CHAPTER 4

Development of a GC/quadrupole-Orbitrap Mass Spectrometer,

Part I: Design and Characterization

SUMMARY

Identification of unknown compounds is of critical importance in GC/MS applications (metabolomics, environmental toxin identification, sports doping, petroleomics, and biofuel analysis, among many others) and remains a technological challenge. Derivation of elemental composition is the first step to determining the identity of an unknown compound by MS, for which high accuracy mass and isotopomer distribution measurements are critical. While recent advances in TOF technology have begun to address the field's need for high mass accuracy measurements, the necessity of continued technological development remains. Here, we report on the development of a dedicated, applications-grade GC/MS employing an Orbitrap mass analyzer, the GC/quadrupole-Orbitrap. Built from the basis of the bench-top Orbitrap LC/MS, the GC/quadrupole-Orbitrap maintains the performance characteristics of the Orbitrap (up to 200,000 resolution with low-to-sub-ppm mass errors, scan/injection time pipelining, automatic gain control), enables quadrupole-based isolation for sensitive analyte detection in challenging matrices, and includes numerous analysis modalities to facilitate structural elucidation. We detail the design and construction of the instrument, discuss its key figures-of-merit, and demonstrate its performance for the characterization of unknowns.

INTRODUCTION

The success of small molecule analysis, especially metabolomics, by gas chromatography/mass spectrometry (GC/MS) ultimately hinges on the ability to identify, through elemental composition annotation and structural characterization, unknown peaks.[1] As a mature analytical technique, sample analysis workflows and post-possessing methods for GC/MS are well-established; analyses typically follow a reproducible, standardized path of spectral deconvolution and peak identification using retention time and spectral matching to existing, and immense, spectral reference libraries.[2,3] There is no shortage of software platforms to assist in these tasks[4-11]. Even so, in a single GC/MS analysis of a complex sample, often only ~100 of the 200-500 observed mass spectral features can routinely be identified;[1,12-14] the rest remain unknown, along with their potential importance to the research question being studied.

The first, and most critical, step to identifying an unknown is elemental composition derivation. For this, high mass accuracy mass measurements are key[15] (though not alone sufficient[16]). Fiehn has shown that a simple two step approach of 1) calculation of candidate compositions from the mass alone, and 2) elimination of false candidates by comparison of empirical and theoretical isotopomer distributions, is a highly effective tool to nearly unambiguously determine elemental compositions from high mass accuracy data.[16,17] Thus, as GC/MS has transitioned in recent years from the routine analysis of knowns, to the comprehensive analysis of complex biological and environmental samples containing multitudinous unknowns, instrumentation for GC/MS has likewise shifted from low resolution, unit mass accuracy detectors to state-of-the-art, time-of-flight (TOF) systems.[18,19]

GC/TOF systems are the highest resolution and mass accuracy GC/MS systems currently available commercially (excluding magnetic sector instruments, which are not suitable, at high resolution, for non-targeted analysis[20,21]). High-end TOF systems offer good mass accuracy (~10

ppm) and solid mass resolution (~1-7k), and are, in general, fast scanning (~20 Hz), an important trait for sampling narrow GC peaks. Two recently introduced systems further this development with tandem MS capabilities[22-24], and mass resolution up to 50,000 through a multi-pass flight path.[25-28] While these newer instruments have not yet been vetted in the peer-reviewed domain, TOF systems, like all mass analyzers, have both strengths and weaknesses. For example, TOFs possess unmatched acquisition rates (necessary for GCxGC[29,30]) and low isotopomer abundance error[16,31], an important filter for eliminating false elemental formulae. However, though capable of high repetition rates, TOFs suffer from signal saturation and a limited dynamic range that limits sensitivity.[32,33] Likewise, while capable of accurate mass, the precision of mass measurements can be poor[14] (60% RSD[24]) and the accuracy is dependent upon internal calibration. For example, in an inter-laboratory comparison by Bristow et al.,[34] 24% of participating TOF instruments had mean mass errors exceeding 10 ppm. By contrast, in this same study, FT-MS systems, such as the Orbitrap, maintained mass errors of ~1 ppm.

The now-widespread[35-40] Orbitrap FT-MS[41,42] perhaps stands poised to address some of the limitations of high resolution GC/TOFs. Introduced commercially into the LC/MS market in 2005, Orbitraps routinely measure mass with low-to-sub-ppm mass errors[43,44] (without or with internal calibration), high resolution (up to 200,000[42,45] and 600,000[46]), and moderate isotopomer abundance errors (3-10%).[47,48] Despite the arguably transformative impact the Orbitrap analyzer has had in proteomics and related fields, no Orbitrap-based system dedicated to GC/MS has been developed. Recently, we reported on a proof-of-principle modification of an electron transfer dissociation-enabled quadrupole linear ion trap (QLT)-Orbitrap[37] for GC/MS to assess the merits of the Orbitrap as a detector for GC.[49] While a rather crude implementation with several shortcomings, the performance of the GC/QLT-Orbitrap, with respect to mass accuracy, resolution, and selectivity in the m/z domain, suggested that a well-designed and optimized

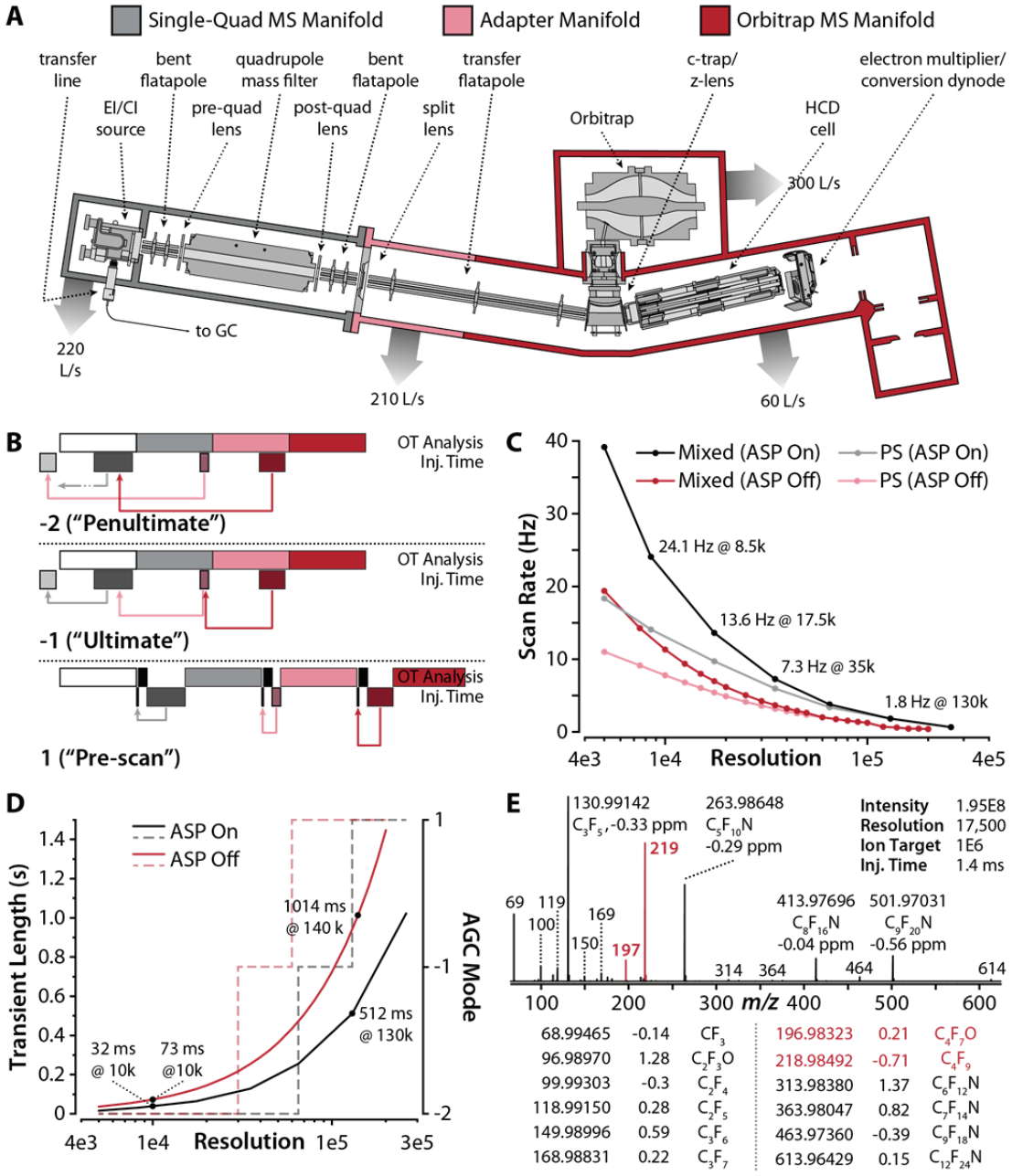
GC/Orbitrap instrument could be similarly transformative in GC/MS-based fields, especially in fields like metabolomics that rely on the unambiguous identification of unknowns.[1,17,50]

Here, building on the lessons learned in our previous work, we introduce the next generation of Orbitrap for gas chromatographic applications, the GC/quadrupole-Orbitrap. Using the bench-top Orbitrap LC/MS[35,38] as a base platform, we have developed an applications-grade, GC/MS that not only enables high mass accuracy and high resolution analysis, but does so at scan rates amenable to the time-scale of gas chromatographic separations. The GC/quadrupole-Orbitrap permits a broad cadre of enabling analysis modalities to address the needs of applications ranging from trace analyses, to the structural characterization of unknown metabolites in metabolomic analyses, including electron ionization, chemical ionization with any desired reagent gas, targeted and scheduled SIM, and targeted, scheduled, and data-dependent MS/MS. Additionally, we have developed an advanced data-dependent acquisition algorithm for MS/MS of alkylsilylated analytes, molecular ion-directed acquisition (MIDA), which is detailed in the accompanying article.[51] Herein, we discuss the design considerations for, and construction of, the instrument; detail optimization of the hardware, electronics, and firmware; and, benchmark instrument figures-of-merit whilst demonstrating performance in key applications.

EXPERIMENTAL PROCEDURES

Reagents. Unless otherwise specified, all reagents were purchased from Sigma-Aldrich (St. Louis, MO). Methanol and water (Optima LC/MS grade), and pyridine, methylene chloride, and iso-octane (GC/pesticide grade), were purchased from Fisher Scientific (Fair Lawn, NJ). All compressed gases (methane, helium, and nitrogen) were ultra-high purity grade and purchased from Airgas (Madison, WI).

Instrument construction. The GC/quadrupole-Orbitrap MS instrument consists of a single-quadrupole GC/MS (DSQ II; Thermo Fisher Scientific, Austin, TX) and a bench-top Orbitrap MS (Exactive; Thermo Fisher Scientific, Bremen, Germany), coupled together via a custom adapter manifold, as shown in **Figure 4.1A** in grey, dark red, and light red, respectively. Construction proceeded by removing the Orbitrap MS's atmospheric pressure (AP) inlet and associated ion optics (including, the source, ion transfer tube, tube lens, skimmer, 90° multipole, split lens, and transfer octapole), and sealing the AP inlet manifold openings with custom flanges. The two-stage source turbo-molecular pump was then removed and replaced with a smaller, turbo-molecular pump (60 L/s; Pfeiffer, Asslar, Germany) to reflect the decreased pumping requirements after removal of the AP inlet. Into this vacated space, the beam-type CAD collision cell (HCD cell, for MS/MS) and electron multiplier/conversion dynode (EM) were relocated from the opposite side of the Orbitrap MS's c-trap and from the single-quadrupole MS, respectively. Connections for the EM to the external electrometer and dynode power supply were made via a custom top-flange. On the opposite side of the Orbitrap MS's c-trap, the custom adapter manifold joined the single-quadrupole MS manifold to the manifold previously housing the HCD cell. The EI/CI combination source, curved pre-filter flatapole, pre-quadrupole lens, and quadrupole mass filter of the single-quadrupole MS remained as before. The gap between the quadrupole mass filter of the single-quadrupole MS and c-trap of the Orbitrap MS was bridged via new ion optic elements: a post-quadrupole lens and curved flatapole of identical design to the pre-quadrupole devices, custom split lens for ion gating, and long transfer flatapole. The adapter manifold was fitted with a turbo-molecular pump (210 L/s; Pfeiffer) to bring the number of isolated pumping stages from source-to-Orbitrap to four (including, the 220 L/s electron ionization/chemical ionization, EI/CI, source and 300 L/s Orbitrap turbo-molecular pumps). Electrical connections for the additional ion optics elements were made through feed-throughs



68.99465	-0.14	CF_3	196.98323	0.21	C_4F_7O
96.98970	1.28	C_2F_3O	218.98492	-0.71	C_4F_9
99.99303	-0.3	C_2F_4	313.98380	1.37	$C_6F_{12}N$
118.99150	0.28	C_2F_5	363.98047	0.82	$C_7F_{14}N$
149.98996	0.59	C_3F_6	463.97360	-0.39	$C_9F_{18}N$
168.98831	0.22	C_3F_7	613.96429	0.15	$C_{12}F_{24}N$

Figure 4.1. (A) Schematic of the GC/quadrupole-Orbitrap instrument. The DSQ II, adapter, and Exactive manifolds are colored in grey, light red, and dark red, respectively. Ion optic components and four pumping stages (as grey outward arrows) are labeled. **(B)** Scan cycle and AGC modes. Modes are dynamically selected by the instrument. The -2 and -1 modes permit ion accumulation during mass analysis, scaling the injection time based on two- or one-scan prior, respectively. Pre-scan mode acquires a dedicated scan between every analytical scan, delaying ion accumulation until after mass analysis. **(C)** Scan rate (Hz) as a function of mass analysis resolution, ASP, and scan mode. **(D)** Transient length (s) and AGC mode selection as a function of mass analysis resolution and ASP. **(E)** Single-scan positive EI FC-43 calibration spectrum. The average mass error with external calibration was 0.24 ppm (σ = 1.04, $error_{rms}$ = 1.06, n = 38).

installed in the adapter manifold. The HCD cell is plumbed directly with ultra-high purity nitrogen collision gas, which also provides for collisional cooling in the attached c-trap. The EI/CI source was interfaced to the GC via a heated transfer line. All component parts are labeled in **Figure 4.1A**.

Electronics and firmware modifications. Most original electronic boards and power supplies from both instruments were retained to drive new and original ion optic and vacuum components. EI/CI source elements (including filament, calibration gas inlet control, heaters, and lenses) and the EM retained single-quadrupole MS electronic control, except for EI/CI source lens 2. Control of source lens 2 was redirected to a spare DC output on the Orbitrap MS's ion optics supply board. The Orbitrap MS's electronics retained control over the same components, where present (e.g., c-trap, Orbitrap, HCD cell), as well as some single-quadrupole MS components, and all new ion optic devices. The Orbitrap MS's RF and DC outputs, originally controlling AP inlet devices, were repurposed to drive the bent flatapoles and pre-/post-quadrupole mass filter lenses. The split lens electronics were modified to provide a deflection potential of ±50 V, rather than ±350V, and drive the new split lens device. Two new circuit boards were added to the system, a quadrupole RF/DC board (Thermo Fisher Scientific), to control the quadrupole mass filter, and an additional ion optic DC/RF supply board (Thermo Fisher Scientific), to provide RF/DC control for the transfer flatapole.

The Orbitrap MS's instrument firmware (written in Python and based on Thermo Q Exactive firmware version 2.0 Build 146201) was adapted to permit simultaneous control of both component instruments via a single data system. Communication between the two data systems was established through a transmission control protocol (TCP) link. Code to control all dynamic devices (i.e., those that change potential during a scan cycle) was written to permit control directly via the Orbitrap MS's firmware and electronics to ensure fast response. Static devices

controlled by the single-quadrupole MS were programmatically controlled through commands sent to the single-quadrupole MS via the TCP link. Code was written to enable injection, automatic gain control, manual and automated signal optimization, ion optic and mass analyzer calibration, quadrupole isolation, and MS/MS capabilities.

Sample preparation and gas chromatography/mass spectrometry. All GC/MS experiments were performed using a Trace Ultra gas chromatograph (Thermo Electron, Milan, Italy) equipped with a CTC Analytics PAL autosampler (Zwingen, Switzerland). Compounds were separated on a 30 m x 0.25 mm (ID) x 0.25 μm (d_f) Crossbond 5% diphenyl/95% dimethyl polysiloxane column (Restek Rxi-5Sil MS, Bellefonte, PA) using helium as the carrier gas. The gas chromatograph was interfaced to the quadrupole-Orbitrap instrument (*vide supra*) via a heated transfer line. The instrument was characterized, and regularly tuned and calibrated with perfluorotributylamine (FC-43; Scientific Instrument Services, Ringoes, NJ), introduced into the transfer line by a calibration gas module. All MS experiments employed advance signal processing (ASP).

Octafluoronaphthalene. Solutions of 1 pg/μL octafluoronaphthalene (OFN) were prepared in iso-octane containing 0%, 1%, 2%, and 5% (v/v) diesel fuel, obtained from a local gas station. Aliquots (1 μL) were analyzed in triplicate. Samples were injected via splitless injection (1 min splitless, 50 mL/min split thereafter) via the hot-needle technique at an injector temperature of 220 °C, and separated at 1.2 mL/min helium using the following oven gradient: isothermal at 40 °C for 1 min, 30 °C/min to 165 °C, 120 °C/min to 275 °C, and isothermal at 275 °C for 5 min. The transfer line and source temperatures were 275 °C and 220 °C, respectively. The MS was configured for SIM of a 5 Th window around the OFN monoisotopic peak at m/z 272 (m/z 269.5 – 274.5) and a scan range of m/z 106-300. EI (-70 eV) was used. A

resolution of 70,000 (m/Δm), relative to m/z 200, was employed with a 250 ms maximum injection time and AGC (automatic gain control) target of 1E6 charges.

EPA 8270 pesticides analysis. A 25-50 µg/mL working solution in methylene chloride of 93 EPA 8270 semi-volatile organic pollutants was prepared by combination of the following standards, all obtained from Restek: SV Internal Standard Mix, Benzoic Acid Mix, Revised B/N Surrogate Mix, 8270 Benzidines Mix #2, 1,4-dioxane, 8270 MegaMix, and Acid Surrogate Mix (4/89 SOW). The working solution was serially diluted to the following concentrations (amounts on column in parentheses) in methylene chloride: 25-50 µg/mL (2.5-5 ng), 2.5-5.0 µg/µL (250-500 pg), 250-500 ng/mL (25-50 pg), 25-50 ng/mL (2.5-5.0 pg), 5-10 ng/mL (0.5-1.0 pg), 2.5-5.0 ng/mL (250-500 fg), 1.25-2.5 ng/mL (125-250 fg), 0.5-1.0 ng/mL (50-100 fg), 375-750 pg/mL (37.5-75.0 fg), 250-500 pg/mL (25-50 fg), 125-250 pg/mL (12.5-25 fg), 50-100 pg/mL (5-10 fg), and 25-50 pg/mL (2.5-5 fg). One microliter each was analyzed in triplicate using EI (-70 eV) at an injection split ratio of 10:1 (12 mL/min split flow for 1.2 mL/min helium column flow). The following oven program was used: 1 min isothermal at 80 °C, 25 °C/min to 280 °C, 5 °C/min to 320 °C, and 1 min isothermal at 320 °C. The injector, transfer line, and source temperatures were 270, 280, and 250 °C, respectively. Full scan analyses employed a scan range of m/z 50-450, resolution of 17,500, an AGC target of 1E6 charges, and a maximum injection time of 250 ms. Targeted SIM analyses targeted the monoisotopic peak, or most-abundant ion, of all 93 pesticides with 3 Th isolation windows based on a scheduled inclusion list (**Appendix 7 Tables 7.1A-B**). All other parameters were the same as the full-scan analyses except 100 ms maximum injection times were used.

Structural characterization of an unknown fatty acid methyl ester (FAME). Bacterial fatty acids were methyl esterified using sodium methoxide in anhydrous methanol at RT for 2 h. The reaction was quenched with 2 N HCl and FAMEs were extracted with hexane. For hydrogenation

experiments, extracted FAMEs were dried under nitrogen, and subsequently hydrogenated in chloroform:methanol (2:1 v/v) with 5% Pt on charcoal as detailed by Montanari, *et al*.[52] A bacterial acid methyl ester (BAME) mix standard, containing 26 FAMEs in methyl caproate, was used for chromatographic and MS source optimization. Samples in hexane (1 μL) were injected via the hot-needle technique at various split ratios depending on sample concentration, with an injector temperature of 250 °C, helium flow rate of 1 mL/min, and the following oven program: 1 min isothermal at 150 °C, 15 °C/min to 250 °C, 1 min isothermal at 250 °C, 80 °C/min to 320 °C, and 2 min isothermal at 320 °C. The transfer line and source temperatures were 280 °C and 250 °C, respectively. Samples were ionized via EI or positive CI (PCI) using ACN as the reagent gas (-70 eV). Full-scan analyses employed a scan range of 75-400 Th, resolution of 17,500, AGC target of 1E6 charges, and maximum injection time of 100 ms. Targeted MS/MS analyses employed a 5 Th isolation width, normalized collision energy of 25 eV, resolution of 17,500, AGC target of 1E6 charges, and maximum injection time of 250 ms.

For ACN PCI, 250 μm (I.D.) fused silica capillary connected an ACN reservoir (6 mL) directly to the source of the MS through the heated transfer line. A two-holed ferrule was used to permit entry of both the GC column and ACN capillary into the transfer line. While the column extended into the source, the ACN capillary was set back approximately 5 cm from the source to prevent interference with the GC eluent. A medium-flow metering valve (Swagelok, Solon, OH) was placed between the reservoir and transfer line to regulate the flow of ACN into the source. A source pressure of 7.1E-5 Torr, ~0.2 ms reagent injection time (at a 1E6 AGC target), and a m/z 42 (protonated ACN)-to-m/z 54 (1-methyleneimino-1-ethenylium, or MIE) ratio of 5:1 were found to be optimal for generation of molecular ion MIE-adducts of unsaturated FAMEs.

Analysis of Arabidopsis thaliana *polar extracts.* Polar metabolite extracts of *A. thaliana* were prepared and analyzed as detailed in the accompanying article.[51] Briefly, wild-type *A.*

thaliana were grown in liquid culture in stable isotope-enriched media ($^{12}C^{14}N$, $^{12}C^{15}N$, $^{13}C^{14}N$, and $^{13}C^{15}N$). Following 12 d growth, plants were harvested, flash frozen, and macerated to a fine powder. Aliquots of ~250-350 mg power were extracted as reported by Fiehn,[53] and polar fractions subjected to methoxyamination, and silylation with either N-(*t*-butyldimethylsilyl)-N-methyltrifluoroacetamide (MTBSTFA) or N-Methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) (Thermo Scientific, Bellafonte, PA). Samples were ionized with either EI or methane PCI using full scan MS or molecular-ion directed acquisition (MIDA)-MS/MS.

Data analysis. Data were manually queried within Xcalibur Qual Browser 2.3.23 (Thermo Fisher Scientific) or processed automatically through processing methods within Xcalibur Quan Browser (version 2.3.23, Thermo Fisher Scientific). Unless otherwise indicated, data were extracted with 5-10 ppm mass error tolerances and peak areas (area-under-the-curve, AUC) determined via the ICIS peak detection algorithm. Regression and statistical analyses were performed in Origin (version 8.5.1 SR2 build 315, Origin Lab, Northampton, MA). Isotopomer abundance or ratio errors were calculated as the percent difference from the theoretical isotopomer abundance or ratio errors expected for a given elemental formula. For EPA 8270 pesticide analysis, the ratio of first isotopomer to the sum of the monoisotopomer and first isotopomer was used. For Arabidopsis thaliana metabolomic analysis, the percent errors of the abundances of the first, second, and third isotopomers relative to the monoisotopomer abundance were used, conforming to the standard used by the Seven Golden Rules Excel macro.[17,51]

RESULTS AND DISCUSSION

Design considerations. We chose the bench-top Orbitrap ("Exactive") LC/MS as the base platform for our GC/quadrupole-Orbitrap (GC/q-Orbitrap) instrument for its potential to address the shortcomings identified in our earlier work:[49] namely, 1) the presence of an

atmospheric pressure inlet and high levels of ion-molecule reactions, 2) insufficient scan speed (maximum 6.5 Hz at 7,500 resolution), and 3) the absence of automatic gain control. In addition to the bench-top Orbitrap's design simply being amenable to removal of its AP inlet to eliminate excess sources of water, numerous technological advances implemented in recent Orbitrap family instruments also address these shortcomings.

First, a shorter, more efficient ion transmission path, compared to the quadrupole linear ion trap (QLT)-Orbitrap implementation, decreases the duration of ion injection and, by extension, the overall duration of ion storage prior to mass analysis. As the duration of ion storage is directly related to the extent of ion-molecular reactions that occur,[54] this aspect also reduces the potential effects of water remaining in the system (present in bath or carrier gases, adsorbed on the manifold, etc.) by reducing injection times. Second, the substantial redesign of the bench-top Orbitrap's firmware into a modern, object-oriented codebase (using the Python language) permits dramatically greater scan rates by multiplexing (or, pipelining) ion injection and mass analysis times.[35,38] Unlike with the nested scan cycle developed in our earlier work[49] to multiplex ion injection and mass analysis, the new design's ability to "keep time" during mass analysis allows for the scheduling of ion injection just prior to end of mass analysis, with injection times determined through automatic gain control calculations that rely on predictive measures.[38] For calculated injection times up to the length of the Orbitrap transient, the parallel accumulation of ions and mass analysis means that ion injection has no effect on instrument duty cycle. Additionally, scheduling ion injection just prior to the end of mass analysis again minimizes overall duration of ion storage and thus, extent of water reactions. Third, modifications to the Orbitrap analyzer, preamplifier detection circuit, and Orbitrap central electrode voltage result in higher resolution at same transient length, or shorter mass analysis times at the same resolution. The latter modification, an increased central electrode voltage (from 3.5 kV to 5 kV), itself yields

20% greater resolution at the same transient length.[38] The former modifications enable the implementation of advanced signal processing (ASP) techniques, which utilize the phase information present in the real component of the FT of a time-domain signal, in combination with the magnitude spectrum, to increase resolution 1.8-2*x* at the same transient length[55,56]. Taken together, these advancements yield scan rates for the bench-top Orbitrap[35] and newer sister-instrument, the bench-top quadrupole-Orbitrap,[38] ~3-10x greater than were possible in our GC/QLT-Orbitrap implementation.

Construction. GC/q-Orbitrap construction is detailed in the **Experimental Procedures** section, but will be described in brief here. The GC/q-Orbitrap MS (**Figure 4.1A**) was constructed from two component instruments, a bench-top Orbitrap LC/MS (dark red), mentioned above, and a single-quadrupole GC/MS (**grey**). The component instruments were joined together via a custom adapter manifold (light red). The single-quadrupole MS provided the hardware, electronics, and firmware to interface to a gas chromatograph, to control the various components of the combination electron ionization (EI)/chemical ionization (CI) source (except source lens 2, *vide infra*), to regulate the introduction of calibration and reagent gases, and to permit the exchange of EI or CI ion volumes through the single-quadrupole manifold's probe interlock (not illustrated in the figure). The transfer ion optics, lenses, and mass resolving quadrupole native to the single-quadrupole MS were retained, albeit under control of the Orbitrap instrument's electronics and firmware, to provide transmission of ions from the source and isolation capabilities. The adapter manifold was coupled at the distal end of the single-quadrupole's manifold, connecting the EI/CI source to the c-trap of the Orbitrap instrument via additional transfer flatapoles and ion lenses. The displaced electron multiplier/conversion dynode (EM), of the single-quadrupole MS, and HCD cell, of the Orbitrap instrument, were then relocated on the opposite side of the c-trap to provide the possibility of an additional mass analysis modality, and

MS/MS capabilities, respectively. Relocation of these devices to the opposite side of the c-trap was made possible by removal of all ion optics associated with the atmospheric pressure inlet of the Orbitrap instrument and closure of the inlet with custom flanges. **Figure 4.1A** labels each component part, as well as indicates four pumping stages, two of which were added to service the new adapter region (210 L/s) and the region housing the HCD cell (60 L/s). All devices from the bent pre-quadrupole flatapole onward to the HCD cell, including lens 2 of the source, are controlled directly by the Orbitrap instrument's electronics and firmware. All other devices, including the EM, are controlled by single-quadrupole's electronics and firmware, and are set through commands sent from the Orbitrap instrument to the single-quadrupole MS via a TCP link.

Optimization of ion gating. Ions were initially gated within the EI/CI source using lens 2, as is done in the unmodified single-quadrupole instrument. The communication lag between the single-quadrupole and Orbitrap MS (inherent to the TCP link) prevented use of the single-quadrupole's electronics to control any fast-switching devices, such as source lens 2, where voltages rapidly change during the course of a scan. Thus, to ensure precise ion gating, source lens 2 was rewired directly to the Orbitrap instrument's pre-existing split-gate electronics. While the split-gate electronics provided a precise ion gating waveform as measured by oscilloscope (with mean slew rates of 2.75 and 0.92 V/μs on the rising and falling edge of the gating wavefunction, respectively, and only 10 μs deviation between set and actual ion injection times), systematic investigation of the actual ion flux obtained per millisecond of ion injection indicated substantial non-linearity in the flux obtained for short injection periods. That is, a 1 ms injection duration repeated 10 times yielded less ion signal than a single 10 ms injection. Specifically, for injection times of 100 μs, only 60-74% of the expected ion flux was observed at the detector (**Figure 4.2**).

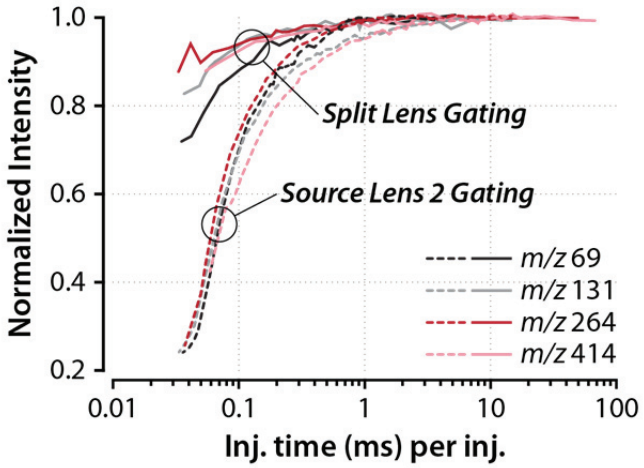


Figure 4.2. Injection time linearity for two different gating devices, source lens 2 and split lens. The normalized intensity for four FC-43 masses is plotted as a function of the injection time per injection. All data is the product of the same cumulative population of ions (i.e., intensity data at 10 ms per injection was the result of 10 x 10 ms injections, while intensity data at 0.1 ms per injection was the product of 1000 x 0.1 ms injections). This plot shows that the nonlinearities in the ion populations transmitted when injection times are short are partially mitigated by switching to the split lens device.

To localize this phenomenon to the ion gating process, the effects of secondary ion gating and of trapping of ions within the c-trap were first ruled-out. Using an oscilloscope, secondary ion gates (i.e., those that "close" after the primary gate, source lens 2, closes) were found to switch potential and settle long enough before/after the primary gate switched to rule out any interaction of the secondary gates in the observed nonlinearities (3.0-3.8 ms and 2.2-5.0 ms on the leading and falling edge of the primary gate pulse, respectively). To study the effect of c-trap trapping, we engaged the EM installed at the far end of the HCD cell. Whether ions were passed directly to EM without trapping in the c-trap, or trapped in the c-trap and then passed to the EM for detection, the non-linearity observed at short injection times during normal operation persisted. These results suggested that the method of ion gating was the likely culprit, with the hypothesis being that ions remaining in the rather lengthy gating region (~10 mm) of source lens 2 were being ejected, or otherwise "lost," when the gate closed. As injection times decrease, the proportion of ions residing in the gating region at any given time, relative to the total transmitted population, increases. Thus, loss of the ions in the gating region when the gate closes has a disproportional effect on measured signal if ion injection times are short, leading to the observed non-linearity in ion flux at short injection times.

To help resolve this issue, a custom split lens device with a gating region of ~1.5 mm was fabricated and installed between the post-quadrupole bent flatapole and the transfer flatapole, as shown in **Figure 4.1A**. The split lens consists of two opposing semi-circular plate lenses, one of which is held at a suitable voltage for ion transmission, while the other pulses between the same transmission voltage (during ion injection) and an attractive deflection voltage (to end ion injection by deflecting the ion beam). With this configuration, non-linearity in the actual ion flux obtained at short injection times, relative to longer injection periods, decreased, albeit mass dependently, such that the percent of expected transmission actually observed was ~90% or

greater for ion injection times above 100 μs (**Figure 4.2**). In other words, the use of the split lens improved ion transmission at short injection times (100 μs) between 26.4-48.4% for ions ranging between 69-414 Th, relative to ion gating with source lens 2.

A side-benefit of the previous analysis was the observation that most of the ion optic elements in the instrument transmitted ions in a mass-dependent manner, as was observed with the split lens above. These devices included the pre- and post-quadrupole bent flatapole DC offsets, the pre- and post-quadrupole lenses, the quadrupole mass filter DC offset, the split lens, and the transfer flatapole DC offset. As a result, the firmware was modified to set ion optic voltages in a mass-dependent manner based on the first mass, in full-scan analysis, or the mean mass, in SIM or MS/MS modes employing quadrupole isolation. Calibration procedures were written to build calibration curves for each mass-dependent devices by ramping the DC offset in question over a range of values for a wide-range of ions in the calibration spectrum. Typical DC and RF set points for positive EI ion transmission are shown for each device in **Figure 4.3**.

Duty cycle optimization and reduction of ion-molecular reactions. While the main benefits (e.g., high scan speed and automatic gain control (AGC)) of the scan cycle-related advances implemented on the bench-top Orbitrap MS were realized without additional modifications, the instrument scan cycle did require optimization to account for the increased reactivity of ions produced by electron-based methods, compared to ions generated by electrospray ionization in LC/MS. To motivate the discussion of the optimization process, the scan modes available on the GC/q-Orbitrap need be explained. Four scan modes are possible; three are depicted in **Figure 4.1B**, while the fourth, not shown, is fixed injection time scanning (with AGC off). The dynamic scanning modes include: the standard “pre-scan” AGC mode (1) where a dedicated, low resolution and short injection time pre-scan is acquired to determine the

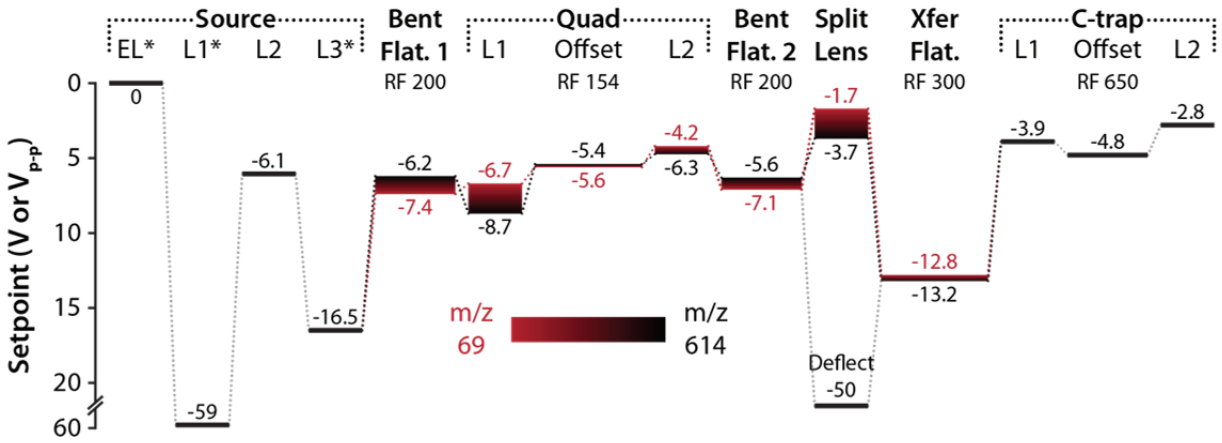


Figure 4.3. Typical DC and RF ion optics settings. Instrument software and electronics were modified to permit control of all devices via a single data system. Communication to DSQ controlled devices (*), including filament regulation (not shown), is achieved via a TCP link between the two data systems. Dynamic components are directly controlled by Exactive electronics. Typical DC and RF voltages for positive mode are shown here. Mass dependent device setpoints are shown as a range for perfluorotributylamine (FC-43) calibrant ions from m/z 69 (red) to 614 (black). The split lens pulses to -50 V for ion gating. All RF devices are cleared between scans.

appropriate injection time for the current analytical scan, and two predictive AGC scan modes utilizing the intensity observed in the ultimate (-1) or penultimate (-2) analytical scan to scale the injection time for the current analytical scan. As illustrated in **Figure 4.1B** (not to scale), the predictive AGC scan modes make use of the aforementioned injection time/mass analysis pipelining to result in increased scan rates relative to the "pre-scan"-only scan mode. In practice, both duty cycle and spectral quality are maximized by allowing the instrument to dynamically select the appropriate scan mode based on a set of selection rules **(Figure 4.1C-D)**. The selection rules depend on the current speed of acquisition (as influenced by transient length/resolution, ion target, and ion flux). At high acquisition rates, the penultimate mode is suitable since the spectrum used for injection time prediction was acquired relatively recently, i.e. still relevant to the current analysis. At lower acquisition rates, the ultimate or even pre-scan mode might be the most suitable depending upon the relevance of the ultimate scan to the current analysis.

These predictive AGC modes (-1 and -2) were developed and optimized for LC-based applications and have been successfully employed in the proteomics literature.[38,40,57] When implemented on the GC/q-Orbitrap, however, it was immediately apparent from spectra of the common GC/MS calibration compound, perfluorotributylamine (FC-43), that ions were being stored longer than expected based on their calculated injection times in the predictive scan modes. Specifically, the relative abundances of two ions within the FC-43 spectrum, m/z 219 ($[C_4F_9]^+$) and its water reaction product (less 2HF), m/z 197 ($[C_4F_7O]^+$), were found to be drastically different between pre-scan (high m/z 219:197 ratio) and predictive AGC modes (low m/z 219:197 ratio) (**Figure 4.1E**, in red). Together, these two ions serve as a hydrometer for the measurement of the level of background water vapor present in the system[49,54] (**Figure 4.4**). However, since the concentration of water vapor in the instrument, at any given time, can be considered to be constant, the m/z 219:197 ratio change between scan modes can only suggest that the ions have

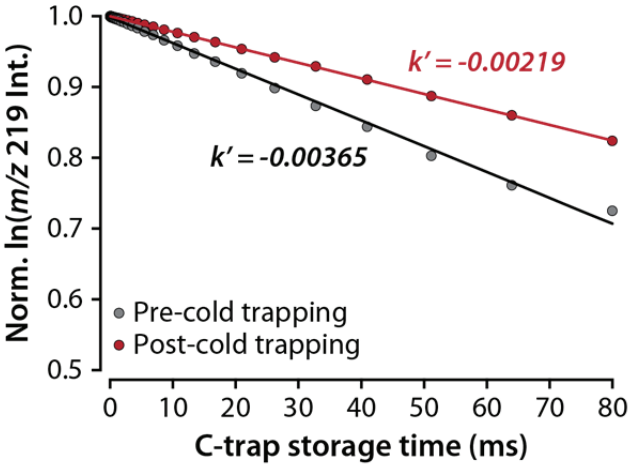


Figure 4.4. Water reaction kinetics for m/z 219 before and after cold-trapping the c-trap nitrogen supply line with liquid nitrogen to reduce the concentration of water vapor in the instrument. Since the concentration of water at any given time under the two conditions can be considered constant, the reaction is pseudo-first order with rate constants, $k' = k[H_2O]$, shown in the plot. The rate constant increases after reduction of water from -0.00354 to -0.00219.

longer to react with background water in the predictive modes, i.e. that the ions are being stored. Careful accountancy of the duration of events in the predictive mode scan cycles revealed that ion injection was occurring, during pipelining, an additional 28-31 ms prior to end of the mass analysis; thus, ions were injected for the AGC-determined period and then stored in the c-trap for ~30 ms prior to injection into the Orbitrap. This additional time was added intentionally in the design of the predictive AGC modes for LC-based instruments to ensure that all devices had completely settled prior to injection of the stored ion packet into the Orbitrap. For the already hydrated, often peptidic, analytes germane to LC/ESI-MS, this excess storage in the c-trap has no effect on the ion population. Despite paying a slight penalty in duty cycle (~4.3 ms added per scan), removal of this additional stabilization time was critical to reducing ion/molecule reactions resulting from trapping in the c-trap and obtaining high quality spectra.

Scan rate and resolution. As shown in **Figures 4.1C-D**, the penalty incurred by eliminating the stabilization period did not preclude achievement of the scan rates expected with the bench-top Orbitrap MS. Scan rates of 13.6 Hz at 17,500 and 7.3 Hz at resolution 35,000 are possible using a dynamically selected scan mode ("mixed") and ASP. Additionally, use of ASP permits the use of pipelined scans up to resolution 130,000 (512 ms transient) compared to only resolution 60,000 (400 ms transient) without ASP. If one considers a scan rate of 4 Hz to be the absolute minimum for practical application of this technology on GC-time scale, the GC/q-Orbitrap is amenable to analyses requiring resolution up to 100,000 (4 Hz with "mixed" scan mode and ASP). Compared to the previous GC/QLT-Orbitrap implementation, this amounts to a 5.7*x* increase in resolution available for analysis if at least a 4 Hz scan rate is required.

The acquisition rates of the GC/q-Orbitrap are comparable to those demonstrated in commercial literature for recently introduced, high-resolution TOF and Q-TOF instrumentation from LECO[25-28] and Agilent[22-24], respectively. In the LECO application notes, acquisition rates of

2-3[26,28], 6[25], and 20[27] spectra/second were demonstrated for mass analysis resolutions between 25-50,000 (relative to m/z 219). Similarly, for the Q-TOF instrument, acquisition rates averaged ~6 Hz in both MS and MS/MS modes at mass analysis resolutions between ~13-15,000 (relative to m/z 272).[24] To our knowledge, no reports utilizing these instruments have been published within the peer-reviewed domain.

Spectral figures-of-merit and quality. In addition to serving as a hydrometer, the FC-43 calibration spectrum provides insight into several other figures-of-merit governing overall spectral quality, as well as the basis for all calibration procedures (including, calibration of ion transmission, mass accuracy, isotopomer distribution accuracy/analyzer injection parameters, ASP, quadrupole isolation, and analyzer mass accuracy dependencies on c-trap RF amplitude, resolution/transient length, and ion space charge) First, the single-scan FC-43 spectrum in **Figure 4.1E** highlights the instrument's typical mass accuracy and precision, demonstrating externally calibrated mean mass errors of 0.24±1.04 ppm for 38 routinely-observed ions ($error_{rms}$ = 1.06 ppm). Second, routine 1-2 ms ion accumulation times at an ion target of 1E6 (i.e., for the injection of 1E6 charges) indicates overall efficiency in the generation and transmission of ions. Third, a full distribution of ions spanning the mass range from m/z 69-614 with significant representation from fragile or reactive ions (m/z 219, *vide supra*, m/z 502, and m/z 614), attests to the "gentleness" of the injection path, the absence of unexpected mass discrimination issues, and, as mentioned previously, low-to-expected levels of background water vapor (a 3:1 ratio of m/z 219:197 is typical for GC-trapping instruments).

While these specifications are tailored to the FC-43 calibration spectrum, high quality FC-43 calibration spectra also imply that spectra of compounds of analytical interest should share in this quality. That is, spectra collected on an instrument that is calibrated to minimize ion/molecule reactions, mass discrimination, and the harshness of ion transmission should

similarly demonstrate little evidence of those adverse effects, to the benefit of spectral quality. For GC/MS, however, one of the primary methods of chromatographic feature annotation is to match experimental spectra against large reference databases of EI spectra[3]. As such, spectra quality is, in practice, dictated by the extent of the match between experimental spectra and the reference spectra. In **Figure 4.5**, experimental single-scan spectra of decafluorobiphenyl (**panel A**) and hexachloroethane (**panel B**) are juxtaposed with unit-resolution reference spectra from the NIST database[3] (in red). Both experimental spectra faithfully reproduce the expected ions and relative abundances present in the reference spectra with no obvious additional or missing ions. The slight left-shift of the experimental spectrum relative to the reference spectrum is due to the negative mass defect of halogenated compounds; in the library spectrum, negative, or positive, mass defects are lost due to peak centroiding to the nearest nominal mass, and mass analysis with unit-mass accuracy analyzers. The GC/q-Orbitrap spectra, on the other hand, possess a maximal mass error of 1.46 ppm, allowing knowledge of the mass defect to be used in conjunction with accurate mass to facilitate spectral annotation. For comparison with **Chapter 3**, **Figure 4.6** compares experimental and NIST reference spectra for FC-43.

While, upon manual inspection, the spectra generated by the GC/q-Orbitrap generally match their associated reference spectrum very well (as in **Figure 4.5**), incompatibility between the collected Orbitrap data and the NIST matching algorithm, or the library search function within Xcalibur Qual Browser, has, as yet, prevented successful automation of spectral matching. For a reason that remains unresolved, during the library searching process, the most abundance ions are deleted from or shifted to different *m/z*'s within the Orbitrap spectrum prior to match score calculations against candidate reference spectra. This error thus prevents the matching process from yielding any useable results. Future development of custom spectral matching software tailored to these data should resolve this issue.

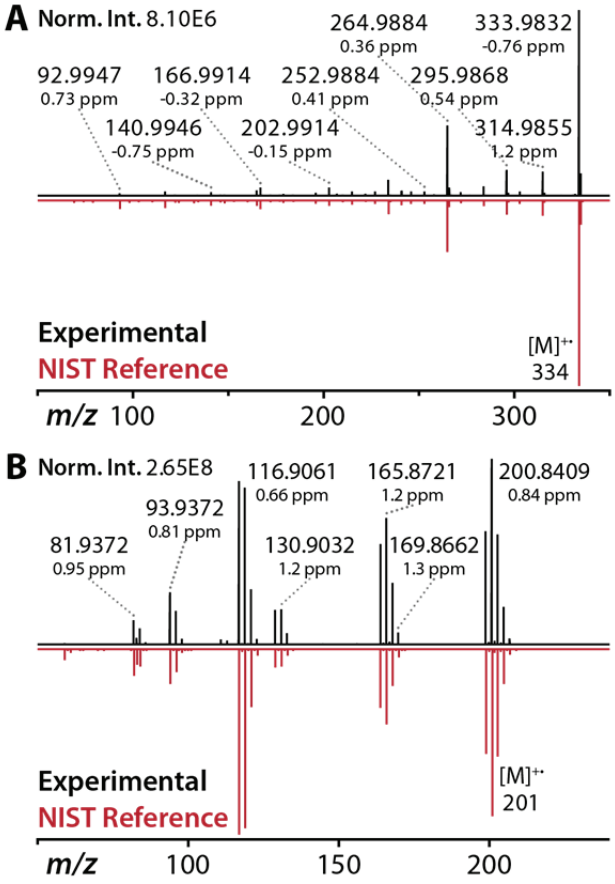


Figure 4.5. Spectral quality. The GC/q-Orbitrap spectra (top, black) for (A) decafluorobiphenyl and (B) hexachloroethane correlate well with NIST reference spectra (bottom, red). Orbitrap spectra were acquired at 17.5k resolution with a 1E6 ion target.

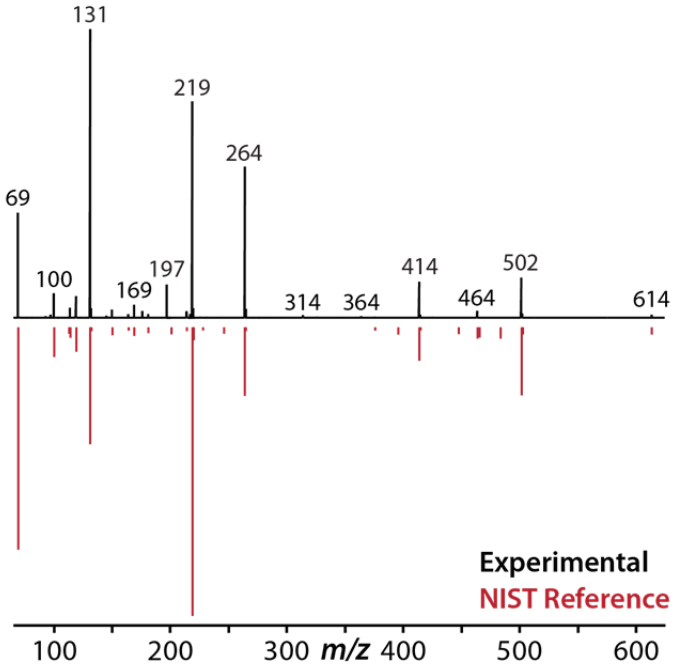


Figure 4.6. Comparison of the GC/q-Orbitrap (black) and NIST (red) FC-43 calibrant spectra. The NIST spectrum was collected by a single-quadrupole instrument in which water reactions (and thus the production of m/z 197) are negligible.

Quadrupole isolation. A quadrupole mass filter was included in GC/q-Orbitrap design to boost instrument sensitivity in targeted and trace analyses. The quadrupole enables experiments such as targeted SIM, targeted MS/MS, or data-dependent MS/MS. While both SIM and MS/MS improve sensitivity for target analytes under high background conditions by enhancing analyte signal-to-noise, the latter also facilitates target identification (targeted MS/MS) and structural elucidation (targeted or untargeted MS/MS). For the purpose of enhancing S/N in trace analyses, two competing processes must be reconciled. First, in order to accumulate ion populations greater than the detection threshold of the analyzer, the target analyte must be efficiently transmitted through the quadrupole, with wider isolation widths usually yielding higher transmission efficiency. Second, in order to successfully exclude matrix background and accumulate target ion populations of sufficient magnitude in the allotted time, tighter isolation windows are indicated. In practice, a compromise must be reached between the two processes based on the quadrupole device's efficiency of target ion transmission. Transmission curves for the GC/q-Orbitrap are plotted in **Figure 4.7A** for four stable ions from FC-43. Here, the percent transmission relative to RF-only quadrupole transmission (q = 0.706) of each ion was measured as a function of isolation width (from 50 Th, to the minimum width for the device, 0.4 Th). A low ion target of 1E5 charges was employed to prevent space-charge effects. At unit resolution (1 Th isolation width, vertical dotted line), isolation transmission efficiency is between 35-75% relative to RF only. While even higher transmission at smaller isolation widths is always desired, given the specifications of this quadrupole, transmission was determined to be acceptable. Indeed, in an analysis of 94 EPA 8270[58]-regulated semi-volatile organic compounds (5 ng on column and an ion target of 1E6), transmission was found not fall off at the rate expected in **Figure 4.7A**. For example, for the isolation of 4-chlorodiphenyl ether's pseudomolecular ion ($[M + H]^+$, 205.0415 Th, methane PCI) (**Figure 4.7B**), a 40% loss in transmission was expected as the isolation width

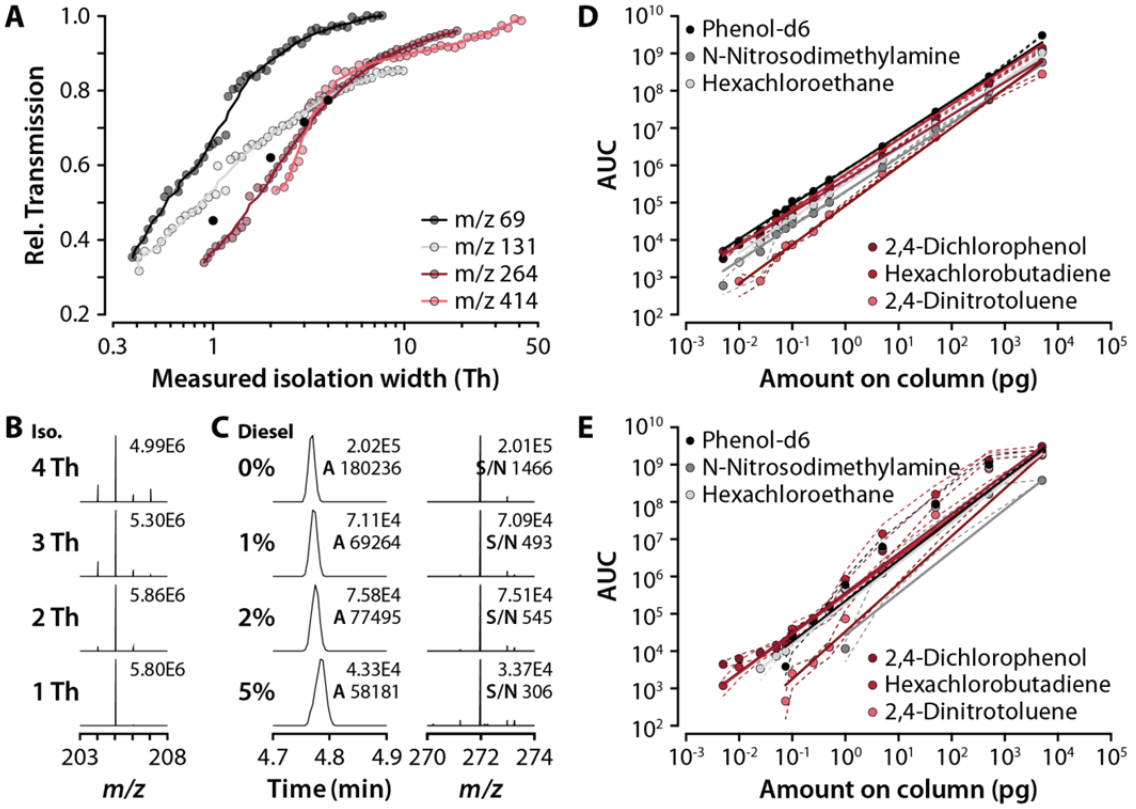


Figure 4.7. (A) Quadrupole isolation transmission efficiency. Transmission relative to RF-only (q = 0.706) operation is plotted for four stable FC-43 ions as a function of measured isolation width using an AGC ion target of 1E5 charges, 17,500 resolution, and <100 ms injection times. The black circles represent expected transmission for 4-chlorodiphenyl isolated from width 1 to 4 Th in panel **B**. (**B**) Isolation of the pseudomolecular ion of 4-chlrodiphenyl (m/z 205.0415) at widths from 1 to 4 Th as part of a targeted SIM analysis of 94 compounds using methane PCI. An ~40% loss in transmission from 4 to 1 Th was expected based on **A**. (**C**) Chromatographic and spectral performance in an analysis of 1 pg octafluoronaphthalene (OFN) in 0-5% v/v diesel. Chromatographic peak areas (denoted "A") and spectral signal-to-noise demonstrate the ability of high resolution/high mass accuracy to maintain performance even at 5% diesel. (**D**) Response curves (peak area versus amount on column) for 6 of 94 compounds (neat) targeted by scheduled SIM (3 Th) over 6 orders-of-magnitude. (**E**) Response curves for 6 of 94 compounds analyzed in full scan mode over 6 orders-of-magnitude. Regression and linearity data for **D** and **E** are available in **Table 4.1**.

was decreased from 4 to 1 Th, yet no loss in signal was observed. In general, however, targeted analyses are performed on samples containing high levels of chemical background in which the target analyte is a very small percentage, not on pure standards as used to benchmark quadrupole transmission efficiency above. Thus, in **Figure 4.7C**, 1 pg of octafluoronaphthalene (OFN) was analyzed neat, as well as in increasingly complex diesel fuel matrices. In SIM mode with a 5 Th isolation window around the molecular ion of OFN (m/z 272), 1 pg of OFN was detected at all diesel concentrations with spectral S/N exceeding 1400 and 300 at 0% and 5% diesel (v/v), respectively, despite chromatographic disturbances and numerous background ions present in the isolation range at 5% (v/v) (see spectra at right). In other experiments where the mass filter was not used (data not shown), no OFN (1 pg) could be detected in the presence of 1% (v/v) diesel. While the quadrupole mass filter successfully eliminates matrix background falling outside of the isolation window, it is the mass selectivity provided by high mass accuracy and high resolution (here, 70,000) mass analysis that enables the differentiation of the ion of interest, with m/z verified by high mass accuracy, from co-isolated, and therefore co-enriched, background.

Linear dynamic range and sensitivity. Again using the 94 compound EPA 8270[58] mix, the achievable linearity, dynamic range, and sensitivity for analysis of these compounds were assessed through replicate injections at amounts ranging from 5 fg to 5 ng, neat, on column. Both full-scan and scheduled targeted-SIM analyses were performed. **Figure 4.7D-E** shows response curves for 6 of the 94 component compounds. Linearity (as measured as the percent relative standard deviation of response factors, i.e., area-under-the-curve normalized by concentration) is enhanced in the targeted SIM analyses compared to the full-scan analyses, where peak areas are underestimated at high concentrations. In the SIM data, the average %RSD of response factors (a measure of both accuracy and precision) was 31.5%, and 26.5%, if the data for 5-10 fg in the 2,4-dichlorophenol dataset were excluded. Linear regression of the data (a measure of precision)

resulted in adjusted R^2 values greater than 0.99 in all cases, suggesting high correlation to a linear model. Responses were typically detected down to 5-10 fg on column for chromatographically well-behaved compounds. In the full-scan data, the average %RSD of response factors was 105%. Linear regression analysis resulted in the adjusted R^2 values, again, greater than 0.99, except for N-nitrosodimethylamine (0.95). In contrast to the SIM data, compounds were typically detected to the ~75-100 fg on column. These data reflect the benefit of quadrupole mass filtering (SIM) in maintaining linearity over a wide dynamic range and extending detection limits. Summary metrics for both datasets are presented in **Table 4.1**.

Isotopomer abundance accuracy. Second in importance only to accurate mass, isotopomer abundances, peak shapes, and/or isotopic fine structure (at very high resolution) can serve as important orthogonal filters for candidate elemental compositions for a particular m/z-value. Typically, the ratio of the first isotopomer to the sum of the mono- and first isotopomers[31], or the relative abundances of the first through third isotopomers[17], relative to theoretically expected ratios or abundances, are used to assess isotopomer abundance accuracy. Using the ratio-based method, in **Figure 4.8A**, isotopomer ratio errors for the 91 of the 94 EPA 8270 analytes are plotted as a function of monoisotopomer abundance, extracted at a tolerance of 5 ppm, for a full scan analysis of 5 ng on column at resolution 17,500. An average error of -2.77 ± 4.96% was observed for all 91 compounds ($error_{rms}$ = 5.65%). Only extracted monoisotopomer peaks with areas less than 1E9 exhibited an isotopomer ratio error outside ±5%. These results are on par with data presented by Abate and colleagues.[31] For a set of 144 trimethylsilylated metabolites analyzed by GC/TOF instrumentation, isotopomer ratio errors of 2.6 ± 2.5% and 2.1 ± 2.6% for the M + 1/M and M + 2/M ratios from molecular and $[M - CH_3]^+$ ions were obtained. They likewise observed decreased isotopomer ratio accuracies for low abundance ions. It should be noted, however, that Orbitrap isotopomer abundance accuracy is inversely related to mass

Table 4.1. Regression and linearity data for SIM and full scan response curves in **Figure 4.7**.

	SIM				Full Scan			
Analyte	**Slope**	**Adj. R2**	**%RSD**	**Det. Limit (pg)**	**Slope**	**Adj. R2**	**%RSD**	**Det. Limit (pg)**
phenol-d6	0.9244	0.9986	24.7	0.005	1.0945	0.9927	95.0	0.075
n-nitrosodimethylamine	0.9223	0.9996	31.4	0.005	1.1207	0.9541	73.0/47.1	1.000/5.000
hexachloroethane	0.9262	0.9996	23.9	0.010	1.0865	0.9991	106	0.025
2,4-dichlorophenol	0.8719	0.9911	47.9/18.2	0.005/0.025	1.0607	0.9971	98.6	0.005
hexachlorobutadiene	0.9448	0.9985	28.5	0.005	1.0429	0.9989	119	0.005
2,4-dinitrotoluene	1.0443	0.9950	32.6	0.010	1.2775	0.9962	142	0.075

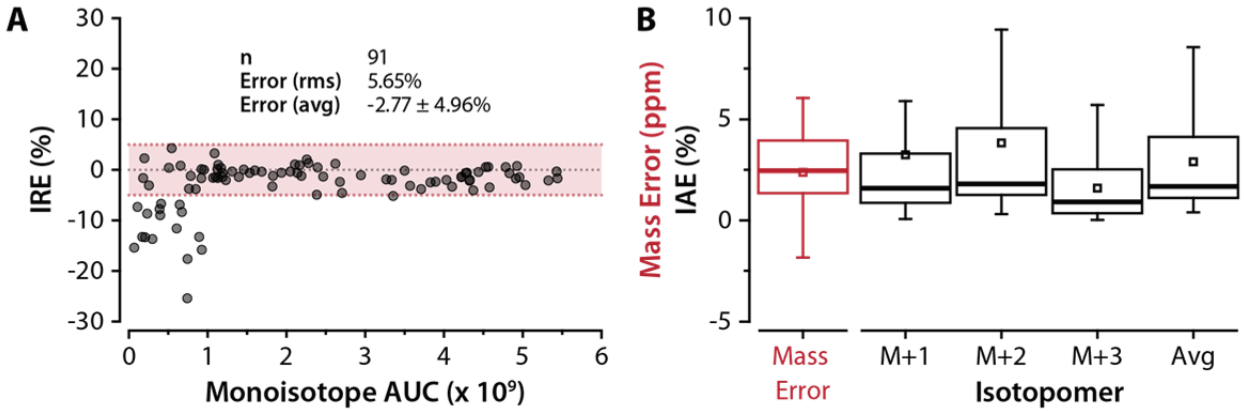


Figure 4.8. (**A**) Percent isotopomer ratio error (IRE) versus extracted ion chromatographic peak area for the monoisotopomeric ion extracted at 5 ppm for 91 of 94 compounds in an analysis at 17,500 resolution. Average IRE was -2.77%. (**B**) Accuracy and precision of mass errors (ppm) and isotopomer abundance errors (IAE, in percent) for 81 putative metabolites and analysis artifacts. M+1, M+2, and M+3 correspond to the individual IAE for the first, second, and third isotopomers, respectively.

analysis resolution. Thus, analyses using resolutions greater than the 17,500 used in this study would likely exhibit lower isotopomer abundance accuracy. Our and other's prior data indicate that isotopomer distribution errors reach ~10% at resolution 100,000.[48,49]

These data indicate that the GC/q-Orbitrap, with the combination of very high mass accuracy, high resolution, and isotopomer abundance errors that are on average less than 5%, will be a powerful tool for the determination of unique elemental composition for mass spectral features of interest. In the accompanying article,[51] this hypothesis is borne out in the ability to uniquely assign elemental compositions to over 80 putative *Arabidopsis thaliana* metabolites and analysis artifacts. For that set of assignments, the distribution of mass and isotopomer abundance errors are displayed in **Figure 4.8B**. Again, isotopomer abundance errors are on average less than 5%, with median errors less than 2.5%. Mass errors average about 2.5%.

Tools for structural characterization of unknowns. The previously discussed figures-of-merit, while providing evidence that the GC/q-Orbitrap will excel in the assignment of elemental composition to mass spectral features, and provide reasonable sensitivity for trace analyses, do not address capabilities of the instrument for structural elucidation. Identification in GC/MS requires comparison of retention time and fragmentation pattern against authentic, preferably internal, reference standards. For a true unknown, a reasonable structural hypothesis is required to guide the purchase and/or synthesis of potential reference standards. While EI-based full-scan spectra are typically rich with ions from which structural inferences can be made, this process can be prohibitively difficult, especially for low-level analytes or analytes that co-elute with other species. As such, alternative analysis modalities can be beneficial to elucidate structure. Numerous such analysis tools have been implemented on the GC/q-Orbitrap instrument, including alternative ionization types (positive or negative chemical ionization with any reagent gas, e.g., methane, acetonitrile (ACN), isobutane, acetone, etc.), targeted-MS/MS

(and targeted-SIM), and advanced data-dependent MS/MS capabilities. **Figure 4.9** provides examples of these analysis modes, and the accompanying article[51] explores the use of advanced data-dependent acquisition in detail.

Figure 4.10 presents a unifying example of how all of the aforementioned analysis modalities can be used in concert to characterize an unknown, here an unknown fatty acid methyl ester (FAME) produced from a bacterial source. The EI spectrum of the unknown in **Figure 4.10A** reveals that the unknown has a molecular ion mass of 310.28663 Th, corresponding to an elemental formula of $C_{20}H_{38}O_2$. While the EI spectrum is indistinguishable from the unknown's cyclic and monoenoic isobars, methyl methylene-octadecanoate (C18:1 CFA) and methyl nonadecenoate (C19:1), the unknown elutes nearly 30 s prior to these isobars. From the chromatogram of the sample (not shown), the unknown is found not to "run with" with any expected FAME present in a bacterial acid methyl ester mix (a standard mix containing many of the fatty acids synthesized by bacteria). The nearest chromatographic peak to the unknown is that of octadecanoate (C18:0) ($C_{19}H_{38}O_2$, $[M]^{+\cdot}$ 298.28663 Th). Neither EI, nor methane PCI, were found to yield any insights into the structure of the unknown, even with use of MS/MS following isolation of the $[M + H]^+$ in methane PCI analyses.

Following hydrogenation of the bacterial FAME sample, however, the retention time of the unknown was found to shift to higher m/z and yield the EI spectrum in **Figure 4.10B**. This spectrum indicates that the unknown is monoenoic, as the molecular ion mass increased to 312.30228 Th, corresponding to the addition of 2 hydrogens for a hydrogenated composition of $C_{20}H_{40}O_2$. While again distinct in spectrum and retention time from to its saturated isobar, nonadecanoate (C19:0), the EI spectrum of the hydrogenated unknown has characteristic fragment ions that suggest a methyl branch at carbon 11.[59,60] This spectrum also matches well with the reference spectrum of methyl 11-methyl-octadecanoate (11-methyl-C18:0). From this

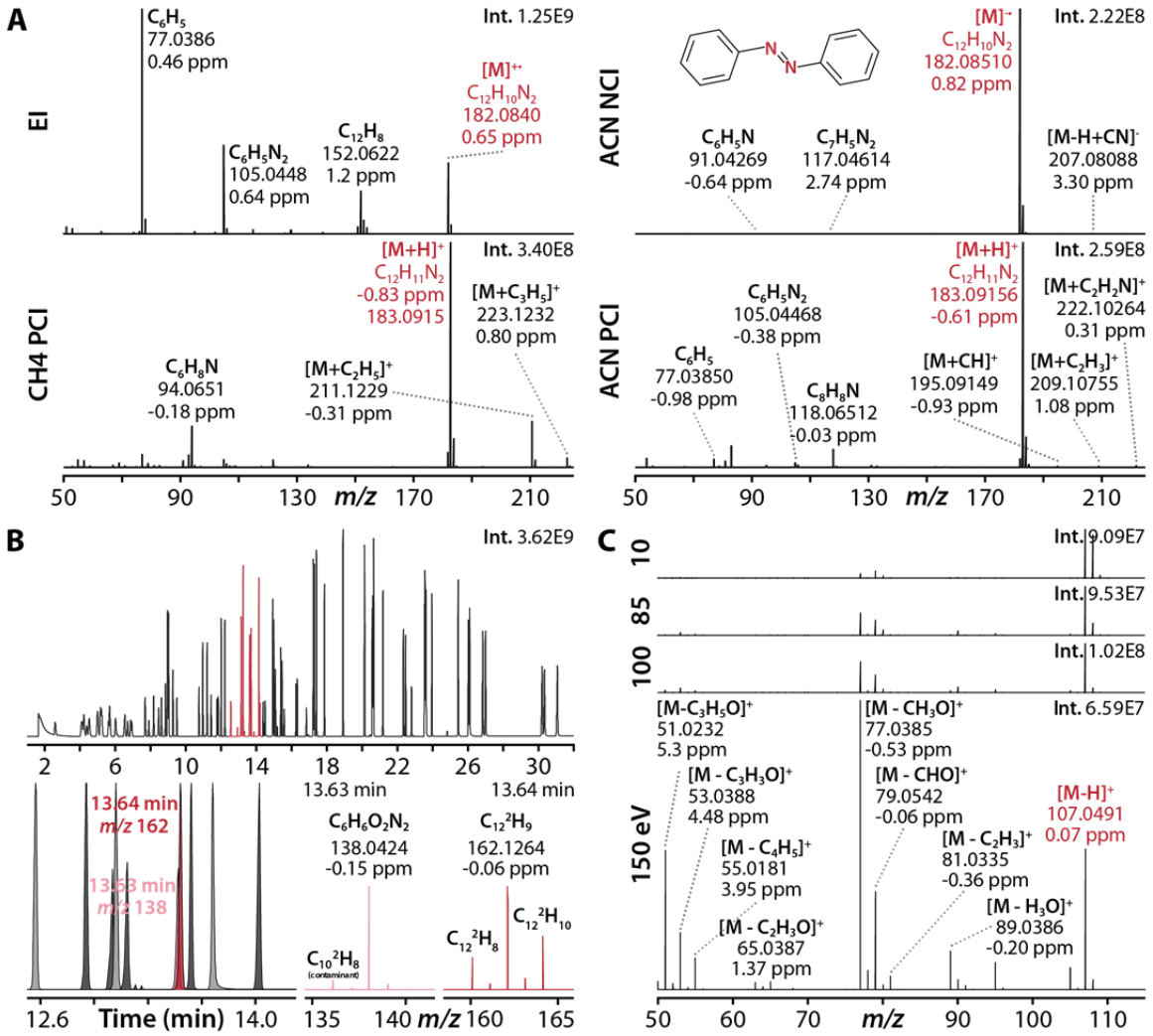


Figure 4.9. Analysis modes enabled on the GC/q-Orbitrap. (**A**) Examples of full scan spectra of 1,2-diphenylhydrazine (as azobenzene) in EI, methane PCI, and acetonitrile PCI and NCI modes. (**B**) Scheduled targeted SIM (4.2 Th isolation width) analysis of 94 EPA 8270 compounds. Extracted ion chromatograms for targeted ions present in the red region of the chromatogram are shown on the bottom left. The SIM spectra for two co-eluting species, 3-nitroaniline at 13.63 min and acenaphthene-d_{10} (in light and dark red, respectively), are shown on the bottom right. The SIM spectrum for 3-nitroaniline is contaminated with a fragment of acenaphthene-d_{10}. (**C**) Scheduled targeted MS/MS (4.2 Th isolation width) of 3-methylphenol in an analysis of 94 EPA 8270 compounds. The effect of increasing collision energy (from 10 to 150 eV) on the distribution of fragment ions is shown.

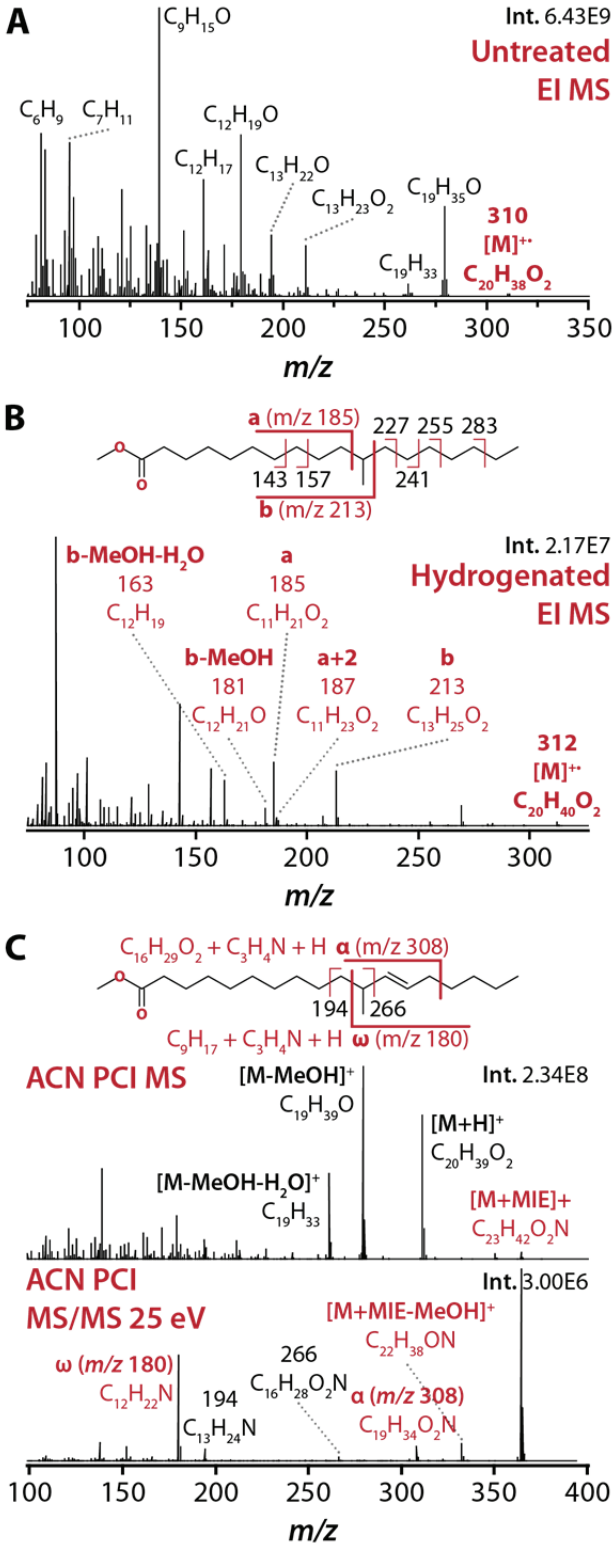


Figure 4.10. Identification of an unknown fatty acid methyl ester. (**A**) EI mass spectrum of the unknown. (B) EI mass spectrum of the unknown following hydrogenation, indicating that the unknown is monoenoic (2 Th mass shift with hydrogenation), and has a branched structure based on characteristic fragments *a* and *b*. The proposed structure of the hydrogenated unknown is depicted above. (B) Acetonitrile PCI MS and MS/MS spectra. The MS spectrum shows several molecular ion adducts, including $[M+MIE]^+$. The $[M+MIE]^+$ ion was isolated and fragmented at 25 eV to generate the MS/MS spectrum. The characteristic α and ω ions localize the double bond. The proposed structure of the unknown is shown above.

information, the unknown is known to possess the general structure of methyl 11-methyl-octadecenoate (11-methyl-C18:1), however the location of the double bond remains unknown.

To localize the double bond, a custom apparatus was set up on the GC/q-Orbitrap to enable PCI with ACN as the reagent gas. In ACN PCI, the reactive species generated by ionization of ACN, 1- methyleneimino-1-ethyenylium (MIE) or $H_2C=N=C=CH_2$, covalently adducts to the double bonds of unsaturated FAMEs to produce a distinct ion, corresponding to $[M + MIE]^+$, in the MS spectrum. Isolation and collisional dissociation of the $[M + MIE]^+$ ion results in diagnostic fragment ions that permit unambiguous localization of double bonds within the FAME.[61-64] The ACN PCI MS spectrum, plotted in the top panel of **Figure 4.10C**, shows the $[M + MIE]^+$ ion of the unknown at m/z 364.32101 with elemental composition, $C_{23}H_{42}O_2N$. Targeted MS/MS of this ion reveals diagnostic α and ω ions that correspond to fragmentation allylic to the double bond (plus a transferred proton). These ions successfully localize the double bond to position 12. Thus, the structure of the unknown can be tentatively assigned as methyl 11-methyl-12-octadecenoate (11-methyl-12-C18:1). While synthesis and side-by-side analysis of this compound (pending) will be required to confirm this assignment, this vignette demonstrates, through a combination of EI, ACN and methane PCI, full scan MS, SIM (not discussed), and targeted MS/MS, the flexibility of the GC/q-Orbitrap to resolve intractable structural characterization and identification challenges.

CONCLUSION

We have described the construction, optimization, and evaluation of a new addition to the GC/MS field, the GC/quadrupole-Orbitrap. This applications-grade, bench-top instrument is the first-ever applications-grade Orbitrap-based instrument dedicated for GC/MS. The GC/quadrupole-Orbitrap robustly generates, transmits, manipulates, and detects EI/CI-generated

ions supplied by a GC in four modes: full scan, SIM with quadrupole isolation, 'all ion fragmentation'-MS/MS with beam-type CAD in the HCD collision cell, and MS/MS with both quadrupole isolation and beam-type CAD. With a dynamic, multiplexed ion injection and mass analysis scheme with automatic gain control, this instrument boasts a 23 Hz (32 ms transient with advanced signal processing (ASP)) or 11 Hz (without ASP) duty cycle at resolution 10,000 (at m/z 200) and resolution 200,000 analysis at 1 Hz (with ASP) or 0.7 Hz (without ASP), all while maintaining low-to-sub-ppm mass errors. Through optimization of scan cycle timing and water sources in the system, ion-molecule reactions are present at levels below those of most GC-ion trap instruments, and resulting spectra match well to reference library spectra. Use of the quadrupole mass filter enables low-femtogram detection limits for common pollutants, with a linear dynamic range spanning 6 orders-of-magnitude. Lastly, the high signal-to-noise, high mass accuracy, and high resolution data generated by this instrument facilitates chemical formulae annotation of unknown spectral features, and enables structural characterization through flexible analysis modalities.

The GC/quadrupole-Orbitrap leaves much room for future development to enhance the capabilities of the instrument for isolation, AGC, and scan speed. For instance, an upgrade of the current mass quadrupole to a high resolution quadrupole rod set with true hyperbolic rods (as opposed to round rods) will provide superior isolation resolution and transmission down to ±0.1 Th isolation widths. Enabling and integrating the electrometer into the current AGC calculations should improve the fidelity of ion injection calculations. And lastly, upgrade of the Orbitrap to the high-field, 20 mm diameter (d20) inner-spindle, Orbitrap mass analyzer will, in combination with advanced signal processing, permit a ~3.5x increase in duty cycle at the same resolution, or significantly higher resolution at the same transient length. Ultra-high resolution (200,000) will then be possible on a GC timescale, ~3.5 Hz with ASP, with fundamentally higher sensitivity.

With the performance metrics of the current, and future, systems exceeding those of any commercially-available GC/MS, the GC/q-Orbitrap has potential to increase the rate of success for annotation and identification of unknowns in analysis of complex biological or environmental samples. By enabling researchers in diverse fields to extract closer to the full story from their GC/MS analyses, the GC/q-Orbitrap has promise to bridge the fundamental technology gap in accurate mass and high resolution instrumentation that exists in small molecule analysis.

REFERENCES

(1) Fiehn, O. Extending the breadth of metabolite profiling by gas chromatography coupled to mass spectrometry. *Trends Analyt Chem* **2008**, *27*, 261-269.

(2) Kind, T.; Wohlgemuth, G.; Lee do, Y.; Lu, Y.; Palazoglu, M.; Shahbaz, S.; Fiehn, O. FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal Chem* **2009**, *81*, 10038-10048.

(3) NIST. *NIST/EPA/NIH Mass Spectral Library and NIST GC Retention Index Database*, 2011; pp.

(4) Bunk, B.; Kucklick, M.; Jonas, R.; Munch, R.; Schobert, M.; Jahn, D.; Hiller, K. MetaQuant: a tool for the automatic quantification of GC/MS-based metabolome data. *Bioinformatics* **2006**, *22*, 2962-2965.

(5) Hiller, K.; Hangebrauk, J.; Jager, C.; Spura, J.; Schreiber, K.; Schomburg, D. MetaboliteDetector: comprehensive analysis tool for targeted and nontargeted GC/MS based metabolome analysis. *Anal Chem* **2009**, *81*, 3429-3439.

(6) Halket, J. M.; Przyborowska, A.; Stein, S. E.; Mallard, W. G.; Down, S.; Chalmers, R. A. Deconvolution gas chromatography/mass spectrometry of urinary organic acids--potential for pattern recognition and automated identification of metabolic disorders. *Rapid Commun Mass Spectrom* **1999**, *13*, 279-284.

(7) Fiehn, O.; Wohlgemuth, G.; Scholz, M. Setup and annotation of metabolomic experiments by integrating biological and mass spectrometric metadata. *Lect Notes Comput Sc* **2005**, *3615*, 224-239.

(8) Luedemann, A.; Strassburg, K.; Erban, A.; Kopka, J. TagFinder for the quantitative analysis of gas chromatography--mass spectrometry (GC-MS)-based metabolite profiling experiments. *Bioinformatics* **2008**, *24*, 732-737.

(9) Tsugawa, H.; Tsujimoto, Y.; Arita, M.; Bamba, T.; Fukusaki, E. GC/MS based metabolomics: development of a data mining system for metabolite identification by using soft independent modeling of class analogy (SIMCA). *Bmc Bioinformatics* **2011**, *12*, 131.

(10) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **2006**, *78*, 779-787.

(11) Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G. XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. *Anal. Chem.* **2012**, *84*, 5035-5039.

(12) Fiehn, O.; Kopka, J.; Dormann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. Metabolite profiling for plant functional genomics. *Nat Biotechnol* **2000**, *18*, 1157-1161.

(13) Fiehn, O.; Kopka, J.; Trethewey, R. N.; Willmitzer, L. Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal. Chem.* **2000**, *72*, 3573-3580.

(14) Kumari, S.; Stevens, D.; Kind, T.; Denkert, C.; Fiehn, O. Applying in-silico retention index and mass spectra matching for identification of unknown metabolites in accurate mass GC-TOF mass spectrometry. *Anal Chem* **2011**, *83*, 5895-5902.

(15) Kim, S.; Rodgers, R. P.; Marshall, A. G. Truly "exact" mass: Elemental composition can be determined uniquely from molecular mass measurement at similar to 0.1 mDa accuracy for molecules up to similar to 500 Da. *Int. J. Mass Spectrom.* **2006**, *251*, 260-265.

(16) Kind, T.; Fiehn, O. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. *Bmc Bioinformatics* **2006**, *7*, -.

(17) Kind, T.; Fiehn, O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *Bmc Bioinformatics* **2007**, *8*, -.

(18) Fjeldsted, J. C. In *Liquid Chromatography Time-of-Flight Mass Spectrometry*; John Wiley & Sons, Inc.: 2008; pp 1-15.

(19) Glish, G. L.; Burinsky, D. J. Hybrid mass spectrometers for tandem mass Spectrometry. *J. Am. Soc. Mass Spectr.* **2008**, *19*, 161-172.

(20) Siuzdak, G. *The expanding role of mass spectrometry in biotechnology*; 2nd ed.; MCC Press: San Diego, CA, 2006.

(21) Thomas, R. A beginner's guide to ICP-MS - Part VII: Mass separation devices - Double-focusing magnetic-sector technology. *Spectrosc.* **2001**, *16*, 22-.

(22) Brady, K. In *Cheese Symposium Proceedings and Presentations* Moorepark, United Kingdom, 2011.

(23) Meng, C.; Sheehan, T. L. July 2011.

(24) Sandy, C. In *Recent Advances in Food Analysis* November 2011.

(25) In *http://www.leco.com/resources/application_notes/pdf/PEG_GC-HRT_LIGHT_CRUDE_OIL_203-821-410.pdf*; Corporation, L., Ed. St. Joseph, MI, 2012.

(26) Alonso, D. E.; Binkley, J.; Siek, K. In *http://www.leco.com/resources/application_notes/pdf/PEG_GC-HRT_ORGANIC_POLLUTANTS_203-821-409.pdf*; Corporation, L., Ed. St. Joseph, MI, 2012.

(27) Binkley, J. In *http://www.leco.com/resources/application_notes/pdf/PEG_GC-HRT_CANNABINOMIMETIC_COMPOUND_SYNTHETIC_CANNABIS_203-821-404.pdf*; Corporation, L., Ed. St. Joseph, MI, 2011.

(28) Alonso, D. E.; Binkley, J.; Siek, K. Comprehensive Analysis of Persistent Organic Pollutants in Complex Matrices Using GC with High-Performance TOF-MS. *Curr Trends Mass Spectrum* **2011**, *July 2011*, 1-6.

(29) Dalluge, J.; Beens, J.; Brinkman, U. A. Comprehensive two-dimensional gas chromatography: a powerful and versatile analytical tool. *J Chromatogr A* **2003**, *1000*, 69-108.

(30) Mondello, L.; Tranchida, P. Q.; Dugo, P.; Dugo, G. Comprehensive two-dimensional gas chromatography-mass spectrometry: a review. *Mass Spectrom Rev* **2008**, *27*, 101-124.

(31) Abate, S.; Ahn, Y. G.; Kind, T.; Cataldi, T. R. I.; Fiehn, O. Determination of elemental compositions by gas chromatography/time-of-flight mass spectrometry using chemical and electron ionization. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 1172-1180.

(32) Blom, K. F. Estimating the Precision of Exact Mass Measurements on an Orthogonal Time-of-Flight Mass Spectrometer. *Anal Chem* **2001**, *73*, 715-719.

(33) Wu, J.; McAllister, H. Exact mass measurement on an electrospray ionization time-of-flight mass spectrometer: error distribution and selective averaging. *J Mass Spectrom* **2003**, *38*, 1043-1053.

(34) Bristow, A. W. T.; Webb, K. S. Intercomparison study on accurate mass measurement of small molecules in mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 1086-1098.

(35) Geiger, T.; Cox, J.; Mann, M. Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation. *Mol Cell Proteomics* **2010**, *9*, 2252-2261.

(36) Makarov, A.; Denisov, E.; Lange, O.; Horning, S. Dynamic Range of Mass Accuracy in LTQ Orbitrap Hybrid Mass Spectrometer. *J. Am. Soc. Mass Spectr.* **2006**, *17*, 977-982.

(37) McAlister, G. C.; Berggren, W. T.; Griep-Raming, J.; Horning, S.; Makarov, A.; Phanstiel, D.; Stafford, G.; Swaney, D. L.; Syka, J. E. P.; Zabrouskov, V.; Coon, J. J. A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer. *J. Proteome Res.* **2008**, *7*, 3127-3136.

(38) Michalski, A.; Damoc, E.; Hauschild, J. P.; Lange, O.; Wieghaus, A.; Makarov, A.; Nagaraj, N.; Cox, J.; Mann, M.; Horning, S. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol Cell Proteomics* **2011**, *10*, mcp.M111.011015.

(39) Michalski, A.; Damoc, E.; Lange, O.; Denisov, E.; Nolting, D.; Mueller, M.; Viner, R.; Schwartz, J.; Remes, P.; Belford, M.; Dunyach, J.-J.; Cox, J.; Horning, S.; Mann, M.; Makarov, A. Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap

Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Mol Cell Proteomics* **2011**, *mcp.O111.013698*.

(40) Olsen, J. V.; Schwartz, J. C.; Griep-Raming, J.; Nielsen, M. L.; Damoc, E.; Denisov, E.; Lange, O.; Remes, P.; Taylor, D.; Splendore, M.; Wouters, E. R.; Senko, M.; Makarov, A.; Mann, M.; Horning, S. A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed. *Mol. Cell. Proteomics* **2009**, *8*, 2759-2769.

(41) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Cooks, R. G. The Orbitrap: a new mass spectrometer. *J Mass Spectrom* **2005**, *40*, 430-443.

(42) Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal Chem* **2000**, *72*, 1156-1162.

(43) Olsen, J. V.; de Godoy, L. M. F.; Li, G. Q.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **2005**, *4*, 2010-2021.

(44) Wenger, C. D.; McAlister, G. C.; Xia, Q.; Coon, J. J. Sub-part-per-million precursor and product mass accuracy for high-throughput proteomics on an ETD-enabled orbitrap mass spectrometer. *Mol Cell Proteomics* **2010**.

(45) Makarov, A.; Denisov, E.; Kholomeev, A.; Baischun, W.; Lange, O.; Strupat, K.; Horning, S. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* **2006**, *78*, 2113-2120.

(46) Makarov, A.; Denisov, E.; Lange, O. Performance Evaluation of a High-field Orbitrap Mass Analyzer. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1391-1396.

(47) Erve, J. C.; Demaio, W.; Talaat, R. E. Rapid metabolite identification with sub parts-per-million mass accuracy from biological matrices by direct infusion nanoelectrospray ionization after clean-up on a ZipTip and LTQ/Orbitrap mass spectrometry. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 3015-3026.

(48) Erve, J. C. L.; Gu, M.; Wang, Y.; DeMaio, W.; Talaat, R. E. Spectral Accuracy of Molecular Ions in an LTQ/Orbitrap Mass Spectrometer and Implications for Elemental Composition Determination. *J. Am. Soc. Mass Spectr.* **2009**, *20*, 2058-2069.

(49) Peterson, A. C.; McAlister, G. C.; Quarmby, S. T.; Griep-Raming, J.; Coon, J. J. Development and characterization of a GC-enabled QLT-Orbitrap for high-resolution and high-mass accuracy GC/MS. *Anal Chem* **2010**, *82*, 8618-8628.

(50) Kind, T.; Fiehn, O. Advances in structure elucidation of small molecules using mass spectrometry. *Bioanal Rev* **2010**, *2*, 23-60.

(51) Peterson, A. C.; Minkoff, B. B.; Bailey, D. J.; Westphall, M. S.; Sussman, M. R.; Coon, J. J. Development of a GC/quadrupole-Orbitrap mass spectrometer, Part II: New

approaches for unambiguous elemental composition assignment and structural characterization in untargeted GC/MS-based metabolomics. *Anal Chem* **2012**, *[In press]*.

(52) Montanari, C.; Sado Kamdem, S. L.; Serrazanetti, D. I.; Etoa, F.-X.; Guerzoni, M. E. Synthesis of cyclopropane fatty acids in Lactobacillus helveticus and Lactobacillus sanfranciscensis and their cellular fatty acids changes following short term acid and cold stresses. *Food Microbiology* **2010**, *27*, 493-502.

(53) Fiehn, O. In *Arabidopsis Protocols*; 2nd ed.; Salinas, J., Sanchez-Serrano, J. J., Eds.; Humana Press: Totowa, NJ, 2006; Vol. 323; pp 439-447.

(54) Creaser, C. S.; West, S. K.; Wilkins, J. P. Reactions of perfluorotri-n-butylamine fragment ions in the quadrupole ion trap: the origin of artefacts in the perfluorotri-n-butylamine calibration spectrum. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 538-540.

(55) Lange, O.; Damoc, E.; Wieghaus, A.; Makarov, A. In *Proc. 59th Conf. Amer. Soc. Mass Spectrom* Denver, CO, 2011.

(56) Lange, O.; Makarov, A.; Denisov, E.; Balschun, W. In *58th Conf Amer Soc Mass Spectrom.* Salt Lake City, Utah, 2010.

(57) Second, T. P.; Blethrow, J. D.; Schwartz, J. C.; Merrihew, G. E.; MacCoss, M. J.; Swaney, D. L.; Russell, J. D.; Coon, J. J.; Zabrouskov, V. Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures. *Anal Chem* **2009**, *81*, 7757-7765.

(58) EPA. *EPA Method 8270C: Semivolatile Organic Compounds by Gas Chromatography/Mass Spectrometry (GC/MS)*, 1996; pp 1-54.

(59) Apon, J. M. B.; Nicolaides, N. The Determination of the Position Isomers of the Methyl Branched Fatty Acid Methyl Esters by Capillary GC/MS. *J. Chromatogr. Sci.* **1975**, *13*, 467-473.

(60) Ryhage, R.; Stenhagen, E. Mass spectrometric studies. IV. Esters of monomethyl-substituted long chain carboxylic acids. *Ark. Kemi* **1960**, *15*, 291-304.

(61) Van Pelt, C. K.; Brenna, J. T. Acetonitrile Chemical Ionization Tandem Mass Spectrometry To Locate Double Bonds in Polyunsaturated Fatty Acid Methyl Esters. *Anal. Chem.* **1999**, *71*, 1981-1989.

(62) Oldham, N. J.; Svatos, A. Determination of the double bond position in functionalized monoenes by chemical ionization ion-trap mass spectrometry using acetonitrile as a reagent gas. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 331-336.

(63) Michaud, A. L.; Diau, G. Y.; Abril, R.; Brenna, J. T. Double bond localization in minor homoallylic fatty acid methyl esters using acetonitrile chemical ionization tandem mass spectrometry. *Anal Biochem* **2002**, *307*, 348-360.

(64) Michaud, A. L.; Yurawecz, M. P.; Delmonte, P.; Corl, B. A.; Bauman, D. E.; Brenna, J. T. Identification and characterization of conjugated fatty acid methyl esters of mixed double bond geometry by acetonitrile chemical ionization tandem mass spectrometry. *Anal Chem* **2003**, *75*, 4925-4930.

CHAPTER 5

Development of a GC/quadrupole-Orbitrap Mass Spectrometer,
Part II: New Approaches for Unambiguous Elemental Composition Assignment
and Structural Characterization in Untargeted GC/MS-based Metabolomics

SUMMARY

Identification of unknown peaks in gas chromatography/mass spectrometry (GC/MS)-based untargeted metabolomics remains an intractable issue. Identification of these peaks is necessary to permit discovery of novel or unexpected metabolites that may elucidate disease processes and/or further our understanding of how genotypes relate to phenotypes. Here, we introduce two new technologies and an analytical workflow that can facilitate the identification of unknown peaks. First, our newly-introduced GC/quadrupole-Orbitrap mass spectrometer provides high mass accuracy, high resolution, and high sensitivity analysis on the fast-GC timescale. Second, with an "intelligent" data-dependent algorithm, termed molecular-ion directed acquisition (MIDA), implemented on this instrument, we maximize the information content generated from unsupervised MS/MS and SIM by directing the mass spectrometer to target the ions of greatest information content, that is, the most-intact ionic species. We combine these technologies with ^{13}C- and ^{15}N-metabolic labeling, multiple derivatization and ionization types, and heuristic filtering of candidate elemental compositions to achieve: 1) structurally-rich MS/MS of intact ion species for structural elucidation, 2) knowledge of carbon and nitrogen content for every ion in MS and MS/MS spectra, 3) relative quantification between alternatively labeled samples, and 4) unambiguous annotation of elemental composition for all queried features.

INTRODUCTION

Broadly, the goal of metabolomics is to characterize, through identity and abundance, all low-molecular-weight metabolites present in a biological system. As with other members of the 'omics' family, approaches for mass spectrometry (MS)-based metabolomics can be of two varieties: targeted and untargeted (akin to the "shotgun" or "discovery" approaches of the proteomics domain). Targeted approaches are geared toward obtaining absolute quantification of a limited set of known metabolites against internal, authentic reference standards, which simultaneously confirm endogenous metabolite identity. Untargeted approaches, however, attempt comprehensive analysis of the metabolome through unbiased investigation into all detected mass spectral features in a given analysis.[1] Both the advantage and disadvantage of the untargeted approach is the generation of data for metabolites previously unknown to the experimenter and/or metabolomics community; if such data can be correctly interpreted, and the identity of the originating metabolites determined, the prospect of discovering novel, or unexpected, metabolites that can elucidate the link between genotype and phenotype, or provide disease biomarkers, awaits.[2] Interpretation of these data, however, is anything but trivial, and despite its centrality to the metabolomics experiment, spectral interpretation and identification remain the most challenging aspects of the analysis.[3]

The challenges hampering spectral interpretation and identification in metabolomics are manifold. First, the targets of a metabolomic analysis are often chemically indistinct from the reagents used to prepare the metabolomic extract. As a result, many, and sometimes most, detected mass spectral features are not of metabolic origin, but rather originate from the reagents, side reactions, or contaminants associated with sample preparation, chemical noise, or sample degradation occurring at some point in the experimental workflow.[4-6] In liquid chromatography (LC)/MS-based metabolomics, for example, less than 10-20% of detected mass spectral features

in a given analysis are thought to be of biological origin.[5-7] Similarly, within the derivatization-based gas chromatography (GC)/MS metabolic platform, multiple MS peaks per analyte, due in part to incomplete derivatization, and multiple analytes per MS peak, due to degradation and side reactions, further complicate matters;[4] thus, in a typical GC/MS study, only 5-15% of mass spectral features can be assigned metabolomic identity.[8] Separating the features of interest from those that are spurious is necessary not only to improve analytical depth[9] and drive understanding of the metabolome, but also to prevent erroneous biological conclusions based on non-metabolic signals.[4] Second, every analyte that reaches the ionization source will further generate numerous ionic species due to various processes (adduction, in-source fragmentation, rearrangements, etc.[10,11]), as well as the its inherent isotopic structure. Depending on the ionization type and chemical nature of the analyte, it can be challenging to group ions arising from the same analyte and obtain accurate readouts of the mass, isotopomer abundances, and overall intensity in congested spectra containing multiple co-eluting species.[6] And third, even for features of metabolic origin with clean mass spectra, true unknowns and novel derivatives of known analytes cannot be annotated via database searching. Rarely does the database searching approach provide unique identification for even known metabolites present in the database.[12-15] Furthermore, without reliable MS peak annotation, the gold-standard of identification against an authentic reference standard is not possible, and interpreting MS/MS fragmentation of a true unknown to gain structural insight is a daunting, if not impossible task requiring a high level of expertise.

In vivo stable isotope incorporation techniques show promise in addressing many of these challenges.[2] Stable isotope labeling (SIL), via metabolic, tagging, and dynamic exchange approaches, has long been used by the proteomics community for structural elucidation and quantification.[16,17] Though only recently fully adopted by the metabolomics community, several groups have used metabolic labeling with stable isotopes (generally, ^{13}C and ^{15}N) in untargeted,

mostly LC/MS-based, studies.[6,18-27] The SIL approach offers a means of discriminating between true metabolite signals and the spurious background discussed above; true metabolites will show paired mass spectral features, whereas background and contaminants will appear as singletons. Furthermore, the *m/z* difference between all paired features denotes the number of carbons (or nitrogens, etc.) present in the feature and, by extension, the metabolite. This information facilitates elemental formula annotation by dramatically reducing the number of candidate formulae possible for a given mass.[22-24] Additionally, if MS/MS is employed, spectral interpretation and structural elucidation is greatly aided by at least partial knowledge of the elemental formula of each fragment ion. Lastly, SIL provides an internal standard for every metabolite to assess recovery,[26] and provide relative quantification.[18,21,25] While these topics have been well-studied within the LC/MS-based untargeted metabolomics domain in a wide-variety of biological systems, to our knowledge only one group,[27] has utilized this technology to assist identification in GC/MS-based applications. There are, however, several reports of stable isotope labeling for the study of metabolic flux ("fluxomics"[2,28]) using GC/MS.[29-33]

Another strategy to reduce candidate elemental compositions is through filtering using a set of heuristic rules developed by Kind and Fiehn.[14] The so-called Seven Golden Rules apply a number of chemical rules (SENIOR, LEWIS, etc.), elemental ratios, and elemental probabilities, and evaluate the accuracy of the experimental mass isotopomer abundance distribution.[13] In conjunction with a mass spectrometer having high mass and isotope ratio accuracy, these rules allow assignment of the correct formula 98% of the time for compounds present in a database.[14] The Seven Golden Rules strategy is straightforward for LC/MS-based metabolomics studies, where soft electrospray ionization (ESI) generally results in an abundant, intact pseudomolecular ion ($[M + H]^+$, $[M + Na]^+$, $[M + K]^+$, etc.). By contrast, most GC/MS-based studies utilize electron ionization (EI) at -70 eV, a hard ionization technique, to make use of the immense EI

spectral libraries available.[3] When annotating a true unknown, the use of standard EI can hinder the calculation of an elemental composition by the above, or any other, method because molecular ions are generally of very low abundance or absent altogether. In addition to various techniques such as lower electron energies,[10] high-mass range ion optics tuning,[34] and cold-EI using supersonic molecular beams,[35] pseudomolecular and/or nearly intact ions can be enhanced through softer ionization types, like methane or ammonia chemical ionization (CI),[10,11] and through use of alternatives to trimethylsilyl (TMS) derivatization.[15,36]

While some metabolites can be analyzed by GC/MS without derivatization (such as those contained in plant essential oils[37] or other volatile organic compounds[38]), the vast majority of metabolites require derivatization prior to GC/MS to enhance their volatility, reduce the polarity of their functional groups, and improve their chromatographic separation.[3,39-41] Silylation is, by far, the most popular method of metabolite derivatization, with TMS derivatization by N-Methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA) the most oft-employed.[41] One alternative, *tert*-butyldimethylsilyl (*t*BDMS) derivatization, has become increasingly popular for the purpose of structural elucidation, due to EI spectra that are dominated in the high-*m/z* range by abundant $[M - t\text{-butyl}]^+$ ions (resulting from preferred loss of the *t*BDMS *t*-butyl moiety).[15,36,39,42] Numerous groups have demonstrated the comprehensive profiling of *t*BDMS-derivatized metabolites, including uncommon ones, with use of unit-resolution mass spectrometers.[15,36,43] Using softer ionization, such as methane or ammonia positive CI (PCI), both TMS- and *t*BDMS-derivatized compounds produce relatively high abundance pseudo-molecular $[M + H]^+$ ions compared to EI at -70 eV,[34,43] in addition to the aforementioned $[M - t\text{-butyl}]^+$ ion (for *t*BDMS), $[M - CH_3]^+$, and chemical ionization adducts.[11,43] While CI can show biases against certain analyte classes, depending on the reagent gas used,[11] in general, the resulting pattern of intact, or nearly intact ion species can greatly improve identification of intact analyte masses, MS feature annotation, and

structural elucidation of unknowns through strategies such as Seven Golden Rules and MS/MS.[44,45]

In this report, we introduce two new technologies that hold promise to further facilitate unambiguous assignment of elemental composition in untargeted, GC/MS-based metabolomic analyses. The first, our newly-introduced GC/quadrupole-Orbitrap mass spectrometer (see accompanying article[46]), enables highly flexible GC/MS analysis with high mass accuracy, resolution, sensitivity, and scan speed. The second, an "intelligent" data-dependent acquisition paradigm for untargeted metabolomics, termed "molecular-ion directed acquisition" or MIDA, maximizes the information content arising from data-dependent MS/MS and SIM by directing the instrument to sample the ions of greatest information content. Using polar metabolites from *Arabidopsis thaliana* extracts as a model system, we combine these technologies with ^{13}C- and ^{15}N-*in vivo* metabolic labeling to enable MS and MS/MS-level annotation and relative quantification, the use of multiple derivatization and ionization conditions, and heuristic rules-based filtering of molecular formulae. We demonstrate the unsupervised, consistent acquisition of structurally-rich MS/MS for intact ion species from nearly all MS features over multiple analyses, knowledge of the carbon and nitrogen content in all MS and MS/MS peaks, and finally, unambiguous assignment of elemental compositions for all queried features.

EXPERIMENTAL PROCEDURES

Reagents. Unless otherwise specified, all reagents were purchased from Sigma-Aldrich (St. Louis, MO). Methanol and water (Optima LC/MS grade) and pyridine (GC grade) were purchased from Fisher Scientific (Fair Lawn, NJ). All compressed gases (methane, helium, and nitrogen) were ultra-high purity grade and purchased from Airgas (Madison, WI).

Stable isotope metabolic labeling, plant growth, and tissue harvest. Stable isotope-labeled growth media variants were prepared by adding the following, per liter, to 1X Murashige and Skoog M059 solution: 1.5mL 1 M $CaCl_2$, 0.75 mL 1 M $MgSO_4$, 0.825 g NH_4NO_3, 0.96 g KNO_3, 0.5 g MES, and 10 g D-glucose. For ^{15}N-labeled plants, NH_4NO_3 and KNO_3 were substituted with $^{15}NH_4{}^{15}NO_3$ and $K^{15}NO_3$. For ^{13}C-labeled plants, D-glucose was substituted with ^{13}C-labeled D-glucose. For ^{15}N/^{13}C-labeled plants, both substitutions were made. Stable isotope-labeled salts and D-glucose were obtained from Cambridge Isotope Laboratories (Andover, MA).

Wild-type *Arabidopsis thaliana* Columbia seeds (Lehle Seeds) were grown in liquid culture. Briefly, 30 mg of sterilized seeds were added to autoclaved, cooled medium in magenta cubes (Nalgene, Rochester, NY), and stratified in the dark at 4 °C for 48 h, after which the magenta cubes were placed at RT under continuous light, shaking at 100 rpm, for 12 d.

Tissue was harvested by removing the plant tissue from the magenta cube with forceps, spinning vigorously for 5 s, flash freezing in liquid nitrogen, and macerating in a pre-frozen mortar and pestle. The tissue powders ($^{12}C^{14}N$, $^{12}C^{15}N$, $^{13}C^{14}N$, and $^{13}C^{15}N$) were immediately collected in separate, pre-frozen 50 mL conical tubes and stored at -80 °C.

Extract preparation and derivatization. Aliquots of ~250-350 mg homogenized powder were extracted as detailed by Fiehn.[47] Briefly, each aliquot of macerate was combined with cold, degassed extraction solution ($CHCl_3/CH_3OH/H_2O$, 1:2.5:1 v/v/v), and 200 ng each of internal standards glycerol-d_8 (Isotec, Miamisburg, OH), benzoic-d_5 acid (Isotec), DL-alanine-2,3,3,3-d_4, and glycine-$^{13}C_2$. The samples were incubated with agitation at 4 °C, followed by sonication in ice. Solids were pelleted by centrifugation, and the supernatant was extracted with H_2O. The aqueous polar and organic lipophilic phases were separated by centrifugation. The upper polar phase was dried under reduced pressure, and subsequently subjected to methoxyamination (20 mg/mL O-methoxyamine HCl in pyridine, 90 min, 30 °C) and silylation

with either N-(*t*-butyldimethylsilyl)-N-methyltrifluoroacetamide (MTBSTFA + 1% TBDMCS; 30 min, 70 °C) or N-Methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA 1% TMCS; 30 min, 37 °C) (Thermo Scientific, Bellafonte, PA). Samples were incubated for 2 h at RT prior to GC/MS analysis.

For relative quantitation studies, $^{12}C^{14}N$ and $^{13}C^{14}N$ samples were first mixed in a 1:1 (v/v) ratio to assess the overall concentration difference between the two samples. Using this information, the samples were mixed again in a concentration-adjusted 1:1 ratio. Serial dilutions of the $^{13}C^{14}N$ sample into the $^{12}C^{14}N$ sample were performed from the 1:1 mix to generate 1:2, 1:5, 1:10, and 1:20 $^{13}C^{14}N$:$^{12}C^{14}N$ mixes.

Gas chromatography. All GC/MS experiments were performed on a Trace Ultra gas chromatograph (Thermo Electron, Milan, Italy) equipped with a CTC Analytics PAL autosampler (Zwingen, Switzerland). Samples (1 μL) were injected via the hot-needle technique at a split ratio of 1:20, 1:50, or 1:100, and injector temperature of 230 °C. Compounds were separated on a 30 m x 0.25 mm (ID) x 0.25 μm (df) Crossbond 5% diphenyl/95% dimethyl polysiloxane column (Restek Rxi-5Sil MS, Bellefonte, PA) with helium carrier gas (1 mL/min) using one of the following oven programs: (1) 5 min isothermal at 50 °C, 80 °C/min to 80 °C, 10 °C/min to 275 °C, 80 °C/min to 320 °C, and 10 min isothermal at 320 °C; or (2) 5 min isothermal at 70 °C, 5 °C/min to 320 °C, and 5 min isothermal at 320 °C. The gas chromatograph was interfaced to the source of the mass spectrometer (*vide infra*) via a transfer line held at 250 °C.

Mass spectrometry. All experiments were performed on a GC/quadrupole-Orbitrap mass spectrometer developed in-house through collaboration with Thermo Fisher Scientific (Bremen, Germany). The design, construction, and characterization of this instrument are described in detail in the accompanying article[46]. Briefly, a standalone Orbitrap MS (Exactive, Thermo Fisher Scientific) was coupled to a single-quadrupole GC/MS (DSQ II) through a custom

adapter manifold mounted in place of the Orbitrap instrument's HCD collision cell, connecting the DSQ's quadrupole to the Orbitrap's c-trap via two flatapoles and added split lens for ion gating. The atmospheric pressure inlet and associated ion optics of the Orbitrap MS were removed, and the HCD cell and DSQ detector setup, consisting of a conversion dynode and electron multiplier, were fitted on the opposite side of the c-trap. Instrument firmware (based on Thermo Q Exactive firmware version 2.0 Build 146201) and electronics were subsequently extensively modified to permit simultaneous control of both component instruments via a single data system.

The GC/quadrupole-Orbitrap's ionization source was configured for either electron ionization (EI) or methane positive chemical ionization (PCI) (electron energy -70 eV, source temperature 200 °C, 2 mL/min methane reagent gas, when applicable). Unless otherwise noted, full scan MS mass spectra were acquired at a resolution (m/Δm) of 17,500 (relative to *m/z* 200), an automatic gain control (AGC) target of 500,000 charges, maximum injection time of 100 ms, and scan range of 70-850 Th. The instrument was routinely tuned and calibrated using FC-43 calibration gas supplied through a calibration gas module. Mass accuracy was checked, and calibrated if necessary, approximately every 24 h.

Molecular-ion directed acquisition (MIDA). The quadrupole-Orbitrap's Python-based firmware was further modified to enable MIDA (pseudocode available in **Appendix 8.1**). Prior to an experiment, the algorithm was informed of the ionization type in use (methane PCI or EI), the sample derivatization reagent employed (MSTFA or MTBSTFA), the mass tolerance to be used by the algorithm in parts-per-million (ppm), the minimum signal-to-noise of mass spectral peaks to be considered as the initial peak in a spectral pattern, the member of the pattern to subsequently target by MS/MS or SIM, the number of targets per MS spectrum to subsequently target, and the duration of time to exclude targets from MS/MS or SIM analysis. These

parameters were accessible to the user via the status tree of the instrument's Tune application. Typical parameters employed for MIDA were as follows: mass tolerance of 10 ppm; minimum S/N of 100; $[M + H]^+$ or $[M - t\text{-butyl}]^+$ target ion for methane PCI and EI, respectively; 1 target per MS spectrum; and, no dynamic exclusion of targets (0 s exclusion). MS spectra were acquired with the parameters listed in the previous section. MIDA MS/MS scans were acquired with an isolation width of 5 Th, normalized collision energy of 25 eV, resolution of 17,500, AGC target of 500,000 charges, maximum injection time of 100 ms, and scan range of 65-850 Th. MIDA SIM scans used the same parameters as MS/MS scans except that the isolation width and scan range were 20 Th in order to capture the entire isotopic envelope of interest. All other method-specific parameters were taken from the associated instrument method file used for the MS analysis.

Data analysis. Manual curation of chromatograms and mass spectra, isotopic distribution simulations, and calculation of elemental compositions were performed with Xcalibur Qual Browser 2.3.23 (Thermo Fisher Scientific) and an in-house developed application MetabID. Unless otherwise indicated, a mass tolerance of 5 ppm was used. MetabID is the offline version of the MIDA algorithm, written in C#/.NET in Visual Studio 2010 (Microsoft, Redmond, WA) with the C# Mass Spectrometry Library (available at https://github.com/dbaileychess/CSMSL). The program processes each spectrum in one, or a set of MS or MIDA MS/MS analyses, using the same parameters as the online algorithm discussed above. Following processing, it displays the top-scoring pattern match for each spectrum overlaid on an extracted chromatogram of targeted ion intensities, as well as outputs the data in a tabular format. The data display permits facile identification and exclusion of background ions picked for MS/MS, supplies a short-list of potentially interesting spectra for manual annotation, and, when metabolically-labeled samples are processed together, allows instantaneous assignment of the number of carbon and nitrogen

present in each pattern target. MetabID also facilitates the assessment of the algorithm accuracy rate for determining the correct target ion in a given spectrum by submitting a set of 100 high-scoring spectra to the user for confirmation.

Candidate elemental compositions were generated within a mass tolerance of 5-10 ppm using the element constraints, $C_{0\text{-}150}H_{0\text{-}150}O_{0\text{-}50}N_{0\text{-}50}S_{0\text{-}50}P_{0\text{-}50}Si_{0\text{-}10}$, and subsequently filtered using the Seven Golden Rules[14] program (available at http://fiehnlab.ucdavis.edu/projects/Seven_Golden_Rules/), which was modified for *t*-butyldimethylsilylated compounds, when applicable. If batch mode of the Seven Golden Rules program was employed, neutralized masses of interest were submitted along with their isotopomer abundances. The program automatically calculated appropriate element constraints in the calculation of candidate compositions. The isotopomer abundance error tolerance was set to 15% if all three isotopes were present and higher if isotopes were not present. Filtering by the number of carbon and nitrogen present was performed on the final list of candidates produced by the Seven Golden Rules program.

Relative quantification of $^{13}C^{14}N$:$^{12}C^{14}N$ samples was performed on the $[M - t\text{-butyl}]^+$ (EI) or $[M + H]^+$/ $[M - t\text{-butyl}]^+$ (CI) isotopomer cluster(s) for 28 randomly selected compounds using a "consensus spectrum" comprising the isotopomer intensities of each member of the cluster averaged over the first half of the peak elution profile (to accommodate isotope swing[48]). The approximate contribution of each species to the isotopomer cluster was determined by the method of least squares for over-determined systems.[49] Given the system,

$$Ax = b$$

Equation 5.1

where A is an m x n matrix of theoretical isotopomer abundance distributions for n chemically-distinct species present in the isotopomer cluster at m m/z values, and b is a m-dimensional matrix of solutions (i.e., the experimentally measured abundance at each m/z value of the cluster). An approximate solution is given by

$$x = (A^T A)^{-1} A^T b$$

Equation 5.2

where x is an n-dimensional matrix of relative proportions of each species comprising the experimental isotopomer cluster. A^T denotes the matrix transpose of the matrix A. Since the incorporation of ^{13}C in the *A. thaliana* model was not 100%, every isotopomer cluster was considered to contain species having 1 or 2 fewer ^{13}C than the fully-incorporated species. The contributions of these species were summed to produce the full contribution of the $^{13}C^{14}N$ sample relative to the $^{12}C^{14}N$ sample.

All software and data (as Thermo .raw files) referenced in the text will be publicly available on our website (http://www.chem.wisc.edu/~coon/).

RESULTS AND DISCUSSION

Molecular-ion directed acquisition (MIDA). MIDA is an 'intelligent,' data-dependent acquisition (DDA) method for directing real-time tandem mass spectrometric events. In traditional, intensity-based DDA, an intensity filter is used to determine the targets of subsequent MS/MS scans.[3,50] In GC/MS, where significant fragmentation occurs upon ionization (in both EI and CI), triggering MS/MS events based on the most-intense species in the spectrum often results in the fragmentation of low-m/z, low-information content ions. As a result, most use of MS/MS

in GC/MS analysis relies on targeted, scheduled methods, like selected reaction monitoring (SRM), which are not amenable to untargeted "discovery" or profiling applications.[3] Thus, to maximize the information content from untargeted MS/MS analysis and establish a "chain-of-evidence" between a precursor and its fragment ions, necessary for the structural elucidation of unknowns,[3] the instrument must be directed at preferred analytical targets, such as the molecular/pseudomolecular ion, or otherwise most-intact ionic species, of each analyte.

The spectral processing algorithm employed by MIDA directs the instrument to these ions by exploiting the expected adducts that form during the methane PCI ionization process, as well as the characteristic fragmentation patterns of derivatization reagents commonly employed in metabolomics studies. We have developed templates, collectively comprising the mass differences resulting from fragmentation of, and adduction to, the molecular ion species, for the following combinations of ionization and derivatization: 1) methane PCI with *t*BDMS derivatization, 2) methane PCI with TMS derivatization, and 3) EI with *t*BDMS derivatization. Shown in **Figure 5.1A**, members of each template have a set mass difference, as denoted by the arrows, from a "template initiator" ion, the lowest *m/z* member of the template. To ensure the specificity of the MIDA algorithm for its intended target, most members of a template must be present for the template to be considered complete (denoted as solid lines). However, since the particular chemistry of the analyte will dictate the presence of certain ions, some template members are not required for a template to be considered complete (shown as dotted lines); this ensures adequate sensitivity of the algorithm. Any required member of the template can be targeted by the subsequent MS/MS analysis, as specified by the user. For the combination of methane PCI and *t*BDMS, for example, the template is made up five ions, three required ions ($[M - C_4H_9]^+$, $[M - CH_3]^+$, and $[M + H]^+$), and two optional species ($[M + C_2H_5]^+$ and $[M + C_3H_5]^+$). The template initiator ion, $[M - C_4H_9]^+$, and the $[M - CH_3]^+$ ($\Delta m = 42.04695$ Da) ion correspond

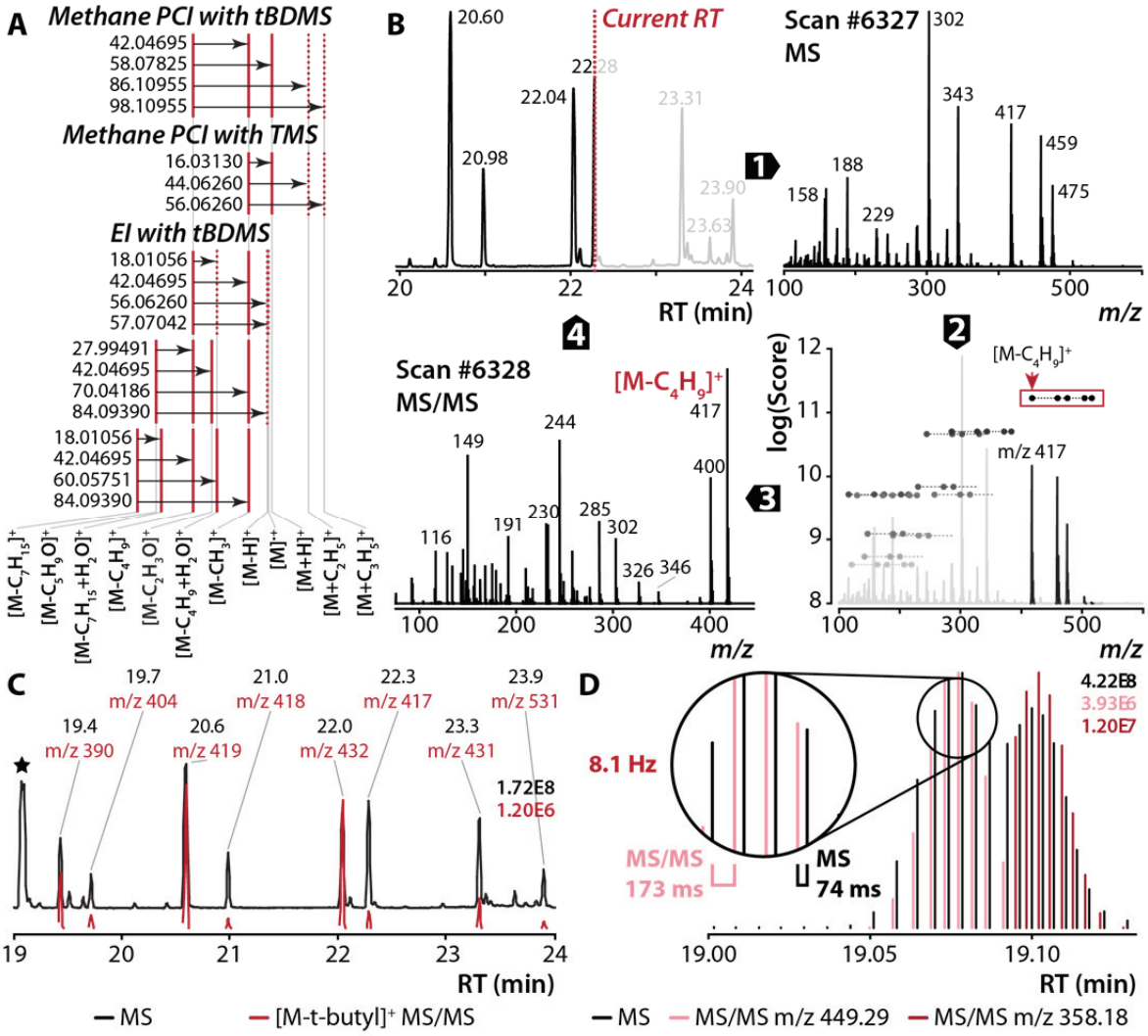


Figure 5.1. Molecular-ion directed acquisition (MIDA). (**A**) Templates for each combination of ionization and sample derivatization type. Mass differences between the template initiator ion (left-most ion) and each template member are shown by black arrows. Required members of the template are denoted by solid lines, optional members by dotted lines. For each ionization/derivatization type, the associated template(s) are scanned over each scan to identify the analyte species designated below. Any required member of a template can be isolated MS/MS or SIM. (**B**) The MIDA process includes four steps: 1) acquisition of an MS scan; 2) analysis of the MS spectrum using the appropriate template, and scoring of the matching templates to identify the best scoring template (boxed in red); 3) acquisition of a MS/MS or SIM scan on the target member of the highest scoring template, $[M - t\text{-butyl}]^+$ at *m/z* 417; and 4) acquisition of further MS/MS or SIM scans in order of decreasing template match scores (not shown), or acquisition of the next MS scan. (**C**) Section of a MIDA-MS/MS analysis showing the MS profile (black), and full elution profiles of MS/MS scans (red) of the $[M - t\text{-butyl}]^+$ of selected analytes. (**D**) Detail of the starred peak in **C** showing the MS profile (black) and MS/MS profiles for the $[M - t\text{-butyl}]^+$ of two species (light and dark red). The inset shows the elapsed time for MS/MS (light red), and MS (black) scans over the first analyte. The overall acquisition rate is 8.1 Hz.

to the loss of a *t*-butyl and methyl moiety, respectively, from the *t*BDMS groups derivatizing the molecule. The remaining three ions result from proton transfer and adduct formation reactions during methane PCI:

$$M + CH_5^+ \text{ (or } C_2H_5^+) \rightarrow [M + H]^+ + CH_4 \text{ (or } C_2H_4)$$

Equation 5.3

and

$$M + C_2H_5^+ \text{ (or } C_3H_5^+) \rightarrow [M + C_2H_5]^+ \text{ (or } [M + C_3H_5]^+)$$

Equation 5.4

respectively, with mass differences, from the initiator, of 58.07825, 86.10955, and 98.10955 Da. The two adduct ions can be of low abundance, or not present, depending on the overall abundance of the analyte, and, thus, are optional in this template. The species to which each template member corresponds for this and other templates are shown at the bottom of **Figure 5.1A**.

The steps of the MIDA process are explained by the example in **Figure 5.1B**. Prior to the start of a GC/MS analysis, the user: 1) informs the instrument of the template in use by specifying the ionization method and sample derivatization type (here, PCI and *t*BDMS); 2) sets the member of the template to target ($[M - C_4H_9]^+$); and, 3) establishes the mass error tolerance (10 ppm) and S/N threshold (100) to enforce for matching the templates to the acquired MS spectra. Following the acquisition of MS spectrum #6327 at 22.28 min (**step 1**), the on-board instrument computer then scans the user-selected template over the MS spectrum (**step 2**). At each potential "template initiator" ion with S/N greater than the S/N threshold of 100, the

spectrum is queried for *m/z* values falling within 10 ppm of each member of the template. If all required members of the template are found, a dot product-based score (explained below) is calculated based on the *m/z* and intensity of each template member. In **step 2**, we see all of the completed templates for MS spectrum #6327 stratified by score and *m/z*, with the highest-scoring template (initiating with *m/z* 417) correctly corresponding to the $[M - C_4H_9]^+$ through $[M + C_3H_5]^+$ ions of the current analyte (asparagine – 3TBS). In **step 3**, the instrument is then directed to acquire MS/MS spectrum #6328 on the user-selected template member, ($[M - C_4H_9]^+$), of the highest scoring template match, on *m/z* 417. The instrument then proceeds to target other templates, in order of decreasing score, if multiple data-dependent events are specified, or acquires the next MS spectrum, if there are not (**step 4**). Due to the time constraints of gas chromatography, as well as its high separation efficiency, a single MS/MS event (top 1) without any dynamic exclusion of previously selected ions was found to be both necessary and sufficient. This process repeats throughout an analysis to yield MS and MS/MS data for nearly all eluting analytes present in the sample. The section of a MIDA analysis shown in **Figure 5.1C** illustrates that for each peak, the targeted ion of the template is profiled over the entirety of its elution (red traces; n.b., only a selection of MS/MS traces are shown for clarity).

The algorithm to score the templates was empirically developed, and is based on a dot-product of the *m/z* and intensity of each member of the template. Initially starting as a straight dot-product score, it was heavily weighted in the *m/z*-domain to promote the selection of the correct series of ions. The scores for templates utilizing methane PCI and EI are given by the following equations (**5.5**, **5.6**), respectively:

$$score_{PCI} = \sum_{j=1}^{n} \frac{I_j}{2} M_j^3$$

Equation 5.5

$$score_{EI} = \sum_{j=1}^{n} \frac{I_j}{2} M_j^5$$

Equation 5.6

where I_j and M_j are the S/N and *m/z*, respectively, of the j^{th} member of n total template members. The EI score was weighted more heavily in the *m/z*-domain, and multiple templates were developed (as seen in **Figure 5.1A**, all of which are scanned over each MS spectrum), because EI spectra contain fewer ions that can direct the algorithm to the correct species in each spectrum. For EI with *t*BDMS derivatization, the algorithm is generally targeted at the $[M - C_4H_9]^+$ ion, which is characteristically the most abundant ion in the high-*m/z* range of the spectrum (usually >5% of base peak[15]), using the presence of a $[M - CH_3]^+$ ion and, occasionally, the molecular ion. A template defined by two peaks does not provide high specificity within a spectrum, especially with the rather common mass difference of 42 Da (C_3H_6). Thus, to ensure accuracy, a greater bias in the *m/z*-domain and additional templates with specificity to certain classes of analytes were required.

To assess the accuracy of the algorithm, we employed a "crowd-sourcing" technique. An offline version of the MIDA algorithm analyzed a MIDA analysis, post-acquisition, and submitted a selection of approximately 100 spectra per analysis to the user to manually annotate as correct or incorrect. For each template, two separate researchers graded three analyses and the accuracy results were averaged. This approach to assessing accuracy was necessary because no

"gold-standard," other than manual annotation, exists for the analysis of these spectra in the absence of library reference spectra. Using this technique, the MIDA algorithm had an accuracy rate of 93.6% and 91.3% for *t*BDMS derivatization with methane PCI and EI, respectively. The accuracy rate fell slightly, to 88.3%, for TMS derivatization with methane PCI.

Another important consideration of any such "real-time" algorithm is the amount of overhead or inter-scan time required to recall the previous mass spectrum, run the algorithm, and report the results back to the instrument to set up the next scan. Given the fast time-scale of the GC/MS analysis, lengthy calculations are contraindicated, and can result in missing the targets of MS/MS analysis once determined, or missing the elution of analytes altogether. Using timing statements embedded in the online algorithm and scan counting, we estimated the time required for each stage of the MIDA process. If the time required to accumulate ions is ignored by forcing 1 ms injection times (a realistic situation only during high ion flux conditions when mass analysis-injection time multiplexing (pipelining) is employed), MIDA (utilizing 17,500 resolution for both MS and MS/MS scans) proceeds at a rate of 9.3 Hz (108 ms per scan), while regular DDA is approximately 16% faster, proceeding at 11 Hz (91 ms per scan). As a point of reference, MS-only acquisition runs at 13 Hz (77 ms), 16% and 28% faster than DDA and MIDA, respectively.

The rather surprising time requirements of MIDA do not originate from the algorithm itself, but from the process of recalling all of the peaks in previous MS spectrum following its acquisition. Once the peaks are recalled, the analysis of the spectrum and setup of the subsequent scan were found to take about 390 μs. Supporting this, if the MIDA algorithm was supplied with a randomly generated spectrum produced *in silico*, MIDA proceeded at an increased rate of 10.3 Hz, thus, requiring only 6.7 ms longer per scan compared to regular DDA. We hypothesize that the inefficiency of recalling peaks from the prior MS spectrum is due to the use of built-in

software functions that pass information from the base firmware, where the spectrum resides, to the software scripting level (Python), where the MIDA algorithm resides. During DDA based on intensity, the function is called only once to pass information for the highest-intensity peak in an intensity-ordered list. In MIDA, because every peak in the spectrum is required, this same function has to be called for every peak in the spectrum (perhaps 1000 times for a complex spectrum) in order of intensity before the collected peaks can be reordered by *m/z* for the MIDA algorithm. Modification and optimization of the functions to pass spectral information tailored to the needs of the MIDA algorithm (to, perhaps, pass spectral information in bulk) would likely be a simple fix to this issue.

Despite the lengthier scan cycle of MIDA, relative to DDA, **Figure 5.1D** demonstrates that the attainable duty cycle, even when ion accumulation times are factored in, is sufficient to properly capture the elution profile of two close-eluting species (black), as well as sample both analytes' molecular ion species with MS/MS (light and dark red). In the inset, details of the scan times and overall scan rates are given for the first, less intense, analyte (light red). Each MS/MS scan averaged about 173 ms (about 65 ms of which were attributable to additional ion accumulation time), while each MS scan averaged about 74 ms in length (ion accumulation times greater than the length of the mass analysis were negligible in this case). Overall, this corresponded to a duty cycle of 8.1 Hz. For the more abundant second analyte (dark red), the time required for MIDA clearly decreases.

While "intelligent" data-dependent and decision tree-based acquisition methods have been developed and utilized within the proteomics domain[51-53], no such methods relying on real-time spectral analysis have been developed for small molecule analysis likely due to the great chemical diversity encompassed therein. However, several methods for data-dependent acquisition based on criteria other than intensity have been developed for small molecules

(though only used on LC/MS instruments), including neutral loss-,[54] mass defect-,[55] and isotope pattern-dependent acquisition.[56-58] In the latter, which is most similar to the MIDA approach, the instrument is directed to target for MS/MS ions having distinct, user-selected isotopic patterns resulting from the presence of natural isotopes (Cl, Br, etc.) or incorporated stable-isotope labels (^{13}C, ^{15}N, ^{2}H, etc.). To the best of our knowledge, ions demonstrating the correct isotopic pattern are simply selected in order of decreasing intensity. This has been used successfully to perform data-dependent MS/MS on chlorine and bromine-containing drugs,[58] and glutathione conjugates containing ^{13}C-labels.[56,57] All of these methods are specific for certain classes of molecules (those having chlorine or exhibiting a certain neutral loss, etc.), and cannot be applied in a generic manner, which is required for untargeted metabolomics and a feature of MIDA. Additionally, all of these methods are geared toward LC/MS applications where the use of soft ionization generally results in the presence of intact precursor ions, which can be subsequently targeted for MS/MS. Outside of the domain of peer-reviewed literature, one report exists where the pattern of methane-PCI adduct ions was used to select pseudomolecular ions for MS/MS on an ion trap-based GC/MS.[59]

MIDA with metabolic labeling for untargeted metabolomics and structural elucidation. To assess the utility of MIDA in an untargeted, metabolomic profiling context, the benefits of a high resolution and accurate mass instrument, the structural information gained by intelligent acquisition of MS/MS spectra, and the advantages of metabolic labeling for assigning elemental compositions were combined to study polar metabolites from *Arabidopsis thaliana*. Identical polar fraction extracts of *A. thaliana* grown under four metabolic labeling conditions were analyzed: natural abundance ($^{12}C^{14}N$), ^{13}C-enriched ($^{13}C^{14}N$), ^{15}N-enriched ($^{12}C^{15}N$), and both ^{13}C- and ^{15}N-enriched ($^{13}C^{15}N$). Using both derivatization and ionization conditions, samples were analyzed using MIDA-MS/MS or full-scan MS. **Figure 5.2** depicts typical data

acquired by this analysis strategy. MS spectra from either full-scan MS or MIDA-MS/MS analysis immediately allow assignment of the number of nitrogens and carbons comprising the underivatized molecule (**Figure 5.2A**). Using MIDA-MS/MS, in **Figure 5.2B-C**, the algorithm's recognition of fragmentation/adduction patterns, rather than specific *m/z* or intensity values, permits the acquisition of MS/MS spectra of the $[M - CH_3]^+$ of each differentially-labeled species over its entire elution period in four separate runs. Comparison of these MS/MS spectra again allows immediate assignment of the number of nitrogens and carbons present in each fragment ion. This powerful information, along with high mass accuracy *m/z* measurements, substantially reduces the number of potential elemental composition candidates for a given peak and facilitates its structural characterization, as shown by the suggested structures for eight peaks in the spectrum in **Figure 5.2C**.

With this set of spectral data in-hand for a given analyte, as well as isotopomer abundance information for the ion targeted by MIDA-MS/MS, the workflow illustrated in **Figure 5.3** was developed to generate unique elemental compositions assignments and tentative identifications for each putative metabolite. Using the spectral data discussed above, a list of candidate elemental compositions within ±5 ppm of the MIDA-targeted ion was generated using lax constraints on the number of carbon, nitrogen, oxygen, hydrogen, sulfur, phosphorus, and silicon allowable in the candidate compounds. This initial list was filtered for the presence of silicon (required given the sample preparation and the templates used by the MIDA algorithm), and then subjected to further attrition by the Seven Golden Rules[14] using a 15% isotopomer abundance error (IAE) threshold. This set of heuristic filters for molecular formulae employs the LEWIS and SENIOR chemical rules, accuracy of isotopomer abundance patterns, elemental ratios (H/C and [N, O, P, S]/C), elemental ratio probabilities, the presence of derivatizable functional groups, and the presence of elemental formulae in the PubChem database

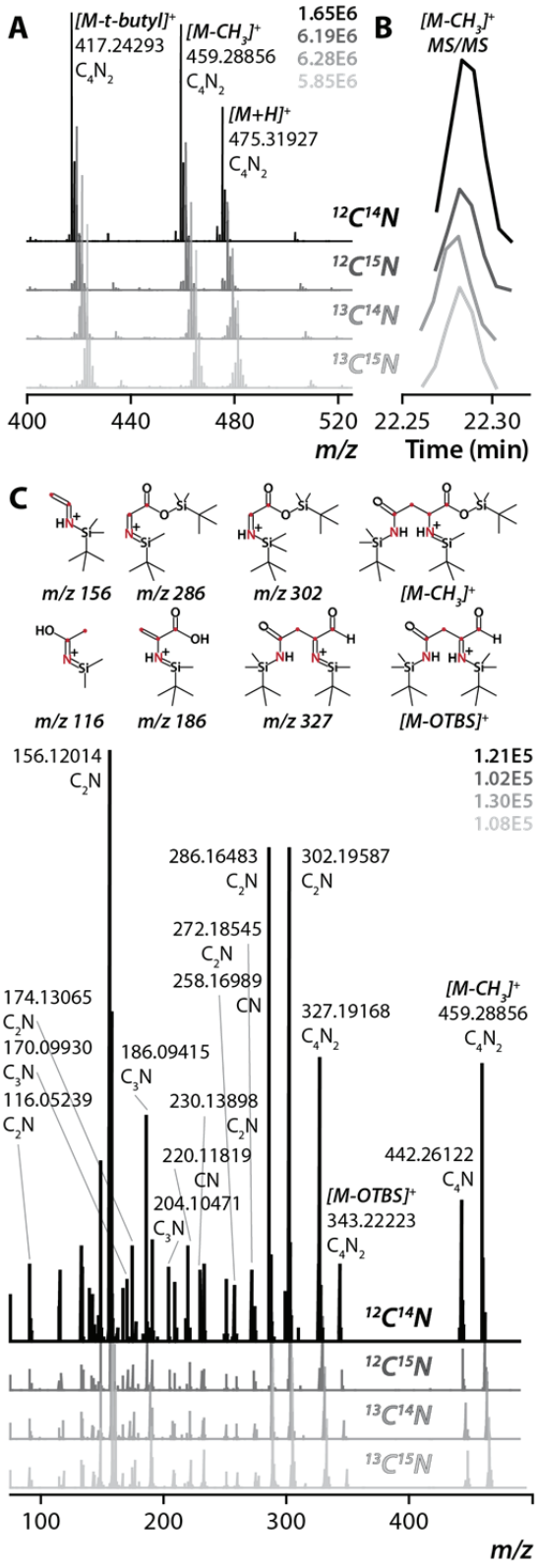


Figure 5.2. Typical MIDA-MS/MS with metabolic labeling data. (**A**) Partial MS spectrum showing [M – *t*-butyl]$^+$, [M – CH_3]$^+$, [M + H]$^+$, and methane PCI adduct ions (unlabeled) for asparagine – 3TBS under 4 different metabolic labeling states. Comparison of the *m/z* shifts of similar ions between the 4 states allows assignment of the number of carbons and nitrogens in each ion. (**B**) Profiles of the MIDA- MS/MS of [M-CH_3]$^+$ over the entire elution profile of asparagine – 3TBS for all 4 labeling states. (**C**) MS/MS of asparagine – 3TBS [M-CH_3]$^+$ ion from all four labeling states. Comparison of the *m/z* shifts of similar ions between the 4 spectra allows assignment of the number of carbon and nitrogen in each fragment ion. Above, structures proposed using the knowledge of the number of carbons and nitrogens are shown as red circles or red letters, respectively, for 8 ions in the MS/MS spectrum. The MS/MS spectrum confirms the structure of asparagine – 3TBS.

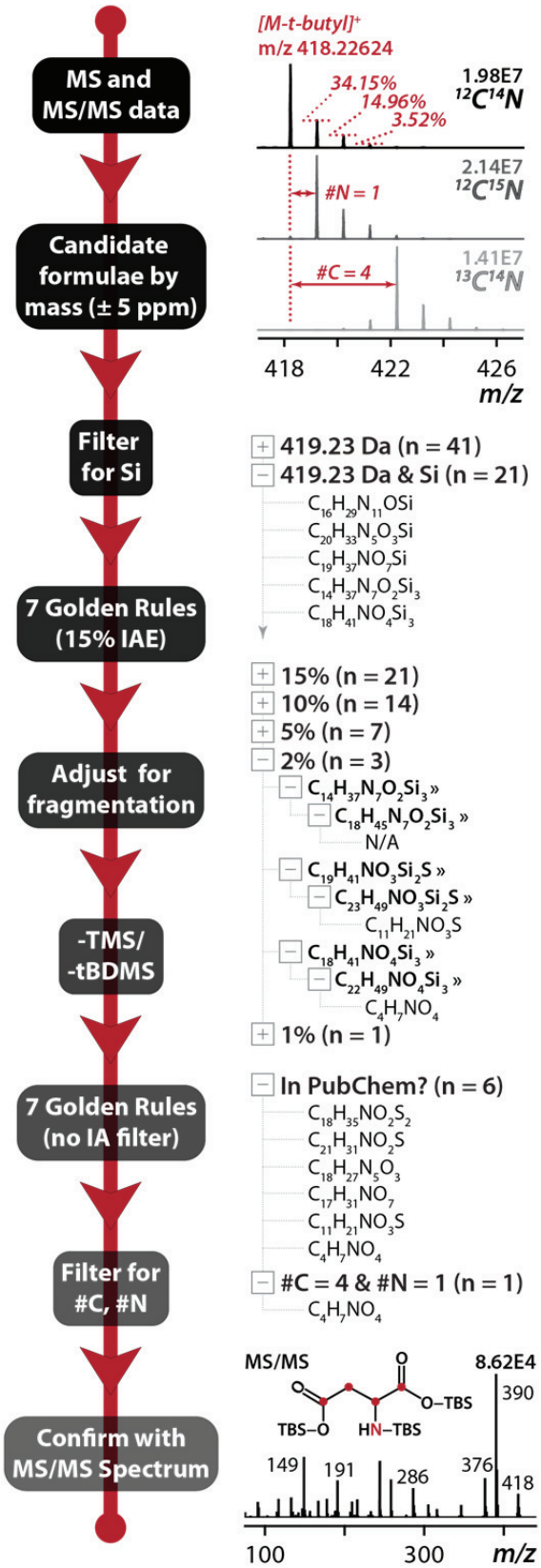


Figure 5.3. Workflow for spectral annotation and structural confirmation. From top to bottom, first, the ion type selected by MIDA for MS/MS, the *m/z*, the abundance of the first – third isotopomers, and the number of carbons and nitrogens present are noted from the MS spectrum. Candidate formulae are then generated within ±5 ppm tolerance of the neutralized mass of the ion, and filtered for Si to result in a list of 21 formulae (out of 41). Candidates are submitted for filtering by the Seven Golden Rules with a 15% isotopomer abundance error (IAE) threshold. All 21 formulae meet the 15% threshold, 14 meet a 10% threshold, 7 at 5%, 3 at 2%, and 1 at 1%. All formulae meeting the 15% threshold are made intact by addition of C_4H_8 (shown in the second level under the formulae meeting the 2% IAE threshold). Silylation groups are removed from the intact formulae, as shown in the third level under the formulae meeting the 2% IAE threshold. The de-silylated formulae are re-filtered by the Seven Golden Rules. The 6 formulae present in PubChem are further filtered by the number of nitrogen and carbon present in the analyte (4 carbons and 1 nitrogen) to yield a single formula, $C_4H_7NO_4$, which is confirmed using the MS/MS spectrum, and tentatively identified as aspartate – 3TBS.

(http://pubchem.ncbi.nlm.nih.gov/). The resultant list was filtered for compositions meeting the 15% IAE threshold, and then each elemental composition was adjusted to be fully intact. In other words, using the example to the right of the workflow in **Figure 5.3** where the spectral information from a neutralized $[M - C_4H_9]^+$ molecule (419.23 Da) was used to generate the initial list of candidates, each remaining candidate with less than 15% IAE was made 'intact' by adding C_4H_8 to each formula (to account for the loss of *t*-butyl, less one hydrogen which was already added to neutralize the initial ion). Thus, the composition $C_{18}H_{41}NO_4Si_3$ became $C_{22}H_{49}NO_4Si_3$. The intact candidates were then stripped of their *t*BDMS groups to reveal a list of native (but possible methoxyaminated) molecules ($C_{22}H_{49}NO_4Si_3$ with loss of 3 *t*BDMS became $C_4H_7NO_4$). After resubmission to the Seven Golden Rules (with a 100% IAE threshold), the remaining elemental compositions that were present in the PubChem database were filtered by the number of carbons and nitrogens known to be present from the spectral data. If no matching compounds were found and methoxyamination was suspected (note, the $=N\text{-}O\text{-}CH_3$ moiety added by methoxyamination did not contain ^{13}C or ^{15}N and, thus, was "invisible" by our metabolic labeling method), the compounds were de-methoxyaminated by subtraction of NCH_3 and re-filtered for nitrogen and carbon. The remaining candidate (or rarely, candidates) was (were) then confirmed by annotation of the associated MS/MS data.

Using this strategy, we have made confident elemental composition assignments and suggested plausible identifications for over 80 compounds, some of which we successfully identified as artifacts of the analysis (e.g., hydroxylamine, and portions of the benzoic and carbonic acid populations). **Table 5.1** shows a selection of the identified compounds from analyses using TMS derivatization and methane PCI. The mean mass error for annotated mass spectral features was 2.4 ppm (σ = 2.1 ppm, median = 2.5 ppm); the mean percent error for isotopomer abundances (including, first, second, and third isotopomer abundance errors) was

Table 5.1. Selected compounds tentatively identified by our workflow in methane PCI with TMS derivatization analyses showing the reduction of candidates with each filtering step (see **Figure 5.3**).

RT	Ion	m/z	#C	#N	MS/MS	# by Mass	#Pub-Chem	#N & #C	Native Formula	Proposed ID
8.15	[M-CH3]+	174.05840	3	0	TRUE	1	1	1	C3H4O3	pyruvic acid
8.31	[M-CH3]+	234.11611	n/a	n/a	TRUE	10	1	n/a	H3NO	hydroxylamine
8.93	[M-CH3]+	218.10275	3	1	TRUE	2	1	1	C3H7NO2	alanine
9.33	[M+H]+	235.08182	2	0	FALSE	9	2	1	C2H2O4	oxalic acid
9.38	[M+H]+	176.07407	2	0	FALSE	3	0	1	C2H2O3	glyoxylic acid
9.45	[M-CH3]+	220.08219	1	0	TRUE	2	0	1	CH2O3	carbonic acid
10.52	[M+H]+	262.16547	5	1	TRUE	6	2	1	C5H11NO2	valine
10.85	[M-CH3]+	189.08734	1	2	TRUE	5	3	1	CH4N2O	urea
10.99	[M-CH3]+	179.05241	7	0	TRUE	3	2	1	C7H6O2	benzoic acid
11.22	[M-CH3]+	262.14713	2	1	TRUE	4	2	1	C2H7NO	ethanolamine
11.25	[M+H]+	315.10260	0	0	TRUE	29	7	1	H3O4P	phosphate
11.28	[M-CH3]+	260.14987	6	1	TRUE	4	2	1	C6H13NO2	leucine
11.30	[M-CH3]+	293.14190	3	0	TRUE	9	3	1	C3H8O3	glycerol
11.58	[M+H]+	276.18144	6	1	TRUE	6	2	1	C6H13NO2	isoleucine
11.61	[M+H]+	196.07900	6	1	FALSE	2	2	1	C6H5NO2	nicotinic acid
11.64	[M-CH3]+	244.11824	5	1	TRUE	3	1	1	C5H9NO2	proline
11.70	[M-CH3]+	245.06599	4	0	FALSE	7	2	1	C4H4O4	maleic acid
11.75	[M-CH3]+	276.12620	2	1	TRUE	17	8	1	C2H5NO2	glycine
11.85	[M-CH3]+	247.08178	4	0	TRUE	7	2	1	C4H6O4	succinic acid
12.05	[M-CH3]+	307.12089	3	0	TRUE	18	4	1	C3H6O4	glyceric acid
12.34	[M-CH3]+	245.06572	4	0	TRUE	19	5	1	C4H4O4	fumaric acid
12.45	[M-CH3]+	306.13719	3	1	TRUE	10	3	1	C3H7NO3	serine
12.79	[M-CH3]+	320.15303	4	1	TRUE	13	3	1	C4H9NO3	threonine
13.61	[M-CH3]+	320.15340	4	1	TRUE	27	7	1	C4H9NO3	allothreonine
13.87	[M-CH3]+	349.13159	5	0	FALSE	32	9	1	C5H8O5	citramalic acid
14.09	[M-CH3]+	335.11585	4	0	TRUE	29	7	1	C4H6O5	malate
14.36	[M-CH3]+	267.08654	7	0	FALSE	10	1	1	C7H6O3	hydroxybenzoic acid
14.48	[M+H]+	350.16327	4	1	TRUE	23	7	1	C4H7NO4	aspartic acid
14.50	[M+H]+	274.12904	5	1	TRUE	7	2	1	C5H7NO3	pyroglutamic acid
14.55	[M-CH3]+	332.15310	5	1	FALSE	15	4	1	C5H9NO3	hydroxyproline
14.61	[M-CH3]+	304.15744	4	1	TRUE	21	7	1	C4H9NO2	4-aminobutyric acid
14.91	[M-CH3]+	322.11348	3	1	FALSE	44	11	1	C3H7NO2S	cysteine
14.94	[M-CH3]+	409.17171	4	0	FALSE	66	18	1	C4H8O5	threonic acid
15.62	[M-CH3]+	333.18443	5	2	TRUE	14	5	1	C5H12N2O2	ornithine
15.67	[M-CH3]+	348.14728	5	1	TRUE	24	7	2	C5H9NO4	glutamic acid
15.75	[M+H]+	310.16485	9	1	TRUE	10	2	1	C9H11NO2	phenylalanine
16.23	[M-CH3]+	333.14831	4	2	TRUE	20	7	1	C4H8N2O3	asparagine
16.92	[M-CH3]+	419.20368	5	2	TRUE	56	13	1	C5H10N2O3	glutamine
17.00	[M-CH3]+	361.23460	4	2	FALSE	15	6	1	C4H12N2	putrescine
17.21	[M+H]+	436.16378	6	0	TRUE	189	1	1	C6H6O7	2-oxalosuccinic acid
17.39	[M-CH3]+	347.16298	5	2	TRUE	47	12	1	C5H10N2O3	glutamine
17.75	[M-CH3]+	447.18690	7	0	FALSE	104	22	1	C7H10O5	shikimic acid
17.85	[M-CH3]+	465.16041	6	0	TRUE	157	23	1	C6H8O7	citrate
17.92	[M-CH3]+	358.18007	6	3	TRUE	17	6	1	C6H11N3O2	arginine [-NH3]
18.14	[M-CH3]+	422.14991	6	0	TRUE	75	1	1	C6H8O7	2-(glycoloyloxy)succinic acid
18.45	[M+H]+	229.11627	10	2	FALSE	6	1	1	C10H8N2	beta-indole-3-acetonitrile
18.53	[M-CH3]+	431.17790	4	4	TRUE	86	19	1	C4H6N4O3	allantoin
18.87	[M-CH3]+	356.16388	6	3	FALSE	23	7	1	C6H9N3O2	histidine
18.92	[M+H]+	363.23144	6	2	FALSE	17	5	1	C6H14N2O2	lysine
19.09	[M-CH3]+	382.16870	9	1	TRUE	35	7	1	C9H11NO3	tyrosine
19.16	[M-CH3]+	449.16596	6	0	TRUE	127	22	1	C6H8O6	ascorbic acid
19.47	[M-CH3]+	435.18698	6	0	FALSE	90	19	1	C6H10O5	1,6-anhydroglucose
20.54	[M+H]+	613.30795	6	0	FALSE	384	45	1	C6H12O6	inositol
20.61	[M-CH3]+	441.16231	5	4	FALSE	109	31	1	C5H4N4O3	uric acid
21.67	[M+H]+	421.21589	11	2	TRUE	55	11	1	C11H12N2O2	tryptophan
21.92	[M-CH3]+	353.12384	11	0	TRUE	41	9	1	C11H12O5	sinapic acid

2.9% (σ = 2.9%, median = 1.7%) (**Figure 5.4A**). While isotopomer abundance accuracy suffered for low abundance compounds, or when the ion used for annotation contained interference from other ion species or chemical noise, mass errors were independent of abundance (**Figure 5.4B**). While authentic reference standards for each of these putatively identified compounds would be required for their confirmation, and for the identification of structural isomers, the metabolic labeling approach concomitant with accurate mass measurements, adequate isotopic abundance accuracy, and structural information from MS/MS spectra, leaves little room for ambiguity. In all cases but three, knowledge of the number of nitrogens and carbons present in the parent molecule permitted a unique result among compounds present in the PubChem database. In the first case, the two remaining compositions for a derivatized $[M - CH_3]^+$ ion at *m/z* 348.14728, containing 5 carbons and 1 nitrogen, were $C_5H_9NO_4$, corresponding to amino acid glutamic acid, and C_5HN, corresponding to 2,4-pentadiynenitrile, a compound thought to be formed in the atmosphere of Saturn's moon Titan or in the interstellar medium.[60] Occam's razor, the MS/MS spectrum, and the unlikelihood of 2,4-pentadiynenitrile's derivatization with four TMS groups, make glutamic acid the clear choice in this case of ambiguity. Similarity, in the second case (or, rather, set of cases), analysis of several mass spectral features having a derivatized $[M - CH_3]^+$ ion at *m/z* 554.26355 containing 6 carbons and no nitrogens resulted in two compositions remaining after filtering, $C_6H_{12}O_6$ – 5 TMS, 1 MeOX (likely corresponding to several isomers of glucose) and $C_6H_4O_2$ – 6 TMS, 1 MeOX, of which only the former was chemically possible. The third case was similarly unambiguous: $C_6H_{10}O_4$ – 3 TMS, an expected fragment of a di-/tri-saccharide, following cleavage of the glycosidic bond, versus C_6H_2 – 4 TMS, which is not chemically possible.

The success of the SIL strategy is mirrored by the results of the only other study utilizing this technology to aid identification of unknown spectral features in GC/MS-based metabolomics.

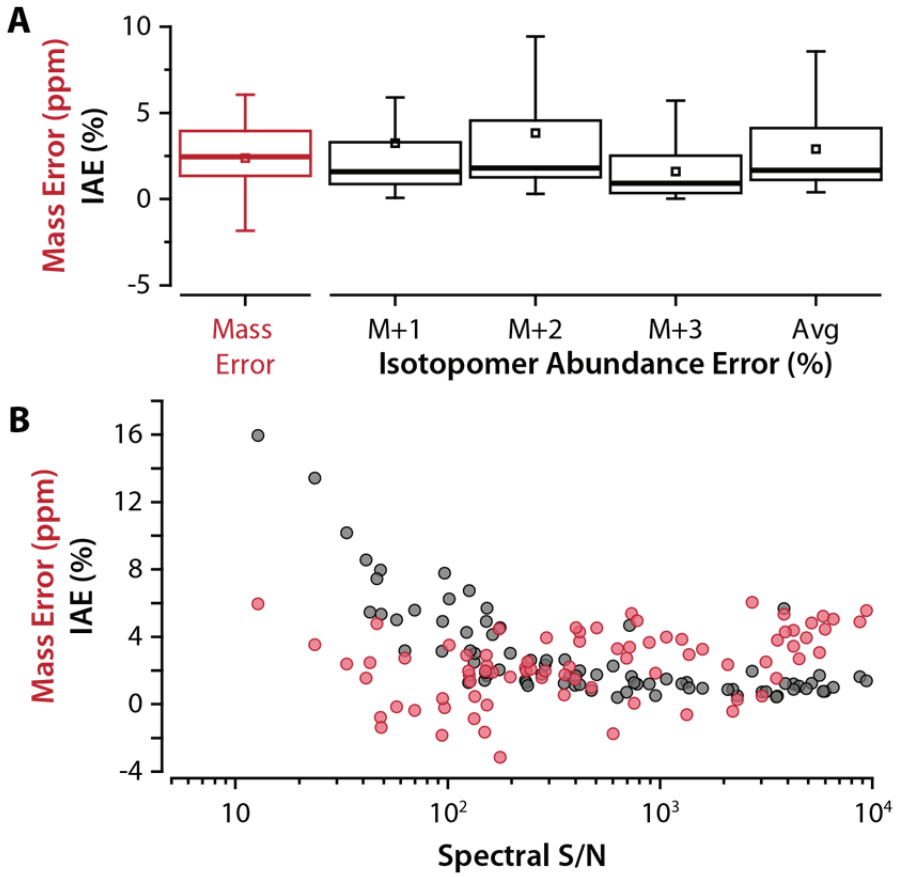


Figure 5.4. (**A**) Boxplots of mass error in parts-per-million and isotopomer abundance error (IAE) in percent for all annotated mass spectral features (n = 81). 'M+1', 'M+2', and 'M+3' represent the IAE for the first, second, and third isotopomers of a given monoisotopic peak, while 'Avg' is the average IAE for all isotopomers. **(B)** Relationship of mass error and IAE with the spectral signal-to-noise (S/N) of the annotated feature. Mass accuracy is independent of spectral S/N, while IAE increases for low S/N features.

Herebian and colleagues[27] studied the metabolome of *Corynebacterium glutamicum*, a species of bacteria used for the industrial-scale production of glutamic acid,[61] through both ^{13}C and ^{15}N metabolic labeling. Using this strategy, they were able to re-classify several compounds, previously considered part of their *C. glutamicum* metabolite library, as artifacts introduced during sample preparation. Additionally, several hitherto unidentified MS peaks were elucidated using nitrogen and carbon counts, as well as knowledge of the number of methoximes and TMS groups present (via further experiments with d_3-methoximes, d_9-TMS, and ethyldimethylsilylation[62]). However, unlike in this study, where a unique elemental formula was obtained in nearly all cases, Herebian *et al*. could not arrive at a unique result, or found no result, in several cases. We believe that the use of high mass accuracy instrumentation and heuristic filtering of candidate elemental compositions in this study were the features that enabled unambiguous elemental formulae annotation.

Relative quantification via metabolic labeling and MIDA-SIM. Besides attempting to catalog and annotate, with elemental composition and putative identifications, all components in a given metabolomic extract, metabolomic studies are also concerned with identifying components that differ in abundance between two or more biological conditions.[2,63] In the absence of an authentic internal reference standard, this might be accomplished by comparing peak areas between separate analyses of two biological conditions.[64] If the two conditions are differentially metabolically labeled, however, samples can be mixed, analyzed simultaneously, and relative quantitative information gathered for each analyte within the same analysis, while retaining the enhanced ability to assign elemental composition using the knowledge of number of carbon and nitrogens present in each peak. This strategy would obviate concerns of incomparability between separate analyses due to variations in analysis conditions or instrument performance. Additionally, such an approach permits relative quantification of numerous, natural-abundance

samples against a common, metabolically labeled sample, enabling large-scale relative quantification experiments.[2]

To establish feasibility of this approach for small molecule metabolites with GC/MS, the $^{13}C^{14}N$-labeled TBS-derivatized sample was serially diluted into the $^{12}C^{14}N$-labeled TBS-derivatized sample at five different ratios (1:1, 2:1, 5:1, 10:1, and 20:1 $^{12}C^{14}N$:$^{13}C^{15}N$), and the mixes analyzed with methane PCI and EI (initially, with full scan MS only). For approximately 28 compounds in each analysis, the $^{12}C/^{13}C$ ion pair cluster, corresponding to the $[M - C_4H_9]^+$ of each compound, was manually extracted. The method of least squares for over-determined systems[49] was then employed to estimate the relative contribution of each species present in the extracted ion cluster based on the theoretical isotopomer abundance distributions for each species in isolation (**Figure 5.5**). In this study, it was not sufficient to simply use the peak heights of the ^{12}C- and ^{13}C- monoisotopic peaks for relative quantitation as has been previously reported.[21] Due to the native metabolites (i.e., without the addition of TBS groups) under study having only between 1 and 11 carbon atoms, incomplete incorporation of ^{13}C in the $^{13}C^{14}N$-labeled sample, and the nature of EI and CI-based fragmentation, the obtained ion clusters for each pair were greatly overlapped, and may have contained as many as 5 different species in EI and 6 different species in CI. As a result, the height of a given $^{12}C^{14}N$-monoisotopic peak might contain contributions from ions less 1 or 2 hydrogens (mostly applicable in CI). Additionally, if very few carbons are present in the native metabolite, the $^{12}C^{14}N$ isotopomer cluster might contain contributions from the $^{13}C^{14}N$-sample ions containing 1 or 2 ^{12}C (due to incomplete incorporation). Likewise, the height of the monoisotopic peak of the $^{13}C^{14}N$ sample always contains contributions from $[M - C_4H_9]^+$ ions having incomplete incorporation of ^{13}C. In all cases, it was assumed that $[M - C_4H_9]^+$ ions with 1 and 2 unincorporated carbons were present in the $^{13}C^{14}N$ sample; the contributions of these species were summed to arrive at the total $^{13}C^{14}N$-$[M - C_4H_9]^+$ ion

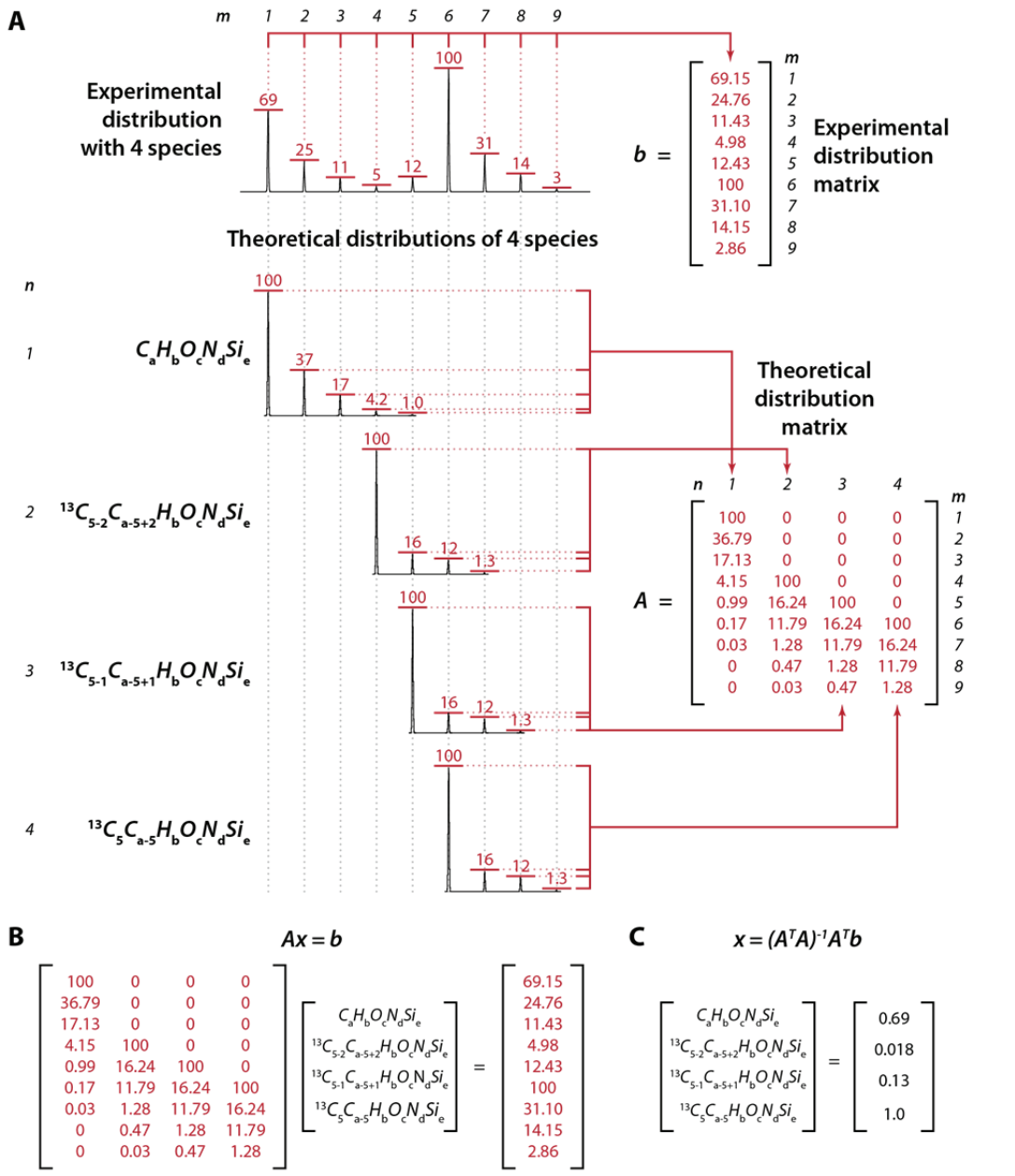


Figure 5.5. Process for determining the approximate contribution of each species in an experimental isotopomer distribution by the method of least squares for over-determined systems. **(A)** An experimental distribution containing 4 distinct species (*n*) has ions at 9 *m/z* values (*m*). The abundance at each *m/z* makes up the *m*-dimensional solution matrix, *b*. Each of the 4 species has a theoretical abundance distribution given by their elemental formulae. The theoretical abundance of the *n* species at all *m m/z* values makes up the *m* x *n*-dimensional matrix *A*. **(B)** Equation relating the matrix of theoretical distributions for the 4 species and the matrix of experimentally measured abundances at the 9 *m/z* values. The combination of these 4 theoretical distributions in some proportion, denoted by the matrix *x*, yields the experimental distribution. **(C)** The formula for the least squares approximation of the matrix *x*. Solving for *x* yields a matrix describing the approximate contribution of each of the *n* species in the experimental cluster.

contribution. On the other hand, $[M - C_4H_9]^+$ ions less 1 or 2 hydrogens were not considered part of the $^{12}C^{14}N$ $[M - C_4H_9]^+$ ion contribution. Once the $[M - C_4H_9]^+$ contribution for both samples was established, we normalized the $^{13}C/^{12}C$ ratio to the ratio obtained in the 1:1 mixture analysis for that compound. The results of this experiment under EI full-scan conditions are shown in black in **Figure 5.6** (for CI full-scan data, see **Figure 5.7**).

It is important to note that, with complex isotopic clusters such as these, knowledge of the elemental formula of the compound is critical to performing relative quantitation. Determining the contribution of each species in the cluster requires that the theoretical isotopomer abundance distribution is known, which in turn requires knowledge of the elemental composition of the peak. Given that the $^{12}C/^{13}C$ pair also serves to assist assignment of elemental composition by signifying the number of carbons present in the analyte, the two samples act as internal standards for each other (in the sense of aiding both identification and quantification). We note that this approach is unlikely to have the accuracy or precision of identification or quantification via an authentic internal reference standard due to the aforementioned problem of incomplete incorporation. With this caveat, the data, while showing slight overestimation of mixing ratios, especially at large dilution ratios (i.e., 20:1), provides sufficient accuracy and reproducibility to detect and make reliable estimates of the relative abundance of analytes between two samples.

One trend in these data is that the accuracy of relative quantitation decreases with lower abundance analytes and as the dilution ratio increases (**Figure 5.7**), a well-known phenomenon in relative quantification.[65] Indeed, in an analogous experiment, Giavalisco and colleagues[21] demonstrated slightly better quantitative accuracy and precision using LC/FT-ICR-MS-based relative quantification of various ratios of mixed $^{12}C^{14}N$ and $^{13}C^{15}N$-labeled *A. thaliana* extracts. One explanation for this discrepancy, besides the obvious differences of instrumentation, is that

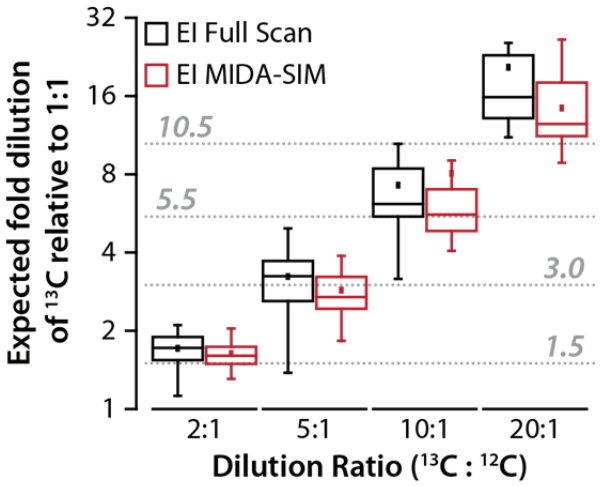


Figure 5.6. Relative quantification with MIDA-SIM. Accuracy and precision of quantification for dilutions of the $^{13}C^{14}N$ sample into the $^{12}C^{14}N$ sample relative to a 1:1 mix. Data from 28 features extracted from EI full scan or EI MIDA-triggered SIM data are shown in black and red, respectively. The improvement of S/N with use of MIDA-SIM enhances quantification. The target ratio at each dilution is denoted by a dotted grey line.

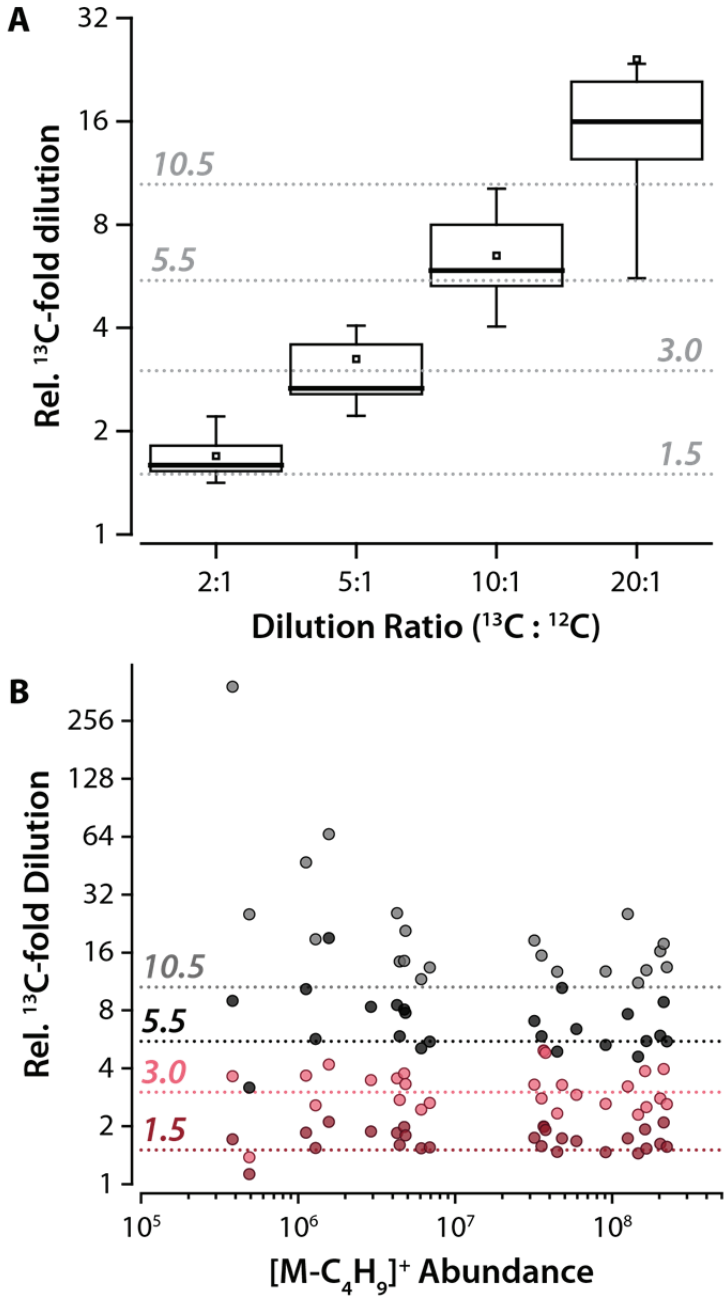


Figure 5.7. (A) Accuracy and precision of quantification for dilutions of the $^{13}C^{14}N$ sample into the $^{12}C^{14}N$ sample relative to a 1:1 mix from 28 features extracted from methane PCI full scan data. The target ratio at each dilution is denoted by a dotted grey line. **(B)** Effect of abundance of the $^{12}C^{14}$-[M - C^4H^9]$^+$ ion of the extracted $^{12}C^{14}N$: $^{13}C^{14}N$ pair on the accuracy and precision of quantification (target dilution ratios for each dilution, i.e., 1.5, 3.0, 5.5, and 10.5, are shown as horizontal dotted lines). The accuracy and precision of quantification decrease (manifesting as over-estimated dilution ratios) for features having low abundance. Data from 28 features extracted from EI full scan analyses.

the abundance of the pseudomolecular ion will have greater S/N when soft ionization techniques (ESI) are employed, which will in turn yield better quantitative accuracy. One method of increasing S/N is to selectively enrich the population of interest in the gas-phase via selected ion monitoring (SIM). Using a wide 20 Th isolation window to capture the entire $^{12}C/^{13}C$ ion pair cluster, we modified the MIDA algorithm to trigger a SIM scan on the algorithm-selected $[M - C_4H_9]^+$ ion, rather than perform MS/MS. We first quantified the S/N enhancement for isolated ions relative to the preceding full scan, finding the average enhancement over ~116k measurements to be 1.8-fold (± 3.4-fold). In accordance with our hypothesis, relative quantification accuracy and precision also improved relative to full-scan quantification, as seen in red in **Figure 5.6**. These data indicate that gas-phase enrichment through the untargeted MIDA-SIM approach can improve relative quantification by reducing some of the quantification bias resulting from insufficient analyte signal. Furthermore, incorporation of a MIDA-SIM scan into the MIDA-MS/MS workflow described above ensures high quality data for both identification and relative quantification purposes for nearly all analytes across multiple samples.

CONCLUSION

Due to the chemical diversity represented by the metabolome (compared to the polymeric natures of the genome and proteome), unknown peak annotation and their subsequent structural elucidation in untargeted GC/MS-based metabolomics remain intractable issues. According to Fiehn, these gaps must be bridged through advances at all stages of the analysis if GC/MS is to realize its full potential within the metabolomics toolbox.[64] Herein, we have detailed the development and use of two such technologies and an analysis workflow that help to address this need in a data-driven manner. Our newly-introduced GC/quadrupole-Orbitrap mass spectrometer provides high resolution, high mass accuracy, and high sensitivity MS data that permit the

reliable use of strict filters for candidate elemental formulae. Additionally, stable-isotope labeling, in conjunction with our Molecular-ion Directed Acquisition (MIDA) approach for MS/MS, guarantees not only information-rich MS/MS spectra for intact, or nearly-intact ionic species, but also immediate readout of the number of carbons and nitrogens present in each precursor and product ion species. Taken together, these data-driven approaches permitted unambiguous assignment of elemental composition to all queried MS features in this study. While we did not employ the standard methods of chromatographic deconvolution, retention index correlation, and spectral database searching, our technology and analysis workflow are complementary to all existing approaches and can be easily incorporated into any standard workflow to further advance the tools available to the GC/MS-based metabolomics community.

REFERENCES

(1) Fiehn, O. Metabolomics--the link between genotypes and phenotypes. *Plant Mol Biol* **2002**, *48*, 155-171.

(2) Birkemeyer, C.; Luedemann, A.; Wagner, C.; Erban, A.; Kopka, J. Metabolome analysis: the potential of in vivo labeling with stable isotopes for metabolite profiling. *Trends Biotechnol* **2005**, *23*, 28-33.

(3) Kind, T.; Fiehn, O. Advances in structure elucidation of small molecules using mass spectrometry. *Bioanal Rev* **2010**, *2*, 23-60.

(4) Xu, F.; Zou, L.; Ong, C. N. Experiment-originated variations, and multi-peak and multi-origination phenomena in derivatization-based GC-MS metabolomics. *TrAC Trends in Analytical Chemistry* **2010**, *29*, 269-280.

(5) Jankevics, A.; Merlo, M. E.; de Vries, M.; Vonk, R. J.; Takano, E.; Breitling, R. Separating the wheat from the chaff: a prioritisation pipeline for the analysis of metabolomics datasets. *Metabolomics* **2012**, *8*, 29-36.

(6) Bueschl, C.; Krska, R.; Kluger, B.; Schuhmacher, R. Isotopic labeling-assisted metabolomics using LC-MS. *Anal Bioanal Chem* **2012**.

(7) Keller, B. O.; Sui, J.; Young, A. B.; Whittal, R. M. Interferences and contaminants encountered in modern mass spectrometry. *Anal. Chim. Acta* **2008**, *627*, 71-81.

(8) Hummel, J.; Strehmel, N.; Selbig, J.; Walther, D.; Kopka, J. Decision tree supported substructure prediction of metabolites from GC-MS profiles. *Metabolomics* **2010**, *6*, 322-333.

(9) Scheltema, R. A.; Decuypere, S.; Dujardin, J. C.; Watson, D. G.; Jansen, R. C.; Breitling, R. Simple data-reduction method for high-resolution LC–MS data in metabolomics. *Bioanalysis* **2009**, *1*, 1551-1557.

(10) McLafferty, F. W.; Turecek, F. *Interpretation of mass spectra*; 4th ed.; Univ Science Books: Sausalito, CA, 1993.

(11) Richter, W. J.; Schwarz, H. Chemical Ionization—A Mass-Spectrometric Analytical Procedure of Rapidly Increasing Importance. *Angewandte Chemie International Edition in English* **1978**, *17*, 424-439.

(12) Kind, T.; Wohlgemuth, G.; Lee do, Y.; Lu, Y.; Palazoglu, M.; Shahbaz, S.; Fiehn, O. FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal Chem* **2009**, *81*, 10038-10048.

(13) Kind, T.; Fiehn, O. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. *Bmc Bioinformatics* **2006**, *7*, -.

(14) Kind, T.; Fiehn, O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *Bmc Bioinformatics* **2007**, *8*, -.

(15) Fiehn, O.; Kopka, J.; Trethewey, R. N.; Willmitzer, L. Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal. Chem.* **2000**, *72*, 3573-3580.

(16) Konermann, L.; Pan, J.; Liu, Y. H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem Soc Rev* **2011**, *40*, 1224-1234.

(17) Bantscheff, M.; Lemeer, S.; Savitski, M. M.; Kuster, B. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem* **2012**, *404*, 939-965.

(18) Engelsberger, W. R.; Erban, A.; Kopka, J.; Schulze, W. X. Metabolic labeling of plant cell cultures with K(15)NO3 as a tool for quantitative analysis of proteins and metabolites. *Plant Methods* **2006**, *2*, 14.

(19) Kluger, B.; Bueschl, C.; Lemmens, M.; Berthiller, F.; Haubl, G.; Jaunecker, G.; Adam, G.; Krska, R.; Schuhmacher, R. Stable isotopic labelling-assisted untargeted metabolic profiling reveals novel conjugates of the mycotoxin deoxynivalenol in wheat. *Anal Bioanal Chem* **2012**.

(20) Giavalisco, P.; Li, Y.; Matthes, A.; Eckhardt, A.; Hubberten, H. M.; Hesse, H.; Segu, S.; Hummel, J.; Kohl, K.; Willmitzer, L. Elemental formula annotation of polar and lipophilic metabolites using (13) C, (15) N and (34) S isotope labelling, in combination with high-resolution mass spectrometry. *Plant J* **2011**, *68*, 364-376.

(21) Giavalisco, P.; Kohl, K.; Hummel, J.; Seiwert, B.; Willmitzer, L. 13C isotope-labeled metabolomes allowing for improved compound annotation and relative quantification in liquid chromatography-mass spectrometry-based metabolomic research. *Anal Chem* **2009**, *81*, 6546-6551.

(22) Giavalisco, P.; Hummel, J.; Lisec, J.; Inostroza, A. C.; Catchpole, G.; Willmitzer, L. High-resolution direct infusion-based mass spectrometry in combination with whole 13C metabolome isotope labeling allows unambiguous assignment of chemical sum formulas. *Anal Chem* **2008**, *80*, 9417-9425.

(23) Hegeman, A. D.; Schulte, C. F.; Cui, Q.; Lewis, I. A.; Huttlin, E. L.; Eghbalnia, H.; Harms, A. C.; Ulrich, E. L.; Markley, J. L.; Sussman, M. R. Stable isotope assisted assignment of elemental compositions for metabolomics. *Anal Chem* **2007**, *79*, 6912-6921.

(24) Rodgers, R. P.; Blumer, E. N.; Hendrickson, C. L.; Marshall, A. G. Stable isotope incorporation triples the upper mass limit for determination of elemental composition by accurate mass measurement. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 835-840.

(25) Wu, L.; Mashego, M. R.; van Dam, J. C.; Proell, A. M.; Vinke, J. L.; Ras, C.; van Winden, W. A.; van Gulik, W. M.; Heijnen, J. J. Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly 13C-labeled cell extracts as internal standards. *Anal Biochem* **2005**, *336*, 164-171.

(26) Mashego, M. R.; Wu, L.; Van Dam, J. C.; Ras, C.; Vinke, J. L.; Van Winden, W. A.; Van Gulik, W. M.; Heijnen, J. J. MIRACLE: mass isotopomer ratio analysis of U-13C-labeled extracts. A new method for accurate quantification of changes in concentrations of intracellular metabolites. *Biotechnol Bioeng* **2004**, *85*, 620-628.

(27) Herebian, D.; Kupper, U.; Schomburg, D.; Marner, F. J. In vivo labeling with stable isotopes as a tool for the identification of unidentified peaks in the metabolome analysis of Corynebacterium glutamicum by GC/MS. *Biol Chem* **2007**, *388*, 865-871.

(28) Cascante, M.; Marin, S. Metabolomics and fluxomics approaches. *Essays Biochem* **2008**, *45*, 67-81.

(29) Price, N. P. J. Acylic Sugar Derivatives for GC/MS Analysis of 13C-Enrichment during Carbohydrate Metabolism. *Anal. Chem.* **2004**, *76*, 6566-6574.

(30) Furch, T.; Preusse, M.; Tomasch, J.; Zech, H.; Wagner-Dobler, I.; Rabus, R.; Wittmann, C. Metabolic fluxes in the central carbon metabolism of Dinoroseobacter shibae and Phaeobacter gallaeciensis, two members of the marine Roseobacter clade. *BMC Microbiology* **2009**, *9*, 209.

(31) Fürch, T.; Hollmann, R.; Wittmann, C.; Wang, W.; Deckwer, W.-D. Comparative study on central metabolic fluxes of <i>Bacillus megaterium</i> strains in continuous culture using ¹³C labelled substrates. *Bioprocess and Biosystems Engineering* **2007**, *30*, 47-59.

(32) Fürch, T.; Wittmann, C.; Wang, W.; Franco-Lara, E.; Jahn, D.; Deckwer, W.-D. Effect of different carbon sources on central metabolic fluxes and the recombinant production of a hydrolase from Thermobifida fusca in Bacillus megaterium. *J Biotechnol* **2007**, *132*, 385-394.

(33) Hiller, K.; Metallo, C. M.; Kelleher, J. K.; Stephanopoulos, G. Nontargeted Elucidation of Metabolic Pathways Using Stable-Isotope Tracers and Mass Spectrometry. *Anal. Chem.* **2010**, *82*, 6621-6628.

(34) Abate, S.; Ahn, Y. G.; Kind, T.; Cataldi, T. R. I.; Fiehn, O. Determination of elemental compositions by gas chromatography/time-of-flight mass spectrometry using chemical and electron ionization. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 1172-1180.

(35) Amirav, A.; Gordin, A.; Poliak, M.; Fialkov, A. B. Gas chromatography-mass spectrometry with supersonic molecular beams. *J Mass Spectrom* **2008**, *43*, 141-163.

(36) Birkemeyer, C.; Kolasa, A.; Kopka, J. Comprehensive chemical derivatization for gas chromatography-mass spectrometry-based multi-targeted profiling of the major phytohormones. *J Chromatogr A* **2003**, *993*, 89-102.

(37) Rubiolo, P.; Sgorbini, B.; Liberto, E.; Cordero, C.; Bicchi, C. Essential oils and volatiles: sample preparation and analysis. A review. *Flavour and Fragrance Journal* **2010**, *25*, 282-290.

(38) Huang, M.; Sanchez-Moreiras, A. M.; Abel, C.; Sohrabi, R.; Lee, S.; Gershenzon, J.; Tholl, D. The major volatile organic compound emitted from Arabidopsis thaliana flowers, the sesquiterpene (E)-beta-caryophyllene, is a defense against a bacterial pathogen. *New Phytol* **2012**, *193*, 997-1008.

(39) Schummer, C.; Delhomme, O.; Appenzeller, B. M.; Wennig, R.; Millet, M. Comparison of MTBSTFA and BSTFA in derivatization reactions of polar compounds prior to GC/MS analysis. *Talanta* **2009**, *77*, 1473-1482.

(40) Zaikin, V.; Halket, J. M. *A handbook of derivatives for mass spectrometry*; IM Publications, 2009.

(41) Halket, J. M.; Waterman, D.; Przyborowska, A. M.; Patel, R. K. P.; Fraser, P. D.; Bramley, P. M. Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J. Exp. Bot.* **2005**, *56*, 219-243.

(42) Baker, A.; Dodd, C. D.; Parsons, R. Identification of amino compounds synthesized and translocated in symbiotic Parasponia. *Plant, Cell & Environment* **1996**, *19*, 1249-1260.

(43) Wood, P. L.; Khan, M. A.; Moskal, J. R. Neurochemical analysis of amino acids, polyamines and carboxylic acids: GC-MS quantitation of tBDMS derivatives using ammonia positive chemical ionization. *J Chromatogr B Analyt Technol Biomed Life Sci* **2006**, *831*, 313-319.

(44) Van Pelt, C. K.; Brenna, J. T. Acetonitrile Chemical Ionization Tandem Mass Spectrometry To Locate Double Bonds in Polyunsaturated Fatty Acid Methyl Esters. *Anal. Chem.* **1999**, *71*, 1981-1989.

(45) Brenna, J. T. In *Lipid Analysis and Lipidomics*; AOCS Publishing: 2006; pp 157-172.

(46) Peterson, A. C.; Hauschild, J.-P.; Quarmby, S. T.; Krumwiede, D.; Lange, O.; Grosse-Coosmann, F.; Horning, S.; Westphall, M. S.; Coon, J. J.; Griep-Raming, J. Development of a GC/quadrupole-Orbitrap mass spectrometer, Part I: Design and characterization. *Anal Chem* **2012**, *[In press]*.

(47) Fiehn, O. In *Arabidopsis Protocols*; 2nd ed.; Salinas, J., Sanchez-Serrano, J. J., Eds.; Humana Press: Totowa, NJ, 2006; Vol. 323; pp 439-447.

(48) Werner, R. A.; Brand, W. A. Referencing strategies and techniques in stable isotope ratio analysis. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 501-519.

(49) Anton, H. *Elementary Linear Algebra*; 9th ed.; John Wiley and Sons, Inc.: Hoboken, 2005.

(50) Mann, M.; Hendrickson, R. C.; Pandey, A. Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem* **2001**, *70*, 437-473.

(51) Bailey, D. J.; Rose, C. M.; McAlister, G. C.; Brumbaugh, J.; Yu, P.; Wenger, C. D.; Westphall, M. S.; Thomson, J. A.; Coon, J. J. Instant Spectral Assignment for Advanced Decision Tree-Driven Mass Spectrometry. *Proc Natl Acad Sci U S A* **2012**, *In press.*

(52) Graumann, J.; Scheltema, R. A.; Zhang, Y.; Cox, J.; Mann, M. A Framework for Intelligent Data Acquisition and Real-Time Database Searching for Shotgun Proteomics. *Mol Cell Proteomics* **2012**, *11*, mcp.M111.013185.

(53) Swaney, D. L.; McAlister, G. C.; Coon, J. J. Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat Meth* **2008**, *5*, 959-964.

(54) Castro-Perez, J.; Plumb, R.; Liang, L.; Yang, E. A high-throughput liquid chromatography/tandem mass spectrometry method for screening glutathione conjugates using exact mass neutral loss acquisition. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 798-804.

(55) Bloomfield, N.; Leblanc, Y.; Office, U. P., Ed. USA, 2009.

(56) Ma, L.; Wen, B.; Ruan, Q.; Zhu, M. Rapid Screening of Glutathione-Trapped Reactive Metabolites by Linear Ion Trap Mass Spectrometry with Isotope Pattern-Dependent Scanning and Postacquisition Data Mining. *Chemical Research in Toxicology* **2008**, *21*, 1477-1483.

(57) Lim, H.-K.; Chen, J.; Cook, K.; Sensenhauser, C.; Silva, J.; Evans, D. C. A generic method to detect electrophilic intermediates using isotopic pattern triggered data-dependent high-resolution accurate mass spectrometry. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 1295-1311.

(58) Cuyckens, F.; Balcaen, L.; De Wolf, K.; De Samber, B.; Van Looveren, C.; Hurkmans, R.; Vanhaecke, F. Use of the bromine isotope ratio in HPLC-ICP-MS and HPLC-ESI-MS analysis of a new drug in development. *Analytical and Bioanalytical Chemistry* **2008**, *390*, 1717-1729.

(59) Ragsdale, J.; Conoley, M. *A demonstration of the use of data-dependent scanning for characterization and MS/MS method development with a bench-top external source ion trap mass spectrometer*, ThermoFinnigan Application Note 9175. Thermo Electron, Mountain View, CA, USA, 2002.

(60) Trolez, Y.; Guillemin, J. C. Synthesis and characterization of 2,4-pentadiynenitrile--a key compound in space science. *Angew Chem Int Ed Engl* **2005**, *44*, 7224-7226.

(61) Eggeling, L.; Bott, M. *Handbook of Corynebacterium glutamicum*; CRC, 2005.

(62) Herebian, D.; Hanisch, B.; Marner, F. J. Strategies for gathering structural information on unknown peaks in the GC/MS analysis of Corynebacterium glutamicum cell extracts. *Metabolomics* **2005**, *1*, 317-324.

(63) Fiehn, O.; Kopka, J.; Dormann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. Metabolite profiling for plant functional genomics. *Nat Biotechnol* **2000**, *18*, 1157-1161.

(64) Fiehn, O. Extending the breadth of metabolite profiling by gas chromatography coupled to mass spectrometry. *Trends Analyt Chem* **2008**, *27*, 261-269.

(65) Carrillo, B.; Yanofsky, C.; Boismenu, D.; Latterich, M.; Kearney, R. E. In *American Society for Mass Spectrometry* Seattle, WA, 2006.

CHAPTER 6

Appendix to Chapter 2:

Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics

SUPPLEMENTAL RESULTS AND DISCUSSION

6.1. PRM parameter evaluation

To determine the optimal isolation parameters for PRM analyses, we performed triplicate data-dependent MS analyses of a tryptic yeast digest using isolation widths ranging from ±0.2 to ±1 Th on the QqOrbi (**Figure 6.1**). From these experiments, we elected to employ two isolation conditions in our PRM characterization. The first isolation width, ±0.2 Th, was utilized because it represents the highest-achievable quadrupole isolation resolution (*i.e.*, highest selectivity) on the QqOrbi platform. The second isolation width, ±1 Th, was selected since this isolation width netted the greatest average number of peptide-spectrum matches (PSMs), was comparable with isolation conditions down to ±0.4 Th, and was similar to isolation conditions used for SRM on the QqQ. In similar experiments, we determined AGC target values and maximum injection times for PRM based upon the analysis conditions resulting in the greatest number of PSMs. The AGC target value range examined for MS/MS scans spanned $1x10^5$ to $1x10^6$; maximum injection times were examined in the range of 30 to 250 ms. For our characterization of PRM, an MS/MS AGC target of $2x10^5$ and maximum injection time of 120 ms were selected.

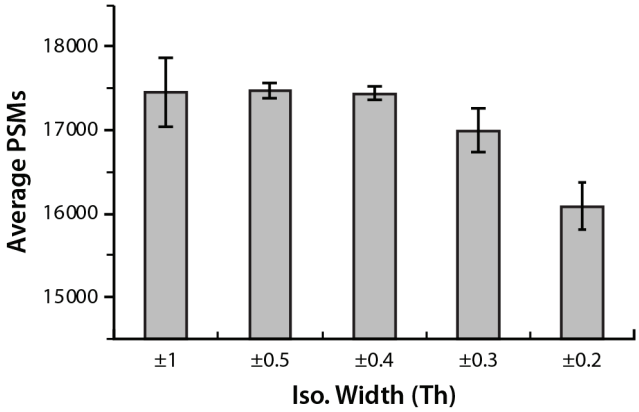


Figure 6.1. Average number of peptide-spectrum matches for triplicate data-dependent (top 10) analysis of a tryptic yeast digest at isolation widths from ±0.2 to ±1 Th. These data were used to determine the optimum parameters for the characterization of PRM.

6.2. Detection criteria for PRM and SRM.

Tables 6.1A-D catalog the maximum XSC scores measured for each replicate, concentration, and PRM isolation width, as well as demonstrate some aspects of the data which were taken into account in our data analysis. In several cases, especially in neat analyses, XSCs at the lowest concentrations for a given peptide showed no noise (score of zero for the entirety of the XSC) but the first detectable concentration did not meet the acceptance threshold. These non-threshold meeting concentrations were accepted, along with the first threshold-meeting concentration, and used in batch data comparisons (*e.g.,* all neat versus all matrix-containing data) but not in comparisons of individual experiments. Two peptides, TVFSSTQLcVLNDr (#23) and SGWTcTQPGGr (#4), under matrix-containing conditions, had XSCs that exceeded threshold at all 6 peptide concentrations (from 200 nM to 2 pM) but were only detected at the top 4-5 concentrations (200 nM to 200 pM or 20 pM, respectively) in the equivalent neat experiments. These concentrations were excluded, despite exceeding the threshold, due the high likelihood that they contained spurious signals from co-isolated peptide or matrix ions.

6.3. Empirical comparison of QqQ SRM and QqOrbi PRM.

As discussed in the Results and Discussion section on QqQ SRM measurement precision, dynamic range, and linearity, we believe that neat SRM measurement precision is artificially depressed due to the inability of the SRM detection criteria to exclude data skewed by small amounts of chemical interference at the lowest detected concentrations (**Chapter 2 Figure 2.7**). In the comparison with QqOrbi PRM, this supposition results in the unlikely situation of PRM demonstrating similar measurement precision to SRM in the neat case, but being outperformed by SRM in the matrix-containing case. We hypothesize, given appropriately-sensitive detection criteria, that SRM would exhibit superior measurement precision in the neat case as well. To

support this hypothesis, we extracted only the neat SRM data that had corresponding matrix-containing measurements from the SRM dataset. This extracted set of neat SRM data is thus truncated to the dynamic range quantified under matrix-containing conditions and therefore excludes the problematic lower concentrations believed to be skewed by chemical noise. Comparing this truncated neat SRM data to neat PRM data (after pairing by concentration and peptide), we then found that neat SRM exhibits superior measurement precision compared to ±0.2 Th neat PRM dataset, and to both PRM datasets when the data were not paired (**Table 6.6B**). This analysis suggests that, if appropriately-sensitive SRM detection criteria are utilized, SRM possesses measurement precision superior to PRM under both neat and matrix-containing conditions.

6.4. QqOrbi SIM detection criteria.

Positive target peptide detection in SIM was based on the following criteria: the presence of the monoisotopic mass of the intact target peptide in at least 2 of 3 replicates within ±1 min of expected retention time (based on the average retention time observed in 200 nM analyses; **Chapter 2 Figure 2.2**), chromatographic signal-to-noise greater than 3, and mass error within ±10 ppm for all precursor-derived ions observed. We did not require the observation of isotopes of the precursor or that isotopes, when present, were of the appropriate abundance relative to the monoisotopic peak.

Figure 6.2A shows representative extracted ion chromatograms (XICs, ±5 ppm) and single-scan spectra for peptide SFEGTDYGk (#12) isolated at ±2 Th under neat and matrix-containing conditions. In the absence of matrix (left), this peptide is observed at all six concentrations with very few interfering ions. At the lowest two concentrations, 20 and 2 pM, nearby chemical interferences are easily resolved from the ions of interest due to the high

resolution of the mass analysis. In the presence of matrix (right), the peptide is observed over a considerably-reduced three orders-of-magnitude in concentration. While analytically, SIM is not amenable to analysis of samples containing high levels of matrix, this expected result demonstrates an important characteristic of high resolution and accurate mass MS data: the combination of high mass accuracy and high resolution provides sufficient selectivity to make the determination of the presence or absence of a target species nearly a binary event. In other words, in many cases, a target species is either present and correctly identified, 1, or absent, 0, with marginal, or possibly correct, cases (that is, between 0 and 1) occurring very rarely. This binary property, a characteristic of PRM data as well, results in data that are easily processed via automated detection algorithms. The paucity of co-isolated chemical noise falling within ±5 ppm of the target ion at the lowest three concentrations (2-200 pM) lies in stark contrast to the large amount of chemical noise present if the same data were treated as having low mass accuracy with mass measurements reliable to ±250 ppm (**Figure 6.2B**). At low mass accuracy, the binary quality of the data, as described above, is lost.

6.5. QqOrbi SIM measurement precision, dynamic range, and linearity.

As in our analysis of QqOrbi PRM in the main text, we assessed the effect of matrix and isolation width on SIM measurement precision, dynamic range, and linearity. SIM measurement precision, unlike PRM precision, was significantly lower in the presence of matrix than in its absence (**Table 6.2A**). Similar to PRM however, SIM measurement precision was independent of isolation width, but did show a significant dependence on target peptide concentration. The SIM data exhibited two concentrations where significant deviations in precision were present: the first, due exclusively to neat experiments at the lowest concentration studied, 2 pM, and the second in the mid-concentration range due to the contribution of matrix-containing experiments,

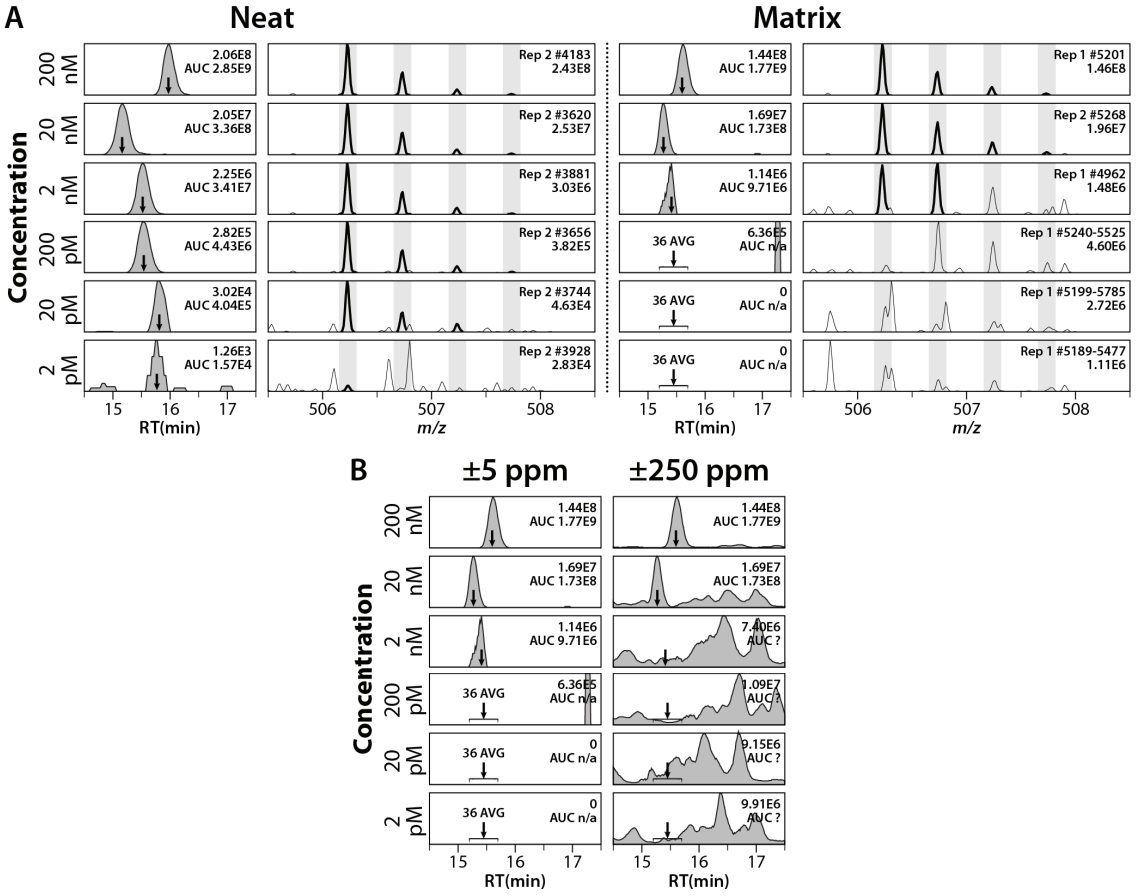


Figure 6.2. (A) Extracted SIM ion chromatograms (XIC, ±5 ppm; 7-point boxcar smoothed) and single-scan (unless otherwise noted) spectra for peptide SFEGTDYGk (#12) under neat (left) and matrix-containing (right) conditions from 2 pM to 200 nM (corresponding to 2 amol to 200 fmol on column with 1 µL injected in every experiment). Target peptide-associated signals in each spectrum are highlighted (in bold stroke). Black arrows in XICs indicate the retention time at which the associated spectrum was acquired. In matrix-containing conditions from 2 pM to 200 pM, spectra shown are an average of 36 scans as designated by the ranged labeled in the associated XIC. **(B)** Comparison of relative noise levels of high and low mass accuracy XICs for peptide SFEGTDYGk (#12) isolated at ±2 Th under matrix-containing conditions from 2 pM to 200 nM. Low resolution/mass accuracy XICs (±250 ppm, right) were generated from the same data as the high resolution/mass accuracy XICs (±5 ppm, left) by simply changing the extraction mass tolerance.

which typically yielded results only at the top 3 concentrations (2, 20, and 200 nM) (**Table 6.2B**). In conjunction with our PRM results, these results demonstrate that the precision of replicate SIM measurements, compared to PRM on the same platform, is significantly adversely affected by matrix interferences and low quantities of peptide loaded on column.

The quantified dynamic range observed in SIM experiments fell into two categories much like measurement precision did above: a 10^2 dynamic range (42%) category arising almost exclusively from matrix-containing experiments (93%), and a $\geq 10^4$ dynamic range category (46%) made up of only neat experiments. All measurements over a dynamic range of 10^5 resulted from neat SIM experiments (**Chapter 2 Figure 2.6**). Matrix interferences resulted in a ~2.5 order-of-magnitude truncation in the quantifiable dynamic range ($10^{1.9}$ versus $10^{4.4}$, matrix versus neat, for both isolations taken together). Furthermore, tighter isolation widths did not mitigate the extent of sensitivity depression: the mean loss of sensitivity due to matrix interferences at ±1 Th was 2.5 versus 2.4 orders-of-magnitude at ± 2 Th (**Tables 6.3A-B**). All things being equal (*i.e.*, comparing ±1 Th isolation widths between SIM and PRM), in the presence of matrix, PRM resulted in a significantly greater quantifiable dynamic range compared to SIM ($10^{2.7}$ versus $10^{2.0}$) and was significantly less susceptible to matrix effects (0.6 versus 2.5 orders-of-magnitude of lost sensitivity; **Table 6.3B**). Under neat conditions, however, PRM was significantly less sensitive than SIM ($10^{3.0}$ versus $10^{4.4}$).

In SIM experiments, like in PRM, the linearity of measurement response (accuracy) was not significantly influenced by isolation width or the presence of matrix: mean %RSDs employing ±1 and ±2 Th isolation widths, respectively, were 34.0 and 32.3%, neat, and 36.1 and 34.1%, matrix (**Tables 6.4A-B**). Linearity improved and significant trends based on presence of matrix in SIM experiments emerged upon exclusion of the lowest detected concentration (associated with lower measurement precision) from each experiment (**Tables 6.4C-F**).

Calculation of adjusted %RSDs to account for the disparate concentration ranges quantified in neat and matrix-containing SIM experiments revealed that, while linearity in PRM was unaffected by the presence of matrix, the presence of matrix in SIM experiments had a significantly detrimental effect on the linearity of measurement response: at ±1 Th, 7.9% RSD neat versus 17.5% RSD matrix, and at ±2 Th, 8.6% RSD neat versus 17.2% RSD matrix (**Tables 6.4G-H**). Isolation width did not cause significant deviations in linearity using adjusted %RSDs. Thus, in addition to yielding a greater quantifiable dynamic range and possessing superior resistance to matrix effects, by this analysis, PRM is also correlated with greater measurement accuracy/linearity in the presence of matrix interference compared to SIM.

The observation that SIM was more susceptible to matrix interference than PRM was expected given the nature of the two methods and, to a lesser extent, the wider isolation widths employed in SIM experiments (used to observe multiple isotopes from the target peptide). SIM, much like SRM, results in spectra with information spanning only a narrow mass-to-charge range, much less information than is generated in PRM. Thus, the presence of just one matrix ion in a SIM spectrum can be much more detrimental to data interpretation than the presence of a high level of matrix in a PRM spectrum, even with the high mass accuracy (<5 ppm mass error) and moderately high mass resolution employed here (17500 at *m/z* 200). While a higher resolution mass analysis might improve the selectivity of SIM by resolving matrix ions away from the ions of interest (of course, at the high cost of decreased instrument duty cycle), PRM will likely always produce superior results in the presence of matrix since the probability of matrix ions interfering with all, or even the majority of, PRM product ion channels is much less than in a SIM spectrum.

A potential concern of PRM is that it is less sensitive than SIM under neat/low background conditions. Fundamentally, PRM, which requires ions to be transported slightly

further (in a QqOrbi) and splits the available charge into numerous product ion channels, will result in more ion losses and lower signal-to-noise spectra compared to SIM where the intact species is analyzed directly. Despite this fundamental disadvantage, the dynamic range and sensitivity of PRM can be tailored to a particular application through experiment-specific optimization of AGC targets, maximum injection times, isolation widths, and, to a limited extent, mass analysis resolution.

6.6. Empirical comparison of QqOrbi SIM and QqQ SRM.

As in our comparison of SRM and PRM, we only considered the fourteen peptides analyzed in both datasets, paired by peptide and concentration, in our comparison of SIM and SRM. **Table 6.6C** summarizes the results of the comparison of SIM and SRM dynamic range and linearity. Under neat conditions, SIM, at both isolation widths, quantified over a wider dynamic range and demonstrated greater measurement accuracy (using the adjusted %RSD metric) than did SRM. However, under matrix conditions, SRM demonstrated greater linearity compared to SIM due to high levels of matrix interference. Despite SIM's susceptibility to matrix interferences, however, SRM did not outperform SIM under matrix conditions with regard to dynamic range: SIM quantified over a dynamic range of, on average, $10^{2.0}$ and $10^{1.9}$ at ±1 and ±2 Th, respectively – statistically no different from SRM. This result suggests that, when high levels of matrix background are present, a single stage of isolation in combination with highly resolved data (SIM) is statistically equivalent, in terms of sensitivity, to two stages of isolation, dissociation, and low resolution mass analysis in combination (SRM).

Table 6.1A. Extracted score chromatogram maxima for all neat PRM ±1 Th experiments.

#	Peptide (Threshold)	# Orders	200 nM			20 nM			2 nM			200 pM			20 pM			2 pM		
			1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
17	AETLVQAr (8)	4	30.25	30.25	30.25	30.25	30.25	30.25	30.25	30.25	30.25	22.5	29.25	22.25	12.5	15.5	18.5	0	0	0
14	AFSQNSVLIk (10)	3	46.25	46.25	46.25	46.25	46.25	46.25	46.25	41.25	46.25	32.5	25.5	32.5	0	0	5	0	0	0
21	APcQAGDLr (9)	4	41.25	40	40.25	41.25	41.25	39	33.25	41.25	41.25	29.25	36.25	30.25	9	14.75	14	0	0	0
3	ARPAcVDAr (9)	1	18.25	19.5	24.5	10.75	10.75	10.75	2	0	2	0	0	0	0	0	0	0	0	0
8	DIQFGSQIk (9)	3	30.25	30.25	30.25	30.25	30.25	30.25	30.25	29.25	29.25	26.25	25.5	22.25	0	6	0	0	0	0
5	EGQLAAGTcEIVTLDr (16)	3	102	98	97.75	101.25	102.5	99.75	69	81	85	33	31.75	30.75	0	0	0	1	0	0
18	FLNPEWk (7)	4	22.25	23.75	22.25	22.25	22.25	22.25	22.25	22.25	22.25	22.25	19.25	20.5	11	9	7	0	0	0
11	FSDLTEEEFr (10)	4	48.5	48.5	47.25	48.5	48.5	48.5	48.5	48.5	48.5	46.25	46.25	46.25	14	21	13	0	0	0
24	FSEVSADk (8)	4	30.75	29.25	29.25	29.25	30.25	30.25	29.25	29.25	29.25	29.25	28.5	29.25	11	11.5	19	0	0	0
16	GALDGEAPr (9)	4	30.25	30.25	31.5	30.25	30.25	30.25	30.25	30.25	30.25	30.25	29.25	30.25	14	11	11	0	0	0
25	GLYEGTGr (8)	4	22.25	23.25	22.25	23.25	22.25	23.25	23.25	22.25	22.25	19.5	20.25	22.25	15	11	9	0	0	0
1	GVSAFSTWEk (10)	3/4	38.25	38.25	39.5	38.25	39.5	38.25	38.25	38.25	39.5	35.5	36.25	37.25	6.5	7.5	8.5	0	0	0
2	HFLTLAPIk (9)	3/4	45	43	45	44.25	45	44	43.25	44.25	42	33	35.25	27.25	3.5	3	4	0	0	0
9	HTGTPLGDIPYGk (13)	3	86.25	94.5	87	89	89	86	57.75	49.75	60.75	11.5	25.75	21.5	0	0	0	0	0	0
19	LEQNPEESQDIk (12)	3/4	72.5	68.25	68.25	69.5	68.25	70	59.75	57.25	68.25	48.25	40.25	47.25	8	8	8	0	0	0
6	LWSLAEIATSDLk (13)	3/4	80.75	80.75	81.75	80.75	79.25	80.75	52.25	59.25	63.25	38.25	57.25	50.25	0	11	11	0	0	0
10	NSWGTDWGEk (10)	3/4	46.5	46.5	46.5	46.5	46.5	46.5	46.5	46.5	46.5	37.5	35.5	36.75	8	7	8	0	0	0
20	SEASSSPPVVTSSSHSr (17)	4	79.5	78.5	76.75	80.75	78.5	91	74.5	89.5	75.25	62.25	56.25	62.25	25.5	25.5	17.5	11	0	0
7	SEDEDEDEGDATr (13)	3	72.25	73.5	71.25	55.25	25	49.25	0.5	0	62.25	33.5	1.5	39.25	2.5	2.5	7	2	1.5	0
12	SFEGTDYGk (9)	3/4	37.25	37.25	37.25	37.25	37.25	37.25	37.25	37.25	36.5	28.5	28.5	28.5	13.5	7.5	7	0	0	0
4	SGWTcTQPGGr (11)	4	57.25	50.25	50.25	50.25	49	57.25	47.25	48.5	48.5	31.75	33.25	31	19	12	12	8	0	2.5
13	TLNGIQLAr (9)	4	38.25	38.25	38.25	38.25	38.25	37.5	38.25	38.25	38.25	28.5	28.5	25.5	13	14.5	13.5	7	0	4
23	TVFSSTQLcVLNDr (14)	3/4	87	85.25	89	87.25	85.5	86.75	81.5	79.25	80.25	50.5	72.25	66.25	7	6	0.5	0	0	0
22	TWFQNQr (7)	4	22.25	22.25	22.25	22.25	22.25	22.25	22.25	22.25	21.5	19.5	21.5	19.5	9.5	9	9	0	0	0
15	WPGYLNGGr (9)	3/4	37.25	37.25	37.25	37.25	37.25	37.25	37.25	37.25	37.25	33	33	34.5	7	13	8	0	0	0

Does not meet threshold, reject concentration
Meets threshold, but reject concentration based on high background levels
Does not meet threshold, but accept concentration based on poor peptide retention
Does not meet threshold, but accept concentration based on zero-noise argument
Meets threshold, accept concentration

Table 6.1B. Extracted score chromatogram maxima for all matrix PRM ±1 Th experiments.

#	Peptide (Threshold)	# Orders	200 nM			20 nM			2 nM			200 pM			20 pM			2 pM		
			1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
17	AETLVQAr (8)	3	30.25	30.25	30.25	29.25	29.25	30.25	22.25	22.25	22.25	20.5	14	15	1.5	1.5	2	1.5	1.5	0.5
14	AFSQNSVLIk (10)	2	39.25	37.25	37.25	37.25	36.5	36.5	32.5	25.5	27.5	0	0	8	0	0	4	0	0	1.5
21	APcQAGDLr (9)	4	42	43.25	42	41.25	39.5	40	37	39.5	38.25	27.25	18.25	21.25	3.5	10.5	10.5	3	6.5	1.75
3	ARPAcVDAr (9)	1	26.5	30	23.25	12.25	11.25	12.25	0	2	3.75	0	0	4	3	0.25	0.25	0	3	0
8	DIQFGSQIk (9)	3	29.25	30.25	29.25	29.25	29.25	29.25	26.25	26.25	27.25	11	6	11.5	0.5	0.5	0.5	0.5	0.5	0.5
5	EGQLAAGTcEIVTLDr (16)	3	100.25	97.5	100.5	98	98.75	99.75	84.5	78.5	88.5	34.75	34.75	23.75	9	12	6	6.5	6.5	5
18	FLNPEWk (7)	3	22.25	22.25	22.25	22.25	21.5	22.25	21.5	20.5	21.5	9	5	9	3	6	3	0	3	3
11	FSDLTEEEFr (10)	3	47.25	46.25	47.25	47.25	46.25	48.25	46.25	46.25	46.25	32.5	33.25	10.5	6	9	15	7	1	1.5
24	FSEVSADk (8)	2	29.25	29.25	29.25	29.25	29.25	25.25	21	24	19.5	12	5	5	4	4	5	4	5	4
16	GALDGEAPr (9)	3	30.25	30.25	30.25	30.25	30.25	30.25	29.25	29.25	30.25	21.25	25.25	24.5	8.5	10	3.5	7.25	6.5	3.25
25	GLYEGTGr (8)	3	22.25	22.25	22.25	22.25	22.25	22.25	19.25	19.25	22.25	17.5	17.5	12.5	8.5	6.5	4.25	5	3.5	6
1	GVSAFSTWEk (10)	3	37.25	38.25	38.25	38.25	38.25	38.25	38.25	39	31.25	12.5	16.25	20.25	6.25	4.75	7.75	6.25	5.25	3.5
2	HFLTLAPIk (9)	2	43.25	38.25	41.25	41	41.25	41.25	31.75	27	26.5	4.5	4	8	0.25	0.25	0	0.25	0	0.25
9	HTGTPLGDIPYGk (13)	2	86	86.5	84.5	73.25	73.25	81.5	42.5	51.25	40.25	10	15.5	9.25	2.5	4.25	4.25	3	5.25	5.25
19	LEQNPEESQDIk (12)	3	68.75	68.25	70	72.25	61.5	69.75	54.25	64.25	52	30.25	18.25	18	1.25	2	8.5	2	1.75	2
6	LWSLAEIATSDLk (13)	3	81.25	80.25	80.75	78.5	79.25	78.75	49.5	45.5	59.25	18	15	12.5	12	9.25	12	13	21	9
10	NSWGTDWGEk (10)	2	47.75	46.5	44.5	46.5	37.5	46.5	31.5	24.75	11	3	7	0.25	0	0.25	3	0.85	0.5	0
20	SEASSSPPVVTSSSHSr (17)	3	76	77.5	79	76.25	63.25	64.5	65.5	72.25	72.25	15.75	29.25	29.5	12.25	11	13	4	13	6.5
7	SEDEDEDEGDATr (13)	1/2	72.75	72.25	64.5	29.25	61.25	69.5	2	3	70	1	6	6	5	1	2.5	2.5	2.5	10
12	SFEGTDYGk (9)	3	33.5	36.5	37.25	36.5	29.25	31.5	28.75	29.5	29.25	18.75	12.5	18.5	7.25	7	7.25	7	9	7
4	SGWTcTQPGGr (11)	3	50	48.5	50.25	49.25	48.5	48.25	35.5	44.5	33.75	19.25	15.25	15.25	11.25	11.25	11.25	8.25	11.25	11.25
13	TLNGIQLAr (9)	3	36.5	37.5	38.5	36.5	37.75	38.5	29.75	29.75	34.75	20.75	17.75	17.75	7.75	6.75	11.75	6	5.75	6
23	TVFSSTQLcVLNDr (14)	3	84	84	85.25	84	87.5	84	80.25	82.5	79.75	23.5	35.5	38.5	13.25	18.5	17	18.5	13.5	9.75
22	TWFQNQr (7)	3	21.5	22.25	21.5	21.75	21.75	21.75	21.5	19.75	21.75	10.75	10.75	10.75	5.25	3.25	6.25	2.5	1.5	1.5
15	WPGYLNGGr (9)	3	37.75	34.5	35.25	36	34.5	37.25	35	36	36	16	15	13	4	6.25	5	5.25	5	5.25

Does not meet threshold, reject concentration
Meets threshold, but reject concentration based on high background levels
Does not meet threshold, but accept concentration based on poor peptide retention
Does not meet threshold, but accept concentration based on zero-noise argument
Meets threshold, accept concentration

Table 6.1C. Extracted score chromatogram maxima for all neat PRM ±0.2 Th experiments.

#	Peptide (Threshold)	# Orders	200 nM			20 nM			2 nM			200 pM			20 pM			2 pM		
			1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
17	AETLVQAr (8)	2/3	24.25	22.25	29.25	22.25	22.25	22.25	18.5	19.25	18.5	3	5	4	0	0	0	0	0	0
14	AFSQNSVLIk (10)	3	46.25	46.25	46.25	47.25	46.25	46.25	36.5	28.25	36.5	14.5	15.5	14	8	0	0	0	0	0
21	APcQAGDLr (9)	3	36.25	38.25	40	38.25	37	39	29	28	26.25	17	20	15	0	0	0	0	0	0
3	ARPAcVDAr (9)	0/1	2	11.25	9	2	3.75	2	0	0	0	0	0	0	0	0	0	0	0	0
8	DIQFGSQIk (9)	2/3	30.25	30.25	31.75	30.25	29.25	29.25	22.25	22.5	24.25	6	6	6	0	0	0	0	0	0
5	EGQLAAGTcEIVTLDr (16)	3	103.25	103.25	103.25	101.25	97.75	102.75	84.25	88.5	92.5	38.75	34.75	33.75	11	0	10	0	0	11
18	FLNPEWk (7)	2/3	22.25	22.25	22.25	22.25	22.25	21.5	13.5	18.5	14.5	2	2	6	0	0	0	0	0	0
11	FSDLTEEEFr (10)	3/4	48.5	48.5	48.5	48.5	48.5	48.75	47.5	47.25	47.25	46.25	46.25	42.25	6	8	6	0	0	0
24	FSEVSADk (8)	2	28.5	28	24	28.5	28.5	25.5	18	9	11	0	0	0	0	0	0	0	0	0
16	GALDGEAPr (9)	2	29.25	27.5	29.25	29.25	29.25	25.25	14	18	18	6	6	5	0	0	0	0	0	7
25	GLYEGTGr (8)	2	19.5	16.5	16.5	16.5	15.5	15.5	15	4	15	0	0	0	0	0	0	0	0	0
1	GVSAFSTWEk (10)	3	39.5	39.5	38.25	38.25	39.5	38.25	37.25	37.25	37.25	23.5	25.5	15.5	0	0	0	0	0	0
2	HFLTLAPIk (9)	3	45	45	45	45	42.5	43.25	32	32.25	31	6	10	11	0	0	0	0	0	0
9	HTGTPLGDIPYGk (13)	2	88.5	94.5	92	89	85.5	94.5	59.75	53	53.75	11	13	11.75	0	0	4	0	0	0
19	LEQNPEESQDIk (12)	3/4	70.75	72.5	74.5	72	72.5	72.5	59.75	60	63.25	37.25	35.25	42.25	8	8	8	0	0	0
6	LWSLAEIATSDLk (13)	3	79.5	78.5	78.5	77.5	78.75	77.25	58.25	61.25	63.25	56.5	45.25	38.25	0	0	0	0	0	0
10	NSWGTDWGEk (10)	3	46.5	46.5	46.5	46.5	46.5	46.5	46.5	46.5	44.5	28.5	14.5	25	7	0	0	0	0	0
20	SEASSSPPVVTSSSHSr (17)	3/4	90.5	94.75	97	82	94.75	84	77.25	69.75	66.75	52.25	49.25	48.25	11.5	11.5	18.5	0	0	0
7	SEDEDEDEGDATr (13)	3	68.5	74.75	69.5	45.5	50.25	72.75	63.25	0	67.25	2.5	33.5	33.5	2.5	2.5	0	0	2.5	2.5
12	SFEGTDYGk (9)	3	32.5	37.25	29.25	36.5	36.5	29.25	24.5	23.5	28.5	7.5	13	13	0	0	0	0	0	0
4	SGWTcTQPGGr (11)	3	51.75	60.25	59	50.25	51	52.75	35.75	36.25	47.25	25	25.75	30	0	0	0	0	0	6
13	TLNGIQLAr (9)	3	38.25	38.25	38.25	37.5	38.25	38.25	28.5	23.5	26.5	7	17	13	1	0	0	0	0	0
23	TVFSSTQLcVLNDr (14)	3	98.75	89	91.5	89	87.75	89	82	81.5	82.25	72.5	72.25	64.5	0.5	3	7.5	0	2.5	0
22	TWFQNQr (7)	3	21.5	22.25	22.25	21.5	21.5	21.5	19.5	19.5	13.5	9	9	9	0	0	0	0	0	0
15	WPGYLNGGr (9)	3	37.25	37.25	36.25	36.75	37.25	36.5	30	30	33	15	15	8	0	0	0	0	0	0

Does not meet threshold, reject concentration
Meets threshold, but reject concentration based on high background levels
Does not meet threshold, but accept concentration based on poor peptide retention
Does not meet threshold, but accept concentration based on zero-noise argument
Meets threshold, accept concentration

Table 6.1D. Extracted score chromatogram maxima for all matrix PRM ±0.2 Th experiments.

#	Peptide (Threshold)	# Orders	200 nM 1	200 nM 2	200 nM 3	20 nM 1	20 nM 2	20 nM 3	2 nM 1	2 nM 2	2 nM 3	200 pM 1	200 pM 2	200 pM 3	20 pM 1	20 pM 2	20 pM 3	2 pM 1	2 pM 2	2 pM 3
17	AETLVQAr (8)	2	22.25	22.25	22.25	22.25	22.25	22.25	15.25	19.5	13.5	4	6	0	0	0	0	0	0	0
14	AFSQNSVLIk (10)	2/3	37.25	46.25	46.25	39.25	46.25	37.25	25.5	32.5	27.5	8	8	8	0	0	0	0	0	0
21	APcQAGDLr (9)	3	38.25	37	37	37	37	38.25	27.25	26.25	34.75	14	15	8.5	3	1.5	3.5	2.5	2	2
3	ARPAcVDAr (9)	0/1	12.25	11.25	12.25	2	3.75	3.75	0	0	0	0	0	0	0	0	0	0	0	0
8	DIQFGSQIk (9)	2	29.25	29.25	29.25	29.25	26.25	29.25	12	12.5	11.5	0.5	0.5	0.5	0	0.5	0.5	0.5	0.5	0.5
5	EGQLAAGTcEIVTLDr (16)	3	103	102.75	103.75	97.75	97.75	96.25	78.75	80.5	84	30.75	32.75	32.75	16	15	10	11	10.75	11
18	FLNPEWk (7)	2	21.5	21.5	18.5	18	14.5	19.5	11.5	7	7	0	0	0	0	0	0	0	0	3
11	FSDLTEEEFr (10)	3	48.5	48.5	48.5	47.5	48.5	48.5	46.25	46.25	46.25	33.25	23.25	18.75	15	3	8	3	3	3
24	FSEVSADk (8)	2	28	29.25	28.5	29.25	29.25	27	7	13	15.5	0	0	0	0	0	0	0	0	0
16	GALDGEAPr (9)	2	29.25	29.25	28.5	29.25	29.25	29.25	15.5	14.5	14.5	5	5	0.5	0.5	0	0	0	0	0
25	GLYEGTGr (8)	2	17.25	19.25	21.5	19.5	18.5	19.5	15	6	9	0	0	0	0	0.5	0	0	0	0
1	GVSAFSTWEk (10)	2	38.25	38.25	38.25	38.25	37.25	38.25	30.25	29.25	24.25	8	5	0.5	0.5	0.5	1	0.5	0	0
2	HFLTLAPIk (9)	2	36.25	38	43	35.25	36.25	35.5	12	12	10	0	0	0	0	0	0	0	0	0
9	HTGTPLGDIPYGk (13)	2	92	92	89	72.5	80.25	75	42.25	24.75	41	8.5	8.5	6.5	2.5	2.5	3	4.25	4.25	4.25
19	LEQNPEESQDIk (12)	3	72.5	70	72.75	70	72.25	74.25	51.25	50.25	53.25	19.75	20.5	29.5	2	1.25	1.5	1.5	2	0.5
6	LWSLAEIATSDLk (13)	3	79.5	80.5	80.25	79.25	78.75	78.25	56.25	57.25	46.5	24.25	18.25	16	12	12	12	12	15	12
10	NSWGTDWGEk (10)	2	46.5	46.5	46.5	46.5	46.5	46.5	35.5	34.5	39.5	7	8	7	0.25	0	2	0	0	2
20	SEASSSPPVVTSSSHSr (17)	3	97	93.75	80.25	84	79.25	93.25	66.25	67.25	64.25	25.5	18.5	30.25	11.75	11	11	4	11	13
7	SEDEDEDEGDATr (13)	1/2	84.75	69.25	68.25	50.25	64.5	55.25	69.5	2.25	2.5	0	6	2.5	0	2.5	2.5	2.5	0	2.5
12	SFEGTDYGk (9)	2	36.5	36.5	29.25	32.5	29.25	29.25	23.5	20.5	20.5	10	5	0.5	0.75	0.75	5	0.75	0.75	0.75
4	SGWTcTQPGGr (11)	3	50.5	60.25	50	58	56.25	48.5	34.75	35.5	35	19.25	19.25	17.25	1.25	2	1	1	0.25	0.25
13	TLNGIQLAr (9)	2	36.5	36.5	37.5	29.5	28.5	29.5	22.5	20.5	18.5	8	8	7	1	1	1	1	3	1
23	TVFSSTQLcVLNDr (14)	3	86.75	87.25	89.25	86.25	87	89	76.25	81.5	80.25	36.5	35.25	46.75	18	18	14	20.5	12.75	18.5
22	TWFQNQr (7)	2	21.5	21.5	21.5	21.5	21.5	21.5	13.5	13.5	15.5	5	5	5	0.25	0.25	0.25	0.25	0.25	1
15	WPGYLNGGr (9)	2	35.25	36.75	34	34	34	33.5	30	26	30	7	0	0.25	1	0.25	0	2	0.25	0.25

Does not meet threshold, reject concentration
Meets threshold, but reject concentration based on high background levels
Does not meet threshold, but accept concentration based on poor peptide retention
Does not meet threshold, but accept concentration based on zero-noise argument
Meets threshold, accept concentration

Table 6.2A. QqOrbi measurement precision by batch category.

Category	n	Mean %RSD	Median %RSD
All data	810	15.0	7.8
All neat	484	13.5	6.4
All matrix	326	17.1	9.9
PRM neat	213	15.3	7.3
PRM matrix	179	14.1	9.2
SIM neat	271	12.2	6.1
SIM matrix	147	20.9	11.4
PRM ±0.2 Th	181	16.8	8.7
PRM ±1 Th	211	13.0	7.5
SIM ±1 Th	211	15.9	7.0
SIM ±2 Th	207	14.5	8.2
All PRM	392	14.7	8.2
All SIM	418	15.2	7.3

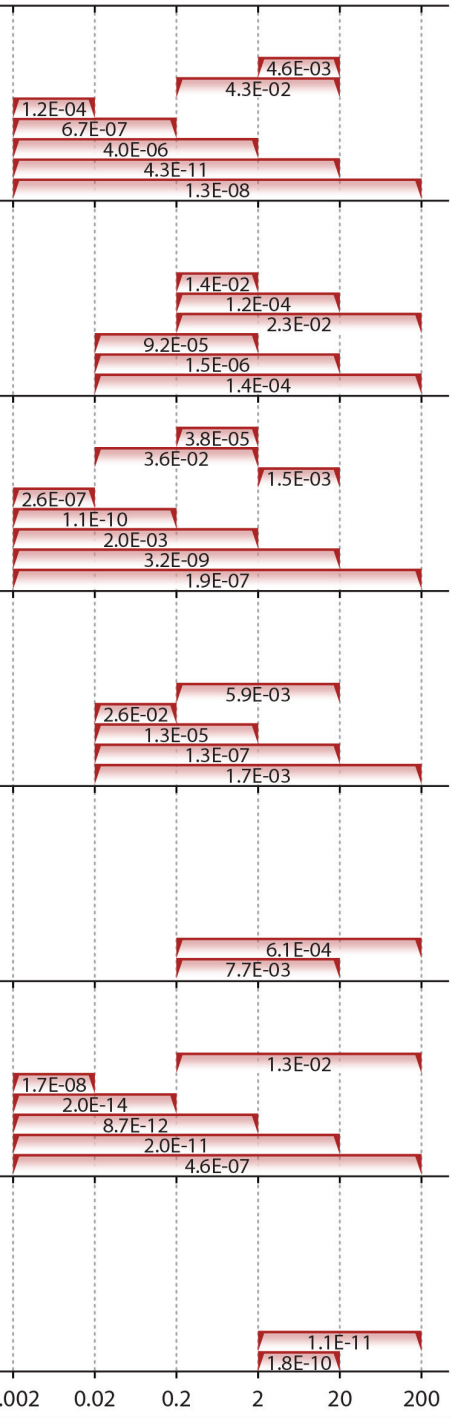
Welch's t-test on log-transformed data

testing...	Category 1	Category 2	p-value
Matrix	**All neat**	All matrix	**2.98E-06**
Matrix	PRM neat	PRM matrix	4.20E-01
Matrix	**SIM neat**	SIM matrix	**4.91E-08**
Isolation width	PRM ±0.2 Th	PRM ±1 Th	5.77E-02
Isolation width	SIM ±1 Th	SIM ±2 Th	8.73E-01
Experiment type	All PRM	All SIM	4.87E-01
Experiment type	PRM ±1 Th	SIM ±1 Th	7.79E-01
Experiment type	PRM neat	**SIM neat**	**1.91E-02**
Experiment type	**PRM matrix**	SIM matrix	**6.74E-03**

Bold p-value signifies statistically significant comparison (α = 5E-02). **Bold** category signifies significantly greater measurement precision.

Table 6.2B. QqOrbi measurement precision by concentration.

	Conc (nM)	n	Mean %RSD	SD	One-way ANOVA p-values with Bonferroni correction
All	0.002	30	31.5	17.5	
	0.02	72	17.6	19.5	
	0.2	123	14.7	18.0	
	2	187	17.0	22.6	
	20	198	11.7	20.5	
	200	200	13.1	21.3	
PRM	0.002	2	19.6	1.8	
	0.02	26	30.4	24.9	
	0.2	71	17.9	15.0	
	2	93	11.0	13.1	
	20	100	12.1	22.2	
	200	100	14.4	26.6	
SIM	0.002	28	32.3	17.9	
	0.02	46	10.4	10.3	
	0.2	52	10.4	20.7	
	2	94	23.0	27.9	
	20	98	11.2	18.7	
	200	100	11.9	14.2	
PRM Neat	0.002	n/a	n/a	n/a	
	0.02	22	32.4	24.4	
	0.2	44	17.1	17.4	
	2	47	10.8	17.1	
	20	50	11.1	25.0	
	200	50	14.7	23.0	
PRM Matrix	0.002	2	19.6	1.8	
	0.02	4	19.6	28.2	
	0.2	27	19.3	10.3	
	2	46	11.2	7.3	
	20	50	13.1	19.2	
	200	50	14.1	30.0	
SIM Neat	0.002	28	32.3	17.9	
	0.02	46	10.4	10.3	
	0.2	47	7.4	18.1	
	2	50	8.3	14.0	
	20	50	10.5	23.3	
	200	50	12.5	14.9	
SIM Matrix	0.002	n/a	n/a	n/a	
	0.02	n/a	n/a	n/a	
	0.2	5	38.6	24.5	
	2	44	39.6	30.4	
	20	48	12.0	12.4	
	200	50	11.2	13.5	



Only p-values for significantly different comparsions shown (α = 5E-02). ANOVA performed on log-vtransformed data.

Table 6.2C. QqOrbi measurement precision by peptide.

Peptide	n	Mean %RSD	SD	n	Mean %RSD	SD	n	Mean %RSD	SD
		All			*Neat*			*Matrix*	
1	33	8.0	7.9	20	7.8	9.2	13	8.2	5.8
2	31	11.9	11.0	19	9.7	10.0	12	15.4	11.9
3	20	39.2	23.7	10	41.5	26.0	10	36.9	22.4
4	35	8.7	11.3	19	6.6	10.2	16	11.2	12.3
5	32	11.7	8.9	18	13.9	8.4	14	8.9	9.1
6	29	12.6	10.1	19	14.3	11.9	10	9.4	4.3
7	26	77.3	52.4	14	80.5	51.9	12	73.6	54.9
8	33	12.3	11.8	20	10.0	11.1	13	15.8	12.3
9	27	11.9	12.0	17	14.4	14.3	10	7.7	4.4
10	33	16.4	19.1	21	11.9	11.2	12	24.3	26.9
11	36	10.5	9.8	22	10.5	10.6	14	10.7	8.9
12	34	13.9	20.9	21	12.8	19.0	13	15.6	24.4
13	34	14.7	13.7	21	13.8	11.7	13	16.2	16.8
14	33	14.5	13.7	20	12.9	14.8	13	17.1	12.1
15	34	11.6	13.9	21	8.6	8.1	13	16.3	19.5
16	33	12.3	12.8	20	9.7	9.2	13	16.2	16.6
17	34	11.8	14.2	21	10.8	15.1	13	13.5	13.2
18	34	15.0	14.8	21	14.8	17.4	13	15.2	9.9
19	36	9.3	14.5	20	6.0	8.6	16	13.3	19.1
20	36	11.5	14.6	22	8.4	14.5	14	16.4	14.0
21	35	10.7	10.0	19	6.1	5.5	16	16.2	11.4
22	34	15.2	23.3	21	9.4	10.8	13	24.5	33.8
23	34	12.3	17.8	19	13.9	20.1	15	10.3	14.8
24	32	14.4	16.9	20	15.0	20.3	12	13.5	9.5
25	32	10.8	16.0	19	10.0	16.3	13	11.9	16.1

One-way ANOVA with Bonferroni correction

Category	Peptide	# Significant Comparisons*	Significant Peptides
All	**3**	23/24	**1-2, 4-6, 8-25**
All	**7**	23/24	**1-2, 4-8, 8-25**
Neat	**3**	18/24	**1-2, 4, 8, 10-12, 14-17, 19-25**
Neat	**7**	23/24	**1-2, 4-8, 8-25**
Matrix	**3**	6/24	**1, 4-5, 9, 23, 25**
Matrix	**7**	22/24	**1-2, 4-6, 8-21, 23-25**

Only statististically significant comparisons (α = 5E-02) are listed; %RSD of the listed peptide is significantly greater in all comparisons (decreased measurement precision). ANOVA performed on log-transformed data.
*Number of significant comparisons out of the total number of comparisons.

Table 6.3A. QqOrbi dynamic range by batch category.

Category	n	Mean	SD
All data	200	2.9	1.2
All neat	100	3.7	1.1
All matrix	100	2.2	0.7
PRM neat	50	3.0	0.8
PRM matrix	50	2.4	0.7
SIM neat	50	4.4	0.8
SIM matrix	50	1.9	0.6
PRM ±0.2 Th	50	2.4	0.7
PRM ±1 Th	50	3.0	0.8
SIM ±1 Th	50	3.2	1.4
SIM ±2 Th	50	3.1	1.4
All PRM	100	2.7	0.8
All SIM	100	3.2	1.4

Welch's t-test

testing...	Category 1	Category 2	p-value
Matrix	**All neat**	All matrix	**1.19E-23**
	PRM neat	PRM matrix	**6.34E-04**
	SIM neat	SIM matrix	**1.17E-30**
Isolation width	PRM ±0.2 Th	**PRM ±1 Th**	**7.41E-05**
	SIM ±1 Th	SIM ±2 Th	7.81E-01
Experiment type	All PRM	**All SIM**	**4.81E-03**
	PRM ±1 Th	SIM ±1 Th	3.91E-01
	PRM neat	**SIM neat**	**2.01E-14**
	PRM matrix	SIM matrix	**2.88E-04**

Bold p-value signifies statistically significant comparison (α = 5E-02). **Bold** category signifies significantly greater number of concentration orders-of-magnitude quantified.

Table 6.3B. QqOrbi depression of dynamic range by experimental characteristics.

Category	Comparison	Mean	SD	1-sample t-test (H_o: μ = 0)
PRM ±0.2 Th	**neat** vs. matrix	0.40	0.58	**2.01E-03**
PRM ±1 Th	**neat** vs. matrix	0.60	0.58	**2.53E-05**
SIM ±1 Th	**neat** vs. matrix	2.52	1.00	**5.02E-12**
SIM ±2 Th	**neat** vs. matrix	2.44	1.08	**4.59E-11**
PRM neat	**±1 Th** vs. ±0.2 Th	0.68	0.95	**1.45E-03**
PRM matrix	**±1 Th** vs. ±0.2 Th	0.48	0.51	**8.75E-05**
SIM neat	±2 Th vs. ±1 Th	-0.12	0.33	8.30E-02
SIM matrix	±2 Th vs. ±1 Th	-0.04	0.20	3.27E-01
Neat ±1 Th	PRM vs. **SIM**	-1.20	0.58	**2.31E-10**
Matrix ±1 Th	**PRM** vs. SIM	0.72	0.94	**7.79E-04**

Welch's t-test

Category 1	Category 2	p-value
PRM ±1 Th (neat vs. matrix)	**SIM ±1 Th (neat vs. matrix)**	**4.59E-10**
PRM ±0.2 Th (neat vs. matrix)	PRM ±1 Th (neat vs. matrix)	2.27E-01
SIM ±1 Th (neat vs. matrix)	SIM ±2 Th (neat vs. matrix)	7.88E-01

Bold p-value signifies statistically significant comparison (α = 5E-02).
Bold category signifies significantly greater number of concentration orders-of-magnitude quantified.

Table 6.4A. QqOrbi measurement accuracy by batch category.

Category	n	Mean %RSD	SD
All data	214	34.6	20.1
All neat	115	35.3	18.1
All matrix	99	33.7	22.4
PRM neat	65	36.9	20.6
PRM matrix	51	32.3	22.5
SIM neat	50	33.3	14.0
SIM matrix	48	35.2	22.3
PRM ±0.2 Th	57	35.1	20.1
PRM ±1 Th	59	34.6	22.9
SIM ±1 Th	49	35.1	17.3
SIM ±2 Th	49	33.4	19.8
All PRM	116	34.9	21.5
All SIM	98	34.2	18.5

Welch's t-test on log-transformed data

testing...	Category 1	Category 2	p-value
Matrix	All neat	All matrix	7.99E-02
	PRM neat	PRM matrix	9.71E-02
	SIM neat	SIM matrix	4.42E-01
Isolation width	PRM ±0.2 Th	PRM ±1 Th	8.33E-01
	SIM ±1 Th	SIM ±2 Th	3.72E-01
Experiment type	All PRM	All SIM	8.82E-01
	PRM ±1 Th	SIM ±1 Th	4.89E-01
	PRM neat	SIM neat	7.28E-01
	PRM matrix	SIM matrix	5.87E-01

Bold p-value signifies statistically significant comparison (α = 5E-02). **Bold** category signifies significantly greater linearity (lower %RSD).

Table 6.4B. QqOrbi measurement accuracy by individual experiment.

Category	n	Mean %RSD	SD
Neat PRM ±0.2 Th	32	37.1	16.6
Neat PRM ±1 Th	35	36.4	23.4
Neat SIM ±1 Th	26	34.0	12.0
Neat SIM ±2 Th	26	32.3	15.5
Matrix PRM ±0.2 Th	28	35.7	27.7
Matrix PRM ±1 Th	28	34.2	25.0
Matrix SIM ±1 Th	25	36.1	21.1
Matrix SIM ±2 Th	25	34.1	23.0

Welch's t-test on log-transformed data

testing...	Category 1	Category 2	p-value
Matrix	Neat PRM ±0.2 Th	Matrix PRM ±0.2 Th	2.75E-01
	Neat PRM ±1 Th	Matrix PRM ±1 Th	5.24E-01
	Neat SIM ±1 Th	Matrix SIM ±1 Th	6.25E-01
	Neat SIM ±2 Th	Matrix SIM ±2 Th	5.70E-01
Isolation width	Neat PRM ±0.2 Th	Neat PRM ±1 Th	5.85E-01
	Neat SIM ±1 Th	Neat SIM ±2 Th	3.86E-01
	Matrix PRM ±0.2 Th	Matrix PRM ±1 Th	9.49E-01
	Matrix SIM ±1 Th	Matrix SIM ±2 Th	5.59E-01
Experiment type	Neat PRM ±1 Th	Neat SIM ±1 Th	6.81E-01
	Matrix PRM ±1 Th	Matrix SIM ±1 Th	6.66E-01

Bold p-value signifies statistically significant comparison (α = 5E-02). **Bold** category signifies significantly greater linearity (lower %RSD).

Table 6.4C. QqOrbi measurement accuracy by batch category, excluding lowest detected concentration.

Category	n	Mean %RSD	SD
All data	203	21.2	16.7
All neat	112	26.4	17.0
All matrix	91	14.8	13.9
PRM neat	62	23.2	16.8
PRM matrix	47	12.0	8.8
SIM neat	50	30.3	16.6
SIM matrix	44	17.8	17.5
PRM ±0.2 Th	54	18.3	11.6
PRM ±1 Th	55	18.4	17.7
SIM ±1 Th	47	24.4	15.8
SIM ±2 Th	47	24.6	20.3
All PRM	109	18.3	14.9
All SIM	94	24.5	18.1

Welch's t-test on log-transformed data

testing...	Category 1	Category 2	p-value
Matrix	All neat	**All matrix**	**6.45E-13**
	PRM neat	**PRM matrix**	**6.49E-08**
	SIM neat	**SIM matrix**	**6.26E-07**
Isolation width	PRM ±0.2 Th	PRM ±1 Th	8.81E-01
	SIM ±1 Th	SIM ±2 Th	9.01E-01
Experiment type	**All PRM**	All SIM	**3.01E-02**
	PRM ±1 Th	SIM ±1 Th	9.67E-02
	PRM neat	SIM neat	**1.97E-04**
	PRM matrix	SIM matrix	2.31E-01

Bold p-value signifies statistically significant comparison (α = 5E-02). **Bold** category signifies significantly greater linearity (lower %RSD).

Table 6.4D. QqOrbi measurement accuracy by individual experiment, excluding lowest detected concentration.

Category	n	Mean %RSD	SD
Neat PRM ±0.2 Th	31	22.2	10.7
Neat PRM ±1 Th	33	23.9	20.7
Neat SIM ±1 Th	26	29.5	10.4
Neat SIM ±2 Th	26	30.8	20.8
Matrix PRM ±0.2 Th	26	14.0	11.1
Matrix PRM ±1 Th	26	12.5	9.0
Matrix SIM ±1 Th	23	18.0	18.2
Matrix SIM ±2 Th	23	17.1	16.4

Welch's t-test on log-transformed data

testing...	Category 1	Category 2	p-value
Matrix	Neat PRM ±0.2 Th	**Matrix PRM ±0.2 Th**	**1.05E-03**
	Neat PRM ±1 Th	**Matrix PRM ±1 Th**	**1.44E-04**
	Neat SIM ±1 Th	**Matrix SIM ±1 Th**	**3.26E-04**
	Neat SIM ±2 Th	**Matrix SIM ±2 Th**	**2.86E-04**
Isolation width	Neat PRM ±0.2 Th	Neat PRM ±1 Th	8.64E-01
	Neat SIM ±1 Th	Neat SIM ±2 Th	7.81E-01
	Matrix PRM ±0.2 Th	Matrix PRM ±1 Th	7.72E-01
	Matrix SIM ±1 Th	Matrix SIM ±2 Th	9.64E-01
Experiment type	**Neat PRM ±1 Th**	Neat SIM ±1 Th	**3.90E-03**
	Matrix PRM ±1 Th	Matrix SIM ±1 Th	4.93E-01

Bold p-value signifies statistically significant comparison (α = 5E-02). **Bold** category signifies significantly greater linearity (lower %RSD).

Table 6.4E. QqOrbi measurement accuracy by batch category, statistical comparison of all data versus data excluding lowest detected concentration.

Comparison	Welch's t-test p-value
All data	**2.16E-16**
All neat	**5.29E-06**
All matrix	**2.25E-14**
PRM neat	**3.52E-06**
PRM matrix	**4.81E-10**
SIM neat	1.79E-01
SIM matrix	**3.71E-06**
PRM ±0.2 Th	**5.10E-07**
PRM ±1 Th	**3.10E-07**
SIM ±1 Th	**6.55E-04**
SIM ±2 Th	**5.76E-03**
All PRM	**5.61E-13**
All SIM	**1.18E-05**

Bold p-value signifies statistically significant comparison (α = 5E-02). Data excluding lowest concentration demonstrates greater linearity (lower %RSD) in all comparisons. T-test performed on log-transformed data.

Table 6.4F. QqOrbi measurement accuracy by individual experiment, statistical comparison of all data versus data excluding lowest detected concentration.

Comparison	Welch's t-test p-value
Neat PRM ±0.2 Th	**1.37E-04**
Neat PRM ±1 Th	**3.22E-03**
Neat SIM ±1 Th	1.18E-01
Neat SIM ±2 Th	5.41E-01
Matrix PRM ±0.2 Th	**5.07E-05**
Matrix PRM ±1 Th	**2.74E-06**
Matrix SIM ±1 Th	**3.88E-04**
Matrix SIM ±2 Th	**1.32E-03**

Bold p-value signifies statistically significant comparison (α = 5E-02). Data excluding lowest concentration demonstrates greater linearity (lower %RSD) in all comparisons. T-test performed on log-transformed data.

Table 6.4G. QqOrbi measurement accuracy, as adjusted %RSD, by batch category.

Category	n	Mean Adj. %RSD	SD
All data	214	13.9	15.2
All neat	115	11.4	13.5
All matrix	99	16.7	16.6
PRM ±0.2 Th	57	15.6	15.8
PRM ±1 Th	59	14.1	21.2
SIM ±1 Th	49	12.6	8.9
SIM ±2 Th	49	12.8	10.5
All PRM	116	14.8	18.7
All SIM	98	12.7	9.7

Welch's t-test on log-transformed data

testing...	Category 1	Category 2	p-value
Matrix	**All neat**	All matrix	**4.28E-04**
Isolation width	PRM ±0.2 Th	PRM ±1 Th	1.13E-01
	SIM ±1 Th	SIM ±2 Th	5.64E-01
Experiment type	All PRM	All SIM	4.55E-01
	PRM ±1 Th	SIM ±1 Th	5.74E-01

Bold p-value signifies statistically significant comparison (α = 5E-02). **Bold** category signifies significantly greater linearity (lower adjusted %RSD).

Table 6.4H. QqOrbi measurement accuracy, as adjusted %RSD, by individual experiment.

Category	n	Mean Adj. %RSD	SD
Neat PRM ±0.2 Th	31	14.4	9.2
Neat PRM ±1 Th	34	13.4	21.8
Neat SIM ±1 Th	25	7.9	3.3
Neat SIM ±2 Th	25	8.6	8.0
Matrix PRM ±0.2 Th	26	17.1	21.2
Matrix PRM ±1 Th	25	15.0	20.9
Matrix SIM ±1 Th	24	17.5	10.2
Matrix SIM ±2 Th	24	17.2	11.1

Welch's t-test on log-transformed data

testing...	Category 1	Category 2	p-value
Matrix	Neat PRM ±0.2 Th	Matrix PRM ±0.2 Th	8.91E-01
	Neat PRM ±1 Th	Matrix PRM ±1 Th	6.29E-01
	Neat SIM ±1 Th	Matrix SIM ±1 Th	**8.96E-06**
	Neat SIM ±2 Th	Matrix SIM ±2 Th	**9.91E-04**
Isolation width	Neat PRM ±0.2 Th	Neat PRM ±1 Th	1.01E-01
	Neat SIM ±1 Th	Neat SIM ±2 Th	6.57E-01
	Matrix PRM ±0.2 Th	Matrix PRM ±1 Th	5.33E-01
	Matrix SIM ±1 Th	Matrix SIM ±2 Th	6.17E-01
Experiment type	Neat PRM ±1 Th	Neat SIM ±1 Th	9.81E-02
	Matrix PRM ±1 Th	Matrix SIM ±1 Th	6.21E-02

Bold p-value signifies statistically significant comparison (α = 5E-02). **Bold** category signifies significantly greater linearity (lower adjusted %RSD).

Table 6.5A. QqQ measurement precision by batch category and individual experiment.

Category	n	Mean %RSD	SD
All data	105	12.4	21.0
All neat	60	17.5	26.3
All matrix	45	5.6	5.3

Welch's t-test on log-transformed data

testing...	Category 1	Category 2	p-value
Matrix	All neat	**All matrix**	**1.19E-03**

1-sample t-test on peptide/conc. matched log-transformed data (H_o: $\mu = 0$)

Comparison	n	Mean	SD	p-value
neat vs. matrix	45	0.9	5.0	1.87E-01

Bold p-value signifies statistically significant comparison (α = 5E-02). **Bold** category signifies significantly greater measurement precision.

Table 6.5B. QqQ measurement precision by concentration.

	Conc (nM)	n	Mean %RSD	SD	One-way ANOVA p-values with Bonferroni correction
Neat	0.002	n/a	n/a	n/a	
	0.02	5	77.7	37.5	1.1E-03
	0.2	13	31.4	24.0	3.4E-09
	2	14	11.0	10.2	1.5E-11, 3.7E-02, 8.1E-03
	20	14	4.3	3.4	9.8E-07, 5.5E-05
	200	14	2.7	1.3	6.3E-11, 1.2E-12
Matrix	0.002	n/a	n/a	n/a	
	0.02	n/a	n/a	n/a	
	0.2	5	12.4	8.0	
	2	12	5.6	4.7	
	20	14	4.6	5.6	
	200	14	4.0	1.6	3.44E-02
				Conc (nM)	0.002 0.02 0.2 2 20 200

Only p-values for significantly different comparsions shown (α = 5E-02). ANOVA performed on log-vtransformed data.

Table 6.5C. QqQ measurement accuracy and dynamic range by batch category and individual experiment.

Category	n	Mean %RSD	SD
All data	28	29.1	15.2
All data*	26	16.1	7.3
All neat	14	42.1	10.0
All neat*	14	15.7	9.3
All matrix	14	16.4	4.3
All matrix*	12	16.6	4.2
All neat (order-matched)	14	16.2	10.8
All matrix (order-matched)	14	16.1	4.1
All neat (Adjusted %RSD)	14	13.0	2.9
All matrix (Adjusted %RSD)	14	8.0	3.4
Neat (orders)	14	3.3	0.6
Matrix (orders)	14	3.2	0.5

Paired t-test on log-transformed data

Category 1	Category 2	p-value
All neat	**All matrix**	**1.20E-07**
All neat*	All matrix*	3.20E-01
All neat (orders matched)	All matrix (orders matched)	3.10E-01
All neat (Adjusted %RSD)	**All matrix (Adjusted %RSD)**	**4.83E-05**
Neat (orders)	Matrix (orders)	**1.97E-05**

Welch's t-test on log-transformed data

Category 1	Category 2	p-value
All data	**All data***	**2.04E-04**
All neat	**All neat***	**5.42E-06**
All matrix	All matrix*	8.90E-01

Bold p-value signifies statistically significant comparison (α = 5E-02).
Bold category signifies significantly greater linearity (lower %RSD) or or dynamic range. *denotes exclusion of lowest concentration data.

Table 6.6A. Comparison of QqOrbi PRM and QqQ measurement precision data by batch category and individual experiment.

Batch Category	n	Mean %RSD	SD
All QqQ data	105	12.4	21.0
All QqOrbi PRM data	229	12.1	13.6
QqQ neat	60	17.5	26.3
QqOrbi neat PRM	124	12.9	16.1
QqOrbi neat PRM ±0.2 Th	56	14.3	16.7
QqOrbi neat PRM ±1 Th	68	11.8	15.7
QqQ matrix	45	5.6	5.3
QqOrbi matrix PRM	105	11.1	9.8
QqOrbi matrix PRM ±0.2 Th	49	11.6	9.2
QqOrbi matrix PRM ±1 Th	56	10.6	10.4
QqQ neat (matrix-matched)	45	6.5	6.6

Welch's t-test on log-transformed data

Category 1	Category 2	p-value
All QqQ data	All QqOrbi PRM data	**1.17E-02**
QqQ neat	QqOrbi neat PRM	9.85E-01
QqQ neat	QqOrbi neat PRM ±0.2 Th	6.05E-01
QqQ neat	QqOrbi neat PRM ±1 Th	6.79E-01
QqQ matrix	QqOrbi matrix PRM	**1.02E-05**
QqQ matrix	QqOrbi matrix PRM ±0.2 Th	**1.33E-05**
QqQ matrix	QqOrbi matrix PRM ±1 Th	**1.68E-04**
QqQ neat (matrix-matched)	QqOrbi neat PRM	**9.21E-04**
QqQ neat (matrix-matched)	QqOrbi neat PRM ±0.2 Th	**8.25E-04**
QqQ neat (matrix-matched)	QqOrbi neat PRM ±1 Th	**1.20E-02**

Bold p-value signifies statistically significant comparison (α = 5E-02).
Bold category signifies significantly greater precision (lower %RSD).

Table 6.6B. Comparison of QqOrbi PRM and QqQ measurement precision data by peptide and concentration-matched category for individual experiments.

Peptide/Concentration-matched Category	n	Mean %RSD	SD	Paired t-test p-value
QqQ neat	54	12.8	18.7	1.21E-01
QqOrbi neat PRM ±0.2 Th	54	13.4	15.0	
QqQ neat	60	17.5	26.3	1.06E-01
QqOrbi neat PRM ±1 Th	60	8.3	7.4	
QqQ matrix	42	5.4	5.4	**1.65E-05**
QqOrbi matrix PRM ±0.2 Th	42	11.0	8.6	
QqQ matrix	45	5.6	5.3	**2.23E-03**
QqOrbi matrix PRM ±1 Th	45	8.8	6.5	
QqQ neat (matrix-matched)	44	6.4	6.6	**3.79E-03**
QqOrbi neat PRM ±0.2 Th	44	12.2	14.9	
QqQ neat (matrix-matched)	45	6.5	6.6	6.05E-01
QqOrbi neat PRM ±1 Th	45	6.9	5.9	

Bold p-value signifies statistically significant comparison (α = 5E-02).
Bold category signifies significantly greater precision (lower %RSD).
Paired t-test performed on log-transformed data.

Table 6.6C. Comparison of QqQ SRM and QqOrbi SIM dynamic range and linearity.

		SRM	SIM ±1 Th	SIM ±2 Th
means				
Neat	**# orders**	3.3	4.8	4.7
	%RSD	42.09	30.48	26.78
	%RSD$_{adj}$	12.97	6.41	5.69
Matrix	**# orders**	2.2	2.0	1.9
	%RSD	16.09	32.06	31.86
	%RSD$_{adj}$	8.03	15.70	17.02
p values (SRM vs SIM)				
Neat	**# orders**		**7.18E-08‡**	**1.13E-07‡**
	%RSD		**8.43E-03‡**	**1.95E-03‡**
	%RSD$_{adj}$		**1.38E-06‡**	**1.49E-06‡**
Matrix	**# orders**		2.72E-01	1.65E-01
	%RSD		**6.86E-03***	**1.01E-02***
	%RSD$_{adj}$		**2.84E-03***	**1.14E-02***

‡SIM data show significantly greater linearity or dynamic range (α = 5E-02).

*SRM data show significantly greater linearity or dynamic range (α = 5E-02).

CHAPTER 7

Appendix to Chapter 4:

Development of a GC/quadrupole-Orbitrap Mass Spectrometer,

Part I: Design and Characterization

SUPPLEMENTAL TABLES

Table 7.1A. Inclusion list for targeted SIM experiments.

Analyte	Target *m/z*	Start (min)	End (min)
1,4-Dioxane	88.05	0.00	1.80
N-Nitrosodimethylamine	74.05	1.60	1.90
Pyridine	79.04	1.60	2.00
2-Fluorophenol	112.03	2.00	2.50
Phenol-d6	99.07	2.50	2.88
Phenol	94.04	2.50	2.88
Aniline, Bis(2-chloroethyl)ether	93.06	2.70	2.95
2-Chlorophenol	128.00	2.86	3.00
1,3-/1,4-/1,2-Dichlorobenzene	145.97	3.00	3.32
1,4-Dichlorobenzene-d4	149.99	3.04	3.18
Benzylalcohol	79.05	3.13	3.22
2-/4-/3-Methylphenol	108.06	3.22	3.45
Bis(2-chloroisopropyl)ether	121.04	3.25	3.35
N-Nitroso-di-n-propylamine	113.11	3.33	3.46
Hexachloroethane	200.84	3.46	3.56
Nitrobenzene-d5	128.06	3.51	3.59
Nitrobenzene	123.03	3.51	3.60
Isophorone	82.04	3.60	3.82
2-Nitrophenol	139.03	3.80	3.90
2,4-Dimethylphenol	107.05	3.80	3.90
Benzoic acid	105.03	3.82	4.21
Bis(2-chloroethoxy)methane	93.01	3.92	3.99
2,4-Dichlorophenol	161.96	3.99	4.11
1,2,4-Trichlorobenzene	179.93	4.10	4.19
Naphthalene-d8	136.11	4.16	4.26
Naphthalene	128.06	4.18	4.28
4-Chloroaniline	127.02	4.22	4.32
Hexachlorobutadiene	224.84	4.30	4.50
4-Chloro-3-methylphenol	142.02	4.50	4.82
2-/1-Methylnaphthalene	141.07	4.78	5.02
Hexachlorocyclopentadiene	236.84	5.00	5.09
2,4,6-/2,4,5-Trichlorophenol	195.92	5.07	5.23
2-Fluorobiphenyl	172.07	5.18	5.30
2-Chloronaphthalene	162.02	5.27	5.40
2-Nitroaniline	138.04	5.37	5.50
1,4-/1,3-/1,2-Dinitrobenzene	168.02	5.50	5.80
Dimethylphthalate	163.04	5.57	5.70
2,6-Dinitrotoluene	165.03	5.65	5.71
Acenaphthylene	152.06	5.71	5.80

Table 7.1B. Inclusion list for targeted SIM experiments, continued.

Analyte, cont'd.	Target *m/z*	Start (min)	End (min)
3-Nitroaniline	138.04	5.80	5.88
Acenaphthene-d10	162.13	5.84	5.92
Acenaphthene	153.07	5.89	5.96
2,4-Dinitrophenol	184.01	5.92	6.00
4-Nitrophenol	139.03	5.96	6.06
2,4-Dinitrotoluene	165.03	6.03	6.09
Dibenzofuran	168.06	6.06	6.14
2,3,5,6-/2,3,4,6-Tetrachlorophenol	231.88	6.11	6.30
Diethylphthalate	149.02	6.25	6.38
4-Chlorophenylphenylether	204.03	6.35	6.47
Fluorene	165.07	6.35	6.47
4-Nitroaniline	138.04	6.41	6.49
4,6-Dinitro-2-methylphenol	198.03	6.44	6.52
N-Nitrosodiphenylamine (diphenylamine)	169.09	6.50	6.58
1,2-Diphenylhydrazine (azobenzene)	182.08	6.55	6.64
2,4,6-Tribromophenol	329.77	6.60	6.80
4-Bromophenylphenylether	247.98	6.80	6.95
Hexachlorobenzene	283.81	6.93	7.09
Pentachlorophenol	265.84	7.09	7.31
Phenanthrene-d10	188.14	7.24	7.40
Phenanthrene, Anthracene	178.08	7.33	7.52
Carbazole	167.07	7.52	7.85
Di-n-butylphthalate	149.02	7.85	8.25
Fluoranthene, Pyrene	202.08	8.25	8.90
Benzidine	184.10	8.60	8.74
Pyrene-d10	212.14	8.70	8.90
p-Terphenyl-d14	244.20	8.89	9.24
3,3'-Dimethylbenzidine	212.13	9.24	9.51
Benzylbutylphthalate	149.02	9.37	9.49
Bis(2-ethylhexyl)adipate	129.05	9.47	9.85
3,3'-Dichlorobenzidine	252.02	9.85	10.12
Benz(a)anthracene, Chrysene	228.09	10.05	10.75
Chrysene-d12	240.17	10.06	10.50
Bis(2-ethylhexyl)phthalate	149.02	10.10	10.23
Di-n-octylphthalate	149.02	10.75	11.30
Benzo(b/k)fluoranthene, Benzo(a)pyrene	252.09	11.30	13.00
Perylene-d12	264.17	12.30	13.50
Indeno(1,2,3-cd)pyrene, Benzo(g,h,i)perylene	276.09	13.50	16.00
Dibenz(a,h)anthracene	278.11	13.50	14.85

CHAPTER 8

Appendix to Chapter 5:

Development of a GC/quadrupole-Orbitrap Mass Spectrometer,

Part II: New Approaches for Unambiguous Elemental Composition Assignment

and Structural Characterization in Untargeted GC/MS-based Metabolomics

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

8.1. Pseudocode for Molecular Ion-Directed Acquisition (MIDA) logic

```
if last scan was MS1 scan:
{
   Get list of mass differences specified by the pattern(s) selected by the
      user;
   Get all peaks from MS1 spectrum with S/N > 3;
   Sort peaks by ascending m/z;
   for each pattern to match against the spectrum:
   {
      for each peak in the spectrum:
      {
         Get the current peak;
         if the current peak's S/N > user's S/N threshold:
         {
            Populate lists with minimum values and maximum values for each
               mass "bin" specified by the members of the pattern given
               the current peak's m/z and user-selected mass error
               tolerance;
```

```
Save the minimum and maximum value of the first member of the
   current pattern;
// The next member of the pattern must fall between the
// current minimum and maximum.
Add the current peak to temporary list of matched pattern
   members;
Calculate the initial score:
{
   if the ionization type is EI:
      score = (intensity / 2) * (current peak's m/z ^ 5);
   else:
      score = (intensity / 2) * (current peak's m/z ^ 3);
}
for each peak with an m/z > current peak m/z:
{
   Get the peak;
   if the peak's m/z < current minimum:
   {
      // The bin has not yet been reached.
      Continue to the next peak in the list;
   }
   else:
   {
      if the peak's m/z > current maximum:
      {
         // The bin was passed over.
```

```
    if the current bin number is < required number of
       bins:
    {
       // The pattern is incomplete.
       Discard the temporary list of matched pattern
          members;
       Break out of loop;
    }
    else:
    {
       // There are enough members to make a pattern, so
       // check the next bin.
       Increment the bin;
       if there are still bins left in the pattern:
       {
          Update the current minimum and maximum;
          Continue to next peak in the list;
       }
       else:
       {
          // There are no more bins, save the temporary
          // list.
          Add the temporary list of peaks in pattern to
             a master list of patterns for this spectrum
             if the score of this pattern is not already
             in the master list;
          Break out of loop;
       }
    }
}
```

```
else:
{
    // The peak is inside a bin.
    Add the peak in the bin to the list of temporary
        pattern members;
    Update score:
    {
        if the ionization type is EI:
        {
            score +=
                (intensity / 2) * (current peak's m/z ^ 5);
        }
        else:
        {
            score +=
                (intensity / 2) * (current peak's m/z ^ 3);
        }
    }
    Increment the bin;
    if there are no bins left in the pattern after
        incrementing:
    {
        // There are no more bins, save the temporary
        // list.
        Add the temporary list of peaks in a pattern to a
            master list of patterns for this spectrum if
            the score of this pattern is not already in
            the master list;
        Break out of loop;
    }
```

```
                    else:
                    {
                        Update the current minimum and maximum;
                    }
                }
            }
        }
        if we break out of the loop AND there are still bins left from the
            last peak AND the minimum number of required bins has been met:
        {
            // Special case of pattern falling a very end of the peak list,
            // save the temporary list.
            Add the temporary list of peaks in a pattern to a master list
                of patterns for this spectrum if the score of this pattern
                is not already in the master list;
        }
        for each temporary pattern added to the master list of patterns:
        {
            Add the target peak of the pattern (the one selected for MS/MS
                or SIM as specified by the user) to a list of targets;
        }
    }
}
Sort the list of targets by the score of the pattern from which the
    target was derived;
```

```
    for each data-dependent event (e.g. ddTop1 = 1, ddTop5 = 5):
    {
        if there are target peaks found by the preceding algorithm:
        {
            Set up a new SIM or MS2 scan using the highest scoring target peak
                m/z;
        }
        else:
        {
            // No targets were found by the algorithm, do normal data-
            // dependent mode.
            Set up a new MS2 scan using the most intense peak in the spectrum;
        }
        Add the new scan as the next scan in the instrument scan queue;
    }
}
```