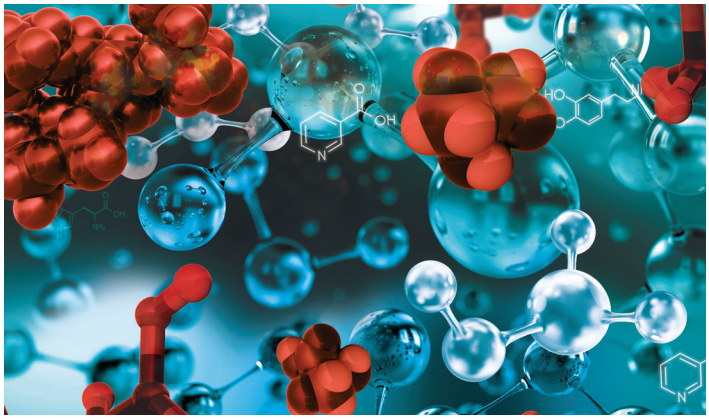
TECHNICAL NOTE 65962

Intelligent data acquisition for automated data reduction leading to increased metabolome annotation

Authors: Ioanna Ntai, Sven Hackbusch, Amanda Souza

Keywords: Metabolomics, metabolome, metabolites, lipidomics, lipidome, lipids, small molecule, annotation, identification, features, compounds, fragmentation, data-dependent acquisition, data independent acquisition, inclusion list, exclusion list, iterative, MS/MS spectra, MS^n spectra, AcquireX, Deep Scan workflow, Orbitrap Exploris MS, Orbitrap Tribrid MS



Goal

Demonstrate the use of Thermo Scientific™ AcquireX™ intelligent acquisition for automated annotation of non-biological and redundant features, resulting in more comprehensive fragmentation coverage of unique sample-relevant compounds.

Introduction

Untargeted metabolomics aims to detect all metabolites in a biological system without bias. The development of fast-scanning high-resolution, accurate-mass spectrometers has enabled the routine detection of thousands of features from a single sample[1]. Given the developments in hardware and the availability of various metabolomics software solutions, most researchers can easily obtain a list of features from any sample. In metabolomics, a feature is defined as a chromatographic peak with a unique retention time and mass-to-charge ratio. Translating a long feature list to metabolite identifications is a necessary step for the biological interpretation of the results, in order to provide biochemical insights into the system under study[2]. Metabolite identification is the most arduous and time-consuming step in this process and is recognised as a current bottleneck in the broad implementation of metabolomics[3].

Confident elemental formula calculation, chemical structure elucidation and name assignment rely on high quality MS and MS/MS data. There are mainly two approaches for collecting fragmentation data: data-dependent acquisition (DDA) and data-independent acquisition (DIA)[4]. During DDA, single precursor ions are isolated with narrow isolation windows and fragmented separately, resulting in fragmentation spectra of high quality that can be easily interpreted or matched to a spectral library. Precursor ions are prioritized in order of intensity, potentially leaving ions of lower intensity without fragmentation due to cycle time limitations. DIA approaches can provide fragmentation for all precursor ions simultaneously, but result in chimeric fragmentation spectra, that require special deconvolution software. Coeluting species with similar molecular weight often share fragment ions, resulting in poor quality spectral matches[5] and complicating the identification process.

Collecting more fragmentation data does not necessarily result in more compounds identified. Even when fragmentation for every ion is collected using a DIA approach, metabolite identification is still limited to just a small percentage of the features detected because most features do not represent unique metabolites of biological origin[1]. Many features detected may correspond to solvent, plasticizer and other background ions introduced during sample handling. Additionally, each small molecule typically produces redundant features, resulting from naturally occurring isotopes, and adducts, dimers and fragment ions commonly generated during electrospray ionization. Fragmentation spectra of background ions and redundant signals detract from metabolite identification efforts, highlighting the need to focus fragmentation on sample-relevant ions.

Traditionally, background contaminants and redundant signals have been annotated after data collection[6,7]. Stable isotope labeling approaches have been used to facilitate annotation of biologically relevant features. Several approaches have been published and include mixing of labeled and unlabeled samples to improve quantitation and identification of metabolites of biological origin. Isotopic Ratio Outlier Analysis (IROA)[8] involves parallel cultures, in which cells are grown in either 5% or 95% ^{13}C enriched media. Metabolites are then mixed in a 1:1 ratio, delivering a diagnostic mirrored isotopic pattern for naturally occurring compounds. The distance between corresponding peaks in this mirrored pattern reveals the carbon content of that metabolite, aiding in identification. Peaks that do not show this pattern are not biologically derived and can be ignored. Credentialing, invented by Mahieu et al in 2014[9], includes two parallel *E. coli* cultures; one uses unlabeled glucose, and the other features ^{13}C glucose as its sole carbon source. The two cultures are mixed at two distinct ratios, 1:1 and 1:2. The resulting LC-MS data are searched for pairs of co-eluting peaks with intensities reflecting the mixing ratio. Features not passing the ratio-criteria are filtered out as not sample-relevant.

Both IROA and the credentialing approach offer a mechanism to annotate features of non-biological origin, thus enabling identification efforts to focus on the truly sample-relevant features. However, these two methods still have major limitations. Both are based on cell culture and are therefore not amenable to other biologically relevant samples such as biofluids and tissue samples. Additionally, the use of isotopically labeled media may be cost-prohibitive and introduces additional complexity in the data, complicating and prolonging the data processing steps of the analysis. We set out to develop automated, knowledge-driven data reduction capabilities onboard a Thermo Scientific™ Orbitrap™ mass spectrometer that identifies non-biological and redundant features during acquisition. AcquireX intelligent acquisition[10], without relying on cell-based assays and isotopically labeled media, automatically annotates background contaminants and redundant signals on-the-fly. This data-informed approach maximizes the number of biologically relevant metabolites interrogated by MS/MS resulting in high confidence identifications and deeper metabolome coverage.

AcquireX — a new acquisition paradigm

AcquireX intelligent data acquisition is available on both the Thermo Scientific™ Orbitrap Exploris™ series of mass spectrometers[11] and Thermo Scientific™ Orbitrap™ Tribrid™[12] series of mass spectrometers and provides a fully automated approach to perform exhaustive fragmentation of sample-relevant compounds. There are four major workflows available in the AcquireX acquisition that utilize automatically generated ion exclusion or inclusion lists or both[13]. This automated process provides knowledge-driven, real-time precursor selection for subsequent data dependent fragmentation.

Figure 1 illustrates the Deep Scan workflow and is described in more detail here. The first step in the sequence is the injection of a blank sample with a full scan method. This can be a solvent blank or an extraction blank. Once the blank injection is finished, feature detection occurs and an exclusion list is automatically generated

recording *m/z*, retention time and intensity for each ion. This first blank injection is used to annotate background ions and exclude them from fragmentation. The next step is to inject the sample using a full scan method. This can be a pooled sample or any sample for which exhaustive fragmentation of all sample-relevant compounds is desired. Once the LC-MS injection of the sample is finished, peak detection occurs and an inclusion list is generated, automatically recording *m/z*, retention time and intensity for each ion. After both inclusion and exclusion lists have been generated, the sample is injected to collect fragmentation data. Following the rules of data-dependent acquisition, a single precursor is selected using a narrow isolation window and resulting in a highly pure spectrum. Precursors are selected in order of decreasing abundance, starting with those matching entries on the inclusion list. Once the injection is completed, inclusion and exclusion lists are updated, based on what was fragmented during this first injection, so that different precursor ions may be selected during subsequent injections. By doing this with every re-injection, more unique compounds of lower abundance are fragmented, digging deeper into the sample.

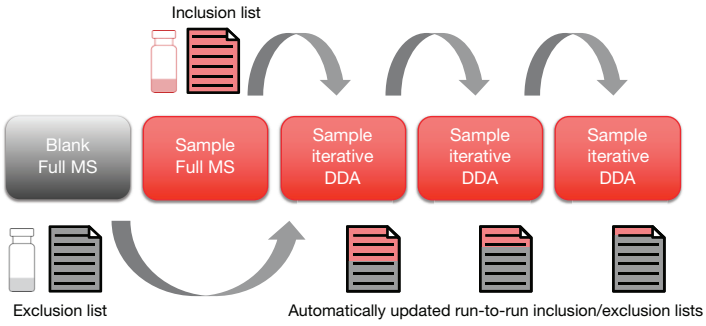


Figure 1. The AcquireX Deep Scan workflow utilizes automatically generated exclusion and inclusion lists through a series of iterative injections to improve MS/MS sampling and focus acquisition on true sample components.

Excluding background ions to reduce the acquisition of irrelevant spectra

Background ions include artifacts, such as electronic noise from the mass spectrometer, and contaminants, such as solvent impurities, coating agents from sample containers, and carry over from previous injections. They can be the most abundant peaks in the mass spectra and include both constant background ions, present throughout the chromatogram, and those background ions that elute as a chromatographic peak at a specific time (Figure 2). In the AcquireX Deep Scan workflow, the blank injection is utilized to distinguish background ions, i.e. those non-biological features that do not originate from the sample. If an ion is detected in both the blank and the sample within a defined

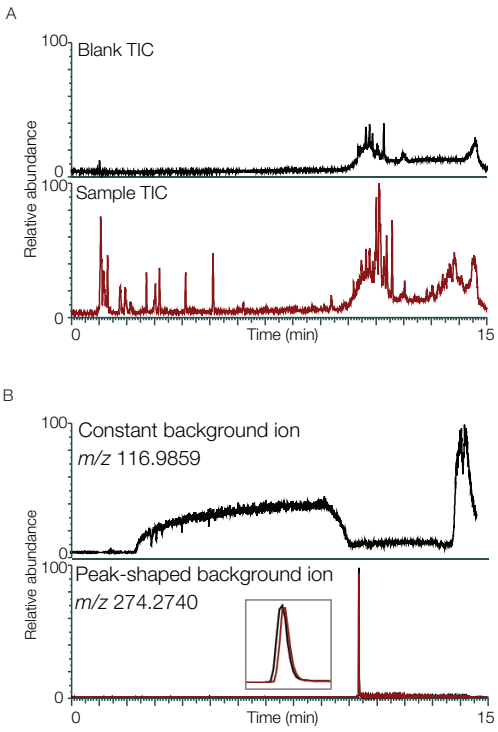


Figure 2. Background ions are highly prevalent in metabolomics experiments. A) Total ion chromatogram of blank (top) and NIST SRM1950 human plasma sample (bottom). B) Selected ion chromatogram of constant background ion *m/z* 116.9859 detected during LC-MS analysis of NIST SRM1950 human plasma sample (top). The bottom panel shows an overlay of the selected ion chromatogram of peak-shaped background ion *m/z* 274.2740 detected in both the blank injection (black trace) and the NIST SRM1950 human plasma injection (red trace) with similar intensity.

m/z and retention time tolerance, it will be denoted as non-biological and excluded from fragmentation when its intensity in the blank and the sample are similar, like in the case of background ion *m/z* 274.2740 in Figure 2. If its intensity is higher in the sample by a user-defined factor than the blank, then it is considered a unique biological feature and is targeted for fragmentation.

During traditional data-dependent MS/MS, ions are selected based on abundance, without any knowledge of biological relevance or type of ion. Often, irrelevant spectra resulting from background ions dominate the duty cycle, limiting the capacity of the instrument to acquire informative spectra. For example, the constant background ion *m/z* 116.9859 from Figure 2 was fragmented 252 times in a single DDA injection. In a representative DDA experiment, >76% of MS/MS spectra acquired by the instrument could be attributed to background ions (Figure 3). By enabling the automatic generation and implementation of a background exclusion list based on real-time feature detection in LC-MS data, AcquireX Deep Scan practically eliminates the acquisition of background ion MS/MS spectra in the same sample analysis, allowing for the fragmentation of more true sample components.

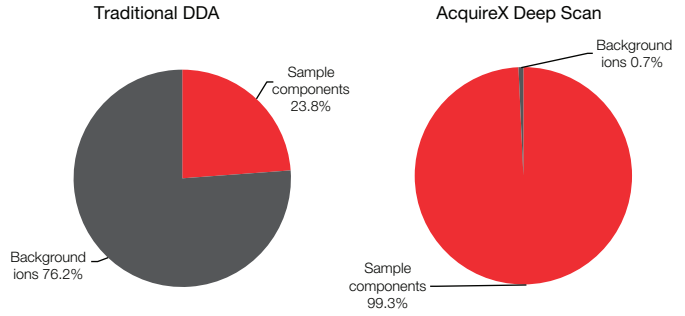


Figure 3. During traditional DDA, background ions dominate the duty cycle and >76% of MS/MS scans could be attributed to background ions, when NIST SRM1950 human plasma was analyzed with 3 DDA injections. With AcquireX Deep Scan, acquisition is focused on true sample components and fragmentation of background ions is minimized. The full circle represents the total number of MS/MS scans in the analysis of NIST SRM 1950 after 3 injections. The MS/MS scans corresponding to background ions and sample components are colored grey and red, respectively.

Minimizing redundant fragmentation and enabling fragmentation of low abundance compounds

Small molecules typically produce more than one feature. Figure 4 illustrates the various features corresponding to tryptophan detected in the LC/MS analysis of plasma. The protonated molecular ion is detected at *m/z* 205.09698. Features corresponding to the naturally occurring isotopes of tryptophan are highlighted with green boxes at *m/z* 206.10039 and *m/z* 207.10385 for the ^{13}C and $^{13}C_2$ isotopes, respectively. Naturally occurring isotopes are well defined, can be calculated based on elemental composition, and thus, can easily be excluded from fragmentation. Unintentional fragmentation may also generate additional features, often resulting from water or ammonia losses. Another common type of feature is produced from the adduction of molecules with a metal cation, such as sodium or potassium, or with another small molecule of the same or different chemical composition resulting in homo- or heterodimers, respectively. Fragmentation of in-source fragments or dimers of the compound as exemplified in Figure 4B for tryptophan, result in spectra very similar to that of the molecular ion. However, fragmentation of

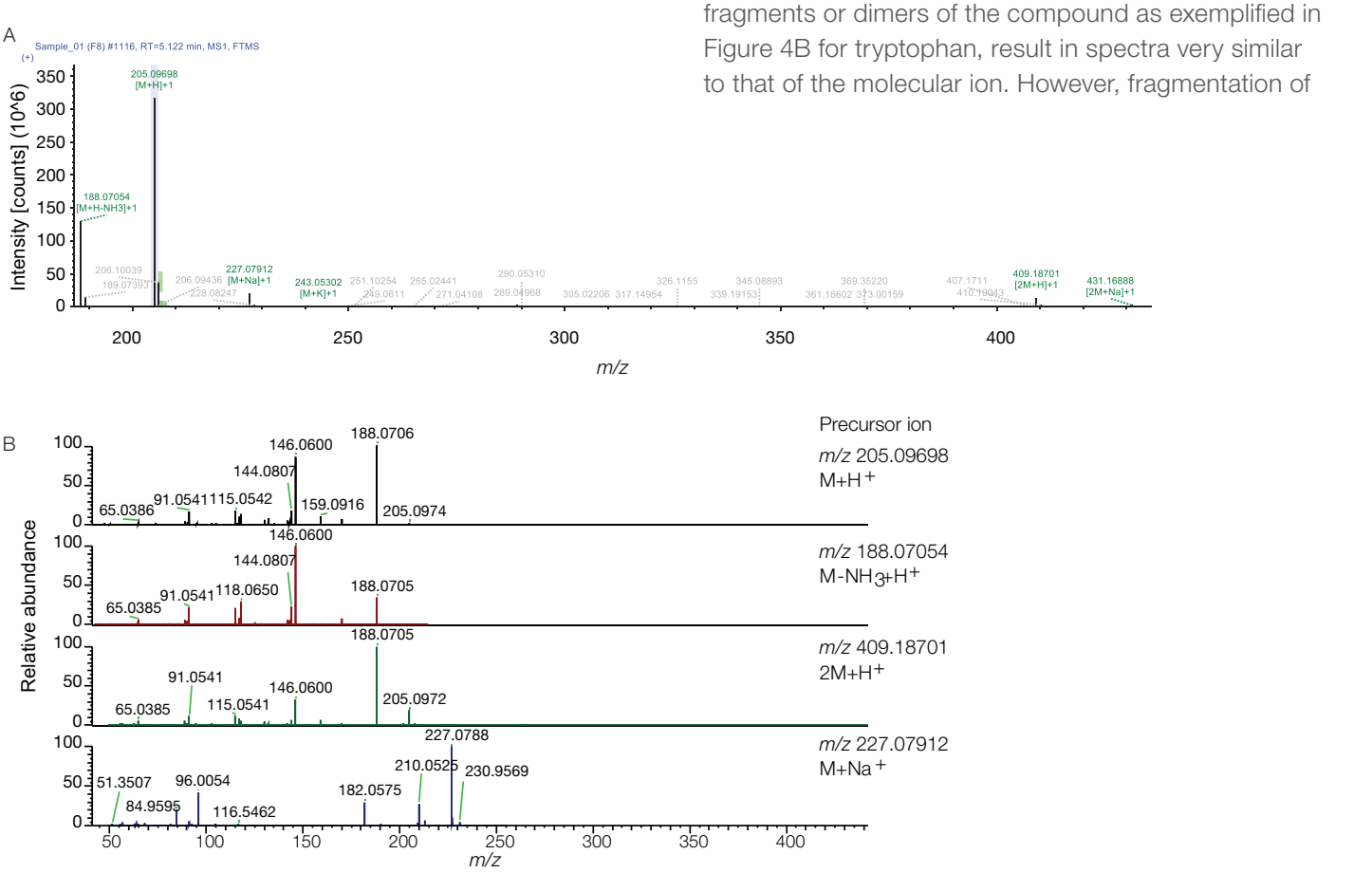


Figure 4. A) MS spectrum of tryptophan, as annotated by Thermo Scientific™ Compound Discoverer™ software. The molecular ion is highlighted in lavender. Additional features detected include naturally occurring isotopes (green boxes), in-source fragment resulting from ammonia loss, adducts with sodium and potassium and dimers. B) the MS/MS spectra of in-source fragment *m/z* 188.07054 and dimer *m/z* 409.18701 are very similar to that of the molecular ion, but sodium adduct *m/z* 227.07912 results in a spectrum that is hard to interpret and cannot be matched to the library spectrum of the molecular ion.

sodium adducts is often unproductive, resulting in noisy spectra that are hard to interpret. Fragmentation spectra of sodium adducts are very different than that of the molecular ion, often resulting in no library spectral match, as spectral libraries are mostly populated with fragmentation spectra of molecular ions only. With AcquireX intelligent acquisition, the user has the option of selecting specific ion species, such as protonated molecular ions, and prioritizing them over adducts such as the sodium adducts, that are less likely to produce a quality spectrum.

Highly abundant compounds, in the form of a parent ion or any of its accompanying features, such as isotopes and adducts, may prevent the fragmentation of metabolites of lower abundance. By populating the inclusion list with the preferred ion for each metabolite, more compounds can be sampled by MS/MS in a single run. Additionally, by automatically updating inclusion and exclusion lists after each injection during analysis, we can ensure that compounds not selected for MS/MS will be prioritized during a subsequent injection. In traditional DDA workflows, each injection is independent of the previous one, resulting in redundant fragmentation spectra. With the AcquireX Deep Scan workflow, inclusion and exclusion lists are automatically updated after each injection, minimizing redundant fragmentation and allowing for more analytes of lower abundance to be sampled with subsequent injections (Figure 5). In the analysis of non-smokers' urine (NIST SRM3673), traditional DDA obtained fragmentation for the most abundant compounds and re-injection of the same sample did not significantly affect the number and intensity of precursors fragmented. Unlike traditional DDA workflows, AcquireX Deep Scan avoids redundant fragmentation, in favor of interrogating new compounds, resulting in a 10-fold decrease in average precursor intensity of compounds fragmented from injection 1 to injection 4 of the urine SRM.

To further demonstrate the ability of the AcquireX Deep Scan workflow to generate high quality MS/MS spectra for low abundance metabolites, a series of experiments was performed by spiking the isotopically labeled Avanti SPLASH® LIPIDOMIX® standard mixture into bovine liver extract at varying concentrations, and analyzed using traditional DDA and the AcquireX Deep Scan workflow. Figure 6 shows the detection of 15:0-18:1(d7) phosphoethanolamine (PE) spiked at a concentration of 0.05 µg/mL in bovine liver extract. It is one of the least abundant ions in the spectrum, with ions from the sample matrix significantly more abundant. The pink triangle indicates that the PE ion was selected for MS/MS and panel B shows the acquired spectrum that contains the characteristic neutral loss (141 Da) for that lipid class. Traditional DDA does acquire MS/MS for this spiked-in standard at high concentrations, but AcquireX Deep Scan is able to trigger MS/MS at concentrations 10x lower than DDA (Figure 6, panel C). With the use of automatically updated inclusion and exclusion lists, the AcquireX Deep Scan workflow effectively increases the dynamic range of precursor sampling for MS/MS in comparison to DDA.

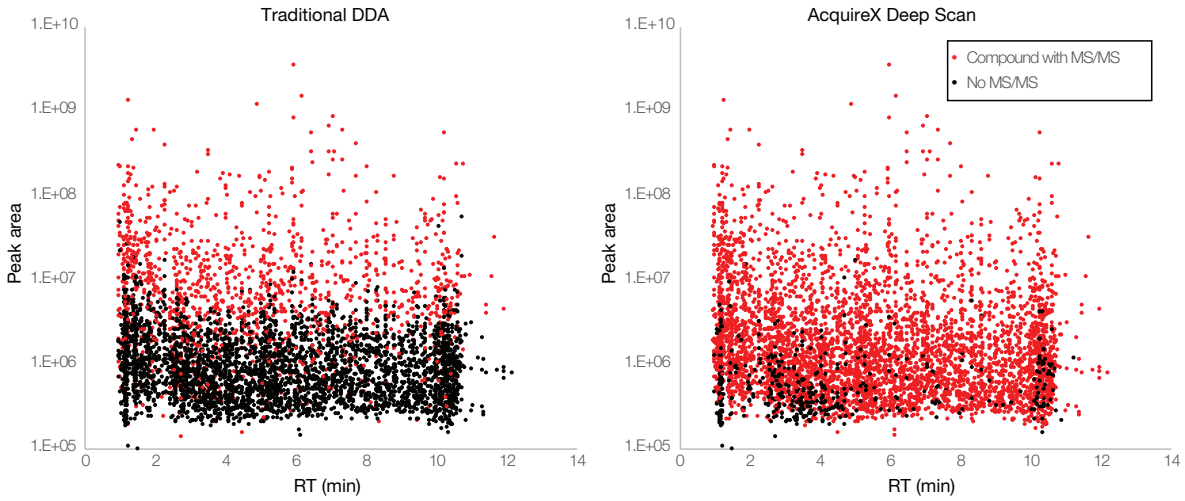
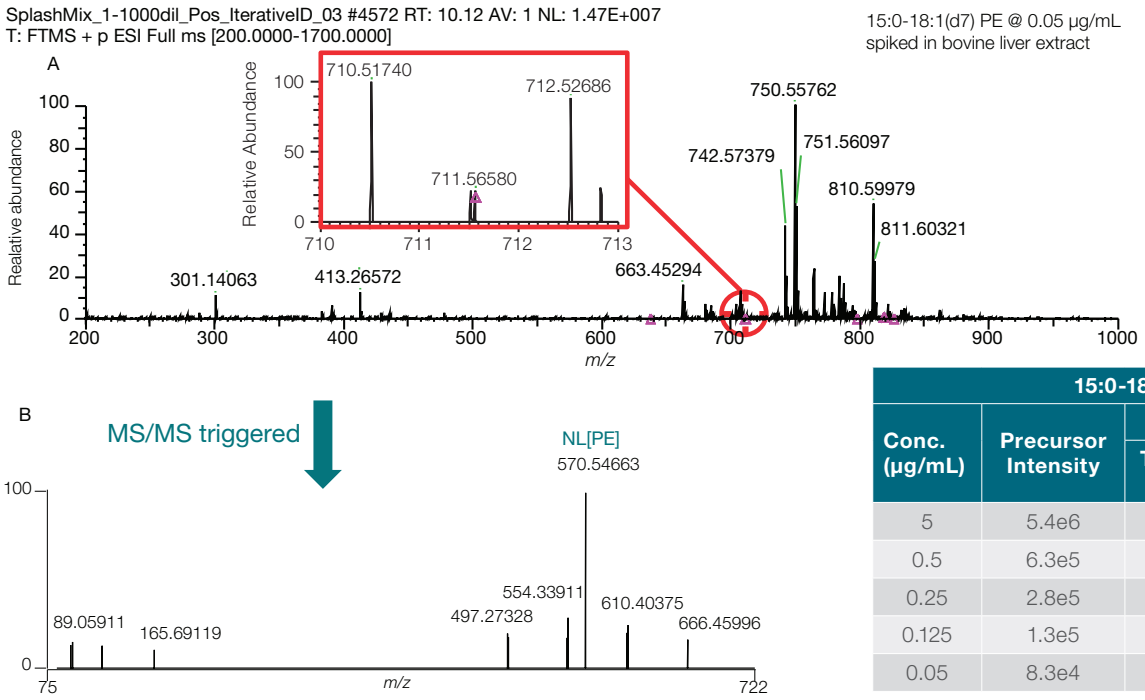


Figure 5. Comparison of traditional DDA (left) and the AcquireX Deep Scan workflow (right) for SRM3673 non-smokers' urine shows increased depth of MS/MS fragmentation coverage with the AcquireX acquisitions, as illustrated after four injections. Each dot represents a detected compound from the urine SRM, with their peak area as the y-axis and retention time as the x-axis. Compounds that have no fragmentation data associated are black, and those with fragmentation data collected are red.



15:0-18:1(d7) PE			
Conc. (µg/mL)	Precursor Intensity	Triggered for MS² (3 inj.)	
		Traditional DDA	AcquireX Deep Scan
5	5.4e6	✓	✓
0.5	6.3e5	✓	✓
0.25	2.8e5	×	✓
0.125	1.3e5	×	✓
0.05	8.3e4	×	✓

Figure 6. AcquireX Deep Scan workflow increases the dynamic range of precursor sampling by MS/MS. A) Detection of 15:0-18:1(d7) PE (theoretical *m/z* 711.56642) in bovine liver extract at a concentration of 0.05 µg/mL. Inset is the zoomed-in spectrum and the pink triangle indicates the ion selected for MS/MS. B) MS/MS spectrum of 15:0-18:1(d7) PE at that concentration. C) Table summarizes the dilution series experiment performed, showing the AcquireX Deep Scan workflow efficiently fragments this standard at a 10-fold lower concentration than traditional DDA.

Getting fragmentation for more unique compounds

As detailed above, the AcquireX Deep Scan workflow selects precursors intelligently by excluding background ions and targeting unique metabolites of biological relevance. Additionally, by automatically updating inclusion and exclusion lists between injections, compounds not previously selected for MS/MS will be prioritized during a subsequent injection. As a result, the AcquireX Deep Scan workflow not only yields more fragmentation of unique compounds in comparison to traditional DDA after one injection, but the MS/MS coverage of the metabolome also significantly increases with subsequent injections. This is independent of sample type and complexity, as illustrated in Figure 7 for up to 3 reinjections.

Improved MS/MS sampling translates to more compound annotations

The AcquireX Deep Scan workflow results in the interrogation of more unique precursors in a sample, which in turn translates to more compounds annotated. One way to confidently annotate compounds is through a spectral match against a library, such as the Thermo Scientific™ mzCloud spectral library (Figure 8). Additionally, even for spectra not matching the mzCloud library, de novo structure elucidation by chemical formula and structure assignment of fragment ions is possible[14].

AcquireX Deep Scan efficiently collects more data allowing for multiple dissociation techniques and multi-stage fragmentation in the same injection without compromising metabolome coverage. Duty cycle is no longer a limitation, as any compounds not selected during the first injection can be prioritized during subsequent injections. Orbitrap Tribrid mass spectrometers offer multistage fragmentation (MS^n) and the flexibility of complementary dissociation techniques to increase the probability of generating information-rich spectra and enable the differentiation of isomers (Figure 9).

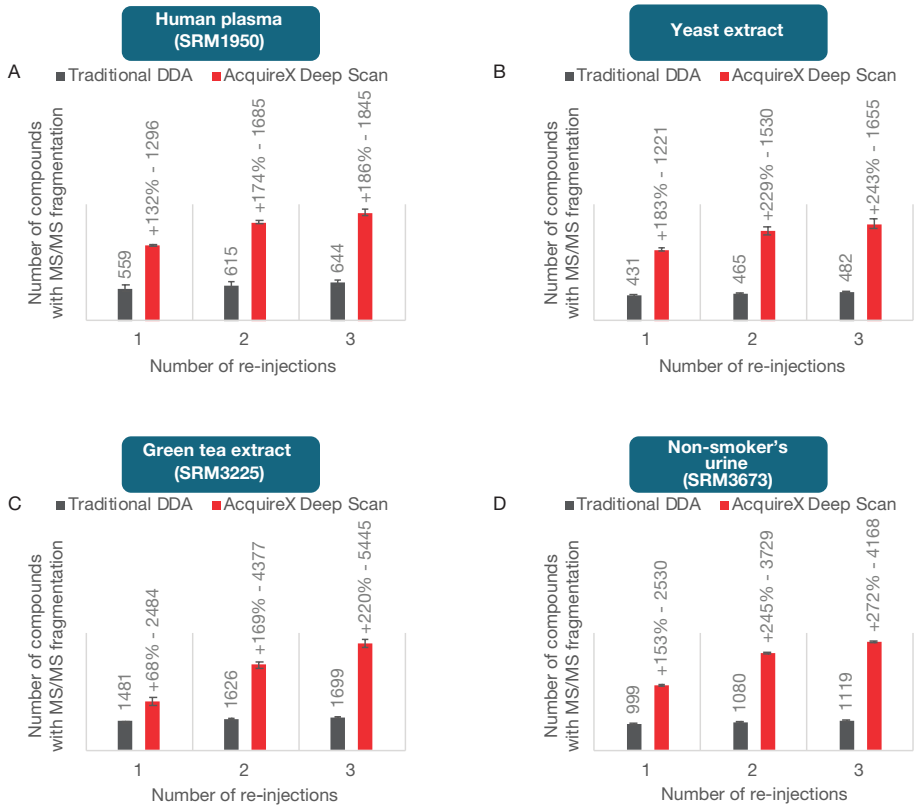


Figure 7. AcquireX Deep Scan workflow increases the number of unique compounds with fragmentation data independent of sample type or complexity. When compared to traditional DDA, after 3 injections, the number of unique compounds fragmented approximately tripled for human plasma (A), yeast extract (B), green tea extract (C), and urine (D).

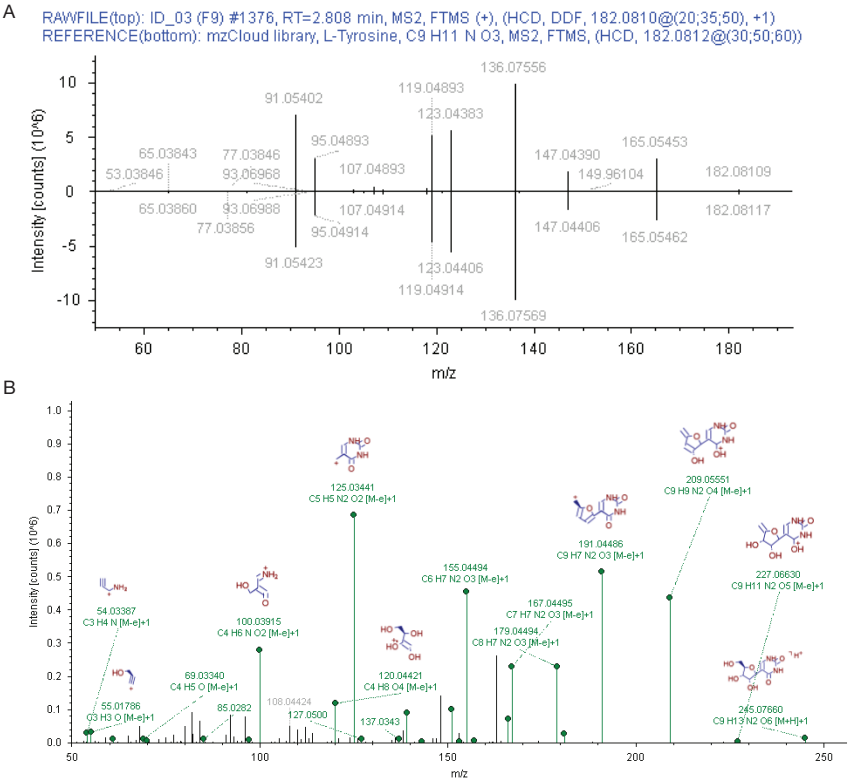


Figure 8. AcquireX Deep Scan workflow results in MS/MS spectra of high quality that can be matched to a spectral library like the example of tyrosine shown in panel A, or can be used for de novo structure elucidation by assigning chemical formulas and structures to fragment ions, such as in the case of pseudouridine in panel B.

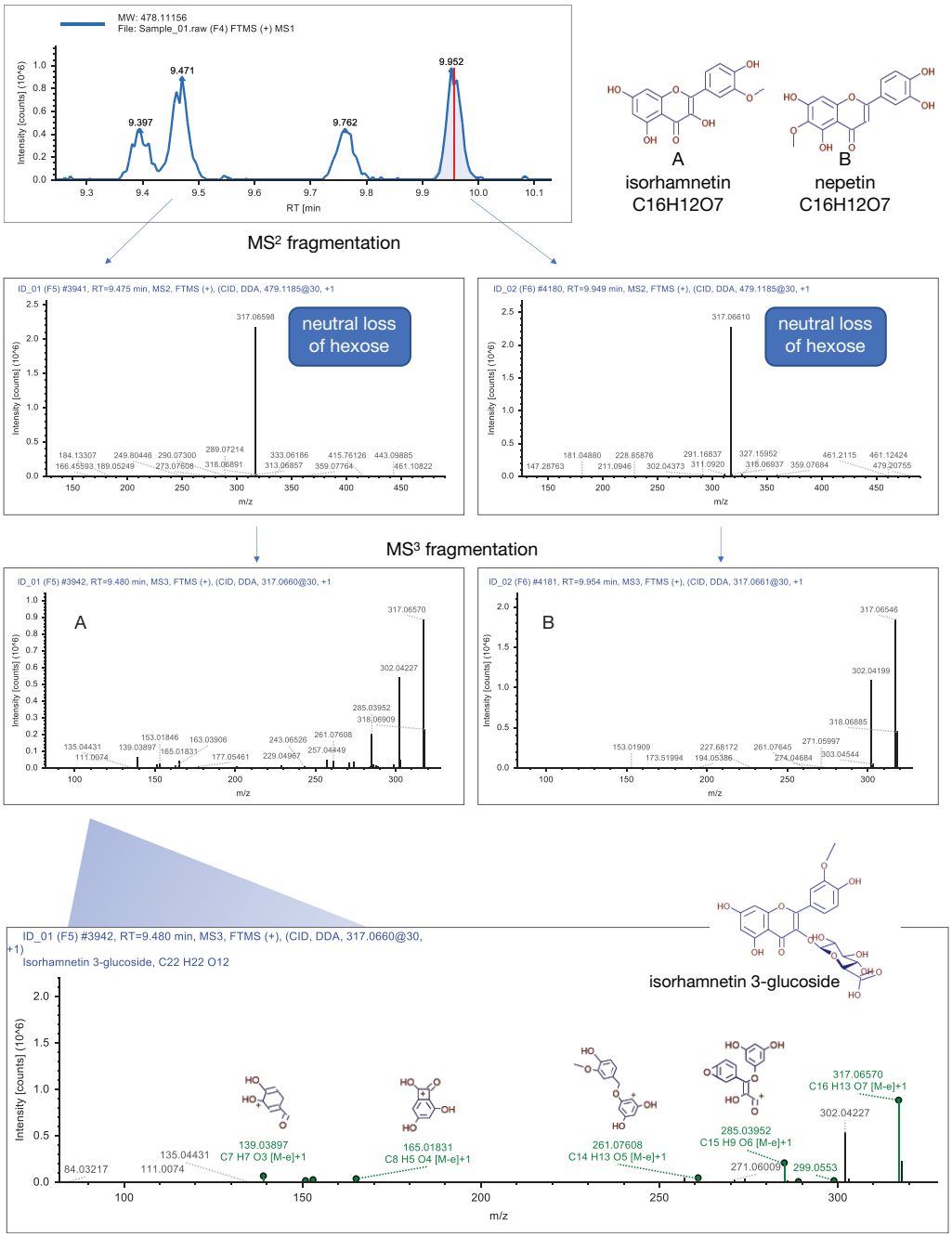


Figure 9. AcquireX Deep Scan workflow more efficiently selects precursors enabling the collection of MS^n fragmentation for more confident annotations of isomeric species. Shown here are two isomeric compounds (flavonoids) detected in the green tea extract (SRM3225) with the same monoisotopic precursor mass, and MS^2 spectra, but different MS^3 spectra. Searching the MS^3 spectra against the mzCloud spectral library provided matches to isorhamnetin (A) and nepetin (B), respectively.

thermo scientific

Conclusions

The recently developed AcquireX intelligent data acquisition can automatically determine features corresponding to background contaminants and compound degeneracy, such as isotopes, adducts, and dimers, enabling efficient MS/MS and MS^n sampling of unique biologically relevant metabolites. Unlike traditional DDA, during which fragmentation of background ions dominate the duty cycle, the AcquireX Deep Scan workflow selects precursors intelligently by excluding background ions and targeting unique metabolites of biological relevance for fragmentation. By excluding background and degenerate signals, the total number of fragmentation targets is reduced without compromising metabolite coverage, effectively increasing the dynamic range of MS/MS sampling. By focusing acquisition on biologically relevant compounds, more cycle time can be spent collecting multistage (MS^n) fragmentation data to provide additional structural information and confidence in compound annotations, without affecting experiment length. Ultimately, AcquireX intelligent acquisition automatically omits fragmentation of non-biological and redundant features, resulting in comprehensive MS^n coverage regardless of sample type and complexity.

References

1. Mahieu, N. G. and Patti, G. J. (2017). Systems-level annotation of a metabolomics data set reduces 25,000 features to fewer than 1,000 unique metabolites. Anal. Chem. 89, 10397-10406.
2. Sindelar, M. and Patti, G. J. (2020). Chemical Discovery in the Era of Metabolomics. J. Amer. Chem. Soc. 142, 9097-9105.
3. Viant, M. R., Kurland, I. J., Jones, M. R. and Dunn, W. B. (2017). How close are we to complete annotation of metabolomes? Curr. Opin. Chem. Biol. 36, 64-69.
4. Fenaille, F., Barbier Saint-Hilaire, P., Rousseau, K. and Junot C. (2017). Data acquisition workflows in liquid chromatography coupled to high resolution mass spectrometry-based metabolomics: Where do we stand? J Chromatogr A. 1526, 1-12.
5. Guo, J. and Huan, T. (2020). Comparison of Full-Scan, Data-Dependent, and Data-Independent Acquisition Modes in Liquid Chromatography-Mass Spectrometry Based Untargeted Metabolomics. Anal. Chem. 92, 8072-8080.
6. Souza, A. and Tautenhahn, R. (2018). Technical Note 65204: Features or compounds? A data reduction for untargeted metabolomics to generate meaningful data.
7. Souza, A., Ntai, I. and Tautenhahn, R. (2018). Application Note 65362: Accelerated unknown compound annotation with confidence: from spectra to structure in untargeted metabolomics experiments.
8. de Jong, F. A. and Beecher C. (2012). Addressing the current bottlenecks of metabolomics: Isotopic Ratio Outlier Analysis, an isotopic-labeling technique for accurate biochemical profiling. Bioanalysis. 4, 2303-14.
9. Mahieu, N. G., Huang, X., Chen, Y.-J. and Patti, G. J. (2014). Credentialing features: a platform to benchmark and optimize untargeted metabolomic methods. Anal. Chem. 86, 9583–9589.
10. Smartnote (2019). AcquireX Intelligent Data Acquisition Technology. (https://assets.thermofisher.com/TFS-Assets/CMD/brochures/sn-65392-ms-acquirex-intelligent-data-acquisition-sn65392-en.pdf).
11. Bromirski, M. and Robitaille, A. (2020). White Paper 65937: Grant Application Resource: Accelerating high-confidence insights for small- and large-molecule research using the Orbitrap Exploris 240 mass spectrometer.
12. Thermo Fisher Scientific (2018). White Paper 65364: Grant Application Resource: Thermo Scientific Orbitrap ID-X Tribrid mass spectrometer for metabolomics, lipidomics and structural elucidation.
13. Thermo Scientific Metabolomics brochure 2021.
14. Smart note. mzLogic data analysis algorithm. https://assets.thermofisher.com/TFS-Assets/CMD/brochures/sn-65387-ms-mzlogic-data-analysis-algorithm-sn65387-en.pdf.

Find out more at **thermofisher.com/acquireX**

ThermoFisher SCIENTIFIC