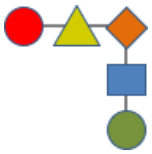
TruQuant Yeast Extract Semi-targeted QC Workflow Kit



Measurement and Normali*zation System for semi-targeted MS analysis*

Identification*, *quantitation*, *suppression-correction*, *normalization* and *Platform QC

of metabolites across time and instruments

Kit contents

Materials and tools for the analysis of 90 experimental samples

Unique fully-labeled *S. Cerevisiae* Yeast Extract*

- 3 vials of lyophilized IROA-IS (Internal Standard); U-^{13}C, 95%
- 3 vials of lyophilized IROA-LTRS (Long-Term Reference Standard); paired U-^{13}C, 95% and 5%; mixed 1:1
- ClusterFinder™ software (including library of IROA peaks in the LTRS and their physical characteristics)
- User manual

Storage: -80ºC, protected from light

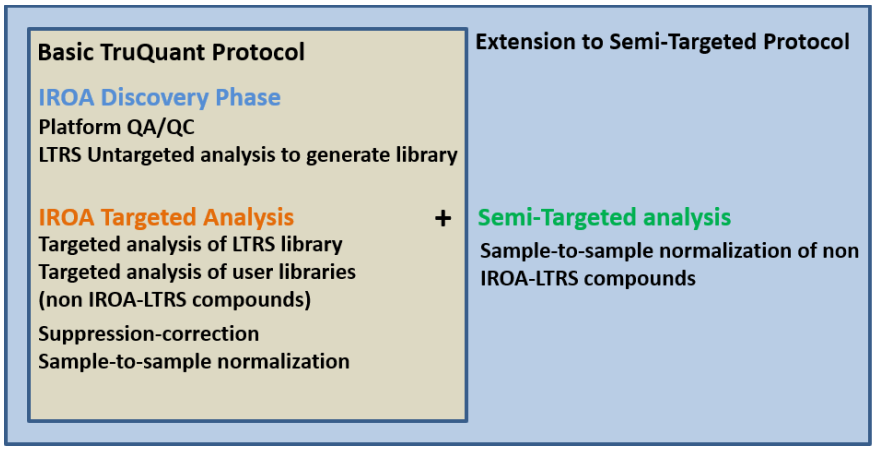
Summary of Benefits of the TruQuant Measurement System

1. **Daily instrument QA/QC.**
2. **No false data. Reports only compounds of biological origin; excludes artifactual peaks.**
3. **Accurate compound formula ID for MS alone; complete ID with addition of SWATH, MS/MS, or IM.**
4. **Fragments** have the IROA ratio pattern derived from their parent peaks and can be **identified** using the **"peak correlation"** ClusterFinder module.
5. **Semi-targeted.** Software searches for the highly characterized 1000+ known IROA LTRS peaks and also will search for unlabeled compound peaks using user generated natural abundance libraries.
6. **Suppression-corrected measurements for significantly better quantitation.**
7. **Reproducibility. Sample normalization to a universal standard for complete comparability.**
8. **Batch-to-batch correction**
9. **ClusterFinder software solution** builds libraries, IDs/quantitates compounds, corrects for ion-

The IROA TruQuant Workflow (ref 1-7, 9) is a protocol based on demonstrated principles that serve to correct the systemic errors introduced during sample preparation (using Dual-MSTUS normalization) or within the mass spectral instrumentation (using suppression-correction algorithms).

*Prepared from proprietary U-5%/U-95% ^{13}C IROA-labeled glucoses specially produced for IROA Technologies by Cambridge Isotope Laboratories, Inc.

We are pleased to extend the TruQuant Protocol to a Semi-Targeted Protocol



In Discovery Phase:

- the IROA-LTRS (containing both 5% and 95% 13C isotopolog ladders) is analyzed using an untargeted analysis with MSMS or SWATH to differentiate isobaric compounds. The formulae for all IROA peaks are determined by their peak (isotopolog) patterns.
- The characteristics of parent, fragment and adduct ions for all LTRS peaks and many additional compounds are available through a ClusterFinder pre-built library (without RT).
- A curated collection of historical LTRS peaks may be saved as a method specific LTRS library with RT. Once established this will become the dominant library.

In Semi-Targeted Phase:

- The properties of compounds identified in the LTRS samples are used to identify all of the same compounds in the experimental samples which contain a chemically identical internal standard (IROA-IS – 95% 13C isotopolog ladders only) in a Targeted analysis to achieve a non-sparse data set. In a second pass ClusterFinder will additionally collect peaks of any user requested compounds (1.1% 12C) which are not accompanied by an internal standard. These compounds cannot be corrected for ion losses (suppression) but can be normalized.

3 STEP PROCESS:

1. LC-MS analysis of experimental samples resolvated with IROA-IS (experimental sample + IROA IS = analytical sample) are randomly analyzed in quantitative mode (without msms).

Within the random analytical injections, the IROA-LTRS is injected approx. once for every 10 analytical samples. The LTRS is analyzed in qualitative mode using msms untargeted analysis to collect characteristics of parent, adduct and fragment ions which are collected in a library format. The LTRS is also used as a QA/QC Standard to account for fluctuations in mass and RT drift.

2. Generate "dictionary" of RT, m/z, formula and physical characteristics from the analysis of IROA-LTRS using ClusterFinder software to validly identify all compounds in the IROA-LTRS.

3. Use ClusterFinder and the LTRS dictionary to identify experimental samples spiked with the same compounds (IROA-IS) in a Targeted analysis to achieve a non-sparse data set and quantitate, suppression-correct and normalize all compounds in original experimental samples. ClusterFinder also identifies and quantitates all user defined additional metabolites of interest in a Targeted search. These metabolites can be normalized but cannot be suppression-corrected.

Table of Contents

The IROA Peaks and Envelopes

Distinct signatures in the molecules of a biological sample are key for identification and quantitation.

The key to understanding the IROA methodology is that both ^{12}C and ^{13}C isotopes are created to be uniformly present at approximately 5% for one isotope and approximately 95% for the second isotope. The molecules labeled at 5% ^{13}C have a strongly enhanced M (monoisotopic)+1 and the molecules labeled at 95% ^{13}C a strongly enhanced M-1, creating a mirror-image of one another (see Figure 1). Using traditional comprehensive (>98%) labeling, the monoisotopic peak of most compounds can usually be detected even if its intensity is low, but the M+1/M-1 minor peaks can be easily lost. Where the ^{13}C is increased to 5% or 95%, the M+1 and M-1 peaks for a six-carbon molecule such as arginine in Figure 1 become significantly larger, namely 32% of the height of the monoisotopic peak. Whereas if the ^{13}C is present at only 1.1%, the height of the M+1 is only approximately 6% of the height of monoisotopic peak.

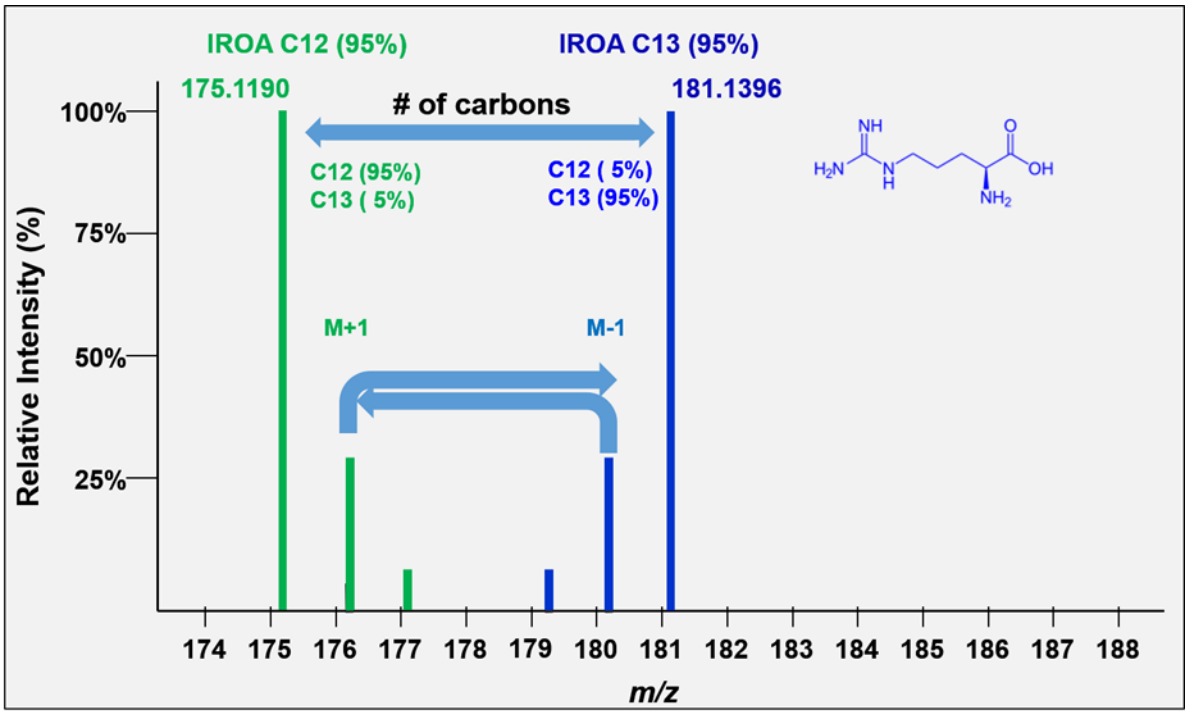
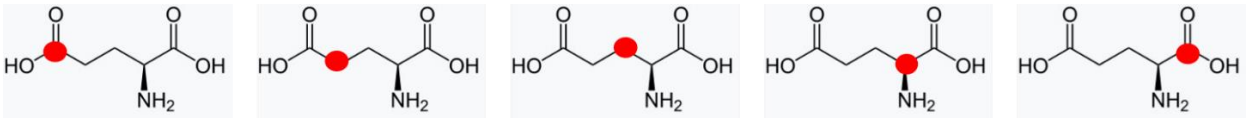


Figure 1. The IROA peaks. Molecule shown is the 6-carbon molecule arginine.
Green: Arginine C12 envelope peaks labeled with U-5% ^{13}C.
Blue: Arginine C13 envelope peaks labeled with U-95% ^{13}C.

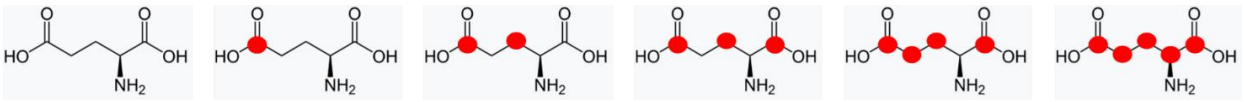
A short tutorial on isotopomers and Isotopologs

- **Isotopomers** are molecules with the same number of isotopes (regardless of position). There are 5 isotopomers of glutamic acid ($C_5H_9NO_4$) containing one ^{13}C.



Therefore, all isotopomers have the same number of isotopic atoms which will all have the same mass and will appear in a single Mass Spectral peak.

- **Isotopologs** are molecules with different numbers of isotopes. There are at least 6 isotopologs of glutamic acid, and aside from the first and last (the two monoisotopic isotopologs), they are not positionally defined.



Therefore, all isotopologs have the same structure but different number of isotopes and will appear in multiple Mass Spectral peaks, a collection of peaks separated by the mass of the appropriate extra neutron (in the case of carbon this is 1.00335 amu). Furthermore, most isotopolog peaks likely contain multiple isotopomers. Every chemical formula has N+1 carbon isotopologs where N is the number of carbons in its formula, beginning with its C12 monoisotopic (M) and ending with its C13 monoisotopic. This collection of isotopologs is referred to as an "isotopolog Ladder". Based on the number of carbons in the molecule and the weight of either the C12 or C13 monoisotopic peaks, the pattern of peaks in the isotopolog ladder is a unique identifier for every molecular formula (for molecules below 20 carbons).

The isotopolog ladder for all isomers of glutamic acid is shown in Figure 2 below.

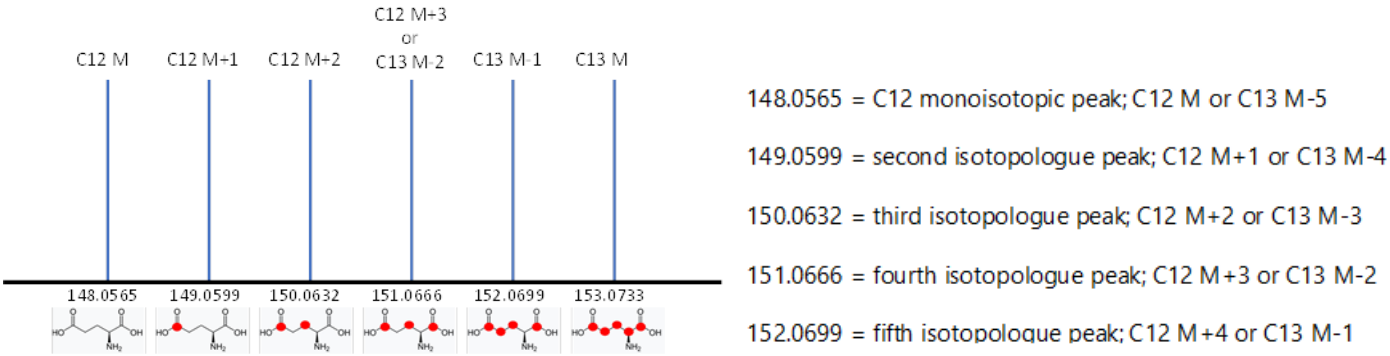


Figure 2. The Isotopomeric ladder for Glutamic acid (C11) as seen in RP pos LC-MS. Each isotopolog represents a specific mass by a molecular formula. In the case of carbon isotopologs, these are seen as a collected ladder of peaks extending from the C12 monoisotopic peak to the C13 monoisotopic peaks that differ in mass by 1.00335 amu. All molecules with six carbons will share this ladder but because each formula has a different mass each will begin and end at different masses. Therefore, each isotopolog ladder is unique to and representative of a single formula. (This is true for masses below 800, at a minimum.)

Specifically:

1) The isotopolog ladders for any formula are unique to that formula, but are common to all its isomers, e.g., Stereo, D/L, Structural, etc. However, the shape of the isotopolog peaks is defined by the relative percentages of the isotopes.

2) The Isotopomeric ladder for Glutamic acid (Figure 2) is shared by all molecules that have the same formula (C5H9NO4, e.g. isobaric compounds), including Glutamic acid, O-Actetyl serine, Threo-3-methyl aspartate, and N-Carboxymethyl alanine.

3) It is the job of chromatography (LC, GC, IM, etc.) to separate these isobaric compounds.

Isotopolog patterns are isotopic envelopes

Each IROA or natural abundance compound will show up in a mass spec as a collection of peaks that represent a mixture of isotopomers and isotopologs. For simplicity, we refer to the collection of peaks from the C13 side or the C12 side as an "isotopic envelope" (see Figure 4 D). Each molecular formula has a unique collection of possible peak positions, referred to as an "Isotopic Ladder" (see Figure 3).

Figure 3 shows the isotopolog ladder for tryptophan, which contains an equal concentration of "molecules" (1:1) for the right and left side set of peaks; natural abundance on the left, 95% ^{13}C on the right.

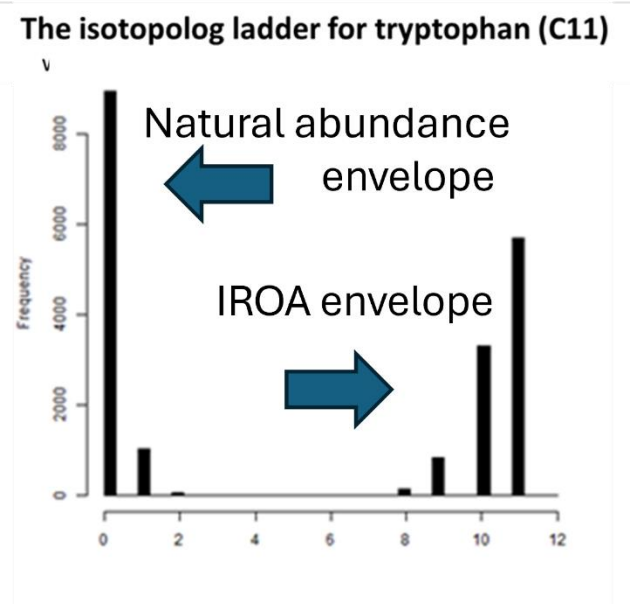


Figure 3. The Isotopomeric ladder for Tryptophan (C11).

The relative heights of each isotopolog peak in a ladder is determined by the isotopic balances of the source materials. The height of the peaks of isotopically defined compounds (enriched in a single element such as carbon) may be effectively calculated by the binomial expansion[1] of the expression ($^{12}C\%$ + $^{13}C\%)^N$ where N equals the number of carbons, and $^{12}C\%$ and $^{13}C\%$ equals the relative isotopic abundance.

[1] This is technically a polynomial expansion in which the dominance of carbon makes the remaining terms less important.

In this image we see a situation common in IROA, namely a ladder that has the isotopic signatures contributed from two sources; the first source on the left is the natural abundance compound, and the second source labeled at 95% ^{13}C on the right is its internal standard. These are presented at exactly equal concentrations of molecules from both sources however they are distributed very differently.

- Note that the height of the base peak is never indicative of concentration; rather **the sum of all peaks from each collection must be considered**.
- In the case of tryptophan, the base peak is still the C13 monoisotopic peak and represents only about half molecules in the internal standard.
- The base peak for an IROA compound with more than 20 carbons will no longer be the C13 monoisotopic peak. Instead, depending on the number of carbons, it will become one of the lower mass isotopologs. This is because as the number of atoms in a molecule increases, the probability that the entire molecule contains at least one heavy isotope also increases.

For this reason we need to consider **isotopic envelopes**.

Isotopic Envelopes

All molecules with the same number of carbons will show the same pattern of peaks but will differ in the mass of their monoisotopic peaks according to the remainder of the formula. We refer to these peak height patterns as the peak "*isotopic envelopes*". These envelopes are diagnostic for each formula.

The IROA carbon envelope shapes are readily and exactly calculable. The defining feature of the IROA carbon envelope is the mass of both monoisotopic peaks and the mass difference between them. The mass difference between the monoisotopic peaks is always a multiple of the mass of a neutron (~1.00335 amu). The additional peaks discussed above contribute to the extended isotopic envelope (M+2, M+3 etc., M-2, M-3 etc., see Figure 4) and the IROA ClusterFinder software can easily identify these peaks by their mass difference (the mass difference between a ^{13}C and ^{12}C isotope).

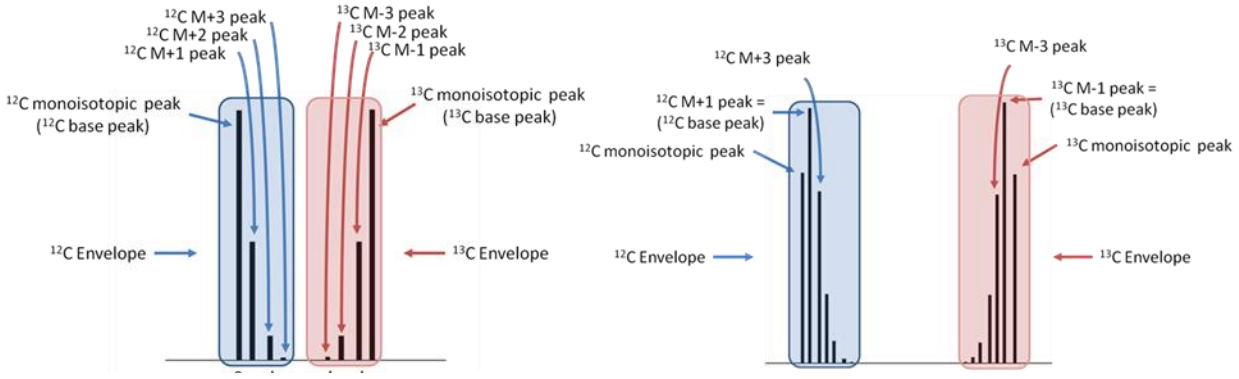


Figure 4. IROA isotopic envelopes (clusters) illustrating parts for a 9 carbon and 27 carbon molecule labeled with 5%13C:95%13C. Note the base peak shift. The monoisotopic peak for the 27-carbon molecule (right side of figure) is not the most abundant peak in the envelope.

Isomers, isotopomers, and isotopolog patterns become complex very quickly. No two isotopologs are the same, and any isotopolog will generally contain more than one isotopomer and each contains different collections of isotopomers. We find this exact language to be useful however cumbersome.

For the IROA-based Protocols we work with three different envelopes; the natural abundance envelope for experimental samples (~1.1% ^{13}C, starting from the ^{12}CM and going up its isotopolog ladder); the 5% ^{13}C envelope (in the LTRS only, starting from the ^{12}CM and going up its isotopolog ladder), and the 95% ^{13}C envelope (the Internal Standard and also in the LTRS and starting from the ^{13}CM and going down the isotopolog ladder (see figures 3, 4 A, 4 D and 7). In all these cases the percent of molecules in the monoisotopic peaks is a function of the number of carbons in the molecule.

Many people ignore the natural abundance M+1 isotopolog in calculating concentration (the AUC for this peak) because the error is reasonably small, and likely less than the variances caused by other factors (suppression, fragmentation, or injector errors). However, in the case of the 5% and 95% envelopes the percent of isotopologs in the monoisotopic peak varies much more strongly and to assure accurate quantitation their losses should not be ignored.

For instance, in a 6-carbon molecule, only 73% of the population will be in the M peak, for a 10-carbon molecule, only 60% of the population will be in the M peak, and for a 20-carbon molecule the M peak contains only 35% of the population. Therefore, if you are comparing ^{12}CM and ^{13}CM monoisotopic peaks of equal heights in an experimental sample (where the collected C13 peaks make up the internal standard and the C12 peaks are natural abundance) then for a 5 carbon molecule the concentration of a natural abundance peak is actually 137% of the ^{13}CM peak, for a 10 carbon the concentration of the ^{12}CM peak is actually 167% the concentration of the standard, and for a 20 carbon molecule the equal-heighted ^{12}CM peak is 286% the concentration of the standard. Note that not only is the error dramatically increasing and variable as the number of carbons varies but it makes these molecules incomparable. On the other hand, if one calculates the ratios based on the sum of the isotopologs from each envelope they will always be exactly equal and comparable.

Summary

- All **isotopomers** are isotopic isomers that share a single mass. There is no positional constraint.
- All **isotopologs** contain different numbers of isotopes but otherwise share a formula.
- In **IROA** there are always contributions from two different sources, usually an analyte and an internal standard. The peaks from both sources will superimpose onto the same ladder.
- An **isotopolog ladder** exists between the two monoisotopic peaks (from 2 different sources), and this ladder is diagnostic for the molecular formula that it is derived from.
- The collection of peaks donated by each source is grouped into a collection of isotopologs that is most easily identified as an **isotopic envelope**.
- **Accurate quantitation** requires summing all of the peaks in each envelope.

The IROA Standards

In the TruQuant Workflow two standards are employed: the **IROA-LTRS** and the **IROA-IS.**

The IROA Long Term Reference Standard (IROA-LTRS)

The IROA-LTRS, a Long-Term Reference Standard, is a fully labeled *S. cerevisiae* yeast cell extract IROA separately labeled at both 5% and 95% U-^{13}C and mixed 1:1, thus every compound is equally present at both 5% C13 and 95% C13 isotopomeric balances (equally true, at other isotopic balances, as needed). The result is a collection of over five hundred metabolites* that exhibit a unique, balanced, and symmetrical collection of peaks, with each half a mirror image of the other as represented by the peak pairs of the molecule arginine in Figure 1.

The LTRS is always constituted as the same carefully controlled mixture of compounds. Compounds may be visible at different concentrations either because their concentrations are below the "limit of detection" (LOD), or because they are severely suppressed below the LOD; however, in any given LTRS batch, each individual compound is always present at the same concentration in all aliquoted LTRS samples. All the compounds present in the LTRS sample are highly defined.

Because of the symmetry of the IROA peaks in the LTRS, LTRS samples can be completely cataloged; even peaks deep into mass spectral noise at extremely low levels well below what would otherwise be possible can be identified and characterized. The triple-redundancy of the Basic IROA peak guarantees consistent interpretation and identification in every analysis.

The LTRS samples are analyzed to find, identify, and collect all identifying physical characteristics for all of the compounds contained within it with extreme accuracy and sensitivity (see Figure 5). For every triply redundant IROA peak, the physical information may include but is not limited to information from the primary ms scans: the retention time (RT), ^{12}C monoisotopic mass, ^{13}C monoisotopic mass, number of carbons contained in the molecule (Figure 5 A'); in-source and post-source fragmentation characteristics (Figure 5 A3'); any physical characteristics gleaned from other methods applied to the effluent stream, for instance, IR, UV); various post source fragmentation methodologies, including for instance, Collision-induced dissociation (CID), Electron-Capture Dissociation (ECD), SWATH, etc., whether Directed (Data dependent Acquisition – DDA), Independent (Data Independent Acquisition (DIA), such as MSe, SWATH, etc. (Figure 5 A2'), Ion Mobility (IM) (Figure 5 A1'); or other collectable techniques that may provide information to support the identification of this IROA peak.

---

*IROA-LTRS contains a broad spectrum of metabolites including amino acids, organic acids, peptides, vitamins, carbohydrates, and co-enzymes. Over four hundred metabolites have been characterized in the IROA-LTRS Standard using ClusterFinder software. The final number is expected to reach between 700-800 as different analytical approaches enable the identity of more compounds.

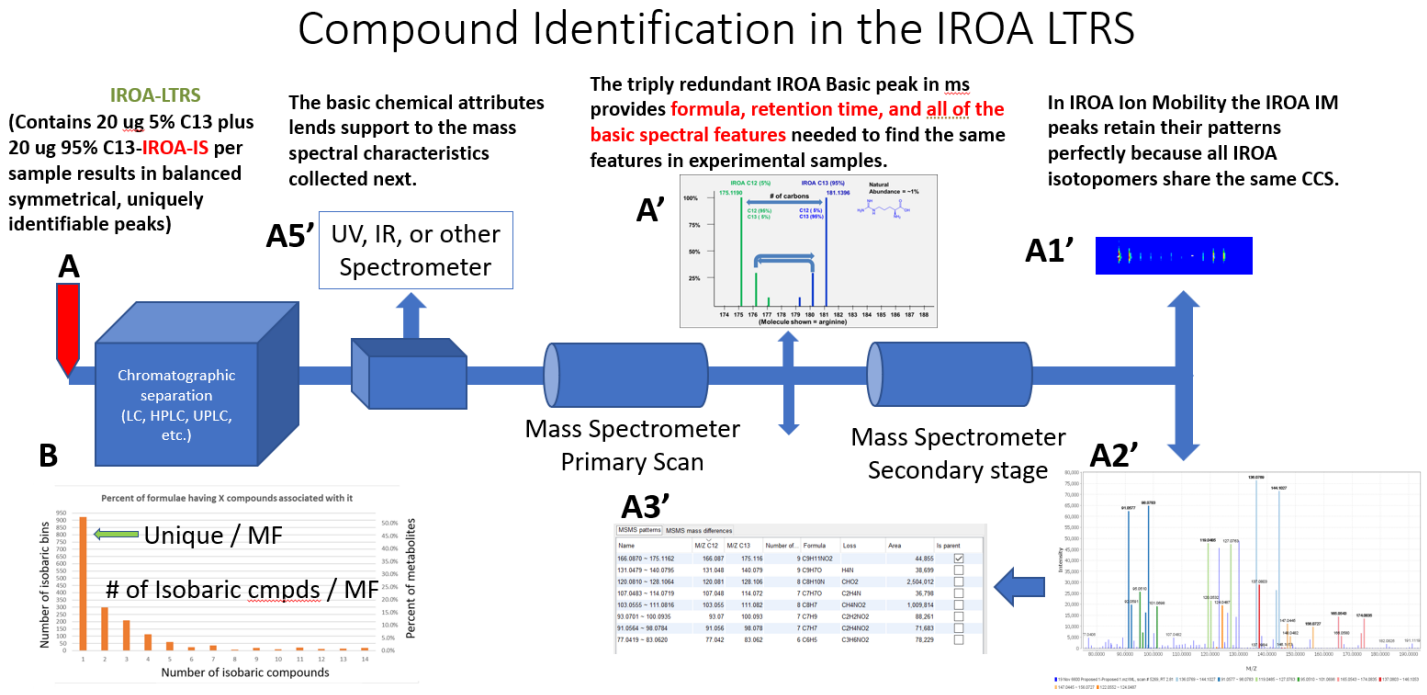


Figure 5. While the formula for any peak is available by analysis of the isotopolog ladder (see 3 A'), isobaric compounds can be discriminated either by its msms (A 2'), or CCS (A 1'). Note that in both cases the entire IROA patterns are seen and therefore, artifactual peaks may be automatically removed and the formula for all fragments are known allowing structure elucidation for unknown peaks (A 3').

Because all the compounds present in the LTRS samples may be cataloged, and because they are consistent in a given LTRS, their chromatographic behavior, ionization efficiency, Ion Mobility (IM) characteristics (Figure 5 A1'), fragmentation behavior (Figure 5 A2' and A3'), etc. may be evaluated and these values used to correct for any day-to-day variances, when the analytical system is similar, or even if it is very dissimilar.

Since the majority of these compounds will be found even across very different analytical platforms, i.e. with different chromatographic, ionization, or detection systems, the IROA characteristics of the IROA primary scan and the IROA secondary chemical characteristics, as seen in Ion Mobility, SWATH, or other fragmentation systems assure that every compound in LTRS may be mapped from any analytical system to any other analytical system; thereby providing a mechanism for directly comparing the complete LTRS chemical composition of any two LTRS samples, and through them any clinical or experimental samples they support.

The IROA-LTRS supports three major functions every day:

1) ***Library Building Standard*** – The IROA-LTRS (injected randomly within the analytical sample set) and IROA-IS (spiked into samples) have the same concentration of compounds and should have extremely reproducible RTs and peak characteristics across all samples.

   The LTRS sample is a single "always the same" sample containing 2 sets of identical compounds, one with 5% patterns and the other set having 95% ^{13}C patterns, mixed 1:1. Because of this, the isotopolog patterns seen for every compound is highly symmetrical as each compound exhibits both strong $^{12}CM+1$ and $^{13}CM-1$ peaks, respectively. Where the ladder is detected both envelopes are examined. The mass

difference between their respective monoisotopic peaks and the height of the ^{12}CM+1 and ^{13}CM-1 peaks all indicate the same number of carbons in the molecule, the peak is considered triply redundant, and the formula is locked down and virtually guaranteed to be correct. For the 52% of metabolite formulae that are unique this is sufficient to identify the compound with some certainty (see Figure 5 B). For the remaining 48% of isobaric compounds some additional elucidation will be required. Here again IROA provides a unique solution. If the entire isotopolog ladder (or even just it's two monoisotopic peaks) can be captured and fragmented using a "SWATH"-like approach all the true fragments will have IROA patterns, and therefore formulae, while stray ions from the fragmentation window can be ignored. This ability to exclude stray peaks and to know the formulaic relationship between two fragments goes a long way toward elucidation of any isobaric peak's identity and the removal of false data or unknowns

The IROA-LTRS is used to build a triply redundant dictionary (library) of RT, *m/z* and physical characteristics including fragmentation data. All peaks are named according to the IROA-LTRS database. The adducts and fragments for each compound are indexed individually. These compound IDs are loaded into ClusterFinder's internal databases as a separate editable IROA-LTRS database.

A library of IROA peaks in the LTRS and their physical characteristics is provided in the ClusterFinder™ software and distributed with the TruQuant kits. This library is distributed without retention time information; however, it is updateable by the user to assure that every compound in it will be reproducibly named according to the library.

2) **Validated Compound Identification** – The IROA-LTRS sample and ClusterFinder software and databases are used to create a daily RT library of all the compounds to be found in the IROA-IS so that their identification is reproducible and assured. The daily library is used as the basis of a targeted search of the IS in the analytical samples, and to quantify the natural abundance isotopologues of each experimental compound. For each compound, once the IROA-IS is found in any sample, a value, even <LOD, will be returned; therefore, the targeted search should yield a non-sparse dataset.

Since the chemical makeup and chromatographic behavior of the IROA-LTRS sample is identical to the IROA-IS, it is possible to use the in-depth and informationally strong, triply redundant chemical information obtained from the IROA-LTRS sample and apply it to identify compounds in the experimental sample. The dictionary catalog of all peak pairs, their RT, number of carbons, and IM and fragmentation characteristics provides information where each of these same IROA peaks will be found in the experimental samples through the use of the IROA-IS. The experimental natural abundance peaks are easily located and quantitated as they will co-locate with their corresponding IROA peaks at a mass that is the mass of the IROA C13 monoisotopic peak less the number of carbons it contains times the mass of a neutron.

All compound identities are validated in the IROA-LTRS where secondary physical characteristics such as fragmentation (ms/ms), ion mobility (ims/ms – CCS) may be collected to assure identity without lessening the quantitative aspects of the IS in the analytical sample.

3) ***QC Standard*** – In addition to providing for a rigorous mechanism for naming all the peaks in the LTRS and using this information to infer peak identities for their equivalent peaks in the IS (and analytical samples), the LTRS is a "standard sample", i.e. it is always the same, so it provides additional significant

benefits whenever it is used as a standard sample. It provides a daily instrument qualification check which could be the basis for a lab's QA/QC program. Using such a standard sample consistent results should be seen daily with respect to: 1) the number of peaks seen, 2) the daily chromatographic attributes of each peak (RT, shape and monoisotopic mass), and 3) source performance (rates of in-source fragmentation or other loss to parent ions, sensitivity, general transmission performance, etc.). In the IROA Workflow, the LTRS is injected multiple times during the analysis of the analytical samples, providing the QA/QC continuously throughout the experiment.

Following LC-MS, the paired compound IROA peaks (U-95% ^{13}C or U-5% ^{13}C) can be readily identified by the IROA ClusterFinder software. The IROA peaks represent actual compounds, fragments and adducts and can be discriminated from **unsigned artifacts and noise which can be removed from the data, eliminating false discoveries**. As a composite sample, sample-to-sample and analytical variance is removed and during MS analysis the identical compounds (labeled with either U-95% ^{13}C or U-5% ^{13}C) experience the same ionization efficiency and suppression.

Over 1000 peaks can be detected in the IROA-LTRS in pos mode (see Figure 6 D). The resulting IROA-LTRS dictionary of compounds is used to identify compounds in the IROA-IS saving time, effort and related costs.

Summary

- The IROA-LTRS is injected periodically i.e. approximately 10 sample intervals
- It is analyzed qualitatively to support more accurate compound identification, i.e. with msms or other techniques.
- The IROA-LTRS is used to build a "dictionary" of RT, *m/z* and physical characteristics including fragmentation data.
- The IROA-LTRS is used as a QC Standard to account for fluctuations in chromatography, mass, or retention time drift, source, or instrument malfunctions.
- IROA-LTRS is always the same and the catalog of all IROA peaks found in each daily **IROA-IS.**
- LTRS analysis provides a way to quantitate the performance characteristics for the instrumentation for every day's analysis and a mechanism for correcting any instrumental error.
- Since there is no need to accurately quantify the compounds in the IROA-LTRS, relative quantitation is sufficient, fewer scans are needed to establish quantitative accuracy, and these scans may be better used to support compound identity verification. Thus, the IROA-LTRS may be used with alternating scans devoted to collection of secondary compound identification information, such as fragmentation, Ion Mobility, etc., while in the analysis of the analytical samples keeping all scans as primary will ensure better peak shape and, thus, better quantitation.

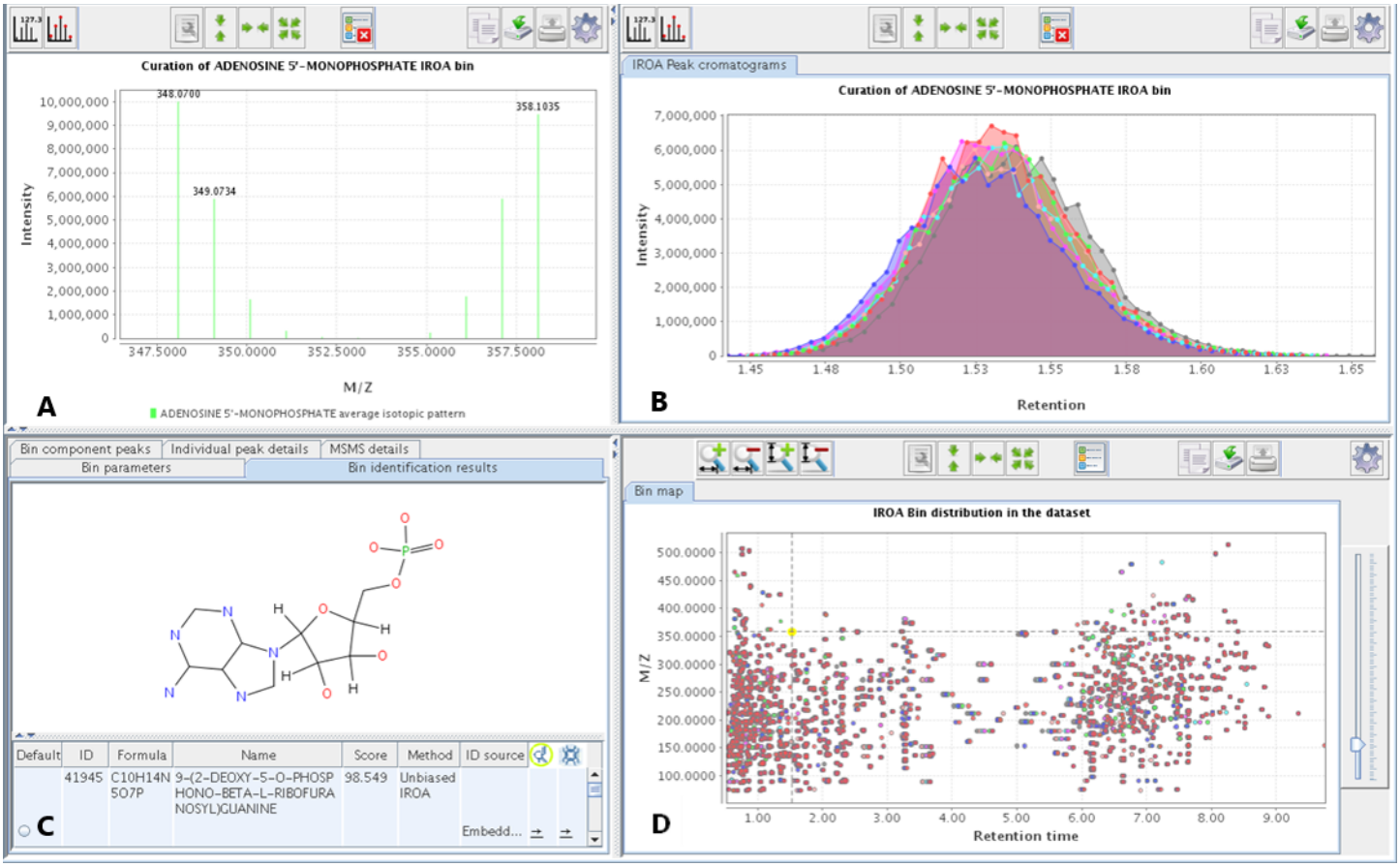


Figure 6. ClusterFinder analysis of IROA-LTRS samples. A. Average isotopic pattern of adenosine 5'monophosphate IROA-IS; B. IROA peak chromatograms of adenosine 5'monophosphate IROA-IS; C. Identification panel for adenosine 5'monophosphate; D. IROA-IS peaks identified in the dataset.

The IROA Internal Standard (IROA-IS)

The **IROA-IS** is chemically identical to the U-95% ^{13}C component of the IROA-LTRS. **The IROA-IS together with ClusterFinder software are used to co-locate, identify, and quantitate 400 to 600[3] biochemicals in experimental samples, depending upon chromatographic mode(s) employed.**

The unique IROA labeling pattern (Figure 7 A) again ensures that the monoisotopic peaks and the *carbon envelope* of the associated isotopic peaks (M-1 etc.) can be detected during LC-MS. The carbon envelope differentiates the IROA-IS from natural abundance peaks (Figure 7 B) and is used to identify compounds of interest and exclude artifacts that may look otherwise similar.

[3] The IM CCS handling is not yet fully functional in the current version of ClusterFinder. Many peaks are fragments or adducts of a parent compound.

The IROA-IS is a true Internal Standard and can be spiked into any natural abundance experimental sample (cells, tissue biopsy, plant material, blood, etc.) and all the IS peaks may be easily identified using the ClusterFinder software according to the presence of their characteristic M-1 peak and associated carbon envelope. It provides enough information for complete identification and quantitation of samples without the need for chromatographic base-line correction.

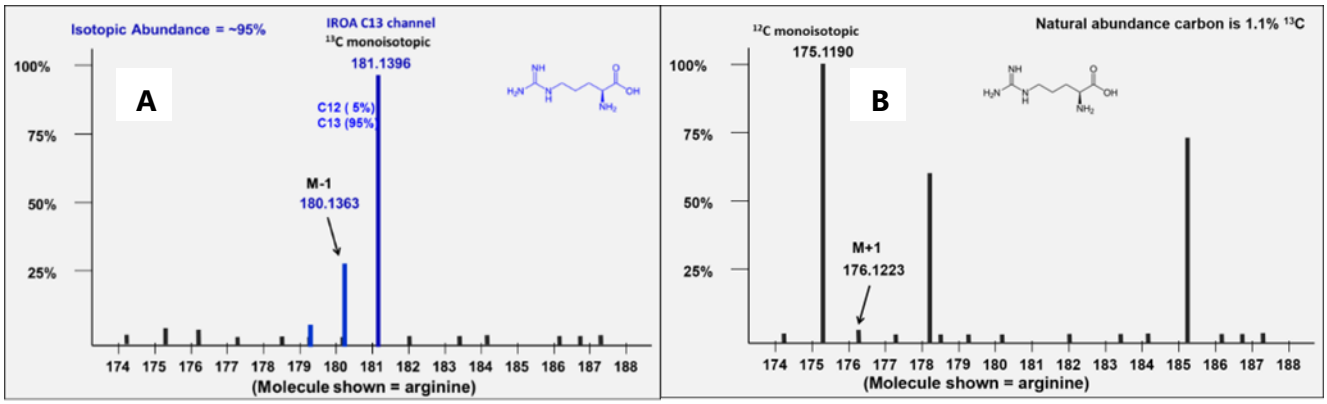


Figure 7 A. Representation of the IROA-IS (U-95% ^{13}C envelope) for arginine, blue peaks; black peaks represent noise/artefacts. B. Natural abundance arginine, monoisotopic peak shown at 175.1190 and M+1 peak at 173.1223; remaining black peaks represent noise/artifacts.

- The IS is added to the experimental samples (the combined IS/experimental sample is called the analytical sample). The IROA-IS is used to find and align all the peaks.
- The alignment of all samples ensures reduced variability in day-to-day measurements.
- Since the chemical makeup and therefore chromatographic behavior of the LTRS sample is identical to the IS that is added to the Experimental samples and analyzed within the same batch, it is possible to use the in-depth, informationally-strong, triply redundant chemical identification information obtained from the LTRS sample and apply it to the Experimental samples using their Phenotypic (1.1% C13 and 95% C13) IROA signal (see Figure 8 C').
- The characteristics for each compound in the IROA-IS allow the calculation of a suppression-corrected area for each compound. Suppression correction and normalization are implemented in ClusterFinder using the qualities inherent in the IS peaks. The user is always presented with raw, suppression corrected, and normalized data so the best choice may always be available according to any experimental design.
- The normalization of the total area under the curve (AUC) for all natural abundance suppression-corrected peaks to the total AUC of their corresponding IROA-IS peaks is a "Dual MSTUS[4]" algorithm that allows for normalization of samples not only within a single day's run, but since the IROA-IS is equivalent every day; all normalizations are to a true

[4] MSTUS described in "Normalization strategies for metabonomic analysis of urine samples" Warrack et. al (ref 8)

Standard, i.e. that is the same every day. Normalization overcomes sample-to-sample variation. (See **Appendix B** for a discussion on suppression correction and normalization.)

- The IROA-IS can be used to build a Retention Index to track "unknown" compounds not identified in the IROA-IS.

IROA TruQuant Workflow

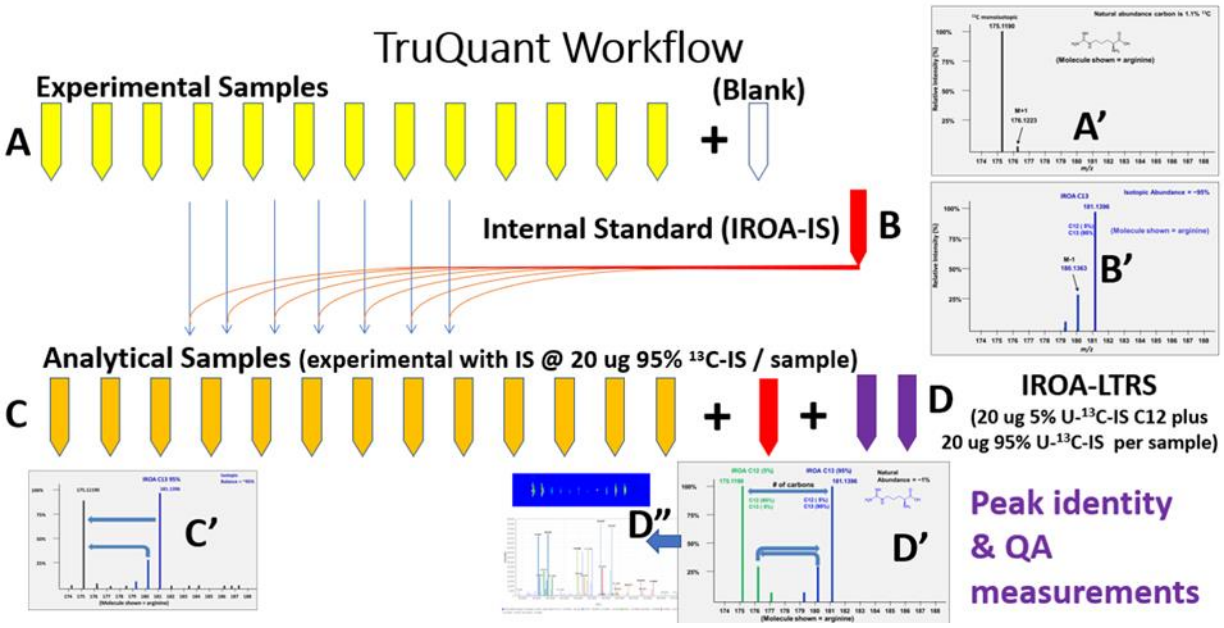


Figure 8. The IROA TruQuant measurement system is based on a well characterized Long-Term Reference Standard (D = LTRS) and a carefully matched Internal Standard (B = IS) to provide instrument and process QA/QC, validated compound identification and quantitation.

(A) Experimental samples (plasma, urine, cells, etc.) are mixed with (B) a complex 400+ compound Internal Standard (IS), fully labeled with U-95% ^{13}C (B'). Experimental/IS samples are analyzed using LC-MS, injecting the LTRS (D) intermittently, approx. every 10 samples. The LTRS has the same concentration and is chemically identical to the IS but also contains a 1:1 mixture of fully labeled U-5% and 95% U-^{13}C metabolites, producing a U-shaped pattern of carbon envelopes (D'). The relative height of the M+1, the relative height of M-1, and the mass distance between the monoisotopic peaks all provide confirmation of the number of carbons in each biological molecule resulting in a triply redundant quality control check point. The IROA peaks represent actual compounds, fragments and adducts and can be discriminated from unsigned artifacts and noise which can be removed from the data, eliminating false discoveries. As a composite sample, sample-to-sample and analytical variance is removed and during MS analysis the identical compounds (labeled with either U-95% ^{13}C or U-5% ^{13}C) experience the same ionization efficiency and suppression.

Following analysis, the resulting IROA-LTRS dictionary of compounds is used to identify compounds in the IROA-IS saving time, effort and related costs. The IROA-IS together with ClusterFinder software are used to co-locate, identify, and quantitate 400 to 600 biochemicals in experimental samples, depending upon chromatographic mode(s) employed. A library of over 420 compounds (including their fragments and adducts) have been identified in the LTRS and stored in the ClusterFinder software. We expect this number to increase as more compounds are identified.

The Internal Standard (IS), fully labeled with U-95% ^{13}C (B') is also added to a blank sample. The characteristic for each compound in the IROA-IS allow the calculation of a suppression-corrected area for each compound.

Unambiguous ID and quantitation (ms2) in a single injection. Complete identification of compounds is achieved with the addition of IM or SWATH. The IROA IM peaks retain their patterns perfectly because all IROA isotopomers share the same CCS (D'). In IROA msms fragmentation, such as SWATH, the IROA peaks retain their patterns (D2') because wide windows are used. Since all fragments retain their IROA character, their formulae and the relationships between them (D3') are determinable.

A Note on Normalization and pilot experiment

We all know that sample amounts can be highly variable due to fluctuations in sample preparation and delivery. A "pre-prep" normalization should be performed on each sample prior to crash. Our recommendation is that a protein reading be done on each sample homogenate, and equal amounts of protein be delivered as the "prep sample". While this may not always be appropriate it is always better to collect all the information one can on every sample and attempt a prep-normalization.

This protocol delivers approximately 2 micrograms each of U-95% ^{13}C and U-5% ^{13}C (IROA-LTRS) and 2 micrograms U-95% ^{13}C (IS) per sample which should be appropriate for most MS platforms, **however if this is the first time you are using your chosen matrix (blood, cells, tissues, urine etc.) with this method your first experiment should be a "calibration" experiment**, i.e. test different amounts of your standard prep with the same amount of IROA-IS to figure out how much to balance with the IROA-IS. **Please see "Pilot Experiment: Internal Standard (IS) Calibration" in Appendix A.**

The Process

The workflow, as described in Figure 8 above, has considered the addition of the internal standard to the dried sample after the sample has been fully prepared, i.e., as an "Injection Standard;" however, the internal standard may be added to sample before sample preparation, i.e., as a "Recovery Standard". Either approach can be performed, however, there are two provisos 1) a larger concentration of the internal standard is required if it is used as a "Recovery Standard", and 2) the losses during sample preparation will mean that the constancy of the internal standard is no longer true, We believe that the next step, the "Normalization" algorithm will correct for sample preparation losses, and therefore think the use of the internal standard as an Injection Standard is more efficient[6].

A) Prepare the experimental/IROA-IS and IROA-LTRS samples

PLEASE NOTE: THE MORE POLAR COMPOUNDS IN THE IS AND LTRS WILL NOT DISSOLVE IN RELATIVELY NON-POLAR SOLVENTS. THESE REAGENTS MAY NEED TO BE DISSOLVED IN DIFFERENT SOLVENTS FOR EACH CHROMATOGRAPHIC CONDITION. THE SOLVENT EMPLOYED SHOULD BE DETERMINED IN YOUR CALIBATION PILOT EXPERIMENT.

1. As determined in the Calibration experiment, the re-solvation solvent may be either the initial chromatographic solvent (minus acid) or dH2O. (Where the chromatographic solvent contains acid then dH2O is preferable to reduce degradation prior to injection.)
2. Add 900 µL re-solvation solvent to the IS amber vial, mix, and then vortex. The material should go into solution freely. This is the Internal Standard solution (IROA-IS).

6 Manuscript in preparation.

3. Hold the IROA-IS on ice until you are ready to re-solvate your samples.
4. Prep all experimental samples according to the SOP / method of choice of the laboratory. For each sample, the "calibrated aliquot amount" i.e. amount determined in the calibration step is then delivered to an Eppendorf capped sample tube.
5. Dry all aliquots under a gentle nitrogen stream.
6. Create experimental analytical samples by adding 30 µL IROA-IS to each dried experimental sample. Vortex to ensure all of the dried material is dissolved.
7. Make up a blank sample, i.e. an empty tube that has no experimental material but will receive an aliquot of Internal Standard (IS). This will not only serve as a QA/QC sample but a sample which will provide important qualitative and quantitative information. This sample, referred to as the "IS-Only" sample, should be injected three or four times randomly within the collection of analytical samples.
8. Add 30 µL of the re-solvation solvent (same solvent as is used for IS) to the vial containing the IROA Long Term Reference Standard (LTRS). Pipette up and down to ensure solubilization of IROA-LTRS material. Keep on ice until use.

B) Experimental analysis

1. Randomly analyze all samples by injecting either 4 or 5 µL of material (analytical sample, blank or LTRS) using the injection volume as determined in the Calibration experiment using the chromatographic method of choice. The analysis may be repeated in multiple modes, i.e., positive reverse phase, negative reverse phase, positive HILIC, or negative HILIC. There is no need to collect msms data for the analytical samples.
2. Inject the IROA LTRS before and after analytical samples and also inject approximately every 10 samples. Collect msms data on only one or two LTRS injections.
3. Build a ClusterFinder Project (ref 9) and establish all of the modes of analysis that were used in the experiment.
4. Attach all of the experimental and LTRS files to their appropriate tabs and associate each file according to its role in the experimental design.
5. Associate any relevant libraries with the mode in which the LTRS datasets are located.
6. Build a daily Dictionary using IROA-LTRS data files, by using ClusterFinder to identify compounds, fragments, and adducts in the LTRS injections.
7. Identify compounds, fragments, and adducts in the experimental samples using ClusterFinder software and the IROA-LTRS dictionary (see **Note 1**).
8. Identify compounds, fragments, and adducts in the experimental samples using ClusterFinder software and user-generated libraries.
    a. The TruQuant Workflow has been extended to allow the identification and quantitation of compounds in experimental samples that are not contained in the internal standard but are represented in libraries that Users may upload into the ClusterFinder software.
    b. ClusterFinder suppression correction algorithms will correct for most of the variance (i.e., instrumentation error, injector, source etc.) and then normalization algorithms correct for a large part of the remaining variance (see **Appendix B** for a longer discussion).
    c. Export the results as tab separated value (tsv) files. Raw data, suppression-corrected and suppression-corrected/normalized data will be available for export.

d. The quantitative output of targeted analysis, together with experiment design may serve as input for the statistical and biochemical interpretation. Employ your favorite statistical package.

C) Establish the LTRS Library

1. After you have completed at least 4 to 5 experiments, copy all of the LTRS samples from all of the experiments into a single newly created ClusterFinder project. As you would do in any normal ClusterFinder project organize the LTRS datasets by their mode of analysis, e.g., RPpos, RPneg, HILICpos, HILICneg, etc. (see **Note 2**).
2. Run a deep analysis on all of the LTRS injections that have been accumulated in each mode and sort the resulting bins by mass.
3. Examine (curate) each bin to determine its quality (see **Note 3**). Each bin will be named according to the most common isobar for each formula. If the identification tab is clicked all of the isobars and other possible names will be shown to the user. If the user selects the name that they know to be accurate and does this for all of the found compounds, they can build a library specific to their mode of analysis that will have appropriate RT information that will be used in all future ClusterFinder projects. If you establish an empty library and add known compounds to it one at a time this will also work but we believe it is best to just "bite the bullet," and curate a full dataset in one concerted action.
4. The identification of unusual or unknown peaks may generally be solved by using ClusterFinder's Correlation tab to identify all of the other peaks that are highly correlated with the unknown peak in question.
5. When the curation is completed simply create a new library, generally name according to the mode, and add all of the entries in a single action to it or any other library associated with the mode.
6. Export the resultant dictionary in comma separated value (CSV) format, or compound exchange format (CEF), or NIST (National Institute of Standards and Technology) format for storage or transport to other ClusterFinder installations.

APPENDIX A: Pilot Experiment: Internal Standard (IS) Calibration

The IROA TruQuant WorkFlow Kit contains enough materials for 3 sets of 30 experiments. For each type of experimental sample, i.e. plasma, urine, tissue etc., a **calibration experiment** should be performed to optimize the amount of experimental sample that you will use with the kit (the quantity of your experimental sample is optimized against the quantity of 30 µL IROA Internal Standard used in the method). You only need to do this once for any sample type or SOP.

Note: The 30 µL of IS specified here means that the amount of all compounds in the IS are exactly equal to the amount that is used for the LTRS. For accurate analytical quantitation it is optimal that the LTRS concentrations are mirrored in the IS.

The calibration of IS to your SOP simply requires the generation of a single large portion of prepared "prepped" experimental sample (the sample type you plan to use for your experiment, i.e. plasma, urine or tissue, etc.), approximately 30X the normal sample size routinely used is sufficient.

PLEASE NOTE: THE MORE POLAR COMPOUNDS IN THE IS AND LTRS WILL NOT DISSOLVE IN RELATIVELY NON-POLAR SOLVENTS. THESE REAGENTS MAY NEED TO BE DISSOLVED IN DIFFERENT SOLVENTS FOR EACH CHROMATOGRAPHIC CONDITION.

Setting up the Calibration Experiment:

1. Use the lab's extraction protocol/SOP on the sample of choice (plasma, tissue etc.), and create a large aliquot by pooling together the extracts. Mix well to create a single homogeneous sample.

2. Deliver varying aliquots of the pooled sample, ranging from a quarter of the normal amount delivered to 3 times the normal amount to Eppendorf capped sample tubes. (For example: aliquots of 0.25X, 0.5X, 0.75X, 1X, 1.5X, and 3X.) Make triplicate samples.

3. Dry aliquots under a gentle nitrogen stream.

4. Be sure to make up a blank sample, i.e. an empty tube that has no experimental material but will receive an aliquot of Internal Standard (IS). This will not only serve as a QA/QC sample but a sample which will provide important qualitative and quantitative information. This sample, referred to as the "IS-Only" sample, should be injected three or four times randomly within the collection of experimental samples.

5. Resolvate a single vial of IROA LTRS with 30 µL of a solvent that will work for injections (based on the chromatographic condition) but is reasonably polar to maximize the number of compounds that will go into solution. (The compounds in the LTRS and IS are mostly polar.) Thoroughly mix to ensure complete solvation. Note: It should go into solution freely. Keep on ice until ready to use. Inject 4 and 5 µL of the LTRS to

determine the best injection size. An injection size that gives you a total of several hundred peaks is the goal.

6. Add 900 µL of the same solvent to a single vial of IS. Thoroughly mix to ensure a single homogeneous solution. Note: Again, it should go into solution freely. Keep on ice until ready to use.

7. Create calibration analytical samples by adding 30 µL IROA-IS to each dried calibration sample. Vortex to ensure material is dissolved.

8. Analyze samples using the chromatographic method you would normally use using injection volumes determined in step 5. The analysis may be repeated in multiple modes, i.e. positive reverse phase, negative reverse phase, positive HILIC, or negative HILIC. The IROA LTRS should be injected before and after analytical samples and should also be injected approximately every 10th injection. (If your injection size requires more than the original 30 µL volume then pool two LTRS prior to the initial injection.)

9. The higher concentrated samples may overload the column, so we do not recommend running them in random order but rather run then from most dilute to most concentrated to reduce any carry-over effects. The randomly run LTRS and IS-Only samples will provide evidence of any carry-over.

10. There is no need for msms in the Calibration (experimental) samples, only do msms on the LTRS. Since msms tends to retard the peaks, run msms on only one or two LTRS injections.

11. Use the ClusterFinder software to find and identify all IROA peaks in the LTRS in an unbiased analysis (ref 9). These IROA peaks will be named.

12. Run a Semi-targeted analysis of the IS containing calibration samples based on the peak parameters determined in the LTRS sample.

13. Use the ClusterFinder software to generate a calibration report. This report will list all of the compounds found in the calibration samples and graphically show the point (and concentration) where the calibration samples C12 MSTUS signals for each concentration are equal to the Internal Standard C13 MSTUS. This aliquot size will be used with all future samples based on the same SOP.

14. Export all of the compound identification, calibration, and QA/QC data from the LTRS.

15. This is the amount of sample that will most accurately be measured using the IS in the future, i.e. well balanced by the standard 30 µL of IS. This is illustrated in Figure 9 below.

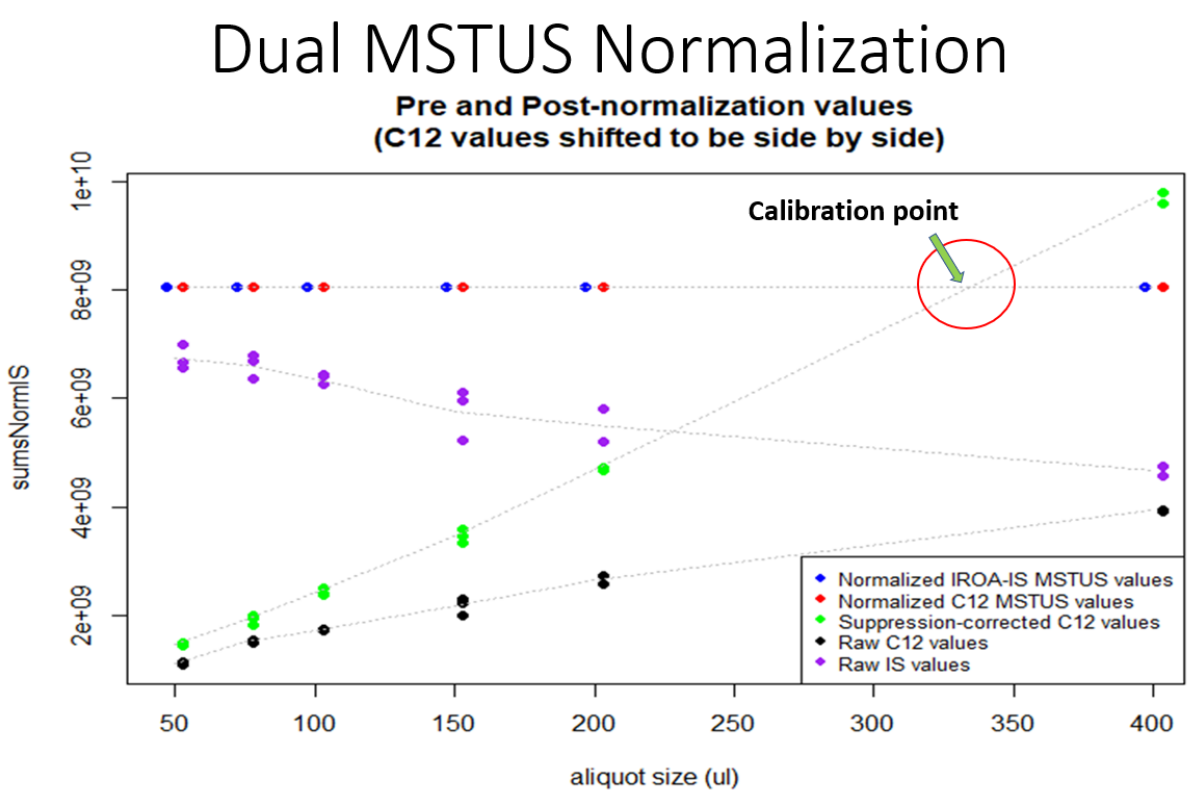


Figure 9. Suppression-correction and normalization for IS related peaks. The intersection ("calibration point") of the suppression corrected MSTUS values (dark blue values), and the MSTUS C13 values for normalized data (green values) represents the aliquot size that will have optimal quantitation.

What you will learn from a calibration experiment:

ClusterFinder (CF) software algorithms find and interpret the entire isotopic envelope for all isotopic balances, natural abundance, U-13C 95%, and U-13C 5%. Because of the nature of the labeling in these situations we now understand that all the peaks of the full isotopic envelope need to be summed to determine the actual quantity of material on either side.

However, for accurate quantitation that is not all that needs to be considered. If we have a sample and inject various aliquots, we expect that the data should reflect the aliquot size injected, but does it? For instance, by injecting varying aliquot sized samples, we would expect that the signal we see in the ms for each compound should increase by the same order, but in fact it does not! Everyone should do this very simple experiment to understand how data on their instrumentation will look and how far the datapoints deviate from the expected curve. In almost all cases, it will be an eye-opener.

What causes this discrepancy is usually the result of a number of different things which end up suppressing the signal that is seen by the detector. While a large portion of this variance is likely ion suppression, significant portions of the variances seen sample-to-sample come from other phenomena. In the following sections we will discuss how the suppression correction algorithms correct for most of the variance (i.e., instrumentation error, injector, source etc.) and then normalization algorithms correct for a large part of the remaining variance.

Consider a model Calibration experiment shown in Figure 10. The different sized aliquots discussed above (1X, 2x, 4X, etc.) would be expected to provide a graph similar to what is seen in Figure 10 A but in reality, results in a graph similar to what is seen in Figure 10 C (blue)[5]. If we put an IROA internal standard into each sample at the same concentration we expect to see a straight horizontal line (Figure 10 B), but what we see is seen in Figure 10 D (blue). Thus, in this experiment a large portion of the error is concentration dependent and therefore largely driven by ion suppression. Casual observation of Figures 10 B and 10 C suggests that the suppression seen on the natural abundance and internal standards are similar, and Figure 10 E proves that they are perfectly correlated. Thus, if we know what the unsuppressed value was for every compound we can correct the data seen in Figure 10 C (blue) to its suppression corrected values seen in Figure 10 C & D (red). It is important to note that that this correction is applied to each compound in each sample independently, i.e., suppression-correction is compound specific. Four additional features arise from this correction.

First, it is possible to examine the suppression that each compound undergoes in each injection. The result of this analysis is shown in Figure 10 F which illustrates the distribution of average suppression among all the MSTUS compounds in the samples in the most concentrated sample (here, approximately 2 times "normal" injection volume). The average compound is subjected to a 40% suppression within a range (<5% suppression to >85% suppression). Above 80% suppression the AUCs are negatively correlated to their aliquot size.

Secondly, the suppression-correction process can be applied to recover from any source of loss of ions where the loss is equally applied to all peaks within the isotopolog ladder after the time that the two envelopes are in place. This includes injector error, but does not include fragmentation error, except for the parent compound, since the breakage of C13-C13 bonds requires more energy than breaking a C12-C12 bond.

Thirdly, for the suppression-correction process to work effectively the majority of internal standard peaks need to be findable and to be sufficiently above noise so that their areas can be effectively determined. While this may seem obvious, some researchers have used the internal standard in too dilute a concentration (i.e. without calibration) and therefore their results were less than optimal. That is why we recommend that everyone who uses IROA reagents perform a calibration of the internal standard to their generic SOP samples. It only needs to be done once for each sample preparation method. The calibration determines the aliquot size of the sample that most closely matches the concentrations of internal standard that the user intends to use. This is centered around the intersection of the suppression corrected sample line with the suppression corrected internal standard line. In Figure 9 this Calibration point intersection is indicated with a red circle. Also seen in Figure 9, the suppression-correction is sensitive enough that this allows one to operate in both highly suppressed and highly dilute conditions.

Fourthly, a MSTUS normalization may be applied to this suppression-corrected data that results in more stringent normalization than current methods. This is presumably due to the removal of suppression error in the MSTUS raw data (See **Note 3**). Here, the natural abundance MSTUS values are normalized to the MSTUS suppression-corrected internal standards data. Because this normalization uses two MSTUS values we refer to this as a "Dual

[5] Because the suppressibility is a function of chemical structure each compound will be different. Graphed here are the MSTUS values of each injection, which provides a commonly-based response for all samples.

MSTUS ™" normalization method (see Figure 9). If the same internal standard is used for two different collections of samples (different experiments) and they are all normalized to the same internal standard (ideally with largely the same MSTUS collection of IROA internal standards) then the samples should be more mathematically comparable[6]. In general, sample-to-sample normalizations are performed to correct for sample size (or other) differences, and for losses during sample preparation. As can be seen in Figure 10 the Dual MSTUS normalization has made vastly different sample sizes all appear to be identical (red and blue dots) and corrected raw data (black and purple) for losses (green and red).

An important distinction needs to be made between the suppression-correction processes, and the Dual MSTUS normalization processes, namely that the suppression-correction is performed compound by compound while the Dual MSTUS is performed sample by sample. This has important implications: 1) it is not possible to do a suppression-correction without an internal standard that can be used to calculate the level of suppression for a compound, but 2) the sample-to-sample Dual MSTUS normalization may be applied to any peak in the sample even if it does not have an accompanying internal standard.

As noted above, together suppression-correction and normalization correct for two of the major sources of error in most experiments, the former corrects for error introduced within the instrumentation (in the injector, source, lensing and detectors), and the latter corrects for sample preparation variances (size difference, sample losses during preparation etc.). As such these two algorithms can remove a large portion of the error in most MS based experiments (see **Note 4**).

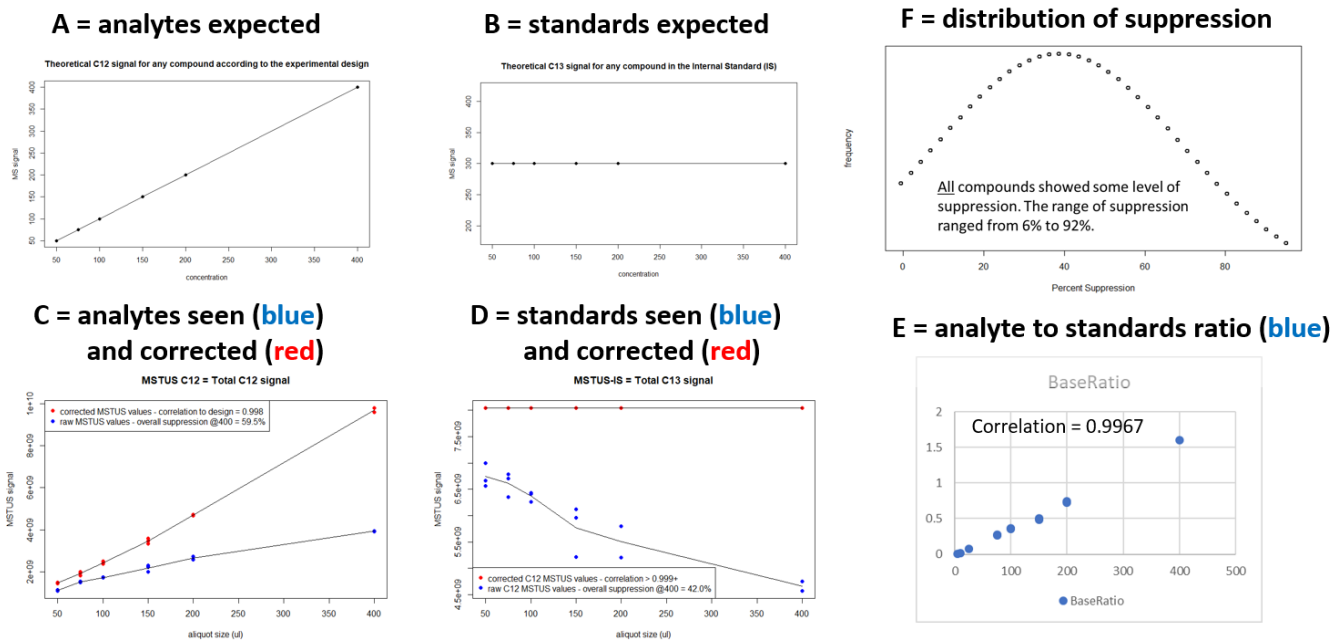


Figure 10. An experimental illustration of suppression correction. Suppression is more common that most researchers assume, but the loss is suffered equally by all isotopologs so the C12/C13 ratio is unaffected and therefore affords a route to correction.

[6] Manuscript in preparation.

APPENDIX B: Algorithms for Suppression Correction and Normalization

Fundamental to the IROA concepts (and inherent in the name Isotopic Ratio Outlier Analysis) is the fact that the ratio of the C-12 envelope to the C-13 envelope is **unaffected by suppression** even though both the C-12 and C-13 isotopomeric sets may be strongly suppressed. This has afforded a mechanism for **suppression correction** that has been built into ClusterFinder. Once suppression is corrected, a Dual MSTUS (ref 8) algorithm is employed to provide a very accurate mechanism for the **normalization of samples** against sample-to-sample variances. This version of ClusterFinder outputs three values: 1) the raw (suppressed) values observed; 2) a suppression-corrected value; and 3) a normalized (suppression-corrected and normalized) value.

1. **The algorithm for suppression correction** starts by determining the true ratio of the total number of molecules in their respective C12 and C13 envelopes and then multiplying this ratio with the "unsuppressed value" of the molecule at hand. Assuming the ratio properly reflects the relative populations of the two envelopes in question, it is worth considering briefly what this "unsuppressed value" means and how it can be determined[7]. Based on the fact that every sample has an equal aliquot of the internal standard, and therefore the same concentration of each internal standard compound:

    a. The "least unsuppressed value" can be the largest C13 value seen within an experiment for each compound. This value can then be used within an experiment and will recover the effects of suppression as seen in the experiment.

    b. The "least unsuppressed value" may be the largest C13 value seen in a sample that is otherwise a blank (i.e., containing no sample).

    c. A more accurate, and enduring value for the "unsuppressed value" can be determined in a specific experiment in which the internal solution is serially diluted and analyzed in the absence of any other sample.

    d. If there is a quantitatively accurate determination of the concentration of the internal standard for the peak in question, then the ratio may simply be multiplied by this concentration to obtain an approximately accurate concentration value.

It should be noted that options a and b require no further data to be generated external to the experiment, while options c and d rely on experimentally established quantitative values, that once established may be used at any

[7] Manuscript in preparation.

time if an identical aliquot size of the same internal standard is used to resolvate the dried sample. Options c and d represent solutions that can be repeatedly used for longer periods of time, over many experiments, or even for very large studies using hundreds or thousands of samples[8].

2. **The MS Total Usable Signal (MSTUS) normalization algorithm** (Warrick ref 8) simply assumes that the overall chemical composition of all samples is close enough that the sum of all the verifiable compounds for comparable samples will be "reasonably" constant. Thus, according to this theory if the sum of these verifiable compounds differs then it is more likely that this is due to differences in the physical size (density or similar property) of the original samples. The algorithm was devised to normalize urine samples which often demonstrate very different concentrations or dilutions but has been shown to be equally effective for most classes of solid or liquid samples. In MSTUS, for every sample a "Normalization Factor" (NF) is developed that when the AUCs for all peaks as individually multiplied they will sum to a "common value." The "common value" is determined arbitrarily by the experimentalist in every experiment. In the IROA Workflow, we have considered and implemented several key features that have strengthened the MSTUS normalization procedure with the modification we call "Dual MSTUS":

    a. All verified peaks are, by definition, peaks that are found in both the IS and the sample, i.e., show up in both the C12 envelope and the C13 envelope of a compound's isotopolog ladder in both the LTRS and experimental samples. This is a very rigorous test of biological relevance.

    b. All the C13 peaks for each compound represent the same amount of material in each sample because they are derived by addition of equal aliquots of the internal standard.

    c. The amplitudes of both the C12 and C13 envelopes suppressed equally (see Figure 10E) are corrected for suppression losses by correcting each according to the same correction factor.

    d. Each sample is represented by two equally important MSTUS values, the C12 MSTUS (the sum of all C12 envelopes), and the C13 MSTUS value (the sum of all C13 envelopes).

    e. The NF is developed to make the C12 MSTUS equal to the C13 MSTUS (which we know to be constant).

Since the C13 MSTUS is always present at the same concentrations this normalization not only avoids any arbitrary quality in the NF but also means that every suppression corrected and normalized sample can be directly compared to any other similarly treated sample, i.e., has the same amount of IS in it.

There is one additional aspect of this algorithm to consider, namely, in practice some samples have more compounds than others. Since the rules of Dual MSTUS require that every C12 envelope must have a C13 envelope equivalent, the NF is a simple relative measurement; however, if you want to directly compare two or more samples the comparison should only use those compounds that are common to all the samples that you wish to compare. The reality is that this is not that much of a restriction as the IROA software algorithms

[8] Manuscript in preparation.

satisfactorily creates non-sparse datasets, but nevertheless it is a factor worth considering when comparing samples.

Notes

1. The use of IROA IS allows the resolution and accurate quantitation of compounds that are coeluting. The presence of higher quantities of compounds brought about by higher volumes of the sample seem to be well tolerated and results in a larger number of compounds being found in both the LTRS and IS containing samples. Note that if during the semi-targeted analysis for every compound the IS signal for that compound is seen but none of its natural abundance isotopologs (NA) are detected the software will report a zero to provide a complete, i.e. non-sparse, dataset, and indicate that this is truly a "below limit of detection" situation. If neither the IS nor its NA are detected the field is left blank even though it is apparent that both have been suppressed to "below limit of detection" (because it is certain that the IS side should have been present).
2. The IROA TruQuant Workflow (Figure 8) is a semi-targeted analysis that should find 1000+ peaks in positive mode, and 500+ peaks in negative mode. A dictionary of these specific peaks, their major fragments, and adducts is imbedded into the ClusterFinder database and may be used to identify all peaks reproducibly; however, it has no RTs because they are necessarily unknown. This means it will not always name isobaric compounds correctly. After several LTRS injections are made (best over 4-5 experiments) a special laboratory library that has RTs may be created that will correctly assign all isobaric compounds. All of the information associated with each of the peaks may be edited by the user. When the user's retention time for a specific compound is associated with the library entry then it will always identify the peak correctly. If additional peaks are identified they may also be added to ClusterFinder's databases. Until these lab specific databases are established the daily LTRS Discovery results can be used but will need curation for their isobaric peaks.
3. The original definition of MSTUS (MS Total Usable Signal, see ref 8) has been extended here to a novel MSTUS variant "Dual MSTUS". In MSTUS only the signals from known compounds are used. In IROA we have two separate extracts in each physical sample, so we calculate the MSTUS values for each (C12 and C13) and adjust them, so they are normalized to the value of the C13 Internal Standard. As a "standard" this enables more accurate comparisons across variable conditions better than the original definition including the correction for all instrumentation derived error.

References

1. Charkoftaki G, Aalizadeh R, Santos-Neto A, Tan WY, Davidson EA, Nikolopoulou V, Wang Y, Thompson B, Furnary T, Chen Y, Wunder EA, Coppi A, Schulz W, Iwasaki A, Pierce RW, Cruz CSD, Desir GV, Kaminski N, Farhadian S, Veselkov K, Datta R, Campbell M, Thomaidis NS, Ko AI; Yale IMPACT Study Team; Thompson DC, Vasiliou V. "An AI-powered patient triage platform for future viral outbreaks using COVID-19 as a disease model." Hum Genomics. 2023 Aug 29;17(1):80. PMID: 37641126; PMCID: PMC10463861. doi:10.1186/s40246-023-00521-4.

2. Petrova B, Lacey TE, Culhane AJ, Cui J, Raskin A, Mishra A, Lehtinen MK, Kanarek N. "Metabolomics of Mouse Embryonic CSF Following Maternal Immune Activation." bioRxiv preprint https://doi.org/10.1101/2023.12.06.570507

3. Lippa KA, Aristizabal-Henao JJ, Beger RD, Bowden JA, Broeckling C, Beecher C, Clay Davis W, Dunn WB, Flores R, Goodacre R, Gouveia GJ, Harms AC, Hartung T, Jones CM, Lewis MR, Ntai I, Percy AJ, Raftery D, Schock TB, Sun J, Theodoridis G, Tayyari F, Torta F, Ulmer CZ, Wilson I, Ubhi BK. "Reference materials for MS-based untargeted metabolomics and lipidomics: a review by the metabolomics quality assurance and quality control consortium (mQACC)." Metabolomics. Apr 9;18(4):24. (2022) PMID: 35397018; PMCID: PMC8994740, doi: 10.1007/s11306-021-01848-6..

4. Habet V, Li N, Qi J, et al. "Integrated Analysis of Tracheobronchial Fluid from Before and After Cardiopulmonary Bypass Reveals Activation of the Integrated Stress Response and Altered Pulmonary Microvascular Permeability." The Yale Journal of Biology and Medicine. Mar;96(1):23-42. (2023) PMID: 37009190; PMCID: PMC10052603. doi:10.59249/kfyz8002.

5. Piqueras, M.D.C., Myer, C., Junk, A., Bhattacharya, S.K. "Isotopic Ratio Outlier Analysis (IROA) of Aqueous Humor for Metabolites." In: Bhattacharya, S. (eds) Metabolomics. Methods in Molecular Biology, vol 1996. (2019) Humana, New York, NY. https://doi.org/10.1007/978-1-4939-9488-5_15

6. Qiu Y, Moir RD, Willis IM, Seethapathy S, Biniakewitz RC, Kurland IJ. "Enhanced Isotopic Ratio Outlier Analysis (IROA) Peak Detection and Identification with Ultra-High-Resolution GC-Orbitrap/MS: Potential Application for Investigation of Model Organism Metabolomes" Metabolites. (2018) Jan 18;8(1):9. PMID: 29346327; PMCID: PMC5875999, doi: 10.3390/metabo8010009..

7. Clendinen CS, Stupp GS, Ajredini R, Lee-McMullen B, Beecher C, Edison AS., "An overview of methods using (13)C for improved compound identification in metabolomics and natural products." Front Plant Sci. 2015 Aug 25;6:611. doi: 10.3389/fpls.2015.00611. PMID: 26379677; PMCID: PMC4548202.

8. Warrick BM, Hnatyshyn S, Ott K-H, Reily MD, Sanders M, Zhang H, and Drexler DM, "Normalization strategies for metabolomic analysis of urine samples" J Chrom B 877 (2009) 547-662, doi:10.1016/j.jchromb.2009.01.007.

9. Beecher C, Raskind A, de Jong F, ClusterFinder Manual (2024), IROA Technologies.