Small Molecule Isotope Resolved Formula Enumerator: a Tool for Assigning Isotopologues and Metabolites in Fourier Transform Mass Spectra

Hunter Moseley, Joshua Mitchell, Robert M Flight

Submitted date: 19/01/2019 • Posted date: 21/01/2019
Licence: CC BY-NC 4.0
Citation information: Moseley, Hunter; Mitchell, Joshua; Flight, Robert M (2019): Small Molecule Isotope Resolved Formula Enumerator: a Tool for Assigning Isotopologues and Metabolites in Fourier Transform Mass Spectra. ChemRxiv. Preprint.

Improvements in Fourier Transform Mass Spectrometry (FT-MS) enable increasingly more complex experiments in the field of metabolomics. What is directly detected in FT-MS spectra are spectral features (peaks) that correspond to sets of adducted and charged forms of specific molecules in the sample. The robust assignment of these features is an essential step for MS-based metabolomics experiments, but the sheer complexity of what is detected and a variety of analytically-introduced variance, errors, and artifacts has hindered the systematic analysis of complex patterns of observed peaks with respect to isotope content. We have developed a method called SMIRFE that detect small biomolecules and determines their elemental molecular formula (EMF) using detected sets of isotopologue peaks sharing the same EMF. SMIRFE does not use a database of known metabolite formulas, instead a nearly comprehensive search space of all isotopologues within a mass range is constructed and used for assignment. This search space can be tailored for different isotope labeling patterns expected in different stable isotope tracing experiments. Using consumer-level computing equipment, a large search space of 2000 daltons was constructed and assignment performance was evaluated and validated using verified assignments on a pair of peak lists derived from spectra containing unlabeled and 15N-labeled versions of amino acids derivatized using ethylchloroformate. SMIRFE identified 18 of 18 predicted derivatized EMFs and each assignment was evaluated statistically and assigned an e-value representing the probability to occur by chance.

File list (1)

ACS_SMIRFE_algo_manuscript_v2.4_combined.pdf (2.58 MiB)
Small Molecule Isotope Resolved Formula Enumerator: a Tool for Assigning Isotopologues and Metabolites in Fourier Transform Mass Spectra

Joshua M. Mitchell¹²,³ and Robert M. Flight¹²,³ Hunter N.B. Moseley¹²,³,⁴,*

¹Department of Molecular & Cellular Biochemistry
²Markey Cancer Center
³Resource Center for Stable Isotope Resolved Metabolomics
⁴Institute for Biomedical Informatics

University of Kentucky, Lexington KY, United States

Metabolomics, Fourier transform mass spectrometry, assignment

ABSTRACT: Improvements in Fourier Transform Mass Spectrometry (FT-MS) enable increasingly more complex experiments in the field of metabolomics. What is directly detected in FT-MS spectra are spectral features (peaks) that correspond to sets of adducted and charged forms of specific molecules in the sample. The robust assignment of these features is an essential step for MS-based metabolomics experiments, but the sheer complexity of what is detected and a variety of analytically-introduced variance, errors, and artifacts has hindered the systematic analysis of complex patterns of observed peaks with respect to isotope content. We have developed a method called SMIRFE that detect small biomolecules and determines their elemental molecular formula (EMF) using detected sets of isotopologue peaks sharing the same EMF. SMIRFE does not use a database of known metabolite formulas, instead a nearly comprehensive search space of all isotopologues within a mass range is constructed and used for assignment. This search space can be tailored for different isotope labeling patterns expected in different stable isotope tracing experiments. Using consumer-level computing equipment, a large search space of 2000 daltons was constructed and assignment performance was evaluated and validated using verified assignments on a pair of peak lists derived from spectra containing unlabeled and ¹⁵N-labeled versions of amino acids derivatized using ethylchloroformate. SMIRFE identified 18 of 18 predicted derivatized EMFs and each assignment was evaluated statistically and assigned an e-value representing the probability to occur by chance.

INTRODUCTION

Advances in Fourier-transform mass spectrometry (FT-MS) provide substantial simultaneous improvements in mass accuracy, mass resolution, and sensitivity. (Eliuk and Makarov, 2015). Theoretically, these combined capabilities provide several analytical and interpretative improvements including: i) the ability to resolve distinct isotopologues with identical unit masses but different accurate masses, which in turn enable natural abundance correction for multi-isotope labeling (Moseley, 2010) (Carree et al., 2013); ii) improved assignment accuracy (Kind and Fiehn, 2006); and iii) the detection of compounds in the sub-femtomolar range (Eyles and Kaltashov, 2004) (Dettmer et al., 2007). For the metabolomics field, these improvements permit the use of multi-isotope-labeled precursors in stable isotope-resolved metabolomics experiments (Yang et al., 2017), providing richer information to: elucidate unknown metabolic pathways (Creek et al., 2012) (Higashi et al., 2014), quantify relative fluxes through connected metabolic pathways (Hiller et al., 2010), identify multiple pools of metabolites in different compartments (Fan et al., 2012), and identify the active metabolic pathways under various cellular conditions (Sellers et al., 2015). In general, FT-MS generates more informative spectra that better represent the chemical constituents of complex biological samples, enabling a more complete modeling of cellular metabolism and a better understanding of physiological and pathological processes at the mechanistic level, which ultimately facilitates the identification of potential therapeutic targets (Fan et al., 2009) and the quantification of differential drug responses (Harris et al., 2012).

Despite these advantages, leveraging the capabilities of FT-MS instruments to achieve these interpretive improvements remains difficult. First, when deployed in high-throughput environments, the volume and complexity of data produced by FT-MS requires automated tools for data reduction, quality control, feature assignment, and downstream analyses. These difficulties are amplified when FT-MS is used in conjunction with direct infusion (sometimes referred to as direct injection with electrospray), due to the lack of orthogonal information about the metabolites such as retention times from chromatography. Despite this disadvantage, direct infusion allows for simpler analytical setups, the ability to use smaller amounts of sample for analysis, and a substantial improvement in the rate of data collection (Chekmeneva et al., 2017). In most cases, the rate of acquisition of FT-MS data often outpaces the rate at which it can be analyzed and meaningfully interpreted. Therefore, new analysis tools are needed to accelerate meaningful interpretation of FT-MS data.

Currently the assignment of FT-MS observed spectral features (peaks) to elemental chemical formulas (EMFs), representing one or more compounds (presumably metabolites) in a biological sample, remains a difficult step in the FT-MS pipeline especially for untargeted metabolomics experiments. In existing untargeted FT-MS metabolomics pipelines, assignment tools such as LipidSearch (Peake et al., 2013) and PREMISE (Lane et al., 2009) have been used to assign observed spectral features based primarily on m/z matching. These methods utilize databases that are typically constructed using formulas of known or predicted metabolites, have low computational overhead, and can be tailored to the biological system being studied.
Additionally, existing public metabolite databases, such as LipidMaps (Fahy et al., 2007), KEGG (Kanehisa et al., 2016), HMDB (Wishart et al., 2018), and non-metabolite focused databases such as ChemSpider (Pence and Williams, 2010) and PubChem (Kim et al., 2016), can be queried with MS-determined molecular masses to provide possible assignments. However, existing metabolite databases are incomplete (Schrimpe-Rutledge et al., 2016) (Mitchell et al., 2014) and tailored m/z databases limit discovery—a stated goal of many untargeted analyses and non-metabolite databases can provide assignments that are not possible in a biological system. This in turn introduces assignment bias (Moseley, 2013) and difficulty with disambiguating multiple assignments, and are statistically error-prone when limited to MS1-only m/z matching without chromatography due to the lack of cross-validating evidence (Kind and Fiehn, 2006). These limitations become more pronounced when artifactual spectral features are present, which we have observed in most spectra generated from FT-MS instruments (Mitchell et al., 2018). An alternative approach that circumvents the limitations of database-based assignment methods is the direct calculation of elemental molecular formulas from the observed masses of peaks (Kind and Fiehn, 2007). These rules however are not foolproof, are not easily applied to isotopologue peaks, and are typically a last resort when a database lookup fails to return a probable assignment. At high molecular weight, the number of possible elemental formulas remains very large even with these heuristic rules (Watson, 2013).

These limitations ultimately hamper the utility of m/z-only based assignment methods for untargeted metabolomics experiments; however, the lack of any real alternative means that these methods remain popular. The combinatorial mathematics underlying natural abundance probabilities used for natural abundance correction of stable isotope tracing experiments provides a possible avenue to develop untargeted assignment methods that can utilize the detection of distinct isotopologues as a source of cross-validating information within an MS1 spectrum. We have developed a new assignment method, Small Molecule Isotope Resolved Molecular Formula Enumerator (SMIRFE), that integrates this methodology while leveraging the improved mass accuracy and mass resolution of FT-MS instruments to provide cross-validated assignments for FT-MS spectra without additional orthogonal information nor databases of known metabolites.

**Figure 1 – Overview of SMIRFE Algorithm**

**Materials and Methods**

**SMIRFE Algorithm Overview**

The number of elemental molecular formulas (EMFs) and corresponding isotope-resolved molecular formulas (IMFs) for an EMF search space is prohibitively large to recalculate for every spectrum to be assigned. For CHONPS stable isotopes alone, very large IMF search spaces greater than 10^39 are typical for molecules of 1400 daltons or less. Instead, SMIRFE first generates a representation of partial EMFs in the search space given a set of possible labeled isotopes and a description of the search space. The resulting partial EMF search space, which can still be quite large, is specific for that search space and labeled isotopes but can be reused between similar experiments. With this partial EMF search space, untargeted EMF assignment followed by targeted IMF assignment on an input peaklist can be performed efficiently. The organization of these steps is represented in Fig 1 and will be discussed in more detail.

**SMIRFE Nomenclature**

The output from SMIRFE is a mapping of peaks in a spectrum to distinct IMFs and a measurement of the statistical reliability of that assignment. Throughout SMIRFE, it is necessary to keep track of which groups of IMFs belong to which EMFs. On the surface, this seems straightforward as each IMF belongs to exactly one EMF, but this becomes more complicated when multiply labeled forms of multiply adducted forms of EMFs are present. As illustrated in Figure 2, all IMFs belonging to an EMF are kept in a superclique object, which is then further subdivided into cliques which contain specific adducted versions of an EMF, which are further subdivided into subcliques containing all identically adducted and identified IMFs. The term ‘identically labeled’ within SMIRFE describes IMFs that have the same number of enriched isotopes from the labeling source. For example [Na]-[13C2]-glucose and [Na]-[13C3]-glucose can belong to the same subclique, if two of the three [13C] atoms in [Na]-[13C]-glucose are from labeling and the third atom is due to natural abundance. Although every compound’s isotopologues will correspond to exactly one superclique and one or more cliques and subcliques, the reverse is not always true. For example, the C6H12O6 superclique will contain all IMFs of glucose and IMFs for isomers of glucose.

**Figure 1 – Flowchart of the SMIRFE Algorithm. SMIRFE can be divided into two main components: the generation of an EMF search space and an isotope component lookup and a peaklist assignment algorithm. The component lookup table stores the NAPs and masses of all isotope combinations for each element to be used during assignment and IMF enumeration. The high abundance IMF database represents all likely IMFs in the EMF search space. Peaklist assignment uses this database of IMFs to generate possible EMFs present in the spectrum which are then used to inform a targeted IMF search for each EMF. The untargeted component of the search is highlighted in green and the targeted component is highlighted in purple.**
SMIRFE makes use of the natural abundance probability (NAP) and relative natural abundance probability (relative or relNAP) of IMFs both for IMF generation and for assignment scoring. The natural abundance probability of an IMF is the probability of observing the set of isotopes present in that IMF by random chance alone assuming that all isotopes have a probability of occurring equal to their natural abundance, i.e. their expected relative frequency in the biosphere. Labeling introduces a non-natural source of specific isotopes and thus relative NAP must be used to quantify the probability of observing a labeled IMF when one or more isotopes can be labeled. These values are calculated as follows:

\[ \text{NAP}_{\text{rel}}(k_1, k_2, \ldots, k_m) = \prod_{i=1}^{m} \text{NAP}_{\text{rel}}^{k_i} \]

(Moseley, 2010) (Carreer et al., 2013)

Where for a given number of nuclei \( E_{\text{max}} \) for an element \( E \) with a set of stable isotopes \( k_i \) with known natural abundances \( (N_A) \), the elemental NAP for a given combination of isotopes is calculated with Equation 1. The NAP of an IMF (Equation 2) is the product of each element’s NAP in the IMF. Without labeling, this NAP is referred to as the absolute NAP. When labeling is present, labeled isotopes must be removed from the calculation of NAP for the element(s) of the labeling isotope(s). This effectively changes \( E_{\text{max}} \) and the isotope counts for a labeled IMF. NAP values for isotope combinations can be calculated and stored in a lookup table for later use. Therefore within SMIRFE, any isotope combination has its NAP calculated only one time. Also, while absolute NAP and relative NAP are interchangeable without labeling, comparisons between relative NAP is always used when comparing IMFs.

**Isotope Component Lookup Table Creation**

Throughout SMIRFE, highly accurate absolute NAPs and masses for the set of isotopes for any given element are needed to calculate various NAPs and relative masses \( (m/z) \) of IMFs. Although the total number of isotope combinations for each element is large, it is much smaller than the number of possible IMFs and can be calculated once and stored for future use. The first step in both the untargeted search space generation and in the actual assignment algorithm is to generate this lookup table. The lookup table is sub-divided by element and for each element every possible combination of isotopes for that element is stored along with the mass of that component, its absolute NAP, and the number of atoms it contains. The masses in this table are calculated using the mpmath high-precision math library (Fredrick Johansson, 2018) in Python and the absolute NAPs are calculated using the multinomial calculation shown in Equation 1. The number of components stored in this table grows exponentially as the search size increases, but the table is only megabytes in size and can be stored in JavaScript Object Notation (JSON) format.

**Untargeted EMF Search Space Creation**

SMIRFE requires an internal representation of an EMF search space that can be queried to produce probable high NAP IMF assignments for a given peak. This search space consists of several components. First is a high level description of the search space provided by the end user that specifies the max count for each element in the space, the maximum mass, the min NAP and the set of labeled elements in the search space (e.g. C:100, H:230, O:40, max_mass:1600, min_NAP:0.4, labeled:{‘C’}). From this description, SMIRFE constructs an N-
dimensional integer lattice where each unique N-tuple of integers represents exactly one EMF in the space. For example, \((6,12,6)\) would represent the EMF \(\text{C}_6\text{H}_{12}\text{O}_6\) in the above search space. The number of tuples in a possible search space grows exponentially with increasing dimensionality, but the enumeration of integer lattices can be done quickly using existing functions from the NumPy Python package (Walt et al., 2011). The set of all unique tuples is equal to the set of all possible EMFs in the search space; however, many of these EMFs are not consistent with known rules for valid chemical structures. For example, \((1,100,0)\) corresponds to the EMF \(\text{C}_1\text{H}_{100}\) which cannot exist due to invalid valence electrons. As EMFs are generated during lattice enumeration, EMFs that fail to comply with known hard rules about valence electrons are removed. Heuristics based on patterns observed in the HMDB and KEGG further restrict possible EMFs.

The second component of the search space representation is the set of high abundance IMFs for each possibly valid EMF. The IMF generation procedure can be modeled as a depth first search on the complete tree of possible isotope components possible with that EMF, with each level in the tree representing the components of a given element. Components of labeled elements are treated as having a NAP of 1, while components of unlabeled elements below the min_NAP are pruned prior to traversal. A depth first search is then performed to find all paths through the tree for which the product of the NAP of all nodes in the path exceeds the min_NAP and the sum of the mass of the nodes in the path is below the max_mass. Every such path represents exactly one IMF for the given EMF. The IMF and its mass is then stored in an SQLite (Hipp and 2015) database for use in SMIRFE. As NAP can only decrease along a path and mass can only increase along the path, traversal can be short-circuited whenever a nascent path has a NAP below the min_NAP or mass above the max_mass. Once all IMFs have been enumerated for all EMFs, an index is built for the IMF database on IMF mass. These two components combined with the adduct description provided to the assignment algorithm completes the internal representation of the search space needed for SMIRFE.

Preliminary EMF Search

Searching all peaks in a spectrum against all possible IMFs for any sufficiently large search space quickly becomes computationally intractable. For example, \(\text{C}_6\text{H}_{12}\text{O}_6\) has 1,119,744 possible isotopologues; however, without extensive labeling, most of these isotopologues are too low in abundance to be detected based on natural abundance probabilities. Searching all of these isotopologues would not only be wasteful, but would also generate many false assignments. This search can be constrained by observing that for a given labeled version of an EMF, there is only one most-abundant IMF that must be present if that labeled EMF exists in the sample. For example, without labeling, the monoisotopic version of \(\text{C}_6\text{H}_{12}\text{O}_6\) is the most abundant IMF of glucose and must be observed if unlabeled glucose is present in the sample. If only \(m+\text{^{13}C}_1\) glucose was observed and no \(^{13}C\) labeling source is present, it can be inferred that this assignment is spurious and should be ignored. However, if a \(^{13}C\) labeling source is present, detection of the \(m+\text{^{13}C}_1\) glucose isotopologue is possible without observing the monoisotopic version.

By searching each peak in a spectrum against the untargeted EMF search space generated previously and testing if each returned IMF is a most abundant IMF of a possible labeled EMF based on the user-provided labeling pattern, a set of possible EMFs in the spectrum can be generated. The set of possible adducts is user specified (e.g. Na⁺, H⁺, K⁺, NH₄⁺), and, for every peak, the database query is performed for each adduct as the adduct’s mass must be accounted for during searching. This set of possible EMFs informs the more targeted IMF search in the next step of SMIRFE assignment. To further eliminate spurious assignments, only peaks present in most scans (>90%) are used for building this possible EMF set. This effectively eliminates the possibility that noise is included in this search and restricts the search to peaks corresponding to high-abundance IMFs. IMFs that do not correspond to the most abundant are stored as a source of random assignments needed for statistical testing (i.e. even false positives are useful in the right statistical situation).

Targeted IMF Searching

With possible EMFs identified, the next step is a more targeted search for the IMFs of those EMFs. This is done at the subclique level and for every possible subclique for every clique for that EMF, the possible IMFs are enumerated in order of descending relative NAP (to a minimum relative NAP). For each IMF, peaks with matching m/z are tentatively assigned to that IMF. If there are multiple possible matching peaks (very rare), the peak with the closest matching mass is assigned.

Given the limitations of m/z-based matching, IMFs after the first IMF must meet additional requirements to be tentatively assigned. First, the nth IMF may only be assigned if the scans in which the nth IMF was observed overlap significantly (>90%) with the scans in which the n-1th IMF was observed. Second, since labeling may only increase the relative abundance of an IMF but never decrease it, IMFs should be observed in the order in which they are enumerated (i.e. it does not make sense to see an IMF with a lower relative NAP but not see an IMF of the same subclique with a higher relative NAP). Third, the nth IMF must have at least one intensity ratio with another IMF in the subclique with a score greater than the 95th percentile of the random intensity ratio scores of the non-matching, non-most abundant IMFs stored from the EMF search step. The third rule is omitted in the case where an IMF may be more heavily labeled. For example, if \(^{13}\text{C}\) is presumed labeled and the \(m+\text{^{13}C}_2\) subclique of Na-glucose is being enumerated, the \(m+\text{^{13}C}_1\) isotopologue might be observed but does not have to have a matching intensity ratio as it could represent either \(^{13}\text{C}_2\) from labeling and \(^{13}\text{C}_1\) from natural abundance (in which case it should have a matching intensity ratio) or \(^{13}\text{C}_1\) from labeling and no \(^{13}\text{C}\) from natural abundance (in which case it may not have a matching intensity ratio). If at any point one of these rules is broken, further enumeration of the subclique is terminated.

Anytime a subclique contains two or more IMFs, it is possible to evaluate the likelihood of observing a set of IMFs that match the intensity ratios observed for that subclique at random. This is reported as an e-value and is calculated using equation 4.

\[ E \sim \frac{\log_{10}(\text{median}(R)) - \log_{10}(\text{median}(R))}{\sigma_{\log_{10}(R)}} \times j \in \text{scans, } i_1 > 0, i_2 > 0 \]

\[ R = \log_{10}(i_{1,j}) - \log_{10}(i_{2,j}) \]

\[ \sigma_{\log_{10}(R)} = \text{variance}(R) \]

4
Equation 3 - The score for a pair of isotopologues A, B (S_{A,B}) in the same subclique can be assigned an E value E, which represents the probability that random pairs of isotopologues would have the same distribution of chi squared statistics as the observed subclique. This calculation requires constructing the score stat vector S for all unique pairs of isotopologues in the subclique (I). The variance for these statistics is the variance of the random pairs of isotopologues constructed earlier from the invalid isotopologue assignments. A combined chi-squared statistic is then calculated for the subclique constructed earlier from the invalid isotopologue assignments. A robust estimate both peak m/z and peak intensity.

\[ \chi_{A,B} = \frac{(\log_{10}(NAP_A) - \log_{10}(NAP_B) - M_{\text{log\_ratios}})^2}{\sigma^2_{\text{log\_ratios}}} \]

\[ S_{A,B} = 1 - \text{chi}_2\text{cdf}(\chi_{A,B}, df = 1) \]

Equation 4 - Every subclique containing at least two isotopologues can be assigned an E value E, which represents the probability that random pairs of isotopologues would have the same distribution of chi squared statistics as the observed subclique. This calculation requires constructing the score stat vector S for all unique pairs of isotopologues in the subclique (I). The variance for these statistics is the variance of the random pairs of isotopologues constructed earlier from the invalid isotopologue assignments. A combined chi-squared statistic is then calculated for the subclique by multiplying S by its transpose with which the E value can be calculated. This formulation enables the future correction of potential correlation between isotopologue pair statistic values and does not require modification if the underlying isotopologue pair scoring function is changed.

Dataset for Validation

The method for generating our validation dataset was adapted from previously published method for performing ethylchloroformate (ECF) amino acid derivatization (Yang et al., 2017).

Reagents

Unlabeled amino acid standards were purchased from Sigma Aldrich as a mixture of acidic and neutral amino acids and basic amino acids (A6407, A6282). 15N-labeled amino acids mixture was purchased from Cambridge Isotope Laboratories (NLM-6695). Ethylchloroformate was purchased from also purchased from Sigma Aldrich.

Preparation of Amino Acid Standard Solutions

2 μL of 2.5 mM Sigma A6407 was added to 2 μL of 2.5mM Sigma A6282 and 5 μL of mM unlabeled glutamine were combined to yield a sample containing unlabeled and labeled amino acids. In the hood, 100 μL of H2O/EtOH/Pyridine (6:3:1 by volume) was added to each sample and vortexed to let the sample dissolve. 5μL of ECF was then added to sample. The samples were then vortexed for 30 seconds and then spun down. Under the hood, 100 μL of CHCl3 (RT) were added and shaken vigorously at 3000rpm for 3 minutes with the Disruptor Genie. The samples were then centrifuged at 21100g on a Thermo Scientific Legend Micro 21R for 10 minutes 4C. The bottom CHCl3 layer was then transferred to a new glass vial and capped. 10 μL of 7M NaOH was added into the remaining aqueous phase to adjust the pH to 10. An additional 5 μL of ECF was then added and the samples centrifuged and extract again to ensure complete derivatization. The samples were then vortexed again and then diluted 10x in 90% acetonitrile (in water) with 0.2 μM of tetraethylammonium and 2) μM NaCl in 0.5mL snap-cap tubes. These were then loaded onto 384-well plate that was pre-washed twice with ddH2O and 1x with acetonitrile for direct infusion FTMS.

Mass Spectrometry Configuration

FTMS analysis was performed using a Tribrid Fusion Orbitrap interfaced with an Advion Triversa Nanomate. The Nanomate’s operating voltage was 1.5kV and the head pressure was 0.5 psi. Spectra were acquired in positive mode. The maximum ion time for automatic gain control was set to 100ms, 5 microscans were acquired per scan and total acquisition time was approximately 5 minutes, resulting in approximately 100 total scans. Spectra were acquired for m/z’s in the 100 to 1000 range selected using quadrupole isolation. Aggregate spectra were constructed from the scans using an in-house method that normalizes scans and uses inter-scan peak correspondence to robustly estimate both peak m/z and peak intensity.

Figure 3 – ECF Replicate Spectra

Figure 3 – ECF Replicate Spectra. Shown in panels A and B are the aggregate level MS1 spectra for ECF replicates 1 and 2 respectively. These visualizations were generated using Xcalibur. All the expected ECF derivatives have masses below 400 m/z and most of the peak density in these spectra occur below 400 m/z as well. Few peaks occur past 500 m/z. Both spectra contain over 100 scans and each were acquired using a microscan setting of 5 and a resolution of 500K. Additional details are provided in materials and methods regarding spectral acquisition. Although fuzzy sites occur in these spectra, they are otherwise free of obvious artifacts.
Manual Inspection of Spectra

Both reference spectra were examined to verify the presence of peaks corresponding to expected derivatives of the amino acids known to be present in the sample and the absence of any large spectral defects. Although fuzzy site FT-MS artifacts are present in the spectrum (Mitchell et al., 2018), these artifacts do not overlap with any of the important non-artifactual peaks. Peaks corresponding to derivatives of 19 of 20 amino acids (or 18 of 19 as isoleucine and leucine are isomers) were observed. No peaks were observed that would correspond to the expected Na+ or H+ adducts of cysteine. The regions of spectra where peaks corresponding to these amino acids are expected to occur are shown in Supplemental Figure 1.

RESULTS

EMF Search Space Growth

The number of possible IMFs in a given EMF search space, without limitations on NAP, valid EMF, and mass, quickly makes a brute force enumeration of all IMFs impossible. For example, the search space C: 100, N: 7, O: 40, H: 230, P: 3, S: 3, contains 1.021x10^29 possible IMFs! However, since the partial EMF search space must only contain highly abundant IMFs for those partial EMFs, the number of possibilities can be greatly reduced by enforcing a minimum relative NAP for the IMF entries in the database. When labeling is considered, the effect of the NAP filter is reduced, but the database remains a manageable size and is sufficiently performant to be queried efficiently with observed m/z’s from a spectrum. For example, the search space used to assign our reference spectra with a min_NAP of 0.4 contains 13,045,817 IMFs with no labeling, but increases to 53,102,054 with 15N-labeling. The size of SMIRFE search spaces with respect to min_NAP, mass and labeling is shown in Figure 4.

Assignments for the ECF Derivatized Spectra

Using a C:130, N:7, O:40, P:3, S:3, H:230, max_mass: 1600, min_NAP: 0.4, labeled_isotopes: 15N, the EMF search space was constructed using SMIRFE for the ECF derivatized spectra. This database was used to identify likely EMFs in the spectrum, which were then followed by a targeted IMF search for those EMFs. Na+ and H+ adducts were considered for every peak. From previous studies, the expected derivatives of each amino acid with ECF is known (Yang et al., 2017). Of the 20 amino acids present in the solution (representing 19 derivatives with distinct formulas – leucine and isoleucine are isomers), the expected derivatives of 18 out of 19 of those formula-distinct amino acid species were assigned by SMIRFE. No assignments were made to cysteine which was confirmed to be absent in the spectrum and likely was lost during sample processing. The 15N-labeled versions of these derivatives were also identified as well as multiple adducted versions of both labeled and unlabeled versions. The distinct IMFs identified for each amino acid are summarized in Supplemental Tables 1 and 2.
Assignment Ambiguity

Intensity ratio scoring combined with scan subsetting and a tight m/z match tolerance significantly narrows the number of possible assignments for a peak. Without these filters, the number of possible assignments for a peak can be large even with high mass accuracy (Figure 5). At low m/z, when the number of possible EMFs and therefore IMFs is relatively small, this can result in unique IMF assignments for those peaks; however, at higher m/z when the number of possible assignments is very large, unique assignments are rare. Multiple assignments can be disambiguated by their e-value in many cases. In our test case, every expected EMF has at least one assignment that had the best e-value for a peak and many of the observed IMFs for the expected EMFs were the best assignment for their peak (82.4% for sample 1 and 82.6% for sample 2).

Assignment m/z error

From the set of IMF assignments for expected EMFs in our example dataset, the patterns in mass error can be investigated. Both spectra were acquired using the same Thermo Tribrid Fusion instrument for which the stated mass accuracy is 1 ppm. The observed mass errors fall within this specification for 100% of the assigned IMFs for the validated EMFs. For the set of all assignments, validated plus unvalidated assignments and including all possible assignment for each peak, only 34 of 24215 fall outside the 1 ppm error specification. When only the best assignment for each peak is considered (lowest E-value), only 1 of 975 assignments have a ppm error over 1 ppm. These results are consistent with previous findings that high mass accuracy alone cannot provide unambiguous metabolite assignments (Kind and Fiehn, 2006).

The mass error across m/z is not constant (as might be expected from miscalibration) and instead changes with m/z. During manual investigation of these spectra, it was observed that peaks near intense peaks in m/z space often had a larger mass error than peaks far away from other intense peaks. This effect can be attributed to coulombic interactions between peaks and can be quantified with Equation 5:

\[ \text{Repulsion} = \sum_{j=0}^{n} \frac{\text{sign}(f_i - f_j)}{|f_i - f_j|^2} \]

The repulsion expression is derived from Coloumb’s law but as the distance between ion clouds in the orbitrap is not known, the average distance between two clouds can be estimated by the different in their precession frequencies \(f_i\) and \(f_j\). In our example, we cannot directly obtain the precession frequencies either, but a proxy frequency can be derived from the observed m/z and the estimated digital resolution. The force on an ion in cloud \(i\) depends also on the number of ions in cloud \(j\) (assuming they are all charge \(z=1\)). Summing this force across all other ion clouds for each ion cloud \(i\) gives a net repulsion term for each cloud. Ions with higher repulsion experience more of a net interaction with other ions that can potentially introduce error into their observed m/z’s. These repulsive mass errors can exceed 1 ppm in size (Philip Remes, 2016). The mass error with respect to m/z and repulsion on the validated IMF assignments are shown in Figure 6.

DISCUSSION

SMIRFE Algorithm

Although more untargeted than existing assignment methods, SMIRFE is not a completely untargeted assignment tool (i.e. can search all possible EMFs and all possible IMFs for any number of any elements for all possible labelling patterns). Enumerating all possible IMFs even for relatively small search spaces quickly becomes intractable both in terms of runtime and...
storage space. The hybrid untargeted / targeted approach employed by SMIRFE leverages the best of both methodologies with the main limitation being the need of the end-user to specify the EMF search space to search. Given that most biomolecules are CHONPS-based, a search space for CHONPS plus expected adducts (in this case Na+ and H+ adducts) only with element counts large enough to cover 99%+ of known biomolecules in the desired mass range will be applicable for most experiments. When an additional element is needed, this is usually known ahead of time and can be trivially added (i.e. if you are looking for vanadium containing metabolites, you know to add vanadium to the search space).

In these more complicated search spaces, the hybrid untargeted / targeted assignment method employed by SMIRFE becomes advantageous. Since the untargeted EMF database must only contain the most abundant IMFs for the EMFs in the search space, the runtime and storage cost of expanding a search space is minimized. Additionally, since these databases are reusable, the cost of building a large database is amortized across all experiments that can utilize it. Despite these improvements in performance, large heavily-labeled EMF search spaces will take considerable resources to generate and triply- or quadruply-labeled full-size search spaces are likely impractical to enumerate.

As mentioned previously, m/z-based matching alone is error prone without cross-validating information supporting the assignment of an IMF. SMIRFE utilizes the relative intensity ratios observed between pairs of assigned IMF peaks and compares these ratios to the ratio of the relative NAP of the assigned IMFs as a source of cross-validating information. While absolute peak intensities are not necessarily trustworthy in mass spectrometry, the relative peak intensities can be made trustworthy through proper preprocessing. Scan subsetting effectively enforces that IMFs of the same EMF appear in the same scans. This additional rule filters out nonsense groups of peaks that happen to match on intensity and mass.

Unique to SMIRFE is the ability to quantify statistically the fit of the generated assignments. This is achieved using the e-value calculation that comes from the clique score. Although the ‘correct’ assignment is not always the best scoring assignment, the e-value still captures how likely an assignment that is compatible with the peaks could have occurred at random. These e-values are useful in downstream analyses built upon SMIRFE assignments.

Implications for Experimental Design

The ability to assign a FT-MS spectrum depends highly on the quality of the spectrum. As SMIRFE assigns FT-MS spectra in a fundamentally different manner than targeted tools, spectra

---

Figure 6 – Mass Error vs. m/z and Repulsion. For the set of IMFs corresponding to expected EMFs in our test case, the mass error of those assignments can be calculated. The Fusion instrument has a stated 1ppm accuracy and most of the observed mass errors fall within this specification. The observed mass errors are not constant (and thus not easily correctable) and they change with respect to m/z (A – ECF Replicate 1, C – ECF Replicate 2). Mass error does correlate with the measured repulsion (see eq. 4) for the peak which estimates the net columbic forces on the ion cloud corresponding to that IMF (B – ECF Replicate 1, D – ECF Replicate 2). Correcting for both an m/z dependent and repulsion dependent m/z error without standards in an untargeted context, while difficult, could substantially improve the ability to differentiate plausible IMF assignments.
can be collected in a way that optimizes them for assignment by SMIRFE. First, additional MS1 scans should be acquired whenever possible. Additional MS1 scans improve aggregate peak characteristics that in turn enable better intensity ratio comparisons, but also improve the likelihood that scan subsetting will eliminate spurious assignments. Since SMIRFE cannot currently use MS2 data in any manner to improve assignment accuracy, MS2 time can be used for additional MS1 scans. Second, since SMIRFE attempts to assign sets of IMFs of an EMF, it is necessary to have enough dynamic range to observe at least two or more IMFs for each EMF. A dynamic range of at least 1000 is sufficient for many examples, but higher dynamic range is always advantageous. While not always possible, ensuring that one or more peaks do not dominate the spectrum ensures that the effective dynamic range of a spectrum is maximized. However, because overloading of ion traps and orbitraps can result in space charge effects that limit mass accuracy and resolution, effective dynamic range must be balanced against effective mass resolution. Alternatively, rather than acquiring a spectrum for the entire m/z domain in a single acquisition, multiple spectra for subdomains of m/z can be acquired, assigned and reassembled to yield an assigned spectrum for the desired m/z domain. This splitting should provide each sub-spectrum with improved effective dynamic range at the cost of increased processing times.

Mass Accuracy and SMIRFE

Despite the use of NAP-based cross-validating information to support SMIRFE generated IMF assignments, the m/z of a peak is still important for determining which IMFs it could possibly match. For targeted assignment tools, a mass accuracy of 1 ppm is typically sufficient to achieve unique assignments (although not necessarily correct assignments) up to approximately 800 m/z. Due to the large number of possible assignments that SMIRFE considers, a 1 ppm accuracy does not always translate into unambiguous assignments even when additional information is considered.

Furthermore, a 1 ppm mass error while small in absolute terms is very large in relation to the resolving power of FT-MS instruments. If the mass error and resolution were on similar scales, the score of a clique could incorporate the “goodness” of the m/z matching to further disambiguate multiply assigned peaks. Our investigation of the mass error trends in our example spectra revealed that the mass error depends both on m/z and on proximity of peaks to other intense peaks in a spectrum. Thus, the 1 ppm designation of “accuracy” is illusionary, since multiple sources of error can sum above 1 ppm. Further studies are needed to develop robust methods of correcting both the m/z and repulsion component of the mass error for the IMF assignments. Ideally this correction needs to be feasible with limited knowledge about the composition of the sample to enable its use in an untargeted experiment.

Conclusions

SMIRFE is a novel assignment algorithm that enables an unprecedented level of untargeted assignment for direct-infusion FT-MS spectra. SMIRFE uses a hybrid untargeted / targeted assignment approach that considers effectively all EMFs in a specified search space in a preliminary untargeted search that informs a more targeted IMF search. This hybrid approach min-
SMIRFE is part of United States patent application US20180011990A1.

ACKNOWLEDGMENT

We would like to thank Qing Jun Wang (University of Kentucky) who prepared the amino acid replicates and acquired the spectra. We would also like to thank Jim Carreau who developed the initial prototype of SMIRFE.

ABBREVIATIONS

EMF, elemental molecular formula; IMF, isotopically resolved molecular formula; NAP, natural abundance probability.

REFERENCES

Carreer, W., Flight, R. and Moseley, H. (2013) A Computational Framework for High-Throughput Isotopic Natural Abundance Correction of Omics-Level Ultra-High Resolution FT-MS Datasets. Metabolites 3, 853.

Chekmeneva, E., Dos Santos Correia, G., Chan, Q., Wijeyesekera, A., Tin, A., Young, J.H., Elliott, P., Nicholson, J.K. and Holmes, E. (2017) Optimization and Application of Direct Infusion Nanoelectrospray HRMS Method for Large-Scale Urinary Metabolic Phenotyping in Molecular Epidemiology. J Proteome Res 16, 1646-1658.

Creek, D.J., Chokkathukalam, A., Jankevics, A., Burgess, K.E.V., Breitling, R. and Barrett, M.P. (2012) Stable Isotope-Assisted Metabolomics for Network-Wide Metabolic Pathway Elucidation. Analytical Chemistry 84, 8442-8447.

Dettmer, K., Aronov, P.A. and Hammock, B.D. (2007) MASS SPECTROMETRY-BASED METABOLICOMICS. Mass spectrometry reviews 26, 51-78.

Eliuk, S. and Makarov, A. (2015) Evolution of Orbitrap Mass Spectrometry Instrumentation. Annual Review of Analytical Chemistry 8, 61-80.

Eyles, S.J. and Kaltashov, I.A. (2004) Methods to study protein dynamics and folding by mass spectrometry. Methods 34, 88-99.

Fahy, E., Sud, M., Cotter, D. and Subramaniam, S. (2007) LIPID MAPS online tools for lipid research. Nucleic Acids Research 35, W606-W612.

Fan, T.W., Lane, A.N., Higashi, R.M., Farag, M.A., Gao, H., Bousamra, M. and Miller, D.M. (2009) Altered regulation of metabolic pathways in human lung cancer discerned by (13)C stable isotope-resolved metabolomics (SIRM). Mol Cancer 8, 41.

Fan, T.W.M., Lorkiewicz, P., Sellers, K., Moseley, H.N.B., Higashi, R.M. and Lane, A.N. (2012) Stable isotope-resolved metabolomics and applications for drug development. Pharmacology & therapeutics 133, 366-391.

Fredrick Johansson, e.a. (2018) mpmath: a Python library for arbitrary-precision floating-point arithmetic.

Harris, D.M., Li, L., Chen, M., Lagunero, F.T., Go, V.L. and Boros, L.G. (2012) Diverse mechanisms of growth inhibition by luteolin, resveratrol, and quercetin in MIA PaCa-2 cells: a comparative glucose tracer study with the fatty acid synthase inhibitor C75. Metabolomics 8, 201-210.

Higashi, R.M., Fan, T.W.M., Lorkiewicz, P.K., Moseley, H.N.B. and Lane, A.N. (2014) Stable Isotope Labeled Tracers for Metabolic Pathway Elucidation by GC-MS and FT-MS. Methods in molecular biology (Clifton, N.J.) 1198, 147-167.

Hillier, K., Metullo, C.M., Kelleher, J.K. and Stephanopoulos, G. (2010) Nontargeted elucidation of metabolic pathways using stable-isotope tracers and mass spectrometry. Anal Chem 82, 6621-8.

Hipp, R., et. al. and (2015) SQLite.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. (2016) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Research 44, D457-D462.

Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A., Wang, J., Yu, B., Zhang, J. and Bryant, S.H. (2016) PubChem Substance and Compound databases. Nucleic Acids Research 44, D1202-D1213.

Kind, T. and Fiehn, O. (2006) Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. BMC Bioinformatics 7, 234.

Kind, T. and Fiehn, O. (2007) Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics 8, 105.

Lane, A.N., Fan, T.W., Xie, Z., Moseley, H.N. and Higashi, R.M. (2009) Isotopomer analysis of lipid biosynthesis by high resolution mass spectrometry and NMR. Anal Chim Acta 651, 201-8.

Mitchell, J.M., Fan, T.W.M., Lane, A.N. and Moseley, H.N.B. (2014) Development and in silico evaluation of large-scale metabolite identification methods using functional group detection for metabolomics. Frontiers in Genetics 5, 237.

Mitchell, J.M., Flight, R.M., Wang, Q.J., Higashi, R.M., Fan, T.W.M., Lane, A.N. and Moseley, H.N.B. (2018) New methods to identify high peak density artifacts in Fourier transform mass spectra and to mitigate their effects on high-throughput metabolomic data analysis. Metabolomics 14, 125.

Moseley, H.N. (2010) Correcting for the effects of natural abundance in stable isotope resolved metabolomics experiments involving ultra-high resolution mass spectrometry. BMC Bioinformatics 11, 139.

Moseley, H.N.B. (2013) ERROR ANALYSIS AND PROPAGATION IN METABOLICOMICS DATA ANALYSIS. Computational and Structural Biotechnology Journal 4, e201301006.

Peake, D.A., Yokoi, Y., Wang, J. and Yingying, H. (2013) A New Lipid Software Workflow for Processing Orbitrap-based Global Lipidomics Data in Translational and Systems Biology Research in Scientific, T.F. (Ed).

Pence, H.E. and Williams, A. (2010) ChemSpider: An Online Chemical Information Resource. Journal of Chemical Education 87, 1123-1124.

Philip Remes, J.C., Mike Senko, Romain Huguet, Vlad Zabrouskov, Nina Soltero, Shannon Eliuk (2016) Improved Control of Ion Population for Orbitrap Mass Analysis, Thermo Fisher Scientific.

Schrimpf-Rutledge, A.C., Codreanu, S.G., Sherrod, S.D. and McLean, J.A. (2016) Untargeted metabolomics strategies—Challenges and Emerging Directions. Journal of the American Society for Mass Spectrometry 27, 1897-1905.
Sellers, K., Fox, M.P., Bousamra, M., 2nd, Slone, S.P., Higashi, R.M., Miller, D.M., Wang, Y., Yan, J., Yuneva, M.O., Deshpande, R., Lane, A.N. and Fan, T.W. (2015) Pyruvate carboxylase is critical for non-small-cell lung cancer proliferation. *J Clin Invest* **125**, 687-98.

Walt, S.v.d., Colbert, S.C. and Varoquaux, G. (2011) The NumPy array: a structure for efficient numerical computation. *Computing in Science & Engineering* **13**, 22-30.

Watson, D.G. (2013) A Rough Guide to Metabolite Identification Using High Resolution Liquid Chromatography Mass Spectrometry in Metabolomic Profiling in Metazoans. *Computational and Structural Biotechnology Journal* **4**, e201301005.

Wishart, D.S., Feunang, Y.D., Marcu, A., Guo, A.C., Liang, K., Vázquez-Fresno, R., Sajed, T., Johnson, D., Li, C., Karu, N., Sayeeda, Z., Lo, E., Assempour, N., Berjanskii, M., Singhal, S., Arndt, D., Liang, Y., Badran, H., Grant, J., Serra-Cayuela, A., Liu, Y., Mandal, R., Neveu, V., Pon, A., Knox, C., Wilson, M., Manach, C. and Scalbert, A. (2018) HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Research* **46**, D608-D617.

Yang, Y., Fan, T.W.M., Lane, A.N. and Higashi, R.M. (2017) Chloroformate derivatization for tracing the fate of Amino acids in cells and tissues by multiple stable isotope resolved metabolomics (mSIRM). *Analytica Chimica Acta* **976**, 63-73.
Calculation of Search Space Sizes

The number of IMFs in a search space can be calculated from the number of possible elements of a given type in the search space definition and the number of isotopes for that element. Equivalent duplicate combinations of isotopes must not be recounted to have a correct estimate, for example $^{12}\text{C}^{13}\text{C}$ is not distinct from $^{12}\text{C}^{13}\text{C}$ and a “missing” carbon.

\[
\text{Possible IMFs} = \prod_{E} \sum_{\text{allE}}^{{\text{MaxE}}} \text{num\_stable\_isotopes}_E^k
\]

The \text{allE} variable is the set of all elements in the search space, and \text{E} is the current element being considered. \text{MaxE} is the max atom count of that element in the search space and \text{k} is the number of atoms of that element being considered at any one time. For elements with two stable isotopes, the sum becomes $2^0 + 2^1 + 2^2 \ldots + 2^{\text{MaxE}}$, the product of each sum for each element is the number of possible IMFs in the search space. This enumeration includes formulas that are not chemically possible but nonetheless exist in the space.

Supplemental Figure 1 – Regions of Spectra Showing Absence of Cysteine

Supplemental Figure 1 - Regions of Spectra Showing Absence of Cysteine: The H+ adduct of ECF derivatized cysteine has a theoretical monoisotopic mass of 294.100586, while the Na+ adduct has a theoretical monoisotopic mass of 316.082531. Shown in panels A, B are the regions of spectra corresponding to the hydrogen-adducted cysteine derivative in the amino acid (AA) samples 1 and 2 respectively. Panels C, D show the sodium-adduct region in AA samples 1 and 2 respectively. For the hydrogen adduct, in both A and B the peaks at 294.10598 or 294.10595 are intense but are too far from the expected m/z to be reasonably assigned to a cysteine derivative. In C and D, the peak at 316.082 has an m/z very close to the expected m/z of the sodium adduct but this peak is not very intense. If this peak was the
monoisotopic isotopologue of the cysteine derivative, the instrument has insufficient dynamic range to observe the less abundant isotopologues.

**SUPPLEMENTAL TABLE 1: ASSIGNMENT RESULTS FOR AMINO ACID SPECTRUM 1**

| Amino Acid | Adduct | Observed Mass | Assignment Rank | Number of Assignments | Absolute Mass Error | E Value | Isotopologue Formula |
|------------|--------|---------------|----------------|-----------------------|---------------------|---------|---------------------|
| Glycine Na | 198.0736959 | 1 | 1 | -1.73E-05 | 2.97E-06 | 12C7,1H13,14N1,16O4 |
|            | 199.0707249 | 1 | 1 | -1.14E-05 | 2.97E-06 | 12C7,1H13,15N1,16O4 |
|            | 199.0770304 | 1 | 1 | 3.06E-06 | 2.97E-06 | 12C6.13C1,1H13,14N1,16O4 |
|            | 200.07408 | 1 | 1 | -1.17E-05 | 8.41E-06 | 12C6.13C1,1H13,15N1,16O4 |
|            | 200.0779317 | 1 | 1 | -8.12E-06 | 2.97E-06 | 12C7,1H13,14N1,16O3.18O1 |
|            | 200.0804207 | 1 | 1 | -3.24E-05 | 8.41E-06 | 12C5.13C2,1H13,14N1,16O4 |
|            | 201.074962 | 1 | 1 | -3.46E-06 | 8.41E-06 | 12C7,1H13,15N1,16O3.18O1 |
|            | 201.0773893 | 1 | 1 | 3.39E-05 | 8.41E-06 | 12C5.13C2,1H13,15N1,16O4 |
| Alanine Na | 212.0893101 | 1 | 1 | 1.86E-05 | 4.55E-06 | 12C8,1H15,14N1,16O4 |
|            | 213.0863416 | 1 | 1 | 2.20E-05 | 0.000232433 | 12C8,1H15,15N1,16O4 |
|            | 213.0926203 | 1 | 1 | 6.32E-05 | 4.55E-06 | 12C7.13C1,1H15,14N1,16O4 |
|            | 214.0896961 | 1 | 2 | 2.23E-05 | 0.000232433 | 12C7.13C1,1H15,15N1,16O4 |
|            | 214.0935393 | 1 | 1 | 3.44E-05 | 4.55E-06 | 12C8,1H15,14N1,16O3.18O1 |
|            | 214.0960062 | 1 | 1 | 3.21E-05 | 4.55E-06 | 12C6.13C2,1H15,14N1,16O4 |
|            | 215.0905958 | 1 | 1 | 1.28E-05 | 0.000232433 | 12C8,1H15,15N1,16O3.18O1 |
|            | 215.0930661 | 1 | 2 | 7.17E-06 | 0.000232433 | 12C6.13C2,1H15,15N1,16O4 |
| Serine Na  | 228.084206 | 1 | 2 | 3.72E-05 | 0.095682723 | 12C8,1H15,14N1,16O5 |
|            | 229.0812568 | 1 | 1 | 2.13E-05 | 0.384927895 | 12C8,1H15,15N1,16O5 |
|            | 229.0875587 | 1 | 2 | 3.94E-05 | 0.095682723 | 12C7.13C1,1H15,14N1,16O5 |
|            | 230.0846166 | 1 | 1 | 1.64E-05 | 0.454987864 | 12C7.13C1,1H15,15N1,16O5 |
|            | 230.0884647 | 1 | 1 | 2.36E-05 | 0.095682723 | 12C8,1H15,14N1,16O4.18O1 |
|            | 231.0855184 | 1 | 1 | 4.74E-06 | 0.886669761 | 12C8,1H15,15N1,16O4.18O1 |
| Proline Na | 238.1049503 | 1 | 1 | 2.85E-05 | 2.10E-06 | 12C10,1H17,14N1,16O4 |
|            | 239.1019827 | 1 | 1 | 3.09E-05 | 0.009527526 | 12C10,1H17,15N1,16O4 |
|            | 239.1082758 | 1 | 2 | 5.78E-05 | 2.10E-06 | 12C9.13C1,1H17,14N1,16O4 |
|            | 240.1053548 | 1 | 3 | 1.37E-05 | 0.009527526 | 12C9.13C1,1H17,15N1,16O4 |
|            | 240.1091928 | 1 | 1 | 3.09E-05 | 2.10E-06 | 12C10,1H17,14N1,16O3.18O1 |
|            | 240.1116853 | 1 | 2 | 3.14E-06 | 2.10E-06 | 12C8.13C2,1H17,14N1,16O4 |
|            | 241.1062592 | 1 | 1 | -5.40E-07 | 0.009527526 | 12C10,1H17,15N1,16O3.18O1 |
|            | 241.1087143 | 1 | 3 | 9.00E-06 | 0.009527526 | 12C8.13C2,1H17,15N1,16O4 |
| Valine Na  | 240.1205808 | 1 | 1 | 4.80E-05 | 7.29E-05 | 12C10,1H19,14N1,16O4 |
|            | 241.1176063 | 1 | 1 | 5.74E-05 | 0.00011425 | 12C10,1H19,15N1,16O4 |
|            | 241.1238882 | 1 | 2 | 9.54E-05 | 7.29E-05 | 12C9.13C1,1H19,14N1,16O4 |
|            | 242.1209571 | 1 | 3 | 6.15E-05 | 0.00011425 | 12C9.13C1,1H19,15N1,16O4 |
|            | 242.1247883 | 1 | 1 | 8.55E-05 | 7.29E-05 | 12C10,1H19,14N1,16O3.18O1 |
| Compound | Formula | Mass Da | Charge | Molar Mass | PPM Error | Retention Time |
|----------|---------|---------|--------|------------|-----------|----------------|
| Threonine Na | 242.1272574 | 1 | 2 | 8.11E-05 | 7.29E-05 | 12C8.13C2,1H19,14N1,16O4 |
| | 243.1218538 | 1 | 1 | 5.49E-05 | 0.00011425 | 12C10.1H19.15N1,16O3.18O1 |
| | 243.1243180 | 1 | 3 | 5.53E-05 | 0.00011425 | 12C8.13C2,1H19,15N1,16O4 |
| | 242.0998336 | 1 | 2 | 5.98E-05 | 0.000862401 | 12C9.1H17,14N1,16O5 |
| | 243.0968783 | 1 | 1 | 5.49E-05 | 0.00011425 | 12C10.1H19.15N1,16O4 |
| | 243.0968783 | 1 | 3 | 5.53E-05 | 0.00011425 | 12C8.13C2,1H19,15N1,16O4 |
| | 244.1002402 | 1 | 2 | 4.29E-05 | 1.40E-06 | 12C8.13C2,1H17,14N1,16O5 |
| | 244.1002402 | 1 | 1 | 4.29E-05 | 1.40E-06 | 12C8.13C2,1H17,14N1,16O5 |
| | 244.1040810 | 1 | 1 | 5.74E-05 | 0.00011425 | 12C9.1H17,15N1,16O4.18O1 |
| | 244.1040810 | 1 | 2 | 5.74E-05 | 0.00011425 | 12C9.1H17,15N1,16O4.18O1 |
| | 245.1011224 | 1 | 1 | 5.08E-05 | 1.40E-06 | 12C9.1H17,15N1,16O4.18O1 |
| | 245.1011224 | 1 | 2 | 3.64E-05 | 1.40E-06 | 12C7.13C2,1H17,15N1,16O4.18O1 |
| Cysteine | Not Observed | | | | | |
| Leucine / Iso-leucine Na | 255.1332318 | 1 | 1 | 8.19E-05 | 0.000245615 | 12C11,1H21,15N1,16O4 |
| | 256.1365881 | 1 | 3 | 8.05E-05 | 0.000245615 | 12C10.13C1,1H21,15N1,16O4 |
| | 257.1374679 | 2 | 2 | 9.08E-05 | 0.000245615 | 12C11,1H21,15N1,16O3.18O1 |
| | 257.1399208 | 1 | 3 | 0.000102666 | 0.000245615 | 12C9.13C2,1H21,15N1,16O4 |
| Asparagine Na | 255.0951077 | 1 | 2 | 3.46E-05 | 0.855371888 | 12C9.1H16,14N2,16O5 |
| | 256.0984141 | 1 | 2 | 8.32E-05 | 0.855371888 | 12C8.13C1,1H16,14N2,16O5 |
| | 257.0891289 | 1 | 1 | 8.32E-05 | 0.413738984 | 12C9.1H16,15N2,16O5 |
| | 258.0925295 | 1 | 1 | 3.74E-05 | 0.413738984 | 12C8.13C1,1H16,15N2,16O5 |
| Aspartate Na | 284.1103927 | 1 | 3 | 6.53E-05 | 2.76E-06 | 12C11,1H19,14N1,16O6 |
| | 285.1074049 | 1 | 1 | 8.80E-05 | 0.088684845 | 12C11,1H19,15N1,16O6 |
| | 285.1135558 | 1 | 4 | 0.000157041 | 2.76E-06 | 12C10.13C1,1H19,14N1,16O6 |
| | 286.1107792 | 1 | 3 | 6.85E-05 | 0.088684845 | 12C10.13C1,1H19,15N1,16O6 |
| | 286.1145409 | 1 | 1 | 0.0001621 | 2.76E-06 | 12C11,1H19,14N1,16O5.18O1 |
| | 286.1170814 | 1 | 2 | 8.64E-05 | 2.76E-06 | 12C9.13C2,1H19,14N1,16O6 |
| Glutamine Na | 269.1107418 | 1 | 2 | 5.06E-05 | 4.02E-07 | 12C10.1H18,14N2,16O5 |
| | 270.1078199 | 1 | 2 | 4.54E-05 | 8.57E-06 | 12C10.1H18,14N1.15N1,16O5 |
| | 270.1141079 | 1 | 2 | 3.93E-05 | 4.02E-07 | 12C9.13C1,1H18,14N2,16O5 |
| | 271.1047843 | 1 | 1 | 7.78E-05 | 0.000240697 | 12C10.1H18,15N2,16O5 |
| | 271.1110665 | 1 | 2 | 0.000115598 | 0.25218173 | 12C9.13C1,1H18,14N1.15N1,16O5 |
| | 271.1149699 | 1 | 1 | 6.75E-05 | 4.02E-07 | 12C10.1H18,14N2,16O4.18O1 |
| | 271.1173965 | 1 | 1 | 0.000105589 | 8.57E-06 | 12C8.13C2,1H18,14N2,16O5 |
| | 272.1081345 | 1 | 3 | 8.26E-05 | 0.000240697 | 12C9.13C1,1H18,15N2,16O5 |
| | 273.1090291 | 1 | 1 | 7.82E-05 | 0.000240697 | 12C10.1H18,15N2,16O4.18O1 |
| | 273.1115011 | 1 | 3 | 7.07E-05 | 0.000240697 | 12C8.13C2,1H18,15N2,16O5 |
| Lysine H | 319.1863015 | 2 | 2 | 6.16E-05 | 0.096679318 | 12C14,1H26,14N2,16O6 |
| | 320.1896696 | 2 | 2 | 4.82E-05 | 0.096679318 | 12C13.13C1,1H26,14N2,16O6 |
| | 321.1803932 | 2 | 3 | 3.98E-05 | 0.206780265 | 12C14,1H26,15N2,16O6 |
| | 322.1837512 | 2 | 3 | 3.64E-05 | 0.206780265 | 12C13.13C1,1H26,15N2,16O6 |
| Na | 341.1681970 | 1 | 4 | 0.000110256 | 5.90E-06 | 12C14,1H26,14N2,16O6 |
| Molecule                  | Property | Value | Charge | Mass | Formula          |
|--------------------------|----------|-------|--------|------|------------------|
| Glutamate Na             | pKa      | 4.15  | 1      | 342.1653339 | 12C14,1H26,14N1,15N1,16O6 |
|                         | pKa      | 4.17  | 1      | 342.1715823 | 12C13,13C1,1H26,14N2,16O6 |
|                         | pKa      | 4.15  | 1      | 343.1622588 | 12C14,1H26,15N2,16O6 |
|                         | pKa      | 4.12  | 1      | 343.1723885 | 12C14,1H26,14N2,16O5.18O1 |
|                         | pKa      | 4.18  | 1      | 343.1748228 | 12C12,13C2,1H26,14N2,16O6 |
|                         | pKa      | 4.16  | 1      | 344.1656351 | 12C13,13C1,1H26,15N2,16O6 |
|                         | pKa      | 4.18  | 2      | 345.1665349 | 12C14,1H26,15N2,16O5.18O1 |
|                         | pKa      | 4.18  | 1      | 345.1689477 | 12C12,13C2,1H26,15N2,16O6 |
| Glutamate Na             | pKa      | 298.1260879 | 1 | 2.02E-05 | 0.162541309 | 12C12,1H21,14N1,16O6 |
|                         | pKa      | 299.1230193 | 1 | 0.000123715 | 2.32E-09 | 12C12,1H21,15N1,16O6 |
|                         | pKa      | 299.1293178 | 1 | 0.000145106 | 0.162541309 | 12C11,13C1,1H21,14N1,16O6 |
|                         | pKa      | 300.1263943 | 1 | 0.00010351 | 2.32E-09 | 12C11,13C1,1H21,15N1,16O6 |
|                         | pKa      | 300.130265 | 1 | 8.81E-05 | 0.369560639 | 12C12,1H21,14N1,16O5.18O1 |
|                         | pKa      | 301.1272967 | 1 | 9.13E-05 | 2.32E-09 | 12C12,1H21,15N1,16O5.18O1 |
|                         | pKa      | 301.1297723 | 1 | 8.04E-05 | 2.32E-09 | 12C10,13C2,1H21,15N1,16O6 |
| Methionine Na            | pKa      | 272.0926229 | 1 | 7.71E-05 | 0.0027612 | 12C10,1H19,14N1,16O4,32S1 |
|                         | pKa      | 273.0896592 | 1 | 7.56E-05 | 0.000331506 | 12C10,1H19,15N1,16O4,32S1 |
|                         | pKa      | 273.0959633 | 1 | 9.15E-05 | 0.0027612 | 12C9,13C1,1H19,14N1,16O4,32S1 |
|                         | pKa      | 274.0884097 | 1 | 8.62E-05 | 0.0027612 | 12C10,1H19,14N1,16O4,34S1 |
|                         | pKa      | 274.0930105 | 1 | 7.92E-05 | 0.000331506 | 12C9,13C1,1H19,15N1,16O4,32S1 |
|                         | pKa      | 274.0968566 | 1 | 8.84E-05 | 0.007232483 | 12C10,1H19,14N1,16O3.18O1,32S1 |
|                         | pKa      | 275.0854578 | 1 | 7.29E-05 | 0.000331506 | 12C10,1H19,15N1,16O4,34S1 |
| Histidine H              | pKa      | 328.1502213 | 3 | 9.05E-05 | 0.033634505 | 12C14,1H21,14N3,16O6 |
|                         | pKa      | 329.1535742 | 3 | 9.25E-05 | 0.033634505 | 12C13,13C1,1H21,14N3,16O6 |
|                         | pKa      | 330.1544391 | 2 | 9.11E-05 | 0.023476844 | 12C14,1H21,15N3,16O6 |
|                         | pKa      | 331.1413254 | 2 | 9.11E-05 | 0.023476844 | 12C14,1H21,15N3,16O6 |
|                         | pKa      | 332.1446704 | 2 | 0.000100938 | 0.023476844 | 12C13,13C1,1H21,15N3,16O6 |
| Phenylalanine Na         | pKa      | 350.1321456 | 1 | 9.87E-09 | 0.00011048 | 12C14,1H21,14N3,16O6 |
|                         | pKa      | 351.1292574 | 1 | 9.87E-09 | 0.00011048 | 12C14,1H21,14N2.15N1,16O6 |
|                         | pKa      | 351.1355161 | 1 | 9.87E-09 | 0.00011048 | 12C13,13C1,1H21,14N3,16O6 |
|                         | pKa      | 352.1325474 | 1 | 9.84E-05 | 0.000221287 | 12C13,13C1,1H21,14N2.15N1,16O6 |
|                         | pKa      | 352.1364507 | 1 | 5.04E-05 | 9.87E-09 | 0.00011048 | 12C14,1H21,14N3,16O5.18O1 |
|                         | pKa      | 352.1383332 | 1 | 9.87E-09 | 0.00011048 | 12C12,13C2,1H21,14N3,16O6 |
|                         | pKa      | 353.123266  | 2 | 9.48E-05 | 0.001384663 | 12C14,1H21,15N3,16O6 |
|                         | pKa      | 354.1266488 | 2 | 6.69E-05 | 0.001384663 | 12C13,13C1,1H21,15N3,16O6 |
|                         | pKa      | 355.1275527 | 2 | 5.31E-05 | 0.005957809 | 12C14,1H21,15N3,16O5.18O1 |
| Phenylalanine Na         | pKa      | 288.1205249 | 1 | 7.35E-05 | 0.000103897 | 12C14,1H19,14N1,16O4 |
|                         | pKa      | 289.1175705 | 1 | 9.32E-05 | 0.000592449 | 12C14,1H19,15N1,16O4 |
|                         | pKa      | 289.1238486 | 1 | 7.35E-05 | 0.000135082 | 12C13,13C1,1H19,14N1,16O4 |
|                         | pKa      | 290.1209271 | 1 | 9.14E-05 | 0.000592449 | 12C13,13C1,1H19,15N1,16O4 |
|                         | pKa      | 290.124766  | 1 | 7.35E-05 | 0.0001078 | 12C14,1H19,14N1,16O3.18O1 |
| Amino Acid | Adduct | Observed Mass | Assignment Rank | Number of Assignments | Absolute Mass Error | E Value | Isotopologue Formula |
|------------|--------|---------------|-----------------|-----------------------|---------------------|---------|---------------------|
| Glycine    | Na     | 198.0737033   | 1               | 1                     | -2.47E-05           | 0.000625043 | 12C7,1H3,14N1,16O4 |
|            |        | 199.0707318   | 1               | 1                     | -1.83E-05           | 1.58E-05 | 12C7,1H3,15N1,16O4 |
|            |        | 199.077029    | 1               | 1                     | 4.43E-06            | 0.000625043 | 12C6.13C1,1H3,14N1,16O4 |
|            |        | 200.0740865   | 1               | 1                     | -1.82E-05           | 1.58E-05 | 12C6.13C1,1H3,15N1,16O4 |
|            |        | 200.0779392   | 1               | 1                     | -1.56E-05           | 0.000625043 | 12C7,1H3,14N1,16O3,18O1 |
|            |        | 200.0804153   | 1               | 1                     | -2.70E-05           | 0.000625043 | 12C5.13C2,1H3,14N1,16O4 |
|            |        | 201.0749682   | 1               | 1                     | -9.73E-06           | 1.58E-05 | 12C7,1H3,15N1,16O3,18O1 |
|            |        | 201.0774326   | 1               | 1                     | -9.41E-06           | 1.58E-05 | 12C5.13C2,1H3,15N1,16O4 |
| Alanine    | Na     | 212.0893161   | 1               | 1                     | 1.26E-05            | 7.17E-09 | 12C8,1H5,14N1,16O4 |
|      | Mass (Da) | Charge | % Intensity | m/z | Formula | Mw  |
|------|----------|--------|-------------|-----|---------|-----|
| Serine Na | 213.0863448 | 1 | 1 | 1.88E-05 | 0.000576917 | 12C8,1H15,15N1,16O4 |
|       | 213.0926138 | 1 | 1 | 6.97E-05 | 7.17E-09 | 12C8,1H15,14N1,16O4 |
|       | 214.0896991 | 1 | 2 | 1.93E-05 | 0.000576917 | 12C7.13C1,1H15,15N1,16O4 |
|       | 214.0935338 | 1 | 1 | 3.99E-05 | 7.17E-09 | 12C8,1H15,14N1,16O3.18O1 |
|       | 214.0960064 | 1 | 1 | 3.19E-05 | 7.17E-09 | 12C6.13C2,1H15,14N1,16O4 |
|       | 215.0905984 | 1 | 1 | 1.01E-05 | 0.000576917 | 12C8,1H15,15N1,16O3.18O1 |
|       | 215.0930687 | 1 | 2 | 4.55E-06 | 0.000576917 | 12C6.13C2,1H15,15N1,16O4 |
| Proline Na | 228.0842124 | 1 | 2 | 3.08E-05 | 0.012432375 | 12C8,1H15,14N1,16O5 |
|       | 229.0812634 | 1 | 1 | 1.48E-05 | 0.092379263 | 12C8,1H15,15N1,16O5 |
|       | 229.0875626 | 1 | 2 | 3.56E-05 | 0.012432375 | 12C7.13C1,1H15,14N1,16O5 |
|       | 230.0846174 | 1 | 1 | 1.56E-05 | 0.152464377 | 12C7.13C1,1H15,15N1,16O5 |
|       | 230.088475 | 1 | 1 | 1.33E-05 | 0.012432375 | 12C8,1H15,14N1,16O4.18O1 |
|       | 230.090991 | 1 | 1 | -3.81E-05 | 0.352363626 | 12C6.13C2,1H15,14N1,16O5 |
|       | 231.0855113 | 1 | 1 | 1.18E-05 | 0.152464377 | 12C8,1H15,15N1,16O4.18O1 |
| Valine Na | 238.1049581 | 1 | 1 | 2.07E-05 | 2.46E-08 | 12C10,1H17,14N1,16O4 |
|       | 239.1019884 | 1 | 1 | 2.52E-05 | 0.000275554 | 12C10,1H17,15N1,16O4 |
|       | 239.1082783 | 1 | 2 | 5.53E-05 | 2.46E-08 | 12C9.13C1,1H17,14N1,16O4 |
|       | 240.1053635 | 1 | 3 | 4.95E-06 | 0.000275554 | 12C9.13C1,1H17,15N1,16O4 |
|       | 240.1092013 | 1 | 1 | 2.24E-05 | 2.46E-08 | 12C10,1H17,14N1,16O3.18O1 |
|       | 240.1116809 | 1 | 2 | 7.48E-06 | 2.46E-08 | 12C8.13C2,1H17,14N1,16O4 |
|       | 241.1062569 | 1 | 1 | 1.69E-06 | 0.000275554 | 12C10,1H17,15N1,16O3.18O1 |
|       | 241.1087282 | 1 | 3 | -4.94E-06 | 0.000275554 | 12C8.13C2,1H17,15N1,16O4 |
| Threonine Na | 240.1205866 | 1 | 1 | 4.22E-05 | 0.128022286 | 12C10,1H19,14N1,16O4 |
|       | 241.1176125 | 1 | 1 | 5.12E-05 | 2.99E-06 | 12C10,1H19,15N1,16O4 |
|       | 241.1238887 | 1 | 2 | 9.49E-05 | 0.128022286 | 12C9.13C1,1H19,14N1,16O4 |
|       | 242.1209607 | 1 | 3 | 5.78E-05 | 2.99E-06 | 12C9.13C1,1H19,15N1,16O4 |
|       | 243.1218519 | 1 | 1 | 5.68E-05 | 2.99E-06 | 12C10,1H19,15N1,16O3.18O1 |
|       | 243.1243141 | 1 | 3 | 5.92E-05 | 2.99E-06 | 12C8.13C2,1H19,15N1,16O4 |
| Cysteine | Not Observed | | | | |
| Leucine / Isoleucine H | 232.1543248 | 1 | 1 | 9.78E-06 | 0.999995517 | 12C11,1H21,14N1,16O4 |
|       | 233.1513521 | 1 | 1 | 1.74E-05 | 0.97225498 | 12C11,1H21,15N1,16O4 |
| Compound          | Formula               | Charge | Purity         | Molecular Weight | Isotopic Mass   | Molecular Formula |
|-------------------|-----------------------|--------|----------------|------------------|----------------|------------------|
| Asparagine Na     | 255.1332371           | 1      | 1.93E-05       | 0.99995517       | 12C10.13C1,1H21,14N1,16O4 |
|                   | 256.136593            | 1      | 3.75E-05       | 0.97225498       | 12C10.13C1,1H21,15N1,16O4 |
|                   | 257.1399176           | 1      | 0.000105836    | 1.34E-05         | 12C9.13C2,1H21,15N1,16O4 |
| Aspartate Na      | 255.0951121           | 1      | 3.02E-05       | 0.775428645      | 12C9.1H16,14N2,16O5 |
|                   | 256.0984422           | 1      | 5.50E-05       | 0.775428645      | 12C8.13C1,1H16,14N2,16O5 |
|                   | 257.0891273           | 1      | 8.49E-05       | 0.017920166      | 12C9.1H16,15N2,16O5 |
|                   | 258.0925187           | 1      | 4.82E-05       | 0.017920166      | 12C8.13C1,1H16,15N2,16O5 |
| Glutamine Na      | 263.1255046           | 1      | 4.41E-05       | 0.139689595      | 12C11.1H19,15N1,16O6 |
|                   | 264.1287968           | 1      | 0.000106726    | 0.139689595      | 12C10.13C1,1H19,15N1,16O6 |
|                   | 265.1104013           | 1      | 5.68E-05       | 1.05E-06         | 12C11.1H19,14N1,16O6 |
|                   | 265.1074123           | 1      | 8.06E-05       | 0.0003516835     | 12C11.1H19,15N1,16O6 |
|                   | 265.1136573           | 1      | 0.000155531    | 1.05E-06         | 12C10.13C1,1H19,14N1,16O6 |
|                   | 266.1107874           | 1      | 6.04E-05       | 0.003516835      | 12C10.13C1,1H19,15N1,16O6 |
|                   | 266.1145799           | 1      | 0.00012316     | 1.05E-06         | 12C11.1H19,14N1,16O5,18O1 |
|                   | 266.1170948           | 1      | 7.29E-05       | 1.05E-06         | 12C9.13C2,1H19,14N1,16O6 |
|                   | 267.1115962           | 1      | 0.000141704    | 0.003516835      | 12C11.1H19,15N1,16O5,18O1 |
|                   | 267.114133            | 1      | 6.96E-05       | 0.003516835      | 12C9.13C2,1H19,15N1,16O6 |
| Lysine            | 319.186309            | 2      | 5.40E-05       | 0.032332202      | 12C14.1H26,14N2,16O6 |
|                   | 320.1896734           | 2      | 4.45E-05       | 0.032332202      | 12C13.13C1,1H26,14N2,16O6 |
|                   | 321.1803857           | 2      | 4.71E-05       | 0.046232267      | 12C14.1H26,15N2,16O6 |
|                   | 322.1837339           | 2      | 5.38E-05       | 0.046232267      | 12C13.13C1,1H26,15N2,16O6 |
|                   | 341.1682083           | 1      | 9.90E-05       | 5.03E-07         | 12C14.1H26,14N2,16O6 |
|                   | 342.1653121           | 1      | 3.01E-05       | 5.03E-07         | 12C14.1H26,14N1,15N1,16O6 |
|                   | 342.1715903           | 1      | 7.18E-05       | 5.03E-07         | 12C13.13C1,1H26,14N2,16O6 |
|                   | 343.1622679           | 1      | 0.000109195    | 0.000127583      | 12C14.1H26,15N2,16O6 |
|                   | 343.1723799           | 1      | 0.00017241     | 5.03E-07         | 12C14.1H26,14N2,16O5,18O1 |
| Glutamate Na | 343.174817 | 1 | 5 | 0.000199936 | 5.03E-07 | 12C12.13C2,1H26,14N2,16O6 |
| | 344.1656444 | 1 | 8 | 8.75E-05 | 0.000127583 | 12C13.13C1,1H26,15N2,16O6 |
| | 345.166534 | 1 | 3 | 8.80E-05 | 0.000127583 | 12C14,1H26,15N2,16O5.18O1 |
| | 345.1689626 | 1 | 6 | 0.000124085 | 0.000127583 | 12C12.13C2,1H26,15N2,16O6 |
| | 298.1261013 | 1 | 3 | 6.76E-06 | 0.376348014 | 12C12,1H21,14N1,16O6 |
| | 299.1229898 | 1 | 2 | 0.000153205 | 9.12E-08 | 12C12,1H21,15N1,16O6 |
| | 299.1293193 | 1 | 4 | 8.80E-05 | 0.000127583 | 12C11.13C1,1H21,15N1,16O6 |
| | 300.1202897 | 1 | 1 | 6.34E-05 | 0.704166485 | 12C12,1H21,14N1,16O5.18O1 |
| | 301.1273099 | 1 | 1 | 7.81E-05 | 9.12E-08 | 12C12,1H21,15N1,16O5.18O1 |
| | 301.1297725 | 1 | 3 | 8.01E-05 | 9.12E-08 | 12C10.13C2,1H21,15N1,16O6 |
| Methionine Na | 272.092628 | 1 | 2 | 7.20E-05 | 0.000175939 | 12C10,1H19,14N1,16O4,32S1 |
| | 273.0896731 | 2 | 4 | 6.18E-05 | 0.0955851 | 12C10,1H19,15N1,16O4,32S1 |
| | 273.095668 | 1 | 4 | 8.80E-05 | 0.000175939 | 12C9.13C1,1H19,14N1,16O4,32S1 |
| | 274.0884138 | 1 | 3 | 8.20E-05 | 0.000175939 | 12C10,1H19,14N1,16O4,34S1 |
| | 274.093015 | 2 | 4 | 7.47E-05 | 0.0955851 | 12C9.13C1,1H19,15N1,16O4,32S1 |
| | 274.0968631 | 1 | 1 | 8.19E-05 | 0.0001552495 | 12C10,1H19,14N1,16O3.18O1,32S1 |
| | 275.0854558 | 2 | 2 | 7.49E-05 | 0.175367468 | 12C10,1H19,15N1,16O4,34S1 |
| H | 328.1502189 | 3 | 5 | 9.29E-05 | 0.092895179 | 12C14,1H21,14N3,16O6 |
| | 329.1535689 | 3 | 8 | 9.78E-05 | 0.092895179 | 12C13.13C1,1H21,14N3,16O6 |
| | 330.1544958 | 2 | 2 | 9.70E-05 | 0.230771567 | 12C14,1H21,14N3,16O5.18O1 |
| | 331.1413182 | 2 | 5 | 9.83E-05 | 0.000641318 | 12C14,1H21,15N3,16O6 |
| | 332.1446673 | 2 | 5 | 0.000104128 | 0.000641318 | 12C13.13C1,1H21,15N3,16O6 |
| Histidine Na | 350.132156 | 1 | 7 | 0.00010011 | 4.61E-10 | 12C14,1H21,14N3,16O6 |
| | 351.1292322 | 1 | 6 | 5.88E-05 | 4.61E-10 | 12C14,1H21,14N2.15N1,16O6 |
| | 351.1355268 | 1 | 13 | 8.42E-05 | 4.61E-10 | 12C13.13C1,1H21,14N3,16O6 |
| | 352.1325194 | 4 | 6 | 0.000126419 | 0.783902465 | 12C13.13C1,1H21,14N2.15N1,16O6 |
| | 352.1364362 | 1 | 8 | 6.49E-05 | 4.61E-10 | 12C14,1H21,14N3,16O5.18O1 |
| | 352.1388712 | 1 | 9 | 9.46E-05 | 4.61E-10 | 12C12.13C2,1H21,14N3,16O6 |
| | 353.1232754 | 3 | 8 | 8.54E-05 | 0.054450655 | 12C14,1H21,15N3,16O6 |
| | 354.1266272 | 3 | 8 | 8.85E-05 | 0.054450655 | 12C13.13C1,1H21,15N3,16O6 |
| | 355.1275036 | 3 | 5 | 0.000102206 | 0.164854564 | 12C14,1H21,15N3,16O5.18O1 |
| Phenylalanine Na | 288.1205313 | 1 | 2 | 9.75E-05 | 1.86E-06 | 12C14,1H19,14N1,16O4 |
| | 289.1175764 | 1 | 3 | 8.73E-05 | 5.17E-07 | 12C14,1H19,15N1,16O4 |
| | 289.123853 | 1 | 5 | 0.000130642 | 1.86E-06 | 12C13.13C1,1H19,14N1,16O4 |
| | 290.1209334 | 1 | 7 | 8.52E-05 | 5.17E-07 | 12C13.13C1,1H19,15N1,16O4 |
| | 290.1247589 | 1 | 2 | 0.00011492 | 1.86E-06 | 12C14,1H19,14N1,16O3.18O1 |
| | 290.1272397 | 1 | 5 | 9.88E-05 | 1.86E-06 | 12C12.13C2,1H19,14N1,16O4 |
| | 291.1218115 | 1 | 3 | 9.72E-05 | 5.17E-07 | 12C14,1H19,15N1,16O3.18O1 |
|         | M  | A  | Name            | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
|---------|----|----|----------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Arginine |    |    | Arginine H     | 291.1242835 | 1  | 5  | 8.99E-05 | 5.17E-07 | 12C12.13C2,1H19,15N1,16O4 |
|         |    |    | Tyrosine       | 275.1713078 | 2  | 2  | 7.38E-05 | 0.036570786 | 12C11,1H22,14N4,16O4 |
|         |    |    | Tyrosine       | 276.1683757 | 1  | 1  | 4.09E-05 | 0.036570786 | 12C11,1H22,14N3,15N1,16O4 |
|         |    |    | Tyrosine       | 276.17468 | 2  | 2  | 5.65E-05 | 0.036570786 | 12C10.13C1,1H22,14N4,16O4 |
|         |    |    | Tyrosine       | 279.1594783 | 1  | 5  | 4.29E-05 | 0.074386212 | 12C11,1H22,15N4,16O4 |
|         |    |    | Tyrosine       | 280.1628303 | 1  | 5  | 4.58E-05 | 0.074386212 | 12C10.13C1,1H22,15N4,16O4 |
|         |    |    | Tyrosine       | 354.1546342 | 4  | 5  | 9.44E-05 | 0.194330657 | 12C17,1H23,14N1,16O7 |
|         |    |    | Tyrosine       | 355.1517371 | 2  | 2  | 2.63E-05 | 0.035385186 | 12C17,1H23,15N1,16O7 |
|         |    |    | Tyrosine       | 355.1579812 | 4  | 5  | 0.00010213 | 0.194330657 | 12C16.13C1,1H23,14N1,16O7 |
|         |    |    | Tyrosine       | 356.1550026 | 2  | 2  | 0.000115675 | 0.035385186 | 12C16.13C1,1H23,15N1,16O7 |
| Tyrosine |    |    | Tyrosine       | 376.1365903 | 1  | 7  | 8.25E-05 | 3.59E-07 | 12C17,1H23,14N1,16O7 |
|         |    |    | Tyrosine       | 377.133661 | 2  | 7  | 4.67E-05 | 0.001643498 | 12C17,1H23,15N1,16O7 |
|         |    |    | Tyrosine       | 377.1398802 | 1  | 15 | 0.00014744 | 3.59E-07 | 12C16.13C1,1H23,14N1,16O7 |
|         |    |    | Tyrosine       | 378.1369907 | 2  | 15 | 7.18E-05 | 0.001643498 | 12C16.13C1,1H23,15N1,16O7 |
|         |    |    | Tyrosine       | 378.1407967 | 2  | 8  | 0.000121048 | 3.59E-07 | 12C17,1H23,14N1,16O6.18O1 |
|         |    |    | Tyrosine       | 378.1432217 | 1  | 11 | 0.000160735 | 3.59E-07 | 12C15.13C2,1H23,14N1,16O7 |
|         |    |    | Tyrosine       | 379.1378583 | 5  | 13 | 9.43E-05 | 0.001643498 | 12C17,1H23,15N1,16O6.18O1 |
|         |    |    | Tyrosine       | 379.1403316 | 2  | 11 | 8.57E-05 | 0.001643498 | 12C15.13C2,1H23,15N1,16O7 |
|         |    |    | Tryptophan     | 305.1495484 | 1  | 3  | 3.52E-05 | 2.20E-06 | 12C16,1H20,14N2,16O4 |
|         |    |    | Tryptophan     | 306.1528972 | 1  | 3  | 4.12E-05 | 2.20E-06 | 12C15.13C1,1H20,14N2,16O4 |
|         |    |    | Tryptophan     | 327.1314879 | 1  | 3  | 3.99E-05 | 7.48E-08 | 12C16,1H20,14N2,16O4 |
|         |    |    | Tryptophan     | 328.1285425 | 2  | 5  | 2.03E-05 | 3.08E-08 | 12C16,1H20,14N1.15N1,16O4 |
|         |    |    | Tryptophan     | 328.1348273 | 1  | 6  | 5.54E-05 | 7.48E-08 | 12C15.13C1,1H20,14N2,16O4 |
|         |    |    | Tryptophan     | 329.1255065 | 1  | 4  | 9.12E-05 | 0.000141009 | 12C16,1H20,15N2,16O4 |
|         |    |    | Tryptophan     | 329.1318135 | 1  | 5  | 0.000104103 | 0.753050157 | 12C15.13C1,1H20,14N1.15N1,16O4 |
|         |    |    | Tryptophan     | 329.135674 | 1  | 3  | 9.88E-05 | 7.48E-08 | 12C16,1H20,14N2,16O3.18O1 |
|         |    |    | Tryptophan     | 329.1381252 | 1  | 6  | 0.000112314 | 7.48E-08 | 12C14.13C2,1H20,14N2,16O4 |
|         |    |    | Tryptophan     | 330.1288598 | 1  | 10 | 9.26E-05 | 0.000141009 | 12C15.13C1,1H20,15N2,16O4 |
|         |    |    | Tryptophan     | 331.132228 | 1  | 7  | 7.93E-05 | 0.000141009 | 12C14.13C2,1H20,15N2,16O4 |
