Published in final edited form as:

Anal Chem. 2016 March 01; 88(5): 2747–2754. doi:10.1021/acs.analchem.5b04263.

Isotopic Ratio Outlier Analysis (IROA) of the *S. cerevisiae* metabolome using accurate mass GC-TOF/MS: A new method for discovery

Yunping Qiu¹, Robyn Moir², Ian Willis², Chris Beecher³, Yu-Hsuan Tsai⁴, Timothy J. Garrett⁵, Richard A. Yost^{4,5}, and Irwin J. Kurland¹

¹Stable Isotope and Metabolomics Core Facility, Diabetes Center, Department of Medicine, Albert Einstein College of Medicine, Bronx, NY 10461, USA

²Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York 10461, USA

³IROA Technologies, Ann Arbor, MI 48105

⁴Department of Chemistry, University of Florida, Gainesville, FL, 32611

⁵Department of Pathology, Immunology, and Laboratory Medicine, University of Florida, Gainesville, FL, 32611

Abstract

Isotopic Ratio Outlier Analysis (IROA) is a ¹³C metabolomics profiling method that eliminates sample-to-sample variance, discriminates against noise and artifacts, and improves identification of compounds, previously done with accurate mass LC/MS. This is the first report using IROA technology in combination with accurate mass GC-TOFMS, here used to examine the *S. cerevisiae* metabolome. *S. cerevisiae* was grown in YNB media, containing randomized 95% ¹³C, or 5% ¹³C glucose as the single carbon source, in order that the isotopomer pattern of all metabolites would mirror the labeled glucose. When these IROA experiments are combined, the abundance of the heavy isotopologues in the 5% ¹³C extracts, or light isotopologues in the 95% ¹³C extracts, follows the binomial distribution, showing mirrored peak pairs for the molecular ion. The mass difference between the ¹²C monoisotopic and the ¹³C monoisotopic equals the number of carbons in the molecules. The IROA GC/MS protocol developed, using both Chemical and Electron Ionization, extends the information acquired from the isotopic peak patterns for formulae generation, a process that can be formulated as an algorithm, in which the number of carbons, as well as the number of methoximations and silylations, are used as search constraints. In Electron Impact (EI/IROA) spectra, the artifactual peaks are identified and easily removed, which has the potential to generate “clean” EI libraries. The combination of Chemical Ionization (CI) IROA and EI IROA affords a metabolite identification procedure that enables the identification of co-eluting metabolites, and allowed us to characterize 126 metabolites in the current study.

Supporting Information

Supporting Information Available: This material is available free of charge via the Internet at <http://pubs.acs.org>. Chromatography and mass spectrum of three co-eluting authentic standards; all CI IROA spectra and retention times; metabolites confirmed by retention times; metabolites detected with both with methoximation and ethoximation; the formulae generated from these CI IROA spectral patterns; summary of the coefficient of variation for those 126 detected metabolites.

TOC image

Use of gas chromatography-accurate mass GC-TOF/MS with Isotope Ratio Outlier Analysis (IROA), we could identify co-eluting metabolites and artifacts in the chemical ionization mode (GC-CI), and identify mass fragments from electron ionization (GC-EI) containing carbons only of biological origin, which allows the generation of “clean” EI libraries without artifactual fragments.

Introduction

Gas chromatography/mass spectrometry has long been the premier tool for small molecule analyses¹⁻⁷. Electron Impact (EI) molecular fragmentation, and extremely high resolution chromatographic separations, are both considered highly reproducible, and have allowed the formation of extensive GC/MS spectral libraries that are extremely useful in the area of metabolite profiling. None-the-less, despite the fact that over 1 million GC/MS EI spectra exist between the Wiley, NIST, Fiehn and Golm libraries, the majority mass spectral tags (MSTs) seen during an average biological experiment may remain unidentified⁸. Typically, in a GC-MS metabolomics experiment only 5-15% of the features found can be identified⁸.

Despite the wealth of GC/MS spectral libraries, it is often difficult to identify and quantitate peaks because of a general lack of 1) authenticated reference compounds⁸, and 2) stable isotope labeled reference substances⁹. Furthermore, unit mass GC/MS spectra lack the mass accuracy needed for the structure elucidation of metabolites in samples that have spectra not identifiable in known databases. Thus for exploring metabolic networks, there is a need for a global method to characterize and discriminate metabolites (that arise from metabolic networks) from chemical artifacts (plasticizers, airborne and waterborne contaminants, silylation artifacts, etc.) and noise, and to support quantitation.

Many approaches have been used to aid unknown compound identification, recently including the use of accurate mass GC/MS¹⁰⁻¹². Fiehn *et al.* derivatized a plant leaf extract containing many unknown compounds with N-tert-Butyldimethylsilyl-N-methyltrifluoroacetamide (MTBSTFA), and sought to ascertain the molecular weight (for elemental composition calculation) by locating M-57 EI fragments¹. Kind and Fiehn established that, in general, high mass accuracy (<1 ppm) alone is not sufficient for unknown metabolite formulae candidates with complex elemental compositions, and that the use of isotopic abundance patterns greatly eliminates false candidate formulae¹³. Peterson *et al.* have used accurate mass GC-Orbitrap and isotopic abundance patterns for extending metabolite identification based on a GC-Orbitrap workflow to annotate metabolites in *Arabidopsis thaliana* extract¹². In addition to MTBSTFA and MSTFA derivatization, the

authors used ^{13}C and ^{15}N labeling to obtain the number of carbon and nitrogen groups in the compound, and used positive chemical ionization (PCI) and EI data to increase the structural information. However, the methods they developed for the GC-Orbitrap MS are not easily adaptable to other accurate mass GC-Time of Flight-MS (GC-TOF/MS) instruments. High Resolution Filtering (HRF) is another GC-Orbitrap based method, without ^{13}C labeling, that attempts to annotate every measured m/z peak in an EI mass spectrum, leveraging accurate mass to discern between possible identifications¹⁴.

Isotopic Ratio Outlier Analysis (IROA) is an isotope labeling technology to generate heavy biomolecules with ^{13}C in a target cell population¹⁵⁻¹⁷. While mass spectrometry has been used for examining the yeast metabolome with stable isotope labeling¹⁸⁻²², this is the first application of IROA technology for the study of yeast. *S. cerevisiae* is a simple model organism that can easily accommodate IROA labeling, using randomized 95% ^{13}C glucose or 5% ^{13}C glucose (in YNB media) as a single carbon source. This enables global profiling, via IROA labeling of all metabolites with the same isotopomer distribution as the labeled glucose. When these IROA experiments are combined, the abundance of the heavy isotopologues in the 5% ^{13}C samples (M+1, M+2, etc., the ^{12}C envelope), or light isotopologues in the 95% ^{13}C samples (M-1, M-2, etc., the ^{13}C envelope), follows the binomial distribution, showing mirrored peak pairs for the molecular ion in all metabolites of biologic origin. The mass difference between the ^{12}C monoisotopic peak, and the ^{13}C monoisotopic peak, indicates the number of carbons (n) in the metabolite's carbon backbone. The peaks in the ^{13}C envelope can be also considered as internal standards for validation of mass assignments, and have been used in IROA LC/MS studies for normalization between control (^{13}C) and treated (^{12}C) groups¹⁷. These unique mirror patterns enable the IROA peak pairs to discriminate between metabolites of biological origin vs. artifacts, an advance from previous yeast stable isotope labeling work (see references 18-22). IROA profiling was successfully used in LC-MS based global metabolomics analysis of *Caenorhabditis elegans*¹⁷.

In GC-MS, the complexity of the IROA signal contains additional information from the presence of the derivatization groups, the IROA peak pair spectral signature for GC/MS is similar to the IROA peak pair spectral signature used for LC/MS^{15,17,23}, but differs in having additional mass isotopomer information from the ratio of the (M0+n+1)/(M0+n) peaks (^{13}C envelope), which may be used for calculating the number of silylation groups per metabolite. The M0+n+1 and higher peaks, normally have very small contributions from the metabolite (deuterium, ^{15}N , or ^{17}O , and negligible ^{14}C), and the majority contribution to these peaks comes from the derivatization reagents. Therefore, an analysis of the intensity of these peaks allows the direct determination of the number of times the molecule is derivatized. This is a tremendous benefit, previously unavailable for IROA, which aids in formulae elucidation from either Chemical Ionization (CI) or EI data. In addition, the IROA patterns for CI accurate mass GC/MS have, by the nature of the IROA experiment, generated a ^{13}C internal standard for every metabolite. Our accurate mass GC-TOF/MS workflow utilized both EI and PCI on metabolite extracts derivatized with BSTFA, and deuterated BSTFA was also used for verification of the number of silylation groups per metabolite. The IROA EI pattern also may be used to determine the number of carbons in any EI fragment, which may be used to support structure elucidation of an unknown, as well as confirming

which parts of the EI fragmentation spectrum is of biologic origin, allowing one to remove the non-biological, artifactual, peaks. In summary, GC/MS IROA has the potential for satisfying unmet needs for small molecule identification, and is independent of the accurate mass GC/MS platform used.

Material and Methods

Yeast strains (*Saccharomyces cerevisiae*) were grown from OD 0.2 to OD 2.5 in IROA YNB medium, containing 5% or 95% ^{13}C glucose (IROAtech Inc.). After harvesting with rapid filtration, as per Crutchfield *et al.* and Bennett *et al.*^{19,24}, yeast was transferred to a 2 ml screw-capped polypropylene tube. The samples were added with similar amount of glass beads and 1mL pre-cooled (-20°C) extraction solution (4:4:2 (v:v:v) methanol:acetonitrile:water). 10 μL of 1mM of chlorophenylalanine in water and 5 μL of 10mM cholesterol in methanol were spiked into the yeast suspensions as internal standards, as both of these metabolites are not found in yeast, and cover polar and non-polar metabolites. Metabolite extraction was performed with a bead beater machine for 4 cycles of 30 seconds. After centrifugation, the supernatant was transferred to a new tube. The extraction was repeated twice with 500 μL of extract solution, and all supernatants were combined.

Yeast extracts, either from 5% ^{13}C extracts, 95% ^{13}C extracts, or combined 5% ^{13}C and 95% ^{13}C (pooled) were dried under gentle nitrogen flow. The dried samples were derivatized with a two-step derivatization procedure. First, the samples were treated with 50 μL of methoxyamine hydrochloride (MOX, 15 mg/mL in pyridine) or ethylhydroxylamine hydrochloride (EOX, 15 mg/mL in pyridine) at 30°C for 90 minutes. The silylation step was done with 50 μL of N,O-Bis(trimethylsilyl)trifluoroacetamide (BSTFA, containing 1% TMCS, Supelco, PA, USA) or with BSTFA_d9 (CDN isotopes, Canada) at 80°C for 60 minutes. After the samples were cooled down at room temperature, the samples were analyzed by GC-TOF/MS (GCT Premier, Waters, USA). Separation was performed with a DB-5MS column (30-meter, $0.25\text{ mm} \times 0.25\text{ }\mu\text{m}$) coupled with 10-meter guard column. Helium was used as carrier gas at a flow of 1mL/min. The oven program is as the following: starting at 60°C for 1min, the temperature was allowed to rise by $10^{\circ}\text{C}/\text{min}$ to 320°C and held at 320°C for an additional 3 min before returning to starting conditions.

To avoid saturation, and to obtain low intensity metabolites, 1 μL samples were injected with 2 different split ratios (splitless and 1:10 split). The injection temperature was set to 270°C . The transfer line temperature was set to 250°C . The data was collected with the scan time of 0.3s and inter-scan delay of 0.1s. The mass range was set to 35 to 650 for EI, and 80 to 1200 for PCI. 95% methane/5% ammonia was used as the PCI reagent gas. The lock mass was 218.9856 for EI and 286.0027 for PCI. Five sets of samples and derivatization were analyzed: 1) the pooled IROA extracts (combined 5% ^{13}C and 95% ^{13}C metabolite extracts, derivatized with MOX and unlabeled BSTFA and analyzed, in both EI and PCI modes (Fig. 1), 2) the 5% ^{13}C metabolite extract derivatized with MOX and unlabeled BSTFA, and analyzed in both EI and PCI modes (Fig. 1); 3) the 95% ^{13}C metabolite extract derivatized with MOX and unlabeled BSTFA, and analyzed in both EI and PCI modes; 4) the 5% ^{13}C metabolite extract derivatized with MOX and BSTFA_d9 and analyzed in the PCI mode (Fig.

1); 5). The 5% ^{13}C metabolite) extract derivatized with EOX and unlabeled BSTFA and analyzed in the PCI mode.

Data analysis

Raw data files (.raw) generated from PCI and EI were analyzed in Refiner MS (Genedata Expressionist, Basel, Switzerland) software, separately. A threshold of 50 counts/scan was set to be the minimum intensity as signal. Peaks eluting within the same retention time window, signifying a metabolite signature (cluster) were outputted from Refiner MS. IROA Peak pairs were selected based on the characteristics of mirrored isotopologue pairs.

Metabolites annotation for EI data was performed with 5% ^{13}C sample with unlabeled BSTFA derivatization, by comparing the mass spectrum and retention time seen to the Fiehn library, and NIST libraries using Refiner MS (Genedata Expressionist, Basel, Switzerland) software. Matching between PCI, EI and deuterated derivatives were performed in MassLynx software (V4.1, Waters). Elemental composition calculation was also performed in MassLynx software. NIST MS search v2.2 software was used to obtain potential identities by matching the EI spectrum (Ringo, NY, USA).

Unlike EI, positive CI is a softer ionization method, which produces fewer fragments and much higher intensity of high mass ions. The CI gas used was 5% ammonium (NH_3 -CI) together with 95% methane. Our results indicate that under our current positive CI (PCI) conditions, most of the metabolites observed have $[\text{MH}]^+$ ions (as shown in Table S1). Many metabolites generated detectable $[\text{M}-\text{CH}_3]^+$ peaks in the PCI spectrum due to the loss of CH from one silylation group (Figure S2). The $[\text{MH}]^+$ peaks and $[\text{M}-\text{CH}_3]^+$ peaks were confirmed in the both ^{12}C and ^{13}C mirrored isotopomers of the typical IROA sample. Despite there only being 5% ammonium, $(\text{MH})^+ + \text{NH}_3$ peak pair are prevalent among the CI spectra found (Figure S2), in accordance with the findings of Warren²⁵. Our PCI results similarly showed that NH_3 -CI mass spectra were generally dominated by $[\text{MH}]^+$ and/or the adduct $[\text{MH} + \text{NH}_4]^+^{25}$ (Figure S2).

Results

^{12}C IROA and ^{13}C IROA GC/MS mirror pair signatures co-elute

S. cerevisiae cells were cultured in YNB media (IROAtech Inc., Methods) containing either 95% ^{13}C glucose, or 5% ^{13}C labeled glucose (see Fig. 1). Accurate mass GC-TOF/MS was done on extracts of each condition and on a 1:1 mixture of the ^{12}C and ^{13}C extracts. Fig. 1 shows the IROA mirror pairs for a 6 carbon molecular ion (isoleucine) assessed from the 1:1 mixture (left panel). A shift of 18 amu, indicating 2 TMS derivatization groups on this molecular ion, was seen when the 5% ^{13}C *S. cerevisiae* extract alone was derivatized with d₉-deuterated BSTFA (unlabeled- SiC_3H_9 vs. labeled- SiC_3D_9). The total ion chromatogram (TIC) of the unlabeled BSTFA derivatized samples under PCI show that the metabolites from the 95% ^{13}C extracts and 5% ^{13}C extracts co-elute, with no peak shape change (Fig. 2). On the other hand, a significant alteration of the chromatography is seen when the 5% ^{13}C extract is derivatized with deuterated d₉-BSTFA (Fig. 2).

Accurate mass IROA GC/MS generates “clean” EI spectra, devoid of artifacts and noise

Figure 3 illustrates the uniqueness of accurate mass IROA GC/MS for characterizing of EI fragments. Only metabolites that are generated in metabolic pathways can form IROA mirror pairs, thus enabling the removal of non-relevant peaks from EI library spectra. Furthermore, since for each IROA pair the number of carbons in the EI fragment is determined, this enables the determination of the exact formula of the fragments, which may support structural elucidation of unknowns. The EI spectra from the 5% ^{13}C *S. cerevisiae* extracts can be used to search standard EI libraries such as NIST, and is illustrated in Figure 3 for malate. Comparison of the spectra from the 5% ^{13}C *S. cerevisiae* extracts (Fig. 3A, green) with the combined 95% ^{13}C and 5% ^{13}C *S. cerevisiae* extracts (Fig. 3A, red) distinguish the 5 fragmentations expected from EI fragmentation of malate (Fig. 3B) from artifacts. For example, the EI m/z 117 fragment should contain fragments containing 2 carbons of the original malate molecule (Fig. 3B). The EI IROA shows a mirror pair of the 117 and 119 m/z fragments, confirming 2 carbons are in this fragment. The EI 233 m/z fragment has 3 carbons and the EI IROA shows a mirror pair of the 233 and 236 m/z fragments. The m/z 245, 319 and 335 fragments have 4 carbons each (Fig. 3B), and all show IROA mirror pairs confirming the presence of 4 carbons in each of fragment (m/z 245:249, m/z 319:323 and m/z 335:339).

Use of EI and PCI IROA for metabolite identification, and discovery

Fig. 4 shows the algorithm we have developed, for formulae elucidation. For metabolites not in any database, the first steps for identification of unknowns using the algorithm is to find the number of carbons in the metabolite, the ^{12}C monoisotopic mass (M_0), and the number of derivatization groups. Normally, determining possibilities for elemental composition analysis would be the first step, within the mass accuracy for a given molecular ion. However, even with high mass accuracy with less than 1ppm, there are frequently too many possibilities without the use of PCI IROA to first find the number of carbons in the metabolite¹³.

The number of carbons (n) in an unknown metabolite can be calculated from the m/z difference of the PCI IROA mirror pairs. Consider mass M_0 as the mass of the ^{12}C monoisotopic derivatized peak. The ^{13}C monoisotopic derivatized peak has a mass of M_0 plus the mass of the additional neutrons, one per carbon in the original metabolite. Since both monoisotopic ions are equally derivatized, the derivatization reagent does not enter into the calculation for n . The number of carbons (n) from the biological part of the fragment is always the mass difference between the two monoisotopic peaks divided by the mass of a neutron (1.00335).

The derivatization adds informational complexity in a way unique from (underivatized) LC/MS IROA. Consider the effect of the BSTFA derivatization: it adds a mass equal to an unknown multiple of the silylation groups to both the ^{12}C monoisotopic mass (M_0) and ^{13}C monoisotopic mass (M_0+n). Every silylation group consists of one Si, nine hydrogens and three carbons ($-\text{SiC}_3\text{H}_9$). The natural abundance of the isotopes of hydrogen, silicon and carbon are 0.015% for ^2H , 4.67% for ^{29}Si , and 1.1% for ^{13}C . For the ^{13}C monoisotopic mass

(M0+n) the only appreciable signals above it (M0+n+1, etc.) are due to the isotopic contribution of the derivatization groups.

Each silylation group ($-\text{SiC}_3\text{H}_9$) would generate about 8.54% of (M0+n+1)/(M0+n), assuming one substitution of a higher mole fraction for each element in the silylation group. This is calculated as the sum of the probabilities for one labeled substitution in the silylated group for each element:

$$\text{Percentage of (M0+n+1)/(M0+n) \% generated by each silylation group} = 3 \times 1.1\% \times (0.989)^2 / (0.989)^3 + 4.67\% / 0.9223 + 9 \times 0.015\% \times (0.9985)^8 / (0.9985)^9 = 8.54\%.$$

Similarly, a doubly derivatized molecule would show a relative height for the M0+n+1 of ~17%, a triply derivatized molecule will have a ~25% relative height for its M0+n+1, etc. With the percentages so widely spaced it is easy to determine the number of derivatization groups that are carried in any given molecule.

In order to assure that this was accurate and to confirm the number of silylation groups, deuterated BSTFA_d9 derivatization was used for the 5% ^{13}C *S. cerevisiae* extracts. Note that the deuterated_d9 BSTFA has a time shift relative to unlabeled BSTFA, as shown in Fig. 2, requiring inspection of peaks within that time shift for the match deuterated BSTFA_d9 spectra, examining for peak pairs of $[\text{MH}]^+$ and $(\text{M}-\text{CD}_3)^+$ (see Fig. 1, 18 amu for 2TMS_d9 derivatizations). The use of the deuterated reagent is not required, but can be done for confirmation of the estimate from the (M0+n+1)/(M0+n) ratio.

In addition to silylation groups, ketone groups that are protected by the methoximation reaction can be detected, by performing an ethoximation of the same sample in parallel. Compared with methoximation, each ethoximation reaction adds 14 amu (equivalent to one CH_2 group) to the protected metabolites. From the differences in the molecular weights between ethoximation and methoximation, the number of ketone groups can be calculated. Each methoximation reaction would add 1 nitrogen, 1 carbon, and 3 hydrogens to the derivative. Due to the additional CH_2 groups, the ethoximation derivatives are eluted later than their methoximation counterparts. A number of 17 metabolites were detected to be protected by methoximation/ethoximation reactions (Table S2).

Knowing the number of silylation groups, the number of methoximation groups, and the carbon chain length, the possibilities for the elemental composition are dramatically reduced. After subtraction of the mass of the derivatized groups the underivatized monoisotopic mass can be determined (see Table S3). In the event that there is more than one formula, as shown by Kind and Fiehn²⁶ the search can be further narrowed using the nitrogen rule (odd nominal mass implies odd number of nitrogens), the hydrogen/carbon element ratio rule (most typical ratios found between $2 > \text{H/C} > 0.5$), as well as examining other heteroatom ratio distributions (N/C, O/C, P/C and S/C).

As part of the algorithm detailed in Fig. 4, the formula(e) containing the derivatized groups were searched against the NIST library to find possible structures on the basis of their EI spectral patterns. If no possible hits show up for EI spectra in the NIST library, formulae derived from CI spectra, after trimethylsilyl and methoximation groups are subtracted out,

were used to search chemical libraries such as Chempider (<http://www.chemspider.com/>) and PubChem Compound (<http://www.ncbi.nlm.nih.gov/pccompound/>).

Fig. 5 shows how to use the spectral information contained within PCI IROA and EI IROA peaks, in combination, to identify new metabolites, as well as to eliminate artifacts (chemicals of non-biological origin). Figure 5 and Figure S1 shows that on the *S. cerevisiae* IROA experiment, 3 compounds were seen to co-elute closely around one retention time on PCI ($^+HM1$ m/z 320:325, $^+HM2$ m/z 365:370, and $^+HM3$ 393:400). IROA pairs at m/z 377:384 and m/z 349:354 are due to the loss of one carbon unit ($C-H_3$) H^+ from a silylation group on $^+HM3$, and $^+HM2$, respectively. However, the spectra around m/z 444.1859 and 428.1496 do not show IROA mirror peak patterns, which flags them as artifacts. Further elucidation of the nature of the 3 compounds was obtained with EI (see Supplemental Fig. 1). $^+HM1$ m/z 320:325 was validated with EI to be α -ketoglutarate, $^+HM2$ m/z 365:370 was validated as 2-hydroxyglutaric acid. α -ketoglutaric acid has been seen to co-elute with 2-hydroxyglutaric acid in this oven program, and this agreed with the size of the molecular ion seen on CI.

The presence of the unknown 7 carbon metabolite ($^+HM3$) was confirmed on both EI and PCI, with m/z 393.2 on PCI. In Fig. 6 this 7 carbon unknown metabolite underwent analysis by the algorithm depicted in Fig. 4 using accurate mass PCI and EI IROA spectra. The number of silylation groups estimated from the $(M_0+n+1)/(M_0+n)$ ratio with PCI IROA was confirmed by the use of deuterated d9_BSTFA (Fig. 6A), narrowing the choices for the formulae for identification. The PCI mass shift seen for the 5% ^{13}C *S. cerevisiae* extracts derivatized with unlabeled vs deuterated d9_BSTFA, showed a mass shift of 27 daltons, from m/z 393 to 420, which indicates there are 3 trimethylsilyl (TMS) groups in this compound (each of the 3 methyl groups on the silyl group has a mass shift of 9 daltons, Fig. 6A). The elemental composition of the derivatized compound suggested the formula as $C_{16}H_{37}O_5Si_3$ (Fig. 6B). This formula ($C_{16}H_{36}O_5Si_3$) was then used in a NIST search to search for possible hits. There are 3 compounds (Fig. 6C) that were found, and only 2-isopropylmalic acid matched the EI-IROA spectrum with the masses of 275, 349, and 377 (Fig. 6D). To confirm the identification of these 3 co-eluting metabolites, authentic standards were analyzed in both PCI and EI modes (Fig. S1)

In our study, we were able to identify 126 metabolites of biological origin with IROA mirror peak pairs of their molecular ions $(MH)^+$ in the PCI data (see Fig. S2). Most of the molecular ions come along with $(M-CH_3)^+$ ion peaks, and some molecular ions have $(M+NH_4)^+$ ion peaks. A total of 82 metabolites were annotated using the procedure described above with EI NIST and Fiehn libraries and the PCI data (see Table S1 and Table S2). The same running conditions (oven program and column) as used for the Fiehn GC/MS library²⁷ were applied in our study, allowing us to use retention time matching for confirmation of metabolite identities. Among these 82 annotated metabolites, 56 of them were matched with both Fiehn EI spectra and their Fiehn library retention times (Table S1). 4 metabolites were confirmed with authentic standards (Table S1, 2 metabolite standards overlapped with 2 found from a Fiehn library retention time match). Another 24 could be annotated with a potential identity (Table S3) using our protocol (Fig. 4), with either an EI NIST match and/or a Chempider formulae match.

In addition to compound identification, ^{13}C monoisotopic peaks were used to normalize the ^{12}C monoisotopic peaks. As shown in Table S4, ^{13}C monoisotopic peaks normalization significantly lowered the coefficient of variation (CV) for those 126 detected metabolites (chemical ionization mode). Over 50% of metabolites were detected with a CV lower than 10%, with 5 replicates. However, only 13.5% of metabolites have a CV lower than 10% in the raw (unnormalized) ^{12}C monoisotopic peaks, and 26.2% of metabolites have a CV lower than 10% with normalization to the chlorophenylalanine internal standard.

Discussion

IROA utilizes ^{13}C labeling to generate mirror pairs of isotopologue spectra for all metabolites of biological origin (Fig. 1). This is the first publication of IROA methodology used in conjunction with accurate mass GC/MS data, and the first use of IROA to annotate the *S. Cerevisiae* metabolome. Accurate mass PCI spectra reveal the number of carbons (n) in the original biological molecules and the examination of the $(\text{M}0+\text{n}+1)/(\text{M}0+\text{n})$ ratio allows one to determine the number of derivatization groups attached, narrowing the possible choices for identification, a complementary approach to that used by LC/MS IROA¹⁷. Accurate mass IROA EI GC/MS was shown to generate/identify “clean” EI spectra, devoid of artifacts and noise.

Our method resolves issues which have held back the power inherent in using accurate mass GC/MS for chemical identification. A recent study by the Suizdak group²⁸, suggested possible drawbacks to GC/MS profiling including thermal degradation of metabolites, and differences in the derivatization efficiency between metabolites. They also noted moderate levels of identification (60% of all plasma GC/MS peaks), a common problem in global untargeted GC/MS metabolomic studies, indicating the need for a platform which could expand identifiability of GC mass spectrometric peaks. IROA experiments, by their nature, contain a ^{13}C standard for every metabolite detected, which obviates any differences metabolites might have with respect to derivatization efficiency or thermal degradation.

Co-eluting chemicals during a GC/MS run have EI spectral patterns that may not allow for easy identification of the co-eluting components, especially if some of the co-elutes are of low abundance. Processing approaches have focused on discerning co-eluting metabolites based on slight retention difference or their peak shapes²⁹⁻³¹. Here, we use both the GC/MS PCI spectra (reflecting the molecular ion) and the GC/MS EI spectra (reflecting MS/MS fragment structure). These IROA spectral signatures can be found in both PCI and EI modes and have complementary information. EI IROA can help characterize the structure/identity of mass fragments by determining the number of carbons they contain, and reveals which parts of the EI databases (NIST, Wiley, and Golm etc.) have derivatization or chemical artifacts. PCI IROA spectra are relatively sparse, enabling the determination of distinct co-eluting chemicals, separate from artifacts, by the presence of the CI IROA isotopologue spectra, as well as providing the chain length of the carbon backbone for the molecular ion IROA mirror pairs. The combination of accurate mass EI and PCI effectively identifies a large number of metabolites in the *S. cerevisiae* metabolome, and Fig. S2 lists the 126 PCI IROA isotopologue spectra found, including their NH_4 adducts $(\text{M}+\text{NH}_4)^+$ and the loss of methyl groups $(\text{M}-\text{CH}_3)^+$. The PCI IROA spectral patterns in Fig. S2, along with their

retention times, suggest that small molecules can be effectively identified from CI IROA databases.

Coon's laboratory¹² has developed a method, termed molecular ion directed acquisition (MIDA, not to be confused with Mass Isotopomer Distribution Analysis³²³³), for use only on a GC-Orbitrap MS, to maximize the information content generated from unsupervised tandem MS (MS/MS) and selected ion monitoring (SIM) by directing the MS to target the ions of greatest information content, that is, the most-intact ionic species. The spectral processing algorithm employed by MIDA directs the instrument to these ions by exploiting the expected adducts that form during the methane PCI process, as well as the characteristic EI fragmentation patterns of commonly employed derivatization reagents. A rather complex MIDA algorithm and workflow utilizes this information along with ¹³C- and ¹⁵N-metabolic labeling, multiple derivatization and ionization types, and heuristic filtering of candidate elemental compositions to achieve identifications.

In contrast, IROA is a platform independent, accurate mass spectrometric method can be used for quantitation, identification and elucidation of unknown chemical identities of biological origin. The m/z distance, using accurate mass GC/MS, between the uniformly labeled ¹²C spectra and the uniformly labeled ¹³C spectra define the number of the carbons in the molecule. This simplifies the algorithm for identification and quantification, and further can be used with any accurate mass GC/MS machine, not just a GC-Orbitrap MS. Quantification with an accurate mass GC/MS IROA experiment can be obtained from the comparison of the 95% ¹³C IROA labeled isotopologues (control group) and the 5% ¹³C IROA labeled isotopologues (experimental group), as has been shown for LC/MS IROA¹⁷. Tobias and Oliver defined seven golden rules to select the most likely and chemically correct molecular formulas.²⁶ These rules, along with IROA determination of carbon number for the molecular ion, and estimation of the number of silylation groups (x) from the (M0+n+1)/(M0+n) CI intensity ratio, were very helpful in reducing the possibilities generated by Mass Lynx for elemental composition in our GC-TOF/MS accurate mass assessments (Fig. 6). The ability of the (M0+n+1)/(M0+n) CI intensity ratio to determine the number of derivatization groups was validated by use of d9_BSTFA.

As the same oven program and column as used in the Fiehn GC/MS library was applied in our study²⁷, the annotated metabolites (Table S1) were compared with the metabolite retention times from the Fiehn library, and were found to be within a 0.2-0.5 min difference. Metabolites labeled with d9_BSTFA, used for validation of the number of silylation groups, show retention time differences are known to be dependent on both the number of deuterium atoms and the structural position, as described by Huang and Regnier³⁴. The differences in resolution between deuterated and non-deuterated derivatives of metabolites cause differences for the metabolites alignment. However, based on the estimated number of silylation groups (x) from the relative intensity of (M0+n+1)/(M0+n), mass shifts based on M0+n+9x was used to locate the deuterated derivatives for d9_BSTFA. In addition, the deuterated molecular ion can be confirmed from the loss of one (deuterated) C-D₃ group from Si-(C-D₃)₃, using the (MH)⁺-19 peak.

Conclusions

This study is the first use of IROA technology, with accurate mass GC-TOF/MS based metabolomics, to examine the *S. cerevisiae* metabolome. Our IROA study uses randomized 95% ^{13}C and 5% ^{13}C glucose as a substrate for labeling the *S. cerevisiae* metabolome. The IROA method for untargeted metabolomic profiling simply depends on finding a combination of nutrients that can be labeled with randomized 5% or 95% ^{13}C for the model organism, or tissue culture model, and randomized ^{13}C media is available for bacterial and mammalian cells. IROA may become a tool for the study of evolutionary metabolomics, allowing the comparison of whole metabolomes of different taxa with one method. The creation of paired IROA peaks, which set this study apart from all other stable isotope labeling studies used for yeast metabolome characterization, yielded many benefits, including the ability to determine the number of carbon atoms in the original molecule, the number of derivatization groups and the ability to discriminate between peaks of biological origin and artifacts. EI IROA can help characterize the structure/identity of mass fragments by determining their carbon chain length. The EI IROA peak pattern easily identifies artifacts in known GC/MS spectral libraries, which has the potential to generate “clean” spectral libraries not dominated by silylation and other artifacts, for increased metabolite identification. Determination of the number of carbons for a metabolite with PCI IROA helps to flag metabolites of unknown identity to further investigate. The combination of EI and CI IROA also helps identify co-eluting metabolites, another bane of EI spectral library matching. Our catalogue of IROA CI paired peaks is a tentative step for the formation of CI spectral libraries, and the wider adoption of CI GC/MS spectra for use in metabolomics, and LC/MS IROA molecular ion patterns have the same potential. IROA allows the use of different accurate mass platforms (LC/MS and GC/MS) to catalogue and quantify the diversity of metabolites seen, potentially allowing the integration of information from different platforms in a common framework.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank for the financial support by the Einstein-Mount Sinai Diabetes Research Center grant P60DK020541, SECIM RCMRC Pilot and Feasibility grant U24 DK097209-01A1 and the Einstein-Montefiore CMCR U19 AI091175.

References

1. Fiehn O, Kopka J, Trethewey RN, Willmitzer L. *Anal Chem.* 2000; 72:3573–3580. [PubMed: 10952545]
2. Peterson AC, Hauschild JP, Quarmby ST, Krumwiede D, Lange O, Lemke RA, Grosse-Coosmann F, Horning S, Donohue TJ, Westphall MS, Coon JJ, Griep-Raming J. *Anal Chem.* 2014; 86:10036–10043. [PubMed: 25208235]
3. Kurland IJ, Accili D, Burant C, Fischer SM, Kahn BB, Newgard CB, Ramagiri S, Ronnett GV, Ryals JA, Sanders M, Shambaugh J, Shockcor J, Gross SS. *Ann N Y Acad Sci.* 2013; 1287:1–16. [PubMed: 23659636]

4. Qiu Y, Cai G, Zhou B, Li D, Zhao A, Xie G, Li H, Cai S, Xie D, Huang C, Ge W, Zhou Z, Xu LX, Jia W, Zheng S, Yen Y. *Clin Cancer Res*. 2014; 20:2136–2146. [PubMed: 24526730]
5. Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L. *Nat Biotechnol*. 2000; 18:1157–1161. [PubMed: 11062433]
6. Qiu Y, Cai G, Su M, Chen T, Liu Y, Xu Y, Ni Y, Zhao A, Cai S, Xu LX, Xu LX, Jia W. *J Proteome Res*. 2010; 9:1627–1634. [PubMed: 20121166]
7. Qiu Y, Cai G, Su M, Chen T, Zheng X, Xu Y, Ni Y, Zhao A, Xu LX, Cai S, Jia W. *J Proteome Res*. 2009; 8:4844–4850. [PubMed: 19678709]
8. Hummel J, Strehmel N, Selbig J, Walther D, Kopka J. *Metabolomics*. 2010; 6:322–333. [PubMed: 20526350]
9. Kopka J. *J Biotechnol*. 2006; 124:312–322. [PubMed: 16434119]
10. Abate S, Ahn YG, Kind T, Cataldi TR, Fiehn O. *Rapid Commun Mass Spectrom*. 2010; 24:1172–1180. [PubMed: 20301109]
11. Kumari S, Stevens D, Kind T, Denkert C, Fiehn O. *Anal Chem*. 2011; 83:5895–5902. [PubMed: 21678983]
12. Peterson AC, Balloon AJ, Westphall MS, Coon J. *J Anal Chem*. 2014; 86:10044–10051.
13. Kind T, Fiehn O. *BMC Bioinformatics*. 2006; 7:234. [PubMed: 16646969]
14. Kwiecien NW, Bailey DJ, Rush MJ, Cole JS, Ulbrich A, Hebert AS, Westphall MS, Coon J. *J Anal Chem*. 2015; 87:8328–8335.
15. de Jong FA, Beecher C. *Bioanalysis*. 2012; 4:2303–2314. [PubMed: 23046270]
16. Clendinen CS, Stupp GS, Ajredini R, Lee-McMullen B, Beecher C, Edison AS. *Front Plant Sci*. 2015; 6:611. [PubMed: 26379677]
17. Stupp GS, Clendinen CS, Ajredini R, Szewc MA, Garrett T, Menger RF, Yost RA, Beecher C, Edison AS. *Anal Chem*. 2013; 85:11858–11865. [PubMed: 24274725]
18. Wu L, Mashego MR, van Dam JC, Proell AM, Vinke JL, Ras C, van Winden WA, van Gulik WM, Heijnen J. *J Anal Biochem*. 2005; 336:164–171.
19. Bennett BD, Yuan J, Kimball EH, Rabinowitz JD. *Nat Protoc*. 2008; 3:1299–1311. [PubMed: 18714298]
20. Weindl D, Wegner A, Jager C, Hiller K. *J Chromatogr A*. 2015; 1389:112–119. [PubMed: 25748542]
21. Mashego MR, Wu L, Van Dam JC, Ras C, Vinke JL, Van Winden WA, Van Gulik WM, Heijnen JJ. *Biotechnol Bioeng*. 2004; 85:620–628. [PubMed: 14966803]
22. Blank LM, Desphande RR, Schmid A, Hayen H. *Anal Bioanal Chem*. 2012; 403:2291–2305. [PubMed: 22543713]
23. Edison AS, Clendinen CS, Ajredini R, Beecher C, Ponce FV, Stupp GS. *Integr Comp Biol*. 2015; 55:478–485. [PubMed: 26141866]
24. Crutchfield CA, Lu W, Melamud E, Rabinowitz JD. *Methods Enzymol*. 2010; 470:393–426. [PubMed: 20946819]
25. Warren CR. *METABOLOMICS*. 2013; 9:S110–S120.
26. Kind T, Fiehn O. *BMC Bioinformatics*. 2007; 8:105. [PubMed: 17389044]
27. Kind T, Wohlgemuth G, Lee do Y, Lu Y, Palazoglu M, Shahbaz S, Fiehn O. *Anal Chem*. 2009; 81:10038–10048. [PubMed: 19928838]
28. Fang M, Ivanisevic J, Benton HP, Johnson CH, Patti GJ, Hoang LT, Uritboonthai W, Kurczy ME, Siuzdak G. *Analytical chemistry*. 2015
29. Likic VA. *BioData Min*. 2009; 2:6. [PubMed: 19818154]
30. Seifi H, Masoum S, Seifi S. *J Chromatogr A*. 2014; 1365:173–182. [PubMed: 25219524]
31. Ni Y, Qiu Y, Jiang W, Suttlemyre K, Su M, Zhang W, Jia W, Du X. *Anal Chem*. 2012; 84:6619–6629. [PubMed: 22747237]
32. Hellerstein MK, Neese RA. *Am J Physiol*. 1992; 263:E988–1001. [PubMed: 1443132]
33. Kelleher JK, MM T. *American Journal of Physiology*. 1992; 262:E118–E125. [PubMed: 1733242]
34. Huang X, Regnier FE. *Anal Chem*. 2008; 80:107–114. [PubMed: 18052339]

Fig. 1.

Flow chart illustrating the use of IROA and GC/MS for global discrimination of metabolites that arise from *S. cerevisiae* metabolic networks. The *S. cerevisiae* cells were grown in YNB media containing either 95% ^{13}C /5% ^{12}C labeled glucose or 95% ^{12}C /5% ^{13}C glucose for 10+ generations to fully label the yeast cells. At an OD of 2.5, the cells were harvested, extracted and the 95% ^{13}C and 5% ^{13}C extracts equally combined, and then derivatized with unlabeled BSTFA or deuterated (d_9) BSTFA. An example of GC/MS Positive Chemical ionization (PCI) spectra of metabolites for the combined extracts show mirrored peak pairs for the molecular ion (isoleucine, left panel), which will not be seen for chemical contaminants/artifacts because they did not arise from a metabolic pathway. The m/z distance between the uniformly labeled ^{12}C peak (M_0), and the uniformly labeled ^{13}C peak (M_0+n) is the number of carbons in the metabolite backbone (n). Stable isotope derivatization reagents can be used with IROA to acquire chemical information for further compound formula elucidation. The right panel shows a shift of 18 amu, indicating 2 TMS derivatization groups, when the 5% ^{13}C *S. cerevisiae* extract was derivatized with d_9 deuterated BSTFA.

Fig. 2.

Labeled and unlabeled metabolites generated in the IROA experiment co-elute. Extracts of the 95% ^{13}C /5% ^{12}C glucose and 95% ^{12}C /5% ^{13}C labeled glucose (IROA) experiments (Fig. 1) were combined in a 1:1 ratio, and accurate mass GC-TOF MS was done. Total ion chromatogram (TIC) of the BSTFA derivatized samples under Chemical Ionization GC-TOF/MS show that the ^{13}C monoisotopic peak co-eluted with ^{12}C monoisotopic peak (purple and green). Deuterated d9 BSTFA derivatized samples had a time shift with respect to the chromatography of non-deuterated BSTFA derivatives (red and green). The chromatogram below is enlarged between the retention times of 15 to 19 min.

Fig. 3.

Characterization of the metabolic components of the EI spectra for 95% ^{13}C and 5% ^{13}C extracts derivatized with unlabeled BSTFA, using the IROA spectra of malate detected as an example. Malate was present at retention time 12.48 min (Table S1). IROA EI spectra confirm that the EI mirrored isotopologue fragments are of metabolic origin, devoid of derivatization artifacts and chemical noise.

A. Comparison of 95% ^{12}C experiment (5% ^{13}C) mass spectrum (green) with the IROA combined extracts from *S. cerevisiae* incubated with 95% ^{12}C and 95% ^{13}C reagents (red). The enlarged areas show the IROA pairs that emphasize the “clean” parts of the EI spectrum, which reflect the specific fragmentations described in B.

B. All of the IROA pairs shown in Fig. 3A have the specific fragmentations illustrated for the BSTFA derivatization of malate. These 5 fragmentations correspond to the 5 IROA isotopologue mirrored spectra seen at 117:119 (2 carbon fragment), 233:236 (3 carbon fragment), 245:249 (4 carbon fragment), 319:323 (4 carbon fragment) and 335:339 (4 carbon fragment).

Fig. 4.

Flowchart illustrating the process for chemical formula elucidation and unknown metabolite identification using accurate mass IROA GC/MS, with a Waters Premier GC-TOF MS and MassLynx using BSTFA derivatization as an example. After determining the number of carbons in the metabolite (n), the number of silylation groups (x) can be estimated from the percentage $(M_0+n+1)/(M_0+n)$, with each silylation group contributing 8.54%. The number of silylation groups (x) and the number of methoximation groups (y), can be confirmed using deuterated silylation reagents for x, and comparison of the CH₂ shift seen with ethoximation vs methoximation. The total number carbons in the metabolite silylation/methoximation derivatives can be calculated as $n+3x+y$ (each silylation group from BSTFA contributes 3 carbons, methoximation adds one carbon). The selected formula can be used for structure interpretation with NIST software or search against websites. The potential hits can be confirmed with authentic standards.

Fig. 5.

Accurate mass IROA GC-TOF/MS identifies co-eluting compounds of known and unknown identities using a combination of Electron Ionization (EI) and Positive Chemical Ionization (PCI) modes. Using the *S. cerevisiae* IROA spectra from the experiment shown in Fig. 1, 3 compounds were seen to co-elute closely around one retention time on PCI ($^+HM1$ m/z 320:325, $^+HM2$ m/z 365:370, and $^+HM3$ 393:400). IROA pairs at m/z 377:384 and m/z 349:354 is the loss of one carbon unit ($C-H_3$) H^+ from a silylation group on $^+HM3$, and $^+HM2$, respectively. Further elucidation of the nature of the 3 compounds was obtained with EI (see Supplemental Fig. 1). $^+HM1$ m/z 320:325 was validated with EI to be α -ketoglutarate, $^+HM2$ m/z 365:370 was validated as 2-hydroxyglutaric acid. The presence of the unknown 7-carbon metabolite was confirmed on both EI and PCI ($^+HM3$), with m/z 393.2 on PCI (see Fig. 6 for identification process).

Fig. 6.

IROA Unknown Metabolite Identification using accurate mass PCI and EI modes on a Waters GC-TOF/MS.

A. Using PCI for the 5% ^{13}C extract derivatized with d9_BSTFA, m/z 420.3623 was detected, with IROA isotopomers decreasing at m/z 421.365 etc. This indicates a shift of 27 amu from 393.2034, seen from the extract of the combined 95% ^{13}C and 5% ^{13}C experiments derivatized with unlabeled BSTFA. This indicates there are 3 trimethylsilyl (TMS) groups in the $^+\text{HM3}$ compound (each silyl group has a 9 daltons shift due to 9 deuterium on 3 methyl groups).

B. Using the algorithm detailed in Fig. 4, the elemental composition function in the MassLynx software indicates the compound formula as $\text{C}_{16}\text{H}_{37}\text{O}_5\text{Si}_3$. There was no ketone group that could be methoximated, as no nitrogen was detected in the compound formula. After removal of 3TMS group, the compound should have a chemical formula of $\text{C}_7\text{H}_{12}\text{O}_5$. C. Compounds fitting this formula were found using a NIST search. Three compounds were selected based on $\text{C}_7\text{H}_{12}\text{O}_5$, and 2-isopropylmalic acid had the correct EI NIST spectra (which is an intermediate for leucine synthesis).

D. EI IROA spectral pairs from the extract of the combined 95% ^{13}C and 5% ^{13}C experiments confirm the identity of 2-isopropylmalic acid for the proposed C_7 metabolites, from EI peaks at m/z 275, 349 and 377. The spectral peak for 2-isopropylmalic acid, shown at m/z 275, has an EI IROA pair showing it represents a 6 carbon fragment. The EI fragment at m/z 349 has a complex IROA envelope, due to the co-elution of 2-hydroxyglutaric acid and 2-isopropylmalic acid (see Fig. 6C and Fig. S1). The other peak pair shown at 377.1680: 384.1948 is also likely a fragment of 2-isopropylmalic acid (see Fig. 6C). The confirmation of the identity of 2-isopropylmalic acid indicates the power of combining CI IROA (Fig 5) and EI IROA for identifications.