Supplementary Information for:

Isotopic Ratio Outlier Analysis Global Metabolomics of *Caenorhabditis elegans*

Gregory S. Stupp[1‡], Chaevien S. Clendinen[1‡], Ramadan Ajredini[1‡],
Mark A. Szewc[2], Timothy Garrett[3], Robert F. Menger[4], Richard A. Yost[4], Chris Beecher[5]
& Arthur S. Edison[1*]

1) Department of Biochemistry & Molecular Biology, University of Florida, Gainesville, FL; 2) Thermo Fisher Scientific, Somerset, NJ; 3) Department of Pathology, Immunology, and Laboratory Medicine; 4) Department of Chemistry, UF; 5) IROA Technologies, Ann Arbor, MI

‡These authors contributed equally to this study.

*Corresponding author: Arthur S. Edison: aedison@ufl.edu

Table of Contents

Supplementary Methods

Isotopically Labeled Bacteria Growth

500 µl of an overnight culture of *E. coli* MG1655 growing in LB media was inoculated into 500 ml of M9 Minimal Media (M9MM) supplemented with 2 g/L of isotopically labeled glucose (either 95% or 5% ^{13}C) in a 2 L flask and allowed to shake at 37 °C for 16 hours. The isotopic distribution of ^{13}C in the glucose was randomly generated by algal growth on either 95% or 5% ^{13}C-enriched CO_2 (Cambridge Isotope Laboratories). The resulting *E. coli* bacteria was then pelleted in 500 ml bottles at 6,000 rpm for 15 min and resuspended in an equal weight of S-complete[1] to make a 50% MG1655 stock and was stored at -20 °C.

M9 Minimal Media

M9MM was prepared combining 12.8 g Na_2HPO_4, 3.0 g KH_2PO_4, 0.5 g of NaCl, and 1.0 g of NH_4Cl in a 1L Erlenmeyer flask. The solution was brought to 950 mL using doubly distilled water (ddH_2O), and the pH was adjusted to 7.4 with NaOH. The solution was autoclaved on slow exhaust and stored at 4°C. The following compounds were dissolved in H_2O and added to the autoclaved media using a 10 mL sterile syringe and a 0.22 micron filter: 2.0 g of glucose (unlabeled, 95%, or 5% labeled), 490 mg of $MgSO_4.7H_2O$, 15 mg $CaCl_2 \bullet 2H_2O$, 10 mg of (unlabeled) thiamine, and 10 mg of $FeSO_4 \bullet 7H_20$. After filtering the solution into the media, the test tube was rinsed and filtered with ddH_2O into the media also and brought up to 1 L. Prior to inoculation, 1 mL of 10 mg/mL of (unlabeled) nystatin was added.

S-Complete Buffer

S-complete buffer is prepared by dissolving 5.85 g of NaCl, 6 g of KH_2PO_4, and 1 g K_2HPO_4 in 1 L of ddH_2O. The solution was autoclaved and allowed to cool to 50°C. The following sterile components were added: 1 ml of 5mg/ml of (unlabeled) cholesterol, 10 ml of 1 M (unlabeled) potassium citrate (pH 6.0), 10 ml of trace metals,[1] 3 ml of $CaCl_2$, and 3 ml of $MgSO_4$.

Isotopically Labeled Worm Growth

350,000 unlabeled larval stage 1 worms (L1s) were grown in 35 mL of S-complete with 2.1 mL (3% by volume) of 95% or 5% ^{13}C labeled *E. coli* MG1655 and 35 µL of 5 mg/mL cholesterol in ethanol. The standard maturation process was allowed to progress from L1 to adults with eggs at a worm density of 10,000 worms per mL (wpm). The worms were synchronized in a bleach solution, and 1.8 million each of 1st generation 5% and 95% ^{13}C eggs were obtained. The eggs were hatched overnight in 30 mL of M9 buffer (M9b),[1] and then washed three times in M9b. The resulting larvae were then fed 7 mL (2% by volume) of 95% or 5% ^{13}C labeled E. coli MG1655 and grown in 170 mL of S-complete with 170 µL of 5 mg/mL cholesterol in ethanol from L1 to young adults (45 hours) at a worm density of 10,000 wpm. The worms were then washed twice in M9b and used for heat shock perturbations.

Heat Shock Perturbation

Worms were washed from bacteria by sucrose centrifugation with 30% sucrose and then washed twice with water. Their guts were cleared by shaking in M9b at 22 °C for 30 min and then washed three times in doubly distilled water. At this point, the 5% ^{13}C worms, used for the test experiments, were split into 4 equal batches of 225,000 worms each, while the 95% ^{13}C worms, used for the control experiments, were not split. Worms were incubated at 22 °C (control worms) or 33 °C (heat shock worms) for 30 min with shaking at 250 rpm and then transferred to a 22 °C incubator with shaking for 90 more min at a worm density of 18,000 wpm. Worms were then pelleted by centrifugation at 2,000 rpm for 2 min and the worms and supernatant were collected.

Combining and Preparing Samples

The control worm solutions were split evenly into four batches. The worms were pelleted by centrifugation at 2,000 rpm for 2 min and the supernatant was collected. 10 mL of each control supernatant and heat shock supernatant were mixed together and filtered through a 0.20 micron nitrocellulose filter. The worm pellets were mixed together (one control with one test) and resuspended in 4 mL of methanol (total volume 5 mL).

The filtered supernatants were frozen at -20 °C for 24 hours and then frozen at -80 °C for 8 hours. They were then lyophilized for two days. The residue was resuspended in 1 mL of ddH_2O in microcentrifuge tubes, separated by centrifugation, and re-lyophilized. The residue was resuspended in 100 µL of LC-MS grade water and analyzed by LC-MS, as described below.

Each pellet sample was split evenly into five 2.0 mL polypropylene microcentrifuge vials with approximately 500 µL of 0.15–0.21 mm diameter glass beads in each vial. The pellets were homogenized in a Mini-Beadbeater-8 for 3 cycles of 30 sec with 1 min on ice. The vials were centrifuged at high speed for 30 sec and the supernatant was removed and combined. The beads were then washed with 500 µL of methanol, combined into one tube for each sample, homogenized again for 1 min, and the supernatant added to the first set of homogenates. The samples were dried under N_2 gas, resuspended in 200 µL of methanol, transferred to microcentrifuge tubes, separated by centrifugation and re-dried under nitrogen. The residue was resuspended in 100 µL of LC-MS grade water and analyzed.

Data Processing and Importing into MATLAB

Thermo RAW files were converted into the open data format mzXML using MSConvert, which is included as part of the ProteoWizard package (http://proteowizard.sourceforge.net/index.shtml). The .mzXML files were then imported into MATLAB (R2013ab) for analysis. The raw peak data for each scan was converted to centroided peak lists using the MATLAB Bioinformatics Toolbox.

Finding and Identifying IROA Peaks

IROA peaks were identified using a custom MATLAB script that works by finding peaks separated by the carbon isotope spacing (1.0034 u). Furthermore, the relative intensities of the isotopomers follow a binomial distribution as determined by the number of carbons and percent incorporation of ^{13}C; thus any peak groupings that

exhibit the appropriate isotopic spacing and relative intensities were further analyzed. In detail, the algorithm works as follows.

Throughout the methods, the 95% ^{13}C sample will be referred to as the control and 5% ^{13}C will be referred to as the test. Although this convention will be utilized throughout the entirety of this work, the relative labeling can be altered such as for a reciprocal labeling experiment.

Synthetic IROA peaks for singly charged ions are generated to check if the intensities of the isotopic peaks within a measured IROA peak follow the binomial distribution. The synthetic distributions are generated by using the binomial probability function to calculate the predicted intensities of all isotope peaks within an IROA peak for a compound with certain ^{13}C abundances and number of carbons (Equation 1). The intensity for the *i*-th isotopic peak in an IROA peak is given by the formula

$$c_i = \binom{n}{i} {p_a}^i (1 - p_a)^{n-i} + \binom{n}{i} {p_b}^i (1 - p_b)^{n-i} \quad (1)$$

where $\binom{n}{i}$ is the binomial coefficient, p_a is the percent ^{13}C abundance for the control sample, p_b is the percent ^{13}C abundance for the test sample, and n is the total number of carbons. In this naming convention c_0 corresponds to the isotope peak in which all carbons in the isotope are ^{12}C and none are ^{13}C, and c_n corresponds to the isotope peak in which all carbons are ^{13}C.

Additionally, cutoffs for the minimum (n_{min}) and maximum (n_{max}) possible number of carbons for a molecule of mass *m* were calculated. Using the entire HMDB database, a convex hull was created using the actual number of carbons and masses for all compounds. The search range for a given mass was then confined by the boundaries of the convex hull.

Each scan is analyzed in turn starting at the peak of greatest detected *m/z*. For each iteration, the adjacent peak in the same scan is the starting point, until there are no more peaks, at which point the next scan is analyzed. The software searches for peaks separated from the c_n peak by the carbon isotope spacing (1.0034 u) within a mass accuracy of ±0.002 u. The search is confined by n_{min} and n_{max} to minimize spurious matches. The furthest peak found from the ^{13}C peak matching these criteria is termed the c_0 peak, and the number of carbons (*n*) is subsequently calculated from the difference between c_n and c_0. Furthermore, all isotope peaks between the c_n and c_0 peaks with the appropriate isotope spacing are included within the identified IROA peak.

The percent ^{13}C abundances (*p*) for the tentative IROA peak for each portion of the IROA peak (test and control isotopic peaks) are calculated using the following equation:

$$p = \frac{\binom{n}{i}}{\frac{c_i}{c_{i+1}}\binom{n}{i+1}+\binom{n}{i}} \quad (2)$$

for the first 2 (when $n < 7$) or 3 (when $n \geq 7$) isotopic peaks in each portion of the IROA peak. The percent ^{13}C abundance is calculated from the weighted average (by intensity) of the results for each isotopic peak.

If p is within constraints, a synthetic IROA peak is generated with the predicted percent abundances to estimate the true relative heights of the isotopic peaks to the C_0 and C_n peaks. The tentative IROA peak is analyzed separately (test and control portions) so the isotopic peak ratios can be checked without interference by fold changes in the experiment. Isotopic peaks that are predicted to have an intensity of ≥ 10% of the respective most intense isotopic peak are checked for proper intensities, as are isotopic peaks that are predicted to have at least 90% of their intensities from either the test or control (i.e. isotopic peaks that are predicted that have significant overlap, usually peaks near the middle of the IROA peak, are not checked). Lastly, each isotopic peak is checked to determine if the peak has previously been identified as belonging to an already found IROA peak. The fold change for an IROA peak is calculated by taking the ratio of the sum of all test isotopic peaks to the sum of all control isotopic peaks, using the actual intensities for all isotopic peaks that could be found in the mass spectra and that do not have significant overlap between test and control portions. The fold change is then normalized by multiplying by a normalization factor (described below).

After finding all of the IROA peaks within a chromatographic run, those that elute across several scans are grouped (analogous to finding peaks in an elution chromatogram for a given mass). Elution chromatograms are built using all IROA peaks containing the same

number of carbons and having masses within 5 ppm. Each chromatogram is peak picked by finding local minima and maxima and the intensities of all IROA peaks within each chromatographic peak are summed (constrained by ±5 seconds).

Afterwards, each IROA peak is queried in a modified, local copy of the HMDB database in which 111 ascarosides,[2] relevant to the biological system being interrogated, have been added. The local copy of the database was supplemented with missing KEGG IDs by matching compounds with their ChEBI ID.[3] An IROA peak matches a compound in the database if it contains the same number of carbons and the monoisotopic masses are within ±5 ppm. If more than one compound in the database fits these criteria, the number of additional possible compounds is recorded, along with the names of all possible compounds. If multiple matches exist, the compound selected is that with the lowest KEGG ID.

Normalization of IROA Peaks

Normalization is conducted to correct for errors caused by unequal numbers of worms in combined test and control batches and for unequal isotope labeling efficiencies. The method used was median fold change normalization which adjusts the median of the log fold changes of all IROA peaks within each sample to zero. This normalization procedure assumes that less than 50% of all IROA peaks are asymmetrically changing due to biological variation and so IROA peaks affected purely by dilution (unequal numbers of worms) or isotope labeling bias will all have the same fold changes.

Grouping IROA Peaks Across Samples

IROA peaks that are detected across several samples are said to originate from the same compound provided the peaks fall within the following search parameters: their masses are within a 5 ppm window, their elution times are within ±5 seconds, and they contain the same number of carbons. The ionization mode is taken into account and so a proton is added or subtracted to the mass depending upon the ionization mode. In the current implementation of our MATLAB script, adducts other than a proton are ignored.

Supplementary References

(1) Stiernagle, T. *WormBook* **2006**, 1–11.
(2) Reuss, von, S. H.; Bose, N.; Srinivasan, J.; Yim, J. J.; Judkins, J. C.; Sternberg, P. W.; CC, S. F. *J. Am. Chem. Soc.* **2012**, *134*, 1817–1824.
(3) Hastings, J.; de Matos, P.; Dekker, A.; Ennis, M.; Harsha, B.; Kale, N.; Muthukrishnan, V.; Owen, G.; Turner, S.; Williams, M.; Steinbeck, C. *Nucleic Acids Res* **2013**, *41*, D456–63.

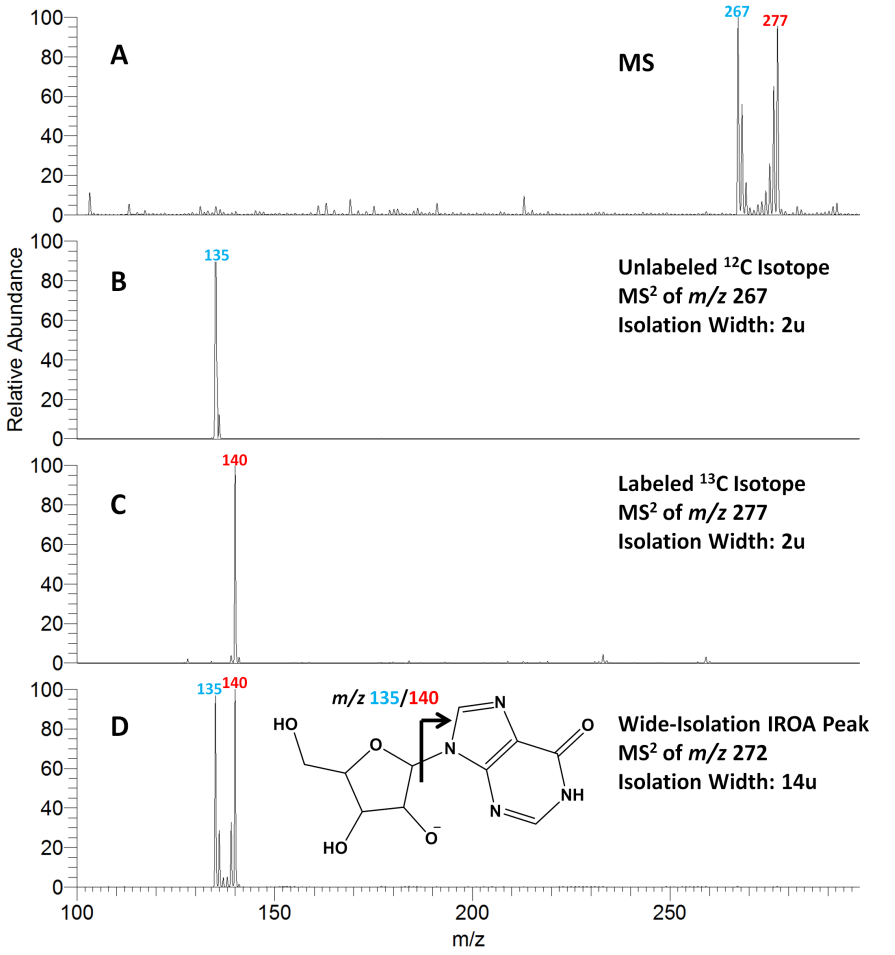

Figure S-1: Mass spectrometric analysis of the isotopic peaks resulting from inosine. A) The IROA peak detected from LC-MS analysis on the LTQ Velos. Data-dependent LC-MS/MS spectra were then collected from the heat-shock nematodes for B) the unlabeled ^{12}C and C) the fully labeled ^{13}C isotopic peaks for inosine. Additionally, a targeted wide isolation MS/MS experiment (Panel C) was conducted isolating and fragmenting the entire IROA peak, yielding the number of carbons in the fragment ion. The structure of the $[M-H]^-$ ion and proposed fragmentation pathway is also displayed in C.

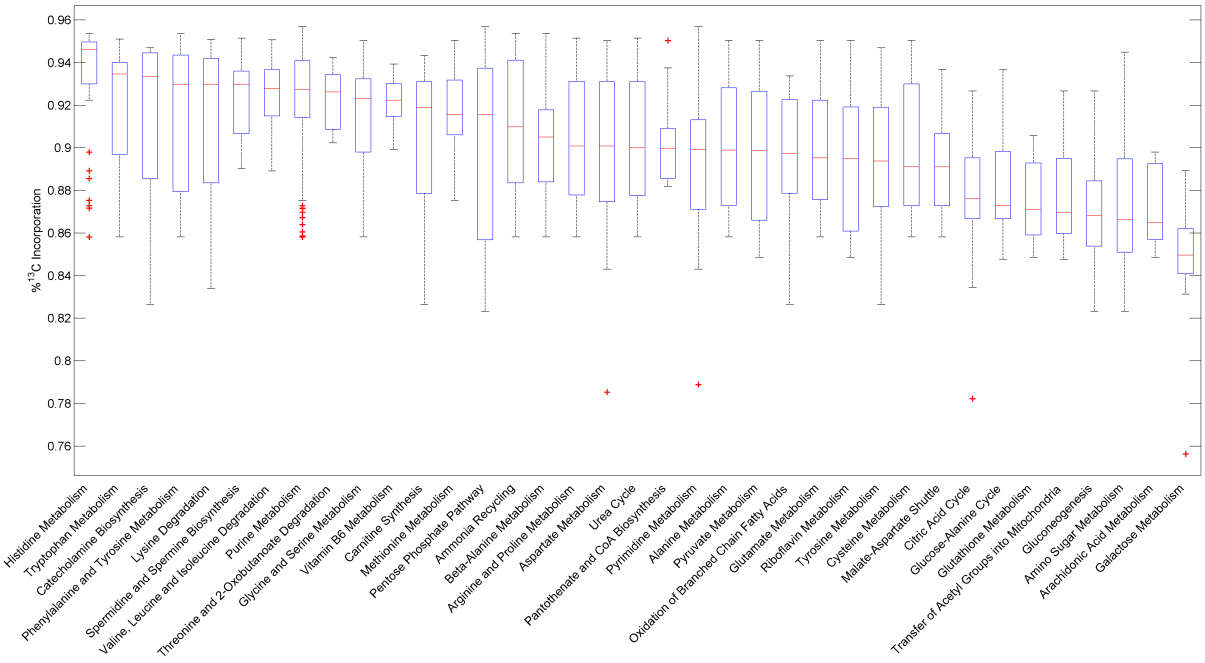

Figure S-2: Box plot of the %^{13}C abundances of 95% IROA peaks of compounds in pathways. Red crosses indicate outlying IROA peaks.

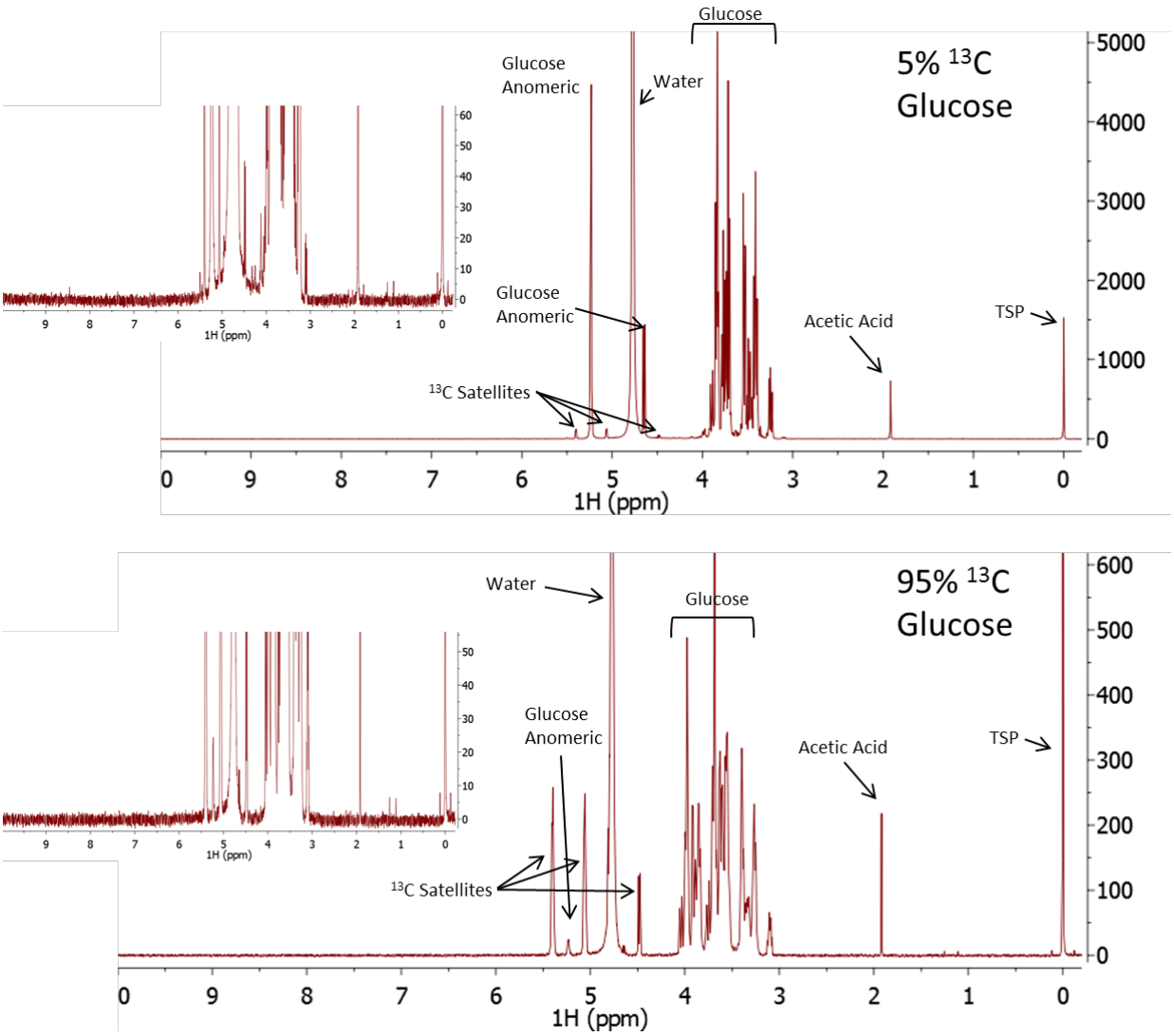

Figure S-3: NMR Analysis of pure 5% and 95% ^{13}C glucose. 1D-^{1}H NMR spectra are shown with TSP added for quantification. Expansions of the baseline are shown in the insets. The contaminating peak at 1.9 ppm was assigned as acetic acid and was quantified at about 0.5% w/w in both samples. NMR spectra were acquired on an Avance III 500 MHZ spectrometer in the AMRIS facility at the University of Florida. Spectra were acquired in D_2O with an 11 ppm spectral width, 30° pulse, 3 sec acquisition time, and 3 sec delay. Processing and quantification were done in MestreNova.

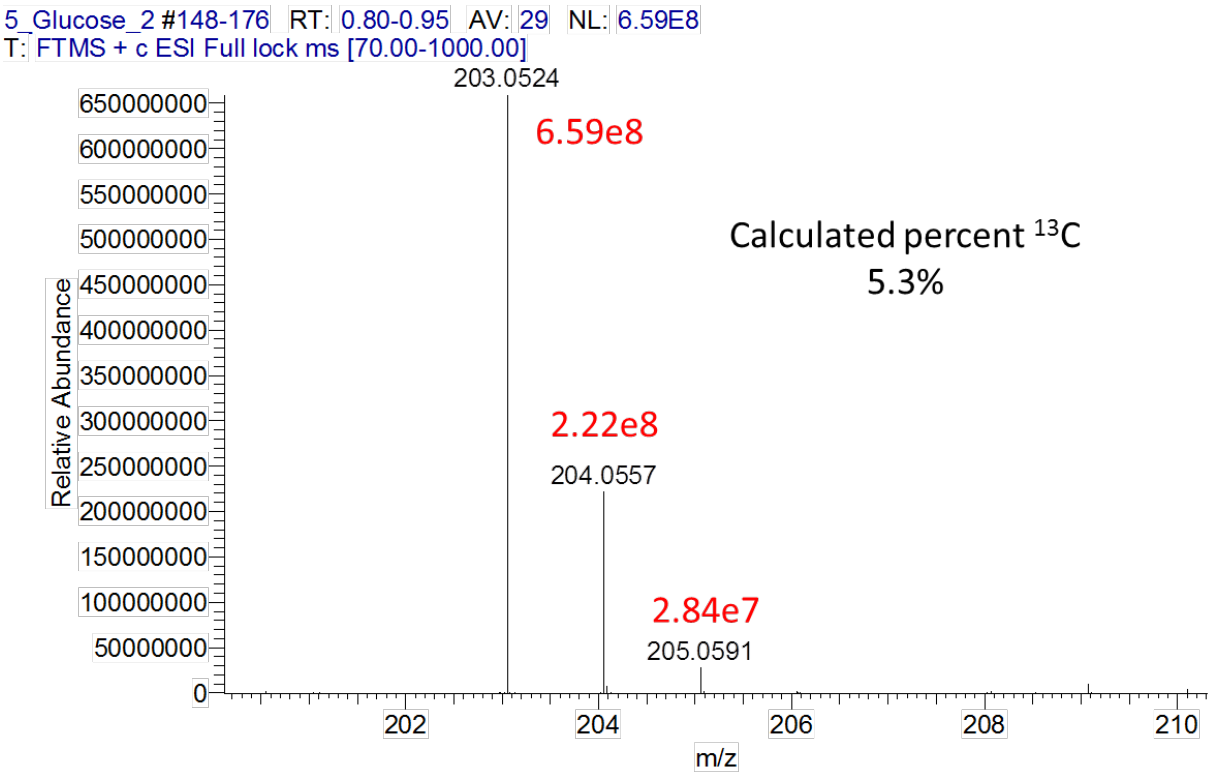

Figure S-4: LC-MS Analysis of 5% ^{13}C labeled glucose. The intensities for each isotopic peak are shown in red.

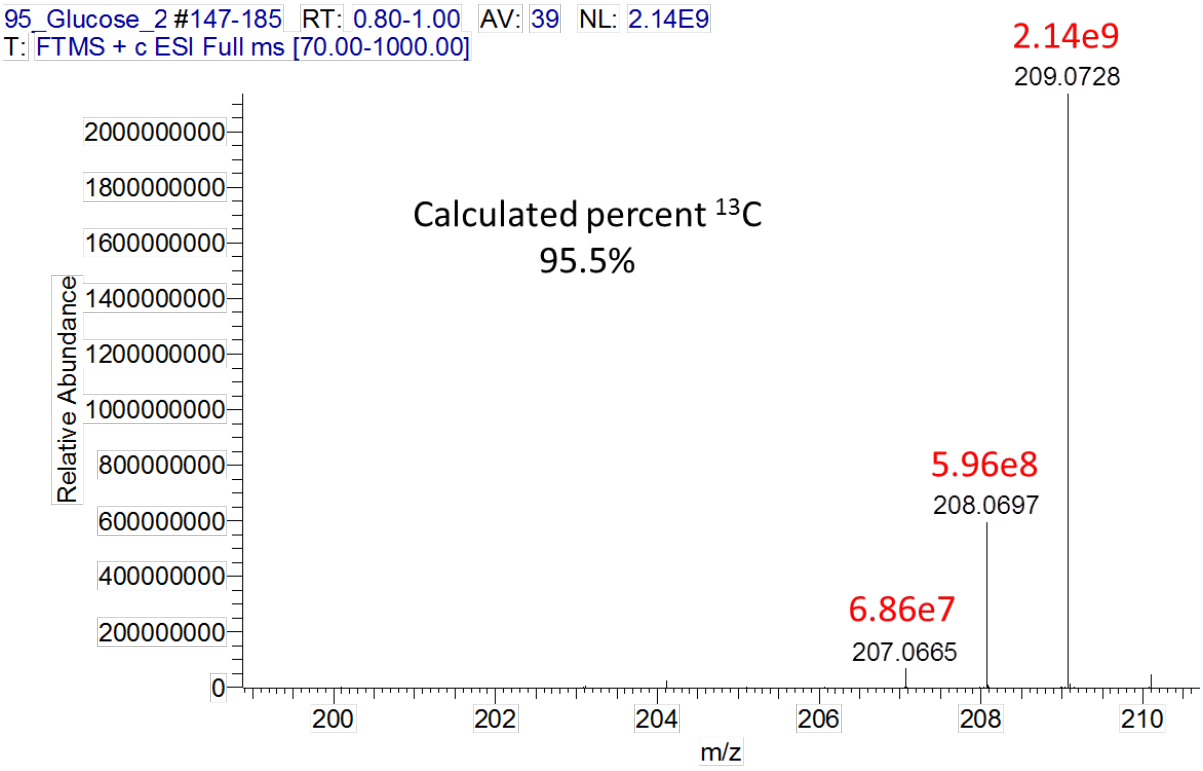

Figure S-5: LC-MS Analysis of 95% ^{13}C labeled glucose. The intensities for each isotopic peak are shown in red.

Figures S-6 – S-16: LC-MS/MS confirmation of compounds using the Thermo Scientific Q Exactive. A) Extracted ion chromatograms for the unlabeled ^{12}C (top) and fully labeled ^{13}C (bottom) isotopologues of the metabolite and B) the MS spectrum averaged across the aforementioned chromatographic peaks obtained from the initial UPLC-MS experiment. C) Extracted ion chromatograms for the unlabeled ^{12}C isotopologue (top), fully labeled ^{13}C isotopologue (middle), and most abundant MS/MS transition for the ^{12}C isotopologue (bottom), and D) the MS/MS spectrum obtained from the targeted HPLC-MS/MS experiment. E) The reference library spectrum obtained from the Metlin Metabolite MS/MS Database or the HMDB. The name for each compound is indicated above each figure.

Figure S-6: Xanthine

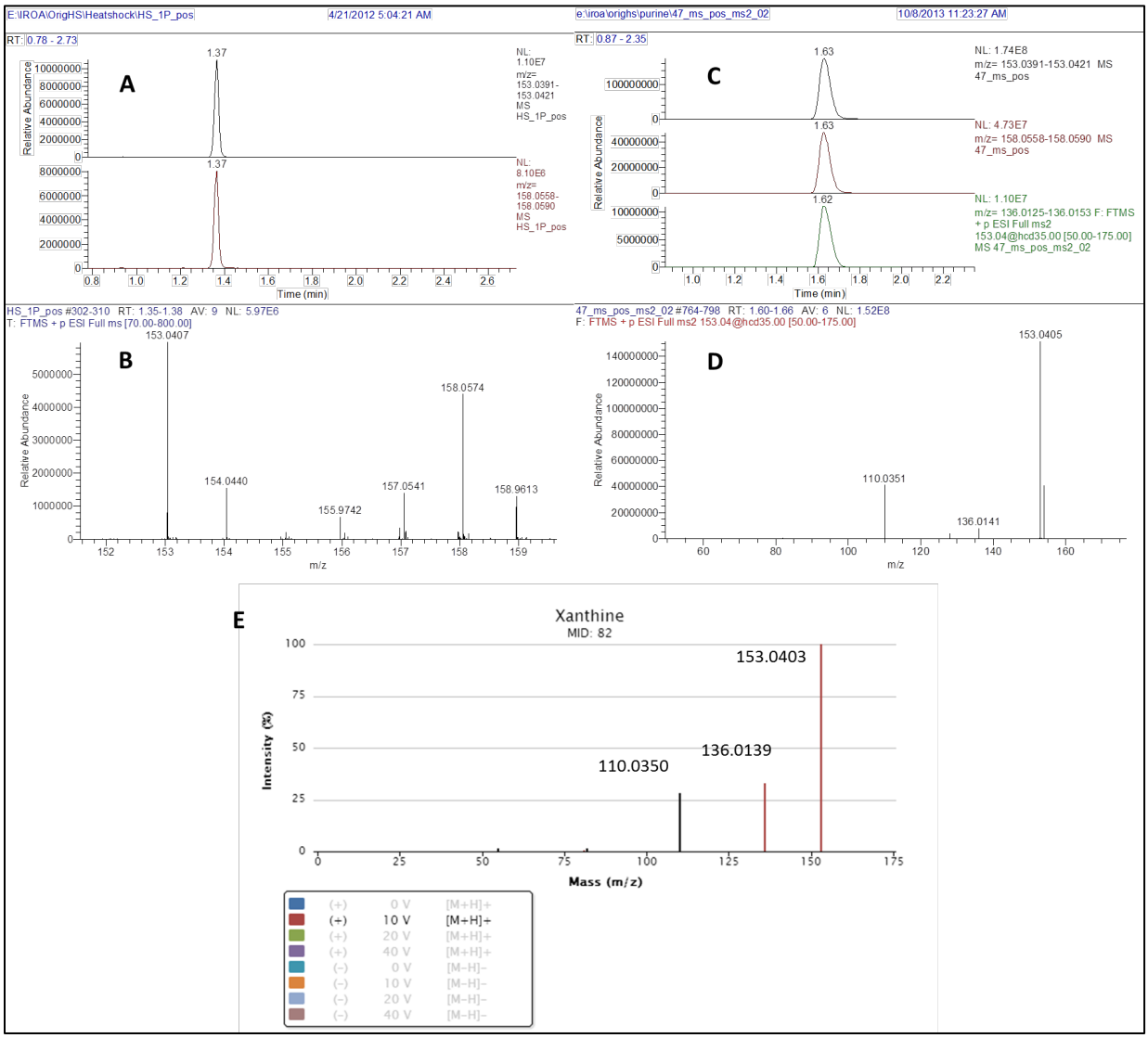

Figure S-7: Adenosine 5’-monophosphate

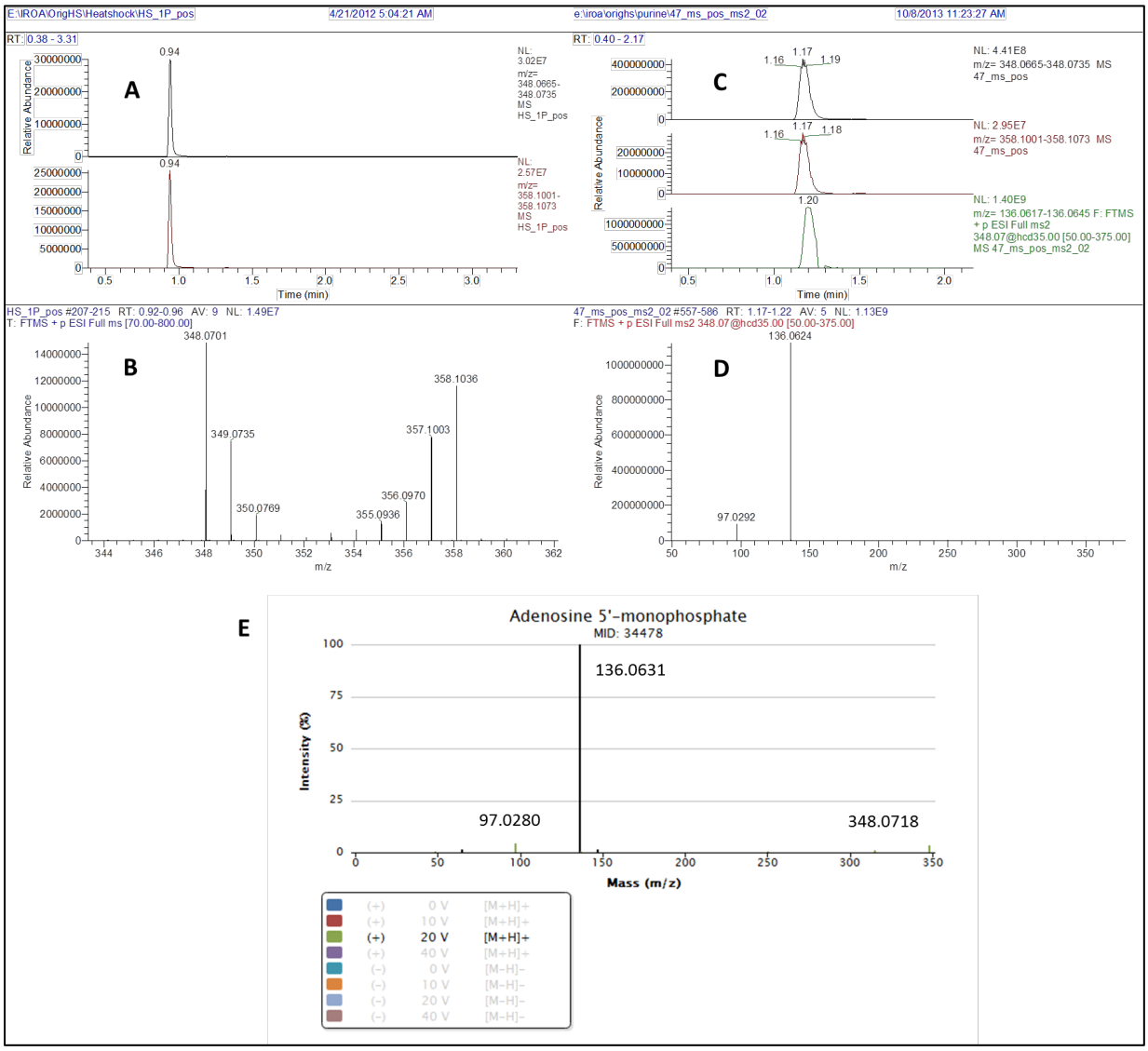

Figure S-8: Adenosine

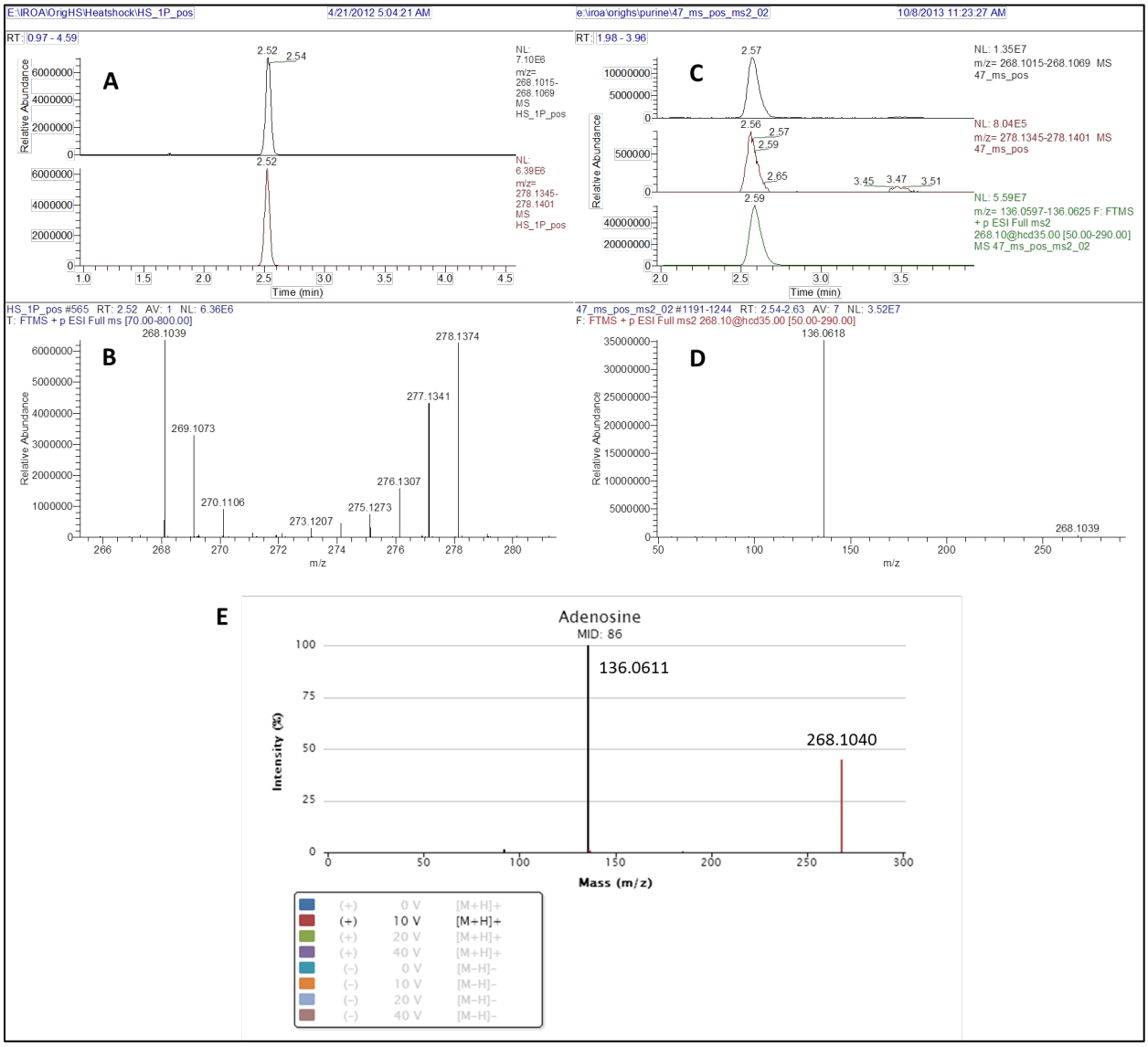

Figure S-9: Guanine

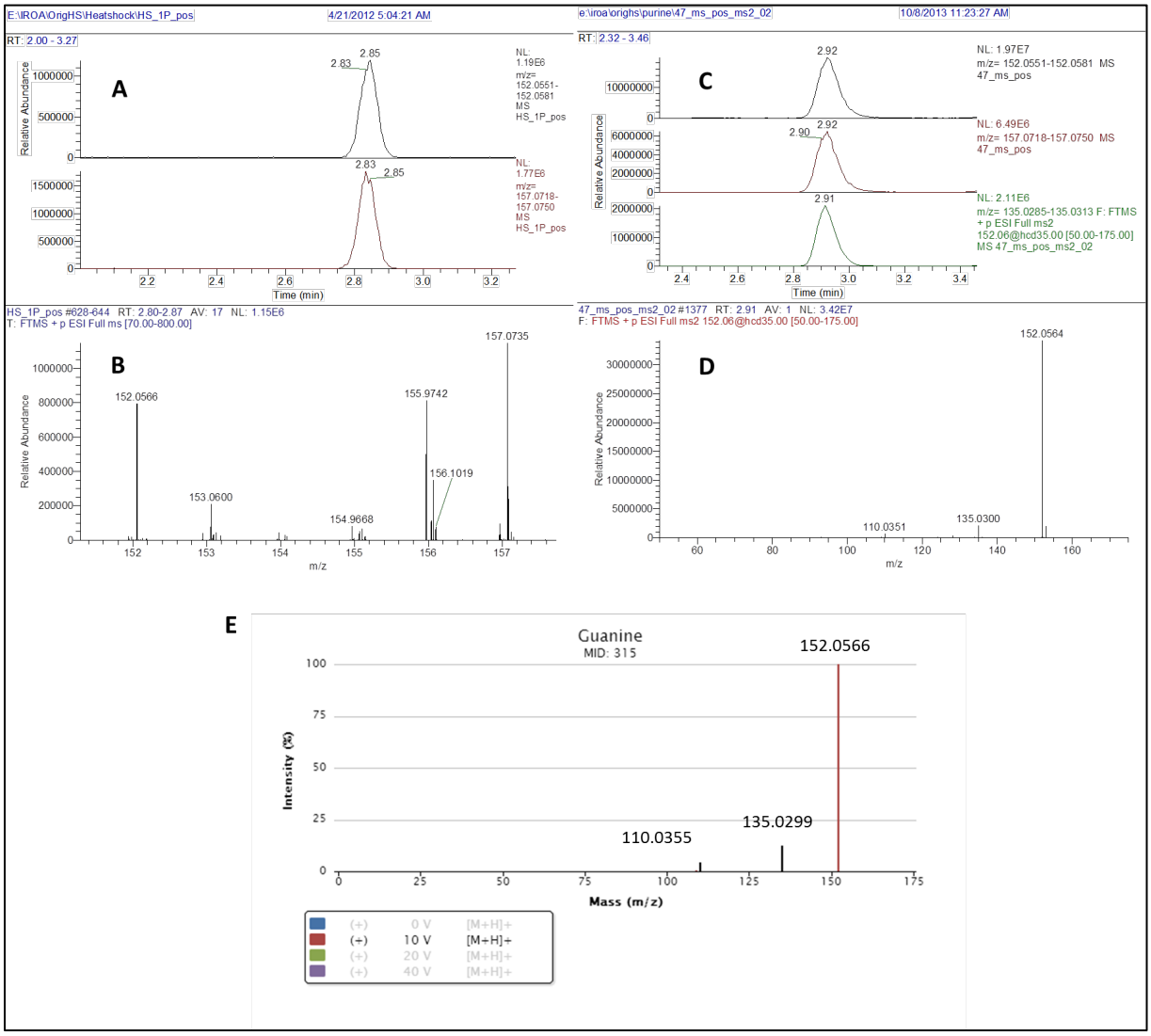

Figure S-10: Hypoxanthine

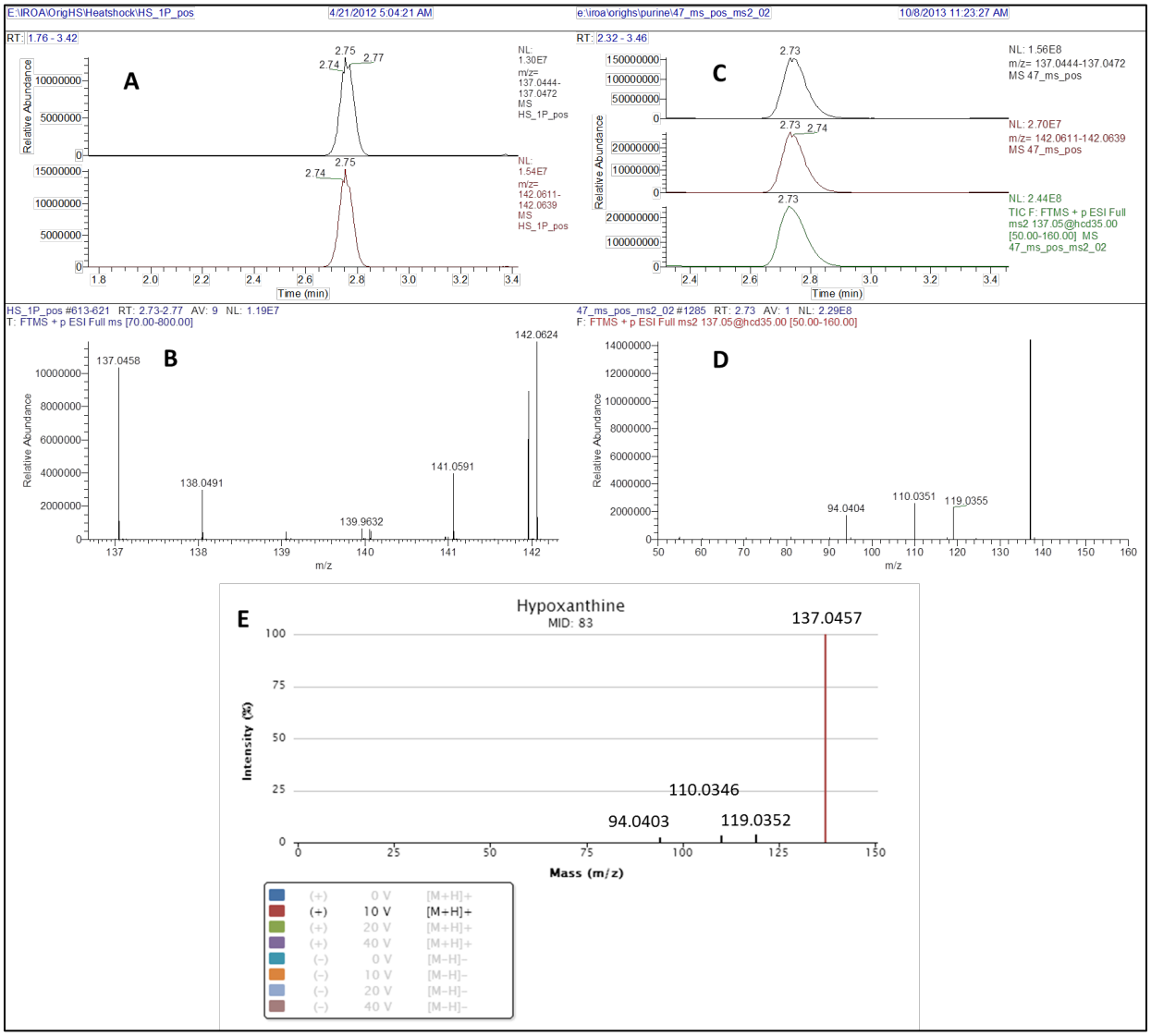

Figure S-11: Guanosine

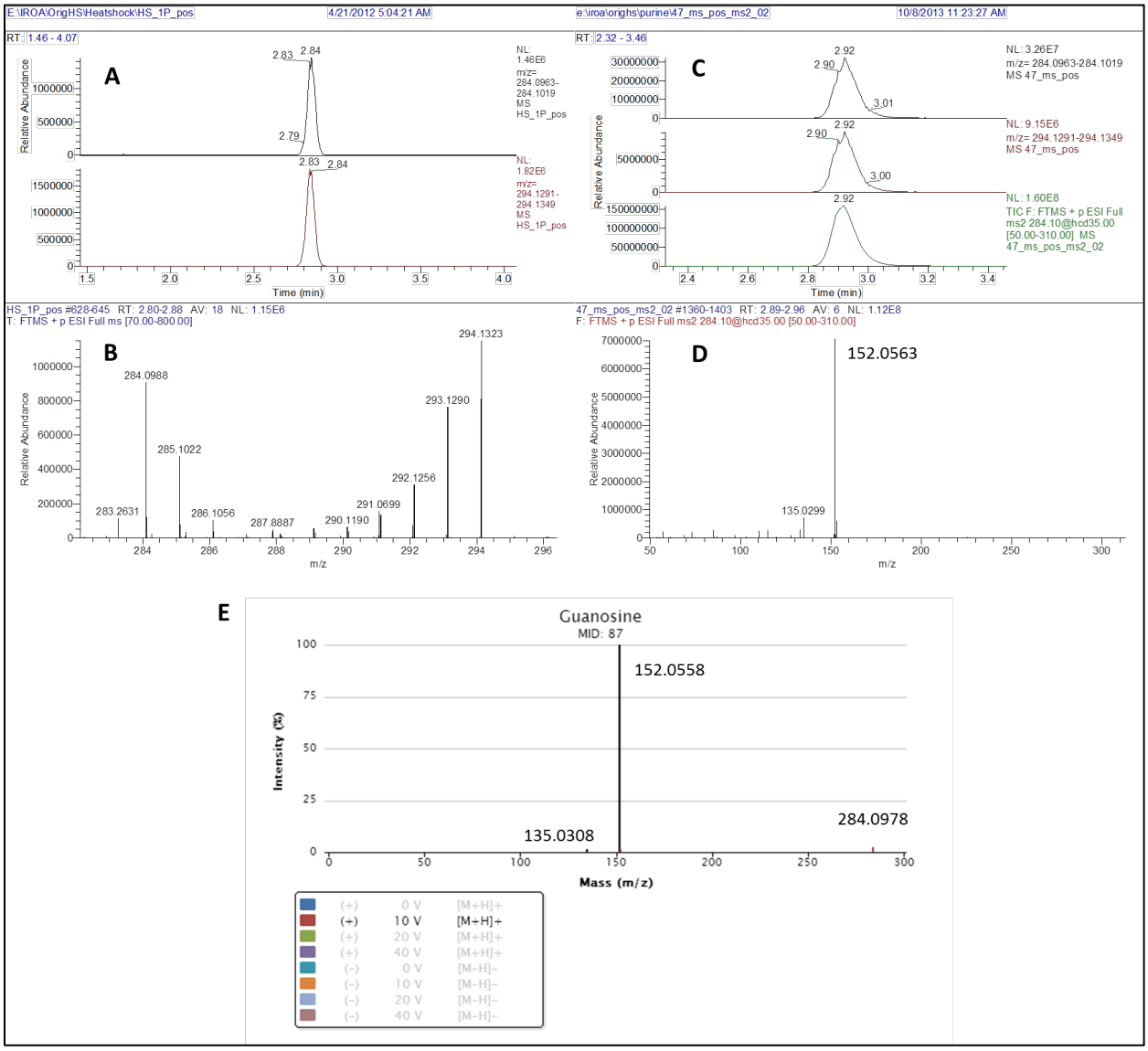

Figure S-12: Guanosine 5'-monophosphate

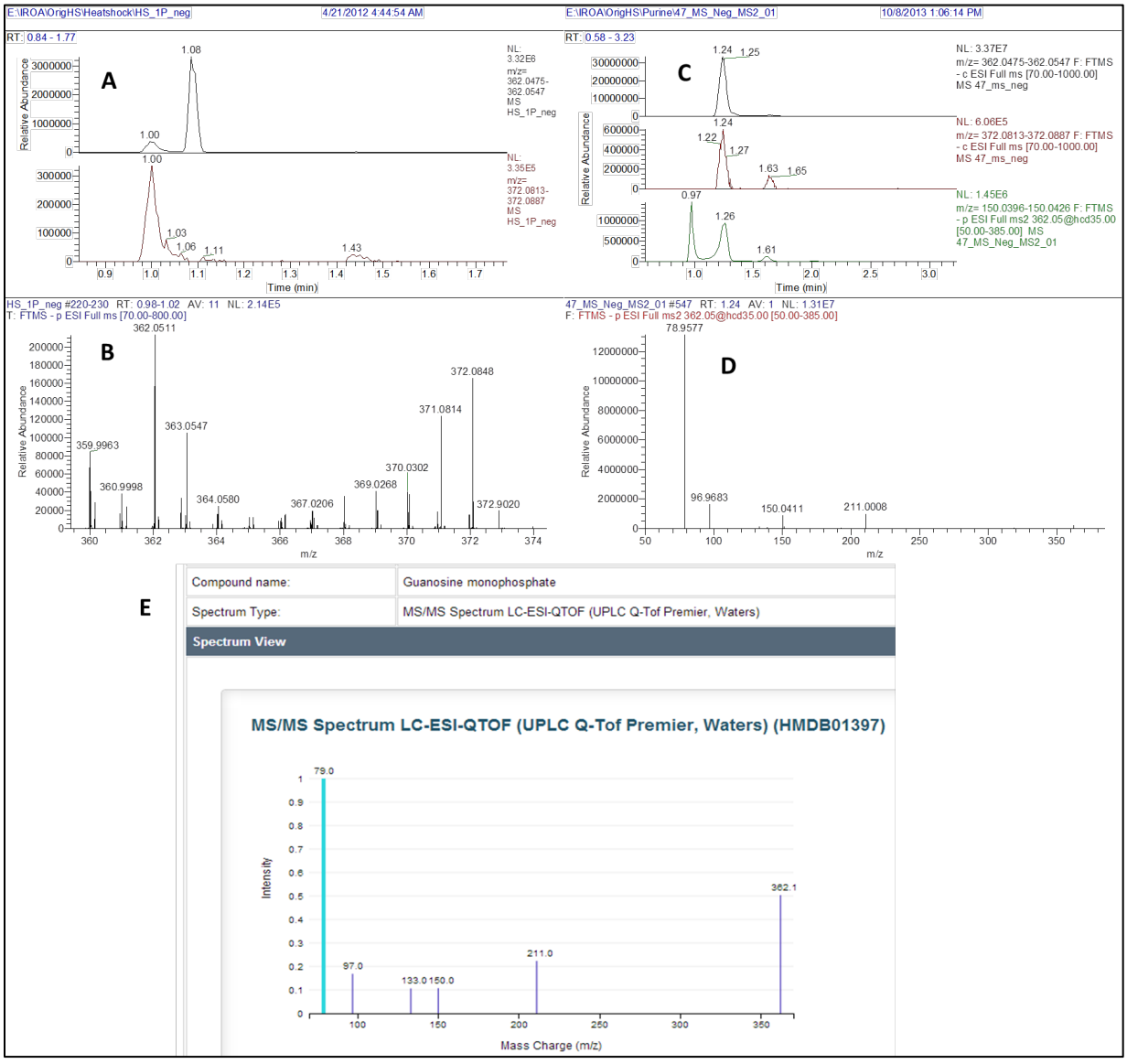

Figure S-13: Adenine

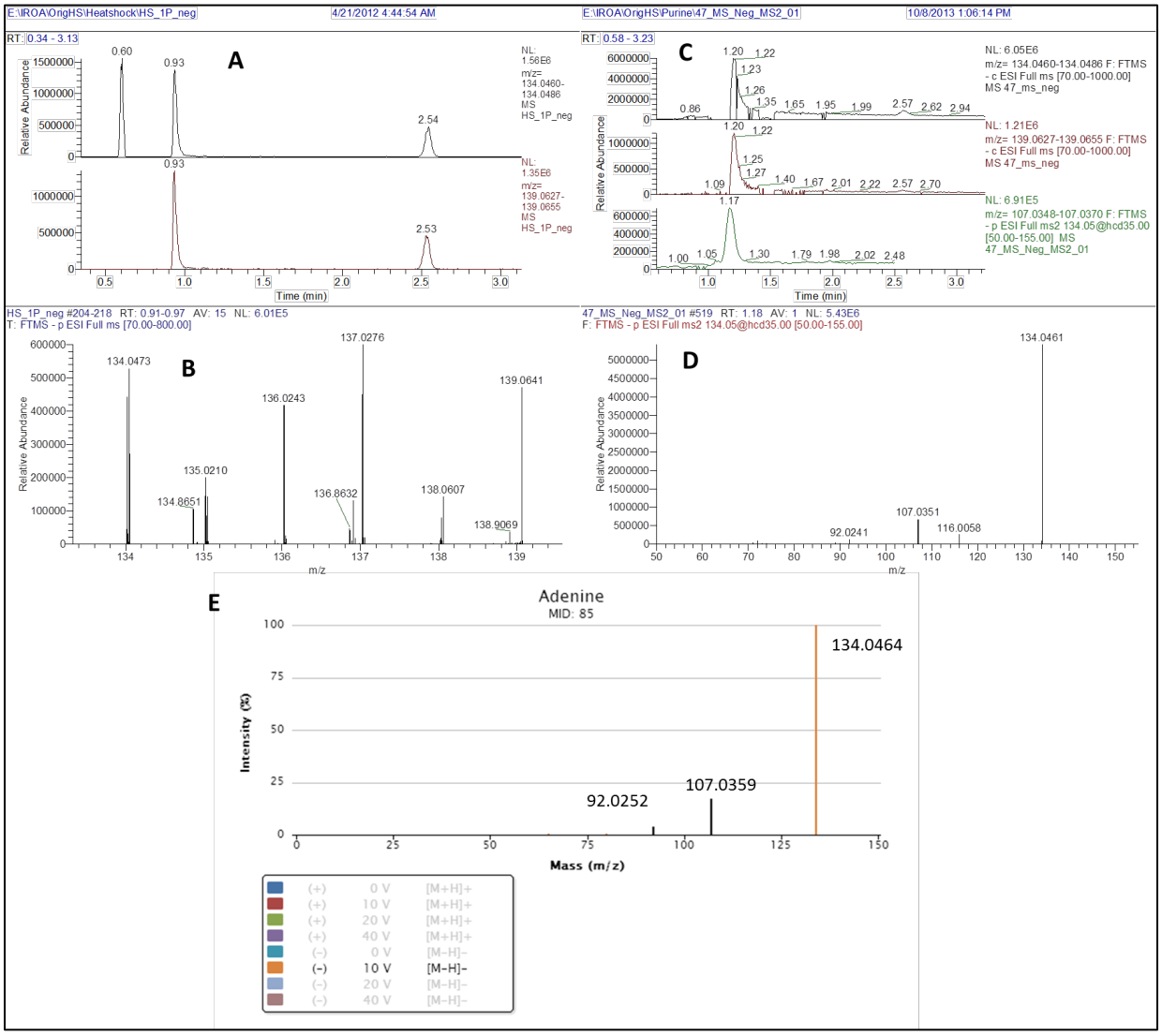

Figure S-14: Inosine

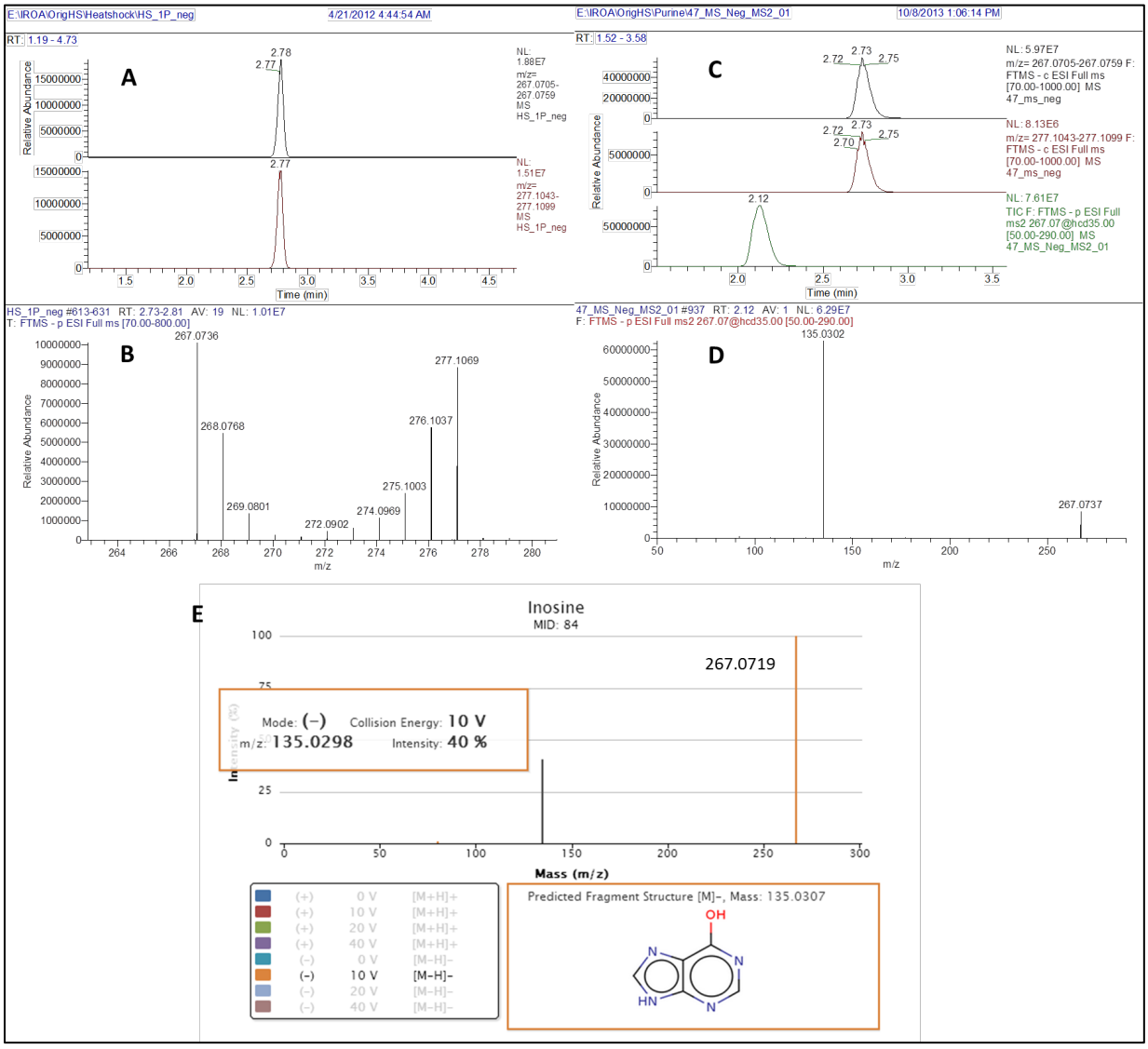

Figure S-15: Allantoic acid

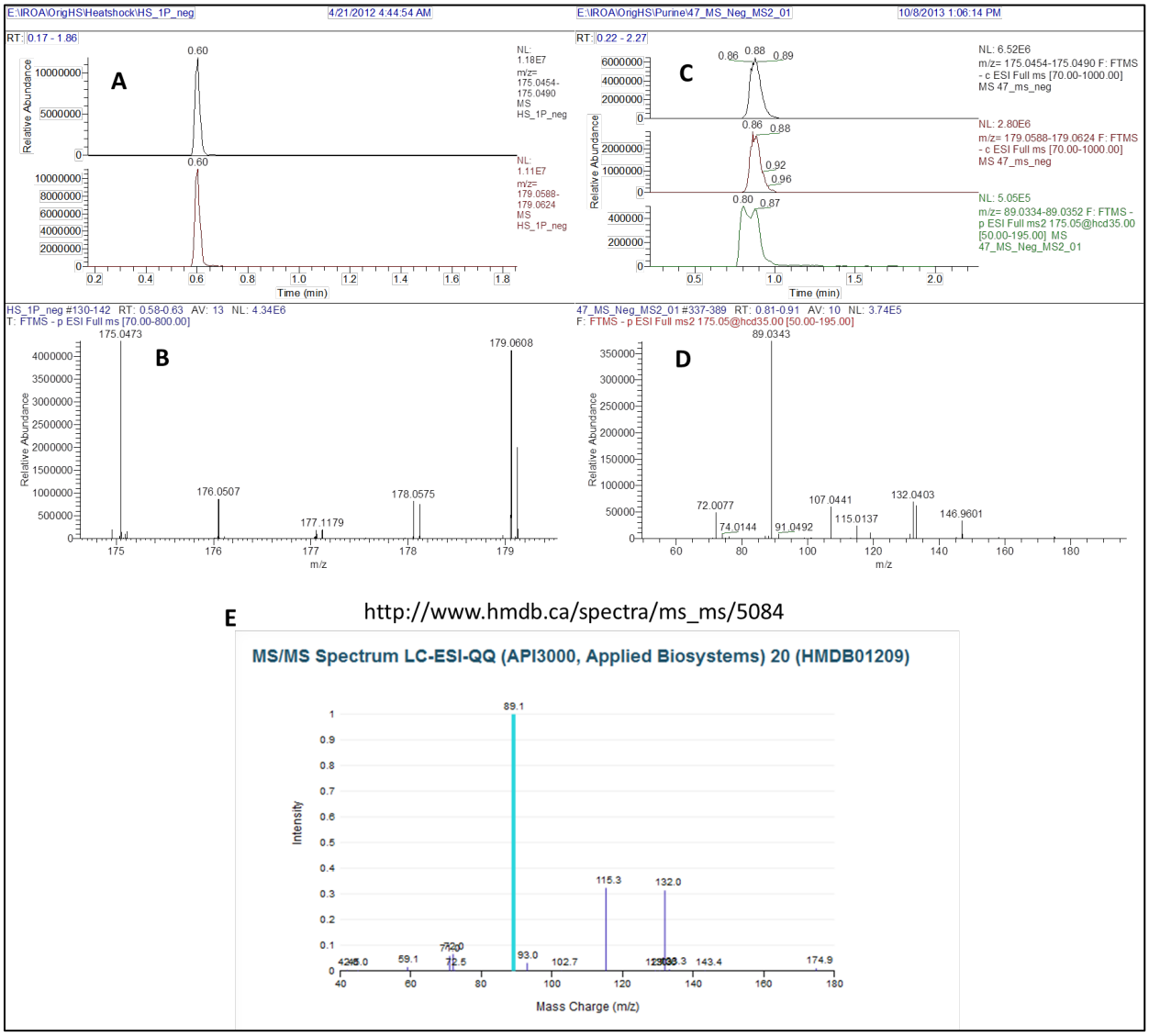

Figure S-16: Xanthosine