

# PyINETA: Open-Source Platform for INADEQUATE-JRES Integration in NMR Metabolomics

Rahil Taujale,<sup>◆</sup> Mario Uchimiya,<sup>\*,◆</sup> Chaevien S. Clendinen, Ricardo M. Borges, Christoph W. Turck, and Arthur S. Edison<sup>\*</sup>



Cite This: *Anal. Chem.* 2024, 96, 19029–19037



Read Online

ACCESS |



Metrics & More



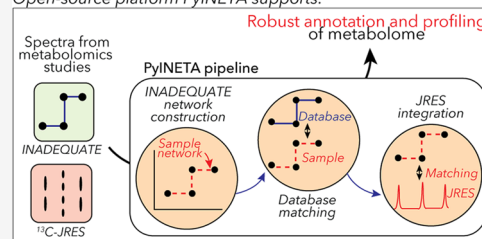
Article Recommendations



Supporting Information

**ABSTRACT:** Robust annotation of compounds is a critical element in metabolomics. The  $^{13}\text{C}$ -detection NMR experiment incredible natural abundance double-quantum transfer experiment (INADEQUATE) stands out as a powerful tool for structural elucidation, but this valuable experiment is not often included in metabolomics studies. This is partly due to the lack of a community platform that provides structural information based on INADEQUATE. Also, it is often the case that a single study uses various NMR experiments synergistically to improve the quality of information or balance total NMR experiment time, but there is no public platform that can integrate the outputs of INADEQUATE with other NMR experiments. Here, we introduce PyINETA, a Python-based INADEQUATE network analysis. PyINETA is an open-source platform that provides structural information on molecules using INADEQUATE, conducts database searches using an INADEQUATE library, and integrates information on INADEQUATE and a complementary NMR experiment  $^{13}\text{C}$  *J*-resolved experiment ( $^{13}\text{C}$ -JRES).  $^{13}\text{C}$ -JRES was chosen because of its ability to efficiently provide relative quantification in a study of the  $^{13}\text{C}$ -enriched samples. Those steps are carried out automatically, and PyINETA keeps track of all the pipeline parameters and outputs, ensuring the transparency of annotation in metabolomics. Our evaluation of PyINETA using a model mouse study showed that PyINETA successfully integrated INADEQUATE and  $^{13}\text{C}$ -JRES. The results showed that  $^{13}\text{C}$ -labeled amino acids that were fed to mice were transferred to different tissues and were transformed to other metabolites. The distribution of those compounds was tissue-specific, showing enrichment of specific metabolites in the liver, spleen, pancreas, muscle, or lung. PyINETA is freely available on NMRbox.

Open-source platform PyINETA supports:



Robust annotation of compounds is a critical task in metabolomics. In NMR metabolomics, compound annotation is primarily based on chemical shifts of  $^1\text{H}$ ,  $^{13}\text{C}$ , or both. Two-dimensional (2D) experiments increase the confidence level of annotation, providing correlations between protons or protons and carbons in molecules.<sup>1</sup> Although  $^1\text{H}$ -detection 2D experiments have been successfully implemented in metabolomics for this purpose,<sup>2</sup>  $^{13}\text{C}$ -detection NMR can complement  $^1\text{H}$  NMR and improve the quality of information.<sup>3,4</sup>  $^{13}\text{C}$  NMR has a broader chemical shift range and fewer overlapping peaks than  $^1\text{H}$  NMR, which is ideal for metabolomics samples that are complex mixtures of compounds.<sup>5</sup>  $^{13}\text{C}$  NMR can directly detect quaternary carbons, which leads to a broader coverage of carbon information in molecules. Most importantly from a perspective of structural elucidation,  $^{13}\text{C}$  NMR can directly extract the backbone structure of molecules,<sup>6,7</sup> essential information in structural elucidation.

Among various  $^{13}\text{C}$  NMR experiments, INADEQUATE (incredible natural abundance double-quantum transfer experiment)<sup>8</sup> stands out as a powerful tool for structural elucidation. This experiment unambiguously detects  $^{13}\text{C}$ – $^{13}\text{C}$  connectivity and extracts networks of carbons in molecules. INADEQUATE suffers from low natural abundance of  $^{13}\text{C}$ – $^{13}\text{C}$  couplings in

molecules (i.e., less than 1 in every  $10^4$  C–C bonds), but this experiment can benefit from isotopic enrichment and becomes applicable to metabolomics samples.<sup>7</sup> Although one could apply INADEQUATE to many samples in a metabolomics study and profile the metabolome differences between samples,<sup>7</sup> INADEQUATE requires a relatively long time for data collection, and this approach is not always practical especially when spectrometer time is limited. Thus, it is useful to have a profiling experiment that requires less instrument time but can be easily used with INADEQUATE. Although an obvious choice is a simple one-dimensional (1D)  $^{13}\text{C}$  experiment to profile all samples in a study,  $^{13}\text{C}$ -enriched samples have complicated peak shapes and more overlap than experiments at natural abundance  $^{13}\text{C}$ .

One approach to the problem is to use a 2D  $^{13}\text{C}$  *J*-resolved experiment ( $^{13}\text{C}$ -JRES), which separates chemical shifts from

Received: July 29, 2024

Revised: October 6, 2024

Accepted: November 12, 2024

Published: November 20, 2024



coupling constants into different dimensions. A 1D projection of 2D  $^{13}\text{C}$ -JRES is free from multiplets and can be collected quickly enough for efficient profiling. The output can be statistically processed and linked to 2D spectra for a representative sample such as internal pooled sample, a mixture of aliquots of all the study samples, for annotation. It has been shown that a combination of  $^{13}\text{C}$ -JRES for profiling and INADEQUATE for annotation can achieve both reducing overall experiment time and maintaining the quality of structural information for a metabolomics study.<sup>9</sup>

In addition to the robust compound annotation, the benefit of introducing INADEQUATE is its suitability for computational tasks. Clendinen et al. developed INETA (INADEQUATE network analysis) that computationally constructs networks of backbone carbons in molecules using the INADEQUATE rules. In INADEQUATE, two directly bonded carbon atoms resonate at their natural frequencies along the acquisition dimension and at the sum of their frequencies along the indirect double-quantum dimension. This leads to pairs of peaks that are symmetric along a diagonal with slope 2, and these pairs of INADEQUATE peaks are then linked vertically to expand the network. INETA used the constructed networks to search an internal INADEQUATE database, which was simulated using assigned  $^{13}\text{C}$  chemical shifts and chemical structures of compounds deposited in Biological Magnetic Resonance Bank (BMRB).<sup>10</sup> INETA was previously used to annotate the endo- and exometabolomes of  $^{13}\text{C}$ -enriched *Caenorhabditis elegans*.<sup>7</sup>

Despite the clear advantages of being able to annotate metabolites using INADEQUATE, there are several obstacles to its routine use. First, samples need to be isotopically labeled with  $^{13}\text{C}$ . Many microorganisms and plants can be uniformly enriched using a carbon source such as  $^{13}\text{C}$ -glucose at modest cost.<sup>4</sup> This is more challenging for human studies but select targeted pathways using isotope tracers with ex vivo tissue slices or cell cultures are regularly studied.<sup>11–13</sup> Second, access to a high-sensitivity cryogenic  $^{13}\text{C}$  NMR probe is necessary. Such probes are made commercially and can be accessed through large NMR facilities with user programs such as The National High Magnetic Field Lab (<https://nationalmaglab.org/>) or The Network for Advanced NMR (<https://usnan.nmrhub.org/>). Third, the previous software developed to perform INETA was written using Mathematica, which is not open source.<sup>7</sup> Finally, the previous software was not developed to integrate INADEQUATE with the  $^{13}\text{C}$ -JRES data.

Here, we propose PyINETA, an open-source platform that can automatically integrate INADEQUATE and JRES data. In addition to the functions that were originally implemented in INETA, our new PyINETA seamlessly transfers INADEQUATE information to JRES, providing compound information for individual JRES peaks. The pipeline is run on Python, and researchers can freely implement this open-source platform to various metabolomics studies. We evaluated the applicability of PyINETA using a model mouse study, in which metabolites originating from  $^{13}\text{C}$ -labeled diet were examined.

As the number of metabolomics publications increases rapidly, the transparency of studies is becoming more critical than ever. This is especially true for compound annotation where the basis for annotation is required with significant rigor<sup>14</sup> but is not always reported in publication.<sup>15</sup> PyINETA is designed to report all of the annotation steps, providing a community platform that ensures the reproducibility and transparency of compound annotation in metabolomics.

## EXPERIMENTAL SECTION

**Development of PyINETA.** PyINETA performs a series of tasks, including importing data, peak-picking, constructing networks, database matching for INADEQUATE spectra, and transferring the annotation information to JRES peaks. The pipeline requires two input file types: a configuration file and experimental NMR spectra. The configuration file contains all of the information relating to parameters used for the analysis. Adjustable parameters in PyINETA are summarized in Supporting Table S1. After the input files are loaded, the pipeline initializes a PyINETA class object. This object contains chemical shift values for the direct and double-quantum dimensions, along with an intensity matrix. The input spectra are those prepared by NMRPipe.<sup>16</sup>

Next, PyINETA defines the peaks for INADEQUATE spectra. For this step, it was pointed out that using a single intensity threshold is not sufficient to resolve peaks that are close together.<sup>7</sup> To solve this issue, PyINETA uses multiple thresholds. The peak-picking algorithm begins with the highest intensity threshold (parameter “PPmax”) for collecting peaks. Then the pipeline proceeds by collecting data points at different slices until it reaches the minimum intensity threshold (“PPmin”). The number of slices between PPmax and PPmin is defined by a parameter “steps”. In every slice, PyINETA finds the local maximum for each peak that reaches the threshold and defines an area for that peak. PyINETA then combines peaks collected from all slices, where the pipeline applies threshold parameters “PPCS” and “PPDQ” to group nearby peaks. Any peaks that have their center of mass within PPCS of one another along the direct dimension and within PPDQ along the double-quantum dimension are considered to be a single peak.

Once the peaks are defined, the pipeline creates INADEQUATE networks. In INADEQUATE, pairs of directly bonded  $^{13}\text{C}$  atoms resonate at the sum of their frequencies in the vertical double-quantum dimension. This leads to pairs of peaks that are symmetric along a diagonal with slope of 2. Based on this, the defined peaks are screened. To meet the definition of horizontally aligned peaks, the difference in chemical shift between peaks in the double-quantum dimension needs to be less than a threshold (DQT). Next, for each pair of horizontal peaks, the difference between the sum of their chemical shifts in the direct dimension and the mean of their chemical shifts in the double-quantum dimension needs to be less than a threshold (SumXY). Also, for each pair, the difference in the absolute chemical shift distance from a line of  $Y = 2X$  between peaks needs to be less than a threshold (SDT). When multiple horizontal networks originate from sequential carbons in a single compound, those horizontal networks can be linked by vertically aligned peaks of a shared carbon. To meet the definition of vertically aligned peaks, two peaks need to have a chemical shift difference less than a threshold (CST) in the direct dimension.

Once networks are created, PyINETA starts database matching. In this pipeline, every peak in every network is compared with peaks in a simulated database in the PyINETA package. This simulated database is created using experimental spectra deposited in BMRB (details are in “Generation of a Simulated INADEQUATE Database” section). First, when the distance between sample peaks and database peaks is less than a threshold (CSMT), peaks are considered matched. Second, when the number of chemical shift matches between database networks and sample networks is more than a threshold (NCMT), they are considered to be matched networks.

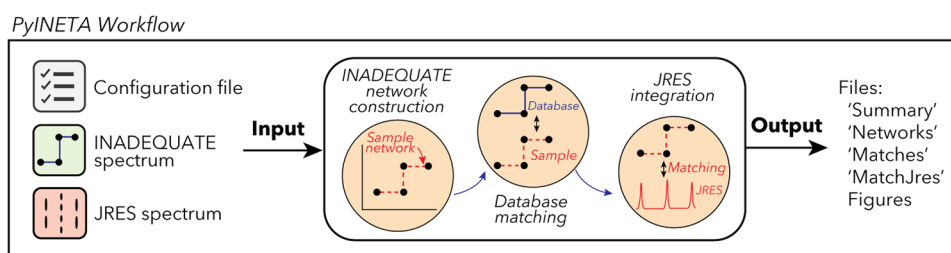


Figure 1. Workflow of PyINETA.

Matched networks are subsequently analyzed along the double-quantum dimension. For this step, when the difference in chemical shift between database networks and sample networks is less than a threshold (DQMT), database networks are considered matched networks. Since all of the database entries that satisfy the criteria are considered matched networks in this scheme, it is possible to find multiple matches for any given network. To evaluate the resulting matches, two scores, the hit score and coverage score, are assigned to each matched network. The hit score quantifies the proportion of peaks in sample networks that matched a specific database entry, whereas the coverage score represents the proportion of peaks in a database entry that matched those in a sample network. For both hit and coverage scores, 1 is the maximum value. For JRES peaks, peak area values are calculated (Peak\_Width\_1D) and the presence and absence of peaks corresponding to INADEQUATE networks are defined using a threshold (Intensity\_threshold\_1D).

Finally, a summary file reports all of the major statistics about the number of peaks that passed every step. Results from each step are also saved as pickle files (i.e., an object serialization mechanism in Python).

The developed pipeline is installed on NMRbox.<sup>17</sup> The source code is also available on GitHub (<https://github.com/edisonomics/PyINETA.git>) along with instructions and example data sets.

**Generation of a Simulated INADEQUATE Database.** We constructed a simulated INADEQUATE database using structural information and experimental 1D <sup>13</sup>C spectra deposited in BMRB,<sup>10</sup> following the scheme proposed previously.<sup>7</sup> It is important to note that in BMRB, there are cases where more than one chemical shift value is assigned to a single carbon when assignments are uncertain. To incorporate this validity information, we calculate an index “ambiguity score” for every compound in the PyINETA database. This index is calculated by  $CA/CT$ , where  $CA$  is the number of carbons that have more than one chemical shift assignments and  $CT$  is the total number of carbons in a molecule. The ambiguity score ranges from 0 (all carbons have no ambiguity) to 1 (all carbons are ambiguous) by definition. For example, in BMRB, all carbons for adenosine have unique chemical shifts. In this case, the ambiguity score for adenosine is 0 in the PyINET database. On the other hand, all carbons of glucose in the BMRB database have more than one chemical shift value to take care of anomers. For this case, the ambiguity score of glucose is 1 in the PyINETA database. The simulated database in this study contains 1973 entries, covering 1209 metabolites, which is available as a json file in the package. A list of compounds available in the PyINETA database, along with PyINETA ambiguity scores and BMRB entry identifications, is provided as a [Supporting File](#). Users can set a threshold for the ambiguity score (parameter “Ambiguity”) in the configuration file, and PyINETA does not

use database compounds that are more ambiguous than the threshold. When users further need to validate the BMRB information, an optional argument “-s singlePlot” is available to plot specific database compounds. Users can add new compounds to the PyINETA database using the module ‘gen\_PyINETAdb.py’. Input format for this module is either NMR-STAR<sup>18</sup> or tables with chemical shifts and structural information. The ambiguity information for peaks in input files is retained as ambiguity scores and reported in the final PyINETA output.

**In Vivo <sup>13</sup>C Labeling and Sample Preparation.** All animal experiments and protocols have been reviewed and approved by the Institutional Animal Care and Use Committee of the Max Planck Institute of Psychiatry. Three mice were fed with a diet that contained <sup>13</sup>C-labeled amino acids, including eight essential and eight nonessential amino acids ([Supporting Tables S2 and S3](#)). After this feeding, the mouse tissues were collected. Dried tissue samples were resuspended in 50  $\mu$ L of deuterated water and methanol (1:4 volume ratio) ([Supporting Table 4](#)), and the distribution of metabolites originating from those amino acids was analyzed. Complete procedures are described in [Supporting Information page S3](#).

**NMR Data Collection and Processing.** Data for INADEQUATE and <sup>13</sup>C-JRES experiments were collected on a Bruker Avance Neo 900 MHz with a 5 mm TXO cryoprobe (Bruker), using NMR tubes with a diameter of 1.7 mm (Bruker). All of the NMR spectra were processed using NMRPipe.<sup>16</sup> Detailed experimental and NMRPipe processing parameters are listed in [Supporting Tables S5 and S6](#), respectively. Additional data processing for JRES spectra was conducted on MATLAB R2022b (MathWorks). Further details are given in [Supporting Information page S4](#). The ALATIS numbering system<sup>19</sup> was used for describing carbon numbers.

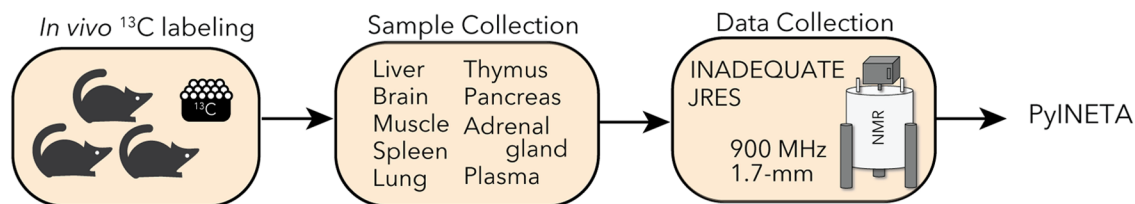
## RESULTS AND DISCUSSION

**PyINETA Provides a Flexible Environment for INADEQUATE-JRES Integration.** Since PyINETA uses various parameters to perform a series of tasks, we utilize a system of “configuration file” (Figure 1, left), which contains all the parameters that will be used in the pipeline ([Supporting Information page S5](#) for an example configuration file). Users can manage and overview the pipeline using this single file. In optimizing parameters, users do not need to run the whole pipeline, and each step can be run separately by defining the “-s” option. This saves computational time.

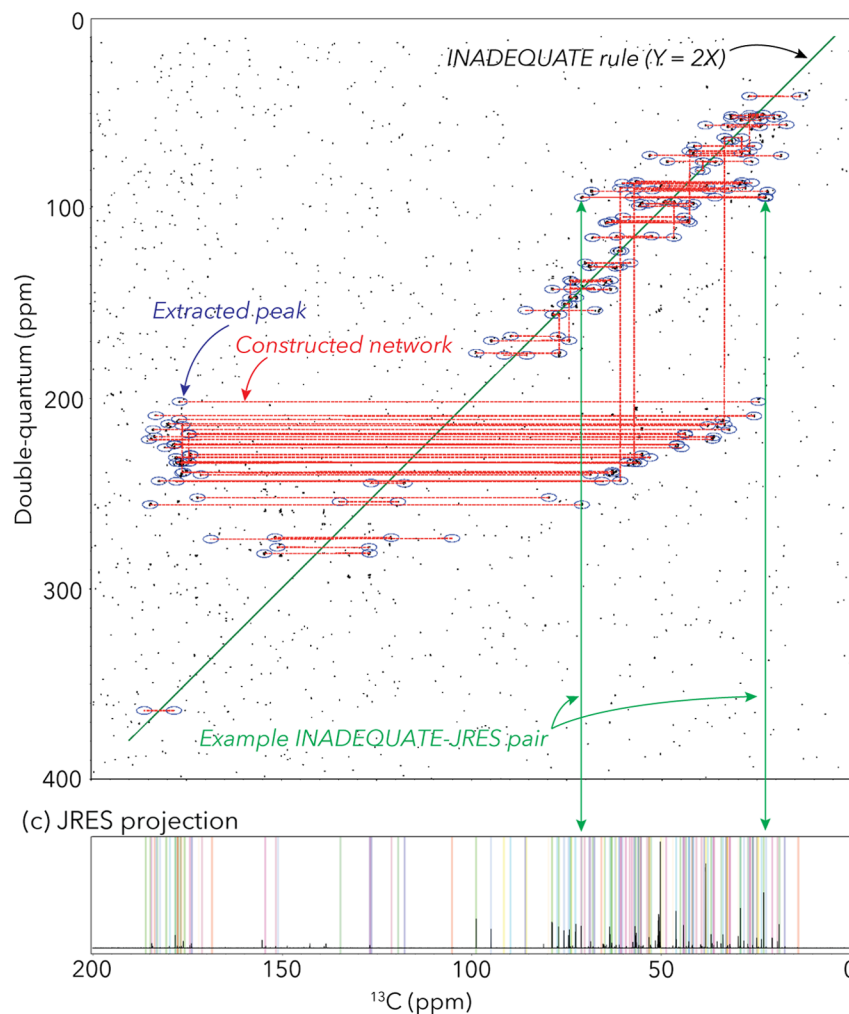
Using parameters in a configuration file and an input INADEQUATE spectrum, the pipeline constructs INADEQUATE networks and searches for the constructed networks in an internal database (Figure 1, middle). The internal database is based on <sup>13</sup>C chemical shift and chemical structure of metabolites deposited in BMRB,<sup>7</sup> one of the largest databases of assigned experimental NMR data from small molecules.<sup>10</sup> When



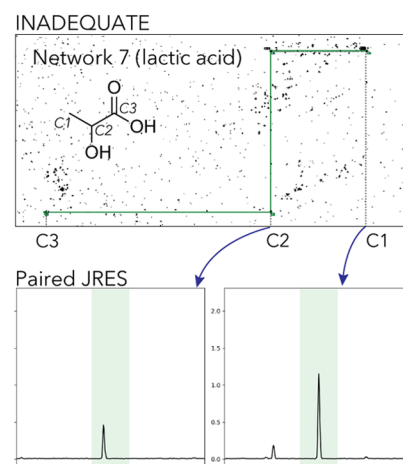
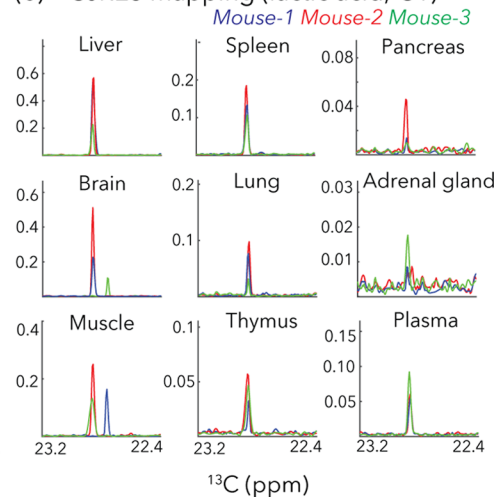
## (a) Experimental Workflow



## (b) PyINETA output: INADEQUATE



## (d) INADEQUATE-JRES pair

(e) <sup>13</sup>CJRES mapping (lactic acid, C1)

**Figure 2.** (a) Experimental workflow of this study. (b) Example of INADEQUATE networks for a mouse liver sample constructed by PyINETA. (c) JRES projection for the same sample. Peaks picked by PyINETA are highlighted in different colors. (d) INADEQUATE-JRES pair found by PyINETA. A compound name was also assigned using a database, which is lactic acid for this pair. (e) Mapping of a JRES peak in different tissues. JRES peaks for C1 for lactic acid are shown here. All of the plots for (b–d) are from the original outputs from PyINETA, with a slight graphical modification. The chemical structure was drawn using ChemDraw V23.

users have a compound(s) of interest that are not deposited with BMRB, they can manually add those compounds to the internal database. This includes other experimental databases that provide <sup>13</sup>C peaks assigned to known structures and computational data of chemical shifts from putative compounds.

Finally, as a key component, PyINETA integrates the information on INADEQUATE and <sup>13</sup>C-JRES (Figure 1, middle). PyINETA reads a raw <sup>13</sup>C-JRES spectrum and creates a projection. Then, peaks are picked from the projection, peaks between INADEQUATE and JRES are matched, and compound information based on INADEQUATE is transferred to JRES. We made this component optional so that users can still

use this pipeline when only INADEQUATE spectra are available.

After this processing, results are reported as a set of output files (Figure 1, right). “Summary” overviews the number of peaks passed every step in INADEQUATE processing steps (Supporting Information page S10 for an example file). “Networks” provides chemical shift values for network peaks (Supporting Information page S11). “Matches” shows matched database entries, compound names, and peak connectivity information (Supporting Information page S13). The Matches file also contains confidence scores for those matches (i.e., ambiguity score, hit score, and coverage score; see Experimental

**Table 1. List of Compounds and Corresponding INADEQUATE Networks Detected by PyINETA for a Mouse Liver Sample**

class	compound <sup>a</sup>	BMRB entry ID	ambiguity score	network ID in PyINETA	backbone carbons extracted <sup>b</sup>
amino acid	alanine	bmse000028	0	3, 41	C1–C2–C3
	glutamine	bmse000038	0	17	C1–C3
	leucine	bmse000042	0.3	32	C3–C5
	lysine	bmse000043	0	9, 10, 22	C2–C1–C3–C5
	threonine	bmse000049	0	6	C1–C2
	glutamic acid	bmse000037	0	27	C2–C4
	isoleucine	bmse000041	0	1, 2	C1–C3, C2–C4
	valine	bmse000052	0.4	4, 5	C1–C3–C2
	proline	bmse000047	0	15	C1–C3
	amino sugar	D-glucuronate <sup>c</sup>	bmse000440	0	57
amino alcohol	O-phosphoryl-ethanamine	bmse000308	0	33	C1–C2
amino sulfonic acid	hypotaurine	bmse000452	0	25	C1–C2
	taurine	bmse000120	0	30	C1–C2
pyrimidine	barbituric acid	bmse000346	0	59	C2–C1–C3
amine	betaine	bmse000069	0.8	54	C4–C5
	choline	bmse000285	0.8	44	C4–C5
	ethanolamine	bmse000276	0	36	C1–C2
	putrescine	bmse000109	0	12	C3–C2–C1–C4
organic acid	lactic acid	bmse000208	0	7, 55	C1–C2–C3
	chloroacetic acid	bmse000367	0	38	C1–C2
	Uk-13			13	C(25.8)–C(183.0)
	Uk-16			16	C(28.0)–C(60.9)–C(182.2)
	Uk-20			20	C(31.7)–C(32.5)
	Uk-28			28	C(37.6)–C(52.6)
	Uk-34			34	C(43.7)–C(53.6)
	Uk-35			35	C(43.9)–C(174.3)–C(59.2)
	Uk-39			39	C(49.3)–C(46.8)–C(68.2)
	Uk-40			40	C(52.7)–C(61.8)
	Uk-42			42	C(55.2)–C(173.8)
	Uk-45			45	C(60.6)–C(61.5)
	Uk-46			46	C(68.9)–C(62.0)
	Uk-49			49	C(63.5)–C(78.8)
	Uk-51			51	C(65.7)–C(177.6)
	Uk-58			58	C(77.3)–C(89.7)
Uk-62			62	C(117.6)–C(126.4)	
Uk-64			64	C(121.1)–C(151.6)	

<sup>a</sup>Only a conservative list of metabolites based on the inspection of the original output is shown. The original outputs are in [Supporting Information pages S11 and S13](#). <sup>b</sup>For numbering of carbons, the ALATIS numbering system<sup>19</sup> was used, except for unknown compounds where chemical shift values were indicated in parentheses. <sup>c</sup>Only partial structural information was available in the BMRB entry.

Section for details), and users can evaluate the reliability of annotation. The INADEQUATE-JRES integration step is summarized in “MatchJres” ([Supporting Information page S24](#)).

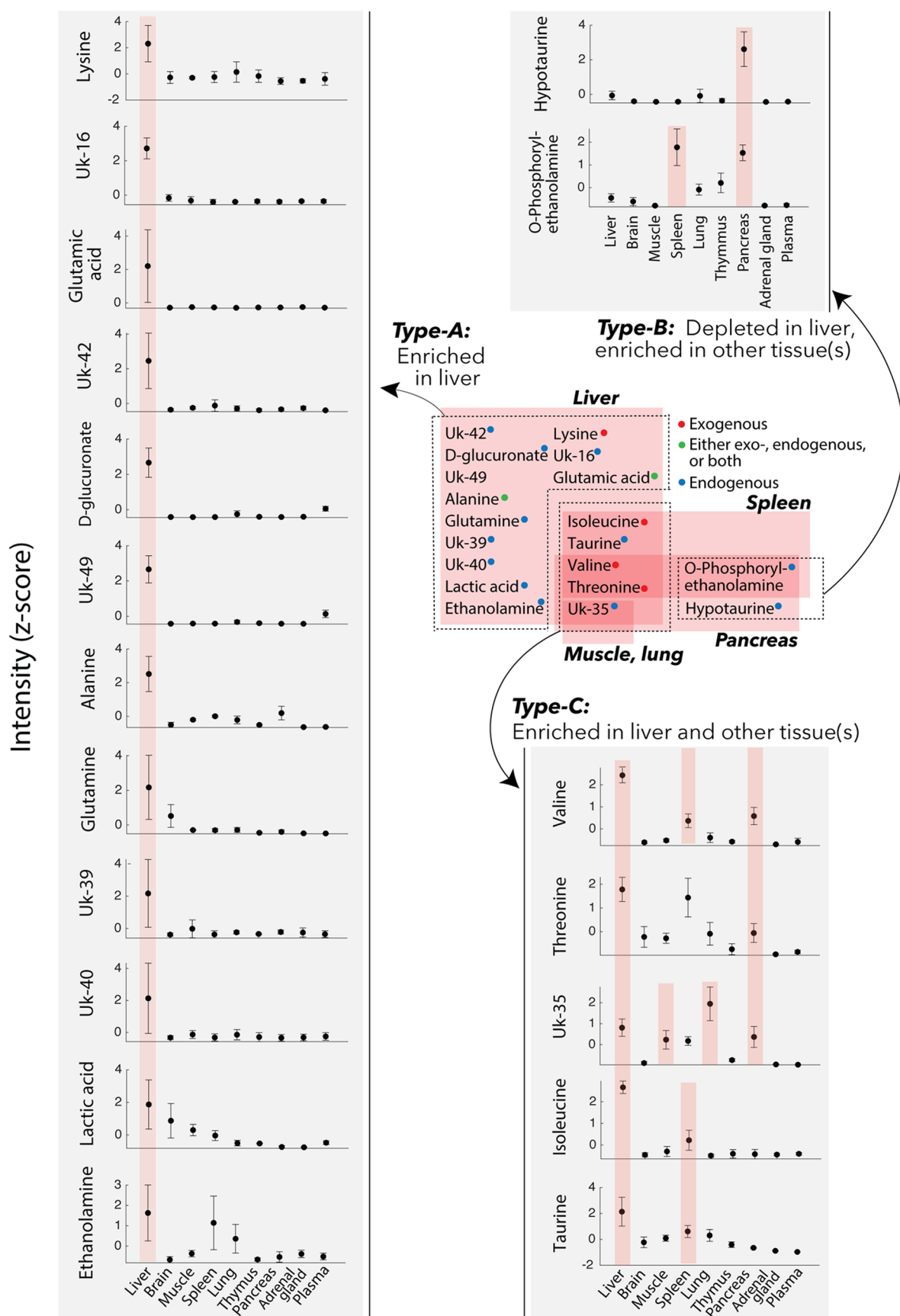
In addition to those output files, PyINETA provides figures for individual networks and matched compounds for INADEQUATE (examples will follow in the next section). Similarly, PyINETA creates figures for matched peaks for INADEQUATE and JRES. This capability was implemented to enable users to further validate the results of automatic annotation. The annotation information made by this new PyINETA is consistent with that of the original study INETA<sup>7</sup> ([Supporting Information page S32 and Table S7](#)).

**<sup>13</sup>C-JRES Profiling Showed Clear Tissue-Specific Spectral Patterns.** We applied <sup>13</sup>C-JRES profiling to a mouse study where a fate of a diet was investigated ([Figure 2a](#)). Three mice were fed with a diet that contained <sup>13</sup>C-labeled amino acids, including eight essential and eight nonessential amino acids ([Supporting Table 3](#)). After this feeding, mouse tissues were collected and the distribution of metabolites originating from those amino acids were analyzed in several

tissues including liver, adrenal gland, lung, muscle, pancreas, plasma, brain, spleen, and thymus.

JRES spectra in those tissues were consistent among the three mice ([Supporting Figure S1](#)), and higher intensities were observed in liver samples.

**PyINETA Integrated INADEQUATE and JRES Information Automatically.** Since the profiling results were consistent between mice and the liver samples had the highest intensities, we used one of the liver samples (Sample ID, 30) for the evaluation of INADEQUATE-JRES integration in PyINETA. From an INADEQUATE spectrum collected for the liver sample ([Supporting Figure S2a](#)), PyINETA constructed 67 INADEQUATE networks ([Figure 2b](#) and [Supporting Information page S11](#) for a list of all networks). The majority (52 out of 67) of the networks had two peaks, whereas 15 networks were longer, containing 8 peaks at maximum in a network (Network 8) ([Supporting Information page S11](#) for a complete list). Networks with just two carbons could reflect the original structure of the compound, including the case where networks are separated by heteroatoms in molecules. They could also



**Figure 3.** Distribution of compounds in the mouse tissues. Those are compounds originated from the  $^{13}\text{C}$ -labeled diet the mice were fed. Gray insets: peak intensities based on JRES (z-scored). Error bars, standard deviation ( $n = 3$ ). When a compound in a specific tissue is enriched compared with any other tissue, it is highlighted in pink (ANOVA with multiple comparison; a complete statistical summary is in Supporting Figure S3).

originate from compounds with longer backbones when whole networks were not created computationally through parameter choices or factors such as signal-to-noise in the original data. We observed both situations in our output. For example, choline is a compound with a backbone of a single network (C–C) and was found in a single horizontal network of Network 44 (Supporting Figure S2b). On the other hand, lactic acid, which has a network of three carbons (C–C–C), was found in two separate horizontal networks (Networks 7 and 55) because of missing vertical connection of two horizontal networks (Supporting Figure S2c,d). Those broken networks can be manually inspected or improved by relaxing the tolerance parameter for vertical network construction (CST; Supporting Table S1).

PyINETA then searched for those 67 networks in the database, and 46 networks matched at least one candidate compound (Supporting Information page S13). The rest of the 21 networks did not have any matched compound, indicating that they are compounds that are not in BMRB (“unknown compound” hereafter).

One network can potentially match more than one compound in the database when compound structures are similar. Also, a single compound can potentially exist in more than one network, as described above. We further investigated the results using PyINETA’s function of output figures and excluded matches with less confidence due to partial structural similarity. As a result, the matched networks were those for 21 compounds (Table 1). They included amino acids (alanine, glutamine, leucine, lysine, threonine, glutamic acid, isoleucine, valine, and proline), an amino sugar (D-glucuronate), an amino alcohol (O-phosphorylethanolamine), amino sulfonic acids (hypotaurine and taurine), a pyrimidine (barbituric acid), amines (betaine, choline, ethanolamine, and putrescine), and organic acids (lactic acid and chloroacetic acid). Most of them showed no ambiguity in annotation (Table 1), except for four compounds (leucine, valine, betaine, and choline). We further investigated the original BMRB entries for those four compounds, and we confirmed that the ambiguity originates from di- or trimethyl carbons for which chemical shifts are too close to resolve. For this situation, the annotation based on INADEQUATE is still acceptable, since those residues create overlapping networks with the identical network structure and only a slight difference in chemical shifts. They also included 16 unknown compounds (Table 1).

Next, PyINETA analyzed a JRES spectrum collected for the same sample (Supporting Figure S2e). Among 67 INADEQUATE networks, 65 of them had JRES peaks in the corresponding regions (Figure 2c and Supporting Information page S24 for a complete list). PyINETA then transferred compound information based on INADEQUATE to JRES peaks (Figure 2d).

**PyINETA Revealed the Distribution of Metabolites Originating from a Diet in Different Tissues.** Since PyINETA transformed compound information to JRES peaks, we were able to examine the distribution of a specific metabolite in different tissues based on JRES (Figure 2e). We further extended this analysis to other metabolites and examined the distribution of the metabolites in different tissues. For this analysis, we used representative JRES peaks (no overlapping peaks with a minimum intensity of 0.1), and 19 compounds are included in the following analysis. We found three different categories (Figure 3). Among the 19 compounds, 13 of them were enriched in the liver compared with other tissues (Compound Type-A) (Figure 3). Compounds in this category

are amino acids (lysine, glutamic acid, alanine, and glutamine), an organic acid (lactic acid), and an amino sugar (D-glucuronate). On the other hand, two metabolites were depleted in liver but enriched in other tissue(s) (Type-B) (Figure 3); they included an amino alcohol (O-phosphorylethanolamine) in pancreas and spleen, and an amino sulfonic acid (hypotaurine) in pancreas. Finally, four compounds were enriched in both liver and other tissue(s) (muscle, spleen, lung, or pancreas; type-C; Figure 3). They are amino acids (valine, threonine, and isoleucine) and an amino sulfonic acid (taurine). Since INADEQUATE detects  $^{13}\text{C}$ – $^{13}\text{C}$  coupling in molecules that occurs less than 0.01% in natural abundance, here we interpret that the metabolites in our results originated from the  $^{13}\text{C}$  that were fed to the mice and the effects of natural abundance metabolites that were originally present in tissues are negligible.

There could be different sources of the metabolites in those three categories. Lysine, isoleucine, valine, and threonine are essential amino acids and were included in the  $^{13}\text{C}$ -labeled amino acids in the diet, suggesting that those amino acids were directly distributed from the diet to tissues (i.e., exogenous; Supporting Figure S4). Alanine and glutamic acid were also contained in the diet, but also, they are nonessential amino acids, which could be biosynthesized (i.e., endogenous), leaving a possibility that those amino acids were either exogenous, endogenous, or both (Supporting Figure S4). On the other hand, another nonessential amino acid glutamine was not included in the diet, indicating that glutamine was exclusively endogenous in this system. For metabolites other than proteinogenic amino acids (D-glucuronate, lactic acid, ethanolamine, taurine, O-phosphorylethanolamine, and hypotaurine), they are exclusively endogenous (Supporting Figure S4).

Liver plays a central role in amino acid metabolism, and this was reflected by our results. Net uptake of alanine predominantly occurs in liver,<sup>20</sup> consistent with the observation of enriched alanine in liver (Type-A pattern). Alanine is further used in liver to produce other metabolites including glutamic acid,<sup>21</sup> which was also in the Type-A pattern. Alanine also serves as a major precursor for gluconeogenesis which occurs in liver.<sup>21,22</sup> Those transformation processes suggest that the rate of alanine input was exceeding that of transformation, resulting in the enriched alanine observed in this study. Similarly, liver is one of the dominant tissues that take up glutamine.<sup>21,22</sup> On the other hand, branched-chain amino acids (BCAAs) isoleucine, valine, and threonine were not exclusively high in liver but were also abundant in other tissues (type-C). This could be due to the fact that BCAAs can escape catabolism in liver because of low activity of BCAA transferases and inefficient uptake of BCAAs in liver.<sup>20,22,23</sup> Other compounds that were abundant in liver could also be a reflection of those importance in liver; lactic acid is a precursor for gluconeogenesis which occurs in liver,<sup>24</sup> taurine one of the abundant amino acids with diverse physiological functions in liver,<sup>25</sup> and glucuronic acid a compound that is used in glucuronidation in liver.<sup>26</sup> On the contrary, hypotaurine was depleted in the liver but enriched in the pancreas. High level of hypotaurine biosynthesis occurs in pancreas in mouse.<sup>27</sup>

**PyINETA Also Revealed the Distribution of Unknown Compounds Originating from the Diet.** PyINETA was useful even when compounds are not in the database. Out of 67 networks, 21 did not match any of the entries in BMRB. Even under that situation, we were able to track those compounds, revealing those backbone structures and distribution in different tissues (Figure 3 and Table 1). For example, Uk-16 is a compound that is not in the database, but PyINETA extracted its



backbone structure and chemical shift information. Because INADEQUATE and JRES are already linked by PyINETA, we were able to trace this unknown compound using JRES and revealed the distribution in different tissues. Uk-16 was exclusively enriched in liver compared with other tissues, indicating that this is a compound that is actively processed in liver but is not in the BMRB.

In NMR metabolomics, database matching is primarily focused on peaks that match database compounds and peaks that did not match database compounds are usually not retained. On the other hand, PyINETA treats matched and unmatched networks equally and provides structural information. Since PyINETA has a capability of adding new entries to the internal database, users can make use of the obtained knowledge on unknown compounds in future studies.

Tracking metabolites using stable isotopes to understand metabolic pathway and flux has been an active field since its establishment.<sup>11</sup> Despite the success and tremendous value of this approach to track targeted compounds,<sup>12</sup> investigating unknown compounds in this framework is a laborious task, and effort has been made to develop untargeted approaches regardless of analytical platform.<sup>28</sup> PyINETA is capable of handling unknown compounds and can contribute to tackling this challenge in this field.

**Practical Aspects of the PyINETA Pipeline in NMR Metabolomics.** <sup>13</sup>C NMR experiments that directly detect the carbon backbones of molecules, including INADEQUATE, rely on <sup>13</sup>C–<sup>13</sup>C couplings. Although collecting INADEQUATE spectra from natural abundance samples has been used in natural products chemistry, this approach requires a large amount of sample and is not practical in metabolomics. Because of this reason, a more practical strategy one can take is isotopic enrichment. Isotopic enrichment is an established approach in metabolomics for organisms such as bacteria,<sup>6,29</sup> nematodes,<sup>7</sup> terrestrial plants,<sup>4</sup> and marine phytoplankton.<sup>9</sup> Although there are sample types that are more challenging including human studies, tracking metabolites using isotope tracers in cell cultures or tissue slices is a regular approach.<sup>11–13</sup> This field has dedicated to developing tools to analyze backbone carbons for labeled samples; they include, but not limited to, the extraction of backbone spin systems using a <sup>13</sup>C–<sup>13</sup>C constant-time total correlation spectroscopy (CT-TOCSY) experiment,<sup>6</sup> introduction of indirect covariance processing to improve backbone information in CT-TOCSY,<sup>30</sup> and the development of a web-based query system to obtain compound names based on CT-TOCSY (TOCCATA, TOCSY Customized Carbon Trace Archive).<sup>29</sup> They also include the introduction of INADEQUATE to metabolomics samples to extract networks of backbone carbons unambiguously.<sup>7</sup> PyINETA can be practically used for systems in metabolomics. Complementing the previous valuable tools, PyINETA provides an open-source environment for structural information and a way to integrate information of <sup>13</sup>C experiments, as was seen in our model mouse experiment.

**Ensuring the Reproducibility of Compound Annotation in Metabolomics.** PyINETA keeps track of all of the parameters used in the pipeline as a configuration file. Also, the results from individual steps are saved as pickle files. Those pickle files contain all the information on the PyINETA class that is required to reproduce the results. Because of this system, the annotation information from PyINETA is completely reproducible. Users can also deposit those files to databases

such as Metabolomics Workbench<sup>31</sup> along with original data to ensure the reproducibility of compound annotation in a study.

## CONCLUSIONS

PyINETA removes current stumbling blocks in the field of metabolomics, making the best use of <sup>13</sup>C NMR and improving the transparency and reproducibility of compound annotation. In addition to the example study presented here, PyINETA can be expanded to any system that can be labeled to address specific questions in various research fields. PyINETA is installed on NMRbox<sup>17</sup> and publicly available.

## ASSOCIATED CONTENT

### Data Availability Statement

All the raw NMR data, NMRPipe processing scripts, processed data, values for Figure 3, PyINETA output files, and MATLAB scripts used in this study are deposited to Metabolomics Workbench with Study ID ST003304.

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.4c03966>.

Additional information for the pipeline; biological experiment; NMR data collection and processing; and results including the original spectra and statistical analysis (PDF)

List of compounds in the PyINETA database, including corresponding BMRB entry identifications and ambiguity scores (XLSX)

## AUTHOR INFORMATION

### Corresponding Authors

Mario Uchimiyama — Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, United States; [orcid.org/0000-0003-4802-1428](https://orcid.org/0000-0003-4802-1428); Email: [mario.uchimiyama@uga.edu](mailto:mario.uchimiyama@uga.edu)

Arthur S. Edison — Institute of Bioinformatics, University of Georgia, Athens, Georgia 30602, United States; Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, United States; Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602, United States; [orcid.org/0000-0002-5686-2350](https://orcid.org/0000-0002-5686-2350); Email: [aedison@uga.edu](mailto:aedison@uga.edu)

### Authors

Rahil Taujale — Institute of Bioinformatics, University of Georgia, Athens, Georgia 30602, United States; Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, United States

Chaevien S. Clendinen — The Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99354, United States

Ricardo M. Borges — Instituto de Pesquisas de Produtos Naturais, Universidade Federal do Rio de Janeiro, 21941-902 Rio de Janeiro, RJ, Brazil

Christoph W. Turck — Max Planck Institute of Psychiatry, Proteomics and Biomarkers, 80804 Munich, Germany; Key Laboratory of Animal Models and Human Disease Mechanisms of Yunnan Province, and KIZ/CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; National Resource Center for Non-human Primates, and National



Research Facility for Phenotypic & Genetic Analysis of Model Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650107, China

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.analchem.4c03966>

### Author Contributions

◆ R.T. and M.U. contributed equally to this work. A.S.E. and C.T. designed the study; R.T. developed PyINETA; C.W.T. supervised mouse experiments and sample collection; C.S.C. prepared the mouse samples for NMR; M.U. conducted the NMR experiments; M.U., R.T., R.M.B., and A.S.E. analyzed the data. M.U., R.T., and A.S.E. wrote the manuscript with all authors' input.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank John Glushka for providing suggestions on INADEQUATE and JRES experiments and Laura Morris and NMRbox staff members for installing PyINETA on NMRbox. This project was supported by NSF Network for Advanced NMR (NAN) (1946970) and an NIH MIRA award 5R35GM148240 awarded to A.S.E. and the Max Planck Society (C.W.T.).

## REFERENCES

- (1) Bingol, K.; Bruschweiler, R. *Anal. Chem.* **2014**, *86* (1), 47–57.
- (2) Bingol, K.; Li, D. W.; Zhang, B.; Bruschweiler, R. *Anal. Chem.* **2016**, *88* (24), 12411–12418.
- (3) Edison, A. S.; Le Guennec, A.; Delaglio, F.; Kupce, E. Practical Guidelines for <sup>13</sup>C-Based NMR Metabolomics. In *NMR-Based Metabolomics*, Methods in Molecular Biology; Springer, 2019; Vol. 2037, pp 69–95.
- (4) Clendinen, C. S.; Stupp, G. S.; Ajredini, R.; Lee-McMullen, B.; Beecher, C.; Edison, A. S. *Front. Plant Sci.* **2015**, *6*, No. 611.
- (5) Clendinen, C. S.; Lee-McMullen, B.; Williams, C. M.; Stupp, G. S.; Vandenberg, K.; Hahn, D. A.; Walter, G. A.; Edison, A. S. *Anal. Chem.* **2014**, *86* (18), 9242–9250.
- (6) Bingol, K.; Zhang, F.; Bruschweiler-Li, L.; Bruschweiler, R. *J. Am. Chem. Soc.* **2012**, *134* (21), 9006–9011.
- (7) Clendinen, C. S.; Pasquel, C.; Ajredini, R.; Edison, A. S. *Anal. Chem.* **2015**, *87* (11), 5698–5706.
- (8) Bax, A.; Freeman, R.; Kempell, S. P. *J. Am. Chem. Soc.* **1980**, *102* (14), 4849–4851.
- (9) Uchimiya, M.; Olofsson, M.; Powers, M. A.; Hopkinson, B. M.; Moran, M. A.; Edison, A. S. *J. Magn. Reson.* **2023**, *347*, No. 107365.
- (10) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Wenger, R. K.; Yao, H. Y.; Markley, J. L. *Nucleic Acids Res.* **2007**, *36*, D402–D408.
- (11) Bartman, C. R.; Faubert, B.; Rabinowitz, J. D.; Deberardinis, R. J. *Nat. Rev. Cancer* **2023**, *23* (12), 863–878.
- (12) Lin, P. H.; Lane, A. N.; Fan, T. W. M. Stable Isotope-Resolved Metabolomics by NMR. In *NMR-Based Metabolomics: Methods and Protocols*, Methods in Molecular Biology; Springer, 2019; Vol. 2037, pp 151–168.
- (13) Hattori, A.; Tsunoda, M.; Konuma, T.; Kobayashi, M.; Nagy, T.; Glushka, J.; Tayyari, F.; Cskimming, D. M.; Kannan, N.; Tojo, A.; Edison, A. S.; Ito, T. *Nature* **2017**, *545* (7655), 500–504.
- (14) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A. N.; Lindon, J. C.; Marriott, P.; Nicholls, A. W.; Reilly, M. D.; Thaden, J. J.; Viant, M. R. *Metabolomics* **2007**, *3* (3), 211–221.
- (15) Powers, R.; Andersson, E. R.; Bayless, A. L.; Brua, R. B.; Chang, M. C.; Cheng, L. L.; Clendinen, C. S.; Cochran, D.; Copié, V.; Cort, J. R.; Crook, A. A.; Eghbalnia, H. R.; Giacalone, A.; Gouveia, G. J.; Hoch, J. C.; Jeppesen, M. J.; Maroli, A. S.; Merritt, M. E.; Pathmasiri, W.; Roth, H. E.; Rushin, A.; Sakallioglu, I. T.; Sarma, S.; Schock, T. B.; Sumner, L. W.; Takis, P.; Uchimiya, M.; Wishart, D. S. *TrAC, Trends Anal. Chem.* **2024**, *171*, No. 117478.
- (16) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6* (3), 277–293.
- (17) Maciejewski, M. W.; Schuyler, A. D.; Gryk, M. R.; Moraru, I. I.; Romero, P. R.; Ulrich, E. L.; Eghbalnia, H. R.; Livny, M.; Delaglio, F.; Hoch, J. C. *Biophys. J.* **2017**, *112* (8), 1529–1534.
- (18) Ulrich, E. L.; Baskaran, K.; Dashti, H.; Ioannidis, Y. E.; Livny, M.; Romero, P. R.; Maziuk, D.; Wedell, J. R.; Yao, H.; Eghbalnia, H. R.; Hoch, J. C.; Markley, J. L. *J. Biomol. NMR* **2019**, *73* (1–2), 5–9.
- (19) Dashti, H.; Westler, W. M.; Markley, J. L.; Eghbalnia, H. R. *Sci. Data* **2017**, *4*, No. 170073.
- (20) Felig, P. *Annu. Rev. Biochem.* **1975**, *44*, 933–955.
- (21) Bröer, S.; Bröer, A. *Biochem. J.* **2017**, *474*, 1935–1963.
- (22) Paulusma, C. C.; Lamers, W. H.; Broer, S.; van de Graaf, S. F. *J. Biochem. Pharmacol.* **2022**, *201*, No. 115074.
- (23) Bifari, F.; Nisoli, E. *Br. J. Pharmacol.* **2017**, *174* (11), 1366–1377.
- (24) Gerich, J. E.; Meyer, C.; Woerle, H. J.; Stumvoll, M. *Diabetes Care* **2001**, *24* (2), 382–391.
- (25) Miyazaki, T.; Matsuzaki, Y. *Amino Acids* **2014**, *46* (1), 101–110.
- (26) Yang, G. Y.; Ge, S. F.; Singh, R.; Basu, S.; Shatzer, K.; Zen, M.; Liu, J.; Tu, Y. F.; Zhang, C. N.; Wei, J. B.; Shi, J.; Zhu, L. J.; Liu, Z. Q.; Wang, Y.; Gao, S.; Hu, M. *Drug Metab. Rev.* **2017**, *49* (2), 105–138.
- (27) Yoon, S. J.; Combs, J. A.; Falzone, A.; Prieto-Farigua, N.; Caldwell, S.; Ackerman, H. D.; Flores, E. R.; DeNicola, G. M. *Cancer Res.* **2023**, *83* (9), 1426–1442.
- (28) Huang, X. J.; Chen, Y. J.; Cho, K.; Nikolskiy, I.; Crawford, P. A.; Patti, G. J. *Anal. Chem.* **2014**, *86* (3), 1632–1639.
- (29) Bingol, K.; Zhang, F. L.; Bruschweiler-Li, L.; Bruschweiler, R. *Anal. Chem.* **2012**, *84* (21), 9395–9401.
- (30) Zhang, F. L.; Bruschweiler-Li, L.; Bruschweiler, R. *J. Magn. Reson.* **2012**, *225*, 10–13.
- (31) Sud, M.; Fahy, E.; Cotter, D.; Azam, K.; Vadivelu, I.; Burant, C.; Edison, A.; Fiehn, O.; Higashi, R.; Nair, K. S.; Sumner, S.; Subramaniam, S. *Nucleic Acids Res.* **2016**, *44* (D1), D463–D470.