

In-source CID ramping (InCIDR) and Co-variant ion analysis of hydrophilic interaction chromatography (HILIC) metabolomics

Xiaoyang Su,^{*,1,2} Eric Chiles,² Sara Maimouni,^{2,3} Fredric E. Wondisford,¹ Wei-Xing Zong,^{2,3} Chi Song^{*,4}

1. Department of Medicine, Rutgers-Robert Wood Johnson Medical School, New Brunswick, NJ, 08901, USA
2. Rutgers Cancer Institute of New Jersey, New Brunswick, NJ 08903, USA
3. Department of Chemical Biology, Ernest Mario School of Pharmacy, Rutgers University, Piscataway, New Jersey 08854, USA.
4. Division of Biostatistics, College of Public Health, The Ohio State University, Columbus, OH 43210, USA

Abstract

A large proportion of the complexity and redundancy of LC-MS metabolomics data comes from adduct formation. To reduce such redundancy, many tools have been developed to recognize and annotate adduct ions. These tools rely on pre-defined adduct lists which are learned empirically from reverse phase LC-MS studies. Meanwhile, hydrophilic interaction chromatography (HILIC) is gaining popularity in metabolomics studies due to better performance on polar compounds. HILIC methods typically use high concentration of buffer salts for improved chromatography performance. It is therefore necessary to analyze the adduct formation in HILIC metabolomics. To this end, we developed co-variant ion analysis (COVINA) to investigate the metabolite adduct formation. Using this tool, we completely annotated 201 adduct and fragment ions of 10 metabolites. Many of the metabolite adduct ions are found to contain cluster ions of mobile phase additives. We further utilized COVINA to find the major ionization forms of metabolites. Our results show that for some metabolites the adduct ion signals can be >200-fold higher than the deprotonated form, offering better sensitivity for targeted metabolomics analysis. Finally, we developed the in-source CID ramping (InCIDR) method to analyze the intensity changes of the adduct and fragment ions of the metabolites. Our analysis demonstrates a promising method to distinguish the protonated/deprotonated ions of the metabolites from the adduct and fragment ions.

Introduction

LC-MS-based metabolomics aims at a comprehensive characterization of small molecule metabolites in biological samples. Modern mass spectrometry can offer both high sensitivity and high mass resolution, making it possible to detect hundreds or thousands of metabolites from a single biological sample.¹⁻³ As a result, the raw data of LC-MS are highly complex. A typical 25-minute LC-MS run on an Orbitrap instrument may contain 4000 spectra, each has >500-1000 peaks in the mass-to-charge ratio (m/z) domain. To reduce the data complexity, untargeted metabolomics data analysis starts with the feature detection.^{4,5} Each LC-MS feature is a combination of an accurate m/z and a retention time. The features are expected to have a decent chromatographic peak shape in the time domain, and may correspond to actual metabolites, isotopic peaks, dimer ions of metabolites, adduct ions, in-source fragment ions or background matrix components. To facilitate the downstream metabolite identification and statistical analysis, this feature list needs to be optimized and annotated. The optimization step removes the duplicate and false positive features.⁶ The annotation step recognizes and annotates the adduct ions and

natural isotopic ions.⁷⁻⁹ These data processing steps dramatically reduce the complexity of the metabolomics data. Mahieu et al. reported the analysis of an *E. coli*. metabolomics dataset in which 25,000 features were reduced to fewer than 1,000 unique metabolites.¹⁰ This example highlights the complexity and redundancy of a typical metabolomics dataset.

The majority of adduct ion annotation tools fall into two categories. One method is to calculate the correlation of peak intensities across samples from the feature list.^{6,8,11,12} The other approach is to read the extracted ion chromatograms (XICs) in the raw data and calculate the correlation of chromatographic peaks.^{7,9} Even though each of the tools has demonstrated its utility, the adduct ions in a metabolomics dataset are often insufficiently annotated. Glutamate, for example, was reported to produce more than 100 spectral peaks, but many of the complex adducts did not have chemical annotation.¹⁰ Lin. et al used isotope labeled samples to facilitate the annotation of metabolites and their adduct ions. They observed ions which have a large m/z but very few carbon atoms. These ions were considered unreasonable and were filtered out in the annotation. In fact, these ions indicated the existence of large adducts.¹² It is noteworthy that many adduct annotation tools rely on a pre-defined list to identify the adduct ions.^{13,14} The incompleteness of adduct lists limit the performance of the adduct annotation tools. This limitation is further amplified in HILIC ESI-MS metabolomics studies. In recent years, hydrophilic interaction liquid chromatography (HILIC) is becoming a popular technique for metabolomics due to its resolving power of very polar cellular metabolites.¹⁵ HILIC typically involves the use of buffer salts to improve separation and peak shape.^{16,17} These salts have significant impact on the ionization of the metabolite molecules. Erngren et al. reported adduct ion formation that contains multiple sodium and potassium formate moieties.¹⁸ These cluster ions are not included in the common adduct lists, making the adduct annotation difficult. We aim to further investigate the adduct ion formation under HILIC condition.

In this work, we set out to analyze the adduct and fragment ion formation of metabolites in HILIC ESI-MS, for which we developed an algorithm named co-variant ion analysis (COVINA). Unlike existing tools that search co-eluting peaks in the extracted feature list, COVINA directly reads mzXML files. By doing so, COVINA can avoid false positives and false negatives in the feature extraction and mis-grouping in the peak alignment and grouping steps. Using this tool, we found a number of cluster adduct ions of the metabolites. To help assign the chemical identity of these adduct ions, the HILIC mobile phase was prepared with ²H-labeled acetic acid. Using this method, we have completely annotated 201 adduct and fragment ions resulted from 10 metabolites. We then used this new HILIC adduct list to annotate untargeted metabolomics datasets using existing tools and observed improved performance.

We also developed the In-source collision induced dissociation (CID) ramping (InCIDR) method to help determine the neutral molecular mass of the metabolites. Adduct annotation tools implement heuristic algorithms to score hypotheses on metabolite neutral masses based on the detection of the expected adduct ions.^{9,19} InCIDR is a completely orthogonal approach that utilizes COVINA to analyze the intensity changes of the adduct and fragment ions under increasing in-source CID energy levels. In general, under a higher in-source CID energy level, the fragment ions will increase in intensity while the adduct ions will decrease in intensity. InCIDR monitors all the co-variant ions and detects such patterns to score the neutral mass hypothesis. Examples show that InCIDR is a robust and promising means of supplementing existing adduct annotation tools.

Experimental

Chemicals. LCMS-grade methanol (A456), acetonitrile (A955), acetic acid (A35) and water (ACROS 61515) were purchased from Fisher Chemicals (Pittsburgh PA). The $^2\text{H}_4$ -acetic acid (99.5% isotopic purity, DLM-12) was purchased from Cambridge Isotope Laboratories (Tewksbury, MA). Other chemicals including ammonium hydroxide (09859), sodium pyruvate (P2256), sodium L-lactate (L7022), leucine (L8912), isoleucine (I5281), NAD hydrate (N7004), ATP disodium salt hydrate (A2383), D-glucose-6-phosphate dipotassium salt hydrate (G7375), D-fructose-6-phosphate disodium salt hydrate (F3726), L-malic acid (112577) and D-fructose-1,6-bisphosphate trisodium salt hydrate (F6803) were purchased from MilliporeSigma (Burlington, MA).

Cell lysate preparation. L3.6 human pancreatic ductal adenocarcinoma cell line was obtained from ATCC. These cells were seeded in 100 mm dishes and were cultured in RPMI supplemented with 10% bovine serum (Fetal Clone III, HyClone), 100 units/ml penicillin and 100 $\mu\text{g}/\text{ml}$ streptomycin. Cells were grown to ~85% confluency for metabolite extraction. Before harvesting the cells, the media was aspirated, then quickly overlaid with 1ml 40:40:20 mixture of methanol:acetonitrile:water with 0.5% (V/V) formic acid. The plates were incubated on ice for 5 min, then 50 μl of 15% NH_4HCO_3 was added to neutralize the formic acid. Cells were scraped into microfuge tubes, and centrifuged for 10 min at 15,000 $\times g$ at 4°C. The supernatant was collected for LC-MS.

Liquid chromatography. The HILIC separation was performed on a Vanquish Horizon UHPLC system (Thermo Fisher Scientific, Waltham, MA) with XBridge BEH Amide column (150 mm \times 2.1 mm, 2.5 μm particle size, Waters, Milford, MA) using a gradient of solvent A (95%:5% H_2O :acetonitrile with 20 mM acetic acid, 40 mM ammonium hydroxide, pH 9.4), and solvent B (20%:80% H_2O :acetonitrile with 20 mM acetic acid, 40 mM ammonium hydroxide, pH 9.4). For the $^2\text{H}_4$ -acetic acid mobile phase experiment, the acetic acid in both mobile phase A and B was replaced with $^2\text{H}_4$ -acetic acid at the same concentration. The gradient was 0 min, 100% B; 3 min, 100% B; 3.2 min, 90% B; 6.2 min, 90% B; 6.5 min, 80% B; 10.5 min, 80% B; 10.7 min, 70% B; 13.5 min, 70% B; 13.7 min, 45% B; 16 min, 45% B; 16.5 min, 100% B and 22 min, 100% B. The flow rate was 300 $\mu\text{l}/\text{min}$. Injection volume was 5 μL and column temperature was 25 °C. The autosampler temperature was set to 4°C and the injection volume was 5 μL .

Mass spectrometry. The mass spectrometry analysis was performed on Thermo Q Exactive PLUS with a HESI source which was set to a spray voltage of -2.7kV under negative mode and 3.5kV under positive mode. The sheath, auxiliary, and sweep gas flow rates of 40, 10, and 2 (arbitrary unit) respectively. The capillary temperature was set to 300°C and aux gas heater was 360°C. The S-lens RF level was 45. The m/z scan range was set to 72 to 1000 m/z under either positive or negative ionization mode. The AGC target was set to 3e6 and the maximum IT was 200ms. The in-source CID energy level was set to 0eV unless otherwise specified. The resolution was set to 70,000 unless using in-source CID ramping. In the data acquisition of InCIDR, each scan cycle is comprised of 8 scan events with in-source CID energy level 0eV, 2eV, 4eV, 6eV, 8eV, 10eV, 15eV and 20eV. InCIDR uses 17.5k resolution to achieve scan rate of 2Hz.

Co-Variant Ion Analysis (COVINA). COVINA is a targeted approach to search for the co-eluting ion species of the metabolite of interest in the raw $mz\text{XML}$ files. When the m/z of a metabolite (base m/z) is specified, COVINA first looks for the scan at the apex of the corresponding XIC (Figure 1A). If multiple peaks exist in one XIC, the scan number can be specified to focus on the metabolite of interest. This selected scan is used as the reference spectrum. COVINA then builds XICs for all the m/z s in the reference spectrum. The Pearson correlation coefficients between the XIC of base m/z and other XICs are calculate and ranked (Figure 1B, Table 1). The adduct, fragment and isotopic ions usually show very good correlation ($R>0.9$)

with the base m/z . These co-variant ions plus the base m/z peak form the pseudo-spectrum of the metabolite (Figure 1C). All XICs in COVINA use scan number instead of retention time as the independent variable.

Table 1. Pseudocode for COVINA

Algorithm COVINA	
1:	procedure COVINA(input parameters: <i>mzXML file(s)</i> , <i>m/z</i> ; optional parameters: <i>scan.number</i> ,
2:	<i>peak.width</i> , <i>mass.tolerance</i> , <i>intensity.threshold</i> , <i>correlation.threshold</i>)
3:	Use the first <i>mzXML file</i> to do
4:	<i>XIC.Base</i> \leftarrow XIC of <i>m/z</i> within <i>mass.tolerance</i>
5:	if <i>scan.number</i> is not specified then
6:	<i>scan.number</i> \leftarrow scan with highest signal in <i>XIC.Base</i>
7:	end if
8:	<i>scan.range</i> \leftarrow <i>scan.number</i> \pm <i>peak.width</i>
9:	<i>mz.table</i> \leftarrow all <i>m/z</i> values in the <i>scan.number</i> -th scan that are above
10:	<i>intensity.threshold</i>
11:	<i>Base.Chromatogram</i> \leftarrow <i>XIC.Base</i> within <i>scan.range</i>
12:	for each value in the <i>mz.table</i>
13:	<i>Query.Chromatogram</i> \leftarrow corresponding XIC within <i>mass.tolerance</i> and
14:	<i>scan.range</i>
15:	if correlation between <i>Base.Chromatogram</i> and <i>Query.Chromatogram</i> >
16:	<i>correlation.threshold</i> then
17:	add this <i>m/z</i> value to the <i>co-variant.ion.list</i>
18:	end if
19:	end for
20:	end do
21:	In each <i>mzXML file</i> do
22:	<i>co-variant.ion.intensities</i> \leftarrow Integrated peak area for each value in the <i>co-</i>
23:	<i>variant.ion.list</i> within <i>scan.range</i>
24:	end do
25:	return <i>co-variant.ion.list</i> , <i>co-variant.ion.intensities</i>
26:	end procedure

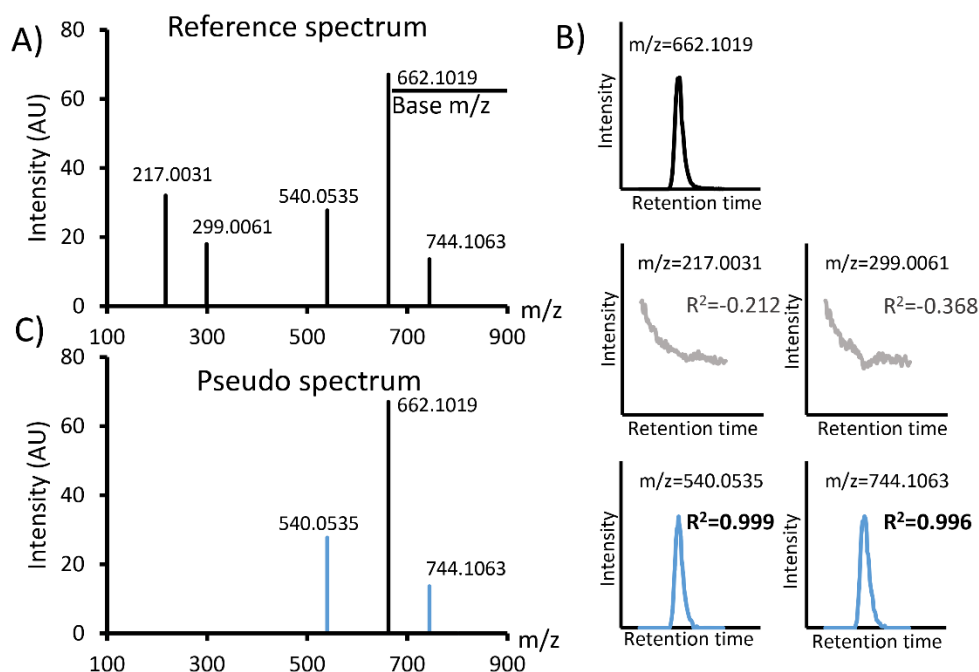


Figure 1. Co-variant ion analysis (COVINA). (A) The reference mass spectrum at the apex of the XIC of the base m/z . (B) The XICs of the base m/z and all other m/z in the reference spectrum. (C) COVINA generates a pseudo-spectrum containing the co-variant ions of the base m/z which are colored in blue.

Code and example data availability. COVINA is implemented in R (version 3.6.1). The code as well as the sample data is available at <https://github.com/XiaoyangSu/COVINA>.

Results and Discussion

Investigating adduct and fragment ion formation using COVINA

A common approach to studying adduct formation is to inject the chemical standard to the mass spectrometer from a syringe pump, either coupled to the normal LC flow or not. This method is conceptually simple, but the adduct signals are mingled with background signals, making complete adduct annotation difficult.¹⁸ Instead, we utilize chromatography to analyze ion species that are co-eluted with the metabolite of interest. The adduct ions and fragment ions should be chromatographically correlated with their parental protonated/deprotonated metabolite ions. COVINA detects such co-elution correlation to filter out persistent background ions and generates high-quality co-variant ion lists.

Another approach used by many common adduct annotation tools is to group related m/z s in the feature lists. The feature lists used by these annotation tools are generated from the peak picking algorithm such as centWave⁵ which produces thousands of features from a typical dataset. Despite significant improvement and refinement of the grouping algorithms, the feature grouping is usually insufficient to put all of the adduct and fragment ions in the same group with the parental metabolite ion. We have tested the grouping performance of CAMERA⁷ and ClueMS⁹. Both tools put the adduct ions resulted from 10 metabolites into more than 20 different groups (Figure S1). In order to systematically study the

patterns of metabolite adduct and fragment formation, it is necessary to optimize the detection of adducts and fragments in the data analysis process.

Unlike other annotation tools, COVINA does not use a feature list generated from a peak picking algorithm. Instead, COVINA takes the input of the m/z value of a specific metabolite (base m/z) and builds the XICs of this m/z and all other m/z in the same spectrum, and calculates the chromatographic correlation between the XIC of base m/z and other XICs. In the output data, COVINA gives a list of highly correlated m/z that includes adduct and fragment ions and isotopic peaks.

Using our COVINA algorithm, we studied the adduct formation of some key metabolites under negative ionization mode. COVINA revealed multiple adducts in each of these metabolites in our chemical standard mixture. For example, Figure 2A is showing the adduct occurrence of pyruvate, lactate, leucine, malate and glucose-6-phosphate (Glc6P). Interestingly, many adducts are showing common m/z differences ($\Delta m/z$, m/z of an adduct – m/z of $[M-H]^-$), suggesting these adducts have the same chemical identities. Previous studies have shown that metabolites may form oligomer adducts ($[2M-H]^-$ or $[3M-H]^-$) or heterodimer adducts ($[M1+M2-H]^-$).²⁰ These complex adducts are less frequently detected in our samples than the common adducts. The commonly observed $\Delta m/z$, 82.003 is an important one. This $\Delta m/z$ is the only one observed in all five metabolites. Moreover, 82.003 is the common difference in some series of $\Delta m/z$, such as 82.003 / 164.006 / 246.009, 97.968 / 179.972 / 261.973 and 157.989 / 239.992 / 321.995 / 403.999 (Figure 2A). This observation suggests that some large adducts such as 321.995 and 403.999 may have repeating units of 82.003. We also calculated the pairwise m/z differences of all the co-variant ions detected in our standard mixture samples. The histogram of these $\Delta m/z$ confirmed that 82.003 is the most common mass difference among co-variant ions (Figure 2B). Multiples of 82.00, such as 164.01 and 239.99 are also common $\Delta m/z$ values, again suggesting 82.003 represents a repeating unit in the adduct ions.

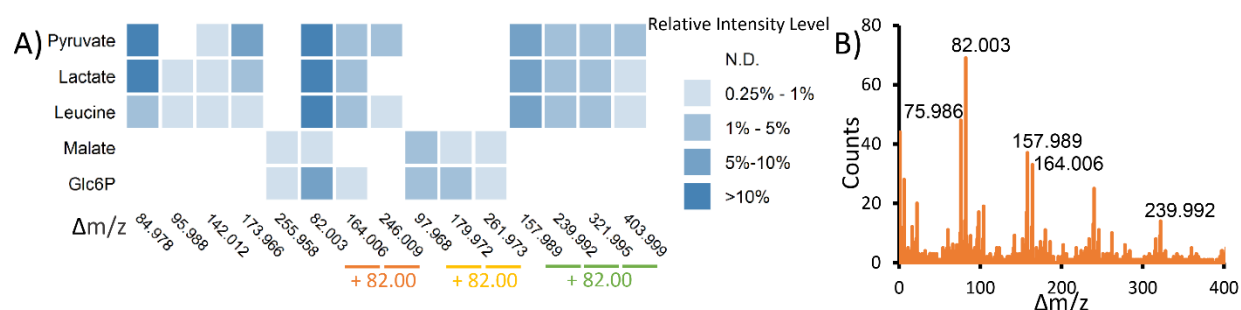


Figure 2. Common adducts found by COVINA. (A) Heatmap showing intensity of the adduct ions relative to the intensity of $[M-H]^-$. The $\Delta m/z$ is calculated as the $m/z(\text{adduct}) - m/z([M-H]^-)$. The $\Delta m/z$ values in the same arithmetic sequence are grouped together. (B) Histogram of $\Delta m/z$ pairwise differences of all the co-variant ions within each metabolite group. The most common $\Delta m/z$ values are highlighted in the plot.

The $\Delta m/z$ 82.003 matches the mass of sodium acetate (NaAc, calculated mass 82.00253). While acetate is used in the mobile phase buffer, the sodium ions were not purposefully added to the mobile phase. Trace amounts of sodium may come from water, glassware or the impurity in the solvent additives. It is a surprising result to see NaAc but not NH_4Ac that presents in many adduct ions. To confirm that NaAc is the repeating unit in many adducts, we prepared mobile phase with deuterium labeled acetate ($[\text{D}_3\text{H}_3]\text{-Ac}^-$). In this labeled mobile phase, many adducts show shifted m/z (Figure 3). Pyruvate has an adduct $[M+157.989]^-$ which shifted to $[M+164.027]^-$ in deuterium labeled mobile phase showing that this adduct

has 2 acetate moieties. Therefore, the chemical identity of this adduct can be determined as $[M+CaAc_2-H]^-$ (Table 2). The high mass accuracy of the Orbitrap mass analyzer lends great confidence to accurate adduct identification. For example, the formula $[M+CaAc_2-H]^-$ matches the detected m/z with a mass accuracy of -1.1 ppm. An alternative formula $[M+KAc_2]^-$ has a mass accuracy of -37.6 ppm and should, therefore, be ruled out. Pyruvate adducts $[M+239.9918]^-$ and $[M+321.9952]^-$ are shifted by 9.056 and 12.075 Da suggesting they have Ac_3 and Ac_4 respectively. They are annotated as $[M+NaCaAc_3-H]^-$ and $[M+Na_2CaAc_4-H]^-$. $[M+173.965]^-$ did not shift in the deuterium labeled mobile phase and, therefore, do not contain an Ac in their adduct moiety. This adduct is annotated as $[M+H_6Si_2O_7-H]^-$. The use of deuterium labeled mobile phase helps to narrow down the possible adduct formulas and leads to confident identification.

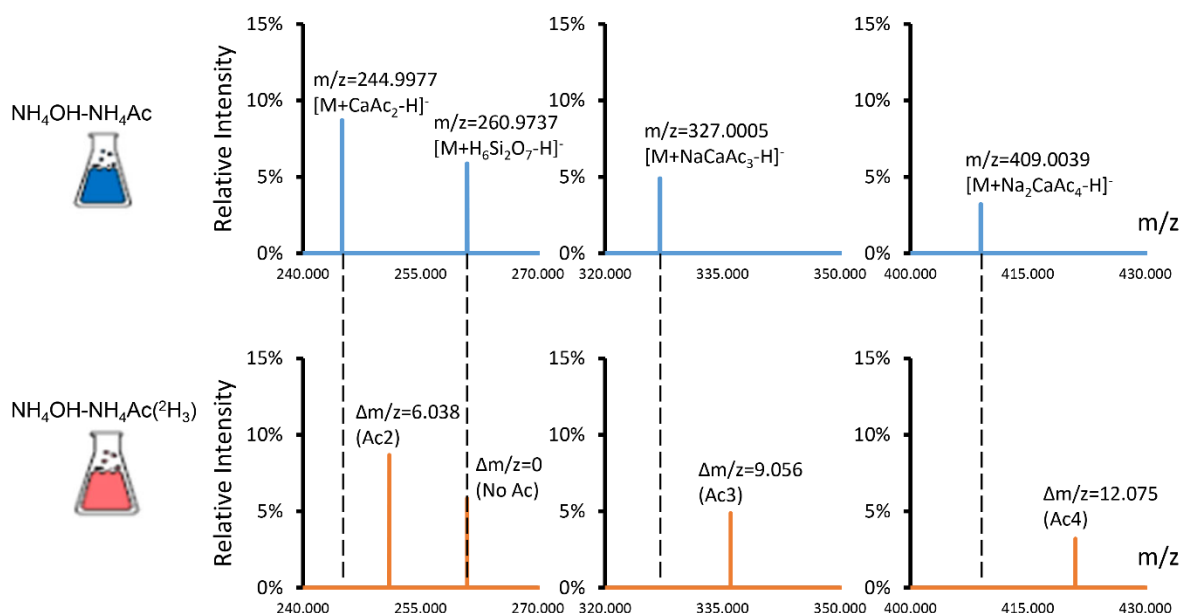


Figure 3. Using ²H-acetic acid to determine the chemical identity of the adducts. The upper panel is showing the adduct ions of pyruvate detected in normal mobile phase. The lower panel is showing the adduct ions of pyruvate detected in mobile phase prepared with [²H₄]-acetic acid. Each acetate moiety in the adduct ion should result in a m/z shift of 3.018. The adduct ions m/z 244.9977, 327.0005 and 409.0039 were shifted by 6.038, 9.056 and 12.075 respectively in the ²H-labeled mobile phase, suggesting that 2, 3 and 4 acetate moieties are in these adduct ions respectively. The m/z 260.9737 did not shift in the ²H-labeled mobile phase.

Similarly to previous reports, complex ions such as adducts of fragments are detected.²¹ For Glc6P, $[M+37.948]^-$ was detected. This m/z matches $[M+Ca-3H]^-$ with a mass accuracy of 2.2 ppm. However, this adduct shifted to $[M+40.965]^-$ in the deuterated mobile phase, indicating 1 Ac group. This adduct is essentially an adduct of erythrose-4-phosphate ($[E4P+CaAc-2H]^-$), a known fragment of Glucose6P.²² Another example of adduct of fragment is $[NAD+57.921]^-$. The $\Delta m/z$ 57.921 does not match any chemical formula as a simple adduct. Using the deuterated mobile phase, it was demonstrated that this ion has 2 Ac groups, and is annotated as $[NAD-nicotinamide+NaCaAc_2-2H]^-$. The combined use of the COVINA and the deuterated mobile phase has led to the complete annotation of 201 ion co-variants produced from 10 metabolites (co-variant ions of lactate are shown in Table 2, see Table S1 for the complete list). The

complete adduct annotations were facilitated by mixed integer linear programming to find the combination for the cluster ions,²³ and also involved manual inspection and curation.

Table 2. Co-variant ions of lactate

Observed m/z	Relative Abundance	Δm	Chemical Annotation	Calculated m/z	Mass accuracy (ppm)	Observed m/z In ² H ₃ -Ac
89.0244	100.00%	0.0000	[M-H] ⁻	89.0244	-0.7	89.0245
90.0276	3.32%	1.0032	¹³ C ₁ -[M-H] ⁻	90.0278	-2.2	90.0278
91.0289	0.69%	2.0045	¹⁸ O ₁ -[M-H] ⁻	91.0287	2.5	91.0287
148.9905	0.40%	59.9661	[M+SiO ₂ -H] ⁻	148.9912	-4.8	148.9910
171.0276	13.20%	82.0032	[M+NaAc-H] ⁻	171.0275	0.6	174.0462
172.0306	0.56%	83.0062	¹³ C ₁ -[M+NaAc-H] ⁻	172.0309	-1.7	175.0496
174.0020	36.30%	84.9776	[M+NaNO ₃ -H] ⁻	174.0020	0.0	174.0019
175.0061	1.00%	85.9817	¹³ C ₁ -[M+NaNO ₃ -H] ⁻	175.0054	4.0	175.0052
177.9570	0.44%	88.9327	[M+FeO(OH)-H] ⁻	177.9570	0.1	177.9572
185.0120	0.87%	95.9876	[M+H ₄ SiO ₄ -H] ⁻	185.0123	-1.7	185.0123
201.0387	6.43%	112.0143	[2M+Na-2H] ⁻	201.0381	3.1	201.0381
202.0426	0.43%	113.0183	¹³ C ₁ -[2M+Na-2H] ⁻	202.0414	6.0	202.0415
231.0369	0.92%	142.0126	[M+MgAc ₂ -H] ⁻	231.0361	3.7	ND*
247.0135	5.70%	157.9892	[M+CaAc ₂ -H] ⁻	247.0136	-0.4	253.0513
253.0312	1.09%	164.0069	[M+Na ₂ Ac ₂ -H] ⁻	253.0306	2.6	259.0681
262.9891	3.02%	173.9648	[M+H ₆ Si ₂ O ₇ -H] ⁻	262.9896	-1.8	262.9896
313.0396	0.51%	224.0152	[M+NaMgAc ₃ -H] ⁻	313.0391	1.3	ND*
329.0170	2.78%	239.9926	[M+NaCaAc ₃ -H] ⁻	329.0167	0.9	338.0732
411.0196	1.37%	321.9953	[M+Na ₂ CaAc ₄ -H] ⁻	411.0198	-0.4	423.0950
493.0235	0.71%	403.9991	[M+Na ₃ CaAc ₅ -H] ⁻	493.0228	1.3	508.1173

* ND: Not detected. m/z 231.0369 and 313.0396 do not have their counterparts detected in ²H₃-Ac mobile phase. They are annotated as Mg²⁺ containing adduct ions. Presumably the ²H-HAc has much lower level of Mg²⁺ so that these adduct ions were not formed.

Our results suggest the adduct formation in HILIC metabolomics is largely affected by the solvent additives as well as other chemicals present in the system. NO₃⁻ adduct, for example, may come from the nitric acid we used to clean the ion transfer tube of the mass spectrometer. Therefore we recommend using COVINA to generate an adduct list specific to each of the different HILIC conditions and instrument being utilized to run them. Nonetheless, we believe our adduct list is helpful to systems running a similar chromatographic condition to what we have evaluated. Using such adduct lists, we can greatly improve the adduct annotation rate in untargeted metabolomics annotation tools such as CliqueMS. From a L3.6 human pancreatic ductal adenocarcinoma cell lysate sample, CliqueMS annotated 23.5% and 53.5% features as adducts using built-in adduct list and COVINA-generated list respectively (Figure S2). We have also manually inspected and confirmed a number of adduct ions detected from L3.6 human PDAC lysate samples. Our results show [M+NaAc-H]⁻ is more common than [M+Na-2H]⁻ (Figure 4A).

Due to the use of a high concentration of salt in the HILIC metabolomics, some metabolites show higher ion counts in their adduct/fragment forms than in the protonated/deprotonated forms. This information is important especially in targeted analysis of metabolomics data. Using COVINA, we investigated such cases in the L3.6 cell lysates. The signal of Acetyl-CoA in the divalent ion form $[M-2H]^{2-}$ is 8-fold higher than the $[M-H]^-$ form. The signal of phosphocholine is 350-fold higher in its acetate adduct form than the $[M-H]^-$ form. S-Adenosyl-L-Methionine (AdoMet) can be detected in several fragment forms that have signals higher than the $[M-H]^-$ of AdoMet, including methylthioadenosine (MTA, 8-fold higher), MTA acetate adduct (35-fold higher) and adenine (55-fold higher) (Figure 4B). These results provide better detection for key metabolites in the targeted metabolomics analysis workflow.

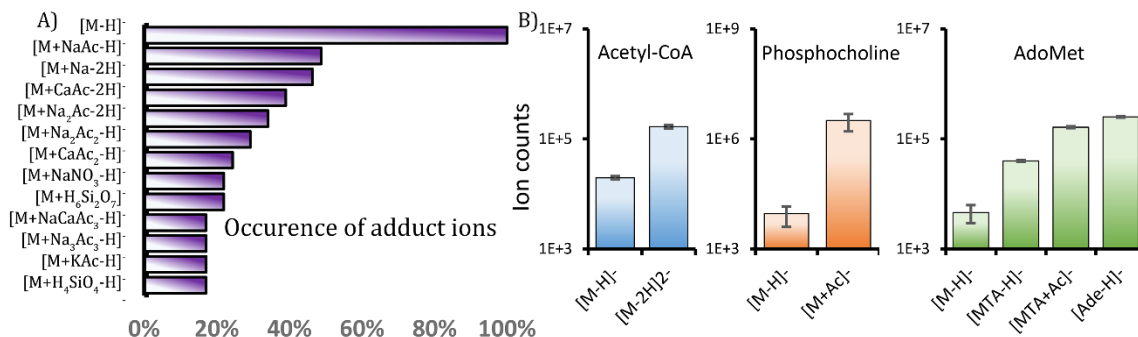


Figure 4. Significant adduct ions and fragment ions of key metabolites in the L3.6 human pancreatic ductal adenocarcinoma cell lysate samples. (A) Comparison of the occurrence of adduct ions. The denominator is number of metabolites with at least one adduct ion detected. (B) Significant co-variant ions that are stronger than the molecular ions. Ion counts are shown as mean \pm standard deviation, n=3.

In essence, COVINA informs us on how metabolite molecules ionize. For instance, if the m/z and retention time of the $[M+H]^+$ are known, we can use COVINA to find all other ionized forms of M. However, not every metabolite has its $[M+H]^+$ form easily detectable. Due to the pKa values, some metabolites are much easier to carry a negative charge than to carry a positive one, such as lactate and pyruvate. Therefore, the LC-MS metabolomics data acquisition is often done under both positive and negative ionization modes to maximize the metabolome coverage. It is possible, however, to detect carboxylic acids such as lactate under positive ionization mode with reasonable sensitivity. We modified COVINA to analyze alternating scans of negative and positive ionization modes (Figure 5A). The XIC of $[M-H]^-$ was used as the reference and all the XICs under negative and positive ionization modes were used to calculate the chromatographic correlation coefficients. Our results show that lactate can be detected under positive ionization mode as $[M-H+2Na]^+$, which has roughly 15% intensity of $[M-H]^-$ under negative mode (Figure 5B).

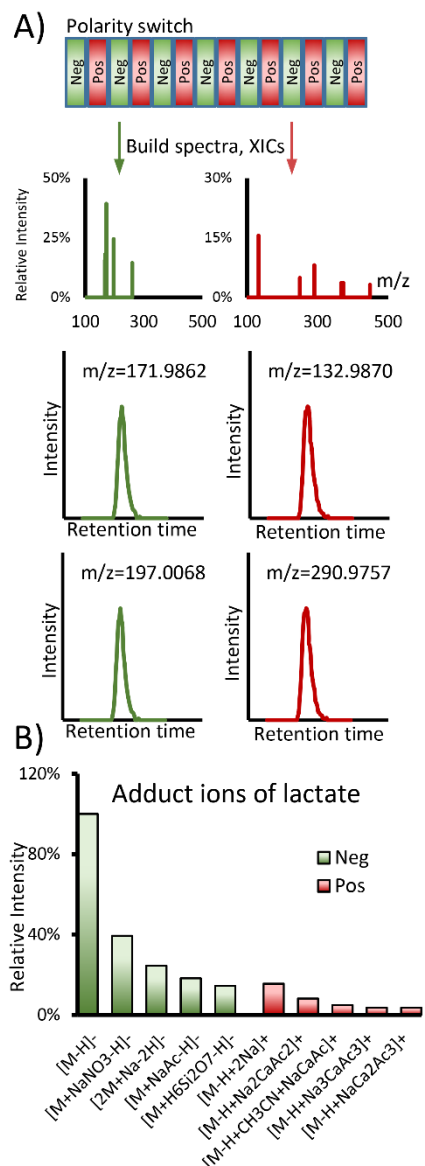


Figure 5. COVINA with polarity switch. (A) Scheme of COVINA under polarity switch mode. The XICs are built separately under each polarity, and the correlation with the reference XIC can be calculated as the standard COVINA method. (B) Example result of lactate adduct ions. Lactate is mainly detected as $[M-H]^-$ under the negative ionization mode. The most abundant cation species of lactate is $[M-H+2Na]^+$ which has a signal that is roughly 15% of the signal of $[M-H]^-$.

Finding the metabolite neutral mass using in-source CID ramping (InCIDR)

For the untargeted metabolomics data processing, the goal of adduct annotation tools is twofold: 1) to identify the adduct ions and exclude them from further analysis and 2) to find the neutral masses of the true metabolites for the downstream structural elucidation and statistical analysis. To find the neutral mass of a metabolite out of many grouped co-eluting ions, the annotation algorithm searches the mass

differences between co-eluting m/z s and matches the gaps to known adducts. The hypotheses of metabolite neutral masses are scored based on the matching results and optionally the empirical frequency of the adducts.^{9,19} From these scores the most plausible neutral mass is returned as the result. Our data suggest that the performance of such algorithms can be improved by having a more complete adduct list generated from COVINA (Figure S3). Nonetheless, because of the heuristic nature of these algorithms, the metabolite neutral mass may not be picked correctly. Other approaches to reinforce the metabolite neutral mass prediction are needed.

Here we propose to use in-source CID ramping (InCIDR) to predict the metabolite neutral mass. Instead of using an empirical adduct list and matching the mass differences, InCIDR predicts the m/z of the protonated or deprotonated metabolite ion by analyzing the co-variant ion intensity change under ramping CID energy levels. InCIDR works on the assumption that when the in-source CID energy level goes higher the adduct ions will collapse and decrease in intensities while the fragment ions will increase in intensities. Lin et al. first described such trends for the fragment ions and used it to filter fragment ions from the feature list.¹² Based on their work, we further extended this analysis to encompass the adduct ions and built the data acquisition and analysis method called In-source CID Ramping (InCIDR). Unlike the original method using different CID energy levels in repeated runs, the data acquisition of InCIDR is a series on alternating scan events with increasing in-source CID energy level (Figure 6A). In this way, there is no need to run the same sample repeatedly at different energy levels, saving instrument time. Second, the single run eliminates the ion intensity variations due to inconsistent injection or chromatography separation. Third, there is no need to group peaks by retention time. Rather, the peaks can be perfectly grouped by matching the scan numbers.

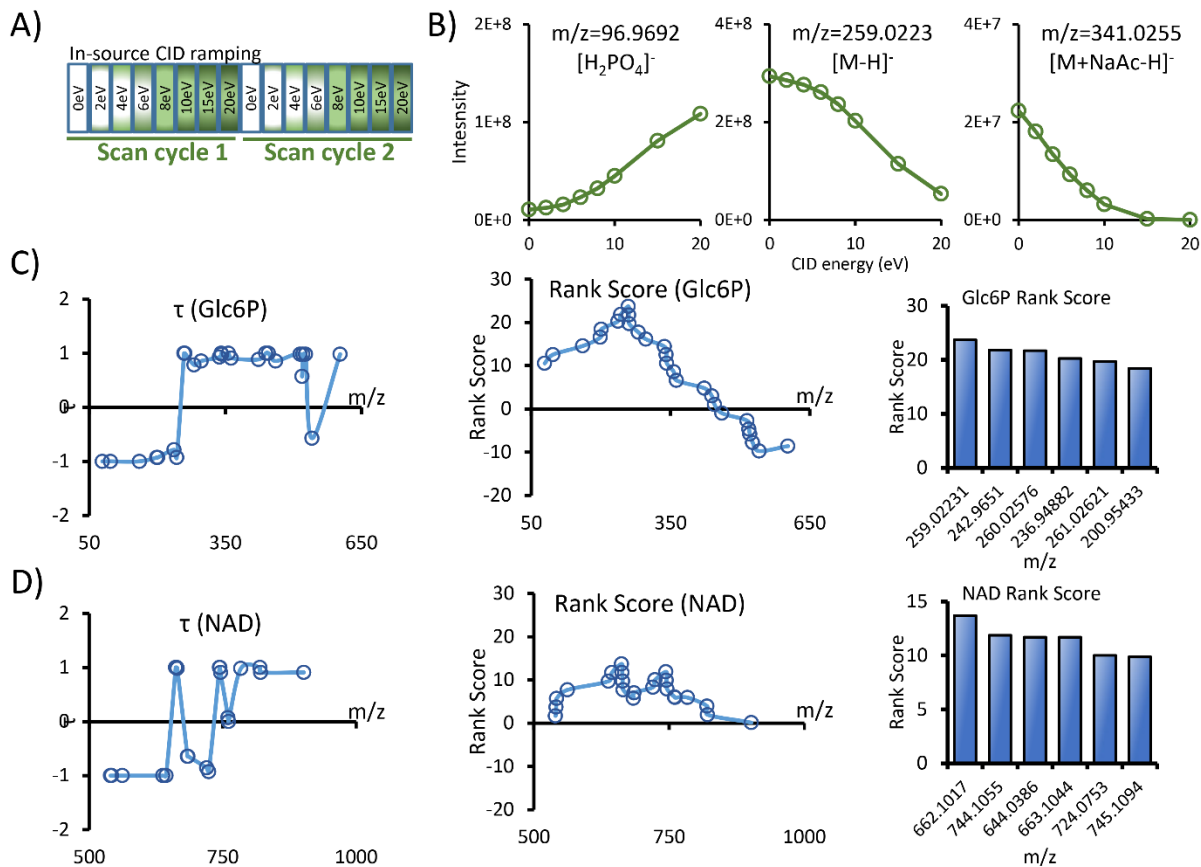


Figure 6. In-source CID ramping (InCIDR). (A) The scan events of InCIDR. We use 8 scan events each cycle to examine the intensity change of the co-variant ions. (B) The trend of intensity change under ramping CID energy levels. (C-D) Example InCIDR results of Glc6P and NAD.

We modified COVINA to work with InCIDR data. Similar to COVINA, InCIDR starts from an m/z of interest and finds all co-variant ions associated with it. This starting m/z can be any m/z in the group of co-variant ions. InCIDR will analyze the intensity of the co-variant ions under all in-source CID energy levels and use the trends to predict which one of them is the protonated or deprotonated ion. To validate the basic assumption of InCIDR, we investigated the co-variant ions of Glc6P. As predicted, the m/z 96.9692, which is the $[H_2PO_4]^-$ fragment of Glc6P, increases in intensity when the CID energy level goes up. The m/z 341.0255, which is the $[M+NaAc-H]^-$ adduct of Glc6P, decreases in intensity when energy level increases. The m/z 259.0223, which is the deprotonated ion of Glc6P also decreases in intensity when energy level increases due to increased fragmentation. To describe the intensity trend in quantitative terms, we calculated the rank correlation coefficient τ for each of the co-variant ion.

$$\tau = \frac{2}{n(n-1)} \sum_{i < j} \text{sgn}(I_i - I_j)$$

In the above equation, n is the total number of the CID energy levels, which is 8 in this study. $\text{sgn}()$ is the sign function that takes value of 1, -1 or 0 if applied to positive number, negative number or 0 respectively. I_i and I_j are the ion counts of the co-variant ion under the i -th and j -th lowest CID energy level, $1 \leq i < j \leq n$. An adduct that strictly decreases in intensity under higher energy levels gets $\tau=1$. A fragment ion that strictly

increases in intensity under higher energy level gets $\tau = -1$. For Glc6P, all co-variant ions with m/z smaller than $[M-H]^-$ have negative values of τ . All ions but 1 with m/z larger than $[M-H]^-$ have positive values of τ . Therefore, it is possible to use the patterns in τ values to find the m/z of $[M-H]^-$ out of all the co-variant ions. We can calculate the rank score for all the co-variant ions of Glc6P. For a specific m/z , the rank score is the sum of τ values of ions above this m/z minus the sum of τ values of ions below this m/z .

$$\text{Rank Score}(m/z) = \sum \tau_{\geq m/z} - \sum \tau_{< m/z}$$

We predict the $[M-H]^-$ should have the highest rank score out of all the co-variant ions. Indeed, the $[M-H]^-$ 259.0223 showed the highest rank score. In another example, we investigated the co-variant ions of NAD using InCIDR. We observed abnormal adduct ions that increase in intensity when ramping up the in-source CID energy level, such as $[M+Na-2H]^-$ and $[M\text{-Nicotinamide}+NaCaAc_2-2H]^-$. The possible reason for the $[M+Na-2H]^-$ adduct ions increasing in intensity is that some even larger adduct ions decompose into this sodium adduct. The $[M\text{-Nicotinamide}+NaCaAc_2-2H]^-$ is an adduct ion of a fragment of NAD, so the higher in-source CID energy level may promote the NAD fragmentation and formation of this ion. Nonetheless, the rank scores of the co-variant ions showed m/z 662.1017 has the highest one, which is $[M-H]^-$ of NAD. The rank correlation coefficients and the rank scores for Glc6P and NAD are shown in Table S2. Using these two examples, we show that the calculation of rank correlation coefficients and the rank scores gives InCIDR robustness in finding the deprotonated and protonated ions and therefore the neutral mass of the metabolites.

Conclusion

In this work, we investigated the metabolite ionization forms in HILIC-ESI-MS metabolomics. Using COVINA and a stable isotope labeled mobile phase, we have detected and annotated a number of adduct ions and fragment ions of the metabolites. HILIC metabolomics datasets contain a large number of adduct ions, especially cluster adduct ions. A complete HILIC adduct list greatly improves the performance of existing annotation tools such as CliqueMS. We also investigated the intensity changes of adduct and fragment ions using InCIDR. We show that the prediction of the metabolite neutral mass is possible by scoring the intensity trends of the co-variant ions under ramping in-source CID energy levels.

Acknowledgement

We thank members of the Rabinowitz lab at Princeton University for suggestions and constructive discussion. This research is supported by NIH grants P30 CA072720, and R01CA224550, R01CA129536, R01CA232246 (to W.X.Z.). Chi Song was partially supported by NIH/NCATS grant UL1TR002733.

Supporting Information

Figure S1. The grouping performance of the adduct annotation tools

Figure S2. Statistics on adduct annotation

Table S1. The complete annotation of co-variant ions of 10 metabolites in the standard mixture

Table S2. The rank correlation coefficients and the rank scores for Glc6P and NAD calculated by InCIDR

File S1. The extended adduct list for CliqueMS (Negative ionization mode)

File S2. The extended adduct list for CliqueMS (Positive ionization mode)

Author Information

*Corresponding author. Email: xs137@rwjms.rutgers.edu

*Corresponding author. Email: song.1188@osu.edu

Reference

- (1) Fiehn, O. The Link between Genotypes and Phenotypes. *Plant Mol. Biol.* **2002**, *48*(1-2), 155–171.
- (2) Wang, T. J.; Larson, M. G.; Vasan, R. S.; Cheng, S.; Rhee, E. P.; McCabe, E.; Lewis, G. D.; Fox, C. S.; Jacques, P. F.; Fernandez, C.; et al. Metabolite Profiles and the Risk of Developing Diabetes. *Nat. Med.* **2011**, *17* (4), 448–453. <https://doi.org/10.1038/nm.2307>.
- (3) Jang, C.; Chen, L.; Rabinowitz, J. D. Metabolomics and Isotope Tracing. *Cell* **2018**, *173* (4), 822–837. <https://doi.org/10.1016/j.cell.2018.03.055>.
- (4) Smith, C. A.; Want, E. J.; O’Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **2006**, *78* (3), 779–787. <https://doi.org/10.1021/ac051437y>.
- (5) Tautenhahn, R.; Bottcher, C.; Neumann, S. Highly Sensitive Feature Detection for High Resolution LC/MS. *BMC Bioinformatics* **2008**, *9* (1), 504. <https://doi.org/10.1186/1471-2105-9-504>.
- (6) DeFelice, B. C.; Mehta, S. S.; Samra, S.; Čajka, T.; Wancewicz, B.; Fahrman, J. F.; Fiehn, O. Mass Spectral Feature List Optimizer (MS-FLO): A Tool To Minimize False Positive Peak Reports in Untargeted Liquid Chromatography-Mass Spectroscopy (LC-MS) Data Processing. *Anal. Chem.* **2017**, *89* (6), 3250–3255. <https://doi.org/10.1021/acs.analchem.6b04372>.
- (7) Kuhl, C.; Tautenhahn, R.; Böttcher, C.; Larson, T. R.; Neumann, S. CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets. *Anal. Chem.* **2012**, *84* (1), 283–289. <https://doi.org/10.1021/ac202450g>.
- (8) Uppal, K.; Walker, D. I.; Jones, D. P. XMSannotator: An R Package for Network-Based Annotation of High-Resolution Metabolomics Data. *Anal. Chem.* **2017**, *89* (2), 1063–1067. <https://doi.org/10.1021/acs.analchem.6b01214>.
- (9) Senan, O.; Aguilar-Mogas, A.; Navarro, M.; Capellades, J.; Noon, L.; Burks, D.; Yanes, O.; Guimerà, R.; Sales-Pardo, M. CliqueMS: A Computational Tool for Annotating in-Source Metabolite Ions from LC-MS Untargeted Metabolomics Data Based on a Coelution Similarity Network. *Bioinformatics* **2019**, 1–9. <https://doi.org/10.1093/bioinformatics/btz207>.
- (10) Mahieu, N. G.; Patti, G. J. Systems-Level Annotation of a Metabolomics Data Set Reduces 25 000 Features to Fewer than 1000 Unique Metabolites. *Anal. Chem.* **2017**, *89* (19), 10397–10406. <https://doi.org/10.1021/acs.analchem.7b02380>.
- (11) Broeckling, C. D.; Afsar, F. A.; Neumann, S.; Ben-Hur, A.; Prenni, J. E. RAMClust: A Novel Feature Clustering Method Enables Spectral-Matching-Based Annotation for Metabolomics Data. *Anal. Chem.* **2014**, *86* (14), 6812–6817. <https://doi.org/10.1021/ac501530d>.

- (12) Wang, L.; Xing, X.; Chen, L.; Yang, L.; Su, X.; Rabitz, H.; Lu, W.; Rabinowitz, J. D. Peak Annotation and Verification Engine for Untargeted LC-MS Metabolomics. *Anal. Chem.* **2019**, *91* (3), 1838–1846. <https://doi.org/10.1021/acs.analchem.8b03132>.
- (13) Tong, H.; Bell, D.; Tabei, K.; Siegel, M. M. Automated Data Massaging, Interpretation, and e-Mailing Modules for High Throughput Open Access Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **1999**, *10* (11), 1174–1187. [https://doi.org/10.1016/S1044-0305\(99\)00090-2](https://doi.org/10.1016/S1044-0305(99)00090-2).
- (14) Keller, B. O.; Sui, J.; Young, A. B.; Whittall, R. M. Interferences and Contaminants Encountered in Modern Mass Spectrometry. *Analytica Chimica Acta*. 2008, pp 71–81. <https://doi.org/10.1016/j.aca.2008.04.043>.
- (15) Wernisch, S.; Pennathur, S. Evaluation of Coverage, Retention Patterns, and Selectivity of Seven Liquid Chromatographic Methods for Metabolomics. *Anal. Bioanal. Chem.* **2016**, *408* (22), 6079–6091. <https://doi.org/10.1007/s00216-016-9716-4>.
- (16) Teleki, A.; Sánchez-Kopper, A.; Takors, R. Alkaline Conditions in Hydrophilic Interaction Liquid Chromatography for Intracellular Metabolite Quantification Using Tandem Mass Spectrometry. *Anal. Biochem.* **2015**. <https://doi.org/10.1016/j.ab.2015.01.002>.
- (17) Spalding, J. L.; Naser, F. J.; Mahieu, N. G.; Johnson, S. L.; Patti, G. J. Trace Phosphate Improves ZIC-PHILIC Peak Shape, Sensitivity, and Coverage for Untargeted Metabolomics. *J. Proteome Res.* **2018**, *17* (10), 3537–3546. <https://doi.org/10.1021/acs.jproteome.8b00487>.
- (18) Erngren, I.; Haglöf, J.; Engskog, M. K. R.; Nestor, M.; Hedeland, M.; Arvidsson, T.; Pettersson, C. Adduct Formation in Electrospray Ionisation-Mass Spectrometry with Hydrophilic Interaction Liquid Chromatography Is Strongly Affected by the Inorganic Ion Concentration of the Samples. *J. Chromatogr. A* **2019**. <https://doi.org/10.1016/j.chroma.2019.04.049>.
- (19) Jaeger, C.; Méret, M.; Schmitt, C. A.; Lisec, J. Compound Annotation in Liquid Chromatography/High-Resolution Mass Spectrometry Based Metabolomics: Robust Adduct Ion Determination as a Prerequisite to Structure Prediction in Electrospray Ionization Mass Spectra. *Rapid Commun. Mass Spectrom.* **2017**, *31* (15), 1261–1266. <https://doi.org/10.1002/rcm.7905>.
- (20) Mahieu, N. G.; Spalding, J. L.; Gelman, S. J.; Patti, G. J. Defining and Detecting Complex Peak Relationships in Mass Spectral Data: The Mz.Unity Algorithm. *Anal. Chem.* **2016**, *88*, 9037–9046. <https://doi.org/10.1021/acs.analchem.6b01702>.
- (21) Mahieu, N. G.; Spalding, J. L.; Gelman, S. J.; Patti, G. J. Defining and Detecting Complex Peak Relationships in Mass Spectral Data: The Mz.Unity Algorithm. *Anal. Chem.* **2016**, *88* (18), 9037–9046. <https://doi.org/10.1021/acs.analchem.6b01702>.
- (22) Xu, Y. F.; Lu, W.; Rabinowitz, J. D. Avoiding Misannotation of In-Source Fragmentation Products as Cellular Metabolites in Liquid Chromatography-Mass Spectrometry-Based Metabolomics. *Anal. Chem.* **2015**, *87* (4), 2273–2281. <https://doi.org/10.1021/ac504118y>.
- (23) Baran, R.; Northen, T. R. Robust Automated Mass Spectra Interpretation and Chemical Formula Calculation Using Mixed Integer Linear Programming. *Anal. Chem.* **2013**, *85* (20), 9777–9784. <https://doi.org/10.1021/ac402180c>.