Title

Alternative Identification of Glycosides Using MS/MS Matching with an In Silico-Modified Aglycone Mass Spectra Library

Journal

Analytical Chemistry, 95(28)

ISSN

0003-2700

Authors

Rodríguez, Elys P

Li, Yuanyue

Vaniya, Arpana

et al.

Publication Date

2023-07-18

DOI

10.1021/acs.analchem.3c00957

Peer reviewed

HHS Public Access

Author manuscript

Anal Chem. Author manuscript; available in PMC 2024 October 21.

Published in final edited form as:

Anal Chem. 2023 July 18; 95(28): 10618–10624. doi:10.1021/acs.analchem.3c00957.

Alternative Identification of Glycosides Using MS/MS Matching with an In Silico-Modified Aglycone Mass Spectra Library

Elys P. Rodríguez,
Department of Chemistry, University of California Davis, Davis, California 95616, United States; West Coast Metabolomics Center, University of California Davis, Davis, California 95616, United States

Yuanyue Li,
West Coast Metabolomics Center, University of California Davis, Davis, California 95616, United States

Arpana Vaniya,
West Coast Metabolomics Center, University of California Davis, Davis, California 95616, United States

Patrick M. Shih,
Department of Plant and Microbial Biology and Innovative Genomics Institute, University of California Berkeley, Berkeley, California 94720, United States; Feedstocks Division, Joint BioEnergy Institute, Emeryville, California 94608, United States; Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, California 97420, United States

Oliver Fiehn
West Coast Metabolomics Center, University of California Davis, Davis, California 95616, United States

Abstract

Glycosylation of metabolites serves multiple purposes. Adding sugars makes metabolites more water soluble and improves their biodistribution, stability, and detoxification. In plants, the increase in melting points enables storing otherwise volatile compounds that are released by hydrolysis when needed. Classically, glycosylated metabolites were identified by mass spectrometry (MS/MS) using [M-sugar] neutral losses. Herein, we studied 71 pairs of glycosides

Corresponding Author: **Oliver Fiehn** – *West Coast Metabolomics Center, University of California Davis, Davis, California 95616, United States;* ofiehn@ucdavis.edu.

Author Contributions

E.R. and O.F. designed the experiments. A.V. performed experimental measurements. E.R. analyzed the data and interpreted the results. Y.L, created the libraries of glycosides. P.M.S. and O.F. provided resources and supervision. E.P and O.F. wrote the manuscript in interaction with all contributing authors.

The authors declare no competing financial interest.

ASSOCIATED CONTENT

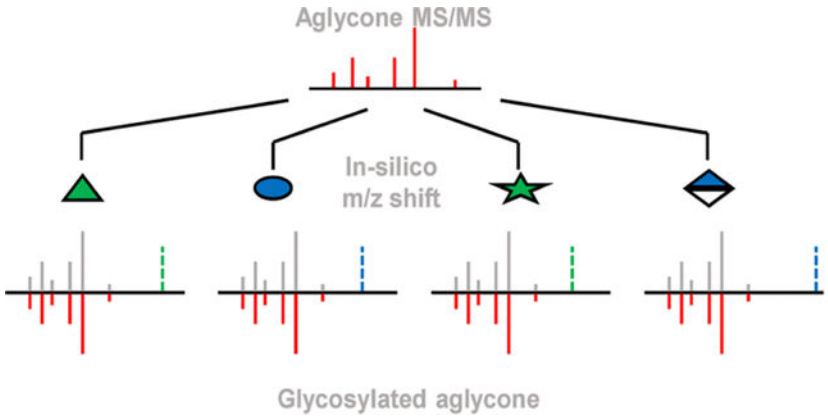
Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.3c00957.

Increment in matching score with higher collision energy; MS1 full scan of 3-glucose-4'-dihydroxypropiophenone; mechanism of collision-induced fragmentation of isolariciresinol-9-O'-glycoside in ESI(−) mode (PDF) Authentic standards data and all identified compounds (XLSX)

with their respective aglycones, including hexose, pentose, and glucuronide moieties. Using liquid chromatography (LC) coupled to electrospray ionization high-resolution mass spectrometry, we detected the classic [M-sugar] product ions for only 68% of glycosides. Instead, we found that most aglycone MS/MS product ions were conserved in the MS/MS spectra of their corresponding glycosides, even when no [M-sugar] neutral losses were observed. We added pentose and hexose units to the precursor masses of an MS/MS library of 3057 aglycones to enable rapid identification of glycosylated natural products with standard MS/MS search algorithms. When searching unknown compounds in untargeted LC-MS/MS metabolomics data of chocolate and tea, we structurally annotated 108 novel glycosides in standard MS-DIAL data processing. We uploaded this new in silico-glycosylated product MS/MS library to GitHub to enable users to detect natural product glycosides without authentic chemical standards.

Graphical Abstract



INTRODUCTION

About 100,000 to one million specialized metabolites are collectively produced by plants. Any single plant species may produce thousands of these metabolites.[1] Owing to the sessile nature of plants, they must rely on innate immunity to protect themselves from the threats of environmental stresses. Such immunity is based on the inherent flexibility of metabolic modifications that are dynamic and largely reversible, such as acetylation, methylation, hydroxylation, prenylation, and glycosylation.[2] Glycosylation is one of the most efficient metabolite modifications in all kingdoms of life because sugars can be added to both N-, O-, or S-heteroatoms and carbon. Upon glycosylation, hydrophobic metabolites become more water soluble, which improves their biodistribution and metabolism, assisting long-distance transport,[2] raising melting points of otherwise volatile compounds,[1,3,4] and improving biodistribution of pharmacokinetic parameters in humans.[5] Many compound classes can be glycosylated, including hormones, sweeteners, alkaloids, flavonoids, antibiotics, and others. It is generally expected that glycosylation leads to earlier retention times in reversed-phase liquid chromatography, but this has only been illustrated with a limited selection of compounds, mainly glucosylated flavonoids.[6–10] Similarly, it is frequently hypothesized that glycosylated compounds show neutral loss product ions of the glycosyl residues in tandem mass spectrometry (MS/MS), leading to 162 Da losses for hexoxides

or 147 Da for rhamnosides and 133 Da for arabinosides.[11] Currently, only 11.9% of the largest Natural Compounds database (COCONUT) comprises glycosylated compounds.[12] Owing to the biological importance of glycosylation in nature, this number of modified compounds in COCONUT is very likely due to underreporting of glycosylated metabolites in classic natural product research. This critical nature of glycosylation is underscored by the presence of more than 100 families of glycosyltransferases across all kingdoms of life. Therefore, methods for comprehensive annotation of glycosylated natural products need to be improved. Unfortunately, glycosylated natural products are mostly absent in mass spectral libraries, although there have been notable developments in the identification of flavonoid glycosides through the use of aglycone's MSMS, such as FlavonoidSearch,[13] GNPS similarity search,[14] SIRIUS,[15] and others,[16,17] especially for glycosides that have been previously studied and characterized. As these tools are mostly trained flavonoids, other types of glycosides require new analytical methods.

We, herein, developed a method for identification of glycosides that does not rely on the neutral loss of a sugar. We present a freely available MS/MS library for rapid identification of glycosides based on their MS/MS fragmentation by full MS mode with HCD fragmentation using orbital ion trap mass spectrometry. We showcase the use of such an approach by detecting 23 novel glycosylated compounds in food products.

EXPERIMENTAL METHODS

Aglycone and Glycoside Standards.

Seventy-one glycoside/aglycone pairs were purchased from Sigma-Aldrich, Cayman Chemicals, and AnalytiCon Discovery with >98% purity. Authentic standards were dissolved in methanol at a concentration ranging from 1 to 100 ppm and stored at 20 °C until analysis.

Preparing Food Samples for Untargeted Metabolomics.

As a test case, dark chocolate and green tea samples were purchased from a local grocery store. Twenty milligrams of lyophilized food samples were homogenized for 30 s at 1500 rpm using a Genogrinder 2010 homogenizer (Spex Sample Prep, Metuchen, NJ). Homogenized samples were extracted using a 1000 L ice-cold 50:50 ethyl acetate/H_2O mixture and centrifuged at 12,000*g* for 10 min. Supernatants were transferred to 1.5 mL Eppendorf tubes and dried down using a Centrivap (LabConCo, Kansas, MO). Dried extract were reconstituted in the LC starting buffer (97/3 H_2O:ACN, v/v), including 10 internal standards (hippuric acid-d5, caffeine-d10, (−)epigallocatechin gallate-d3/d4, trans-cinnamic acid-d5, trans-resveratrol-d4, daidzein-d4, quercetin-d3, genistein-d4, apigenin-d5, 2-hydroxyfluorene-d9, reserpine-d9).

LC-MS/MS Data Acquisition.

Measurements were performed on a ThermoFisher Q Exactive HF instrument. For polar aromatic compounds, 5 μL of diluted samples were separated on a Kinetex UPLC PFP C18 column (50 × 2.1 mm; 1.7 μm). Concentrations ranged from 5 to 1000 pm depending on compound ionization efficiency to give optimal peak shapes and intensities without

column or MS saturation. The column was maintained at 40 °C with a flow rate of 0.4 mL/min. The mobile phases consisted of (A) water (100%) with formic acid (0.1%) and (B) acetonitrile (100%) and formic acid (0.1%). The separation was conducted under the following gradients: 0 min 7% B; 0–4.5 min 30% B; 4.5–6.5 min 100% B; 6.5–7.5 min 100% B; 7.5–8.0 min 7% B; and 8–10.0 min 7% B. The Q Exactive HF mass spectrometer was operated using negative-mode electrospray ionization with a mass range from 120 to 1700 *m*/*z*; sheath gas flow rate 60; aux gas flow rate 25; sweep gas flow rate 2; spray voltage 3.6 kV; capillary at 300 °C; S-lens RF level 50; and aux gas heater at 370 °C. MS^1 parameters: 70,000 resolving power; 1e6 automatic gain control target; 100 ms maximum injection time; and centroided spectra. Data-dependent MS^2 parameters: 17,500 resolving power; 1e5 automatic gain control target; 100 ms maximum injection time; 4 loop count; TopN 4; 1 Da isolation window; 70 *m*/*z* fixed first mass; (stepped) normalized collision energy (NCE) 35–45–65; and centroided spectra. All 80 pairs of authentic chemical standards were analyzed at 65 NCE in negative electrospray mode.

Data Processing and Glycoside/Aglycone MS/MS Matching.

LC-MS/MS data was analyzed by MS-DIAL software. MS-DIAL ver. 3.96[20] was used with 10,000 counts/spectrum as a lower threshold for peak peaking in LC-Orbitrap MS/MS and 80% identification cutoff with an accurate mass filter of 0.01 Da for the precursor and 0.05 Da for MS2 ions. Detailed parameter settings are listed in Supporting Table 2. Mass spectra were exported from MS-DIAL in a text file format and uploaded to MassBank of North America (http://massbank.us) including compound identifier InChI keys, molecular formulas, exact isotope masses, and precursor *m*/*z*. The Vaniya/Fiehn Natural Products Library of 3257 compounds was downloaded from MassBank.us, which comprised of 28,260 spectra in positive and negative modes and different adducts (Figure 1). These spectra were modified as new in silico libraries by adding glycosides to be used as.msp file for MS/MS matching in standard data processing software such as MS-DIAL. To this end, the following neutral precursor mass shifts were added to mimic glycosylated natural products: +162.0533 (glucose or galactose), +176.0319 (glucuronide), +147.0660 (rhamnose), +133.05003 (arabinoside), double glycosylation +324.1066, and rhamnose + glucose (or galactose) +309.1201. Corresponding SMILES and InChI keys were generated for these virtual structures and uploaded to GitHub for public use. All food metabolome spectra with a similarity score >800 were automatically annotated, while spectra with a similarity dot score of 400–600 were manually curated. Annotations of glycosylated metabolites were required to elute earlier than their aglycone counterparts.

RESULTS AND DISCUSSION

Determining Glycosides by Detection of Sugar Loss.

Few plant glycosides are commercially available or present in MS/MS libraries. We first compared MS/MS spectra and retention times of 71 matched pairs of authentic aglycone standards and their corresponding glycosides (Table S1). Only the intrinsically cationic compounds cyanidin and cyanidin-3-*O*-glucoside were analyzed in positive electrospray (ESI); all other compounds showed best ionization efficiency as deprotonated species $[M - H]^-$ in negative ESI mode. Lower abundant formate adducts $[M + HCOO]^-$ were detected

in 10% of the cases. Glycoside spectra and aglycones were compared as shown in Figure 2, giving details of the fragmentation of the rings A, B, and C of luteolin and luteolin-4-*O*-glucoside.[7,8] This example typifies a liable fragmentation of the sugar, leaving the aglycone intact until further fragmentation. The MS2 spectrum for Luteolin-4-*O*-glucoside shows that all of the ions produced are characteristic of the flavonoid subgroups of aglycone ($^{0,2}A^-$, $^{0,2}B^-$,$^{1,3}A^-$,$^{1,3}B^-$,$^{1,3}A^-$) and water and CO losses for each ring [7]. We found that identical fragmentations also applied to luteolin-4-*O*-glucoside that showed a dominant loss of the glycosyl residue (M-162), followed by the fragmentation of aglycones. Next, we studied the 80 authentic chemical standard pairs using 10, 15, 25, 35, and 65% NCEs to determine if MS/MS fragmentation of glycosides always showed a typical neutral loss of a sugar moiety, including for di- and triglycosylated metabolites and glucuronide variants (Figure 3). Accurate masses, retention times, and similarity scores of authentic standards between aglycones and glycosides are given in Table S1. In total, 69% of our 71 pairs of authentic standards of glycosides and corresponding aglycones demonstrated a characteristic neutral loss of a sugar moiety between glycosylated- and non-glycosylated natural products (i.e., glucose, galactose, rhamnose, arabinose, xylose, and fucose). Surprisingly, two out of seven glycosylated pairs of the same aglycone parent molecule did not show the same neutral glycoside loss under identical fragmentation conditions. An example of different spectra of glycosylated regioisomers is given for Isorhapontigenin glycosides (Figure S4). Across all 71 glycosylated compounds, the abundance of characteristic glycosyl neutral loss fragments ranged from 0.3 to 100%. For *O*-glycosides, the aglycone ion often became the most abundant fragment ion (base peak). Conversely, however, for *C*-glycosides, product ions resulting from the neutral loss of 120 and 90 m/z never exceeded 0.8% abundance.[18,19] We concluded that lower abundant fragments may easily become absent from MS/MS spectra or indistinguishable from noise ions if total metabolite levels (and corresponding precursor ion intensities) become too low to obtain high-quality spectra. Hence, confident identification of glycosides in untargeted metabolomics should not solely depend on the presence of corresponding neutral loss fragments.

Identification of Aglycone Residues for Glycosylated Compounds.

As neutral losses of glycoside residues were found to be insufficient to detail the presence of a glycosylated natural product, we investigated the MS/MS spectra for the 71 pairs of aglycones and glycosylated derivatives of these aglycones that were available to us, including four *C*-glycosides. We specifically studied if the MS/MS spectra of the glycosylated intermediates showed fragments that directly indicated the presence of the corresponding specific aglycones. Fragmentation mechanisms and rules were tested for 10 subclasses of flavonoid glycosides, including anthocyanin, coumarin, fatty acyl, flavonoid, isoflavonoid, flavones, steroidal, stilbene, terpene, and lignan glycosides. Initial studies on 15 glycosides showed insufficient fragmentation at 35 NCE. We then tested optimal energies in ten random glycosides ranging from 10 to 65 NCE and found 65 NCE to be optimal to match with aglycone spectra in all cases (Figure S1).

Spectral matching was not dependent on glycoside adducts, as both $[M - H]^-$ and $[M + HCOO]^-$ showed high similarity of glycosides with their corresponding aglycones. This data showed that MS/MS spectra can be used for compound identification even if

formate adducts dominate MS1 precursor ions, and thus, would be preferentially chosen for fragmentation in data-dependent untargeted MS/MS acquisition schemes. However, 31% of the tested glycosides did not show a characteristic neutral loss of a sugar. For these 22 glycosylated products, we only obtained an average MS/MS similarity score of 542, whereas glycosides that showed the neutral sugar loss of a sugar gave a 682 average similarity score. Interestingly, 54/71 glycosylated compounds even showed an in-source fragmentation loss of sugar moieties (Table S1). However, only 14 of these cases showed more than 25% abundance of in-source fragmentation when comparing precursor intensities of the glycosylated intermediate and the in-source fragment aglycone ion based on identical retention times. Eight cases showed more than 70% in-source fragmentation abundance. We did not find a relationship between the abundance of in-source fragmentation and the presence of a glycosyl neutral loss in MS/MS spectra.

The absence or presence of a neutral sugar loss fragment ion was consistent across all $[M - H]^-$, $[M + HCOOH - H]^-$, and $[M + Cl]^-$ adducts except for two cases, 5′-*O*-glucoseisorhapontigenin and 2′-*O*-glucose-phloretin (Table S1). For these two compounds, the neutral sugar loss product ion was exclusively seen in the $[M - H]^-$ or $[M + Cl]^-$ precursor ions, respectively. More interestingly, only one compound yielded sugar fragmentation to yield a negatively charged glucose ion at 179.056 *m*/*z* (Figure S2), concomitantly with a [M-179]-product ion for the glycosyl residue.

When aglycone fragment anions Y_o were highly unstable (e.g., cis-3-hexenol glycoside, 7-glycosyl-11-methylodeoside, isolariciresinol-9-*O*′-glycoside, (−)-erythro-anethole glycol 2-glucoside), glycosylated MS2 spectra did not show the aglycones. Although the intact aglycones Y_o were absent from the spectra, the remaining fragments were still highly similar to the original unmodified aglycones. For example, for isolariciresinol-9-*O*′-glycoside, the most proximal ion to Y_o was observed at 344 *m*/*z*, which is interpreted as an oxolane formation followed by a water loss ($-H_2O$).[7] The same mechanism was observed in the isolariciresinol-9-*O*′-glycoside $[M - H]^-$ MS2 spectrum (Figure S3), with the most proximal ion to Y_o at *m*/*z* 344, not at *m*/*z* 521 nor at *m*/*z* 359. We interpret that the further fragmentation of *m*/*z* 344 leads as opening of the oxolane ring and splitting into a coniferyl alcohol fragment yielding ions at *m*/*z* 163 and 159 (Figure S3).[21] The same fragments were found for the corresponding authentic isolariciresinol aglycones MS2 spectrum (Figure S3). Another example of an absent Y_o aglycone fragment ion was found for 3,4′-dihydropropiophenone glucoside (Figure S2). For both the aglycone and the glycoside, a water loss was observed at the *para* position of the phenol ring to yield *m*/*z* 147 and *m*/*z* 119, respectively, with the carbonyl still attached to the phenol ring.

In Silico Library Creation and Validation.

Owing to the ubiquity of glycosylation in plant biology and metabolic engineering, it is essential to develop methods that rapidly identify glycosylated compounds. To enable the use of existing aglycone MS/MS libraries, we added 12 different sugar combinations for producing in silico structures of mono-, di-, and triglycosides using pentose, deoxypentose, and hexose as sugar units. To this end, we first combined all MassBank.us and NIST20 spectra into a single msp file. We then removed all spectra except for MS/MS spectra

obtained by orbital ion traps. We then added the exact masses of the sugar moieties to the corresponding adduct ions, leading to hypothetical precursor ions of 2,009,850 glycosylated metabolites. As plant glycosylation reactions are known to occur even for unnatural exposome compounds like TNT,[15] we did not constrain this hypothetical library any further. Next, glycosylated spectra were separated into two libraries for positive ESI and negative ESI mass spectra. We validated this newly generated glycoside MS/MS library by matching our experimental glycoside MS/MS spectra against the newly generated library with a precursor window of 3 mDa. The results were subjected to false discovery rate (FDR). Combining test cases for both positive and negative ESI modes, 28 out of our 71 experimentally known glycosides were correctly ranked as top hit in this virtual MS/MS library; 54/71 of the spectra were found in the top-5 ranks, and 61/71 of the spectra gave the correct structures within the top-10 hits with an average similarity dot score of 789. Three glycoside compounds were misidentified as constitutional isomers of their aglycone scaffold, maesopin-4-glucose-4-rhamnoside, oleoside-11-methylester-4-glucoside, and quercetin-3-glucoronide (Table S1). Finally, 12/71 compounds were identified as a near identical isomer, distinguished by a single-bond difference. Many of these hits showed less than 50 units difference in dot score similarities between both isomers, which are caused by slight differences in ion abundances in library spectra. In compound annotation schemas, such marginal differences do not lead to confident identifications but must be validated by retention time differences or ion mobility differences using authentic compounds. Mass spectrometry often poorly performs in recognizing one-bond differences in positional isomers, especially for labile sugar neutral losses or labile fatty acyl neutral losses. Hence, with some justification, such annotations may be counted as true identifications, which bare the position of the glucose residue.

Novel Glycosides in Food.

To find glycosylated food compounds, we examined comprehensive MS2 data from replicates of six chocolate and green tea samples. Herein, we used the Vaniya-Fiehn Natural Products Library of 3057 authentic standards that were run under the same chromatographic and MS/MS conditions, including 469 glycosylated compounds. This library only confirmed 15 glycosylated metabolites in chocolate and green tea samples (Table S2). Using the in silico expansion, LC-data-dependent MS/MS spectra yielded 108 newly identified glycosylated compounds corresponding to 68 unique aglycones (Figure 5 and Table S2). Annotations relied on <3 mDa accurate precursor mass differences and shifts toward earlier retention times than the corresponding aglycones. All retention times and MS/MS data were manually evaluated. Ninety-four compounds showed a weighted dot score >700, while 14 compounds scored at <700 similarity (Table S2), Figure 4. Aglycones included glycosylated flavonoids, benzenoids, benzodioxols, carboxylates, cinnamates, coumarins, isocoumarins, isoflavonoids, linear 1,3-diarylpropanoids, phenols, prenols, and stilbenes (Table S1). This list of aglycones confirms identifications in previous food studies.[22,23] Fifty-four novel glycosylated compounds were found as positional isomers defined by well-separated chromatographic peaks but identical accurate precursor masses and highly similar MS/MS spectra. Most novel glycosylated aglycones were found in both foods, but 21 glycosides were unique to chocolate and three compounds were unique in tea. Structural annotation of glycosylations remained at the level of positional classifications because

different hexoses or pentoses cannot be distinguished by MS/MS alone. Similarly, many aglycones offered different potential glycosylation sites that remained indistinguishable without fractionation and additional NMR investigations.[24] Hence, annotations for these compounds mostly remained at the MSI level 3.[10] Additional glycosides might be found if virtual MS/MS libraries were to be extended to additional precursor mass variants such as chloride and acetate adducts, or to neutral losses of water. We here refrained from such extension as neutral losses of water would coincide with differences between rhamnose and arabinose glycosylations. Future experiments with different LC buffers or variation of electrospray voltages might be needed to extend virtual libraries and identify more glycosylated food compounds (Figure 5).

CONCLUSIONS

We present a novel in silico library for the detection of glycosylated compounds, which expands the very few experimental negative electrospray MS/MS spectra that are publicly available today in public libraries. We validated our observations on experimentally acquired high-resolution mass spectra of chocolate and tea using untargeted metabolomics method and identified 108 novel glycosides that could have only been identified using this approach of expanding MS/MS libraries by precursor accurate mass shifts and matching retention time shifts in comparison to authentic aglycone standards.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Funding was provided by NIH grant 1 T32 GM 136597–1 A1 for ER and NIH R03 OD034497 and USDA 2021–67017-35783 to OF. P.M.S. was supported by the DOE Joint BioEnergy Institute (https://www.jbei.org) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, supported by the U.S. Department of Energy, Energy Efficiency and Renewable Energy, Bioenergy Technologies Office, through contract DE-AC02–05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

Data Availability Statement

All modified libraries are publicly available at GitHub.com/eprod2016/MSMS-Glycoside-Libraries.

REFERENCES

(1). Yu X; Xiao J; Chen S; Yu Y; Ma J; Lin Y; Li R; Lin J; Fu Z; Zhou Q; Chao Q; Chen L; Yang Z; Liu R Nat. Commun. 2020, 11, No. 5586. [PubMed: 33149146]

(2). Wang S; Alseekh S; Fernie AR; Luo J Mol. Plant 2019, 12, 899–919. [PubMed: 31200079]

(3). Whitmore BA; McCann SE; Noestheden M; Dennis EG; Lyons SM; Durall DM; Zandberg WF Molecules 2021, 26, No. 4519. [PubMed: 34361670]

(4). Ohgami S; Ono E; Horikawa M; Murata J; Totsuka K; Toyonaga H; Ohba Y; Dohra H; Asai T; Matsui K; Mizutani M; Watanabe N; Ohnishi T Plant Physiol. 2015, 168, 464–477. [PubMed: 25922059]

(5). Křen V Glycoside vs. Aglycon: The Role of Glycosidic Residue in Biological Activity. In Glycoscience: Chemistry and Chemical Biology; Fraser-Reid; Tatsuta; Thiem, Eds.; Springer: Berlin, Heidelberg, 2008; pp 2589–2644.

(6). Abrankó L; Szilvássy B J. Mass Spectrom. 2015, 50, 71–80. [PubMed: 25601677]

(7). Chen G; Li X; Saleri F; Guo M Molecules 2016, 21, No. 1275. [PubMed: 27669205]

(8). Li Z-H; Guo H; Xu W-B; Ge J; Li X; Alimu M; He D-J J. Chromatogr. Sci. 2016, 54, 805–810. [PubMed: 26896347]

(9). Ma YL; Li QM; Van den Heuvel H; Claeys M Rapid Commun. Mass Spectrom. 1997, 11, 1357–1364.

(10). Schymanski EL; Jeon J; Gulde R; Fenner K; Ruff M; Singer HP; Hollender J Environ. Sci. Technol. 2014, 48, 2097–2098. [PubMed: 24476540]

(11). Kazuno S; Yanagida M; Shindo N; Murayama K Anal. Biochem. 2005, 347, 182–192. [PubMed: 16269127]

(12). Schaub J; Zielesny A; Steinbeck C; Sorokina M Biomolecules 2021, 11, No. 486. [PubMed: 33804966]

(13). Akimoto N; Ara T; Nakajima D; Suda K; Ikeda C; Takahashi S; Muneto R; Yamada M; Suzuki H; Shibata D; Sakurai N Sci. Rep. 2017, 7, No. 1243. [PubMed: 28455528]

(14). Wang M; Carver JJ; Phelan VV; Sanchez LM; Garg N; Peng Y; Nguyen DD; Watrous J; Kapono CA; Luzzatto-Knaan T; Porto C; Bouslimani A; Melnik AV; Meehan MJ; Liu W-T; Crüsemann M; Boudreau PD; Esquenazi E; Sandoval-Calderón M; Kersten RD; Pace LA; Quinn RA; Duncan KR; Hsu C-C; Floros DJ; Gavilan RG; Kleigrewe K; Northen T; Dutton RJ; Parrot D; Carlson EE; Aigle B; Michelsen CF; Jelsbak L; Sohlenkamp C; Pevzner P; Edlund A; McLean J; Piel J; Murphy BT; Gerwick L; Liaw C-C; Yang Y-L; Humpf H-U; Maansson M; Keyzers RA; Sims AC; Johnson AR; Sidebottom AM; Sedio BE; Klitgaard A; Larson CB; Boya P CA; Torres-Mendoza D; Gonzalez DJ; Silva DB; Marques LM; Demarque DP; Pociute E; O'Neill EC; Briand E; Helfrich EJN; Granatosky EA; Glukhov E; Ryffel F; Houson H; Mohimani H; Kharbush JJ; Zeng Y; Vorholt JA; Kurita KL; Charusanti P; McPhail KL; Nielsen KF; Vuong L; Elfeki M; Traxler MF; Engene N; Koyama N; Vining OB; Baric R; Silva RR; Mascuch SJ; Tomasi S; Jenkins S; Macherla V; Hoffman T; Agarwal V; Williams PG; Dai J; Neupane R; Gurr J; Rodríguez AMC; Lamsa A; Zhang C; Dorrestein K; Duggan BM; Almaliti J; Allard P-M; Phapale P; Nothias L-F; Alexandrov T; Litaudon M; Wolfender J-L; Kyle JE; Metz TO; Peryea T; Nguyen D-T; VanLeer D; Shinn P; Jadhav A; Müller R; Waters KM; Shi W; Liu X; Zhang L; Knight R; Jensen PR; Palsson BØ; Pogliano K; Linington RG; Gutiérrez M; Lopes NP; Gerwick WH; Moore BS; Dorrestein PC; Bandeira N Nat. Biotechnol. 2016, 34, 828–837. [PubMed: 27504778]

(15). Dührkop K; Fleischauer M; Ludwig M; Aksenov AA; Melnik AV; Meusel M; Dorrestein PC; Rousu J; Böcker S Nat. Methods 2019, 16, 299–302. [PubMed: 30886413]

(16). Mass Spectrometry-Based Techniques to Elucidate the Sugar Code | Chemical Reviews. 10.1021/acs.chemrev.1c00380. (accessed May 18, 2023).

(17). Polasky DA; Yu F; Teo GC; Nesvizhskii AI Nat. Methods 2020, 17, 1125–1132. [PubMed: 33020657]

(18). Abad-García B; Garmón-Lobato S; Berrueta LA; Gallo B; Vicente F Rapid Commun. Mass Spectrom. 2008, 22, 1834–1842. [PubMed: 18470889]

(19). Cao J; Yin C; Qin Y; Cheng Z; Chen D J. Mass Spectrom. 2014, 49, 1010–1024. [PubMed: 25303391]

(20). Qiu F; Fine DD; Wherritt DJ; Lei Z; Sumner LW Anal. Chem. 2016, 88, 11373–11383. [PubMed: 27934098]

(21). Eklund PC; Backman MJ; Kronberg LÅ; Smeds AI; Sjöholm RE J. Mass Spectrom. 2007, 43, 97–107.

(22). Gandia-Herrero F; Lorenz A; Larson T; Graham IA; Bowles DJ; Rylott EL; Bruce NC Plant J. 2008, 56, 963–974. [PubMed: 18702669]

(23). Pinto G; Aurilla M; Illiano A; Fontanarosa C; Sannia G; Trifuoggi M; Lettera V; Sperandeo R; Pucci P; Amoresano A J Mass Spectrom. 2021, 56, No. e4651. [PubMed: 32893948]

(24). Śliwka-Kaszyńska M; Anusiewicz I; Skurski P Molecules 2022, 27, No. 1032. [PubMed: 35164300]

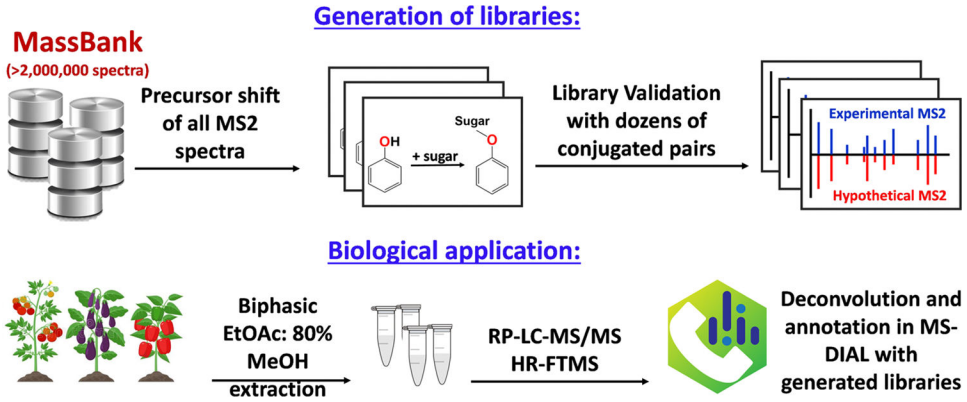


Figure 1.

Modifying MS/MS spectral libraries of natural products of aglycones in negative ionization mode. By computationally adding the mass of a sugar (top panel), 13 libraries were generated to be used in MS-DIAL data processing software (bottom panel).

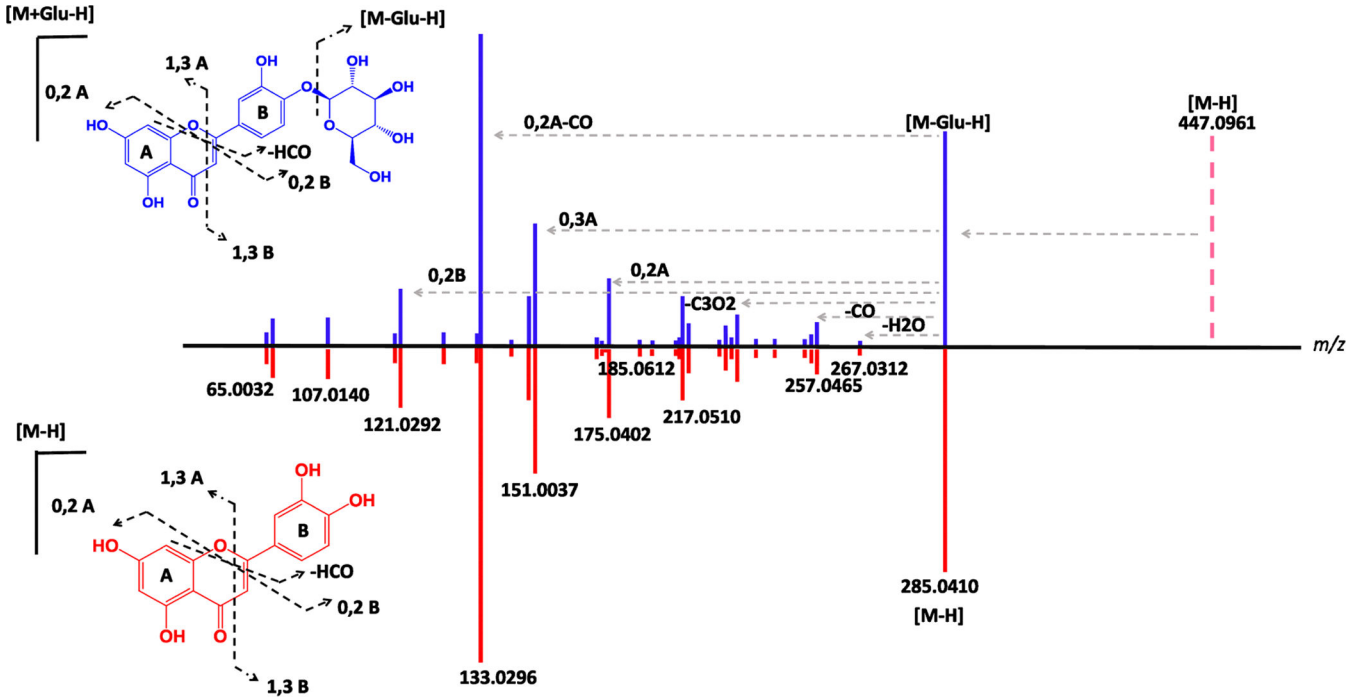


Figure 2.

Head-to-tail MS/MS comparison of an example pair of (A) luteolin-4′-*O*-glucoside and (B) luteolin. Owing to neutral loss of glucose, glycosylated molecules often do not show the precursor mass in MS/MS spectra. Therefore, MS/MS spectra of glycosylated compounds are often highly similar to the corresponding aglycone. Aglycone fragmentations are rationalized per Ma, Y. L.; Li, Q. M.; Van den Heuvel, H.; Claeys, M. (9).

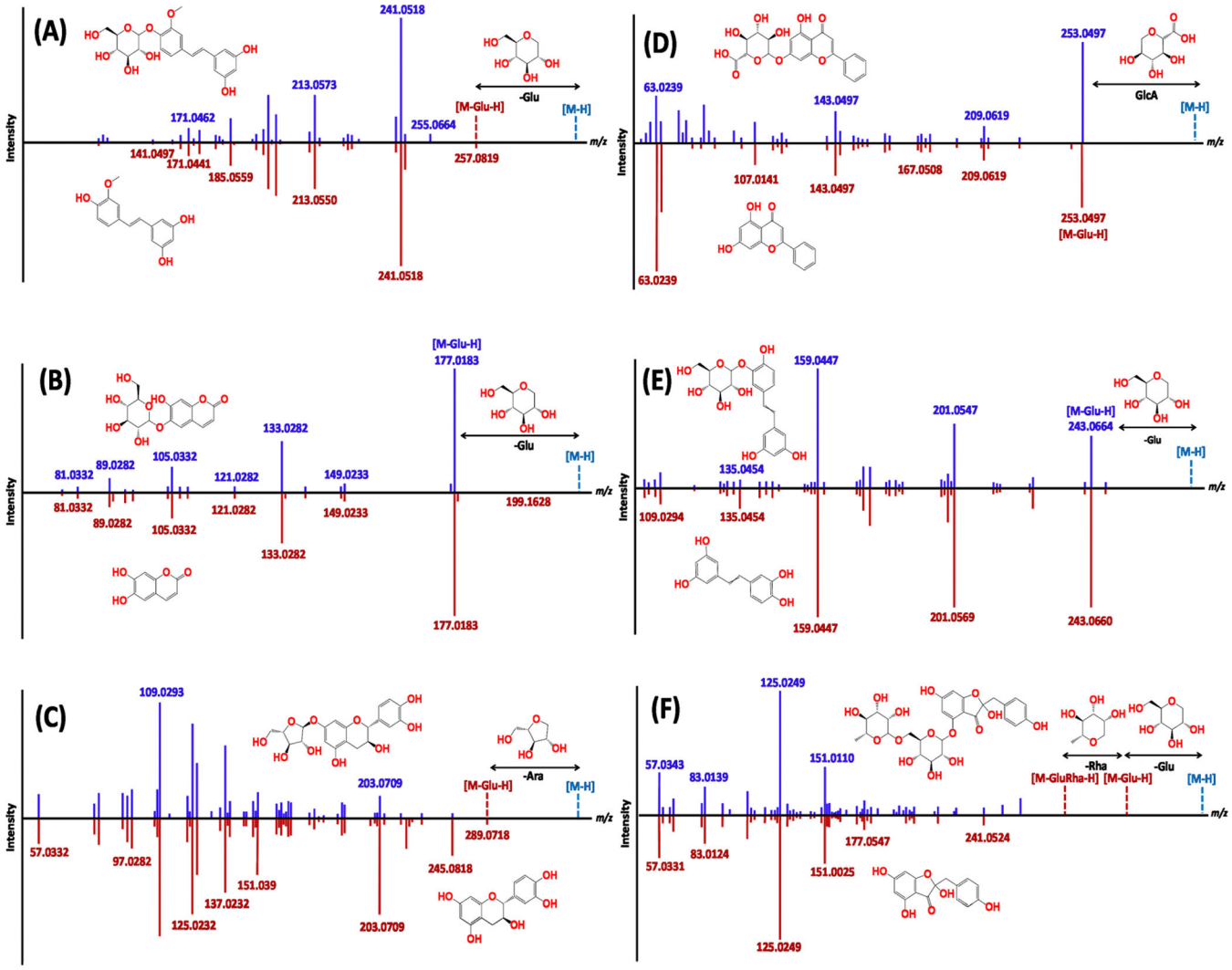


Figure 3.

Matching arabinosides, rhamnosides, glucosides, and combinations against their corresponding aglycones MS/MS spectra. (A) Catechin 7-arabinofuranoside, (B) piceannetanol-3′-glucoside, (C) cyanidin-3-*O*-glucoside, (D) chrysin-7-*O*-glucoronide, (E) naringenin-7-*O*-glucoside, and (F) quercetin-3-rhamnoside. All spectra are uploaded to the public MassBank.us repository.

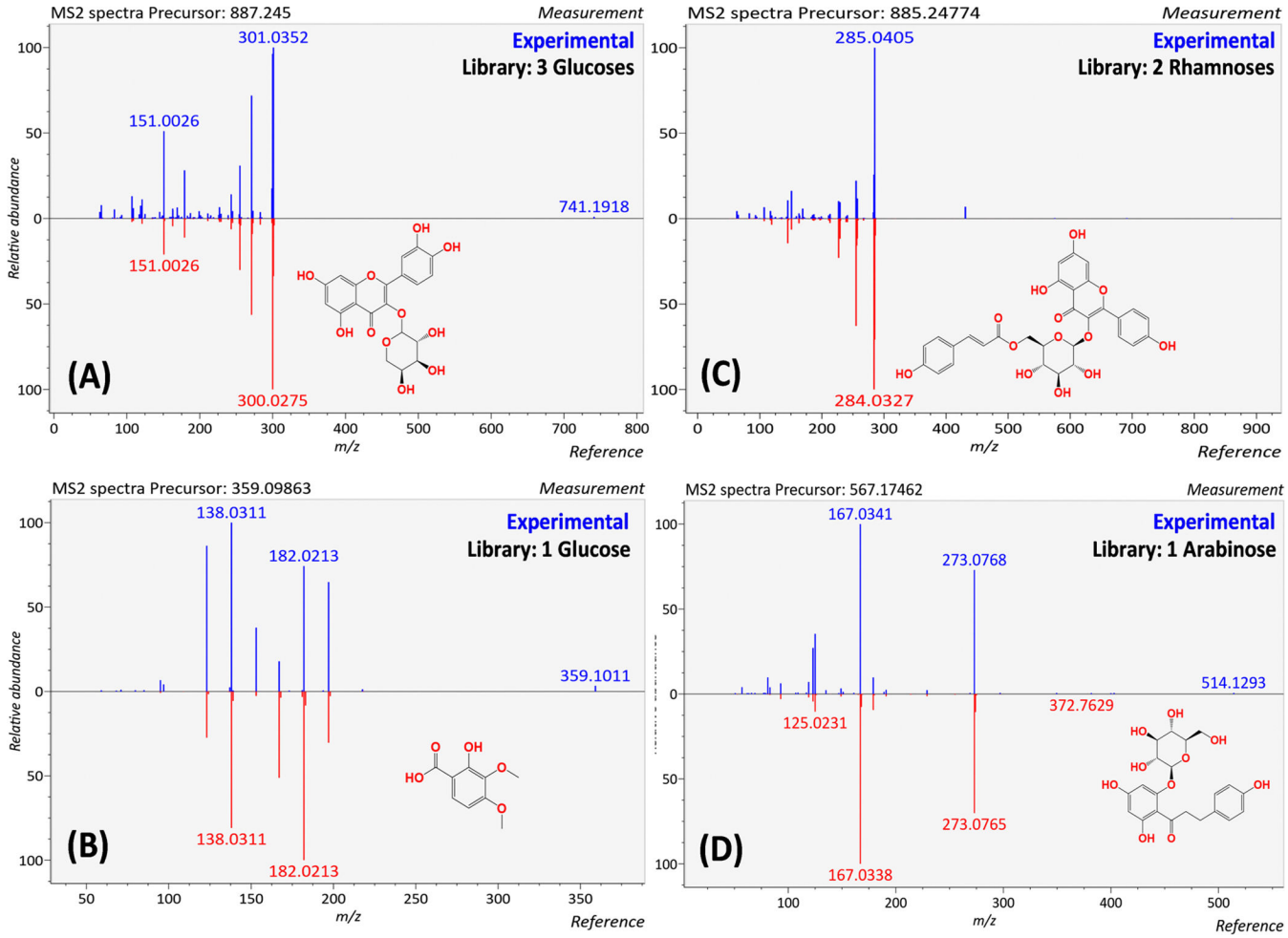


Figure 4.

Examples of experimental tandem mass spectra of glycosides found in food (blue) versus the conjugated libraries spectra (red). Examples in negative mode are (A) avicularin, (B) 2-hydroxy-3,4-dimethoxybenzoic acid, (C) tiliroside, and (D) phlorizin. Aglycone structures are shown.

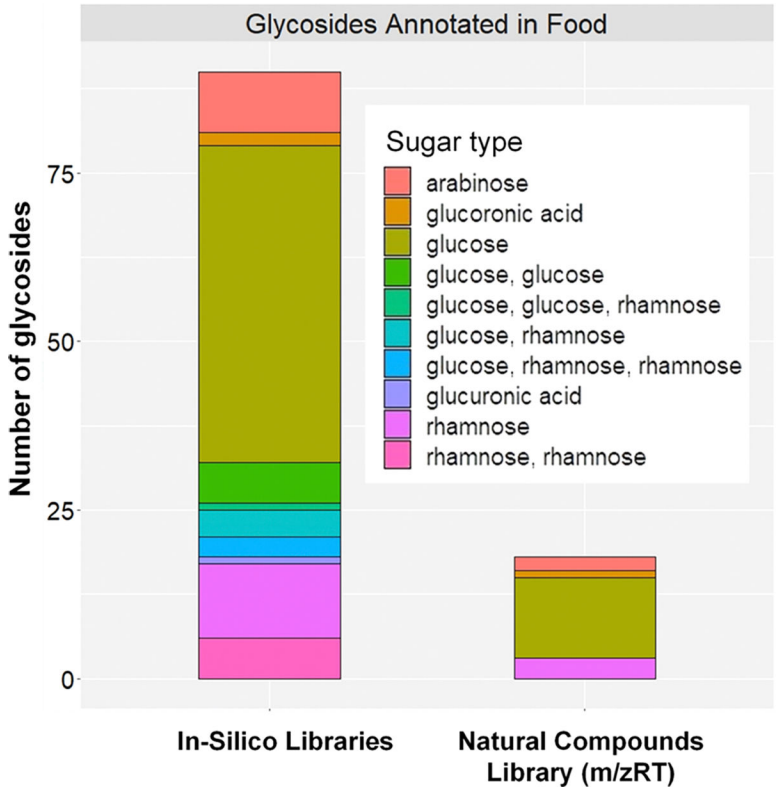


Figure 5.
Bar graph comparison of compounds annotated using novel libraries (left) versus in-house m/zRT libraries (right).