JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

pubs.acs.org/JAFC

GLS-Finder: A Platform for Fast Profiling of Glucosinolates in *Brassica* Vegetables

Jianghao Sun,[†] Mengliang Zhang,[†] and Pei Chen*

Food Composition and Methods Development Laboratory, Beltsville Human Nutrition Research Center, Agricultural Research Service, United States Department of Agriculture, Beltsville, Maryland 20705, United States

Supporting Information

ABSTRACT: Mass spectrometry combined with related tandem techniques has become the most popular method for plant secondary metabolite characterization. We introduce a new strategy based on in-database searching, mass fragmentation behavior study, formula predicting for fast profiling of glucosinolates, a class of important compounds in brassica vegetables. A MATLAB script-based expert system computer program, "GLS-Finder", was developed. It is capable of qualitative and semi-quantitative analyses of glucosinolates in samples using data generated by ultrahigh-performance liquid chromatography–high-resolution accurate mass with multi-stage mass fragmentation (UHPLC–HRAM/MS^n). A suite of bioinformatic tools was integrated into the "GLS-Finder" to perform raw data deconvolution, peak alignment, glucosinolate putative assignments, semi-quantitation, and unsupervised principal component analysis (PCA). GLS-Finder was successfully applied to identify intact glucosinolates in 49 commonly consumed *Brassica* vegetable samples in the United States. It is believed that this work introduces a new way of fast data processing and interpretation for qualitative and quantitative analyses of glucosinolates, where great efficacy was improved in comparison to identification manually.

KEYWORDS: *GLS-Finder, glucosinolates, Brassica, vegetables, UHPLC–HRAM/MS^n*

INTRODUCTION

Glucosinolates are a group of sulfur- and nitrogen-containing secondary metabolites widely distributed in the family of Brassiaceae (syn. Cruciferae), especially for the *Brassica* genus.

The epidemiologic studies showed that the intake of cruciferous vegetables may reduce the risks for occurrence of tumors.[1−4] The primary mechanisms for cancer risk reduction and consumption of cruciferous vegetables are phase I enzyme inhibition and phase II enzyme induction, which may be attributed to the isothiocyanates, the breakdown products of glucosinolates.[5] During food processing, storage, and cooking procedure, glucosinolates are broken down to biologically active compounds, such as isothiocyanates, thiocyanates, indole derivatives, and nitriles, by the mediation of myrosinase.[6−8] Glucosinolates exhibit a common chemical structure of $C_6H_{11}O_5S$ (β-D-1-thioglucopyranose) moiety in the structure. They can be grouped into aliphatic, aromatic, and indolic glucosinolates depending upon whether they originate from aliphatic amino acids (methionine, alanine, valine, leucine, and isoleucine), aromatic amino acids (tyrosine and phenylalanine), or tryptophan.[1,9] To date, about 200 individual glucosinolates have been reported in natural products.[1,9,10]

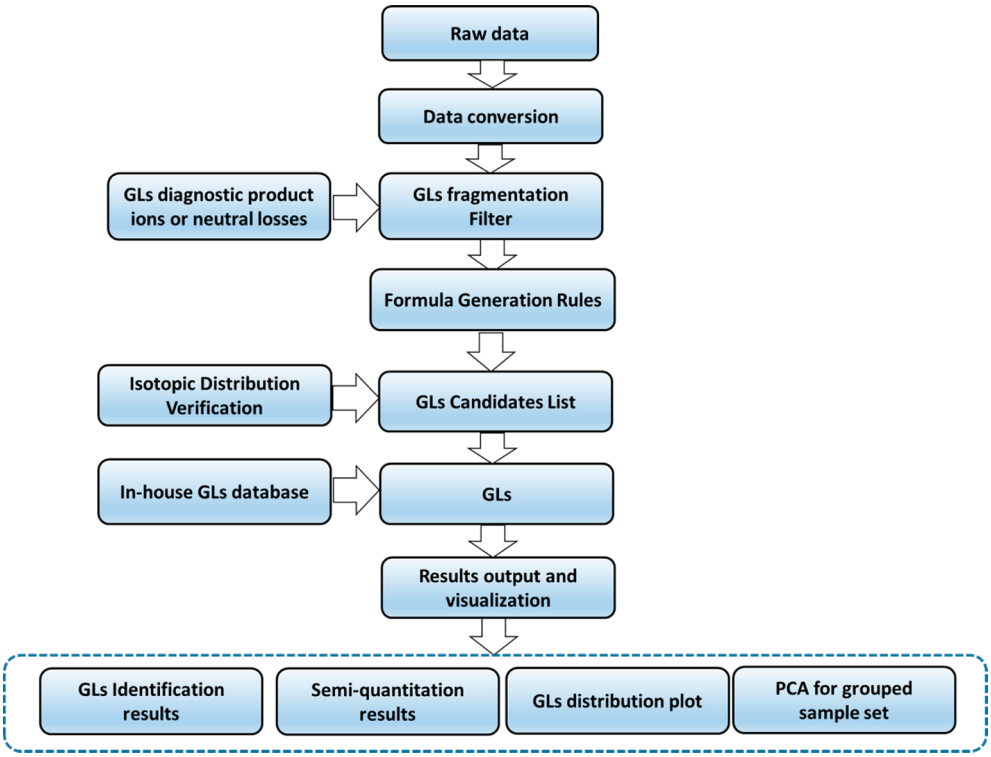
Mass spectrometry (MS) is usually used as one of many key technologies for the identification and structural elucidation of glucosinolates.[11−16] Glucosinolates naturally occur in anion form. Typical MS fragmentation patterns of glucosinolates have been investigated using multi-stage mass spectrometry (MS^n), primarily in the negative ionization mode.[11,13,17−20] With new instrumental technologies, such as ultrahigh-performance liquid chromatography (UHPLC) and high-resolution accurate mass (HRAM), putative identification of glucosinolates can be achieved with much more confidence. Data-dependent acquisition (DDA) using UHPLC–HRAM/MS^n can produce comprehensive data from a single run with minimal user input and interaction, which yields information-rich data in a relatively short period of time. However, the acquired data are in the form of a very complex data array, including diode array detection (DAD) spectra, full-MS spectra, and multi-stage fragmentation spectra (MS^2–MS^4 or even higher) as a function of the retention time. Manual data mining of targeted compounds from such a complex data array, first and foremost, needs highly trained personnel in the field of study. The manual profiling of glucosinolates in huge sample sets is considered to be labor-intensive and error-prone, and the result of the analysis depends highly upon the skills of the person who performs the data mining. Therefore, interpretation of the UHPLC–HRAM/MS^n data is a major bottleneck in the glucosinolate analysis process. Hence, the concept of the development of an "expert system" for rapid analysis of glucosinolates in foods was conceived.

This study describes the development of the GLS-Finder for rapid profiling of intact glucosinolates in *Brassica* vegetables and allows researchers to gain an overview of glucosinolate distribution in one or multiple samples in a short time. The strategy used for the development of the GLS-Finder and the analysis of glucosinolates with UHPLC–HRAM/MS^n data acquired in-house are presented in this study. GLS-Finder can also be used for metabolomics, if desired. The majority of glucosinolates in 49 commonly consumed microgreen and

Received: March 21, 2016
Revised: May 4, 2016
Accepted: May 14, 2016
Published: May 15, 2016

DOI: 10.1021/acs.jafc.6b01277
J. Agric. Food Chem. 2016, 64, 4407−4415

Scheme 1. GLS-Finder Development Strategy for Fast Identification of Glucosinolates



mature vegetable samples in the United States were successfully analyzed using GLS-Finder, and their results were verified manually.

THEORY

Strategy Used for the Development of GLS-Finder. The strategy used for the development of GLS-Finder is shown in Scheme 1. The first step was to study the mass fragmentation behavior of glucosinolates for different collision-induced dissociation (CID) types with a variety of collision energies and then employ the characteristic product ions and neutral losses (NLs) as a preliminary diagnostic tool [glucosinolate fragmentation filter (GFF)] that was coded in our program for targeting the possible glucosinolate ions. The program finds the precursor ions that can produce the characteristic product ions with intensities above the pre-defined threshold using a GFF. Then, GLS-Finder calculates the formula of the precursor ion using glucosinolate formula generation rules based on the HRAM measurement. The possible false-positive identification is excluded by comparing the experimental and measured isotopic distribution. The possible glucosinolate candidates are compared to the in-house glucosinolate database for identification of the known glucosinolates and annotated on the total ion chromatogram.

Glucosinolate peaks found by the program are integrated automatically using the previous method reported[21] and quantified using internal or external standards. For multiple or grouped samples, peak retention time alignment is provided using the correlation optimized warping (COW) method,[22] and the glucosinolate peak lists are exported to a Microsoft Excel spreadsheet, including m/z, molecular formula, retention time, MS^n data, and putative identification. Several visualization tools, such as glucosinolate extraction ion chromatograms, glucosinolate distribution plots, and principal component analysis (PCA), are provided in GLS-Finder for assisting with the interpretation of the raw data.

Construction of a Glucosinolate In-House Database. An extensive literature search was carried out through SciFinder (American Chemical Society, Washington, D.C.) for previous reports of glucosinolates from natural products. The synthetic glucosinolates and their esters are not included. The semi-systematic name, traditional name, CAS registry number, molecular formula, molecular weight, side chain type, and rings and double bond equivalent (RDBE) values were collected and entered into a spreadsheet using Microsoft Excel 2007 (Table S1 of the Supporting Information). A list of calculated accurate masses ($[M - H]^-$) of the 200 known glucosinolates was created on the basis of their elemental composition. The chemical structure of each glucosinolate was downloaded from SciFinder through the web version for exploring the chemical literature and saved in a ".mol" format.

Construction of Formula Generating Rules for Glucosinolates. The R chains may contain double bonds, oxo, hydroxyl, methoxy, benzoyl, indolyl, and carbonyl or disulfide linkages.[1] From the review of the known glucosinolate structures, the rules for calculating the formula of glucosinolate candidates were implemented by GLS-Finder as follows: (a) only five elements C, H, O, N, and S are included, and the numbers of C, H, O, N, and S are in the ranges of 7−30, 12−40, 6−20, 1−3, and 2−6, respectively; (b) the C/H ratios are set ranging from 0.5 to 1; (c) the RDBE values are set ranging from 2.5 to 16.5; and (d) the mass accuracy threshold is set at 5 ppm according to the instrument performance by external calibration.

Construction of a GFF for Raw Data Mining. In tandem mass spectrometry (MS/MS), especially for triple quadrupole mass spectrometers, the precursor ion scan and NL scan are very useful when a particular product ion or NL is characteristic

DOI: 10.1021/acs.jafc.6b01277
J. Agric. Food Chem. 2016, 64, 4407−4415

for a class of compounds. Although the LTQ Orbitrap mass spectrometer used in our study is not able to perform real-time NL and precursor ion scans, a pseudo-precursor ion or NL ion map can be generated by GLS-Finder in the post-data-processing stage. The mass fragmentation behavior of glucosinolates can be considered characteristic because all of the glucosinolates share a core sulfated isothiocyanate group, which is conjugated to thioglucose, and a further R group. This characteristic structure leads to the similar fragmentation behavior in MS, which can produce common product ions or NLs. The mass fragmentation behaviors of glucosinolates were studied extensively to obtain the characteristic product ions, NLs, or combinations of the two. Then, a GFF for extracting the precursor ions from specified common glucosinolate product ions and NLs was developed in the post-data-processing stage. This filter serves as the first explorative tool for identifying glucosinolate candidates from the raw data.

Construction of Glucosinolate Isotopic Distribution Rules. All glucosinolates share a common structure and possess at least two sulfur atoms. The natural ^{34}S distribution is 4.25%, providing a prominent isotopic A + 2 peak (i.e., ~2 mass heavier than monoisotopic ion A should be observed in most MS spectra of glucosinolates). We calculated the theoretical isotope distribution pattern for the glucosinolate candidates based on the International Union of Pure and Applied Chemistry (IUPAC) and provided a code in our program for comparison to the experimental value. An experimental value was checked with known glucosinolates in broccoli vegetables with the ion intensity from 10^4 to 10^9. The relative ion abundance (RIA) errors were calculated as shown in eq 1

$$\mathrm{RIA}_{\mathrm{error}}\ (\%) = 100 \times \frac{\mathrm{RIA}_{\mathrm{exp}} - \mathrm{RIA}_{\mathrm{theo}}}{\mathrm{RIA}_{\mathrm{theo}}} \quad (1)$$

where the theoretical RIA values are calculated by Xcalibur software (version 2.2.0) at infinite resolution.

EXPERIMENTAL SECTION

Chemicals. (−)-Sinigrin hydrate (≥99.0%) was purchased from Sigma-Aldrich (St. Louis, MO). Glucoerucin and glucoraphanin monopotassium salts were obtained from Chromadex (Irvine, CA). LC/MS-grade methanol, acetonitrile (ACN), and formic acid (99%) were from Sigma-Aldrich (St. Louis, MO). Ultrapure water was produced using a Milli-Q RG system from Millipore (Bedford, MA).

Plant Materials. A total of 30 microgreen vegetable samples were provided by the Food Quality Lab of the Agricultural Research Service (ARS), United States Department of Agriculture (USDA). A total of 19 mature vegetable samples were purchased from local supermarkets in Maryland. All samples were lyophilized immediately upon arrival and ground into powder. Powdered samples (300 mg) were extracted in triplicate with 5 mL of methanol/water (60:40, v/v) using sonication (Branson 1200, Branson Ultrasonics, Danbury, CT) for 60 min at room temperature and then centrifuged at 5000*g* for 15 min (IEC Clinical Centrifuge, Damon/IEC Division, Needham, MA). The supernatant was filtered through a 17 mm (0.45 μm) polyvinylidene fluoride (PVDF) syringe filter (VWR Scientific, Seattle, WA), and 2 μL of the extract was used for each UHPLC−HRAM/MS^n analysis.

UHPLC−Photodiode Array (PDA)−Electrospray Ionization (ESI)/HRAM/MS^n Conditions for Glucosinolate Profiling. The UHPLC−HRAM/MS^n system consists of a LTQ Orbitrap XL mass spectrometer with an Accela 1250 binary pump, a PAL HTC autosampler, a PDA detector (ThermoFisher Scientific, San Jose, CA), and an Agilent G1316A column compartment (Agilent, Palo Alto, CA). Separation was carried out on a Hypersil Gold AQ RP-C_{18} UHPLC column (100 × 3.0 mm inner diameter, 1.9 μm, ThermoFisher Scientific) with an UltraShield pre-column filter (Analytical Scientific Instruments, Richmond, CA) at a flow rate of 0.3 mL/min. The mobile phase consists of a combination of A (0.1% formic acid in water, v/v) and B (0.1% formic acid in acetonitrile, v/v). The linear gradient was from 4 to 15% B (v/v) at 10 min to 70% B at 20 min and held at 70% B to 25 min. The re-equilibration time for the initial gradient is 5 min. The MS parameters were optimized using a standard solution of (−)-sinigrin hydrate (100 ng/mL) infused with the initial mobile phase at a flow rate of 0.3 mL/min. The optimized conditions were set as follows: sheath gas at 70, aux and sweep gas at 15, spray voltage at 4.8 kV, capillary temperature at 270 °C, capillary voltage at 15 V, and tube lens at 70 V. The mass range was from 100 to 1500 amu with a resolution of 30 000, FTMS AGC target at 2×10^5, FT-MS/MS AGC target at 1×10^5, isolation width of 1.5 amu, and maximum ion injection time of 500 ms. The most intense ion was selected for the data-dependent scan to offer their MS^2−MS^4 product ions, with a normalization collision energy at 35%.

Raw Data Conversion. GLS-Finder was developed in MATLAB R2012b (MathWorks, Inc., Natick, MA), which currently supports MS data formats netCDF and mzXML. Raw data from different manufacturers of mass spectrometers can be converted into these two common data formats using commercial or open-source tools. In the presented work, the raw data files acquired were converted to mzXML files using ProteoWizard 3.0.6965 (http://proteowizard.sourceforge.net/) with a binary encoding precision of 64 bits, no zlib compression, and MS level from 1 to 5 (means full-scan and MS^2−MS^5 data were converted). A peak-picking filter was used to convert the raw data from profile to centroid. The calculation was performed on a Dell Precision T3610 eight-core Intel Xeon 3.00 GHz processor with 16 GB RAM running a Microsoft Windows 7 Enterprise 64-bit operation system (Microsoft Corp., Redmond, WA).

RESULTS AND DISCUSSION

Typical Fragmentation of Aliphatic, Aromatic, and Indolyl Glucosinolate. Both higher energy collisional dissociation (HCD) and CID were used to examine the fragmentation behaviors of the three classes of glucosinolates under different collision energy levels. The first step was to find the characteristic product ions in MS/MS or MS^n fragmentation. Glucheirolin (3-methylsulfonylpropylglucosinolate), glucotropaeolin (benzylglucosinolate), and glucobrassicin (3-indolylmethyl glucosinolate) represent aliphatic, benzyl, and indolyl glucosinolates, respectively. Consistent with previous reports,[15,18,19,23−25] four product ions, mass-to-charge ratio (m/z) of 195, 241, 259, and 275, can be used as diagnostic ions for glucosinolates in the spectra acquired by the LTQ Orbitrap XL mass spectrometer. The ion at m/z 195 is formed from cleavage of the sulfur−aglycone bond; the ion at m/z 241 is produced by loss of H_2O from m/z 259; and the ion at m/z 259 is formed via cleavage of the sulfur−sugar bond, followed by rearrangement of the sulfate group to the glucose residue. The ion at m/z 275 is formed via the NL of R−N═C═O from the $[M - H]^-$ ion. The NL of 242 Da, which originated from the combined loss of sulfur trioxide (SO_3, 80 Da) and anhydroglucose group ($C_6H_{10}O_5$, 162 Da) is another common fragmentation pathway observed for the glucosinolates. Among all of these characteristic product ions and the NLs, observation of product ions at m/z 259 and 275 in MS^2 spectra can be considered as the most indicative of the existence of the glucosinolates. As shown in Figure S1 of the Supporting Information, the occurrence of the two typical ions can be observed for a wide range of collision energies. A universal normalization fragmentation energy of 35% under CID mode with data-dependent acquisitions was used for all further studies of the Brassicaceae vegetables.

Applying the Fragmentation Filter and the Formula Generating Rules for Glucosinolate. Armed with knowl-

DOI: 10.1021/acs.jafc.6b01277
J. Agric. Food Chem. 2016, 64, 4407−4415

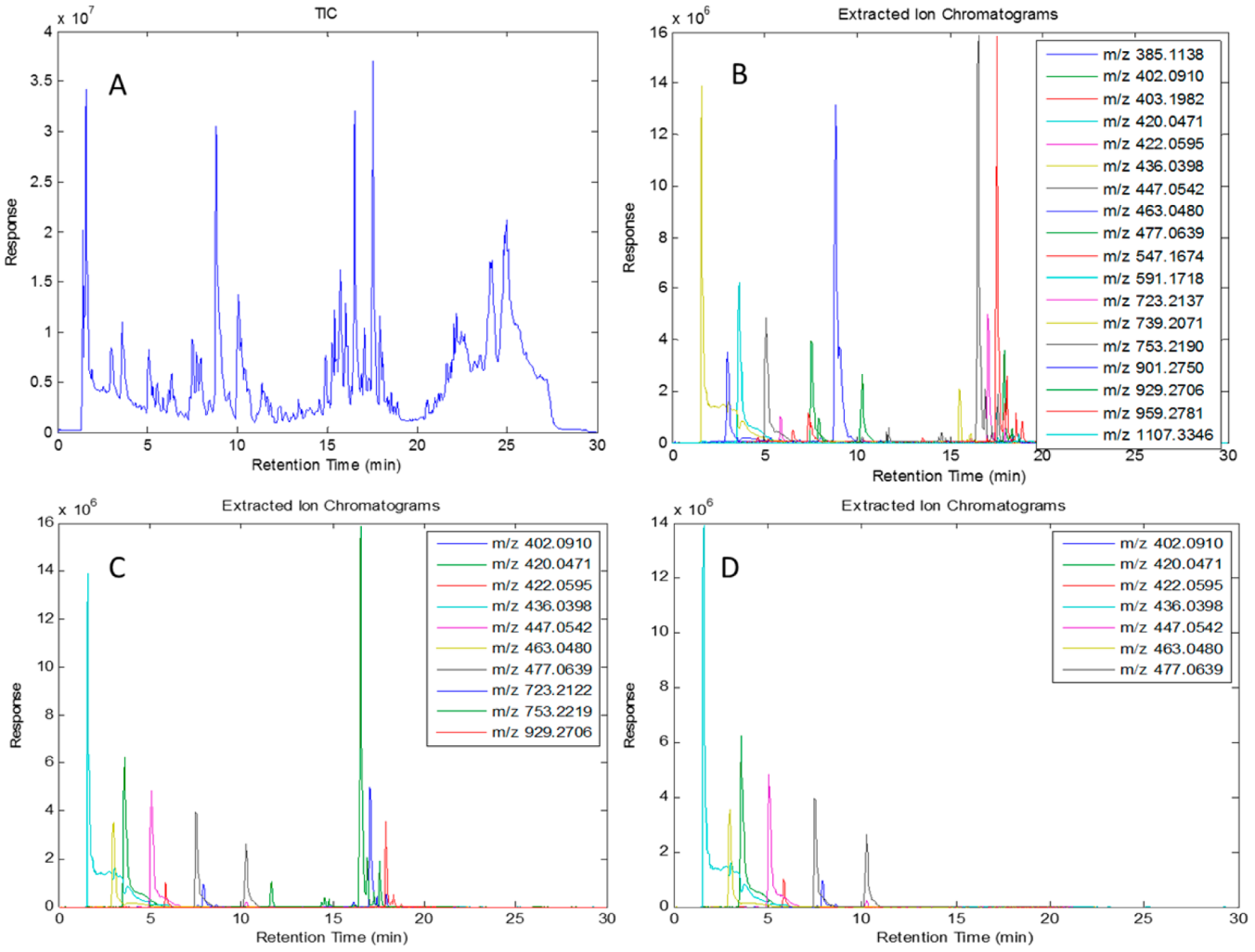


Figure 1. (A) Total ion chromatogram of broccoli microgreen. (B) Extracted ion chromatogram of precursor ions after applying the GFF for NL of 242 Da. (C) Extracted ion chromatogram of precursor ions of glucosinolate candidates after applying the GFF for product ions of m/z 259 and 275. (D) Extracted ion chromatograms for glucosinolate candidates after applying GFF and the glucosinolate formula generation rules.

edge of the fragmentation behavior and diagnostic product ions of glucosinolates, a GFF was used for selecting possible glucosinolate candidates from the raw data. GFF was defined as the selection of characteristic product ions, NLs, or their combinations. From the application of the GFF, GLS-Finder will annotate the scans in the raw data where the typical glucosinolate mass fragmentation occurred. The selection of GFF determines the results of the glucosinolate peak extraction. More accurate results can be achieved if more characteristic product ions and NLs are included. However, the risk of missing minor glucosinolate peaks may increase. The users need to decide the best compromise on how to apply the rules. Using the broccoli microgreen sample as an example, Figure 1A shows the results of applying different GFFs. Figure 1B shows the extracted ion chromatogram of precursor ions that produce typical NL of 242 Da, while Figure 1C shows the extracted ion chromatogram of precursor ions of glucosinolate candidates that produce a NL of 242 Da and product ions of m/z 259 and 275 in MS^2 spectra. False-positive ions may exist during this step, especially if the intensity threshold is set very low at the 10^1 level (as in Figure 1, the intensity threshold is defined as the product ion intensity in the MS^2 scan). The intensity threshold is a user-selectable parameter that can be adjusted to achieve the best compromise between detecting all minor glucosinolates and too many false positives. However, this parameter is not limiting because the false positives can be eliminated easily by applying the glucosinolate formula generating rules. The intensity threshold was set at a very low value in this example to demonstrate the ability of the GLS-Finder at eliminating false positives. Figure 1D shows the extracted ion chromatograms of glucosinolate candidates after applying the rules, and all of the false positives (m/z 723, 753, and 929) were eliminated.

The formula generating rules also provide a more accurate glucosinolate formula compared to the existing formula generation tools included in Xcalibur and are capable of identifying new glucosinolates not present in the database. For example, m/z 509.0751 was a unique glucosinolate found by GLS-Finder in arugula (*Eruca sativa*) that produces characteristic product ions at m/z 259 and 275. Manual generation of formulas using Xcalibur Qual Browser will lead to 18 possible formula hits with less than 5 ppm mass accuracy tolerance if only defining the elements of C, H, O, N, and S. The correct glucosinolate formula ($C_{14}H_{25}O_{14}N_2S_2$) is not the top hit. However, in GLS-Finder, using the glucosinolate formula generating rules, the false formulas are easily eliminated.

Verification of Possible Glucosinolate Candidates Using Isotopic Distribution Rules. If a user still suspects that false positives are included in the results, the isotopic

DOI: 10.1021/acs.jafc.6b01277
J. Agric. Food Chem. 2016, 64, 4407−4415

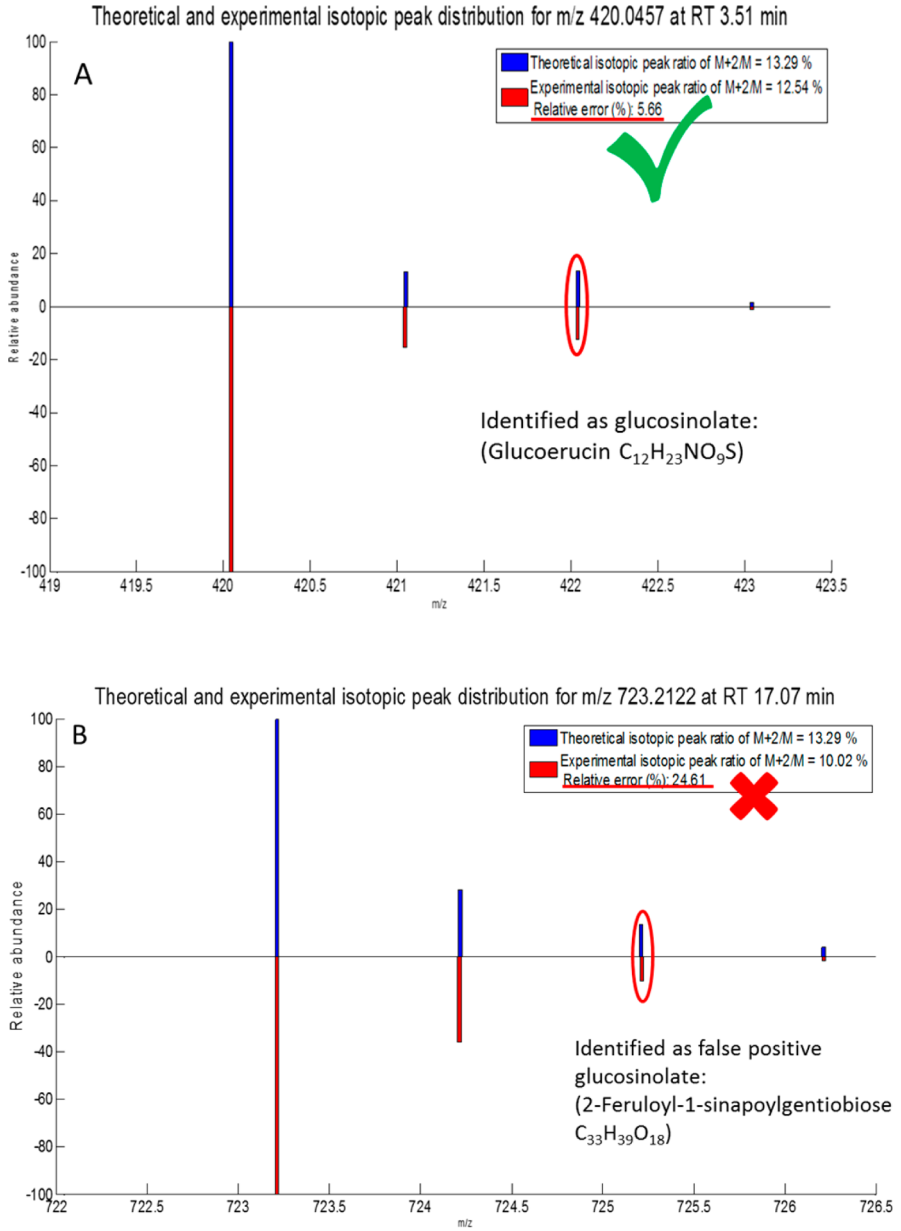
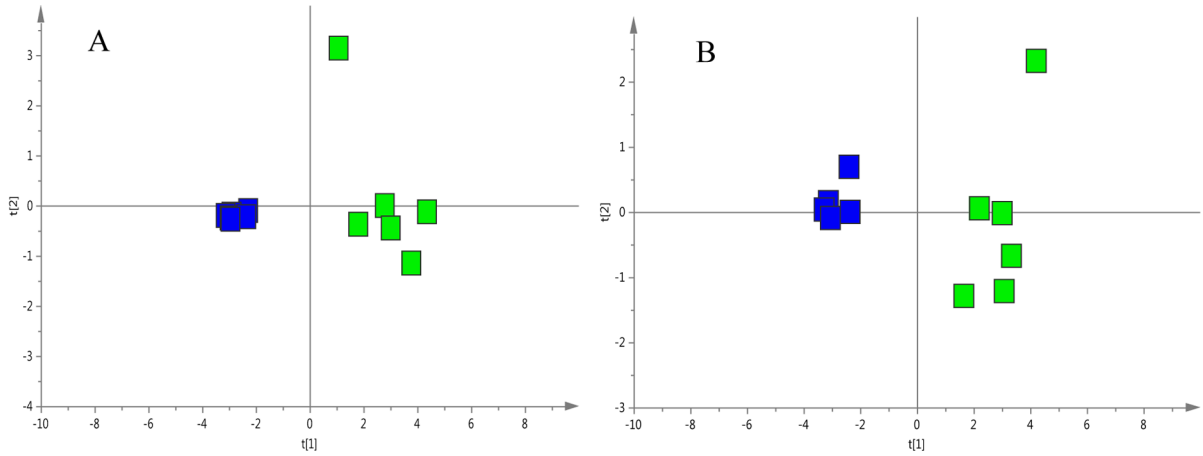


Figure 2. Verification of glucosinolates using isotopic distribution rules.

distribution check function can be used to exclude any remaining false positives. This can be executed automatically or manually in GLS-Finder. In this study, a relatively conservative setting of 20% for positive identification is based on relative isotopic analysis (RIA) of 130 positively identified known glucosinolates with ion intensity from 10^4 to 10^9. As shown in Figure S2 of the Supporting Information, the RIA errors were lower than 15% for glucosinolates containing two, three, and four sulfurs using the mass spectrometer in the lab of the authors. Consequently, for automatic annotation of glucosinolates, the use of the natural isotope patterns has proven to be very useful for excluding false positives. Figure 2A shows that a possible ion at m/z 420.0457 can produce ions m/z 275 and 259 above the predefined intensity threshold. A possible formula with m/z 420.0457 (mass error of 0.06 ppm) was given as $C_{12}H_{23}NO_9S_3$ by GLS-Finder. The measured isotopic distribution was compared to the theoretical distribution with the RIA error of 5.66%. Hence, this peak could be unambiguously identified as a glucosinolate because of the good agreement between theoretical and measured values. Conversely, Figure 2B is an example of a false-positive peak found by the program. The ion at m/z 723.2122 also produces product ions at m/z 275 and 259 (although in low abundance but over the threshold setting of 10^1) in the MS/MS spectrum. It lead to a possible glucosinolate formula of $C_{26}H_{47}O_{15}N_2S_3$ (mass error of 2.30 ppm), However, it does not pass the RIA analysis with the error over 20%, and the formula of this ion most likely has an elemental composition of $C_{33}H_{39}O_{18}$ (2-feruloyl-1-sinapoylgentiobiose) and can be easily excluded as a false positive.

Targeted Glucosinolate Metabolomic Studies for Comparing Glucosinolate Profiles/Levels between Grouped Samples. Metabolomics represents another hot area in plant science. The existing bioinformatic tools for metabolomics are mostly optimized for non-targeted studies and are not flexible enough for targeted analyses aimed at

DOI: 10.1021/acs.jafc.6b01277
J. Agric. Food Chem. 2016, 64, 4407−4415



C

	m/z	Retention	Treatment1	Treatment1	Treatment1	Treatment1	Treatment1	Treatment1	Treatment2	Treatment2	Treatment2	Treatment2	Treatment2	Treatment2
1	388.0735	7.63	2658716	2822291	3570856.6	2427005.5	3130595	2598274	335626.6	85978.628	50835.78	62270.94	338932.3	107829.7
2	402.0891	14.82	3787787	3833190	4226840.5	3334291.7	3816993.7	3583169	622312.8	210132.4	198144.8	319081.7	680978.3	312025.1
3	406.0306	3.6	9952660	12926547	19228187	9170820	12728555	13299261	1130303	285543.62	166292.3	161889.3	1323988	380882
4	420.0447	5.74	26967878	41226749	50465835	29678349	55594643	40854933	14805939	4208211	2633697	4220287	11450038	4204534
5	420.0447	7.17	325899.1	439610.7	590059.13	603678	548225	912105.2	78282.91	23089.699	6435.524	19274.37	82255.31	37174.87
6	422.0231	1.84	13921823	15297133	15903483	9573237	12887706	15274179	3232182	628143.52	285686.2	511723.2	3374967	776093.9
7	434.0614	10.67	820586.2	1086644	1351503.2	798936.96	1195161.6	1259070	87012.98	12011.38	4801.389	11503.82	62846.85	24901.04
8	436.0383	1.91	33175414	41813182	52219378	22563567	45621579	45204266	8380704	2344186.8	1521468	3251968	8851945	3142983
9	447.0533	6.98	6078274	6913928	8270037.9	4112670.2	5046258.8	5679806	1841632	660399.11	386203.5	590229.7	1231521	564865.7
10	463.0486	3.51	12072791	9683146	13966248	8728340.8	10145004	3470454	2664323	2019012.9	1135586	2006004	2460466	1246761
11	477.0636	12.21	22509813	22320168	57291534	9657926.5	13171369	16056507	52532277	44875730	44045477	40547390	40816337	41671395

Figure 3. PCA score plot based on (A) manually integrated peak area of raw data and (B) peak area from GLS-Finder and (C) sample summary report of GLS-Finder. Blue, $CaCl_2$-treated group; green, control group.

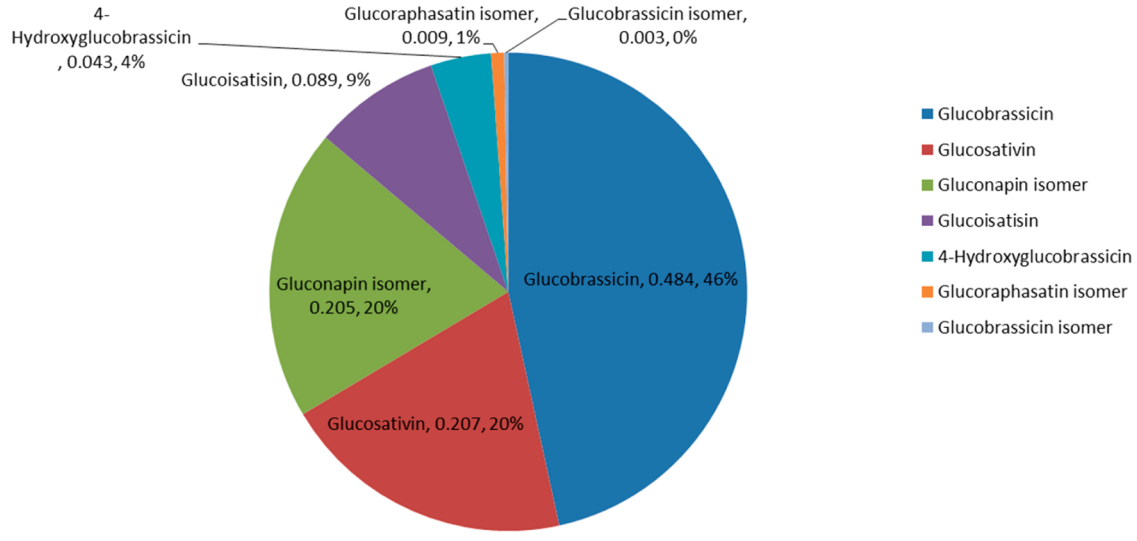


Figure 4. Typical glucosinolate percentage plot from cauliflower with semi-quantitation results (μmol/g of DW).

specific classes of compounds. GLS-Finder, on the other hand, can be used for targeted metabolomics to evaluate glucosinolate levels or profile changes between groups of samples. Figure 3A shows the PCA score plot of two group of broccoli samples with different treatments ($CaCl_2$-treated and control groups) based on 11 glucosinolates and the peak areas manually checked and integrated by a human using Quanbrowser in Xcalibur software from an instrument vendor. Figure 3B is the PCA score plot based on the same 11 glucosinolates and the peak areas exported by GLS-Finder. The first two principal components explained 93.5% (R_2X) total variance with the predictability of 77.2% (Q^2) for using the glucosinolate peak areas from raw data, while the first two components explained 95.1% (R_2X) total variance with the predictability of 78.9%

DOI: 10.1021/acs.jafc.6b01277
J. Agric. Food Chem. 2016, 64, 4407−4415

Glucosinolates found | MS^n data | Possible Identification and CAS number

Feature ID	Mass Feat	Mass Feat	ntensity	Formula	Mass Erro	RDBE	MS2	MS3	MS4		Name	CAS number
1	1.62	388.0361	45599932	C11H18O1	2.796615	3.5	(@388.0361) 259	(@259.0872) 138.980	(@138.980	4) 82.0	2-Hydroxy-3-butenylglucosinol	31362-92-2; 19237-18-4; 585-95
2	1.71	372.0412	9207493	C11H18O9	3.050165	3.5	(@372.0412) 259	(@259.0635) 96.9596		0	2-Methyl-2-propenyl glucosino	956700-01-9; 19041-09-9
3	1.99	402.0514	17321818	C12H20O1	3.672163	3.5	(@402.0514) 259	(@259.1354) 139.111	(@139.11	2) 81.1	2-Hydroxy-3-pentenyl glucosin	1185047-99-7; 956700-09-7; 137
4	2.09	372.0412	31156858	C11H18O9	3.050165	3.5	(@372.0412) 259	(@259.1504) 139.075	(@139.07	9) 81.0	2-Methyl-2-propenyl glucosino	956700-01-9; 19041-09-9
5	2.33	388.0361	4652550	C11H18O1	2.796615	3.5	(@388.0361) 259	0		0	2-Hydroxy-3-butenylglucosinol	31362-92-2; 19237-18-4; 585-95
6	2.52	374.0573	9444107	C11H20O9	1.632024	2.5	(@374.0573) 259	(@259.1338) 97.0321		0	2-Methylpropyl glucosinolate;	38226-91-4; 35535-42-3; 499-24
7	2.81	388.0361	7120505	C11H18O1	2.796615	3.5	(@388.0361) 259	(@259.1045) 97.0223		0	2-Hydroxy-3-butenylglucosinol	31362-92-2; 19237-18-4; 585-95
8	3.02	463.0465	12040203	C12H23O7	3.402048	3.5	(@463.0465) 285	(@285.1239) 97.0878	(@97.087	8) 81.08	1-Methoxyl-3-indolyl glucosino	355116-49-3; 87592-99-2; 83327
9	3.21	386.0562	76935608	C12H20O9	4.506134	3.5	(@386.0562) 259	(@259.2003) 139.078	(@139.07	1) 81.1	1-Pentenyl-glucosinolate; 4-Pe	956700-10-0; 19041-10-2; 6737-
10	3.87	386.0562	3164238	C12H20O9	4.506134	3.5	(@386.0562) 259	(@259.0663) 139.028		0	1-Pentenyl-glucosinolate; 4-Pe	956700-10-0; 19041-10-2; 6737-
11	3.97	386.0562	3122898	C12H20O9	4.506134	3.5	(@386.0562) 259	(@259.1966) 169.127		0	1-Pentenyl-glucosinolate; 4-Pe	956700-10-0; 19041-10-2; 6737-
12	4.73	386.0562	7945918	C12H20O9	4.506134	3.5	(@386.0562) 259	0		0	1-Pentenyl-glucosinolate; 4-Pe	956700-10-0; 19041-10-2; 6737-
13	4.87	438.0531	6839088	C15H20O1	0.530954	6.5	(@438.0531) 259	(@259.1156) 139.084		0	1-Hydroxy-2-Phenyl-ethyl gluc	827322-02-1; 666235-43-4; 6662
14	5.07	447.05	56574860	C12H23O6	0.109607	3.5	(@447.0500) 259	(@259.1135) 139.134	(@139.13	1) 104.	3-Indolylmethyl glucosinolate	4356-52-9
15	5.75	447.0546	6177312	C16H19O9	3.049806	8.5	(@447.0546) 259	0		0	3-Indolylmethyl glucosinolate	4356-52-9
16	5.83	422.0584	17541916	C15H20O9	1.084316	6.5	(@422.0584) 259	(@259.1642) 139.079		0	2-Phenethyl glucosinolate	499-30-9
17	5.96	447.0546	3899087	C16H19O9	3.049806	8.5	(@447.0546) 259	0		0	3-Indolylmethyl glucosinolate	4356-52-9
18	7.13	447.0546	156007.9	C16H19O9	3.049806	8.5	(@447.0546) 401	0		0	3-Indolylmethyl glucosinolate	4356-52-9
19	7.48	477.06	31666584	C12H21O1	0.534524	4.5	(@477.0600) 259	(@259.0482) 139.083		0	7-Methoxyglucobrassicin; 5-Me	1143960-20-6; 87593-00-8; 8332
20	8.57	402.0894	1764749	C13H24O9	0.406636	2.5	(@402.0894) 259	0		0	2-Ethylbutyl glucosinolate; 3-M	1160697-39-1; 956700-00-8; 138
21	10.27	477.0648	12573851	C17H21O1	2.269825	8.5	(@477.0648) 446	(@446.1096) 284.178		0	7-Methoxyglucobrassicin; 5-Me	1143960-20-6; 87593-00-8; 8332
22	13.17	492.0637	2011130	C18H22O1	0.460648	8.5	(@492.0637) 259	0		0	2-Benzoyloxy-3-butenyl glucosi	356534-71-9

Figure 5. Typical report of GLS-Finder on turnip microgreen.

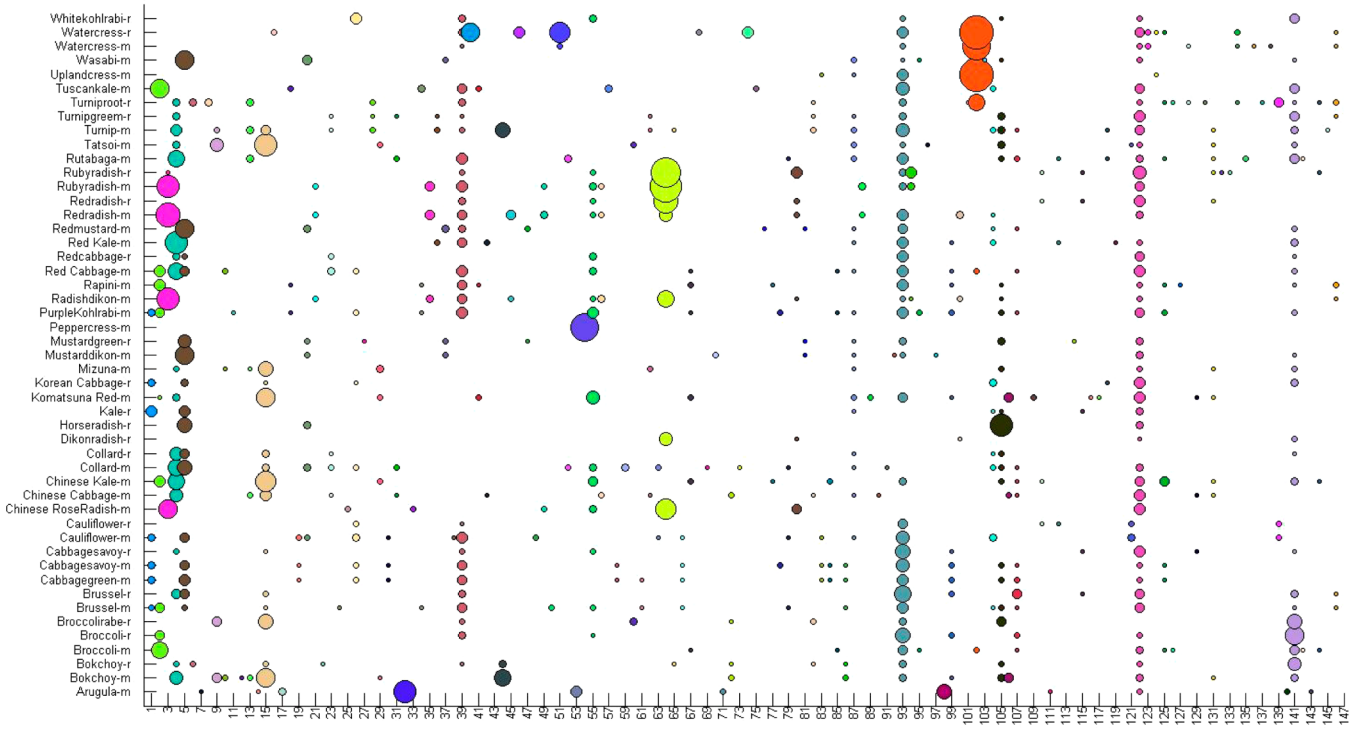


Figure 6. Distribution plot of 146 glucosinolates in the 49 vegetables analyzed. Each dot represents one glucosinolate, and dot size represents the content of glucosinolates (the larger the dot size, the larger the glucosinolate content) based on semi-quantitation results using glucoiberin. Different glucosinolates were labeled in different colors. m, microgreen; r, mature vegetable.

(Q^2) for using the data from GLS-Finder. These two PCA score plots exhibit very similar data clustering trends, and it suggested that GLS-Finder keeps the original data characteristics from the raw data. GLS-Finder can also output a summarized spreadsheet (Figure 3C) with retention times, HRAM data, and peak areas that can be used for further chemometric analysis with commercial third-party software.

Glucosinolate Profiling and Semi-quantification. GLS-Finder can perform automated semi-quantitation using either internal or external standards. It also gave the individual/total glucosinolate peak area ratio of a sample for fast estimation of glucosinolate content. Figure 4 shows a typical glucosinolate profile with the percentage of each individual glucosinolate. The total glucosinolate concentration of cauliflower extract was 1.04 μmol/g of dry weight (DW), which was calculated using

DOI: 10.1021/acs.jafc.6b01277
J. Agric. Food Chem. 2016, 64, 4407−4415

glucoiberin as an external standard curve. This function can give researchers a general idea of the glucosinolate content in a very short time. More accurate results can be obtained with standard curves of individual glucosinolates (if reference standards are available) in a quantitative study if necessary.

Applying GLS-Finder to Different Kinds of Samples. A total of 49 samples, including 30 microgreens and 19 mature vegetables commonly consumed in the U.S. market, were screened using the "GLS-Finder" program. A typical report is shown in Figure 5. The information on glucosinolate retention time, intensity, accurate mass, major MS^{2-n} fragment ions, CAS registry number, and possible identification based on in-house database search were reported in the spreadsheet.

GLS-Finder can identify glucosinolates found in the in-house database very quickly in all samples studied, and it can also help identify new glucosinolates. For example, two peaks with retention times at 1.68 and 2.17 min with a deprotonated ion at m/z 580.0438 were determined to be glucosinolates by the program but not found in the in-house database. They were flagged as potential new glucosinolates. Further manual examination found base fragmentation ions at m/z 406 in their MS^2 spectra. Their MS^3 spectra were identical to the MS^2 spectra of 4-mercaptobutyl glucosinolate.[17] In comparison to 4-mercaptobutyl glucosinolate ($C_{11}H_{20}O_9NS_3$), these two compounds have an additional $-C_{10}H_6OS$ side chain (174.0139 Da) attached to the structure. They are tentatively identified as thieno[3,2-*e*]benzofura-*n*-mercaptobutyl glucosinolate and its isomer. These two glucosinolates have not been reported previously in the literature.

The typical run time for GLS-Finder for one sample (with 30 min elution time) is around 10 min compared to days for manual data mining. The program greatly facilitates data analysis of glucosinolates, displaying them as overlays of multiple extracted ions (Figure 1D). The vegetables show a great diversity in glucosinolate composition and content. Figure 6 provides a general distribution plot of the glucosinolates in the 49 vegetables investigated. The x axis represents the 146 glucosinolates found by GLS-Finder, and the y axis represents each vegetable sample investigated. Each dot represents a glucosinolate found in the vegetable, and the dot size represents the glucosinolate content from the semi-quantitation results. The total content of glucosinolates ranges from 1.04 to 9.87 μmol/g of DW among the 49 vegetables (Table S2 of the Supporting Information). With the aid of this distribution figure, it would be easy to compare glucosinolate profiles between investigated vegetables to find out the common glucosinolate distribution. Also, it can pinpoint unique glucosinolates determined, such as dimeric 4-mercaptobutyl glucosinolate and 10-methylthiodecyl glucosinolate in arugula, 6-methylthiohexyl glucosinolate in turnip root, etc.

In conclusion, a computational program "GLS-Finder" for fast detection and identification of glucosinolates in commonly consumed *Brassica* vegetables using UHPLC−HRMS data has been developed. GLS-Finder is able to process single data, multiple data, and grouped data and provides glucosinolate detection, visualization, peak list alignment, putative assignment, semi-quantitative results, and PCA. It provides additional unique features for finding new potential glucosinolates. GLS-Finder can also be modified to accommodate nominal MS data but with less accurate results. It saves a significant amount of time in data mining. The general design idea of our program is not limited to glucosinolates. It can be extrapolated to other plant secondary metabolites if they share common mass fragmentation behavior.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jafc.6b01277.

Occurrence of characteristic ions under different collision energy and CID type (Figure S1) and RIA error of glucosinolates with ion intensity ranging from 10^4 to 10^9 (Figure S2) (PDF)

In-house database of glucosinolates used for GLS-Finder (Table S1) and major glucosinolates and distributions in *Brassica* microgreen and mature vegetables (Table S2) (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*Telephone: +1-301-504-8144. Fax: +1-301-504-8314. E-mail: pei.chen@ars.usda.gov.

Author Contributions

†Jianghao Sun and Mengliang Zhang contributed equally to this work.

Funding

This research is supported by the Agricultural Research Service of the United States Department of Agriculture and an Interagency Agreement with the Office of Dietary Supplements at the National Institutes of Health.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Fahey, J. W.; Zalcmann, A. T.; Talalay, P. *Phytochemistry* **2001**, *56*, 5−51.

(2) Steinbrecher, A.; Nimptsch, K.; Husing, A.; Rohrmann, S.; Linseisen, J. *Int. J. Cancer* **2009**, *125*, 2179−2186.

(3) Keck, A. S.; Finley, J. W. *Integr. Cancer Ther.* **2004**, *3*, 5−12.

(4) Tawfiq, N.; Heaney, R. K.; Plumb, J. A.; Fenwick, G. R.; Musk, S. R.; Williamson, G. *Carcinogenesis* **1995**, *16*, 1191−1194.

(5) Tang, L.; Zhang, Y. *Curr. Drug Metab.* **2004**, *5*, 193−201.

(6) Song, L.; Thornalley, P. J. *Food Chem. Toxicol.* **2007**, *45*, 216−224.

(7) Fahey, J. W.; Zhang, Y.; Talalay, P. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 10367−10372.

(8) Nestle, M. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 11149−11151.

(9) Clarke, D. B. *Anal. Methods* **2010**, *2*, 310−325.

(10) Agerbirk, N.; Olsen, C. E. *Phytochemistry* **2012**, *77*, 16−45.

(11) Njumbe Ediage, E.; Diana Di Mavungu, J.; Scippo, M. L.; Schneider, Y. J.; Larondelle, Y.; Callebaut, A.; Robbens, J.; Van Peteghem, C.; De Saeger, S. *J. Chromatogr. A* **2011**, *1218*, 4395−4405.

(12) Zhang, J.; Sun, X.; Zhang, Z.; Ni, Y.; Zhang, Q.; Liang, X.; Xiao, H.; Chen, J.; Tokuhisa, J. G. *Phytochemistry* **2011**, *72*, 1767−1778.

(13) Bianco, G.; Lelario, F.; Battista, F. G.; Bufo, S. A.; Cataldi, T. R. *J. Mass Spectrom.* **2012**, *47*, 1160−1169.

(14) Kaufmann, A.; Walker, S. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 1081−1090.

(15) Zandalinas, S. I.; Vives-Peris, V.; Gomez-Cadenas, A.; Arbona, V. *J. Agric. Food Chem.* **2012**, *60*, 8648−8658.

(16) Bianco, G.; Agerbirk, N.; Losito, I.; Cataldi, T. R. *Phytochemistry* **2014**, *100*, 92−102.

(17) Lee, K. C.; Cheuk, M. W.; Chan, W.; Lee, A. W.; Zhao, Z. Z.; Jiang, Z. H.; Cai, Z. *Anal. Bioanal. Chem.* **2006**, *386*, 2225−2232.

(18) Fabre, N.; Poinsot, V.; Debrauwer, L.; Vigor, C.; Tulliez, J.; Fouraste, I.; Moulis, C. *Phytochem. Anal.* **2007**, *18*, 306−319.

DOI: 10.1021/acs.jafc.6b01277
J. Agric. Food Chem. 2016, 64, 4407−4415

(19) Rochfort, S. J.; Trenerry, V. C.; Imsic, M.; Panozzo, J.; Jones, R. *Phytochemistry* **2008**, *69*, 1671−1679.
(20) Maldini, M.; Baima, S.; Morelli, G.; Scaccini, C.; Natella, F. *J. Mass Spectrom.* **2012**, *47*, 1198−1206.
(21) Zhang, M.; Sun, J.; Chen, P. *Anal. Chem.* **2015**, *87*, 9974−9981.
(22) Tomasi, G.; van den Berg, F.; Andersson, C. *J. Chemom.* **2004**, *18*, 231−241.
(23) Cataldi, T. R.; Lelario, F.; Orlando, D.; Bufo, S. A. *Anal. Chem.* **2010**, *82*, 5686−5696.
(24) Lin, L. Z.; Sun, J.; Chen, P.; Zhang, R. W.; Fan, X. E.; Li, L. W.; Harnly, J. M. *J. Agric. Food Chem.* **2014**, *62*, 6118−6129.
(25) Glauser, G.; Schweizer, F.; Turlings, T. C.; Reymond, P. *Phytochem. Anal.* **2012**, *23*, 520−528.

DOI: 10.1021/acs.jafc.6b01277
J. Agric. Food Chem. 2016, 64, 4407−4415