analytical chemistry

Article

pubs.acs.org/ac

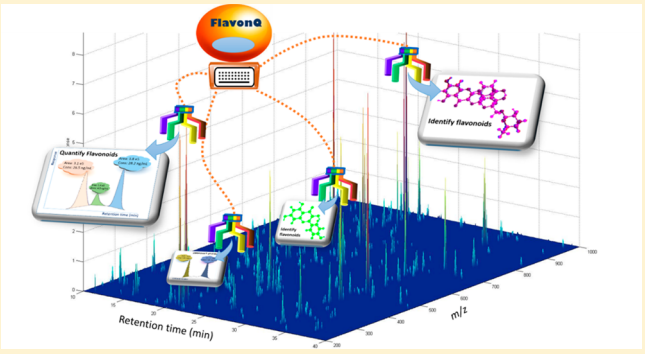
FlavonQ: An Automated Data Processing Tool for Profiling Flavone and Flavonol Glycosides with Ultra-High-Performance Liquid Chromatography−Diode Array Detection−High Resolution Accurate Mass−Mass Spectrometry

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

Food Composition and Methods Development Lab, Beltsville Human Nutrition Research Center, Agricultural Research Services, United States Department of Agriculture, Beltsville, Maryland 20705-2350, United States

Supporting Information

ABSTRACT: Profiling flavonoids in natural products poses a great challenge due to the diversity of flavonoids, the lack of commercially available standards, and the complexity of plant matrixes. The increasingly popular use of ultra-high-performance liquid chromatography−diode array detection−high resolution accurate mass−mass spectrometry (UHPLC-HRAM-MS) for the analysis of flavonoids has provided more definitive information but also vastly increased amounts of data. Thus, mining of the UHPLC-HRAM-MS data is a very daunting, labor-intensive, and expertise-dependent process. An automated data processing tool, FlavonQ, was developed that can transfer field-acquired expertise into data analysis and facilitate flavonoid research. FlavonQ is an "expert system" designed for automated data analysis of flavone and flavonol glycosides, two important subclasses of flavonoids. FlavonQ is capable of data format conversion, peak detection, flavone and flavonol glycoside peak extraction, flavone and flavonol glycoside identification, and production of quantitative results. An expert system was applied to the determination of flavone and flavonol glycosides in nine different plants with an average execution time of less than 1 min. The results obtained by FlavonQ were in good agreement with those determined conventionally by a flavonoid expert.



Flavonoids are phenolic compounds found in nearly every plant and constitute one of the three most important natural pigments.[1] In addition, flavonoids are well-known for their health benefits.[2] The determination of flavonoids is very challenging because more than 5000 have been reported in plants.[3] Although many methods have been developed for the identification of flavonoids, few validated and standardized methods are available.[4] Liquid chromatography (LC) with diode array detector (DAD) and/or mass spectrometric (MS) detection is frequently used for determination of flavonoids.[5−7] However, the majority of the published quantitation methods have, in general, focused on only a few flavonoids (usually less than a dozen) due to the limited availability of reference standards.

Hydrolyzing the glycosides to aglycones is another approach;[8] however, control of the hydrolysis process is, in practice, very difficult as the aglycones themselves are not equally stable under hydrolysis conditions. Quantitation of flavonoids can also be carried out using similar compounds as standards;[9] however, the process can be difficult to employ. To date, methods based on ultra-high-performance liquid chromatography combined with diode array detection and high-resolution accurate-mass mass spectrometry (UHPLC-DAD-HRAM MS) have come closest to being a universal approach for profiling flavonoids in plant materials.[10] However, manual interpretation of information-rich UHPLC-DAD-HRAM MS data is very time-consuming (weeks to months for a single analysis), highly expertise-dependent, and error prone.

The extraction of information from complex LC-MS data has always been a challenging process. The "omics" tools, especially metabolomics tools, are typically applied to food analysis at the molecular level.[11] The currently available metabolomics programs such as XCMS,[12] mzMine,[13] and MAVEN[14] are inadequate when applied to most plant analyses due to the differences in research focus. For example, untargeted metabolomics often focuses on metabolite profiling and biomarker discovery between treatments which is not suitable for the analysis of a class of compounds, such as flavonoids. Targeted metabolomics mainly uses predefined metabolite-specific signals such as selected reaction monitoring (SRM) in

Received: July 9, 2015
Accepted: September 11, 2015
Published: September 11, 2015

DOI: 10.1021/acs.analchem.5b02624
Anal. Chem. 2015, 87, 9974−9981

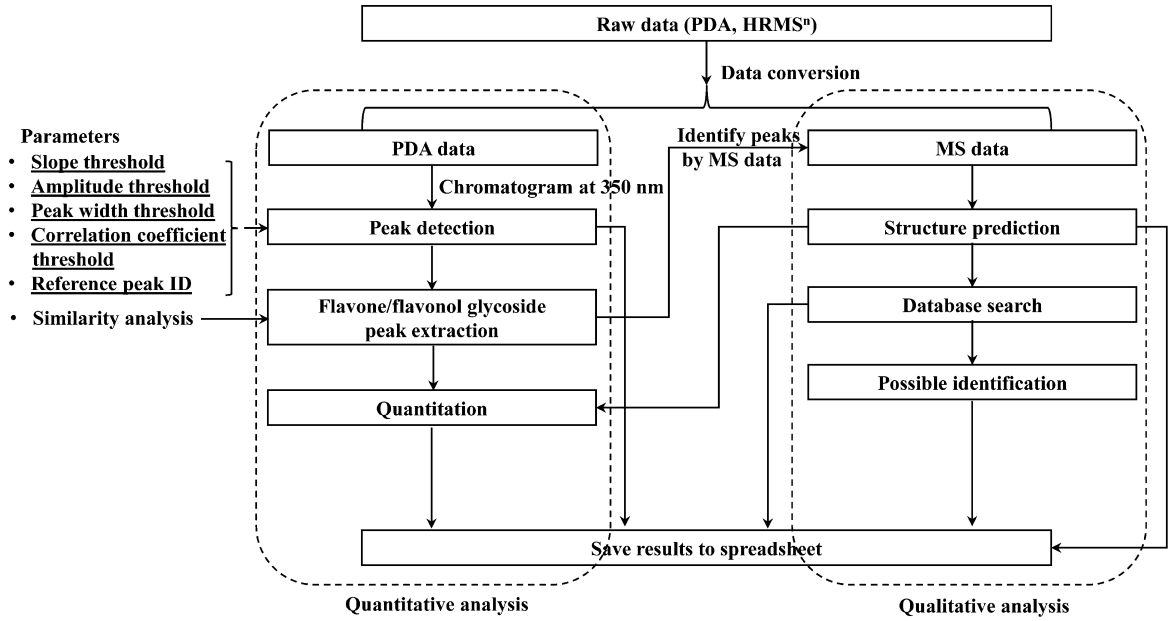


Figure 1. Flowchart showing the general strategy for processing and analysis of HPLC-DAD-HRMSⁿ data for flavone/flavonol glycosides. Only one raw data file is read at a time. Iterations of multiple data files are possible, as necessary.

tandem mass spectrometry for known metabolites, not designed to elucidate unknown chemical identities.[15]

A universal method for the analysis of flavonoids in plants should ultimately identify and quantify all flavonoids. The ideal method should satisfy the following criteria: (a) detect every chromatographic peak above a set threshold (with various widths, heights, and shapes), (b) identify all of the flavonoid peaks (a subset of the peaks from the previous step), (c) elucidate flavonoid structures based on the MS spectra and databases, and (d) provide the results in user-friendly formats.

FlavonQ was developed as the first step toward the ultimate goal for flavonoid analysis and is focused on two subclasses of flavonoids, the flavone and flavonol glycosides. A new strategy was proposed for putative identification and quantitation of flavone/flavonol glycosides using UHPLC-DAD HRAM MSⁿ (multistage MS). The flavone/flavonol glycoside chromatographic peaks were extracted from DAD chromatograms based on their characteristic UV absorbance (UV 350 nm) and were putatively identified and quantified. Compared with DAD or UV detectors, MS is less stable and its response can be greatly affected by the ionization efficiency of the analytes and the matrixes and by environmental conditions.[16,17] UV absorbance at 350 nm was used for quantifying the flavone/flavonol glycosides, providing a simpler algorithm, and more reliable results. This approach was applied to the analysis of flavone/flavonol glycosides in nine different plants with minimal user input.

THEORY

Overview. FlavonQ uses a novel strategy to extract chromatographic peaks and predict structures for flavone/flavonol glycoside. The flowchart of FlavonQ is shown in Figure 1. Here only the strategies of peak detection and peak area measurement, UV−vis spectrum similarity analysis, and the flavone/flavonol glycoside identification process will be introduced in detail. The other steps will be discussed in Results and Discussion.

Peak Detection and Peak Area Measurements. A typical chromatogram from a plant extract may consist of hundreds of peaks.[10] In open source software packages for metabolomics such as XCMS,[12] mzMine,[13] and MAVEN,[14] several peak detection methods have been used to detect peaks from extracted ion chromatograms (EIC). It is noted that the "peak detection" discussed in this work is different from those MS-based metabolomics software. In the MS-based metabolomics software, the peak detection often involves mass slice detection, extraction of ion-specific chromatograms, baseline detection for each EIC, data smoothing, peak picking and grouping, and retention time alignment. In our study, peaks are detected from UV chromatograms (350 nm) and their positions, heights, widths, and areas are measured.

The zero-crossing points of the first or second derivatives of the chromatogram curve are commonly used to detect the borders of individual peaks.[12,13] Inherent background noises can generate many zero-crossing points and may be recognized as "peaks" by peak detection algorithms. Thus, smoothing is need for reliable peak detection. Several smoothing filters are available in FlavonQ. One of the simplest smoothing algorithms is the unweighted sliding average filter. The data points in the signal (y) are replaced by the average of m adjacent points.[18] The weighted moving average filter is similar to the unweighted moving average filter except that it gives different weights to the adjacent points before calculating the average. Usually the central point gains more weights than its neighbors on both sides. Another filter, the pseudo-Gaussian smooth which is a heavier smoothing method, can be considered as three passes of the unweighted moving average filter. It can be used when the signal contains a high frequency of noise. The last smoothing filter, the Savitzky-Golay filter, estimates the data points based on the polynomial fitting by least-squares from the adjacent points.[18] In FlavonQ, users can choose the most suitable filter for peak detection according to the data and noise levels. In this study, the weighted moving average filter was used for smoothing all the chromatograms because it was more efficient in noise reduction.

DOI: 10.1021/acs.analchem.5b02624
Anal. Chem. 2015, 87, 9974−9981

After smoothing, the first derivative of the smoothed chromatogram was used for peak detection by finding the zero-crossing points. Three user-controllable parameters, "slope threshold", "amplitude threshold", and "peak width threshold", are introduced to detect the desired peaks. The "slope threshold" of the zero-crossing points can determine the borders of peaks which is especially useful for the non-Gaussian shape peaks with serious fronting and tailing. It can also be used to ignore the peaks which are too small or too wide by setting at a larger value. The extremely small or narrow peaks can be excluded by the setting of "amplitude threshold" and "peak width threshold". Figure 2A shows an example of the peak detection results using our program for red mustard green sample.

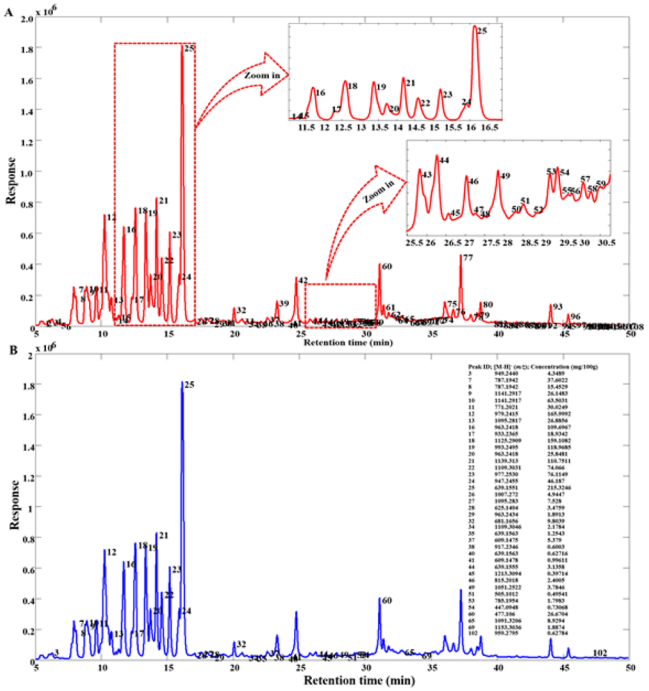


Figure 2. Peak detection result (A) and identification and quantitation result of flavone/flavonol glycosides by FlavonQ (B) for UV chromatogram at 350 nm for red mustard green sample.

The peak area measurement is a crucial step for quantitation. For well-separated peaks, the integration is calculated simply by summing the signals from the start to the end of a peak which are predetermined in peak detection. In practice, background noise is ubiquitous and overlapping peaks are inevitable for complex samples such as crude plant extracts. In manual integrations, the drop perpendicular method and valley method are widely accepted for overlapping peaks. The drop perpendicular method finds the projection of valley to the baseline and integrates the first peak from peak start to the vertical line and the second peak from vertical line to peak end. The valley method integrates each peak by setting start and stop points at the valley between the peaks (Figure S1, Supporting Information). The more state-of-the-art strategy for peak area measurement used in some metabolomics software such as XCMS applies second-derivative Gaussian transformation to define peak borders by zero-crossing points and usually the peak tails which may overlap with other peaks that are not integrated.[12] The ideal method to measure each peak area is to deconvolute peaks with a fitted peak model such as Gaussian or Lorentzian or exponentially modified Gaussian. It is necessary to predetermine the fundamental peak shape as a function of retention time before deconvoluting the overlapping peaks; otherwise, the computation time could be very long for optimization of the peak model by an iterative process. However, it is usually an insurmountable task to predetermine the peak shapes for hundreds of chromatographic peaks. To simplify the task, a "Gaussian fit method" is available in our program which measures peak areas, assuming each peak has a Gaussian shape.

All four peak area measurement methods discussed above are implemented in the program for users to choose, and an example of comparing peak area results by these methods with manual integration in Xcalibur software (Thermo Fisher Scientific, Inc., San Jose, CA) is shown in Table S1, Supporting Information. The peak area results by the drop perpendicular and second-derivative Gaussian transformation methods are similar to those of manual integration. The valley method is problematic for shoulder peaks (e.g., peak ID 17 and 24 in Figure 2A and Table S1). The simple Gaussian fit method overestimates the areas for sharp peaks and underestimates the areas for broadened peaks. In this study, the drop perpendicular method was used to measure the peak areas because all peaks were determined in terms of quercetin-3-*O*-rutinoside (rutin) equivalents, and overlapping tails could be accounted for as one peak if the overlapping peaks were both identified as flavonols/flavones. It is noted that the smoothed chromatograms were used to determine the start and end points for each peak, but the peak area measurement is calculated from the original chromatogram because the smoothing process may distort the peak and compromise the quantitation.

UV−vis Spectrum Similarity Analysis. Among the various classes of flavonoids, flavones and flavonols share a characteristic absorption band in UV; for example, the cinnamoyl structure in the B and C ring has UV absorption in the range from 305 to 390 nm.[19] Although substitutions of different functional groups and/or substitutions at different positions may affect the UV absorption bands (the wavelength of the peak maximum and the absorption coefficient), their UV−vis spectra still differ from other classes of flavonoids and nonflavonoids compounds in general. Thus, similarity analysis between UV−vis spectra can be used to filter out nondesired peaks. The filtering process selects a reference flavone/flavonol glycoside peak and then compares all other peaks with the reference peak. The reference peak can be a spiked standard in the sample, an endogenous flavone/flavonol glycoside peak, or an external standard (such as rutin). In this study, an endogenous flavone/flavonol glycoside peak from each chromatogram was selected as the reference peak for each sample. The similarity of the UV−vis spectra from 200 to 600 nm between all the other peaks, and the reference peak was evaluated by the Pearson product-moment correlation coefficient R:

$$R = 100 \times \frac{\sum_{i=1}^{n}(A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum_{i=1}^{n}A_i - \bar{A})^2} \times \sqrt{\sum_{i=1}^{n}(B_i - \bar{B})^2}}\% \quad (1)$$

where A_i and B_i are two UV−vis spectra. The larger the R, the more similar the two spectra. The background UV−vis spectrum was estimated by averaging the UV−vis spectra at peak start and peak end points and was subtracted from the UV−vis spectrum of each peak. Before calculation of the

DOI: 10.1021/acs.analchem.5b02624
Anal. Chem. 2015, 87, 9974−9981

correlation coefficient, the UV−vis spectra for both the reference peak and the unknown peak were normalized to their unit vector length.[20] By setting a threshold for R, the flavone/flavonol glycoside peaks can be extracted from the chromatograms for further analysis.

Putative Identification of Flavone/Flavonol Glycosides. HRAM/MS^n data were used for putative identification of flavone/flavonol glycosides. First, for each chromatographic peak extracted from the DAD chromatogram using the above-mentioned method, the accurate masses of the top n intense ions in HRMS full scan of that peak were extracted and were considered as potential deprotonated molecules ($[M - H]^-$) of flavone/flavonol glycosides. The number of peaks (n) extracted from each HRMS full scan is predetermined by the user. The possible structures of flavone/flavonol glycosides can be calculated as

$$M_{\text{flavone/flavonol glycoside}} = M_{\text{aglycone}} + a \times M_{\text{glycosyl groups}} + b \times M_{\text{acyl groups}} \quad (2)$$

where the masses (M) are self-explanatory. A total of 12 aglycones, 7 common glycosyl groups, and 14 common acyl groups[21] (Tables S2−S4, Supporting Information) are included in the calculation, and their combinations cover most of the common flavone/flavonol glycoside structures. An ion will be considered a flavone/flavonol only if the difference between the experimentally measured mass and all possible theoretical masses is less than a predetermined threshold. If no reasonable combination exists, the ion will be deleted from the UV−vis output table. It must be noted that only the numbers and the types of aglycones and substitution groups can be calculated; the linkage positions cannot be predicted by FlavonQ. Next, the program searches the MS^n spectra of the peaks that pass the above-mentioned step for characteristic product ions of the 12 aglycones such as m/z 285 (tetrahydroxyflavone, e.g., kaempferol), 301 (pentahydroxyflavone, e.g., quercetin), and 315 (tetrahydroxymethoxyflavone, e.g., isorhamnetin), refining the identification result by confirming the types of aglycones. For example, the presence of product ion m/z 285 indicates that the aglycone is tetrahydroxyflavone (most likely kaempferol).[22] Third, each candidate ion was matched with an existing open-source database such as FooDB[23] or METLIN and a link included in the search-result table. For each data set, the results in each of the above-mentioned steps were saved into a separate spreadsheet.

Quantitative Analysis of Flavone/Flavonol Glycosides. In a previous study, the molar relative response factors (MRRF) of some flavone and flavonol glycosides were computed using rutin as the reference and were found to fall in the range of 0.97 to 1.20, with the exception of hinokiflavone (MRRF = 1.94) and cupressuflavone (MRRF = 1.94).[19] In FlavonQ, an assumed MRRF of 1.0 at UV 350 nm was used for quantitation of all flavone and flavonol glycosides to simplify calculation. If more accurate quantitation number is needed, the user can use the MRRF value table (Table S5, Supporting Information) after they manually verify/identify the individual flavonoids of interest. A calibration curve for rutin was constructed based on rutin standards run at the same time as the samples. The concentrations for the flavones/flavonols identified from the chromatographic peaks in each plant sample were calculated from the external rutin calibration curve. The concentrations of the target flavone/flavonol can be determined as

$$C_{\text{target}} = C_{\text{rutin}} \times \text{MW}_{\text{target}}/\text{MRRF} \times \text{MW}_{\text{rutin}} \quad (3)$$

where C_{rutin} is the concentration of the compound as calculated from the rutin calibration curve, MW_{rutin} is the molecular weight of rutin, and $\text{MW}_{\text{target}}$ is the molecular weight of the target compound, which can be calculated from the accurate precursor ion of the chromatographic peak. Because the MRRF value is assumed as 1 in FlavonQ, the eq 3 is further simplified.

EXPERIMENTAL SECTION

Materials and Reagents. Formic acid, HPLC grade methanol and acetonitrile were purchased from Fisher Scientific. (Pittsburgh, PA). HPLC grade water was prepared from distilled water using a Milli-Q system (Millipore Laboratory, Bedford, MA). Quercetin-3-*O*-rutinoside (rutin) was obtained from Sigma-Aldrich (St. Louis, MO).

Red mustard greens, green onion, giant green onion, Chinese chive, chive, curry leaf, leek, elderflower, rubber, and yu choy were purchased from local food stores in Maryland and were lyophilized immediately after arrival. The dried materials were powdered after lyophilization. Instrument settings and sample preparation are detailed in Supporting Information.

Data Formats. The UHPLC-DAD HRAM MS data sets were acquired as RAW files. The DAD data were converted to text files from RAW files by Xcalibur plug-in tool, MSGet.[24] With an in-house algorithm, text files were read into MATLAB. For the MS data, the RAW files were first converted to mzXML by an open-source software package, ProteoWizard,[25] and then read into MATLAB by the built-in "mzxmlread" function in the MATLAB bioinformatics toolbox. The peak selection scripts were developed and modified based on the script of "findpeaksG.m" provided by Prof. Tom O'Haver (http://terpconnect.umd.edu/~toh/spectrum/findpeaksG.m).

RESULTS AND DISCUSSION

Parameter Settings. FlavonQ allows users to set several parameters to optimize the desired results. The flowchart of optimization process is shown in Figure S2 (Supporting Information), and each step of optimization is visualized for the user's convenience. As mentioned previously, parameters for peak detection include the thresholds for "slope", "amplitude", and "peak width". It is impossible to have a universal setting which can accommodate all different chromatographic profiles due to the fact that peak shapes, background variations, and signal intensity are highly dependent on the sample contents, extraction methods, and instrumental conditions. The primary goal here is to include as many "real" chromatographic peaks as possible. In practice, small values for "slope threshold" (e.g., 0) and "peak width threshold" (e.g., 0.05 min) and a large values for "amplitude threshold" (e.g., 10 000) are recommended as a starting point.

The other important parameters are "reference peak ID" and "correlation coefficient threshold" (R threshold). Because flavone and flavonol glycosides have strong UV absorption at 350 nm and characteristic product ions in MS^n spectra (e.g., m/z 285, 301, 315 and so on),[22] it is relatively easy to find a flavone/flavonol glycoside peak as the reference peak and verify it manually. The UV−vis and MS^n spectra for each peak are listed in the figure window (see Figure 3 as an example) with correlation coefficient R. The threshold of R can be in the range from −100 to 100. The details of R threshold optimization will be discussed later.

DOI: 10.1021/acs.analchem.5b02624
Anal. Chem. 2015, 87, 9974−9981

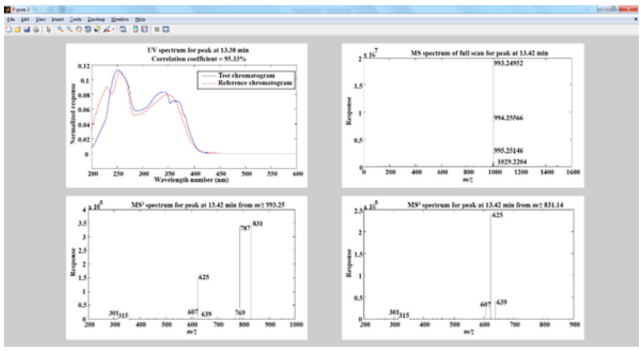


Figure 3. Example of visualization in "FlavonQ" for UV−vis and MS^n spectra examination.

Effect of Reference Peak Selection on UV−vis Spectrum Similarity Analysis. The selection of a reference peak is critical for flavone/flavonol glycoside peak extraction because the similarity analysis algorithm compares the UV−vis spectra of all the other peaks with the UV−vis spectrum of the reference peak. To evaluate the effect of reference peak selection, the similarity analysis of red mustard green sample using different peaks as reference was investigated and the results are shown in Table 2. Six peaks including five flavone/flavonol glycosides (peaks 18, 19, 20, 21, and 25 in Figure 2A) and sinapic acid (a hydroxycinnamic acid, peak ID 42 in Figure 2A) were selected as reference peaks, respectively. Using peaks 18−21 or 25 as the reference peak generated similar results: 39−40 peaks were identified with 1 or 2 false positives (nonflavone/flavonol glycoside peaks were identified as flavone/flavonol glycosides). Using peak 42, only 8 flavone/flavonol peaks were found with 23 false positive peaks. The results showed that the similarity analysis performs well as long as an appropriate flavone/flavonol peak is selected. The user is not required to find the "best" peak as the reference; however, selection of a well-separated major (or base) flavone/flavonol glycoside peak in a chromatogram is recommended.

***R* Threshold Determination.** The *R* threshold determines which of the chromatographic peaks are extracted as flavone/flavonol peaks. With a larger value, *R* is more likely to exclude nonflavone/flavonol peaks but with the risk of excluding some minor flavone/flavonol glycoside peaks. The optimum *R* threshold can eliminate most nonflavone/flavonol peaks and, at the same time, retain all major flavone/flavonol peaks. Figure 4 shows the effect of the *R* threshold on the flavone/flavonol peak extraction results. By increasing the *R* threshold, the number of peaks extracted as flavone/flavonol glycosides decreased, and fewer false positive peaks (nonflavone/flavonol peaks identified as flavone/flavonol peaks) and more false negative peaks (missed flavone/flavonol peaks) were observed (Figure 4A). The peak area ratios of FlavonQ identified flavone/flavonol peaks to total area of human identified flavone/flavonol peaks are in the range of 90.8 to 100% by setting *R* threshold to 85−95%, which indicates that the majority of flavone/flavonol peaks (>90%) have been extracted (Figure 4B).

Two factors affect the determination of optimum *R* threshold: the diversities of flavone/flavonol glycosides in the analyzed sample and the chromatographic separation. The characteristic UV absorption band is indeed affected by the positions and species of substitutions, e.g., the wavelength of maximum UV absorbance of the band around 350 nm for

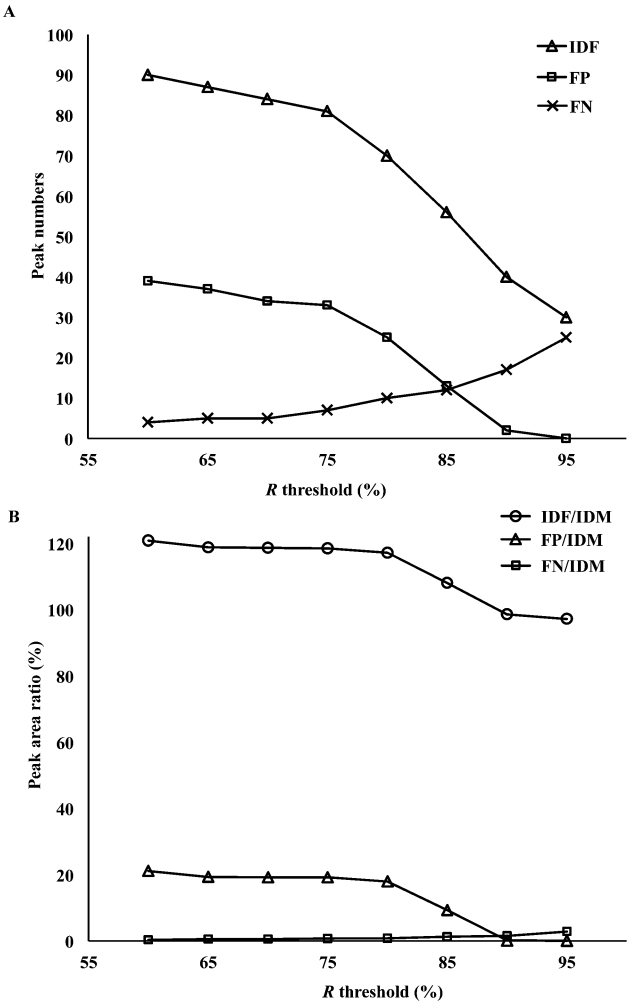


Figure 4. Effect of correlation coefficient (*R*) threshold settings on peak extraction for red mustard green sample: effect on peak numbers (A) and effect on peak areas (B). IDF: peaks identified as flavone/flavonol by FlavonQ; FP: false positive peaks; FN: false negative peaks; IDM: peaks manually identified as flavone/flavonol by a human expert.

different flavonol glycosides in red mustard green can range from 328 to 369 nm,[10] and the *R* values can be different. Compounds which coelute with flavone/flavonol glycosides may distort the UV−vis spectra of chromatographic peaks and, thus, cause inaccurate estimation of *R* values. Similarly, a good chromatographic separation can improve flavone/flavonol identification by reducing the overlapping peaks in the chromatogram. The optimum *R* thresholds for different plants are shown in Table 1. Table 1 presents the numbers of peaks identified as flavones/flavonols by FlavonQ in different plants as compared to the number determined by experts and the number of false positives and negatives for FlavonQ. The areas for the false positive and negative peaks were less than 10% in all plants except in leek, in which a large feruloyl hexose peak was misclassified as a flavone/flavonol peak by FlavonQ because of their similar UV−vis spectra.

Overall the peak extraction result by FlavonQ is reliable and accurate. The determination of the *R* threshold is also dependent on the research purpose. For example, a lower value will pick out all possible flavone/flavonol glycosides (less or no false negatives). Although false positives can be removed by FlavonQ in the following steps (using MS data), a value too

DOI: 10.1021/acs.analchem.5b02624
Anal. Chem. 2015, 87, 9974−9981

Table 1. Peak Selection after Similarity Analysis by FlavonQ on Different Plant Samples[a]

plant name	reference peak	R threshold (%)	N_d	N_h	N_f	N_p	N_n	$area_p/area_h$ (%)	$area_n/area_h$ (%)	$area_f/area_h$ (%)
red mustard green	25	90	108	55	40	2	17	0.1	1.5	98.6
green onion	56	90	106	52	45	2	9	0.7	1.0	99.7
giant green onion	44	75	106	42	38	7	11	9.8	2.8	107.0
Chinese chive	12	50	108	24	36	14	2	8.8	2.0	106.8
chive	10	50	94	31	37	10	4	4.2	1.0	103.2
curry leaf	36	50	77	25	20	2	7	1.0	4.5	96.5
leek	36	50	118	47	53	11	5	17.2 (6.0)[b]	0.8	116.4 (105.2)[b]
elderflower rubber	32	20	49	22	23	3	2	0.2	0.4	99.8
yu choy	29	15	100	31	45	18	4	4.1	7.9	96.2

[a]N_d: Number of detected peaks by FlavonQ; N_h: number of peaks identified as flavone/flavonol by human expert; N_f: number of peaks identified as flavone/flavonol by FlavonQ; N_p: number of false positive peaks by FlavonQ; N_n: number of false negative peaks by FlavonQ; $area_p/area_h$: peak area ratio of false positive peaks and flavonol peaks identified by human expert; $area_n/area_h$: peak area ratio of false negative peaks and flavonol peaks identified by human expert; $area_f/area_h$: peak area ratio of flavonol peaks identified by FlavonQ and flavonol peaks identified by human expert. [b]Data contained a large feruloyl hexose peak which was identified as flavone/flavonol glycoside. $Area_p/area_h$ is 6.0% and $area_f/area_h$ is 105.2% if feruloyl hexose peak was not counted.

Table 2. Comparison of Similarity Analysis Results Using Different Peaks as Reference

reference peak ID	putative identification of reference peak[a]	N_f[b]	N_p[b]	R threshold (%)
18 (RT: 12.59 min)	quercetin 3-feruloylsophorotrioside-7-glucoside	40	1	90
19 (RT: 13.38 min)	quercetin 3-sinapoylsophoroside-7-glucoside	40	2	90
20 (RT: 13.73 min)	isorhamnetin 3-caffeoylsophoroside-7-glucoside	39	1	90
21 (RT: 14.19 min)	kaempferol 3-sinapoylsophorotrioside-7-glucoside	40	2	90
25 (RT: 16.13 min)	isorhamnetin 3-glucoside-7-glucoside	40	2	90
42 (RT: 24.78 min)	sinapic acid	31	23	90

[a]Details of identification refer to ref 10. [b]N_f: Number of peaks identified as flavone/flavonol by FlavonQ. N_p: Number of false positive peaks;

low may still increase the number of false positives and necessitate further manual analysis.

Putative Identification of Flavone/Flavonol Glycoside. After UV−vis peak extraction, the peaks identified as potential flavone/flavonol glycosides were further examined using HRAM MS^n data. Elimination of false-positives and putative identification of each extracted chromatographic peak can be achieved using the following steps:

(1) For each selected peak, the five most intense ions (this default value can be changed if desired) in the full MS scan spectrum of the peak were selected as candidate ions, which may include one or multiple flavone/flavonol glycoside deprotonated molecules.
(2) Possible chemical formulas were generated for the candidate ions, and possible structures of flavone/flavonol glycoside(s) were predicted by eq 2 with the combination of aglycones, glycosyl groups, and acyl groups.
(3) If no candidate ions selected from a peak matched any reasonable combination of aglycones, glycosyl groups, and acyl groups, the peak was considered as a nonflavone/flavonol peak (false positive) and excluded from result.
(4) For peaks passing step 3, FlavonQ refined the result by searching MS^n spectra for characteristic fragment ions (e.g., m/z 285 for "tetrahydroxyflavone", m/z 301 for "pentahydroxyflavone") for further confirmation.
(5) Database searching and further analysis of MS^2 and MS^3 spectra gives clues of the substitution positions and glycosyl/acyl groups which lead to putative identification of the peak.

The structural information such as numbers of hydroxylation and/or methylation of the basic core (flavone) structure, types, and numbers of glycosyl groups and acyl groups are included in the output results. In most cases, multiple possible structures will be generated from a single high resolution accurate mass. For example, peak 22 at 14.59 min in Figure 2A contains a major mass peak of m/z 609.1467 in full scan. Five possible structures were found by inputting m/z 609.1467 to eq 2 with mass error tolerance ±5 ppm: tetrahydroxyflavone, 2 hexosyl; hexahydroxyflavone, 2 rhamnose; pentahydroxymethoxyflavone, 1 pentosyl, 1 rhamnose; pentahydroxydimethoxyflavone, 2 pentosyl; pentahydroxyflavone, 1 rhamnose, 1 hexosyl; tetrahydroxymethoxyflavone, 1 pentosyl, 1 hexosyl. The next step, refining the result by searching MS^n data, found a characteristic fragment ion m/z 285 in MS^2, so only tetrahydroxyflavone with 2 hexosyl was output in the result table. Thus, the peak was most likely a kaempferol with 2 hexsoyl moieties attached. Further manual analysis of MS^2 and MS^3 spectra confirmed the substitution positions and hexosyl groups with literature and plant source research. The peak was putatively identified as kaempferol 3-*O*-glucoside-7-*O*-glucoside. For a larger mass, e.g., m/z 755.1825, the identification was more complicated: 59 possible structures were calculated by FlavonQ, and there were still six candidates left in the refined result with a characteristic ion m/z 285 in MS^2 (see Tables S6 and S7, Supporting Information). In such cases, expertise in the field of flavonoid research is needed to study the MS fragmentation patterns in MS^n spectra to determine the most likely flavone/flavonol glycosides structure among the six candidates. FlavonQ is intended to assist an expert. It is designed to facilitate flavonoid analysis and can save considerable time for researchers by picking out possible flavone/flavonol peaks and offering possible structures to match with MS^n data.

A list of unique deprotonated molecular ions of potential flavone/flavonol glycosides is provided in the program readout

DOI: 10.1021/acs.analchem.5b02624
Anal. Chem. 2015, 87, 9974−9981

which can be used to search existing metabolites in open-source databases, such as FooDB, METLIN, and HMDB.[23,26] METLIN link was implemented for each candidate ion (if available) in the result table (see Figure S4, Supporting Information). However, human analysis of MS^n data is crucial to confirm the structure for final identification of the compounds, if needed.

False negatives may occur when the intensities of deprotonated molecular ions for flavone/flavonol glycosides are not among the top five highest in a peak. This was observed infrequently but was primarily associated with minor chromatographic peaks with either high background noise and/or a coeluting interference. The solution was to adjust the number of the ions selected in FlavonQ, manually check the suspicious peaks, or use the "dynamic exclusion" function in data-dependent acquisition method (MS software) to eliminate interfering ions from the MS spectrum.

Quantitation of Flavone/Flavonol Glycoside. One of the goals of FlavonQ is quantitation of flavone and flavonol glycosides in plant materials. The analysis may be termed semiquantitative, as traditional calibration methods are not feasible. A single external standard calibration curve and molar relative response factors were used in this study. Two regression methods were used for curve fitting in FlavonQ: unweighted and weighted least-squares linear regression. Weighted least-squares linear regression is generally recommended to improve the accuracy of calibration curves at the lower end when the concentration range of calibration is large.[27,28]

In this experiment, the calibration curve of rutin was constructed using weighted least-squares linear regression with a weighting factor of $1/x^2$ (x represents concentration). From Table 1, the quantitation results by FlavonQ covered over 90% of peak areas for flavone and flavonol glycosides in each plant analyzed. The total concentrations of flavone/flavonol glycosides in rutin-equivalents in nine plants are listed in Table 3, and the amounts of individual flavone/flavonol glycosides in different plants are included in the spreadsheets of Supporting Information. An example of the result output is shown in Table S8, Supporting Information. For overlapping peaks containing more than one flavone/flavonol glycoside, the concentrations were computed together. FlavonQ also generates a figure which labels all the flavone/flavonol glycoside peaks and lists the deprotonated molecular ions and concentrations for each peak (Figure 2B). As reported, the flavonoids are strongly influenced by environmental conditions and their variability in foods is very high (e.g., an average relative standard deviation of ±330% for 22 flavonoids was observed for samples collected nationwide at 6-month intervals).[29] Therefore, the quantitation results by FlavonQ is more than sufficient to meet the needs of most users. More accurate quantification results, if needed, can be achieved by using the response factor table (Table S5, Supporting Information) after manual identification of the individual peaks, or by additional individual calibration curves for compounds of interest (if reference standards are available).

Table 3. Total Flavone/Flavonol Glycosides in Rutin-Equivalents

plant name	total flavone/flavonol glycosides rutin-equivalents (mg/100 g dried weight)
red mustard green	1414.50
green onion	176.00
giant green onion	93.50
Chinese chive	90.64
chive	111.85
curry leaf	167.66
leek	90.94
elderflower rubber	87.72
yu choy	61.67

Performance Evaluation. The computation speed of FlavonQ is dependent on the numbers of chromatographic peaks, the numbers of candidate ions to be identified, and the performance of the computer. For example, the red mustard green data contained 108 chromatographic peaks, 34 peaks were selected as flavone/flavonol glycosides, and 5 candidate ions were extracted from each peak for identification. The execution time for red mustard green data was 46.9 s with our computer. It should be noted that the FlavonQ spent about a third of the time (about 16 s) saving results to an Excel spreadsheet, as the output included all results with details.

The output Excel spreadsheet contains four separate sheets: (1) peaks detected by UV−vis, (2) peaks for R > threshold (peaks that passed the UV−vis similarity test), (3) identified peaks (putatively identified peaks using MS data), and (4) matched peaks with a database (peaks that were found from a database). The program allows manual tracking of the results from each step if needed. It is convenient to transfer the results generated from FlavonQ to other software or publications. The average execution time for the analysis of flavone/flavonol glycosides by FlavonQ in nine different plants was less than 1 min, compared to days or weeks needed with the traditional workflow.

The putative identification of flavone/flavonol glycosides in red mustard green by FlavonQ (R threshold: 40%) were compared with the published results of Lin et al.[10] From 277 candidate ions, FlavonQ putatively identified a total of 94 peaks as flavone/flavonol glycosides. In the literature report, 92 flavone/flavonol glycosides were identified. Of these, 77 of them were in agreement with FlavonQ. The 15 peaks that FlavonQ failed to identify (false negatives) were examined manually and were all found to be minor peaks with very low intensities (signal-to-noise ratios are less than 3). For the 17 false positive peaks by FlavonQ, 6 of them were hydroxycinnamic acid derivatives (e.g., sinapoyl hydroxyl feruloyl gentiobiose, sinapoyl hexose, 1,2-diferuloylgentiobiose, trisinapoyl gentiobiose, disinapoyl feruloyl gentiobiose, and disinapoylgentiobiose) which were misidentified as flavone/flavonol glycosides, and the other 11 were all small peaks either coeluted with flavone/flavonol glycosides or partially separated isomers (may not be false positive in such case). The percentage of the total area of 17 peaks to the total area of flavone/flavonol glycosides peaks identified by FlavonQ was 7.7%.

CONCLUSION

FlavonQ was developed as a LC/MS data processing tool for putative identification and quantitation of flavone/flavonol glycosides. The time needed for data analysis is significantly reduced, hours for FlavonQ, with human verification, as compared to days or weeks for manual data-mining. The new data processing strategy represents a comprehensive and potentially ground-breaking workflow that provides both

DOI: 10.1021/acs.analchem.5b02624
Anal. Chem. 2015, 87, 9974−9981

putative identification and quantitation capabilities for flavonoid analysis. The FlavonQ is designed to facilitate flavonoid research, not to replace human experts. Manual investigation and verification of the results generated by FlavonQ can save more than 90% of research time compared to manual data-mining. This research is part of the initiative of the Food Composition and Methods Development Laboratory of the Agricultural Research Services of the United States Department of Agriculture to automate "expert knowledge" to facilitate analysis of bioactive compounds in plant materials. FlavonQ is the first step toward that goal and will be developed into a web-based application to benefit the scientific community in flavonoids research.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.5b02624.

Additional information as noted in text (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Tel: +1 301 504 8144; fax: +1 301 504 8314; e-mail: pei.chen@ars.usda.gov.

Author Contributions

†Contributed equally to this manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

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

■ REFERENCES

(1) Merken, H. M.; Beecher, G. R. *J. Agric. Food Chem.* **2000**, *48*, 577−599.

(2) Yao, L. H.; Jiang, Y. M.; Shi, J.; Tomas-Barberan, F. A.; Datta, N.; Singanusong, R.; Chen, S. S. *Plant Foods Hum. Nutr.* **2004**, *59*, 113−122.

(3) Robards, K. *J. Chromatogr. A* **2003**, *1000*, 657−691.

(4) Luo, C. Y.; Zou, X. L.; Li, Y. Q.; Sun, C. J.; Jiang, Y.; Wu, Z. Y. *Food Chem.* **2011**, *127*, 314−320.

(5) Lin, L. Z.; Harnly, J. M. *J. Agric. Food Chem.* **2007**, *55*, 1084−1096.

(6) Escarpa, A.; Gonzalez, M. C. *Crit. Rev. Anal. Chem.* **2001**, *31*, 57−139.

(7) de Rijke, E.; Out, P.; Niessen, W. M. A.; Ariese, F.; Gooijer, C.; Brinkman, U. A. T. *J. Chromatogr. A* **2006**, *1112*, 31−63.

(8) Bae, H.; Jayaprakasha, G. K.; Jifon, J.; Patil, B. S. *Food Chem.* **2012**, *130*, 751−758.

(9) Mikulic-Petkovsek, M.; Slatnar, A.; Stampar, F.; Veberic, R. *Food Chem.* **2012**, *135*, 2138−2146.

(10) Lin, L. Z.; Sun, J.; Chen, P.; Harnly, J. *J. Agric. Food Chem.* **2011**, *59*, 12059−12072.

(11) Garcia-Canas, V.; Simo, C.; Herrero, M.; Ibanez, E.; Cifuentes, A. *Anal. Chem.* **2012**, *84*, 10150−10159.

(12) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779−787.

(13) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. *BMC Bioinf.* **2010**, *11*, 395.

(14) Lommen, A. *Anal. Chem.* **2009**, *81*, 3079−3086.

(15) Griffiths, W. J.; Koal, T.; Wang, Y.; Kohl, M.; Enot, D. P.; Deigner, H. P. *Angew. Chem., Int. Ed.* **2010**, *49*, 5426−5445.

(16) Pelillo, M.; Bonoli, M.; Biguzzi, B.; Bendini, A.; Toschi, T. G.; Lercker, G. *Food Chem.* **2004**, *87*, 465−470.

(17) Burlingame, A. L.; Boyd, R. K.; Gaskell, S. J. *Anal. Chem.* **1998**, *70*, 647−716.

(18) Savitzky, A.; Golay, M. J. E. *Anal. Chem.* **1964**, *36*, 1627−1639.

(19) Lin, L. Z.; Harnly, J.; Zhang, R. W.; Fan, X. E.; Chen, H. J. *J. Agric. Food Chem.* **2012**, *60*, 544−553.

(20) Zhang, M.; Harrington, P. d. B. *Talanta* **2013**, *117*, 483−491.

(21) Sun, J.; Lin, L.; Chen, P. *Curr. Anal. Chem.* **2013**, *9*, 397−416.

(22) Vukics, V.; Guttman, A. *Mass Spectrom. Rev.* **2010**, *29*, 1−16.

(23) Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J.; Liu, P.; Yallou, F.; Bjorndahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A. *Nucleic Acids Res.* **2013**, *41*, D801−807.

(24) Kazusa DNA Research Insititute, C., Kisarazu, Chiba 292-0818, Japan. Komics Wiki, 2008.

(25) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. *Bioinformatics* **2008**, *24*, 2534−2536.

(26) Tautenhahn, R.; Cho, K.; Uritboonthai, W.; Zhu, Z.; Patti, G. J.; Siuzdak, G. *Nat. Biotechnol.* **2012**, *30*, 826−828.

(27) Almeida, A. M.; Castel-Branco, M. M.; Falcao, A. C. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2002**, *774*, 215−222.

(28) Zhang, M.; Harrington, P. d. B. *Rapid Commun. Mass Spectrom.* **2015**, *29*, 789−794.

(29) Harnly, J. M.; Doherty, R. F.; Beecher, G. R.; Holden, J. M.; Haytowitz, D. B.; Bhagwat, S.; Gebhardt, S. *J. Agric. Food Chem.* **2006**, *54*, 9966−9977.

DOI: 10.1021/acs.analchem.5b02624
Anal. Chem. 2015, 87, 9974−9981