JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

pubs.acs.org/JAFC

Untangling the Chemistry of Port Wine Aging with the Use of GC-FID, Multivariate Statistics, and Network Reconstruction

Dan Jacobson,[†] Ana Rita Monforte,[§] and António César Silva Ferreira*[,†,§]

[†]IWBT – DVO University of Stellenbosch, Private Bag XI, Matieland 7602, South Africa

[§]Escola Superior de Biotecnologia, Universidade Católica Portuguesa, Rua Dr. António Bernardino de Almeida, 4200-072 Porto, Portugal

ABSTRACT: Chromatography separates the different components of complex mixtures and generates a fingerprint representing the chemical composition of the sample. The resulting data structure depends on the characteristics of the detector used, univariate for devices such as a flame ionization detector (FID) or multivariate for mass spectroscopy (MS). This study addresses the potential use of a univariate signal for a nontargeted approach to (i) classify samples according to a given process or perturbation, (ii) evaluate the feasibility of developing a screening procedure to select candidates related to the process, and (iii) provide insight into the chemical mechanisms that are affected by the perturbation. To achieve this, it was necessary to use and develop methods for data preprocessing and visualization tools to assist an analytical chemist to view and interpret complex multidimensional data sets. Dichloromethane Port wine extracts were collected using GC-FID; the chromatograms were then aligned with correlation optimized warping (COW) and subsequently analyzed with multivariate statistics (MVA) by principal component analysis (PCA) and partial least-squares regression (PLS-R). Furthermore, wavelets were used for peak calling and alignment refinement, and the resulting matrix was used to perform kinetic network reconstruction via correlation networks and maximum spanning trees. Network-target correlation projections were used to screen for potential chromatographic regions/peaks related to aging mechanisms. Results from PLS between aligned chromatograms and target molecules showed high X to Y correlations of 0.91, 092, and 0.89 with 5-hydroxymethylfurfural (HMF) (Maillard), acetaldehyde (oxidation), and 4,5-dimethyl-(5*H*)-3-hydroxy-2-furanone, respectively. The context of the correlation (and therefore likely kinetic) relationships among compounds detected by GC-FID and the relationships between target compounds within different regions of the network can be clearly seen.

KEYWORDS: *univariate signal, aging, mechanisms, preprocessing, GC-FID, PCA, PLS, network theory, kinetic network reconstruction*

INTRODUCTION

Port wine is a fortified wine produced in the Douro region of Portugal. After the vinification process, wines are exclusively barrel aged (tawnys) or matured for 2 years in a cask and then bottled (vintage).

The aromatic profile of Port wine changes during aging as the result of several underlying mechanisms. Therefore, if one wants to understand or modulate the sensory attributes of Port, it is important to understand these mechanisms and the interconnections among them. Several of the mechanisms are to a large extent already described as Maillard[1−7] or oxidation;[8−12] nevertheless, the overlaps between these two mechanisms are not well-known.

In Port wines sotolon was recognized as a key molecule in the "perceived age" of barrel storage Port wine and consequently in the aroma quality of the final product. Its concentration can range from a few dozen micrograms per liter in a young wine to 1 mg/L in wines older than 50 years. The odor threshold has been estimated to be 19 μg/L.[13−20]

The Maillard reaction has been suggested by several authors to be responsible for the formation of sotolon as a product of a reaction involving hexoses and pentoses in the presence of cysteine[4] and from the aldol condensation of butane-2,3-dione and hydroxyacetaldehyde.[21] On the other hand, several papers have linked sotolon formation with oxidation.[19,22−25] Both oxygen and temperature influence sotolon concentration, which suggests that its origin involves a connection between oxidation and Maillard mechanisms.[26]

Therefore, wine aging is a complex system, which requires more information to be analyzed to better understand the mechanisms at play. Given this, techniques that are able to capture information about a broader range of compounds participating in the aging process are necessary to achieve a better understanding thereof.

Metabolomics is defined as the study of "as many small metabolites as possible" in a system.[27] In this paper we attempt to describe an example of chemiomics, which we define to be the study of the relationships between as many chemical compounds as possible in a complex chemical (nonenzymatic) system. To accomplish this by chemical profiling two strategies can be employed: (i) "targeted analysis", using a priori knowledge of which compounds to analyze, which requires their identification and manual quantification, or (ii) "untargeted analysis", in which one tries to detect as many compounds as possible to acquire sample fingerprints, which will be submitted to multivariate analyses and network analysis for further contextualization. The identification and quantita-

Received: June 26, 2012
Revised: February 18, 2013
Accepted: February 18, 2013
Published: February 18, 2013

dx.doi.org/10.1021/jf3046544 | J. Agric. Food Chem. 2013, 61, 2513−2521

tion are then performed on the variables that are found to be associated with the principal components/correlation vectors as determined by multivariate and network analysis.

Spectroscopic detectors, such as those based on ultraviolet−visible (UV−vis), Fourier transform infrared (FTIR), or nuclear magnetic resonance (NMR) spectroscopy, are largely employed to obtain chemical fingerprints that can be used for sample classification as well for chemical quantitation.[28−37] The extremely convoluted resulting spectra can be further processed with multivariate statistics (MVA) techniques that compensate, to a certain extent, for the absence of structural information in complex chemical mixtures.

Despite the versatility of these detectors, the absence of structural information, due to the extremely convoluted signal, constitutes a major drawback in obtaining molecular identifications if there is a need to study complex systems that require kinetic contextualization.

Chromatography has long been used for the separation of molecules enabling both the quantitative and qualitative analysis of constituents in complex mixtures. The separation of the different components of a complex mixture generates a fingerprint representing the chemical composition of the sample. Chromatographic fingerprints taken from samples under different experimental conditions can then be used to explain the changes caused by a perturbation.[38] Due to the separation performed by the column, it is possible to identify structures on the basis of the elution time in a given chromatographic profile.

Chromatographic data have a huge number of variables, and principal component analysis (PCA) and partial least-squares (PLS) are MVA visualization techniques that allow for the interpretation of multidimensional data sets. When multivariate analysis involves large data sets, variable selection processes play an important role because they eliminate the less significant or noninformative variables. The overall aim of any variable selection technique is to capture variables from the original data set that are most specifically related to the problem of interest and to exclude those variables that are affected by other sources of variation.

PCA is a nonsupervised technique that decomposes the original variables of a data set into two matrices: the score and the loading matrices. The score matrix contains information about the samples, which are described in terms of their projection onto the principal components. The loading matrix contains information about the variables which are also described in terms of their projection onto the principal components. The loadings can also be interpreted as the contribution of the variables for the observed scores distribution.

Consequently, the use of GC fingerprints with MVA should make it possible to extract considerable amounts of information from complex mixtures. The tandem of GC-MVA is, in our perspective, a middle ground between a rich detector, which provides structural information (NMR), and detectors such as FTIR and UV−vis.

Therefore, the aim of this study is to validate the feasibility of using univariate chromatographic data, in particular, gas chromatography with a flame ionization detector (GC-FID), as a screening procedure to classify complex chemical mixtures, such as wine samples, to identify which compounds are responsible for differences and to perform network reconstructions that may indicate underlying kinetic relationships and mechanisms.

MATERIALS AND METHODS

Reagents. The chemicals 3-octanol (97%) 3-hydroxy-4,5-dimethyl-2(5*H*)-furanone (≥99.5%), 5-methylfurfural (≥99.5%), 5-hydroxymethylfurfural (≥99.5%), acetaldehyde (≥99.5%), ethyl lactate (98%), 5-(ethoxymethyl)furfural (≥99.5%), acetic acid (≥99.5%), 2,3-butanediol (98%), diethyl succinate (≥99.5%), 2-phenylethanol (≥99.5%), diethyl malate (>97%), succinic monoethyl ester (≥99.5%), benzaldehyde (≥99.5%), octanoic acid (≥99.5%), hexanoic acid (≥99.5%), aspartame (>99%), glutamine (>99%), cysteine (>99%), serine (>99%), glycine (>99%), arginine (>99%), γ-aminobutyric acid (>99%), alanine (>99%), tyrosine (>99%), valine (>99%), phenylalanine (>99%), leucine (>99%), ornithine (>99%), lysine (>99%), homoserine (>98%), norvaline (>98%), homocysteine (>99%), 2-sulfanylethanol (98%), tetraphenylborate (>99.5%), iodoacetic acid (>99%), *o*-phthaldialdehyde (>99%), and *n*-alkanes (C11−C22) were obtained from Sigma-Aldrich, USA. *cis*-5-Hydroxy-2-methyl-1,3-dioxane (*cis*-dioxane), *cis*-4-hydroxymethyl-2-methyl-1,3-dioxolane (*cis*-dioxolane), *trans*-4-hydroxymethyl-2-methyl-1,3-dioxolane (*trans*-dioxolane), and *trans*-5-hydroxy-2-methyl-1,3-dioxane (*trans*-dioxane) were synthetized according to the method of Maillard.[39]

Dicloromethane (HPLC grade) was purchased from LabScan, Sowinskiego, Gliwice, Poland. Anhydrous sodium sulfate and methanol (HPLC grade) were obtained from Merck, Darmstadt, Germany.

Port Wine Samples. The 37 samples used in this study were between 2 and 60 years of age: one sample for 2, 14, 19, 23, 35, 40, 42, 48, 54, 57, and 60 years of age, two samples for 7 and 20 years of age, three samples for 4 and 5 years of age, and 10 samples of 10 years of age. All wines were matured in oak barrels. These samples were supplied by the Instituto do Vinho do Porto e do Douro. The wines were made following standard traditional winemaking procedures for Port wine and have been certified.

Analytical Procedure. *Volatiles Extraction.* A liquid−liquid extraction was performed to extract the volatile fraction from each sample. The procedure used was as follows: 5 g of anhydrous sodium sulfate and 50 μL of internal standard (3-octanol) were added to 50 mL of sample and were extracted twice with 5 mL of dicloromethane using a magnetic stir bar for 5 min per extraction, and 2 mL of the resulting organic phase was concentrated under a nitrogen stream four times. The extract was then analyzed by GC (Agilent 5980, USA) with FID detection. Two microliters of the extract was injected. Chromatographic conditions were the following; column BP-21 (50 m × 0.25 mm × 0.25 μm) fused silica (SGE, Portugal); hydrogen (5.0, Air-Liquide, Portugal); 1.2 mL/min flow rate; injector temperature, 220 °C; oven temperature, 40 °C for 1 min programmed at a rate of 2 °C/min to 220 °C, maintained during 30 min; splitless time, 0.5 min; split flow, 30 mL/min.

To facilitate identification, the Kovats index for each peak was calculated as described by Van den Dool and Kratz.[40] This determination was performed on polar phase columns, BP21 (50 m × 0.25 mm × 0.25 μm).

Amino Acid Analysis. Twenty-one amino acids were analyzed in the Port wine samples: aspartic acid, glutamic acid, cysteine, asparagine, histidine, serine, glycine, arginine, threonine, alanine, γ-aminobutyric acid, tyrosine, ethanolamine, valine, methionine, tryptophan, phenylalanine, isoleucine, leucine, ornithine, and lysine. The methodology used was that described by Pipris-Nicolau et al.[41]

Acetaldehyde, Furanic Compounds, and Sotolon Analysis. These analyses were done as described by Silva Ferreira et al.[19]

Data Preprocessing. The ASCII file of chromatographic data obtained from each sample was extracted and a matrix created containing all of the chromatograms. The intensities were normalized by dividing each value by the intensity of the internal standard (3-octanol). The raw data set (GC-FID) was then imported into The UnscramblerX 10.1 (Camo, Sweden), where the first stage of the alignment of chromatograms was performed using correlation optimized warping (GC-FID-COW). This algorithm aligns chromatograms by means of sectional linear stretching and compression, which shifts the peaks of one chromatogram to correlate with those of the

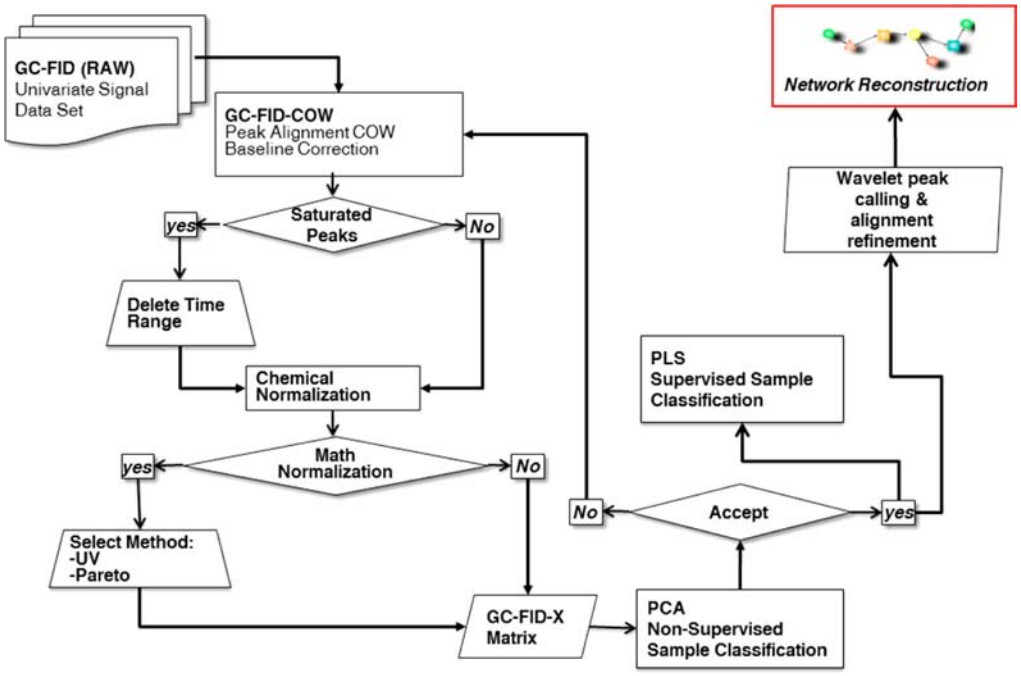

Figure 1. Proposed workflow for univariate (chromatographic) signal processing.

other chromatograms in the data set.[42] The saturated peaks were then removed and the baseline corrected (GC-FID-COW-saturated removed and baseline correction). The resulting matrix (GC-FID-X) was then used for multivariate data analysis as described under Statistical Analysis.

Statistical Analysis. The data were analyzed with PCA and PLS-R using either Qlucore (Lund, Sweden) or SIMCA-P+ 12.0.1 (Umetrics, Norway). PCA shows similarities between samples projected on a plane and makes it possible to determine which variables determine these similarities and in what way. PLS is used to extract factors related to one or more response values. PLS validation was performed by cross-validation method.

Kinetic Network Reconstruction. To attempt to reconstruct the underlying kinetic network, the fingerprint needed to be further compressed to a single value for each putative molecule detected by GC-FID. Thus, each chromatographic peak needed to be replaced with a single value for the intensity and retention time, at the apex of each peak, and therefore a more refined alignment procedure was required. This was achieved as follows: An "average chromatogram" was created by taking the mean of the values at each elution point in the GC-FID-X matrix. The average chromatogram and the sample chromatograms were smoothed with the Savitzky–Golay method (settings: left = 15, right = 15, polynomial degree = 0)[43] as implemented in The Unscrambler X 10.1. The wavelet method of Du et al.,[44] which was originally developed for peak calling in peptide mass spectrometry data, was adapted for finding peak centers in chromatographic data and a Mexican hat wavelet used to determine the location of all of the peaks in the average and sample chromatograms. A custom-built Perl program was integrated with the R-based wavelet method to achieve this. The distances between the locations of all of the peaks in each sample chromatogram and the locations of the peaks in the average chromatogram were calculated. It was observed that there was a correlation between peak height and the amount of peak center shift that occurred across chromatograms, and we thus devised a two-step process for aligning sample peaks to those of the average chromatogram. If the (internal standard normalized) height of the average peak was >2 and the distance to the nearest sample peak was <0.3 min, then the intensity value of the sample peak was assigned as the sample value at the average peak location. However, if the (internal standard normalized) height of the average peak was <2 and the distance to the nearest sample peak was <0.15 min, then the intensity value of the sample peak was assigned as the sample value at the average peak retention time. This algorithm was implemented in Perl.

As a result of this process a new matrix was created that contained the retention time of all peaks in the average chromatogram and the intensity values of all of the peak centers from each sample aligned to these average retention times. Thus, a vector was created for each peak (presumably representing a compound) across all samples. An all-against-all comparison was done by calculating the Pearson correlation between each and every peak vector. As such, one is able to track the increase or decrease of compounds (peaks) during the aging process and determine the correlative relationships among them. We applied a Pearson correlation threshold of 0.8 and represented the remaining relationships as a mathematical graph to form a correlation network with the nodes representing peaks and the edges weighted with the Pearson correlations between the peak vectors. To reconstruct the most likely kinetic network underlying the set of chemical reactions involved in the aging process, a maximum spanning tree was created by transforming the edge weights into inverse correlations (by taking the difference between the number 1 and the absolute correlation values) and the subsequent use of a minimum spanning tree (mst) algorithm[45] on the this inverse correlation network. A minimum spanning tree represents the shortest possible path through a graph and, as such, selects for the smallest inverse correlation (i.e., highest correlation) pairs between all nodes in the network. The resulting networks were visualized in Cytoscape.[46]

RESULTS/DISCUSSION

Principal Component Analysis. Our initial goal for the use of PCA was to examine the intrinsic variation in the data set prior to alignment to determine if the volatile fraction of the samples followed a trend related to age. However, when using the GC-FID matrix described above, some samples did not follow the latent age variable described by PC1, namely the 4- and 60-year-old samples (score plot not shown). The analysis global workflow is described in Figure 1.

The loading plot in Figure 2 shows higher levels of acetic acid, 2,3-butanediol, diethyl succinate, diethyl malate, phenylethanol, and succinic monoethyl ester present in the older samples. The esterification process appears to be the most

dx.doi.org/10.1021/jf3046544 | J. Agric. Food Chem. 2013, 61, 2513–2521

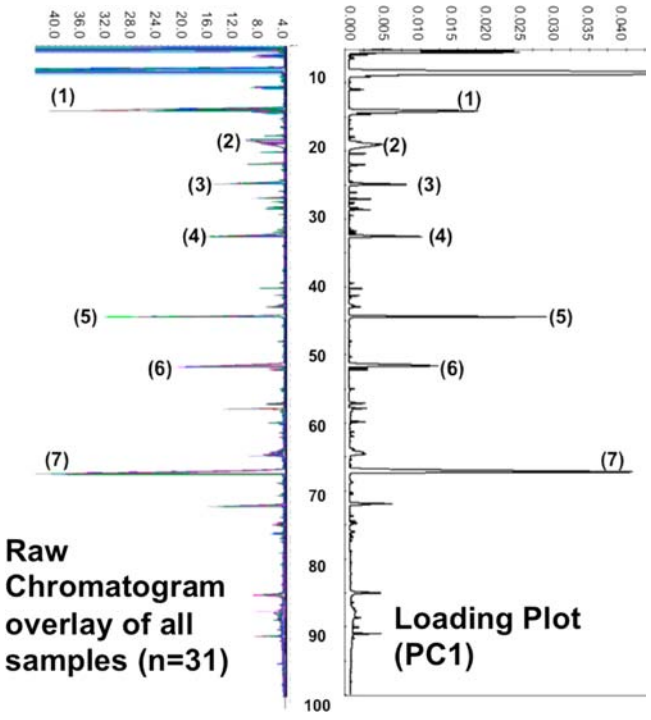

Figure 2. Raw chromatogram overlay of all samples (n = 31) and loading plot (PC1) representing the average GC-FID chromatogram: (1) ethyl lactate, (2) acetic acid, (3) 2,3-butanediol, (4) diethyl succinate, (5) phenylethanol, (6) diethyl malate, and (7) succinic monoethyl ester.

prevalent reaction among the compounds apparent in the loading plot. The organic acids naturally present in grape must, such as malic acid, and those present as the result of fermentation, such as lactic, succinic, and acetic acids, all react with ethanol to yield the esters seen in the loading plot.[47] However, these molecules are out of the detector's linear response range, so they needed to be eliminated. Furthermore, the chromatograms must be aligned because an unavoidable characteristic of all chromatographic data is that the retention times for the peaks in the chromatograms shift slightly from one analysis to another. To address this problem, COW was used to align all of the chromatograms.

A new fingerprint was created (GC-FID-COW) by the removal of saturated peaks and baseline correction to yield the GC-FID-X matrix, which was subsequently analyzed by PCA. The new score plot shows the same latent age variable described by PC1, but the explained variance of the first two components is 74% (Figure 3), and the samples that did not previously follow the age vector now do so after alignment.

The score plots in Figure 3A,B show a clear trend related to wine age, suggesting that the chemical mechanisms are correlated with time across the first principal component, with the first two components explaining 74% of the variance. It appears that the latent age vector remains intact whether the data are mathematically normalized or not.

It is important to note that the data for both the PCA and PLS analyses were not mathematically centered or normalized as is commonly done to give all variables equal impact on the model. Centering is usually used as a matter of convenience for display and mathematically has no impact on the multivariate model. When we viewed the loadings, we chose not to center

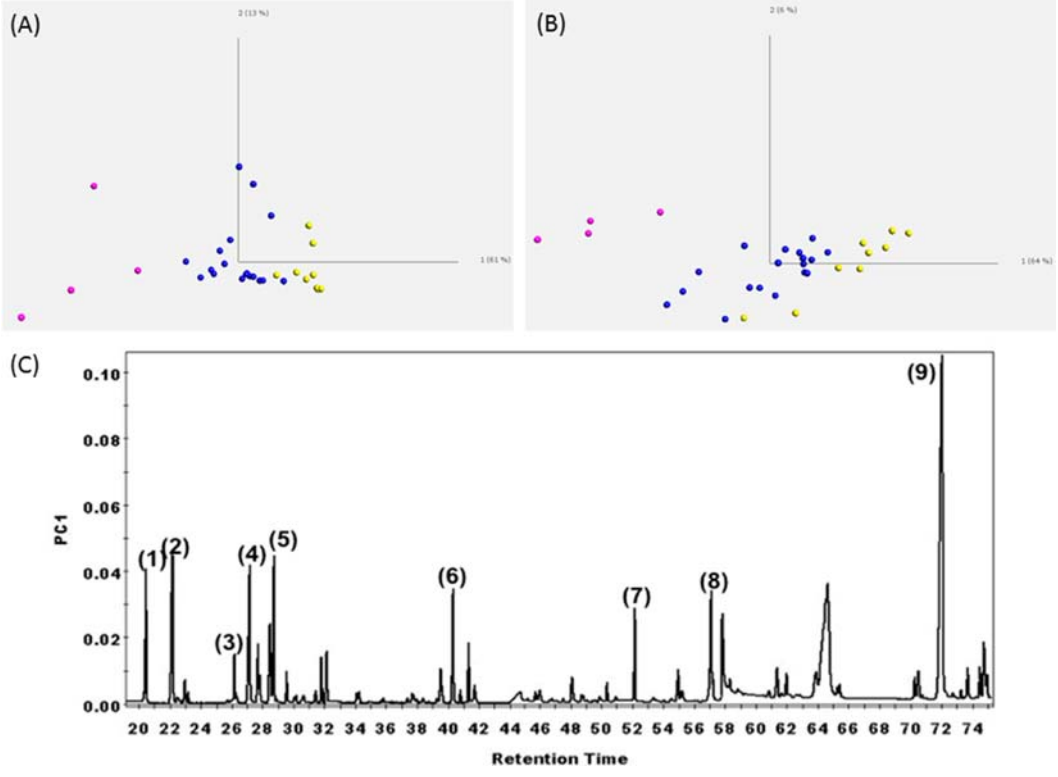

Figure 3. PCA score plots of cleaned and COW-aligned chromatograms: (A) un-normalized; (B) normalized. Colors denote wines of age 2−7 years (yellow), 10−42 years (blue), and 48−60 years (pink). (C) Loading plot of PC1 with nine of the peaks identified as (1) furfural, (2) *cis*-dioxane, (3) benzaldehyde, (4) 5MF, (5) *cis*-dioxolane, (6) *trans*-dioxolane, (7) octanoic acid, (8) unknown, and (9) HMF.

the data to not have negative scores along PC1 and thus to have no negative peaks on the loading plot to keep the loading plot looking as much like a normal chromatogram as possible. We are aware that our choice not to normalize means that peaks with higher intensities will have a larger impact on the model and be more apparent in the loading plots shown in Figures 2 and 3. However, this allows the patterns visible in the loading plots to be recognizable to an analytical chemist and therefore easily read and interpreted as a normal chromatogram. Mathematical normalization unfortunately makes the standard chromatographic patterns unrecognizable as it rescales every peak to the same amount of variance. In addition, as can be seen from Figure 3A,B, normalization does not change the fact that PC1 comprises the age vector with the biggest affect of normalization changing the sample distribution along PC2, which is likely to represent vintage and vinification technology effects. This suggests that the aging of wine largely overwhelms the differences between wines that are present due to the season they were made in or variations among the approaches used to make them (different yeast strains, temperatures, crushing mechanisms, fermentation tanks, barrels used for maturation, etc.). The primary purpose of PCA and PLS in our pipeline is as a graphical user interface for analytical chemists to use as a screening step for univariate data sets. As such, we strove to present the analytical chemist with a multivariate interpretative environment with which they would be as familiar as possible, namely, chromatographic fingerprints with which they have great experience. Thus, we feel that, for this part of the analysis, the visual representation of the chromatographic loading plots outweighs the assignment of equal weights to every variable in the model. Furthermore, we address this variable normalization issue with the use of network reconstruction via Pearson correlation networks and maximum spanning trees. In the network analysis, every peak is analyzed and has an equal opportunity to form a part of the network and Pearson correlation includes vector normalization.

During aging there are likely to be several different mechanisms involved, including oxidation and Maillard reactions. In PC1 the samples correlate with the age of the wine, which points out that the overall kinetic system overrides any one specific mechanism. As such, the connections between the mechanisms at play are more relevant to sample classification than the contributions of any individual mechanism.

After alignment, compounds that appear to correlate with Port wine aging as found in the loading plots were *cis*-dioxane, *cis*-dioxolane, *trans*-dioxolane, *trans*-dioxane, octanoic acid, and HMF as shown in Figure 3C.

The *cis*- and *trans*-dioxane and -dioxolane are formed by the condensation reaction between glycerol and acetaldehyde. These molecules were identified in Port wine by Silva Ferreira et al.,[48] who noted that they increased with age, and, as such, could be used as age markers for Port wine kept under oxidative conditions. Furanic compounds, HMF, 5MF, and furfural, are thought to be products from the Maillard reaction, formed by the fragmentation or cyclization of 3-deoxyosone, a highly reactive intermediary of the reaction.[49] Barrel oak can also be a source of HMF and furfural.[50]

Partial Least-Squares Analysis. PLS analysis was used on the GC-FID-X matrix in an effort to associate specific peaks/fingerprint regions with mechanisms known to be involved in aging. It is worth noting that PLS was not used in its traditional role as a method with which to build calibrated, predictive models (that would therefore be built with training sets and validated with independent test sets). Rather, the goal of our use of PLS-1 was simply as a method with which to perform principal component based regressions in an effort to identify sets of peaks that were associated with known mechanistic markers or potential precursors for volatile compounds. Molecules that are thought to be associated with different mechanisms were selected and quantified from each sample and used as markers to try to find other compounds in GC-FID-X that may be related with the same mechanism. Acetaldehyde and HMF were used as markers for oxidation and the Maillard reaction, respectively. Sotolon was also used in an effort to gather more information about its origin. The concentration of each of the marker molecules was determined for each sample, and the resulting vectors were used as a second data block in PLS.

The resulting PLS coefficient plots show the variables that correlate with each mechanism marker. Some variables have a positive value, which means that these have kinetic vectors which correlate with that of the mechanism marker, and some have negative values, which indicate that they have an inverse correlation with the kinetic vector of the mechanism marker.

For sotolon the correlation is 0.89 (over seven components) for molecules such as furfural, 5-MF, *cis*-dioxane, *cis*-dioxolane, *trans*-dioxane, *trans*-dioxolane, and HMF. We also found some organic acids with negative correlations, which indicate that they were being consumed as sotolon was being formed (Figure 4). The model had correlations of 0.91 for HMF (a Maillard

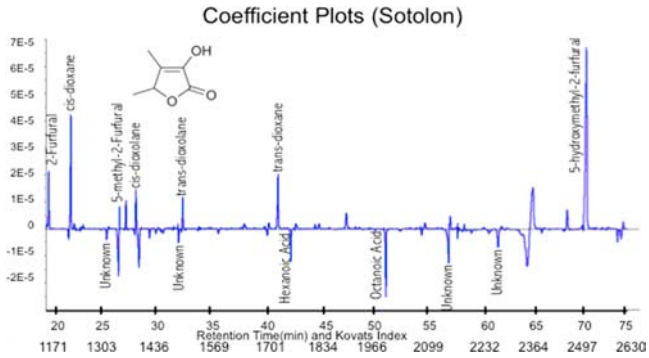

Figure 4. PLS *b* coefficients for sotolon as the *Y* vector with seven latent variables.

reaction marker) and 0.92 to acetaldehyde (an oxidation marker) for seven latent variables. The PLS loading plot for sotolon was very similar to those seen for HMF and acetaldehyde, which means that the mechanisms are correlated and during aging contribute in the same way to the dynamics of the overall process.

Some amino acids, namely, valine, alanine, arginine, glutamine, and aspartate, had relatively high (0.69−0.83) inverse correlations with a number of peaks in the volatile profile, which were themselves correlated to Maillard reaction markers such as HMF and furfural. Thus, it seems likely that these amino acids are major Maillard aroma precursors.

Network Reconstruction. Figure 5 shows the maximum spanning tree derived from the correlation network between all peaks. Each node represents the center of a peak (Kovats index), and each edge represents the best correlation between the peaks. Fold changes between 2- and 60-year-old wines were calculated for each peak and the nodes colored accordingly with

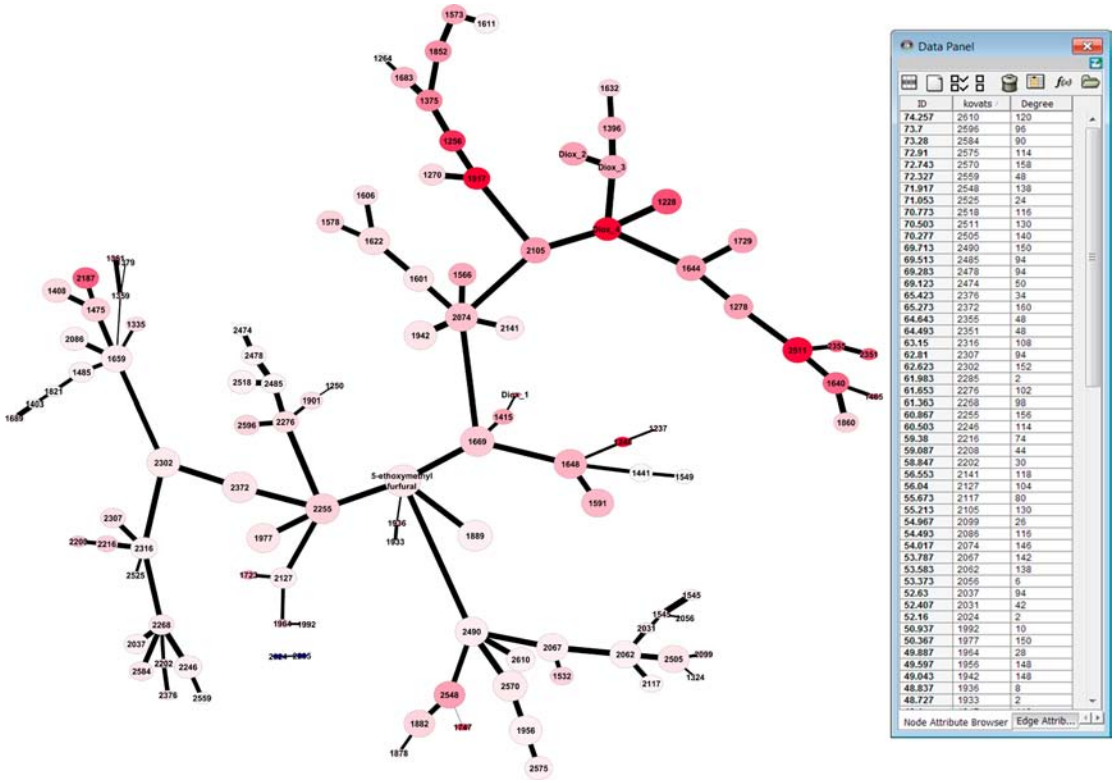

Figure 5. Putative kinetic network. Nodes are colored in shades of red based on the fold change from 2 to 60 years. Node sizes are scaled by the number of other nodes (peaks) that are correlated to them above a Pearson threshold of 0.8. Edge thickness is scaled by the degree of correlation between its two nodes. (Dioxanes in the network are labeled as follows: *cis*-dioxane, Diox 1-cisdioxane; *cis*-dioxolane, Diox 2-cisdioxolane; *trans*-dioxolane, Diox 3-transdioxolane; and *trans*-dioxane, Diox 4-transdioxane.)

shades of red representing increasing concentration and blue representing decreasing concentration. The thickness of each edge has been scaled to represent the level of correlation (thicker lines mean higher correlation values). Furthermore, the size of each node has been scaled to represent the number of correlations it had with other peaks above a threshold of 0.8.

Using pure standards as markers for Maillard and oxidation together with the Kovats index, it is possible to explore and extract more information from the proposed network. In fact, the cyclic acetals of glycerol and ethanal (oxidation products) cluster together on the upper part of Figure 5. In addition, 5-ethoxymethylfurfural, an ester of a major Amadori product (HMF), links two major branches of the network. As such, the network illustrates the aging process with the continuous formation of substances absent in young wines, which explains the aging character of wine. The volatile compounds that relate to 5-ethoxymethylfurfural are coexpressed during aging, thus presenting similar kinetics, and future work will focus on their identification using the respective Kovats index and rich information detectors such as MS. The network representation captures some of the dynamics of the aging process based on the underlying kinetics. In fact, those compounds that are highly correlated to one another (>0.9) are likely to have the same kinetic order, and the network thus enables one to screen molecules according to their kinetic parameters.

We propose that this network is an approximate representation of the underlying chemical reaction network during aging. The higher level of correlation (and therefore the nearer any two peaks are to each other in the network), the higher the probability is that they participate in the same or neighboring reactions. The correlation between compounds drops as you move farther away in a chemical reaction network, as the intervening kinetics of each reaction will cause differences at each step. There are no doubt intermediate compounds for some reactions that were not detected by FID. The network is robust to this missing data as the intervening steps will simply be represented by a lower correlation value of an edge between compounds that were detected. This correlative approach of course is not proof of causation but rather serves as a useful tool for hypothesis generation to prioritize the identification of unknown compounds represented by the peaks.

By using correlation values to targeted compounds (or other variables such as age) we found that we could highlight the regions of the network that are closely associated with them and therefore likely involved in their formation or consumption. To explore regions of the network that may be related to age and particular mechanisms, Pearson correlation values between the peaks and each of the target vectors (sample age, sotolon, acetaldehyde, HMF, glutamate, and alanine) were loaded into cytoscape as node annotations. Alanine and glutamate were selected as target vectors because they were the amino acids best correlated with the GC-FID-X matrix. By sorting the nodes by correlation values and selecting the nodes corresponding to a correlation value with a target above some threshold, portions of the network that correlated with each target vector could be visualized in aqua as shown in Figure 6.

Figure 6A shows the nodes with a 0.86 Pearson correlation to the age of the wines. There is a clearly defined subnetwork that corresponds to age and represents compounds involved in the aging process. It was clear in the PCA diagrams that there are a group of compounds that correspond to aging which are

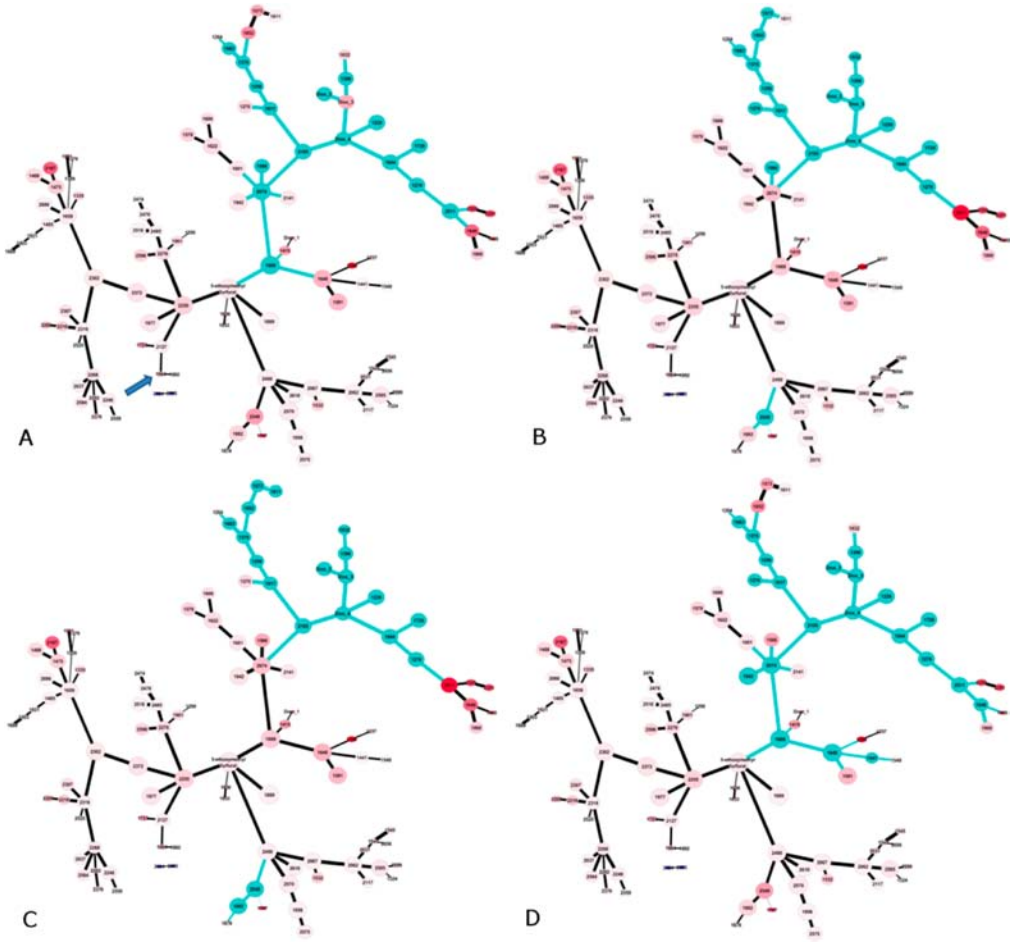

Figure 6. Subnetworks correlating to (A) age, (B) sotolon, (C) HMF, and (D) acetaldehyde. Nodes (compounds) with strong Pearson correlations to these target vectors are colored aqua.

responsible for the first principal component. It is likely that the compounds responsible for the second principal component are due to differences between the vintages of the starting wines. We can see the same pattern in this network, where there are a number of compounds that do not correlate well with age and are likely reflecting vintage differences among the wines.

Panels B, C, and D, respectively, of Figure 6 show the regions of the network (colored as aqua) that correlate (Pearson threshold = 0.86) with sotolon, HMF, and acetaldehyde, respectively. It is clear that there is considerable overlap between the subnetworks correlated with these three compounds, and as such it is possible that there is a mixture of oxidation and Maillard reactions at play in forming these compounds. The subnetwork that correlates with age at a Pearson threshold of 0.86 in Figure 6A clearly overlaps to a very large degree with the subnetworks defined by the correlations to acetaldehyde, HMF, and sotolon. The arrow in Figure 6A shows the node that negatively correlates to both alanine and glutamate (−0.8 Pearson threshold) and as such likely represents the entry point to the volatile network. The fact that the anticorrelation is relatively low (−0.8 to −0.83) probably indicates that there are one to several intermediates between the amino acids and their products' entry into the volatile network.

We believe we have demonstrated that the approach described here using GC-FID univariate data, when used as sample fingerprints, can be used to classify the age of Port wine and to predict potential molecules involved in this process by the deconvolution of time and the kinetics of different aging mechanisms. We began this analysis with the hope that multivariate analysis and network reconstruction would be useful tools with which to study mechanisms related to a perturbation. The PCA score plots allow for sample classification and the visualization of larger peaks that correspond with aging. The PLS loading plots provide the analytical chemist with a familiar set of patterns, namely, virtual chromatograms which point out the larger peaks that appear to be associated with marker compounds for oxidation and Maillard mechanisms. The network reconstruction is very useful in visualizing the relationships between all of the compounds detected via GC-FID and their changes in concentration over time. This view of the data should provide considerably more information in an effort to understand the probable kinetic contexts of the molecules represented by peaks in each chromatogram. Furthermore, it is possible to identify regions of the network that appear to be involved in the formation or consumption of target compounds. As such, the approach described here should indeed be a very powerful tool for the further study of mechanisms and kinetic networks in complex mixtures.

In conclusion, univariate chromatographic signals are less expensive compared to NMR or MS and therefore constitute a valuable tool in a bioanalytical pipeline. However, the use this type of data in an untargeted approach required the

development and use of new data processing methods and graphical user interfaces to extract the maximum amount of information from the data. The approach reported here enables us to (1) minimize the cost of analysis, (2) perform sample classification and contextualization, (3) perform process monitoring of aging or other time series, (4) consider the prospect of building databases from the large amounts of univariate data already available, (5) screen for correlations with known mechanism markers, (6) select biomarkers for identification and further study, and (7) explore the putative kinetic network for a greater understanding of the process being studied.

AUTHOR INFORMATION

Corresponding Author

*Phone: +351 22 558 0001. Fax: +351 22 509 0351. E-mail: asferreira@porto.ucp.pt.

Funding

This research was funded by the project "Wine Metrics: Revealing the Volatile Molecular Feature Responsible for the Wine Like Aroma a Critical Task Toward the Wine Quality Definition" (PTDC/AGR-ALI/121062/2010), partially supported by ESB/UCP plurianual funds through the POS-Conhecimento Program that includes FEDER funds through the program COMPETE (Programa Operacional Factores de Competitividade), by Portuguese national funds through FCT (Fundação para a Ciência e a Tecnologia), Winetech, the Technology and Human Resources Programme, and by the South African National Science Foundation.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Piet Jones, Guy Emerton, and Debbie Weighill for useful discussions about the networks presented in this paper.

REFERENCES

(1) Maillard, L. C. Action des acides amines sur les sucres formation des melanoidines par voie methodique. *Council R. Acad. Sci. Ser. 2* **1912**, *154*, 66−68.

(2) Hodge, J. E. Dehydrated foods, chemistry of browning reactions in model systems. *J. Agric. Food Chem.* **1953**, *1*, 928−943.

(3) Hoffman, T.; Schieberle, P. Evaluation of the key odorants in a thermally treated solution of ribose and cysteine by aroma extract dilution techniques. *J. Agric. Food Chem.* **1995**, *43*, 2187−2194.

(4) Hoffman, T.; Schieberle, P. Identification of potent aroma compounds in thermally treated mixtures of glucose/cysteine and rhamnose/cysteine using aroma extract dilution. *J. Agric. Food Chem.* **1997**, *45*, 898−906.

(5) Pripis-Nicolau, L.; Revel, G.; Bertrand, A.; Maujean, A. Formation of flavor components by the reaction of amino acid and carbonyl compounds in mild conditions. *J. Agric. Food Chem.* **2000**, *48*, 3761−3766.

(6) Marchand, S.; Revel, G. Approaches to wine aroma: release of aroma compounds from reactions between cysteine and carbonyl compounds in wine. *J. Agric. Food Chem.* **2000**, *48*, 4890−4895.

(7) Silva, H. O.; Machado, B. P.; Hogg, T.; Marques, J. C.; Câmara, J. S.; Albuquerque, F.; Silva Ferreira, A. C. Impact of forced-aging process on Madeira wine flavor. *J. Agric. Food Chem.* **2008**, *56*, 11989−11996.

(8) Singleton, V. Oxygen with phenols and related reactions in musts, wines and model systems: observations and pratical implications. *Am. J. Enol. Vitic.* **1987**, *38*, 69−77.

(9) Danilewicz, J. C. Review of reaction mechanisms of oxygen and proposed intermediates reduction products in wine: central role of iron and copper. *Am. J. Enol. Vitic.* **2003**, *54*, 73−85.

(10) Waterhouse, A. L.; Laurie, V. F. Oxidation of wine phenolics: a critical evaluation and hypotheses. *Am. J. Enol. Vitic.* **2006**, *57*, 306−313.

(11) du Toit, W.; Marais, J.; Pretorius, I.; du Toit, M. Oxygen in must and wine: a review. *S. Afr. J. Enol. Vitic.* **2006**, *27*, 76−94.

(12) Kilmartin, P. The oxidation of red and white wines and its impact on wine aroma. *Chem. N. Z.* **2009**, 18−22.

(13) Sulser, H.; Depizzol, J.; Buchi, W. A probable flavoring principle in vegetable-protein hydrolysates. *J. Food Sci.* **1967**, *32*, 611−615.

(14) Dubois, P.; Rigaud, J.; Dekimpe, J. Identification of 4,5-dimethyltetrahydrofuran-2,3-dione in a Flor sherry wine. *Lebensm. Wiss. Technol.* **1976**, *9*, 366−368.

(15) Masuda, M.; Okawa, E.; Nishimura, K.; Yunome, H. Identification of 4,5-dimethyl-3-hydroxy-2(5*H*)-furanone (sotolon) and ethyl 9-hydroxynonanoate in botrytised wine and evaluation of the roles of compounds characteristics. *Agric. Biol. Chem.* **1984**, *48*, 2707−2010.

(16) Silva Ferreira, A. C. *Caractérisation du Vieillissement du vin de Porto. Approche Chimique et Statistique. Rôle Aromatique du Sotolon.* Ph.D. Thesis, Université Victor Segalen Bordeaux 2, 1998.

(17) Câmara, J. S.; Marques, J. C.; Alves, M. A.; Silva Ferreira, A. C. 3-Hydroxy-4,5-dimethyl-2(5*H*)-furanone. *J. Agric. Food Chem.* **2004**, *52*, 6765−6769.

(18) Lavigne, V.; Pons, A.; Darriet, P.; Dubourdieu, D. Changes in the sotolon content od dry white wines during barrels and bottle aging. *J. Agric. Food Chem.* **2008**, *56*, 2688−2693.

(19) Silva Ferreira, A. C.; Barbe, J.; Bertrand, A. 3-Hydroxy-4,5-dimethyl-2(5*H*)-furanone: a key odorant of the typical aroma of oxidative aged Port wine. *J. Agric. Food Chem.* **2003**, *51*, 4356−4363.

(20) Silva Ferreira, A.; Avila, I.; Guedes de Pinho, P. Sensorial impact of sotolon as the "perceived age" of aged port wine. In *Natural Flavors and Fragrances*, 1st ed.; Frey, C., Rouseff, R., Eds.; ACS Symposium Series 908; American Chemical Society: Washington, DC, 2005; pp 141−159.

(21) Hofman, T.; Schieberle, P. Identification of the key odorants in processed ribose-cysteine Maillard mixtures by instrumental analysis and sensory studies. *Spec. Publ. − R. Soc. Chem.* **1996**, *197*, 175−81.

(22) Pham, T. T.; Guichard, E.; Schlich, P.; Charpentier, C. Optimal conditions for the formation of sotolon from α-ketobutyric acid in the French Vin Jaune. *J. Agric. Food Chem.* **1995**, *43*, 2616−2619.

(23) Cutzach, I.; Chatonnet, P.; Dubourdieu, D. Rôle du sotolon dans l'arôme des vins doux naturels, influence des conditions d'élevage et de vieillissement. *J. Int. Sci. Vigne Vin* **1998**, *32*, 223−233.

(24) Silva Ferreira, A. C.; Hogg, T.; Guedes de Pinho, P. Identification of key odorants related to the typical aroma of oxidation-spoiled white wines. *J. Agric. Food Chem.* **2003**, *51*, 1377−1381.

(25) Escudero, A.; Cacho, J.; Ferreira, V. Isolation and identification of odorants generated in wine during its oxidation: a gas chromatography-olfactometric study. *Eur. Food Res. Technol.* **2011**, *211*, 105−110.

(26) Martins, R. C.; Lopes, V. V.; Silva Ferreira, A. C. Port wine oxidation management: a chemoinformatics approach. *60th American Society for Enology and Viticulture Meeting*, 2006

(27) Cevallos-Cevallos, J.; Reyes-De-Corcuera, J.; Etxeberria, E.; Danyluk, M.; Rodrick, G. Metabolomics analysis in food science: a review. *Trends Food Sci. Technol.* **2009**, *20*, 557−566.

(28) Rohman, A.; Che Man, Y. B. Fourier transform infrared (FTIR) spectroscopy for analysis of extra virgin olive oil adulterated with palm oil. *Food Res. Int.* **2010**, *43*, 886−892.

(29) Coimbra, M.; Alves, F. G.; Barros, A.; Delgadillo, I. Fourier transform infrared spectroscopy and chemometric analysis of white wine polysaccharide extracts. *J. Agric. Food Chem.* **2002**, *50*, 3405−3411.

(30) Zhang, J.; Cui, M.; Yun, H.; Yu, H.; Guo, D. Chemical fingerprint and metabolic fingerprint analysis of Danshen injection by

HPLC-UV and HPLC-MS methods. *J. Pharm. Biomed. Anal.* **2005**, *36*, 1029−1035.
(31) Shen, D.; Wu, Q.; Sciarappa, W.; Simon, J. Chromatographic fingerprints and quantitative analysis of isoflavones in tofu-type soybeans. *Food Chem.* **2012**, *130*, 1003−1009.
(32) Ding, Y.; Wu, E.; Liang, C.; Chen, J.; Tran, M.; Hong, C. Discrimination of cinnamon barl and cinnamon twig samples sourced from various countries using HPLC-based fingerprint. *Food Chem.* **2011**, *127*, 755−760.
(33) Gu, H. *NMR and MS-Based Metabolomics: Development and Applications*; Ph.D. Thesis, Purdue University, 2008.
(34) Lee, J.-E.; Hwang, G.-S.; Van Den Berg, F.; Lee, C.-H.; Hong, Y. Evidence of vintage effects on grape wines using H-NMR-based metabolomic study. *Anal. Chim. Acta* **2009**, *19*, 71−76.
(35) Cuadros-Inostroza, A.; Giavalisco, P.; Hummel, J.; Eackard, A.; Willmitzer, L.; Peña-Cortés, H. Discrimination of wine attributes by metabolome analysis. *Anal. Chem.* **2010**, *82*, 3573−3580.
(36) Consonni, R.; Cagliani, L.; Guantieri, V.; Simonati, B. Identification of metabolic content of selected Amarone wine. *Food Chem.* **2011**, *129*, 693−699.
(37) Son, H.; Hwang, G.; Ahn, H.; Park, W.; Lee, C.; Hong, Y. Characterization of wines from grape varieties through multivariate statistical analysis of H NMR spectroscopic data. *Food Res. Int.* **2009**, *42*, 1483−1491.
(38) Gong, F.; Wang, B.; Chau, F.; Liang, Y. Data preprocessing for chromatographic fingerprint of herbal medicine with chemometric approaches. *Anal. Lett.* **2005**, *38*, 2475−2492.
(39) Maillard, B. *Additions Radicalaires de Diols et de leurs Dérivés: Diesters et Acétals Cycliques*; Ph.D. Thesis (no. 325), University Bordeaux 1, Bordeaux, France, 1971.
(40) Van den Dool, H.; Kratz, P. D. A generalization of the retention index system including linear temperature programmed gas-liquid partition chromatography. *J. Chromatogr., A* **1963**, *11*, 463−471.
(41) Pipris-Nicolau, L.; Revel, G.; Marchand, S.; Beloqui, A. A.; Bertrand, A. Automated HPLC method for the measurement of free amino acids including cysteine in musts and wines; first applications. *J. Sci. Food Agric.* **2001**, *81*, 731−738.
(42) Fan, G.; Liang, Y.-Z.; Ying-Sing, F.; Chau, F. Correction of retention time for chromatographic fingerprints of herbal medicines. *J. Chromatogr., A* **2004**, *1029*, 173−183.
(43) Savitzky, A.; Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627−1639.
(44) Du, P.; Kibbe, W. A.; Lin, S. M. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics* **2006**, *22*, 2059−2065.
(45) Dijkstra, E. W. A note on two problems in connection with graphs. *Numer. Math.* **1959**, *1*, 269−271.
(46) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498−2504.
(47) Ribéreau-Gayon, P.; Glories, Y.; Maujean, A.; Dubourdieu, D. *Handbook of Enology: The Chemistry of Wine Stabilization and Treatments*; Wiley: Chichester, UK, 2006; Vol. 2, pp 51−64.
(48) Silva Ferreira, A. C.; Barbe, J.; Bertrand, A. Heterocyclic acetals from glycerol and acetaldehyde in Port wines: evolution with aging. *J. Agric. Food Chem.* **2000**, *50*, 2560−2564.
(49) Martins, S.; Jongen, W.; van Boekel, M. A review of Maillard reaction in food and implications to kinetic modelling. *Trends Food Sci. Technol.* **2001**, *11*, 364−373.
(50) Moutounet, M.; Rabier, P.; Puech, J. L.; Verette, E.; Barillere, M. Analysis by HPLC of extractable substances of oak wood − application to a Chardonnay wine. *Sci. Aliments* **1989**, *9*, 35−51.