Title: Chemiomics: Network Reconstruction and Kinetics of Port Wine Aging.

Ana Rita Monforte[a], Dan Jacobson[b], A. C. Silva Ferreira[a,b]

[a] CBQF – Centro de Biotecnologia e Química Fina – Laboratório Associado, Escola Superior de Biotecnologia, Universidade Católica Portuguesa/Porto, Rua Arquiteto Lobão Vital, Apartado 2511, 4202-401 Porto, Portugal.

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

Corresponding author: António Silva Ferreira

Phone: +351 22 558 0001

Fax: +351 22 509 0351

E-mail address: asferreira@porto.ucp.pt

Running title: Chemiomics: Network Reconstruction and Kinetics of Port Wine Aging.

Abstract

Network reconstruction (NR) has proven to be useful in the detection and visualization of relationships between the compounds present in a Port wine aging dataset. This view of the data provides a considerable amount of information with which to understand the kinetic contexts of the molecules represented by peaks in each chromatogram.The aim of this study was to use NR together with the determination of kinetic parameters to extract more information about the mechanisms involved in Port wine aging. The volatile compounds present in samples of Port wines spanning 128 years in age were measured with the use of GC-MS. After chromatogram alignment a peak matrix was created and all peak vectors compared to one another to determine their Pearson correlations over time. A correlation network was created and filtered based on the resulting correlation values. Some nodes in the network were further studied in experiements on Port wines stored under different conditions of oxygen and temperature in order to determine their kinetic parameters. The resulting network can be divided into three main branches. One is related to compounds that do not directly correlate to age, the second branch contains compounds affected by temperature and the third branch contains compounds associated with oxygen. Compounds clustered in the same branch of the network have similar expression patterns over time; the same kinetic order and are thus likely to be dependent on the same technological parameters. Network construction and visualization provides more information with which to understand the probable kinetic contexts of the molecules represented by peaks in each chromatogram. The approach described here is a powerful tool for the study of mechanisms and kinetics in

complex systems and should aid in the understanding and monitoring of wine quality.

Keywords: Network reconstruction, kinetic parameters, aging, oxygen and temperature.

Introduction

Port wine is a fortified wine produced in a specific area in Northern Portugal known as the Douro region. The diversity of styles offered by Port wines accounts for its uniqueness among other fortified wines and contributes to its recognition around the world. Its remarkable ageing potential means that Port improves in the cask or bottle for much longer than is the case for most wines. This process implies a significant financial cost that must be recovered in the final price of the wine. For this reason, it is of commercial interest to have the maximum amount of information regarding the chemical "fingerprint" for a given age category resulting from the chemical reactions that take place during the process and their dependence on technological parameters.

Kinetic studies in forced aged Port wine prove the existence of compounds dependent on temperature (2-furfural, 5-methylfurfural and hidroxymethylfurfural), oxygen (*cis* and *trans* dioxanes) and on both parameters (sotolon).[1] However, for a holistic vision of the process it is necessary to collect a considerable amount of information about other compounds.

In recent years the concept of biological/chemical "fingerprints" has emerged and comprehensive "omics" technologies (transcriptomics, proteomics and metabolomics) have become a new way of addressing complex systems.[2]

Metabolomics is defined as the study of "as many small metabolites as possible" in a system.[3] High-throughput metabolomics have the potential to identify the functional consequences of induced and natural metabolite variation. However, the data obtained in metabolomics is highly multivariate and multi scale and thus the use of multivariate analysis techniques are

required in order to find trends and/or significant information (biomarkers) in the data.

Network reconstruction has proven to be useful in visualizing the relationships between all the compounds present in a dataset and the changes in their concentration over time.[4] This view of the data provides more information with which to understand the kinetic context of the molecules represented by peaks in each chromatogram and in doing so facilitates the identification of candidates for further study.

The aim of this study is to use network reconstruction and kinetic parameters in order to extract more information related to ageing mechanisms and consequently their dependence on technological parameters such as oxygen and temperature in order to better explain the chemical “fingerprint” for a given port wine category. In order to achieve this, old Port wines (age between 1-129 years) were analyzed by GC-MS and a network constructed from the peaks found in the total ion current (TIC). Some nodes of the network were subsequently identified and quantified in forced aged wines (wines stored on different conditions of oxygen and temperature) to determine their kinetic parameters

Material and Methods

Reagents

Chemicals and standards were obtained from Sigma-Aldrich, USA (a high-purity grade, >99.0%). Dichloromethane and anhydrous sodium sulfate were obtained from Merck, Darmstadt, Germany.

106 *Wine Material*

107 Thirty-four samples of Port wine with ages between 1 and 129 years were
108 used. All wines were maturated in oak barrels and were made following the
109 standard traditional winemaking procedures for Port wine. For the isothermal
110 protocol 16 liters of Port wine made on the year of the experiment (without
111 any oak contact) with pH = 3.4, an amount of dissolved oxygen of 2.5 mg/L, a
112 free SO_2 level of 17 mg/L, 105 g/L of reducing sugars and 20.5% alcohol
113 were used in order to determine kinetic parameters.

114

115 *Isothermal Protocol*

116 Kinetics for each compound detected in Port wine were followed using an
117 isothermal storage method in order to determine the effects of oxygen and
118 temperature. Isothermal storage was performed at different temperatures (20,
119 30, 35 and 40°C) and a series of sequential oxygen saturations inside a
120 hermetically sealed vessel (0 to 10 saturations). Oxygen treatments included
121 0 (F1), 3 (F2), 5 (F3) and 10 (F4) saturations. Each combination of
122 oxygen/temperature was performed in glass vessels containing 500 mL of
123 wine. Oxygen saturation was achhieved by stirring the sample for roughly 1
124 hour until an oxygen concentration of about 8-9 mg/L was reached. The F1
125 group was never supplemented with O_2. F2 was saturated on day 14, 35 and
126 56. F3 was saturated in the beginning of the experiment and at days 14, 28,
127 42 and 56. F4 was saturated with O_2 on days 0, 14, 21, 28, 35, 42, 49, 56 and
128 63.

This forced aging protocol was performed in duplicate. For logistical reasons not all samples were analyzed by GC-MS but the replicate was rather used as a crosscheck procedure.

Dissolved Oxygen Measurements

Oxygen concentrations were measured using a Fibox 3 LCD fiber optic oxygen transmitter, a polymer optical fiber and planar oxygen sensitive spots (5mm sensor spots PSt3), from PreSens Precision Sensing GmbH (Germany). The sensor was positioned in the center of the vessel (500 mL) where it remained in contact with the wine at all times.

Chemical Analysis

The extraction procedure was based on the one described by Silva Ferreira *et al*..[5] Briefly, 50 mL samples of Port were spiked with 50 µL of 3-octanol in a hydro alcoholic solution (1:1, v/v) at 427 mg/L as the internal standard and 5 g of anhydrous sodium sulphate. The wine was extracted twice with 5 mL of dicloromethane. The two organic phases obtained were blended and dried over anhydrous sodium sulphate. Two milliliters of this organic extract was concentrated to 0.4 mL under a nitrogen stream.

GC-MS Analysis

Extracts were analyzed using a Varian 450 gas chromatograph, equipped with a mass detector using Varian 240-MS and Saturn GC-MS workstation software version 5.51. The injector port was heated to 220 °C. The split vent was opened after 30 s. The carrier gas was helium C-60 (Gasin, Portugal), at

1 mL/min, constant flow. The oven temperature was 40 °C (for 1 min), then increased at 2 °C/min to 220 °C, and held for 30 min. All mass spectra were acquired in the electron impact (EI) mode. The ion trap temperatures were 230, 45 and 170 °C. The mass range was m/z 33-350, with a scan rate of 6 scans/s. The emission current was 50 μA, and the electron multiplier was set in relative mode to autotune procedure. The maximum ionization time was 25000 μs, with an ionization storage level of m/z 35. The injection volume was 1 μL, and the analysis was performed in full-scan mode. Compound identification was achieved from comparisons of mass spectra obtained from the sample with those from pure standards injected in the same conditions by comparing the Kovats indices and the mass spectra present in NIST 05 MS Library Database.

Statistical Analysis

Statistical analysis and the developed methodology involving the experimental data were performed with R 2.12.1 for Mac, using the following packages: i) classical multivariate analysis library (mva); ii) main library of Venables and Ripley's (MASS); iii) Harrell miscellaneous (Hmisc); and iv) R-base packages (R-Project, 2012).

Data Pre-processing

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). All of the m/z channels were summed to yield a

179 total ion current (TIC) chromatogram for each sample. The resulting dataset
180 was then imported into The Unscrambler®X 10.1 (Camo, Sweden), where the
181 first stage of the alignment of chromatograms was performed using
182 Correlation Optimized Warping (COW).[6] The saturated peaks were then
183 removed and the baseline corrected. The resulting matrix (GC-MS-TIC matrix
184) was then used for multivariate data analysis as described in the statistical
185 analysis section.
186
187 **Kinetic Network Reconstruction.**
188 In order to attempt to reconstruct the underlying kinetic network, the
189 fingerprint needed to be further compressed to a single value for each
190 putative molecule detected by GC-MS. Thus, each chromatographic peak
191 needed to be replaced with a single value for the intensity and retention-time,
192 at the apex of each peak and therefore a more refined alignment procedure
193 was required. This was achieved as follows: An “average chromatogram” was
194 created by taking the mean of the values at each elution point in the GC-MS-
195 TIC matrix. The wavelet method of Du et al.,[7] which was originally developed
196 for peak calling in peptide mass-spec data, was adapted for finding peak
197 centres in chromatographic data and a Mexican hat wavelet used to
198 determine the location of all of the peaks in the average and sample
199 chromatograms. A custom built Perl program was integrated with the R-
200 based wavelet method to achieve this. The distances between the locations of
201 all of the peaks in each sample chromatogram and the locations of the peaks
202 in the average chromatogram were calculated. It was observed that there was
203 a correlation between peak height and the amount of peak centre 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 greater than 2 and the distance to the nearest sample peak was less than 0.3 minutes 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 less than 2 and the distance to the nearest sample peak was less than 0.15 minutes 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 centres 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 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 amongst them. We applied a Pearson correlation threshold of 0.8 and represented the remaining relationships as a mathematical graph in order to form a correlation network with the nodes representing peaks and the edges weighted with the Pearson correlations between the peak vectors. In order 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

229 the difference between the number 1 and the absolute correlation values) and
230 the subsequent use of a minimum spanning tree (mst) algorithm [8] on the this
231 inverse correlation network. A minimum spanning tree represents the
232 shortest possible path through a graph and, as such, selects for the smallest
233 inverse correlation (i.e highest correlation) pairs between all nodes in the
234 network. The resulting networks were visualized in Cytoscape.[9]
235
236 *Kinetic Parameters*
237 Non-linear regression for model fitting was performed on all data, using a one-
238 step methodology[10] by maximizing the likelihood function for all temperatures
239 and using bootstrap sampling to estimate the prediction residuals sum of
240 squares (PRESS) criterion. [11,12] The lack of fit test was performed to
241 determine the adequacy of the regression model, and the studentised effect
242 was studied at a 5% confidence level.
243 The general rate law for compound formation is:

$$r = \frac{d\,[compound]}{dt} = kobs\,[compound]$$

244 Most of the kinetic models reported in the literature to describe the kinetics of
245 compound formation are zero-, first- or second-order reaction models.

$$\mathrm{C} = \mathrm{C_0} + \mathrm{kt} \quad (\mathrm{zero-order})$$

$$\mathrm{C} = \mathrm{C_0} \exp \mathrm{kt} \quad (\mathrm{first-order})$$

$$\frac{1}{\mathrm{C}} = \frac{1}{\mathrm{C_0}} + \mathrm{kt} \quad (\mathrm{second-order})$$

246 where C_0 is the initial compound concentration, t the time and k the kinetic
247 rate.
248 Therefore the kinetic rate may follow the Arrhenius behavior with temperature.

$$kapp = kref \exp\left[-\frac{Ea}{R}\left(\frac{1}{T}-\frac{1}{Tref}\right)\right]$$

where k_{ref} (day^{-1}) is the kinetic rate at the reference temperature T_{ref} (K), Ea represents the Arrhenius activation energy (J mol^{-1}), and R is the universal gas constant (J mol^{-1} K^{-1}).

It was assumed that the studied compounds tend to follow first-order kinetics[13]

$$C(t) = C0 + kref \exp\left[-\frac{Ea}{R}\left(\frac{1}{T}-\frac{1}{Tref}\right)t\right]$$

Statistical analysis and the developed methodology involving the experimental data were performed with R 2.12.1 for MAC using R-base packages (R-Project).

Results and Discussion

Figure 1 shows the network derived from the correlation between all peaks, it is possible to observe three main branches indicated as A, B and C.

It was observed that branch A (Figure 2) included compounds such as hexanoic (1886), octanoic (2110) and decanoic (2308) acids, which cluster together. The nodes are colored in shades of red representing increasing concentration and blue representing decreasing concentration.

The ANOVA analysis on perturbed wines shows that acids are not dependent on temperature or oxygen.

In fact its concentration during wine ageing is regulated by the esterification/hydrolysis between the acid and the correspondent ester, acids may be formed due to hydrolysis, be lost through chemical esterification, or remain at a near constant equilibrium concentration depending on their initial,

post-fermentation levels.[14] When the system is in balance, there is a constant correlation between the concentrations of the substances present, governed by the mass action law.

Branch B (Figure 3) contains two compounds that are demonstrated to be highly sensitive to temperature due to their high activation energy (HMF (2506) and furfural (1446)), which suggests that the other compounds in the branch may also be sensitive to temperature.

A summary of the ANOVA analysis results is present on each figure as well as the time expression of each compound identified. All the compounds identified tend to increase during ageing and ANOVA shows that significant differences were established across the temperature conditions, meaning that the compounds identified were all dependent on temperature.

It is possible to observe three sub-branches apparently related to activation energy value. The branch of vanillin (2541), γ-ethoxycarbonyl-γ-butyrolactone (2207), diethyl tartrate (2361) and triethyl citrate (2480) have an Ea between 16 and 36 kJ/mol.

The organic acids esters, diethyl tartrate (2361) and triethyl citrate (2480), tend to increase due to the high concentration of the respective acids (tartaric and citric acids), by the reaction between the acid and the ethanol (esterification reaction). According to Ribéreau-Gayon and Peynoud [14] polyprotic acids such as tartaric, citric and malic acid esterified more rapidly than the monoprotic acids acetic, propanoic and butanoic acid and temperature accelerates the reaction.

The γ-ethoxycarbonyl-γ-butyrolactone (2207) is a lactone that probably results from glutamic acid or related compounds, and is dependent on wine ageing

296 and contributes a cherry aroma to wine.[15] The formation of the lactone is an
297 esterification reaction. Vanillin (2541) is a phenolic aldehyde related with
298 barrel ageing. Its sensitivity to temperature in wines that were not subject to
299 wood contact is still unknown.
300 Diethyl glutarate (2392) and ethyl-2-furoate (1525) are located in a second
301 sub branch. These two compounds have activation energies of 74 and 56
302 kJ/mol which is higher than those present in the posterior sub-branch. Ethyl-2-
303 furoate (1525) is also dependent on temperature and its formation is related to
304 the Maillard reaction.
305 In summary, this branch contains compounds related to temperature and it
306 would appear that there is an association between network location and
307 activation energy values.
308 Branch C (Figure 4) contains compounds related to oxidation such as
309 benzaldehyde (1508) (formed by the oxidation of benzyl alcohol) and
310 dioxanes (1486 and 1863) and dioxolanes (1619) (formed by the reaction of
311 ethanal and glycerol). One region of this branch, as shown in Figure 3.6.3.4,
312 seems to idicate a correlation between sotolon (2232) and 3-hydroxy-2-
313 methyl-4H-pyran-4-one (1996) (maltol). These two compounds have
314 somewhat similar chemical structures, i.e. a furanone and a pyranone
315 respectively and both have a very sweet aroma, normally classified in the
316 same group of aromas with flavors of “burnt sugar, caramel and maple”.
317 However, the sensorial impact of each compound are at different orders of
318 magnitude, ppb for sotolon and ppm for maltol.
319 Sotolon is dependent on temperature and oxygen which suggests that maltol
320 is dependent on the same parameters and, in fact, the ANOVA analysis

indicates that the compound is sensitive to both parameters and has a very high activation energy when compared to sotolon. This suggests that maltol is much more sensitive to temperature than is sotolon.

Yaylayan et al observed that maltol formation is related with 1-deoxyosone, an important intermediary of the Maillard reaction.[16] However, there are no studies on the impact of oxidation on its formation.

Other compounds present in this sub-branch (indicated in Figure 4. as C.1.) of the network appear to be connected with sotolon and maltol. These compounds are probably dependent on oxygen and temperature and their identification could be useful in the unraveling of the connection between the Maillard mechanism and oxidation.

These observations would seem to validate the assumptions made that compounds grouped on the same branch of the network have the same time expression and consequently the same kinetic order. It was also observed that compounds grouped together are sensitive to the same technological parameter.

We believe that this network-driven approach to the study of large kinetic systems is proving to be a useful tool for identifying putative compounds of interest and to do mechanistic hypothesis generation which can subsequently be tested experimentally. We would recommend that this type of Chemiomics approach become a standard workflow for the future investigation of complex chemical systems.

Acknowledgments

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) and 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 national funds through FCT (Fundação para a Ciência e a Tecnologia), the Wine Industry Network of Expertise and Technology (Winetech IWBT BI 13/01) andthe Technology and Human Resources Programme and the South African National Science Foundation.

References

1. Martins, R. C.; Monforte, A. R.; Silva Ferreira, A., Port Wine Oxidation Management: A Multiparametric Kinetic Approach. *J. Agr. Food Chem.* **2013**, 61 (22), 5371-5379.

2. Courant, F.; Antignac, J.-P.; Monteau, F.; Le Bizec, B., Metabolomics as a Potential New Approach for Investigating Human Reproductive Disorders. *J. Proteome. Res.* **2013**, 12 (6), 2914-2920.

3. Cevallos-Cevallos, J. M.; Reyes-De-Corcuera, J. I.; Etxeberria, E.; Danyluk, M. D.; Rodrick, G. E., Metabolomic analysis in food science: a review. *Trends Food Sci. Tech.* **2009**, 20 (11–12), 557-566.

4. Jacobson, D.; Monforte, A. R.; Silva Ferreira, A. C., Untangling the chemistry of Port wine aging with the use of GC-FID, multivariate statistics and network reconstruction. *J. Agr. Food Chem.* **2013**, 61, 2513-2521.

5. Silva Ferreira, A. C.; Barbe, J.-C.; Bertrand, A., 3-Hydroxy-4,5-dimethyl-2(5H)-furanone: A Key Odorant of the Typical Aroma of Oxidative Aged Port Wine. *J. Agr. Food Chem.* **2003**, 51, 4356-4363.

6. Gong, F.; Liang, Y.-Z.; Fung, Y.-S.; Chau, F.-T., Correction of retention time shifts for chromatographic fingerprints of herbal medicines. *J. Chromatogr. A* **2004**, 1029 (1–2), 173-183.

7. 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 (17), 2059-2065.

8. Dijkstra, E. W., A note on two problems in connection with graphs. *Numer. Math.* **1959**, 1, 269-271.

394 9. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.;
395 Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T., Cytoscape: A Software
396 Environment for Integrated Models of Biomolecular Interaction Networks.
397 *Genome Res.* **2003**, 13 (11), 2498-2504.
398 10. Arabshahi, A.; Lund, D. B., Considerations in calculating kinetic
399 parameters from experimental data. *J. Food Process. Eng.* **1985**, 7, 239-251.
400 11. Hastie, T.; Tibshirami, R., Discriminant adaptative nearest neighbour
401 classification. *IEEE T. Pattern Anal.* **1996**, 18, 607-616.
402 12. Bates, D. M.; Watts, D. G., Non-linear Regression Analysis and its
403 Applications. John Wiley & Sons: New York, **1998**.
404 13. Baig, S.; Rehman, F., Signal Modeling using Singular Value
405 Decomposition. In Advances in Computer, Information, and Systems
406 Sciences, and Engineering, Elleithy, K.; Sobh, T.; Mahmood, A.; Iskander, M.;
407 Karim, M., Eds. Verlag: Netherlands, **2006**; pp 31-37.
408 14. Ramney, D.; Ough, C. S., Volatile ester hydrolysis or formation during
409 storage of model solutions of wines. *J. Agr. Food Chem.* **1980**, 28, 928-934.
410 15. Wurz, R. E. M.; Kepner, R. E.; Webb, A. D., The Biosynthesis of
411 Certain Gamma-Lactones from Glutamic Acid by Film Yeast Activity on the
412 Surface of Flor Sherry. *Am. J. Enol. Viticult.* **1988**, 39 (3), 234-238.
413 16. Yaylayan, V. A.; Mandeville, S., Stereochemical Control of Maltol
414 Formation in Maillard Reaction. *J. Agr. Food Chem.* **1994**, 42 (3), 771-775.
415

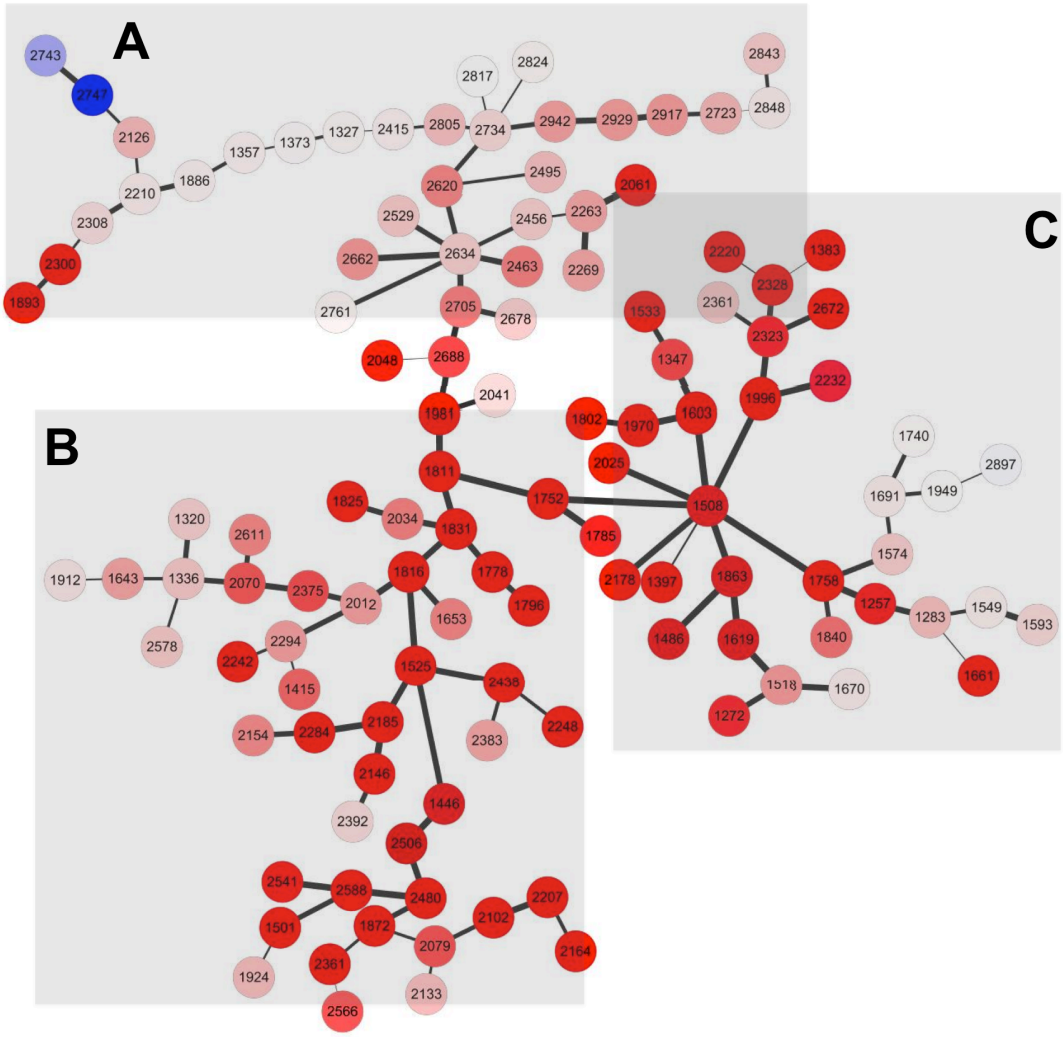
Figure Captions

Figure 1. GC-MS Putative Kinetic Network.

Figure 2. Putative Kinetic Network. Branch A Identification: 2210 – Octanoic acid, , 1886– Hexanoic acid, 2308 – Decanoic Acid.

Figure 3. Putative Kinetic Network. Branch B (2070. – pantolactone, 2392 – diethylglutarate, 2541 – vanillin, 2207 - γ-ethoxycarbonyl-γ-butyrolactone., 1525 – ethyl furoate.

Figure 4. Putative Kinetic Network. Branch C (2232 – sotolon, 1508 – benzaldehyde, 1486 – cis-5-hydroxy-2-methyl-1,3-dioxane, (1996 – Maltol). C.1. Sub-branch related with oxygen and temperatur



\ Figure 1. GC-MS Putative Kinetic Network.

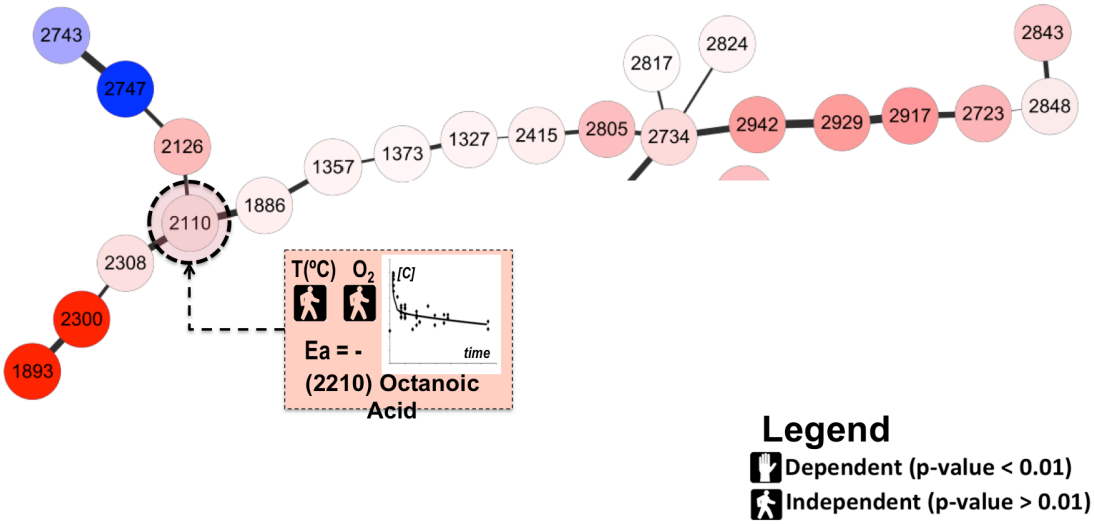


Figure 2. Putative Kinetic Network. Branch A Identification: 2210 – Octanoic acid, , 1886– Hexanoic acid, 2308 – Decanoic Acid.

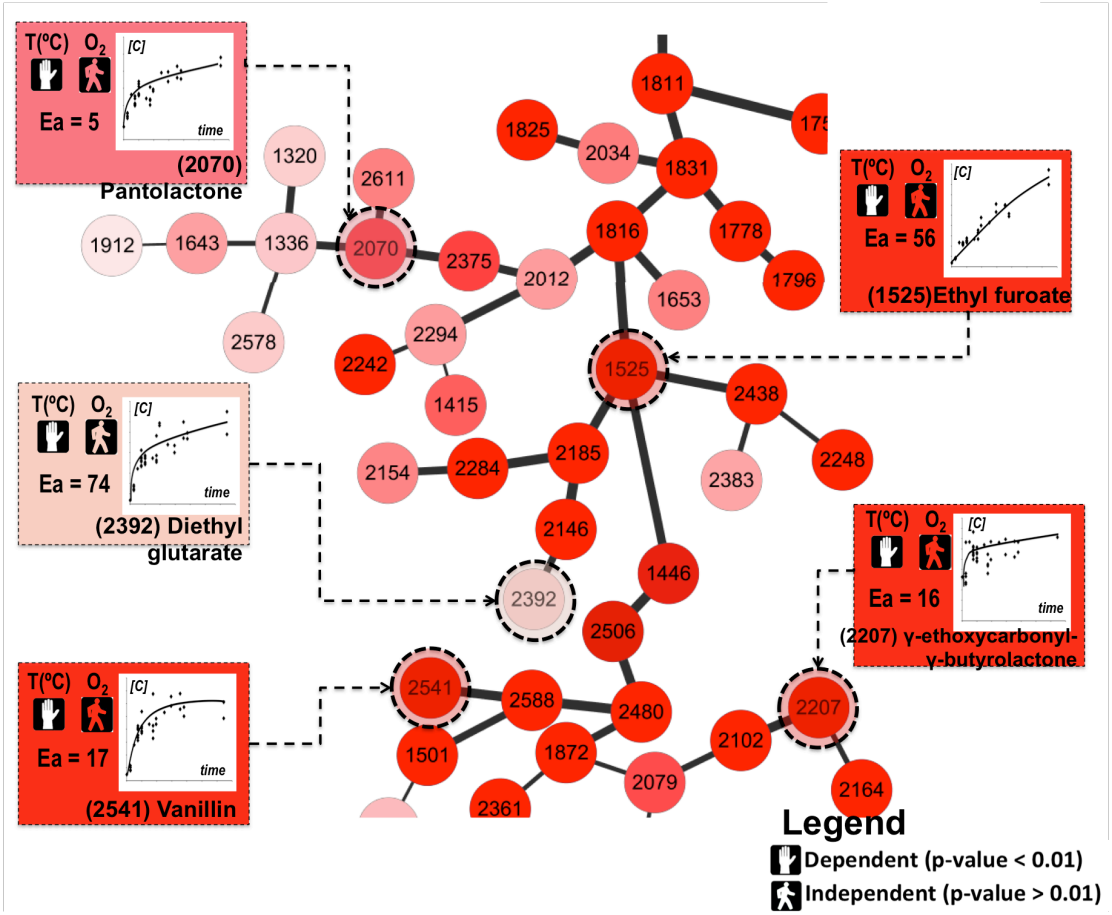


Figure 3. Putative Kinetic Network. Branch B (2070. – pantolactone, 2392 – diethylglutarate, 2541 – vanillin, 2207 - γ-ethoxycarbonyl-γ-butyrolactone., 1525 – ethyl furoate.

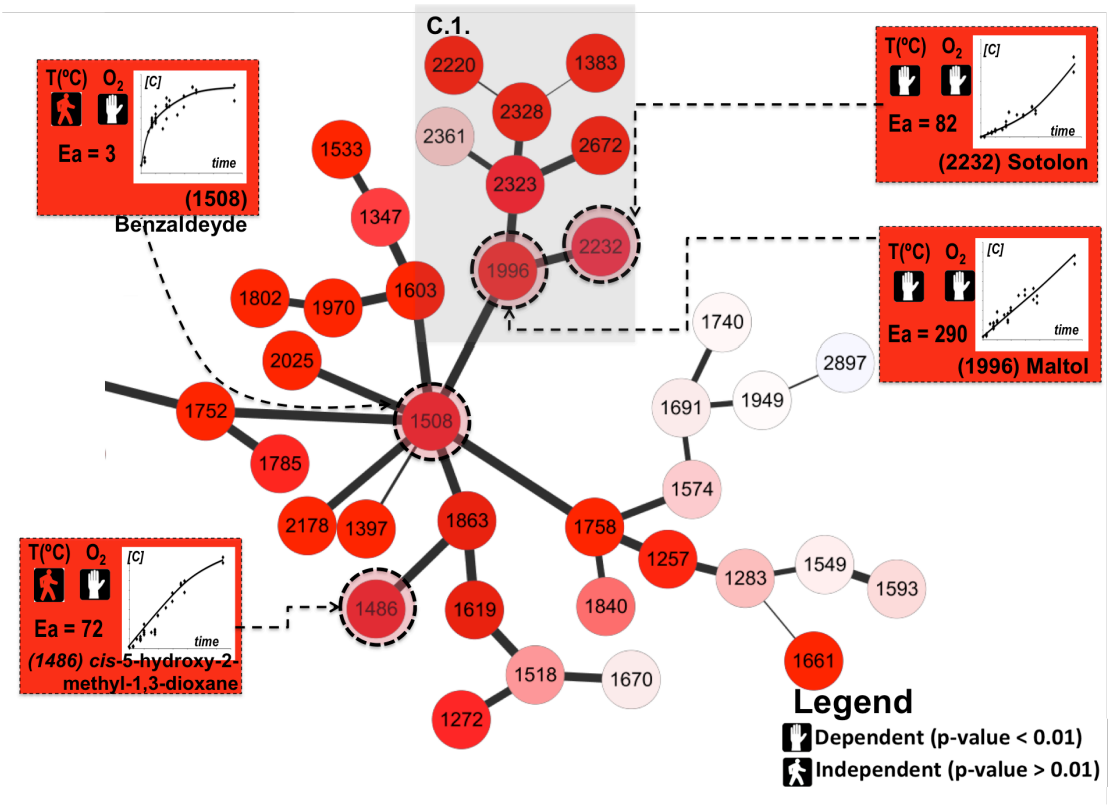
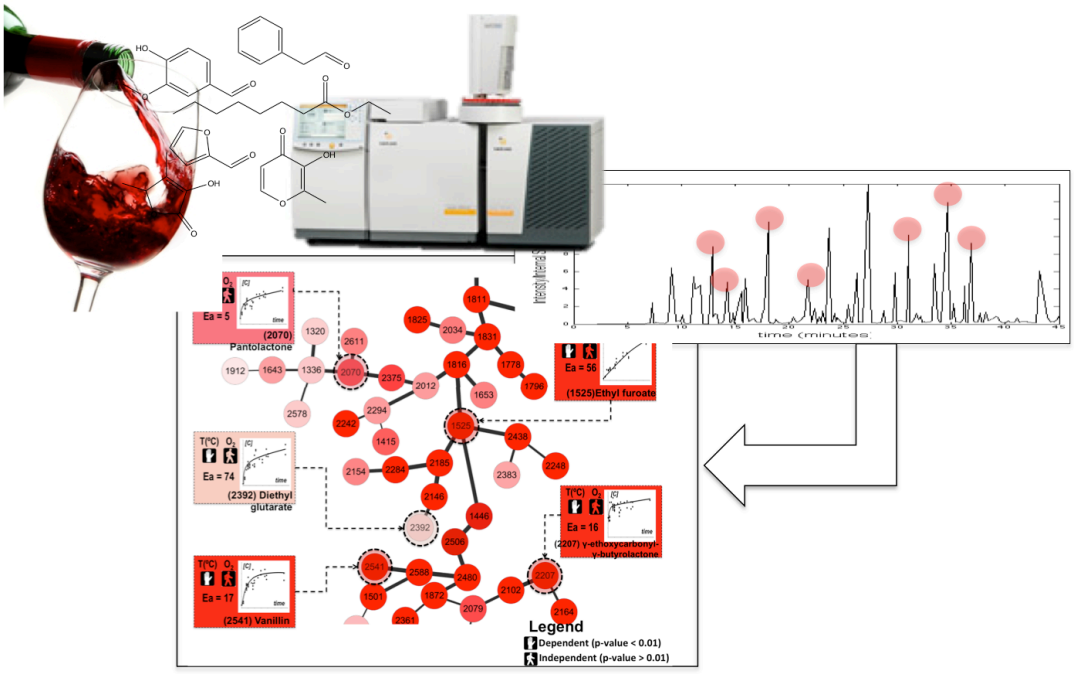


Figure 4. Putative Kinetic Network. Branch C (2232 – sotolon, 1508 – benzaldehyde, 1486 – cis-5-hydroxy-2-methyl-1,3-dioxane, (1996 – Maltol). C.1. Sub-branch related with oxygen and temperature.

For table of contents only



TOC Graphic