Food Chemistry 140 (2013) 210–216

Contents lists available at SciVerse ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Analytical Methods

Amino acid profiles and quantitative structure–property relationship models as markers for Merlot and Torrontés wines ☆

Pablo R. Duchowicz [a,*,1], Miguel A. Giraudo [b,1], Eduardo A. Castro [a,1], Alicia B. Pomilio [c,1,2]

[a] Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA (Universidad Nacional de La Plata, CCT La Plata-CONICET), Diag. 113 y 64, C.C. 16, Suc. 4, B1900 La Plata, Argentina
[b] Carrera de Ciencia y Tecnología de los Alimentos, Universidad Nacional de Lanús, 29 de Setiembre 3901, B1826 Lanús, Provincia de Buenos Aires, Argentina
[c] IBIMOL (ex PRALIB;CONICET, UBA), Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, Junín 956, C1113AAD Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 20 August 2011
Received in revised form 27 December 2012
Accepted 17 February 2013
Available online 24 February 2013

Keywords:
Wines
Aminograms
Quantitative structure–property relationships
Replacement method

ABSTRACT

Quantitative structure–property relationships (QSPRs) were applied to the aminograms obtained by HPLC in our laboratories for Torrontés and Merlot wines. Dragon theoretical descriptors were derived for a set of optimized amino acid structures with the purpose of establishing QSPR models. The statistical Replacement Method was used for designing the best multi-parametric linear regression models, which included structural features selected from a pool containing 1497 constitutional, topological, geometrical or electronic molecular descriptors. Predicted QSPR results were in good agreement with experimental amino acid profiles. The developed QSPR approach showed to be of practical value for distinguishing each wine varietal, and for calculating experimentally non-available amino acid concentrations of Torrontés and Merlot wines. It was also useful for assessing wine authenticity; the models were especially suitable for Merlot and Torrontés wines.

1. Introduction

The determination of wine authenticity is a crucial issue in wine quality control and safety. The fingerprint of food allows its univocal identification and characterization. Therefore, it is necessary to use biomarkers to characterize a wine, as well as other foods. Knowledge of the biomarker concentration profile behavior is further essential during manufacture and control processes of wines (Arvanitoyannis, 2003, chap. 20).

The evolution of amino acids and ammonium during grape ripening had recently been reported and was used to differentiate grape varieties and (organic and nonorganic) cultured systems (Garde-Cerdán et al., 2009). Not only amino acids (Košir & Kidric, 2002; Soufleros, Bouloumpasi, Tsarchopoulos, & Biliaderis, 2003), but also anthocyanins (Košir & Kidric, 2002), aroma compounds (Zalacain, Marín, Alonso, & Salinas, 2007) and metals (Ajtony et al., 2008) had been reported to typify wines.

The amino acid content of grapes is dependent on climate conditions, culture, and soil managements (Molnár-Perl, 2005). But the free amino acid content of wines is closely related to quality, authenticity, and the technology of winemaking. Amino acids had been used as a criterion of authenticity in different countries, thus displacing other so far used criteria, such as those based on anthocyanins. Amino acids had shown to be more stable, safe and reliable (Arvanitoyannis, 2003, chap. 20; Molnár-Perl, 2005). Therefore, the profile of free amino acids and ammonia (aminogram) of white and red wines had previously been determined in our laboratories by high performance liquid chromatography (HPLC) with fluorescence and UV detection (Giraudo et al., 2006).

In the present work, the interest was focused on Torrontés and Merlot Argentine wines. Argentina is the fifth largest world wine producer with a long list of premium varieties (Organisation Internationale de la Vigne et du Vin, 2009), such as the white wine, Torrontés (Morosi, 2008). Torrontés is the most distinctive of all Argentine wines, including both white and red, because Argentina is the only country to produce it. It had been considered a wholly Argentine variety (emblem of Argentine white wine). This grape is part of the *criollo* vines (Atkins, 2009).

Multivariate analysis as principal component analysis (PCA), discriminant analysis (DA), canonical analysis (CA), cluster analysis (CLA), had, in most cases, been used in wine differentiation and classification according to geographical origin (Arvanitoyannis, 2003, chap. 20). Also, a new classification method, called Classification and Influence Matrix Analysis (CAIMAN), based on the influence matrix (or leverage matrix) had recently been proposed for the geographic classification of samples of wine and olive

☆ This work was part of the Ph.D. Thesis of Miguel Giraudo, approved with awards on May 21, 2009.

* Corresponding author.
E-mail addresses: pabloducho@gmail.com, pomilio@ffyb.uba.ar (P.R. Duchowicz), mgiraudo@unla.edu.ar (M.A. Giraudo), eacast@gmail.com, castro@quimica.unlp.edu.ar (E.A. Castro), abpomilio@sinectis.com.ar (A.B. Pomilio).
[1] All authors contributed equally to this work.
[2] Tel./fax: +54 (011) 4814 3952

0308-8146/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodchem.2013.02.064

oil, and the results had been compared with DA, by focusing great attention on the model predictive capabilities (Todeschini, Ballabio, Consonni, Mauri, & Pavan, 2007).

The aim of this work was to obtain Merlot red wine and Torrontés white wine at a pilot scale using industrial parameters, carrying it out by the support of a local wine factory that provided the facilities. The yeast strains, *Saccharomyces cerevisiae* var. *bayanus*, *S. cerevisiae* var. *ellipsoides*, and native flora (yeasts and bacteria) were used.

Since aminograms are not clear enough for non-expertise people to distinguish both wine varietals, a mathematical model was searched in order to discriminate between both Torrontés and Merlot Argentine wines. Therefore, the aim was focused on the design of predictive QSPR models (Hansch & Leo, 1995) (Duchowicz, Giraudo, Castro, & Pomilio, 2011; Pomilio, Duchowicz, Giraudo & Castro, 2010b; Pomilio, Giraudo, Duchowicz & Castro, 2010a), which could serve as suitable tools for estimating aminograms of Merlot red and Torrontés white wines, whose experimental data had been obtained in our laboratories.

In the present analysis, a pool containing 1497 theory-based descriptors (Katritzky, Lobanov, & Karelson, 1995; Todeschini & Consonni, 2000) was explored computed by means of the Dragon software, and established QSPRs for each white and red wine under analysis by means of the Replacement Method technique (Duchowicz, Castro, Fernández, & González, 2005; Duchowicz, Mercader, Fernández, & Castro, 2008a; Duchowicz, Talevi, Bruno-Blanch, & Castro, 2008b).

2. Materials and methods

2.1. Materials

The following reagents and solvents were used for liquid chromatography: Trihydrated sodium acetate, HPLC grade; triethylamine, *p.a.*; disodium EDTA, *p.a.*; 85% phosphoric acid, *p.a.*; acetonitrile HPLC grade (JTBaker, Interchemistry, Buenos Aires, Argentina). The internal standards norleucine and α-aminobutyric acid (AABA) were purchased from Sigma (USA). Water was obtained by filtration through a Milli-Q system, and samples were filtered through Millex-SLCR13 Millipore filters (Biopore, Buenos Aires, Argentina). All solvents were filtered daily through GVWP04700 Millipore filters (Buenos Aires, Argentina).

AccQ•Tag kit contained an ampoule with 17 standards of amino acids each with a concentration of 2.5 mM, except for cystine, which had a concentration of 1.25 mM.

2.2. Instrumentation and chromatographic conditions

An Agilent liquid chromatograph Model 1050 was used, composed of a ternary pump, on-line vacuum degasser, automatic injector, column oven, microbore connectors; UV–Vis variable Model C detector with 1 μl-cell; Model 1046 fluorescence detector and HP 3396 II Integrator.

A C_{18} Hypersil column (200 mm length × 2.1 mm i.d.; 5 μm of particle diameter; pore diameter of 12 nm) and a C_{18} guard column of 3.9 × 20 mm were used. As mobile phase the following ternary system was used: *Solvent A:* 140 mM sodium acetate, 17 mM triethanolamine, pH 5.05, containing 1 mM disodium edetate. *Solvent B:* Acetonitrile, HPLC grade. *Solvent C:* Water, HPLC grade. The purity of the mobile phase was monitored by blind tests by UV and fluorescence detection. Alternative Solvent B for high-pressure mixing systems: 60% acetonitrile.

The system was balanced with 100% Solvent A for 10 min before injecting the sample. Complete separation using a ternary gradient (Millipore Corporation) lasted ca. 35 min, including column rebalance.

Flow rate: 0.33 mL min^{-1}. *Detection:* UV: 248 nm. Fluorescence: excitation at 250 nm and emission at 395 nm. *Oven temperature:* 40 °C. *Injection volume:* 5 μL.

Retention times were recorded. Measuring the relative concentration percent to total concentration of main amino acids assessed a composition profile.

2.3. Merlot red and Torrontés white wines

Merlot red and Torrontés white argentine wines were obtained at a pilot scale using *S. cerevisiae* var. *bayanus*, *S. cerevisiae* var. *ellipsoides* and native flora as yeast strains. Wine making began by crushing the grapes into a must, which was further pressed. The extracted juice was poured into a tank. During the settling, the grape sediment separated from the juice and settled to the bottom of the tank. Over a period of time the juice was transferred into different tanks and allowed to settle further (Giraudo et al., 2006). The period of five weeks was considered in the experimental design to obtain Merlot and Torrontés varietals from the corresponding grape varieties. Representative samples were taken at zero time, and during the next four weeks (one sample per week) to monitor the essential fermentation parameters using the official methodology of the National Wine Institute (Ministry of Agriculture, Livestock, Fishing and Food of Argentina). Parameters to be controlled were: pH, total acidity, reducing sugars, free sulfite and alcoholic degree. Aliquots were obtained from fermentation tanks 1–6 after 1, 2, 3, 5, 7 and 11 weeks of wine fermentation.

2.3.1. Grape varieties and fermentative process for obtaining Merlot and Torrontés wines

Both advanced-fermented Merlot and Torrontés grape varieties were purchased in a local market. The initial population was counted, so that 40 millions CFU mL^{-1} for Merlot, and 10 millions CFU mL^{-1} for Torrontés were found.

About 100–110 kg of each argentine grape variety were placed in separate fermentation tanks, provided by a local wine industry, thus obtaining 60–70 L of must. The initial processing of the must included a defined amount of sulfite added, pH value, total acidity, and sugar content for each fermentation tank and grape variety.

2.3.2. Merlot and Torrontés sample preparation

After the first week, the grape skin was separated out (without stirring) in order to allow sampling up to the bottom of the tank. Samples were obtained in duplicate to perform analyses. Samples for determination of amino acids were stored at −20 °C until analysis. They were filtered through SS blue band filter paper to retain coarse material. Two drops of 1 mol L^{-1} sodium hydroxide per mL of filtrate were added to each filtrate to obtain a pH about 6. If not neutralized, amino acids could not have been derivatized at a lower pH of the filtrate.

Samples were further filtered through Millex LCR 13 filters, and to 200 μL of each sample 4 μL of 2.5 mM AABA (internal standard) and 800 μL of water were added.

Results were expressed as a concentration range in ppm for all amino acids that were present in Merlot red wine and Torrontés white wine samples ($n = 5$ each) for various previously defined processing times.

2.4. Procedures

2.4.1. Analysis of free amino acids

Free amino acids of the wines under study were determined by AccQ • Tag (Waters Corporation), which showed to be the most suitable technique due to the high speed, sensitivity (detection

limit, 2.5–5.0 ng) and resolution of the chromatographic peaks. The Waters commercial kit provided by D'Amico Sistemas SA (Buenos Aires, Argentina) was used for precolumn derivatization.

The internal standard α-aminobutyric acid (AABA) and the corresponding borate buffer were added, and further derivatized according to the specifications of the AccQ · Tag kit of Waters. Once the profile of amino acids was obtained, the concentrations of the main amino acids were calculated by comparison with the AccQ · Tag kit of amino acids' standards and the internal standard concentration.

Five samples each of Merlot red and Torrontés white wines were obtained at a pre-fixed fermentation extent and using the three yeast types mentioned above, and further analyzed by HPLC. Then, mean values of the amino acid content were obtained. Once the chromatograph was optimized with the standards of amino acids (variation coefficient less than 5%), alternately 3 samples and a standard were injected.

Chromatograms were obtained using the manual method of derivatization. If the automated method would have been used, variation coefficients would have been lower. A calibration curve was constructed using the standard as previously reported (Giraudo et al., 2006).

2.4.2. Determination of proteins in Merlot and Torrontés wines

The method of Pyrogallol Red (PROTIU/LCR; Wiener Laboratories, Rosario, Argentina) was used, previous comparison with the Kjeldahl method, because of being an approach for protein determination in very low concentrations.

2.5. Statistical analysis

The statistical program NWA Quality Analyst, release 6.1 (2007) was used to study the accuracy and precision of the method applied.

2.6. QSPR studies

2.6.1. Data set of amino acids

The experimental concentration of each amino acid in the aminogram of Merlot red and Torrontés white wines was expressed as the mean concentration range (ppm), which was then converted into logarithm units for modeling purposes [$\log_{10}$ range (ppm)].

2.6.2. Geometry optimization and theoretical descriptors calculation

The structure of each amino acid was pre-optimized with the molecular mechanics force field (MM+) procedure included in the Hyperchem 6.03 package (Hypercube, Inc.). The levogyre enantiomer was chosen because of its natural occurrence. The Semi-Empirical Method PM3 (Parametric Method-3) was used to refine the resulting geometries by the Polak-Ribière algorithm and a gradient norm limit of 0.01 kcal $Å^{-1}$.

The Dragon software (Milano Chemometrics and QSAR Research Group) enabled to calculate 1497 theoretical molecular descriptors, including descriptors of all types, such as Constitutional, Topological, Geometrical, Charge, GETAWAY (Geometry, Topology, and Atom-Weights AssemblY), WHIM (Weighted Holistic Invariant Molecular) descriptors, 3D-MoRSE (3D-Molecule Representation of Structures based on Electron diffraction), Molecular Walk Counts, BCUT descriptors, 2D-Autocorrelations, Aromaticity Indices, Randic Molecular Profiles, Radial Distribution Functions, Functional Groups, Atom-Centred Fragments, Empirical and Property-based descriptors (Todeschini & Consonni, 2000). Finally, five quantum-chemical descriptors not provided by Dragon were added: molecular dipole moments, total energies, homo–lumo energies, and homo–lumo gap ($\Delta_{\text{homo–lumo}}$).

2.6.3. Model search

The replacement method (RM) (Duchowicz, Mercader, Fernández, and Castro, 2008a; Duchowicz, Talevi, Bruno-Blanch, and Castro, 2008b; Duchowicz et al., 2005) was applied as molecular descriptors selection approach, an algorithm that had been proposed by our research group some time ago, being an efficient optimization tool which generates multi-parametric linear regression QSPR models by searching the set D of D descriptors for an optimal subset d of $d \ll D$ ones with minimum model's standard deviation S.

The quality of the results achieved with this technique was quite close to that obtained by performing an exact (combinatorial) full search (FS) of molecular descriptors, although, of course, requires much less computational work. The RM provided models with better statistical parameters than the Forward Stepwise Regression procedure and similar ones to the more elaborated Genetic Algorithms (Mitchell, 1998). The Matlab 5.0 software was used for all our calculations (The MathWorks, Inc., 2004).

2.6.4. Model validation

The design of a properly validated model constitutes the most important step for every QSPR analysis in order to generate predictive models of global applicability, and not limited to function only correlatively. The Leave-One-Out Cross Validation procedure (*loo*) was practiced over each linear regression. The parameters R_{loo} and S_{loo}, the correlation coefficient and standard deviation of Cross Validation, each measured the stability of the developed QSPRs upon inclusion/exclusion of compounds. R^2_{loo} should be higher than 0.5 for obtaining a validated model according to the literature (Golbraikh & Tropsha, 2002).

The standard practice of validation was applied, that consisted on omitting from the complete molecular set some amino acids, which constituted the 'test set', denoted here as "*val*". This subdivision was performed in order to assess whether the found QSPRs had any predictive ability for estimating the concentrations on the independent test set of "fresh" compounds, that were not involved during the model fitting using the 'training set' compounds, denoted as "*train*". The molecules of both the training and test series were selected, previously to the model search, so that both sets shared similar qualitative structure–property features.

The robustness of the structure–property equations established in this study was checked by the so-called *y*-randomization (Wold, Eriksson, & Clementi, 1995, chap. 5) in order to demonstrate that these equations did not result from happenstance. The *y*-randomization technique consisted on scrambling the concentration values in such a way that they did not account for the respective amino acids. Upon analyzing 1000 cases of *y*-randomization for each developed QSPR, the smallest standard deviation obtained for the model by using this procedure (S_{rand}) was compared to that found when considering the true calibration (S). If $S_{rand} > S$, then it was expected that the correlations found were not fortuitous, and resulted in real structure–property relationships.

Furthermore, the Kubinyi function (*FIT*) (Kubinyi, 1994) had been proposed as a statistical parameter that closely related to the Fisher ratio (F), but avoided the main disadvantage of the latter that was too sensitive to changes in small d values, and poorly sensitive to changes in large d values. The *FIT* criterion had a low sensitivity to changes in small d values and a substantially increasing sensitivity for large d values. The higher the *FIT* was, the better was the quality of the linear regression equation.

3. Results and discussion

Merlot red and Torrontés white argentine wines were obtained at a pilot scale as indicated in Materials and methods. Data of the

Table 1
Data taken from wine fermentation tanks. Numerical values are presented according to increasing fermentation times (weeks: 0–5, 7, 11).

Tank		Grape variety: Merlot (Initial pH 4.15)	Grape variety: Torrontés (Initial pH 3.95)
1 (*S. bayanus*)	Sugar content (g L^{-1})	260, 100, 66, 50, 50, -, -, -	240, 140, 0, 0, 0, -, -, -
	Sulfite (ppm)	added: 20 g, 46.08, 53.76, 64.00, -, 38.40, 53.76, -	added: 20 g, 12.80, 53.76, -, -, 67.84, 52.48, -
	Reducing sugars (%)	-, 6.20, 5.20, -, -, -, 5.00, -	-, < 0.1, < 0.1, -, -, -, < 0.1, -
	Total acidity[a] (g L^{-1})	1.50, 3.75 [2.25][b], -, -, -, -, -, -	3.00, 5.55 [4.00][b], -, -, -, -, -, -
	Alcohol (mL L^{-1})	0, 100, 120, 130, 130, -, -, 13 % v/v	0, 62, 150, 150, 150, -, -, 15% v/v
2 (*S. ellipsoides*)	Sugar content (g L^{-1})	260, 98, 70, 45, 35, -, -, -	240, 140, 0, 0, 0, -, -, -
	Sulfite (ppm)	added: 40 g, 30.72, 25.60, 26.88, -, 25.60, 20.48, -	added: 40 g, 32.00, 46.08, -, -, 122.88, 121.60, -
	Reducing sugars (%)	-, 7.00, 6.60, -, -, -, 3.56, -	-, < 0.1, < 0.1, -, -, -, < 0.1, -
	Total acidity[a] (g L^{-1})	1.50, 6.00 [3.90][b], -, -, -, -, -, -	3.00, 6.45 [3.08][b], -, -, -, -, -, -
	Alcohol (mL L^{-1})	0, 101, 130, 145, 150, -, -, 15% v/v	0, 62, 150, 150, 150, -, -, 15% v/v
3 (Native flora)	Sugar content (g L^{-1})	260, 200, 160, 83, 40, -, -, -	240, 140, 70, 0, 0, -, -, -
	Sulfite (ppm)	added: 100 g, 87.04, 53.76, 53.76, -, 52.48, 64.00, -	added: 100 g, 39.68, 98.56, 11.52, -, 23.04, 16.64, -
	Reducing sugars (%)	-, 16.00, 5.50, -, -, -, 4.00, -	-, 7.00, < 0.1, -, -, - <0.1, -
	Total acidity[a] (g L^{-1})	1.50, 3.00 [1.65][b], -, -, -, -, -, -	3.00, 6.90 [4.00][b], -, -, -, -, -, -
	Alcohol (mL L^{-1})	0, 37, 62, 110, 140, -, -, 14% v/v	0, 62, 105, 150, 150, -, -, 15% v/v

[a] As tartaric acid content.
[b] Volatile acidity as tartaric acid content (g L^{-1}).

initial processing of the must are shown in Table 1 (week: 0). Results about reducing sugars, free sulfite, and acidity obtained at the end of the first week of fermentation are shown in Table 1 (week: 1).

At the second week of fermentation, 40 g of sodium sulfite were added to tanks 4 and 5 (in order to reserve it since fermentation was finished, and the blockage of the secondary fermentation was intended, e.g., malolactic fermentation, as well as to avoid fungal growth.

Free sulfite was controlled after the third and fifth weeks of fermentation. Furthermore, free sulfite and reducing sugars were recorded after seven weeks of wine fermentation, and alcohol (% v/v) was measured at the eleventh week of wine fermentation, as well as the variation of alcohol and sugar during winemaking. All these data are shown in Table 1.

Experimental amino acid concentration values, expressed as a concentration range in ppm, for Merlot and Torrontés wine samples ($n = 5$ each) at different processing times are shown in Tables 1S and 2S, respectively.

3.1. Concentration of the main amino acids as a function of time for each type of wine

Proline and the main amino acid concentrations (isoleucine, arginine, threonine, cystine, alanine), and total proteins were measured during the fermentation time for both Merlot and Torrontés wines, taking into account the three types of yeasts: *S. cerevisiae* var. *bayanus*, *S. cerevisiae* var. *ellipsoides* and native flora.

According to experimental results, proline was the main amino acid in both wines. Unlike most amino acids in must, proline was not taken up by yeast because it could not be metabolized aerobically. Only when low concentrations of amino acids were available in the must, then the yeast would use proline. Anyway, there was an initial consumption, then reaching a plateau.

If there were very few amino acids in the must, during the aerobic phase the yeast could obtain them from glycolysis *via* pyruvate, which reversibly produced ketoacids and then, amino acids. This is an extreme possibility that the yeast can perform as it is being actively reproduced.

The behavior of proline in Torrontés white wine was opposite to that in Merlot red wine. In Merlot red wine proline values fell before reaching a plateau. In white wine the values of proline first increased, and then reached a plateau. Red wine showed a more logical behavior than white wine, in which from a high initial value proline decreased and then, due to the fermentation process, it reached a plateau. It appeared that what happened in Torrontés wine was a consequence of the added sulfite, and also of the partial lysis of the native flora population (Giraudo et al., 2006).

Alanine was the second amino acid in quantitative importance, since this amino acid belongs to the analytes that are absorbed by the yeast after a period of latency (Jackson, 2008).

Isoleucine and arginine were the amino acids that appeared in the third place in both wines, followed by threonine in white wine, and cystine in red wine, all belonging to those amino acids that are slowly absorbed by yeast.

The source of amino acids in the must was provided by the polypeptides produced during the proteolysis of the proteins initially present in the grapes. *Vitis vinifera* contains about 500–1500 ppm of protein. The residual protein in wine is ca. 50–200 ppm (Mazzei & Pachulu, 1995). In the white wine protein values were within the values given by the National Wine Institute of Argentina, while those in red wine were about 10 times higher, probably due to the extra soil fertilization, and the vineyard's age (data provided by the grapes' producer).

Reported free amino acids in wines, determined by several chromatographic techniques, were present in the following concentrations: proline: 700–1500 ppm; alanine: 70–150 ppm; arginine: 20–150 ppm; isoleucine: 10–80 ppm; cystine: 10–100 ppm and threonine: 20–60 ppm (Elfakir, 2005). The results presented herein were in agreement with these values, except for proline, whose value was 10 times higher in red musts (explained by the higher experimental protein percentage) and nearly 8 times higher in white musts. This seemed to be due to the sensitivity of the method for proline quantitation (fluorescence). Any method used showed so high sensitivity for proline, e.g., at the level of femtomols for amino acids, including methodologies that did not quantify it, such as the OPA-FMOC approach (Elfakir, 2005; Pomilio, 1994; Pomilio, 2004).

Amino acid concentration profiles were suitable fingerprints for wine quality. Climate, fertilized soil, geographical location, age of the vines, the role of amino acids as precursors of biogenic amines, etc., were taken into account. The presence/absence of these analytes would also determine wine flavor due to the transformations of the amino acids, thus yielding very particular higher alcohols, aldehydes, esters, diketonic acids, etc.

The content of free amino acids in musts and grapes showed to strongly depend on the grape variety, and the type of yeast, e.g., the native flora consumed less proline than the others.

Table 2
Statistical results for linear QSPR models for Merlot red wine.

Week	*R3u*	*R4e*	Intercept	*R*	*S*	*FIT*	R_{loo}	S_{loo}	S_{rand}	$o(2S)$[b]
0	2.644(±0.3)[a]	−1.690(±0.4)	−0.025(±1)	0.961	0.24	6.45	0.922	0.35	0.25	0
Native flora										
1	2.337(±0.4)	−1.770(±0.4)	0.710(±1)	0.947	0.26	4.64	0.918	0.33	0.28	0
2	2.385(±0.2)	−1.494(±0.3)	−0.0234(±0.7)	0.973	0.18	9.55	0.944	0.27	0.20	0*
3	2.628(±0.2)	−1.344(±0.2)	−0.854(±0.5)	0.988	0.12	22.74	0.975	0.19	0.22	0
4	2.643(±0.2)	−1.405(±0.2)	−0.693(±0.5)	0.988	0.12	22.10	0.971	0.22	0.19	0
Saccharomyces cerevisiae var. ellipsoides										
1	2.426(±0.3)	−1.618(±0.4)	0.268(±1)	0.958	0.23	6.00	0.936	0.28	0.23	0
2	2.444(±0.2)	−1.600(±0.2)	0.0416(±0.7)	0.980	0.16	12.75	0.956	0.25	0.20	0
3	2.467(±0.2)	−1.445(±0.2)	−0.241(±0.5)	0.988	0.12	21.18	0.976	0.18	0.16	0
4	2.606(±0.2)	−1.272(±0.2)	−0.782(±0.6)	0.985	0.14	17.00	0.977	0.17	0.20	0
Saccharomyces cerevisiae var. bayanus										
1	2.341(±0.3)	−1.797(±0.4)	0.745(±1)	0.957	0.23	5.82	0.934	0.29	0.24	0
2	2.532(±0.2)	−1.564(±0.2)	−0.186(±0.6)	0.985	0.14	17.25	0.967	0.22	0.17	0
3	2.213(±0.2)	−1.504(±0.3)	0.376(±0.7)	0.973	0.17	9.41	0.908	0.35	0.17	0*
4	2.678(±0.2)	−1.547(±0.2)	−0.543(±0.6)	0.982	0.16	14.44	0.967	0.22	0.20	0

[a] Regression coefficient and standard deviation (between parentheses).
[b] Number of outliers exceeding 2.*S* with exception of *, which exceeds 2.5.*S*.

3.2. Quantitative structure–property relationship (QSPR) analysis of amino acid profile in Merlot red and Torrontés white argentine wines

General linear regression equations were established on the tested amino acid concentration profiles in order to achieve predictive values on new fresh data. This predictive ability of each mapped QSPR equation was tested carrying out an internal validation (Leave-One-Out Cross Validation Technique), and also by an external validation (using some molecules as a test set, but not for data adjustment). Furthermore, all models designed in this work followed the "Rule of Thumb" (Tute, 1990), which stated that at least five or six data points should be present for each fitting parameter.

The main statistical results found for Merlot and Torrontés wine aminograms, respectively, are shown in Tables 2 and 3. Decision criteria that were simultaneously analyzed for determining the model's size for the training sets investigated, such as the determination of the optimal *d* to be included in each QSPR, were the following: (a) the lowest value for the *S* parameter; (b) the lowest S_{loo} value; (c) the highest *FIT* parameter; (d) the lowest number of outlier amino acids exceeding 2.*S* or 2.5.*S*; and (e) the lowest value for the maximal inter-correlation between descriptors in the model. Therefore, in every developed QSPR the linear regression model involved only two molecular descriptors. A brief description of the molecular descriptors involved in the established QSPR is presented in Table 3S, while Table 4S includes the numerical values for such descriptors.

For the case of Merlot based models, 11 training set amino acids were always used, and isoleucine was left as part of the test set. In Torrontés based models, 12 training amino acids were used, except for the fourth week, for which 11 training compounds were used; phenylalanine was used in the test set. Figs. 1S and 2S plotted the predicted $\log_{10}$ concentration range (ppm) *vs.* experimental $\log_{10}$ concentration range (ppm) for both wines, showing that the proposed models fitted a straight line for both the training and test set data. The predictions for all QSPR models were included in Supplementary Tables 5S and 6S.

All the predictive linear QSPR models shown in Tables 2 and 3 were able to capture the essential structural features of the amino acids related to their concentration in wine, *e.g.*, Torrontés white

Table 3
Statistical results for linear QSPR models for Torrontés white wine.

Week	*MATS3p*	*dip*	*Intercept*	*R*	*S*	*FIT*	R_{loo}	S_{loo}	S_{rand}	$o(2S)$[b]
0	−3.276(±0.9)[a]	0.907(±0.1)	−0.339(±0.4)	0.906	0.26	2.57	0.795	0.39	0.28	0
Native flora										
Week	*RDF030 m*	*R3u*	Intercept	*R*	*S*	*FIT*	R_{loo}	S_{loo}	S_{rand}	$o(2S)$
1	−0.186(±0.07)	3.053(±0.4)	−3.298(±0.7)	0.924	0.29	3.28	0.645	0.60	0.31	0
2	−0.198(±0.06)	3.194(±0.4)	−3.591(±0.6)	0.949	0.25	5.11	0.857	0.42	0.26	0
3	−0.207(±0.07)	3.269(±0.4)	−3.817(±0.7)	0.935	0.29	3.89	0.750	0.55	0.29	0
4	−0.223(±0.07)	2.960(±0.4)	−3.037(±0.7)	0.930	0.28	3.42	0.590	0.64	0.29	0
Saccharomyces cerevisiae var. ellipsoides										
1	−0.258(±0.07)	2.463(±0.4)	−2.046(±0.7)	0.917	0.26	2.95	0.810	0.40	0.27	0
2	−0.199(±0.06)	2.889(±0.3)	−2.930(±0.6)	0.942	0.24	4.42	0.626	0.57	0.26	0
3	−0.198(±0.06)	2.960(±0.3)	−3.095(±0.5)	0.957	0.21	6.17	0.774	0.47	0.25	0
4	−0.196(±0.05)	2.845(±0.3)	−2.834(±0.5)	0.959	0.20	6.06	0.821	0.43	0.21	0*
Saccharomyces cerevisiae var. bayanus										
1	−0.179(±0.06)	2.919(±0.3)	−3.108(±0.6)	0.945	0.24	4.69	0.751	0.49	0.25	0*
2	−0.179(±0.07)	3.080(±0.4)	−3.505(±0.8)	0.924	0.30	3.29	0.722	0.54	0.31	0
3	−0.203(±0.06)	3.051(±0.3)	−3.287(±0.6)	0.953	0.23	5.61	0.763	0.50	0.24	0
4	−0.205(±0.07)	2.839(±0.4)	−2.875(±0.7)	0.936	0.26	3.78	0.620	0.59	0.26	0

[a] Regression coefficient and standard deviation (between parentheses).
[b] Number of outliers exceeding 2.*S* with exception of *, which exceeds 2.5.*S*.

Table 4
Comparison of specificity for developed QSPR models on native flora wine aminograms. The best model is given in bold.

	R	*S*	*R*	*S*
Merlot	**0.961**	**0.24**	0.620	0.67
Torrontés	0.720	0.43	**0.906**	**0.26**

and Merlot red wines. It is worthwhile to mention that these QSPRs involved a combination of 2D- and 3D-type molecular descriptors in order to achieve the best predictions for the aminograms. For the case of Merlot data, the best QSPR found involved *R*3*u* and *R*4*e* 3D-descriptors. The different quantification of native flora and yeast strains profiles was achieved by using different regression coefficients in the linear model. This is a consequence of observing similar numerical variations for the numerical data. In the case of Torrontés data, the zero time considered *MATS*3*p* (2D) and *dip* (3D) descriptors, while the rest of the weeks were characterized by *RDF*030*m* and *R*3*u* 3D-descriptors.

In addition, an alternative validation of the designed QSPR models of this work was carried out. It was corroborated what happened when these structure–property relationships were exposed to independent data, e.g., to predict Merlot (native flora) free amino acid profile through molecular descriptors from the Torrontés (native flora) model and *vice versa*. The results found are shown in Table 4. According to the *R* and *S* parameters, it was observed that the descriptors were specific for the aminograms for which they were calibrated, since the statistical parameters deteriorated the models when applied to other data. Therefore, the herein established QSPR were able to discriminate among the different kind of profiles of these two wine varietals.

Finally, it was concluded that these qualitative and quantitative amino acid parameters could be taken into account as an index of genuinity and/or fingerprint of the produced wines.

4. Conclusions

Suitable experimental conditions for the HPLC analysis of Torrontés and Merlot wines had been obtained in our laboratories. According to our findings, aminograms could be used as fingerprints to characterize each varietal wine. This was a contribution for setting the seal of quality "*Alimentos Argentinos. Una elección natural*", which means "Argentine Food. A natural choice", e.g., Resolution N° 392/05 of the Secretariat of Agriculture, Livestock, Fishing and Food (SAGPyA) of Argentina, (Nimo, 2005) as well as DOC (*denominación de origen controlado*; certified brand denomination) to better position these argentine wines. The aim was to promote and safeguard the authenticity and origin of food, and also to incorporate distinct food attributes.

QSPR analysis of the experimental wine data could be considered as a specific fingerprint for adulteration, if the technological winemaking process is standard and known. This was a new application of the QSPR theory as, to our knowledge, only few studies had been devoted to the quantitation of amino acid profiles in food. A practical application of these QSPRs consisted on estimating some structures, of which experimental values were missing. Furthermore, any wine sample belonging to Torrontés and Merlot varietals would fit the corresponding developed model, thus demonstrating its authenticity.

Acknowledgements

Special thanks are given to the National Research Council of Argentina [Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina], Universidad Nacional de La Plata, Universidad Nacional de Lanús, and Universidad de Buenos Aires (Argentina) for financial support; and Ministerio de Ciencia y Técnica (Mincyt, Argentina) for electronic bibliography facilities. MAG is Chairman of Food Science and Food Technology of Universidad Nacional de Lanús (Argentina). ABP, EAC and PRD are Research Members of CONICET. None of the authors had any personal or financial conflict of interest.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.foodchem.2013.02.064.

References

Ajtony, Z., Szoboszlai, N., Suskó, E. K., Mezei, P., György, K., & Bencs, L. (2008). Direct sample introduction of wines in graphite furnace atomic absorption spectrometry for the simultaneous determination of arsenic, cadmium, copper and lead content. *Talanta, 76*(3), 627–634.

Arvanitoyannis, I. (2003). Part 2: Authenticating and tracing particular foods. 20. Wine authenticity. In M. Lees (Ed.), *Food authenticity and traceability*. 1-85573-526-1 (pp. 426–458). Abington, Cambridge: Woodhead Publishing Limited.

Atkins, S. (2009). Sauce. Argentine wines. *Telegraph*, 29 Apr. <www.telegraph.co.uk>.

Duchowicz, P. R., Castro, E. A., Fernández, F. M., & González, M. P. (2005). A new search algorithm of QSPR/QSAR theories: Normal boiling points of some organic molecules. *Chemical Physics Letters, 412*, 376–380.

Duchowicz, P. R., Giraudo, M. A., Castro, E. A., & Pomilio, A. B. (2011). Quantitative structure-property relationship analyses of aminograms in food: Hard cheeses. *Chemometrics and Intelligent Laboratory Systems, 107*, 384–390.

Duchowicz, P. R., Mercader, A. G., Fernández, F. M., & Castro, E. A. (2008a). Prediction of aqueous toxicity for heterogeneous phenol derivatives by QSAR. *Chemometrics and Intelligent Laboratory Systems, 90*(2), 97–107.

Duchowicz, P. R., Talevi, A., Bruno-Blanch, L. E., & Castro, E. A. (2008b). New QSPR study for the prediction of aqueous solubility of drug-like compounds. *Bioorganic & Medicinal Chemistry, 16*(17), 7944–7955.

Elfakir, C. (2005). 1.2. HPLC. In I. Molnár-Perl (Ed.), *Quantitation of amino acids and amines by chromatography: Methods and protocols. Journal of chromatography library* (70. 0-444 52050-3, pp. 119–267). Amsterdam, Boston, Heidelberg: Elsevier.

Garde-Cerdán, T., Lorenzo, C., Lara, J. F., Pardo, F., Ancín-Azpilicueta, C., & Salinas, M. R. (2009). Study of the evolution of nitrogen compounds during grape ripening. Application to differentiate grape varieties and cultivated systems. *Journal of Agricultural and Food Chemistry, 57*(6), 2410–2419.

Giraudo, M., Vicente, F., Sánchez Tuero, H., Pavesi, R., Markowski, I., Guirin, G., et al. (2006). Estudio de la vinificación en Argentina a partir de los aminoácidos libres en la producción de vino tinto (Merlot) y blanco (Torrontés) utilizando diferentes levaduras (nativa y cultivada). *Revista Enólogos (Spain), 44*(6), 24–29.

Golbraikh, A., & Tropsha, A. (2002). Beware of q^2! *Journal of Molecular Graphics and Modelling, 20*(4), 269–276.

Hansch, C., & Leo, A. (1995). *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*. Washington, DC: American, Chemical Society [book review].

Jackson, R. S. (2008). *Wine Science: Principles and Applications* (3rd ed.), Technology & Engineering. San Diego: Academic Press. ISBN 0-123-73646-3.

Katritzky, A. R., Lobanov, V. S., & Karelson, M. (1995). QSPR: The correlation and quantitative prediction of chemical and physical properties from structure. *Chemical Society Reviews, 24*(4), 279–287.

Košir, I. J., & Kidrič, J. (2002). Use of modern nuclear magnetic resonance spectroscopy in wine analysis: Determination of minor compounds. *Analytica Chimica Acta, 458*(1), 77–84.

Kubinyi, H. (1994). Variable selection in QSAR studies. II. A highly efficient combination of systematic search and evolution. *Quantitative Structure-Activity Relationships, 13*(4), 393–401.

Mazzei, M. A., & Pachulu, M. del R. (Eds.) (1995). *Tabla de Composición Química de los Alimentos*, CENEXA (Centro de Endocrinología Experimental Aplicada, Universidad Nacional de La Plata – CONICET) FEIDEN (Fundación para la promoción de la Educación y la Investigación en Diabetes y Enfermedades de Nutrición), Buenos Aires, Argentina.

Mitchell, M. (1998). *An introduction to genetic algorithms*. 0-262-63185-7. Cambridge, MA: The MIT Press.

Molnár-Perl, I. (2005). *Quantitation of amino acids and amines by chromatography: Methods and protocols. Journal of Chromatography Library* (Vol. 70). Amsterdam, Boston, Heidelberg: Elsevier [ISBN 0-444-52050-3].

Morosi, F. (2008). *World Wine Market*. In: *International Business and Trade*, May 28. <http://bizcovering.com/international-business-and-trade/world-wine-market/>.

Nimo, M. (2005). *A Quality Seal for positioning Argentine Food in the World*. Ministry of Agriculture, Livestock, Fishing and Food of Argentina. <http://www.alimentosargentinos.gov.ar/programa_calidad/diferenciacion/gacetilla_Argentine_Food.htm>.

Organisation Internationale de la Vigne et du Vin (OIV) (2009). *State of Vitiviniculture World Report*, 2002–2008. *State of Conditions Reports*, 2002–2008. *State of Vitiviniculture World Report*, 2009. OIV, 18 Rue d'Aguesseau, 75008 Paris, France, March 2009. <www.oiv.int>.

Pomilio, A. B. (1994). Separación de aminoácidos mediante cromatografía líquida de alta resolución. *Acta Bioquímica Clínica Latinoamericana, 28*(4), 527–554 [ISSN 0325-2957].

Pomilio, A. B. (2004). *Actualización en cromatografía líquida de alta resolución. Aplicaciones clínicas y medioambientales*. 987-98643-7-9. Idecefyn: Serie Química. Buenos Aires.

Pomilio, A. B., Duchowicz, P. R., Giraudo, M. A., & Castro, E. A. (2010b). Amino acid profiles and quantitative structure-property relationships for malts and beers. *Food Research International, 43*(4), 965–971.

Pomilio, A. B., Giraudo, M. A., Duchowicz, P. R., & Castro, E. A. (2010a). QSPR analyses for aminograms in food: Citrus juices and concentrates. *Food Chemistry, 123*(3), 917–927.

Soufleros, E. H., Bouloumpasi, E., Tsarchopoulos, C., & Biliaderis, C. G. (2003). Primary amino acid profiles of Greek white wines and their use in classification according to variety, origin and vintage. *Food Chemistry, 80*(2), 261–273.

Todeschini, R., Ballabio, D., Consonni, V., Mauri, A., & Pavan, M. (2007). CAIMAN (classification and influence matrix analysis): A new approach to the classification based on leverage-scaled functions. *Chemometrics and Intelligent Laboratory Systems, 87*(1), 3–17.

Todeschini, R., & Consonni, V. (2000). *Handbook of molecular descriptors. Book series: Methods and principles in medicinal chemistry*. Weinheim: Wiley-VCH.

Tute, M. S. (1990). History and objectives of quantitative drug design. In C. Hansch, P. G. Sammes, & J. B. Taylor (Eds.), *Comprehensive medicinal chemistry. Quantitative drug design* (4, pp. 1–31). Oxford: Pergamon Press.

Wold, S., Eriksson, L., & Clementi, S. (1995). Statistical validation of QSAR results. In H. van de Waterbeemd (Ed.), *Chemometric methods in molecular design* (pp. 309–338). Weinheim.

Zalacain, A., Marín, J., Alonso, G. L., & Salinas, M. R. (2007). Analysis of wine primary aroma compounds by stir bar sorptive extraction. *Talanta, 71*(4), 1610–1615.