Food Chemistry 123 (2010) 917–927

Contents lists available at ScienceDirect

Food Chemistry

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

Analytical Methods

QSPR analyses for aminograms in food: Citrus juices and concentrates

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

[a] PRALIB (CONICET, UBA), Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, Junín 956, C1113AAD Buenos Aires, 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] Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA (Universidad Nacionald de La Plata, CCT La Plata-CONICET), Diag. 113 y 64, C.C. 16, Suc. 4, B1900 La Plata, Argentina

ARTICLE INFO

Article history:
Received 22 April 2009
Received in revised form 22 March 2010
Accepted 29 April 2010

Keywords:
QSPR theory
Amino acids
Citrus juices and concentrates
Navel and Valencia oranges
Eureka lemon
Aminograms
Multi-variable linear regression
Replacement method

ABSTRACT

Dragon theoretical descriptors were derived for a set of optimised amino acid structures, with the purpose of establishing quantitative structure–property relationship (QSPR) models to predict aminograms for 100% natural fresh juices and concentrates of Navel and Valencia oranges, and Eureka lemon. We used the statistical replacement method technique for designing the best multi-parametric linear regression models, which included structural features selected from a pool containing 1497 constitutional, topological, geometrical, or electronic types of molecular descriptors. The prediction results achieved in this work were in most cases in good agreement with experimental amino acid profiles obtained in our laboratories by a validated HPLC procedure, thus demonstrating the predictive power of the designed QSPR. The developed approach is of practical value, especially when it is not possible to assign the analyte concentration with an accurate degree of certainty.

1. Introduction

Natural citrus juices and concentrates are widely marketed for national and international consumption. Citrus products are in high demand in Europe, USA, and a few other countries, either for direct marketing or as raw material for dairy products (yoghurts, desserts), ice creams, soft drinks, beverages.

The profile of free amino acids and ammonia (aminogram) of 100% natural lemon and orange fresh juices and concentrates from an Argentine juice company has been previously reported from our laboratory (Giraudo et al., 2004). The amino acids were derivatised with 6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate (AQC), and further analysed by high performance liquid chromatography (HPLC) with fluorescence and UV detection. The entire production process and issues of authenticity and quality of food were also reported.

Comparison of the amino acid concentrations, for each citrus concentrate with respect to the natural fresh juice results, showed that the concentration process did not significantly affect the nature and composition of the aminogram, except for alanine, which increased significantly in lemons and decreased in oranges, as well as for proline and arginine, which increased in oranges and decreased in lemons (Giraudo et al., 2004).

An accurate determination of an aminogram, *e.g.*, amino acid concentration profile, is of crucial importance in the food industry. It is thus feasible to find the characteristic fingerprint of food under investigation, a fact that enables its unequivocal identification and characterisation. Furthermore, knowledge of the concentration profile behaviour, for a given food, is essential for manufacture and control processes, allowing further chemical manipulation of food with the main purpose of achieving much better food quality. Therefore, one would be able to design much cheaper food in better yields and with intrinsic characteristics, such as colour, size, enhanced protein composition, health properties.

In recent decades, scientific communities have directed efforts to perform new experiments in a more rational way by systematisation of the available experimental information, this being possible through development of mathematical models derived from the field of computational chemistry. An alternative way to overcome the absence of experimental measurements is based on the ability to formulate quantitative structure–property relationships (QSPR) (Hansch & Leo, 1995; van de Waterbeemd, 1995). This approach has, as an ultimate goal, a model capable of

* Corresponding author. Tel./fax: +54 11 4814 3952.
E-mail addresses: pomilio@ffyb.uba.ar, abpomilio@sinectis.com.ar (A.B. Pomilio), mgiraudo@unla.edu.ar (M.A. Giraudo), pabloducho@gmail.com, duchow@inifta.unlp.edu.ar (P.R. Duchowicz), castro@quimica.unlp.edu.ar, eacast@ gmail.com (E.A. Castro).
[1] All authors contributed equally to this work.

0308-8146/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodchem.2010.04.082

estimating the properties of compounds by relying completely on the assumption that these relationships are consequences of the molecular structure.

Nowadays, the use of QSPR has become one of the most widely applied modern computational approaches. A chemical structure is translated into so-called molecular descriptors, describing different relevant features of the compounds, through a mathematical formula that has been proposed from chemical graph theory, information theory, quantum mechanics (Katritzky, Lobanov, & Karelson, 1995; Todeschini & Consonni, 2000; Trinajstic, 1992). More than one thousand descriptors are now available in the literature, and usually it has to be decided how to select those that characterise the property under consideration in the best possible manner.

Although a large number of structure–property relationships (SPR) have been previously reported for analysing the effects of amino acids on food concentration profiles, very few of them have included analyses based on the QSPR theory. A major reason for this is that, during recent decades, few experimental data on amino acid concentration profiles were available for building a quantitative model. Nowadays, different research groups, who continuously report new data, are surmounting this drawback.

The aim of this work was the design of predictive QSPR models, which could serve as suitable tools for estimating aminograms of various citrus juices and concentrates, whose experimental data were obtained from experiments carried out in our laboratories. Amino acids of 10 samples of Valencia and Navel orange concentrates, five Eureka lemon concentrates and seven fresh juices of each fruit, thus accounting for a total of 36 samples, were analysed by a validated HPLC approach in our laboratories.

In the present analysis, we explored a pool containing more than one thousand (1497) theory-based descriptors computed by means of the Dragon software (Milano Chemometrics and QSAR Research Group), and established QSPR for each natural citrus juice and concentrate under analysis by means of the replacement method technique (Duchowicz, Castro, Fernández, & González, 2005; Duchowicz, Mercader, Fernández, & Castro, 2008; Duchowicz, Talevi, Bruno-Blanch, & Castro, 2008). As it has not always been feasible to accurately estimate the amino acid concentration in food, both experimentalist's expertise and proper design of simple and general QSPR models would allow solutions to this problem.

2. Materials and methods

2.1. Materials

The following reagents and solvents were used: sodium acetate trihydrate, HPLC grade; triethylamine, *p.a.*; disodium EDTA, *p.a.*; 85% phosphoric acid, *p.a.*; acetonitrile HPLC grade (JTBaker, Interchemistry, Buenos Aires, Argentina). 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).

The AccQ·Tag kit contains an ampoule with 17 standards of amino acids, each with a concentration of 2.5 mM, except for cystine, which has a concentration of 1.25 mM.

2.2. Natural citrus juices and concentrates

Samples of 100% natural juices and (cloudy) concentrates of Valencia and Navel oranges, and Eureka lemon, as well as each variety of these citrus fruits, used in the juices' elaboration, were supplied by the Faculty of Food Science of Universidad Nacional de Entre Ríos, Concordia City (Province of Entre Ríos, north-eastern Argentina), after obtaining these juices and concentrates from a local juice company. Brix degrees of these fresh juices, as well as of concentrates, were measured with a Boeckel refractometer, so as to obtain the necessary dilution degree of the latter (Barrett, Somogyi, & Ramaswamy, 2004). Then, to a part of the concentrated lemon juice, 7 parts of water were added; for oranges, to one part of the orange concentrate 9.6 parts of water were added. Brix, as used in citrus processing, generally refers to the percent by weight of sugars in the citrus juice value obtained by a refractometer, corrected for temperature, to which is added a correction for citric acid (Barrett et al., 2004).

2.3. Analysis of free amino acids

Free amino acids present in the food matrices under study were determined by the AccQ·Tag technique of Waters Corporation, which proved to be the most suitable due to its high speed, sensitivity and resolution of chromatographic peaks. We used a Waters commercial kit provided by D'Amico Sistemas SA (Buenos Aires, Argentina) for precolumn derivatisation.

The internal standard, α-aminobutyric acid (AABA) and the corresponding borate buffer, were added, and further derivatised 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.

Ten samples of concentrated orange juice from the same company (5 of each species) and 5 of concentrated lemon juice were chromatographed. Also, 7 samples of each respective fruit (the same varieties) were chromatographed, and mean values of the amino acid content were obtained.

Once the chromatograph was optimised 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 derivatisation. If the automated method were to be used, variation coefficients would have been lower (Boogers, Plugge, Stokkermans, & Duchateau, 2008).

2.4. Instrumental 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 fluorimetric detector and HP 3396 II Integrator.

A C_{18} Hypersil column (200 mm length and 2.1 mm i.d.; 5 μm particle diameter; pore diameter 12 nm) was used. Another C_{18} column with other dimensions, 150 × 3.9 mm, was also used. In each case, a guard column of 3.9 × 20 mm was used.

The following ternary system was used as mobile phase: Solvent A: 140 mM sodium acetate, 17 mM triethanolamine, pH 5.05, containing 1 mM disodium EDTA. Solvent B: acetonitrile, HPLC grade. Solvent C: water, HPLC grade. The quality of the mobile phase was controlled by blind tests for UV and fluorescence detection, thus confirming the purity.

An alternative solvent, B, for high-pressure mixing systems was 60% acetonitrile. For chromatographic operation, the system was balanced with 100% Solvent A for 10 min before injecting the sample. Then, a ternary gradient elution (Millipore Corporation) was performed so that the complete separation lasted *ca.* 35 min, including column rebalance.

Flow rate was 0.33 ml/min. Detection was by UV at 248 nm. Fluorimetric: excitation wavelength = 250 nm and emission wave-

length = 395 nm. Oven temperature was 40 °C. Injection volume was 5 μl.

Retention times and order of elution were compared. A composition profile was established by calculating the relative percentage (to total concentration) of main amino acids.

2.5. Preparation of the internal standard at 2.5 mM concentration

α-Aminobutyric acid (6.45 mg) was added to 0.1 M HCl (25 ml). In the case of norleucine, 8.2 mg were weighed and dissolved in 0.1 M HCl (25 ml). The solutions were kept for 6 months at −20 °C.

2.6. Preparation of the calibration curve with internal standard

A solution of internal standard (40 μl) was mixed with the solution of standard amino acids (40 μl) and water HPLC grade (920 μl). The calibration standard sample, with the internal standard contained, 100 pmols/μl of each amino acid, except for cystine (50 pmols/μl, equivalent to 100 pmols/μl of cysteine).

2.7. Preparation of the calibration standard

(1) Diluted standard (10 μl) was placed in a 6 × 50 mm tube, using a micropipette, and the proper internal standard was added. (2) Borate buffer (70 μl) was added and mixed. (3) AQC reagent (20 μl) was added and quickly vortexed after the addition. (4) The tube content was transferred to a conical vial of the chromatograph, and was closed tightly with parafilm. (5) The vial was placed at 50 °C for 10 min. (6) An aliquot (5 μl) was injected into the chromatograph, accounting for 50 pmols of each amino acid, except for cystine which accounted for 25 pmols.

2.8. Method validation

2.8.1. Precision

Precision is the measure of how close data values are to each other for a number of measurements under the same analytical conditions. The components of precision, *e.g.*, repeatability and reproducibility, were determined according to ICH (1996) and US Pharmacopeia (USP, 2007) recommendations:

For repeatability (intra-day variation), five consecutive injections of 2.5 mM solutions of standard amino acids and a 2.5 mM of AABA internal standard were analysed within the same day. For reproducibility (inter-day variation), the reproducibility was determined by injecting (five times during five consecutive days) derivatised 2.5 mM solutions of amino acids and AABA on the same instrument.

Results were expressed as percent of coefficient of variation (%CV). The %CV values must be ≤5% for the standard and ≤10% for the samples.

2.8.2. Linearity

Several standard curves were prepared over a concentration range of 0.1–15 μM. HPLC analysis was performed under the chromatographic conditions mentioned above. Data from peak area versus amino acid concentration plots were treated by linear least square regression analysis. The standard curves were evaluated for intra-day and inter-day reproducibility. Each experiment was repeated in triplicate.

2.8.3. Accuracy

Accuracy is the measure of how close the experimental value is to the true value.

Recovery studies were performed for 2.5 mM samples of the most important amino acids found in citrus juices and concentrates, by the standard addition method, at concentrations ranging from 50 to 100, and 200% of a mean value of the linearity concentration range. The experiment was performed in triplicate. Recovery (%), and CV (%) were calculated for each amino acid.

2.8.4. Limit of detection (LOD) and limit of quantification (LOQ) values

Sensitivity is the minimum amount of analyte that can produce a significant result. Sensitivity is defined by the limit of detection (LOD), which accounts for the lowest concentration of amino acid that can be detected with a variation coefficient of up to 20%, but not necessarily be quantified in a sample (Nollet, 2004). It accounts for a signal to noise ratio in the range of 2–3. Therefore, solutions of standard amino acids with concentrations ranging from 0.1 to 15 μM were injected, and the LOD was determined for aspartic acid, glycine, proline and isoleucine, using a fluorescence and UV–Vis variable detector. The LOQs of the amino acids were also assessed in the 0.1–15 μM range.

2.8.5. Ruggedness and robustness

Technical data transfer from one laboratory to another requires a clear demonstration that the methodology can be successfully transferred. The USP (2007) defines ruggedness as the degree of reproducibility of test results obtained by the analysis of the same samples under a variety of normal test conditions, such as different laboratories, analysts, instruments, reagent lots, analysis days, elapsed assay times, and different assay temperatures. Therefore, different laboratories and analysts, as well as reagents from different lots and different manufacturers, were used, and the results were compared.

Robustness is defined by both the USP (2007) and ICH (1996) guidelines as a measure of the method capacity to remain unaffected by small but deliberate variations in method parameters, and provides an indication of its reliability during normal use. Therefore, small variations of the chromatographic conditions on the determination of amino acids were checked. Variations of the reproducibility of retention times were assessed. Length of the column, flow, oven temperature and preparation time-injection were the variables of the chromatographic runs that were monitored.

2.8.6. Stability tests

2.8.6.1. Stability of the derivatisation reagent. The stability of the derivatisation reagent, AQC, during analysis was determined by repeated analysis during the course of experimentation on the same day and also after storage of the freshly prepared derivatisation reagent for 1 week, 2 weeks and up to 2 months under refrigeration (4 °C). The occurrence of other fluorescent components was checked.

2.8.6.2. Stability of the derivatised amino acids. AQC reacted rapidly with primary and secondary amino acids (AA) to yield highly stable substituted urea adducts (AQC-AA), which showed a strong fluorescence at 395 nm. The stability of the AQC-AA derivatives was studied immediately after preparing the derivative, and after 5, 10, 20 and 30 days.

2.8.7. Specificity/selectivity

This is the property of the method to produce a measurable signal only due to an analyte, free of interferences from other components in the sample matrix, *e.g.*, degradation products, by-products and metabolites.

The specificity of the method was determined by injecting the 2.5 mM solution of amino acids, and checking peak homogeneity by HPLC with a diode-array detector (DAD) in an inter-laboratory exercise with the HPLC laboratories of the Career of Food Engineering, National University of Entre Ríos, Concordia City, Province of Entre Ríos, Argentina.

When studying specificity/selectivity, at least five reagent blanks were injected in order to observe whether any interfering peak appeared in the area of the amino acid peaks.

2.8.8. System suitability tests

The chromatographic systems used for analyses had to pass the system suitability limits before starting sample analysis. The suitability tests were applied to the chromatographic columns used, as specified by USP (2007), i.e., the efficiency, tailing (asymmetry of peaks), and repeatability of response (%CV) were determined for the 2.5 mM amino acid standard sample to assess the accuracy and precision of the developed HPLC system.

2.9. Statistical analysis

The statistical programme, NWA Quality Analyst release 6.1 (2007), was used throughout the validation method.

2.10. QSPR studies

2.10.1. Data set of amino acids

The experimental data of 100% natural fresh juices and concentrates of Navel and Valencia oranges and Eureka lemon modelled in the present study, were measured in our laboratory. The concentration of each amino acid in the aminogram was expressed as

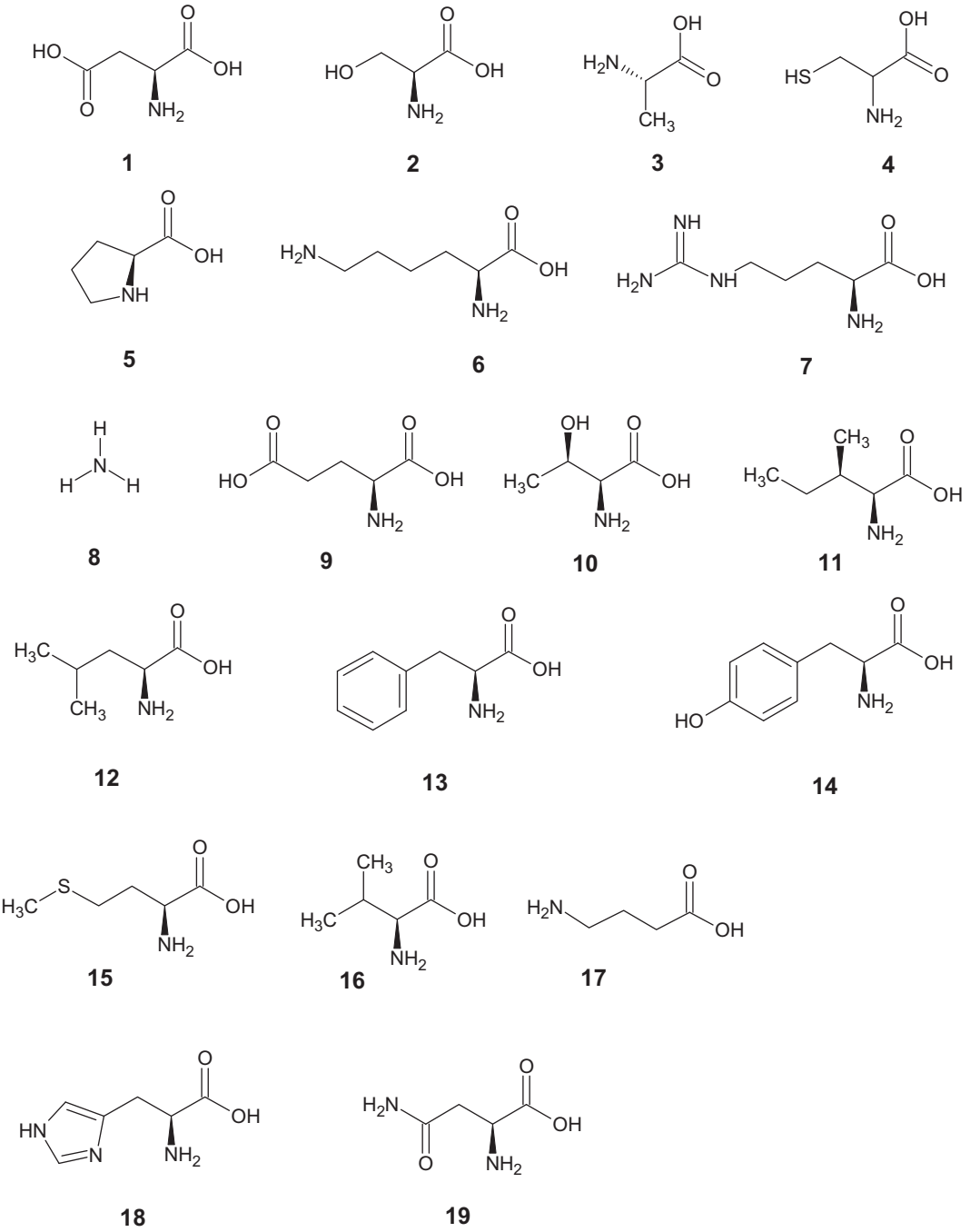

Fig. 1. Molecular structures of the 19 studied amino acids present in citrus juices and concentrates.

the average relative percentage (% rel), which was then converted into logarithm units for modelling purposes [$\log_{10}$(% rel)]. More than 80% of the amino acids appearing in orange juices accounted for aspartic acid, serine, alanine, cysteine, proline, lysine, arginine, ammonia, glutamic acid, and threonine. On the other hand, the amino acids that prevailed in lemon juices were, aspartic acid, arginine, proline, threonine, serine, glutamic acid, alanine, isoleucine, leucine, tyrosine, methionine, ammonia, phenylalanine, and valine. The molecular structures for the 19 amino acids analysed in this work are shown in Fig. 1.

2.10.2. Geometry optimisation and theoretical descriptors calculation

We first pre-optimised the structure of each amino acid with the Molecular Mechanics Force Field (MM+) procedure included in the Hyperchem 6.03 package (Hypercube). We chose the levogyre enantiomer because of its natural occurrence. After that, the resulting geometries were refined by the Semi-Empirical Method PM3 (Parametric Method-3) using the Polak-Ribière algorithm and a gradient norm limit of 0.01 kcal $Å^{-1}$.

We computed 1497 theoretical molecular descriptors, using Dragon software (Milano Chemometrics and QSAR Research Group) 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, we added to the pool, five quantum-chemical descriptors not provided by the Dragon programme: molecular dipole moments, total energies, homo–lumo energies, and homo–lumo gap ($\Delta_{\text{homo–lumo}}$). Total pool of descriptors explored consisted of 1233 theoretical variables.

2.10.3. Model search

We applied the Replacement Method (RM) (Duchowicz et al., 2005; Duchowicz, Mercader, et al., 2008; Duchowicz, Talevi, et al., 2008) as a molecular descriptors selection approach, an algorithm that has been proposed by our research group some years ago, that is an efficient optimisation tool which generates multiparametric 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 is quite close to that obtained by performing an exact (combinatorial) full search (FS) of molecular descriptors, although, of course, it requires much less computational work. The RM provides models with better statistical parameters than does the Forward Stepwise Regression procedure (Draper & Smith, 1998) and similar ones to the more elaborated Genetic Algorithms (Mitchell, 1998). We used the computer Matlab 5.0 system for all our calculations (The MathWorks Inc., 2004).

2.10.4. Model validation

The design of a properly validated model constitutes the most important step that should be born in mind for every QSPR analysis, in order to generate predictive models that involve general applicability and that are not limited to function only correlatively. The theoretical validation that was practised over each linear regression was based on the Leave-One-Out Cross Validation procedure (*loo*). Both parameters, R_{loo} and S_{loo}, the correlation coefficient and standard deviation of cross validation, respectively, measured the stability of the developed QSPR upon inclusion/exclusion of compounds. According to the specialised literature R_{loo} should be higher than 0.50 for obtaining a validated model (Golbraikh & Tropsha, 2002).

We also applied a rigorous and more realistic validation process that consisted in omitting from the complete molecular set (shown in Fig. 1) some amino acids which constituted the 'test set', denoted here as "*val*". The main purpose of performing such a splitting was to assess whether the found QSPR relationships had any predictive capability 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". We selected the molecules composing the training and test series as a previous step to the model search, and this was done in such a way that both sets shared similar qualitative structure–property characteristics for the compounds of the sets. The amino acids included in the test set were **4** (cysteine) and **11** (isoleucine).

In addition, with the purpose of demonstrating that the structure–property equations established in this study did not result by chance, we checked their robustness by means of the so-called *y*-randomization (Wold, Eriksson, & Clementi, 1995). This technique consisted of scrambling the concentration values in such a way that they did not correspond to the respective amino acids. After analysing 5,000,000 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 the one found when considering the true calibration (S). Therefore, if $S_{rand} > S$, then it would be expected that the correlations found were not fortuitous, and resulted in real structure–property relationships.

Finally, the Kubinyi function (*FIT*) (Kubinyi, 1994) has been proposed as a statistical parameter that was 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 greater the *FIT*, the better was the quality of the linear regression equation.

3. Results and discussion

3.1. Aminograms of citrus juices and concentrates

The aminograms of fresh juices and concentrates of Valencia and Navel oranges, and Eureka lemons were obtained by HPLC analysis, using AQC precolumn derivatisation, as detailed under Section 2. The concentrations of the amino acids, expressed as percentages, are shown in Table 1.

3.2. HPLC method validation

3.2.1. Precision

Precision was measured in accordance with ICH (1996) and USP (2007) recommendations. Intra-day variation was assessed by injecting samples on the same day at different time intervals, thus showing a very good repeatability within the same day with a %CV ranging from 2.1% to 5.9% (Table 2). The reproducibility of inter-day injections was higher, ranging from 6.1% to 11.6% (data not shown). For the internal standard, the values were higher, *e.g.*, repeatability was 5.7% (Table 2) and reproducibility was 34.2%.

3.2.2. Linearity

The linear regression data for the calibration plot were indicative of a good linear relationship over a wide concentration range. Therefore, a good linear relationship was obtained with regression coefficients in the range of $r = 0.970–0.990$ over the concentration range studied, as shown in Table 2.

Table 1
Amino acid concentrations in Citrus fresh juices and concentrates (relative percentage of total).

Amino acid	Navel orange			Valencia orange			Eureka lemon		
	Fresh juice, $n = 7$ (% range)	SD	Concentr., $n = 5$ (% range)	Fresh juice, $n = 7$ (% range)	SD	Concentr., $n = 5$ (% range)	Fresh juice, $n = 7$ (% range)	SD	Concentr., $n = 5$ (% range)
Aspartic acid	28.9–36.0	5.00	28.4–30.0	27.1–29.7	1.80	16.1–18.1	21.8–23.9	1.00	27.9
Serine	9.40–9.60	0.14	7.00–7.09	20.3–21.6	1.00	9.20–12.00	21.4–22.4	0.70	22.8
Alanine	7.90–10.0	1.50	12.6–13.3	10.3–10.6	0.21	16.30–17.00	2.20–2.50	0.21	4.80
Cystine	7.90–8.10	0.14	4.20	4.20–5.00	0.57	6.80	3.90–4.10	0.14	4.60
Proline	5.80–8.00	1.60	8.80–9.40	5.50–5.90	0.28	10.0–11.2	9.40–10.00	0.42	3.20
GABA	N.D.	N.D.	N.D.	N.D.	N.D.	N.D	N.D	N.D	N.D.
Lysine	6.20–6.90	0.50	6.40	5.90–6.10	0.14	6.00	0.40–0.60	0.14	0.70
Arginine	5.80–6.30	0.35	13.0–13.8	12.3–12.8	0.35	16.0–16.9	22.9–23.6	0.49	14.4
NH_3 (ammonia)	5.10–5.20	0.07	6.40–7.00	6.90–7.60	0.49	6.70–7.00	1.50–1.90	0.28	1.90
Glutamic acid	2.10–2.20	0.07	1.40–1.60	2.00–2.10	0.07	2.20–3.00	3.60–3.70	0.07	2.10
Threonine	2.20–3.00	0.57	2.80–3.30	2.70–2.90	0.14	3.30–4.40	7.10–7.30	0.14	10.9
Isoleucine	1.40–1.70	0.21	1.20	1.10–1.30	0.14	1.40	2.50–2.90	0.28	1.90
Leucine	1.30–1.50	0.14	1.00	0.90–1.00	0.07	1.20	2.50–2.70	0.14	1.90
Phenylalanine	1.10–1.40	0.21	0.80	0.80–0.90	0.07	0.80	1.50–1.80	0.21	1.60
Tyrosine	0.80–1.10	0.21	0.70	0.80–1.00	0.14	1.10	2.10–2.30	0.14	1.70
Methionine	0.70–0.90	0.11	0.80	1.50–1.60	0.07	0.90	2.40–2.50	0.07	0.70
Valine	0.50–0.70	0.11	0.40	0.30–0.50	0.14	0.60	1.00–1.20	0.14	1.30
Histidine	n.d.	n.d.	1.70–2.10	n.d.	n.d.	1.50–2.10	n.d.	n.d.	n.d.
Asparagine	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.

Concentr., concentrate; SD, standard deviation; n.d., non-detected; N.D., non-determined.

Table 2
Precision data and regression coefficients of the amino acids and internal standard (AABA).

Amino acid	Mean experimental values ($n = 7$) (mM)	%CV	Regression coefficient (r)
Aspartic acid	0.62	4.60	0.972
Serine	0.11	5.50	0.980
Glutamic acid	1.76	4.30	0.972
Glycine	3.50	4.20	0.981
Histidine	2.50	3.80	0.980
NH_3 (ammonia)	1.43	4.70	0.990
Arginine	0.39	5.60	0.975
Threonine	0.51	4.90	0.974
Alanine	1.20	4.10	0.971
Proline	1.35	2.60	0.980
AABA	2.18 (n = 15)	5.70	0.980
Cystine	0.18	2.10	0.986
Tyrosine	2.42	5.30	0.976
Valine	1.31	3.20	0.977
Methionine	1.90	3.80	0.970
Lysine	0.13	5.90	0.976
Isoleucine	2.00	3.90	0.970
Leucine	1.72	4.60	0.980
Phenylalanine	7.30	3.00	0.974

Abbreviations: CV: coefficient of variation. Mean %CV: 4.30.

3.2.3. Accuracy

Recovery studies were performed by the standard addition method for the most important amino acids found in citrus juices and concentrates, in order to assess the accuracy of the proposed method (Table 3). The percent recovery ranged from 85.5% to 99.8%, the mean recovery value being 91.8%. The values of recovery (%), and CV (%) indicate that the method is accurate.

Table 3
Percentage of the recovery of the amino acids under study.

Amino acid	% Recovery	%CV
Asp	93.1	5.6
Ser	85.5	11.0
Glu	99.8	1.3
Arg	93.5	10.1
Ala	88.7	9.1
Pro	92.2	8.3
Cyt	91.2	7.5
Lys	90.5	4.9

3.2.4. Sensitivity

The limit of detection (LOD) accounts for a signal to noise ratio in the range 2–3. After injecting solutions of standard amino acids with concentrations ranging from 0.1 to 13 μM, the LOD was determined for aspartic acid (311 fmols), glycine (293 fmols), proline (278 fmols), and isoleucine (47 fmols), using a fluorescence detector, while the UV–Vis variable detector accounted for pmols, *e.g.*, 47–780 pmols.

Data obtained herein demonstrated the importance of using a fluorescence detector in the analysis of amino acids because of the highest sensitivity achieved for the experimental data.

3.2.5. Limit of quantification (LOQ)

The LOQ for all amino acids under study was close to 0.1 μM, and the linearity extended to at least 15 μM, with a regression coefficient (r) higher than 0.970 (Table 2). Therefore, the method can be used for detection and quantification of amino acids over a very wide range of concentrations.

3.2.6. Ruggedness and robustness

The ruggedness indicates the degree of reliability of the method after deliberate changes in operating parameters, but always external to the analytical test method and test parameters. When small but deliberate internal variations in test parameters should be evaluated, to assess whether the analytical procedure remains unaffected, this is the so-called robustness of the method. Furthermore, ruggedness measures the lack of external influence on the test results, whereas robustness measures the lack of internal influences on the test results.

The analytical method proved to be rugged and robust; however, a well designed experiment should identify test conditions or specification limits that need to be closely controlled, and even test parameters that require further investigation and optimisation.

Results showed that the analytical procedure was not affected by changes of either analysts or instruments. There was no significant change in the retention time of amino acids when reagents

from different lots and different manufacturers were used. The values of regression coefficients (Table 2) indicated the ruggedness of the method.

Concerning the pH of derivatisation, the value should not be changed; otherwise derivatisation will not occur. By contrast, the pH of the mobile phase can be changed in order to improve the separation of amino acids. Retention times were reproducible, with variations within 0.01 min. The peaks obtained had a width of less than 0.1 cm.

A designed experiment is a simple matrix design. The Plancket–Burman designs are the most used in technical transfer of a validated analytical method, because they are suitable for studying the effect of n variables in $(n + 1)$ tests to check the ruggedness/robustness of a procedure. Modifications of column length (from 150 to 200 mm, keeping diameter 2.1 mm, and particle diameter 5 μm constant), flow (from 0.25 to 0.33 ml/min), oven temperature (from 30 to 40 °C), and preparation time-injection (from 20 min to 8 h) were studied. All of these variables were monitored, and both absolute and relative results were coincident. In conclusion, the present method proved to be rugged and robust.

3.2.7. Specificity

This is the property of the method to produce a measurable signal due only to an analyte, free of interferences from other components in the sample matrix, *e.g.*, degradation products, by-products, and metabolites. The specificity of the method was determined by exposing 2.5 mM solutions of amino acids to diode-array detector (DAD) scrutinity for peak homogenity through inter-laboratory work.

When injecting the derivatised amino acid standards, only the peaks of the pure highly fluorescent AQC-AA appeared, except for a small peak detected prior to the derivatised amino acids. This small peak accounts for 6-aminoquinoline (AMQ), which is one of the AQC hydrolysis products (Bosch, Alegría, & Farré, 2006; Cohen & Michaud, 1993). In quantitative studies, AQC should be in enough excess to drive the reaction to completion. This excess of AQC is hydrolysed within 1 min to the weakly fluorescent AMQ, *N*-hydroxysuccinimide, and carbon dioxide, which appeared in the chromatogram without interfering with the amino acid quantification. AMQ fluorescence response is <1% relative to an equivalent amount of amino acid derivative (Cohen & Michaud, 1993).

We obtained the retention time of each peak in a comparative manner with both fluorescence and UV detectors, the former being, as expected, much more sensitive and of higher selectivity, thus allowing limits of detection at the femtomole level. This is in agreement with previous reports on AQC-AA fluorescence detection being about 50–100 times more sensitive than UV detection (Boogers et al., 2008; Cohen, 2000; Oreiro-García, Vázquez-Illanes, Sierra-Paredes, & Sierra-Marcuño, 2005).

There was no degradation of amino acids, and no significant changes in peak area and retention time were observed throughout the running of the chromatograms.

3.2.8. Stability tests

3.2.8.1. Stability of the derivatisation reagent. If the response of 100 accounted for the freshly prepared derivatisation reagent (AQC), after a week the response was 99 at 4 °C, and after 2 weeks under the same conditions was 98. Two months later, a value slightly higher than 90 was obtained. It is noteworthy that no other fluorescent components or degradation products were at any time obtained in the chromatogram.

Therefore, the AQC solution can be stored, without any degradation, over the time interval studied.

3.2.8.2. Stability of the derivatised amino acids (AQC-AA). The fluorescent derivatives, AQC-AA, showed high stability, the response being *ca.* 100% within the first week of preparing the derivative. The loss of stability was not significant until 30 days, as shown in Table 4.

3.2.9. System suitability tests

The results of the system suitability tests assure the adequacy of the proposed HPLC method for routine analysis of amino acids in citrus fresh juices and concentrates. The test of adequacy was ap-

Table 4
Stability (%) of the derivatised amino acids under study.

Derivatised amino acid (AQC-AA)	Zero time	5 days	10 days	20 days	30 days
AQC-Asp	100	101	100	95	80
AQC-Ser	100	103	99	89	81
AQC-Arg	100	99	95	88	69
AQC-Pro	100	100	96	84	75

Table 5
Brief description for theoretical descriptors appearing in the established QSPR.

Descriptor	Category	Type	Brief description	Reference
BELe4	2D	BCUT	Lowest eigenvalue no. 4 of Burden matrix/weighted by atomic Sanderson electronegativities	Benigni, Passerini, Pino, and Giuliani (1999)
RDF050v	3D	Radial Distribution Function	Radial Distribution Function 5.0/ weighted by atomic van der Waals volumes	Consonni, Todeschini, Pavan, and Gramatica (2002)
BELv4	2D	BCUT	Lowest eigenvalue No. 4 of Burden matrix/weighted by atomic van der Waals volumes	Benigni et al. (1999)
Mor02e	3D	3D-MoRSE	3D-MoRSE signal 02 – weighted by atomic Sanderson electronegativities	Gasteiger et al. (1996)
MATS6p	2D	2D-Autocorrelations	Moran Autocorrelation-lag 6/weighted by atomic polarizabilities	Moreau and Broto (1980)
Mor32u	3D	3D-MoRSE	3D-MoRSE signal 3.2 – unweighted	Gasteiger et al. (1996)
BELv7	2D	BCUT	Lowest eigenvalue no. 7 of Burden matrix/weighted by atomic van der Waals volumes	Benigni et al. (1999)
IDDE	2D	Topological	Mean information content on the distance degree equality	Bonchev (1983)
Mor21u	3D	3D-MoRSE	3D-MoRSE signal 21 – unweighted	Gasteiger et al. (1996)
Me	0D	Constitutional	Mean atomic Sanderson electronegativity (scaled on Carbon atom)	Todeschini and Consonni (2000)
MATS8p	2D	2D-Autocorrelations	Moran Autocorrelation-lag 8/weighted by atomic polarizabilities	Moreau and Broto (1980)

plied to the chromatographic columns used, as specified by USP (2007); *e.g.*, efficiency, tailing (asymmetry of peaks), and repeatability of response (%CV) were determined for the amino acid standard sample. The %CV of seven consecutive injections, performed under the precision test was found to be in the range 2.1–5.9%. The tailing factor (t) reflected good peak symmetry. Good separation of the amino acids was obtained with a suitable resolution for quantification. In all cases, columns fulfilled the manufacturer's specifications.

3.3. QSPR studies

3.3.1. General

In each instance, general quantitative structure–property relationship (QSPR) equations were established on the tested 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). Eventually, some molecules of the training set were removed, thus improving the model's quality performance. Furthermore, all models designed in this work followed the "rule of thumb" (Tute, 1990), which stated that at least 5 or 6 data points should be present for each fitting parameter.

For every QSPR of this work, N was the number of amino acids, R the correlation coefficient, *res* the residual for a given amino acid (difference between the experimental and predicted concentration), and p was the significance of the model, *FIT* accounted for the Kubinyi function (Kubinyi, 1994), and S_{rand} was the model's

Table 6
Numerical values of theoretical descriptors involved in the various QSPRs designed in this study.

No.	Amino acid (*levo*)	*BELe*4	*RDF*050*v*	*BELv*4	*Mor*02*e*	*MATS*6*p*	*Mor*32*u*	*BELv*7	*IDDE*	*Mor*21*u*	*Me*	*MATS*8*p*
1	*S*-Aspartic acid	0.246	0.056	0.782	17.366	0.401	−0.076	0.126	2.725	−0.194	1.070	0.000
2	*S*-Serine	0.000	0.038	0.481	18.122	0.463	−0.070	0.000	2.522	−0.140	1.050	0.000
3	*S*-Alanine	0.000	0.012	0.410	16.954	0.000	−0.053	0.000	0.918	−0.252	1.030	0.000
4	*R*-Cysteine	0.000	0.334	0.427	15.568	0.457	−0.123	0.000	2.522	−0.293	1.040	0.000
5	*S*-Proline	0.032	0.002	0.518	20.115	0.565	−0.241	0.092	2.250	−0.358	1.020	0.000
6	*S*-Lysine	0.560	3.999	1.074	29.139	−0.087	−0.094	0.406	2.646	−0.829	1.010	0.349
7	*S*-Arginine	0.645	4.766	1.098	31.058	0.098	0.000	0.572	2.752	−0.313	1.020	−0.238
8	Ammonia	0.000	0.000	0.000	4.420	0.000	−0.042	0.000	0.000	0.214	1.000	0.000
9	*S*-Glutamic acid	0.481	1.372	0.993	20.889	0.082	−0.187	0.253	2.722	−0.443	1.050	0.538
10	Threonine	0.304	0.159	0.855	20.300	0.466	−0.196	0.071	1.750	−0.091	1.040	0.000
11	Isoleucine	0.541	0.605	1.069	19.642	0.326	−0.278	0.279	2.948	−0.526	1.000	0.000
12	Leucine	0.630	1.681	1.097	21.880	0.275	−0.326	0.226	2.725	−0.610	1.000	0.000
13	*S*-Phenylalanine	0.807	1.759	1.037	20.226	−0.131	−0.320	0.394	2.459	−0.590	1.010	0.452
14	*S*-Tyrosine	0.807	2.583	1.054	18.445	−0.180	−0.254	0.437	3.085	−0.686	1.020	0.188
15	*S*-Methionine	0.533	0.840	1.057	21.643	−0.276	−0.125	0.129	2.948	−0.606	1.010	0.461
16	*S*-Valine	0.535	0.165	1.068	20.450	0.477	−0.157	0.084	1.750	−0.487	1.010	0.000
17	GABA	0.107	1.750	0.638	20.608	0.267	−0.039	0.000	2.236	−0.499	1.020	0.000
18	*S*-Histidine	0.578	0.370	1.079	19.802	−0.214	−0.099	0.137	2.732	−0.108	1.030	0.740
19	*S*-Asparagine	0.403	0.573	0.899	17.282	0.374	−0.064	0.127	2.725	−0.281	1.050	0.000

Table 7
Experimental (Exp.) and QSPR predicted (Pred.) average relative percentages for orange and lemon fresh juices and concentrates.

No.	Amino acid (*levo*)	Orange juices								Eureka lemon juices			
		Navel				Valencia							
		Natural		Concentrate		Natural		Concentrate		Natural		Concentrate	
		Exp. % rel	Pred. % rel Eq. (1)	Exp. % rel	Pred. % rel Eq. (2)	Exp. % rel	Pred. % rel Eq. (3)	Exp. % rel	Pred. % rel Eq. (4)	Exp. % rel	Pred. % rel Eq. (5)	Exp. % rel	Pred. % rel Eq. (6)
1	*S*-Aspartic acid	32.5	–	29.2	–	28.4	15.2	17.1	6.20	22.9	23.0	27.9	31.9
2	*S*-Serine	9.50	7.42	7.00	10.0	21.0	18.3	10.6	12.0	21.9	22.6	22.8	14.0
3	*S*-Alanine	8.95	7.31	13.0	11.0	10.5	8.01	16.7	12.0	2.35	1.05	4.80	6.17
4	*R*-Cysteine	8.00[a]	8.73	4.20[a]	7.39	4.60[a]	11.1	6.80[a]	12.0	4.00[a]	9.84	4.60[a]	9.30
5	*S*-Proline	6.90	6.37	9.10	13.0	5.70	4.68	10.6	18.3	9.70	4.56	3.20	4.09
6	*S*-Lysine	6.55	6.42	6.40	6.13	6.00	4.56	6.00	5.95	0.50	0.65	0.70	0.93
7	*S*-Arginine	6.05	6.88	13.4	8.35	12.6	16.1	16.5	10.8	23.3	12.6	14.4	8.49
8	Ammonia	5.15	7.26	6.70	5.12	7.25	8.87	6.85	12.0	1.70	3.21	1.90	1.80
9	*S*-Glutamic acid	2.15	2.10	1.50	1.44	2.05	2.77	2.60	3.45	3.65	5.93	2.10	2.69
10	Threonine	2.60	2.24	3.05	2.52	2.80	5.74	3.85	3.00	7.20	9.01	10.9	9.30
11	Isoleucine	1.55[a]	1.07	1.20[a]	0.75	1.20[a]	2.00	1.40[a]	2.84	2.70[a]	5.35	1.90[a]	1.80
12	Leucine	1.40	1.34	1.00	1.07	0.95	1.15	1.20	1.15	2.60	2.41	1.90	1.80
13	*S*-Phenylalanine	1.25	0.67	0.80	1.00	0.85	0.52	0.80	1.21	1.65	1.78	1.60	0.68
14	*S*-Tyrosine	0.95	1.05	0.70	0.62	0.90	0.86	1.10	1.61	2.20	2.77	1.70	2.30
15	*S*-Methionine	0.80	1.26	0.80	1.24	1.55	2.30	0.90	1.09	2.45	3.46	0.70	0.66
16	*S*-Valine	0.60	0.86	0.40	0.90	0.40	–	0.60	0.80	1.10	1.05	1.30	2.71
17	GABA	–[b]	12.2	–[b]	7.98	–[b]	16.1	–[b]	6.24	–[b]	2.08	–[b]	4.09
18	*S*-Histidine	–[b]	0.81	1.90	0.74	–[b]	3.33	1.80	0.87	–[b]	37.1	–[b]	0.64
19	*S*-Asparagine	–[b]	1.87	–[b]	1.03	–[b]	16.0	–[b]	2.39	–[b]	14.4	–[b]	14.0

[a] Test set amino acid.
[b] Estimated amino acid.

standard deviation of *y*-randomization. The various decision criteria that were simultaneously analysed for determining the model's size of the five training sets investigated, such as the determination of the optimal *d* to be included in each QSPR, are 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*, 2.5*S* or 3*S*; (e) the lowest value for the maximal inter-correlation between descriptors in the model. A brief description for each molecular descriptor appearing in the following structure–property relationships is supplied in Table 5. Numerical values of the theoretical descriptors involved in the various QSPR designed in this study are shown in Table 6.

3.3.2. Natural Navel orange juices

From the set of amino acids shown in Fig. 1, we used 14 of them for calibrating the structure–property relationship. The elimination of **1** (aspartic acid) from this molecular set lowered *S* by one half, and markedly improved both *R* and *FIT* parameters. In addition, this removal allowed us to use one less variable in the QSPR, thus resulting in the following two-descriptor model:

$$\log_{10}(\%\ \text{rel}) = -1.804(\pm 0.2)\cdot BELe4 + 0.239(\pm 0.03)\cdot RDF050v + 0.861(\pm 0.07) \quad (1)$$

$N = 13$, $d = 2$, $N/d = 6.5$, $R = 0.956$, $S = 0.138$, $FIT = 6.228$, $p < 10^{-5}$, outliers (>3*S*) = 0, $R_{loo} = 0.924$, $S_{loo} = 0.182$, and $S_{rand} = 0.210$.

As can be appreciated, the cross validation parameters R_{loo} and S_{loo} took values that were quite close to the training parameters *R* and *S*, while Eq. (1) passed the *y*-randomization test. The ratio of number of observations to number of descriptors used was 6.5, thus tallying with the rule of thumb. This QSPR was employed for predicting the average relative percentages for the test set amino acids, denoted with [a] in Table 7 (**4**, cysteine and **11**, isoleucine). Good predictions were found for both structures. The plot of predicted % rel values, as a function of experimental % rel, shown in

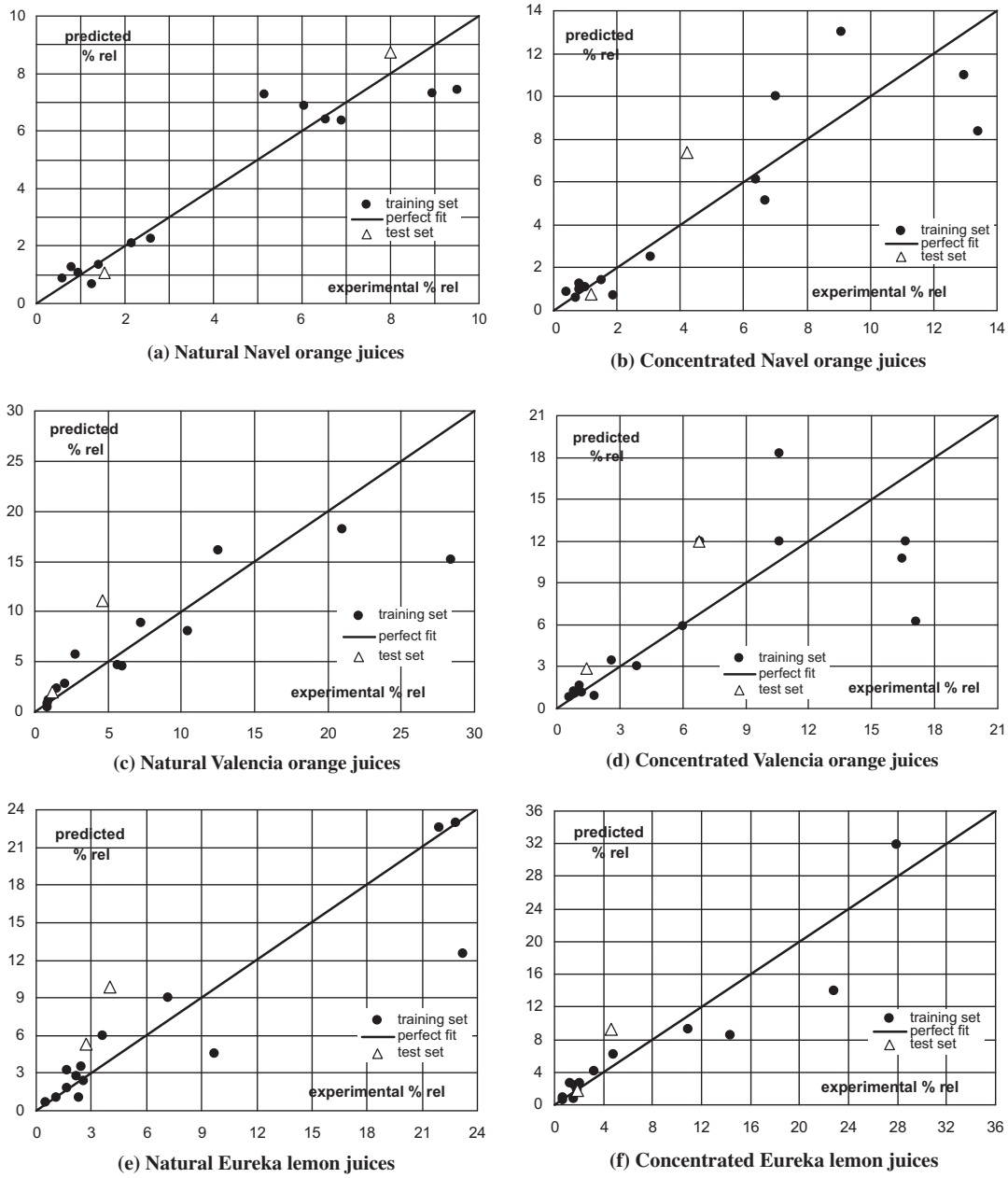

Fig. 2. Observed and QSPR predicted average relative percentages for different citrus juices and concentrates analysed in this work.

Fig. 2(a), suggests that the proposed model approximately fits a straight line, for both the training and test set data. In addition, Eq. (1) was also used to estimate amino acids **17** (GABA), **18** (histidine) and **19** (asparagine), denoted with [b] in Table 7, which had no experimental data available in the aminogram.

3.3.3. Navel orange concentrates

As mentioned above, the removal of **1** (aspartic acid) from the training set, whose experimental % rel deviated from the other values, improved the statistical parameters *S*, *R*, and *FIT*, and allowed us to use one less descriptor in the QSPR:

$$\log_{10}(\%\ \text{rel}) = -2.169(\pm 0.2)\cdot BELv4 + 0.0973(\pm 0.01)\cdot Mor02e + 0.279(\pm 0.2) \quad (2)$$

$N = 14$, $d = 2$, $N/d = 7$, $R = 0.933$, $S = 0.204$, $FIT = 4.119$, $p < 10^{-5}$, outliers (>3*S*) = 0, $R_{loo} = 0.884$, $S_{loo} = 0.269$, and $S_{rand} = 0.350$.

This structure–property relationship had a good statistical behaviour from both calibration and validation points of view. Eq. (2) was used to predict average relative percentages of the test set amino acids **4** (cysteine) and **11** (isoleucine), which were not taken into account for adjustment. Again, the predictions of % rel achieved through QSPR reflected the tendency of variation of the dependent variable, for both the training and test sets of amino acids. This QSPR was also used to estimate the amino acids **17** (GABA) and **19** (asparagine), which had no experimental data available. All predictions are included in Table 7.

3.3.4. Natural Valencia orange fresh juices

The prediction of the aminogram was notably improved when **6** (valine), the amino acid exhibiting the lowest % rel value in Valencia orange juices, was removed from the training set, leading to the following QSPR:

$$\log_{10}(\%\ \text{rel}) = 0.920(\pm 0.2)\cdot MATS6p + 4.009(\pm 0.5)\cdot Mor32u + 1.116(\pm 0.09) \quad (3)$$

$N = 13$, $d = 2$, $N/d = 6.5$, $R = 0.949$, $S = 0.182$, $FIT = 5.323$, $p < 10^{-5}$, outliers (>3*S*) = 0, $R_{loo} = 0.912$, $S_{loo} = 0.238$, and $S_{rand} = 0.194$.

Eq. (3) was used to predict the average relative percentages of the test set amino acids, and also to estimate amino acids having no observed data, leading to acceptable predictions in both cases.

3.3.5. Valencia orange concentrates

An increased concentration in Valencia orange juices enabled us to include the previously removed outlier **6** (valine), due to its higher experimental % rel value when compared with the previous concentration. This allowed this compound to exhibit an experimental concentration more similar to that of the other amino acids. The next QSPR was designed on 15 structures:

$$\log_{10}(\%\ \text{rel}) = 2.920(\pm 0.5)\cdot BELv7 - 2.663(\pm 0.3)\cdot BELe4 + 1.080(\pm 0.1) \quad (4)$$

$N = 15$, $d = 2$, $N/d = 7.5$, $R = 0.923$, $S = 0.219$, $FIT = 3.663$, $p < 10^{-5}$, outliers (>3*S*) = 0, $R_{loo} = 0.875$, $S_{loo} = 0.277$, and $S_{rand} = 0.301$.

As was the case for previous data, Eq. (4) was used to predict average relative percentages of the test set **4** (cysteine) and **11** (isoleucine), and to estimate the unknown amino acids **17** (GABA) and **19** (asparagine).

3.3.6. Natural Eureka lemon fresh juices

The following two-descriptor model was established on 14 amino acids:

$$\log_{10}(\%\ \text{rel.}) = 0.666(\pm 0.09)\cdot IDDE + 2.356(\pm 0.3)\cdot Mor21u + 0.00291(\pm 0.2) \quad (5)$$

$N = 14$, $d = 2$, $N/d = 7$, $R = 0.929$, $S = 0.209$, $FIT = 3.875$, $p < 10^{-5}$, outliers (>3*S*) = 0, $R_{loo} = 0.838$, $S_{loo} = 0.317$, and $S_{rand} = 0.305$.

As can be seen from Table 7, this QSPR was able to acceptably perform predictions on the training set amino acids, having predictive power on the removed test set data.

3.3.7. Eureka lemon concentrates

As for the previously proposed QSPR, the next model accomplished with the empirical rule of thumb, displayed an appropriate N/d ratio:

$$\log_{10}(\%\ \text{rel.}) = 17.832(\pm 3)\cdot Me - 1.332(\pm 0.2)\cdot MATS8p - 17.577(\pm 3) \quad (6)$$

$N = 14$, $d = 2$, $N/d = 7$, $R = 0.942$, $S = 0.194$, $FIT = 4.795$, $p < 10^{-5}$, outliers (>3*S*) = 0, $R_{loo} = 0.904$, $S_{loo} = 0.249$, and $S_{rand} = 0.293$.

All of these predictive linear QSPR models were able to capture the essential structural features of the amino acids that related to their concentration in a given citrus juice. It is notable that these derived QSPRs involved a combination of 2D- and 3D-type molecular descriptors in order to achieve the best predictions for the aminograms.

4. Conclusions

A large worldwide consumption of citrus products justifies an exhaustive chemical study addressed to enhanced food quality and food authenticity control. The aim of our work was to study Argentine orange and lemon juices and concentrates, since these are the most consumed in the country, and largely exported citrus products. Juices and concentrates of Navel and Valencia oranges and Eureka lemons were selected.

Suitable experimental conditions for the analysis of natural juices and concentrates have been previously obtained in our laboratories. Further validated HPLC analyses of these juices and concentrates were carried out herein, taking into account the reported results, but with a higher number of samples, thus allowing a statistical analysis. Owing to the number of samples analysed, *e.g.*, a total of 36 samples, this work can be considered a preliminary and promising study. A proper study should be carried out with a representative number of samples of fresh and concentrated juices from several countries.

The validation of the HPLC method developed herein showed that this approach is accurate, precise, reproducible, specific, and linear over a wide concentration range. All these factors make this method suitable for quantification of amino acids in fresh and concentrated citrus juices. It can therefore be concluded that the method can be used, even in small laboratories, with very high accuracy and precision. The method can also be used for the routine analysis of amino acids in citrus juices and concentrates without interference.

Furthermore, experimental values were subjected to a theoretical study in order to contribute to the classification and identification of these natural juices and concentrates from the fruit varieties mentioned above. The findings may also help to protect against fraudulent products.

Consequently, the representation of the molecular structure of various citrus amino acids through more than one thousand theoretical structural descriptors, encoding various constitutional, topological, geometrical and electronic aspects, allowed the formulation of predictive linear QSPR models for aminograms, accounting for 100% natural fresh juices and concentrates from Navel and Valencia oranges, and Eureka lemons.

This is a new application of the QSPR theory as, to our knowledge, only a few studies are devoted to the quantification of amino acid profiles in food. A practical application of these QSPRs consists

of estimating some structures, whose experimental values are missing. Furthermore, any natural citrus juice and/or concentrate belonging to these fruit varieties will fit the corresponding developed model, thus demonstrating its authenticity.

Acknowledgements

We thank 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, Tecnología e Innovación Productiva (Mincyt, Argentina) for electronic bibliography facilities. M.A.G. is Chairman of Food Science and Food Technology (Director de Carrera de Ciencia y Tecnología de los Alimentos) of Universidad Nacional de Lanús (Argentina). A.B.P., E.A.C. and P.R.D. are Research Members of CONICET.

References

Barrett, D. M., Somogyi, L. P., & Ramaswamy, H. S. (2004). *Processing fruits: Science and technology* (2nd ed.). Boca Raton, FL: CRC Press.

Benigni, R., Passerini, L., Pino, A., & Giuliani, A. (1999). The information content of the eigenvalues from modified adjacency matrices: Large scale and small scale correlations. *Quantitative Structure–Activity Relationships, 18*, 449–455.

Bonchev, D. (1983). *Information theoretic indices for characterization of chemical structures*. Chichester, UK: RSP-Wiley.

Boogers, I., Plugge, W., Stokkermans, Y. Q., & Duchateau, A. L. (2008). Ultra-performance liquid chromatographic analysis of amino acids in protein hydrolysates using an automated pre-column derivatisation method. *Journal of Chromatography A, 1189*, 406–409.

Bosch, L., Alegría, A., & Farré, R. (2006). Application of the 6-aminoquinolyl-*N*-hydroxy-succinimidyl carbamate (AQC) reagent to the RP-HPLC determination of amino acids in infant foods. *Journal of Chromatography B, 831*, 176–183.

Cohen, S. A. (2000). Amino acid analysis using precolumn derivatization with 6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate. *Methods in molecular biology*. Totowa, NJ: Humana Press (pp. 39–47).

Cohen, S. A., & Michaud, D. P. (1993). Synthesis of a fluorescent derivatizing reagent, 6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate, and its application for the analysis of hydrolysate amino acids *via* high-performance liquid chromatography. *Analytical Biochemistry, 211*, 279–287.

Consonni, V., Todeschini, R., Pavan, M., & Gramatica, P. (2002). Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies. *Journal of Chemical Information and Computer Sciences, 42*, 693–705.

Dragon. Milano Chemometrics and QSAR Research Group. <http://www.michem.disat.unimib.it/chm>.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). New York: John Wiley & Sons.

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

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

Gasteiger, J., Sadowski, J., Schuur, J., Selzer, P., Steinhauer, L., & Steinhauer, V. (1996). Chemical information in 3D-space. *Journal of Chemical Information and Computer Sciences, 36*, 1030–1037.

Giraudo, M., Guirin, G., Montesano, J., Markowski, I., Pavesi, R., & Sánchez Tuero, H. (2004). El proceso productivo de los jugos cítricos y la determinación de los aminoácidos libres por RP-HPLC en jugos frescos y concentrados como criterio de genuidad de los mismos. *Alimentaria: Revista de Tecnología e Higiene de los Alimentos*(350), 97–103.

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

Hansch, C., & Leo, A. (1995). *Exploring QSAR: Fundamentals and applications in chemistry and biology*. Washington, DC: American Chemical Society.

Hypercube Inc. *Hyperchem 6.03*. <http://www.hyper.com>.

ICH (International Conference on Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use) (1996). *Validation of analytical procedures: Definitions and terminology. Methodology*. Geneva, Switzerland: ICH.

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*, 279–287.

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

Mitchell, M. (1998). *An introduction to genetic algorithms*. Boston: The MIT Press.

Moreau, G., & Broto, P. (1980). The autocorrelation of a topological structure: A new molecular descriptor. *Nouveau Journal de Chimie, 4*, 359–360.

Nollet, L. M. L. (2004). *Handbook of food analysis* (2nd revised and expanded ed.). *Methods and instruments in applied food analysis* (Vol. 3). New York, USA: Marcel Dekker Inc.

Oreiro-García, M. T., Vázquez-Illanes, M. D., Sierra-Paredes, G., & Sierra-Marcuño, G. (2005). Analysis of neuroactive amino acids from microdialysate samples by fluorescence detection using a modification of the 6-aminoquinolyl-*N*-hydroxy-succinimidyl carbamate method. *Biomedical Chromatography, 19*, 720–724.

The MathWorks Inc. (2004). *Matlab 5.0*. The MathWorks Inc.

Todeschini, R., & Consonni, V. (2000). *Handbook of molecular descriptors*. Weinheim: Wiley–VCH.

Trinajstic, N. (1992). *Chemical graph theory* (2nd revised ed.). Boca Raton, FL: CRC Press.

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* (Vol. 4, pp. 1–31). Oxford: Pergamon Press.

USP (United States Pharmacopeia) (2007). General Chapter 1225. *Validation of compendial methods* (30th ed.). National Formulary 25, Rockville, MD, USA: The United States Pharmacopeial Convention Inc.

van de Waterbeemd, H. (1995). In R. Mannhold et al. (Eds.), *Advanced computer-assisted techniques in drug discovery*. *Methods and principles in medicinal chemistry* (Vol. 3). Weinheim: VCH.

Wold, S., Eriksson, L., & Clementi, S. (1995). Statistical validation of QSAR results. In H. van de Waterbeemd (Ed.), *Chemometrics methods in molecular design*. *Methods and principles in medicinal chemistry* (pp. 309–338). Weinheim: VCH (Chapter 5).