Available online at www.sciencedirect.com
SCIENCE DIRECT®
Food Chemistry 93 (2005) 113–123
Food Chemistry
www.elsevier.com/locate/foodchem

Classification of apple fruits according to their maturity state by the pattern recognition analysis of their polyphenolic compositions

Rosa M. Alonso-Salces [a], Carlos Herrero [b], Alejandro Barranco [a], Luis A. Berrueta [a,*], Blanca Gallo [a], Francisca Vicente [a]

[a] Departamento de Química Analítica, Facultad de Ciencias, Universidad del País Vasco/Euskal Herriko Unibertsitatea, P.O. Box 644, E-48080 Bilbao, Spain
[b] Departamento de Química Analítica, Nutrición y Bromatología, Facultad de Ciencias, Universidad de Santiago de Compostela, Augas Férreas s/n, Campus Universitario, 27002 Lugo, Spain

Received 28 June 2004; received in revised form 13 October 2004; accepted 13 October 2004

Abstract

Polyphenolic profiles of cider apple cultivars were studied in order to differentiate fruits according to their maturity state (ripe or unripe). Thiolysis and direct solvent extracts of freeze-dried apple pulps and peels were analysed by HPLC-DAD. Univariate data treatment did not achieve the mentioned target; thus a multivariate approach was considered. For each apple tissue data set, several chemometric techniques were applied to the most discriminant variables. Cluster analysis allowed a preliminary study of the data structure. Then, supervised pattern recognition procedures, namely linear discriminant analysis, K-nearest neighbours, soft independent modelling of class analogy, and multilayer feed-forward artificial neural networks (MLF-ANN), were used to develop decision rules to classify samples in the established categories. Excellent results were afforded by MLF-ANN applied to the concentrations of total procyanidins and (+)-catechin and the average degree of polymerisation of procyanidins in apple pulp, with success in the prediction ability of 97% and 99% for unripe and ripe categories, respectively.

Keywords: Polyphenols; Apple; Maturity; Pattern recognition analysis; Chemometrics

1. Introduction

Biotransformations of cider apple fruits have to be carried out when they satisfy certain technological quality criteria. In this sense, fruits have to be processed at their optimum maturity state, i.e., the time when they present an adequate chemical composition, which is responsible for their organoleptic and nutritional properties, as well as for the characteristics of their derived products.

Several works have been developed to study the chemical constituents of apple involved in the biochemical transformations that take place during maturation (Ackermann, Fisher, & Amadò, 1992; Blanco, Picinelli, Gutiérrez, & Mangas, 1992b; Mangas et al., 1992), in order to establish the ontogenic stage for technological purposes. Significant changes have been detected in apple composition in the last stages of ripening, such as an accumulation of sugars and pectins, a decrease in the content of organic acids and some amino acids, and a minimum in the total polyphenol and nitrogen concentrations reached just before the final accumulation of sugars and starch degradation (Blanco et al., 1992a).

Traditional criteria used for evaluating the optimum maturity state of apples are the starch content (Le Lezec

* Corresponding author. Tel.: +34 94 601 5505; fax: +34 94 464 8500.
E-mail address: qapbesil@lg.ehu.es (L.A. Berrueta).

0308-8146/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodchem.2004.10.013

& Babin, 1988), the total sugar/total acidity ratio (Board & Woods, 1983), the internal ethylene content (Dilley, 1981), the fruit consistency, as resistance to penetration (Trillot, Masseron, & Tronel, 1993), and the content of soluble solids (Sieguist, 1987). However, other chemical markers, closely related to the quality of apple products, such as sugars, organic acids, amino acids, total polyphenols, and pectins, together with chemometric techniques, have allowed a more precise characterisation of the different apple varieties according to their degree of ripening (Mangas, Moreno, Picinelli, & Blanco, 1998).

Polyphenols play an important role in the biotransformation process of apples since, not only do they contribute to cider flavour and aroma (Lea, 1995), but they also control microorganism metabolism present in the medium, controlling fermentation rates (Cowan, 1999), avoiding the development of some faults in cider (Sponholtz, 1993), participating in cider spontaneous clarification, and inhibiting enzymatic systems such as clarification enzymes (Lea, 1990). Hence, it is important to consider phenolic compounds in order to establish the optimum maturity state of apples to be processed.

Among the main classes of apple polyphenols, flavan-3-ols are preponderant, being present in monomeric forms named catechins, and in oligomeric and polymeric forms, known as procyanidins. The latter contribute to astringency and bitterness of apples (Lea, 1990) and their derived products, and to the formation of hazes and precipitates during apple juice and cider storage, due to their ability to associate with proteins (Mc Manus et al., 1985) and polysaccharides (Renard, Baron, Guyot, & Drilleau, 2001). Hydroxycinnamic acids are the next major class. Together with catechins, they are involved in the browning phenomena that take place during apple fruit processing, being responsible for the yellow or orange coloration of apple products (Amiot, Tacchini, Aubert, & Nicolas, 1992). Dihydrochalcones, flavonols and anthocyanins are minor components that contribute to the pigmentation of apples, and to the potential antioxidant activity of apples and their derived foodstuffs (Ridgway, Oreilly, West, Tucker, & Wiseman, 1996).

Several studies of apple composition during fruit development and maturation (Burda, Oleszek, & Lee, 1990; Lancaster, 1992; Mayr, Treutter, Santos-Buelga, Bauer, & Feucht, 1995; Treutter, 2001) demonstrate that phenolic contents of the different apple tissues are not exclusively dependent on the cultivar, but also on their maturity state. Polyphenolic profiles vary during fruit growth and ripening. Thus, in the first weeks of their ontogenesis, fruits present high levels of polyphenols that decrease throughout fruit development to a minimum, preceding the final accumulation of sugars in the last stage of maturation. Then, polyphenol concentrations remain practically constant or slightly increase (Macheix, Fleuriet, & Billot, 1990; Blanco et al., 1992a). On the other hand, anthocyanin contents increase markedly during maturation. The variation of the procyanidin content during fruit ontogenesis is closely related to the changes detected in the astringency. Main responsible factors for these organoleptic parameter changes are genetic, leading to notable differences between varieties, and physiological, which are directly related to the state of maturation of the fruit. Indeed, astringency decreases or disappears completely during maturation. This decrease generally takes place along with a decrease in the fruit procyanidin content, or with physicochemical changes of the procyanidin molecules, affecting their degree of polymerisation, which plays an important role in their ability to interact with proteins (Macheix et al., 1990).

In the present work, polyphenolic content of Basque cider apple varieties, harvested at two different stages of ripening, were evaluated in order to differentiate ripe and unripe fruits. The aim of this study was to achieve classification rules that would allow prediction of fruit maturity state (ripe or unripe) according to their polyphenolic profiles by pattern recognition analysis of the data.

2. Materials and methods

2.1. Solvents and standard phenolics

Methanol (Romil Chemical Ltd, Heidelberg, Germany) was of HPLC grade. Water was purified on a Milli-Q system from Millipore (Bedford, MA, USA). Glacial acetic acid, formic acid, toluene-α-thiol and fuming hydrochloric acid (37%) provided by Merck (Darmstadt, Germany) and ascorbic acid, by Panreac (Barcelona, Spain), were of analytical quality. All solvents used were previously filtered through 0.45 μm nylon membranes (Lida, Kenosha, WI, USA).

Polyphenol standards were supplied as follows: (+)-catechin, (−)-epicatechin, rutin, phloridzin, 5-caffeoylquinic acid and *p*-coumaric acid by Sigma–Aldrich Chemie (Steinheim, Germany); hyperoside, isoquercitrin, avicularin, quercitrin and ideain chloride by Extrasynthèse (Genay, France). (−)-Epicatechin 4R-benzylthioether and 4-*p*-coumaroylquinic acid were kindly provided by Dr. Guyot (Laboratoire de Recherches Cidricoles, Biotransformation des Fruits et Légumes, INRA, Le Rheu, France), and phloretin-2′-*O*-xyloglucoside and procyanidin B2 by Dr. F.A. Tomás-Barberán (Laboratorio de Fitoquímica, Departamento de Ciencia y Tecnología de los Alimentos, CEBAS (CSIC), Murcia, Spain) and Dr. C. Santos-Buelga (Departamento de Química Analítica, Nutrición y Bromatologia, Universidad de Salamanca, Spain), respectively. Stock standard solutions of (+)-catechin, (−)-epicatechin, (−)-epicate-

chin 4R-benzylthioether, rutin, phloridzin, 5-caffeoylquinic acid and *p*-coumaric acid, at a concentration of 1 mg ml^{-1} and hyperoside, isoquercitrin, quercitrin and ideain, at 0.6 mg ml^{-1}, were prepared in methanol and stored at 4 °C in darkness.

2.2. Samples

Pulp and peel from 14 apple cultivars used in the Basque Country for cidermaking were analysed. The technological groups of the cultivars studied were as follows: *bittersweet*: Geza Miña (GM), Mozoloa (MZ), Patzuloa (PT) and Ugarte (UG); *semiacid*: Bost Kantoi (BK), Manttoni 111 (MN111), ManttoniEM7 (MNEM7) and Urtebi Txiki (UT); *acid*: Errezila (ER), Goikoetxea (GK), Txalaka (TX), Udare Marroi (UM) and Urtebi Haundia (UH); and *bitter-acid*: Moko (MK).

Apples were harvested in the Experimental Orchard of the Diputación Foral de Gipuzkoa in Hondarribia (Guipúzcoa, Spain) during the 1999, 2000 and 2001 seasons. The time lag between the harvest of unripe and ripe fruits was about three weeks. Maturity state of the fruits was tested by the starch index (Planton, 1995). Unripe fruits presented index values between 4 and 8, and ripe fruits, between 9 and 10. For each variety and season, two or three batches of 10 apple fruits were mechanically peeled and cored, and sprayed with an aqueous solution of formic acid (3%, v/v) in order to avoid polyphenol oxidation. Peels (average thickness: 0.8 mm) and pulps were immediately frozen in liquid nitrogen and then freeze-dried. An aliquot of each variety was used to determine the fresh/dry matter ratio. The dried tissues were crushed in closed vials to avoid hydration, using stainless-steel balls (∅1 cm) and an overhead shaker (Reax 2, Heidolph, Schwabach, Germany). A homogeneous powder was obtained, which was stored at room temperature in a desiccator until analysis.

2.3. Analytical procedures

2.3.1. Thiolysis and direct solvent extraction

Freeze-dried apple pulp and peel (0.5 g) were submitted to thiolysis as described by Guyot, Marnet, Sanoner, and Drilleau (2001), and to direct solvent extraction with 30 ml of methanol–water–acetic acid (30:69:1, v/v/v) with ascorbic acid (2 g/l) in an ultrasonic bath during 10 min (Alonso-Salces et al., 2004a). Afterwards, both the thiolysis reaction mixture and the crude solvent extract were filtered through a 0.45μm PTFE filter (Waters, Milford, CA, USA) prior to injection into the HPLC system.

2.3.2. Reversed-phase HPLC analysis

Chromatographic analysis was performed on a Hewlett–Packard Series 1100 system, equipped with a vacuum degasser, a quaternary pump, a thermostatted autosampler, a thermostatted column compartment and a DAD detector, connected to an HP ChemStation software. A reversed phase Nova-Pak C18 (300 × 3.9 mm i.d., 4 μm) column and a Nova-Pak C18 (10 × 3.9 mm i.d., 4 μm) guard column (Waters, Barcelona, Spain) were used. Solvents that constituted the mobile phase were A (acetic acid–water, 10:90, v/v) and B (methanol). The elution conditions applied were: 0–10 min, 0% B isocratic; 10–40 min, linear gradient from 0% to 15% B; 40–60 min, 15% B isocratic; and finally, washing and reconditioning of the column. The flow rate was 0.8 ml min^{-1} and the injection volume was 50 μl of crude extracts or 10 μl of thiolysis media. The system operated at 25 °C. Flavan-3-ols and dihydrochalcones were monitored and quantified at 280 nm, hydroxycinnamic acids at 320 nm, flavonols at 370 nm and anthocyanins at 530 nm. Polyphenol identification, for which standards were available, was carried out by comparison of their retention times and their UV–visible spectra with those of the standards. Some other chromatographic peaks were assigned to a particular polyphenol class according to their UV–visible spectra and bibliographic sources. In this sense, those unknown chromatographic peaks that exhibited flavan-3-ol spectra were designated as CAT-*n*, and those with a spectrum of 5-caffeoylquinic acid as CAA-*n*, of *p*-coumaric as CMA-*n*, of dihydrochalcone as PLD-*n*, of flavonol as QG-*n* and of anthocyanin as CG-*n* (where "*n*" is a number). Quantification was performed by reporting the measured integration areas in the calibration equation of the corresponding standards. Thus, procyanidin B2 and the unknown flavan-3-ols were quantified as (+)-catechin; phloretin-2′-*O*-xyloglucoside and the unknown dihydrochalcones were quantified as phloridzin; aviculin and the unknown flavonols were quantified as rutin; CAA-*n* species were quantified as 5-caffeoylquinic acid; 4-*p*-coumaroylquinic acid and CMA-*n* species were quantified as *p*-coumaric acid and the unknown anthocyanins as ideain.

2.4. Data analysis and chemometric procedures

A peel and pulp data set consisted of a 85 × 27 matrix and a 85 × 18 matrix, in which rows represented apple samples, and columns the concentrations of individual polyphenols determined by HPLC-DAD, the total concentration of procyanidins, and the average degree of polymerisation of procyanidins (DPn). Each sample was represented in the multidimensional space by a data vector, which was an assembly of the 27 features in peel and 18 features in pulp. Data vectors belonging to the same category (ripe or unripe) were analysed using chemometric procedures that have been described in the literature (Latorre, Peña, García, & Herrero, 2000; Padín et al., 2001), such as cluster analysis (CA), linear discriminant analysis (LDA), K-nearest neighbours (KNN),

soft independent modelling of class analogy (SIMCA), and multilayer feed-forward artificial neural networks (MLF-ANN). Statistic and chemometric data analysis was performed by means of the statistical software packages Statgraphics (1994–2000), Parvus (Forina, Lanteri, & Armanino, 2000) SPSS (1989–1999) and WinNN32 (1993–1996).

Cluster analysis is a preliminary way to study data sets in the search for natural groupings among the samples characterised by the values of a set of measured features. Owing to its unsupervised character, CA is a pattern recognition technique that can be used to reveal the structure residing in a data set (Massart & Kaufman, 1983). CA was performed on autoscaled data, sample similarities were calculated on the basis of the squared Euclidean distance and the Ward hierarchical agglomerative method was used to establish clusters.

The classification rules achieved by the supervised chemometric techniques were validated by means of a cross-validation procedure, which was performed by dividing the complete data set into a training set and an evaluation set. Samples were assigned randomly to a training set, consisting of 75% of them, and the test set, composed of the remaining 25% of the samples. Such a division allows for a sufficient number of samples in the training set and a representative number of members among the test set. The same process was repeated four times with different constitutions of both sets to ensure that all the samples had the possibility of inclusion in the evaluation set at least once. The different pattern recognition techniques were applied to the four training-test sets obtained. The reliability of the classification models achieved was studied in terms of recognition ability (percentage of the members of the training set correctly classified) and prediction ability (percentage of the members of the test set correctly classified by using the rules developed in the training step).

In KNN, the inverse square of the Euclidean distance was used as the criterion for calculating the distance between samples, and the number of neighbours (K) was selected after studying the success in classification with different K values, applying this technique to a training set with all the samples.

The model achieved by SIMCA for each category was also evaluated in terms of sensitivity and specificity. The sensitivity of the model is known as the percentage of objects belonging to the category correctly identified by the mathematical model, and its specificity, as the percentage of objects foreign to the category classified as foreign (Meléndez, Ortız, Sánchez, Sarabia, & Iñiguez, 1999).

In MLF-ANN, the target output was assigned as 0 or $(0,1)$ for unripe fruits and 1 or $(1,0)$ for ripe ones, and a sigmoidal function $f(x) = 1/(1 + [\exp(-x)])$ was used as the transfer function. The neural network was trained by means of an algorithm that combined the use of an adaptative learning rate parameter (ALRP) (η) and a momentum (μ) which have been described previously (Padín et al., 2001). The initial values of the weights associated with the connections between neurons were selected randomly in the range -3 to 3. The maximum number of epochs was 2000 and the initial values of η and μ were 0.2 and 0.5, respectively; the target error was 0.1.

3. Results and discussion

3.1. Analytical data

Analytical data obtained by the chromatographic determination of apple polyphenols in pulp and peel for each variety are summarised in Tables 1 and 2. Total polyphenols and procyanidin contents were smaller in unripe apple peels and pulps. Flavonols and anthocyanins in peels, and hydroxycinnamic acids and flavan-3-ols in pulps presented higher concentrations in ripe fruits. The average degrees of polymerisation of procyanidins (DPn) were considerably higher in unripe fruits. From the results obtained by the analysis of the polyphenolic profiles of the studied apple cultivars, it seemed that unripe fruits were in an earlier stage of maturation but later than the minimum reached in total polyphenol concentration prior to the accumulation of sugars aforementioned, since the concentrations of certain classes of polyphenols were lower in unripe fruits than in ripe ones. Furthermore, notable differences were observed in the DPn, these being higher in unripe fruits, which agreed with the observations reported by Macheix et al. (1990). In cidermaking, the use of fruits that are not at the optimum maturity state can lead to an important increase of bitterness and astringency and the incidence of troubles related to procyanidins, such as cloudiness after bottling (Lea, 1990; Siebert, Carrasco, & Lynn, 1996).

Bitterness of the unripe fruits was predicted by the classification system achieved by Alonso-Salces, Herrero, Barranco, Berrueta, and Vicente (2004b), in order to study the influence of apple maturity state on the classification of apple cultivars in technological groups. In most cases, bitterness predictions made for the unripe fruits agreed with the results obtained for the ripe fruits (Alonso-Salces et al., 2004b). This allows us to conclude that differences in polyphenolic composition between bitter and non-bitter varieties are already evident before fruits reach their optimum maturity state. However, regarding some unripe fruits, non-conclusive results were obtained for GM (pulp), and MK (pulp), UG (peel) and PT (peel) were misclassified as non-bitter, because they had lower concentrations of polyphenols than ripe fruits. This can be explained by the fact that the polyphenol content reaches a minimum just before

Table 1
Concentrations (mg kg^{-1} of apple) of flavan-3-ols, hydroxycinnamic acids, dihydrochalcones and flavonols and the average degree of polymerisation of procyanidins (DPn) in apple pulps

Polyphenol	Fruit maturity state							
	Unripe				Ripe			
	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
Flavan-3-oles								
CAT	14	23	nd	91	23	29	0.7	113
EC	86	57	26	209	115	63	48	227
PB2	87	61	32	231	117	74	46	324
CAT-2	9	5	3	22	11	6	5	27
PC	990	545	500	2065	1398	623	672	3041
DPn	7	1	5	9	4.7	0.5	4.0	5.6
Hydroxycinnamic acids								
CQA	240	177	40	630	311	200	61	724
PCQ	19	17	1	59	29	32	1	120
CAA-1	23	19	4	58	24	15	8	58
CMA-2	0.5	1.0	nd	3.4	0.2	0.5	nd	1.5
Dihydrochalcones								
PLXG	19	13	3	45	29	19	5	62
PLG	19	25	5	100	17	12	7	56
PLD-1	6	6	0.4	20	3	2	0.7	7
PLD-2	5	9	1	34	3	2	1	11
Flavonols								
HYP	0.2	0.6	nd	2.2	t		nd	t
IQC	0.4	0.3	nd	1.0	0.4	0.3	nd	0.8
QCI	1	1	nd	5	2	1	0.5	5
QG-1	0.6	0.4	nd	1.6	0.6	0.5	t	1.6

the final accumulation of sugars and the starch degradation, which takes place at the final stage of apple maturation (Blanco et al., 1992a).

3.2. Univariate data analysis

The analysis of variance (ANOVA) performed on apple peel and pulp data sets, constituted of the individual polyphenol concentrations, total procyanidin content and DPn of the fruits disclosed significant differences for all variables between ripe and unripe fruits. A least significant difference (LSD) test ($p < 0.05$) was also carried out on the data matrices of both apple tissues, in order to check that there were no significant differences between seasons. Fisher's test allowed us to detect the most discriminant variables between ripe and unripe fruits (Sharaf, Illman, & Kowalski, 1986). In both tissues, peel and pulp, DPn was the feature that presented a higher Fisher weight ($p < 0.001$). In pulp, the following variables with the highest Fisher values ($p < 0.050$) were the total content of procyanidins (PC) and (+)-catechin (CAT), and in peel ($p < 0.001$), PLD-2 and PLD-1, which are hydroxyphloretin glycosides (Alonso-Salces et al., 2004c) and phloretin-2′-*O*-xyloglucoside (PLXG). Hence, it seems that flavan-3-ols in pulp and dihydrochalcones in peel are the classes of polyphenols that undergo greater changes in the last stage of maturation, whereas the remaining polyphenol concentrations remain practically constant, as noted before in the literature (Burda et al., 1990; Mayr et al., 1995). Despite the differences in these features between ripe and unripe apple fruits, their box and whisker plots showed an overlap between the two classes, indicating insufficient discriminatory ability. Thus, none of the variables measured was able to discriminate between ripe and unripe categories by itself. Therefore, it was necessary to move on to a multivariate data analysis.

3.3. Multivariate data analysis

3.3.1. Cluster analysis

When CA was applied to the complete set of variables, no clear groupings of the samples according to their maturity (ripe or unripe) could be observed, either in peel or in pulp. However, considering only the most discriminant variables in each apple tissue, enough information was provided by these features to achieve a classification of fruits in the established categories. Thus, the variables regarded in peel were DPn, PLD-2, PLD-1 and PLXG, and in pulp, DPn, PC and CAT. Results attained for each apple tissue are presented as a dendrogram (Fig. 1). In pulp, at a similarity level of 0.30, three clusters were identified as follows: A, made up of unripe fruits, B, containing all ripe fruits and the unripe fruits of cultivars GK and MNEM7, and C, consisting of samples of UG variety. In peel, at a similarity

Table 2
Concentrations (mg kg^{-1} of apple) of flavan-3-ols, hydroxycinnamic acids, dihydrochalcones, flavonols and anthocyanins and the average degree of polymerisation of procyanidins (DPn) in apple peels

Polyphenol	Fruit maturity state							
	Unripe				Ripe			
	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
Flavan-3-oles								
CAT	3	3	0.2	11	4	4	0.3	17
EC	38	28	12	94	37	25	11	81
PB2	33	19	12	72	32	17	8	56
CAT-2	4	2	2	7	4	2	2	6
PC	574	171	355	882	629	181	360	929
DPn	9	2	7	16	5.8	0.7	4.9	7.1
Hydroxycinnamic acids								
CQA	20	24	2	81	28	21	5	67
PCQ	2	2	nd	5	3	3	0	12
CAA-1	4	3	1	12	5	3	1	11
CAA-2	1.1	0.7	0.3	2.4	0.4	0.6	nd	1.8
CMA-2	0.5	0.5	nd	2.0	0.5	0.7	nd	2.7
Dihydrochalcones								
PLXG	7	4	1	13	15	9	2	28
PLG	30	23	6	82	44	34	7	123
PLD-1	9	6	2	23	4	3	1	11
PLD-2	27	20	4	66	9	6	1	21
Flavonols								
HYP	23	17	1	53	34	20	4	68
IQC	7	5	1	18	9	6	2	24
AVI	18	13	5	49	21	14	9	52
QCI	9	6	2	22	11	8	3	29
QG-1	11	7	3	23	14	7	5	28
QG-2	1	1	0	6	2	1	0.4	6
QG-3	0.4	1	nd	4	0.5	1	nd	4
Anthocyanins								
IDE	1.1	2.2	nd	7.5	1.6	3.1	nd	9.7
CG-1	0.05	0.1	nd	0.4	0.1	0.1	nd	0.5
CG-2	nd				t		nd	t
CG-3	0.02	0.07	nd	0.27	0.03	0.06	nd	0.22
CG-4	0.02	0.07	nd	0.27	0.02	0.06	nd	0.21

level of 0.75, six clusters were found: clusters A, D, E and F contained unripe fruits, whereas clusters B and C were constituted of ripe samples. Two apple varieties (MNEM7 and PT) presented their unripe samples in a cluster of ripe fruits, and ripe samples of ER variety were inside cluster E of unripe fruits. These samples, that are in clusters not corresponding to their category, were situated in the overlapped region of both classes in a multidimensional plot of the samples in the space defined by the most discriminant variables (Fig. 2). In these plots, a natural separation between unripe and ripe apple fruits could be observed, these results being in accordance with those obtained by CA.

3.3.2. Supervised pattern recognition methods

3.3.2.1. General. LDA, KNN, SIMCA and a MLF neural network were applied to the autoscaled data matrix of each apple tissue, formed of unripe and ripe samples (85 pulp samples and 85 peel samples) and the most discriminant features (3 variables for pulp and 4 variables for peel), in order to achieve a prediction rule for classifying apple fruits according to their maturity state: ripe or unripe.

In KNN, the optimum number of neighbours (K) was studied. For both apple tissues, pulp and peel, the same results were attained for the K-values assayed (3, 5, 7 and 9); none of the samples was misclassified, therefore $K = 5$ was selected.

When using MLF-ANN with the purpose of making predictions, some empirical preliminary trials have to be performed in order to determine an adequate MLF-ANN structure. The MLF-ANN was applied to an input pattern consisting of the autoscaled data matrix. The neural architecture which gave better results then others was a MLF-ANN with three layers: for pulp, an input layer and one hidden layer with 3 neurons each, and an output layer consisting of a neuron with a binary output, and, for peel, an input layer with 4 neurons, one

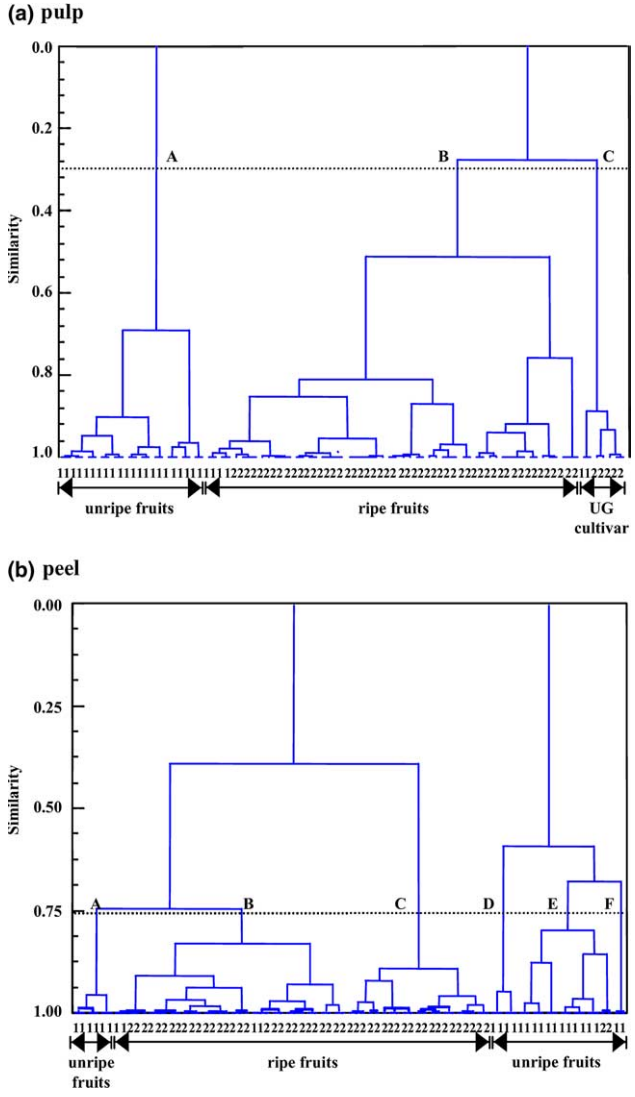

Fig. 1. Dendrograms of cluster analysis for apple data. Sample codes: 1, unripe fruits; 2, ripe fruits.

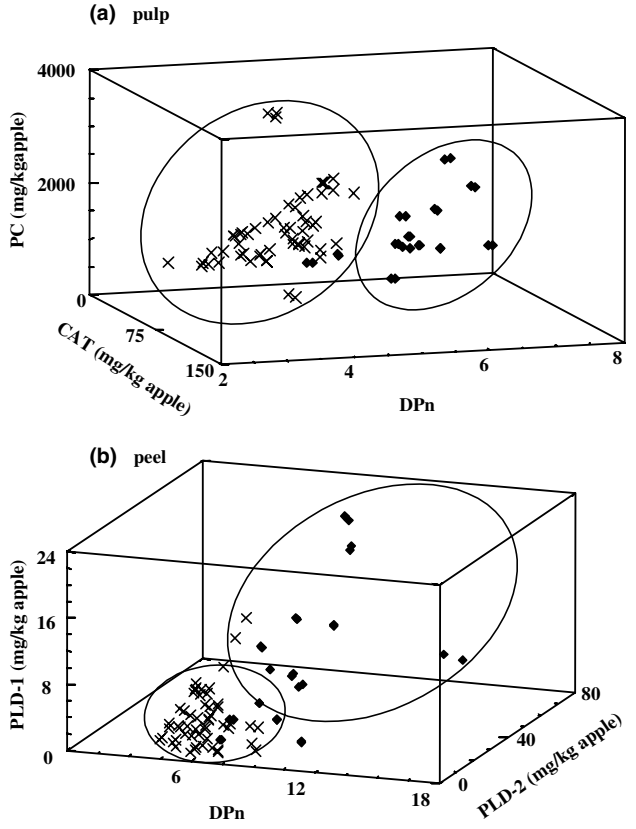

Fig. 2. Projection of apple samples on the multidimensional space defined by the most discriminant features between the two categories. Sample codes: ♦, unripe fruits; ×, ripe fruits.

hidden layer with 9 neurons, and a binary output neuron (Table 3).

Table 4 shows the recognition and prediction abilities afforded by each multivariate technique applied on the pulp and peel data, respectively.

3.3.2.2. Apple pulp. LDA and KNN achieved highly satisfactory classifications of ripe apples, with recognition and prediction abilities of 100%. For unripe apples, these abilities were slightly worse. Thus, the models proposed by these techniques were selective for ripe fruits; that is, the probability that ripe apples were classified as unripe was hardly nought. However, there existed a certain likelihood that unripe fruits were misclassified (8% in LDA and 14% in KNN). These results agreed with those obtained by CA, where some unripe samples were found inside clusters of ripe fruits. The classification accomplished by SIMCA achieved better results for the unripe pulps than for the ripe ones, the recognition and prediction abilities being 98.8% and 96.4%, and of 95.0% and 94.4%, respectively. The unripe model presented a sensitivity of 91% and a specificity of 92%, which meant that it accepted 91% of the unripe samples and 8% of the ripe fruits. The ripe model recognised 56%

Table 3
MLF-ANN architectures assayed and their prediction abilities for ripe and unripe apple fruits

Apple material	MLF-ANN architecture	Prediction ability (%)	RMSE
Pulp	3, 2, 1	92.7	0.04
	3, 3, 1	**97.8**	**0.02**
	3, 5, 1	97.1	0.02
	3, 3, 2	96.3	0.04
Peel	4, 3, 1	88.4	0.09
	4, 5, 1	88.4	0.08
	4, 7, 1	88.4	0.05
	4, 9, 1	**92.0**	**0.05**
	4, 11, 1	90.2	0.09
	4, 7, 2	91.1	0.09
	4, 9, 2	90.2	0.05

Table 4
Classification results for the supervised pattern recognition techniques applied to apple data

Technique	Class	Pulp		Peel	
		Recognition ability (%)	Prediction ability (%)	Recognition ability (%)	Prediction ability (%)
LDA	Unripe	93.1	92.0	93.8	90.3
	Ripe	100.0	100.0	95.1	91.3
KNN (K = 5); inverse squared Euclidean distance	Unripe	94.1	85.7	94.1	92.9
	Ripe	100.0	100.0	98.8	100.0
SIMCA; normal range; $\alpha = 0.05$	Unripe	98.8	96.4	94.1	92.3
	Ripe	95.0	94.4	88.8	79.2
MLF-ANN; $\eta = 0.2$; $\mu = 0.5$; sigmoidal transfer function[a]	Unripe	100.0	96.6	100.0	96.0
	Ripe	100.0	98.7	100.0	88.7

[a] MLF-ANN architecture: $(3 \times 3 \times 1)$ in pulp; $(4 \times 9 \times 1)$ in peel.

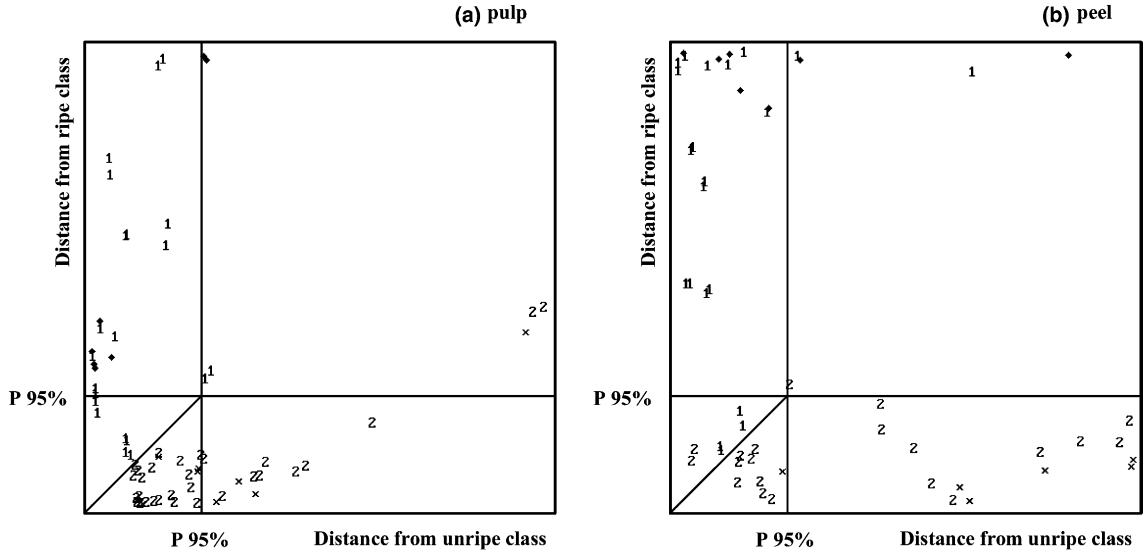

Fig. 3. Coomans plot for the squared SIMCA distances for apple data. Codes: Training set: 1, unripe class; 2, ripe class. Test set: ♦, unripe fruits; ×, ripe fruits.

of the ripe pulps and rejected 84% of unripe ones, these percentages being its sensitivity and specificity, respectively. Hence, there existed a probability (of 8% for ripe fruits and of 16% for unripe fruits) that they were wrongly classified by SIMCA. Fig. 3 represents SIMCA results as a Coomans plot for the squared SIMCA distances obtained from the data set. Classification results afforded by the neural network were excellent for both categories, the prediction abilities being above 96%.

3.3.2.3. Apple peel. LDA results obtained were similar for both categories, showing recognition and prediction percentages above 90%. Most samples misclassified were localised in the overlapped region of the two classes (Fig. 2). KNN achieved better classifications than LDA, presenting for ripe fruits, recognition and prediction abilities close to 100%. With both techniques, LDA and KNN, results were slightly better for the ripe category. Thus, the likelihood that ripe apples were classified as unripe was smaller than in the opposite case. These results tally with those of CA, where unripe fruits of two apple cultivars were present in a group of ripe fruits, and ripe fruits of one other cultivar was included in an unripe cluster. On the other hand, SIMCA afforded considerably more satisfactory results for unripe apples, with successes in recognition and prediction of 94.1% and 92.3%, than for ripe fruits, with 88.8% and 79.2%, respectively. In terms of sensitivity and specificity, SIMCA models for both categories presented a sensitivity of 93%, but the unripe model had a specificity of 63%, whereas the model for the ripe category was of 86%. Hence, the probabilities that SIMCA misclassified samples was considerably high for both categories (37% of ripe peels and 14% of unripe peels). Fig. 3 represents SIMCA results as a Coomans plot for the squared SIMCA distances. MLF-ANN achieved notably better predictions for the unripe fruits (96.0%) than for the ripe ones (88.7%).

When apples are processed to obtain a final product with adequate sensory and nutritional qualities, it is of great importance to ascertain that raw material is at the optimum condition of maturation, so that unripe apples are not used. To accomplish this, suitable tools that can classify unripe apples inside their category with high percentages of hits are required, thus minimising the risk of considering unripe fruits as ripe. In this particular case, the fact that ripe apples could be misclassified is not so relevant. In this sense, the chemometric technique that afforded the better results was the neural network in both apple tissues, presenting prediction abilities for unripe apples above 96% and implying little probability (4%) that unripe apple were mispredicted. These classification systems presented better prediction abilities than other decision rules previously reported (Mangas et al., 1998), which used sugars, organic acids, amino acids, total polyphenols, and pectins as chemical variables, and chemometrics in order to classify cider apples according their degree of ripening.

4. Conclusions

The results attained in this study allow us to conclude that certain polyphenols (procyanidins and (+)-catechin in pulp and dihydrochalcones, such as the xyloglucoside of phloretin and hydroxyphloretin glycosides in peel) and the average degree of polymerisation of procyanidins, jointly with several chemometric techniques, are appropriate for differentiating ripe and unripe apple fruits and for making predictions of the maturity state of apples.

The procedure proposed for establishing the maturity state (ripe/unripe) of the apple cultivars studied, involves the determination of the (+)-catechin concentration, the total procyanidin content and the average degree of polymerisation of procyanidins in apple pulp, and the use of the classification rule developed by the neural network for performing the prediction. This final proposal is due to the fact that apple peel composition depends, to a great extent, on climatology (Lea, 1990) and sun exposive of the fruit, causing differences according to the position of the fruit on the tree and even, in the same fruit, between sun-exposed parts and shaded parts (Awad, De Jager, & Van Westing, 2000). Hence, since pulp predictions were more homogeneous and did not depend so much on external factors, they were considered to be more reliable than those of peel.

Acknowledgements

This research was supported by Gobierno Vasco/Eusko Jaurlaritza (project number PI-1997-19 and PI-1999-106) and Universidad del País Vasco/Euskal Herriko Unibertsitatea (Project Number 171.310-EB013/98). The authors express their gratitude to the Diputación Foral de Gipuzkoa for providing apple samples. Rosa M[a] Alonso-Salces wishes to thank Gobierno Vasco/Eusko Jaurlaritza for a PhD Grant.

Appendix A

Abbreviations

AVI	avicularin
CQA	5-caffeoylquinic acid
CAA-1-2	unknown hydroxycinnamic acids with caffeic acid UV spectra
CAT	(+)-catechin
CAT-2	unknown flavan-3-ol
CG-1,-2,-3, -4	unknown anthocyanins
CMA-2	unknown hydroxycinnamic acid with p-coumaric acid UV spectra
DPn	average degree of polymerization of procyanidins
EC	(−)-epicatechin
HYP	hyperoside
IDE	ideain
IQC	isoquercitrin
PB2	procyanidin B2
PC	total procyanidins
PCQ	4-*p*-coumaroylquinic acid
PLD-1	hydroxyphloretin diglycoside
PLD-2	hydroxyphloretin monoglycoside
PLG	phloridzin
PLXG	phloretin-2′-*O*-xyloglucoside
QCI	quercitrin
QG-1-2-3	unknown flavonols
CA	Cluster analysis
KNN	K-nearest neighbours
LDA	linear discriminant analysis
MLF-ANN	multilayer feed-forward-artificial neural network
RMSE	root medium square error
SIMCA	soft independent modelling of class analogy
DAD	Diode array detector
HPLC	High Performance Liquid Chromatography
nd	not detected
t	traces
SD	standard deviation
BK	Bost Kantoi
ER	Errezila
GM	Geza Miña
GK	Goikoetxea
MN111	Manttoni 111
MNEM7	ManttoniEM7
MK	Moko

MZ	Mozoloa
PT	Patzuloa
TX	Txalaka
UM	Udare Marroi
UG	Ugarte
UH	Urtebi Haundia
UT	Urtebi Txiki

References

Ackermann, J., Fisher, M., & Amadò, R. (1992). Changes in sugars, acids, and amino acids during ripening and storage of apples (Cv. Glockenapfel). *Journal of Agricultural Food Chemistry, 40*, 1131–1134.

Alonso-Salces, R. M., Barranco, A., Corta, E., Berrueta, L. A., Gallo, B. & Vicente, F. (2004a). A validated solid–liquid extraction method for the HPLC determination of polyphenols in apple tissues. Comparison with pressurised liquid extraction. *Talanta*, in press.

Alonso-Salces, R. M., Herrero, C., Barranco, A., Berrueta, L. A., Gallo B. & Vicente, F. (2004b). Technological classification of Basque cider apple cultivars according to their polyphenolic profiles by pattern recognition analysis. *Journal of Agricultural Food Chemistry*, in press.

Alonso-Salces, R. M., Ndjoko, K., Queiroz, E. F., Ioset, J. R., Hostettmann, K., Berrueta, L. A., et al. (2004c). On-line characterisation of apple polyphenols by high performance liquid chromatography coupled with mass spectrometry and ultraviolet absorbance detection. *Journal of Chromatography A, 1046*, 89–100.

Amiot, M. J., Tacchini, M., Aubert, S., & Nicolas, J. (1992). Phenolic composition and browning susceptibility of various apple cultivars at maturity. *Journal of Food Science, 57*, 958–962.

Awad, M. A., De Jager, A., & Van Westing, L. M. (2000). Flavonoid and chlorogenic acid levels in apple fruit: characterisation of variation. *Science Horticulturae, 83*, 249–263.

Blanco, D., Morán, M. J., Gutiérrez, M. D., Moreno, J., Dapena, E., & Mangas, J. J. (1992a). Biochemical study of the ripening of cider apple varieties. *Zeitschrift für Lebensmitteluntersuchung und -Forschung, 194*, 33–37.

Blanco, D., Picinelli, L., Gutiérrez, M. D., & Mangas, J. J. (1992b). Determination of amino acids in ripening apples by high performance liquid chromatography. *Zeitschrift für Lebensmitteluntersuchung und -Forschung, 194*, 134–138.

Board, P. W., & Woods, H. J. (1983). Compositional variations and sensory acceptability of apple juice drink. *Journal of Food Technology, 18*, 763–769.

Burda, S., Oleszek, W., & Lee, C. Y. (1990). Phenolic compounds and their changes in apples during maturation and cold storage. *Journal of Agricultural Food Chemistry, 38*, 945–948.

Cowan, M. M. (1999). Plant products as anti-microbial agents. *Clinical Microbiology Reviews, 12*, 564–582.

Dilley, D. R. (1981). Assessing fruit maturity and ripening and techniques to delay ripening in storage. *Annual Report of Michigan State Horticultural Society, 110*, 132–146.

Forina, M., Lanteri, S. & Armanino, C. (2000). Q-Parvus 3.0. An extendable package of programs for data explorative analysis, classification and regression analysis. Dipartimento Chimica e Tecnologie Farmaceutiche ed Alimentari, University of Genova, Genova, Italy.

Guyot, S., Marnet, N., Sanoner, P., & Drilleau, J. F. (2001). Direct thiolysis on crude apple materials for the HPLC characterization and quantification of polyphenols in cider apple tissues and juices. *Methods in Enzymology, 335*, 57–70.

Lancaster, J. E. (1992). Regulation of skin color in apples. *Critical Reviews in Plant Science, 10*, 487–502.

Latorre, M. J., Peña, R., García, S., & Herrero, C. (2000). Authentification of Galician (N.W. Spain) honeys by multivariate techniques based on metal content data. *The Analyst, 125*, 307–312.

Le Lezec, M. & Babin, J. (1988). Test de regresion de l'almidon des pommes. In: *L'Arboriculture du Val de Loire, November 88* (pp. 18–19).

Lea, A. G. H. (1990). Bitterness and astringency: The procyanidins of fermented apple ciders. In R. L. Rousself (Ed.), *Bitterness in food and beverages* (pp. 123–143). Oxford: Elsevier.

Lea, A. G. H. (1995). Cidermaking. In A. G. H Lea & J. R. Piggott (Eds.), *Fermented beverage production* (pp. 66–96). London: Blackie Academic & Professional.

Macheix, J. J., Fleuriet, A., & Billot, J. (1990). Changes and metabolism of phenolic compounds in fruits. In *Fruit phenolics* (pp. 149–237). Boca Ratón: CRC Press.

Mangas, J. J., Dapena, E., Rodríguez, M. S., Moreno, J., Gutiérrez, M. D., & Blanco, D. (1992). Changes in pectic fractions during ripening of cider apples. *Hortscience, 27*, 328–330.

Mangas, J. J., Moreno, J., Picinelli, A., & Blanco, D. (1998). Characterization of cider apple fruits according to their degree of ripening. *Journal of Agricultural Food Chemistry, 46*, 4174–4178.

Massart, L., & Kaufman, L. (1983). *The interpretation of analytical chemical data by the use of clusters analysis*. New York: Wiley Interscience.

Mayr, U., Treutter, D., Santos-Buelga, C., Bauer, H., & Feucht, W. (1995). Developmental changes in the phenol concentrations of "Golden Delicious" apple fruits and leaves. *Phytochemistry, 38*, 1151–1155.

Mc Manus, J. P., Davis, K. G., Beart, J. E., Gaffney, S. H., Lilley, T. H., & Haslam, E. (1985). Polyphenol interactions. Part I. Introduction: some observations on the reversible complexation of polyphenols with proteins and polysaccharides. *Journal of the Chemical Society, Perkin Transactions II*, 1429–1438.

Meléndez, M., Ortíz, M. C., Sánchez, M., Sarabia, L., & Iñiguez, M. (1999). Chemometric characterization of the clarets and rose wines of the certified denomination of origin Rioja using Cielab parameters. *Química Analítica, 18*, 119–126.

Padín, P. M., Peña, R. M., García, S., Iglesias, R., Barro, S., & Herrero, C. (2001). Characterization of Galician (N.W. Spain) quality brand potatoes: a comparison study of several pattern recognition techniques. *The Analyst, 126*, 97–103.

Planton, G. (1995). Le point sur le test almidon des pommes pour l'aide à la décision de récolte. Août 1995, No. 6. Paris: Infos-CTIFL.

Renard, C. M. G. C., Baron, A., Guyot, S., & Drilleau, J. F. (2001). Interactions between apple cell walls and native apple polyphenols: quantification and some consequences. *International Journal of Biological Macromolecules, 29*, 115–125.

Ridgway, T., Oreilly, J., West, G., Tucker, G., & Wiseman, H. (1996). Potent antioxidant properties of novel apple-derived flavonoids with commercial potential as food additives. *Biochemical Society Transactions, 24*, S391–S391.

Sharaf, M., Illman, D., & Kowalski, B. R. (1986). *Chemometrics*. New York: John Wiley & Sons.

Siebert, K. J., Carrasco, A., & Lynn, P. Y. (1996). Formation of protein-polyphenol haze in beverages. *Journal of Agricultural Food Chemistry, 44*, 1997–2005.

Sieguist, R. (1987). Dates de récolte des pommes de garde: 'Jonalgold', 'Gloster', 'Jonnee', et 'Idared'. *Revue Suisse de Viticulture, Arboriculture, Horticulture* 19, 295–298.

Sponholtz, W. R. (1993). Wine spoilage by micro-organisms. In G. H. Fleet (Ed.), *Wine microbiology and biotechnology* (pp. 397–399). USA: Harwood Academic Publishers.

SPSS for Windows, version 9.0, SPSS Inc., ©1989–1999.

Statgraphics Plus 5.0, Statistical Graphics Corporation, ©1994–2000.

Treutter, D. (2001). Byosinthesis of phenolic compounds and its regulation in apple. *Plant Growth Regulation, 34*, 71–89.

Trillot, M., Masseron, A., & Tronel, C. (1993). Paris: CTIFL-INRA, p. 202.

WinNN32 1.2, Y. Danon, ©1993–1996.