JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

High-Resolution MAS NMR and Chemometrics: Characterization of the Ripening of Parmigiano Reggiano Cheese

LAETITIA SHINTU* AND STEFANO CALDARELLI

JE 2421 TRACES Université Aix Marseille I and III, Site de Saint Jérôme, Service 511, 13013 Marseille, France

We performed a preliminary NMR investigation on grated Parmigiano Reggiano cheese of different ripening ages. Principal component analysis (PCA) and discriminant analysis (DA) were able to successfully group the analyzed samples according to their respective ages.

KEYWORDS: HRMAS NMR; Parmigiano Reggiano cheese; chemometrics; PCA; DA

INTRODUCTION

Analytical techniques and chemometrics are increasingly being employed for the characterization of varieties, quality, or geographical origin of foodstuffs (*1*). In the dairy area, they were applied to discriminate dairy products (*2*, *3*) and to study cheese origin (*4*−*6*) and cheese quality (*7*−*10*) or to characterize the cheese ripening (*11*). And the ripening is a core step in the manufacture of cheese and more particularly so in the case of Parmigiano Reggiano cheese. The organoleptic quality of Parmigiano cheese and thus its price are strongly associated with a long time of ripening. It is thus particularly valuable to well describe the chemical characteristics of Parmigiano cheese at a particular age of ripening to prevent fraudulent claims or adulteration of the cheese quality. However, most of chemical analyses used for cheese study are time-consuming and specific to a given sort of compound. In a previous study, we showed that high-resolution magic angle spinning (HRMAS) NMR spectroscopy (*12*, *13*), although limited in sensitivity, enables us to assign the major constituents in Parmigiano cheese without requiring extraction or other selective steps in the sample preparation. In this contribution, we propose to test the potentiality of the HRMAS NMR spectroscopy combined with multivariate statistics (principal component analysis and discriminant analysis) for the discrimination of the Parmigiano Reggiano cheese age of ripening. The method proposed in this preliminary study (only 15 samples) is a fast one that, while taking into account only the major compounds present in the samples, succeeded in achieving this aim.

MATERIALS AND METHODS

Samples. The samples of grated Parmigiano Reggiano cheese were obtained from the "Consorzio del Formaggio Parmigiano Reggiano". We studied 15 samples with different ages of ripening. The number of samples and experiments for each age of ripening are reported in **Table 1**.

* To whom correspondence should be addressed. E-mail: l.shintu@univ.u-3mrs.fr.

Table 1. Distribution of the Samples According to Their Age of Ripening

no. samples	age of ripening (months)	no. experiments
3 (a, b, c)	4	10
3 (a, b, c)	8	9
3 (a, b, c)	12	11
2 (a, b)	18	7
4 (a, b, c, d)	24	14

The samples were stored at 4 °C until analysis. For each sample, about 20 mg of grated Parmigiano Reggiano cheese were packed into a 4-mm MAS rotor of 50 μL total volume. To stabilize the pH at 6.70 (and thus prevent a pH-dependent variation of some signals) were added 43 μL of 1 M pH 7 phosphate buffer (61 mg of NaH_2PO_4 and 82 mg of Na_2HPO_4 in 1 mL of D_2O) with DSS (2,2-dimethyl-2-silapentane-5-sulfonate sodium salt) as an internal reference.. This mixture was homogenized directly in the rotor with a spatula.

Proton HRMAS NMR Spectroscopy. All proton HRMAS NMR spectra were recorded on a Bruker Avance 400 instrument operating at 400.13 MHz for ^{1}H and using a 4-mm HRMAS ^{1}H/^{13}C probehead. The samples were spun at 5000 Hz. All spectra were performed with presaturation of the water peak (*14*). A typical proton HRMAS NMR spectrum consisted of 1024 transients using 16K data points over a 4800 Hz spectral width with a 7 μs 90° pulse. A recycle time of 2 s was used. Each spectrum was phased and linearly baseline corrected.

Data Pretreatment. For the sake of simplification of the statistical analysis, the frequency region between 0.5 and 9 ppm of each of the 51 spectra (three or four replicates per sample) was split into integrated zones of identical width ("buckets") using the software Mestrec-C v. 3.6.9 (*15*). This procedure greatly reduced the computational nuisances arising from variable redundancy and pseudovariations. The optimal simplification we selected was of 162 regions of 0.05 ppm each. Furthermore, the range corresponding to the solvent residual peak, between 4.70 and 5.10 ppm, was removed to eliminate the variability due to the reproducibility of the water suppression schemes. The reduced datasets were arranged in the form of a 51 × 162 matrix in which each row was an experiment and each column a "bucket". Previous spectral assignment allowed linking each bucket content to a specific compound signal (*16*). The data matrix was imported into STATISTICA v. 6 software (Statsoft Inc) for statistical analysis.

10.1021/jf048141y CCC: $30.25
Published on Web 04/19/2005

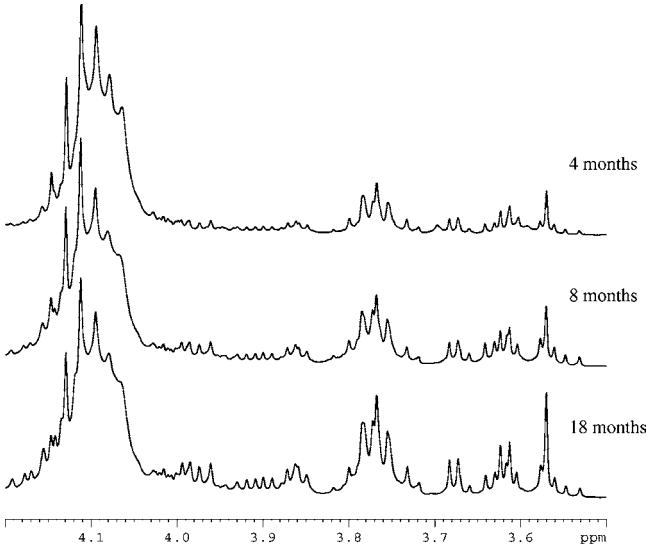

Figure 1. Expansion of ^{1}H HRMAS NMR spectra of samples of 4, 8, and 18 months aging.

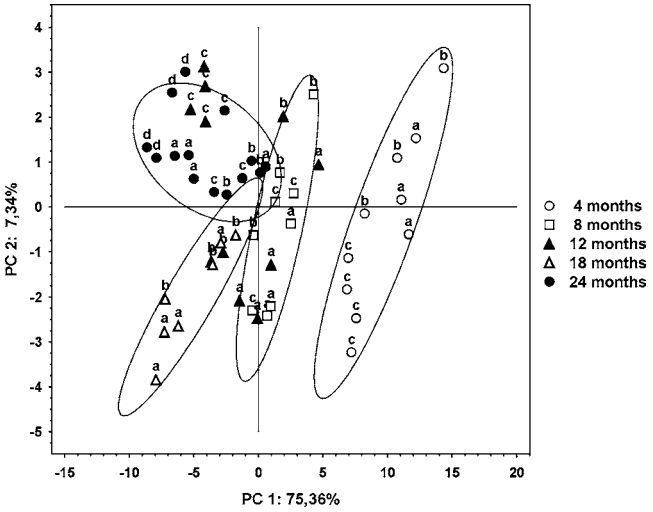

Figure 2. Projections of sample replicates of the five groups on the PC1 vs PC2 plane. (a–d) Different samples in each group.

Chemometrics. The most significant variables were selected with ANOVA, with a *p* level smaller than 0.01 and an *F* factor greater than 9 to minimize signal variations due to spurious spectral sources (baseline or noise instabilities). The selected variables were submitted to a PCA for further reduction. The group separation achieved by this method was not sufficient to completely distinguish all sets of ages. Therefore, the first 10 variables with the highest discrimination power were used for a DA (*17*), in which a new system of (canonical) axes are calculated to maximize the between-group distances, while minimizing those of the within-group. These variables were selected through a stepwise DA performed on the centered variables.

The following procedures were applied to validate the robustness of the results. A first validation was achieved by dividing the collected data in a test set (10 cases) and a training set (the rest of the data). This latter was used to calculate classification functions, which in turn served to estimate the age group of the elements of the test set. To minimize the risk of overfitting due to the presence of replicates in both training and test sets, we assigned a significant fraction of our data collection (10 observations) to the test set, and the procedure was repeated five times, so that each case joined a test set at least once. A case was considered correctly classified if it had a classification probability greater than 85% for the expected group. Conversely, it was considered misclassified if it had a classification probability greater than 85% for an unexpected group. An undetermined classification corresponded to a classification probability smaller than 85% for expected or unexpected groups (*18*). We also proceeded to a "leave-one-out" cross-validation in which all replicates of one sample were removed, and classification was redone according to the same procedure explained above.

RESULTS AND DISCUSSION

A portion of the ^{1}H HRMAS spectra of three samples of different ages (4, 8, and 18 months) of Parmigiano Reggiano cheese is shown in **Figure 1**. It is already clear by inspection that the variations of certain signals were appreciable, so that statistical methods should be able to discriminate among the different samples.

One-way ANOVA was performed on the whole data set (51 observations vs 162 variables), the grouping factor being the age of ripening of each sample. According to the results of the analysis of variance, 68 variables were statistically significant, in agreement with the criteria highlighted in the previous section. These selected variables constituted the initial dataset for the PCA based on the correlation matrix. The projections of the 51 observations on the first two principal component axes are shown in **Figure 2**.

The 4-, 8-, 18-, and 24-month groups are well separated on the PC1 × PC2 factor plane. These two coordinates represent 82.7 % of the total variance. The 4- and 8-month groups have a positive contribution on PC1 and are well distributed along the PC2 direction. On the other hand, both the 18- and 24-month groups have negative contributions on PC1 but opposite sign contributions on PC2 (negative for the 18-month group and positive for the 24-month group). The 12-month group experiments were discriminated neither by the principal PC1 × PC2 plane nor by any of the others. Although not being completely effective in our goal of age discrimination, PCA remains a good starting point to understand the metabolic issues relevant for the discrimination of the different Parmigiano Reggiano cheese samples. For instance, the analysis of the PC1 and PC2 loadings (**Table 2**) shows that the first principal component is mainly negatively correlated with the amino acids regions (coefficients from −0.71 to −0.99). It was thus possible to conclude that the 18- and 24-month groups (negative scores on PC1 axis) have a higher amount of amino acids than the 4-, 8-month groups (positive scores on PC1 axis). The second principal component is mainly associated with the citrulline and tyrosine region (−0.66), unknown compounds between 8.20 and 8.35 ppm (not shown) (from −0.67 to −0.71), and other unknown compounds between 4.55 and 4.60 ppm, labeled B (**Figure 3**). This set of molecules is higher in concentration for the 18-month group compared to that for the 24-month group.

We can also notice that the replicated data points do not gather in the variate space. This means that, from the point of view of our method, they behave almost as independent measures. This behavior is probably due to a combination of limitations in experimental reproducibility and true sample heterogeneity effects. This latter would not be surprising considering the process of cheese making and the fact that the samples treated are solids. It is not possible to push the analysis any further with the limited dataset at our disposal, but this point should be investigated more deeply for the sake of a complete analysis of cheese or other solid samples. For the scope of this work, the important issue is that the ellipsoid of probability for each age group is fairly well covered by the data points at our disposal. However, it is clear that the PCA is not satisfactory for the large number of variables employed and its incapacity to distinguish the 12-month-old samples.

Table 2. Coefficients of the Correlations between Initial Variables and Principal Components

compounds	PC1 loading	PC2 loading	compounds	PC1 loading	PC2 loading
unknown signals	0.54	–0.68	proline + phenylalanine	–0.97	0.00
unknown signals	0.64	–0.67	tyrosine	–0.78	–0.47
unknown signals	0.51	–0.71	citrulline + tyrosine	–0.21	–0.66
unknown signal A	-0.61	–0.17	tyrosine	–0.87	–0.02
phenylalanine	–0.94	0.08	tyrosine	–0.71	0.40
phenylalanine	–0.93	0.00	unknown signal C	–0.71	0.04
phenylalanine	–0.94	0.02	lysine	–0.96	0.09
pyridine	–0.91	–0.12	asparagine	–0.73	0.08
unknown signal B	0.48	–0.64	methionine	–0.90	–0.34
pyroglutamic acid	–0.87	–0.15	pyroglutamic acid	–0.89	–0.19
proline + lactate + FA[a]	–0.77	0.00	glutamic acid	–0.97	–0.12
asparagine + phenylalanine	–0.91	–0.15	Pyro + Glu + unknown signals + FA[a]	–0.97	–0.02
serine	–0.96	0.09	Pyro + Glu + FA[a]	–0.99	–0.02
aspartic acid + methionine	–0.92	–0.16	Glu + FA[a]	–0.92	–0.12
aspartic acid + methionine	–0.97	0.01	methionine + FA[a]	–0.96	–0.16
methionine + alanine	–0.96	0.04	methionine + proline + FA[a]	–0.80	–0.15
lysine + glutamic acid	–0.97	0.10	lysine + FA[a]	–0.89	–0.15
leucine	–0.94	0.03	lysine + FA[a]	–0.94	0.02
isoleucine	–0.88	0.15	lysine	–0.82	–0.35
valine + threonine	-0.93	0.20	lysine	–0.98	–0.10
glycine + unknown signals	–0.92	0.23	alanine + FA[a]	–0.85	–0.20
proline	–0.98	0.02	valine + isoleucine	–0.99	–0.02
proline	–0.97	0.07	valine + isoleucine	–0.95	–0.08

[a] Abbreviations: Pyro = Pyroglutamic acid; Glu = Glutamic acid; FA = Fatty acid.

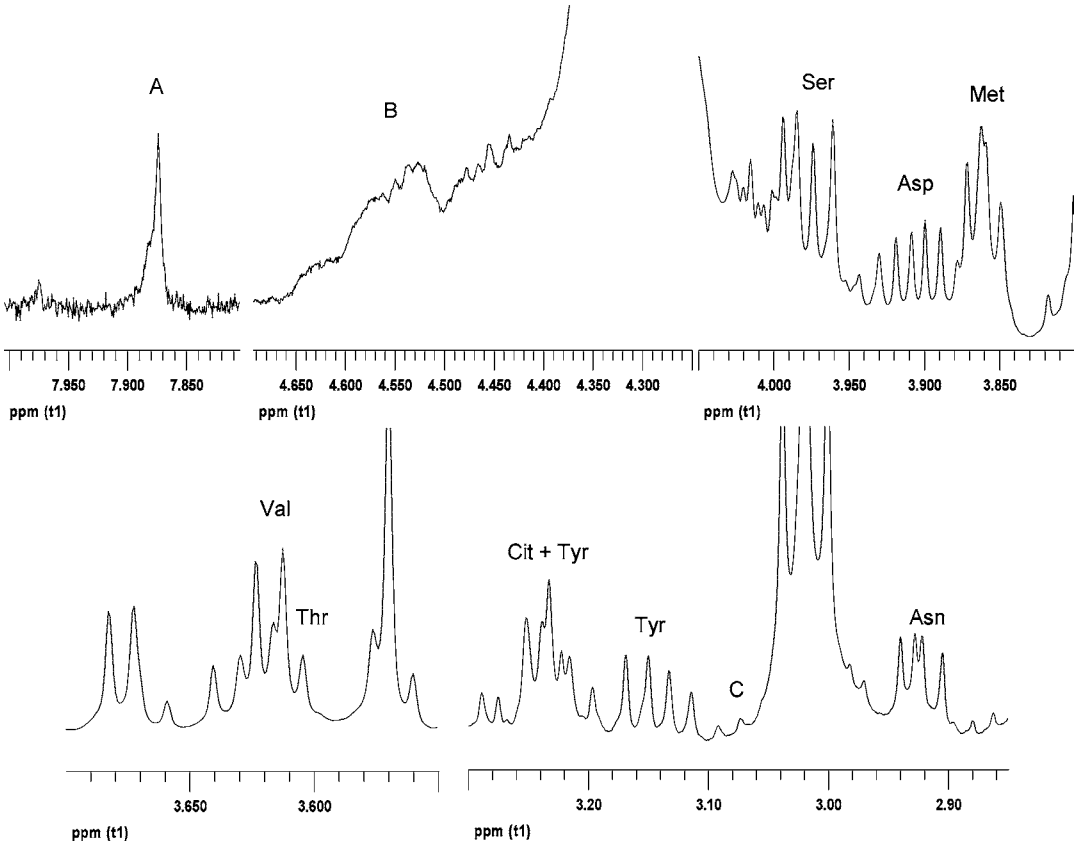

Figure 3. Regions on the spectrum corresponding to the 10 selected variables.

Therefore, we performed a stepwise DA on the 46 variables previously selected. The most discriminant variables were selected according to their Wilk's lambda factors, the smaller ones being retained. In our work, we kept 10 variables of the original 46 (**Table 3**). All of these variables had a very low Wilk's lambda factor and were statistically significant (p level < 0.05).

The regions on the spectrum corresponding to these variables are shown in **Figure 3**.

The discriminant model succeeded in accurately grouping 100% of the experiments (i.e., including replicates). We tested the robustness of these classification functions as illustrated in the methods section. The five test sets used are shown in **Table 4**. For each test set and corresponding training set we calculated

Table 3. Wilk's Lambda Factor and *p* Level for the Selected Discriminant Variables

	Wilk's lambda	*p* level
unknown compound A	0.00105	0.04588
unknown compound B	0.00118	0.00680
serine	0.00177	0.00001
aspartic acid + methionine	0.00111	0.01839
aspartic acid + methionine	0.00132	0.00093
valine + threonine	0.00257	0.00000
citrulline + tyrosine	0.00165	0.00002
unknown compound C	0.00237	0.00000
tyrosine	0.00148	0.00013
asparagine	0.00193	0.00000

Table 4. Distribution of Experiments in the Test Sets

	months of aging						months of aging				
test set	4	8	12	18	24	test set	4	8	12	18	24
1	2	3	2	1	2	4	2	2	3	1	2
2	2	1	2	2	3	5	2	2	1	1	4
3	2	1	3	1	3						

Table 5. Results (in %) of the Classification of Cases Following the Dataset and the Classification Category

set	correct classification	indeterminate classification	misclassification
1	100.0	0.0	0.0
2	98.0	2.0	0.0
3	100.0	0.0	0.0
4	94.2	3.9	1.9
5	100.0	0.0	0.0
mean	98.4	1.2	0.4

Table 6. Squared Mahalanobis Distances from Group Centroids

	4 months	8 months	12 months	18 months	24 months
4 months	0.00	60.29	97.04	276.87	173.89
8 months	60.29	0.00	21.63	90.94	58.97
12 months	97.04	21.63	0.00	80.39	62.68
18 months	276.87	90.94	80.39	0.00	60.98
24 months	173.89	58.97	62.68	60.98	0.00

the number of the correct, undetermined, and incorrect classified observations. **Table 5** shows the classification percentage of the cases according to their classification category.

Our first validation model enabled us to classify correctly 98.4% of all experiments on average. With the "leave-one-out" cross-validation, we obtained 93.3% of correct classification of the samples. Thus, we considered that the selected variables were robust to validate the DA. To further explore the effect of the temporal evolution in statistical terms, the squared Mahalanobis distances were calculated (**Table 6**).

In this metrics, the 8- and 12-month groups were the nearest ones (21.63), whereas the 4- and 18-month groups were the farthest away (276.87).

Once a protocol for group discrimination has been found, it is possible to translate it into relevant metabolic markers by detailing the composition of the new coordinates. To know which variable transformation was required to achieve a discrimination of the groups, we further performed a canonical analysis. The projections of experiment scores on the first two canonical axes and the first and third canonical axes are shown in **Figures 4** and **5**, respectively.

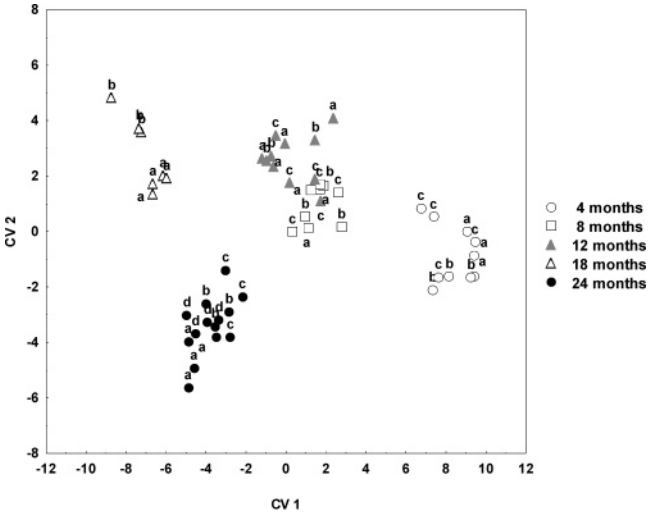

Figure 4. Projections of cases of 4, 8, 12, 18, and 24 months on the CV1 vs CV2 plane (a–d) Different samples in each group.

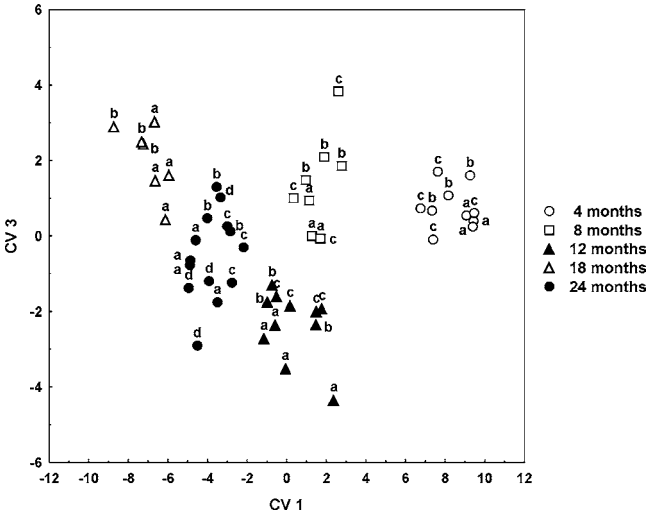

Figure 5. Projections of cases of 4, 8, 12, 18, and 24 months on the CV1 vs CV3 plane (a–d) Different samples in each group.

On the CV1 vs CV2 plane (**Figure 4**), we observed that the 4-, 8-, and 12-month groups and the 18- and 24-month groups were separated by the first canonical variable, with positive and negative scores, respectively. We noted also the location of the groups along CV1 axis follows the increase of the time of ripening until 18 months ripening. The second canonical axis enabled us to discriminate the 24-month groups (negative scores) and the 8-, 12-, 18-month groups (positive scores). Finally, the third canonical axis (**Figure 5**) separated the 4-, 8-, and 18-month groups (positive scores) and the 12-month group (negative scores). A graphical representation of the correlation between the canonical axis and all variables are shown in **Figures 6** and **7**.

Aspartic acid, methionine, and serine were the most (negatively) correlated variables with CV1. Thus the 18- and 24-month groups have higher contents of these amino acids than the other three groups. We can infer that the concentration of these amino acids increases until 18 months of aging and successively it decreases between 18 months and 24 months of aging, in agreement with previous studies (*19*). Citrulline and tyrosine are positively correlated with CV2, implying a concentration of these metabolites lower for the 24 month groups compared to the others. Asparagine has the most significative (and negative) correlation with CV3, which is relevant for the

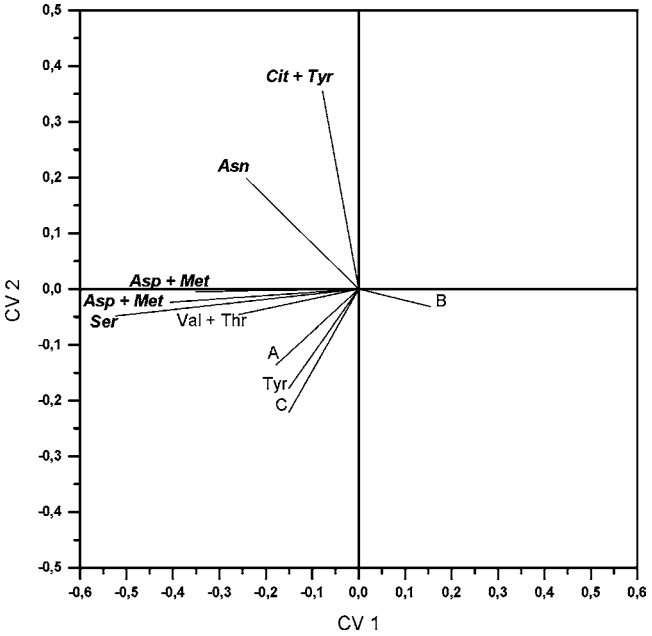

Figure 6. Graphical representation of the coefficients of correlation of the 10 variables with CV1 and CV2.

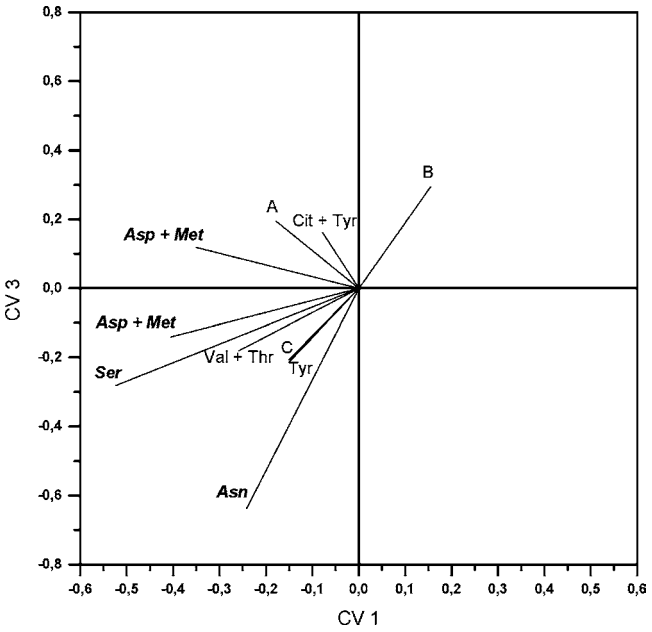

Figure 7. Graphical representation of the coefficients of correlation of the 10 variables with CV1 and CV3.

discrimination of the 12-month group compared to the 4-, 8-, and 18-month groups. On the other hand, CV3 was also positively correlated with compound B at 4.55−4.60 ppm, which was present in higher content for 4-, 8-, 18-month groups than in 12-month group.

In conclusion, HRMAS proton NMR provides a quick tool, despite its limited sensitivity, for characterization of the age of Parmigiano Reggiano cheese. Common statistical tools (PCA and DA) allowed successful grouping of the analyzed samples. In this way, selected free amino acids and other low molecular weight metabolites were found to be among the most relevant compounds characterizing the ripening of Parmigiano Reggiano cheese. This provides useful insight into the metabolic pathway evolution with time. Obviously, these results maintain an exploratory character, due to the rather reduced data set size. However, on the basis of our results of the validation procedures, it can be legitimately hoped that the classification functions derived this way can be extended to describe samples of unknown age of Parmigiano Reggiano cheese. Other kinds of cheese can be promptly submitted to the protocol described here.

ACKNOWLEDGMENT

We are indebted to the Consorzio del Parmigiano Reggiano (Dr. Panari and Dr. Pecorari) for a gift of the sample used in this paper.

LITERATURE CITED

(1) Tzouros, N. E.; Arvanitoyannis, I. S. Agricultural Produces: Synopsis of Employed Quality Control Methods for the Authentication of Foods and Application of Chemometrics for the Classification of Foods According to Their Variety or Geographical Origin. *Crit. Rev. Food Sci. Nutr.* **2001**, *41*, 287−319.

(2) Molher-Smith, A.; Nakai, S. Classification of cheese varieties by multivariate analysis of HPLC profiles. *Can. Inst. Food Sci. Technol. J.* **1990**, *23*, 53−58.

(3) Karoui, R.; Mazerolles, G.; Dufour, E. Spectroscopic techniques coupled with chemometric tools for structure and texture determinations in dairy products. *Int. Dairy J.* **2003**, *13*, 607−620.

(4) Brescia, M. A.; Monfreda, M.; Buccolieri, A.; Carrino, C. Characterisation of the geographical origin of buffalo milk and mozzarella cheese by means of analytical and spectroscopic determinations. *Food Chem.* **2005**, *89*, 139−147.

(5) Noël, Y.; Ardö, Y.; Pochet, S.; Hunter, A.; Lavanchy, P.; Luginbühl, W.; Le Bars, D.; Polychroniadou, A.; Pellegrino, L. Characterisation of protected denomination of origin cheeses: Relationship between sensory texture and instrumental data. *Lait* **1998**, *78*, 569−588.

(6) Pillonel, L.; Tabacchi, R.; Bosset, J. O. Analytical Methods for the Determination of the Geographic Origin of Emmental cheese. Summary of a Screening Study. *Mitt. Lebensmittelunters. Hyg.* **2003**, *94*, 60−69.

(7) Pinho, O.; Ferreira, I. M. P. L. V. O.; Ferreira, M. Discriminate analysis of the volatile fraction from "Terrincho" ewe cheese: correlation with flavour characteristics. *Int. Dairy J.* **2004**, *14*, 455−464.

(8) O'Shea, B. A.; Uniacke-Lowe, T.; Fox, P. F. Objective assessment of cheddar cheese quality. *Int. Dairy J.* **1996**, *6*, 1135−1147.

(9) Gennaro, M. C.; Gianotti, V.; Marengo, E.; Pattono, D.; Turi, R. M. A chemometric investigation of the effect of the cheese-making process on contents of biogenic amines in a semi-hard Italian cheese (Toma). *Food Chem.* **2003**, *82*, 545−551.

(10) Pelaez Puerto, P.; Fresno Baquero, M.; Rodriguez Rodriguez, E. M.; Darias Martin, J.; Diaz Romero, C. Chemometric studies of fresh and semi-hard goats' cheeses produced in Tenerife (Canary Islands). *Food Chem.* **2004**, *88*, 361−366.

(11) Carbonell, M.; Nuñez, M.; Fernández-García, E. Evolution of the volatile components of ewe raw milk le Serena cheese during ripening. Correlation with flavour characteristics. *Lait* **2002**, *82*, 683−698.

(12) Fitch, W. L.; Detre, G.; Shoolery, J. N.; Keifer, P. A. High Resolution ^{1}H NMR in Solid-Phase Organic Synthesis. *J. Org. Chem.* **1994**, *59*, 7955−7956.

(13) Lippens, G.; Bourdonneau, M.; Dhalluin, C.; Warras, R.; Richert, T.; Setharaman, C.; Boutillon, C.; Piotto, M. Study of Compounds Attached to Solid Supports Using High-Resolution Magic Angle Spinning NMR. *Curr. Org. Chem.* **1999**, *3*, 147−170.

(14) Guéron, M.; Plateaux, P.; Decorps, M. Solvent signal suppression in NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **1991**, *23*, 135−209.

(15) http://www.mestrec.com/.

(16) Shintu, L.; Ziarelli, F.; Caldarelli, S. Is high-resolution magic angle spinning NMR a practical speciation tool for cheese samples? Parmigiano Reggiano as a case study. *Magn. Reson. Chem.* **2004**, *42*, 396−401.

(17) Marcotte, D. *Cours de Traitement Statistique des Données Géologiques* (*Courses of Statistical Geological Data Processing*); Chapters III and V. **2003/05/15**, *2004*, 40.

(18) Bourne, R.; Himmelreich, U.; Sharma, A.; Mountford, C.; Sorrell, T. Identification of Enterococcus, Streptococcus, and Staphylococcus by multivariate analysis of proton magnetic resonance spectroscopic data from plate cultures. *J. Clin. Microbiol.* **2001**, *39*, 2916−2923.

(19) Resmini, P.; Pellegrino, L.; Hogenboom, J.; Bertuccioli, M. Gli amminoacidi liberi nel formaggio Parmigiano-Reggiano stagionato (Free amino acids in seasoned Parmigiano Reggiano cheese). *Ricerca triennale sulla composizione e su alcune peculiari caratteristiche del formaggio Parmigiano-Reggiano* **1988**, 41−57.

Received for review November 8, 2004. Revised manuscript received March 18, 2005. Accepted March 21, 2005.

JF048141Y