Determination of fatty acid composition in milk of individual animals by Fourier-Transform Mid-Infrared Spectrometry

M. Ferrand[1], V. Huquet[1], M.R. Aurel[2], S, Barbey[3], F. Barillet[4], F. Bouvier[5], P. Brunschwig[1], K. Duhem [7], F. Faucon[1], D. Gueldry[8], H. Larroque[6], G. Lautier[9], O Leray[10], S. Leverrier[11], I. Palhiere[4], J.L. Peyraud[12], J.M. Trommenschlager[13] & M. Brochard[1]

[1]*Institut de l'Elevage, 149 rue de Bercy, 75595 Paris cedex 12, France*
[2]*UE INRA La Fage*
[3]*UE INRA du Pin-Au-Haras*
[4]*INRA-SAGA*
[5]*UE INRA de Bourges*
[6]*INRA-GABI*
[7]*CNIEL, 42 rue de Châteaudun, 75314 Paris cedex 09, France*
[8]*LIAL MC Aurillac*
[9]*LILCO Surgères*
[10]*Actilait,* [11] *LiLano*
[12]*INRA UMR-PL*
[13]*UR-ASTER INRA Mirecourt*

Abstract

The increasing social demand for healthy products leads more and more dairy companies to select collected herd milks according to the fine milkfat composition and could also lead to the introduction of the composition in fatty acid (FA) as a criterion for milk payment. However, today, there is neither rapid method officially validated in France to determine milk fatty acid composition in routine analysis nor tools to allow adaptation of the fine milkfat composition to the evolving consumers demand.

Consequently it has become today of a major interest to define technical levers that will allow milk producers to orient milk fatty acid profiles as soon as the production stage on-farm. Since then the objectives to measure elementary milk components with sufficient accuracy and to identify the genetic and environmental factors affecting the composition are being pursued through a R&D project, initiated in France in 2008, called PhenoFinLait.

The first step was to develop a cheap and large scale phenotyping system for determination of FA individual milk content. A set of equations was developed to estimate fine FA milk composition from MIR (Mid Infra-Red) spectra usually obtained by milk recording laboratory. For several FA, a variable selection was applied to improve the equations. In the end, 15 to 20 FA are well estimated in the three ruminant species (cow, sheep and goat). Statistical research is ongoing to improve estimation for other FA and normalize this method.

This study is part of PhenoFinlait project funded by Apis-Gène, CASDAR, CNIEL, FranceAgriMer, France Génétique Elevage, French Ministry of Agriculture and French Ministry of Research.

Keywords: cow, ewe, goat, milk, fatty acid, mid-infrared (MIR) spectrometry, genetic algorithms, Partial Least Squares (PLS) regression

1.0 Materiel and methods

1.1 Milk samples

1.1.1 Cow milk samples

First, 154 milk samples from 77 crossbred Holstein X Normande dairy cows were collected in 2008 at the experimental Pin-Au-Haras INRA farm. The cows were a part of a QTL (Quantitative Trait Loci) detection

experiment, and they arose from a cross over two generations (F2) between Normande and Holstein breeds that display numerous differences in particular for milk fat and protein contents.

Milk samples were collected twice, in winter and in summer to take into account the possible effect of feed. Cows were in an average stage of lactation of 160 days (77-235) in winter, and 209 days (126-284) in summer. Each time and for each cow, two milk samples were realized during the morning milking. One was analyzed freshly using MIR spectrometry, the other was frozen at -20°C and analyzed using gas chromatography. Finally, on 154 samples, 150 milks were kept in the study due to missing data.

Secondly, 153 milk samples from 54 Montbéliarde and 42 Prim'Holstein were collected in 2009 at the experimental Mirecourt INRA farm. Depending on calving date, the cows belong either to a grazing system or to a mixed crop dairy system. Milk samples were collected using the same protocol as above, but only a part of the cows were present at both sampling. On 153 samples, 100 milks were analysed by gas chromatography.

1.1.2 Goat milk samples

705 milk samples from 235 Alpine dairy goats were collected in 2008 at the INRA experimental farm of Bourges at three stage of lactation (about 40, 150 and 240 days). The goat's diet was almost similar throughout lactation and was based on grass hay offered ad libitum and a commercial concentrate mixture. These samples were collected in tubes containing a preservative (Bronopol).

For each goat, one sample was analysed by MIR spectrometry, and one other was frozen at -20° C. Among them, 149 samples (about 50 per stage of lactation) with a large variability of spectra were selected to be analysed for milk fatty acid composition by the referenced method.

1.1.3 Ewe milk samples

A first sampling was carried out twice in 2008 in the experimental La Fage INRA farm: milk samples were collected from Lacaune dairy ewes, respectively on March 2008 at 80 days in milk (DIM) for 490 ewes in winter diet (hay, silage and concentrates), and on May 2008 at 152 DIM on average for 493 ewes in spring diet including pastures. At each sampling carried out at the morning milking, 2 milk samples were collected, the first fresh one to be analyzed without delay to provide MIR spectra and the second one to be frozen at-20°C for a possible reference gas chromatography carried out later.

Accounting for somatic cell count, fat content and milk spectra, 75 milk samples were chosen within each sampling period, i.e a total of 150 frozen milk samples to be analyzed by gas chromatography.

A second sampling, using the same design described above, was performed in 2009 in 3 private flocks, the first one composed of Basco-Bearnaise (BB) ewes, and the two other of Manech red faced (MRF) ewes: a total of 103 milk samples, respectively 35 from BB ewes and 78 from MRF ewes, were collected by the end of April 2009 at 120 DIM on average in pasture diet condition. Accounting for fat content, milk spectra and breed, 50 milk samples were chosen to be analyzed by gas chromatography (respectively 20 and 30 for BB and MRF breed).

Finally 200 milks from Lacaune ewes (150 samplings) or from BB or MRF ewes (50 samplings) with both milk spectra and gas chromatography results were included in the present analysis carried out in dairy sheep.

1.2 MIR spectra

After a transport at 4°C to the laboratory (LILANO of St Lo, LILCO of Surgères and LIAL of Aurillac), fresh milk samples were analyzed for milk spectra extraction using MIR spectrometry with defined routine FT-MIR analyzers (Milkoscan FT6000, Foss and Bentley FTS). Spectra have been recorded from 5012 to 926 cm-1. According to Foss (1998), only informative wavelength bands, i.e. bands not spoiled by water molecule, were kept (representing a total of 446 wavelengths). No pre-treatments were applied as suggested by Soyeurt *et al.* (2006).

1.3 Fatty acid composition

Frozen milk samples were analyzed for milk fatty acid composition using gas chromatography according to ISO standards (Kramer, 1997). Quantities of 64 fatty acids were expressed in g/100mL. Outliers were removed by Grubb's test as indicated in the norm ISO 8196.

1.4 Calculation of calibration equations

MIR spectra and milk fatty acid composition of samples presenting a large variability in their composition were retained to calculate the equations. For cow and ewe milk, the samples were divided into calibration and validation sets (cow milk: $n_{calibration}$=175 and $n_{validation}$=75, ewe milk: $n_{calibration}$=140 and $n_{validation}$=60).

The equations were developed by univariate and multivariate PLS regression (Tennehaus, 2002), data being centered but not reduced according to Bertrand *et al.* (2006). For each equation, optimal number of latent variables was chosen according to root mean square error of cross-validation ($RMSEP_{cv}$). To improve equations and quality of estimation, a selection of wavelengths by genetical gorithm was performed before PLS regression in cow and goat milk (Ferrand, 2010). The genetic algorithm used is the algorithm developed by Leardi (1998) which is specific to wavelengths selection. Mutation rate, initial population, and number of variables selected in the solution of initial population were fixed to 1%, 30 and 5 respectively.

GA were performed with MATLAB 7.8 and PLS regressions were performed with the package PLS in R 2.8.1.

To compare and assess the equations, several statistical parameters were computed: mean, standard deviation (Sd), standard error of validation ($SE_{validation}$), validation coefficient of determination ($R^2_{validation}$) and the relative error ($SE_{validation}$/Mean).

$SE_{validation}$ is defined as $\sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N-k-1}}$

with N the number of samples and k the number of latent variables introduced in PLS regression.

We considered that estimation was accurate enough and robust to be applied in routine, when the relative error was under 8%. For relative error in the range of 8 to 12 %, we advise to using these equations with caution. We chose to use this parameter rather than the $R^2_{validation}$ because this latter depends on the standard deviation of our population.

2.0 Results and discussion

The calibrations were validated through the accuracy values obtained by validation on a new dataset in cow and ewe milk and by cross-validation in goat milk (Table 1 to 3). About 10 to 20 fatty acids (depending on the species) of 60 have a relative error below 10%. The estimations are better for the FA present in medium or high concentration, i.e. for the saturated fatty acid (C4:0 to C16:0) and for some monounsaturated fatty acids (cis or trans isomers of C18:1). It is worth noting that in the three species, the relative error for the stearic fatty acid (C18:0) is important

The results are comparable in ewe and cow milk. The estimation of lauric acid (c12) is however better in ewe milk. The accuracy is lower in goat milk. This is certainly linked to the lower level of fat in goat milk. But even for the caprylique (C8:0), capric (C10:0), and lauric fatty (C12:0) acids, whose the contents are more important than in ewe and cow milk, the relative error is important (R.E. >12%). New samples in goat milk are intended in the next weeks to improve the accuracy.

Table 1. Statistical parameters for cow milk validation set (PLS regression only or genetic algorithm (GA) + PLS regression).

	N	Mean	Sd	Relative error (%)	R²
Fat content	70	3.816	0.637	0.32	1.00
C4:0	72	0.149	0.025	5.71	0.88
C6:0	70	0.087	0.015	3.97	0.95
C8:0	70	0.050	0.010	5.00	0.94
C10:0	71	0.111	0.029	6.92	0.93
C12:0	71	0.126	0.037	11.12	0.86
C14:0	72	0.435	0.088	6.10	0.91
C16:0	71	1.271	0.282	6.41	0.92
C18:0	71	0.342	0.099	12.58	0.81
Total 18:1	69	0.780	0.203	6.70	0.93
Saturated	72	2.766	0.510	2.09	0.99
Monounsaturated	69	0.889	0.220	5.80	0.95
Polyunsaturated	69	0.107	0.019	8.06	0.80
Omega 3	70	0.029	0.010	16.24	0.77
Omega 6	70	0.075	0.016	11.23	0.72

Table 2. Statistical parameters for ewe milk validation set (PLS regression).

	N	Mean	Sd	Relative error (%)	R²
Fat content	54	6.802	1.398	0.40	1.00
C4:0	52	0.233	0.035	5.88	0.85
C6:0	54	0.177	0.033	4.21	0.95
C8:0	54	0.175	0.037	4.83	0.95
C10:0	54	0.574	0.147	5.90	0.95
C12:0	54	0.339	0.103	8.57	0.92
C14:0	54	0.821	0.214	6.98	0.93
C16:0	54	1.650	0.345	6.70	0.90
C18:0	55	0.511	0.143	12.62	0.80
Total 18:1	54	1.276	0.414	4.40	0.98
Saturated	54	4.825	0.994	2.31	0.99
Monounsaturated	54	1.389	0.443	3.83	0.99
Polyunsaturated	55	0.238	0.075	7.03	0.95
Omega 3	52	0.069	0.016	13.65	0.66
Omega 6	55	0.137	0.036	12.13	0.79

Table 3. Statistical parameters for goat milk, cross-validation results (PLS regression only or genetic algorithm (GA) + PLS regression).

	N	Mean	Sd	Relative error (%)	R²
Fat content	150	3.310	0.666	0.48	1.00
C4:0	150	0.092	0.025	9.23	0.87
C6:0	150	0.078	0.020	8.97	0.87
C8:0	150	0.080	0.022	12.36	0.78
C10:0	150	0.264	0.071	12.48	0.77
C12:0	150	0.134	0.041	13.36	0.79
C14:0	150	0.307	0.077	9.17	0.85
C16:0	150	0.996	0.197	5.14	0.93
C18:0	150	0.282	0.099	18.14	0.73
Total 18:1	150	0.756	0.176	8.84	0.85
Saturated	150	2.351	0.485	3.55	0.97
Monounsaturated	150	0.798	0.184	8.92	0.84
Polyunsaturated	150	0.128	0.028	12.47	0.65
Omega 3	150	0.018	0.005	19.58	0.44
Omega 6	150	0.109	0.027	13.71	0.65

3.0 Conclusion

These first results show it is possible to obtain accurate estimations for the main fatty acids in individual milk samples of the three species. It was observed that performing a selection of variables prior to the PLS regression permitted to improve accuracy and stabilize equations over the time.

Future researches will focus on other spectrum data pre-treatment procedures, while increasing simultaneously the initial sampling size to get more accurate estimation equations of milk fatty profile.

The advancements of the *PhenoFinLait* program are available on http://www.phenofinlait.fr/.

4.0 References

Bertrand D. and Dufour E., 2006. La spectrométrie infrarouge et ses applications analytiques, Second ed., Tec&Doc Lavoisier, Paris.

Foss, 1998. Reference Manual of Milkoscan FT120 (Type 71200), Denmark.

Kramer J.K.G., Fellner V., Dugan M.E.R., Sauer F.D., Mossoba1 M. M., Yurawecz M. P., 1997. Evaluating acid and base catalysts in the methylation of milk and rumen fatty acids with special emphasis on conjugated dienes and total trans fatty acids. Lipids 32, 1219-1228.

Ferrand M., Huquet B., Barbey S., Barillet F., Faucon F., Larroque H., Leray O., Trommenschlager J.M. and Brochard M., 2010. Determination of fatty acid profil e in cow's milk using mid-infrared spectrometry: Interest of applying a variable selection by genetical gorithms before a PLS regression. Chemometrics and Intelligent Laboratory Systems, Article in Press.

Leardi R. and Lupiañez G., 2002. Genetic algorithms applied to feature selection in PLS regression: how and when to use them. Anal.Chim. Acta. 461, 189-200.

Soyeurt H., Dardenne P, Dehareng F., Lognay G., Veselko G., Marlier M., Bertozzi C., Mayeres P., Gengler N., 2006. Estimating Fatty Acid Content in Cow Milk Using Mid-Infrared Spectrometry. J. Dairy Sci. 89, 3690-3695.

Tenenhaus M., 2002. La regression PLS, Technip, Lassay-les_Chateaux.