Application of FT near spectroscopy for determination of true protein and casein in milk

K. Šustová[1], J. Růžičková[1], J. Kuchtík[2]

[1]Department of Food Technology, Mendel University of Agriculture and Forestry in Brno, Czech Republic
[2]Department of Animal Breeding, Mendel University of Agriculture and Forestry in Brno, Czech Republic

ABSTRACT: Our study deals with a possibility of determining true protein and casein in cow's, ewe's and goat's milk and in ewe's colostrums by FT NIR spectroscopy. Samples of milk were analysed by FT NIR in the reflectance mode with the transflectance cuvette. The values of correlation coefficients of calibration were as follows: cow's protein 0.943; cow's casein 0.964; ewe's protein 0.997; ewe's casein 0.977; goat's protein 0.989; goat's casein 0.890; ewe's colostrum protein 0.983. Calibration was tested using the same set of samples by the cross validation method. The values of correlation coefficients of validation were as follows: cow's protein 0.923; cow's casein 0.910; ewe's protein 0.994; ewe's casein 0.963; goat's protein 0.972; goat's casein 0.814; ewe's colostrum protein 0.871. The NIRS results were compared with reference data and no significant differences between them were found ($P = 0.05$). Results of this study indicate that FT NIR spectroscopy can be used for a rapid analysis of protein and casein in cow's, ewe's and goat's milk and ewe's colostrum.

Keywords: near-infrared spectroscopy; cow's milk; ewe's milk; goat's milk; colostrum; true protein; casein; chemical composition

Proteins constitute a part of the complex system of nitrogenous substances in milk that are called total proteins. From the processing industry aspect, true proteins are more important. Their content in cow's milk ranges between 3.0 and 3.3%. Casein is the most important component of milk proteins while its content fluctuates from 2.4 to 2.6%. True proteins plus non-protein nitrogenous substances are called crude proteins. Non-protein nitrogenous substances amount to 4 to 7% of the total content of nitrogen in milk. Urea accounts for its largest proportion.

In recent years, fluctuations in milk casein content have come to the fore in connection with the increasing production and consumption of cheese. The higher the content of protein in milk, particularly that of casein, the lower the consumption of milk in litres per kilo of cheese under constant conditions of production.

Traditional methods of protein determination in milk are relatively slow and comparatively expensive. Proteins can be assayed by the Kjeldahl method using a conversion factor 6.38. The Kjeldahl method, however, determines total nitrogen in milk and the results are considerably inaccurate owing to the presence of non-protein nitrogenous substances. The proteins determined in this way are called crude protein. Among the instrumental methods of milk protein determination, measurements on a Milko-Scan instrument (O'Sullivan et al., 1999) should be named. This apparatus is calibrated for crude proteins, even though more recent Milko-Scan models can determine casein as well.

Supported by the Ministry of Education, Youth and Sports of the Czech Republic (Research Project No. MSM 6215648905).

Methodical principles of determination of proteins described in our papers are not identical with the earlier presented MIR spectroscopy of milk. These methods are developmentally different (Hanuš et al., 1992a,b).

In recent years, FT NIR spectrometers have come into use in the analysis of milk and milk products. These are relatively new analytical techniques that belong among the non-destructive methods. The samples do not have to be treated or modified before analysis, the measurements are rapid. In the dairy industry, NIR analyses are used for the determination of a number of intermediates and products, such as a rapid analytical method when controlling the production of yoghurt, processed cheese and curds, primarily for the determination of their dry matter, protein and fat content. The sample need not be solubilised as in other methods used for the determination of the content of basic milk components, which enables a rapid analysis of a large number of samples and, if necessary, operative interventions into the production. This technique has also been used increasingly in the control of production of milk powder, butter and milk. When the content components are determined in samples, the NIR spectrometer must be calibrated with a suitable set of calibration standards of the known composition determined by the respective analytical methods. At least 30 calibration samples should be used, covering the entire concentration range (Rodriguez-Otero et al., 1997).

Some of the numerous applications of NIR methods include their use for rapid determinations of milk components; primarily dry matter, fat, proteins and lactose should be mentioned. Lefier et al. (1996) compared the measurements on the FTIR spectrometer 740 and on the Milko-Scan 605. In raw milk, they determined the content of crude proteins and true proteins. An interesting paper dealing with the effect of the composition of feed rations for dairy cows on the spectroscopically determined fat and protein content in milk was published by Purnomoadi et al. (1999). Ru and Glatz (2000) measured the composition of homogenized cow's milk. Fewer papers paid attention to the problems of measuring non-homogenised milk (Tsenkova et al., 2000, 2001; Jankovská and Šustová, 2003). Some authors studied the determination of proteins in raw milk by means of a fibre-optic probe (Kukačková et al., 2000; Turza et al., 2002).

Fourier transformation near infrared spectroscopy (FT NIR) in combination with partial least-squares (PLS) method was used to determine the content of proteins in bovine colostrum. The values of correlation coefficients (*R*) and standard error of calibration (SEC) were computed for protein 0.999 and 0.149, respectively (Navratilová et al., 2006).

Albanell et al. (1999) studied the determination of crude proteins by means of NIR reflectance spectroscopy in homogenised ewe's milk. In goat's milk, the basic components were determined by Diaz-Carrillo et al. (1993), and the content of proteins and casein in the lactation stage was studied by Albanell et al. (2003).

The objective of our work was to study possible applications of FT NIR spectrometry in the determination of true proteins, including casein in cow's, ewe's and goat's milk and in ewe's colostrum.

MATERIAL AND METHODS

Material

For the calibration of the NIR spectrometer, samples of cow's, goat's and ewe's milk and ewe's colostrum were used. The milk samples were taken from dairy cows during lactation; mixed samples were also used. The samples of colostrum were taken in the interval from 2 to 72 hours after delivery. The samples were immediately cooled to a temperature of 6–8°C, stored in cooling containers and transported to the Department of Food Technology, where they were analysed on the very day of sampling. Before the measurement, the samples of milk were heated to 40°C in a water bath and then, after shaking, cooled to 20 ± 1°C. Any possible errors due to milk temperature fluctuations around 20°C are thus included in the calibration.

Reference methods

True protein (TP) and casein were determined spectrophotometrically – Pro-Milk II, Foss Electric, Denmark. The apparatus was calibrated according to Kjeldahl method. Calibration was made separately for cow's, ewe's and goat's milk and for ewe's colostrum.

FT NIR analysis

A wavelength scanning instrument FT NIR Antaris (ThermoNicolet, USA) was used with a

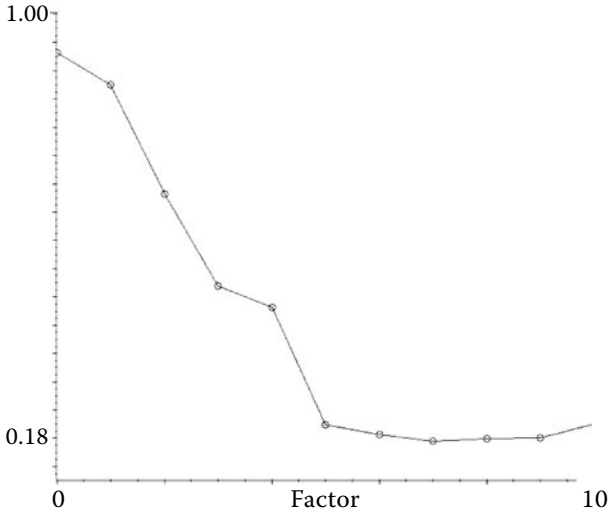

Figure 1. Dependence of the PRESS function on the number of PLS factors used for the calibration of protein estimation in ewe's milk

scanning range from 4 000 to 10 000/cm in the reflectance mode. Samples of milk were transferred to Petri dishes. The measured area was spaced by transflectance cuvette, which defined the constant thickness of sample 0.2 mm. An average of 100 spectral scans was taken for each sample. Diffuse reflectance was recorded as log 1/*R*. Each sample was analysed three times and the average spectrum was used for calibration. The whole spectrum area was tested. The spectra were not modified by any derivative.

A calibration model was created by means of the partial least-squares (PLS) algorithm (Haaland and Thomas, 1988a,b). The same samples were employed for full cross validation by the software FT NIR Reference Analysis. The selection of an optimum number of PLS terms for the calibration was based on the standard error of prediction (SEP) which should be minimised. The statistical parameters (correlation coefficient – *r* and SEP) were used to determinate the final calibration equation.

Correlation coefficients and standard errors of the calibration and validation were calculated in the TQ Analyst operation program. In the construction of calibration models in the TQ Analyst program, three diagnostics were used for the identification of outlying standards which eliminate outlying calibration standards from the set on the basis of spectral or concentration difference: PC Scores, Spectrum Outlier and Leverage methods.

(a) Spectrum Outlier – it calculates the Mahalanobis distance from the average spectrum of each calibration standard of the active method. It seeks spectra which are the most different from other standards, and uses the Dixon or Chauvenet test to estimate whether the difference is significant. For the calculation of the Mahalanobis distance, the concentration and spectral information for each standard and for each component to be determined is used. These diagnostics are based on the spectral information from all standards.

(b) The Principal Component Scores (PC Scores) method depicts graphically how each standard of the PLS, Discriminant Analysis or Distance Match method is represented by means of the PC (during calibration, all important spectral information on the analysed region or regions is condensed into the set of new variables, Principal Components – PC) which were calculated for the method being calibrated. Each PC represents an independent source of spectral variability in the calibration data. The PCs are arranged according to the amount of variability they represent. The first PC describes the greatest variability of calibration spectra. Each additional PC describes the majority of the remaining variability. These diagnostics determine whether the PCs that have been calculated for the calibration method correctly represent the spectral data of each standard. Data points separated from others indicate that the corresponding standard is different from other standards used in the method. The method must be calibrated before the use of the diagnostics.

(c) Leverage: this provides information on the effect of each of the standards on the calibration model, and how accurately the calibration model describes each standard. The data points should be distributed symmetrically over the entire range of the graph thus generated. Every isolated point indicates the difference of the corresponding standard from the others in the method. If the value of the "Leverage" (i.e. the distance of the point from the model centroid) is high, the standard has a significant effect on the calibration model for the given component. The diagnostics can be used only for the PLS (Partial Least-Squares) algorithm. The method being assessed must be calibrated first.

All results were evaluated using the variation statistic analysis (ANOVA). Correlation matrices and regression functions were calculated according to Meloun and Militký (1998) when using the statis-

Table 1. Development of calibration models

Measurement of NIR spectra (Omnic program)
⇩
Transformation of spectra (to TQ Analyst program)
⇩
Addition of reference values
⇩
Modification of spectral data
(without correlation, derivative)
⇩
Partial least squares – PLS
⇩
Selection of wave length
⇩
Calibration: *R*, SEC, CCV, optimal number of PLS factors, PRESS, exclusion of outlying results
⇩
Validation: full cross validation, *R*, SEP, PCV
⇩
Testing: (*z*-test, *t*-test)

tical package Microsoft® Excel 2000 and Unistat 5.1. On the basis of the *F*-test it was established that the variances were homogeneous and therefore the classical paired *t*-test was used for testing the differences.

RESULTS AND DISCUSSION

The calibration models were created by means of the PLS algorithm. The development of the calibration model is described in Table 1. The number of samples, the concentration range of the protein and casein content and the PLS values of the factors for individual calibration models are shown in Table 2. The indicator of the error of the PLS calibration method is the PRESS value (Predicted Residual Error Sum of Squares: the diagnostics that show how the value of the predicted residual error of the sum of squares changes with the number of factors used for the calibration of each component determined using the active method). If the PRESS course is optimal, first a sharp decline occurs, followed by a slow decrease. The optimal number of PLS factors will be found if the PRESS value is minimal. A high number of PLS factors impairs the predictive capability because PRESS also includes the spectral noise. PLS factors are arranged according to the variability quantity they represent. The first factor describes the highest variability of the calibration standards. Each subsequent factor describes the majority of the remaining variability. Nevertheless, the first factor contains the majority of the common information contained in the data. The remaining factors describe more specific information representing small changes in data, which is often important for the analysis. If the trend of the PRESS function is falling sharply, this gives evidence of considerable robustness of the calibration model. The error in the data decreases sharply and the minimum error is determined in the fourth point. This model is mostly characterised by high correlation coefficients, low standard errors, and it can be used for the determination of the given component. The optimal PRESS function was found in the calibration model for casein in ewe's milk (Figure 1).

Figure 2 shows the spectra of samples of cow's, goat's and ewe's milk and ewe's colostrum. The results of the calibration and validation of the method of milk protein and casein determinations are given in Tables 3 and 4. The model reliability was verified by cross-validation with the same group of samples as those used for calibration. The validation verifies the reliability of the calibration model and it is characterised by the standard error of validation

Table 2. Calibration components

True protein	*n*	Max.	Min.	PLS	Transformation
Cow's milk	152	3.97	2.62	12	none
Ewe's milk	76	7.79	4.92	14	none
Ewe's colostrum	29	11.41	5.79	10	none
Goat's milk	64	3.34	1.87	9	none
CASEIN					
Cow's milk	64	3.26	2.10	10	none
Ewe's milk	46	6.50	3.14	7	none
Goat's milk	29	2.49	1.40	6	none

n = number of samples; max. = maximum values; min. = minimum values; PLS = number of factors

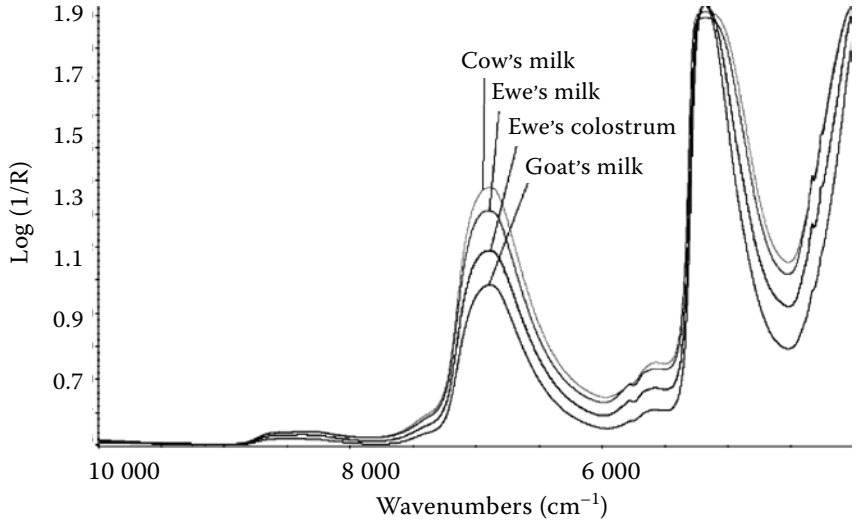

Figure 2. Examples of NIR spectra of cow's milk, ewe's milk, goat's milk and ewe's colostrum

(SEP). The calibration is considered very reliable if the assumed value of the calibration coefficient of variation CCV is lower than 5%, and the value of the prediction coefficient of variation PCV is lower than 10%. In our case, these criteria were met in all models (a slightly higher CCV was in the results of the casein content in goat's milk, and PCV in the results of the protein content in ewe's colostrum). This gives evidence of very reliable calibration.

The linear dependence of the reference results versus results predicted by the PLS algorithm for the determination of true protein and casein in cow's milk is illustrated in Figures 3 and 4. Obviously, there is a good correlation between predicted values and known chemical reference

Table 3. Parameters of the regression function $y'_i = a + bx_i$ for the calibration model

True protein	$a + bx_i$	SEC (%)	CVV (%)	R
Cow's milk	0.373 + 0.888	0.09	2.69	0.943
Ewe's milk	0.039 + 0.994	0.06	0.98	0.997
Ewe's colostrum	0.287 + 0.965	0.34	4.16	0.983
Goat's milk	0.058 + 0.978	0.06	2.28	0.989
CASEIN				
Cow's milk	0.192 + 0.929	0.08	2.92	0.964
Ewe's milk	0.208 + 0.954	0.19	4.18	0.977
Goat's milk	0.439 + 0.792	0.12	5.68	0.890

SEC = standard error of calibration; CCV = calibration coefficient of variation; R = correlation coefficient

Table 4. Parameters of the regression function $y'_i = a + bxi$ for the validation model

True protein	$a + bx_i$	SEP (%)	PCV (%)	R
Cow's milk	0.410 + 0.877	0.10	2.99	0.923
Ewe's milk	0.056 + 0.991	0.09	1.43	0.994
Ewe's colostrum	1.313 + 0.842	0.92	11.22	0.871
Goat's milk	0.162 + 0.937	0.09	3.43	0.972
CASEIN				
Cow's milk	0.339 + 0.874	0.13	4.65	0.907
Ewe's milk	0.268 + 0.941	0.24	5.27	0.963
Goat's milk	0.514 + 0.757	0.16	7.58	0.814

R = correlation coefficient; SEP = standard error of prediction; PCV = prediction coefficient of variation

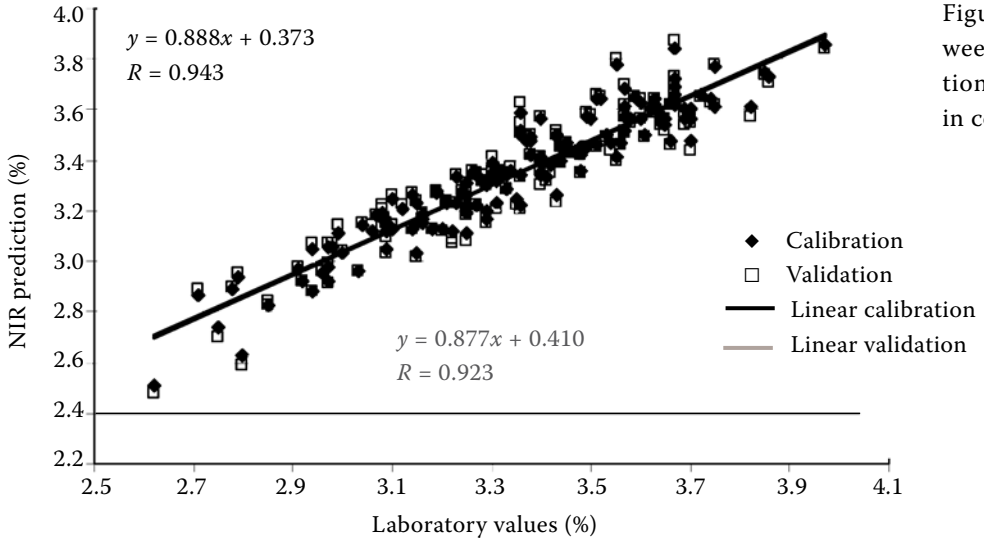

Figure 3. Relationship between calibration and validation results of true protein in cow's milk

values. The correlation coefficients of calibration for true proteins were in the range of 0.997 for ewe's milk to 0.943 for cow's milk, with standard errors of calibration from 0.06% to 0.34%. In calibration models for the determination of the casein content, correlation coefficients from 0.977 in ewe's milk to 0.890 in goat's milk were obtained, with standard errors from 0.08% to 0.19%. Correlation coefficients for validation for true proteins were in the range from 0.994 for ewe's milk to 0.871 for ewe's colostrum. Correlation coefficients of validation for casein ranged from 0.963 (ewe's milk) to 0.814 (goat's milk). Standard errors of validation were in the range of 0.09% to 0.92% for true proteins, and 0.13% to 0.24% for casein. The most reliable model was obtained for true proteins in ewe's milk, and the least reliable for ewe's colostrum. Ewe's colostrum has a quite different composition from milk. True proteins were determined with the PROMILK instrument. This method is based on the formation of an insoluble protein-dye complex, forming after the addition of excess Amido Black 10B to a buffered solution. In colostrum, one cannot expect the same binding of the dye by proteins as in normal milk, and thus the error of the assay can be larger. The error of determination by NIR spectrometry can never be lower than in the reference methods.

Very similar results in the calibration of the protein content in cow's milk (R = 0.965) were obtained by Kukačková et al. (2000); these authors used the fibre-optic technology for the measurements. In other published papers, transmittance techniques were used for reading cow's milk spectra. One of these papers is that by Tsenkova et al. (2000), who obtained the best results for proteins (R = 0.886). Purnomoadi et al. (1999) reported

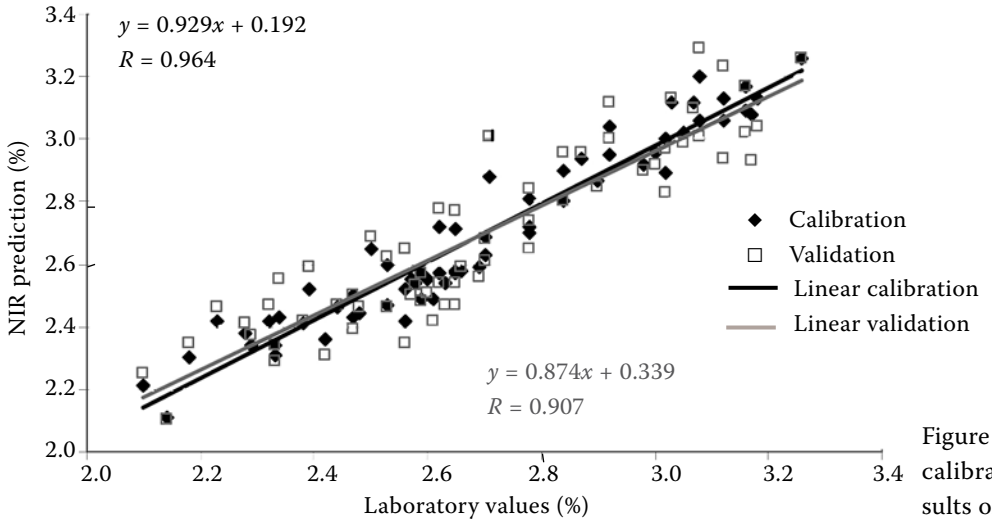

Figure 4. Relationship between calibration and validation results of casein in cow's milk

Table 5. Parameters of the basic components in ewe's milk and ewe's colostrum as estimated by NIR reference values and their mutual comparison by paired *t*-test

True protein	*n*	*x*NIR	*x*REF	SD	T stat	T krit (1)	T krit (2)
Cow's milk	152	3.34	3.34	0.021	$2.74 \cdot 10^{-2}$	1.66	1.98
Ewe's milk	76	6.08	6.08	0.089	$1.92 \cdot 10^{-2}$	1.67	1.99
Ewe's colostrum	29	8.19	8.19	0.341	$1.07 \cdot 10^{-2}$	1.70	2.05
Goat's milk	64	2.63	2.63	0.049	$-4.20 \cdot 10^{-2}$	1.67	1.99
CASEIN							
Cow's milk	64	2.69	2.69	0.037	$-1.58 \cdot 10^{-2}$	1.67	1.99
Ewe's milk	46	4.55	4.55	0.132	$-1.7 \cdot 10^{-15}$	1.68	2.01
Goat's milk	29	2.11	2.11	0.049	$6.09 \cdot 10^{-2}$	1.70	2.05

n = number of samples; *x*NIR = mean of the NIR values; *x*REF = mean of the reference values; SD = standard deviation of mean; T stat = values of paired *t*-test; T krit (1) = table values for α = 0.05; T krit (2) = table values for α = 0.01

R = 0.904 and SEP = 0.07% for the calibration of proteins in milk, and *R* = 0.902 and SEP = 0.09% for the validation. Tsenkova et al. (2001) and Turza et al. (2002) agree on the basis of their results that NIR spectroscopy as such is a very suitable method for the determination of the basic composition of raw, non-homogenised milk. Lefier et al. (1996) calculated the following standard errors of calibration after creating their calibration model: 0.048 for crude proteins and 0.035 for true proteins. From their results and on the basis of other studies, they arrived at the conclusion that the FTIR spectrometer was significantly better than the standard filtration instrument as it provided more spectral information on individual milk samples.

Diaz-Carrillo et al. (1993) paid most attention to the determination of the basic components of goat's milk by the NIR spectroscopy. In their paper they reported high correlation coefficients that are comparable with our results. They also used the PLS method for calibration, but the samples (50) were measured in the transmittance mode. The correlation coefficient of calibration *R* for proteins reached the value 0.96, and SEP amounted to 0.15%. In our work, the *R* values are higher, namely 0.989 with SEP 0.06%.

Albanell et al. (2003) were another team that worked with goat's milk and analysed the possibilities of determining its basic components by NIR spectroscopy. The samples were measured by reflectance and they obtained the following results: calibration in homogenized goat's milk: *R* = 0.96 for proteins and *R* = 0.95 for casein with validations for proteins *R* = 0.91 and casein *R* = 0.91. The correlation coefficients of the calibration in non-homogenised milk were *R* = 0.95 for proteins and *R* = 0.92 for casein, with the validation for proteins of *R* = 0.95 and *R* = 0.92 for casein.

Another paper that analysed the basic composition of ewe's milk is that by Albanell et al. (1999). In their work, they obtained the correlation coefficient of calibration amounting to *R* = 0.92 for proteins and validation of *R* = 0.92.

In our work, we obtained higher values of correlation coefficients than those reported by the above mentioned authors. The testing of the difference between the reference and predicted values confirmed that there was no statistical difference between the methods of determination (Table 5) in any of the components tested. On the basis of these comparisons, NIR spectrometry appears suitable for the determination of the protein and casein content in non-homogenised milk. The optional rapid analysis of milk true proteins, particularly casein, would contribute to an objective assessment of milk quality.

CONCLUSIONS

The results have demonstrated the possibility of determining the milk true protein and casein content in the near-infrared spectral region in raw non-homogenised milk, cow's, goat's and ewe's milk and ewe's colostrum. The use of NIR spectroscopy, despite its high initial investment costs for the purchase of the technology, can create economic savings as it enables the analysis of a large number of samples with minimal additional costs. This method enables the easy and quick control of milk composition, timely interventions

in the production, and thus it can improve the economics of dairy plants, in particular of cheese producers.

REFERENCES

Albanell E., Cáceres P., Caja G., Molina E., Garcouri A. (1999): Determination of fat, protein, and total solids in ovine milk by Near-Infrared Spectroscopy. J. AOAC Int., 82, 753–758.

Albanell E., Caja G., Such X., Rovai M., Salama A., Casals R. (2003): Determination of fat, protein, casein, total solids and somatic cell count in goat's milk by Near-Infrared Reflectance Spectroscopy. J. AOAC Int., 86, 746–752.

Diaz-Carrillo E., Munoz-Serrano A., Alonso-Moraga A., Serradilla-Manrique J.M. (1993): Near Infrared calibrations for goat's milk components: protein, total casein, αS-, βS- and κ- caseins, fat and lactose. J. Near Infrared Spec., 1, 141–146.

Haaland D.M., Thomas E.V. (1988a): Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. Anal. Chem., 60, 1193–1202.

Haaland D.M., Thomas E.V. (1988b): Partial least-squares methods for spectral analyses. 2. Application to simulated and glass spectral data. Anal. Chem., 60, 1202–1208.

Hanuš O., Genčurová V., Žváčková I. (1992a): Tests of Milkofix a new preservative substance for milk samples used for the purposes of an infrared analysis of basic milk composition. Part II. Checks of preservative effects in relation to the infrared analysis. Vet. Med., 37, 33–43. (in Czech)

Hanuš O., Genčurová V., Gabriel B. (1992b): The effect of sample aging on the accuracy of an infrared analysis of basic milk composition. Vet. Med., 37, 149–160. (in Czech)

Jankovská R., Šustová K. (2003): Analysis of cow milk by near-infrared spectroscopy. Czech J. Food Sci., 21, 123–128.

Kukačková O., Čurda L., Jindřich J. (2000): Multivariate calibration of raw cow milk using NIR spectroscopy. Czech J. Food Sci., 18, 1–4.

Lefier D., Grappin R., Pochet S. (1996): Determination of fat, protein and lactose in raw milk by Fourier Transform Infrared Spectroscopy and by analysis with a conventional filter based analyser. J. Assoc. Off. Ana. Chem. Int., 79, 711–717.

Meloun M., Militký J. (1998): Statistické zpracování experimentálních dat. 1. vyd. Praha: Ars Magna, 839 pp.

Navrátilová P., Hadra L., Dráčková M., Janštová B., Vorlová L., Pavlata L. (2006): Use of FT-NIR Spectroscopy for bovine colostrum analysis. Acta Vet. Brno, 75, 57–63.

O'Sullivan A., O'Connor B., Kelly A.M., McGrath M. (1999): The use of chemical and infrared methods for analysis of milk and dairy products. Int. J. Dairy Technol., 52, 139–148.

Purnomoadi A., Batatajoo K., Ueda K., Terada F. (1999): Influence of feed source on determination of fat and protein in milk by near-infrared spectroscopy. Int. Dairy J., 9, 447–452.

Rodriguez-Otero J.L., Centero J., Hermida M. (1997): Application of near-infrared transflectance spectroscopy to the analysis of fermented milks. Milchwissenschaft, 52, 196–200.

Ru Y.J., Glatz P.C. (2000): Application of NEAR Infrared Spectroscopy (NIR) for monitoring the quality of milk, cheese meat and fish. Asian-Aus. J. Anim. Sci., 13, 1017–1025.

Tsenkova R., Atanassova S., Itoh K., Ozaki Y., Toyoda K. (2000): Near infrared spectroscopy for bio-monitoring: Cow milk composition measurement in a spectral region from 1.100 to 2.400 nanometers. J. Anim. Sci., 78, 515–522.

Tsenkova R., Atanassova S., Toyoda K. (2001): Near Infrared Spectroscopy for diagnosis: Influence of mammary gland inflammation on cow's milk composition measurement. Near Infrared Anal., 2, 59–66.

Turza S., Chen J. Y., Terazawa Y., Takusari N., Amari M., Kawano S. (2002): On-line monitoring of rumen fluid in milking cow's by fibre optics in transmittance mode using the longer NIR region. J. Near Infrared Spec., 10, 111–120.

Received: 2006–07–04
Accepted after corrections: 2007–05–14

Corresponding Author

Ing. Květoslava Šustová, Ph.D., Department of Food Technology, Mendel University of Agriculture and Forestry in Brno, Zemědělská 1, 613 00 Brno, Czech Republic
Tel. + 420 545 133 257, e-mail: sustova@mendelu.cz