PIOTR MINKIEWICZ, JERZY DZIUBA*), DOROTA NAŁĘCZ

University of Warmia and Mazury in Olsztyn, Chair of Food Biochemistry
Plac Cieszyński 1, 10-726 Olsztyn-Kortowo, Poland

Application of derivatives of UV spectra and similarity indices to identification of milk proteins separated using reversed-phase high performance liquid chromatography**)

Summary — The method which may be applied as a tool for identification of proteins separated using reversed-phase high performance liquid chromatography, based on similarity indices of UV spectra derivatives, was proposed. Commercial milk protein preparations such as sodium and calcium caseinates, whey protein concentrate and milk protein coprecipitate as well as standards of milk proteins were used in the experiment. Statistical significance of difference between first and second similarity index, evaluated with Student's t test was used as a criterion to choose the mathematical treatment of spectra as well as to predict reliability of identification. First derivatives of spectra revealed the highest value of t parameter. The presence of statistically significant difference between first and second similarity indices always led to correct identification of proteins. The presented method enabled appropriate identification of caseins in commercial milk protein preparations, but failed in the case of whey proteins.

Key words: identification of proteins, reversed-phase HPLC, UV spectroscopy, similarity indices of UV spectra derivatives.

Although mass spectrometry is the most efficient method applied for identification of proteins and peptides separated using reversed-phase high performance liquid chromatography (RP-HPLC) [1], in many cases identification may be performed using relatively simple photodiode-array (PDA) detectors. Such detectors are commonly used [2].

High performance liquid chromatography on-line with UV spectroscopy was applied in numerous experiments concerning identification of proteins and/or peptides [3—13]. In some cases this technique was used as supplementary one for mass spectrometry [9, 11, 13] or electrophoretic methods [10].

Calculation of spectra derivatives in order to enhance resolution is commonly used procedure for data interpretation concerning peptides and proteins. This procedure was applied in most of experiments concerning UV spectra-based identification of peptides and/or proteins separated by HPLC [3, 4, 6—13]. On the other hand common procedure for identification of any substances separated with HPLC is in comparison with standard by means of calculation of the similarity indices of spectra. Similarity index is defined as cosinus of angle between vectors in multidimensional space (named match angle) being mathematical representation of spectra. The calculation of match angles between raw spectra has been introduced as a tool for qualitative analysis of peptides [5, 9].

The aim of present study was to propose a strategy including both calculation of derivatives of UV spectra and similarity indices for identification of milk proteins separated with reversed-phase high performance liquid chromatography. The second goal was to find a criterion to evaluate a method performance.

EXPERIMENTAL

Materials

Standards of milk proteins: β-casein (β-Cn), κ-casein (κ-Cn), α-lactalbumin (α-La), β-lactoglobulin (β-Lg) and bovine serum albumin (BSA) have been purchased from Sigma. As regards α_{s1}-casein (α_{s1}-Cn) and α_{s2}-casein (α_{s2}-Cn) they were obtained using anion-exchange chromatography [14] from whole casein precipitated from the skim milk of individual Jersey cows. The commercial milk protein preparations: sodium caseinate, calcium

*) To whom all correspondence should be addressed, e-mail: dziuba@uwm.edu.pl

**) A paper sent to Editorial Office between research works to be published in the special issue of "Polimery" devoted to biopolymers (No 1/2003).

caseinate, whey protein concentrate and milk protein coprecipitate have been purchased from the Lacpol (Poland).

Methods

The chromatographical analysis was carried out with the Shimadzu assembly consisting of two LC-10AD pumps, SIL-10AD autosampler, SCL-10AD controller, CTO-10AS column oven and SPD-M10AW PDA detector. The Jupiter C_{18} column (Phenomenex) of 250×4.6 mm size was used. The Class-VP 5.03 software (Shimadzu) was used for data acquisition and processing.

Protein samples were dissolved, reduced and diluted as described by Visser *et al.* [15]. The concentration of protein standards and milk protein preparations was within the range 1—2 mg/mL and 4—6 mg/mL respectively. The injection volume was 50 μL. The protein separations were carried out in acetonitrile (ACN) gradient. Solvents A and B consisted of ACN, water and trifluoroacetic acid (TFA) in the volume ratio 100:900:1 and 900:100:0.7 respectively [15]. We have used the deionized water (MilliQ system, Millipore), ACN and TFA of HPLC grade (Baker). All solutions were filtered through the nylon filter with 0.45 μm pore diameter. The separations were carried out using the flow rate 0.8 mL/min at the temperature 30°C. The following gradient was used: 30% B at the start, 35% B after 10 min, 39% B after 12 min, 46% B after 40 min. After the completion of the gradient, column was washed and equilibrated as described previously [12]. Spectra of proteins were taken from the chromatograms. The UV spectra were recorded within the wavelength range 190—300 nm. Data acquisition time was 45 min.

The used program calculates only the first or second derivative of function. Due to this fact the following order of calculations was maintained: 2nd derivative was calculated using 2nd derivative option; 3rd derivative — using 2nd derivative followed by 1st derivative; 4th derivative — using 2nd derivative applied twice and 5th derivative — 2nd derivative calculated twice and followed by 1st derivative. We have constructed six separate libraries containing raw spectra and 1st, 2nd, 3rd, 4th and 5th derivatives of milk protein standards spectra.

The spectra or derivatives of spectra were compared with the standard spectra using similarity indices calculated according to the equation:

$$SI = \frac{\sum_{\lambda_i} \frac{d^m A_1}{d\lambda^m}(\lambda_i) \times \frac{d^m A_2}{d\lambda^m}(\lambda_i)}{\sqrt{\sum_{\lambda_i}\left[\frac{d^m A_1}{d\lambda^m}(\lambda_i)\right]^2} \times \sqrt{\sum_{\lambda_i}\left[\frac{d^m A_2}{d\lambda^m}(\lambda_i)\right]^2}} \quad (1)$$

where: *SI — similarity index;* λ *— wavelength;* A_1, A_2 *— absorbance, at the wavelength* λ_i*, of standard and checked substance respectively;* $d^m A/d\lambda^m$ *—* m^{th} *derivative of spectrum;* $m = 0$ *(for raw spectra),* 1, 2, 3, 4 *or* 5.

The spectra and derivatives of spectra were compared with those of standards within the wavelength range 270—300 nm as recommended in our previous article [12].

The statistical significance of the difference between first (highest) similarity index and the second similarity index was evaluated using Student's *t* test.

RESULTS AND DISCUSSION

The chromatograms of commercial milk protein preparations are shown in the Fig. 1. The values of retention time of protein standards are close to those shown

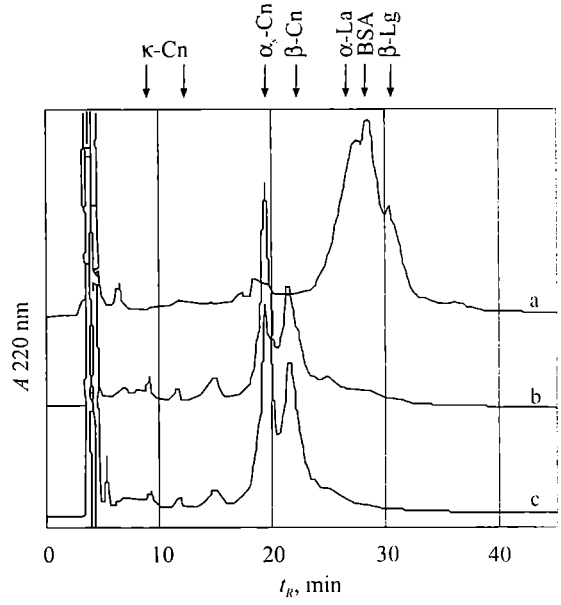

Fig. 1. Chromatograms of milk protein preparations: (a) whey protein concentrate, (b) milk protein coprecipitate, (c) sodium caseinate. The chromatograms of calcium caseinate were nearly identical with the chromatogram (c). Arrows indicate retention times of protein standards

previously [12]. The chromatograms of commercial preparations revealed drastical decrease of resolution as compared with the standard ones. The standards of α-La and β-Lg being main components of a whey protein concentrate and present in a milk protein coprecipitate were well separated within the conditions used [12]. The decrease of resolution is probably due to the chemical changes in proteins during production of preparations. Production of all preparations involves heating. Heat treatment of milk proteins leads among others to partial deamidation [16]. This reaction should lead to a change of apparent hydrophobicity of peptides or proteins and, hence to a change of their retention times [17]. The most common protein modification occurring during heating of milk and dairy products is reaction with sugars named as Maillard reaction [18—20]. In its advanced

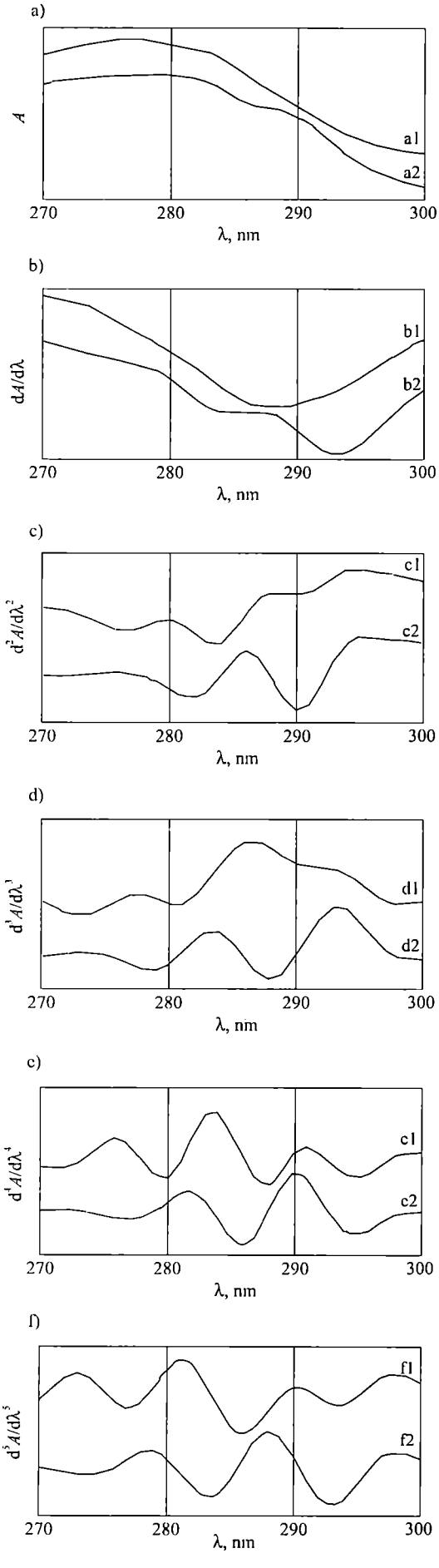

Fig. 2. UV spectra of protein standards: (a) raw spectra, (b) first derivatives of spectra, (c) second derivatives of spectra, (d) third derivatives of spectra, (e) fourth derivatives of spectra, (f) fifth derivatives of spectra; a1, b1, c1, d1, e1 and f1 — spectra of α_{s1}-casein; a2, b2, c2, d2, e2 and f2 — spectra of α-lactalbumin

stage this reaction leads to diverse products originating from one protein, which may reveal different retention times. Whey proteins are more sensitive to any heat-induced changes than caseins [21].

Table 1. The average similarity indices calculated for standards of milk proteins

	First $SI \pm SD$ ($n = 21$)	Second $SI \pm SD$ ($n = 21$)	t
Raw spectra	1.000±0.000	0.998±0.002	3.990**
1st derivatives	0.999±0.001	0.995±0.005	4.246**
2nd derivatives	0.994±0.011	0.977±0.022	3.134*
3rd derivatives	0.957±0.136	0.935±0.139	0.521
4th derivatives	0.929±0.172	0.907±0.176	0.414
5th derivatives	0.756±0.360	0.693±0.335	0.592

*) Difference significant at the level 0.01.
**) Difference significant at the level 0.001.

Figure 2 shows the UV spectra of α_{s1}-Cn and α-La as examples of the influence of mathematical treatment on the differences between spectra of various proteins. The figure illustrates the fact that second, third and fourth derivatives (Fig. 2c, 2d and 2e respectively) calculated from closely related raw spectra (Fig. 2a) may differ. The differences between first and second similarity indices are shown in Table 1. Calculation of derivatives of spectra leads to an increase of resolution, but causes a decrease of a signal to the noise ratio [22, 23]. The difference between first and second similarity indices may be used as a measure of method resolution. Decrease of a signal to the noise ratio leads to an increase of standard deviation. Thus the value of parameter indicating statistical significance of difference between first and second similarity indices is the best criterion for choice of mathematical treatment of spectra. In our case the first derivatives of UV spectra has given the highest t value. The data set used for this calculation have included among others the pairs of proteins whose UV spectra are impossible to discriminate such as α_{s1}- and α_{s2}-caseins [12].

The experiment carried out with commercial milk protein preparations allowed us to answer the following question: how to check that protein identification based on the UV spectrum is reliable? The data in Table 2 show that all incorrect identifications based on the spectra revealed lack of statistical significance of difference be-

Table 2. Identification of proteins in milk protein preparations

Identification based on:			
Retention time	Raw spectra	First derivatives of spectra	Second derivatives of spectra
κ-Cn	κ-Cn or BSA	κ-Cn[1)] SI_1 = 0.999±0.001 SI_2 = 0.997±0.002 (n = 18)	κ-Cn SI_1 = 0.984±0.019 SI_2 = 0.974±0.016 (n = 18)
α_s-Cn	α_s-Cn	α_s-Cn[*)2)] SI_1 = 1.000±0.000 SI_3 = 0.995±0.004 (n = 9)	α_s-Cn[*)2)] SI_1 = 0.997±0.002 SI_3 = 0.971±0.004 (n = 9)
β-Cn	α_s-Cn or β-Cn	β-Cn[**)] SI_1 = 0.999±0.001 SI_2 = 0.997±0.001 (n = 9)	β-Cn[**)] SI_1 = 0.995±0.002 SI_2 = 0.982±0.003 (n = 9)
α-La	α-La	α-La	α-La
BSA	α-La	α-La	α-La
β-Lg	α-La	α-La	α-La

*) Difference between first and second similarity indices was significant at the level 0.05.

**) Difference between first and second similarity indices was significant at the level 0.01.

1) For the κ-Cn peak with longer retention time (see Fig. 1); difference between first and second similarity indices was significant at the level 0.01.

2) This peak at the chromatograms of commercial milk preparations contained both α_{s1}- and α_{s2}-Cn (see Fig. 1). Due to this fact we have considered the third similarity index (assigned to the β-Cn) instead of the second one.

tween the first and second similarity indices. The presence of difference significant at the level at least 0.05 led always to the correct identification. The data in Table 2 show that first and second derivatives of spectra are better tools for protein identification than the raw spectra. As we have pointed out previously [12] raw spectra of proteins are sensitive to the concentration changes. The data in Table 1 were taken from the chromatograms obtained using similar protein concentration. Whey proteins (α-La, β-Lg and BSA) could not be identified using UV spectra. The changes of proteins during production of preparations led probably not only to loss of resolution revealed by RP-HPLC separations, but also to the changes of their spectral properties.

Some limitations of protein identification using UV spectra have been discussed previously [12]. Figure 3 illustrates an additional possible source of an error. This figure shows the two second derivatives of α-La spectrum calculated for the same peak at the same chromatogram *via* different procedures. At the spectrum "a" the depths of minima (especially that at *ca.* 290 nm) have decreased as compared with spectrum "b". Spectra of all proteins revealed similar behavior. Spectrum "a" has been incorrectly identified as β-Lg spectrum (SI = 0.918). Calculation of the third derivative of spectra, as the first derivative followed by second derivative (the order of

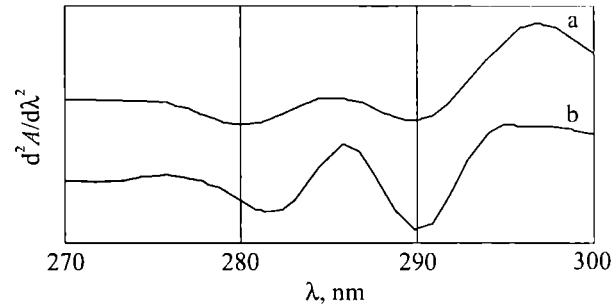

Fig. 3. Second derivatives of α-lactalbumin UV spectra: (a) first derivative calculation repeated twice, (b) second derivative calculated directly

calculations opposite to this described in "Methods"), led to appropriate identification of proteins but first similarity indices were within the range 0.94—0.95 when the same spectrum (inserted into library) was used for calculation *via* both ways. The phenomenon described above may be tentatively explained by the error propagation during numerical calculations. The fact that numerical procedures equivalent from the mathematical point of view may lead to different results, is known for many years [24]. It implies for the importance to maintain the same order of calculation of first and second derivatives of standard spectra and identified substances spectra.

CONCLUSIONS

Similarity indices of UV spectra derivatives may be used as a tool for protein identification. We can expect the correct identification when the difference between first and second similarity indices is statistically significant. First and second derivatives of spectra provide the most precise identification of proteins. The procedure of numerical calculations may be a source of identification error.

ACKNOWLEDGMENT

This work was supported from the funds of UWM within the project 0712.203.

REFERENCES

1. Mann M., Hendrickson B. C., Pandey A.: *Ann. Rev. Biochem.* 2001, **70**, 437.
2. Riordon J. R.: *Anal. Chem.* 2000, **72**, A483.
3. Grego B., Nice E. C., Simpson R. J.: *J. Chromatogr.* 1986, **352**, 359.
4. Palladino D. E. H., Cohen K. A.: *J. Chromatogr. Sci.* 1990, **29**, 91.
5. Young P. M., Gorenstein M. V.: *LC-GC Int. Mag. Separ. Sci.* 1994, **12**, 832.
6. Perrin E., Miclo L., Driou A., Linden G.: *J. Chromatogr. B* 1995, **664**, 267.

7. Zhao Q., Sannier F., Garreau I., Lecoeur C., Piot J. M.: *J. Chromatogr. A* 1996, **723**, 35.
8. Bartolomé B., Moreno-Arribas V., Pueyo E., Polo C.: *J. Agr. Food Chem.* 1997, **45**, 3374.
9. Macaud C., Zhao Q. Y., Ricart G., Piot J. M.: *J. Liq. Chromatogr. Relat. Technol.* 1999, **22**, 401.
10. Darewicz M., Dziuba J., Mioduszewska H., Minkiewicz P.: *Acta Aliment. Hung.* 1999, **28**, 339.
11. Minkiewicz P., Slangen C. J., Dziuba J., Visser S., Mioduszewska H.: *Milchwissenschaft* 2000, **55**, 14.
12. Dziuba J., Nałęcz D., Minkiewicz P.: *Anal. Chim. Acta* 2001, **449**, 243.
13. Miclo L., Perrin E., Driou A., Papadopoulos V., Boujrad N., Vanderesse R., Bondier J. F., Desor D., Linden G., Gaillard J. L.: *FASEB J.* 2001, **15**, U466.
14. Thompson M. P.: *J. Dairy Sci.* 1966, **49**, 792.
15. Visser S., Slangen C. J., Rollema H. S.: *J. Chromatogr.* 1991, **548**, 361.
16. Gaucheron F.: *Milchwissenschaft* 2001, **56**, 183.
17. Wilce M. J. C., Aguilar M. I., Hearn M. T. W.: *Anal. Chem.* 1995, **67**, 1209.
18. Yayalayan V. A.: *Trends Food Sci. Technol.* 1997, **8**, 13.
19. Martins S. I. F. S., Jongen W. M. F., van Boekel M. A. J. S.: *Trends Food Sci. Technol.* 2001, **11**, 364.
20. Fayle S. E., Healy J. P., Brown P. A., Reid E. A., Gerrard J. A., Ames J. M.: *Electrophoresis* 2001, **22**, 1518.
21. Singh H., Creamer L. K.: "Proteins" in "Advanced Dairy Chemistry", vol. 1 (Ed. Fox P. F.), Elsevier Applied Science, London 1992, p. 621—656.
22. Padros E., Dunach M., Morros A., Sabes M., Manosa J.: *Trends Biochem. Sci.* 1984, **9**, 508.
23. Bray M. R., Carriere A. D., Clarke A. J.: *Anal. Biochem.* 1994, **221**, 278.
24. Peitgen H.-O., Jürgens H., Saupe D.: "Fractals for the Classroom", Springer-Verlag, New York 1992 (Polish edition entitled: "Fraktale i chaos", PWN, Warszawa 1995).

W kolejnym zeszycie ukażą się następujące artykuły:

- Multihydroksylowe silnie rozgałęzione polietery
- Poliimidy. Cz. I. Synteza i właściwości fizykochemiczne
- Wybór metody olejowania kauczuku i właściwości olejowanego 1,4-*cis*-polibutadienu otrzymywanego wobec katalizatora neodymowego (*j. ang.*)
- Właściwości termiczne kauczuków butadienowo-akrylonitrylowych
- Ocena właściwości adhezyjnych utlenionych powierzchni stopów aluminium w świetle teorii adsorpcji
- Wpływ procesu wytłaczania z rozdmuchiwaniem folii z poliamidu na jej strukturę i właściwości
- Modelowanie wytłaczania wielowarstwowego tworzyw. Cz. I. Przepływy lepkie
- Właściwości skrobi i polimeru syntetycznego zawierającego grupy karboksylowe — *Komunikat szybkiego druku* (*j. ang.*)
- Kinetyka polimeryzacji etylenu pod wpływem katalizatora tytanowego na bimetalicznym nośniku $MgCl_2(THF)/AlEt_2Cl$ — *Komunikat szybkiego druku* (*j. ang.*)