Protein secondary structure and near infrared spectroscopy

P. Robert,[a] D. Renard,[a] N. Mouhous[b] and E. Dufour[c]

[A]*LPCM, INRA, BP 71627, 44316 Nantes Cedex 03, France.*
[b]*LEIMA, INRA, BP 71627, 44316 Nantes Cedex 03, France.*
[c]*ENITA Clermont Ferrand, Marmilhat 63370 Lempdes, France.*

Introduction

The effects of various alcohols on proteins have been extensively studied using circular dichroism, fluorescence and mid infrared spectroscopy. It has been shown that alcohol molecules denature the native state of proteins and induce α-helical structures. The hydrophobic interactions stabilising the native structure of proteins are weakened by solvents of low polarity. Simultaneously, local hydrogen bonds are strengthened, resulting in denaturation and stabilisation of extended α-helical structures. Hirota-Nakaoka and Goto recently examined the additive contribution of various chemical groups constituting alcohol molecules.[1] Whereas the hydrocarbon groups of alcohol molecules promote the transition between native and α-helical states of proteins, the hydroxyl groups of alcohols have a low potential in α-helix formation. In addition, the authors observed that halogen groups favoured the α-helix transition depending on their type and number. Using both CD and mid infrared spectroscopy, Dong *et al*. demonstrated that the non-native α-helical structures induced in β-lactoglobulin by 2,2,2 trifluoroethanol (TFE) are unstable.[2] The strong Amide I band, observed at 1654 cm^{-1} just after dissolution in TFE, decreases in intensity as a function of time. Concomitantly, two new absorption bands appear near 1620 and 1695 cm^{-1}. These results suggested that the non-native α-helices readily converts to inter-molecular β-sheet aggregates.

In the mid infrared region, the elucidation of the secondary structure of protein may be hindered by the strong absorption band of water at about 1640 cm^{-1}. An alternative to this difficulty might be the use of near infrared spectroscopy in the 2000–2400 nm region. In this spectral region, combination bands at 2050 and 2180 nm have been tentatively assigned to peptide groups.[3] While the absorption band at 2050 nm is thought to describe a combination band of Amide A and Amide II, the band at 2180 nm has been assigned to a combination of either the Amide A mode with Amide III or the first overtone of Amide I with Amide III. Although the combination bands observed in the near infrared region involve amide bands sensitive to the secondary structure of proteins, the potential use of near infrared spectroscopy still needs to be clarified. In 1991, Kamishikiryo *et al*. carried out a study on the stability of 2170 nm for bovine serum albumin in solution.[4] Surprisingly, the authors did not observe any changes in the 2170 nm absorption band under the influence of pH or temperature. Different results were obtained for the denaturation of lysozyme at high pressures and temperatures.[5,6] The second derivative spectra revealed significant changes near 2208, 2170 and 2055 nm that were ascribed to changes in the secondary structure. Kamishikiryo-Yamashita *et al*. conducted a closer inspection of the 2170 nm absorption band using nine model proteins.[7] They observed that the relative extent of α-helix, β-sheet and random structures to the absorption intensity of 2170 nm was 2 : 1 : 1.

In the present work, we first examined the potential of near infrared spectroscopy to characterise the secondary structure of protein in the solid state. The near infrared spectra of a set of 12 well-known

Near Infrared Spectroscopy: Proceedings of the 9th International Conference

proteins were studied in relation to the corresponding mid infrared spectra. The mid and near infrared spectra were, simultaneously, studied by using the generalised canonical correlation analysis. In a second step, the aggregation of β-lactoglobulin in water / ethanol solution was investigated by recording near infrared spectra for a period of 300 minutes.

Experimental

Samples

β-lactoglobulin was purified from bovine milk by the Maillart and Ribadeau-Dumas method.[8] β-casein and κ-casein were purified from bovine milk as previously described.[9,10] The other proteins were purchased from Sigma and used without further purification. The percentage of α-helix, β-sheet and random structures (Table 1) were the reference data used in the generalised canonical correlation analysis.

β-lactoglobulin was dissolved in water (30 mg mL^{-1}, 10 mg mL^{-1}) and the pH adjusted to 8 with NaOH solutions (0.1 or 1 M). The protein concentrations were determined from the optical density at 278 nm, using the value of 9.6 for the specific absorption coefficient. The aggregation of β-lactoglobulin (2 mL) was induced by addition of ethanol (2 mL).

Infrared spectra

The mid infrared spectra were recorded on a Fourier transform spectrometer IFS25 (Bruker) equipped with a DTGS detector. They were recorded between 1580 and 1720 cm^{-1} at 2 cm^{-1} intervals with an ATR cell made of a ZnSe crystal. Each spectrum was the result of 200 scans. The near infrared spectra were recorded between 2000 and 2400 nm (4 nm intervals) either on an InfraAlyzer 500 (Technicon) for the set of model proteins or on NIRSystems 6500 (Foss) for the β-lactoglobulin aggregation. While the spectra of proteins in the solid state were obtained in the reflectance mode, the near infrared spectra of β-lactoglobulin in water / ethanol solution were recorded in the transmittance

Table 1. Percentages of α-helices, β-sheets and unordered structures.

	Code	α-helix	β-Sheet	Unordered
Myoglobin	MYO	78	00	22
Bovine serum albumin	SAB	66	03	31
Lysozyme	LYS	46	19	35
Ribonuclease A	RNA	23	46	31
Chymotrypsinogen A	CHA	12	50	38
α-Chymotrypsin	CHY	11	50	39
Trypsin	TRY	07	45	48
β-Lactoglobulin	BLG	06	46	48
Papain	PAP	28	24	48
Ovalbumin	AC	25	26	49
β-Casein	CASB	10	20	70
κ-Casein	CASK	07	37	56

mode. In the particular case of β-lactoglobulin aggregation, the temperature was controlled and adjusted to 30°C.

Data treatments

Both the mid and near infrared spectra were centred and normalised as follows :

$$X_{i,j} = (S_{i,j} - av_i) / Sd_i$$

where $S_{i,j}$ is the absorbance value at the wavelength j for the spectrum i, av_i is the average absorbance of the spectrum i and Sd_i is the standard deviation of the absorbance values of the spectrum i. $X_{i,j}$ represents the centred and normalised data.

Second derivative of the centred and normalised data were assessed by applying the finite-difference method. The gaps used for the mid and near infrared spectra were 10 cm^{-1} and 10 nm, respectively.

Infrared spectra were studied by applying principal component analysis and generalised canonical correlation analysis. While principal component analysis makes it possible to describe a set of n spectra, the generalised canonical analysis assesses the correlations between k tables of n samples. The generalised canonical correlation analysis allows a comparison and description of n samples by taking k different data tables into account. In the present work, three data tables corresponding to mid infrared spectra, near infrared spectra and reference data of 12 model proteins were jointly analysed.

Results and discussion

Myoglobin, β-lactoglobulin and β-casein are known to present mainly α-helix, β-sheet and unordered structures, respectively. The second derivative spectra obtained in the near infrared region (data not shown) revealed common absorption bands at about 2055, 2170 and 2350 nm. While the absorption band at 2055 nm was previously assigned to a combination of Amide A with Amide II, the peak at 2170 nm corresponds to the combination band of Amide I with Amide III. The absorption band observed at 2350 nm characterises CH stretch / CH deformation. As pointed out by Kamishikiryo-Yamashita *et al*.,[7] the intensity of 2170 nm was higher in myoglobin than in β-lactoglobulin. The intensity of the absorption band observed at 2055 nm for β-casein was smaller than for the other two proteins. In the CH combination region, the three proteins exhibited specific bands. Peaks at 2203, 2267 and 2300 nm were observed in the case of β-lactoglobulin. While myoglobin revealed an absorption band at 2289 nm, β-casein was identified by a peak at 2269 nm. As the differences between myoglobin, β-lactoglobulin and β-casein could be relevant to either the secondary structure or the specific nature of each protein, a generalised canonical correlation analysis, taking mid infrared, near infrared and reference data into account, was applied to the entire set of model proteins. The similarity map of the near infrared spectra discriminated the samples according to their secondary structure (Figure 1). While the first canonical variate made it possible to identify α-helix structure (myoglobin, bovine serum albumin), the second canonical variate separated β-sheet (β-lactoglobulin, trypsin, α-chymotrypsin) and unordered structures (β-casein, κ-casein). The spectral patterns associated with the canonical variates are shown in Figure 2. On the first spectral pattern, negative peaks characteristic of α-helices were observed at 2056, 2172, 2239, 2289 and 2343 nm. The bands at 2172 and 2289 nm had already been observed in the second derivative spectrum of myoglobin. The positive peaks at 2205, 2264 and 2313 nm were tentatively assigned to β-sheet structures. In the particular case of β-lactoglobulin, the second derivative spectrum revealed absorption bands at 2203, 2267 and 2300 nm. The spectral pattern associated with the canonical variate 2, exhibited a positive peak at 2265 nm that was thought to be characteristic of unordered structure.

The β-lactoglobulin aggregation in water / ethanol solution was investigated by recording near infrared spectra every 2 minutes for a 300 minutes period. The final concentrations of β-lactoglobulin

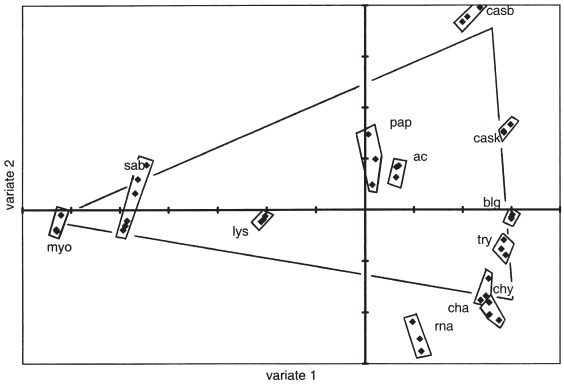

Figure 1. Secondary structure of proteins in the solid state. Similarity map of near infrared spectra defined by canonical variates 1 and 2.

were 5 and 15 mg mL^{-1}. Two separated principal component analyses, corresponding to the 5 and 15 mg mL^{-1} protein concentrations, were applied to the second derivative spectra recorded during the aggregation process. Kinetic curves were obtained by plotting scores of principal component 1 v. time. In the particular case of 15 mg mL^{-1} protein concentration, the kinetic curve indicated that half the aggregates were formed in the first 20 minutes (Figure 3). A similar result was previously obtained by studying in the mid infrared region the intensity increase of 1620 cm^{-1}, an absorption band assigned to inter-molecular β-sheet structures. The spectral pattern associated with principal component 1 opposed peaks at 2174 and 2210 nm (Figure 4). The decrease in 2174 nm was indicative of a decrease in α- helices during β-lactoglobulin aggregation. The peak at 2210 nm was close to an absorption band characteristic of β-sheet structures (2205 nm) and was believed to characterise inter-molecular β-sheets. In its native state, β-lactoglobulin mainly presents intra-molecular β-sheet structures (46%). With the addition of ethanol, the hydrophobic interactions stabilising the native structure are weakened and α-helices are formed. The α-helical structures induced by ethanol are unstable and converted

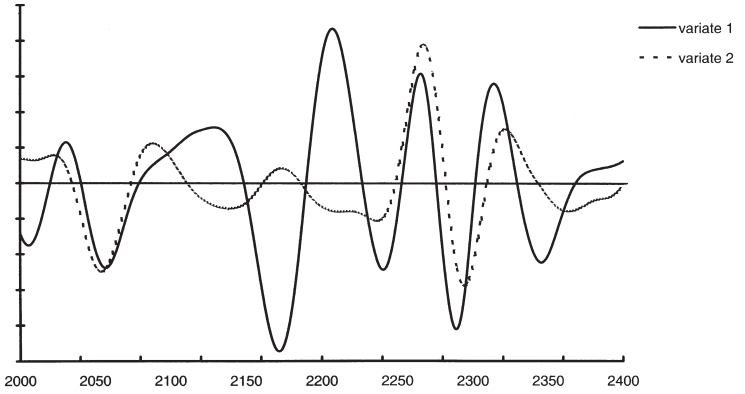

Figure 2. Secondary structure of proteins in the solid state. Spectral patterns associated to canonical variates 1 and 2.

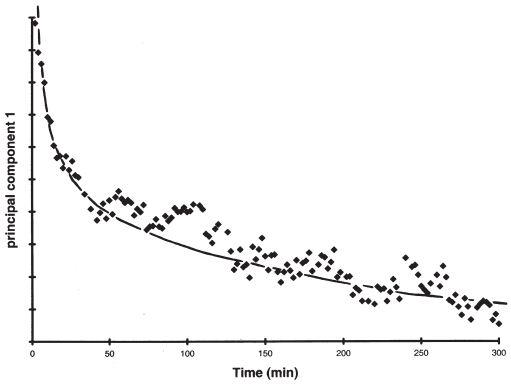

Figure 3. β-lactoglobulin aggregation (15 mg mL^{-1}). Kinetic curve obtained by plotting principal component 1 v. time.

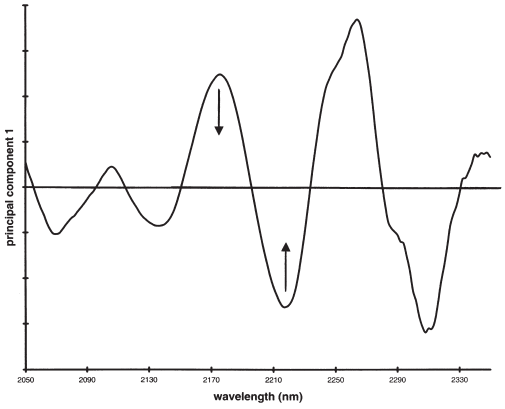

Figure 4. β-lactoglobulin aggregation (15 mg mL^{-1}). Spectral pattern associated to principal component 1.

to inter-molecular β-sheet aggregates. While the decrease in α-helices was responsible for the decrease in the absorption band at 2174 nm, the formation of aggregates might correspond to the development of a band at 2210 nm. At 5 mg mL^{-1} protein concentration, a longer time was necessary for the conversion of α-helices to β-sheet aggregates. Half the aggregates were formed in about 80 minutes.

Conclusion

The potential of near infrared spectroscopy to study the secondary structure of proteins was demonstrated. Specific wavelengths were assigned to α-helix, β-sheet and unordered structures in the 2000–2400 nm region. The conformational changes occurring during the aggregation of β-lactoglobulin in water / ethanol solution were investigated by recording near infrared spectra for a period of 300 minutes. While the disappearance of α-helices corresponded to a decrease of the

2174 nm absorption band, the formation of aggregates, through inter-molecular β-sheets, was responsible for an increase at 2210 nm.

References

1. N. Hirota-Nakaoka and Y. Goto, *Bioorg. Med. Chem.* **7,** 67 (1999).
2. A. Dong, J. Matsuura, M.C. Manning and J.F. Carpenter, *Archives of Biochemistry and Biophysics* **355,** 275 (1998).
3. A.J. Sadler, J.G. Horsch, E.Q. Lawson, D. Harmatz, D.T. Brandau and C.R. Middaugh, *Anal. Biochem.* **138,** 44 (1984).
4. H. Kamishikiryo, K. Hasegawa and T. Matoba, *Nippon Shokuhim Kogyo Gakkaishi* **38,** 850 (1991).
5. R.K. Cho, Y. Ozaki, J.J. Ahn and M. Iwamoto, 5th International Conference on Near Infrared Spectroscopy, in *Near Infrared Spectroscopy, Bridging the Gap between Data Analysis and NIR Applications*, Ed by K.I. Hildrum, T. Isakson, T. Næs and A. Tandberg. Ellis Horwood, Chichester, UK, pp. 333–338 (1992).
6. R.K. Cho, J.H. Lee, Y. Ozaki and M. Iwamoto, 6th International Conference on Near Infrared Spectroscopy, in Leaping Ahead with Near Infrared Spectroscopy, Ed by G.D. Batten, P.C. Flinn, L.A. Welsh and A.B. Blakeney. Near Infrared Spectroscopy Group, RACI, North Melbourne, Australia, pp. 483–486 (1994).
7. H. Kamishikiryo-Yamashita, M. Tatara, H. Takamura and T. Matoba, *Nippon Shokuhim Kogyo Gakkaishi* **41,** 65 (1994).
8. P. Maillart and B. Ribadeau-Dumas, *J. Food Sci.* **53,** 743 (1988).
9. C.A. Zittle and J.H. Custer, *J. Dairy Sci.* **46,** 1183 (1963).
10. G. Brignon, B. Ribadeau-Dumas and J.C. Mercier, *FEBS Lett.* **71,** 111 (1976).