Protein-water interaction monitored by analysis of near infrared spectra

Bogusława Czarnik-Matusewicz[a], Koichi Murayama[b], Yuqing Wu[c], and Yukihiro Ozaki[d]

aFaculty of Chemistry, University of Wrocław, F. Joliot-Curie 14, 50-383 Wrocław, Poland

bDepartment of Physiology, School of Medicine, Gifu University, Gifu 500-8705, Japan

cKey Laboratory for Supramolecular Structure and Materials of the Ministry of Education, Jilin University, 130023, P. R. China

dDepartment of Chemistry and Research Center for Near-Infrared Spectroscopy, School of Science and Technology, Kwansei-Gakuin University, Sanda, 669-1337, Japan

Introduction

There is no doubt that water environment is particularly essential for function and stability of biomolecules. Therefore, of great interest is development of experimental or/and theoretical approaches that allow one to extract valuable information about the properties and structures of both biomolecules and surrounding water. Near-infrared (NIR) spectroscopy should play a significant role in this field as NIR spectroscopy enables one to explore the structure and hydration of biomolecules simultaneously. However, in the NIR region spectral features arising from biomolecules in aqueous solutions often remain hidden in broad envelops built from intense spectral features of water, and therefore, are hard to be interpreted. Our previous studies on protein denaturation by means of NIR spectroscopy have shown that overtone and combination bands of proteins are very useful in monitoring subtle changes in protein structures[1-4]. To investigate the fine intensity changes stemmed from rearranged structures of the proteins the original NIR spectra were analyzed by both exploratory principal component analysis (PCA) and two-dimensional correlation spectroscopy (2DCOS).

Measurements of the mean density of the hydration shell of proteins by x-ray and neutron scattering provide a strong argument against the existence of a first hydration shell with a ≈10% larger average density than that of the bulk solvent[5]. Moreover, the packing of water changes when a protein undergoes denaturation. The disruption of the native conformation of a protein has very strong correlation with perturbation of the structure of water surrounding the protein[6-7]. The close relation between the structure of water and its NIR band shape is well recognized[8]. The question is whether NIR spectroscopy provides new insight into water structure at different stages of unfolding process.

The purpose of this article is to demonstrate the potential of PCA and 2DCOS methods to investigate fine intensity changes in the FT-NIR spectra of aqueous solutions of β-lactoglobulin (BLG), which arise from the rearranged structure of the protein-water system. Analysis of temperature perturbed NIR spectra of BLG aqueous solutions done by the combination of the homo- and hetero- 2DCOS and PCA methods is reported for selected temperatures. The most important conclusion of the study is that unfolding events accompanied by changes in the hydration surface of

Near Infrared Spectroscopy: Proceedings of the 11th International Conference.

protein can be monitored from initial stages until final point through the intensity variations of a band near 7000 cm^{-1} assigned to the first overtone of the stretching vibration of water. Obtained results have revealed that NIR spectroscopy has an important role to play in providing information about water function in protein unfolding.

Experimental

Bovine milk BLG was purchased from Sigma Chemical Co. and was used without further purification. Five different aqueous solutions of BLG with concentrations of 1, 2, 3, 4, and 5 wt % were prepared using a phosphate buffer at pH 6.6. NIR spectra of the BLG solutions were measured with a 2 cm^{-1} resolution with a Magna 760 FTIR/NIR spectrometer equipped with a PbSe detector. To yield a high signal-to -noise ratio, 512 interferograms were coadded.

For investigation of the temperature-dependent hydration process, the spectra were collected with an increment of 5 °C between 25 and 45 °C and with an increment of 2 °C between 45 and 81 °C. A quartz cell of 1 mm pathlength was employed for the experiment. The temperature of solutions with accuracy ±0.2 °C was controlled directly in the cell before each collection of data.

Prior to the PCA and 2DCOS calculations the raw spectra were subjected to following pretreatment steps: offset to common absorption value at 8000 cm^{-1}, truncation into two ranges (4900 - 4180 cm^{-1} and 8000 - 5325 cm^{-1}), smoothing by Lorentzian function of 40 cm^{-1} for the low frequency range and 120 cm^{-1} for the high frequency range. Finally, the spectra were normalized over the protein concentration and arranged into 23 sets for both frequency ranges. All pretreatment steps were performed by GRAMS/32 (Galactic Industries Corp.). Each set corresponding to one temperature consisted of the spectrum of buffer and the five spectra of the BLG solutions with the different concentrations. The dynamic spectra, necessary for the calculation of the 2D correlation spectra were calculated by using the spectrum of buffer as a reference spectrum. The 2D correlation spectra were calculated by using software written in our laboratory[1], while PCA was accomplished by commercially available chemometrics software (Unscrambler 6.1, Camo AS). Usually raw data have a distribution that is not optimal for PCA, and therefore the data have to be preprocessed to extract the meaningful information from the PCA results. In the present case the double-centring procedure was utilized. In the first step the average absorption value calculated for each spectrum was subtracted from its each absorption point. Next, the average absorption value calculated for each wavenumber of the one-centred data was subtracted from appropriate absorption point of every spectrum.

Results and discussion

Figure 1 shows a spectrum of water and the five spectra of BLG-buffer solutions in the (a) 4900 –4180 cm^{-1} and (b) 8000 - 5325 cm^{-1} regions. Strong bands due to water dominate these two regions obscuring almost totally any peaks arising from the protein. The insets show that the absorption changes due to the increase in the concentration of BLG can hardly be detected visually. Bands assigned to BLG can be observed only after subtracting the spectrum of buffer from the spectra of the BLG solutions as shown in the top in Figure 1 (a) and (b). However, the subtraction was executed without any quantitative criterion. This procedure should be performed with an accurate control of the subtraction factor to obtain data suitable for both PCA and 2DCOS. Since so precise subtraction operation cannot be achieved due to the intense absorption of water, the unsubtracted data were explored by the two data analysis methods. Murayama et al.[3] has shown that both 2D correlation analysis and PCA provide essentially similar information about analyzed spectra and could play a decisive role for any doubtful features obtained by one of the two methods.

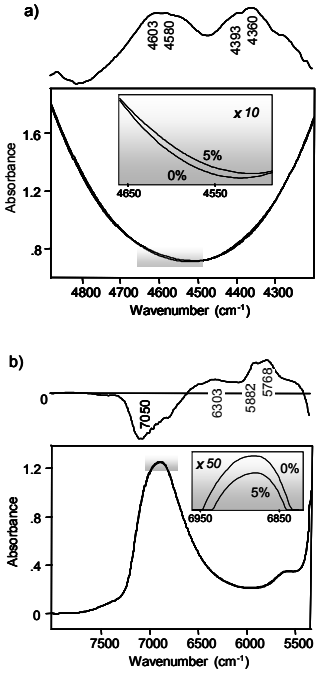

Figure 1. FT-NIR spectra at 25 ºC in the a) 4900 - 4180 cm^{-1} and b) 8000 - 5325 cm^{-1} regions of buffer and 5 wt. % solution of BLG. Insets: Enlargement of the spectra in the 4650-4500 and 6950-6850cm^{-1} regions. Top: Difference spectra.

In Figure 2 the synchronous spectrum in the 4900 - 4180 cm^{-1} region constructed from the concentration-dependent spectral variations at 25 °C is presented. A power spectrum (dashed line) along the diagonal line in the synchronous spectrum and a PC1 loadings plot (solid line) are shown on the top of the synchronous spectrum. Throughout this paper, shaded and light areas represent negative and positive peaks, respectively. The central part of the synchronous map is dominated by two groups of bands. Note that the power spectrum and PC1 loadings plot are very similar to the difference spectrum in Figure 1(a). The synchronous spectrum in Figure 2 shows very close spectrum pattern to those of synchronous spectra of human serum albumin[2] and ovalbumin[4]. Bands around 4600 cm^{-1} have been assigned to the combination of the amide vibrations (amide B/II) and those in 4400-4350 cm^{-1} region to the combination of the stretching/bending mode in the side chains.[2-4,10] Despite the strong screening of the region above 4800 cm^{-1} by a huge absorption arising from the combination mode of water both the power spectrum and the loadings plot have additional peaks at 4820 and 4848 cm^{-1}, respectively. As for the assignment of these peaks, there is some controversy. There are two alternative propositions. One of them assigns these peaks to the combination modes of protein[1,9,10] while the other to the second overtones of bending vibration of water[4]. Additional argument for the second assignment can be obtained from the 2D hetero correlation spectra discussed later in this paper.

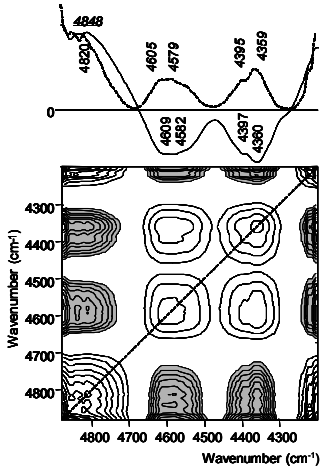

Figure 2. Synchronous 2D NIR correlation spectra in the 4900 - 4180 cm^{-1} region with a power spectrum along the diagonal line (dashed line) and a PC1 loadings plot (solid line) on the top of the 2D spectrum at 25 °C.

Figure 3 depicts the synchronous spectrum together with a power spectrum along the diagonal line (dashed line) and a PC2 loadings plot constructed from the concentration-perturbed NIR spectra in the 8000 - 5325 cm^{-1} region. The loadings plot enables to identify independent absorption areas controlled by protein and water vibrations. The positive area corresponds to the bands of different water species (first overtone of water stretching mode), while the negative part is concerned with the absorptions attributed to the protein vibrations[1-4]. Two negative features at 5768 and 5882 cm^{-1} are assigned to the first overtone of the C-H stretching modes. Mainly the first overtone of the amide A mode is responsible for the negative peaks at 6342 and 6473 cm^{-1}. These peaks observed in the loadings plot have counterparts in the power spectrum.

Of note is the corresponding asynchronous spectrum shown in Figure 4 where a slice spectrum extracted at 6996 cm^{-1} is also presented on the top. The slice spectrum develops two negative peaks at 6726 and 7124 cm^{-1} assignable to different species of water. The asynchronous spectrum uncovers the heterogeneity of the water structure. As one can see, there are distinct two types of differently bonded molecules. Due to the difference in strength of hydrogen bonds two distinct cross-peaks are observed in the asynchronous spectrum. The balance of the water-water interaction itself is disrupted by the introduction of protein molecules. In addition, the water bands have strong asynchronicity against the C-H bands of BLG, showing strong cross-peaks visible in Figure 4.

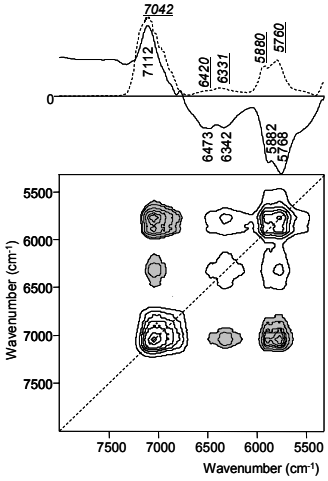

Figure 3. Synchronous 2D NIR correlation spectrum in the 8000 – 5325 cm^{-1} region with a power spectrum along diagonal line (dashed line) and a PC2 loadings plot (solid line) on the top.

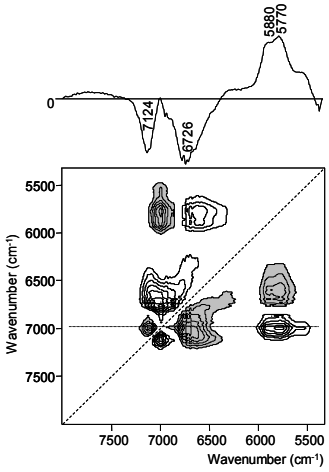

Figure 4. Asynchronous 2D NIR correlation spectrum in the 8000 – 5325 cm^{-1} region with a slice spectrum at 6996 cm^{-1} on the top.

The absorption variations in the high frequency region provide much more details concerning the changes in water structure due to the interaction with protein than the low frequency region. However, the 4900 - 4180 cm^{-1} region plays a very important role in the 2D heterospectral correlation spectroscopy because the intensity variations in the spectral region dominated by protein bands (4900 - 4180 cm^{-1}) are confronted with the intensity changes in the region almost totally controlled by water absorption (8000 – 5325 cm^{-1}). The heterospectral correlation analysis in studies of protein systems was previously adopted by Schultz et al.[9] to investigate correlations between bands in the mid-IR and NIR regions. 2D heterospectral analysis was also attempted for the IR and Raman data of the amide III region of the same aqueous solutions of BLG. Detailed band assignments for the amide III regions have been proposed[11]. In the present study different regions chosen from the IR spectra are subjected to the 2D heterospectral analysis. Such approach enables elegant separation of the bands into those attributed to protein and those to different water species.

Positive peaks in a synchronous heterospectral correlation map highlight locations of bands attributed to the same components of a system, whereas negative peaks point absorption ranges shared by bands due to diverse components. Figure 5 presents the synchronous heterospectral correlation spectra calculated from the NIR data at 25 °C. Examining the synchronous 2D heterospectral correlation map allows one to assign bands to the protein or water. In the low frequency region the unambiguous protein peaks are located around 4360 and 4605 cm^{-1}. According

to the above rule, intensive positive cross-peaks at (4360, 6332), (4360, 5882), and (4360, 5768) cm^{-1} reveal that the bands at 6332, 5882, and 5768 cm^{-1} are assigned to the vibrations of BLG. In contrast, negative cross-peak at (4360, 7045) cm^{-1} elucidates that the band at 7045 cm^{-1} is assigned to water. A positive cross-peak emerged at (4841, 7045) cm^{-1} reveals that the peak at 4841 cm^{-1} should be assigned to water species. Detailed analysis of its position as a function of temperature should clear remaining uncertainties as to its assignment. It is still unresolved whether this band arises from the combination mode of water or the second overtone of its bending vibration.

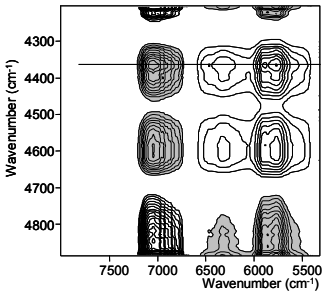

Figure 5. Synchronous 2D NIR heterospectral correlation spectra in the 4900 - 4180 cm^{-1} and 8000 – 5325 cm^{-1} regions at 25°C.

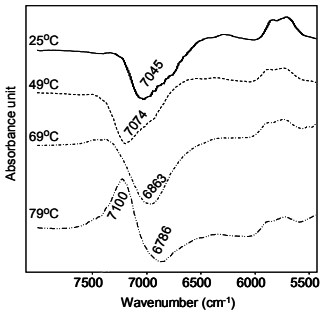

Figure 6. Slice spectra extracted at 4360 cm^{-1} from the synchronous 2D NIR heterospectral correlation spectra at selected temperatures.

Very informative is the comparison of slice spectra extracted at 4360 cm^{-1} from the synchronous 2D heterospectral correlation maps at different temperatures, i.e, at different stages of unfolding process. Figure 6 shows slice spectra extracted at 4360 cm^{-1} for selected four temperatures. These spectra give a view on the absorption changes in the region controlled by water structure. From the analysis of the slice spectra the protein-water interaction can be explored. An explanation of the role of water in the thermal unfolding process has a fundamental meaning in the understanding of this process. The structure of water in this intriguing process should be considered as a mixture of two structural domains whose numbers in liquid phase change with temperature and pressure[12]. According to the model liquid water is composed of dynamically transforming microdomains of two very different bonding types. One type is the regular tetrahedral water-water bonding similar to that in ordinary ice Ih, whereas the other is a more dense nonregular tetrahedral bonding similar to that in ice II. This balance undergoes also changes due to interaction of water with groups of different polarity. Generally, protein is composed of a large number of amide and side chains groups with different polarity. Depending on whether the protein is in its native or its unfolded state different kinds of groups are exposed to water. Therefore, the process of the protein unfolding can be monitored by the absorption changes of water bands. The NIR spectral variations attributed to water absorption can serve as an efficient tool for exploring the protein folding.

Coming back to the analysis of the spectra in Figure 6, of note is the obvious changes of the position and shape of the peak due to water. From 25 °C to 49 °C this peak gradually shifts to a higher frequency most probably mainly due to a direct effect of the temperature increase on water structure. With temperatures increase the balance of water structure is shifted into the direction of more disordered structure characterized by weaker hydrogen bonds. Probably at this stage temperature has stronger influence than protein in modification of the structure of water. However, at higher temperatures, this peak is no longer shifted into a higher frequency with temperature increase. In consequence of protein unfolding that starts above 50 °C the structure of water is also

altered. This effect of the increasing exposition of the non-polar groups towards water triggers the shift of the water peak into the opposite direction. Further heating of the BLG solutions leads to next stages of unfolding that cause additional changes in the protein structure. The water-protein interaction on one hand and increasing temperature on the other hand have contrary effects on water properties. The slice spectrum at 79 °C uncovers very clearly the heterogeneity of the water structure. As one can see, there are distinct two sub-peaks. A band at 6786 cm^{-1}, according to the two-state model of water[12], is assigned to the regular tetrahedral water-water bonded molecules, whereas the other band at 7100 cm^{-1} to a more dense nonregular tetrahedral configuration. It is noted that the cross-peaks observed in the asynchronous spectrum (Figure 4) and the peaks detected in the slice spectra both confirm the concept of heterogeneous structure of water in specific pathways modified by the interaction with proteins.

Important conclusion of this paper is that NIR spectra of aqueous solution of protein examined by the two exploring methods (2DCOS and PCA) have a great potential to establish stages of protein unfolding through the analysis of the absorption changes in the region of water absorption. Additionally, it was shown that the bands assignments can be essentially enhanced by employing 2D heterospectral correlation spectroscopy to the analysis of NIR spectra.

References

1. Y. Wang, K. Murayama, Y. Myojo, R. Tsenkova, H. Hayashi, Y. Ozaki, *J. Phys. Chem.* **102**, 6655 (1998).
2. Y. Wu, B. Czarnik-Matusewicz, K. Murayama, and Y. Ozaki, *J. Phys. Chem. B* **104**, 5840 (2000).
3. K. Murayama, B. Czarnik-Matusewicz, Y. Wu, R. Tsenkova, and Y. Ozaki, *Appl. Spectrosc.* **54**, 978 (2000).
4. K. Murayama and Y. Ozaki, *Biospectrosc.* **67**, 394 (2002).
5. D.I. Svergun, S. Richard, M.H.J. Koch, Z, Sayers, S. Kuprin, and G. Zaccai, *Proc. Natl. Acad. Sci. USA* **95**, 2267 (1998).
6. T. Head-Gordon, Proc. *Natl. Acad. Sci. USA* **92**, 8308 (1995).
7. A. Pertsemlidis, A.M. Saxena, A.K. Soper, and T. Head-Gordon, R.M. Glaeser, *Proc. Natl. Acad. Sci. USA* **93**, 10769 (1996).
8. V.H. Segtnan, S. Šašic, T. Isaksson, and Y. Ozaki, *Anal. Chem.* **55** 29 (2001).
9. C.P. Schultz, H. Fabian, H.H. Mantsch, *Biospectrosc.* **4**, S19 (1998).
10. J. Wang, M.G. Sowa, M.K. Ahmed, and H.H.Mantsch, *J. Phys. Chem.* **98**, 4748 (1994).
11. Y.M. Jung, B. Czarnik-Matusewicz, and Y. Ozaki, *J. Phys. Chem. B* **32**, 7812 (2000).
12. G. W. Robinson, J. Urquidi, S. Singh, and C.H. Cho, *Cell. Mol. Biol*. **47**, 757 (2001). (and citations herein)