Analysis of the Protein Behavior in Molecular Crowding Environments Investigated by Raman Spectroscopy

Chikashi Ota

2016

Chapter 1. General Introduction

1-1. Behavior of a Protein Molecule in the Molecular Crowding Environments

Proteins are one of the most important biomolecules for life, and each protein plays a crucial, unique role in a crowded environment such as a cell or other biological environments.[1-3] However, in vitro studies of proteins are often conducted in dilute solution using low protein concentrations. Therefore, characterizing the state of proteins in such crowded environments is an essential subject that has not been completely understood. Some reports have shown that the molecular behaviour of proteins in crowded environments is different from that in vitro (the ideal solution).[4-6] The comportment of proteins in crowded environments is dominated mainly by protein-crowder and protein-protein interactions. The former type of interaction is primarily an entropic effect such as exclude volume,[2] whereas the latter interaction is primarily a soft chemical interaction involving, for example, hydrogen bonding or van der Waals interactions.[6] Nonetheless, while these interactions have been defined, the details of these interactions are not fully revealed.

1-2. Highly Concentrated Liquid Formulation for the Antibody Solution

Understanding of the interactions between proteins in the crowding environment is also getting important from the industrial point of view. Recently, many biopharmaceuticals have got strong potential as drugs for their high efficacy, and in the biopharmaceutical industry, liquid formulations are becoming increasingly popular.[7] For

such formulations, highly concentrated (>100 mg/mL) protein solutions are required. At high concentrations, however, the increased viscosity is accompanied by an increased risk of protein aggregation or denaturation. These changes may result in side effects, and therefore the optimization of drug formulations is of great importance. Accordingly, a comprehensive understanding of the interactions of proteins in nonideal solutions (e.g., highly concentrated solutions) is essential. In particular, protein–protein interactions that could cause oligomerization or aggregation in highly concentrated solution should be thoroughly investigated.

Currently, relatively few techniques are available to characterize protein-protein interactions in high-concentration solutions. The second virial coefficient (B_{22}), as an indicator of deviation from ideality, or the interaction parameter (k_D), as deduced from dynamic light scattering (DLS), can give information about protein-protein interactions. However, since short range attractive interactions between the surfaces of proteins would be expected to increase at highly concentrated solution, the approach on the basis of the colloidal model cannot be applied to the highly concentration study.[7,8] Therefore, the colloidal approximation can be available only in the case when the concentration is relatively low and the molecules are mono-disperse, while this approach is powerful to describe the behavior of molecules simply

In highly concentrated solutions, the interactions between molecules are highly complex and the situation is not easy to understand. At such concentrations, conformational information or knowledge about the interactions of each functional group is getting more important than lower concentrations to elucidate how the crowding effect works and stresses on the molecules. For this purpose, the direct analytical method for the conformational knowledge is highly required.

1-3. Protein Structure Analysis by Raman Spectroscopy

Small-angle X-ray scattering and small-angle neutron scattering are powerful analytical methods for investigating the structure of molecules in high-concentration solutions directly,[9,10] however the knowledge from this method is the size of the molecule on the colloidal model and it is difficult to obtain information about protein-protein interactions by using these techniques. In contrast, Raman spectroscopy is a powerful method for the study of highly concentrated solutions that can provide conformational and interaction information about each functional group, especially of the aromatic residue such as Phe, Tyr and Trp of proteins.[11,12]

Raman spectroscopy is one of the vibrational spectroscopies and a valuable method for investigating the structures and interactions of the molecules.[13,14] Since the Raman scattering of water is relatively lower than the organic molecules due to the optical selection rule of this method, the spectrum can be observed with higher sensitivity than water background. In addition, Raman spectroscopy has an advantage that it can reveal the detail behavior of molecules from functional groups level. Consequently, detailed information about each type of amino acid residue can be obtained from the Raman spectrum of a protein.

In the structure analysis of proteins by Raman spectroscopy, amide I or III band is often focused to investigate the secondary structures, because the amide I or III band of proteins is sensitive to the strength of hydrogen-bonding interactions (C=O···HN), and then the information about protein secondary structures can be obtained from the location of this band.[13,14] In addition, unique information of proteins can be received by specific bands from disulfide bonds or aromatic residues, Trp, Tyr and Phe.[15-19] These

bands are clearly observed in the Raman spectra, since the optical selection rule of Raman effect is originally from the change of the polarizability induced by the incident light. In contrast, these bands are silent in infrared spectroscopy due to the different selection rule.

Although the Raman spectroscopy could work powerfully for the protein analysis, it has the disadvantage that the Raman scattering itself is relatively lower compared to other spectroscopy. In fact, it is difficult to measure Raman spectra at low concentrations (<10 mg/mL), especially for non-resonance scattering.[13] However, when we focus on the highly concentrated solution (>10 mg/mL) only, this disadvantage can turn into a great advantage, since the Raman spectrum is not disturbed by optical effects such as the multi-scattering or re-absorption effect even in the highly concentrations over 100 mg/mL. (Other spectroscopic techniques are not available for the highly concentrated solution over 100 mg/mL.) In this manner, Raman spectroscopy is one of the best methods for analyzing high-concentration solutions.

The relationship between Raman band characteristics (wavenumber, band intensity ratio, and full width at half-maximum) and the protein secondary structure were well established from a large number of previous studies, however, these assignment is only in the case when the experimental condition is in a simple model system and the interpretation often lacks the environmental effect such as the molecular crowding effect.

For these reasons, the aim of the study in this thesis is to establish a methodology for the understanding of the protein behavior in the crowding / highly concentrated solution by using Raman spectroscopy and to describe their behavior in detail. In addition, using the established method, the practical application for the evaluation of the

protein-protein interaction in the highly concentrated antibody solution has been suggested.

1-4 Organization of This Thesis

This thesis consists of three chapters as described below.

In Chapter 2, the structure and interactions of lysozyme molecules were investigated over a wide range of concentrations (2.5–300 mg/mL) by using Raman spectroscopy to elucidate the marker bands and establish the fundamental method. In Chapter 3, using the established method in Chapter 2, the concentration-dependent measurements of BSA solutions at pH7.0 and 3.0 by Raman spectroscopy were carried out to discuss the entropic effect and the soft chemical interaction under these different conditions as an advanced study. In addition, H–D exchange measurements in the presence of another protein as a molecular crowder were performed to further characterize the entropic effect. In Chapter 4, on the basis of the knowledge of the studies of a model system in Chapter 2 and 3, the protein-protein interactions were discussed in a highly concentrated antibody solution as a practical system and the application studies were performed to show the practical use.

1-5. References

1. Ellis, R. J., Minton, A. P. Cell biology: Join the crowd. *Nature*, **2003**, *425*, 27-28.

2. Minton, A. P. Excluded volume as a determinant of macromolecular structure and reactivity. *Biopolymers*, **1981**, *20*, 2093-2120.

3. Harada, R., Sugita, Y., Michael, F. Protein crowding affects hydration structure and dynamics. *J. Am. Chem. Soc*. **2012**, *134*, 4842-4849.

4. Inomata, K., Ohno, A., Tochio, H., Isogai, S., Tenno, T., Nakase, I., Takeuchi, T., Futaki, S., Ito, Y., Hiroaki, H., Shirakawa, M. High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature*, **2009**, *458*, 106.

5. Feig, M., Sugita, Y. Variable interactions between protein crowders and biomolecular solutes are important in understanding cellular crowding. *J. Phys. Chem. B*, **2012**, *116*, 599-605.

6. Miklos, A. C., Sarkar, M., Wang, Y., Pielak, G. J. Protein crowding tunes protein stability. *J. Am. Chem. Soc.* **2011**, *133*, 7116-7120.

7. Uchiyama, S. Liquid formulation for antibody drugs. *Biochim. Biophys. Acta*, **2014**, *1844*, 2041-2052.

8. Yadav, S., Liu, J., Shire, S. J., Kalonia, D. S. Specific interactions in high concentration antibody solutions resulting in high viscosity. *J. Pharm. Sci*. **2010**, *99*, 1152-1168.

9. Stradner, A., Sedgwick, H., Cardinaux, F., Poon, W. C. K., Egelhaaf, S. U., Schurtenberger, P. Equilibrium cluster formation in concentrated protein solutions and colloids. *Nature*, **2004**, *432*, 492-495.

10. Godfrin, P. D., Zarraga, I. E., Zarzar, J., Porcar, L., Falus, P., Wagner, N. J., Liu, Y. Effect of hierarchical cluster formation on the viscosity of concentrated monoclonal antibody formulations studied by neutron scattering. *J. Phys. Chem. B,* **2016**, *120*, 278-291.

11. Takeuchi, H. UV Raman markers for structural analysis of aromatic side chains in proteins. *Anal. Sci*, **2011**, *27*, 1077-1086.

12. López-Peña, I., Leigh, B. S., Schlamadinger, D. E., Kim, J. E. Insights into protein structure and dynamics by ultraviolet and visible resonance Raman spectroscopy. *Biochemistry*, **2015**, *54*, 4770-4783.

13. Wen, Z. Raman spectroscopy of protein pharmaceuticals. *J. Pharm. Sci*. **2007**, *96*, 2861-2871.

14. Oladepo, S. A., Xiong, K., Hong, Z., Asher, S. A., Handen, J., Lednev, I. K. UV resonance Raman investigations of peptide and protein structure and dynamics. *Chem. Rev*. **2012**, *112*, 2604-2628.

15. Siamwiza, M. N., Lord, R. C., Chen, M. C., Takamatsu, T., Harada, I., Matsuura, H., Shimanouchi, T. Interpretation of the doublet at 850 and 830cm^{-1} in the Raman spectra

of Tyrosyl residues in proteins and certain model compounds. *Biochemistry*, **1975**, *14*, 4870-4876.

16. Takeuchi, H., Harada, I. Normal coordinate analysis of the indole ring. *Spectrochim. Acta*, **1986**, *142A*, 1069-1078.

17. Miura, T., Takeuchi, H., Harada, I. Characterization of individual tryptophan side chains in proteins using Raman spectroscopy and hydrogen deuterium exchange kinetics. *Biochemistry*, **1988**, *27*, 88-94.

18. Xu, M., Ermolenkov, V. V., He, W., Uversky, V. N., Fredriksen, L., Lednev, I. K. Lysozyme fibrillation: deep UV Raman spectroscopic characterization of protein structural transformation. *Biopolymers*, **2005**, *79*, 58-61.

19. Xu, M., Shashilov, V. A., Ermolenkov, V. V., Fredriksen, L., Zagorevski, D., Lednev, I. K. Hen egg white lysozyme fibrillation:_a deep UV resonance Raman spectroscopic study. *J. Biophotonics*, **2008**, *1*, 215-229.

Chapter 2. The Molecular Interaction of a Protein in Highly Concentrated Solution Investigated by Raman Spectroscopy

1. Introduction

Understanding of the interactions between proteins in the crowding environment is also getting important not only from the scientific point of view but also from the industrial one. Information about the behavior of proteins in non-ideal solutions may suggest ways to improve the stability and increase the shelf life of biopharmaceuticals. In the biopharmaceutical industry, high-concentration (>100 mg/mL) protein formulations are required.[1] At high concentrations, the increased viscosity is accompanied by increased risk of protein aggregation or denaturation. These changes may result in side effects, so evaluating the function of proteins under such conditions is important. In particular, the behavior of functional groups involved in protein function should be investigated in detail. Small-angle X-ray scattering and small-angle neutron scattering are powerful analytical methods for investigating the structure of molecules in high-concentration solutions, but obtaining information about individual functional groups by means of these techniques is difficult.[2–5]

Raman spectroscopy is a powerful tool for detailed investigation of the behavior of functional groups in proteins. Specifically, Raman spectroscopy has two important advantages: information about each type of amino acid residue can be obtained from the Raman spectrum of a protein, and Raman spectroscopy is one of the best methods for analyzing high-concentration solutions. In fact, measuring Raman spectra at low concentrations (<10 mg/mL) is difficult, especially for non-resonance scattering.

In this study, we used Raman spectroscopy to analyze high-concentration solutions of hen egg-white lysozyme as a model protein. In particular, by investigating the relationship between the behavior of selected amino acid residues and the distance between lysozyme molecules, we were able to obtain detailed information about the structure and interactions of the molecules at high concentrations. In addition, from a practical point of view, we identified marker bands that can be used to evaluate high-concentration solutions of the protein.

2. Experimental Methods

2-1. Materials

Hen egg-white lysozyme was purchased from Wako Pure Chemical Industries (cat. no. 100834, Osaka, Japan). A buffer solution of lysozyme was prepared by mixing the enzyme with tris(hydroxymethyl)aminomethane (CAS no. 1185-53-1, lot no. 35433, Nacalai Tesque, Kyoto, Japan) and 35% HCl (code 18320-15, 500 ml, lot no. V5P7712, Nacalai Tesque). The concentration of the buffer solution was 20 mM, and the final pH of the solution was adjusted to 8.0 by means of a LAQUA F-72 pH meter (HORIBA, Kyoto, Japan). Test solutions of lysozyme were prepared at concentrations of 2.5, 5, 10, 20, 50, 100, 200, and 300 mg/mL by dilution of the buffer solution with water purified with a Milli-Q laboratory water purifier (Millipore, Billerica, Massachusetts, USA). The solution of lysozyme was passed through a Millipore SLLGH13NL filter (pore size, 0.2 μm) to remove residual contaminants. All measurements were performed at room temperature (~20 °C).

2-2. Raman Spectroscopy

Raman spectra were obtained with an ARAMIS Raman microscope equipped with Labspec 5 software (HORIBA Jobin Yvon, Paris, France). The spectrograph of the microscope had the Czerny–Turner configuration, and the focal length was 460 mm. The optical configuration was the back-scattering geometry. A thermoelectrically cooled Synapse charge-coupled device camera (HORIBA) was used as the detector.

The entrance slit of the spectrometer was set to 100 μm. The dispersive element had a grating of 600 lines/mm, which provides a wavenumber resolution of ca. 3 cm^{-1}. Excitation was accomplished with the 632.8 nm line of a He–Ne laser, and the laser power was 6 mW at the sample surface. The excitation intensity was monitored with a TQ8210 optical power meter (Advantest, Tokyo, Japan). The scattered light was collected through a multipass cell holder (HORIBA Jobin Yvon) with a visible macro lens (focal length, 40 mm) for high sensitivity. The optical cell was a 10 mm × 10 mm quartz cell (Hellma Analytics, Müllheim, Germany). The Raman shift was calibrated by means of the atomic emission lines of a neon lamp. Classical least-square calculations and Gaussian–Lorentz fitting calculations were carried out with the Labspec 5 software.

3. Results and Discussion

3-1. Raman Spectrum of Lysozyme

In the wavenumber region from 200 to 2000 cm^{-1} of the Raman spectrum of lysozyme (Fig. 1a, 20 mg/mL), there are many bands for each functional group, and from this “fingerprint region,” information about changes in the conformations and microenvironments of amino acid residues can be obtained. The relationship between Raman band characteristics (wavenumber, band intensity ratio, and full width at half-maximum) and the protein secondary structure are well established.[6,7] The strong

and broad bands in the wavenumber region from 2800 to 4000 cm^{-1} of the spectrum of lysozyme were assigned to C–H and O–H vibrations of lysozyme and water, respectively (Fig. 1b). As the lysozyme concentration was increased, changes in several of the Raman bands were observed (Fig. S1), as described in detail below.

3-2. Amide I Band

The location of the amide I band of proteins depends mainly on the strength of hydrogen bonding interactions (C=O···H) involving amide groups and the strength of dipole–dipole interactions between carboxyl groups. Thus, the location of this band provides information about protein secondary structure.[8,9] In the Raman spectrum of lysozyme, the amide I band overlaps with the band for the bending mode of water at 1640 cm^{-1}, and we subtracted the contribution of the bending mode band, which we obtained by calculating the ratio of the CH stretching mode of lysozyme at 2940 cm^{-1} and the OH stretching mode of water at 3420 cm^{-1}. In the lysozyme concentration range from 10 to 300 mg/mL (Fig. 2), the location of amide I was not markedly affected by concentration, which implies that the overall structure of lysozyme was unaffected by concentration.

3-3. Width of the Trp Band

The Raman band for the ring-breathing mode of the indole ring of Trp appears at 760 cm^{-1} (Fig. 3a).[10] We found that the width of the band did not vary with concentration in the range from 2.5 to 300 mg/mL (Fig. 3b), which means that the structure of the indole ring was unaffected by changes in concentration.

The Raman band at 1555 cm^{-1} is due to the C=C stretching vibration of the

C_2=C_3–C_β–C_α moiety of the indole ring (Figs. 3c and S2), and the width of this band is a conformational marker and provides information about the torsion angle ($\chi^{2,1}$) of the C_2=C_3–C_β–C_α moiety.[11] In contrast to the width of the 760 cm^{-1} band, the width of the 1555 cm^{-1} band increased with increasing concentration up to 100 mg/mL (Fig. 3d). This result indicates that as the concentration increased, molecular interactions, such as long-range repulsion, between lysozyme molecules increased, which caused $\chi^{2,1}$ to change. We expected the band to continue to increase in width at concentrations of >100 mg/mL, but no such increase was observed. One possible reason for this result is that in this concentration region, the protein existed in a metastable state.

3-4. Ratio of Tyr and Trp Band Intensities

In the Raman spectra of lysozyme in the wavenumber region from 810 to 960 cm^{-1} (Fig. 4a, 10–300 mg/mL; the spectra for the 2.5 and 5 mg/mL solutions are excluded from the discussion because of the low signal-to-noise ratio), the doublet at 856 and 837 cm^{-1} is due to Fermi resonance between the ring-breathing vibration of Tyr and the overtone of an out-of-plane ring-bending vibration of the Tyr benzene ring. The ratio of the intensity of the two bands of the doublet (I_{856}/I_{837}) is known to be a marker of hydration.[12] A low I_{856}/I_{837} ratio indicates that the OH group of Tyr acts as a strong hydrogen donor, and an increase in the ratio indicates that the OH group is acting as both a hydrogen acceptor and a hydrogen donor.

We compared the Raman spectra in the region from 820 to 890 cm^{-1} for 20 and 300 mg/ml solutions of lysozyme and found that the I_{856}/I_{837} ratios at the two concentrations were different (Fig. 4b). To calculate the ratio at each concentration, we fitted the 837 and 856 cm^{-1} bands and four additional bands (870, 877, 900, and 935 cm^{-1}) with a

Gaussian–Lorentz function, and normalized the bands by the intensity of the 877 cm^{-1} band (the fitting results are indicated by the blue curves in Fig. 4a). Comparison of the I_{856}/I_{837} ratios at concentrations from 10 to 300 mg/mL showed that the ratio increased up to a concentration of 100 mg/mL (Fig. 4c). This result implies that in the lower-concentration region, the OH group of Tyr acted as a hydrogen donor, mainly to the OH^- ions that are present in water at pH8. In the high-concentration region, however, as the distance between the lysozyme molecules decreased, neighboring protein molecules surrounded the OH group of Tyr. In this altered microenvironment, the OH group acted as an acceptor of hydrogens from electropositive donors (such as the side chain NH_3^+ groups of amino acid residues of neighboring protein molecules). At concentrations of >100 mg/mL, the I_{856}/I_{837} ratio of Tyr remained constant, indicating that most of the protein molecules interacted with the side chains of neighboring protein molecules in this concentration range.

The N_1–H site of the indole ring of Trp can act as a proton donor. When this site is strongly hydrogen bonded, the band at 877 cm^{-1}, which is attributed to a benzene υ12-like vibration from the phenyl ring coupled with N_1–H bond motion,[13] shifts to a lower wavenumber.[14,15] We found that although the position of the band at 877 cm^{-1} of lysozyme did not change with concentration, the band broadened toward the lower wavenumber region as the lysozyme concentration was increased (Fig. 4b). This result reveals that at high concentration, the hydrogen bonding around the N_1–H site got stronger. We found that the ratio of the intensities of the Trp bands at 870 and 877 cm^{-1} (I_{870}/I_{877}) increased with increasing concentration up to 100 mg/mL (Fig. 4c). This result indicates that the microenvironment around Trp at high concentration differed from that at low concentration. At low concentration, the N_1–H site was surrounded mainly by

water, so the site acted as a hydrogen donor, mainly to OH^- present at pH8. In contrast, at high concentration, the distance between the lysozyme molecules decreased, and the N_1–H site was surrounded by other protein molecules. As a result, the site acted both as a hydrogen acceptor from electropositive donors such as –OH and NH_3^+ groups of amino acid residues and as a hydrogen donor to electronegative acceptors such as $–COO^-$ groups of amino acid residues. Our results indicate that the microenvironment of the N_1–H site strengthened the hydrogen bonding around the site. In other words, the hydrogen bonding interaction between Trp and water was shifting to the side chain interaction with neighborhood protein.

At concentrations of >150 mg/ml, the I_{870}/I_{877} ratio remained constant, which implies that the lysozyme molecules could not get any closer at concentrations of >150 mg/ml and thus that the environment around the N_1–H site did not change much.

3-5. CH and OH Stretching Region

Bands for CH and OH stretching appear in the high-wavenumber region (2600–4000 cm^{-1}, Fig. 5). As the lysozyme concentration was increased, the intensity of the CH stretching bands (2940 cm^{-1}) increased. Therefore, we expected that lysozyme concentration would be linearly related to the ratio of the intensities of the CH and OH stretching bands. We used the CH_3 band at 2940 cm^{-1} as representative of the amount of lysozyme and the OH band at 3420 cm^{-1} as representative of the amount of water. Although water is the main contributor to the OH band at 3420 cm^{-1}, lysozyme also contains OH groups, and their contribution to the band must be taken into account. To precisely calculate the ratio of the intensity of the lysozyme CH stretching band and the water OH stretching band (I_{2940}/I_{3420}) while considering the contribution of the

lysozyme OH band, we used a classical least-squares (CLS) analysis method.[16]

The error in the ratios calculated by means of the CLS method was less than 0.2% at all the concentrations. At concentrations up to 50 mg/mL, the ratio increased linearly with increasing concentration (Fig. 6a; the line in the figure was calculated by the CLS method with data for the 0–50 mg/mL concentration range). However, at concentrations of >100 mg/mL, the relationship was no longer linear. Previous results have also shown that there is some deviation from linearity at concentrations of >150 mg/mL.[17]

The difference (ΔR) between the measured value (R_{real}) and the calculated value (R_{linear}) is plotted against concentration in Fig. 6b. The reason for the non-linearity may have to do with the difference in the volume (number) of molecules in the excitation volume at low and high concentrations. The Raman intensity ratio may be linearly related to the volume (number) of lysozyme molecules in the excitation volume at low concentration, because the volume of lysozyme molecules in the excitation volume is negligible; whereas at high concentration, the volume of molecules in the excitation volume is not negligible. In other word, that the numbers of water molecules and lysozyme molecules are not linearly related to the mass concentration and therefore that I_{2940}/I_{3420} is not linearly related to the mass concentration.

To evaluate this possibility, we changed the horizontal axis from mass concentration of lysozyme to volume fraction of lysozyme (Φ_L):

$$\Phi_L = \frac{V_L}{V_w + V_L}, \tag{1}$$

where V_{w} and V_{L} are the volumes of water and lysozyme, respectively. For the density of native lysozyme (d_{L}), which is necessary to change the axis, we used a value of 1.4 g/cm^3 (Fig. 6c).[18] (The black line in Fig. 6c was calculated from the data point up to 50 mg/mL by means of the CLS fitting method in the region from 0 to 50 mg/mL.) The

resulting plot still deviated from linearity at concentrations of >200 mg/mL. To reproduce the measured values, we introduced the coefficient N_L:

$$\Phi_L = \frac{N_L V_L}{V_w + N_L V_L} \tag{2}$$

The coefficients for the 200 and 300 mg/mL concentrations were 1.04 and 1.09, respectively.

3-6. The Distance between Lysozyme Molecules

To understand the molecular structure of lysozyme at high concentration, we focused on characteristic Raman bands that are sensitive to changes in the structure and microenvironment around amino acid residues, such as changes in hydrogen bonding. Our analysis of the amide I band of lysozyme did not show any significant changes in the overall molecular structure with changes in concentration, but the Raman bands of Trp and Tyr, which are sensitive to hydrogen bonding, showed characteristic changes as the protein concentration was increased.

To evaluate molecular interactions at high concentration, we must consider the relationship between the changes in the Raman spectra and the distance between the protein molecules. To estimate the distance between lysozyme molecules at each concentration, we calculated the mean nearest-neighbor distance[19,20] on the basis of particle size and particle dimension number. For this calculation, lysozyme was approximated by a simple hard spherical particle with a radius (R) of 2 nm.[21] The mean interparticle surface separation $\langle r \rangle$ for non-interacting hard spheres is given by

$$\langle r \rangle = 2R(\lambda - 1), \tag{3}$$

where λ is the mean distance between the centers of mass of nearest neighbors obtained

from a conditional pair distribution function.[19,20] Mean distance λ can be expressed in terms of the volume fraction (ϕ),

$$\lambda = 1 + \frac{(1-\phi)^3}{24\phi(1-\phi/2)}, \tag{4}$$

which is given by

$$\phi = (\frac{4\pi R^3}{3M}) N_A C, \tag{5}$$

where M is molecular weight (14,000) and C is protein concentration (mg/mL).

We plotted the calculated $\langle r \rangle$ values for lysozyme molecules as a function of concentration (Fig. S3). At concentrations up to 50 mg/mL, the distance between lysozyme molecules decreased exponentially, whereas above 50 mg/mL, $\langle r \rangle$ decreased more slowly. At about 50 mg/mL, the distance between lysozyme molecules was approximately the same order of magnitude as the size of the protein. This result is consistent with previous results indicating that the distance between serum albumin molecules is equal to the size of the molecules at concentrations of >50 mg/mL; note however that albumin is not exactly the same size and shape as lysozyme.[1]

When we plotted the width of the Trp band at 760 cm^{-1} as a function of the distance between molecules, we observed no change in the bandwidth with the change in distance (Fig. 7a). In contrast, there was a steep rise in the width of the band at 1555 cm^{-1} as the distance decreased from 50 to 10 nm (Fig. 7b); that is, the bandwidth started to change even when the distance between lysozyme molecules was still relatively long. This result implies that the increase in the conformational flexibility of the protein was due mainly to an electrostatic repulsive interaction that acted even over long distances.

The I_{856}/I_{837} ratio for Tyr and the I_{870}/I_{877} ratio for Trp increased with decreasing distance (increasing concentration) (Fig. 7c). In particular, there was a steep rise in the ratios when the distance between lysozyme molecules decreased from <10 nm to 1 nm, and at distances longer than 1 nm, the ratios became saturated. Compared to the change in the 1555 cm^{-1} band, this change started to occur when the distance between lysozyme got relatively short. This result indicates that side chain interactions became important as the lysozyme molecules approached each other.

The I_{2940}/I_{3420} ratio was linearly related to volume fraction at low concentrations (Figs. 6c, d), but above $\phi = 0.125$ (C = 200 mg/mL), there was still some deviation from linearity. To produce a linear relationship between I_{2940}/I_{3420} ratio and volume fraction, N_L must be >1. This result indicates that the volume of lysozyme and water in high-concentration solutions were apparently different from the volume in low-concentration solutions. We plotted ΔR as a function of the distance between lysozyme molecules with N_L = 1 (Fig. 7d), which showed that the apparent volume of lysozyme molecules started to change when the distance between the molecules fell below 1 nm.

We suggest two interpretations of these results. One is that the increase in the apparent volume of lysozyme was due to the formation of clusters.[2,22–28] Cametti *et al.* suggested that in concentrated solutions, clusters of lysozyme molecules form.[28] These investigators also found that the hydration number of the protein decreases as the concentration rises above 100 mg/mL, as indicated by dielectric relaxation measurements. They attributed this behavior to removal of loosely bound water as the clusters form, whereas tightly bound water around the protein persists. This behavior also means that when the protein concentration is low, the molecules are rather

monodisperse and that the average distance between the monomers is much larger than the hydration layer thickness. As the protein concentration increases, however, protein clusters form, and the average distance between proteins gets smaller as the amount of loosely bound water decreases. In other words, the lysozyme molecules in high-concentration exclude the loosely bound water, which might cause the decrease in water molecules around lysozyme. Therefore, the volume occupied with lysozyme molecules increases all the more with the decrease of water molecules around lysozyme.

Our results support this interpretation. The behavior of the I_{2940}/I_{3420} ratio shows that N_L was >1 at concentrations of >100 mg/mL, which means that the volume occupied with lysozyme increased all the more as the concentration increased. In particular, N_L started to increase at concentrations of >100 mg/mL, and this result is similar to the result observed by means of dielectric relaxation measurements. Therefore, we attributed this behavior to the formation of protein clusters.

Another possible interpretation of our results involves changes in the water structure around the protein molecules. At high concentrations, the water between lysozyme molecules consists of only a few layers[28] and there is only a small amount of bulk water. So, most of the water molecules interact with protein molecules. This may cause the structural change of the water.

The increase of N_L at high concentrations meant the relative decrease of the volume of water, which is explained by the fact that water molecules near lysozyme molecules are more ordered than bulk water[29] and generally have a lower volume.[30] Therefore, at high protein concentrations, the increase in the proportion of water molecules that are near lysozyme molecules causes the decrease in the volume of water. A more detailed interpretation of this phenomenon will require further study involving other analysis

methods as well as theoretical calculations.

On the basis of our results, we propose that the molecular interactions occur via the following processes, the contributions of which strongly depend on protein concentration. Specifically, the molecular interactions change as the order of magnitude of the distance between the lysozyme molecules changes as shown in Fig. 8:

(i) C < 10 mg/mL (d > 10 nm): This state is the precursor state of the increase of the potential energy of a system. The electrostatic repulsive interaction works dominantly to increase the potential energy.

(ii) 10 mg/mL < C < 100 mg/mL (1 nm < d < 10 nm): In this intermediate state, hydrogen bonding interactions between protein molecules start to stabilize the protein, and electrostatic repulsive interactions strongly destabilized the protein.

(iii) C > 100 mg/mL (d < 1 nm): In this metastable state, both the electrostatic repulsive interaction and the hydrogen bonding interaction are saturated and other molecular interactions such as van der Waals interactions begin to contribute. The repulsive and attractive interactions are well-balanced.

In conclusion, our results show that although the distance between molecules changed by more than an order of magnitude over the tested range of concentrations (2.5–300 mg/mL), the secondary structure of lysozyme did not change. However, the molecular interactions of the protein molecules changed in a stepwise process as the order of magnitude of the distance between the molecules changed.

Our results can be expected to be helpful for the development and quality control of high-concentration protein formulations. In particular, the information about the behavior of Trp and Tyr residues, which are related to protein function, will be important. For example, the width of the 1550 cm^{-1} band of Trp can serve as a marker

of the crowdedness of protein molecules in solution. In addition, the I_{856}/I_{837} ratio of Tyr and the I_{870}/I_{877} ratio of Trp can act as markers of side chain interactions.

Additional work involving more variables, such as pH, salt concentration, additives, and solvent, is required and is currently underway in our laboratory. These future studies can be expected to improve our understanding of the structure and function of proteins under real-word conditions.

4. Conclusions

We investigated the structure and interactions of lysozyme molecules over a wide range of concentrations (2.5–300 mg/mL) by using Raman spectroscopy. In this concentration range, the amide-I band was not affected by concentration, but the width of the Trp band at 1555 cm^{-1}, the I_{856}/I_{837} ratio of Tyr, the I_{870}/I_{877} ratio of Trp, and the I_{2940}/I_{3420} ratio changed as the concentration was increased. These results indicate that although the distance between molecules changed by more than an order of magnitude over the tested concentration range, the secondary structure of lysozyme did not change. In contrast, interactions between the molecules did change in a stepwise process as the order of magnitude of the distance between the molecules changed. Our results also indicate that marker bands can be used to evaluate high-concentration solutions of protein and that the use of Raman spectroscopy can be expected to lead to progress in our understanding not only of the basic science of protein behavior under concentrated (*i.e.*, crowded) conditions but also of practical processes involving proteins, such as in the field of biopharmaceuticals.

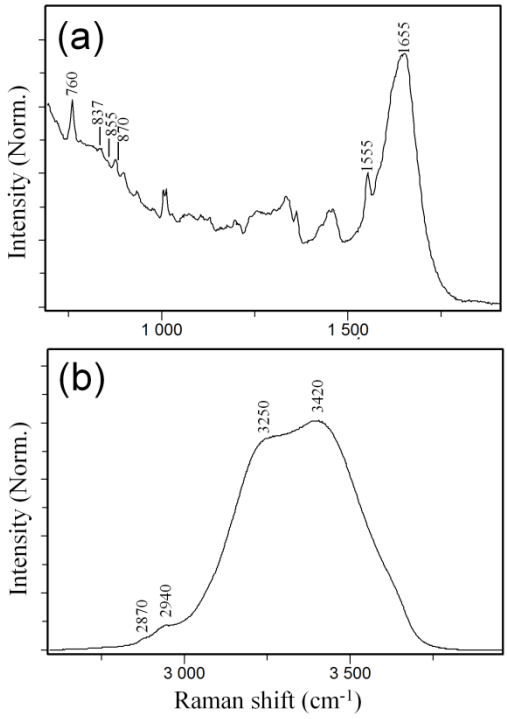

Figure 1. Raman spectrum of 20 mg/mL lysozyme solution (a) in the fingerprint region and (b) in the CH and OH stretching region.

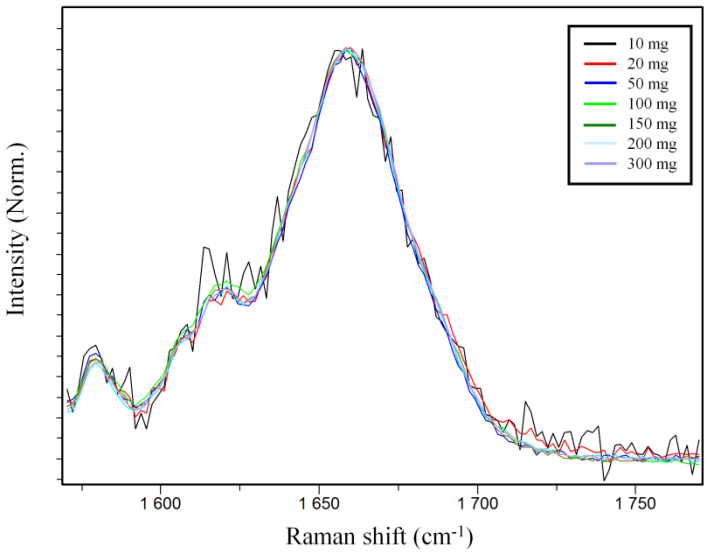

Figure 2. Amide I band of lysozyme solution of each concentration.

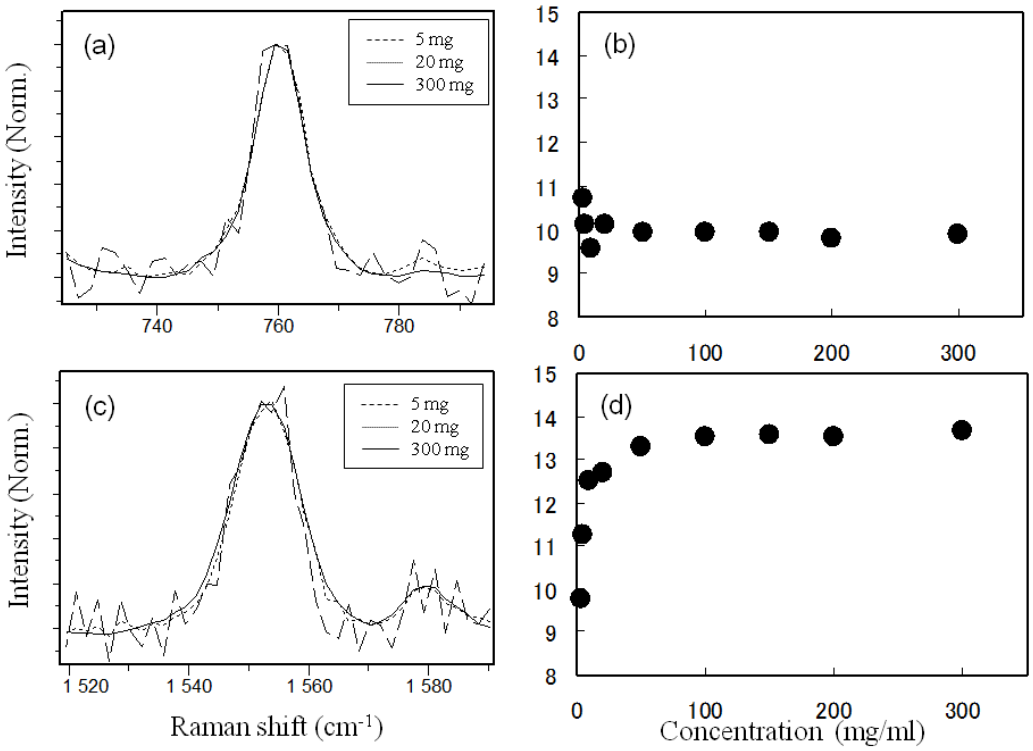

Figure 3. (a) Raman bands of Trp at 760 cm^{-1} of lysozyme solutions (5, 20, 300 mg/mL). (b) The band width at 760 cm^{-1} as a function of concentration. (c) Raman bands of Trp at 1550 cm^{-1} of lysozyme solutions (5, 20, 300 mg/mL). (d) The band width at 1550 cm^{-1} as a function of concentration.

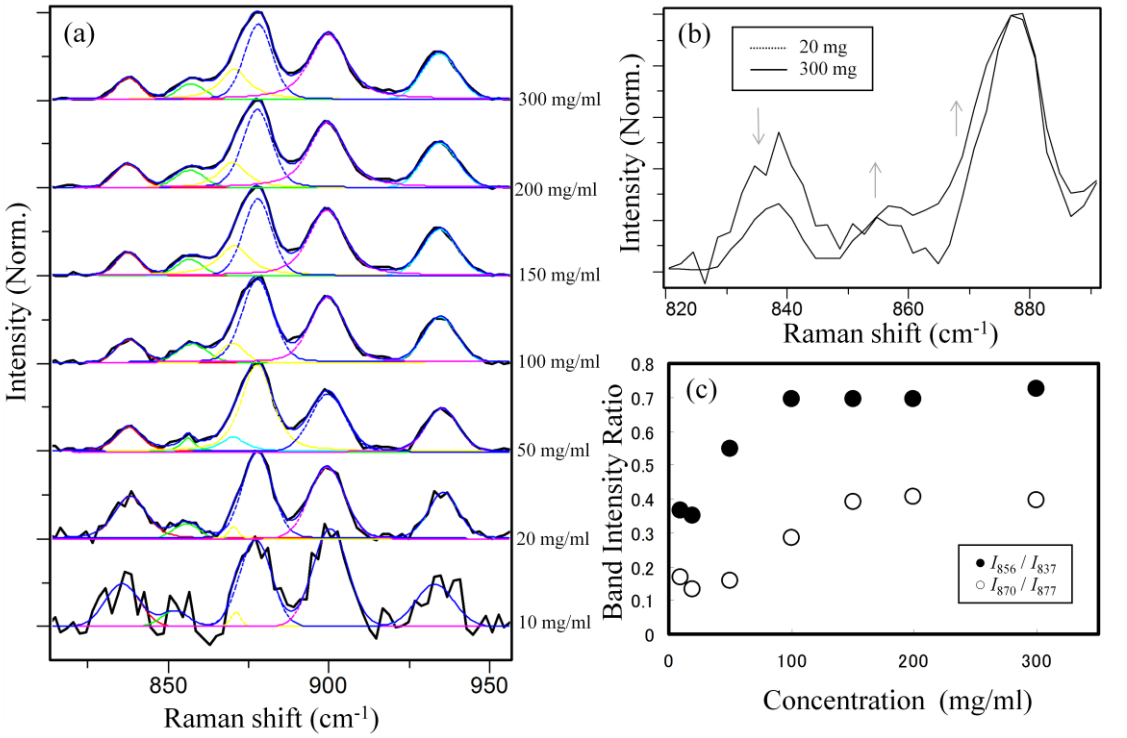

Figure 4. (a) Raman spectrum of lysozyme solutions in the range 810 cm^{-1} to 960 cm^{-1}. The concentration is 10, 20, 50, 100, 150, 200, 300 mg/mL. (b) An overlay of the 877 cm^{-1} band for 20 mg/mL solutions and for 300 mg/mL solutions. (c) The band intensity ratio of I_{856}/I_{837} (●) and I_{870}/I_{877} (○) as a function of concentration.

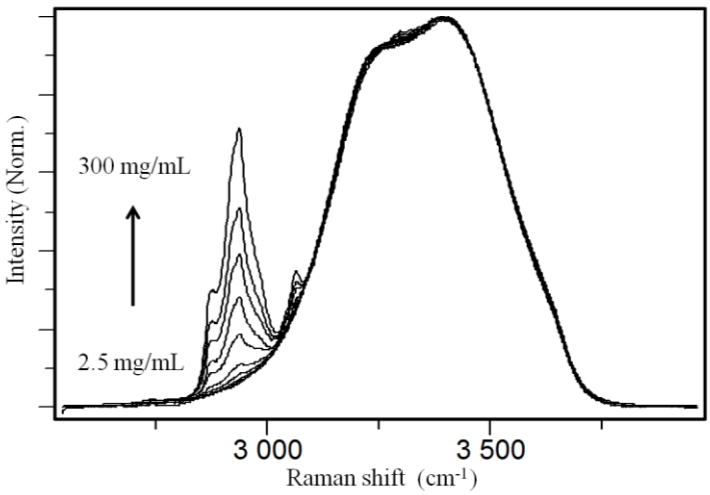

Figure 5. Raman spectrum in the CH and OH stretching region of lysozyme solution and buffer solution. The concentration is 2.5, 5, 10, 20, 50, 100, 150, 200, 300 mg/mL respectively.

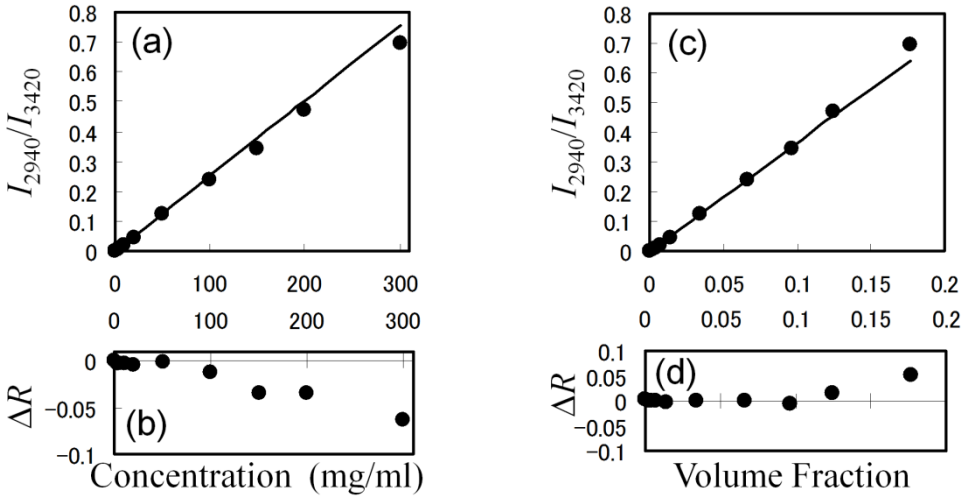

Figure 6. (a) The band intensity ratio I_{2940}/I_{3420} as a function of concentration. (b) The difference from the linearity ΔR as a function of concentration. (c) The band intensity ratio I_{2940}/I_{3420} as a function of volume fraction. (d) The difference from the linearity ΔR as a function of volume fraction.

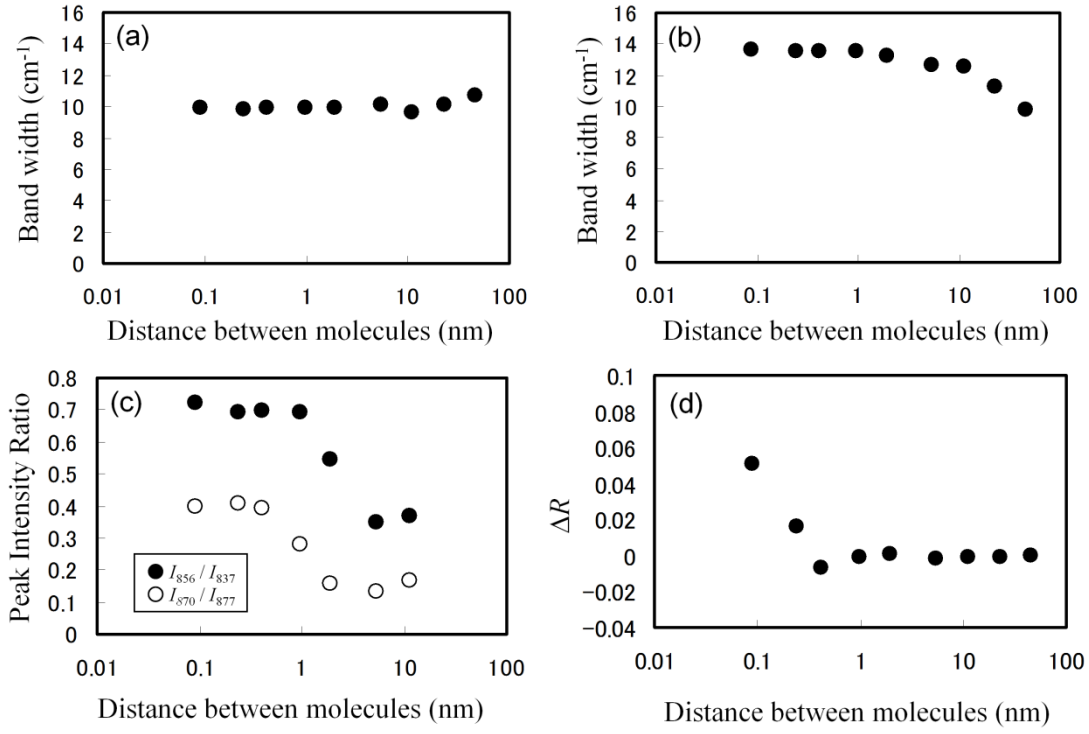

Figure 7. (a) The band width of Trp at 760 cm^{-1}, (b) The band width of Trp at 1550 cm^{-1}, (c) The band intensity ratio of I_{856}/I_{837} (●) and I_{870}/I_{877} (○), (d) ΔR as a function of distance between molecules.

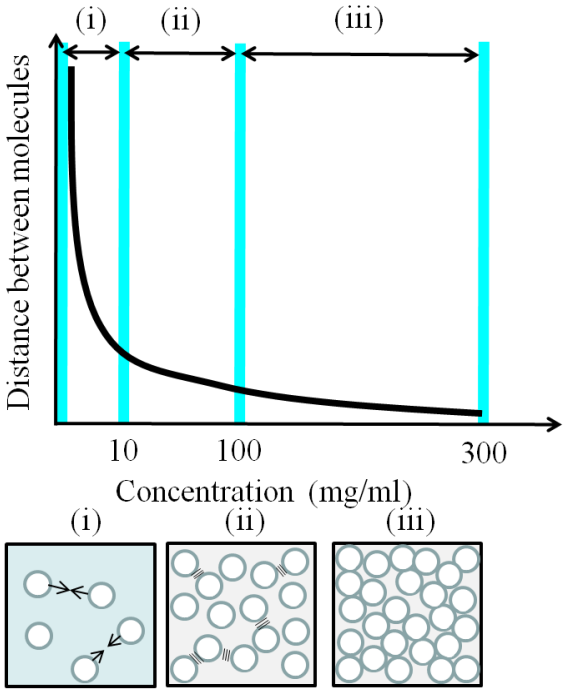

Figure 8. Schematic figure of protein solution at each concentration. (i) C<10 mg/mL , (ii) 10 mg/mL<C<100 mg/mL, (iii) C>100 mg/mL. Arrow lines (→) represent the electrostatic repulsive interaction. Dotted lines (///) represent the hydrogen bonding interaction.

5. Supporting Information

5-1. Supporting Information 1

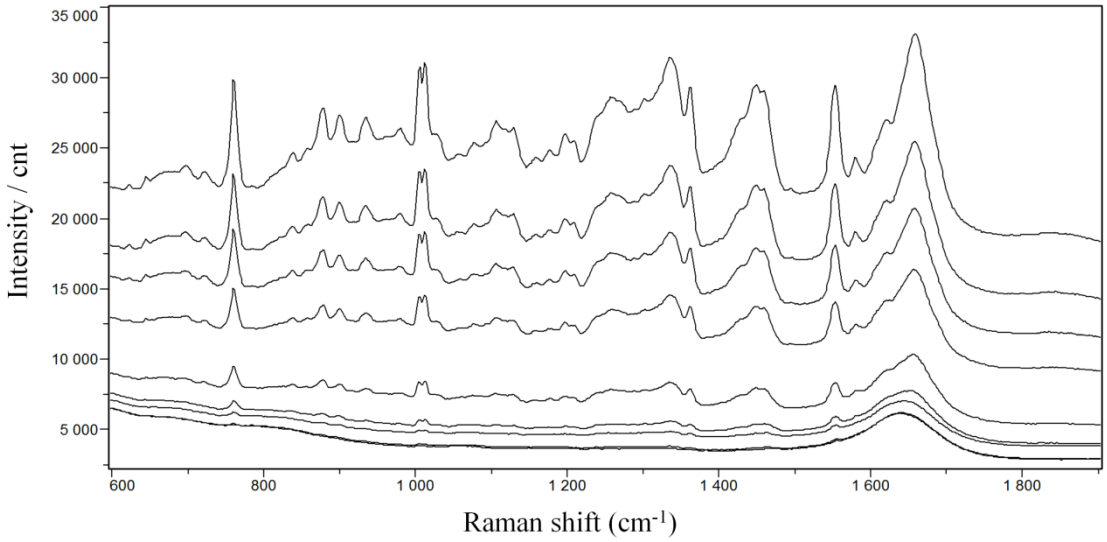

Figure S-1. Raman spectrum in the fingerprint region of lysozyme solution. The concentration is 2.5, 5, 10, 20, 50, 100, 150, 200, 300 mg/mL from the bottom respectively.

5-2. Supporting Information 2

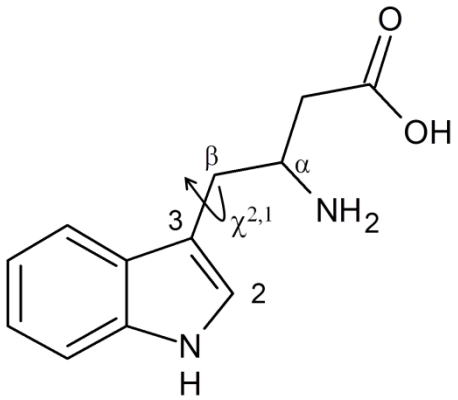

Figure S-2. The molecular structure of Trp

5-3. Supporting Information 3

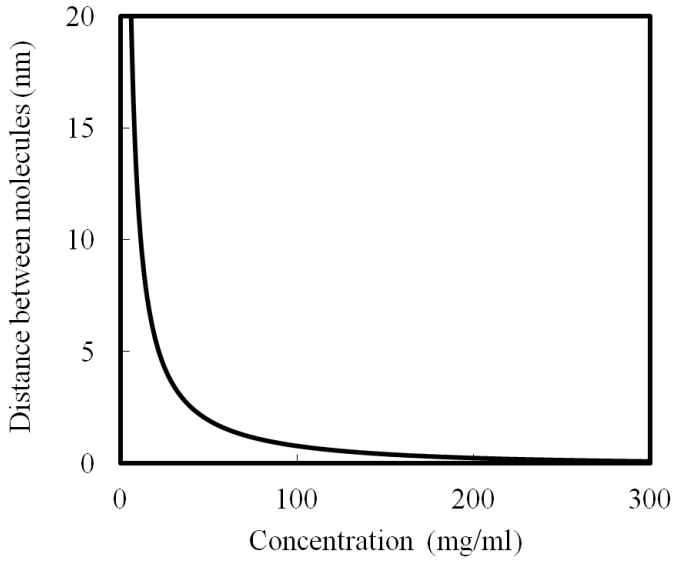

Figure S-3. Distance between molecules as a function of concentration.

6. References

1. Laue, T. Proximity energies: a framework for understanding concentrated solutions. *J. Mol. Recognit.* **2012**, *25*, 165-173.

2. Stradner, A., Sedgwick, H., Cardinaux, F., Poon, W. C. K., Egelhaaf, S. U., Schurtenberger, P. Equilibrium cluster formation in concentrated protein solutions and colloids. *Nature*, **2004**, *432*, 492-495.

3. Sato, T., Fukasawa, T., Shimozawa, T., Komatsu, T., Sakai, H., Ishiwata, S. Protein–protein interactions in solution and their interplay with protein specific functions. *J. Phys. Soc. Jpn.* **2012**, *81*, SA002.

4. Velev, O. D., Kaler, E. W., Lenhoff, A. M. Protein interactions in solution characterized by light and neutron scattering: comparison of lysozyme and chymotrypsinogen. *Biophys. J.* **1998**, *75*, 2682-2697.

5. Narayanan, J., Liu, X. Y. Protein interactions in undersaturated and supersaturated

solutions: a study using light and x-ray scattering. *Biophys. J.* **2003**, *84*, 523-532.

6. Wen, Z. Raman spectroscopy of protein pharmaceuticals. *J. Pharm. Sci.* **2007**, *96*, 2861-2871.

7. Oladepo, S. A., Xiong, K., Hong, Z., Asher, S. A., Handen, J., Lennev, I. K. UV resonance Raman investigations of peptide and protein structure and dynamics. *Chem. Rev*. **2012**, *112*, 2604-2628.

8. Maiti, N. C., Apetri, M. M., Zagorski, M. G., Carey, P. R., Anderson, V. E. Raman spectroscopic characterization of secondary structure in natively unfolded proteins: α-Synuclein. *J. Am. Chem. Soc.* **2004**, *126*, 2399-2408.

9. Apetri1, M. M., Maiti, N. C., Zagorski, M. G., Carey, P. R., Anderson, V. E. Secondary structure of α-synuclein oligomers: characterization by Raman and atomic force microscopy. *J. Mol. Biol.* **2006**, *355*, 63-71.

10. Lord, R. C., Yu, N. Laser excited Raman spectroscopy of biomolecules : I. Native lysozyme and its constituent amino acids. *J. Mol. Biol.* **1970**, *50*, 509-524.

11. Takeuchi, H. Raman structural markers of tryptophan and histidine side chains in proteins. *Biopolymers*, **2003**, *72*, 305-317.

12. Siamwiza, M. N., Lord, R. C., Chen, M. C., Takamatsu, T., Harada, I., Matsuura, H., Shimanouchi, T. Interpretation of the doublet at 850 and 830cm^{-1} in the Raman spectra of Tyrosyl residues in proteins and certain model compounds. *Biochemistry*, **1975**, *14*, 4870-4876.

13. Takeuchi, H., Harada, I. Normal coordinate analysis of the indole ring. *Spectrochim. Acta*, **1986**, *142A*, 1069-1078.

14. Miura, T., Takeuchi, H., Harada, I. Characterization of individual tryptophan side chains in proteins using Raman spectroscopy and hydrogen deuterium exchange kinetics. *Biochemistry*, **1988**, *27*, 88-94.

15. Miura, T., Takeuchi, H., Harada, I. Raman spectroscopic characterization of

tryptophan side chains in lysozyme bound to inhibitors: role of the hydrophobic box in the enzymatic function. *Biochemistry*, **1991,** *30,* 6074-6080.

16. Malinowski, E. R. *Factor Analysis in Chemistry* (2nd ed.); Wiley-Interscience: New York, 1991.

17. Samanta, S. R., Walrafen, G. E. Raman intensities and interactions in aqueous lysozyme solutions. *J. Chem. Phys.***1978**, *68*, 3313-3315.

18. Fischer, H., Polikarpov, I., Craievich, A. F. Average protein density is a molecular-weight-dependent function. *Protein Sci.* **2004,** *13*, 2825-2828.

19. Torquato, S. Mean nearest-neighbor distance in random packings of hard D -dimensional spheres. *Phys. Rev. Lett.* **1995**, *74*, 2156-2159.

20. Torquato, S., Lu, B., Rubinstein, J. Nearest-neighbor distribution functions in many-body systems. *Phys. Rev. A*, **1990**, *41*, 2059-2075.

21. Parmar, A. S., Gottschall, P. E., Muschol, M. Pre-assembled clusters distort crystal nucleation kinetics in supersaturated lysozyme solutions. *Biophys. Chem.* **2007**, *129*, 224-234.

22. Stradner, A., Cardinaux, F., Suhurtenberger, P. A small-angle scattering study on equilibrium clusters in lysozyme solutions. *J. Phys. Chem. B*, **2006**, *110*, 21222-21231.

23. Shukla, A., Mylonas, E., Di Cola, E., Finet, S., Timmins, P., Narayanan, T. Svergun, D. I. Absence of equilibrium cluster phase in concentrated lysozyme solutions. *Proc. Natl. Acad. Sci. USA,* **2008**, *105*, 5075-5080.

24. Porcar, L., Falus, P., Chen, W., Faraone, A., Fratini, E., Hong, K., Baglioni, P., Liu, Y. Formation of the dynamic clusters in concentrated lysozyme protein solutions. *J. Phys. Chem. Lett.* **2010**, *1*, 126-129.

25. Kowalczyk, P., Ciach, A., Gauden, P. A. Terzyk, A. P. Equilibrium clusters in concentrated lysozyme protein solutions. *J. Colloid Interface Sci.* **2011**, *363*, 579-584.

26. Barhoum, S., Yetiraj, A. NMR detection of an equilibrium phase consisting of monomers and clusters in concentrated lysozyme solutions. *J. Phys. Chem. B*, **2010**, *114*, 17062-17067.

27. Liu, Y., Porcar, L., Chen, J., Chen, W., Falus, P., Faraone, A., Fratini, E., Hong, K. Baglioni, P. Lysozyme protein solution with an intermediate range order structure. *J. Phys. Chem. B*, **2011**, *115*, 7238-7247.

28. Cametti, C., Marchetti, S., Onori, G. Lysozyme hydration in concentrated aqueous solutions. effect of an equilibrium cluster phase. *J. Phys. Chem. B*, **2013**, *117*, 104-110.

29. Feig, M., Pettitt, B. M. Crystallographic water sites from a theoretical perspective. *Structure*, **1998**, *6*, 1351-1354.

30. Merzel, F., Smith, J. C. Is the first hydration shell of lysozyme of higher density than bulk water? *Proc. Natl. Acad. Sci. USA*, **2002**, *99*, 5378-5383.

Chapter 3. The Behavior of Bovine Serum Albumin Molecules in Molecular Crowding Environments Investigated by Raman Spectroscopy

1. Introduction

Biomolecules are often in crowded environments such as in a cell or blood. Proteins are one of the most important biomolecules for life, however, characterizing the state of proteins in such crowded environments is a subject that is not still fully understood. There are reports that the molecular behavior of proteins in crowded environments is different from that in vitro (the ideal solution).[1–3] The behavior of proteins in crowded environments is dominated mainly by protein-crowder and protein-protein interactions. The former type of interaction is primarily an entropic effect such as exclude volume, whereas the latter interaction is primarily a soft chemical interaction involving, for example, hydrogen bonding or van der Waals interactions. Nonetheless, while these interactions have been defined, the details of these interactions are not fully understood. In this study, the behavior of bovine serum albumin (BSA) in crowded environments was investigated by Raman spectroscopy as a model system to examine the entropic effect and soft chemical interactions.

Currently, relatively few techniques are available to investigate directly the molecular behavior of macromolecules in highly crowded solutions. In-cell NMR spectroscopy or small-angle X-ray/neutron scattering approaches are powerful analytical methods for investigating the structure of molecules in molecular crowded environments or highly concentrated solutions.[4–7] Unfortunately, these methods require

highly sophisticated experimental setups and access to large equipment resources. In addition, interpretation of the results is often challenging and complex. Raman spectroscopy is a powerful tool for detailed investigation of the behavior of molecules in highly crowded environments because the Raman spectrum is not disturbed by optical effects such as the multi-scattering effect.[8] Specifically, for protein science, Raman spectroscopy has some important advantages. Not only information about each amino acid type, but also information about the secondary and tertiary structures of a protein can be obtained from Raman spectra. In addition, the water Raman band provides information about the hydrogen bond structure of hydration water molecules of a protein.

Serum albumin is the most abundant protein in the bloodstream and is present in blood at ~50 mg/mL.[9] In a dilute solution at pH7.0, BSA has a rigid structure and behaves as a hard sphere. In contrast, at lower pH (e.g., pH3), the packing of BSA is weaker and the structure adopts a softer, less rigid state.[5] Although the basic properties of BSA in dilute solutions (in vitro) have been investigated in detail and are well known, the behavior of BSA in crowded environments remains poorly understood. In this study, we performed concentration-dependent measurements of BSA solutions at pH7.0 and 3.0 by Raman spectroscopy, and discuss the entropic effect and the soft chemical interaction under these different conditions. In addition, H–D exchange measurements in the presence of another protein as a molecular crowder were performed to further characterize the entropic effect.

2. Experimental Methods

2-1. Materials.

2-1-1. Concentration Dependence Study

BSA (> 98%) was purchased from Sigma Aldrich (Product no. A3983-109, St. Louis, MO, USA). A buffer solution of BSA was prepared by mixing trisodium citrate dihydrate (Code 31404, Nacalai Tesque, Kyoto, Japan) with citrate acid monohydrate (Code 09106-15, Nacalai Tesque). The concentration of the buffer solution was 20 mM. The water for the buffer was purified with a Milli-Q laboratory water purifier (Millipore, Billerica, MA, USA). The final pH of the solutions was adjusted to either 7.0 or 3.0 and measured by a LAQUA F-72 pH meter (Horiba, Ltd., Kyoto, Japan). The BSA solutions were prepared at concentrations of 10, 20, 30, 40, 50, 60, 80, 100, 200, and 300 mg/mL. All measurements were carried out at room temperature (~20 °C).

2-1-2. H-D Exchange Study

H-D exchange measurements were performed to investigate the entropic effect on the BSA solution in the highly crowded environment. BSA solutions in D_2O (> 99.9%) (Product no. 151882, Sigma Aldrich) were prepared at 100 mg/mL. In the H–D exchange study, Tk-RNase H2 fragment 20 (F20) from a hyperthermophile, *Thermococcus kodakarensis*, was added as a crowder. F20 is a derivative from Tk-RNase H2 with residues 1–176. F20 exists as a monomer and has high stability. F20 was prepared as described previously.[10,11] The concentration of F20 in the BSA solution was 50 mg/mL. To examine the temperature dependence of the H–D exchange measurement, the test BSA solutions were incubated at 25, 35, 40, 45, and 50 °C overnight in the presence and absence of F20.

2-2. Raman Spectroscopy

Raman spectra were measured with a LabRam Evolution Raman microscope equipped with the Labspec 6 software (Horiba Jobin Yvon, Paris, France). The spectrograph of the system had the Czerny–Turner configuration and the focal length was 800 mm. Raman spectra were measured in the back-scattering geometry. A thermoelectrically cooled Synapse charge-coupled device camera (Horiba, Ltd.) was used as the detector.

The entrance slit of the spectrometer was set to 100 μm. The dispersive element had a grating of 600 lines/mm, which provides a wavenumber resolution of ca. 2 cm^{-1}. For the concentration-dependent measurements, the excitation was accomplished with the 532 nm line (20 mW at the sample surface). For the H–D exchange measurements in the presence of F20, the excitation was accomplished with the 785 nm line and the laser power was 50 mW at the sample surface.

The scattered light was collected using a multipass cell holder (Horiba Jobin Yvon) with a visible macro lens (focal length = 40 mm). A 115-QS micro cell (Hellma Analytics, Müllheim, Germany) was used for the micro cuvette. The Raman shift was calibrated by means of the atomic emission lines of a neon lamp. Raman spectra were normalized at the maximum intensity in the measurement range.

2-3. Spectroscopic Analysis

Principal components analysis (PCA) is a multivariate analysis method that is useful for resolving the spectrum of a pure component from several types of spectra.[12] PCA has a unique property to detect the spectra of minute chemical species in local environments or different interactions from those of dominant bulk species. The number of chemical species in the protein solution was estimated by PCA. The accuracy of the

predictive ability of the model was evaluated by rank analysis using a reduced-eigenvalue (REV) plot. REV has the same statistical meaning as the normal eigenvalue (EV), however REV is more suitable for the evaluation of minute eigenvalues since it takes into account the degree of freedom.[12] A REV plot is calculated by the following equation:

$$REV_j = \frac{EV_j}{(N - j + 1)(M - j + 1)}, \quad (1)$$

where j is the factor level, and the parameters N and M represent the size of the spectral matrix ($N \times M$). Of note, the analysis of basis factors by means of eigenvalue analysis does not take into account concentration (intensity) information. Thus, the cross-validation (CV) technique, which takes into account both spectral and intensity information, can be employed to cross-check the accuracy of the results of REV.[13]

Root-mean-square prediction error of cross-validation (RMSECV), which is CV with leave-one-out cross-validation,[13] was also calculated according to the following equation:

$$RMSECV = \sqrt{\frac{\sum_{j=1}^{n} (\hat{c_j} - c_j)^2}{n}}, \quad (2)$$

where $\hat{c_j}$ is the predicted concentration, c_j is the actual concentration, as determined by the reference method, and n is the number of samples used in the calibration model. The accuracy of the factor analysis of PCA increases by use of both REV and RMSECV. Although PCA is useful for investigating the mathematical structure of data, interpretation of the results can be complex, since the PCA loading has negative bands owing to the relative quantitative change, which cannot be treated as real chemical component spectra. Multivariate curve resolution alternating least squares (MCR-ALS)

is a complementary method because MCR-ALS is based on nonnegative matrix factorization and the results obtained with MCR-ALS have more direct physical meaning.[14] Specifically, if the number of species can be estimated, the results from MCR-ALS have more realistic chemical meaning. In the matrix form, MCR-ALS analysis can be written as

$$D = CS + E, \tag{3}$$

where D is the original data matrix, S is the pure matrix of the pure spectra, C is the score (the concentration) of each species, and E is the matrix of the residuals. The contribution ratio of each score C_R is the normalized concentration, which is the quotient upon division of the concentration of the sample (C_j) by the sum of concentrations of all components as described below.

$$C_R(j) = \frac{C_j}{\sum_{j=1}^{h} C_j}. \tag{4}$$

3. Results

3-1. Raman Spectrum of BSA

The "fingerprint region" between 400 and 2000 cm^{-1} of the Raman spectrum of a protein provides secondary and tertiary structural information or the microenvironment of each functional group. The interpretation of the relation between the Raman band and the protein conformation is well established.[15,16] As shown in Fig. 1a, the "fingerprint region" of the Raman spectrum of BSA at pH7.0 has many bands, including the amide I band at 1650 cm^{-1}, the CH_2 stretching band at 1450 cm^{-1}, and the bands of hydrophobic residues such as Phe at 1002 cm^{-1} and Tyr at 850 and 825 cm^{-1}. In the higher wavenumber region, strong broad bands from 2800 to 4000 cm^{-1} were

observed, which were assigned to the CH and OH stretching modes of BSA and water, respectively (Fig. 1b). These bands were used to describe in structural detail the concentration dependence and crowding effect behavior of BSA.

3-2. CH and OH Stretching Bands

Increasing the protein concentration of a solution leads to increases in the viscosity of the solution.[17,18] In addition, some previous studies have shown that with increasing viscosity, the structure of hydrogen bonds involving water molecules may change.[19–21] The OH stretching Raman band is sensitive to changes in the geometry of hydrogen bonds.[22–24] Therefore, we focused on the OH stretching band at ~2800 to 4000 cm^{-1}. The CH and OH stretching bands at each concentration (20, 40, 80, 100, and 300 mg/mL) at pH7.0 and pH3.0 are shown in Fig. S1, and described in Supporting Information 1. To investigate the interaction between the protein and water molecules in highly concentrated solutions in detail, the data set of these concentration-dependent spectra were subject to PCA.

For estimating the number of chemical species in the BSA solution at pH7.0, the REV and RMSECV were calculated, as described in Supporting Information 2. The REV and RMSECV of BSA between 0 and 300 mg/mL at pH7.0 are shown in Figs. S2a and S2b, respectively. The results at pH7.0 indicate that there were three species present. Based on the results of PCA, the MCR-ALS analysis was applied to this concentration-dependent data set of the concentration dependence measurements. The abstract spectra of each component are shown in Figs. 2a, 2c and 2e, and the normalized concentrations, C_R, of each component are shown in Figs. 2b, 2d, and 2f. Fig. 2a shows the spectrum with two main bands at 3240 and 3420 cm^{-1}, which is the typical spectrum

of bulk water. In Fig. 2c, the spectrum has two features: the CH stretching band at ~2927 cm^{-1} of the protein, and a water band with two main broad bands at 3290 cm^{-1} and at ~3460 cm^{-1}, which are mainly from the hydrated waters of the protein. As shown in Figs. 2b and 2d, the C_R of the hydrated water component increases as the concentration increases (Fig. 2d), whereas the C_R of bulk water decreases (Fig. 2b). The profile of the third component was of unique with respect to the concentration changes (Fig. 2f). Up to 80 mg/mL, the C_R of this component decreased as the concentration increased and over 80 mg/mL, the C_R went down to almost zero. This result indicates that the third component is influenced by and associated with the highly concentrated conditions. The abstract spectrum (Fig. 2e) has two features: the CH band at around 2870 and 2932 cm^{-1} , which may be from the citrate acid in the solution and a broad band near 3270 and 3415 cm^{-1} associated with the O–H stretching of water. Since the third component decreased with the increase of the concentration and the spectrum of this component is similar to the bulk water spectra, this component can be assigned to the interfacial water around the hydrated protein. These profiles indicate that as the concentration of the protein increases, the protein hydrates the neighboring free water and as the concentration increases further the amount of free water around the proteins decreases, which is especially the case for a protein concentration >100 mg/mL. The behavior of the hydration of free water correlates (to some degree) with the solution viscosity. A previous study showed that the viscosity of a BSA solution at pH7.0 increases nonlinearly over 100 mg/mL. Thus, this observation supports the behavior of the water Raman bands observed in this study.[17]

For the estimation of the number of the chemical species in the BSA solution at pH3.0, REV and RMSECV were calculated, as described in Supporting Information 2.

The REV and RMSECV of the BSA concentrations from 0 to 300 mg/mL at pH3.0 are shown in Figs. S2c and S2d, respectively. The results of both REV and RMSECV at pH3.0 indicate that there were only two main species, which is different to the results observed at pH7.0. Based on the PCA results, MCR-ALS analysis was applied to this concentration dependence data set. The abstract spectra of each component are shown in Figs. 3a and 3c, and the C_R of each component is shown in Figs. 3b and 3d. Fig. 3a shows the typical spectral features of bulk water. In Fig. 3c, the spectrum has two features: the CH band at ~2930 cm^{-1} from the protein, and a water band with three main bands at 3220 cm^{-1}, 3290 cm^{-1}, and 3430 cm^{-1}, which mainly arise from the hydrated water of the protein. Interestingly, this hydrated protein component at pH3.0 gives rise to a water band that has greater intensity when compared with the main band observed at pH7.0 (Fig. 2c). This observation indicates that the amount of hydrated water associated with the protein at pH3.0 is larger than that at pH7.0.

3-3 Amide I Band

In previous studies examining the effects of molecular crowding environments, protein molecules often adopt confined and compact structures.[1,2] Here, we carried out an in-depth examination of BSA in a homogenous crowded environment by examining the amide I band, which represents the secondary structure of a molecule.

The amide I band of proteins is sensitive to the strength of hydrogen-bonding interactions (C=O···HN) involving amide groups. Thus, the information about protein secondary structures can be obtained from the location of this band.[25,26] In the Raman spectrum of BSA, the amide I band overlaps with the water bending band at 1640 cm^{-1} (Fig. 1a). We therefore applied PCA to resolve these two bands. To estimate the number

of chemical species, REV and RMSECV were calculated for the estimation of the number of the chemical species as described in Supporting Information 3. The detailed results at pH7.0 and 3.0 are presented in Figs. S3a–S3d. The results of both REV and RMSECV showed that there are two chemical species in this data set at both pH values. MCR-ALS analysis with two-component approximation was performed based on the results of PCA. The results for pH7.0 are shown in Figs. S4a and S4b. The abstract spectra were (a) the amide I band of the protein and (b) the OH bending band. The band position of the resolved spectra of the protein was ~1660 cm^{-1}, which means that BSA adopts primarily an α-helical secondary structure with a partial β-sheet structure. The open circles shown in Fig. S4c represent the C_R of the protein and the closed circles represent the C_R of water. As shown in the figure, as the C_R of the protein increases, the C_R of water decreases. The analysis at pH3.0 is presented in Fig. S5. Clearly, the MCR-ALS results at pH3.0 are nearly the same as the results at pH7.0. Based on this quantitative information, the OH bending band was subtracted from the original spectrum recorded at each concentration. In Fig. 4, the amide I bands of each concentration (10, 50, 100, and 300 mg/mL) at (a) pH7.0 and (b) pH3.0 are shown. This result shows that at both pH values, there were no specific changes to the amide I band at each concentration, thereby indicating that over this concentration range there were no significant changes to the secondary structure of BSA.

3-4. Tyr Band

The band intensity ratio I_{850}/I_{825} is a known marker of hydration.[27] A I_{850}/I_{825} ratio roughly indicates that the OH group of Tyr acts as a hydrogen donor or acceptor. However, in recent studies, the interpretation of these bands has changed slightly. The

first interpretation of the ratio of the Tyr doublet I_{850}/I_{825} was suggested in 1975 by Siamwiza et al.[27] This rule was derived from studying the simple molecule "p-cresol", and subsequently it has been revealed that this interpretation cannot be adapted to Tyr derivatives or macromolecules containing Tyr residues, such as proteins. In addition, it has been reported that the ratio I_{850}/I_{825} of Tyr is a rather qualitative marker of the hydrophobic or hydrophilic environments of Tyr.[28]

We compared Raman spectra in the region of 800–870 cm^{-1} over the BSA concentration range of 20–300 mg/mL at pH7.0 (Fig. 5a). To calculate the ratio at each concentration, we fitted the 825 and 850 cm^{-1} bands with a Gaussian–Lorentz function (the fitting results are indicated by the blue curves in Fig. 5a). Comparison of the I_{850}/I_{825} ratios for BSA samples ranging from 10 to 300 mg/mL showed that the ratio was essentially constant over this concentration range (Fig. 5b).

We also compared the I_{850}/I_{825} ratio of the BSA samples at pH3.0. In contrast to the results at pH7.0, a change in the ratio was observed as the BSA concentration increased. The concentration dependence of the I_{850}/I_{825} ratio shows that the ratio started to increase from ca. 1.4 to 1.6 at concentration ≥20 mg/mL and was saturated at concentrations ≥100 mg/mL, which indicates that Tyr residues at concentrations ≥100 mg/mL at pH3.0 are located in the local environments, similar to those at pH7.0. Since there was not a significant change of the amide I band as the protein concentration increased, the observed increase of the I_{850}/I_{825} ratio indicates the partial environmental change around Tyr residues, which may be due to the excluded volume effect with the increase of the concentrations.

3-5. Phe Band

The band intensity of Phe at 1002 cm^{-1} arises from a symmetrical stretching mode of the benzene ring of Phe. The relative intensity of the band arising from Phe is caused by the local environmental change of Phe residues from hydrophilic to hydrophobic conditions.[29,30] In previous studies, the increase in band intensity at 1002 cm^{-1} represents a maker of the hydrophobic effect in which hydrophobic parts oligomerize or aggregate without changes in secondary structure.

The Raman band at 1450 cm^{-1} is a CH_2 stretching band arising from the amino acids of a protein and the band intensity reflects the amount of protein and is not sensitive to the local environments. Thus, the band at 1450 cm^{-1} can be used as an internal standard of protein concentration and the intensity is linearly proportional to the concentration of the protein.

The concentration dependencies of the two bands, 1002 and 1450 cm^{-1} at pH7.0 are shown in Fig. 6a and the relative intensity of the I_{1002}/I_{1450} ratio at pH7.0 is plotted in Fig. 6b. As the concentration increased in the lower concentration range, there was no specific change of the I_{1002}/I_{1450} ratio; however, at concentrations ≥80 mg/mL the ratio began to increase. This behavior indicates a change of the local environment of Phe residues, i.e., an increase of the hydrophobic effect surrounding Phe residues. These changes may be due to the interaction at the hydrophobic protein-protein interface around Phe, induced by the attractive short range interactions. The results at pH3.0 differed to those observed at pH7.0. The concentration dependencies of the two bands, 1002 and 1450 cm^{-1} at pH3.0 are shown in Fig. 6c, and the relative intensity of the I_{1002}/I_{1450} ratio at pH3.0 is presented in Fig. 6d. The ratio at pH3.0 started to increase at > 60mg/mL and got higher than the one at pH7.0. It is of interest that the increase in the I_{1002}/I_{1450} ratio began at concentrations lower than observed for data recorded at pH7.0.

This result shows that at pH3.0, protein molecules interacted with each other at lower concentrations.

3-6. Effect of Crowding

Although the behavior of the hydrophobic residues of Tyr and Phe was affected by the concentration effect from these experimental results, the amide I band was not sensitive to the concentration effect which can induce the excluded volume effect. It would be of interest to perform the H–D exchange experiment with and without the crowder and we could observe the subtle change in the amide I band to ascertain the excluded volume effect more directly. For this purpose, the hyperthermophile protein F20, which is temperature stable, was added as a molecular crowder to the BSA samples. In Fig. S6, the Raman spectrum of F20 (50 mg/mL) in D_2O is shown. The water bending band of D_2O located at 1200 cm^{-1} was subtracted from the spectrum. The amide I band at 1650 cm^{-1}, the CH_2 stretching band at 1450 cm^{-1}, and the bands arising from hydrophobic residues such as Phe at 1002 cm^{-1} were observed. Because the bending band of D_2O was located at 1200 cm^{-1} and was not in overlap with the amide I band at ~1650 cm^{-1}, the analysis of the amide I band was straightforward.

To investigate subtle changes to the secondary structure of BSA in the presence of F20, a temperature-dependent series of H–D exchange measurements between 25 and 50 °C was performed. In Fig. S7, the amide I bands of F20 at 25 (red) and 50 °C (blue) are shown. Over this temperature range, no specific change in the amide I band of F20 was observed, which indicates that the secondary structure of F20 is stable over this temperature range. Fig. 7 shows the 780–880 cm^{-1} spectral region (this region probes the local environment of Tyr residues) of the Raman spectra of BSA between 25 and

50 °C before (red) and after the addition of F20 (blue). No specific interaction between BSA and F20 was observed, and the local environment of the Tyr residues of BSA did not change over this temperature range. Therefore, F20 functioned as simple hard spheres and the addition of F20 predominantly influenced the entropic effect (the excluded volume effect).

The amide I band at each temperature in (a) BSA and (b) BSA with F20 are shown in Fig. 8. The band position of the amide I band shifted slightly to a lower wavenumber, which indicates that H–D exchange occured.[31] To analyze the thermally induced H–D exchange dynamics in more detail, the H–D exchange spectra at each temperature were subject to PCA and MCR-ALS analysis.

To estimate the number of components in the BSA solution in the absence and present of F20, REV and RMSECV were calculated from these profiles and the detailed results are shown in Figs. S8a–S8d. The results of both REV and RMSECV show that there were two components in the H–D exchange dynamics of the amide I band. MCR-ALS analysis with two-component approximation was performed based on the results of the PCA. The abstract spectra of each component and the contribution ratio C_R of each component at each temperature are shown in Figs. 9a–9d. Figs. 9a and 9c show the abstract spectra of the amide I band before (blue) and after (red) H-D exchange. The amide I bands of both systems shifted to a lower wavenumber because of H–D exchange. Interestingly, the concentration profile of each system was different. In the H–D exchange dynamics of the BSA sample, the H–D exchange started at slightly above 25 °C and the temperature of the intermediate state (T_i) was ca. 32 °C. In contrast, the T_i of BSA in the presence of F20 was different. The H–D exchange did not start until ~35 °C and the T_i was ~40 °C. This difference between the two systems shows that

the addition of F20 as a crowder prevents H–D exchange of solvent labile groups of BSA. In other words, the excluded volume effect suppresses BSA dynamic fluctuations and thus stronger intramolecular hydrogen bonds are formed that reduce H–D exchange of the hydrogen atom of the amide bond. Thus, the results show that the crowding effect functions by forcing BSA to adopt a more compact state that stabilizes the secondary structure elements and therefore yields stronger intramolecular hydrogen bonds.

4. Discussion

4-1. The Difference in the Behavior of BSA between pH7.0 and pH3.0

From these results, it was confirmed that the hydrophobic residues, e.g., Tyr and Phe, behave differently between pH7.0 and pH3.0. On the basis of some previous researches, it is a well-known fact that BSA at pH7.0 has a N form, which is a rigid globular structure, while BSA at around pH3.0 has an E (Expanded) form which partially loses the packing of the molecule.[5] As the concentration increases, the distance between surfaces of the molecules decreases, which can be calculated as shown in the Fig. S9. Since the pI of BSA is at pH ca. 4.7, the electric repulsion will work powerfully at both pH7.0 and 3.0. Therefore, the excluded volume effect from the electric repulsive interaction will work well. Unexpectedly, the amide I band at both pH does not change apparently, although the concentration changes. This result can imply that the excluded volume effect does not change the secondary structure but stabilizes the secondary structure of the BSA. In addition, it is of interest that the local environment of Tyr at pH7.0 does not change as the concentration increases, while the local environment of Tyr at pH3.0 partially changes. This difference of Tyr will be caused by the rigidity of BSA at pH7.0 and 3.0. At pH7.0, the packing of BSA is highly fastened, thus, there is

almost no change of the local segmental motion of Tyr with the increase of the excluded volume effect. On the other hand, at pH3.0, the packing of BSA is partly loosen and the excluded volume effect will induce the re-packing of BSA, therefore, this re-packing effect will restore BSA to the native-like conformation and the local environment of Tyr becomes the similar state at pH7.0. Therefore, these results show that BSA at pH7.0 can be treated as a hard sphere, thus the behavior of the local motion around Tyr is not sensitive to the excluded volume effect. In contrast, BSA at pH3.0 is relatively softer, therefore, the behavior of the local motion around Tyr is partially sensitive to the excluded volume effect (Fig. 10a).

As the distance between the molecules gets close to the size of the molecule, the soft chemical interaction (the short range attractive interaction) such as the van der waals interaction or the CH–π interaction will also work well. In the case of BSA, the distance between the molecules almost becomes the same size over 50 mg/mL. The relative intensity of Phe (I_{1002}/I_{1450}) at pH7.0 (Fig. 6b) slightly increases ≥80 mg/mL, which is the concentration when the short range attractive interaction works well. In other words, this result indicates that the change will be caused by the short range attractive interaction, which may be induced by the interaction at the hydrophobic protein-protein interfaces around Phe. Of note is that the relative intensity of Phe (I_{1002}/I_{1450}) at pH3.0 (Fig. 6d) starts to increase from ≥60 mg/mL, which is the lower concentration than the one at pH7.0 and saturates at ca. 1.4, which is a higher value than in the case of pH7.0. The difference of the behavior of Phe can be due to the rigidness of the structure again, although the electric repulsive interaction works well at both pH. At pH7.0, the packing of the molecule is highly fastened, thus, the molecule can be treated as a charged colloidal sphere, therefore, the electric repulsive interaction can

compete the short range attractive interaction. In this manner, the short range attractive interaction gently works. On the other hands, at pH3.0, the packing of the molecule is partly loosen, thus, the hydrophobic surfaces can be attracted mutually with the aid of the flexibility of each segment, which will lead to the short range interaction between the molecules (Fig. 10b).

On the basis of these discussions, the difference of the behavior of the hydrophobic residues of both Tyr and Phe at the crowding environment can be explained from the view point of the rigidness or softness of the molecule. The results show that the “hard” molecule is not sensitive to the crowding effect such as the excluded volume effect, while the “soft” molecule is sensitive to the one.

4-2. The Role of the Interfacial Water around BSA Molecules

The behaviors of the hydration water at both pH7.0 and 3.0 are re-focused on the basis of the knowledge of the structure of BSA at both pH. In the neighboring water molecules around protein of the N form (pH7.0), there can be seen a clear difference between the hydration water and the interfacial free water, and the interfacial water decreases and gets hydrated with the increase of the concentration, as shown in Fig. 2. In the E form (pH3.0), the amount of the hydration of water seems to be larger than that at pH7.0 from the results of the hydrated water Raman bands of the BSA in Fig. 3. This difference can be explained by the difference of the structure of the BSA, since the E form is partly loosen and has larger area which can be hydrate a larger number of neighboring water molecules. In addition, there is not a clear difference between the hydration water and the interfacial free water, in contrast to the case of pH7.0. This behavior will be due to the fact that the structure at pH3.0 is softer and more flexible,

thus the interface between the hydrated protein and bulk water will not be divided clearly as in the case of pH7.0. In this manner, the behavior of the hydrated and interfacial waters at pH7.0 and 3.0 can be highly relevant to the molecular structure as in the case of the behavior of the hydrophobic groups of Tyr and Phe. In other words, the different behavior of the hydrophobic groups at pH7.0 and 3.0 is also related to that of the hydrated and interracial waters. Therefore, the results can also indicate that the local environment of the hydrophobic groups is kept stable, while the interfacial water around the BSA molecules exists. On the other hand, the molecule is sensitive to the excluded volume and the soft chemical interaction, while the interfacial water around the BSA molecules does not exist. Thus, the role of the interfacial water will keep the local environment of hydrophobic groups stable.

4-3. A Description of BSA in Crowded Environments from a Raman Spectroscopic Viewpoint

From a spectroscopic analysis viewpoint, the hydration effect was clearly observed and the difference between pH7.0 and pH3.0 was confirmed. The amide I band showed negligible change with increasing protein concentration; however, the amide I band was sensitive to the excluded volume effect, and the influence of this entropic effect on protein backbone fluctuations was observed by H–D exchange analysis. The I_{850}/I_{825} ratio of Tyr probes the local environmental change associated with the packing of the molecule. The I_{1002}/I_{1450} ratio of Phe probes the protein-protein interfacial interactions between hydrophobic regions of a protein. In this manner, the spectroscopic analysis of the amide I band, the I_{850}/I_{825} of Tyr, and the I_{1002}/I_{1450} of Phe can provide a detail description of proteins in different molecular crowding

environments, and should enable assessment of molecular crowding environments. By using these various Raman spectroscopic probes, a detailed molecular description of BSA under crowding conditions has been presented. The results of this study showed that the behavior of BSA in different molecular crowding environments is dominated by a balanced interplay of the entropic effect, soft chemical interactions, and the hydration effect.

5. Conclusions

In summary, Raman spectroscopic data have revealed that the entropic effect, soft chemical interactions, and hydration effect are tightly balanced in crowded environments (Fig. 11). The key molecular findings of BSA in crowded environments are:

(i) The crowding effect strongly affects the secondary structure of a protein by maintaining a compact state and strengthens intramolecular hydrogen bonds in a concentration-dependent manner, as determined from the analysis of the amide I band and the H-D exchange experiment.

(ii) From the water, Tyr and Phe Raman bands, an increase in protein concentration induces cooperatively a change in the local environment of hydrophobic residues associated with the packing of the molecule or the intermolecular interaction. The interfacial free water layer surrounding the protein maintains the local environment of hydrophobic residues.

(iii) On the basis of these knowledge, it was found that the “hard” molecule is not sensitive to the crowding effect such as the excluded volume effect, while the “soft” molecule is sensitive to the one.

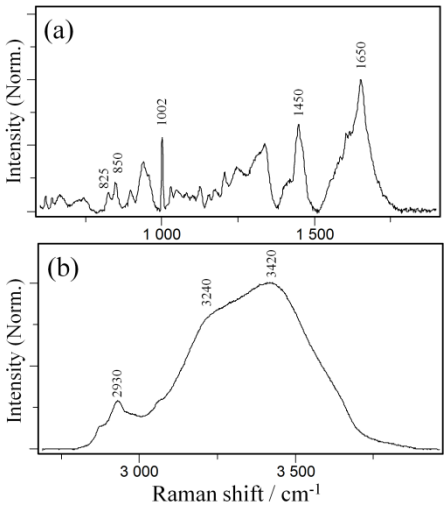

Figure 1. Raman spectra of the 100 mg/mL BSA sample at pH7.0 (a) in the fingerprint region and (b) in the CH and OH stretching regions.

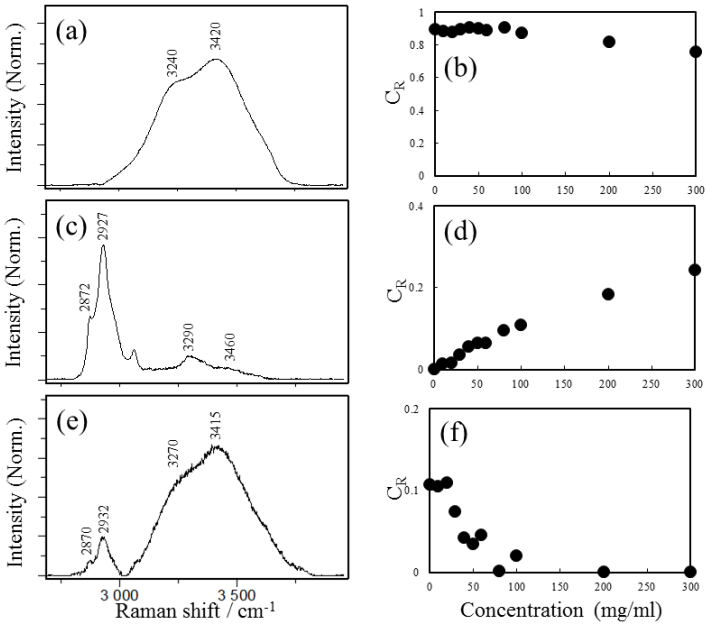

Figure 2. Abstract Raman spectra of each component (a, c, e) and the normalized concentration C_R of each component (b, d, f) in the CH and OH stretching regions of the BSA sample at pH7.0.

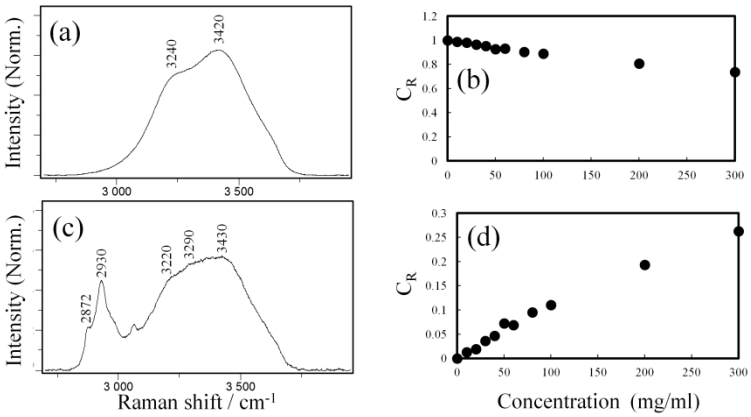

Figure 3. Abstract Raman spectra of each component (a, c) and the normalized concentration C_R of each component (b, d) in the CH and OH stretching regions of the BSA sample at pH3.0.

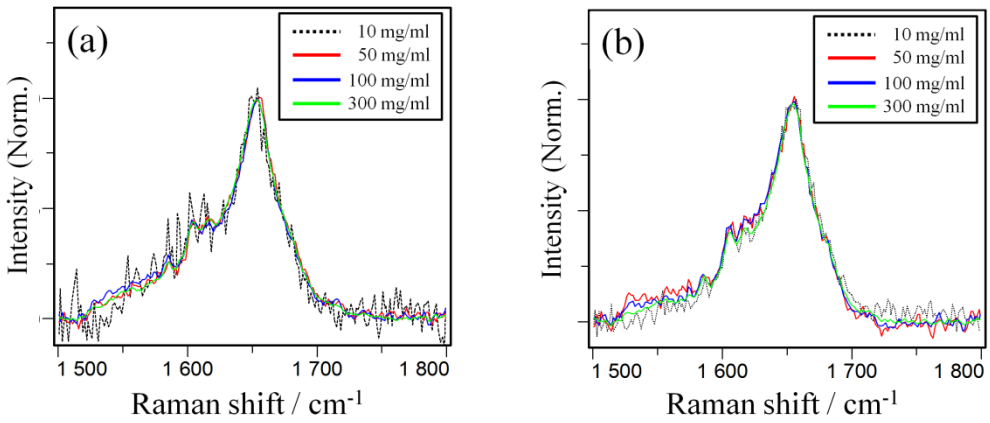

Figure 4. Overlay of the amide I band of BSA at each concentration (10, 50, 100 and 300 mg/mL) and (a) pH7.0 or (b) pH3.0.

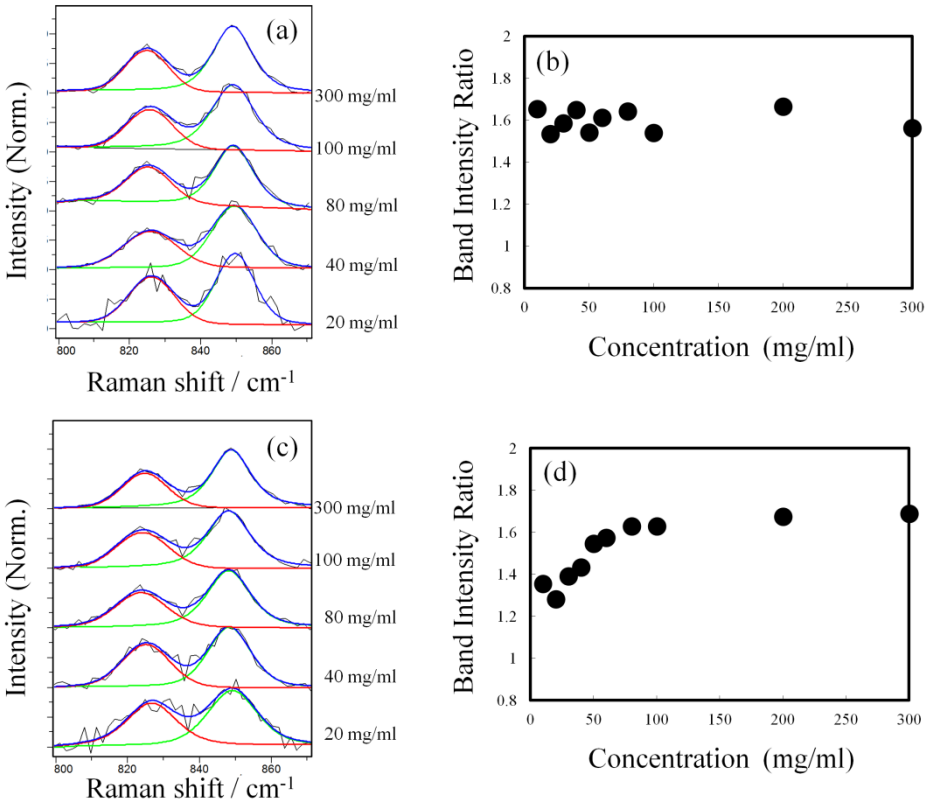

Figure 5. Raman spectra of BSA solutions in the range of 800–880 cm^{-1} at (a) pH7.0 and (c) pH3.0. The concentrations are 20, 40, 80, 100, and 300 mg/mL. The concentration-dependent band intensity ratio I_{850}/I_{825} of BSA at (b) pH7.0 and (d) pH3.0.

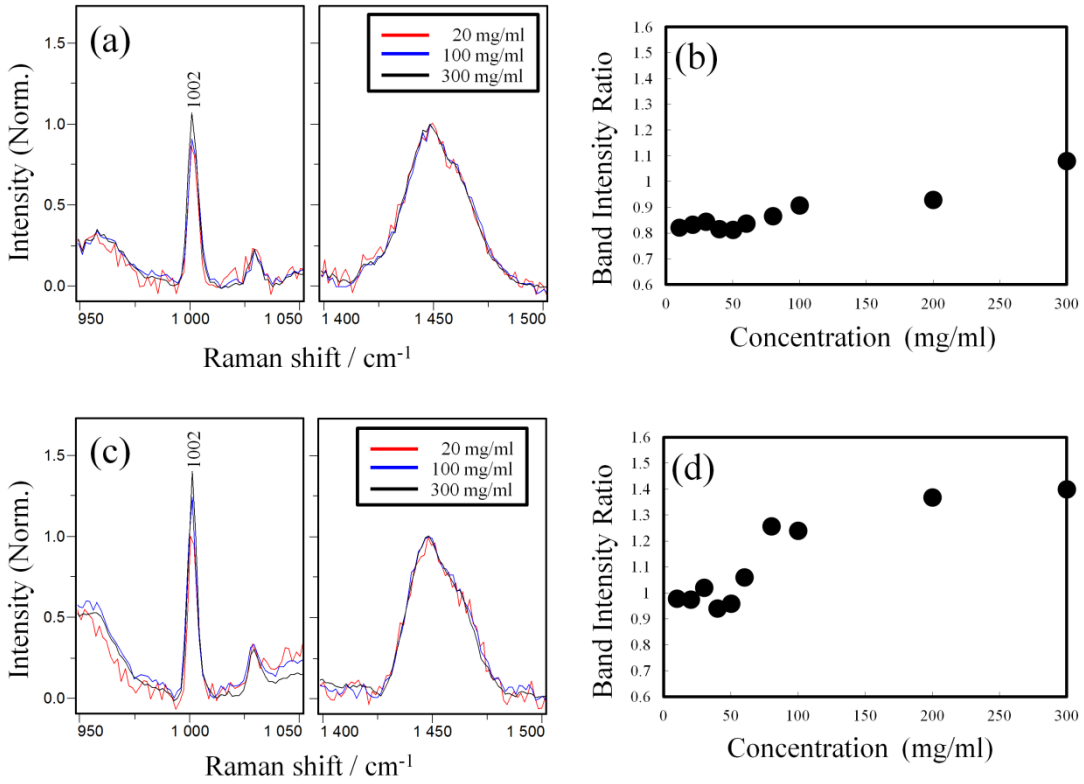

Figure 6. Raman spectra of BSA solutions between 950–1050 and 1400–1500 cm^{-1} at (a) pH7.0 and (c) pH3.0. The concentrations are 20, 100, and 300 mg/mL. The concentration-dependent band intensity ratio I_{1002}/I_{1450} of BSA at (b) pH7.0 and (d) pH3.0.

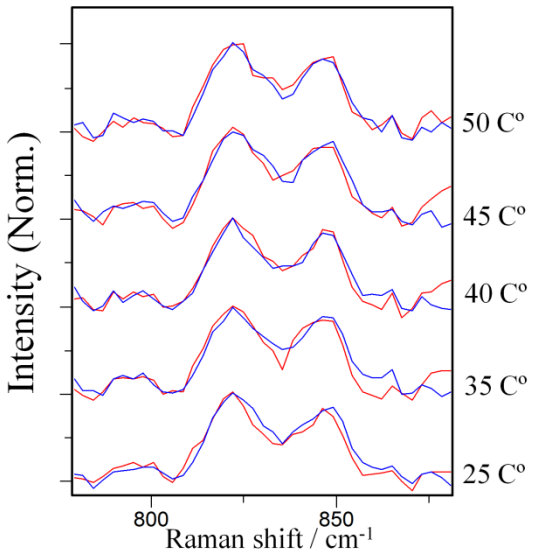

Figure 7. Overlay of the Tyr band ratio I_{850}/I_{825} of (a) BSA (red) and (b) BSA with F20 (blue) over the temperature range of 25–50 °C.

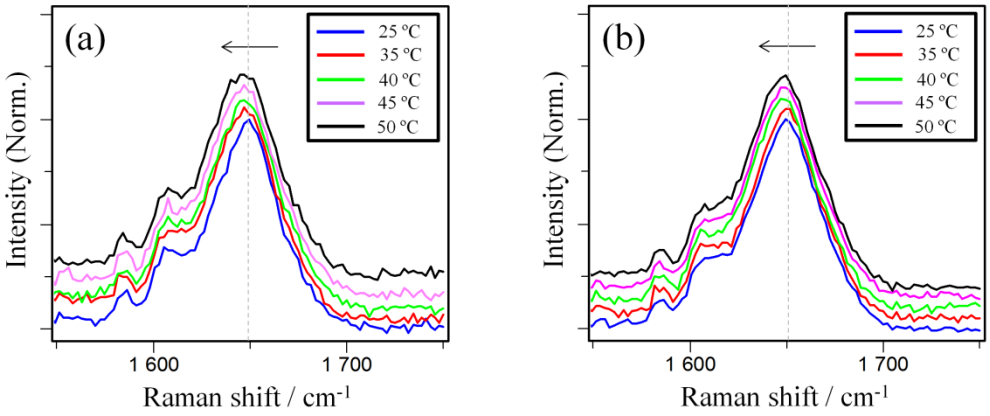

Figure 8. Overlay of the amide I band of (a) BSA and (b) BSA with F20 over the temperature range of 25–50 °C.

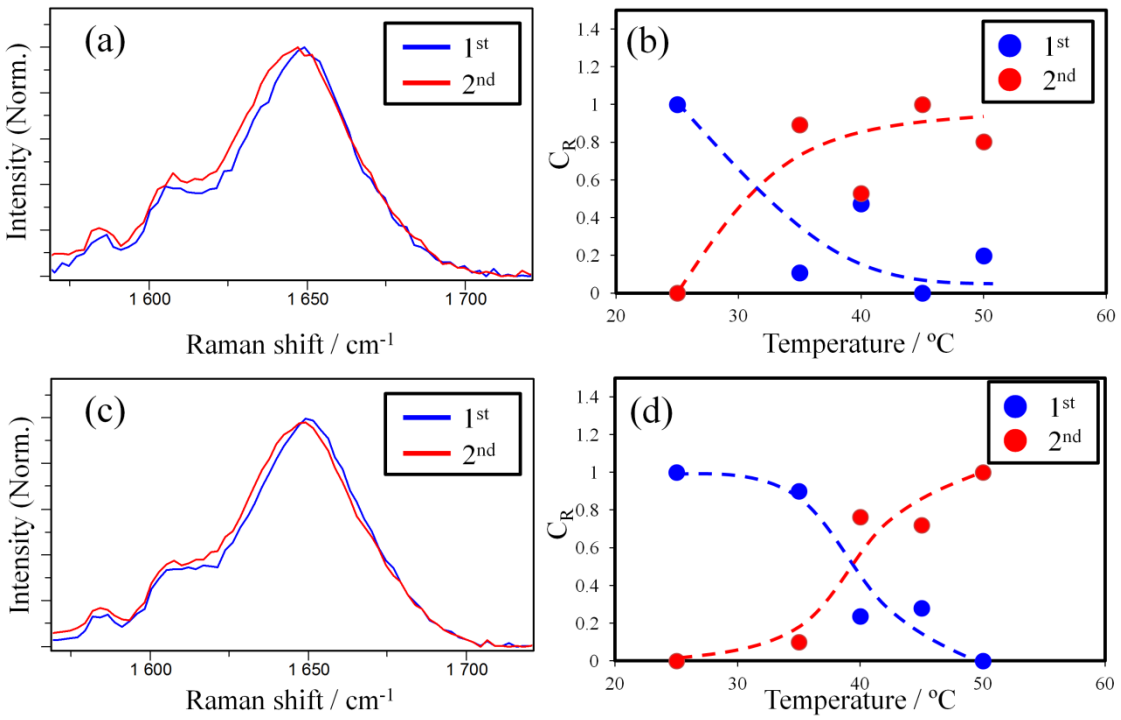

Figure 9. Abstract spectra in the amide I region of (a) BSA and (c) BSA with F20. The scores of (b) BSA solution and (d) BSA solution with F20 by MCR-ALS analysis of the temperature-dependent measurements between 25 and 50 °C. First (blue) and second (red) mean before and after H–D exchange, respectively.

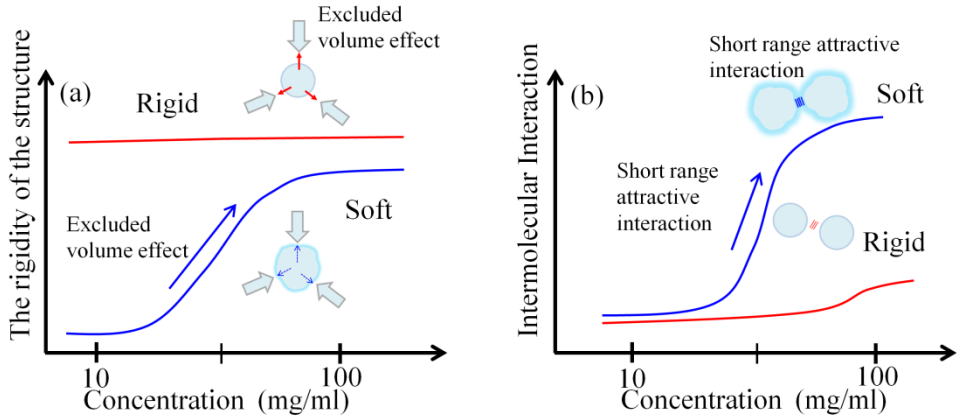

Figure 10. Schematic of (a) the rigidity of the structure and (b) the intermolecular interaction of “rigid” and “soft” molecule with increasing concentration.

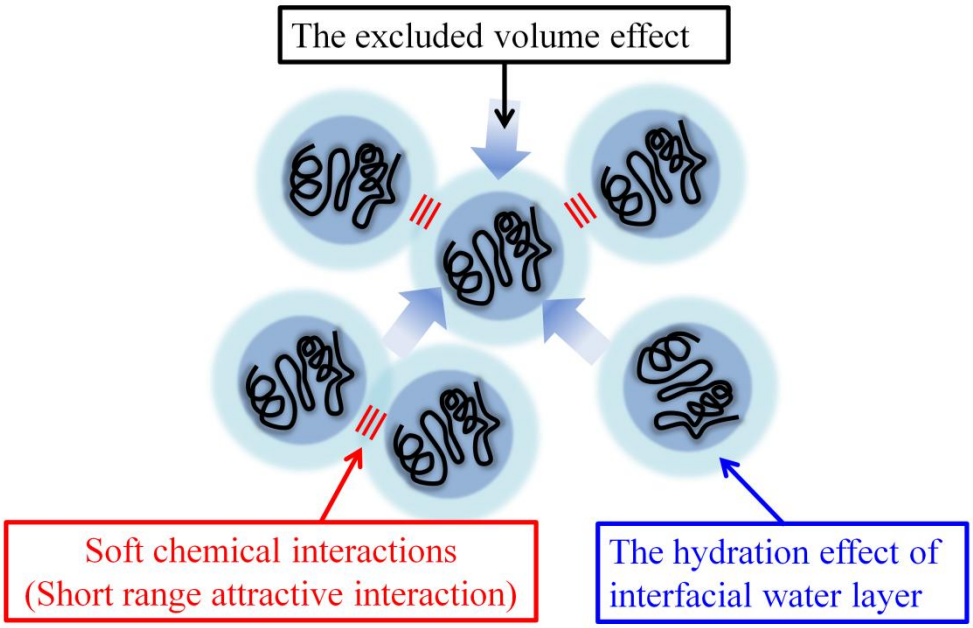

Figure 11. Schematic of the molecular crowding effects on the BSA solution.

6. Supporting Information

The dependence of protein concentration on Raman spectra in the CH and OH stretching region are supplied as Supporting Information 1. The results of PCA (REV, RMSECV) are supplied as Supporting Information 2, 3, and 8. The abstract spectra and the normalized score of the MCR-ALS analysis of amide I band are shown in Supporting Information 4 and 5. The Raman spectra of F20 and its overlay of the amide I band at 25 °C and 50 °C are shown in Supporting Information 6 and 7. The distance between the molecules of the BSA solution as a function of concentration is supplied as Supporting Information 9.

6-1. Supporting Information 1

Raman spectra in the CH and OH stretching regions of BSA at 20, 40, 100, 200, and 300 mg/mL and (a) pH7.0 or (b) pH3.0 are shown in Fig. S1. As the BSA concentration increased, the intensity of the CH stretching bands (2930 cm^{-1}) also increased.

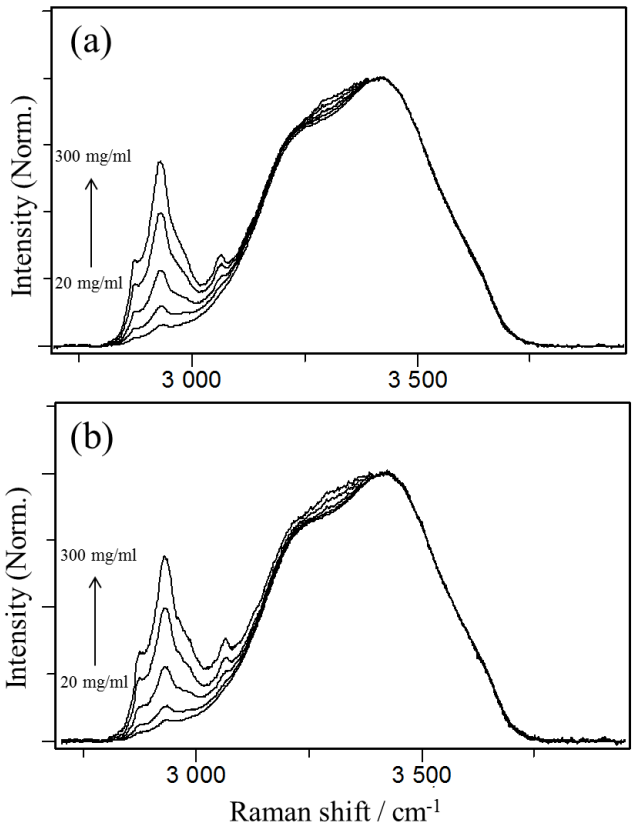

Figure S1. Raman spectra in the CH and OH stretching regions of BSA samples at (a) pH7.0 and (b) pH3.0. The BSA concentrations are 20, 40, 100, 200, and 300 mg/mL.

6-2. Supporting Information 2

To estimate the number of chemical species, REV and RMSECV in the CH and OH stretching regions of BSA samples at pH7.0 and pH3.0 were calculated. In Fig. S2a, the REV of the BSA solution at pH7.0 is shown. Because the changes in the eigenvalue against the factor level were large, we have plotted the ordinate axis using a logarithmic scale. As the number of the factor increased, the REV score decreased. The first and second loading had relatively larger scores and the one after the third score had a smaller score. Because a smaller REV score indicates that the factor is from a minute species or represents noise, this result indicates that the first and second loadings are the main species. In addition, after the third loading, the eigenvalue decreased rapidly and the loading value was essentially at the noise level. This result indicates that the factors after the third factor can be attributed to random noise and the third loading may be assigned to the minor species.

The RMSECV score at pH7.0 is plotted in Fig. S2b. As shown, after the third loading, the RMSECV score is almost constant, which indicates that factors after the

third one can be attributed to random noise and the third loading is very small. Since the third loading is a minor species, the detail of RMSECV was performed by use of ΔRMSECV, which is the difference of RMSECV against the maximum RMSECV. Factors that decrease RMSECV by < ca. 2% of the maximum RMSECV are usually rejected as error factors. In Fig. S2c, the ΔRMSECV (%) at pH7.0 is plotted. The dotted line in Fig. S2c shows 2% of the ΔRMSECV, which can be rejected as error factors. The score shows that the third loading is over 2 %, which indicates the third loading is a real minor species.[34]

Therefore, both REV and RMSECV for pH7.0 showed that there are two chemical species and a very minor species in this data set of the CH and OH stretching region.

The results of the data recorded at pH3.0 contrast those observed at pH7.0. In Fig. S2c, the REV of the BSA solution at pH3.0 is shown. The first and second loadings had relatively large scores and this result indicates that the first and second loadings are the main species. In contrast, the one after the third score had a score similar to that of the noise level and there was not so much difference between the third one and the other scores after that. This result implies that the one after the third loading can be considered as noise.

The RMSECV score at pH3.0 is plotted in Fig. S2d. As shown in the figure, after the second loading, there is no specific improvement of the RMSECV score and the score is almost constant, which indicates that factors after the second factor can be attributed to random noise. To ascertain the detail, the ΔRMSECV (%) is also calculated. As shown in Fig. S2f, the ΔRMSECV (%) at pH3.0 is plotted. The score shows that the third loading is nearly close to 2 %, which indicates the third loading can be assigned to an error factor. Hence, both REV and RMSECV for the data recorded at pH3.0 showed that there are two chemical species in this data set of the CH and OH stretching region.

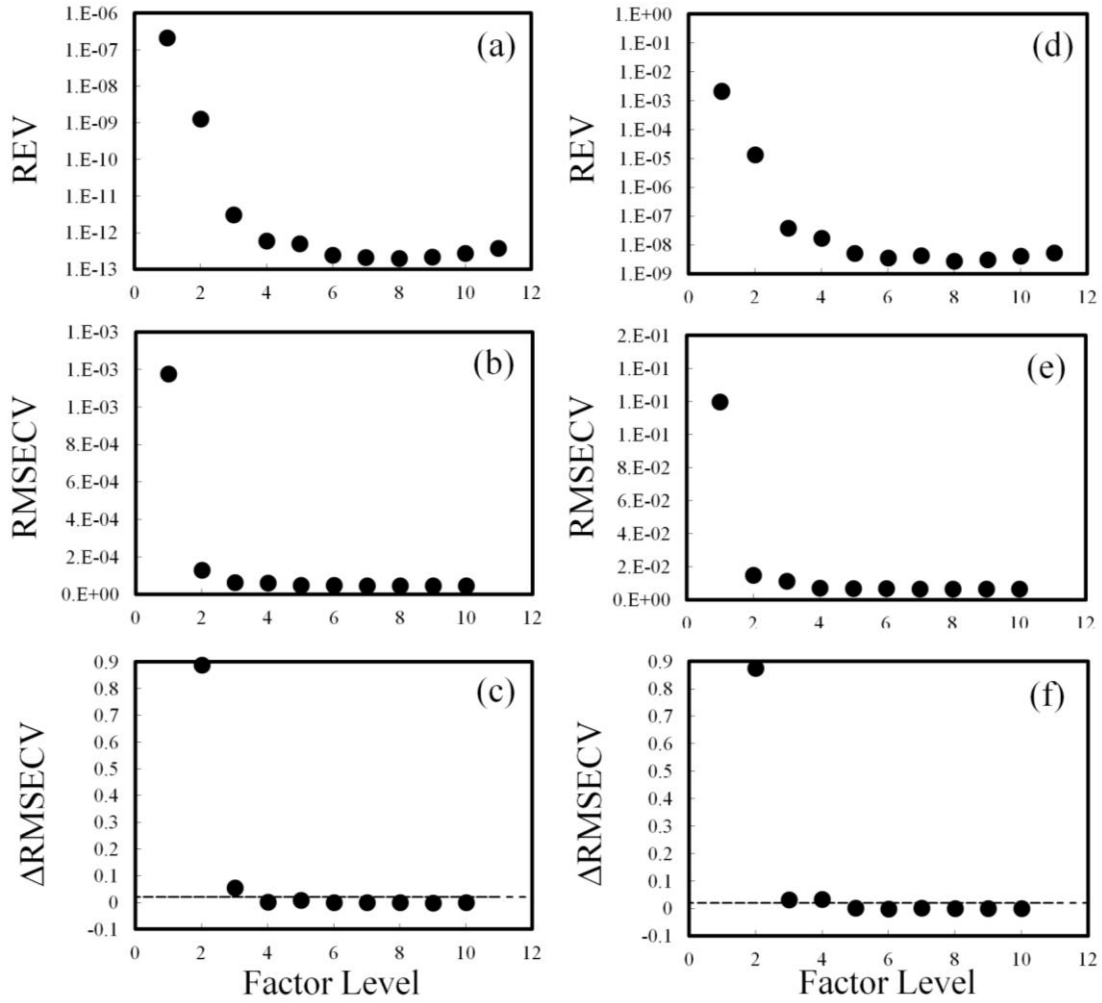

Figure S2. Plots of reduced eigenvalues (REV), root-mean-square prediction error of cross-validation (RMSECV) and the ΔRMSECV against the factor level of the PCA in the CH and OH stretching region of the BSA samples at (a, b and c) pH7.0 and (d, e and f) pH3.0.

6-3. Supporting Information 3

To estimate the number of the chemical species, REV and RMSECV in the amide I band region of the BSA samples at pH7.0 and pH3.0 were calculated. In Fig. S3a, the REV of the BSA sample at pH7.0 is shown. As the number of the factors increased, the REV score decreased. After the second loading, the eigenvalue decreased rapidly and the loading value was essentially the same as the noise level. This result indicates that factors after the second factor can be attributed to random noise. The RMSECV score at pH7.0 is also plotted in Fig. S3b. As shown, the RMSECV score was clearly constant after the second loading. Thus, both REV and RMSECV showed that

there are two chemical species in this data set of the amide I band.

In Figs. S3c and S3d, the REV and RMSECV of the BSA sample at pH3.0 are shown. The REV and RMSECV showed almost the same results as observed for pH7.0 data. Therefore, this also confirmed that at pH3.0 there are two chemical species in this data set of the amide I band.

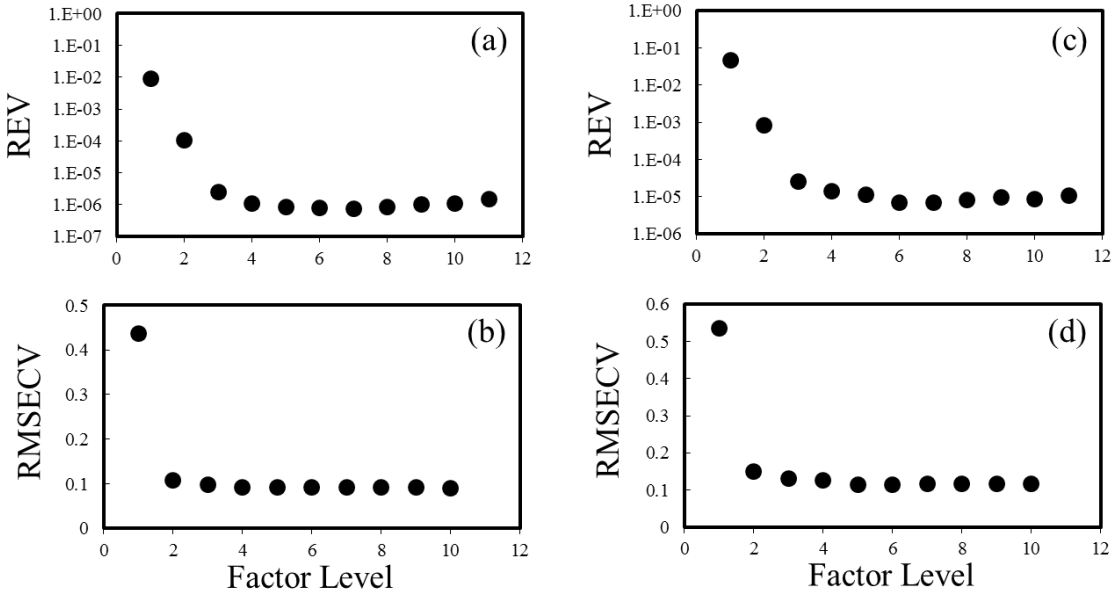

Figure S3. Plots of reduced eigenvalue (REV) and root-mean-square prediction error of cross-validation (RMSECV) against the factor level of the PCA in the amide I band region of the BSA sample at (a, b) pH7.0 and (c, d) pH3.0.

6-4. Supporting Information 4

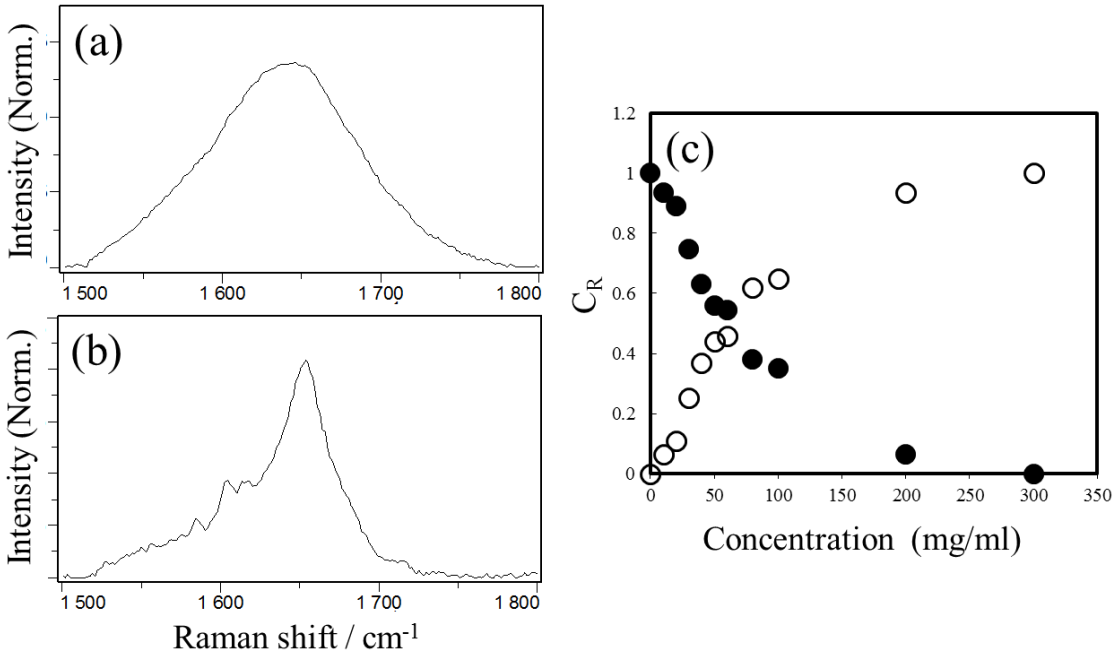

Figure S4. Abstract Raman spectra of (a) the water bending band of the BSA solution at pH7.0 and (b) the amide I band of BSA at pH7.0 by MCR-ALS analysis. (c) The normalized concentration C_R of each component. The filled circles represent the contribution ratio C_R of the water band and the open circles represent the C_R of the amide I band.

6-5. Supporting Information 5

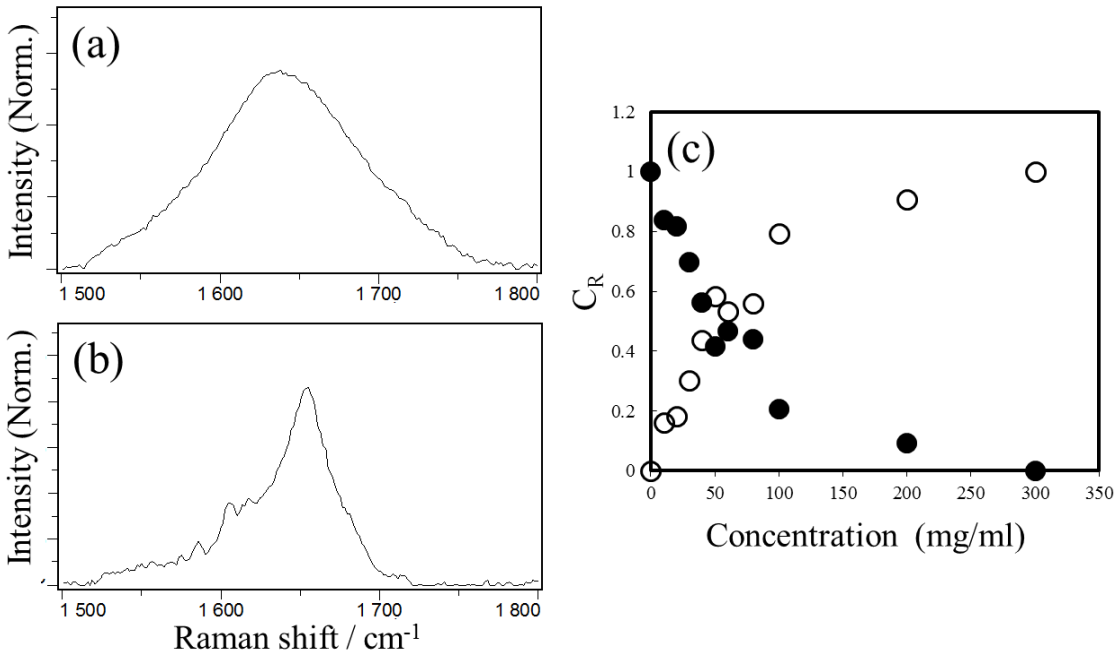

Figure S5. Abstract Raman spectra of (a) the water bending band of the BSA sample at pH3.0 and (b) the amide I band of the BSA sample at pH3.0 by MCR-ALS analysis. (c) The normalized concentration C_R of each component. The filled circles represent the contribution ratio C_R of the water band and the open circles represent the C_R of the amide I band.

6-6. Supporting Information 6

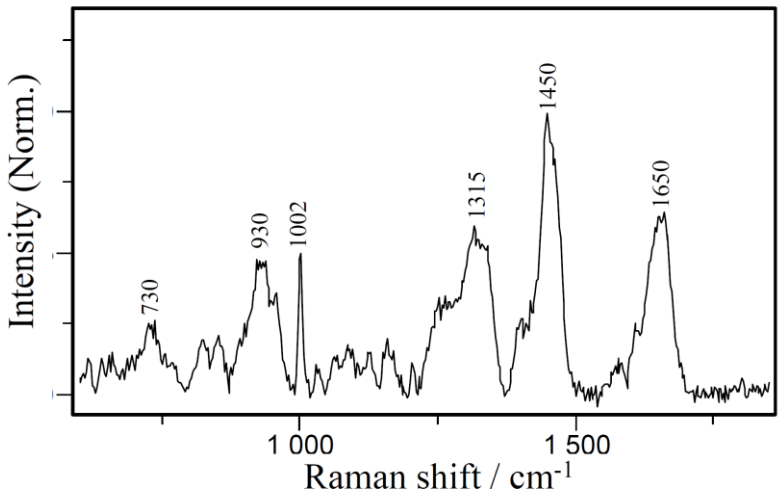

Figure S6. A Raman spectrum of F20 (50 mg/mL) dissolved in D_2O. The D_2O bending band has been subtracted from the spectrum.

6-7. Supporting Information 7

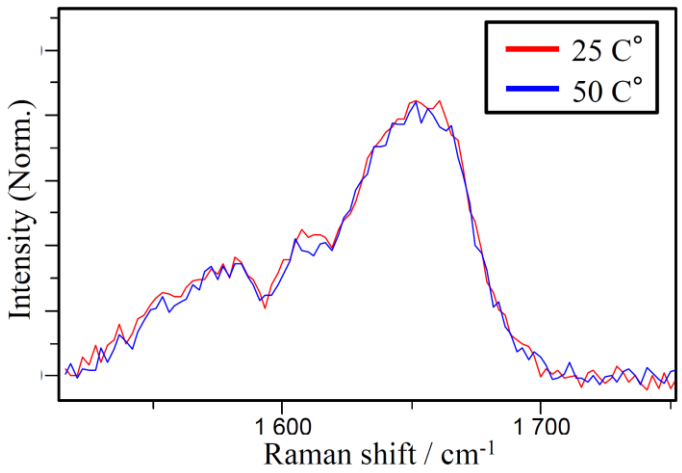

Figure S7. An overlay of the amide I band of the F20 solutions at 25 (red) and 50 °C (blue).

6-8. Supporting Information 8

To estimate the number of chemical species in the H-D exchange experiments, REV and RMSECV in the amide I region of the BSA sample before and after the addition of F20 were calculated. In Fig. S8a, the REV of the BSA sample before the addition of F20 is shown. After the second loading, the eigenvalue decreased rapidly and the loading value was almost the same as the noise level. This result indicates that factors after the second factor can be attributed to random noise. The RMSECV score is also plotted in Fig. S8b. The RMSECV shows a sharper result than the REV. There was a break at the second factor in the RMSECV plots, again suggesting that there are two main species in the data set. Thus, both REV and RMSECV showed that there are two chemical species in this data set of the H–D exchange experiment.

In Fig. S8c, the REV of the BSA solution after the addition of F20 is shown. After the second loading, the eigenvalue decreased rapidly and the loading value was almost the same as the noise level. This result indicates again that factors after the second factor can be attributed to random noise. The RMSECV score is also plotted in Fig. S8d. As shown in the figure, after the second loading, the RMSECV score is clearly constant. Thus, both REV and RMSECV showed that there are two chemical species in this data set in the H–D exchange experiment of the amide I band after the addition of F20.

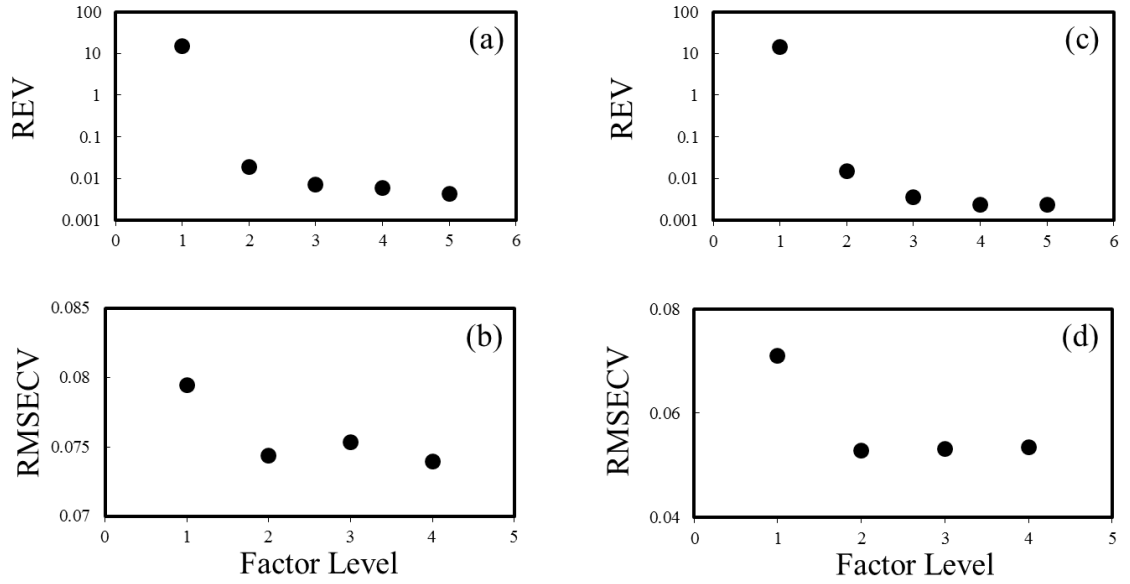

Figure S8. Plots of reduced eigenvalue (REV) and root-mean-square prediction error of cross-validation (RMSECV) against the factor level of the PCA in the amide I region of a BSA sample (a, b) before and (c, d) after the addition of F20.

Supporting Information 9

The distance between surfaces of the molecules of the BSA solution can be calculated as the mean nearest-neighbor distance on the basis of particle size and particle dimension number. The BSA molecule can be approximated by a simple hard spherical particle with a radius (*R*) of 3.48 nm.[5] The mean interparticle surface separation (distance between the molecules) $\langle r \rangle$ for non-interacting hard spheres is given by

$$\langle r \rangle = 2R(\lambda - 1), \tag{5}$$

where λ is the mean distance between the centers of mass of the nearest neighbors obtained from a conditional pair distribution function. Mean distance λ can be expressed in terms of the volume fraction (ϕ),

$$\lambda = 1 + \frac{(1-\phi)^3}{24\phi(1-\phi/2)}, \tag{6}$$

which is given by

$$\phi = (\frac{4\pi R^3}{3M}) N_A C \,, \tag{7}$$

where M is molecular weight (64,000) and C is protein concentration (mg/mL).

The calculated $\langle r \rangle$ values for BSA molecules as a function of concentration were plotted as shown in Fig. S9. At concentrations up to 50 mg/mL, the distance between the BSA molecules $\langle r \rangle$ decreases exponentially, whereas over 50 mg/mL, $\langle r \rangle$ decreases more slowly. At about 50 mg/mL, the distance between the BSA molecules is approximately the same order of magnitude as the size of the molecule.

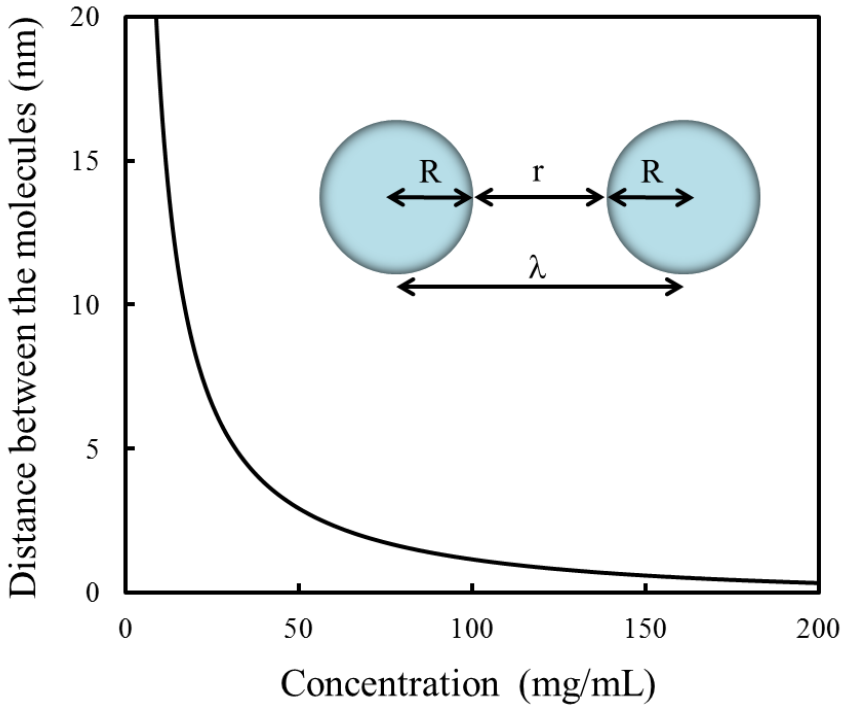

Figure S9. Distance between surfaces of the molecules of the BSA solution as a function of concentration.

7. References

1. Ellis, R. J., Minton, A. P. Cell biology: Join the crowd. *Nature*, **2003**, *425*, 27-28.

2. Minton, A. P. Excluded volume as a determinant of macromolecular structure and

reactivity. *Biopolymers,* **1981**, *20*, 2093-2120.

3. Harada, R., Sugita, Y., Michael, F. Protein crowding affects hydration structure and dynamics. *J. Am. Chem. Soc*. **2012**, *134*, 4842-4849.

4. Stradner, A., Sedgwick, H., Cardinaux, F., Poon, W. C. K., Egelhaaf, S. U., Schurtenberger, P. Equilibrium cluster formation in concentrated protein solutions and colloids. *Nature,* **2004**, *432*, 492-495.

5. Barbosa, L. R. S., Ortore, M. G., Francesco S. F., Mariani, P., Bernstorff, S., Itri, R. The importance of protein-protein interactions on the pH-induced conformational changes of bovine serum albumin: A small-angle X-ray scattering study. *Biophys. J.* **2010**, *98*, 147-157.

6. Südhof, T. C. Neuroligins and neurexins link synaptic function to cognitive disease. *Nature,* **2008**, *455*, 903-911.

7. Sakakibara, D., Sasaki, A., Ikeya, T., Hamatsu, J., Hanashima, T., Mishima, M., Yoshimasu, M., Hayashi, N., Mikawa, T., Wälchli, M., Smith, B. O., Shirakawa, M., Güntert, P., Ito, Y. Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature,* **2009**, *458*, 102-105.

8. Ota, C., Noguchi, S., Tsumoto, K. The molecular interaction of a protein in highly concentrated solution investigated by Raman spectroscopy. *Biopolymers,* **2015**, *103*, 237-246.

9. Carter, D., Ho, J. X. *Advances in Protein chemistry*; Academic Press: New York, 1994; Vol. 45, pp 153-215.

10. Shima, K., Nagao, Ai., Okada, J., Sano, S., Takano, K. Slow unfolding pathway of the hyperthermophilic Tk-rnase H2 examined by pulse proteolysis using mutant proteins. *Biochem. Anal. Biochem*. **2015**, *4*, 213.

11. Okada, J., Koga, Y., Takano, K., Kanaya, S. Slow unfolding pathway of hyperthermophilic Tk-rnase H2 examined by pulse proteolysis using the stable protease Tk-subtilisin. *Biochemistry,* **2012**, *51*, 9178-9191.

12. Malinowski, E. R. *Factor Analysis in Chemistry* (2nd ed.); Wiley-Interscience: New York, 1991.

13. Lobo, A. P., Valles, B. S., Tascón, N. F., Madrera, R. R. Calibration models for routine analysis of cider by mid-infrared spectroscopy. *Lebensm. Wiss. Technol.* **2006**, *39*, 1026-1032.

14. Tauler, R. Application of non-linear optimization methods to the estimation of multivariate curve resolution solutions and of their feasible band boundaries in the investigation of two chemical and environmental simulated data sets. *Anal. Chim. Acta*, **2007**, *595*, 289-298.

15. Wen, Z. Raman spectroscopy of protein pharmaceuticals. *J. Pharm. Sci.* **2007**, *96*, 2861-2871.

16. Oladepo, S. A., Xiong, K., Hong, Z., Asher, S. A., Handen, J., Lednev, I. K. UV resonance Raman investigations of peptide and protein structure and dynamics. *Chem. Rev.* **2012**, *112*, 2604-2628.

17. Shire, S. J., Kalonia, D. S. Viscosity analysis of high concentration bovine serum albumin aqueous solutions. *Pharm. Res.* **2011**, *28*, 1973-1983.

18. Binabaji, E., Ma, J., Zydney, A. L. Intermolecular interactions and the viscosity of highly concentrated monoclonal antibody solutions. *Pharm. Res.* **2015**, *32*, 3102-3109.

19. Feig, M., Pettitt, B. M. Crystallographic water sites from a theoretical perspective. *Structure*, **1998**, *61*, 351-1354.

20. Thompson, H., Soper, A. K., Ricci, M. A., Bruni, F., Skipper, N. T. The three-dimensional structure of water confined in nanoporous vycor glass. *J. Phys. Chem. B*, **2007**, *111*, 5610-5620.

21. Chandler, D. Hydrophobicity: two faces of water. *Nature*, **2002**, *417*, 491.

22. Chen, H., Lin, H., Chen, H., Mai, F., Liu, Y., Lin, C., Chang, C., Tsai, H., Yang, C.

Innovative strategy with potential to increase hemodialysis efficiency and safety. *Sci. Rep.* **2014**, *4*, 1-5.

23. Shen Y. R., Ostroverkhov, V. Sum-frequency vibrational spectroscopy on water interfaces: polar orientation of water molecules at interfaces. *Chem. Rev.* **2006,** *106*, 1140-1154.

24. Davis, J. G., Gierszal, K. P., Wang, P., Ben-Amotz, D. Water structural transformation at molecular hydrophobic interfaces. *Nature*, **2012**, *491*, 582-585.

25. Maiti, N. C., Apetri, M. M., Zagorski, M. G., Carey, P. R., Anderson, V. E. Raman spectroscopic characterization of secondary structure in natively unfolded proteins: α-Synuclein. *J. Am. Chem. Soc.* **2004**, *126*, 2399-2408.

26. Apetri, M. M., Maiti, N. C., Zagorski, M. G., Carey, P. R., Anderson, V. E. Secondary structure of α-synuclein oligomers: characterization by Raman and atomic force microscopy. *J. Mol. Biol.* **2006**, *355*, 63-71.

27. Siamwiza, M. N., Lord, R. C., Chen, M. C., Takamatsu, T., Harada, I., Matsuura, H., Shimanouchi, T. Interpretation of the doublet at 850 and 830cm^{-1} in the Raman spectra of Tyrosyl residues in proteins and certain model compounds. *Biochemistry,* **1975**, *14*, 4870-4876.

28. Hernández B., Coïc, Y., Pflüger, F., Kruglik S. G., Ghomi M. All characteristic Raman markers of tyrosine and tyrosinate originate from phenol ring fundamental vibrations, *J. Raman Spectrosc.* **2016**, *47*, 210-220.

29. Xu, M., Ermolenkov, V. V., He, W., Uversky, V. N., Fredriksen, L., Lednev, I. K. Lysozyme fibrillation: Deep UV Raman spectroscopic characterization of protein structural transformation. *Biopolymers*, **2005**, *79*, 58-61.

30. Xu, M., Shashilov, V. A., Ermolenkov, V. V., Fredriksen, L., Zagorevski, D., Lednev, I. K. Hen egg white lysozyme fibrillation: a deep UV resonance Raman spectroscopic study. *J. Biophotonics,* **2008**, *1*, 215-229.

31. Wu Y., Murayama, K., Ozaki, Y. Two-dimensional infrared spectroscopy and principle component analysis studies of the secondary structure and kinetics of hydrogen-deuterium exchange of human serum albumin. *J. Phys. Chem. B*, **2001**, *105*, 6251-6259.

Chapter 4. The Assessment of the Protein–Protein Interactions in a Highly Concentrated Antibody Solution by Using Raman Spectroscopy

1. Introduction

Many biopharmaceuticals have strong potential as drugs due to their high efficacy, and in the biopharmaceutical industry, liquid formulations are becoming increasingly popular.[1] For such formulations, highly concentrated (>100 mg/mL) protein solutions are required. At high concentrations, however, the increased viscosity is accompanied by an increased risk of protein aggregation or denaturation. These changes may result in side effects, and therefore the optimization of drug formulations is of great importance. Accordingly, a comprehensive understanding of the interactions of proteins in nonideal solutions (e.g., highly concentrated solutions) is essential. In particular, protein–protein interactions that could cause oligomerization or aggregation in highly concentrated solution should be thoroughly investigated.

Currently, relatively few techniques are available to characterize protein–protein interactions in high-concentration solutions. The second virial coefficient (B_{22}), as an indicator of deviation from ideality, or the interaction parameter (k_D), as deduced from dynamic light scattering (DLS), can give information about protein-protein interactions but these factors are applicable only in a dilute solution.[1,2] In highly concentrated solutions, attractive interactions such as dipole-dipole interactions would be expected to increase, therefore, it is extremely important to understand such interactions in detail.

Small-angle X-ray scattering and small-angle neutron scattering are powerful

analytical methods for investigating the structure of molecules in high-concentration solutions,[3] but obtaining information about protein–protein interactions by using these techniques is difficult. Raman spectroscopy is a powerful tool for detailed investigation of the behavior of functional groups in proteins.[4,5] Specifically, Raman spectroscopy has two important advantages: information about each type of amino acid residue can be obtained from the Raman spectrum of a protein, and Raman spectroscopy is one of the best methods for analyzing high-concentration solutions.[6] In this study, on the basis of our previous study of a model system,[6] we used Raman spectroscopy to analyze the protein-protein interactions in a highly concentrated antibody solution as a practical system.

2. Experimental Methods

2-1. Materials

2-1-1. Concentration Dependence Measurement

Purified IgG from rabbit serum (>95%) was purchased from Sigma Aldrich (product no. I5006, St. Louis, MO, USA). A buffer solution of IgG was prepared by mixing trisodium citrate dihydrate (code 31404, 500 mg, lot no. M2E8017, Nacalai Tesque, Kyoto, Japan) with citrate acid monohydrate (code 09106-15, 50 mg, lot no. M2G9914, Nacalai Tesque). The concentration of the buffer solution was 20 mM and the concentration of sodium chloride (code 31320-05, 500 g, lot no. V1N6638, Nacalai Tesque) was 150 mM. The final pH of the solution was adjusted to 6.0 by using a LAQUA F-72 pH meter (HORIBA, Kyoto, Japan). Test solutions of IgG were prepared at concentrations of 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, and 200 mg/mL by dilution of the buffer solution with water that was purified with a Milli-Q laboratory

water purifier (Millipore, Billerica, MA, USA). All measurements were performed at room temperature (~20 °C).

2-1-2. Application Study

To investigate the pH effect on the concentration dependence measurement, test solutions of IgG at pH4.0 were prepared. The concentrations were 10, 20, 30, 40, 60, 80, 100 mg/mL. The concentration of the buffer solution was 20 mM and the concentration of sodium was 150 mM as was the same as at pH6.0.

For the measurement of the salt effect on the molecular interaction, the salt dependence measurements were performed. The concentration of IgG was 50 mg/mL (pH6.0). The salt concentrations of these IgG solutions were 10, 50, 100, 500, 1000 mM NaCl.

For the storage stability test, the fresh 100 mg/mL IgG solution (pH6.0, 150 mM NaCl) was preserved at 4 ºC for 3 month in a darkroom.

2-2. Raman Spectroscopy

Raman spectra were obtained with an XploRA Raman microscope equipped with Labspec 6 software (HORIBA Jobin Yvon, Paris, France). The spectrograph of the microscope had the Czerny–Turner configuration, and the focal length was 200 mm. Raman spectra were measured in the back-scattering geometry. A thermoelectrically cooled Synapse charge-coupled device camera (HORIBA) was used as the detector.

The entrance slit of the spectrometer was set to 100 μm. The dispersive element had a grating of 1800 lines/mm, which provides a wavenumber resolution of ca. 2 cm^{-1}. Excitation was accomplished with the 532 nm line, and the laser power was 40 mW at

the sample surface. The excitation intensity was monitored with a TQ8210 optical power meter (Advantest, Tokyo, Japan). The scattered light was collected through a multipass cell holder (HORIBA Jobin Yvon) with a visible macro lens (focal length, 40 mm) for high sensitivity. The optical cell was a 115-QS micro cell (Hellma Analytics, Müllheim, Germany). The Raman shift was calibrated by means of the atomic emission lines of a neon lamp.

2-3. Spectroscopic Analysis

Principal components analysis (PCA) is a multivariate analysis method that has proven effective for distinguishing the spectra of a pure component from several kinds of spectra. PCA has also proven useful for detecting the spectra of minute chemical species in different local environments or interactions from those of dominant bulk species.[7] The number of chemical species in the antibody solution was estimated by using PCA. The accuracy of the predictive ability of the model was evaluated by rank analysis using a reduced-eigenvalue (REV) plot. REV intrinsically has the same statistical meaning as the normal eigenvalue (EV), but REV is better for the evaluation of minute eigenvalues because it takes into account the degree of freedom.[7] A REV plot is obtained by using the following equation:

$$REV_j = \frac{EV_j}{(N - j + 1)(M - j + 1)}, \quad (1)$$

where j is the factor level, and the parameters N and M represent the size of the spectral matrix ($N \times M$). Of note, the analysis of basis factors by means of eigenvalue analysis does not take into account concentration (intensity) information. Therefore, the cross-validation (CV) technique, which takes into account both spectrum and intensity information, is useful to ascertain the accuracy of the results of REV.[8]

Root-mean-square prediction error of cross-validation (RMSECV), which is CV with leave-one-out cross-validation,[8] is also carried out by using the following equation:

$$RMSECV = \sqrt{\frac{\sum_{j=1}^{n} (\hat{c}_j - c_j)^2}{n}}, \quad (2)$$

where $\hat{c}_j$ is the predicted concentration, c_j is the actual concentration, as determined by the reference method, and n is the number of samples used in the calibration model. The use of both REV and RMSECV increases the accuracy of the factor analysis of the PCA. Although PCA is useful to investigate the mathematical structure of data, interpretation of the results can be complex. This is because the loading spectra have negative bands due to the relative quantitative change, which cannot be treated as pure component spectra. Multivariate curve resolution (MCR) is a complementary approach because MCR is based on nonnegative matrix factorization and the results obtained with MCR have more direct physical meaning.[9] Specifically, if the number of species is known, the results from MCR are more realistic. In matrix form, MCR analysis can be written as

$$D = CS + E, \quad (3)$$

where $\boldsymbol{D}$ is the original data matrix, $\boldsymbol{S}$ is the pure matrix of the pure spectra, $\boldsymbol{C}$ is the concentration of each species, and $\boldsymbol{E}$ is the matrix of the residuals not explained by the chemical species of conformations in $\boldsymbol{C}$ and $\boldsymbol{S}$ and should be close to the experimental error. The contribution ratio C_R is the normalized concentration, which is the quotient upon division of the concentration of the sample (C_j) by the sum of concentrations of all components as described below.

$$C_R(j) = \frac{C_j}{\sum_{j=1}^{h} C_j}. \quad (4)$$

3. Results

3-1. Raman Spectrum of IgG

In the wavenumber region from 200 to 2000 cm^{-1} of the Raman spectrum of IgG (Fig. 1a, 100 mg/mL), there are many bands for each functional group, and from this "fingerprint region," information about conformational changes and changes in the microenvironments of the amino acid residues can be obtained. The relationships between Raman band characteristics (wavenumber, band intensity ratio, and full width at half-maximum) and protein secondary structure are well established.[4,5] The strong and broad bands in the wavenumber region from 2800 to 4000 cm^{-1} of the spectrum of IgG are assigned to the CH and OH stretching modes of IgG and water, respectively (Fig. 1b). As the IgG concentration was increased, changes in several of the Raman bands were observed, as described in detail below.

3-2. Amide I Band

The location of the amide I band of proteins depends mainly on the strength of the hydrogen-bonding interactions (C=O···HN) involving amide groups and the strength of the dipole-dipole interactions between carboxyl groups. Thus, the location of this band provides information about protein secondary structure.[10,11] In the Raman spectrum of IgG, the amide I band overlaps with the band for the bending mode of water at 1640 cm^{-1}. We therefore used PCA to separate these two bands. To estimate the number of the chemical species, REV and RMSECV were calculated and the detailed analysis results of these values are shown in Fig. S1a and b, respectively. The results of both REV and RMSECV show that there are two chemical species in this concentration data set of the amide I band.

MCR-ALS analysis with two-component approximation was performed based on the result of the PCA. As shown in Fig. 2a and b, the abstract spectra are the amide I band of the protein and the OH bending band. The band position of the resolved spectra of the protein is around 1670 cm^{-1}, which means that this IgG has mainly β-sheet and partially α-helical and non-helical structure. The red circles shown in Fig. 2c represent the contribution ratio C_R of the protein and the blue circles represent the C_R of water. As expected, when the C_R of the protein increases, the C_R of water decreases proportionally. Using this quantitative information, the OH bending band was subtracted from the original spectra of each concentration. In Fig. 2d, the Amide I bands of each concentration (10, 20, 40, 60, 80, 100, 200 mg/mL) are shown. This result shows that there is not specific change of Amide I band of each concentration, therefore, at these concentrations, there is not significant change of the secondary structure.

In a highly concentrated solution, the structure of the hydrogen bond of water could change, however, the water bending spectrum would not change, because it is not sensitive to the hydrogen-bonded structure.[12] In contrast, the OH stretching band is sensitive to the hydrogen-bonded structure, and therefore, next we focused on the spectra of the OH stretching band at around 2800 to 4000 cm^{-1}.

3-3. CH and OH Stretching Bands

To analyze the interaction between water and protein in a highly concentrated solution in depth, we subjected the data set of the concentration dependence spectra in the CH and OH stretching regions to PCA. The concentrations analyzed are 0 mg/mL (buffer solution) and 10 to 200 mg/mL. The Raman spectrum in the CH and OH stretching regions of the IgG solution at 10, 40, 80, 100, and 200 mg/mL are shown in

Fig. S2.

For the estimation of the number of the chemical species, REV and RMSECV were calculated. The REV and RMSECV of these data (0–200 mg/mL) and the detailed analysis results of these values are shown in Figs. S3a and S3b, respectively. The results of both REV and RMSECV indicate that there are two main species with four minute species, suggesting that highly concentrated solutions can contain various minute species. Detailed interpretation of these results, however, is complex, because the characterization of these four species is difficult from PCA. Therefore, to simplify the analysis of these data, the range of the analysis was divided into two parts: (i) lower concentrations (0–60 mg/mL) and (ii) higher concentrations (60–200 mg/mL).

3-3-1. Lower Concentrations (0–60 mg/mL)

For the lower concentrations, the data set included 10–60 mg/mL and the buffer solution. The results of the REV and RMSECV are shown in Figs. S3c and S3d, respectively. Both results indicate the presence of two main species and one minute one.

On the basis of the result of PCA, the MCR-ALS analysis was applied to this lower concentration data set. The abstract spectra of each component are shown in Figs. 3a,c,e and the concentration of each component is shown in Figs. 3b,d,f. In Fig. 3a, the spectra have two features: the CH band at around 2940 cm^{-1} from the protein, and a water band, with two main bands at 3290 cm^{-1} and at around 3570 cm^{-1}, which is mainly from the hydrated water of the protein. Figure 3c shows the typical spectral shape of bulk water, which has two main bands at 3240 and 3420 cm^{-1}. As shown in Figs. 3b and 3d, the concentration of the hydrated protein component increases as the concentration increases, whereas the concentration of the bulk water decreases. The third component

has a unique profile with respect to concentration changes (Fig. 3f). Up to 30 mg/mL, the concentration ratio of this component increases; however, over 30 mg/mL, the concentration ratio decreases as the concentration increases. This behavior indicates that the third component is displaced as the protein-protein interaction increases. The abstract spectra of this component are also of interest, comprising two main bands: one at around 3560 cm^{-1} and the other at the lower region at around 3150 cm^{-1}. The former band represents weakly hydrogen-bonded water which may be the disordered water around the hydrophilic pocket and the latter band represents strongly hydrogen-bonded water which may be the "ice-berg" water around hydrophobic residue.[13–17] For these results, the third component can be assigned to the interfacial water between the bulk water and the protein.

3-3-2. Higher Concentrations (60–200 mg/mL)

For the higher concentrations, the data set included 60–200 mg/mL and the buffer solution. The results of the REV and RMSECV are shown in Figs. S3e and S3f, respectively. The results of both REV and RMSECV show there are two main species at these concentrations.

Therefore, MCR-ALS analysis was applied to the higher concentration data set with two-component approximation, bulk water and protein and its hydrated water as shown in Figs. 4a,c. As the concentration increases, the C_R of the water decreases, whereas the C_R of the protein increases (Figs. 4b,d). Of note, there is a difference in the water Raman bands of the resolved abstract spectra of the second species between the lower and higher concentrations as shown in Fig. S4. The main difference between these Raman spectra is the spectral shape at 3550 cm^{-1}. The band intensity at 3550 cm^{-1} at the

higher concentrations is lower than that at the lower concentrations, indicating that at the higher concentrations, the amount of the interfacial water around the protein is smaller compared with that at the lower concentrations. This is because as the concentration increases, the protein hydrates the neighboring water molecules, and therefore there is less interfacial water around the protein as the concentration increases. In terms of the thermodynamics, this behavior of the interfacial water can be explained. It means that these dehydrated water molecules act favorably with the gain in entropy when they are displaced from the interface. This entropic gain compensates the entropy loss from the excluded volume effect with the increase of concentration.

In summary of the results of CH and OH stretching bands, the amount of the interfacial water begins to decrease at the concentration greater than 30 mg/mL.

3-4. Fingerprint Region

To obtain detailed information concerning the functional groups of the protein, we focused on the fingerprint region of the Raman spectra. The fingerprint region of the Raman spectra is rich in information about the protein, because the Raman bands of amide III, from the peptide backbone, and each functional group, such as Phe, Tyr, and Trp, are found in this region.[4,18] Interpretation of the region, however, is not easy because the Raman bands of each functional group overlap. Therefore, rather than focus on each specific Raman band, the better approach is to focus on the entire fingerprint region and to interpret changes over a larger spectral region. For these reasons, we applied PCA to the whole fingerprint region (980–1500 cm^{-1}).

The data set included concentrations ranging from 10 to 200 mg/mL and the buffer solution. The results of the REV and RMSECV and the detailed analysis result of these

values were plotted by using PCA (Figs. S5a,b). The results of both REV and RMSECV show that there are three main species at these concentrations.

We also applied MCR-ALS analysis to the fingerprint region data set. The C_R of each component is shown in Figs. 5a–c and the abstract spectra of each component are shown in Fig. 5d. The abstract spectra were overlayed and normalized at 1240 cm^{-1}. The abstract spectrum of the first, second and third components are shown as a black, red and blue line. The band at 1240 cm^{-1} is assigned to the Amide III band of the IgG molecules. The band at 1415 cm^{-1} and the one at 1090 cm^{-1} are assigned to the COO^- and the ν (C–C) stretch vibration of citric acid, respectively.[21,22]

Of note is that the band intensity at 1415 cm^{-1} and the one at 1090 cm^{-1} of citric acid are clearly different. (Fig. 5d) This result indicates that the spectrum from the buffer solution is dominant in the first loading, whereas the spectrum from the IgG molecules is dominant in the second loading. In other words, the first-loading spectrum is mainly dominant at the lower concentration and the second-component spectrum is mainly dominant at the higher concentration. As expected, the C_R of the first-component spectra is dominant at the lower concentration and decreases as the protein concentration of the solution increases. In contrast, the C_R of the second-component spectra decreases as the concentration of the solution increases. Interestingly, the third component has a unique profile, because the ratio of this component increases at concentrations greater than 80 mg/mL. The spectrum of the third component has another unique feature: the band intensity of Phe at 1004 cm^{-1} of the third-component spectra is greater than that of the second-component spectra, although the other main bands such as the Amide III band at 1240 cm^{-1} have almost the same intensity, which suggests that there is not significant change of the secondary structure. In Fig. 5e, the

band intensity ratio I_{1004}/I_{1240} was replotted.

This increase in the relative intensity of Phe indicates that not the effect of the increase of concentration but the change of the interaction around Phe caused this behavior. Previous studies have shown that changes in the relative intensity of Phe are caused by changes in the local environment of Phe from hydrophilic conditions to hydrophobic ones as the amount of available water molecules surrounding Phe decreases.[19,20] These results are consistent with our findings, since at high concentrations, the hydrated water around the IgG molecules decreased as shown by the result of the water Raman band analysis.

In addition, hydrocarbons residues such as Phe and $-CH_2-$ are attracted each other and self-assembling due to their high polarizability and atomic density.[23] Thus, the Phe residue will be surrounded by its neighboring hydrocarbons residues. This will cause the CH–π interaction of Phe. From the previous theoretical study, the CH–π interaction of Phe is known to increase polarizability.[24] This result can explain the increase in the relative intensity of Phe, since the intensity of Raman scattering is proportional to the polarizability. While π–π stacking could also increase the polarizability of Phe, this interaction is rarely seen in proteins. Hence, the CH–π interaction is one of the strong candidates of this cause.

The average distance between the C donor and the π acceptor in a CH–π interaction is less than 1 nm, which means that in the highly concentrated solutions the distance between the molecules is likely less than 1 nm. In such a state, with the distance between the proteins being so small, the protein molecules may have been in the initial stages of aggregation. In other words, the band intensity ratio I_{1004}/I_{1240} can be used as a maker of the short range attractive interaction such as the CH–π interaction.

3-5. Band Intensity Ratio of Tyr

The band intensity ratio I_{856}/I_{830} is a known marker of hydration.[25] A low I_{856}/I_{830} ratio indicates that the OH group of Tyr acts as a strong hydrogen donor, and an increase in the ratio indicates that the OH group is acting as both a hydrogen acceptor and a hydrogen donor. Our previous study of a highly concentrated solution of lysozyme showed that the ratio of these bands (I_{856}/I_{830}) can be used as a maker of Tyr side chain interaction.[6] In the case of IgG, many Tyr residues are part of the Fab region; therefore, we would expect the I_{856}/I_{830} ratio to be sensitive to Tyr side chain interactions with neighboring protein molecules, originating from dipole–dipole interactions. Such interactions was estimated from a theoretical calculation.[26]

We compared the Raman spectra in the region from 820 to 890 cm^{-1} for 10 and 200 mg/mL solutions of IgG and found that the I_{856}/I_{830} ratios at these two concentrations differed (Fig. 6a). Another band at 877 cm^{-1} is attributed to a benzene 12-like vibration from the phenyl ring coupled with N_1–H bond motion.[4,6] To calculate the ratio at each concentration, we fitted the 830 and 856 cm^{-1} bands and another 877 cm^{-1} band with a Gaussian–Lorentz function, and normalized the bands by the intensity of the 877 cm^{-1} band (the fitting results are indicated by the blue curves in Fig. 6a). Comparison of the I_{856}/I_{830} ratios at concentrations ranging from 10 to 200 mg/mL shows that the ratio starts to increase at 30 mg/mL and becomes saturated at concentrations of >80 mg/mL (Fig. 6b). This result suggests that the side chain begins to interact at a concentration of 30 mg/mL as the distance between the molecules decreases and become saturated at concentrations greater than 80 mg/mL, because the molecules are tightly packed at this concentration.

3-6. Bandwidth of Trp

The Raman band for the ring-breathing mode of the indole ring of Trp appears at 760 cm^{-1} (Fig. 7a). We found that the width of the band did not vary with concentration in the range of 10 to 200 mg/mL (Fig. 7b), indicating that the conformation of the indole ring was unaffected by changes in concentration. The Raman band at 1555 cm^{-1} is due to the C=C stretching vibration of the C_2=C_3–C_b–C_a moiety of the indole ring,[27,28] and this band is a conformational marker that provides information about the torsion angle ($v_{2,1}$) of the C_2=C_3–C_b–C_a moiety. In contrast to the width of the 760 cm^{-1} band, the width of the 1555 cm^{-1} band increases with increasing concentration up to 80 mg/mL as shown in Fig. 7c,d. The increase of the width of this band can be explained by the excluded volume effect, because in highly concentrated solution, the entropy ΔS decreases and Trp residue avoid the unfavorable steric repulsion to other amino acid residue, which increases the variety of the conformation of Trp. From our previous study[6], the width of this band at 1555 cm^{-1} can be used as a maker band of the molecular interactions, thus this result indicates that as the concentration increases, molecular interactions, such as long-range repulsion, between IgG molecules increases. We expected the band to continue to increase in width at concentrations of >80 mg/mL, but no such increase is observed. One possible reason for this result is that in this concentration range, the protein may be in a metastable state.

4. Discussion

4-1. Relationship between the Distance between Molecules and Their Interactions

To evaluate molecular interactions at high concentration, we must consider the

relationship between changes in the Raman spectra and the distance between the protein molecules. To estimate the distance between IgG molecules at each concentration, we calculated the mean nearest-neighbor distance on the basis of particle size and particle dimension number. Although the true shape of an antibody is not that of a sphere, IgG can be approximated by a simple hard spherical particle with a radius (R) of 4.4 nm.[29] The mean interparticle surface separation $\langle r \rangle$ for non-interacting hard spheres is given by

$$\langle r \rangle = 2R(\lambda - 1), \tag{5}$$

where λ is the mean distance between the centers of mass of the nearest neighbors obtained from a conditional pair distribution function (Fig. S6a). Mean distance λ can be expressed in terms of the volume fraction (ϕ),

$$\lambda = 1 + \frac{(1-\phi)^3}{24\phi(1-\phi/2)}, \tag{6}$$

which is given by

$$\phi = (\frac{4\pi R^3}{3M}) N_A C, \tag{7}$$

where M is molecular weight (150,000) and C is protein concentration (mg/mL).

We plotted the calculated $\langle r \rangle$ values for IgG molecules as a function of concentration (Fig. S6b). At concentrations up to 50 mg/mL, the distance between the IgG molecules decreases exponentially, whereas above 50 mg/mL, $\langle r \rangle$ decreases more slowly. At about 50 mg/mL, the distance between the IgG molecules is approximately the same order of magnitude as the size of the protein. By using this calculated result, the horizontal axis of the concentration was converted to the one of the distance

between the molecules.

In Figs. 8a–c, the width of the band at 1555 cm^{-1}, the I_{856}/I_{830} ratio for Tyr, and the I_{1004}/I_{1240} ratio for Phe are shown as a function of the distance between molecules. In Fig. 8a, there is a steep rise in the width of the band at 1555 cm^{-1} as the distance decreased from 50 to 5 nm; that is, the bandwidth starts to change even when the distance between the IgG molecules is still relatively long. This result implies that the increase in the conformational flexibility of the protein is due mainly to an electrostatic repulsive interaction that acts even over long distances. In Fig. 8b, the I_{856}/I_{830} ratio for Tyr increases with decreasing distance (increasing concentration). In particular, there is a steep rise in the ratios when the distance between the IgG molecules decreases from <5 nm to 2 nm, and at distances longer than 2 nm, the ratios becomes saturated. Compared with the change in the 1555 cm^{-1} band, this change begins when the distance between the IgG molecules is relatively short. This result indicates that side chain interactions become important as the IgG molecules approach each other. In Fig. 8c, the I_{1004}/I_{1240} ratio for Phe drastically increases at distances of less than 2 nm. This result indicates the CH–π interaction occurs when the IgG molecules are within a distance of half of their radius (i.e., almost attached to each other).

On the basis of our results, we propose that the molecular interactions take place in a stepwise process that depends on the distance between the molecules. In the tested concentration range (10–200 mg/mL), the secondary structure of IgG did not change; however, the molecular interactions of the protein molecules changed as the concentration changed, as shown in Fig. 9. The boundary value of the distance is almost the same as the van der Waals radius of the IgG molecule. The details are shown below.

4-1-1. C < 30 mg/mL ($d_M > R$)

In this state, the distance between the IgG molecules is larger than *R*. The IgG molecules remain disperse in solution and interfacial water still exists around the protein. Most of the IgG molecules interact with the surrounding water and they are kept stable by the hydrogen interactions with water. Under these conditions, the electrostatic repulsive interactions still predominate.

4-1-2. 30 mg/mL < C < 80 mg/mL ($R/2 < d_M <$ R)

In this concentration range, the distance between the IgG molecules is less than *R* but larger than *R*/2. In this range, the IgG molecules start to interact with neighboring IgG molecules as seen from the assessment of the peak intensity ratio I_{856}/I_{830}. The amount of the interfacial water hydrated with IgG molecules decreases as the concentration increases, as shown by the water Raman band. In other words, the interaction between the IgG molecules and the water has changed to interactions among the IgG molecules themselves. Therefore, in this range, short-range attractive interactions, such as the dipole–dipole interaction, begin to occur between the proteins with the displacement of the interfacial water around protein.

4-1-3. C > 80 mg/mL ($R/2 > d_M$)

At these concentrations, the distance between the IgG molecules is less than *R*/2. The IgG molecules are tightly packed and almost all of them interact with neighboring IgG molecules as demonstrated by the side chain interaction from the assessment of the peak intensity ratio I_{856}/I_{830}. There is almost no interfacial water because the water molecules are hydrated with IgG molecules. In addition, CH–π interactions start to

occur because the distance between the molecules is less than a few nanometers according to the Phe band analysis. Therefore, in addition to the dipole–dipole interaction between proteins, this second type of short-range attractive interaction begins. This state is the precursor state to aggregation.

In conclusion of the concentration dependence measurement, our results show that although in the tested range of concentrations (10–200 mg/mL), the secondary structure of IgG did not change, as the concentration changed, short-range attractive interactions between IgG molecules began to occur, which correlated with the amount of interfacial water present. This correlation between short-range interaction and hydrated water also suggests that pH and cosolute adjustment is highly effective in stabilizing proteins in high-concentration solution.

4-2. The application of the marker bands

On the basis of the knowledge discussed above, we performed additional experiments (i) pH effect, (ii) Salt effect and (iii) the storage stability test to demonstrate the practicality of the marker bands.

4-2-1. pH Effect

The band intensity ratio I_{856}/I_{830} of Tyr can be a marker for the evaluation of the side chain interactions for many kinds of protein molecules. IgG molecules, in particular, have a number of Tyr residues in their Fab region, suggesting that this band would be sensitive to the side chain interactions. To investigate the pH effect on the protein-protein interaction, the concentration dependence measurement at pH4.0 was carried out and compared to the one at pH6.0. The concentration dependence of I_{856}/I_{830}

of Tyr at pH6.0 and 4.0 are shown in Figs. 10a and b. The ratio I_{856}/I_{830} at pH6.0 starts to increase from 30 mg/mL to 80 mg/mL and the averaged inclination (*dR/dC*) of the ratio I_{856}/I_{830} with respect to the concentration is 4.6×10^{-3} as shown in Fig. 10a. On the other hand, as shown in Fig. 10b, the ratio I_{856}/I_{830} at pH4.0 starts to increase from less than 10 mg/mL steeply and get saturated at around 40 mg/mL and the averaged inclination *dR/dC* at pH4.0 is 2.2×10^{-2}. Of note is that the I_{856}/I_{830} gets saturated in the lower concentration than pH6.0 and the *dR/dC* at pH4.0 is larger than the one at pH6.0.

This result indicates that there is potentially an increased tendency to form aggregation at pH4.0, in other words, the cohesive strength will be relatively strong at pH4.0, since IgG molecules are positively charged at pH4.0[30] and their intramolecular interaction destabilizes. Thus, it could be expected that the molecules unfolds partially and the hydrophobic part of the molecule is exposed to the solvent. This change could lead to the oligomerization or aggregation. In contrast, at pH6.0, IgG molecules are more stable, thus the cohesive strength will be relatively weak. In this manner, the averaged inclination *dR/dC* can be a useful maker to assess the cohesive strength. (Fig. 10c)

Assuming that the concentration induced oligomerization from the monomer of the IgG molecule is the equilibrium reaction, this can be expressed shown below (eq8).

$$M \underset{\leftarrow}{\overset{K_N}{\rightarrow}} N , \tag{8}$$

where *M* is the monomer of the IgG molecule and *N* is the oligomer or aggregation. On the basis of the equilibrium reaction shown above, K_N is defined as the equilibrium constant of the monomer for high oligomer species. Although the K_N is intrinsically not the same as the averaged inclination *dR/dC* and the equilibrium equation needs the assumptions, the *dR/dC* may have some correlations with K_N.

4-2-2. Salt Effect

To investigate the salt effect on the protein-protein interaction and demonstrate the practicality of the ratio I_{856}/I_{830} of Tyr, the salt concentration dependence measurements were performed. The IgG solutions were 50 mg/mL at pH6.0 and the salt concentrations were 10, 50, 100, 500, 1000 mM. The ratio I_{856}/I_{830} of Tyr of each concentration is shown in Fig. 11. It is of interest that the ratio has a higher value when the salt concentration is less than 50 mM and more than 500 mM, while the ratio has a minimum at 100 mM. This result indicates that the side chain interactions works powerfully when the salt concentration is less than 50 mM and more than 500 mM. In other words, at 100 mM, the IgG molecules are more stable. This result implies that at lower concentrations, the electric repulsive interaction is dominant and it destabilizes the proteins while at higher concentrations, the salting-out effect induces the attractive protein-protein interaction. The previous experimental research supports this result.[30,31]

In this way, the ratio I_{856}/I_{830} of Tyr can be used to assess the protein stability and it was confirmed that at 100 mM, the salt of NaCl stabilized the protein effectively. This band will also be used for the optimization of the drug formulation.

4-2-3. Storage Stability Test

The band at 1004 cm^{-1} of Phe could be useful for the storage stability test, since its intensity increases when the molecules get in the pre-aggregation state. To demonstrate the practicality of this band, we performed the storage stability test (3 month in 4°C). In Fig. 12a, the spectra of 100 mg/mL IgG solution, just after the preparation (red) and after the storage stability test (blue) are shown respectively. The spectra are normalized

at 1240 cm^{-1}. Both solutions are transparent apparently. It is of interest that the band intensity at 1004 cm^{-1} of the one after 3 month in 4 °C is much larger. This indicates there are amount of the aggregation of the IgG solution after the storage stability test, although it is transparent. The band intensity ratio I_{1004}/I_{1240} of this sample are plotted in Fig. 12b. The averaged band intensity ratio I_{1004}/I_{1240} of the fresh sample in the range of normal state (10–90 mg/mL) and in the pre-aggregation state (100–200 mg/mL) shown in Fig. 5e are also plotted as references. The error bars in this Figure is 3 standard deviation of each data set. Compared to these two regions, the value of the sample after the storage stability test are 4 times larger than those fresh samples. In this manner, the band intensity ratio I_{1004}/I_{1240} will be a practical maker band to evaluate the aggregation.

As shown above, these application studies have shown the practical potential of Raman spectroscopy in highly concentrated antibody solution. Additional work involving more quantitative evaluation of the aggregation of antibodies will be carried out in the near future.

On the basis of these results of this application studies, it is also concluded that Raman spectroscopy is a valuable method for the study of highly concentrated solutions that can provide conformational and interaction information about each functional group, especially of the aromatic residue such as Phe, Tyr and Trp. While conventional colloidal approaches are powerful to describe the behavior of molecules simply, this is only the case when the concentration is relatively low and the molecules are monodisperse. As the concentration increases, molecules interact with each other, leading to oligomerization or aggregation. At such concentrations, conformational information or knowledge about the interactions of each functional group is important, and can be obtained by using Raman spectroscopy. Therefore, Raman spectroscopy is a

useful complementary method for the study of highly concentrated solutions.

5. Conclusions

We used Raman spectroscopy to perform concentration-dependent measurements of highly concentrated antibody solutions over a wide range of concentrations (10–200 mg/mL). Our analysis of the amide I band, I_{856}/I_{830} of Tyr, relative intensity of Phe at 1004 cm^{-1}, and the OH stretching region at around 3000 cm^{-1} together showed that at these concentrations, the secondary structure of the IgG molecules does not change; however, short-range attractive interactions around Tyr and Phe begin to occur as the distance between the IgG molecules decreases. Our analysis of the OH stretching region at around 3000 cm^{-1} further showed that these short-range attractive interactions correlate with the amount of hydrated water around the IgG molecules. Thus a conformation-based approach using Raman spectroscopy can provide valuable information especially regarding highly concentrated solutions, whereas conventional colloidal approaches work well for dilute solutions.

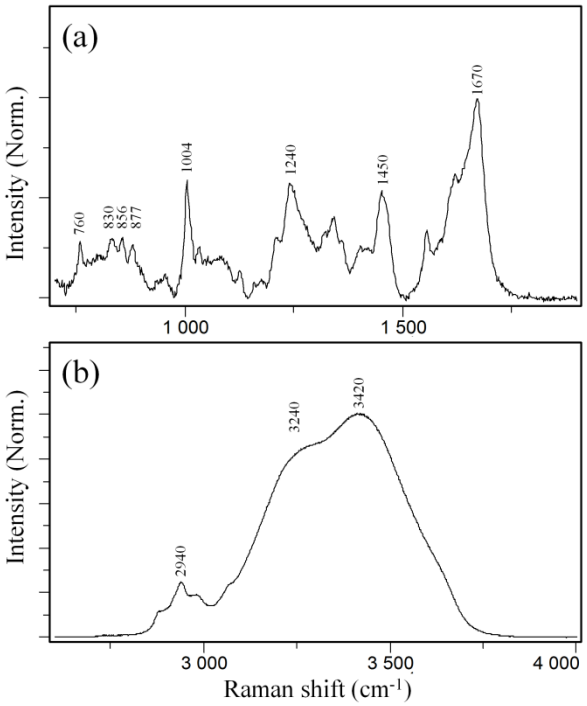

Figure 1. Raman spectrum of the 100 mg/mL IgG solution (a) in the fingerprint region and (b) in the CH and OH stretching regions.

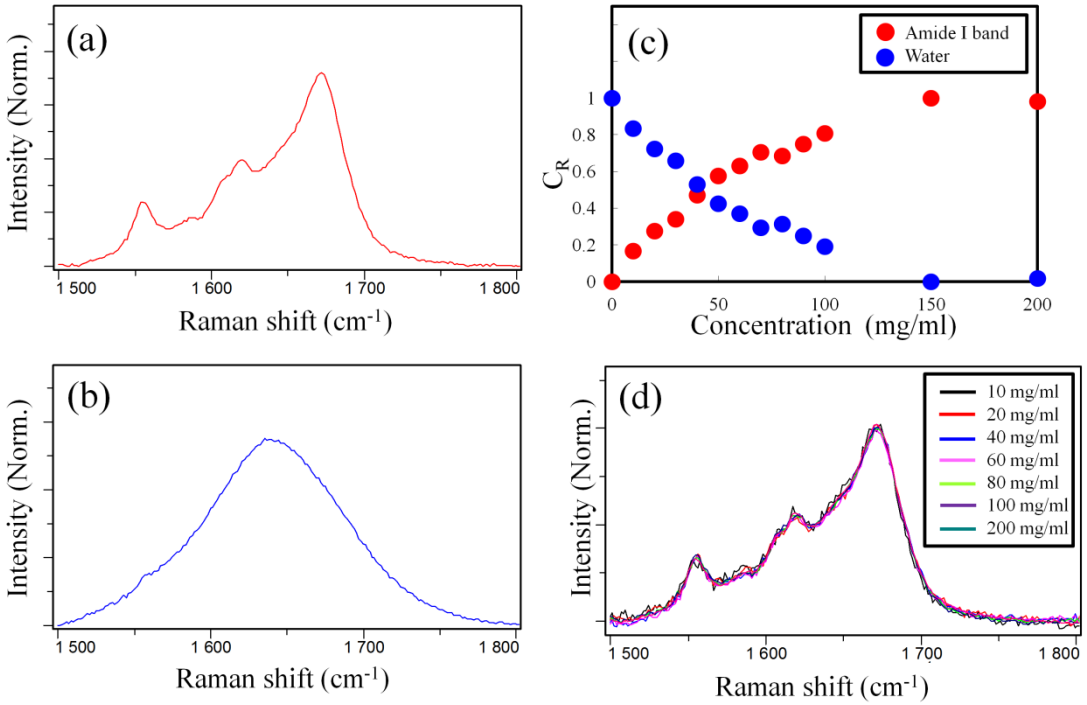

Figure 2. Abstract Raman spectra of (a) the amide I band of the IgG solution (red) and (b) water (blue) by MCR-ALS analysis. (c) The normalized concentration C_R of each component and (d) an overlay of the amide I band of the IgG molecules of each concentration (10, 20, 40, 60, 80, 100, 200 mg/mL).

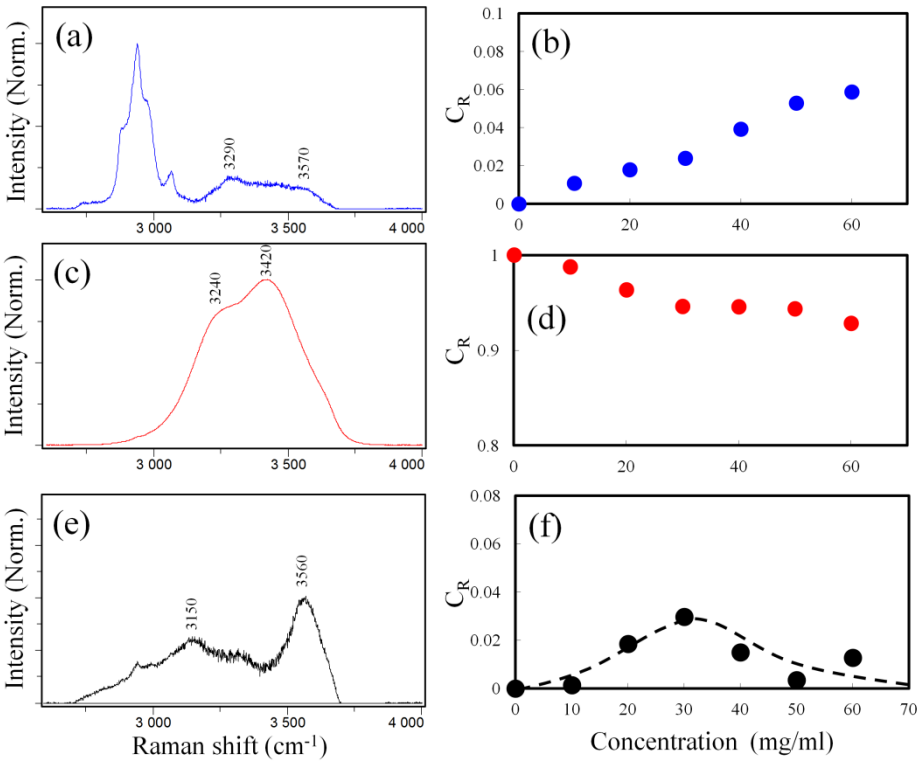

Figure 3. Abstract Raman spectra of each component (a,c,e) and the normalized concentration C_R of each component (b,d,f) at the lower concentrations (0–60 mg/mL).

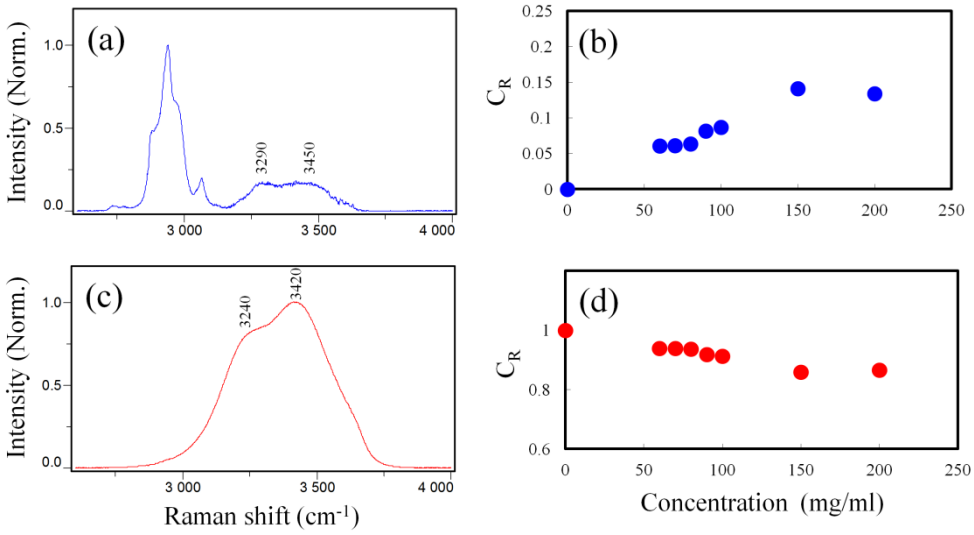

Figure 4. Abstract Raman spectra of each component (a,c) and the normalized concentration C_R of each component (b,d) at the higher concentrations (60–200 mg/mL).

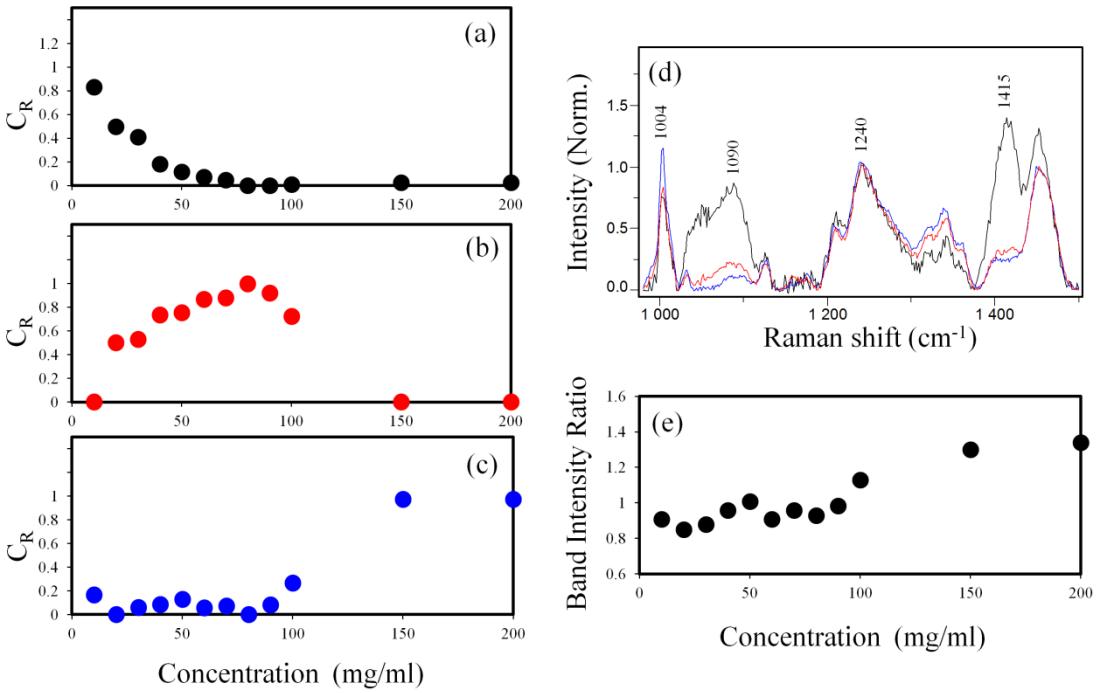

Figure 5. The normalized concentration C_R of each component (a,b,c) in the fingerprint region and (d) an overlay of the abstract Raman spectra of each component. (e) The concentration dependence of the band intensity ratio I_{1004}/I_{1240} of Phe.

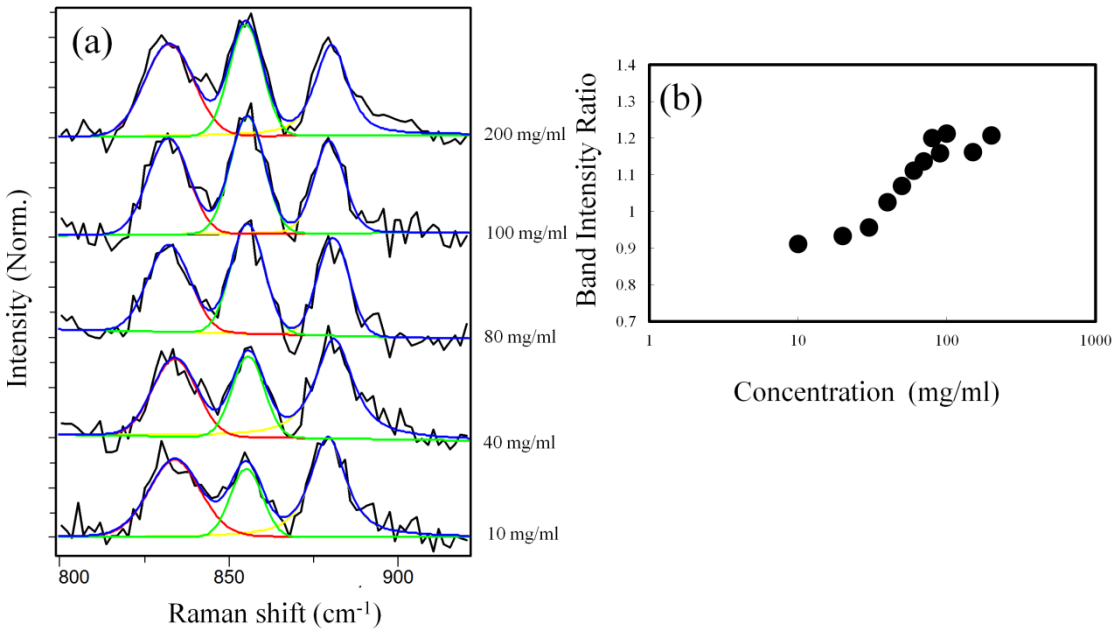

Figure 6. (a) Raman spectrum of IgG solutions in the range of 800 to 925 cm^{-1}. The concentrations are 10, 40, 80, 100, and 200 mg/mL, respectively. (b) The concentration dependence of the band intensity ratio I_{856}/I_{830}.

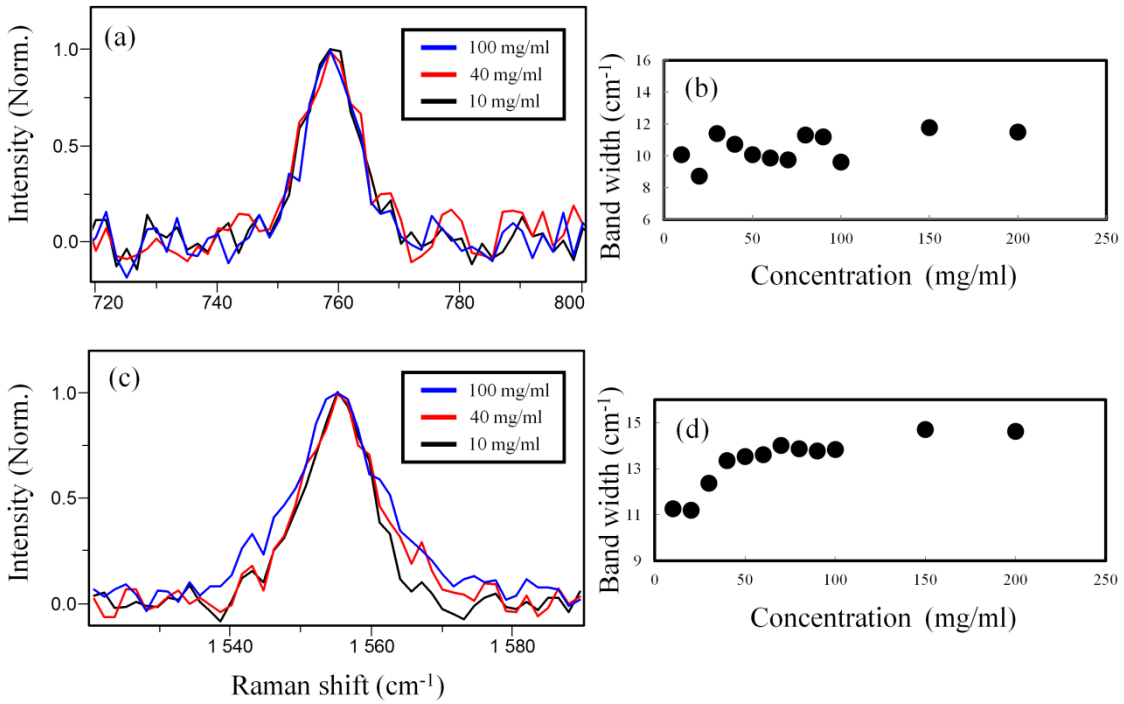

Figure 7. (a) An overlay of a Raman band of Trp at 760 cm^{-1} and (b) its bandwidth at each concentration (10–200 mg/mL). (c) An overlay of a Raman band of Trp at 1555 cm^{-1} and (d) its bandwidth at each concentration (10–200 mg/mL).

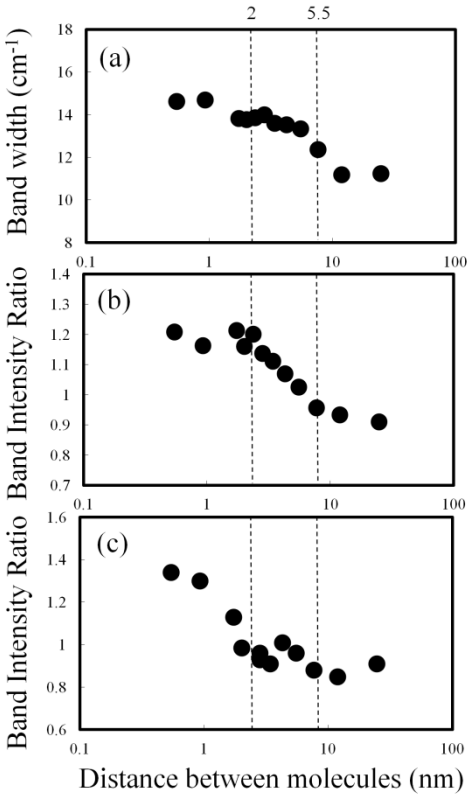

Figure 8. (a) Bandwidth of Trp at 1555 cm^{-1}, (b) band intensity ratio I_{856}/I_{837} of Tyr, and (c) band intensity ratio I_{1004}/I_{1240} of Phe as a function of the distance between molecules.

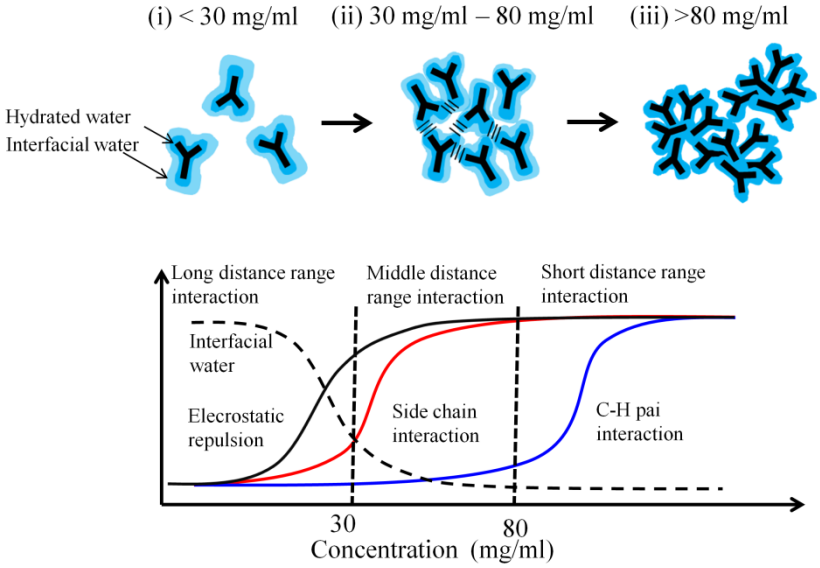

Figure 9. Schematic representation of the IgG solution at each concentration. (i) C < 30 mg/mL, (ii) 30 mg/mL < C < 80 mg/mL, (iii) C > 80 mg/mL. Relationship between the protein–protein interaction and the distance between molecules.

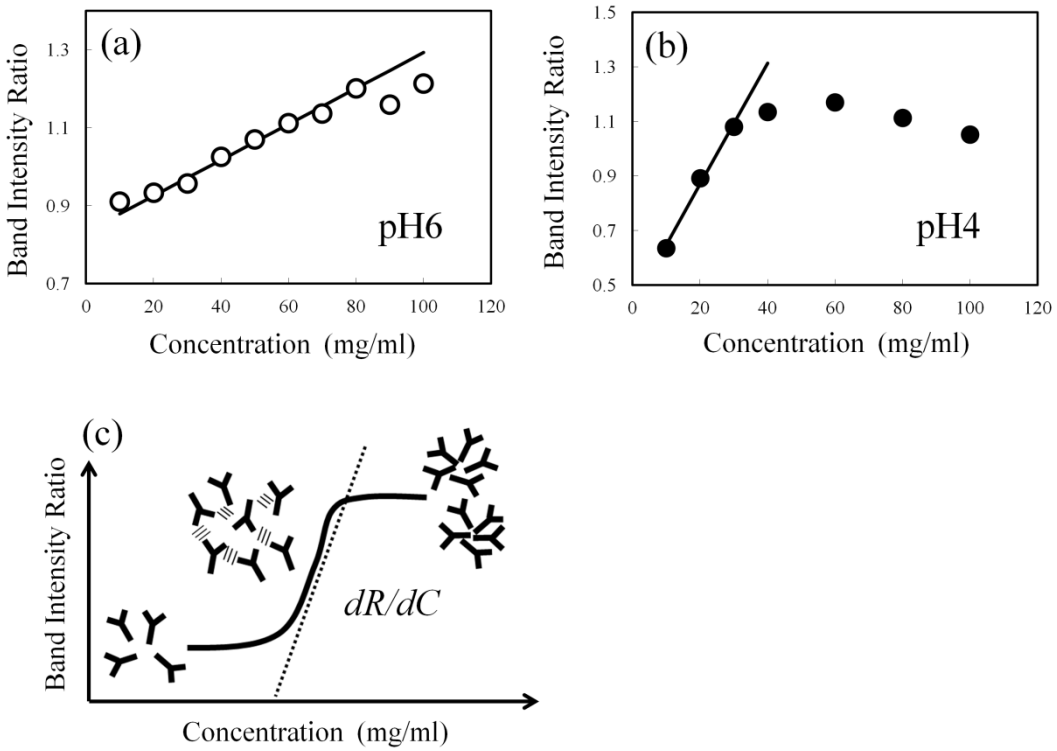

Figure 10. Band intensity ratio I_{856}/I_{830} of Tyr of the IgG solution at (a) pH6.0 and (b) pH4.0 as a function of the concentration. (c) Schematic representation of the concentration dependence of the IgG molecules and the definition of *dR/dC*.

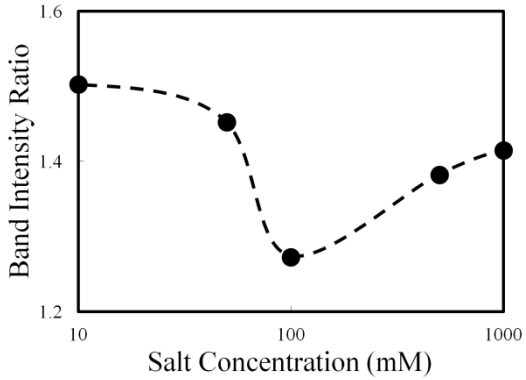

Figure 11. Salt concentration dependence of band intensity ratio I_{856}/I_{830} of Tyr of the IgG solution.

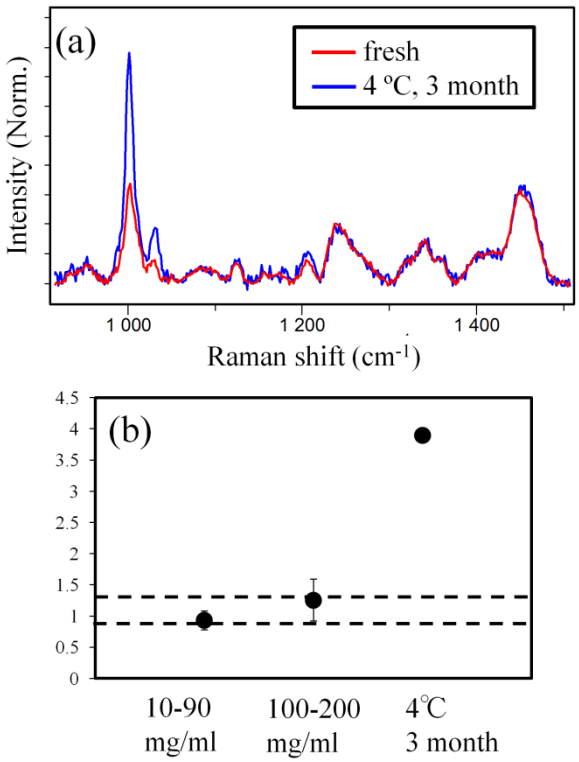

Figure 12. (a) Raman spectra of the 100 mg/mL IgG solution just after the preparation (red) and after the storage stability test (4 °C, for 3 months) (blue). (b) Band intensity ratio I_{1004}/I_{1240} of Phe of each concentration and after the storage stability test.

6. Supporting Information

The results of PCA (REV, RMSECV) are supplied as Supporting Information 1, 3, 5 (Figs. S1, S3, S5). Concentration dependence of Raman spectra in CH and OH stretching region (Fig. S2), the abstract spectra of MCR-ALS analysis (Fig. S4) and distance between the molecules of the IgG solution as a function of concentration (Fig. S6) are also supplied as Supporting Information 2, 4, 6.

6-1. Supporting Information 1

To estimate the number of the chemical species, REV and RMSECV in the amide I band region were calculated. In Fig. S1a, the REV is shown. Given that the changes in eigenvalue against factor level were drastic, we plotted the ordinate axis using a

logarithmic scale. As shown in the figure, as the number of the factor increases, the REV score decreases. A small REV score indicates that the factor is from a minute species or is noise. After the second loading, the eigenvalue decreases rapidly and the loading value is almost the same as the noise level. This result indicates that the factors after the second factor can be attributed to random noise. The RMSECV score was also plotted in Fig. S1b. As shown in the figure, after the second loading, the RMSECV score is almost constant. Both REV and RMSECV, thus show that there are two chemical species in this concentration data set of the amide I band.

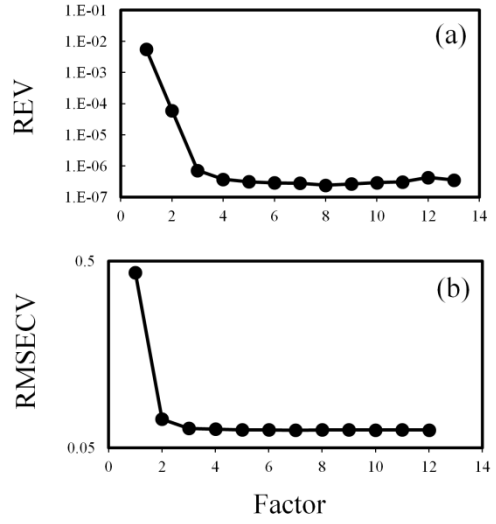

Figure S1. Plots of (a) reduced eigenvalue (REV) and (b) root-mean-square prediction error of cross-validation (RMSECV) against the factor level of the PCA in the amide I band region.

6-2. Supporting Information 2

We subjected the data set of the concentration dependence spectra in the CH and OH stretching regions (0 mg/mL (buffer solution) and 10 to 200 mg/mL) to PCA. The Raman spectrum in the CH and OH stretching regions of the IgG solution at 10, 40, 80, 100, and 200 mg/mL are shown in Fig. S2. As the IgG concentration increases, the intensity of the CH stretching bands (2940 cm^{-1}) also increases.

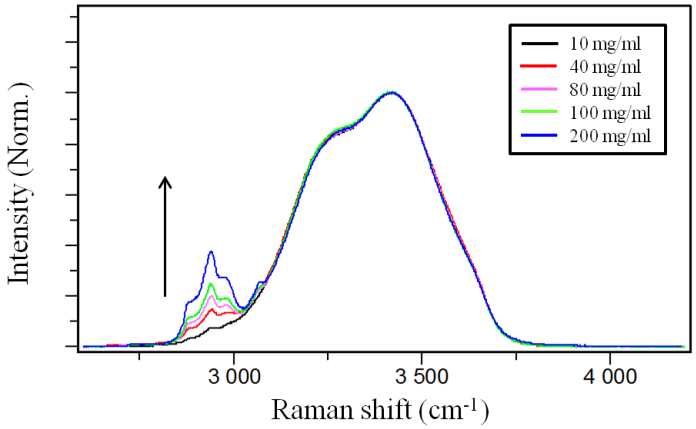

Figure S2. Raman spectrum in the CH and OH stretching regions of the IgG solution and the buffer solution. The concentrations are 10, 40, 80, 10, 50, 100, and 200 mg/mL, respectively.

6-3. Supporting Information 3

For the estimation of the number of the chemical species, REV and RMSECV of the data (0–200 mg/mL) in the CH and OH stretching regions were calculated as shown in Figs. S3a and S3b. Both the REV and RMSECV for the first and second factors are more than two orders of magnitude larger than that after the second factor, indicating that there are two main species in these plots. In addition, there appears to be four minute species, suggesting that highly concentrated solutions can contain various minute species.

6-3-1. Lower Concentrations (0–60 mg/mL)

For the lower concentrations, the data set included 10–60 mg/mL and the buffer solution. The results of the REV and RMSECV are shown in Figs. S3c and S3d. The REV after the third factor is almost constant, suggesting that the factors after the third factor can be attributed to random noise. Of note, although the third factor is small, it may represent a minute species. The RMSECV results are similar to the REV results, with two main species represented in the RMSECV plots. The RMSECV after the third factor does not improve the value, again indicating that the factors after the third factor can be attributed to random noise. Taken together, these results show that there are three species in the data set at these concentrations: two main species and one minute one.

6-3-2. Higher Concentrations (60–200 mg/mL)

For the higher concentrations, the data set included 60–200 mg/mL and the

buffer solution. The results of the REV and RMSECV are shown in Figs. S3e and S3f. The REV after the second factor is almost constant, suggesting that factors after the second factor can be attributed to noise components. The RMSECV shows similar results to those of REV. The RMSECV after the second factor is two orders of magnitude smaller than after the second factor, again indicating that factors after the second factor can be attributed to noise. Taken together, these results show that there are two main species at these concentrations.

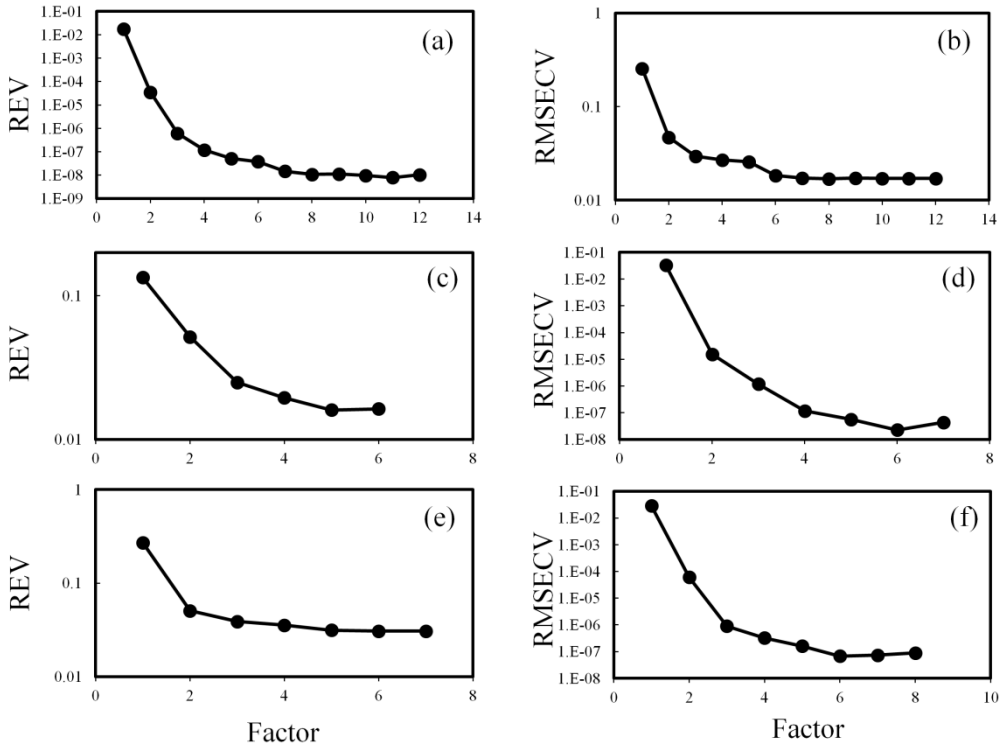

Figure S3. Plots of (a,c,e) REV and (b,d,f) RMSECV against the factor level of the PCA in the CH and OH stretching regions. Concentrations: (a,b) 10–200 mg/mL; (c,d) 10–60 mg/mL; (e,f) 60–200 mg/mL.

6-4. Supporting Information 4

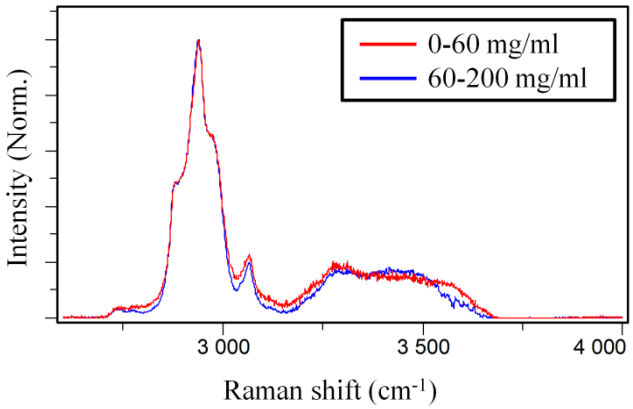

Figure S4. An overlay of abstract Raman spectra of a component of the protein with hydrated water at the lower (10–60 mg/mL) and the higher concentrations (60–200 mg/mL).

6-5. Supporting Information 5

The results of the REV and RMSECV in the fingerprint region were plotted by using PCA (Figs. S5a, b). Since the REVs for the first through the third factor are more than one order of magnitude larger than the REV for the fourth factor, there are three main species in the data set. In addition, the REV after the third factor is almost constant, indicating that factors after the third factor can be attributed to random noise. The RMSECV shows a sharper result than the REV. There is a break at the third factor in the RMSECV plots, again suggesting that there are three main species in the data set. Thus, both results show that there are three main species at these concentrations.

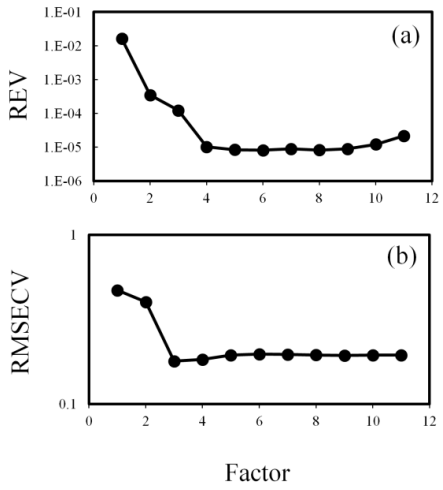

Figure S5. Plots of (a) REV and (b) RMSECV against the factor level of the PCA in the fingerprint region.

6-6. Supporting Information 6

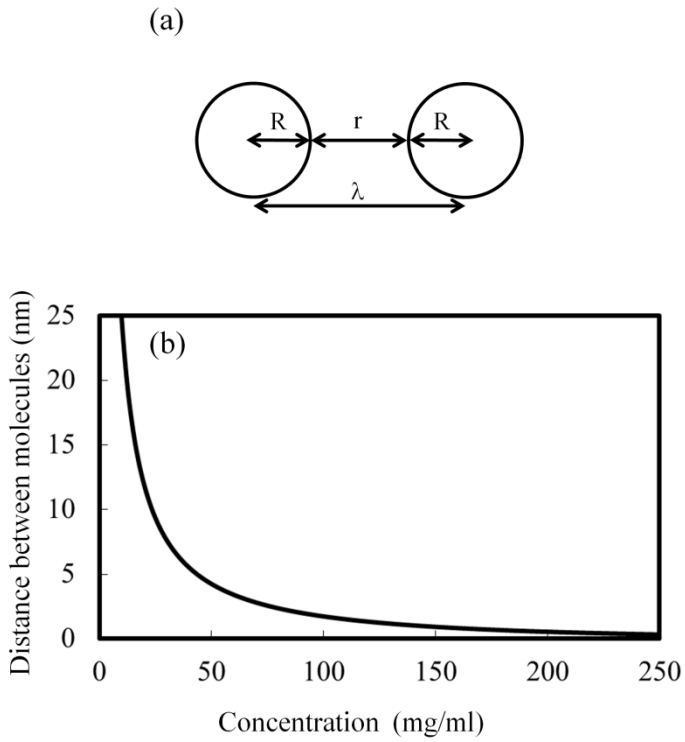

Figure S6. (a) Schematic representation of the model of hard spherical particles for the calculation of the mean nearest-neighbor distance. (b) Distance between the molecules of the IgG solution as a function of concentration.

7. References

1. Uchiyama, S. Liquid formulation for antibody drugs. *Biochim. Biophys. Acta*, **2014**, *1844*, 2041-2052.

2. Yadav, S., Liu, J., Shire, S. J., Kalonia, D. S. Specific interactions in high concentration antibody solutions resulting in high viscosity. *J. Pharm. Sci*. **2010**, *99*, 1152-1168.

3. Stradner, A., Sedgwick, H., Cardinaux, F., Poon, W. C. K., Egelhaaf, S. U., Schurtenberger, P. Equilibrium cluster formation in concentrated protein solutions and colloids. *Nature*, **2004**, *432*, 492-495.

4. Wen, Z. Raman spectroscopy of protein pharmaceuticals. *J. Pharm. Sci*. **2007**, *96*, 2861-2871.

5. Oladepo, S. A., Xiong, K., Hong, Z., Asher, S. A., Handen, J., Lednev, I. K. UV resonance Raman investigations of peptide and protein structure and dynamics. *Chem. Rev*. **2012**, *112*, 2604-2628.

6. Ota, C., Noguchi, S., Tsumoto, K. The molecular interaction of a protein in highly concentrated solution investigated by Raman spectroscopy. *Biopolymers*, **2015**, *103*, 237-246.

7. Malinowski, E. R. *Factor Analysis in Chemistry* (2nd ed.); Wiley-Interscience: New York, 1991.

8. Lobo, A. P., Valles, B. S., Tascón, N. F., Madrera, R. R. Calibration models for routine analysis of cider by mid-infrared spectroscopy. *Lebensm. Wiss. Technol.* **2006**, *39*, 1026-1032.

9. Tauler, R. Application of non-linear optimization methods to the estimation of multivariate curve resolution solutions and of their feasible band boundaries in the investigation of two chemical and environmental simulated data sets. *Anal. Chim. Acta*, **2007**, *595*, 289-298.

10. Maiti, N. C., Apetri, M. M., Zagorski, M. G., Carey, P. R., Anderson, V. E. Raman spectroscopic characterization of secondary structure in natively unfolded proteins: α-Synuclein. *J. Am. Chem. Soc*. **2004**, *126*, 2399-2408.

11. Apetri, M. M., Maiti, N. C., Zagorski, M. G., Carey, P. R., Anderson, V. E. Secondary structure of α-synuclein oligomers: characterization by Raman and atomic force microscopy. *J. Mol. Biol*. **2006**, *355*, 63-71.

12. Vinaykin, M., Benderskii, A. V., Vibrational sum-frequency spectrum of the water bend at the air/water interface. *J. Phys. Chem. Lett*. **2012**, *3*, 3348-3352.

13. Chen, H., Lin, H., Chen, H., Mai, F., Liu, Y., Lin, C., Chang, C., Tsai, H., Yang, C.

Innovative strategy with potential to increase hemodialysis efficiency and safety. *Sci. Rep*. **2014**, *4*, 1-5.

14. Shen, Y. R., Ostroverkhov, V. Sum-frequency vibrational spectroscopy on water interfaces: polar orientation of water molecules at interfaces. *Chem. Rev*. **2006**, *106*, 1140-1154.

15. Davis, J. G., Gierszal, K. P., Wang, P., Ben-Amotz, D. Water structural transformation at molecular hydrophobic interfaces. *Nature*, **2012**, *491*, 582-585.

16. Englert, L., Biela, A., Zayed, M., Heine, A., Hangauer, D., Klebe, G. Displacement of disordered water molecules from hydrophobic pocket creates enthalpic signature: Binding of phosphonamidate to the S1'-pocket of thermolysin. *Biochim. Biophys. Acta*, **2010**,*1800*, 1192-1202.

17. Phillip, W. S., Matthew, R. L., Demetri, T. M., George, M. W. Is it the shape of the cavity, or the shape of the water in the cavity? *Eur. Phys. J. Special Topics*, **2014**, *223*, 853-891.

18. Lord, R. C., Yu, N. Laser excited Raman spectroscopy of biomolecules : I. Native lysozyme and its constituent amino acids. *J. Mol. Biol.* **1970**, *50*, 509-524.

19. Xu, M., Ermolenkov, V. V., He, W., Uversky, V. N., Fredriksen, L., Lednev, I. K. Lysozyme fibrillation: deep uv Raman spectroscopic characterization of protein structural transformation. *Biopolymers*, **2005**, *79*, 58-61.

20. Xu, M., Shashilov, V. A., Ermolenkov, V. V., Fredriksen, L., Zagorevski, D., Lednev, I. K., Hen egg white lysozyme fibrillation_a deep UV resonance Raman spectroscopic study. J. *Biophotonics*, **2008**, *1*, 215-229.

21. Gelden, J. D., Gussem, K. D., Vandenabeele, P., Moens, L. Reference database of Raman spectra of biological molecules. *J. Raman Spectrosc.* **2007**, *38*, 1133-1147.

22. Sarbon, N. M., Badii, F., Howell, N. K. Preparation and characterization of chicken skin gelatin as an alternative to mammalian gelatin. *Food Hydrocolloids*, **2013**, *30*, 143-151.

23. McCammon, J. A., Wolynes, P. G., Karplus, M. Picosecond dynamics of tyrosine side chains in proteins. *Biochemistry*, **1979**, *18*, 927-942.

24. Brandl, M., Weiss, M. S., Jabs, A., Suhnel, J., Hilgenfeld, R. C-H π-interactions in proteins. *J. Mol. Biol*. **2001**, *307*, 357-377.

25. Siamwiza, M. N., Lord, R. C., Chen, M. C., Takamatsu, T., Harada, I., Matsuura, H., Shimanouchi, T. Interpretation of the doublet at 850 and 830cm^{-1} in the Raman spectra of Tyrosyl residues in proteins and certain model compounds. *Biochemistry*, **1975**, *14*, 4870-4876.

26. Chari, R., Jerath, K., Badkar, A. V., Kalonia, D. S. Long and short range electrostatic interactions affect the rheology of highly concentrated antibody solutions. *Pharm. Res*. **2009**, *26*, 2607-2618.

27. Takeuchi, H., Harada, I. Normal coordinate analysis of the indole ring. *Spectrochim. Acta*, **1986**, *142A*, 1069-1078.

28. Miura, T., Takeuchi, H., Harada, I. Characterization of individual tryptophan side chains in proteins using Raman spectroscopy and hydrogen deuterium exchange kinetics. *Biochemistry*, **1988**, *27*, 88-94.

29. Scherer, T. M., Liu, J., Shire, S. J., Minton, A. P. Intermolecular interactions of IgG_1 monoclonal antibodies at high concentrations characterized by light scattering. *J. Phys. Chem. B*, **2010**, *114*, 12948-12957.

30. Saito, S., Hasegawa, J., Kobayashi, N., Tomitsuka, T., Uchiyama, S., Fukui, K. Effects of ionic strength and sugars on the aggregation propensity of monoclonal antibodies: influence of colloidal and conformational stabilities. *Pharm. Res*. **2013**, *30*, 1263-1280.

31. Thomas, M. S. Role of cosolute−protein interactions in the dissociation of monoclonal antibody clusters, *J. Phys. Chem. B*, **2015**, *119*, 13027-13038.

Chapter 5. Conclusions

The present study aims at establishing a methodology for the understanding of the protein behavior in the crowding / highly concentrated solution by using Raman spectroscopy and describing their behavior in detail. In addition, using these established methods, the practical application and its potential for the evaluation of the protein-protein interaction in the highly concentrated antibody solution have been shown.

In Chapter 2, the structure and interactions of lysozyme molecules over a wide range of concentrations (2.5–300 mg/mL) have been investigated by using Raman spectroscopy. In this concentration range, the amide I band was not affected by concentration, but the width of the Trp band at 1555 cm^{-1}, the I_{856}/I_{837} ratio of Tyr, the I_{870}/I_{877} ratio of Trp, and the I_{2940}/I_{3420} ratio changed as the concentration was increased. These results indicate that although the distance between molecules changed by more than an order of magnitude over the tested concentration range, the secondary structure of lysozyme did not change. In contrast, interactions between the molecules did change in a stepwise process as the order of magnitude of the distance between the molecules changed. These results also indicate that fundamental marker bands can be used to evaluate high-concentration solutions of protein and that the use of Raman spectroscopy can be expected to lead to progress in our understanding not only of the basic science of protein behavior under concentrated (*i.e.*, crowded) conditions but also of practical processes involving proteins, such as in the field of biopharmaceuticals.

In Chapter 3, using the established method in Chapter 2, the concentration-dependent measurements of BSA solutions at pH7.0 and 3.0 by Raman spectroscopy has been carried out to discuss the entropic effect and the soft chemical interaction under these different conditions as an advanced study. Raman spectroscopic data have revealed that the entropic effect, soft chemical interactions, and hydration effect are tightly balanced in crowded environments. In addition, it has been found that the “hard” molecule is not sensitive to the crowding effect such as the excluded volume effect, while the “soft” molecule is sensitive to the one.

In Chapter 4, on the basis of the knowledge of the study of a model system of lysozyme and BSA in Chapters 2 and 3, the protein-protein interactions has been discussed in a highly concentrated antibody solution as a practical system and the application studies have been performed to show the practical use. In addition, in these studies, the importance of the conformational information or knowledge about the interactions of each functional group has been shown and these knowledge can be obtained by using Raman spectroscopy. Therefore, it has been concluded that Raman spectroscopy is a useful complementary method for the study of highly concentrated antibody solutions.

In this thesis, the methodology for the understanding of the protein behavior in the crowding / highly concentrated solution has been established by using Raman spectroscopy and the detail behavior of proteins in the crowding environment has been described in the model and practical system. These knowledge and methodology will likely lead to progress in understanding and describing the molecules in more real and complex molecular crowding environments in the near future.

List of Publications

Chapter 2.

C. Ota, S. Noguchi, K. Tsumoto, The Molecular Interaction of a Protein in Highly Concentrated Solution Investigated by Raman Spectroscopy. *Biopolymers* **2015**, 103, 237-246.

Chapter 3.

C. Ota, K. Takano, Behavior of Bovine Serum Albumin Molecules in Molecular Crowding Environments Investigated by Raman Spectroscopy. *Langmuir*, **2016**, 32, 7372-7382.

Chapter 4.

C. Ota, S. Noguchi, S. Nagatoishi, K. Tsumoto, Assessment of the Protein–Protein Interactions in a Highly Concentrated Antibody Solution by Using Raman Spectroscopy. *Pharm. Res.* **2016**, 33, 956-969.

Acknowledgements

I would like to highly express my gratitude to my supervisor, Professor Kazufumi Takano, Department of Biomolecular Chemistry, Kyoto Prefectural University, for his constructive advice about my studies and for his kind help in revising the published papers as well as this thesis.

I am grateful to another supervisor, Professor Kouhei Tsumoto, Department of Bioengineering, School of Engineering and Laboratory of Medical Proteomics, Institute of Medical Science, the University of Tokyo, for his excellent comments and kind advice on my studies.

I greatly thank Dr. Satoru Nagatoishi, Department of Bioengineering, School of Engineering, the University of Tokyo, for his valuable comments and kind help on my studies.

I express my sincere thanks to my colleague, Mr. Shintarto Noguchi, Advanced R&D Center, Horiba, Ltd., for his kind help and advice on my study and useful discussions.