Journal of
Applied Crystallography
ISSN 0021-8898

Received 16 August 2006
Accepted 26 February 2007

Small-angle light and X-ray scattering measurements of a protein–oligosaccharide complex mucin in solution

Yasushi Watanabe[a]* and Yoji Inoko[b]

[a]National Food Research Institute, 2-1-12 Kannondai, Tsukuba, Ibaraki 305-8642, Japan, and [b]Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan. Correspondence e-mail: yasuw@affrc.go.jp

The molecular assembly and the chain conformation of intact bovine submaxillary mucin (BSM) in solution over wide-ranging concentrations were characterized by using low-angle laser light scattering and small-angle X-ray scattering (SAXS) methods. The specific refractive index increment of BSM was estimated to be 0.152 ml g^{-1} and was used to determine the molecular weight of BSM by low-angle laser light scattering photometry combined with high-performance gel chromatography. The total molecular weight of BSM was 55 million and the molecular weight of the main fractionated components was about 2 million. Fractal analysis of the SAXS data revealed that the intact BSM molecule is a chain with excluded volume (fractal dimension 1.67) at concentrations of 1.4 and 3.6 mg ml^{-1} and a chain with a Gaussian chain character (fractal dimension 2) at concentrations of 7.2–15 mg ml^{-1}. Moreover, the Kratky plots of the SAXS data showed that the chain conformation of BSM molecules is a Gaussian (unfolded) structure in solution. The estimated cross-sectional radius of gyration value lay in the range 0.65–0.76 nm, which is reasonable for a long thin shape.

1. Introduction

Mucins are the glycoproteins in epithelial mucous secretions that coat the oral cavity and the respiratory, gastrointestinal and urogenital tracts of animals; their protective function is due to their solution properties that are related to the molecular structure (Carraway & Fregien, 1995). It is important to clarify the molecular assembly and chain conformation of the conjugates in solution in order to understand the physiological functions. The intact mucin glycoprotein is very large, over 1 million g mol^{-1} (Sheehan & Carlstedt, 1984; Shogren *et al.*, 1986; Malmsten *et al.*, 1992; Shi & Caldwell, 2000). Mucins contain a large number of oligosaccharide side chains covalently bound to a protein backbone. The macromolecules typically contain 60–90 wt% carbohydrate. The oligosaccharides may contain about 20 monosaccharides. The typical structure of a mucin molecule consists of several subunits connected by disulfide bridges as shown in Fig. 1. Since mucins have very high molecular weights and high carbohydrate contents, it is very difficult to characterize their structures by gel electrophoresis or ultracentrifugation by sedimentation equilibrium. In this study, in order to elucidate further the solution structure of bovine submaxillary mucin (BSM) molecules, the molecular assembly and the chain conformation of intact BSM in solution were characterized by using low-angle laser light scattering (LALLS) and small-angle X-ray scattering (SAXS) methods.

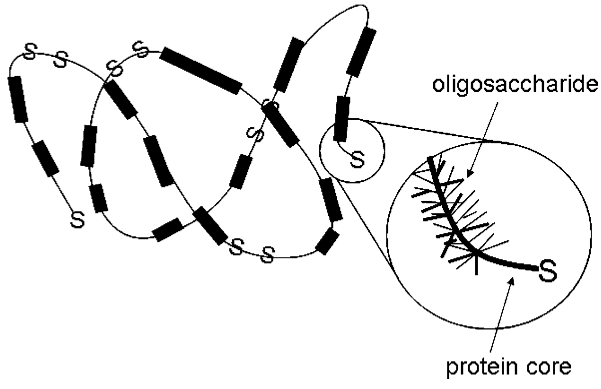

Figure 1
A schematic drawing of a mucin molecule. Several subunits bind with terminal cysteine-rich domains (S) by disulfide bridges (S–S). Thick bars represent the highly glycosylated domains.

2. Materials and methods

2.1. Materials

Purified BSM was purchased as type I-S from Sigma Aldrich Inc., St Louis, MO, USA. All other chemicals were of analytical grade.

2.2. LALLS measurements

The system of LALLS photometry consists of a high-performance gel chromatography (HPGC) pump system equipped with a sample loop having an internal volume of 100 µl, HPGC columns and two sequential detectors, a low-angle laser light scattering photometer (TSK LS-8000, Tosoh, Tokyo), and a differential refractometer (TSK RI-8000, Tosoh). The HPGC column was a TSK-GEL GCSWXL column (40 × 6 mm internal diameter, Tosoh) to estimate the total weight average molecular weight of BSM because the BSM molecules

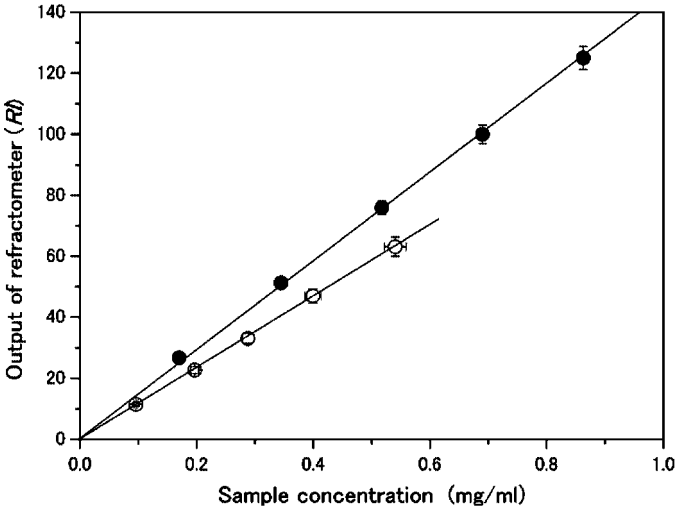

Figure 2
Linear plots of the outputs of refractive index *versus* sample concentrations. Closed and open circles show the data points of BSA and BSM, respectively. The error bars of some data points are within the symbols.

were eluted as a single peak in this short column, while a GF-7M HQ HPGC column (300 × 7.5 mm internal diameter) equipped with a GF-1G 7B pre-column (50 × 7.5 mm internal diameter, Shodex, Tokyo) was used for the estimation of the molecular weight distribution of BSM. The temperature of the columns and the flow-through cell in the low-angle laser light scattering photometer was kept at 297 K using a column jacket and a metallic cell holder through which constant-temperature water was circulated. The eluent was a 50 m*M* sodium phosphate buffer solution, pH 6.8. The flow rate was 0.3 ml min^{-1} in all experiments. Data handling has been described previously in connection with the molecular weight determination of proteins (Takagi, 1990; Watanabe & Takagi, 1993; Watanabe, 2005). The molecular weight of the protein, *Mw*, was obtained on the basis of

$$Mw = k_1(\partial n/\partial c)^{-1}(LS)(RI)^{-1}, \quad (1)$$

where $\partial n/\partial c$ is the specific refractive index increment expressed in terms of the weight concentration of the protein; k_1 is a constant, which depends on the experimental conditions; and (*LS*) and (*RI*) are the outputs of the low-angle laser light scattering photometer and the differential refractometer, respectively. The constant, k_1, was determined using bovine serum albumin (BSA) of known molecular weight and specific refractive index [*Mw* = 66 300 (monomer), $\partial n/\partial c$ = 0.187 ml g^{-1} (Fasman, 1976)]. The concentration of BSA was estimated spectrophotometrically assuming that the absorbance value of a 1% solution at 280 nm was 6.78 in a 1 cm light-path cuvette (Fasman, 1976). The weight concentration of BSM was determined by a dry weight method (Nozaki, 1986).

2.3. SAXS experiments

Synchrotron-radiation SAXS measurements were performed with an optics system at the beamline BL-10C station at the Photon Factory of the High Energy Accelerator Research Organization as previously described (Ueki *et al.*, 1983; Inoko, 1999; Watanabe *et al.*, 1999). A wavelength (λ) of 0.1488 nm was used. The temperature of the cell with a 1 mm light path and a pair of 15 µm quartz windows was kept constant at 297 K using the metallic cell holder through which constant-temperature water was circulated. Data were collected for 10 min with a position-sensitive proportional counter at a sample-to-detector distance of 900 mm. The signals obtained were corrected for solvent scattering and normalized to the beam intensity to yield the net scattering intensity $I(q)$, where q [= $(4\pi/\lambda)\sin\theta$, where 2θ is the scattering angle] is the modulus of the scattering vector. The q value was calibrated using a diffraction pattern of dried chicken collagen.

3. Results and discussion

3.1. Molecular assemblies determined by using the LALLS system

The molecular assembly of BSM was estimated through molecular weight determination using light scattering methods in this study. Since molecular weight depends on the value of the specific refractive index increment ($\partial n/\partial c$) of the sample protein [see equation (1)], the $\partial n/\partial c$ value must be measured independently using a precision differential refractometer. Therefore, the value for BSM was estimated under the present sample conditions (50 m*M* sodium phosphate buffer, pH 6.8, at 297 K). The output of the refractometer (*RI*) can be expressed as

$$(RI) = k_2(\partial n/\partial c)c, \quad (2)$$

where k_2 is a proportional constant and c is the weight concentration of the sample protein. The constant, k_2, was determined using a protein of known specific refractive index such as BSA ($\partial n/\partial c$ = 0.187 ml g^{-1}; Fasman, 1976). Fig. 2 gives the plots of the (*RI*) value *versus* the sample concentration for BSA and BSM, showing a good linear relationship which is a necessity for a reliable $\partial n/\partial c$ calculation. The slope of the linear plots corresponds to the $\partial n/\partial c$ value. Thus, the $\partial n/\partial c$ value of BSM can be estimated to be 0.152 ± 0.01 ml g^{-1} (n = 5). Using this $\partial n/\partial c$ value and equation (1), the estimation of the apparent total weight average molecular weight of BSM was made for solutions ranging in concentration from 0.1 to 0.54 mg ml^{-1}. Fig. 3 shows the sample-concentration dependence of the apparent weight average molecular weight of BSM. The molecular weight at zero sample concentration was calculated as 55 ± 3 million (10^6). To estimate the molecular weight distribution of BSM in solution, an HPGC/ LALLS method was applied next. Fig. 4 shows typical chromatograms of (*LS*) and (*RI*) and the molecular weight distri-

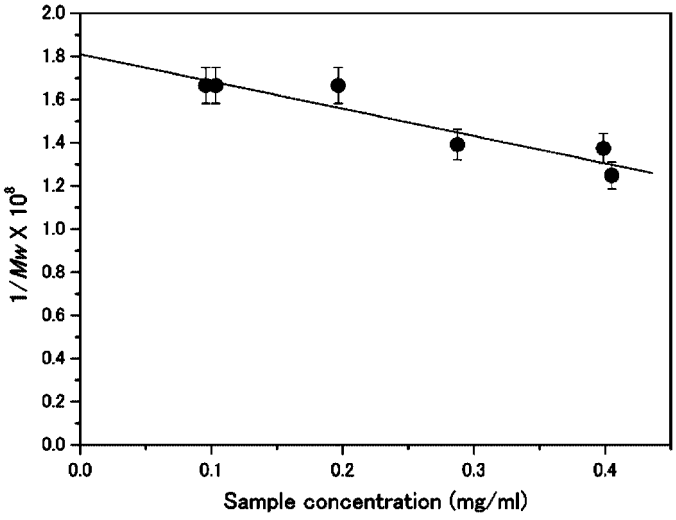

Figure 3
Plots of the reciprocal of the apparent molecular weight against the sample concentration of BSM. The line was obtained by least-squares analysis.

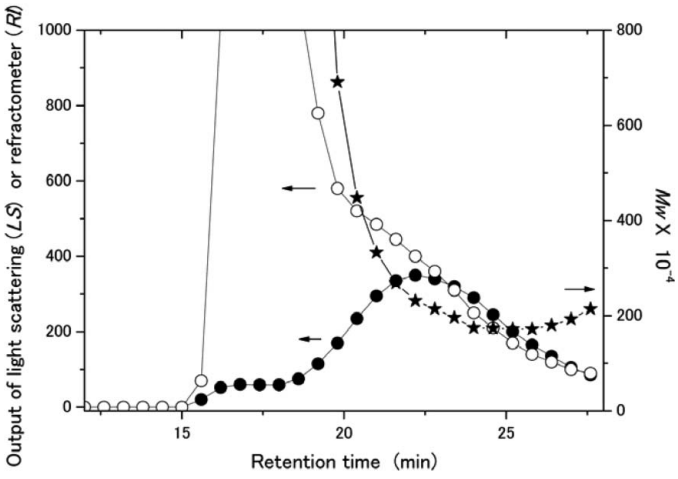

Figure 4
Typical chromatograms of (*LS*) (open circles) and (*RI*) (filled circles) and the molecular weight (filled stars) distribution of BSM. The concentration of the sample applied to the HPGC columns was 2 mg ml^{-1}.

bution of BSM. A small peak was observed at 15–20 min in the (*RI*) chromatography curve. The molecular weight value of this minor component was over several million. The main components of BSM were eluted at the peak top of about 22 min in the (*RI*) chromatograph, and the molecular weight was about 2 million under the present physiological conditions.

The literature $\partial n/\partial c$ values were reported to be 0.105 ml g^{-1} (BSM, 0.175 *M* Tris–HCl buffer, pH 7.4) (Shi & Caldwell, 2000), 0.16 ml g^{-1} (human cervical mucin, 0.2 *M* NaCl, 1 m*M* Na_2EDTA, 1 m*M* sodium phosphate buffer, pH 6.5) (Sheehan & Carlstedt, 1984) and 0.135 ml g^{-1} (porcine submaxillary mucin, 6 *M* guanidine hydrochloride) (Shogren *et al.*, 1986). The discrepancies will depend on the sample purity, carbohydrate composition in the samples and experimental conditions. Therefore, the $\partial n/\partial c$ value was estimated under the same conditions as that of the light scattering experiments in this study. The total molecular weights of intact BSM have been reported to be 1.6 million in 0.175 *M* Tris–HCl buffer, pH 7.4 (Shi & Caldwell, 2000) and 4.0 million in water and aqueous NaCl solutions (Bettelheim *et al.*, 1962). On the other hand, total molecular weights of human cervical, rat gastric, porcine gastric and porcine submaxillary mucins were reported to be 10, 10, 15 and 17 million, respectively (Sheehan & Carlstedt, 1984; Shogren *et al.*, 1986; Malmsten *et al.*, 1992). The total molecular weight of BSM estimated in the present study is larger than the previous values for BSM and other mucins. Since the molecular weight of the subunit of BSM has been reported to be 0.4 million in 0.2 *M* NaCl (Tettamanti & Pigman, 1968), the total molecular weights will correspond to the molecular assemblies of the subunits of BSM. Moreover, for the same reason, it can be considered that the main components of intact BSM molecules will also be the oligomers with molecular weight of about 2 million.

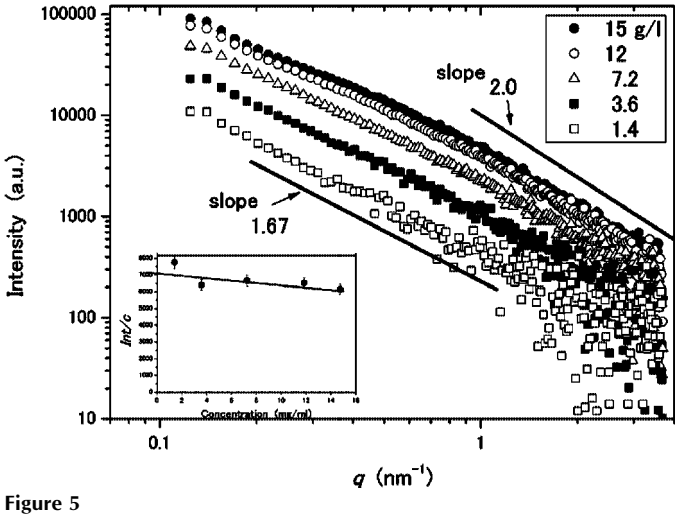

Figure 5
SAXS patterns at different sample concentrations. The sample concentrations increase from the bottom (1.4 mg ml^{-1}) to the top (15 mg ml^{-1}) as shown in the figure. The sample-concentration dependence of the intensity (at q = 0.12 nm^{-1}) of the SAXS divided by the concentration c is presented in the inset figure.

3.2. Chain conformation characterized by SAXS experiments

In this study, since we have focused our attention on the structural characterization of the intact mucin in solution, non-fractionated sample preparations were used for solution SAXS experiments. Fig. 5 shows the typical SAXS patterns of BSM aqueous solutions. The sample-concentration dependence of the corrected intensity (at q = 0.12 nm^{-1}) divided by the weight concentration suggests no drastic change in molecular weight distribution of BSM molecules in solution under the present conditions, as shown in the inset to Fig. 5.

The notion of a fractal has been introduced quite successfully in physical chemistry (Mandelbrot, 1983) and has been used to interpret the scattering results. The scattering intensity will scale as

$$I(q) \simeq q^{-D}. \tag{3}$$

This scaling law has frequently been used to determine the fractal dimensions (D) from the log–log plot of I *versus* q. The Gaussian chain has a fractal dimension of two. On the other hand, a chain with excluded volume has a fractal dimension of 5/3 (= 1.67), which is smaller than 2 but larger than 1 (corresponding to rods) (Dewey, 1997). In this study, the slope of the log I *versus* log q plot (Fig. 5) in the q range of the SAXS experiments can be estimated to be 1.67 at concentrations of 1.4 and 3.6 mg ml^{-1}. Furthermore, the 2.0 fractal feature appears at the higher concentrations (7.2–15 mg ml^{-1}). The selected q range was sufficiently large for this protein (the radius of gyration R_g > 20 nm) to test this relation. The fractal dimension obtained indicates that the BSM molecule is a chain with excluded volume at concentrations of 1.4 and 3.6 mg ml^{-1} and the Gaussian chain character appears clearly at the higher concentrations (7.2–15 mg ml^{-1}). The expanded chains are found in dilute solutions in good solvents. The effective interaction between structural units of macromolecules is always repulsive and thus chains become expanded with decreasing conformational entropy. The structural unit may correspond to the highly glycosylated region of a mucin molecule (Fig. 1) since the terminal sialic acids in the oligosaccharide chains have anionic properties. The conformational entropy decrease makes a retracting force that balances the repulsive excluded volume force under equilibrium conditions. The expanded chain at lower concentrations changes to a chain with an ideal (Gaussian) chain character when the sample concentration is increased. This observation may be due to some partial weak interaction between mucin molecules.

Furthermore, as shown in Fig. 6, Kratky plots (Kratky, 1982) of the SAXS data show the Gaussian (ideal, random coil or unfolding) chain structure, although detailed conformational analysis can not be done with these plots because the raw data points are scattered at lower concentrations (the inset to Fig. 6). As shown in Fig. 7, the cross-sectional radius of gyration R_{gc} can be estimated by the straight region ($q^2 \times R_{gc} < 1$) of cross-sectional Guinier plots according to

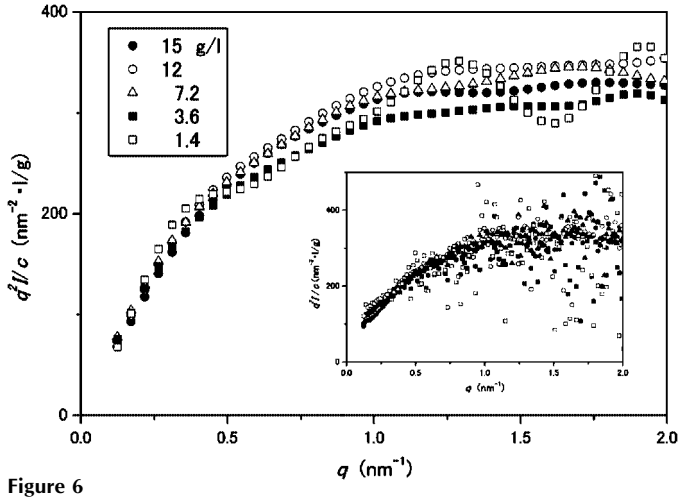

Figure 6
Kratky plots of the SAXS data. The data were obtained with a fast Fourier transform smoothing method (Press *et al.*, 1988). Raw data are shown in the inset figure. Other conditions are the same as those of Fig. 5.

$$Iq \simeq \exp(-R_{gc}^2 q^2/2). \quad (4)$$

The estimated R_{gc} value lies in the range 0.65–0.76 nm. When the R_g value of mucin can be estimated to be 20–170 nm (Sheehan & Carlstedt, 1984), the chain length will be calculated to be 70–590 nm assuming its rod-like shape. This estimation is reasonable since the large long structures of mucin molecules have already been pictured by electron microscopy (Sheehan *et al.*, 1986).

4. Conclusions

We found that the main components of intact BSM molecules are the oligomers with molecular weight of about 2 million because the molecular weight of a BSM subunit is several hundred thousand (Tettamanti & Pigman, 1968; Watanabe, unpublished results), containing a small amount of very large molecular weight species. The minor supramolecules may prevent structural analysis by methods such as gel electrophoresis or ultracentrifugation by sedimentation equilibrium. Fractal analysis of the first solution small-angle measurements revealed that the intact BSM molecule is a chain with excluded volume at lower concentrations (1.4 and 3.6 mg ml^{-1}) and a Gaussian chain character will appear in the chain conformation with increasing sample concentration (7.2–15 mg ml^{-1}). Kratky plots of the SAXS data show that the chain conformation of BSM molecules is a Gaussian (ideal, random coil, unfolded) chain structure in solution. The estimated cross-sectional radius of gyration value is reasonable for a long thin shape. These results show that small-angle scattering measurements are useful for characterization of intact flexible supramolecules such as mucins in solution.

We thank Drs J. Miyake and C. Nakamura for the use of the light-scattering equipment. This work was partially supported by a project grant from the Ministry of Agriculture, Forestry and Fisheries of Japan. The SAXS experiment was performed under the approval of the Photon Factory (proposal Nos. 03 G139 and 05 G084).

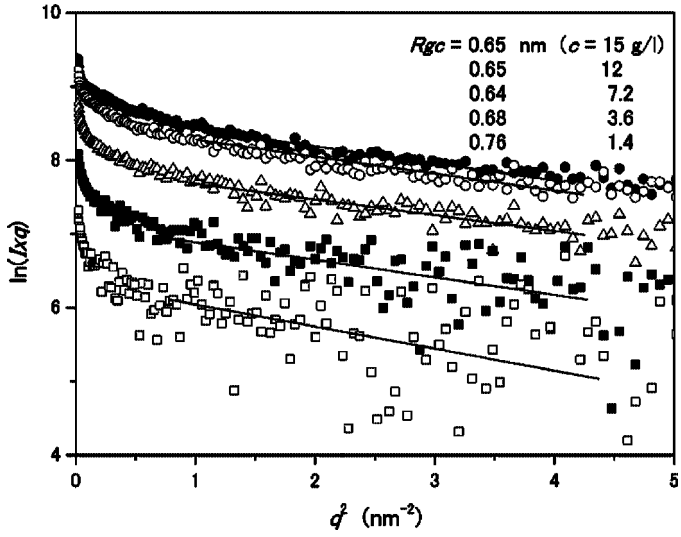

Figure 7
Cross-sectional Guinier plots of the SAXS data. Conditions are the same as those of Fig. 5. The estimated R_{gc} values are indicated in the figure.

References

Bettelheim, F. A., Hashimoto, Y. & Pigman, W. (1962). *Biochim. Biophys. Acta*, **63**, 235–242.
Carraway, K. L. & Fregien, N. (1995). *Trends Glycosci. Glycotech.* **7**, 31–44.
Dewey, T. G. (1997). *Fractals in Molecular Biophysics*. Oxford University Press.
Fasman, G. D. (1976). *Handbook of Biochemistry and Molecular Biology 2*. Cleveland, OH: CRC Press.
Inoko, Y. (1999). *KEK Proc.* **99–1**, 2–5.
Kratky, O. (1982). *Small Angle X-ray Scattering*, edited by O. Glatter & O. Kratky, pp. 361–386. New York: Academic Press.
Malmsten, M., Blomberg, E., Claesson, P., Carlstedt, I. & Ljusegren, I. (1992). *J. Colloid Interface Sci.* **151**, 579–581.
Mandelbrot, B. B. (1983). *The Fractal Geometry of Nature*. New York: W. H. Freeman.
Nozaki, Y. (1986). *Arch. Biochem. Biophys.* **249**, 437–446.
Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1988). *Numerical Recipes in C: The Art of Scientific Computing*. New York: Cambridge University Press.
Sheehan, J. K. & Carlstedt, I. (1984). *Biochem. J.* **221**, 499–504.
Sheehan, J. K., Oates, K. & Carlstedt, I. (1986). *Biochem. J.* **239**, 147–153.
Shi, L. & Caldwell, K. D. (2000). *J. Colloid Interface Sci.* **224**, 372–381.
Shogren, R. L., Jamieson, A. M., Blackwell, J. & Jentoft, N. (1986). *Biopolymers*, **25**, 1505–1517.
Takagi, T. (1990). *J. Chromatogr.* **506**, 409–416.
Tettamanti, G. & Pigman, W. (1968). *Arch. Biochem. Biophys.* **124**, 41–50.
Ueki, T., Hiragi, Y., Tagawa, H., Kataoka, M., Muroga, Y., Matsushita, T. & Amemiya, Y. (1983). *Photon Factory Activity Report* (1982/83), edited by K. Kohra, pp. VI70–VI71. Tsukuba: National Laboratory for High Energy Physics.
Watanabe, Y. (2005). *Protein J.*, **24**, 167–174.
Watanabe, Y., Sano, Y. & Inoko, Y. (1999). *Jpn. J. Appl. Phys.* **38**, 180–182.
Watanabe, Y. & Takagi, T. (1993). *J. Chromatogr. A*, **653**, 241–246.