Protein Interactions in Solution Characterized by Light and Neutron Scattering: Comparison of Lysozyme and Chymotrypsinogen

O. D. Velev, E. W. Kaler, and A. M. Lenhoff
Center for Molecular and Engineering Thermodynamics, Department of Chemical Engineering, University of Delaware, Newark, Delaware 19716, USA

ABSTRACT The effects of pH and electrolyte concentration on protein-protein interactions in lysozyme and chymotrypsinogen solutions were investigated by static light scattering (SLS) and small-angle neutron scattering (SANS). Very good agreement between the values of the virial coefficients measured by SLS and SANS was obtained without use of adjustable parameters. At low electrolyte concentration, the virial coefficients depend strongly on pH and change from positive to negative as the pH increases. All coefficients at high salt concentration are slightly negative and depend weakly on pH. For lysozyme, the coefficients always decrease with increasing electrolyte concentration. However, for chymotrypsinogen there is a cross-over point around pH 5.2, above which the virial coefficients decrease with increasing ionic strength, indicating the presence of attractive electrostatic interactions. The data are in agreement with Derjaguin-Landau-Verwey-Overbeek (DLVO)-type modeling, accounting for the repulsive and attractive electrostatic, van der Waals, and excluded volume interactions of equivalent colloid spheres. This model, however, is unable to resolve the complex short-ranged orientational interactions. The results of protein precipitation and crystallization experiments are in qualitative correlation with the patterns of the virial coefficients and demonstrate that interaction mapping could help outline new crystallization regions.

INTRODUCTION

Knowledge of the interactions between protein molecules in solution is fundamental to understanding protein functions in nature and in all practical processes involving proteins. One particularly important process is crystallization, which is a major step in protein separation and characterization. At present, protein crystallization may still be considered an art rather than a science, with optimal crystallization conditions generally determined by trial and error and intuition rather than from a firm theoretical basis (George and Wilson, 1994; Durbin and Feher, 1996). The onset of protein precipitation or crystallization and the morphology of any separated phase will be predominantly determined by the mechanism of molecular approach, reorientation, and incorporation, which is governed by the strength and range of the colloidal interactions (Durbin and Feher, 1996). Further theoretical and experimental study of these interactions is therefore the key to better understanding of protein crystallization (Durbin and Feher, 1996; Rosenberger, 1996). This area has developed slowly, however, and it was only recently that a pioneering study of George and Wilson (1994; see also George et al., 1997) demonstrated the relation between the protein osmotic second virial coefficient (a basic parameter representing the integral of the intermolecular potential over distance; Hunter, 1989) and protein crystallization behavior. One reason for the incomplete development of this field is the paucity of systematically developed experimental data.

Received for publication 8 January 1998 and in final form 17 August 1998.
Address reprint requests to Dr. Eric W. Kaler, Department of Chemical Engineering, University of Delaware, Newark, DE 19716-3119. Tel.: 302-831-3553; Fax: 302-831-4466; E-mail: kaler@che.udel.edu.
© 1998 by the Biophysical Society
0006-3495/98/12/2682/16 $2.00

The most versatile tools for investigating the bulk colloidal properties of protein solutions are scattering methods (Chen and Bendedouch, 1986; Hunter, 1989; Kaler, 1995), although a number of techniques have recently allowed measurement of the interactions between adsorbed layers in thin films (Leckband and Israelachvili, 1993; Blomberg et al., 1994), on particles, or even between single molecules attached to atomic force microscope tips (Florin et al., 1994). Interprotein forces in solution have been studied by light, x-ray, and neutron scattering techniques (Delaye and Tardieu, 1983; Chen and Bendedouch, 1986; Veretout et al., 1989; Skouri et al., 1992; Eberstein et al., 1994; Coen et al., 1995; Georgalis et al., 1995; Muschol and Rosenberger, 1996; Bonnete et al., 1997), and the scattering data can be related to the protein-protein interactions and separation behavior (Muschol and Rosenberger, 1995; Coen et al., 1995; Ducruix et al., 1996; Lafont et al., 1997; Gripon et al., 1998). A protein of particular interest is hen egg lysozyme, which crystallizes readily and for which solubility and crystallization parameters are known. Some aspects of the interactions in lysozyme solutions have been studied separately by static light scattering (SLS) (George and Wilson, 1994; Muschol and Rosenberger, 1995; Rosenbaum et al., 1996; George et al., 1997), x-ray scattering (Ducruix et al., 1996), and small-angle neutron scattering (SANS) (Giordano et al., 1991, 1992; Niimura et al., 1994; Gripon et al., 1996, 1997). Scattering methods have also been applied in an attempt to characterize the size and the growth dynamics of the initial aggregates in the crystallization process (Boue et al., 1993; Niimura et al., 1995; Tanaka et al., 1996). To date, however, the available data on lysozyme interactions do not cover the effect of all basic variables such as pH or ionic strength. In addition, the measurements have often

been carried out in the presence of high concentrations of a specific salt or buffer.

Thus a number of important aspects remain open: 1) whether the virial coefficients of different proteins below the isoelectric point follow a similar characteristic pattern as a function of pH; 2) as the scattering methods measure the interaction parameters indirectly, large and sometimes unpredictable errors can occur, therefore the reliability of the data should be checked by at least two independent methods; and 3) more data are required to clarify the correlation observed between the virial coefficient values and the crystallization/precipitation patterns. In the present study, we describe and compare interaction and crystallization data obtained with solutions of lysozyme and α-chymotrypsinogen. Chymotrypsinogen is a globular protein with a molecular mass of 25,670 Da (Creighton, 1993) and an isoelectric point at pH ~9.5. It has almost twice the molecular weight of lysozyme, with lower but less uniformly distributed surface charge. Previous research on a related protein, chymotrypsin (Coen et al., 1995), has shown the importance of the attractive electrostatic forces induced by the uneven charge distribution.

To compare the trends in the virial coefficient patterns of these two proteins, we systematically varied both pH and electrolyte concentration over a wide range, with only NaCl used as the major electrolyte. Thus we investigated not only the magnitude of the virial coefficients, but also the general pattern of their behavior. The measurements were carried out by SLS and SANS. To estimate the effect of the measured parameters on these conditions, simple crystallization tests were also performed. These data are used as a basis for discussion of the relation between the virial coefficients and protein crystallization and the similarities and differences between the two proteins studied.

MATERIALS AND METHODS

Materials

Hen egg white lysozyme (6× crystallized, L-6876) and α-chymotrypsinogen from bovine pancreas (C-4879) were obtained from Sigma. Solutions for SLS were prepared with deionized water from a Millipore Milli-Q system. The samples for SANS were dissolved in D_2O (Cambridge Isotope Laboratories). NaCl from Sigma was used to adjust the electrolyte concentration. A minimal amount of citrate buffer (citric acid from Fisher Scientific) in a ratio of buffer/NaCl of 1:10 was added to maintain a stable pH of the batch electrolyte solutions. Thus the predominant contribution to the electrolyte composition is from NaCl, a source of monovalent ions with minimal specific interactions with proteins. The pH of the batch electrolyte and of the protein solutions was measured by a Chemcadet 5984 digital meter and adjusted by the addition of small volumes of 0.1 M NaOH or HCl. The meter readings in the D_2O solutions were corrected by subtracting 0.45 units to take into account the difference between pH and pD measured by a glass electrode (Glasoe and Long, 1960).

Special care was taken to minimize protein aggregation and dust. All work was performed in a FS-209c certified, clean airflow bench. Glassware was treated with detergent, stored in Nochromix oxidizing solution, abundantly washed with deionized water, and dried in a clean environment shortly before each experimental run. The protein solutions were prepared less than 2 h before the measurements. The electrolyte solutions were filtered through 20-nm cutoff inorganic Anotop filters (Whatman). The protein solutions were filtered through 100-nm filters of the same brand and immediately diluted and sealed in 1-ml ampules. The protein concentrations were determined with a Perkin-Elmer Lambda 4B spectrophotometer at $\lambda = 280$ nm, using extinction coefficients of 2.64 liter/g cm (Sophianopoulos et al., 1962) and 2.0 liter/g cm (Smith, 1970) for lysozyme and chymotrypsinogen, respectively. All measurements were performed at 25 ± 0.1°C.

Methods

SLS data at a scattering angle of 90° were obtained with a Brookhaven light scattering apparatus, equipped with an Ar-ion laser (wavelength $\lambda = 488$ nm) and BI9000AT correlator. The absolute Rayleigh ratios of the samples were calculated by calibration with pure benzene before each experimental point. The disturbance from dust was minimized by using the built-in statistical dust rejection function and setting a tight rejection ratio (1.05). Each final data point is based on the average of not fewer than 50 statistically consistent measurements.

SANS experiments were performed with the 30-m instrument at the National Institute of Standards and Technology at Gaithersburg, MD. Quartz cells of 5-mm pathlength were filled following the same procedures as for the light scattering samples. The neutron wavelength was fixed at 6 Å, and three sample-to-detector distances (1, 4.5, and 13 m) were used to cover a wide range of scattering angles. The scattering from the pure electrolyte solutions and standard samples was also measured, and the data were normalized and put on an absolute scale with NIST software.

Lysozyme crystallization and precipitation were studied by examining solutions stored in closed sterile vials at 4°C. The crystallization of chymotrypsinogen was carried out by centrifugal concentration in a Sorvall RC-70 centrifuge equipped with a Sorvall SW-5 rotor. Samples of the solutions (5 ml each) were introduced into centrifuge tubes of diameter 13 mm and rotated for 44 h at 33,000 rpm and 25°C. The theory and detailed procedures are reported elsewhere (Lenhoff et al., 1997). Thin layers of samples from the crystallized solutions were deposited on microscope slides and quickly observed (within 20 min) by transmitted polarizing microscopy (Nikon Optiphot-2) for the presence of crystalline phases.

Scattering data processing

The scattering intensity from protein solutions, I, is given in general by the product of the structure factor, $S(q)$, the form factor, $P(q)$, and the molecular protein concentration, ρ (Hunter, 1989; Belloni, 1991; Kaler, 1995):

$$I(q) = \rho P(q) S(q) \tag{1}$$

where the magnitude of the scattering vector q is a function of the scattering angle θ and the wavelength λ:

$$|q| = \frac{4\pi}{\lambda} \sin\left(\frac{\theta}{2}\right) \tag{2}$$

The form factor is an orientation-averaged function of the size and shape of the protein molecules. The structure factor, $S(q)$, is the Fourier transform of the radial distribution function of the molecules in solution, $g(r)$, and samples the averaged distribution of intermolecular distances (in the formulae hereafter the dependence on the relative angular position of the molecules is omitted):

$$S(q) = 1 + 4\pi\rho \int_0^\infty (g(r) - 1) \frac{\sin(qr)}{qr} r^2 \mathrm{d}r \tag{3}$$

The radial distribution function is related to the potential of mean force $W(r)$:

$$g(r) = \exp(-W(r)/kT) \tag{4}$$

In the light scattering regime, the effective protein radius and the mean distances between the molecules in solution (Fredericks et al., 1994) are much lower than the wavelength of the laser used (488 nm). Thus the form factor is effectively equal to unity for most of the accessible q values. This allows us to measure the scattered intensity at an angle of 90°, which minimizes eventual effects from dust particles. For neutron scattering, the form factor varies with q but approaches unity at $q \rightarrow 0$. In this limit, the structure factor reduces to (Goodwin, 1981; Chen and Bendedouch, 1986)

$$S(q \rightarrow 0) \approx I(0) = kT\left(\frac{\partial \rho}{\partial \pi}\right)_T \cong (1 + 2B'_{22}\rho)^{-1} \quad (5)$$

where π is the osmotic pressure of the protein solution and B'_{22} is the second virial coefficient, which samples the net effect of the interaction energy over the whole range of accessible configurations:

$$B'_{22} = -2\pi \int_0^\infty (g(r) - 1)r^2 \mathrm{d}r \quad (6)$$

When the protein molecules are approximated as spheres of radius r_p, $g(r) \equiv 0$ over the range $0 < r < r_p$, because of steric repulsion. The value of the above integral in this range is equal to four times the molecular volume of the protein and makes up the so-called excluded volume contribution to B'_{22}. The value of the integrand in the range $r_p < r < \infty$ is a complex function of the intermolecular interactions present at different separations. A positive virial coefficient reflects the presence of predominantly repulsive interactions and vice versa.

The SLS data were processed by the classical method of Zimm (1948), whereby

$$\frac{KC_p}{R_\theta} = \frac{1}{M_W} + 2B_{22}C_p \quad (7)$$

where M_w is the molecular weight of the protein and C_p is its w/v concentration. The second virial coefficient as determined experimentally using Eq. 7 has units of mol ml/g^2 and is related to that given in Eqs. 5 and 6 by $B_{22} = B'_{22}/M_w^2 N_A$ (N_A is Avogadro's number). The Rayleigh ratio, R_θ, is the normalized scattering intensity, which was calibrated by using pure benzene and subtracting the background scattering from the pure electrolyte solution. K is a constant that is calculated from the optical properties of the system (Utiyama, 1972):

$$K = \frac{4\pi^2 n_0^2}{N_A \lambda^4}\left(\frac{\mathrm{d}n}{\mathrm{d}C}\right)^2 \quad (8)$$

where n_0 is the refractive index of the solvent, ($\mathrm{d}n/\mathrm{d}C$) is the refractive index increment of the protein, and λ is the wavelength of the laser in vacuum. The values of ($\mathrm{d}n/\mathrm{d}C$) used were 0.20 ml/g for lysozyme (Fredericks et al., 1994, confirmed by our independent measurements) and 0.192 for chymotrypsinogen (Smith, 1970). The value of $R_\theta(90°)$ for benzene was set at 38.6×10^{-6} cm^{-1}, interpolated from the data of Coumou et al. (1960, 1964).

Once the data are plotted according to Eq. 7, the values of the slope and the intercept give the second virial coefficient and the molecular weight of the dissolved protein, which is independently known from its amino acid composition and therefore can be used as a check on the experimental precision.

RESULTS

Static light scattering

Lysozyme

The virial coefficient measurements for lysozyme span the pH region from 3 to 10.6 at four different electrolyte concentrations—0.005, 0.1, 0.3, and 0.5 M NaCl. Measurements at the higher electrolyte concentrations of 0.1, 0.3, and 0.5 M were made to determine the behavior of B_{22} upon the approach to protein crystallization conditions (Alderton et al., 1945; Ataka and Shoji, 1986; Howard et al., 1988), whereas the measurements at 0.005 M NaCl probed the interactions in the low-electrolyte region, where long-range electrostatics are significant. The data were obtained with protein solutions diluted below 10 mg/ml. This value is lower than the reported lysozyme solubility values at all salt concentrations in the pH range 4–7.5 at 25°C (Howard et al., 1988), which reduces the risk of interference from higher order terms in the virial expansion and of experimental errors arising from concentration-induced protein aggregation. Examples of typical experimental measurements plotted according to Eq. 7 are shown in Fig. 1. The lines show the results of least-squares linear regression with both the slope and the intercept varied. The lines should meet at the same point, which is related to the reciprocal of the protein molecular weight (Eq. 7). This is reasonably well accomplished for all of the data plots, with the exception of that at the lowest electrolyte concentration and pH. The average intercept, excluding the point at pH 3 and I = 0.005 M, corresponds to a protein molecular weight of 14,600. This value is closer to the lysozyme molecular weight from amino acid sequencing (14,320) than to the molecular weights determined by other physical methods (e.g., sedimentation; Creighton, 1993, p. 266).

The virial coefficients obtained (Fig. 2) are highly positive (corresponding to repulsive interactions) at low electrolyte concentrations and low pH. As pH increases, the coefficients decrease, dropping sharply above pH 9 to reach a negative value at pH ~10.5. The virial coefficients are slightly positive at I = 0.1 and 0.3 M electrolyte and pH 3, but they become increasingly negative at intermediate and

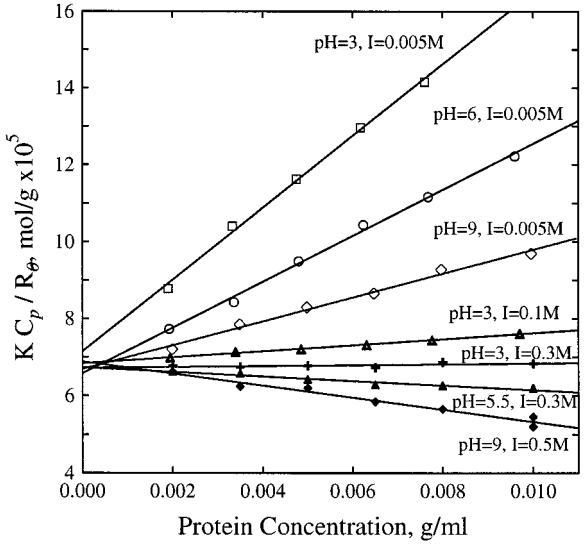

FIGURE 1 Typical plots of lysozyme SLS data according to Eq. 7. The variation of the virial coefficient (given by the slope of the lines) with pH and electrolyte concentration is evident.

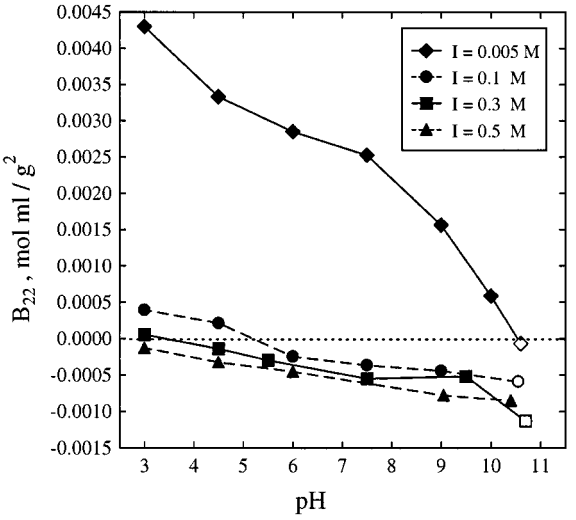

FIGURE 2 A summary of virial coefficients of lysozyme obtained from the SLS data at varying pH and four electrolyte concentrations (each point corresponds to one data set similar to those plotted in Fig. 1). The open symbols denote experimental conditions where the onset of aggregation was detected.

high pH. All of the coefficients at $I = 0.5$ M are negative and decrease slowly with pH. The primary data at pH 10.5–10.6 typically demonstrate a negative deviation from the expected common intercept value (cf. Fig. 1), which formally corresponds to higher molecular weight and is probably an indication that some protein aggregation may occur during the time of measurement (2–3 h).

For all pH values, increasing the electrolyte concentration leads to a decrease in the virial coefficients. This dependence has been studied in more detail at pH 4.5 (Fig. 3). The virial coefficients decrease monotonically with increasing electrolyte concentration, the slope becoming smaller at higher concentrations. The result of a single measurement carried out with 0.05 M NaCl and 0.05 M sodium acetate buffer at pH 4.5 is also shown (after recalculation of the degree of acetic acid ionization) and is in good agreement with the values for samples containing NaCl as the only major electrolyte.

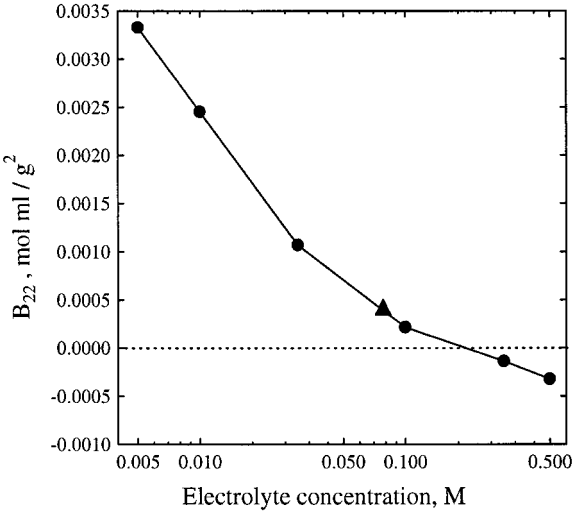

FIGURE 3 Lysozyme virial coefficients measured at pH 4.5 as a function of electrolyte concentration. The single point marked with a triangle was obtained in an acetate buffer environment.

Chymotrypsinogen

As for the lysozyme, the measured chymotrypsinogen solutions were dilute enough to be below the solubility limit of the protein. Chymotrypsinogen tends to deteriorate at pH values above 5, partly because of the autocatalytic self-cleavage initiated by the presence of a small amount of chymotrypsin in the samples. The self-cleavage and self-digestion process can be suppressed by adding phenylmethylsulfonyl fluoride (PMSF), an inhibitor of the enzymatic activity of chymotrypsin (Gold and Fahrney, 1964). The protein solutions containing PMSF indeed gave more stable SLS data with time than those without inhibitor, but PMSF also caused significant changes in the values of the measured virial coefficients, e.g., the coefficient decreased by almost an order of magnitude at $I = 0.3$ M. Although PMSF has been reported not to bind to chymotrypsinogen at pH 7 (Gold and Fahrney, 1964), it can possibly adsorb or change the molecular configuration at other experimental conditions. To avoid such artefacts, further measurements were made in solutions without PMSF. Consistent results from pure protein solutions could be obtained by simply making the measurements reasonably quickly, before self-digestion or dimerization could influence the data. The molecular weight obtained from the intercepts (~24,450) agrees well with the literature values determined from sedimentation (23,660) and from structural data (25,670; Creighton, 1993, p. 266).

A summary of all of the virial coefficient data obtained is presented in Fig. 4 (the lines are included to guide the eye). The values obtained correlate quite well with the data of Coen et al. (1995) for chymotrypsin, which have a similar structure. The virial coefficients at low pH and $I = 0.005$ M are positive but much smaller in magnitude than those of lysozyme. Increasing the pH at low ionic strength leads to a drastic decrease of the virial coefficients, which reach highly negative values at pH 6.8. Increasing the electrolyte concentration at pH 3 reduces the virial coefficients until they become slightly negative at $I = 0.3$ M. The data for $I = 0.3$ M electrolyte are practically independent of pH and are similar to the high I behavior of lysozyme, and the values at the intermediate concentration of $I = 0.1$ M lie between the other two limiting cases. The cross-over of the lines at a pH of ~5.2 suggests that the virial coefficients at this pH are almost unaffected by the value of the ionic strength. Finally, to ensure that the small quantity of citrate buffer used does not interfere with protein-protein interactions, a single measurement of a solution containing 0.1 M acetate buffer was made. The values for the two buffers were essentially the same (Fig. 4).

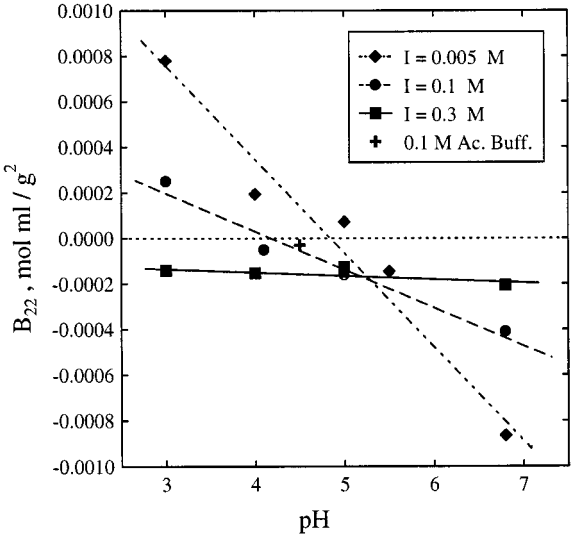

FIGURE 4 Summary of the chymotrypsinogen SLS data at three different electrolyte concentrations as a function of pH (each point corresponds to one data set similar to the lines plotted in Fig. 1).

Small-angle neutron scattering

Lysozyme

SANS spectra $I(q)$ were obtained for six lysozyme solutions of 10 mg/ml at pH 4.5, 6, and 9, and electrolyte concentrations of 0.005 and 0.3 M (Fig. 5, *A–C*). For the q range between 0.004 and 0.05 $Å^{-1}$, the curves are well separated, and the intensity values increase with decreases in the corresponding virial coefficients. This behavior at low q reflects the role of the structure factor and thus depends on protein-protein interactions. The neutron scattering spectra overlap in the q range from 0.05 to 0.18 $Å^{-1}$ and diverge again for higher q values. In the overlap region $S(q)$ is $\sim$1.0, so the scattering reflects the form factor, $P(q)$, of the lysozyme. A reasonable form factor approximation (the molecular dimensions of lysozyme are well known) should fit the data in this region. Two such approximations are plotted in Fig. 5 *A*. The first approximation is that of a sphere of radius 16 Å, which has a volume equivalent to the lysozyme molecular volume. The second approximation is that of an ellipsoid with semimajor and semiminor axes, $a = 22.5$ Å and $b = 15$ Å, which correspond to the lysozyme dimensions determined crystallographically (Creighton, 1993). The value at the intercept, $P(0)$, was calculated based on data from the literature (Gripon et al., 1996). The formulae and the parameters used are described in the Appendix. Although both of these approximations match the experimental results, a slightly more realistic fit is obtained with the form factor for an ellipsoidal particle.

Both the spherical and the ellipsoidal approximations begin to deviate significantly from the data at $q \geq 0.18$ $Å^{-1}$, which is approximately where higher order terms in the expansion of the scattering function become significant (Stuhrmann and Fuess, 1976) and the effects of hydration layers may be important (Svergun et al., 1998). The experimental plots above this q range diverge slightly again, and this does not appear to be related to experimental artifacts. The data at $q > 0.18$ $Å^{-1}$ probably reflect scattering from the fine details of the lysozyme outer hydrated structure, and the differences among the curves are likely to be a result of small changes in the conformation of the outer chains resulting from the difference in pH or electrolyte concentration, or from ion binding to the protein surface.

Once the form factor has been modeled, the structure factor can be extracted by dividing $I(q)$ by the form factor (Eq. 1). The virial coefficients B_{22} are extracted from the structure factor obtained from Eq. 5, using an extrapolation of the data to $q \rightarrow 0$ via a polynomial least-squares fit approximation (note that as the value of $P(0)$ is known, these data are obtained without fitting parameters). The values of B_{22} obtained are presented in Table 1. A graphical presentation of correspondence between the second virial coefficient values by SLS and SANS is shown in Fig. 6. The data by SANS are systematically lower, which may be attributed to the lower lysozyme virial coefficients in D_2O as compared to H_2O (Gripon et al., 1997). However, given the possible significant sources of error (particularly at low electrolyte concentration) and the absence of any adjustable parameters in the calculations, the two sets of estimates of B_{22} are in satisfactory agreement.

Chymotrypsinogen

SANS studies were carried out on three chymotrypsinogen samples of protein concentration 10 mg/ml: 1) pH 3, $I =$ 0.005 M; 2) pH 3, $I = 0.3$ M; and 3) pH 7, $I = 0.005$ M. The first two samples at pH 3 exhibited satisfactory time stability, as reflected in the overlap of the three data sets collected during consecutive time periods at different sample-to-detector distances. The scattering data from these samples are plotted in Fig. 7. As in the case of lysozyme, three regions in the $I(q)$ curves can be differentiated. For $q < 0.05$ $Å^{-1}$, the scattering profiles from the two samples are different as a result of the differences in the protein-protein interactions in the solution, which are reflected in the different forms of the structure factor, $S(q)$.

At higher q values the curves overlap, and in this region the scattering is determined by the form factor of the protein molecules, $P(q)$, primarily reflecting the molecular structure. The intensity values in the intermediate q range ($0.05 < q < 0.13$ $Å^{-1}$) could again be described by an effective sphere model for the single-particle scattering function as described in the Appendix. The fitted form-factor curve is plotted in Fig. 7. The deviations at high q probably reflect details of the intramolecular structure, as discussed earlier for lysozyme. After $P(q)$ is fitted, the virial coefficients can be obtained by extrapolating the data to the limit of $q \rightarrow 0$, as described above. The values obtained are compared with those from SLS in Table 2. There is general agreement between the values, although they are not as close quantitatively as in the case of lysozyme.

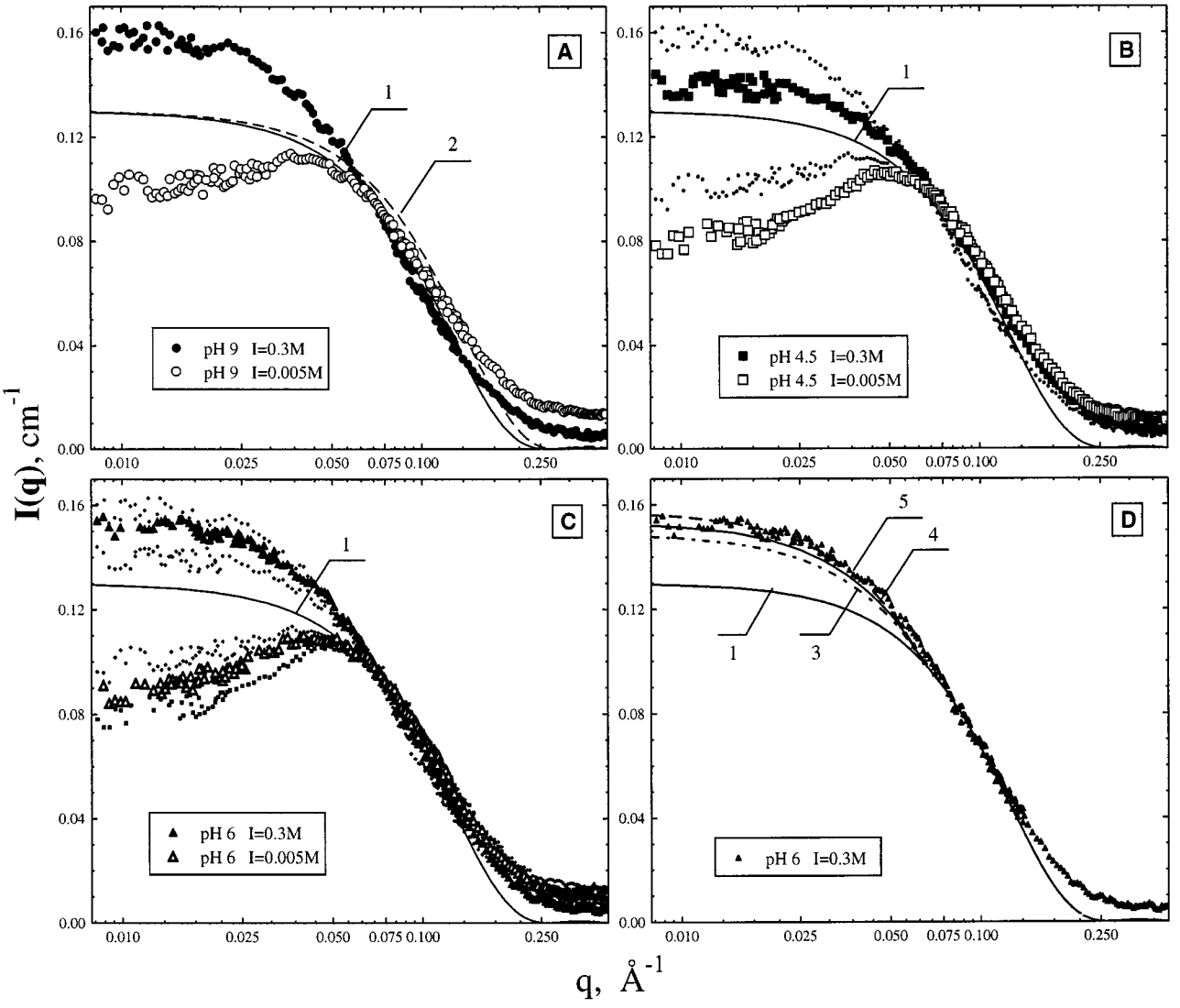

FIGURE 5 SANS spectra from lysozyme solutions. (*A*) Data points at pH 9. (*B*) Points at pH 4.5 (*large symbols*), compared with points at pH 9 (*small symbols*). (*C*) Points at pH 6 (*large symbols*), compared with data at pH 4.5 and 9 (*small symbols*). (*D*) Data at pH 6, $I = 0.3$ M, compared to the theoretical curves of total scattering intensity. Theoretical model curves: (1) Form factor in the ellipsoidal approximation; (2) form factor in the spherical approximation; (3) total scattering at pH 4.5 (structure factor calculated from Eqs. 3, 4, and 10 with the fitted parameters from SLS and multiplied by the form factor); (4) total scattering at pH 6; (5) total scattering at pH 9. Curve 1 is a hypothetical case of scattering without protein interaction and is shown in plots *A–D* for comparison.

The third neutron scattering sample, at pH 7 and $I = 0.005$ M, changed with time. Three data sets were obtained—two in the medium q region, at ~2 and ~10 h after sample preparation, and one in the high q region, 11 h after preparation. The data are plotted in Fig. 8 along with the data at pH 3. The scattering curves show a dramatic increase in the intensity with time in the q region below 0.06 $Å^{-1}$, whereas the intensity in the q region between 0.07 $Å^{-1}$ and 0.12 $Å^{-1}$ decreases.

This behavior can be explained by protein aggregation over the course of SANS experiment, which is longer than SLS experiments. The aggregation leads to an increase in the intensity in the low q region because the aggregates scatter strongly in this region. This increase is accompanied by an intensity decrease in the high q region due to the depletion of the monomolecular species. An exact description of a system of polydisperse aggregates is complex, and the parameters involved outnumber those that can be fitted based on the available data. However, an estimate of the degree of aggregation and the size of the aggregates is possible. If only a single type of spherical cluster of radius R_{agg} is formed (each containing n protein molecules of radius R) and the protein mass balance is taken into account, the scattering function from the solution will be given by (cf. Eq. A1 from the Appendix)

$$I(q) = (1 - \beta)S(q)V^2\Delta\rho_s^2(3j_1(qR)/(qR))^2 + n\beta S(q)V^2\Delta\rho_s^2(3j_1(qR_{agg})/(qR_{agg}))^2 \quad (9)$$

TABLE 1 Comparison of lysozyme virial coefficients obtained by SLS, SANS, and theoretical fitting

pH	Electrolyte conc. (M)	B_{22} (mol ml × 10^4/g^2) From SLS	From SANS	Calculated
3	0.1	3.95	—	10.1
	0.5	−1.26	—	−2.09
4.5	0.005	33.3	31.0	145
	0.1	2.15	—	4.27
	0.3	−1.39	−2.46	−2.80
	0.5	−3.24	—	−4.72
6	0.005	28.5	21.0	112
	0.1	−2.46	—	1.77
	0.3	−3.60*	−5.56	−4.02
	0.5	−4.54	—	−5.45
7.5	0.1	−3.65	—	0.3
	0.3	−5.50	—	−4.63
9	0.005	15.6	17.0	77.7
	0.1	−4.41	—	−1.00
	0.3	−5.28*	−6.46	−5.14
	0.5	−7.75	—	−6.09

The fitted parameters were $A_H = 13.8kT$, $d_p = 36.8$ Å.
*Interpolated value.

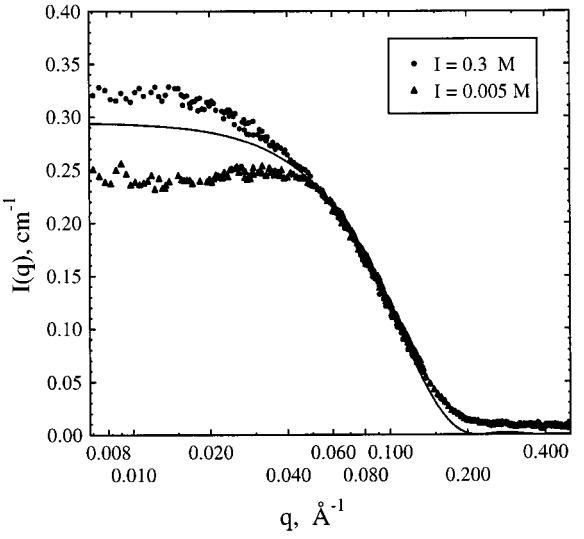

FIGURE 7 Intensity I versus scattering vector q for SANS measurements on chymotrypsinogen at pH 3 and two different electrolyte concentrations. The solid curve is the form factor function fitted by approximating the protein shape by a sphere. Positive deviations from this curve in the low q region indicate attractive interactions and negative deviations indicate repulsions.

Here β is the fraction of monomolecular protein converted into aggregates. The effective radius of the aggregates R_{agg} is an unknown function of n, but the simple physical restriction $R_{agg} < nR$ will always hold.

Fits of the data for the 2 h scattering curve at pH 7 with Eq. 9 were carried out with progressively increasing values of n in the q region above 0.05 Å^{-1}, where the scattering depends only on the size and shape of the species ($S(q) \equiv 1$). The physical restriction above was met at the lowest value of $n = 4$, with the fitting parameters taking the values $\beta = 0.33$ and $R_{agg} = 40$ Å. As seen from the theoretical fitting curves shown in Fig. 8 and expanded in the inset, the scattering intensity at q values higher than ~0.12 Å^{-1} is determined mostly by the presence of the monomers and increases at lower values because of the strong scattering from the aggregates. The value obtained for the effective size of the tetramer is twice the diameter of a single molecule, which is reasonable, although the real physical picture likely includes polydisperse aggregates. The negative deviation of the theoretical curve from the data at low q values reflects the fact that $S > 1$ because of the strongly attractive interactions, which are not accounted for in the model. Thus the estimate above suggests that ~30% of the original protein sample aggregated in a period of 2 h into clusters comprising four or more molecules each. The fraction of monomers incorporated into clusters increased with time, as evidenced by the continued increase in intensity at low q, due to the increasing number of aggregates, and the con-

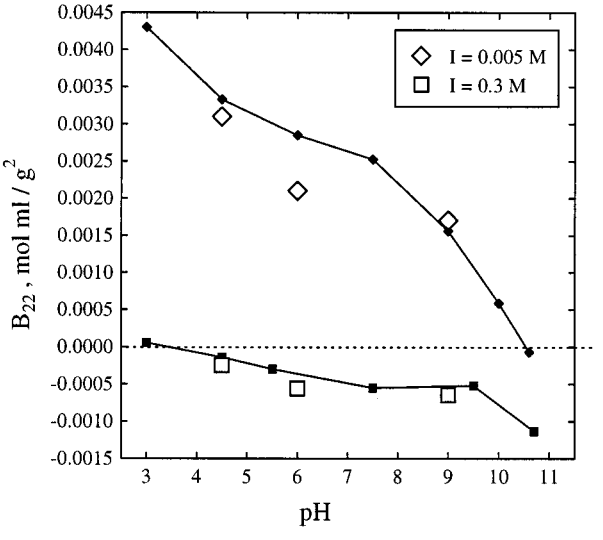

FIGURE 6 The values of the virial coefficients for lysozyme obtained by SANS (*large symbols*) compared to the corresponding SLS data (*small symbols and lines*).

TABLE 2 Comparison of chymotrypsinogen virial coefficients obtained by SLS, SANS, and theoretical fitting

pH	Electrolyte conc. (M)	B_{22}, mol ml/g^2 × 10^4 From SLS	From SANS	Calculated
3	0.005	7.8	4.46	90.8
	0.1	2.50	—	3.21
	0.3	−1.40	−1.8	0.08
5.25	0.005	−1.75*	—	−1.75
	0.1	−1.75*	—	−1.95
	0.3	−1.75*	—	−2.22
6.8	0.005	−8.65	−4.8[#]	−6.08
	0.1	−4.10	—	−2.52
	0.3	−2.05	—	−2.43

The fitted parameters were $A_H = 10.1kT$, $d_p = 43.5$ Å.
*Common interpolated value.
[#]Estimated by the idealized model (Eq. 9).

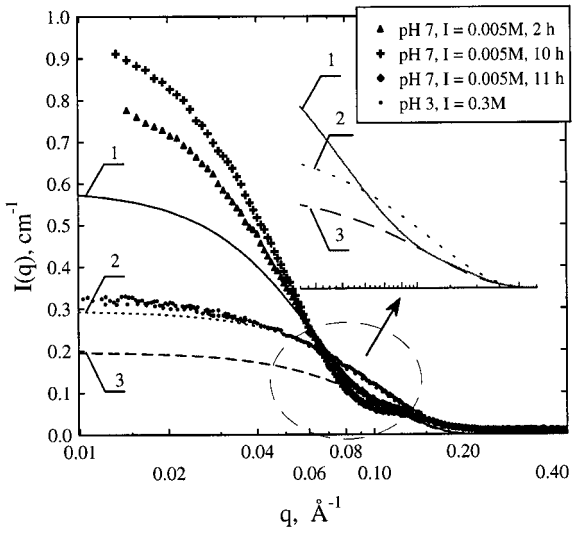

FIGURE 8 SANS chymotrypsinogen data at pH 7. As the sample changes with time, the three sets of data are shown as separate plots. The data at pH 3 and $I = 0.3$ M are also shown for comparison. The lines are (1) the plot of the fit according to Eq. 9; (2) the form-factor fit as in Fig. 3; and (3) the plot of the first term in Eq. 9 only (i.e., the scattering from the remaining monomolecular protein alone). The same three theoretical lines are shown expanded in the inset.

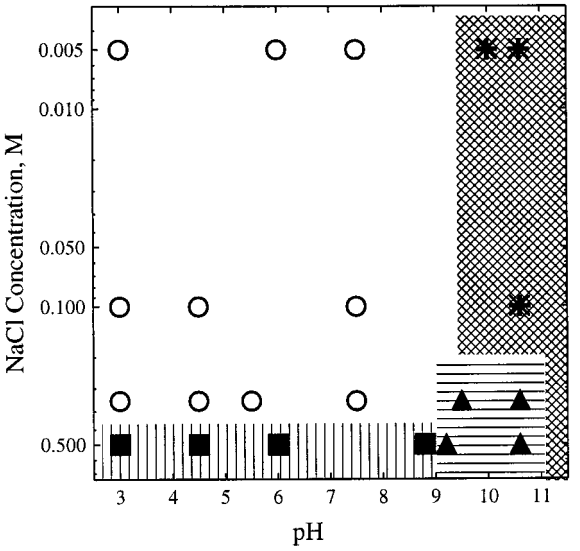

FIGURE 9 The lysozyme crystallization/precipitation pattern at 4°C mapped as a function of pH and electrolyte concentration. ○, No precipitation, crystallization, or any visible changes; *, ▨, partial precipitation; ■, ▥, compact tetragonal crystals; ▲, ▤, needle-like orthorhombic crystals.

comitant decrease in intensity at moderate q, due to depletion of the monomers.

Crystallization results

Lysozyme

Refrigerated solutions of lysozyme were monitored for up to 2 months after preparation, and four different patterns of crystallization or precipitation behavior were observed (Fig. 9):

1. Absence of precipitation or crystallization. This is characteristic for the low to medium range of electrolyte concentrations and pH values. The solutions remain stable and do not exhibit any inhomogeneity.

2. Partial precipitation. This was observed for pH values close to the lysozyme isoelectric point, pI (> 11), at high electrolyte concentration and even at lower pH values when the electrolyte concentration was also low. In this case some low-density, milky, or translucent precipitate formed in a few hours or days. Microscopic observation using crossed polarizers showed no traces of crystalline phases.

3. Compact, rhomboid-like crystals. These crystals grew in the pH region 3–9 and at an electrolyte concentration of 0.5 M. Both the nucleation and the growth were slow, with up to 2–4 weeks required to observe crystals. Good quality symmetrical crystals up to a few millimeters long were obtained (Fig. 10 *A*). X-ray crystallography of one of the crystals obtained at pH 4.5 indicated tetragonal symmetry, similar to that reported earlier for crystallization in this pH region (Alderton et al., 1945; Artymiuk et al., 1982).

4. Needle-like crystals (Fig. 10 *B*). These were observed between pH 9 and 10.5 and electrolyte concentrations above 0.3 M. This is the most intriguing region of lysozyme behavior because the crystals grew in only a few days. When observed by microscopy, these crystals demonstrated uniform regular cross section and rotation of the polarized light. X-ray investigation of a sample obtained at pH 10.5 and $I = 0.3$ M showed resolution better than 2.8 Å, which is typical for medium-quality protein crystals. The crystal was in an orthorhombic space group with lattice constants 30.39, 57.72, and 67.81 Å. This is a comparatively rare type of lysozyme crystal modification that has been reported previously (Alderton et al., 1945; Artymiuk et al., 1982; Tanaka et al., 1996), but for which the crystallographic structure seems to have been resolved only to 6-Å resolution (Artymiuk et al., 1982).

Chymotrypsinogen

Chymotrypsinogen could not be crystallized in refrigerated solutions because this protein exhibits retrograde solubility, i.e., the solubility increases with decreasing temperature. The centrifugal method used as an alternative yielded either a crystalline precipitate or a dense transparent liquid phase (concentrated in protein) at the bottom of the centrifuge tubes. Immediately after centrifugation the top portion of the samples was carefully withdrawn, leaving ~0.5 ml of the bottom fraction together with the separated phase. The amount of aqueous phase available for redissolution of the sedimented crystals was thus decreased by a factor of 10. Therefore, if the crystalline phase is thermodynamically stable outside the centrifuge, it will reach equilibrium with the surrounding solution fairly quickly. The precipitate was

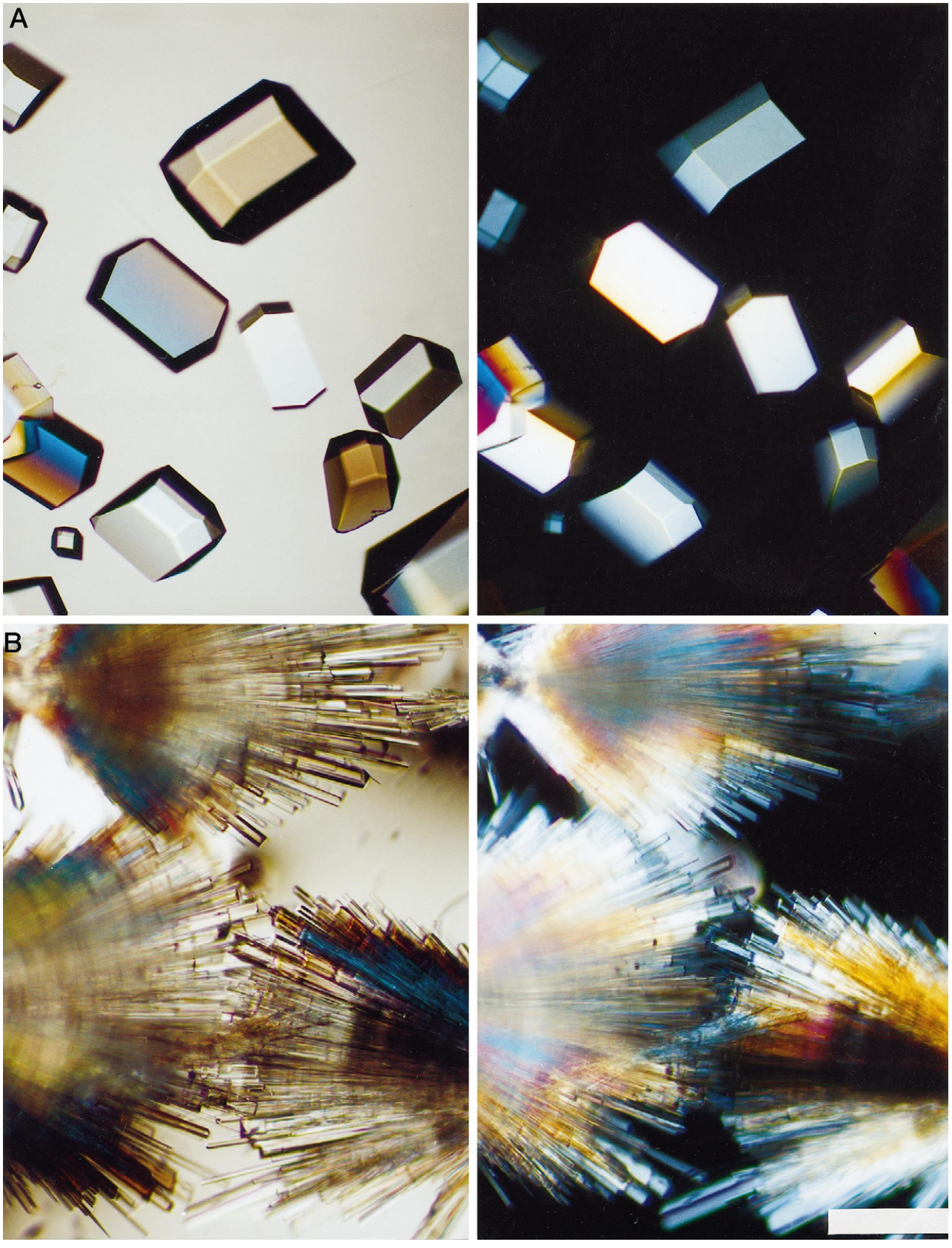

FIGURE 10 Micrographs of lysozyme crystals grown at (*A*) pH 3, $I = 0.5$ M, and (*B*) pH 10.6, $I = 0.3$ M. The left frames were obtained without polarizers. The right frames show the appearance of the sample between crossed polarizers, so that only the crystalline phase appears bright. The scale bar at bottom right is 300 μm.

considered stable if it remained intact for a prolonged period or was marked as unstable if it dissolved in the surrounding concentrated solution.

The results from the microscopic observation of small samples from the concentrated fractions are summarized in Table 3. Each system was studied for at least two concentrations, 10 mg/ml and 5 mg/ml or 2 mg/ml. No difference between the concentrated and the diluted samples was found, except for the proportionally smaller amount of crystalline material eventually collected. Only the sample at pH 5.25 exhibited crystallization at the higher salt concentration. In contrast, crystals formed in all of the samples at the lower electrolyte concentration. There was, however, a pronounced difference among these three samples. The precipitate at pH 3 dissolved in a period of ~30 min. The other two samples were stable but appeared different in polarized illumination. The sample at pH 5.25 demonstrated rotation of the polarized light and appeared bright on the darkened background, whrease the sample at pH 7 showed no optical anisotropy, although the precipitate consisted of crystals with clearly defined edges (Fig. 11).

DISCUSSION

The values obtained for the virial coefficients have been verified by two independent methods and agree well with the few common points of literature data at pH 4.5 (George and Wilson, 1994; Muschol and Rosenberger, 1995; George et al., 1997). Thus we believe our data to be precise and reliable enough to serve as a basis for testing theoretical predictions. The mechanisms of protein-protein interaction in solution are known in principle, but because of their complexity the typical quantitative models are based mainly on approximate expressions. Here we first compare these approximate theories with the data and explore their advantages and limitations. In the next subsection, the virial coefficient data are discussed from the viewpoint of protein crystallization. Finally, a comparison between the virial coefficient and crystallization patterns of lysozyme and chymotrypsinogen is presented.

Protein-protein interactions in solution

Precise calculation of the interaction energies of protein molecules is a computationally difficult procedure, mainly because the complex geometry and the uneven charge distribution should be taken into account by explicitly calculating the angular dependence of the interactions, together with the complex short-range electrostatic and hydration effects (Roth et al., 1996; Neal et al., 1998). The simplest crude approximation of the data is the calculation of the interactions between spheres of radius and electrostatic properties similar to those of the protein molecules. The protein-protein interaction energy $W(r)$ can be approximated by a simple DLVO type of model (Vilker et al., 1980; Haynes et al., 1992; Coen et al., 1995) as a sum of the contributions from the charge-charge electrostatic repulsion ($W_{q\text{-}q}$), the charge-dipole electrostatic attraction ($W_{q\text{-}\mu}$), the dipole-dipole attraction ($W_{\mu\text{-}\mu}$), the van der Waals attraction (W_{disp}), and all complex short-range interactions, denoted as W_{SR}:

$$W(r) = W_{q-q}(r) + W_{q-\mu}(r) + W_{\mu-\mu}(r) + W_{disp}(r) + W_{SR}(r) \quad (10)$$

$W_{q\text{-}q}$ is the leading and longest range electrostatic term, and when the protein is below its pI (as is the case for both proteins under the conditions studied here), this term will decrease with increasing pH, together with the overall positive charge, reaching zero at the isoelectric point. Whereas the charge-charge repulsion is universally important for the interactions between protein molecules in solution, $W_{q\text{-}\mu}$ and $W_{\mu\text{-}\mu}$ are significant for proteins that have large enough dipole moments. $W_{q\text{-}\mu}$ also vanishes at the isoelectric point. In contrast, $W_{\mu\text{-}\mu}$ does not become zero at pI if the dipole moment of the molecules does not vanish at this point. The van der Waals attraction, W_{disp}, becomes important at small separations. W_{SR} denotes other short-range interactions such as hydrophobic interactions, ion depletion, hydrogen bonding, and specific electrostatics. Because these complex interactions are still incompletely understood and explicit expressions are unavailable or intractable, they were not included in the calculation scheme.

The interplay of these intermolecular forces can explain the qualitative trends in the virial coefficient data. The behavior at low ionic concentrations is governed by the long-range electrostatic interactions. The protein is strongly positively charged at low pH (far from the isoelectric point), so the repulsion between the molecules dominates in the positive (repulsive) virial coefficient (Retailleau et al., 1997). The net positive charge decreases with increasing pH, increasing the importance of the long-range attractive van der Waals and electrostatic forces. The repulsive and attractive interactions approach a balance above pH ~9 for lysozyme and pH 5.25 for chymotrypsinogen, and the virial coefficients above those pH values become negative. The data for chymotrypsinogen, where above pH 5.2 the virial coefficients are more attractive at lower electrolyte concentrations, indicate the presence of attractive electrostatic forces. In a DLVO-type model these forces can be approximated as charge-dipole and dipole-dipole electrostatic attraction because chymotrypsinogen has a significant dipole moment.

TABLE 3 Results of centrifugal crystallization experiments with chymotrypsinogen

pH	Electrolyte conc. (M)	Crystals formed	Stable*	Rotate polarized light
3	0.005	Yes	No	Yes
	0.3	No	—	—
5.25	0.005	Yes	Yes	Yes
	0.3	Yes	Yes	Yes
7	0.005	Yes	Yes	No
	0.3	No	—	—

*Insoluble in the remaining 0.5 ml of concentrated mother liquor.

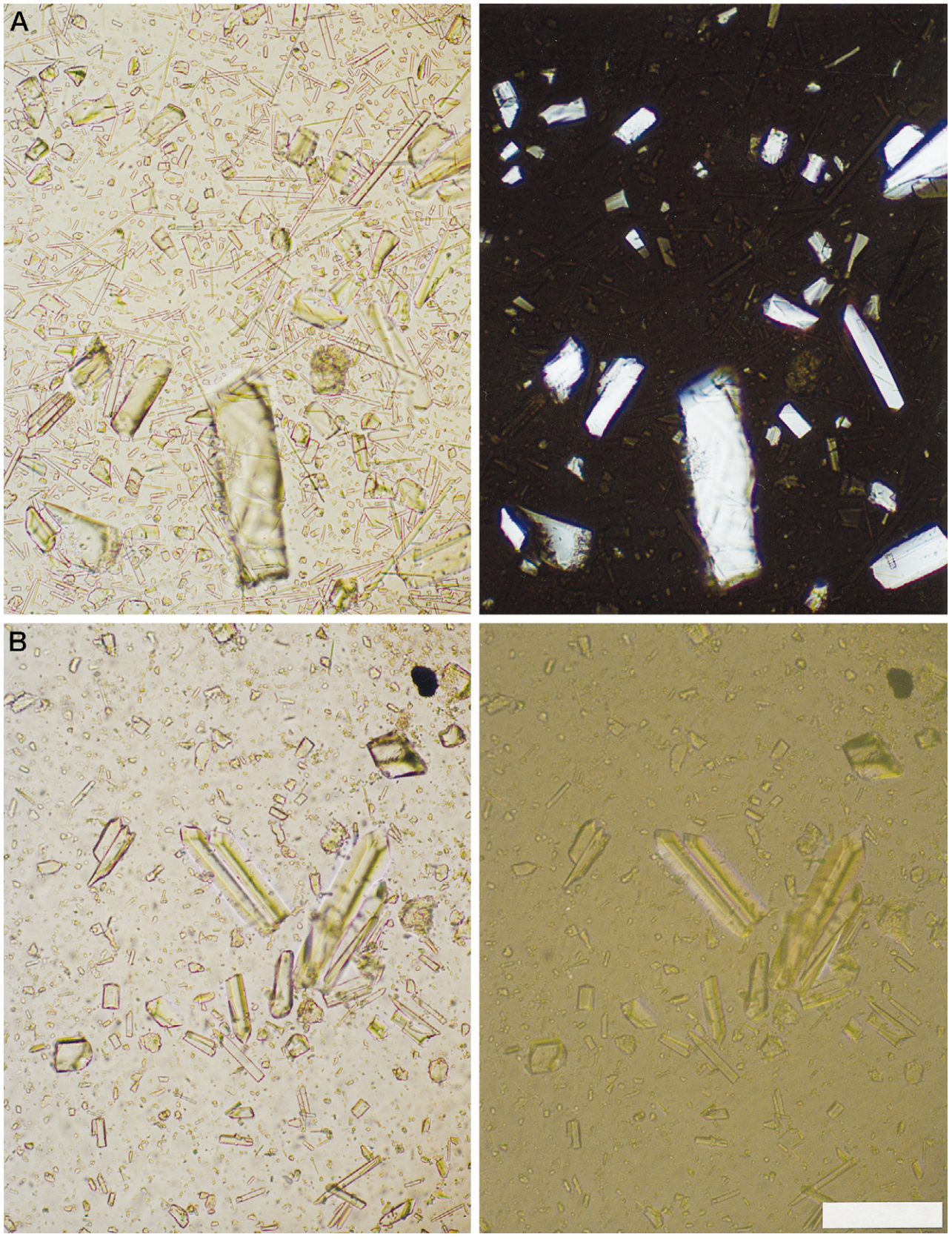

FIGURE 11 Micrographs of chymotrypsinogen crystals obtained by the centrifugal method at $I = 0.005$ M. (*A*) pH 5.25; (*B*) pH 7. The left frames are obtained without polarizers, and the right frames with crossed polarizers. The scale bar at bottom right represents 300 μm.

The addition of electrolyte both suppresses the strength and reduces the range of the electrostatic interactions, and it is not surprising that the virial coefficients generally decrease with increasing NaCl concentration. The negative virial coefficients at 0.3 and 0.5 M electrolyte are small in magnitude irrespective of pH. The van der Waals attraction should be dominant in this range. Theoretical studies (Rosenbaum et al., 1996; Neal et al., 1998) suggest that short-range interactions can be crucial in determining the overall value of the virial coefficients of proteins in solution. Although long-range electrostatic forces are suppressed, short-range electrostatics, together with the solvation and hydrophobic affects, may still contribute significantly to the interactions.

A more quantitative analysis was performed numerically, where the virial coefficients were evaluated on the basis of the hard sphere excluded volume and contributions given in Eq. 10 without the complex short-range W_{SR} term (similar to the procedures of Vilker et al., 1980; Haynes et al., 1992; Coen et al., 1995). The protein charge, z_p, and dipole moment, μ_p, were calculated for any given pH by the numerical procedure of Barlow and Thornton (1986; see also Takashima and Asami, 1993). The coordinates of the protein atoms were taken from the Protein Data Bank (PDB) files 194L (tetragonal lysozyme crystal at 1.4 Å resolution) and 2CGA (the file describes a chymotrypsinogen dimer, so only the coordinates for chain A were used). The degree of polarization of the amino acid residues at the defined pH was calculated on the basis of the Henderson-Hasselbalch equation, using the pK_a values from Stryer (1988).

The fitting parameters in the numerical procedure were the Hamaker constant, A_H, and the effective protein diameter, d_p. In contrast to previous studies, we did not allow A_H to vary for the fit of each experimental point, as there is no physical reason for the Hamaker constant to change. The objective function to be minimized was chosen as the sum of the absolute values of the differences between the measured and the calculated values of B_{22}. Using this objective function instead of the sum of the squares of the deviations attenuates the effect of deviating points at low pH and improves the fit across the whole pH region. The nontrivial parameters used are the relative permittivity of the protein interior, $\epsilon_s = 4$, and the smallest distance of approach allowed between the spherical molecules, $d_s = 3$ Å (Vilker et al., 1980, and Coen et al., 1995, interpreted this as corresponding to a layer of bound water between the proteins, leading to an excluded volume contribution of $2\pi(d_p + d_s)^3/3$). The overall interaction decays quickly with distance, so a finite upper limit of 200 nm was used in the calculations.

The results of the fitting procedure for lysozyme are shown in Table 1. Reasonable agreement is obtained within the 0.3–0.5 M range of electrolyte concentrations. The fitted molecular diameter is close to the effective diameter of the lysozyme molecule. The procedure could not fit the data at $I = 0.005$ M with reasonable accuracy, so these points were excluded from the optimization procedure. The large discrepancy between theory and experiment for the low electrolyte concentration may be related to electrostatic effects such as screening by ions released from the ionizable groups on the protein surface.

The virial coefficients represent integrals of $W(r)$, so in principle different forms of the potential of mean force could match the data. A more rigorous test of the applicability of the theory is the modeling of the $I(q)$ SANS spectral data by calculation of $S(q)$. Using the same sphere-sphere interaction model with the parameters obtained by fitting the SLS data (bottom of Table 1), the structure factor for lysozyme was calculated by numerical integration of Eq. 3 after inserting Eqs. 4 and 10, and the $S(q)$ obtained was used in Eq. 1. The resulting curves are plotted against the experimental data in Fig. 5 *D*. We find that the model fits the data reasonably well, given that the SANS structure data model curves are based only on parameters from the SLS fit.

The fitting procedure for chymotrypsinogen yields the results given in the last column of Table 2. The fitted effective protein diameter, $d_p = 43.5$ Å, is in accord with the crystallographic data and the neutron scattering results. The Hamaker constant obtained, $A_H = 10.1kT$, is close to that obtained for lysozyme and fitted data reported earlier (Coen et al., 1995). The model gives a picture of the trends in the virial coefficient change that is close to being quantitatively correct and that in particular captures the crossover point in the mid-pH region. Thus both repulsive and attractive electrostatics together with the dispersion attraction determine the trends in the virial coefficients.

Overall, the fits of both the SLS and SANS data demonstrate that the sphere-sphere interaction model can be used as a starting point to describe protein-protein interactions. The usefulness of this approach should not be overestimated, however. The calculations deviate from the data at both low and high pH values. Moreover, the value of the fitted Hamaker constants, $A_H = 10.1$–$13.8kT$, is ~3–5 times higher than that expected theoretically (Roth et al., 1996), and one may wonder whether the values of d_p and the closest molecular separation correspond to real physical dimensions in the protein system. Thus although the spherical approximation may work well at large separations, it fails to calculate precisely the complex short-range orientation-dependent interactions, because all of these interactions are parametrically incorporated into the van der Waals interaction term (hence the high value of A_H). The real physics of the system remains undetermined by this approach. Detailed calculations of the virial coefficient that account for the three-dimensional protein conformation and angular dependence of interactions are reported elsewhere (Neal et al., 1998).

Besides the agreement of fitted virial coefficients with experimental data, greater insight emerges from comparing the patterns of virial coefficients versus pH and ionic strength, which show a number of significant differences between the two proteins. For lysozyme, all of the curves at the same ionic strength are situated above each other for all pHs, whereas for chymotrypsinogen these lines all intersect

at around pH 5.2. The curves for chymotrypsinogen become independent of pH at an electrolyte concentration of 0.3 M, whereas the lysozyme virial coefficients still change somewhat with pH, even for $I = 0.5$ M. Finally, the magnitudes of the virial coefficients of lysozyme are higher in the repulsive region, but the values for both proteins are approximately the same in the attractive (negative) region. These differences can be explained by considering the magnitude and the distribution of the electrostatic charge in the protein molecules. The smaller lysozyme molecule is highly charged (about +8 at pH 7), and the oppositely charged groups are uniformly distributed across the surface (Barlow and Thornton, 1986; see also Rosenberger et al., 1996). As a result, lysozyme has a relatively low dipole moment (74 D at pH 7). The larger chymotrypsinogen molecule is more weakly charged at pH 7 (+4 charges) but has a much higher dipole moment of 518 D (all values are calculated based on sequence and coordinate data as described above).

The contrast between chymotrypsinogen and lysozyme implies that all proteins can be systematically divided into two groups with respect to their long-range electrostatic behavior. The first group includes proteins like lysozyme, for which repulsion is dominant at all pH values below the isoelectric point. The second group encompasses proteins like chymotrypsinogen, for which electrostatic repulsion controls at low pH but attractive electrostatics become dominant close to the isoelectric point. This finding has important consequences from the viewpoint of protein stability in solution. Proteins like lysozyme should be stable toward aggregation at lower electrolyte concentration in almost the entire pH region below the pI. Proteins like chymotrypsinogen, on the other hand, may be inclined to precipitate or crystallize when the electrolyte concentration is decreased at a pH where the dipolar attraction is predominant.

The above conclusion is supported by the neutron scattering data at pH 7 and low electrolyte concentration, which suggest that chymotrypsinogen has aggregated under these conditions. This is not surprising, bearing in mind the attractive interactions reflected in the negative virial coefficients. This aggregation and eventual precipitation (albeit not observed in our system) are in accordance with the correlation of George et al. (1994, 1997) relating lower virial coefficients to precipitation. It has traditionally been assumed that protein precipitation close to the isoelectric point is driven by hydrophobic and van der Waals attractions. The data for chymotrypsinogen at higher pH suggest, however, that electrostatic attraction is present and may contribute to protein aggregation. This effect may be significant for the precipitation of proteins with chymotrypsinogen-like virial coefficient patterns, whereas it may be insignificant for proteins with a virial coefficient pattern such as that seen for lysozyme.

Correlation between interaction data and crystallization patterns

Our scattering data were obtained under precrystallization conditions, i.e., the solutions are too dilute for crystallization or precipitation to occur at 25°C. The formation of lysozyme precipitate or crystals at 4°C is due mainly to lowering of the solubility. Once the solubility boundary is crossed, a condensed protein phase will start to grow, the structure of which is determined by the intermolecular interactions in the solution. The magnitude of the virial coefficients in the transition from 25°C to 4°C should decrease in a smooth and linear manner (Gripon et al., 1997), so the observed trends of the virial coefficients should remain approximately unchanged.

There are four qualitatively distinct regions in the plot of B_{22} of lysozyme as a function of pH and electrolyte concentration (Fig. 2) that correspond roughly to the four crystallization/precipitation patterns mapped in Fig. 9. The first is the region of low electrolyte concentration and low pH, where the virial coefficients are positive and no changes in the protein solutions occur with time at 4°C. The second region is at low and medium electrolyte concentrations, but high pH, where the pI is approached and B_{22} falls sharply because of the decrease in electrostatic repulsion. The dominance of the attractive interactions leads to rapid formation of precipitates.

At high electrolyte concentrations and low pH, the virial coefficients become slightly negative. The repulsive electrostatic element in this region is suppressed but still measurable, as can be seen by the dependence of the B_{22} on pH (Fig. 2). The slow growth of tetragonal crystals under these conditions is in agreement with literature reports (Ataka and Shoji, 1986).

The most interesting feature of the lysozyme crystallization map is the rapid growth of needle-like crystals with orthorhombic symmetry in the region above pH 9. In the virial coefficient data pattern (Fig. 2), this is the region where the virial coefficients decrease sharply because of the suppressed electrostatic repulsion. The changed balance between the repulsive and attractive interactions in this region obviously leads to the formation of another crystalline form, although the exact difference in the growth mechanisms is still unknown. Thus appropriately scaled interactions between the protein molecules can lead to rapid growth of high-quality crystals.

For chymotrypsinogen, the centrifugal method used allows the parameters (pH and electrolyte concentration) to be adjusted and then the protein concentration to be increased until the crystallization boundary is eventually crossed. The first set of crystallization data was obtained in the region of dominant electrostatic repulsion at pH 3. The second set, at pH 5.25, was at a point where the repulsive and attractive electrostatic forces are approximately balanced, and the third data set was at pH 7, where attractive electrostatics dominate. The differences among these three cases are clearly revealed by the different crystallization results at low electrolyte concentrations. Chymotrypsinogen crystallizes even at the lowest pH, where the measured virial coefficients are positive, but the crystals obtained in this way are metastable structures, created by the locally high values of the protein chemical potential at the bottom of the

tubes. This metastability drives redissolution of the crystals once the external centrifugal field is removed.

In accordance with the virial coefficient concept, the chymotrypsinogen solutions at pH 5.25 and 7, which exhibit negative virial coefficients, yield stable crystals that do not redissolve in the restricted amount of mother liquid. As demonstrated by the polarized light observations, the symmetry of these crystals is different. The lack of optical anisotropy of the crystals at pH 7 suggests that their unit cell may have a high degree of symmetry, compared to the optically active crystals at pH 5.25. As with lysozyme, there may be a transition from one crystal form to another in the region of higher pH, where the virial coefficients decrease sharply. The crystallographic structure of one of these phases will be reported elsewhere (Pjura et al., manuscript submitted for publication).

The only crystallizing condition for chymotrypsinogen found at the highest electrolyte concentration ($I = 0.3$ M) is at pH 5.25. The absence of crystallization at higher and lower pH may be purely a kinetic effect, due to the rapid supersaturation caused by the centrifugation. Some residual repulsive and attractive electrostatics still exist, as demonstrated by the independence of the virial coefficient on the electrolyte concentration. These electrostatics are approximately in balance at pH 5.2, so it is likely that delicately balanced and complex short-range forces govern the ordering. It is possible that this particular cross-over point may be conducive to crystallization of proteins exhibiting a pattern of virial coefficient behavior similar to that of chymotrypsinogen.

George et al. (1994, 1997) have shown experimentally that proteins typically crystallize when their virial coefficients are in the range from -1×10^{-4} to -8×10^{-4} mol ml g^{-2}. Our data agree well with this result. Overall, using virial coefficients as a predictor for protein crystallization conditions is a simple and attractive concept, but although most crystallizing conditions might be associated with slightly negative virial coefficients, it is doubtful whether all cases where slightly negative virial coefficients are observed will eventually be crystallizing conditions. For example, no crystallization was observed with chymotrypsinogen in two of the cases at $I = 0.3$ M, even though the virial coefficients in the solutions were negative. A feature of concentrated protein solutions that needs further understanding in the context of virial coefficients is the existence of liquid-liquid phase equilibria (Coen et al., 1995; Muschol and Rosenberger, 1997), which may be conducive to crystallization (Muschol and Rosenberger, 1997). Important progress has been made recently in modeling of phase separation in protein solutions, based on parametric descriptions of the intermolecular interactions (Rosenbaum and Zukoski, 1996; Lomakin et al., 1996; Malfois et al., 1996; ten Wolde and Frenkel, 1997), and further advances may help to elucidate some of the above issues. Unfortunately, at this stage these models assume an idealized molecular shape and simple potentials and thus are not able to predict the delicate nature of protein-protein complementarity in the crystal contacts.

Based on the results obtained, a simple semiquantitative strategy for managing a predictable crystallization protocol can be formulated. It appears sensible first to vary a parameter that strongly affects the virial coefficient until a slightly attractive coefficient is measured. Once this point is reached, the protein solubility can be reduced by changing factors that do not affect the interaction strength very much, e.g., decreasing temperature or concentrating the solution under isotonic conditions. In these experiments, such a path has been followed by first changing the pH and electrolyte concentration and then lowering the temperature or centrifuging.

The observed correlation between the different regions in the protein virial coefficient plot and the crystallization patterns suggests that interaction mapping might be used as a general predictor of protein crystallization and precipitation patterns and can outline areas where different behavior could be expected. It is reassuring that the pattern of virial coefficients is related to the type and biological specificity of a given protein and therefore can be used as a "signature" of that protein in considering its crystallization specificity. Both proteins studied here happen to display virial coefficients at high ionic strengths that fall in the "crystallization slot" (George et al., 1994, 1997). It remains to be seen if less easily crystallized proteins have virial coefficients that are remote from the crystallization slot at high electrolyte concentrations. Another problem worthy of exploration concerns the extent to which the virial coefficient changes might be related to the corresponding morphological changes of the separated protein phase.

CONCLUSIONS

This paper describes the results of investigations into the virial coefficients of lysozyme and chymotrypsinogen by both SLS and SANS. The virial coefficients are positive at low electrolyte concentration and pH. They become negative when the pH and electrolyte concentration are increased. Whereas the virial coefficients for lysozyme always decrease with increasing electrolyte concentration, with chymotrypsinogen there is a cross-over point at pH $\sim$5.2, above which the negative virial coefficients increase with increasing ionic strength. The different patterns of virial coefficient change are caused by the difference in the magnitude and distribution of the charges of the two proteins. The data can be approximately described by existing simplistic DLVO models, but their real value is in serving as a basis to test exact theoretical descriptions of the complex, orientation-dependent interactions in protein solutions.

The measured values of the virial coefficients can be correlated with protein crystallization and precipitation behavior, in agreement with previous observations (George et al., 1994, 1997). The significance of our results and the questions that are raised lie in the extent to which the trends

observed in the virial coefficient can be understood and manipulated to achieve desirable phase behavior. The differences observed between lysozyme and chymotrypsinogen indicate the potential complexity of the protein-protein interactions, which are inadequately captured by the kinds of potential of mean force models discussed here. In addition, the virial coefficient patterns of more proteins need to be explored extensively to allow general conclusions to be drawn. In particular, methods such as the ones described here may be useful in probing the efficacy of different chemical additives in yielding optimal crystallization conditions.

APPENDIX

The formula for the form factor of a sphere is (Glatter and Kratky, 1982; Hunter, 1989)

$$P(q) = V^2\Delta\rho_s^2(3j_1(qR)/(qR))^2 \quad \text{(A1)}$$

where V is the volume of the molecule, $\Delta\rho_s$ is the scattering contrast between the molecule and the solvent (D_2O), and $j_1(x)$ is the first-order spherical Bessel function given by

$$j_1(x) = (\sin(x) - x\cos(x))/x^2$$

The corresponding formula for $P(q)$ for a prolate ellipsoid is

$$P(q) = V^2\Delta\rho_s^2\left(\int_0^1 3j_1(u)/(u)\mathrm{d}\varphi\right)^2 \quad \text{(A2)}$$

where

$$u = q[a^2\varphi^2 + b^1(1 - \varphi^2)]^{1/2}$$

and a and b are the semimajor and semiminor axes of the ellipsoid.

The value of the lysozyme scattering density factor at 10 mg/ml, $P(0) = 0.132\ \mathrm{cm}^{-1}$, was calculated from the data of Gripon et al. (1996). In the case of chymotrypsinogen, only the expression for the form factor of a sphere, Eq. A1, was used, with both the radius, R, and the scattering contrast, $\Delta\rho_s$, being fitted. The value obtained for the effective protein radius, R, was 20.5 Å, which is reasonably close to both half of the actual crystallographic protein dimensions of $40 \times 40 \times 50$ Å (Creighton, 1993) and the effective radius of a sphere of volume equivalent to the molecular volume of chymotrypsinogen (19.4 Å, Creighton, 1993). The fitted scattering contrast, $\Delta\rho_s = 3.65 \times 10^{-14}\ \mathrm{cm\ Å}^{-3}$, is close to that of lysozyme ($3.30 \times 10^{-14}\ \mathrm{cm\ Å}^{-3}$).

We acknowledge the support of National Institute of Standards and Technology of the U.S. Department of Commerce and the National Science Foundation in providing the neutron research facilities and support for the neutron scattering experiments. The assistance of D. Iampietro and L. Ryan in the neutron scattering studies is gratefully acknowledged. We are thankful to P. Pjura for useful discussions and for the introduction to the centrifugal crystallization, to S. Leonard and A. Gittis for x-ray diffraction characterization of the lysozyme crystals, and to D. Asthagiri for kindly providing the program for the protein charge and dipole moment calculations.

This study was supported by grants from the National Science Foundation (BES-9510420) and NASA (NAG8–1346).

REFERENCES

Alderton, G., W. H. Ward, and H. L. Fevold. 1945. Isolation of lysozyme from egg white. *J. Biol. Chem.* 157:43–58.

Artymiuk, P. J., C. C. F. Blake, D. W. Rice, and K. S. Wilson. 1982. The structures of the monoclinic and orthorhombic forms of hen egg-white lysozyme at 6 Å resolution. *Acta Crystallogr. B.* 38:778–783.

Ataka, M., and T. Shoji. 1986. The growth of single crystals of lysozyme. *Biopolymers.* 25:337–350.

Barlow, D. J., and J. M. Thornton. 1986. The distribution of charged groups in proteins. *Biopolymers.* 25:1717–1733.

Belloni, L. 1991. Interacting monodisperse and polydisperse spheres. *In* Neutron, X-Ray and Light Scattering. P. Linder and Th. Zemb, editors. Elsevier Science, Amsterdam. 135–155.

Blomberg, E., P. M. Claesson, J. C. Fröberg, and R. D. Tilton. 1994. Interaction between adsorbed layers of lysozyme studied with the surface force technique. *Langmuir.* 10:2325–2334.

Bonnete, F., M. Malfois, S. Finet, A. Tardieu, S. Lafont, and S. Veesler. 1997. Different tools to study interaction potentials in γ-crystallin solutions—relevance to crystal growth. *Acta Crystallogr. D.* 53:438–447.

Boue, F., F. Lefaucheux, M. C. Robert, and I. Rosenman. 1993. Small angle neutron scattering study of lysozyme solutions. *J. Crystal Growth.* 133:246–254.

Chen, S.-W., and D. Bendedouch. 1986. Structure and interactions of proteins in solution studied by small angle neutron scattering. *Methods Enzymol.* 130:79–116.

Coen, C. J., H. W. Blanch, and J. M. Prausnitz. 1995. Salting out of aqueous proteins: phase equilibria and intermolecular potentials. *AIChE J.* 4:996–1004.

Coumou, D. J. 1960. Apparatus for the measurement of light scattering in liquids: measurement of the Raylegh factor of benzene and of some other pure liquids. *J. Colloid Sci.* 15:408–417.

Coumou, D. J., E. L. Mackor, and J. Hijmans. 1964. Isotropic light-scattering in pure liquids. *Trans. Faraday Soc.* 60:1539–1547.

Creighton, T. E. 1993. Proteins: Structure and Molecular Properties. W. H. Freeman and Co., New York.

Delaye, M., and A. Tardieu. 1983. Short-range order of crystallin proteins accounts for eye lens transparency. *Nature.* 302:415–417.

Ducruix, A., J. P. Guilloteau, M. Ries-Kautt, and A. Tardieu. 1996. Protein interactions as seen by solution x-ray scattering prior to crystallogenesis. *J. Crystal Growth.* 168:28–39.

Durbin, S. D., and G. Feher. 1996. Protein crystallization. *Annu. Rev. Phys. Chem.* 47:171–204.

Eberstein, W., Y. Georgalis, and W. Saenger. 1994. Molecular interactions in crystallizing lysozyme solutions studied by photon correlation spectroscopy. *J. Crystal Growth.* 143:71–78.

Florin, E.-L., V. T. Moy, and H. E. Gaub. 1994. Adhesion forces between individual ligand-receptor pairs. *Science.* 264:415–417.

Fredericks, W. J., M. C. Hammonds, S. B. Howard, and F. Rosenberger. 1994. Density, thermal expansivity, viscosity and refractive index of lysozyme solutions at crystal growth concentrations. *J. Crystal Growth.* 141:183–192.

Georgalis, Y., J. Schüller, J. Frank, M. D. Soumpasis, and W. Saenger. 1995. Protein crystallization screening through scattering techniques. *Adv. Colloid Interface Sci.* 58:57–86.

George, A., Y. Chiang, B. Guo, A. Arabshahi, Z. Cai, and W. W. Wilson. 1997. Second virial coefficient as predictor in protein crystal growth. *Methods Enzymol.* 276:100–110.

George, A., and W. Wilson. 1994. Predicting protein crystallization from a dilute solution property. *Acta Crystallogr. D.* 50:361–365.

Giordano, R., A. Grasso, J. Teixeira, F. Wanderlingh, and U. Wanderlingh. 1991. Small-angle neutron scattering in lysozyme solutions. *Phys. Rev. A.* 43:6894–6899.

Giordano, R., A. Grasso, J. Teixeira, F. Wanderlingh, and U. Wanderlingh. 1992. SANS on lysozyme and lactoglobulin solutions. *Phys. A.* 180: 762–764.

Glasoe, P. K., and F. A. Long. 1960. Use of glass electrodes to measure acidities in deuterium oxide. *J. Phys. Chem.* 64:188–190.

Glatter, O., and O. Kratky. 1982. Small Angle X-Ray Scattering. Academic Press, London. 18–32.

Gold, A. M., and D. Fahrney. 1964. Sulfonyl fluorides as inhibitors of esterases. II. Formation and reactions of phenylmethanesulfonyl α-chymotrypsin. *Biochemistry.* 3:783–791.

Goodwin, J. W. 1981. Concentrated dispersions. *In* Colloidal Dispersions. J. W. Goodwin, editor. The Royal Society of Chemistry, London.

Gripon, C., L. Legrand, I. Rosenman, F. Boue, and C. Regnaut. 1998. Relation between the solubility and the effective solute-solute interaction for C60 solutions and lysozyme solutions. *J. Crystal Growth.* 183:258–268.

Gripon, C., L. Legrand, I. Rosenman, O. Vidal, M-C. Robert, and F. Boue. 1996. Study of protein-protein interactions in undersaturated and supersaturated lysozyme solutions in heavy water as a function of temperature. *C. R. Acad. Sci. Paris.* 322(IIb):565–571.

Gripon, C., L. Legrand, I. Rosenman, O. Vidal, M. C. Robert, and F. Boue. 1997. Lysozyme-lysozyme interactions in under- and super-saturated solutions: a simple relation between the second virial coefficients in H_2O and D_2O. *J. Crystal Growth.* 178:575–584.

Haynes, C. A., K. Tamura, H. R. Körfer, H. W. Blanch, and J. M. Prausnitz. 1992. Thermodynamic properties of aqueous α-chymotrypsin solutions from membrane osmometry measurements. *J. Phys. Chem.* 96:905–912.

Howard, S. B., P. J. Twigg, J. K. Baird, and E. J. Meehan. 1988. The solubility of hen-egg lysozyme. *J. Crystal Growth.* 90:94–104.

Hunter, R. J. 1989. Foundations of Colloid Science, Vol. II. Calderon Press, Oxford. 675–873.

Kaler, E. 1995. Small-angle scattering from complex fluids. *In* Modern Aspects of Small-Angle Scattering. H. Brumberger, editor. Kluwer Academic Publishers, Dordecht. 239–353.

Lafont, S., S. Veesler, J. P. Astier, and R. Boistelle. 1997. Comparison of solubilities and molecular interactions of BPTI molecules giving different polymorphs. *J. Crystal Growth.* 168:132–140.

Leckband, D., and J. Israelachvili. 1993. Molecular basis of protein function as determined by direct force measurements. *Enzyme Microb. Technol.* 15:450–538.

Lenhoff, A. M., P. E. Pjura, J. G. Dilmore, and T. S. Godlewski, Jr. 1997. Ultracentrifugal crystallization of proteins: transport-kinetic modeling, and experimental behavior of catalase. *J. Crystal Growth.* 180:113–126.

Lomakin, A., N. Asherie, and G. B. Benedek. 1996. Monte Carlo study of phase separation in aqueous protein solutions. *J. Chem. Phys.* 104: 1646–1656.

Malfois, M., F. Bonnete, L. Belloni, and A. Tardieu. 1996. A model of attractive interactions to account for fluid-fluid phase separation of protein solutions. *J. Chem. Phys.* 105:3290–3300.

Muschol, M., and F. Rosenberger. 1995. Interactions in undersaturated and supersaturated lysozyme solutions: static and dynamic light scattering results. *J. Chem. Phys.* 107:1953–1962.

Muschol, M., and F. Rosenberger. 1996. Lack of evidence for prenucleation aggregate formation in lysozyme crystal growth solutions. *J. Crystal Growth.* 167:738–747.

Muschol, M., and F. Rosenberger. 1997. Liquid-liquid phase separation in supersaturated lysozyme solutions and associated precipitate formation/crystallization. *J. Chem. Phys.* 167:738–747.

Neal, B. L., D. Asthagiri, and A. M. Lenhoff. 1998. Molecular origins of osmotic second virial coefficients of proteins. *Biophys. J.* (in press).

Niimura, N., Y. Minezaki, M. Ataka, and T. Katsura. 1994. Small angle neutron scattering from lysozyme in unsaturated solutions, to characterize the pre-crystallization process. *J. Crystal Growth.* 137:671–675.

Niimura, N., Y. Minezaki, M. Ataka, and T. Katsura. 1995. Aggregation in supersaturated lysozyme solution studied by time-resolved small angle neutron scattering. *J. Crystal Growth.* 154:136–144.

Retailleau P., M. Ries-Kautt, and A. Ducruix. 1997. No salting-in of lysozyme chloride observed at low ionic strength over a large range of pH. *Biophys. J.* 73:2156–2163.

Roth, C. M., B. L. Neal, and A. M. Lenhoff. 1996. Van der Waals interactions involving proteins. *Biophys. J.* 70:977–987.

Rosenbaum, D., P. C. Zamora, and C. F. Zukoski. 1996. Phase behavior of small attractive colloidal particles. *Phys. Rev. Lett.* 76:150–153.

Rosenbaum, D. F., and C. F. Zukoski. 1996. Protein interactions and crystallization. *J. Crystal Growth.* 168:752–758.

Rosenberger, F. 1996. Protein crystallization. *J. Crystal Growth.* 168: 40–54.

Rosenberger, F., P. G. Vekilov, M. Muschol, and B. R. Thomas. 1996. Nucleation and crystallization of globular proteins—what do we know and what is missing? *J. Crystal Growth.* 168:1–27.

Skouri, M., J-P. Munch, B. Lorber, R. Giege, and S. Candau. 1992. Interactions between lysozyme molecules under precrystallization conditions studied by light scattering. *J. Crystal Growth.* 122:14–20.

Smith, M. H. 1970. Peptides and proteins. *In* Handbook of Biochemistry. H. A. Sober, editor. The Chemical Rubber Co., Cleveland, OH. Sect. C, 71–98.

Sophianopoulos, A. J., C. K. Rhodes, D. N. Holcomb, and K. E. van Holde. 1962. Physical studies of lysozyme. I. Characterization. *J. Biol. Chem.* 237:1107–1112.

Stuhrmann, H. B., and H. Fuess. 1976. A neutron small-angle scattering study of hen egg-white lysozyme. *Acta Crystallogr. A.* 32:67–74.

Stryer, L. 1988. Biochemistry. W. H. Freeman and Co., New York.

Svergun, D. I., S. Richard, M. H. J. Koch, Z. Sayers, S. Kuprin, and G. Zaccai. 1998. Protein hydration in solution: experimental observation by X-ray and neutron scattering. *Proc. Natl. Acad. Sci. USA.* 95:2267–2272.

Takashima, S., and K. Asami. 1993. Calculation and measurement of the dipole moment of small proteins: use of protein data base. *Biopolymers.* 33:59–68.

Tanaka, S., M. Yamamoto, K. Kawashima, K. Ito, R. Hayakawa, and M. Ataka. 1996. Kinetic study on the early stage of the crystallization process of two forms of lysozyme crystals by photon correlation spectroscopy. *J. Crystal Growth.* 168:44–49.

ten Wolde, P. R., and D. Frenkel. 1997. Enhancement of protein crystal nucleation by critical density fluctuations. *Science.* 277:1975–1978.

Utiyama, H. 1972. Calibration and correction factors. *In* Light Scattering from Polymer Solutions. M. B. Huglin, editor. Academic Press, London.

Veretout, F., M. Delaye, and A. Tardieu. 1989. Molecular basis of eye lens transparency. Osmotic pressure and x-ray analysis of α-crystallin solutions. *J. Mol. Biol.* 205:713–728.

Vilker, V. L., C. K. Colton, and K. A. Smith. 1980. The osmotic pressure of concentrated protein solutions: effect of concentration and pH in saline solutions of bovine serum albumin. *J. Colloid Interface Sci.* 79:548–566.

Zimm, B. H. 1948. The scattering of light and the radial distribution function of high polymer solutions. *J. Chem. Phys.* 16:1093–1099.