ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Structural Biology 142 (2003) 56–65

Journal of Structural Biology

www.elsevier.com/locate/yjsbi

Light scattering as a diagnostic for protein crystal growth—A practical approach

W. William Wilson*

Department of Chemistry, Mississippi State University, Box 9573, Mississippi State, MS 39762, USA

Received 5 February 2003

Abstract

Static and dynamic light scattering are discussed as particularly useful tools for studying various aspects of protein crystal growth. Specific applications for prenucleation assays as well as for monitoring postnucleation growth processes are presented. Protein–protein interactions determined by light scattering, which serve as a predictor for favorable crystallization conditions as well as for protein solubility behavior, are detailed. Several precautions regarding the practical aspects of light scattering and interpretation of data are also discussed.

Keywords: Dynamic light scattering; Static light scattering; Protein crystal growth

1. Introduction

The term ''protein crystallogenesis'' has been used to describe the various aspects of the conversion of individual protein molecules in solution to a macroscopic crystal suitable for diffraction studies which ultimately lead to a three-dimensional structure for the protein. In a general sense, the entire process can be divided into a protein crystal growth (PCG) stage followed by a protein crystal analysis (PCA) stage. This natural division has fostered distinct research communities with different motives regarding protein crystallogenesis. The PCG-ers include researchers who strive to understand the basic physicochemical aspects of PCG, such as ''how'' do protein crystals grow (nucleation and postnucleation kinetics) as well as ''why'' do specific solution conditions lead to crystallization and others not (thermodynamics of protein–protein interactions). On the other hand, many PCA-ers, especially those in the commercial sector, admittedly discount the ''how'' and ''why'' in favor of high-throughput screening technologies that may produce ''hits'' in a relatively short time while consuming minimum protein. This approach is understandable from an economic standpoint and in view of the fact that most proteins ever crystallized were done so using a screening method. However, the human proteome dictated by about 30 000 genes will yield an avalanche of 100 000 to 300 000 proteins (Kempner and Felder, 2002), most with unknown structure and function. Many of these ''new'' proteins will defy crystallization using traditional trial-and-error screening methods, so it will be incumbent upon researchers to develop a different, more fundamental, strategy that will improve the probability of successful crystallization. This paper describes light scattering as an analytical diagnostic that has proven useful for understanding particular fundamental aspects of PCG. Specifically, static light scattering (SLS) and dynamic light scattering (DLS) will be presented as tools that PCG-ers can utilize as a predictor and monitor during both prenucleation and postnucleation growth studies. Although not a review for every application of SLS and DLS for PCG, it is hoped that this summary will provide the reader a useful starting point for further inquiry.

2. Methods

2.1. Static light scattering

The basic SLS experiment measures the average intensity of light scattered by a protein solution of

* Fax: 1-662-325-1618.
E-mail address: bilwil@ra.msstate.edu.

1047-8477/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S1047-8477(03)00038-8

defined concentration in excess of that scattered by background sources (solvent, crystallizing agents, stray light, etc.). A complete description of the SLS experiment and theory is given by Kratochvil (1987). Several versions of SLS instrumentation are commercially available (see individual Websites). Alternatively, with current technology for solid-state lasers and detectors, the motivated researcher can easily assemble an adequate apparatus.

As the term implies, SLS (also referred to as classical or Rayleigh light scattering) can be used to obtain the so-called 'static' parameters of the protein such as weight-average molecular weight, second virial coefficient, and sometimes the radius of gyration. The working equation appropriate for most protein molecules is

$$\frac{Kc}{R_{90}} = \frac{1}{M_w} + 2Bc + \cdots, \tag{1}$$

where K is an optical constant given by

$$K = \frac{4\pi^2 n^2 (dn/dc)^2}{N_A \lambda^4} \tag{2}$$

and n is the refractive index of the solvent, dn/dc is the refractive index increment for the protein/solvent pair ($\mathrm{ml \cdot g^{-1}}$), N_A is Avogadro's number ($\mathrm{mol^{-1}}$), λ is the wavelength of the incident vertically polarized light in vacuum (cm), R_{90} is the excess Rayleigh factor at a scattering angle of 90° ($\mathrm{cm^{-1}}$), c is the concentration of the protein ($\mathrm{g \cdot ml^{-1}}$), M_w is the weight-average molecular weight of the protein ($\mathrm{g \cdot mol^{-1}}$), and B is the second virial coefficient ($\mathrm{mol \cdot ml \cdot g^{-2}}$). In order to determine the absolute R_{90} values for the protein solutions, the SLS instrument can be calibrated using toluene, which has an established R_{90} value of $14.06 \times 10^{-6}\,\mathrm{cm^{-1}}$ at 633 nm. Eq. (1) indicates that a plot of Kc/R_{90} versus c allows the determination of M_w (usually independently known for a carefully purified protein) and B, discussed below. Note that the radius of gyration, R_G, is not normally determined by SLS since most protein molecules are smaller than $\lambda/20$ (~30 nm), and there is no measurable angular dependence of the scattered light intensity, a requirement for R_G measurement.

For PCG studies, clarification of small sample volumes (≤10 μl) by filtration is important to minimize protein consumption. A low swept volume (~0.8 μl) filter housing (Upchurch, A318) combined with 0.2- or 0.1-μm, pore size membrane filters (Nuclepore, 13 mm, polycarbonate) is usually adequate for dust removal by a single-pass filtration. To the novice, this step may seem insignificant compared to the effort expended to isolate and purify the protein; however, frustration will undoubtedly be the reward to those with a casual hand. The old adage that "light scatterers die a dusty death" has been realized by more than one practitioner of this art.

2.2. Dynamic light scattering

Since its introduction in the 1960s, DLS (also referred to as quasi-elastic light scattering or photon correlation spectroscopy) has become the more popular (if not trendy) of the light scattering methods. DLS is more forgiving since fluctuations about the average intensity are measured rather than the absolute average intensity as required for SLS. Experimental and theoretical aspects of DLS are well documented (Berne and Pecora, 1976; Brown, 1993; Cummins and Pike, 1973, 1976; Degiorgio et al., 1980; Earnshaw and Steer, 1983; Johnson and Gabriel, 1994; Schmitz, 1990). The scattered intensity fluctuations are primarily due to the random diffusive motion of the protein molecules into and out of a focused laser beam. One way to mathematically describe the time variation of such a random physical process is via a correlation function, and for DLS the appropriate form for the experimental intensity correlation function, $C(\tau)$ is (Cummins, 1983)

$$C(\tau) = B\left[1 + a(C'(\tau))^2\right], \tag{3}$$

where B is the background, a is a constant which depends on the spatial coherence of the detected scattered light, and $C'(\tau)$ is called the normalized scattered electric field correlation function. For a singly diffusing protein species,

$$C'(\tau) = e^{-\Gamma\tau}, \tag{4}$$

where Γ is the decay constant which contains the protein molecular information via $\Gamma = D(c)q^2$, $D(c)$ being the concentration-dependent translational diffusion coefficient of the protein and q the scattering vector given by

$$q = \frac{4\pi n}{\lambda} \sin\frac{\theta}{2}, \tag{5}$$

where θ is the scattering angle, usually 90°. Larger values for $D(c)$ indicate smaller particle size and cause the correlation function to decay rapidly, whereas smaller values for $D(c)$ correspond to larger particles and result in a slowly decaying function. Since globular proteins are often assumed to be spherical in shape, the Stokes–Einstein relation can be used to estimate the apparent hydrodynamic radius, R_h(app) of the protein molecule by

$$R_h(\mathrm{app}) = \frac{kT}{6\pi\eta D(c)}, \tag{6}$$

where k is the Boltzmann constant, T is the absolute temperature, and η is the solution viscosity. Thus, analysis of $C(\tau)$ from DLS provides an appealing way of estimating the apparent size of the protein solute.

The diffusion coefficient has a protein concentration dependence usually described by (Altenberger and Deutch, 1973; Anderson and Reed, 1976; Hess and Klein, 1976; Phillies, 1975)

$$D(c) = D_0(1 + k_D c + \cdots), \tag{7}$$

where D_0 is the infinite dilution value for $D(c)$ and k_D is called the diffusion virial coefficient. The true equivalent spherical hydrodynamic radius of the protein molecules can be estimated using D_0 in

$$R_h = \frac{kT}{6\pi\eta D_0}. \tag{8}$$

3. Prenucleation assays

3.1. State of aggregation of the protein stock

It has been established (Bhamidi et al., 1999; Skouri et al., 1995; Thibault et al., 1992; Thomas et al., 1998a,b; Veesler et al., 1994) that both charge heterogeneity and size heterogeneity are detrimental to PCG experiments. A common way to assay for protein purity is SDS (sodium dodecyl sulfate)–PAGE (polyacrylamide gel electrophoresis). The method is quite sensitive for detecting charge heterogeneity in a protein sample, which is usually interpreted as due to the presence of proteins different from the target protein. However, since SDS denatures the protein molecules, information about the size heterogeneity of the target protein is lost. DLS provides an attractive alternative for quickly assaying a protein solution for size heterogeneity. Three cases which commonly occur for protein solutions are (I) the protein in its pure monomeric state (or whatever its natural single-sized state may be), (II) the protein monomer accompanied by lower order oligomers (dimers, trimers, etc.), and (III) the protein monomer with a relatively low concentration of significantly higher order protein aggregates. Case (I) presents the least complicated analysis for $C(\tau)$, which from Eqs. (3) and (4) becomes

$$C(\tau) = B[1 + ae^{-2\Gamma\tau}]. \tag{9}$$

One way of obtaining $D(c)$ from $C(\tau)$ is by linear regression on

$$\ln\left[\frac{C(\tau) - B}{B}\right] = \ln a - 2\Gamma\tau, \tag{10}$$

from which the slope gives Γ and thus $D(c)$, which is subsequently used to estimate R_h(app) via Eq. (6).

Cases (II) and (III) are both described as solutions that are polydispersed (sometimes the term paucidispersed is used when a few discrete molecular-sized species are present). The expression for $C'(\tau)$ in Eq. (4) must be modified to account for the distribution of different species,

$$C'(\tau) = \int_0^\infty f(\Gamma)e^{-\Gamma\tau}\,d\Gamma, \tag{11}$$

where $f(\Gamma)d\Gamma$ represents the fraction of the scattered light intensity due to protein species having Γ in the range Γ to $\Gamma + d\Gamma$. The problem then becomes devising a way of extracting the distribution $f(\Gamma)$ from the experimentally measured $C(\tau)$. Solving for the distribution of Γ's gives the distribution of $D(c)$'s, which is then used to estimate the distribution in R_h(app)'s. A general review of various methods used to analyze $C(\tau)$ from DLS has been given by Chu (1983). One approach that is often used for analyzing $C(\tau)$ is the method of moments or cumulants (Koppel, 1972), for which Eq. (4) is modified to give

$$\ln C'(\tau) = -\langle\Gamma\rangle\tau + \frac{\mu_2}{2!}\tau^2 - \frac{\mu_3}{3!}\tau^3 + \cdots, \tag{12}$$

where $\langle\Gamma\rangle$ is the average decay constant and the coefficients in the series are called cumulants. A quantity that is often used to specify the degree of polydispersity (referred to as the "poly" value) is the normalized variance defined as $\mu_2/\langle\Gamma\rangle^2$. For a carefully prepared monomeric protein solution (Case (I)), the measured poly should be $\leqslant 0.01$, and the PCG-er can be reasonably sure that the protein sample has essentially no size heterogeneity. For solutions described by cases (II) and (III) above, the poly may typically be 0.05–0.5, depending on the extent of size heterogeneity. Although the poly is quite sensitive to the presence of size heterogeneity, its value is nondescriptive of the species actually present in solution.

An alternative approach for estimating size heterogeneity is to analyze $C(\tau)$ using one or more of a host of particle size distribution (PSD) algorithms developed specifically for DLS applications (Goldin, 1991; Stepenek, 1993). Commercial DLS apparatus will generally offer at least one version of a PSD analysis as part of the operating software. Typical outputs of PSD analysis include the relative abundance (number, weight, and intensity averages) of different protein species as a function of particle size (nm). In addition, a "variance" (analogous to poly) is reported as a measure of sample size heterogeneity.

For the primary purpose of using DLS to improve PCG experiments, it is not necessary to employ overly sophisticated data reduction schemes. As an assay for the presence of size heterogeneity in the protein sample, the cumulants method of data analysis, which gives an average value for the diffusion coefficient and a measure of the variance of the distribution, is most often adequate.

3.2. Precautions

Using PSD results from batch DLS experiments to report absolute states of protein aggregation is risky and should be done with caution. For solutions of type (III) described above, the difference in diffusion coefficients of the protein monomer and large aggregates is often great enough that a reasonable estimate of size distribution is

obtainable. As a rule of thumb, the size difference between two species needs to be about a factor of 5 or greater to obtain reliable resolution of PSD by batch DLS. For solutions of type (II), PSD analysis by DLS is insufficiently accurate to resolve lower order oligomerization by batch experiments. This shortcoming of DLS is often not a major issue for PCG since it is the mere presence of aggregates that is desirable to know rather than their actual size distribution. If quantitative resolution of type (II) solutions is required, then the best way is to couple SLS and DLS as online detectors with analytical size-exclusion chromatography (Casay, 1986; Helfrich and Jones, 1999; Jones, 2002). This approach resolves the different-sized protein species before light scattering detection, simplifying data analysis.

When performing prenucleation studies by DLS, it is important to ascertain whether there is a contribution to $C(\tau)$ by components of the solvent. Some crystallizing agents such as PEGs and higher concentrations (1.5–2.5 M) of ammonium sulfate and sodium potassium tartrate, for example, will by themselves give a well-defined decay constant, Γ_s. When protein monomer is dissolved in such a solution, the resulting $C(\tau)$ has the form (Dubin, 1970)

$$C(\tau) = \underset{(I)}{\langle I_s\rangle^2 e^{-2\Gamma_s\tau}} + \underset{(II)}{2\langle I_s\rangle\langle I_p\rangle e^{-(\Gamma_s+\Gamma_p)\tau}} + \underset{(III)}{\langle I_p\rangle^2 e^{-2\Gamma_p\tau}} + \text{constant}, \tag{13}$$

where $\langle I_s\rangle$ and $\langle I_p\rangle$ are the total intensities of light scattered by solvent components and protein monomer, respectively. In addition to the individual contributions from the solvent components (I) and protein monomer (III), there is a cross term (II) due to self-beating in homodyne detection employed by DLS. Analysis of the $C(\tau)$ by cumulants analysis gives an R_h(app) that is a weighted average of the three contributions and a poly value that can easily be misinterpreted as due to a nonmonomeric state of the protein. PSD analysis of $C(\tau)$, in our experience, often gives a too-small R_h(app) value due to solvent components and a too-large R_h(app) value corresponding to the protein which, once again, can be misinterpreted as an indication of protein aggregation. One approach to avoid this ambiguity is to independently measure the $C(\tau)$ from the solvent alone to obtain a value for Γ_s and then fit the protein solution $C(\tau)$ according to Eq. (13) with Γ_s held constant. Fig. 1 shows the results of such an approach for equine serum albumin (ESA) monomer with ammonium sulfate (AS) as the crystallizing agent. The R_h(app) values determined via Eqs. (13) and (6) as a function of AS concentration clearly show that a near-monomeric state of the ESA is preserved, whereas PSD analysis indicates aggregation. It is always judicious to check the $C(\tau)$ from solvent components when performing DLS measurements.

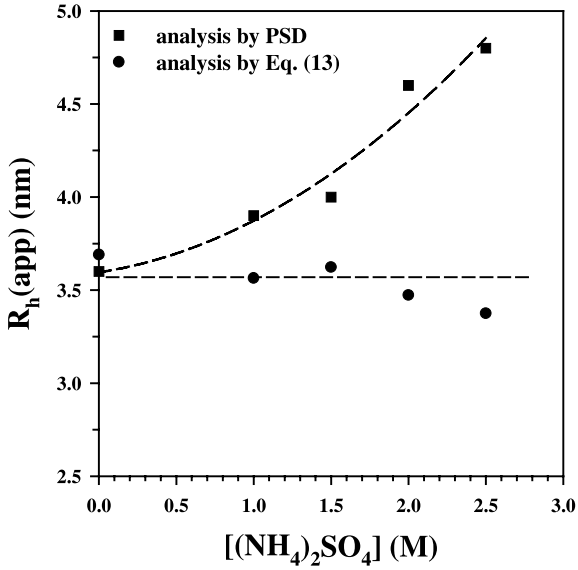

Fig. 1. R_h(app) versus AS concentration for ESA in 0.1 M NaAc buffer, pH 5.6, 22 °C using two methods of data analysis.

3.3. Protein–protein interactions and the crystallization slot

It is well established both experimentally and theoretically (Bonnete et al., 1999; Ducruix et al., 1996; George et al., 1997; George and Wilson, 1994; Malfois et al., 1996; Neal et al., 1998; Neal et al., 1999; Rosenbaum et al., 1996) that a range of values of the second virial coefficient, B, in Eq. (1) is correlated with solution conditions that are favorable to crystallization. The "crystallization slot," shown in Fig. 2, corresponds to B values in the range of about -1×10^{-4} to -8×10^{-4} ($\text{mol}\cdot\text{ml}\cdot\text{g}^{-2}$) and indicates protein–protein interactions that are slightly to moderately attractive. PCG

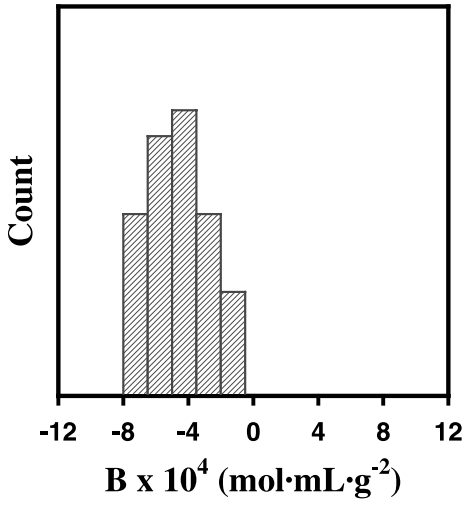

Fig. 2. The crystallization slot representing solution conditions, as described by B, favorable for crystallization.

experiments conducted in solution conditions at more negative B values have a greater risk of forming an amorphous solid phase because of correspondingly stronger protein–protein attractions. On the other hand, experiments at more positive values, at which the net protein–protein interactions are repulsive, typically require protein concentrations that are impractically high to cause phase separation of any kind. The crystallization slot can be used as an effective guide by PCG-ers to direct changes in a particular solution parameter (pH, temperature, crystallization agent concentration, etc.) that will increase the probability of a successful crystallization trial. An example of utilizing the crystallization slot for thaumatin I is shown in Fig. 3. The slot is superimposed on a plot of B versus crystallization agent (NaK tartrate) concentration and defines a range of NaK tartrate concentrations appropriate for crystallization at the specified temperature and other solution conditions. For thaumatin I and other proteins, it has been found that conducting PCG experiments at the less negative end of the crystallization slot typically results in fewer, larger crystals while experiments at the more negative end of the slot give a larger number of smaller crystals, as shown in Fig. 4. It is important to point out that conducting PCG experiments under conditions that correspond to the crystallization slot does not guarantee a successful crystallization trial. However, working under conditions outside the slot most assuredly reduces the probability of a desirable outcome.

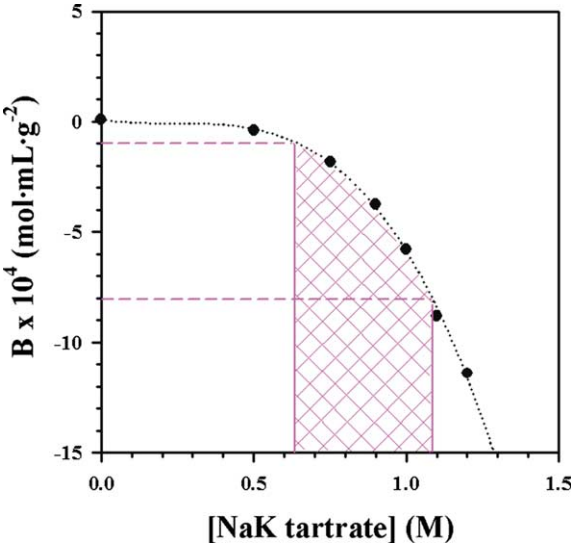

Fig. 3. B versus NaK tartrate concentration for thaumatin I in 0.05 M ADA buffer, pH 6.5, 22 °C, with crystallization slot overlay.

A promising application of the crystallization slot is its utilization for crystallizing membrane-bound proteins. A report by Hitscherich et al. (2000) showed that tetragonal and trigonal crystals of OmpF porin (a bacterial outer membrane protein) formed within a narrow range of B values that was within the crystallization slot for soluble proteins. The implication of this work is that efforts to crystallize membrane proteins should be directed toward placing the protein–detergent complexes within the crystallization slot as defined by the second virial coefficient. Loll et al. (2002) performed B measurements on protein-free detergent micelles and found that the B behavior is a good predictor for the behavior of the protein–detergent complexes under similar solution conditions.

The direct and most often reported way of determining B values to compare with the crystallization slot is by SLS using Eq. (1), in which scattered intensity versus protein concentration data are collected. Alternative methods for predicting crystallization from light scattering experiments have been investigated and involve measuring the diffusion coefficient by DLS versus protein concentration. One approach (Feher and Kam,

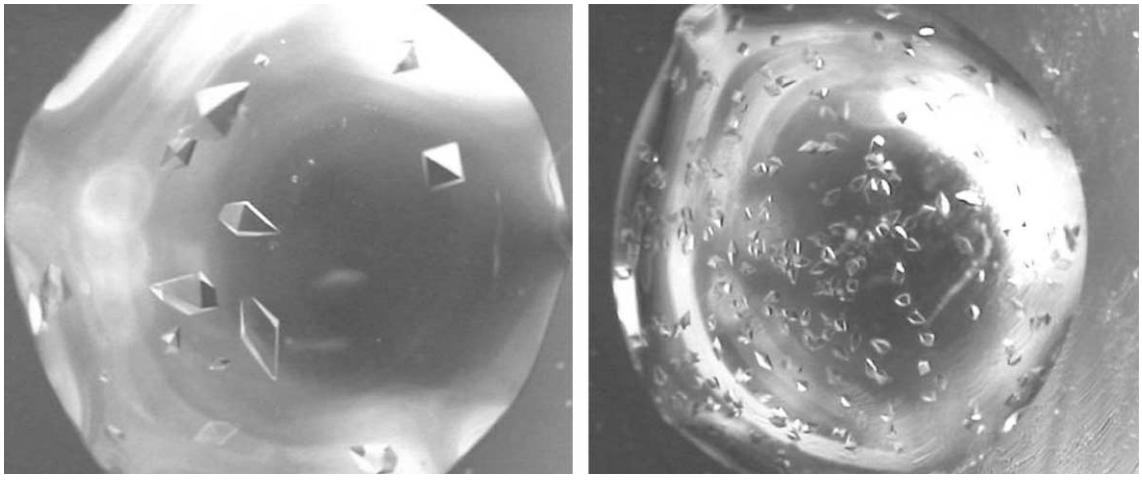

Fig. 4. Thaumatin I crystals grown in 0.5 M ADA buffer, pH 6.5, 22 °C, with 0.75 (left) and 1.0 M (right) NaK tartrate as crystallizing agent and with $B = -1.8 \times 10^{-4}$ and -5.7×10^{-4} $(\text{mol} \cdot \text{ml} \cdot \text{g}^{-2})$, respectively.

1985; Kam et al., 1978) used lysozyme to develop a thermodynamic model which distinguished between praggs and craggs (Wilson, 1990), which are amorphous aggregates that ultimately grow to form a precipitate and highly structured aggregates that eventually form nucleation centers for crystal growth, respectively. Details of the method are beyond the scope of this paper, but the main requirement was to measure the diffusion coefficient of the protein in the presence of a crystallizing agent as a function of protein concentration. A set of so-called dilution curves was generated by calculating a mean value for $D(c)$ for a particular size distribution of protein aggregates. The experimental dilution curve was compared to the calculated set of dilution curves and a judgment made as to whether the aggregates formed in a particular solution condition were praggs or craggs.

Other researchers have attempted to correlate B with the diffusion virial coefficient, k_D (see Eq. (7)) by measuring $D(c)$ as a function of c for proteins in various crystallizing agents (Chiang, 1994; Zukoski et al., 2002). Theoretical considerations (Cummins and Pike, 1973) show that k_D comprises a thermodynamic term, which incorporates B, and a concentration-dependent frictional coefficient term. It is the interplay between these two terms that determines the overall concentration dependence of D in monodisperse protein solutions, and although B and k_D are certainly correlated, a general analytical model allowing accurate values of B to be extracted from k_D has not been presented.

3.4. Precaution

The value for k_D in Eq. (7) can be positive, negative, or near zero, depending on the particular properties of the protein solution. For proteins dissolved in a crystallizing agent, k_D is usually negative so that when the DLS correlation function is analyzed either by cumulants or by PSD algorithms, the R_h(app) presented will be larger than anticipated for the protein monomer (because $D(c)$ is considerably smaller than D_0). The larger R_h(app) can be misinterpreted as unequivocal evidence of aggregation of the protein, when, in fact, it is not.

3.5. Protein–protein interactions and solubility behavior

Protein solutions used to measure B values by SLS are typically well undersaturated, usually having concentrations in the 1–5 mg/ml range. However, results presented by Muschol and Rosenberger (1995) for lysozyme with NaCl as the crystallizing agent showed that the slope of the Kc/R versus c plot (from which B is determined) was essentially constant for undersaturated solutions (<10 mg/ml) to well beyond saturation (>60 mg/ml). This finding implied that the protein–protein pair potentials manifest in undersaturated (even dilute) protein solutions extend into supersaturation regions. Subsequent studies (Rosenbaum and Zukoski, 1996) showed that the lysozyme phase boundary could be constructed using an adhesive hard sphere potential along with B values to model protein interactions. Experimental data correlating B with about 25 solubility (s) values for lysozyme in various crystallizing agents along with additional data for ovalbumin, for ESA, and for lysozyme in H_2O and D_2O have been reported (Demoruelle et al., 2002; George et al., 1997; Gripon et al., 1997; Guo et al., 1999). The empirical correlation between B (a dilute solution parameter) and s (a phase transition parameter) was initially met with considerable consternation by some in the PCG and PCA communities. After all, solubility depends on the binding energy between closely spaced protein molecules having very specific mutual orientations in the crystal lattice. By contrast, the second virial coefficient is a statistical average of interactions between two protein molecules in the liquid phase accounting for all relative separation distances and orientations, with each configuration weighted by a Boltzmann factor. Nevertheless, the experimental data demanded that a simple theoretical relation between B and s must exist, and subsequent publications (Haas et al., 1999; Ruppert et al., 2001) provided the theoretical basis for the direct link between B and s. Major findings of these studies were that protein interactions were strongly anisotropic and that crystallization conditions had little effect on the interaction distance or the anisotropy between the protein molecules.

The correlation between B and s offers PCG-ers a distinct advantage for determining the solubility behavior of proteins. From the practical standpoint of setting up PCG trials, it is desirable to know how a protein's solubility depends on a particular solution parameter. For example, for any rational PCG approach, it is important to know if and to what extent the protein's solubility has normal, retrograde, or no temperature dependence. In other cases, knowledge of the dependence of s on solution pH, ionic strength, type and concentration of crystallization agent, etc., will allow systematic PCG trials to be conducted. Since it is impractical to try and determine absolute values for s by equilibrium crystallization studies as a function of so many solution variables, the measurement of B by SLS provides a realistic alternative for obtaining the s behavior. One example demonstrating such an application is shown in Fig. 5, in which the temperature dependencies of B, and thus the same trends for s, are shown for lysozyme, ESA, and thaumatin I in their respective crystallizing agents. Lysozyme is found to have a strong normal $s(T)$, ESA has a moderate retrograde $s(T)$, and thaumatin I shows almost no $s(T)$ for the solution conditions tested. If absolute values for the protein solubility are required, then at least one (preferably two)

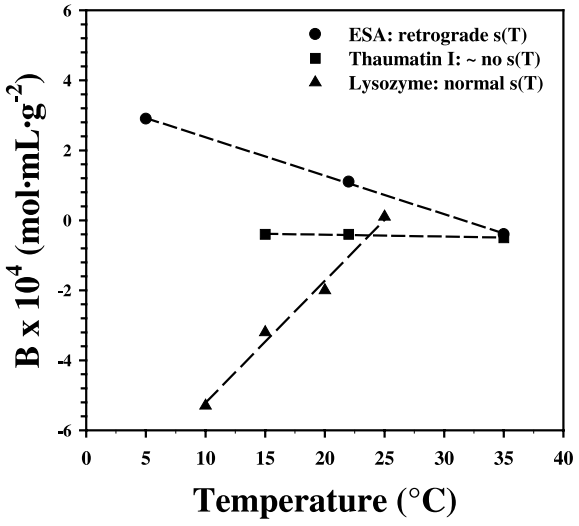

Fig. 5. B versus temperature for ESA in 0.1 M NaAc buffer, pH 5.6, with 0.5 M AS; thaumatin I in 0.05 M ADA buffer, pH 6.5, with 0.5 M NaK tartrate; and lysozyme in 0.1 M NaAc buffer, pH 4.2, with 2% NaCl (w/v), showing retrograde $s(T)$, approximately no $s(T)$, and normal $s(T)$, respectively.

absolute measurement of B and s for identical solution conditions must be made to establish the scale between the two parameters. Then, using the analytical expression relating B and s (Haas et al., 1999), a measured value for B can be used to calculate the corresponding value for s under essentially any solution condition desired.

4. Nucleation and postnucleation assays

In a typical PCG experiment, protein monomer is exposed to a crystallizing agent(s) using one of many techniques (Ducruix and Giege, 1992; McPherson, 1999; Unge, 1999). At this point, the supersaturation, $\sigma = c/s$, becomes the important thermodynamic parameter. If $\sigma \leqslant 1$, then the protein solution will usually be stable for an extended period of time. If $\sigma > 1$ by initial preparation in a batch experiment, by a slow equilibration process (vapor diffusion, liquid–liquid diffusion, etc.), by a sting method (very spatially localized changes in T, pH, etc.), or by some other novel design, then after an induction period the conversion of protein monomer to protein aggregates will occur. The process can be represented as

$$A_j + A_k \underset{k^{\mathrm{b}}_{j+k}}{\overset{k^{\mathrm{f}}_{j,k}}{\rightleftarrows}} A_{j+k} \quad (j,k = 1,2,3), \quad K_{j,k} = \frac{K^{\mathrm{f}}_{j,k}}{K^{\mathrm{b}}_{j+k}}, \tag{14}$$

where A_j is an aggregate consisting of j monomers, $k^{\mathrm{f}}_{j,k}$ and k^{b}_{j+k} are the forward and backward rate constants, and $K_{j,k}$ is the equilibrium constant for the formation of an ordered aggregate (cragg) A_{j+k}. With time, as the craggs form and dissociate, a critical-sized one or more (nuclei) will form which will sustain their growth and ultimately form crystals, if σ is high enough. The reported applications of light scattering, primarily DLS, to study nucleation and postnucleation growth of proteins are numerous, each with its own unique findings and recommendations. See, for example, Kadima et al. (1990), Skouri et al. (1991), Wessel and Ricka (1998), Georgalis and Saenger (1999), Tanaka et al. (1999), Habel et al. (2001), Juarez-Martinez et al. (2001), Saridakis et al. (2002), and Vekilov et al. (2002). Two specific applications in which SLS or DLS have been used as monitors of the process are discussed below.

4.1. Determination of a critical size aggregate

With a proper choice of σ, aggregation of the protein can be induced as depicted by Eq. (14). A general approach for using DLS to study the process is to prepare protein solutions with variable σ and then perform DLS measurements to determine R_{h}(app) as a function of time. At low supersaturation conditions (metastable region), aggregation occurs, but no nuclei are formed so that crystallization does not proceed, even over long periods of time. The critical solution supersaturation necessary to obtain crystals varies with the protein. For higher supersaturation conditions, under which, after an induction period, aggregation eventually leads to crystals, PSD analysis of the time-dependent DLS data allows an estimate for the critical size aggregate, R_{c}(app). It has been found that R_{c}(app) depends on the value for σ (Malkin and McPherson, 1994), i.e., for pumpkin seed globulin $R_{\mathrm{c}}(\mathrm{app}) \sim 45\,\mathrm{nm}$ at $\sigma = 1.3$ and $\sim 64\,\mathrm{nm}$ at $\sigma = 1.2$ and for apoferritin $R_{\mathrm{c}}(\mathrm{app}) \sim 39\,\mathrm{nm}$ at $\sigma = 2.5$ and $\sim 51\,\mathrm{nm}$ at $\sigma = 2.1$. These results also allowed estimates of the interfacial free energies between a nucleus and a bulk solution. A similar study on lysozyme (Mikol et al., 1989) showed that for a particular value of σ the R_{c}(app) was about 3.2 nm, indicating that only a few lysozyme molecules composed the nucleus.

4.2. Automated dynamic control of PCG

Both SLS and DLS have good sensitivity (albeit low resolution) for the appearance of craggs in a supersaturated protein solution, and this feature has led researchers to devise PCG apparatus incorporating light scattering monitors. The general approach is to detect changes in the light-scattering signal (increase in average intensity or intensity spikes for SLS, increase in R_{h}(app) and poly for DLS) as a function of time after the PCG trial has begun. Such changes are interpreted as the onset of aggregation/nucleation. It is paradoxical that in order to have measurable changes in the SLS or DLS signal, one has to create in solution an undesirable situation, i.e., enough aggregates/nuclei to cause such a change. If the intent is to produce a very few single large crystals,

then all but a very few of the nuclei must be quickly dissociated before postnucleation growth occurs. This is best done by lowering σ, which can be accomplished by quickly adjusting a particular solution parameter. Operationally, the dynamically controlled protein crystal growth (DCPCG) apparatus detects aggregation/nucleation by light scattering and then automatically imposes a change in T, for example, that lowers σ enough that most craggs/nuclei dissociate. The few (it is hoped) remaining nuclei then proceed to grow, preferably under near constant σ conditions (Schall et al., 1996).

Several applications of DCPCG have been reported. A single optical fiber was used to deliver a laser beam directly into an ~40-μl droplet in a hanging-drop vapor-diffusion apparatus (Casay and Wilson, 1992). SLS and DLS were utilized to monitor the onset of aggregation of lysozyme, and T changes were used to reverse the process. Another device (Ansari et al., 1996) comprised two monomode optical fibers and two GRIN microlenses (one for laser beam delivery and one for detection of scattered light) to provide a compact and noninvasive backscatter (161.5°) probe for performing DLS measurements on several variations of a hanging-drop vapor-diffusion apparatus. In addition, the device was demonstrated as a DCPCG monitor for the prenucleation, nucleation, and postnucleation growth stages using temperature as a control parameter for lysozyme. A rather sophisticated DCPCG apparatus envisioned for microgravity studies using SLS as a monitor and either temperature or water evaporation rate as a control parameter has been reported by Bray et al. (1998).

The initial successes reported for DCPCG were predominately using temperature as a control parameter for lysozyme, and it was conjectured that this would be a general method for most other proteins as well. As indicated in Fig. 5, the $B(T)$ and $s(T)$ is quite large for lysozyme with NaCl as crystallizing agent, but this behavior is more of an exception than the norm. While many proteins may have some $s(T)$ dependence in particular crystallization agents, very few have large enough $s(T)$ to make them good candidates for DCPCG. Another practical difficulty with DCPCG for some proteins is that once nuclei are formed, they are difficult to dissociate even when σ is lowered because the kinetics of dissociation are so slow. Thus, the successful application of DCPCG for a particular protein depends on finding a solution control parameter (T, pH, etc.) that can be used to vary σ enough so that formation and disassociation of the protein nuclei can truly be controlled on a useful time scale.

5. Conclusions

Static and dynamic light scattering have been presented as analytical tools that have broad application in the arena of protein crystal growth. Recent studies from both SLS and DLS have had a major impact on uncovering the fundamentals of thermodynamic and kinetic aspects of PCG. Advances in technology have resulted in commercially available instrumentation that makes light scattering feasible even for researchers with limited or no experience in the field. As with any analytical method, especially for light scattering, it is prudent to learn the true capabilities of a method and to not rely on "black box" results that can be overinterpreted. As PCG experiments evolve to microchip formats, the innovative researcher will undoubtedly find new ways to incorporate light scattering as a noninvasive diagnostic.

Acknowledgments

I am especially grateful to Dr. Heather McDonald, who was instrumental in helping with the mechanics of the manuscript. Partial support for this work from NASA Grants NAG8-1566 and NAG8-1837 is also acknowledged.

References

Altenberger, A.R., Deutch, J.M., 1973. Light scattering from dilute macromolecular solutions. J. Chem. Phys. 59, 894–898.

Anderson, J.L., Reed, C.C., 1976. Diffusion of spherical macromolecules at finite concentration. J. Chem. Phys. 64, 3240–3250.

Ansari, R., Suh, K., Arabshahi, A., Wilson, W., Bray, T., DeLucas, L., 1996. A fiber optic probe for monitoring protein aggregation, nucleation and crystallization. J. Cryst. Growth 168, 216–226.

Berne, B.J., Pecora, R., 1976. Dynamic Light Scattering. Wiley, New York.

Bhamidi, V., Hanson, B., Edmundson, A., Skrzypczak, E., Schall, C., 1999. The influence of a homologous protein impurity on lysozyme crystal growth. J. Cryst. Growth 204, 542–552.

Bonnete, F., Finet, S., Tardieu, A., 1999. Second virial coefficient: variations with lysozyme crystallization conditions. J. Cryst. Growth 196, 403–413.

Bray, T., Kim, L., Askew, R., Harrington, M., Rosenblum, W., Wilson, W., DeLucas, L., 1998. New crystallization systems envisioned for microgravity studies. J. Appl. Crystallogr. 31, 515–522.

Brown, W. (Ed.), 1993. Dynamic Light Scattering: The Method and Some Applications. Clarendon Press, Oxford.

Casay, G., Wilson, W., 1992. Laser scattering in a hanging drop vapor diffusion apparatus for protein crystal growth in a microgravity environment. J. Cryst. Growth 122, 95–101.

Casay, G.A., 1986. Photon Correlation Spectroscopy as an On-line Detector for Size Exclusion Chromatography. Mississippi State Univ., Mississippi State. (M.S. thesis).

Chiang, Y., 1994. Determination of Diffusion Virial Coefficients and Osmotic Second Virial Coefficients of Lysozyme by Laser Light Scattering. Mississippi State Univ., Mississippi State. (M.S. thesis).

Chu, B., 1983. In: Earnshaw, J., Steer, M. (Eds.), The Application of Light Scattering to the Study of Biological Motion. Plenum, New York, p. 53.

Cummins, H.Z., 1983. In: Earnshaw, J., Steer, M. (Eds.), The Application of Laser Light Scattering to the Study of Biological Motion. Plenum, New York, pp. 171–207.

Cummins, H.Z., Pike, E.R. (Eds.), 1973. Photon Correlation and Light Beating Spectroscopy. Plenum Press, New York.

Cummins, H.Z., Pike, E.R. (Eds.), 1976. Photon Correlation Spectroscopy and Velocimetry. Plenum Press, New York.

Degiorgio, V., Corti, M., Giglio, M., 1980. Light Scattering in Liquids and Macromolecular Solutions. Plenum Press, New York.

Demoruelle, K., Guo, B., Kao, S., McDonald, H., Nikic, D., Holman, S., Wilson, W., 2002. Correlation between the osmotic second virial coefficient and solubility for equine serum albumin and ovalbumin. Acta Crystallogr. D 58, 1544–1548.

Dubin, S.B., 1970. Quasi-elastic Light Scattering from Macromolecules. Massachusetts Institute of Technology, Cambridge, MA. (Ph.D. dissertation).

Ducruix, A., Giege, R., 1992. In: Crystallization of Nucleic Acids and Proteins—A Practical Approach. Oxford University Press, New York, pp. 73–97.

Ducruix, A., Guilloteau, J., Ries-Kautt, M., Tardieu, A., 1996. Protein interactions as seen by solution X-ray scattering prior to crystallogenesis. J. Cryst. Growth 168, 28–39.

Earnshaw, J.C., Steer, M.W. (Eds.), 1983. The Application of Laser Light Scattering to the Study of Biological Motion. Plenum Press, New York.

Feher, G., Kam, Z., 1985. In: Wycoff, H., Hirs, C., Timasheff, S. (Eds.), Methods in Enzymology, vol. 114. Academic Press, New York, pp. 77–112.

Georgalis, Y., Saenger, W., 1999. Light scattering studies on supersaturated protein solutions. In: Science Progress, vol. 82. Northwood, United Kingdom, pp. 271–294.

George, A., Chiang, Y., Guo, B., Arabshahi, A., Cai, Z., Wilson, W., 1997. Second virial coefficient as a predictor in protein crystal growth. In: Methods in Enzymology, vol. 276. Academic Press, New York, pp. 100–110.

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

Goldin, A.A., 1991. Processing of experimental data in photon correlation spectroscopy under homodyne measuring scheme. Opt. Spectrosk. 71, 485–489.

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

Guo, B., Kao, S., McDonald, H., Asanov, A., Combs, L., Wilson, W., 1999. Correlation of second virial coefficients and solubilities useful in protein crystal growth. J. Cryst. Growth 196, 424–433.

Haas, C., Drenth, J., Wilson, W., 1999. Relation between the solubility of proteins in aqueous solutions and the second virial coefficient of the solution. J. Phys. Chem. B 103, 2808–2811.

Habel, J., Ohren, J., Borgstahl, G., 2001. Dynamic light scattering analysis of full-length human RPA 14/32 dimer: purification, crystallization and self-association. Acta Crystallogr. D 57, 254–259.

Helfrich, J.P., Jones, W.R., 1999. High-sensitivity protein aggregate characterization using flow-mode static and dynamic laser light scattering detection coupled to a size exclusion chromatograph. Am. Biotechnol. Lab. 17, 26–28.

Hess, W., Klein, R., 1976. Theory of light scattering from a system of interacting Brownian particles. Physica 85A, 509–527.

Hitscherich, C., Kaplan, J., Allaman, M., Wiencek, J., Loll, P., 2000. Static light scattering studies of OmpF porin: implications for integral membrane protein crystallization. Protein Sci. 9, 1559–1566.

Johnson, C.S., Gabriel, D.A., 1994. Laser Light Scattering. Dover, New York.

Jones, W.R., 2002. Dynamic and Rayleigh laser light scattering detection modes in size exclusion chromatography as a means to characterize therapeutic proteins and as a methods development tool. Am. Biotechnol. Lab. 20, 36–38.

Juarez-Martinez, G., Garza, C., Castillo, R., Moreno, A., 2001. A dynamic light scattering investigation of the nucleation and growth of thaumatin crystals. J. Cryst. Growth 232, 119–131.

Kadima, W., McPherson, A., Dunn, M., Jurnak, F., 1990. Characterization of precrystallization aggregation of canavalin by dynamic light scattering. Biophys. J. 57, 125–132.

Kam, Z., Shore, H., Feher, G., 1978. On the crystallization of proteins. J. Mol. Biol. 123, 539–555.

Kempner, M.E., Felder, R.A., 2002. Proteomics: diagnostics and specimen biorepositories. Am. Genomic/Proteomic Technol. 2, 6–8.

Koppel, D., 1972. Analysis of macromolecular polydispersity in intensity correlation spectroscopy: the method of cumulants. J. Chem. Phys. 57, 4814–4820.

Kratochvil, P., 1987. Classical Light Scattering from Polymer Solutions. Elsevier, Amsterdam.

Loll, P., Hitscherich, C., Aseyev, V., Allaman, M., Wiencek, J., 2002. Assessing micellar interaction and growth in detergent solutions used to crystallize integral membrane proteins. Cryst. Growth Des. 2, 533–539.

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

Malkin, A., McPherson, A., 1994. Light scattering investigations of nucleation processes and kinetics of crystallization in macromolecular systems. Acta Crystallogr. D 50, 385–395.

McPherson, A., 1999. Crystallization of Biological Macromolecules. Cold Spring Harbor Laboratory Press, New York.

Mikol, V., Hirsch, E., Giege, R., 1989. Monitoring protein crystallization by dynamic light scattering. FEBS Lett. 258, 63–66.

Muschol, M., Rosenberger, F., 1995. Interactions in under- and supersaturated lysozyme solutions—Static and dynamic light scattering results. J. Chem. Phys. 103, 10424–10436.

Neal, B., Asthagiri, D., Lenhoff, A., 1998. Molecular origins of osmotic second virial coefficients of proteins. Biophys. J. 75, 2469–2477.

Neal, B., Asthagiri, D., Velov, O., Lenhoff, A., Kaler, E., 1999. Why is the osmotic second virial coefficient related to protein crystallization? J. Cryst. Growth 196, 377–387.

Phillies, G.D.J., 1975. Continuum hydrodynamic interactions and diffusion. J. Chem. Phys. 62, 3925–3932.

Rosenbaum, D., Zamora, P., Zukoski, C., 1996. Phase behavior of small attractive colloidal particles. Phys. Rev. Lett. 1, 150–153.

Rosenbaum, D., Zukoski, C., 1996. Protein interactions and crystallization. J. Cryst. Growth 169, 752–758.

Ruppert, S., Sandler, S., Lenhoff, A., 2001. Correlation between the osmotic second virial coefficient and the solubility of proteins. Biotechnol. Prog. 17, 182–187.

Saridakis, E., Dierks, K., Moreno, A., Dieckmann, M., Matthias, W., Chayen, N., 2002. Separating nucleation and growth in protein crystallization using dynamic light scattering. Acta Crystaollogr. D 58, 1597–1600.

Schall, C., Riley, J., Li, E., Arnold, E., Wiencek, J., 1996. Application of temperature control strategies to the growth of hen egg-white lysozyme crystals. J. Cryst. Growth 165, 299–307.

Schmitz, K., 1990. An Introduction to Dynamic Light Scattering by Macromolecules. Academic Press, San Diego.

Skouri, M., Delsanti, M., Munch, J., Lorber, B., Giege, R., 1991. Dynamic light scattering studies of the aggregation of lysozyme under crystallizing conditions. FEBS Lett. 295, 84–88.

Skouri, M., Lorber, B., Giege, R., Munch, J., Candau, J., 1995. Effect of macromolecular impurities on lysozyme solubility and crystallizability: dynamic light scattering, phase diagram, and crystal growth studies. J. Cryst. Growth 152, 209–220.

Stepenek, P., 1993. Data analysis in dynamic light scattering. In: Brown, W. (Ed.), Dynamic Light Scattering: The Method and Some Applications. Clarendon Press, Oxford, pp. 177–241.

Tanaka, S., Ito, K., Hayakawa, R., Ataka, M., 1999. Size and number density of precrystalline aggregation in lysozyme crystallization process. Chem. Phys. 111, 10330–10337.

Thibault, F., Langowski, J., Leberman, R., 1992. Optimizing protein crystallization by aggregate size distribution analysis using dynamic light scattering. J. Cryst. Growth 122, 50–59.

Thomas, B.R., Carter, D., Rosenberger, F., 1998a. Effect of microheterogeneity on horse spleen apoferritin crystallization. J. Cryst. Growth 187, 499–510.

Thomas, B.R., Vekilov, P.G., Rosenberger, F., 1998b. Effects of microheterogeneity in hen egg-white lysozyme crystallization. Acta Crystallogr. D 54, 226–236.

Unge, T., 1999. In: Bergfors, T. (Ed.), Protein Crystallization Techniques, Strategies, and Tips. International University Line, La Jolla, CA, pp. 7–17.

Veesler, S., Marcq, S., Lafont, S., Astier, J., Boistelle, R., 1994. Influence of polydispersity of protein crystallization: a quasi-elastic light scattering study applied to α-amylase. Acta Crystallogr. D 50, 355–360.

Vekilov, P., Feeling-Taylor, A., Petsev, D., Galkin, O., Nagel, R., Hirsch, R., 2002. Intermolecular interactions, nucleation, and thermodynamics of crystallization of hemoglobin C. Biophys. J. 83, 1147–1156.

Wessel, T., Ricka, J., 1998. Micro-volume dynamic light scattering and simultaneous video microscopic observation for screening of protein crystal growth. SPIE 3199, 299–305.

Wilson, W., 1990. Monitoring crystallization experiments using dynamic light scattering: assaying and monitoring protein crystallization in solution. Methods: Companion Methods Enzymol. 1, 110–117.

Zukoski, C., Mirarefi, A., Kulkarni, A., 2002. Dynamics and Crystal Nucleation in Protein Solutions. American Crystallographic Association, P120A, San Antonio, TX.