Acta Cryst. (1994). D**50**, 361–365

Predicting Protein Crystallization from a Dilute Solution Property

BY ABRAHAM GEORGE AND W. WILLIAM WILSON

Department of Chemistry, PO Box CH, Mississippi State University, Mississippi State, MS 39762, *USA*

(*Received* 21 *October* 1993; *accepted* 31 *January* 1994)

Abstract

A dilute solution parameter obtained from static light-scattering measurements is proposed as a predictor for protein crystallization experiments. The osmotic second virial coefficients, B_{22}, have been measured for a variety of proteins in solvents that are known to promote crystallization and the values for B_{22} were found to lie within a fairly narrow range which we refer to as a crystallization slot. Solution conditions which were known not to favor crystallization of the proteins resulted in B_{22} values well outside the crystallization slot.

Introduction

The growing of suitable crystals for X-ray or neutron diffraction has always been one of the major barriers to protein structure determination. The basic problem that all protein crystallographers face is that there are more than 20 solution variables that can influence crystallization (McPherson, 1985). The odds against finding the correct combination of solution conditions to promote crystallization are seemingly overwhelming. As a result, most protein crystals have been grown by a trial-and-error method in which many solution conditions were investigated in the hope of finding one that would yield usable crystals. One advancement in utilizing the trial-and-error method is the development of a microscale methodology which literally allows the use of microlitre volumes of sample. These techniques have been applied towards dialysis, liquid–liquid diffusion and vapor diffusion. An equally important advance in protein crystal growth technology is the statistical design of experiments to reduce the quantity of protein required to find the best solution conditions for crystal growth (Carter, Baldwin & Frick, 1988; Carter, 1990). Both of these advances have increased the efficiency of crystal growth screening, but the process remains somewhat haphazard because there is no reliable, quantitative assay to predict the outcome of crystal growth experiments. Such an assay would facilitate the use of much more powerful statistical analysis and optimization procedures.

There is a genuine interest among protein crystallographers for the development of a pre-crystallization assay as a way to test protein solutions to determine the likelihood for either crystal or amorphous precipitate formation. Such an assay could greatly reduce the time and the number of screening experiments necessary to find a crystallizing condition. Having a universal predictor will allow crystallographers to 'fine tune' existing crystallization conditions or discover new solution conditions to crystallize difficult proteins.

Any assay that is to be truly utilized as a universal predictor for protein crystallization should have certain features.

(*a*) The assay should consistently give a discriminating response. For a solution condition leading to crystallization, the assay result should lie within a certain range of values. On the other hand, for a solution condition leading to precipitation or non-crystallization, the assay result should clearly lie outside the range of values for crystallization.

(*b*) The assay should require a minimum amount of protein and should be non-destructive so that essentially all of the protein is recoverable.

(*c*) The assay should accommodate virtually any solvent condition that may be required for crystallization of a particular protein.

(*d*) The assay should be non-invasive so as to prevent protein interaction with foreign bodies.

(*e*) The assay should be able to be routinely performed by laboratory technicians.

There have been recent reports utilizing laser scattering for the specific purpose of studying crystallizing protein solutions (Wilson, 1990). Most of these reports emphasize the use of dynamic laser scattering (DLS) (Berne & Pecora, 1976) to study the change in the state of aggregation of protein solutions during nucleation and post-nucleation growth. A number of attempts have been made to monitor the protein aggregate size as a function of time. The DLS data were subjected to various analysis schemes (Chu, 1983) which gave an estimate of the particle-size distribution in the protein aggregate mixture. Results from these types of experiments have been used to postulate the existence of a critical nucleus to sustain growth (Mikol, Hirsch & Giegé, 1989), and to differentiate between the formation of craggs (aggregates which eventually result in crystals) and praggs

(aggregates which eventually result in precipitation) (Kadima, McPherson, Dunn & Jurnak, 1990). DLS is a sensitive detection method, that is, the formation of small amounts of aggregates in a protein solution can be easily detected, but the method has low resolution so that the actual aggregate sizes are determined with uncertainty. The ability of DLS to differentiate between the presence of craggs or praggs in a protein solution from an aggregate size distribution estimated by DLS at a particular protein concentration has not been demonstrated.

As a predictor for protein crystallization, DLS has relied basically on two types of measurements. The first is the behavior of the translational diffusion coefficient, D_T, of the protein as a function of protein concentration (Feher & Kam, 1985; Mikol, Hirsch & Giegé, 1990) while the second is the behavior of the aggregate size growth curves (particle-size distributions) as a function of time (Georgalis, Zouni, Eberstein & Saenger, 1993; Malkin & McPherson, 1993*b*). Both measurements are generally performed at relatively high protein concentrations, *i.e.* near saturation or supersaturated condition. Interpretation of the DLS data under these conditions is non-trivial because (1) the value for D_T is determined by a combination of thermodynamic and hydrodynamic factors, both of which are concentration dependent, and (2) the inherently low resolution of DLS makes it difficult to accurately and reproducibly estimate particle-size distributions, especially in non-stationary systems such as aggregating protein solutions.

We present here the results from static light-scattering (SLS) experiments that may reveal the discovery of a dilute solution parameter, the osmotic second virial coefficient, that predicts protein crystallization. Measurements have been performed on a chosen set of proteins under a variety of known crystallizing and non-crystallizing conditions so that the generality of the predictor could be verified.

Experimental

Sample preparation

The samples used for this study were ribonuclease A, canavalin, concanavalin A (con-A), lysozyme, ovalbumin, ovostatin, α-chymotrypsin, bovine serum albumin (BSA) and satellite tobacco mosaic virus (STMV). Con-A, lysozyme and BSA were purchased from Boehringer Mannheim Biochemicals, while ovalbumin, ribonuclease A and α-chymotrypsin were obtained from Sigma Chemical Company. Canavalin and STMV were provided by Dr A. McPherson from the University of California at Riverside, and the ovostatin was obtained from Dr M. Pusey at Marshall Space Flight Center in Huntsville, Alabama. All of the samples, with the exception of STMV, were subjected to a strict preparation protocol. Each protein was analyzed by high-performance size-exclusion chromatography (HPSEC) using one Biosep 4000 and one Biosep 3000 analytical column in series (each 300 × 7.8 mm, Phenomenex) with 20 m*M* sodium phosphate buffer, pH 7.0, containing 0.1 *M* sodium chloride and 0.05%(*w*/*v*) sodium azide as the mobile phase. Almost all of the proteins tested gave chromatograms that revealed the presence of protein oligomers or other impurities. For purposes of interpreting the light-scattering data, whether SLS or DLS, it is desirable to remove such components so that the starting solutions are essentially composed of protein monomers. The monomer fraction was collected by using either a Biosep 4000 or a Biosep 3000 prep size column (600 × 25 mm) with the same mobile phase as for the analytical columns. Each of the fractionated proteins was then exhaustively dialyzed at room temperature against a solvent reported in the literature as being either a crystallizing or a precipitating condition. The membrane for dialysis was a Spectra Por/molecular porous with a molecular weight cutoff of 5000 Da. Protein concentrations during dialysis were dilute enough to prevent the formation of either praggs or craggs. The solvent conditions for the various proteins are listed in Table 1. After dialysis, the protein solutions were concentrated using an Amicon ultrafiltration system. The final concentrations, determined spectrophotometrically (see Table 1) were kept well below the protein's solubility. As a final test, 10–20 μl of the concentrated protein solution was subjected to HPSEC using the analytical columns. A single well defined peak representing protein monomers was obtained in all cases indicating that the purification steps were successful. Solvents and protein solutions for light scattering were filtered using 0.2 μm pore-size Anotop Plus filters from Whatman in a closed-loop filtration system (Casey & Wilson, 1992).

Static light scattering

All SLS measurements were performed using a DAWNF laser photometer from Wyatt Technology. The light source was a 5 mW He–Ne laser with wavelength, λ, of 632.8 nm and vertical polarization. Since the molecular size of each of the samples used was smaller than $\lambda/20$, no angular dependence for the excess scattered intensity was expected and all SLS data were recorded at an angle of 90°. The working equation (Kratochvil, 1987) used to interpret the SLS data was

$$\frac{Kc}{R_{90}} = \frac{1}{M} + 2B_{22}c, \qquad (1)$$

Table 1. *Crystallization details*

Protein	$E_{280}^{1\%}$	Crystallization conditions	$B_{22} \times 10^4$ (mol ml g^{-2})	References
Lysozyme	26.3	40 m*M* NaAc, pH = 4.6, 2% NaCl, 298 K	−3.0	Mikol *et al.* (1990)
Canavalin	6.8	50 m*M* Phosphate, pH = 7.0, 0.7% NaCl, 298 K	−0.8	McPherson (1982)
Concanavalin A	13.0	50 m*M* Tris–Ac, pH = 7.0, 1.0 *M* $(NH_4)_2SO_4$, 298 K	−2.5	Mikol *et al.* (1990)
Concanavalin A	13.0	10 m*M* Sodium cacodylate, pH = 6.0, 0.1 *M* NaCl, 298 K	−1.9	Mikol *et al.* (1990)
BSA	6.6	50 m*M* Potassium phosphate, pH = 6.2, 52% saturated $(NH_4)_2SO_4$, 298 K	−2.0	Carter (1992)
Ovostatin	7.98	0.1 *M* Imidazole, pH = 7.5, 7.5% PEG 8000, 293 K	−7.1	Pusey (1992)
Ribonuclease A	22.0	50% *n*-Propanol, pH = 5.0, 297 K	−4.1	King *et al.* (1956)
α-Chymotrypsin	20.0	0.1 *M* NaAC, pH = 4.6, 10% PEG 3350, 298 K	−8.4	Gaier *et al.* (1981)
STMV		12.5% SAS, pH = 6.5, 298 K	−1.8	Malkin & McPherson (1993*a*)
Ovalbumin	26.9	50 m*M* Sodium cacodylate, pH = 5.4, 43% SAS, 2% methanol	−6.1	Miller *et al.* (1983)

where

$$K = \frac{4\pi^2 n_o^2 (dn/dc)^2}{N_A \lambda^4}, \qquad (2)$$

and n_o = refractive index of the solvent; (dn/dc) = refractive index increment for the protein/solvent pair; N_A = Avogadro's number; λ = wavelength of the incident light in vacuum; c = concentration of the protein (g ml^{-1}); R_{90} = excess Rayleigh ratio at an angle of 90°.

The absolute R_{90} values were determined from the excess scattered intensities (intensities above that due to solvent and background) by calibration of the photometer using toluene which has a known R_{90} value of 14.06×10^{-6} cm^{-1} at 632.8 nm. Equation (1) indicates that a plot of Kc/R_{90} *versus* c yields the parameters M, molecular weight of the protein, and B_{22}, the osmotic second virial coefficient. The highest protein concentrations used in this work varied from about 5–6 mg ml^{-1} for small proteins such as lysozyme and ribonuclease A down to 0.1 mg ml^{-1} or even less for large scatterers such as ovostatin or STMV.

Results and discussion

Typical SLS results are shown in Fig. 1 for BSA in four different solvent conditions at 298 K. The solvent composition was 50 m*M* potassium phosphate buffer at pH = 6.2 with either (*a*) 20, (*b*) 37, (*c*) 52 or (*d*) 60%(*v*/*v*) saturated ammonium sulfate (SAS) as the crystallizing agent. The second virial coefficient, B_{22} in (1), was obtained from the slope of each of the four data sets and was found to vary systematically from a positive value at 20% SAS to a negative value at 60% SAS. A reported crystallization condition for BSA at this temperature and pH value was with 52% SAS which gave a B_{22} value of -2.0×10^{-4} mol ml g^{-2}.

Similar experiments were performed for each of the protein/solvent pairs studied and the measured values for B_{22} are listed in Table 1. The values range between about -1×10^{-4} and -8×10^{-4} mol ml g^{-2} and a histogram representation of the results is shown in Fig. 2. It was very encouraging to see that the limited number of experiments was resulting in a well defined histogram. These results indicate that for every protein crystallization condition tested, the value of the second virial coefficient lay within a fairly narrow range which will be referred to as the crystallization slot. It is important to note that a variety of crystallizing agents were tested including inorganic salts, PEG's and alcohols on proteins ranging in molecular weight from about 1.4×10^4 (lysozyme) to 7×10^5 (ovostatin). Although it is premature to say that this is a completely general assay, it has shown remarkable consistency for the samples tested thus far.

An obvious question is what values for B_{22} are obtained for solution conditions that do not promote crystallization? Fig. 3 shows results for lysozyme in (*a*) 40 m*M* sodium acetate buffer, pH 4.3, T = 298 K, a non-crystallizing condition; (*b*) 40 m*M* sodium acetate buffer with 2.0%(*w*/*v*) sodium chloride, pH 4.3, T = 298 K a crystallizing condition; (*c*) 40 m*M* sodium acetate buffer with 1.25 *M* ammonium sulfate, pH 4.3, T = 298 K, a precipitating condition. The values for B_{22} from data sets *a*, *b*, and *c* are 12×10^{-4}, -3×10^{-4} and -9×10^{-4} mol ml g^{-2},

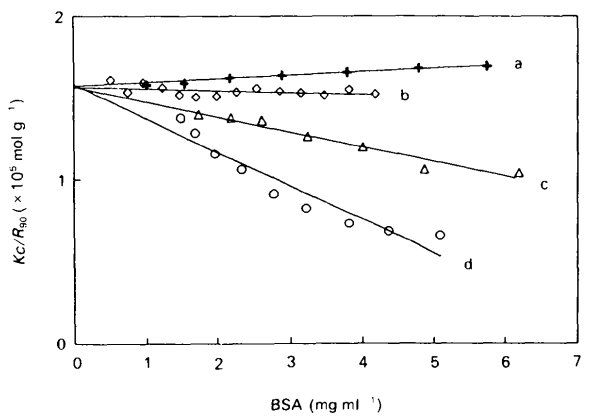

Fig. 1. SLS data for BSA in 50 m*M* potassium phosphate buffer, pH = 6.2, T = 298 K with (*a*) 20, (*b*) 37, (*c*) 52, (*d*) 60%(*v*/*v*) SAS.

respectively. Solution condition *a* lies clearly outside the crystallization slot, while condition *c* is on the fringe and may indicate that 1.25 *M* ammonium sulfate is a borderline crystallizing/precipitating agent. In general, solution conditions which eventually produce praggs give B_{22} values that are large and negative. For example, con-A in 10 m*M* sodium cacodylate buffer with 200 m*M* spermine, pH = 6.0, $T = 298$ K gave $B_{22} = -47 \times 10^{-4}$ mol ml g^{-2}. Con-A in the same buffer but with 150 m*M* ethylenediamine sulfate, pH = 7.0, $T = 298$ K gave $B_{22} = -29 \times 10^{-4}$ mol ml g^{-2}. Both of these precipitating conditions (Mikol, Hirsch & Giegé, 1990) led to B_{22} values well beyond the crystallization slot. The second virial coefficient, as presented in (1), is an empirical/parameter attempting to account for the non-ideality in protein solutions. Qualitatively, B_{22} is a measure of two-body (protein–protein) interactions in a dilute solution condition. Positive values for B_{22} generally indicate that repulsive forces between protein molecules dominate so that protein–solvent interactions are favored over those between protein molecules, and the solvent in this case is referred to as a 'good solvent' in a thermodynamic sense. On the other hand, as attractive interactions between protein molecules become stronger, B_{22} becomes negative and the solvent is said to be 'poor'. A special situation occurs when the repulsive and attractive forces between protein molecules are equal. The value for B_{22} becomes zero and the solvent is termed ideal or a 'theta solvent'.

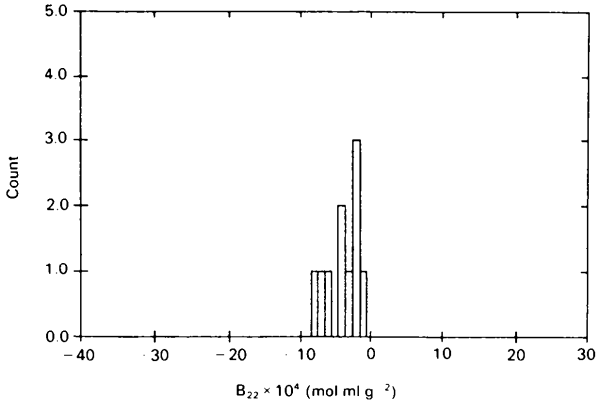

Fig. 2. Histogram representing crystallization slot obtained from second virial coefficient measurements on various proteins in crystallizing solvents.

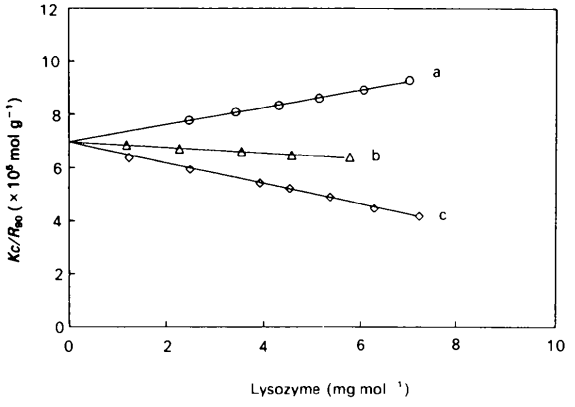

Fig. 3. SLS data for lysozyme in a (*a*) non-crystallizing solvent, (*b*) crystallizing solvent, (*c*) precipitating solvent. See text for solvent conditions.

A more quantitative understanding of the second virial coefficient is given by thermodynamic equilibrium solution theory in which B_{22} is related to the potential of mean force, W_{22}, which describes all of the forces of interaction between two protein molecules in infinitely dilute solution (Stighter & Hill, 1959). In their description, W_{22} accounts for hard-sphere interaction between pairs of protein molecules, charge–charge interactions, charge–dipole interactions, charge-induced dipole interactions, dipole–dipole interactions, dipole-induced dipole interactions and dispersion/van der Waals interactions. In short, B_{22} reflects, through W_{22}, the total thermodynamic environment for protein molecules diluted in a given solvent.

Referring to Fig. 2, the data show that solvent conditions which are known to promote protein crystallization are grouped within a narrow range of B_{22} values that are somewhat negative. These solvent conditions can be referred to as being 'moderately poor', *i.e.* the solvent has to be poor enough (slightly negative B_{22} values) to eventually promote the formation of craggs at high enough protein concentration but not so poor (larger negative B_{22} values) that praggs are produced which lead to an amorphous structure. Thus, B_{22} is clearly a dilute solution parameter that has a predictive character regarding protein crystallization. The identification of a crystallization slot for solvent character could be an important discovery regarding a more systematic approach to protein crystallization. The database for the histogram presented in Fig. 2 needs to be expanded to confirm the universality of the assay. Even so, it is strongly postulated that any protein/solvent pair which corresponds to a B_{22} value outside the crystallization slot has a low probability for producing protein crystals.

This research was supported in part by grant number NAG8-965 from the National Aeronautics and Space Administration and a sub-contract from the Center for Macromolecular Crystallography at the University of Alabama in Birmingham, USA.

References

BERNE, B. J. & PECORA, R. (1976). *Dynamic Light Scattering*. New York: John Wiley.

CARTER, C. W. JR (1990). *Methods: a Companion to Methods in Enzymology*, Vol. 1, pp. 12–24. New York: Academic Press.

CARTER, C. W. JR, BALDWIN, E. T. & FRICK, L. (1988). *J. Cryst. Growth*, **90**, 60–73.

CARTER, D. (1992). Marshall Space Flight Center, Huntsville, Alabama, USA. Personal communication.

CASEY, G. & WILSON, W. (1992). *J. Cryst. Growth*, **122**, 85–101.

CHU, B. (1983). *The Application of Light Scattering to the Study of Biological Motion*, edited by J. EARNSHAW & M. STEER, p. 53. New York: Plenum Press.

FEHER, G. & KAM, Z. (1985). *Methods Enzymol.* **114**, 77–111.

GAIER, J. R., TULINSKY, A. & LIENER, I. E. (1981). *J. Biol. Chem.* **256**, 11417–11419.

GEORGALIS, Y., ZOUNI, A., EBERSTEIN, W. & SAENGER, W. (1993). *J. Cryst. Growth*, **126**, 245–260.

KADIMA, W., MCPHERSON, A., DUNN, M. F. & JURNAK, F. A. (1990). *Biophys. J.* **57**, 125–132.

KING, M. V., MAGDOFF, B. S., ADELMAN, M. B. & HARKER, D. (1956). *Acta Cryst.* **9**, 460–465.

KRATOCHVIL, P. (1987). *Classical Light Scattering from Polymer Solutions*. Amsterdam: Elsevier.

MCPHERSON, A. (1982). *Preparation and Analysis of Protein Crystals*. New York: Wiley.

MCPHERSON, A. (1985). *Methods Enzymol.* **114**, 112–120.

MALKIN, A. J. & MCPHERSON, A. (1993*a*). *J. Cryst. Growth*, **128**, 1232–1235.

MALKIN, A. & MCPHERSON, A. (1993*b*). *J. Cryst. Growth*, **133**, 29–37.

MIKOL, V., HIRSCH, E. & GIEGÉ, R. (1989). *FEBS Lett.* **258**, 63–66.

MIKOL, V., HIRSCH, E. & GIEGÉ, R. (1990). *J. Mol. Biol.* **213**, 187–195.

MILLER, M., WEINSTEIN, J. N. & WLODAWER, A. (1983). *J. Biol. Chem.* **258**, 5864–5866.

PUSEY, M. L. (1992). Marshall Space Flight Center, Huntsville, Alabama, USA. Personal communication.

STIGHTER, D. & HILL, T. W. (1959). *J. Chem. Phys.* **63**, 55.

WILSON, W. W. (1990). *Methods: a Companion to Methods in Enzymology*, Vol. 1, pp. 110–117. New York: Academic Press.