Macromolecules **1984,** *17,* 1723-1726 **1723**

(15) Berger, A. E., submitted for publication.
(16) Cotter, M. A. *Phys. Rev. A* **1974,** *10,* 625-636.
(17) Sattinger, D. H. *Bull. Am. Math. Soc.* **1980,** *3,* 779-819. (See theorem on page 782.)
(18) Sattinger, D. H. "Topics in Stability and Bifurcation Theory"; Springer-Verlag: New York, 1973; pp 141-144.
(19) Crandall, M. G.; Rabinowitz, P. H. *Arch. Rational Mech. Anal.* **1973,** *52,* 161-180.

Statistical Analysis of Quasi-Elastic Light Scattering Data

G. J. Wei and Victor A. Bloomfield*

Department of Biochemistry, University of Minnesota, St. Paul, Minnesota 55108.
Received November 18, 1983

ABSTRACT: We often observe a linear correlation between the first and the second cumulants when analyzing autocorrelation functions obtained in quasi-elastic light scattering (QLS). For time-invariant samples, this is explained by a theory based on random noise effects. The theory is supported by evidence from computer simulation and from experimental QLS study of polystyrene latex spheres. Using these results, we propose a method for reducing scatter of QLS data from a reacting system and illustrate it by a simulated example.

Introduction

Quasi-elastic light scattering (QLS) has become a convenient and reliable tool for obtaining hydrodynamic properties of macromolecules.[1,2] With the cumulant method of data analysis,[3] one gets a *z*-averaged diffusion coefficient from the first cumulant and a relative variance or "quality factor" from the second cumulant, which reflects the width of the size distribution of the sample.

In our routine QLS measurements of various types of samples, we often observe a pattern in the variation of the diffusion coefficient D and the quality factor $Q = \langle (D - \langle D \rangle)^2 \rangle / \langle D \rangle^2$. For a given sample under identical experimental conditions, repeated measurements show that variations in D and Q are linearly correlated.

In this paper we show how this observation can be explained by a statistical theory of the effect of random noise on the least-squares fitting process. Using this experimental observation and the theoretical explanation, we suggest a method to improve data from QLS measurements on time-dependent samples.

Theory

Let $\epsilon(0,\sigma^2)$ represent Gaussian random numbers with a mean value of zero and a standard deviation of σ. Then, for a "noise-infected" quadratic function

$$f(x) = b_0 + b_1 x + b_2 x^2 + \epsilon(0,\sigma^2) \qquad (1)$$

the correlation among the expected values β_0, β_1, and β_2 (corresponding to b_0, b_1, and b_2), obtained from least-squares fitting, is given by the Variance Matrix,[4] $\mathbf{V}$:

$$\mathbf{V} = \begin{pmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{pmatrix} = \sigma^2(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1} \qquad (2)$$

where

$$\mathbf{x} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & x_n & x_n^2 \end{pmatrix} \qquad (3a)$$

$$\mathbf{x}^{\mathrm{T}} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \\ x_1^2 & x_2^2 & \dots & x_n^2 \end{pmatrix} \qquad (3b)$$

and x_i is the ith data point.

The variance of each fitted parameter is given by the corresponding diagonal element of $\mathbf{V}$. The off-diagonal elements indicate the correlations between fitted parameters. If, for example, β_2 is plotted against β_1 for a large number of fits, the least-squares slope for the plot would be V_{23}/V_{22}.

In light scattering experiments, the noise appears in the exponential function, while the least-squares fitting is usually carried out after taking the logarithm. The theory needs to be modified as follows:

$$f'(x) = \ln\,[\exp(b_0 + b_1 x + b_2 x^2) + \epsilon] \qquad (4)$$

This may be expanded about $\epsilon = 0$ as

$$f'(x) = (b_0 + b_1 x + b_2 x^2) + \epsilon \exp[-(b_0 + b_1 x + b_2 x^2)] + \mathcal{O}(\epsilon^2) \qquad (5)$$

If terms of higher order than ϵ can be ignored, the Variance Matrix for the new set of fitted parameters (β_0', β_1', β_2') would be (see Appendix)

$$\mathbf{V}' = \begin{pmatrix} V_{11}' & V_{12}' & V_{13}' \\ V_{21}' & V_{22}' & V_{23}' \\ V_{31}' & V_{32}' & V_{33}' \end{pmatrix} = \sigma^2(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}(\mathbf{x}^{\mathrm{T}}\mathbf{w}\mathbf{x})(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1} \qquad (6)$$

where $\mathbf{w}$ is an $n \times n$ diagonal matrix with the diagonal elements

$$w_{ii} = \exp[-(b_0 + b_1 x_i + b_2 x_i^2)] \qquad (7)$$

Again a plot of β_2' vs. β_1' should yield a slope, β, given by V_{23}'/V_{22}'.

In cumulant analysis, the logarithm of the photocurrent autocorrelation function, $C(t)$, is fitted to a quadratic equation:

$$y(t) = \tfrac{1}{2} \ln C(t) = a - bt + ct^2/2 \qquad (8)$$

where $b = q^2\langle D \rangle$, $c = q^4\langle (D - \langle D \rangle)^2 \rangle$, q is the scattering wavevector $= (4\pi\tilde{n}/\lambda) \sin(\theta/2)$, λ is the wavelength of laser light, $\tilde{n}$ is the refractive index of the medium, θ is the scattering angle, and Q is the quality factor $= \langle (D - \langle D \rangle)^2 \rangle / \langle D \rangle^2 = c/b^2$.

Let b', c', and Q' be the "true" values of b, c, and Q, respectively, and $\Delta b = b - b'$, $\Delta c = c - c'$, and $\Delta Q = Q - Q'$ the corresponding differences. We have used the term "true" value to indicate the value that one would obtain in the absence of random noise. It may include systematic

0024-9297/84/2217-1723$01.50/0

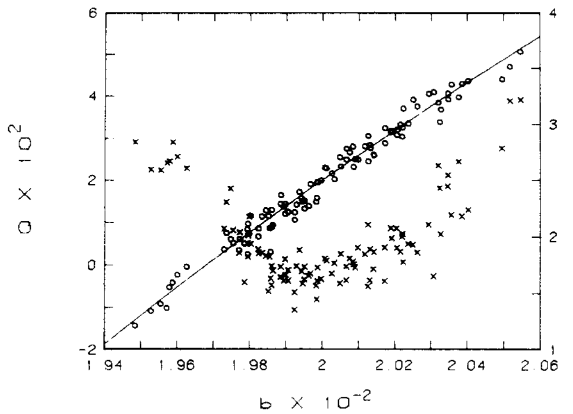

Figure 1. Computer simulation of the effect of random noise on cumulant fit. Random noise is added to an autocorrelation function $C(i) = \exp(-0.04i + 8 \times 10^{-6}i^2)$, and the sum is analyzed by the cumulant method to obtain b and Q, plotted as circles. The true values are $b' = 0.02$ and $Q' = 0.02$. The solid line represents the theoretical prediction (eq 10), and the crosses show the variance of the fit.

error and other errors not considered here.

The theory predicts that the variations in the first and the second cumulants due to random noise are linearly related with a proportionality constant β:

$$\beta = (\Delta c)/(\Delta b) = -2(V_{23}'/V_{22}') \qquad (9)$$

Here a factor of −2 is included to account for the minus sign in the t term and for the $^1/_2$ in the t^2 term of eq 8. In practice, the relative variance, or quality factor, is a more meaningful quantity:

$$Q = c/b^2 = (c' - \Delta c)/b^2 = (c' - \beta b' + \beta b)/b^2 \qquad (10)$$

$$\mathrm{d}Q/\mathrm{d}b = (2\beta b' - \beta b - 2c')/b^3 \qquad (11)$$

Therefore over a small range of variation in b near b', the plot of Q vs. b will be very close to a straight line with a slope of $(\beta b' - 2c')/(b')^3$. This explains the correlation between Q and b observed experimentally.

Experimental Section

1. Computer Simulation. A computational experiment was carried out to test the theory. Normally distributed (Gaussian) random numbers were generated with the RND function of the Tektronix computer (Model 4052), followed by averaging. These numbers were superimposed on a synthesized correlation function with $a' = 0$, $b' = 0.02/T$, and $Q' = 0.02$:

$$C(i) = \exp(-0.04i + 8 \times 10^{-6}i^2) \qquad i = 1, 2, ..., n \qquad (12)$$

Here the channel time T is introduced to make b' and c' unitless and sample-independent. The final noise-affected correlation functions were analyzed by the cumulant method.

The results for the case where the magnitude of random noise is 1% of the first channel are shown in Figure 1. Over a small range of variation, typical of good-quality experimental data, the quality factor Q varies linearly with the first cumulant b.

The Variance Matrix **V**′ for the given condition is calculated to be

$$\mathrm{V}' = \begin{pmatrix} 1.235 \times 10^{1} & -9.785 \times 10^{-1} & 1.207 \times 10^{-2} \\ -9.785 \times 10^{-1} & 7.918 \times 10^{-2} & -9.926 \times 10^{-4} \\ 1.207 \times 10^{-2} & -9.926 \times 10^{-4} & 1.275 \times 10^{-5} \end{pmatrix}$$

and $\beta = 2.504 \times 10^{-2}$.

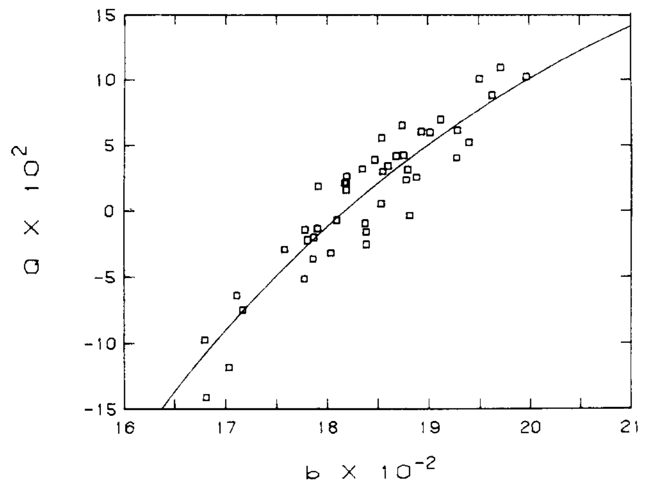

Figure 2. Variation of Q and b obtained from QLS of a sample of polystyrene latex spheres of nominal radius 1090 Å at 20 °C. The solid line was plotted according to eq 10, with $b' = 1825$ s^{-1}, $T = 20$ μs, and $Q' = 0.005$.

The theoretical prediction (eq 10) is plotted as the solid line in Figure 1, which is in good agreement with the experimental results. Also plotted are the variance of the least-squares fit, which shows the expected minimum near the true value (b, Q).

2. QLS of Polystyrene Latex Spheres. An aqueous suspension of polystyrene latex spheres was studied with our photon-counting, digital-correlation apparatus.[5,6] The light intensity was kept low (5×10^4 counts/s) and averaging time short (100 s) to allow a wider range of variation (base line count = 5×10^6, first-channel count = 2×10^6 above base line). Forty-five autocorrelation functions were acquired consecutively, and the values of Q and b were determined by the cumulant analysis. The results are plotted in Figure 2.

The theoretical line is calculated with eq 10, with $b' = 1825$ s^{-1}, in accordance with the size of the latex spheres used (1090 Å). A value of 0.005 is chosen for Q' to reflect the dispersion of the size distribution of about 7%. Then c' is determined by $c' = Q'b'^2$ to be 16653. To obtain β in this case, the Variance Matrix **V**′ is constructed for $n = 56$ (64-channel correlator with the last 8 channels delayed), yielding a slope of 4.436×10^{-2}, which is set equal to βT (see Discussion). With $T = 20$ μs from the experimental setting, $\beta = 2218$. The theoretical prediction is represented by the solid line in Figure 2. The agreement is satisfactory.

Applications

1. Time-Independent Samples. When random noise or counting error is the major cause for the observed fluctuation in the diffusion coefficients, it is possible to construct a diagram like Figure 1 and use the following rules to determine a more precise value of D. (1) The mean values of D and Q are good approximations to the true values due to the symmetry of the distribution. (2) The true values lie close to the minimum of the variance of the fit.

In principle, given enough time for the data averaging process, random counting error may be reduced to achieve the same accuracy as by the method just described. However, when the sample contains a small amount of strongly scattering contaminant, such as dust, the method of averaging several D's and Q's obtained in short runs would give a more reliable approximation to the true values than a single long run. Runs affected by dust (recognizable by transient high scattering intensity) can be discarded.

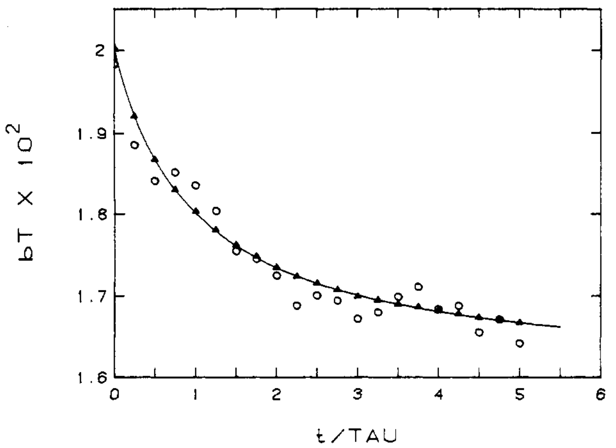

Figure 3. Reduction of scatter of data in a simulated kinetic experiment. The solid line is the true time course of $b = q^2\langle D\rangle$ in the reaction; (○) are the data that exhibit substantial scatter due to noise, and (▲) are (○) treated by eq 13. TAU = $1/kA_0$ in eq 14. $\gamma = 0.8$. t is the reaction time, and T is the correlator sample time.

2. Time-Dependent Samples. The finding of the precise relationship between c and b can be very useful in cases where repeated measurements of a sample are not possible due to the time dependence of the sample. One example is the use of QLS to follow the kinetics of a reaction by measuring the average diffusion coefficient as a function of time (e.g., ref 7). The scatter of data may be reduced significantly by "normalizing" the second cumulant according to eq 10, which is rearranged in the following way:

$$b'(t) = (c' + \beta b - Qb^2)/\beta \qquad (13)$$

Here β is estimated from eq 9; $b(t)$ and $Q(t)$ are quantities measured at reaction time t; $c'(t)$ may be determined experimentally or from calculation, depending on the reaction system.

For a simple case where one reaction component scatters much more strongly than the rest, $c'(t)$ is only weakly dependent on time and can be obtained experimentally by repeated measurements of the reaction mixture, excluding one or more components of low scattering power. For more complex cases, $c'(t)$ will depend on the composition of the sample mixture, which is time dependent. If the composition of the sample can be derived from $b'(t)$, it is possible to find an expression for $c'(t)$ in terms of $b'(t)$. Then eq 13 can be solved by a numerical iteration method to generate a self-consistent $b'(t)$.

Figure 3 shows an example of how this method can be useful in reducing the scatter of kinetic data. For a simple bimolecular reaction of the form A + B → C, with reactant concentrations equal, the z-averaged diffusion coefficient $\langle D\rangle$ varies with time as

$$\langle D\rangle = D_A(1 + \gamma k A_0 t) \qquad (14)$$

where D_A is the diffusion coefficient of A (the major scattering reactant; i.e., the molecular weight of A is much greater than that of B), $\gamma = D_C/D_A$ = ratio of diffusion coefficients, k is the rate constant of reaction, and $A_0 = B_0$ = initial concentration of reactants.

Using eq 14, we programed our computer to generate a series of autocorrelation functions, each corresponding to a particular $\langle D\rangle$ at time t. After random noise was superimposed on these correlation functions, they were analyzed by the cumulant method. Without normalization to a predetermined quality factor, the data (open circles) are scattered about the true curve (solid line). When these data were normalized according to eq 13, the result (solid triangles) shows almost no scatter and agrees very well with the true reaction profile.

Discussion

Since in a light scattering experiment, the noise is in the exponential autocorrelation function, a direct nonlinear fit appears to be statistically more appropriate. However, due to its ease and speed, linear least-squares fit to the logarithm of the autocorrelation function is far more popular. This is exemplified by the widely used cumulant procedure.[3]

On the basis of the statistical theory, the variations in the fitted cumulants should be linearly correlated. This is demonstrated by the computer simulation experiment and by the actual light scattering data from solutions of polystyrene latex spheres (Figures 1 and 2).

The correlation coefficient for the c vs. b plot (not shown, but similar to Figure 1) is 0.988. This suggests a high degree of linear correlation between b and c. A comparable plot for a vs. b is also linear (not shown), with a correlation coefficient of 0.990. It should also be mentioned that the magnitude of a is quite small, such that e^a has a mean of 1.0 and a standard deviation of 0.003.

Normally, if the noise is in a linear function itself, the Variance Matrix **V** would depend only on the functional form of $f(x)$ (eq 1–3). In other words, no matter what values b_0, b_1, and b_2 may take, the same **V** results. But in the case of light scattering, an additional perturbation term $\epsilon e^{-(b_0+b_1x+b_2x^2)}$ (see eq 5) is included in the consideration. The Variance Matrix **V**′ becomes dependent on b_0, b_1, and b_2. However, the dependence is rather weak. The slope of the c vs. b plot changes from 2.507×10^{-2} for monodisperse samples ($Q = 0$) to 2.357×10^{-2} for polydisperse ones ($Q = 0.3$).

The slope β from the computer simulation (2.494×10^{-2}) is lower than that predicted from the Variance Matrix (2.504×10^{-2}) by only 0.4%. The agreement is remarkable. This indicates that the truncation of terms of higher order than ϵ is justified for the present condition.

An important point should be made here: β is in fact βT, where T is the channel time. We have for convenience set $T = 1$ in eq 9. A study of the dependence of βT on n shows $\beta T = 2.5/n$, or $\beta = 2.5/nT$, for the prescribed condition where nT equals $2/b'$. That is, nT is twice the decay time of the autocorrelation of the scattered field, or 4 times that of $C(i)$ in eq 12. With $nT = 2/b'$, $\beta = 1.25b'$, which means β is only dependent on the property of the sample, and not on the correlator used. This argument may be easily extended to other experimental conditions where nT equals $1/b'$ or $3/b'$, etc.

Although the relationship between Q and b is more complex than that between c and b, data have been presented in Q vs. b plots because Q is widely used as the standard measure of relative dispersion of the data. When the noise level is low and the variations in b and c are small, Q vs. b plots are also linear.

When this method is applied to time-dependent measurements, the choice of a proper $c'(t)$ may be difficult. Qualitatively, one large uncertainty in $c'(t)$ is the constant offset arising from all sources other than random noise. Fortunately, because of the fixed linear relationship between c' and b', the main effect of this uncertainty is to produce a constant shift in $b'(t)$; it does not alter the time dependence of $b'(t)$ significantly. Therefore the kinetic parameters obtained would still be acceptable.

We have occasionally observed experimentally what these calculations confirm: the quality factor can be negative due to random noise, although one would expect it to be positive from the greater contribution of slow relaxation processes at long channel times. The true value of Q should indeed be positive, but its random variation may extend to the negative domain.

Acknowledgment. This research was supported in part by NSF Grants PCM 78-06777 and 81-18107. We thank Professor S. Weisberg and the Statistics Clinic of the University of Minnesota for helpful advice on the use of the Variance Matrix.

Appendix. Derivation of Eq 6

For a function $f(x) = b_0 + b_1x + b_2x^2 + \epsilon$, as in eq 1, its matrix representation is

$$\mathbf{F} = \mathbf{xB} + \mathbf{E}$$

where

$$\mathbf{F} = \begin{pmatrix} f_1 \\ f_2 \\ \cdot \\ \cdot \\ \cdot \\ f_n \end{pmatrix} \qquad \mathbf{x} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & x_n & x_n^2 \end{pmatrix} \qquad \beta = \begin{pmatrix} b_0 & b_1 & b_2 \end{pmatrix}$$

and

$$\mathbf{E} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_n \end{pmatrix}$$

and $\mathbf{x}^{\mathrm{T}}$ is the transpose of the matrix $\mathbf{x}$. The estimated values (β_0, β_1, β_2) for the corresponding parameters (b_0, b_1, b_2) are given by (see eq 2.1.17 of ref 4)

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = (\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}\mathbf{x}^{\mathrm{T}}\mathbf{F}$$

using the relationships Var ($\mathbf{kx}$) = $\mathbf{k}$[Var ($\mathbf{x}$)]$\mathbf{k}^{\mathrm{T}}$, and $(\mathbf{AB})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$

$$\begin{aligned} \mathrm{Var}\ (\beta) &= [(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}\mathbf{x}^{\mathrm{T}}][\mathrm{Var}\ (\mathbf{F})][(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}\mathbf{x}^{\mathrm{T}}]^{\mathrm{T}} \\ &= (\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}[\mathbf{x}^{\mathrm{T}}\ \mathrm{Var}\ (\mathbf{F})\mathbf{x}](\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1} \end{aligned}$$

Since Var ($\mathbf{F}$) = Var ($\mathbf{xB}$) + Var ($\mathbf{E}$) = Var ($\mathbf{E}$) = σ^2

$$\mathrm{Var}\ (\beta) = \sigma^2(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1} \qquad \text{(see eq 2.3.1 of ref 4)}$$

Now, if the noise appears in the exponential function

$$f(x) = \exp(b_0 + b_1x + b_2x^2) + \epsilon$$

and a linear least-squares fit is done on ln $f(x)$

$$\begin{aligned} f'(x) &= \ln\ (e^{b_0+b_1x+b_2x^2+\epsilon}) \\ &= (b_0 + b_1x + b_2x^2) + \epsilon e^{-(b_0+b_1x+b_2x^2)} \end{aligned}$$

the Variance Matrix becomes

$$\mathrm{Var}\ (\beta') = \begin{pmatrix} \beta_0' \\ \beta_1' \\ \beta_2' \end{pmatrix} = (\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}[\mathbf{x}^{\mathrm{T}}\ \mathrm{Var}\ (\mathbf{F}')\mathbf{x}](\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}$$

Since Var ($\mathbf{F}'$) = $\mathbf{W}$[Var ($\mathbf{E}$)]$\mathbf{W}^{\mathrm{T}}$ = $\sigma^2\mathbf{WW}^{\mathrm{T}}$ = $\sigma^2\mathbf{W}^2$ where $\mathbf{W}$ is a diagonal matrix with elements

$$W_{ii} = \exp[(b_0 + b_1x_i + b_2x_i^2)]$$

$\mathrm{Var}\ (\beta') = \sigma^2(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}(\mathbf{x}^{\mathrm{T}}\mathbf{w}^2\mathbf{x})(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}$.

Registry No. Polystyrene, 9003-53-6.

References and Notes

(1) B. J. Berne and R. Pecora, "Dynamic Light Scattering with Applications to Chemistry, Biology and Physics", Wiley, New York, 1976.
(2) V. A. Bloomfield, *Annu. Rev. Biophys. Bioeng.*, **10**, 421–50 (1981).
(3) D. E. Koppel, *J. Chem. Phys.*, **57**, 4814–20 (1972).
(4) See, for example, N. R. Draper and H. Smith, "Applied Regression Analysis", Wiley, New York, 1967, Chapter 2.
(5) G. J. Wei and V. A. Bloomfield, *Anal. Biochem.*, **101**, 245–53 (1980).
(6) C. H. Pletcher, R. M. Resnick, G. J. Wei, V. A. Bloomfield, and G. L. Nelsestuen, *J. Biol. Chem.*, **256**, 7433–8 (1980).
(7) J. A. Benbasat and V. A. Bloomfield, *J. Mol. Biol.*, **95**, 335–57 (1975).

Static and Dynamic Solution Properties of Pullulan in a Dilute Solution

Tadaya Kato,* Tsunehisa Katsuki, and Akira Takahashi

Department of Industrial Chemistry, Faculty of Engineering, Mie University, Tsu, Mie 514, Japan. *Received December 6, 1983*

ABSTRACT: Pullulan, an extracellular polysaccharide, has been studied in the dilute region by both static and dynamic light scattering techniques. The samples used are fractions with narrow molecular weight distribution over the range of molecular weight from 5×10^3 to 2.4×10^6. The radii of gyration (R_G) have been estimated from the scattered light intensity with the usual Zimm plot. The hydrodynamic radii (R_H) have been obtained from the translational diffusion coefficient (D_0) at infinite dilution. The characteristic exponents for the molecular weight dependence of polymer chain dimensions in good solvents are determined as ν_G for the R_G values and ν_H for the R_H values, respectively. The ν_G value ($\nu_G = 0.58$) falls close to the asymptotic value predicted theoretically. The ν_H value ($\nu_H = 0.52$) is, however, smaller than the ν_G value obtained. The ratios ($\rho = R_G/R_H$) estimated from the experimental results for each fraction used are discussed. The concentration dependence coefficients (k_D) of the diffusion coefficient are also discussed in terms of the hydrodynamic theories. It is confirmed that pullulan aqueous solutions behave as typical solutions of linear polysaccharides in good solvents.

Introduction

Polymer chain dimensions in dilute solutions have been determined conventionally by both static and dynamic light scattering experiments. For samples having sufficiently high molecular weight, the dimensions extrapolated to the infinite dilution limit depend on the molecular weight, M, at a constant temperature as a simple power law:

$$R_G \sim M^{\nu_G} \qquad (1)$$