Modeling Dynamic Light Scattering of Supercoiled DNA

Konstantin Klenin, Markus Hammermann, and Jörg Langowski*

Division Biophysics of Macromolecules, German Cancer Research Center, Im Neuenheimer Feld 280, D-69120, Heidelberg, Germany

Received August 25, 1999; Revised Manuscript Received November 29, 1999

ABSTRACT: A new computational scheme, based on the Brownian dynamics and Monte Carlo techniques, was used to calculate the dynamic structure factor of supercoiled SV40 DNA (5243 base pairs) in 0.1 M sodium chloride solution. All model parameters are known independently, and none are fitted. The results are in good agreement with earlier dynamic light scattering measurements. The calculated dynamic structure factor, as a function of time, could be very well approximated by a simple analytical formula. The analysis of the experimental data based on this formula allows one to obtain the distribution function for relaxation times of intramolecular motions, if additional information about the DNA structure is available from Monte Carlo simulations. The range of relaxation times for supercoiled SV40 DNA comprises approximately 2 orders of magnitude, with the upper boundary exceeding 1 ms.

Introduction

Dynamic light scattering (DLS) studies of DNA solutions provide valuable information about DNA structure and dynamics. However, the extraction of this information from experimental data is an ill-posed problem allowing a variety of solutions, only one of which corresponds to physical reality. Therefore, it is important to have a DNA model that enables one to calculate the scattering properties of the molecule. The experimental measurements can then be used as a criterion for the adequacy of such a model.

The primary objective of the present study is to verify that the Brownian dynamics (BD) method correctly predicts the dynamic structure factor of supercoiled DNA. In the BD model, the DNA molecule is approximated as a chain of beads diffusing in a viscous medium (water) under the influence of a certain interbead interaction potential.[1] The technique of BD simulation is widely used to analyze various DNA properties[2−7] including those available from DLS measurements.[8−10] Recently, a number of computer programs have been developed in which the particular properties of supercoiled DNA are taken into account.[11−15] The main problem here, however, is that the relaxation times of internal motions in supercoiled DNA are quite large (up to the several milliseconds[15]), and the simulated times, restricted by the available computer capacities, cannot significantly exceed this value. This leads to difficulties in collecting good statistics. The first attempt to calculate the dynamic structure factor of supercoiled DNA by BD simulations was done by Hammermann et al.,[16] who used additional assumptions based on the analytical theory of Berg[17] and Soda.[18] This theory, however, was developed for circular but not supercoiled DNA, and its applicability in this case is disputable.

In the present study, we have developed a new computational scheme involving a combination of Brownian dynamics and Monte Carlo (MC) methods. No additional assumptions, other than those intrinsic to the BD model, were made. Following this scheme, we were able to calculate the dynamic structure factor for a sufficiently large initial time interval allowing a reasonable further extrapolation. The parameters used correspond to naturally supercoiled SV40 DNA (5243 bp) in dilute solution at a NaCl concentration of 0.1 M. All model parameters are known independently, and none are fitted. The simulated dynamic structure factor is in good agreement with earlier DLS measurements. The experimental data were then analyzed based on the new information provided by the simulations. In particular, a distribution function of relaxation times of internal motions could be obtained.

Preliminary Theoretical Considerations

Consider a DNA molecule in a dilute solution. Let the molecule axis be represented parametrically as $\mathbf{r} = \mathbf{r}(s, t)$, where t is the time and s is a dimensionless length parameter varying from 0 to 1. The vertically polarized monochromatic light scattered by such a molecule in the horizontal direction is given by the electric field

$$E(\mathbf{q}, t) = \int_0^1 \mathrm{d}s\, \mathrm{e}^{i\mathbf{q}\mathbf{r}(s,t)} \tag{1}$$

where $\mathbf{q}$ is the scattering vector (we use here arbitrary units and omit the frequency-dependent phase factor). The dynamic structure factor is defined as

$$S(\mathbf{q}, t) = \langle E(\mathbf{q}, 0)E^*(\mathbf{q}, t)\rangle \tag{2}$$

where the asterisk indicates the complex conjugate and the averaging is performed over the set of initial conformations. The dynamic structure factor normalized to 1 at $t = 0$,

$$g^{(1)}(\mathbf{q}, t) = \frac{S(\mathbf{q}, t)}{S(\mathbf{q}, 0)} \tag{3}$$

is of particular interest. The experimentally observed quantity is the intensity of light $I(\mathbf{q}, t)$ scattered by a small volume of the DNA solution. The intensity autocorrelation function normalized to 1 at $t = \infty$ can be shown to be related to $g^{(1)}(\mathbf{q}, t)$ as follows:[19]

$$g^{(2)}(\mathbf{q}, t) \equiv \frac{\langle I(\mathbf{q}, 0)I(\mathbf{q}, t)\rangle}{\langle I(\mathbf{q}, 0)\rangle^2} = 1 + \alpha[g^{(1)}(\mathbf{q}, t)]^2 \tag{4}$$

10.1021/ma9914467 CCC: $19.00
Published on Web 01/26/2000

where α is some equipment-dependent constant, which, in the ideal case, is equal to 1.

Let us rewrite eq 1 in the form

$$E(\mathbf{q}, t) = e^{i\mathbf{q}\mathbf{r}_c(t)}F(\mathbf{q}, t) \quad (5)$$

where $\mathbf{r}_c(t)$ is the center of mass of the molecule and

$$F(\mathbf{q}, t) = \int_0^1 ds\, e^{i\mathbf{q}\mathbf{R}(s,t)} \quad (6)$$

with $\mathbf{R}(s, t) = \mathbf{r}(s, t) - \mathbf{r}_c(t)$ representing the molecule axis in the local coordinate system attached to the center of mass. We assume that the diffusion of the center of mass is only weakly correlated to the internal motion, that is,

$$S(\mathbf{q}, t) = \langle e^{i\mathbf{q}(\mathbf{r}_c(0)-\mathbf{r}_c(t))}F(\mathbf{q}, 0)F^*(\mathbf{q}, t)\rangle \approx \langle e^{i\mathbf{q}(\mathbf{r}_c(0)-\mathbf{r}_c(t))}\rangle\langle F(\mathbf{q}, 0)F^*(\mathbf{q}, t)\rangle \quad (7)$$

This assumption is tested and verified later in this work (see Results and Discussion).

The exponential function in eq 7 can be substituted by its average over all possible angles between the vectors $\mathbf{q}$ and $(\mathbf{r}_c(0) - \mathbf{r}_c(t))$:

$$\langle e^{i\mathbf{q}(\mathbf{r}_c(0)-\mathbf{r}_c(t))}\rangle = \langle {}^1/_2\int_{-1}^1 e^{iq|\mathbf{r}_c(0)-\mathbf{r}_c(t)|\cos\theta}\, d(\cos\theta)\rangle = \left\langle \frac{\sin(q|\boldsymbol{r}_c(0) - \boldsymbol{r}_c(t)|)}{q|\mathbf{r}_c(0) - \mathbf{r}_c(t)|}\right\rangle = \int_0^\infty P_c(r, t)\,\frac{\sin(qr)}{qr}dr \quad (8)$$

where $q = |\mathbf{q}|$ and

$$P_c(x, t) = \langle \delta(x - |\mathbf{r}_c(0) - \mathbf{r}_c(t)|)\rangle \quad (9)$$

is the distribution function of the net distance traveled by the center of mass in the time interval t.

The internal motion factor in eq 7, which we denote as $Q(\mathbf{q}, t)$, can be treated in the same way:

$$Q(\mathbf{q}, t) \equiv \langle F(\mathbf{q}, 0)F^*(\mathbf{q}, t)\rangle = \int_0^\infty P_R{}^{(2)}(R, t)\frac{\sin(qR)}{qR}\, dR \quad (10)$$

$$P_R{}^{(2)}(x, t) = \langle \int_0^1 ds_1 \int_0^1 ds_2 \delta(x - |\mathbf{R}(s_1, 0) - \mathbf{R}(s_2, t)|)\rangle \quad (11)$$

where $P_R{}^{(2)}(R, t)$ is the dynamic site-to-site distance distribution function in the local coordinate system. The attribute *dynamic* means that the site positions are taken at different times.

The exact expression for the $S(\mathbf{q}, t)$ function (obtained without the assumption that the center of mass diffuses independently) can similarly be represented as

$$S(\mathbf{q}, t) = \int_0^\infty P_r(r, t)\frac{\sin(qr)}{qr}\, dr \quad (12)$$

$$P_r(x, t) = \langle \int_0^1 ds_1 \int_0^1 ds_2 \delta(x - |\mathbf{r}(s_1, 0) - \mathbf{r}(s_2, t)|)\rangle \quad (13)$$

Equation 12 for $t = 0$ is known as the Debye formula.[20]

The distributions $P_c(r, t)$, $P_R{}^{(2)}(R, t)$, and $P_r(r, t)$ are available directly from the BD simulations (provided the overall time interval is not very large). For $P_c(r, t)$, however, we will use an analytical expression. If we assume that the center of mass moves as a usual Brownian particle with the diffusion coefficient D, then

$$P_c(r, t) = \frac{4\pi r^2}{(4\pi Dt)^{3/2}}\exp\left(-\frac{r^2}{4Dt}\right) \quad (14)$$

and eq 8 gives

$$\langle e^{i\mathbf{q}(\mathbf{r}_c(0)-\mathbf{r}_c(t))}\rangle = e^{-q^2Dt} \quad (15)$$

Combining eqs 3, 7, 10, and 15, we get

$$g^{(1)}(\mathbf{q}, t) = e^{-q^2Dt}\frac{Q(\mathbf{q}, t)}{Q(\mathbf{q}, 0)} \quad (16)$$

The following empirical model function is often used for $g^{(1)}(\mathbf{q}, t)$:[21−23]

$$g^{(1)}(\mathbf{q}, t) = e^{-q^2D_{tr}t}(1 - a + ae^{-q^2D_{in}t}) \quad (17)$$

where D_{tr}, D_{in}, and a are fitted parameters depending on q but not on t. Although D_{tr} is simply the translational diffusion coefficient, $D_{tr} = D$, the interpretation of D_{in} and a is more difficult. These parameters are called the internal diffusion coefficient and the (relative) amplitude of internal motion, respectively. The independence of D_{in} from t is an approximation, which may be good or bad, whereas a is time-independent. The comparison of eqs 16 and 17 for $t \to \infty$ shows that

$$1 - a = \frac{Q(\mathbf{q}, \infty)}{Q(\mathbf{q}, 0)} = \frac{\langle F(\mathbf{q}, 0)F^*(\mathbf{q}, \infty)\rangle}{Q(\mathbf{q}, 0)} = \frac{\langle F(\mathbf{q}, 0)\rangle\langle F^*(\mathbf{q}, 0)\rangle}{Q(\mathbf{q}, 0)} \quad (18)$$

where $\langle F(\mathbf{q}, 0)\rangle$ can be found as

$$\langle F(\mathbf{q}, 0)\rangle = \int_0^\infty P_R{}^{(1)}(R)\frac{\sin(qR)}{qR}\, dR \quad (19)$$

$$P_R{}^{(1)}(x) = \langle \int_0^1 ds \delta(x - |\boldsymbol{R}(s, 0)|)\rangle \quad (20)$$

The function $P_R{}^{(1)}(R)$ represents the average distribution of sites around the center of mass. Note that the functions $P_R{}^{(1)}(R)$ and $P_R{}^{(2)}(R,0)$, required for calculation of a, can be obtained simply from MC simulations, which contain no information about the dynamics.

An alternative way to find a is the evaluation of $Q(\mathbf{q}, \infty)$ from eq 10 using the function $P_R{}^{(2)}(R,\infty)$, which can be expressed through $P_R{}^{(1)}(R)$ as follows:

$$P_R{}^{(2)}(R,\infty) = \int_0^\infty dR_1 P_R{}^{(1)}(R_1)\int_0^\infty dR_2 P_R{}^{(1)}(R_2)\,{}^1/_2\int_{-1}^1 d(\cos\theta)\delta(R - \sqrt{R_1{}^2 + R_2{}^2 - 2R_1R_2\cos\theta}) = \frac{R}{2}\int_0^\infty dR_1\frac{P_R{}^{(1)}(R_1)}{R_1}\int_{|R-R_1|}^{R+R_1} dR_2\frac{P_R{}^{(1)}(R_2)}{R_2} \quad (21)$$

In deriving eq 21 we use the fact that $|\mathbf{R}_1 - \mathbf{R}_2|^2 = R_1{}^2 + R_2{}^2 - 2R_1R_2\cos\theta$, where $\mathbf{R}_1 \equiv \mathbf{R}(s_1, 0)$, $\mathbf{R}_2 \equiv \mathbf{R}(s_2, 0)$, $R_1 = |\mathbf{R}_1|$, $R_2 = |\mathbf{R}_2|$, $\cos\theta = \mathbf{R}_1\mathbf{R}_2/R_1R_2$, and the three random variables, R_1, R_2, and $\cos\theta$, are mutually independent.

Expanding the sine in eq 10 in a Taylor series, we get

$$Q(\mathbf{q}, t) = 1 - \frac{q^2}{3!}2R_G{}^2 + \sum_{m=2}^{\infty}\frac{(-1)^m q^{2m}}{(2m+1)!}\langle R^{2m}(t)\rangle_2 \quad (22)$$

where R_G is the mean-square radius of gyration and

$$\langle R^{2m}(t)\rangle_2 = \int_0^\infty P_R{}^{(2)}(R,t)R^{2m}\,dR = \left\langle \int_0^1 ds_1 \int_0^1 ds_2 |\mathbf{R}(s_1,0) - \mathbf{R}(s_2,t)|^{2m}\right\rangle \quad (23)$$

is the $(2m)$th moment of the distribution $P_R{}^{(2)}(R,t)$ (as can easily be shown, the second moment $\langle R^2(t)\rangle_2$ does not depend on the time and is equal to $2R_G{}^2$). A set of the moments $\langle R^{2m}(t)\rangle_2$ for $m \le M$, where M is some not very large integer, can be found directly from BD simulations.

As follows from eqs 10, 19 and 20, the function $Q(\mathbf{q}, t)$ for $t \to \infty$ after the Taylor expansion of the sine can be expressed as

$$Q(\mathbf{q}, \infty) = \langle F(\mathbf{q}, 0)\rangle\langle F^*(\mathbf{q}, \infty)\rangle = \left[\sum_{m=0}^{\infty}\frac{(-1)^m q^{2m}}{(2m+1)!}\langle R^{2m}\rangle_1\right]^2 \quad (24)$$

where the moments

$$\langle R^{2m}\rangle_1 = \int_0^\infty P_R{}^{(1)}R^{2m}\,dR = \left\langle \int_0^1 ds |\mathbf{R}(s,0)|^{2m}\right\rangle \quad (25)$$

are available from MC simulations.

Comparing eqs 22 and 24, we get

$$\langle R^{2m}(\infty)\rangle_2 = \sum_{n=0}^{m}\frac{(2m+1)!}{(2m-2n+1)!(2n+1)!}\langle R^{2m-2n}\rangle_1\langle R^{2n}\rangle_1 \quad (26)$$

For small q, it follows from eqs 16 and 22 that

$$g^{(1)}(\mathbf{q}, t) = e^{-q^2Dt}\left[1 - \frac{q^4}{5!}(\langle R^4(0)\rangle_2 - \langle R^4(t)\rangle_2)\right] \quad (27)$$

As one can see from eq 27, at small scattering vectors, the amplitude of internal motions is proportional to q^4:

$$a = \frac{q^4}{5!}(\langle R^4(0)\rangle_2 - \langle R^4(\infty)\rangle_2) \quad (28)$$

As a summary of this introductory section, we note that there are several possible ways to obtain the function $g^{(1)}(\mathbf{q}, t)$ from the BD simulations. The first one is to get the distribution $P_r(r, t)$ (eq 13) for a set of times t_j ($j = 1, 2, ..., N_t$) and then calculate $g^{(1)}(\mathbf{q}, t_j)$ for arbitrary q according to eqs 3 and 12. This is, in principle, an exact method for the BD model, but it does not provide opportunities for further analysis. The other possibility is to get the distributions $P_R{}^{(2)}(R,t_j)$ (eq 11) and, in addition, to calculate the diffusion coefficient D of the center of mass. Then we can find $g^{(1)}(\mathbf{q}, t_j)$ using eqs 10 and 16. In the both cases, we have to approximate the distribution functions $P_r(r, t_j)$ and $P_R{}^{(2)}(R,t_j)$ with a histogram. If we are not interested in very large q values, we can avoid this kind of approximation: instead of the distribution $P_R{}^{(2)}(R,t_j)$, a set of its first several even moments $\langle R^{2m}(t)\rangle_2$ suffices (see eq 22).

For computing the amplitude of internal motions, a, no dynamic data are required and all of the necessary information can be extracted from the relatively less time-consuming MC simulations.

Methods

BD Model. Our BD algorithm is described in detail elsewhere.[14] Here, we present the principal points. The DNA molecule is modeled by a chain of N beads connected through straight elastic segments. To each segment, a local reference frame is attached with the z axis pointing along the segment. We use harmonic potentials with respect to (i) the angles between consecutive segments, (ii) the twist angles between consecutive reference frames, and (iii) the deviations from the equilibrium segment length. The volume interactions between segments are given by the Debye−Hückel potential. In one BD step, the jth bead is displaced by the vector

$$\Delta\mathbf{r}_j = \frac{1}{k_B T}\sum_{k=1}^{N}\mathbf{D}_{jk}\mathbf{F}_k\Delta t + \mathbf{Y}_j \quad (29)$$

where k_B is the Boltzmann constant, T is the absolute temperature, Δt is the time interval, $\mathbf{F}_j$ is the force acting on the jth bead ($\mathbf{F}_j = -\partial U/\partial\mathbf{r}_j$, with U being the total energy), $\mathbf{D}_{jk}$ is the Rotne-Prager 3×3 diffusion tensor depending on the positions of the jth and the kth beads, and $\mathbf{Y}_j$ is a Gaussian random displacement with the covariance $\langle\mathbf{Y}_j \otimes \mathbf{Y}_k\rangle = 2\mathbf{D}_{jk}\Delta t$ (the symbol $\otimes$ denotes the tensor product). Each local reference frame is then rotated about an axis perpendicular to both the old and the new positions of the corresponding segment until the z-axis is oriented properly. After that, the local reference frame is rotated about the z-axis by the angle

$$\Delta\varphi_j = \frac{D_r}{k_B T}T_j\Delta t + \Phi_j \quad (30)$$

where T_j is the torque acting on the jth segment ($T_j = -\partial U/\partial\varphi_j$), D_r is the rotational diffusion coefficient, and Φ_j is a Gaussian random angular displacement with the variance $\langle\Phi_j{}^2\rangle = 2D_r\Delta t$. Finally, second-order corrections are made for the displacements $\Delta\mathbf{r}_j$ and $\Delta\varphi_j$.

MC Model. The geometry and the energy of the MC model are the same as those for the BD model with the following two exceptions. First, the segment length is fixed, and second, there is an additional topological potential, which is equal to ∞ if the given chain conformation is knotted and vanishes otherwise. In most principal points, our MC algorithm coincides with that described by Vologodskii et al.[24] Two types of MC steps are used. (1) In the pivoting step, a subchain bounded by two randomly chosen beads is rotated about its end-to-end vector by the angle β uniformly distributed in an interval $(-\beta_0, \beta_0)$. The β_0 value is chosen in such a way that the acceptance rate of trial steps is $\sim 1/2$. A trial step is accepted with the probability[25]

$$p_{piv} = \min\left[1, \exp\left(-\frac{E_{new} - E_{old}}{k_B T}\right)\right] \quad (31)$$

where E_{old} and E_{new} are the old and the new energy values. To achieve better efficiency, the pivoting step is performed only for subchains satisfying the following condition: either the subchain contour length is less than 100 nm (two persistence lengths) or the end-to-end distance of the subchain is less than 60 nm ($\sim$5 superhelix diameters[26]). (2) In the so-called reptational step, a randomly chosen subchain of five segments is exchanged with a randomly chosen subchain of four segments. By this procedure, each subchain is deformed to assume the former end-to-end distance of its counterpart. The orientation of the center of mass with respect to the end-to-end axis for the both subchains is also reciprocally inherited. Subchains of five segments with an initial end-to-end distance exceeding four segment lengths are excluded from the possible choice. The deformation of the subchains is performed as a series of small moves. In one such move, each subchain segment $\mathbf{s}_j$ is

rotated about the vector $\mathbf{s}_j \times \mathbf{X}$ by a small angle proportional to $|\mathbf{s}_j \times \mathbf{X}|$, where $\mathbf{X}$ is the subchain end-to-end vector. The trial step is accepted with the probability

$$p_{\text{rept}} = \min\left[1, \frac{n_{\text{old}} P_4(\mathbf{X}_5) P_5(\mathbf{X}_4)}{n_{\text{new}} P_4(\mathbf{X}_4) P_5(\mathbf{X}_5)} \exp\left(-\frac{E_{\text{new}} - E_{\text{old}}}{k_B T}\right)\right] \quad (32)$$

where n_{old} is the total number of the subchains of five segments included in the possible choice, n_{new} is the same quantity for the new conformation, $\mathbf{X}_4$ and $\mathbf{X}_5$ are the initial end-to-end vectors of the subchains of four and five segments, and $P_4(\mathbf{X})$ and $P_5(\mathbf{X})$ are the distribution functions for the end-to-end vector of a linear free-joint chain consisting of four and five segments, respectively. In eq 32, the factor $n_{\text{old}}/n_{\text{new}}$ compensates the inequality of the a priori probabilities of the direct and reverse trial steps.[25] The factor $P_4(\mathbf{X}_5)P_5(\mathbf{X}_4)/P_4(\mathbf{X}_4)P_5(\mathbf{X}_5)$ is due to the entropy changes by adjusting the end-to-end distance of the subchains. The necessity of such an entropy factor can be illustrated by a simple example. Let us apply a Metropolis MC procedure to a free particle traveling on the surface of a sphere. If a random, symmetrically distributed displacement of the spherical angles (φ, θ) is taken as a trial step, then an incorrect result will follow, namely, that all the possible θ values are equally probable. In reality, the distribution function of θ is $P_\theta(\theta) = (^1/_2)\sin\theta$. The correct distribution can be obtained, however, if we add to the energy of the system the entropy term $-k_B T \ln P_\theta(\theta)$. In general, a change in the entropy must be taken into account whenever the states of the free system (i.e., the system with the energy $U \equiv 0$) are distributed nonuniformly along the generalized coordinate over which the trial step is performed.

Equation 32 is valid under the following two conditions. First, the total number of segments in the chain is large enough, so that the distribution functions $P_4(\mathbf{X})$ and $P_5(\mathbf{X})$ for linear free-joint chains are applicable. Second, the procedure of readjustment of the end-to-end distances $|\mathbf{X}|$ is "ideal"; i.e., among the entire set of generalized coordinates defining a subchain conformation, only the $|\mathbf{X}|$ coordinate is changed and all of the others are kept constant. It is very difficult to prove analytically that these two conditions are satisfied with sufficient accuracy. Instead of this, the described algorithm was tested numerically. We have simulated a short unbranched supercoiled plasmid with and without the reptational steps. The results for the mean values of the radius of gyration and the writhe were indistinguishable within statistical error (~0.1%).

Although the acceptance rate for the reptational step is very low (~1%), it is crucial for efficiency of the MC simulations of a branched supercoiled DNA. The fraction of this type of step in our calculations was ~$^1/_3$.

Scheme and Parameters of Simulations. The following parameters were used for modeling. The molecule length was 1780 nm, corresponding to SV40 DNA (5243 bp). The hydrodynamic and the electrostatic radii of DNA were both equal to 1.2 nm.[27,28] The linking number difference was $\Delta Lk = -25$ (superhelical density $\sigma = -0.05$),[16] the persistence length L_p = 50 nm, the torsional rigidity $C = 3.0 \times 10^{-19}$ erg cm, the NaCl concentration $I = 0.1$ M,[16] and the temperature T = 293.15 K. The elastic stretch modulus was assumed to be S_{sim} = 63 pN, whereas the experimental value is $S_{\text{exp}} = 1000$ pN.[29] We did not expect any essential dependence of the dynamic structure factor on S_{sim} and took the "softened" potential to reduce the computational time. The equilibrium segment length was $l_0 = 20$ nm, and the time step $\Delta t = 4$ ns. The above parameters correspond to a bead radius $r_b = 3.59$ nm. We have checked for a shorter DNA (~900 nm) that the dependence of the dynamic structure factor on the parameters S_{sim}, l_0, and Δt in the neighborhood of the chosen values is practically negligible.

It should be noted that the parameters used here to describe the physical properties of DNA are known with different accuracy. The most uncertain quantity is the torsional rigidity C, for which the values in the range $(2.0-3.0) \times 10^{-19}$ erg cm are usually reported. We have chosen the highest plausible value for a pure technical reason: the computational efforts in this case are slightly lower. Although the assumed set of parameters might not be optimal in terms of agreement between computation and experiment, finding an optimal set would require a very large amount of CPU time and is, therefore, beyond the scope of the present study.

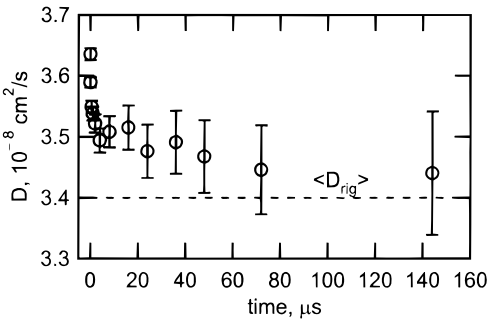

Figure 1. Diffusion coefficient of the center of mass as a function of the time t. The dashed line corresponds to the value $\langle D_{\text{rig}}\rangle = (3.40 \pm 0.01) \times 10^{-8}$ cm^2/s calculated by eq 35.

The scheme of the simulations was as follows. First, we used the MC method to get 400 "independent" starting conformations. Each conformation was obtained from the previous one in 2×10^5 MC steps. For successive conformations, the numbers of branches were correlated with a coefficient less than 1/e.[30] From each starting conformation, a BD trajectory of 36 100 steps was initiated. We began to collect data only after the first 100 BD steps, letting the segment lengths relax. For constructing the histograms for the distributions $P_r(r, t)$, $P_R^{(2)}(R,t)$, and $P_R^{(1)}(R)$, we substituted the integration in eqs 11, 13, and 20 by the summation over "scattering" points (20 scattering points per segment). In the direct calculation of the moments $\langle R^{2m}(t)\rangle_2$ and $\langle R^{2m}\rangle_1$, the scattering points coincided with the beads. The m value was in the range 1−30. The diffusion coefficient of the center of mass was obtained as

$$D(t) = \frac{1}{6t}\langle|\mathbf{r}_c(0) - \mathbf{r}_c(t)|^2\rangle \quad (33)$$

Overall, our simulations took ~2 months CPU time on an IBM SP2.

Results and Discussion

All our results correspond to supercoiled SV40 DNA in aqueous solution at a NaCl concentration of 0.1 M. The diffusion coefficient of the center of mass D as a function of the time is presented in Figure 1. At the initial time instant, $t = 0$, the value of D was obtained in the following way. When t is very small, we can neglect in eq 29 the term proportional to Δt in comparison with the random displacement, which scales, on the average, as $(t)^{1/2}$. From eqs 29 and 33, we have then

$$D(0) = \lim_{\Delta t \to 0} \frac{1}{6\Delta t\, N^2}\langle \sum_{j,k=1}^{N} \mathbf{Y}_j \mathbf{Y}_k\rangle = \frac{1}{3N^2}\sum_{j,k=1}^{N} \text{tr}\langle \mathbf{D}_{jk}\rangle \quad (34)$$

where tr denotes the trace of a tensor. For the Rotne-Prager tensor $\mathbf{D}_{jk}$, eq 34 reduces to the Kirkwood formula.[31]

Much more interesting for us is the limit $t \to \infty$. A quite good approximation for $D(\infty)$ is[32]

$$D(\infty) \approx \langle D_{\text{rig}}\rangle = \langle {}^1/_3 \text{tr}[\sum_{j,k=1}^{N} \mathbf{D}_{jk}^{-1}]^{-1}\rangle \quad (35)$$

where the inverse diffusion tensor $\boldsymbol{D}_{jk}^{-1}$ is defined by $\sum_{k=1}^{N} \mathbf{D}_{jk}^{-1}\boldsymbol{D}_{km} = \delta_{jm}\mathbf{I}$, $\mathbf{I}$ being the unit 3 × 3 tensor.

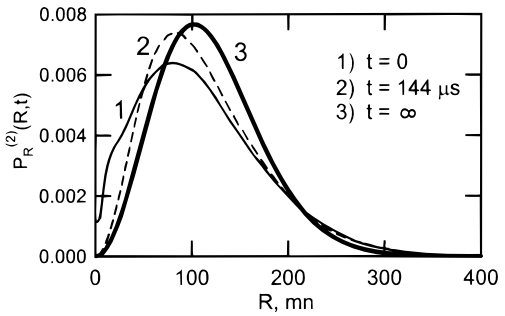

Figure 2. Dynamic site-to-site distance distribution function in the local coordinate system $P_R^{(2)}(R,t)$ for various time intervals t.

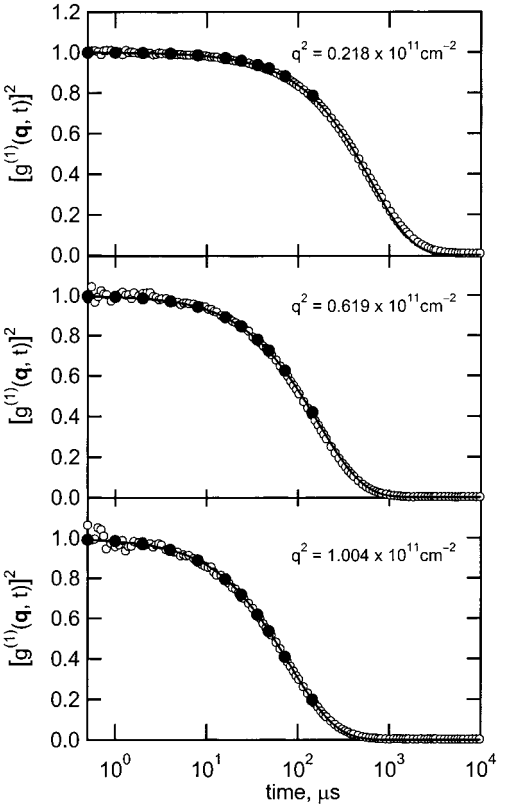

Figure 3. Square of the normalized dynamic structure factor $[g^{(1)}(\mathbf{q}, t)]^2$ vs time t for various scattering vectors q: experiment (○) and simulations (●). The extrapolation curves were calculated according to eqs 16, 22, 38, and 39 with the parameters from Table 1. The experimental data were normalized with an arbitrary constant.

Equation 35 represents the diffusion coefficient of a completely rigid chain, D_{rig}, averaged over chain conformations. In Figure 1, the $\langle D_{rig}\rangle$ value is shown with the dashed line: it is seen that $D(t)$ approaches $\langle D_{rig}\rangle$ for $t \rightarrow \infty$. For $t \geq 20$ μs, $D(t)$ and $\langle D_{rig}\rangle$ are practically indistinguishable because of the large statistical errors. In future calculations, we use the approximation $D(t) = \langle D_{rig}\rangle$ for all values of t.

The evolution of the distribution function $P_R^{(2)}(R,t)$ is illustrated in Figure 2. For $t = \infty$, it was calculated from $P_R^{(1)}(R)$ according to eq 21.

In Figure 3, our data are compared with the results of earlier DLS measurements. The experimental methods are described elsewhere.[16] The comparison between simulations and experiment is based on eq 4. The experimentally observed function $[g^{(2)}(\mathbf{q}, t) - 1]$, properly scaled, should be equal to the simulated function

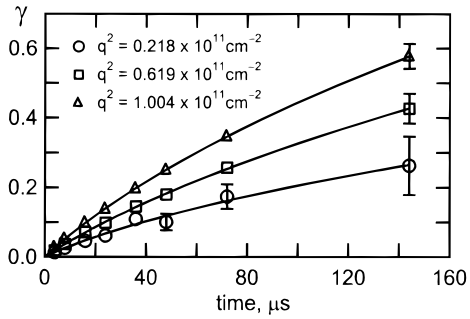

Figure 4. Exponent γ from eq 36 vs time t for various scattering vectors q. The interpolation curves are the best fits for the model function of eq 37. The parameters are $b = 0.155$, $c = 1.06 \times 10^3$ s^{-1} ($q^2 = 0.218 \times 10^{11}$ cm^{-2}); $b = 0.339$, $c = 3.28 \times 10^3$ s^{-1} ($q^2 = 0.619 \times 10^{11}$ cm^{-2}); and $b = 0.410$, $c = 5.61 \times 10^3$ s^{-1} ($q^2 = 1.004 \times 10^{11}$ cm^{-2});

$[g^{(1)}(\mathbf{q}, t)]^2$. Both quantities are plotted in Figure 3 against the time t for various q. The function $[g^{(1)}(\mathbf{q}, t)]^2$ was calculated using eqs 16 and 22 with the time-independent diffusion coefficient of the center of mass $D = \langle D_{rig}\rangle$. In the series of eq 22, the maximal value of m required for the evaluation of $Q(\mathbf{q}, t)$ does not exceed 18, provided that q belongs to the range available in the experiment ($q^2 \leq 1.1 \times 10^{11}$ cm^{-2}).

Now we can answer the question whether there is any essential correlation between translational and internal motions. The values of the function $[g^{(1)}(\mathbf{q}, t)]^2$ presented in Figure 3 were obtained using the assumption that these two types of motion are mutually independent. These values differ from the exact ones, calculated without such an assumption (eqs 3, 12), by no more than 1% (data not shown). Thus, the translational and internal motions are practically uncorrelated, and eq 7 holds with good accuracy.

The next problem to be solved is to find an interpolation/extrapolation procedure allowing the computation of $g^{(1)}(\mathbf{q}, t)$ for arbitrary t. The empirical expression for $g^{(1)}(\mathbf{q}, t)$ (eq 17) can be rewritten in a more general form:

$$g^{(1)}(\mathbf{q}, t) = e^{-q^2Dt}(1 - a + ae^{-\gamma}) \qquad (36)$$

where γ depends on both q and t. The plot of γ against t for different q^2 is shown in Figure 4. These data do not confirm the assumption that $\gamma = q^2D_{in}t$. Instead of this, the relation between γ and t can be remarkably well modeled by a parabola $t = B\gamma + C\gamma^2$, where B and C are some constants. Resolving this expression with respect to γ, we get

$$\gamma = -b + \sqrt{b^2 + ct} \qquad (37)$$

with the new constants b and c depending only on q but not on t. This relation solves, in principle, the problem of interpolation/extrapolation, because the parameters b and c can easily be fitted for any given value of q. We chose, however, another method that makes it possible to obtain a set of parameters independent of both t and q, although the number of required parameters is rather large. The starting point is eq 22. Here, the moment $\langle R^{2m}(t)\rangle_2$ can be expressed as

$$\langle R^{2m}(t)\rangle_2 = \langle R^{2m}(\infty)\rangle_2 + [\langle R^{2m}(0)\rangle_2 - \langle R^{2m}(\infty)\rangle_2]e^{-\gamma_m} \qquad (38)$$

where $\langle R^{2m}(\infty)\rangle_2$ is available from eq 26 and γ_m is a

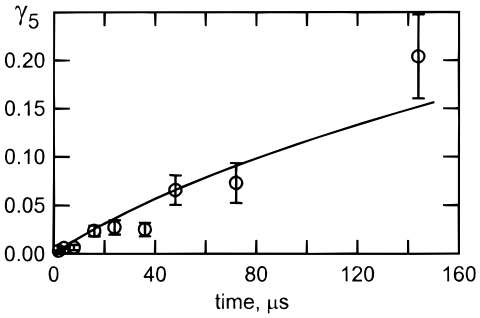

Figure 5. Exponent γ_5 from eq 38 vs time t. The interpolation curve was calculated according to eq 39 with the parameters from Table 1.

Table 1. Complete Set of the Parameters Required for the Calculation of the Normalized Dynamic Structure Factor $g^{(1)}(q, t)$[a]

m	$\langle R^{2m}(0)\rangle_2/(2R_G^2)^m$	$\langle R^{2m}(\infty)\rangle_2/(2R_G^2)^m$	b_m	c_m, s^{-1}
2	2.3289	1.7715	0.28024	2239
3	8.7505	4.5972	0.24654	1890
4	4.5176×10^1	1.5823×10^1	0.23958	1883
5	2.9098×10^2	6.7874×10^1	0.25331	2078
6	2.1993×10^3	3.4755×10^2	0.27633	2445
7	1.8744×10^4	2.0586×10^3	0.30564	2997
8	1.7535×10^5	1.3765×10^4	0.34196	3756
9	1.7667×10^6	1.0194×10^5	0.38403	4733
10	1.8904×10^7	8.2337×10^5	0.42928	5916
11	2.1257×10^8	7.1624×10^6	0.47547	7258
12	2.4916×10^9	6.6432×10^7	0.51978	8672
13	3.0252×10^{10}	6.5162×10^8	0.55935	10046
14	3.7854×10^{11}	6.7146×10^9	0.59218	11256
15	4.8617×10^{12}	7.2296×10^{10}	0.61623	12192
16	6.3877×10^{13}	8.0974×10^{11}	0.63084	12779
17	8.5620×10^{14}	9.4001×10^{12}	0.63592	12972
18	1.1680×10^{16}	1.1276×10^{14}	0.63085	12806

[a] According to eqs 16, 22, 38, and 39 for $q^2 \leq 1.1 \times 10^{11}$ cm^{-2}: $D = 3.40 \times 10^{-8}$ cm^2/s, and $2R_G^2 = 17\,741$ nm^2.

function of t. We assume that this function has a form similar to eq 37:

$$\gamma_m = -b_m + \sqrt{b_m^2 + c_m t} \tag{39}$$

where b_m and c_m are constants. Unfortunately, the plot of γ_m against t can neither confirm nor refute this assumption because of the large statistical errors (an example of such a plot for $m = 5$ is shown in Figure 5.) The complete set of the parameters b_m and c_m ($m \leq 27$) was fitted simultaneously to match the function $Q(\mathbf{q}, t)$ at the available time points and for a large number of q values ($q^2 \leq 2.5 \times 10^{11}$ cm^{-2}). All of the parameters (calculated and fitted) that are necessary for restoring the dynamic structure factor for arbitrary t and for q in the range of the experiment are presented in Table 1. Given these data, the function $g^{(1)}(\mathbf{q}, t)$ can be calculated according to eqs 16, 22, 38, and 39. The examples are shown in Figure 3.

The extrapolation curves were subjected to our traditional analysis for the treatment of experimental data: a two-exponential fit (eq 17) was applied, and the three parameters, a, D_{tr}, and D_{in}, were determined as functions of q. The results are presented in Figures 6 and 7 together with the corresponding experimental values. By fitting the data, we took into account that in the real experiment a small amount of dust, inevitably present in the sample, also contributes to the dynamic structure factor. Because the dust particles are relatively large and slow, this contribution does not change

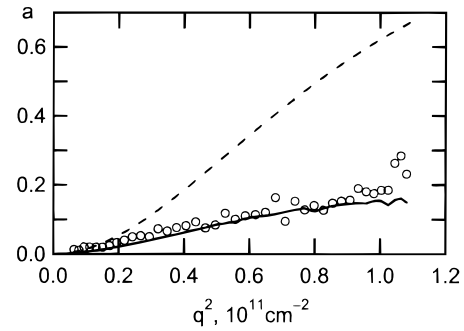

Figure 6. Amplitude of internal motions a as a fitted parameter from eq 40 vs square of the scattering vector q^2: experiment (○) and simulations (—). The true simulated values of a calculated according to eq 18 are shown with the dashed line.

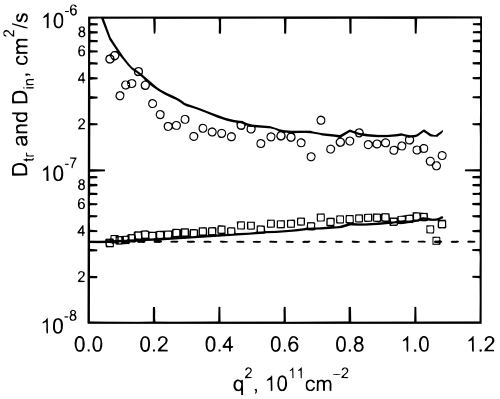

Figure 7. Translational D_{tr} (□) and internal D_{in} (○) diffusion coefficients as fitted parameters from eq 40 vs square of the scattering vector q^2: experiment (□, ○) and simulations (—). The true value of D_{tr}, used for the construction of the simulated dynamic structure factor (eq 35), is shown with the dashed line.

much in the time interval of interest and, hence, can be approximated with a time-independent constant. The model function for $g^{(2)}(\mathbf{q}, t)$ has, therefore, the form

$$g^{(2)}(\mathbf{q}, t) - 1 = [\alpha_{dust} + \alpha_{DNA} e^{-q^2 D_{tr} t}(1 - a + a e^{-q^2 D_{in} t})]^2 \tag{40}$$

with the two additional parameters α_{dust} and α_{DNA}. The data points at large t, where the contribution from dust prevails, were excluded from the fitting.

For the simulated data, the quantities a and D_{tr}, obtained as fitted parameters from eq 40, deviate considerably from their "true" values as given by eqs 18 and 35. The latter are shown in Figures 6 and 7 with the dashed lines. The discrepancy is because the internal diffusion coefficient D_{in} is assumed to be time-independent, whereas in reality, this is not the case. Thus, the traditional treatment of experimental data does not allow the proper separation of the contributions from the translational and internal motions to the dynamic structure factor. Unfortunately, we cannot propose a better fitting procedure. For example, if we substitute the exponent $-q^2 D_{in} t$ in eq 40 by the more realistic time function $b - (b^2 + ct)^{1/2}$ (eq 37), we would have six adjustable parameters. It is too many to provide a reliable estimation of each of them (given the available accuracy of the DLS measurements). Of course, the parameter D_{tr} can be fixed, but it does not help much.

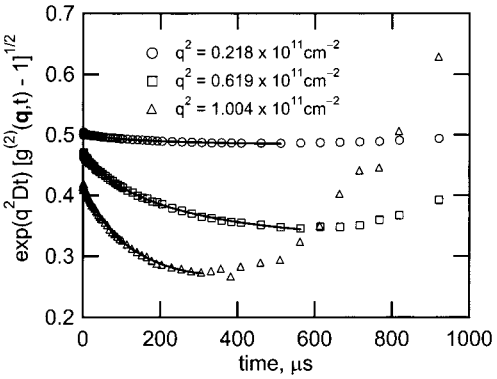

Figure 8. Experimental data, in the form of the left side of eq 41, as functions of the time t for various scattering vectors q. The best fit curves correspond to the following parameters: $a = 0.063$, $b = 0.340$, $c = 3.50 \times 10^3$ s^{-1} ($q^2 = 0.218 \times 10^{11}$ cm^{-2}); $a = 0.358$, $b = 0.772$, $c = 7.81 \times 10^3$ s^{-1} ($q^2 = 0.619 \times 10^{11}$ cm^{-2}); and $a = 0.624$, $b = 0.781$, $c = 10.33 \times 10^3$ s^{-1} ($q^2 = 1.004 \times 10^{11}$ cm^{-2}). The a values were fixed and taken from the MC simulations (dashed line in Figure 6).

This situation changes, however, if we take the a value from the MC simulations. It should be noted that the amplitude of internal motions a can be calculated with significantly higher accuracy than the exponent γ (eq 36), the reason being not only that the MC method is less time-consuming than the BD technique but also that the assumptions and approximations required for modeling dynamic behavior are not involved.

Now, we assume that the parameters a and D_{tr} are known and fixed. The translational diffusion coefficient D_{tr} is available from both DLS measurements and MC simulations. This fact can be used as a test for the MC model. In our case, $D_{tr} = 3.40 \times 10^{-8}$ cm^2/s, and a as a function of q^2 is presented in Figure 6 with the dashed line. Equation 40, after the substitution $-q^2D_{in}t \rightarrow b - (b^2 - ct)^{1/2}$, can be rewritten in the following way:

$$e^{q^2D_{tr}t}\sqrt{g^{(2)}(\mathbf{q}, t) - 1} = \alpha_{dust}e^{q^2D_{tr}t} + \alpha_{DNA}(1 - a + ae^{b-(b^2+ct)^{1/2}}) \quad (41)$$

Here, we have four adjustable parameters: α_{dust}, α_{DNA}, b, and c. The experimental data, represented in the form of the left side of eq 41, are plotted in Figure 8 as functions of the time t for various q. These functions have a characteristic minimum at a time instant t_{min} where the contribution from the dust becomes approximately equal to that of the DNA. The experimental points corresponding to $t > t_{min}$ do not contain much useful information and should therefore be excluded from the fitting. In Figure 8, the best-fit curves are shown only for the intervals that were used for their determination.

The information about the DNA dynamics is contained in the exponent $\gamma = -b + (b^2 + ct)^{1/2}$, which can easily be expressed through the experimentally determined function $g^{(2)}(\mathbf{q}, t)$ by means of eq 41 after fitting of the adjustable parameters. (The values of a and D_{tr} are taken from the MC simulations.) The exponent γ as a function of the time t for various q is presented in Figure 9. The parabolic curves in the form of eq 37 match the experimental points remarkably well, as was the case for the simulated data (cf. Figure 4). Now this kind of functional dependence can be traced up to the essentially higher values of the time t. The dashed lines in Figure 9 reproduce the best fit to the simulated data

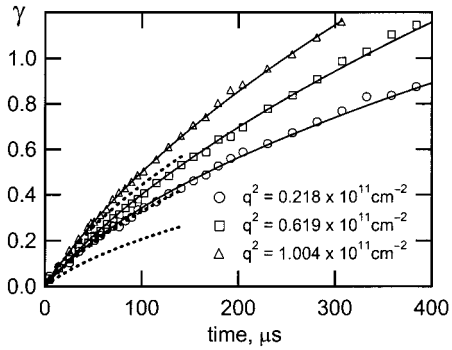

Figure 9. Experimental data for the exponent γ of the internal motion factor vs time t for various scattering vectors q. The parameters of the best fit curves (solid lines) are the same as in Figure 8. The dashed lines reproduce the best fit to the simulated data from Figure 4.

from Figure 4. For small q, the difference between the experiment and the simulations is quite noticeable. One should take into account, however, that the simulated data for small q in this representation have considerable statistical errors and the disagreement does not, in fact, exceed two standard deviations. At the same time, the experimental results at small scattering angles are very sensitive to any possible contaminations of the sample, e.g., by the open form of DNA.

An important problem in the analysis of the DLS data is to obtain the distribution function $\rho(\mathbf{q}, \tau)$ of relaxation times τ for the dynamic structure factor $g^{(1)}(\mathbf{q}, t)$. The function $\rho(\mathbf{q}, \tau)$ is defined by the following relation:

$$g^{(1)}(\mathbf{q}, t) = \int_0^\infty \rho(\mathbf{q}, \tau)e^{-t/\tau}\, d\tau \quad (42)$$

If we are interested only in the internal motion, it is natural to consider the distribution function $\rho_{in}(\mathbf{q}, \tau)$ for the internal motion factor, which can be defined by the formula

$$\frac{Q(\mathbf{q}, t)}{Q(\mathbf{q}, 0)} = 1 - a + a\int_0^\infty \rho_{in}(\mathbf{q}, \tau)e^{-t/\tau}\, d\tau \quad (43)$$

The simple analytical form of the internal motion factor,

$$\frac{Q(\mathbf{q}, t)}{Q(\mathbf{q}, 0)} = 1 - a + ae^{b-(b^2+ct)^{1/2}} \quad (44)$$

makes it possible to restore the function $\rho_{in}(\mathbf{q}, \tau)$ using the table of Laplace transformations:

$$\rho_{in}(\mathbf{q}, \tau) = {}^1/_2\sqrt{\frac{c}{\pi\tau}}\exp\left(b - \frac{c\tau}{4} - \frac{\sqrt{b}}{c\tau}\right) \quad (45)$$

The quantity $a\rho_{in}(\mathbf{q}, \tau)$ as a function of τ for various q is presented in Figure 10. The distribution $\rho_{in}(\mathbf{q}, \tau)$ is very broad. Its upper boundary lies in the range of milliseconds, in agreement with estimations by Jian et al.[15]

The presentation of our results would not be complete without the static structure factor $S(\mathbf{q}, 0) = Q(\mathbf{q}, 0)$ used as a normalizing function in the calculation of $g^{(1)}(\mathbf{q}, t)$. The data shown in Figure 11 demonstrate good agreement between the simulations and the experiment.

Conclusions

Using BD and MC techniques, we have calculated the dynamic structure factor for the supercoiled SV40 DNA in 0.1 M NaCl solution. Two modifications of the

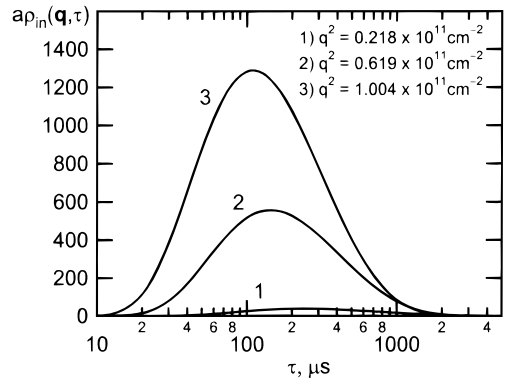

Figure 10. Distribution functions for relaxation times of intramolecular motions $\rho_{in}(\mathbf{q}, \tau)$ weighted by the amplitudes of internal motions a for various scattering vectors q. The curves were obtained using eq 45 with the same parameters as those in Figure 8.

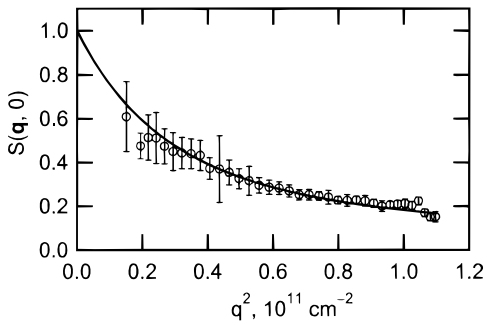

Figure 11. Static structure factor $S(\mathbf{q}, 0) = Q(\mathbf{q}, 0)$ as a function of the square of the scattering vector q^2: experiment (○) and simulations (—). The experimental data were normalized with an arbitrary constant.

computational scheme have been tested: one of them is exact in the frame of the BD model; the other implies that the internal and translational motions of the molecule are mutually independent. This assumption has been justified a posteriori by the fact that the results in the both cases were the same. The diffusion coefficient of the center of mass is, in principle, time-dependent, but from the practical point of view, it can be regarded as a constant equal to the diffusion coefficient of a rigid chain, averaged over chain conformations.

The limited computer capacity did not allow the direct computation of the dynamic structure factor for arbitrary large times. For the times exceeding ~150 μs, the simulated data were available only as extrapolations.

The calculated dynamic structure factor has been compared with the DLS measurements. For this comparison three types of data representation were used: (1) the "raw" experimental output, (2) the results of the two-exponential fit, and (3) the exponent of the internal motion factor. In the first two cases, the agreement between the theory and the experiment was good, whereas the third representation demonstrated a considerable difference between the two sets of data for small scattering vectors.

A very important parameter in the theory of DLS is the amplitude of internal motions a. Its reliable estimation from experimental data is a challenge, to which, at present, we have no solution. However, if the a value is available from MC simulations, then the distribution function for relaxation times of intramolecular motions can easily be extracted from DLS measurements. The range of the relaxation times for the supercoiled SV40 DNA is very broad. It comprises approximately 2 orders of magnitude, with the upper boundary exceeding 1 ms.

Acknowledgment. We thank Mickey Schurr for inspiring discussions that initiated this work.

References and Notes

(1) Ermak, D. L.; McCammon, J. A. *J. Chem. Phys.* **1978**, *69*, 1352−1360.
(2) Allison, S. A.; McCammon, J. A. *Biopolymers* **1984**, *23*, 167−187.
(3) Allison, S. *Macromolecules* **1986**, *19*, 118−124.
(4) Allison, S.; Austin, R.; Hogan, M. *J. Chem. Phys.* **1989**, *90*, 3843−3854.
(5) Chirico, G.; Langowski, J. *Macromolecules* **1992**, *25*, 769−775.
(6) Jian, H.; Vologodskii, A.; Schlick, T. *J. Comput. Phys.* **1997**, *73*, 123−132.
(7) Podtelezhnikov, A.; Vologodskii, A. *Macromolecules* **1997**, *30*, 6668−6673.
(8) Allison, S. A.; McCammon, J. A. *Biopolymers* **1984**, *23*, 363−375.
(9) Allison, S. A.; Sorlie, S. S.; Pecora, R. *Macromolecules* **1990**, *23*, 1110−1118.
(10) Allison, S. A. *Macromolecules* **1991**, *24*, 530−536.
(11) Chirico, G.; Langowski, J. *Biopolymers* **1994**, *34*, 415−433.
(12) Chirico, G.; Langowski, J. *Biophys. J.* **1996**, *71*, 955−971.
(13) Heath, P. J.; Gebe, J. A.; Allison, S. A.; Schurr, J. M. *Macromolecules* **1996**, *29*, 3583−3596.
(14) Klenin, K.; Merlitz, H.; Langowski, J. *Biophys. J.* **1998**, *74*, 780−788.
(15) Jian, H.; Schlick, T.; Vologodskii, A. *J. Mol. Biol.* **1998**, *284*, 287−296.
(16) Hammermann, M.; Steinmaier, C.; Merlitz, H.; Kapp, U.; Waldeck, W.; Chirico, G.; Langowski, J. *Biophys. J.* **1997**, *73*, 2674−2687.
(17) Berg, O. G. *Biopolymers* **1979**, *18*, 2861−2874.
(18) Soda, K. *Macromolecules* **1984**, *17*, 2365−2375.
(19) Chu, B. *Laser Light Scattering*, 2nd ed.; Academic Press: Boston, 1991.
(20) Tanford, C. *Physical Chemistry of Macromolecules*; Wiley: New York, 1961.
(21) Lewis, R.; Huang, J. H.; Pecora, P. *Macromolecules* **1985**, *18*, 944−948.
(22) Langowski, J.; Giesen, U.; Lehmann, C. *Biophys. Chem.* **1986**, *25*, 191−200.
(23) Fishman, D. M.; Patterson, G. D. *Biopolymers* **1996**, *38*, 535−552.
(24) Vologodskii, A. V.; Levene, S. D.; Klenin K. V.; Frank-Kamenetskii, M. D.; Cozzarelli, N. R. *J. Mol. Biol.* **1992**, *227*, 1224−1243.
(25) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, 1987.
(26) Hammermann, M.; Brun, N.; Klenin, K. V.; May, R.; Tóth, K.; Langowski, J. *Biophys. J.* **1998**, *75*, 3057−3063.
(27) Hagerman, P. J.; Zimm, B. H. *Biopolymers* **1981**, *20*, 1481−1502.
(28) Stigter, D. *Biopolymers* **1977**, *16*, 1435−1448.
(29) Baumann, C. G.; Smith S. B.; Bloomfield V. A.; Bustamante, C. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6185−6190.
(30) Langowski, J.; Kapp, U.; Klenin, K.; Vologodskii, A. *Biopolymers* **1994**, *34*, 639−646.
(31) Yamakawa, H. *Modern Theory of Polymer Solutions*; Harper and Row: New York, 1971.
(32) García de la Torre, J.; Bloomfield, V. A. *Q. Rev. Biophys.* **1981**, *14*, 81−139.

MA9914467