Estimating Friction Coefficients of Mixed Globular/Chain Molecules, such as Protein/DNA Complexes

Charles H. Robert
Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331 USA

ABSTRACT Existing methods for predicting translational friction properties of complex molecules start by explicitly building up their three-dimensional shape with spherical subunits. This treatment has been used especially for two types of systems: rigid assemblies and flexible chain molecules. However, many protein/DNA complexes such as chromatin consist of a small number of globular, relatively rigid, bound proteins interspersed by long stretches of flexible DNA chain. I present a higher level of treatment of such macromolecules that avoids explicit subunit modeling as much as possible. An existing analytical formulation of the hydrodynamics equations is shown to be accurate when used with the present treatment. Thus the approach is fast and can be applied to hydrodynamic studies of highly degenerate multiple equilibria, such as those encountered in problems of the regulation of chromatin structure. I demonstrate the approach by predicting the effect of a hypothetical unwinding process in dinucleosomes and by simulating the distribution of sedimentation coefficients for cooperative and random models for a chromatin saturation process.

INTRODUCTION

When analytical ultracentrifugation appeared in the mid-1920s, researchers had the choice of using either a sphere (Stokes equation) or an ellipsoid (Perrin equations) to represent molecules in solution. This limited menu was later augmented by "subunit" modeling techniques (Bloomfield et al., 1967) in which spherical subunits were used to build up the overall shape of a molecule. The subunit approach was derived from treatments by Kirkwood and Riseman (1948) and Kirkwood (1953) of random-coil chain molecules, and it is still used today. More recently it has been exploited in numerical simulations of flexible assemblies (Zimm, 1980; Dìaz et al., 1990), but a whole class of biological macromolecules falls at the limit of analyses of this sort: molecules with both globular regions and long flexible-chain regions, such as many protein/DNA complexes. Structural and functional questions that are currently being asked about these macromolecules can be an order of magnitude more complex than those of earlier hydrodynamics studies. I present here a means of treating such complex macromolecules at a higher level, which enables one to approach the pertinent questions. This article addresses DNA/protein complexes, although similar principles could be applied to other materials.

An example of a such a complex macromolecule is chromatin, the packaged form of DNA in the eucaryotic cell. The structure of the genetic material, especially the degree of packaging, is highly changeable and is regulated by the presence and relative amounts of a vast array of cellular factors. The mechanisms of this control are currently the subject of a great deal of physical and biophysical investigation in molecular biology (Hansen and Ausio, 1992; Leuba et al, 1994; Bustamante et al., 1994). These macromolecules are clearly too large and flexible for routine analysis by crystallography or nuclear magnetic resonance. But they are well suited to hydrodynamic study; in fact, sedimentation analysis has been a traditional tool for elucidating their structural features (van Holde, 1988). The more or less globular nucleosomes (1.5-2 superhelical turns of DNA wrapping a protein core) can be found to be dispersed along the DNA chain at intervals that can range from fewer than 40 bp to upwards of hundreds of basepairs. This spacing alone can control to a large extent the overall three-dimensional structure. The spacing depends in turn on factors such as the degree of saturation by nucleosomes and on the cooperativity of the saturation process.

Existing subunit modeling can be used for fully packaged chromatin because the regions of uncomplexed DNA are short and relatively stiff (e.g., Shaw and Schmitz, 1979). But to use these methods for the more open states would involve specifying explicit arrays of spheres and accounting for the flexibility of the assemblies through a Monte Carlo approach. Beyond the basic question of how best to represent the DNA helix by such subunit arrays (see Garcìa de la Torre et al., 1994), any implementation of this modeling for long chain regions would lead to 1) lots of spheres and 2) the requirement of solving the hydrodynamics equations for each of a multitude of conformations of the flexible array in order to adequately represent the thermally randomized structure. In principle, such exigencies can be accommodated by using more and more computer time for longer and longer chain regions. But in an equilibrium in which proteins can bind at a multitude of competing sites on the DNA, the number of distinct macromolecular species can easily reach into the thousands. Thus existing procedures rapidly become untenable for even the most dedicated facilities.

Received for publication 21 November 1994 and in final form 31 May 1995.

Address reprint requests to Dr. Charles H. Robert at his present address, Laboratoire de Biologie Structurale, Bât. 34-C.N.R.S., 91198 Gif-sur-Yvette, France.

0006-3495/95/09/840/09 $2.00

The "equivalent-spheres" approach I present here reduces a macromolecule with both globular and chain regions to an array of just a few spheres. Rigid globular regions are modelled by one or more subunits as before, but a chain region is represented by a single equivalent sphere whose frictional properties can be either calculated or assigned empirically by using existing hydrodynamic data. The average distance between each pair of the resulting frictional elements is then established by standard chain conformational calculations, and the average hydrodynamics of the system are solved. Because this higher level of treatment virtually eliminates contacting subunits, an earlier approximate solution of the hydrodynamics equations can be employed with a new accuracy. I demonstrate the new approach by simulating the effects of varying the number of turns in dinucleosomes. I also show how the approach can be applied to the problem of cooperativity in nucleosome assembly. A preliminary account of this approach has been presented (Robert and van Holde, 1992).

EQUIVALENT-SPHERES REPRESENTATION OF COMPLEX MACROMOLECULES

A molecule with mixed globular and chain regions will end up being treated as a small set of spherical friction elements. Fig. 1 *A* shows one way of treating a chromatin-like molecule with the present technique. Here each globular nucleosome is represented by a spherical subunit. The major simplification lies in replacing any uninterrupted stretch of free DNA chain by a single equivalent sphere. For a long chain region, the sphere might be described as a "cloud" of thermally condensed chain segments, not unlike the clouds invoked heuristically by Flory (1953) in describing long chain molecules. By definition, the equivalent sphere reproduces the average perturbation of the fluid flow of the chain at long distances; in the current method it is used to estimate the perturbation at all relevant distances. For shorter or stiffer chains, the spherical shape will of course be less descriptive of the true segment distribution. But I show in the first example that the result is quite good even for short chains. In addition, this simple approach helps avoid difficulties that result from more explicit modeling, as discussed below.

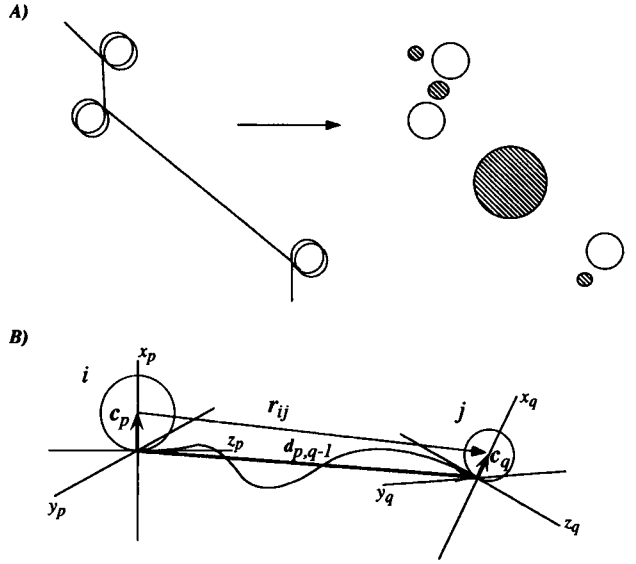

FIGURE 1 (*A*) Schematic of a chromatin-like molecule (—— indicates DNA chain) represented by the present treatment; (*B*) example of the center-to-center vector $\mathbf{r}_{ij}$ between two frictional elements *i* and *j*. Elements *i* and *j* are at segments *p* and *q*, respectively, of a superchain (see Appendix) in which the coordinate system of *q* is rotated with respect to that of *p* by the product of the intervening rotations. Intervening segments of the superchain are indicated schematically by the wavy line.

Globular regions of the complex

Globular regions of a complex are modelled as a rigid, compact assembly of one or more spheres, as in existing subunit approaches. The accuracy of this treatment will depend on factors such as the radii of the spheres being not too disparate (Bloomfield, 1967; Garcìa de la Torre and Bloomfield, 1977). In the examples presented below, each globular region is represented by a single sphere whose Stokes radius has been established empirically in separate studies.

Chain regions

Any uninterrupted chain region is represented by a single equivalent sphere. This sphere is conceptually placed on the axis of the chain it replaces, midway along the contour length. The Stokes radius of the sphere is taken as that of an isolated chain of the same composition and length. If the sedimentation behavior of the chain material is known, the appropriate Stokes radius can be obtained directly because the frictional coefficient, *f*, of a particular chain region is related to its sedimentation coefficient, *s*, by

$$f = \frac{M(1 - \bar{\nu}\rho)}{N_A s} \tag{1}$$

in which the fragment molecular weight is *M*, $\bar{\nu}$ is the partial specific volume of the chain material, ρ is the solution density, and N_A is Avogadro's number. The Stokes radius of a sphere with this frictional coefficient is

$$R = \frac{f}{6\pi\eta} \tag{2}$$

where η is the viscosity of the solvent.

Under typical salt conditions the sedimentation coefficient of a DNA fragment *n* bp long can be estimated from the work of Kovacic and van Holde (1977). The equation presented by these authors is reproduced here for convenience (*M* is the molecular weight of the chain in units of g/mol):

$$s = k(c_1 \log[M] + c_2 + c_3 M + c_4 M^2 + c_5 M^3) \tag{3}$$

where $k = 3.620\ 10^{-13}$ s, c_1 is defined as unity, c_2 is -3.552, c_3 is 3.23×10^{-7} mol/g, c_4 is 1.63×10^{-13} (mol/g)^2, and $c_5 = -7.85 \times 10^{-20}$ (mol/g)^3. The persistence length implicit in this equation is $a_o = 575$ Å, as reported in the original paper. Frictional coefficients calculated by Eq. 3 are not valid for short DNA chains; an equivalent-volume prolate ellipsoid can be used to estimate frictional coefficients for chains <50 bp.

In some situations, other values of the persistence length may apply (e.g., Hagerman, 1988). For a given value a, Eq. 3 can be modified by multiplying parameters c_3, c_4, and c_5 by the ratio a/a_o raised to the first, second, or third power, respectively. If a is not known, one can still examine the sedimentation coefficient for free DNA under the specified conditions and compare with the expected Kovacic and van Holde (1977) values; an apparent persistence length can be defined accordingly. Such reference measurements of the behavior of the naked DNA are typically part of a complete sedimentation study.

CALCULATING THE FRICTIONAL COEFFICIENT OF THE ASSEMBLY

The molecule, reduced now to a few spheres, will be sent translating under a suitable force through the fluid. Here the spheres will tend to shield one another from the full velocity of the passing solvent, decreasing the frictional coefficient of the assembly as a whole. This shielding is typically treated by the Oseen hydrodynamic interaction tensor. From there several different paths can be followed to solve the force equations. Bloomfield et al. (1967) solved the equations analytically by using approximate subunit forces. The limited accuracy of this approximation has been discussed by Garcìa de la Torre and Bloomfield (1977) and Zimm (1980), among others. However, it is quite accurate when used with the present treatment (see Example: Rigid Linear Structures) inasmuch as the spheres are generally noncontacting. Because it is calculated as a function of the average hydrodynamic interactions, the analytical result is further suited to the present approach because the equivalent sphere itself produces the average perturbation of the chain region.

Exact method

Garcìa de la Torre and Bloomfield (1977) presented an exact numerical solution of the hydrodynamics equations together with a somewhat simpler diagonal-matrix approximation that is exact for linear structures. Zimm (1980) reformulated the problem in simpler terms and compared several exact and approximate methods for treating random-coil chains. In the first example to be presented below, the equivalent spheres treatment will be tested for linear structures. The "true" frictional coefficient of the assembly will be calculated by using the conventional subunit modeling techniques with the diagonal method of Garcìa de la Torre and Bloomfield (1977), which is exact for these structures.

Analytical method

Bloomfield et al. (1967) presented an approximate solution of the force equations for an assembly of dissimilar spherical subunits, based on the model of Kirkwood (1953). It is simple and fast. For an assembly of N frictional spheres of radii R_i in a given conformation,

$$f = \frac{6\pi\eta \sum_{i=1}^{N} R_i}{1 + \frac{2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{R_i R_j}{r_{ij}}}{\sum_{i=1}^{N} R_i}} \tag{4}$$

where r_{ij} is the distance between spheres i and j.

For a flexible molecule, the conformational average $\langle 1/r_{ij} \rangle$ is called for because measured quantities (diffusion constant or sedimentation velocity) typically involve the inverse of the friction coefficient (Bloomfield et al., 1967). I will use a Monte Carlo approach to estimate this average in the second example. In more complex cases, calculations of $\langle r_{ij}^2 \rangle$ will be exploited, which can be evaluated in closed form for a given chain model. For short DNA chains, the difference between $\langle 1/r_{ij} \rangle$ and $1/\sqrt{\langle r_{ij}^2 \rangle}$ is negligible. For longer (or more flexible) chains the discrepancy will be greater, reaching a maximum of about 28% for the limiting case of a gaussian distribution. In order to account for this effect to first order, I performed Monte Carlo simulations of chain models used in this work. This produced the end-to-end distance distribution at each of a series of chain lengths. From these distributions the relevant averages were calculated. I then fitted the ratio between $1/\sqrt{\langle r^2 \rangle}$ and $\langle 1/r \rangle$ as a sigmoidal function of the chain length n, for 100, 300, 500, and 1000 bp chains, using $(1 + a_1(a_2 n)^m)/(1 + (a_2 n)^m)$ with a_1 fixed at the limiting value 0.723, finding $a_2 = 0.00204$ and $m = 1.94$ at the least-squares minimum error (standard error of a point 1.5%); this function was used to adjust root mean-square distances before substitution into Eq. 4.

All programs were developed in the Mathematica programming environment (Wolfram Research, Champaign, IL).

CONFORMATIONAL CALCULATIONS

In order to calculate the average distances to use in Eq. 4, it is necessary to describe how the different regions of the macromolecule are connected. A method for handling the general case to any desired degree of detail is described in the Appendix, but detailed conformational calculations are not always required or justified. In such cases one can proceed to calculate the root mean-squared distance between the two frictional elements directly from considerations of the center-to-center vector $\mathbf{r}_{ij}$ (Fig. 1 *B*).

Consider a spherical protein bound near the middle of a length of DNA, where the DNA chain is long compared with the size of the binding site. Treating this assembly yields three frictional elements: one sphere representing the protein and one sphere for each chain region on either side. Among them, three distinct pairwise distances are defined. These are r_{12} and r_{23}, which describe the distance from the protein center to the midpoints of the chain regions, and r_{13}, which spans from the midpoint of one chain region to the other. The distance r_{13} can be estimated from the root mean-square end-to-end distance of the DNA chain between the midpoints of the two chain regions. To estimate it, one can employ a model for DNA of a given contour length, the Porod-Kratky chain, which gives (Cantor and Schimmel, 1980)

$$\langle d^2 \rangle = a^2 2\lambda \left\{ 1 - \frac{1}{\lambda}(1 - e^{-\lambda}) \right\}$$

in which a is again the persistence length of the DNA under the appropriate conditions and λ is the contour length of the chain normalized to the persistence length.

For the remaining two distances, r_{12} or r_{13}, it can be remarked that the center of the sphere representing the protein will not lie on the chain axis unless the protein forms a symmetrical collar around the DNA. However, the vector pointing from the chain axis to the protein center has in general a component along the chain axis and one perpendicular to it, and the average can be broken down accordingly. For example, if the protein introduces no net bend, the mean projection of this center vector on any segment of the chain to either side will be zero, for there will be as many positive as negative terms in the average. Thus r_{12} or r_{23} can be estimated from the square root of the sum of the protein-center/DNA distance squared and the mean squared length of the DNA from the binding site to the midpoint of the chain region, calculated as in r_{13}.

EXAMPLE: RIGID LINEAR STRUCTURES

How well does the present treatment reproduce the frictional coefficients for known systems? Garcìa de la Torre and Bloomfield (1977) presented rigorous solutions of the hydrodynamic equations for linear subunit assemblies. I will apply the present treatment to the same geometries for comparison.

Garcìa de la Torre and Bloomfield (1977) analyzed colinear assemblies of subunits of two sizes, R_1 and R_2 (σ_1 and σ_2 in their paper), with $R_1 > R_2$. The "dumbbell" structure comprises two identical large subunits separated by a linear spacer region of length L, made of contiguous smaller subunits. The "lollipop" is essentially the same as the dumbbell but has only one large subunit. Such structures are shown in Fig. 2 *A*. The authors calculated frictional coefficients of these assemblies for a series of values of subunit radii and spacer lengths. They conveniently represented their solutions by polynomials, which give the frictional

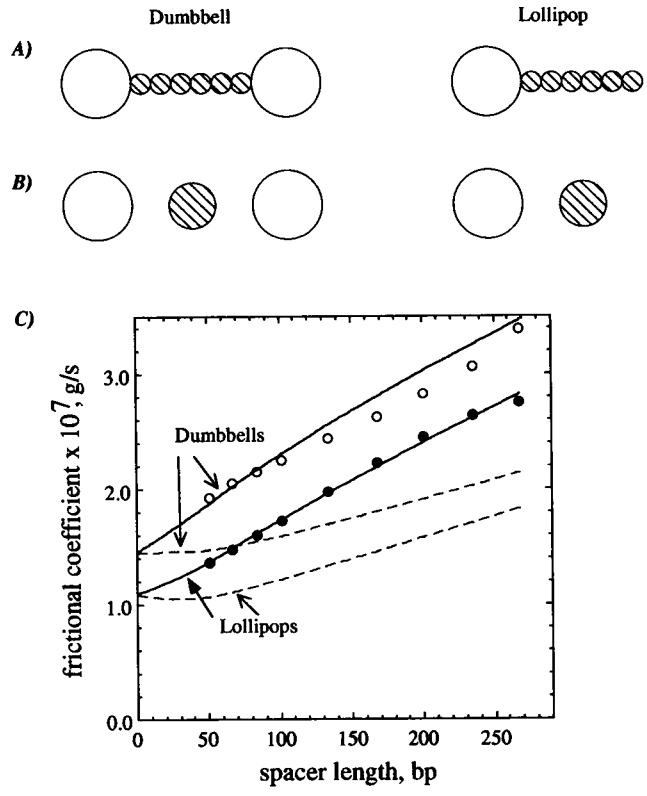

FIGURE 2 (*A*) The two types of subunit assemblies treated exactly by Garcìa de la Torre et al. (1977); (*B*) representation of these assemblies by the present treatment; (*C*) comparison of translational friction coefficient for a series of spacer lengths. Data points are the "exact" values calculated from the best-fit polynomial given in Garcìa de la Torre et al. ○ indicate dumbbell structures; ● indicate lollipop structures; —— indicates the results of Eq. 4 when used with the present treatment; and – – – – indicates the results of Eq. 4 applied in the ordinary way to the full subunit arrays.

coefficients of dumbbells and lollipops as a function of spacer length ($0 < L < 16R_1$). I have used their equations to simulate frictional coefficients for structures in which R_1 = 57 Å and R_2 = 16.5 Å for the spacer spheres. The larger spheres are roughly the size of a nucleosome, and the volume of a linear array of the smaller spheres will be equivalent to that of a cylindrical DNA molecule of the same length and with a diameter of 27 Å (the same value used in the study of Kovacic and van Holde, 1977). The frictional coefficients for the two types of structures are presented as points in Fig. 2 *C* as a function of spacer length; this value is given in bp units ($n = L/$ 3.4 Å).

In their paper, Garcìa de la Torre and Bloomfield (1977) compared their exact results to those obtained by using the Bloomfield et al. (1967) analytical formula (see Eq. 4). They showed it to be a poor predictor of the frictional coefficient for the linear assemblies, in error by as much as 30% from the rigorous values. I have reproduced this comparison by plotting frictional coefficients calculated by Eq. 4 for the arrays shown in Fig. 2 *A* (*dashed curves* in Fig. 2 *C*). Clearly Eq. 4 performs poorly with the explicit subunit modeling, markedly underestimating the frictional coefficient.

Now the present treatment is applied to the assemblies. The spacer region, composed of an array of contacting spheres, is replaced by a single equivalent sphere, which is placed at the center of the chain. Because the dimensions of the assemblies are consistent with the chain model used by Kovacic and van Holde (1977), Eqs. 2 and 3 can be used in the limit of an infinite persistence length to assign the Stokes radius for the rigid spacer region. The resulting sets of frictional elements are shown schematically in Fig. 2 *B*. Equation 4 is then used to calculate the frictional coefficient by using the new set of elements. The calculated values (*solid lines* in Fig. 2 *C*) compare well with the rigorous solutions. Replacing the linker region by an equivalent sphere gives a good approximation to the hydrodynamic properties of the assembly as a whole.

Range of applicability of analytical formulation

It has long been known that Eq. 4 commonly fails to provide accurate estimates of translational frictional coefficients. So why does it work with the present treatment? As Zimm (1980) points out, the analytical solution of the hydrodynamic equations by Kirkwood (1953) essentially assumed that the magnitudes of the forces $\mathbf{F}_i$ used for the velocity perturbations could be approximated by $1/N$ for N identical particles. The corresponding assumption in the Bloomfield et al. (1967) formulation for nonuniform particles (with friction coefficients ρ_i) can be seen to be $\mathbf{F}_i = \rho_i/\Sigma\rho_i$. I will refer to the assumption generally as the K/B approximation. It is correct in the limit of a weak hydrodynamic interaction; i.e., elements at long distances. It is therefore better in some situations than in others.

A way to see in detail the consequences of the K/B approximation is to examine the forces acting on any given particle from the other particles. For an assembly in a particular orientation, rearranging the velocity equation (Eq. 3 in Zimm, 1980) to express the total force on a subunit i gives

$$\mathbf{F}_i = \rho_i \mathbf{u} - \rho_i \sum_{j \neq i} T_{ij} \mathbf{F}_j$$

where $\mathbf{u}$ is the sedimentation velocity of the assembly and $\mathbf{T}$ is the Oseen tensor. Solution of the entire set of $3N+4$ hydrodynamics equations for a unit sedimenting force as prescribed by Zimm (1980) provides the true forces and the sedimentation velocity for an assembly in a given orientation. This can be done for any of the subunit arrays just described. If the true forces are substituted right back into these force equations, the sum of all of the equations gives (of course) exactly unit force, as required for consistency. To see the effects of the K/B approximation, we can substitute the approximate forces into the right-hand side and then add the equations together. The difference is a measure of the inconsistency of the approximation for a given array of subunits. For the subunit arrays in lollipop geometries, the discrepancy of the K/B approximation is on the order of 25%, consistent with the deviations seen in Fig. 2 *B*. Looking at each term in the matrix of perturbations demonstrates that the error arises mainly from contacting subunits, as expected.

The equivalent-spheres treatment, on the other hand, usually results in a small number of completely noncontacting frictional elements, so that in principle the K/B approximation is better suited to it. Indeed, Fig. 3 shows that the K/B forces become essentially indistinguishable from the true forces once the chain region becomes long enough that its equivalent sphere clears the point of contact with the nearest globular region. For a given geometry, this point can be calculated from the interception of the Porod-Kratky equation and Eq. 3 (it is ~11 bp for the linear geometries in these cases).

EXAMPLE: FLEXIBLE DINUCLEOSOMES

A question of current biological interest is how eucaryotic genetic material is bundled into chromatin at the various stages of development of the cell and how it can be accessed by regulatory factors (see e.g., Hansen and Ausio, 1992). Parameters such as the number of turns of DNA on the nucleosome can affect, for example, the superhelical tension in a given topological domain of a chromosome or the through-space distance between different control regions of the DNA. Because of the large size and flexibility of chromatin, such phenomena are not readily approached by crys-

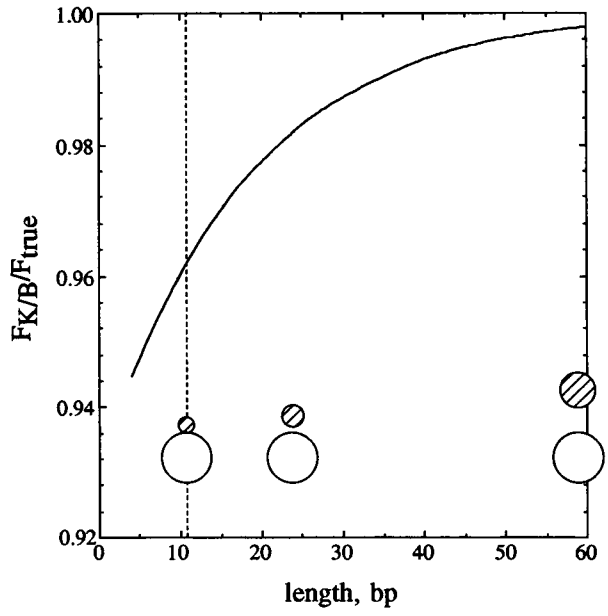

FIGURE 3 Total hydrodynamic force estimated with the Kirkwood (1953)/Bloomfield et al. (1967) approximation, $F_{K/B}$, relative to the true value, F_{true}, calculated by numerical solution of the force equations, for a series of equivalent-spheres representations of lollipop geometries as in Fig. 2. Schematic representations are shown for selected chain lengths. For DNA with a persistence length of 510 Å, the spheres are calculated to come into contact when the chain region is 11 bp (· · ·). Values are calculated for assemblies oriented perpendicular to the direction of sedimentation.

tallographic or spectroscopic means; rather, hydrodynamic techniques have been called on. Thus far the hydrodynamics have been quantitatively interpreted when the linker DNA is short and thus relatively rigid, so that subunit modeling can be exploited (Schmitz and Shaw, 1977). The following examples will show how the present treatment may be of use even when the linkers are longer, as is the case if nucleosome unwinding is hypothesized.

The nucleosome core particle is defined as an octameric protein (molecular weight of 108,500 g/mol) wrapped approximately 1.75 times with 146 bp of DNA (van Holde, 1988). The DNA is tightly bound to the octamer and spirals in a left-handed superhelix around the axis of the nucleosome core at a radius of 45 Å. The average helical repeat for the DNA on the nucleosome can be estimated from the work of Hayes et al. (1991) to be around 10.15 bp/turn, whereas the helical repeat of free DNA is typically about 10.4 bp/turn. The core particle exhibits an apparent frictional coefficient ratio (f/f_o, where f_o is the frictional coefficient for an equivalent-volume sphere) of 1.5 (Ausio et al., 1989), which corresponds to an equivalent Stokes radius of 57 Å.

To demonstrate how the present treatment might be applied to such a system, I placed a spherical friction element of this radius at the center of each nucleosome to represent its frictional properties. If more (or less) than 1.75 turns of DNA was wrapped in the nucleosome, the Stokes radius of the sphere was increased (decreased) by an amount corresponding to the additional volume of DNA wrapped on the helical path, which assumes that the frictional coefficient ratio for the nucleosome would remain approximately constant. The length of the DNA chain between adjacent nucleosomes (the linker length) was calculated from the total length of DNA, less that wrapped on the helical path around the particle, and the linker-DNA equivalent sphere was calculated as described above.

Analytical $1/\sqrt{\langle r^2 \rangle}$ versus Monte Carlo $\langle 1/r \rangle$ calculations

When a flexible system is considered by Eq. 4, the inverse distances from rigid geometry calculations must be replaced by the conformational averages. To estimate the magnitude of the error resulting from replacing $\langle 1/r \rangle$ in Eq. 4 with the more easily calculated $1/\sqrt{\langle r^2 \rangle}$ used for this example, I applied a Monte Carlo approach to generate a sample of dinucleosomes with a linker of 70 bp. The linker DNA was modelled as an isotropically bending chain with a persistence length of 510 Å. The bending direction at each bp was assigned to a random variable scaled between 0 and 2 π. After a few hundred chains, the average squared magnitude of the center-to-center vector, which is the largest span in the dinucleosome and is where the maximum effect of flexibility will be seen, had converged to within 1.5% of its analytical value. For the sample of chains, I then calculated the inverse of the root mean-square distance as well as the average inverse distance between nucleosomes. The discrepancy between the two different averages was found to be less than 3%.

Simulations and comparison with experimental measurements

I have simulated a hypothetical dinucleosome in which the two nucleosomes are positioned at the ends of a DNA template with a total length similar to that just used in the Monte Carlo simulations. The translational diffusion coefficient for this assembly was calculated as a function of the number of superhelical turns and is plotted in Fig. 4. Changing the number of turns obviously changes the geometry of the dinucleosome dramatically, as can be seen in the small figures associated with the plot (which are drawn for clarity as if the DNA chain regions were perfectly stiff). The predicted diffusion coefficients increase, reflecting the progressive compaction as the DNA is wrapped further onto the nucleosomes. Of course, conventional subunit modeling could be used (Schmitz and Shaw, 1977) for the more compact dinucleosomes.

The results show a suggestive correspondence with the salt effect seen in a light scattering study from the Widom group (Yao et al., 1991), for which representative results are indicated on the right-hand scale in Fig. 4. A direct comparison cannot be made because here the total DNA length (354 bp) is shorter than that estimated by Yao et al. inasmuch as the possibility of tails at either end of the dimer have been neglected. There are clearly other possibilities that could account for the salt effect as well, such as a change in the dimensions of the core particle. Nevertheless, this example demonstrates that the current method can be applied to interpreting sedimentation and light-scattering data in relevant structural terms.

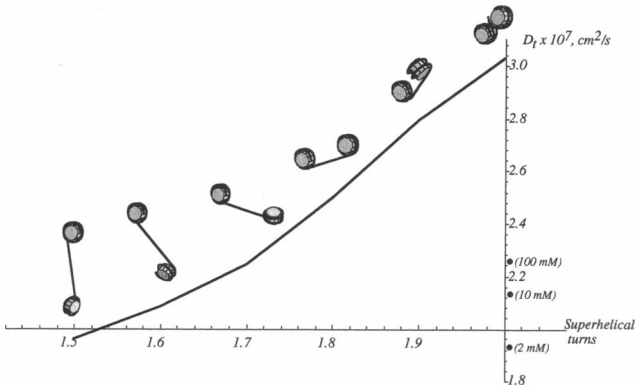

FIGURE 4 Simulation of the translational diffusion coefficient for flexible dinucleosomes (354 bp total DNA length) plotted versus the number of superhelical turns on the nucleosome. ● at right indicate the range of values for dinucleosomes seen experimentally by Yao et al. (1991), where values were reported as translational diffusion coefficients at 23°C with error ± 0.06 cm²/g. Figures are drawn as rigid dinucleosomes for each calculated number of turns: from left, 1.5, 1.6, 1.7, 1.8, 1.9, and 2.0 turns. Other parameters are as given in text.

EXAMPLE: CALCULATION OF CHROMATIN SATURATION INTERMEDIATES

The present method could be of most benefit in using hydrodynamics to help understand highly degenerate protein/DNA binding systems. For the progressive compaction of DNA by nucleosomes this degeneracy can be staggering. Even assuming discrete nucleosome binding at positioning sites, such as those found repeated in the 5S ribosomal RNA gene in various organisms (Hansen and Ausio, 1992), the number of saturation intermediates is 2^N for a template consisting of N sites. So in the saturation of a DNA chain consisting of 12 positioning sites, there are at minimum 4096 possible species, each a flexible assembly with distinct hydrodynamic properties. In order to examine whether the saturation of a given DNA template occurs randomly, or if alternatively there is either cooperative or anticooperative behavior, one must be prepared to account for each of these species in the equilibrium distribution. This would not be feasible with explicit subunit modeling techniques. In this example I will show how the present technique can be applied to such a problem. The system consists of the progressive saturation of a DNA template composed of three nucleosome positioning sequences, which would in principle allow an investigator to explore nearest-neighbor and next-nearest-neighbor effects in the saturation process. The approach can clearly be extended to longer templates as desired.

The nucleosome geometry here is the same as that described in the previous example. The template DNA is taken as 3 × 208 bp of DNA. Nucleosomes are allowed to form only on the first w bp of each segment, where w is calculated as a function of the number of turns of DNA wrapped on the nucleosome; for example, 1.75 turns gives w = 146 bp. The total number of species in equilibrium for such a system is eight, corresponding to the free DNA, the DNA fully occupied with three nucleosomes, and all intermediates. In this example, because of the longer chain distances, the first-order correction described previously (see the section on Calculating the Frictional Coefficient of the Assembly) is applied to distances obtained from the root mean-square calculation.

Table 1 shows the various species together with their predicted translational friction and sedimentation coefficients. In any equilibrium, each species will be present to some extent, but their relative concentrations will depend on the cooperativity of the process (Wyman and Gill, 1990). A perfectly cooperative binding will reveal only the unligated and fully ligated species in complementary proportions. On the other hand, a random or noncooperative assembly process will give statistical ratios between the concentrations of intermediates at different saturations. The procedure of van Holde and Weischet (1978) can be employed to obtain experimentally the sedimentation coefficient distribution at a given nucleosome saturation from a sedimentation velocity run. I present the prediction of such an experiment in Fig. 5. The present approach enables one to predict the large difference between these two models for DNA packaging. The approach can be extended in a straightforward manner to far more complex equilibria.

SUMMARY

I have presented a procedure that can be used to estimate translational friction coefficients of macromolecules with both globular regions and flexible chain regions. Such structures can be difficult to treat efficiently by using existing formalisms. The present treatment exploits existing measurements for those parts of the assembly that present the most difficulty to existing approaches. This results in a procedure that sidesteps the problems that typically compromise analytical hydrodynamics calculations. Consequently, one can approach complex macromolecules while exploiting analytical methods to accurately estimate their frictional properties. This provides several advantages; for instance, all of the computations presented here were able to be performed on a personal computer.

Beyond this convenience, however, the higher level of treatment allows higher level biological questions to be approached. The examples show that the approach is suited to interpreting hydrodynamic studies of chromatin under biologically pertinent conditions. In principle, other protein/

TABLE 1 Nucleosome saturation of a DNA chain of three 208-bp positioning sites (20°C)

Species*	Statistical weight‡		Friction coefficient, 10^{-7} g/s	Sedimentation coefficient, 10^{-13} s
	Random§	Cooperative¶		
Free DNA	$(1-\Theta)^3$	$1-\Theta$	3.75	8.1
{1}	$\Theta(1-\Theta)^2$		3.70	9.5
{2}	$\Theta(1-\Theta)^2$		3.44	10.2
{3}	$\Theta(1-\Theta)^2$		3.40	10.3
{1, 2}	$\Theta^2(1-\Theta)$		2.87	13.8
{1, 3}	$\Theta^2(1-\Theta)$		3.25	12.2
{2, 3}	$\Theta^2(1-\Theta)$		2.58	15.3
{1, 2, 3}	Θ^3	Θ	2.16	20.4

* Species are defined by the occupied site in a chain of three sequential sites.
‡ Given as a function of the average fractional saturation, Θ.
§ Binding to identical and independent sites (Wyman and Gill, 1990).
¶ All-or-nothing, limiting cooperativity model (no intermediates).

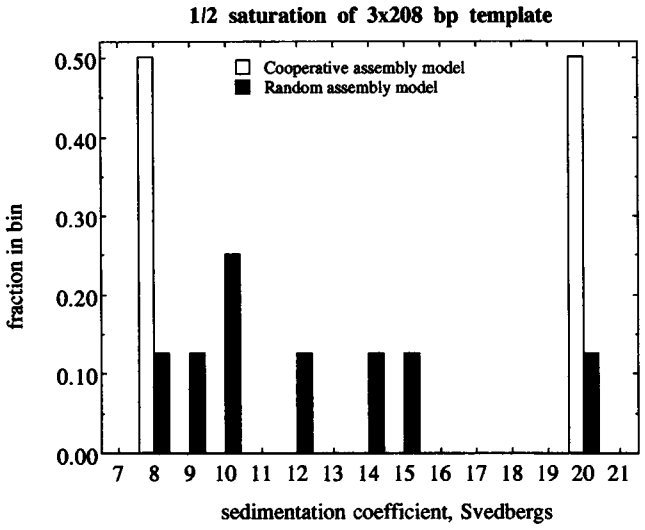

FIGURE 5 Sedimentation coefficient distribution of a 3 × 208 bp DNA template at half saturation by positioned nucleosomes for two different models of the cooperativity of assembly, as calculated from Table 1. Open bars show the predicted result for the limiting case of perfectly cooperative, "all or nothing" assembly (i.e., no populated intermediates); solid bars show the results for a "random," or noninteracting-sites, model for the saturation of identical and independent sites.

DNA systems can be treated, and the treatment may be extended to polyribosome arrays, etc.

I am especially grateful to Ken van Holde for discussions and for his kind support through PHS Grant GM-22916. The completion of this work was made possible through support from the John Simon Guggenheim Memorial Foundation.

APPENDIX: CONFORMATIONAL CALCULATIONS

Defining the "superchain"

The whole molecule consisting of rigid globular regions and flexible-chain regions can be treated as a single superchain of segments in space. Each segment is of type α or β. Type α refers to a single residue in a chain region; type β indicates a segment that traverses a globular region, connecting one or two chain regions. Thus the molecule is a formed by a continuous chain of segment vectors $\boldsymbol{\nu}_k$, each corresponding to one or the other type.

If a particular segment has an associated frictional effect, a center vector $\mathbf{c}_k$ is defined to describe the point at which it acts. As with $\boldsymbol{\nu}_k$, $\mathbf{c}_k$ is either of type α or of type β. For any segment representing a globular (β) region, $\mathbf{c}_k$ points to the center of mass in the reference coordinate system of $\boldsymbol{\nu}_k$. However, in a chain region a center vector is required only at the central residue because this is where the frictional effect of the entire region is attributed. Here $\mathbf{c}$ will be either a zero-length vector or one-half of a single-residue vector, depending on whether an odd or an even number of chain residues makes up the region.

Pairwise distance calculations

Any friction element i or j of the assembly ($1 \leq i, j \leq N$) is associated with a particular segment p or q, respectively, of the superchain ($1 \leq p, q \leq n$). The distance between any two friction elements is just the magnitude of the center-to-center vector $\mathbf{r}_{ij}$. Fig. 1 *B* shows that in general there will be three parts to this vector. These are the vector $\mathbf{d}_{pq-1}$, which spans the intervening superchain, together with the center vectors $\mathbf{c}_p$ and $\mathbf{c}_q$ pointing from the chain axis to the center of each sphere. The complete center-to-center vector between sphere i (at segment p) and sphere j (at segment q) is thus

$$\mathbf{r}_{ij} = -\mathbf{c}_p + \mathbf{d}_{pq-1} + \mathbf{c}_q \tag{5}$$

in which

$$\mathbf{d}_{pq-1} = \sum_{k=p}^{q-1} \boldsymbol{\nu}_k \tag{6}$$

The center vector of the jth element (qth segment) can be represented in the coordinate system of the ith element (the pth segment) by rotating the appropriate reference vector with the transformation matrix $\mathbf{A}_{pq}$ to be defined below. Thus in Eq. 5

$$\mathbf{c}_p = \mathbf{c}_p^o, \quad \text{and}$$

$$\mathbf{c}_q = \mathbf{A}_{pq}\mathbf{c}_q^o$$

The segment vectors composing $\mathbf{d}_{pq-1}$ are similarly defined in terms of their reference vectors by transformations of the form

$$\boldsymbol{\nu}_k = \mathbf{A}_{pk}\boldsymbol{\nu}_k^o$$

The matrices $\mathbf{A}$ are products of stepwise transformation matrices $\mathbf{T}$ for each vector of the superchain:

$$\mathbf{A}_{pk} = \prod_{h=p}^{k} \mathbf{T}_h$$

where $\mathbf{T}_p$ is defined as the identity matrix.

The stepwise matrices are defined according to the particular relationships between segments of the superchain. For current purposes there are two distinct matrix types: $\mathbf{T}_{\alpha\to\alpha}$ describing chain → chain junctions and $\mathbf{T}_{\beta\to\alpha}$ describing globular region → chain junctions. I will present examples of these matrices for isotropically bending chains.

Once $\mathbf{r}_{ij}$ is constructed, the evaluation of $\langle \mathbf{r}_{ij} \cdot \mathbf{r}_{ij} \rangle$ can be performed by numerical evaluation, explicitly summing the overlaps of all vectors in $\mathbf{r}$ with each other by using the averaged transformation matrices.

Transformation matrix for an isotropic chain

A simple model for the DNA chain involves, for each bp step, a fixed axial twist about the z axis (τ) followed by a fixed tilt away from the chain axis (γ), the direction (ϕ) of this bend being uniformly distributed in the xy plane ($0 \leq \phi < 2\pi$). The stepwise matrix for such a model is $\mathbf{T}_{\alpha\to\alpha} = \mathbf{S}_z(-\phi)\mathbf{S}_x(\gamma)\mathbf{S}_z(\tau+\phi)$, in which the matrices $\mathbf{S}_x$, etc., are the standard matrices representing elementary rotations in three-dimensional space. The tilt γ in this equation can be obtained for a given persistence length a from the Porod-Kratky relationship (Cantor and Schimmel, 1980)

$$\cos(\gamma) = 1 - \frac{l_b}{a}$$

in which l_b is the bp height (typically 3.4 Å). The twist angle is assigned from the helical repeat, h (bp/turn), by

$$\tau = \frac{2\pi}{h}$$

For DNA a typical value for τ is 0.6 radians/bp.

Transformation matrices for simple chain-protein junctions

Across a given globular region a deviation of the chain can occur. For example, binding of a protein may bend the DNA. The reference vector for such a segment is constructed from the beginning to the end of the path taken by the DNA chain through the bend. The stepwise transformation matrix associated with such a region must be constructed to rotate the reference chain vector, aligned along the reference z axis, to the new direction after performing the overall axial twist. If the chain is continuous throughout the bound region, the following free residue may be conceived of as possessing all the freedom associated with any other free residue with respect to its previous neighbor, so that the overall transformation matrix consists of the rotation just described, followed by a normal flexible bp→bp transformation.

First imagine that the protein binds without introducing a net bend to the chain. Then the stepwise matrix $\mathbf{T}_{\beta\to\alpha}$ can be taken as the overall axial twist of the DNA chain within the globular region multiplied by the bp→bp transformation at the far end, giving $\mathbf{T}_{\beta\to\alpha} = \mathbf{S}_z\mathbf{T}_{\alpha\to\alpha}$. Now if the bound protein introduces a bend about, say, the x axis (in the coordinate system of the globular segment), the rotation will be written as $\mathbf{S}_x$, so that the overall stepwise matrix is $\mathbf{T}_{\beta\to\alpha} = \mathbf{S}_x\mathbf{S}_z\mathbf{T}_{\alpha\to\alpha}$.

In the case of a nucleosome, the transition includes intermediate rotations about the y axis to generate the superhelix. Furthermore, the x axis rotation angle depends on the number of superhelical turns taken around the protein. But the transformation matrix is formulated in an analogous manner.

REFERENCES

Ausio, J., F. Dong, and K. E. van Holde. 1989. Use of selectively trypsinized nucleosome core particles to analyze the role of histone "tails" in the stabilization of the nucleosome. *J. Mol. Biol.* 206:451–463.

Bloomfield, V., W. O. Dalton, and K. E. van Holde. 1967. Frictional coefficients of multi-subunit structures. 1. Theory. *Biopolymers.* 5:135–148.

Bustamante, C., D. A. Erie, and D. Keller. 1994. Biochemical and structural applications of scanning force microscopy. *Curr. Opin. Struct. Biol.* 4:750–760.

Cantor, C. R., and P. R. Schimmel. 1980. Biophysical Chemistry, Part III: The Behavior of Biological Macromolecules. W. H. Freeman, San Francisco.

Dìaz, F. G, A. Iniesta, and J. Garcìa de la Torre. 1990. Hydrodynamic study of flexibility in immunoglobulin IgG1 using Brownian dynamics and Monte Carlo simulations of a simple model. *Biopolymers.* 30:547–554.

Flory, P. J. 1953. Principles of Polymer Chemistry. Cornell University Press, Ithaca, New York.

Garcìa de la Torre, J., and V. A. Bloomfield. 1977. Hydrodynamic properties of macromolecular complexes. 1. Translation. *Biopolymers.* 16: 1747–1763.

Garcìa de la Torre, J., S. Navarro, and M. C. Lopez Martinez. 1994. Hydrodynamic properties of a double-helical model for DNA. *Biophys. J.* 66:1573–1579.

Hagerman, P. J. 1988. Flexibility of DNA. *Annu. Rev. Biophys. Biophys. Chem.* 17:265–286.

Hansen, J. C., and J. Ausio. 1992. Chromatin dynamics and the modulation of genetic activity. *Trends Biochem. Sci.* 17:187–191.

Hayes, J. J., D. J. Clark, and A. P. Wolffe. 1991. Histone contributions to the structure of DNA in the nucleosome. *Proc. Natl. Acad. Sci. USA.* 88:6829–6833.

Kirkwood, J. G. 1953. The general theory of irreversible processes in solutions of macromolecules. *J. Polymer Sci.* 12:1–14.

Kirkwood, J. G., and J. Riseman. 1948. The intrinsic viscosities and diffusion constants of flexible macromolecules in solution. *J. Chem. Phys.* 16:565–573.

Kovacic, R. T., and K. E. van Holde. 1977. Sedimentation of homogeneous double-strand DNA molecules. *Biochemistry.* 16:1490–1498.

Leuba, S. H., G. Yang, C. H. Robert, K. van Holde, J. Zlatanova, and C. Bustamante. *Proc. Natl. Acad. Sci. USA.* 91:11621–11625.

Robert, C. H., and K. E. van Holde. 1992. Prediction of sedimentation distributions for positioned nucleosomes on DNA templates at arbitrary saturation. *Biophys. J.* 61:66*a*. (Abstr.)

Schmitz, K. S., and B. R. Shaw. 1977. Chromatin conformation: a systematic analysis of helical parameters from hydrodynamic data. *Biopolymers.* 16:2619–2633.

Shaw, B. R., and K. S. Schmitz, K. S. 1979. Conformation of polynucleosomes in low ionic strength solution. *In* Chromatin Structure and Function: Levels of Organization and Cell Function (NATO Advanced Study Institutes Series A: Life Sciences, Vol. 21b). C. A. Nicolini, editor. Plenum Press, New York. 427–439.

van Holde, K. E. 1988. Chromatin. Springer-Verlag, New York.

van Holde, K. E., and W. O. Weischet. 1978 Boundary analysis of sedimentation velocity experiments with monodisperse and paucidisperse solutes. *Biopolymers.* 17:1387–1403.

Wyman, J., and S. J. Gill. 1990 *In* Binding and Linkage: Functional Chemistry of Biological Macromolecules. University Science Books, Mill Valley, CA. 72–73.

Yao, J., P. T. Lowary, and J. Widom. 1991. Linker DNA bending induced by the core histones of chromatin. *Biochemistry.* 30:8408–8414.

Zimm, B. H. 1980. Chain molecule hydrodynamics by the Monte-Carlo method and the validity of the Kirkwood-Riseman approximation. *Macromolecules.* 13:592–602.