Theory of Gel Electrophoresis of DNA

OSCAR J. LUMPKIN,* PHILIPPE DÉJARDIN,† and BRUNO H. ZIMM, Departments of Physics and Chemistry, B-019, University of California (San Diego), La Jolla, California 92093

Synopsis

A theory of the electrophoresis of DNA through gels with large interfiber spacing, such as dilute agarose, is presented. We assume that the DNA molecule moves along its axis through a “tube” in a neutral gel under the influence of the electric field. The tube is random except for possible bias due to the effects of the field. When the field is small, we easily recover the inverse-length dependence of the mobility found previously by de Gennes and by Doi and Edwards. At higher fields, a new effect appears; the tube becomes oriented because the field biases the direction of the leading end of the chain as it moves to form an extension of the tube. This leads to an increase of the mobility with increasing field by adding a field-dependent but length-independent term to the mobility expression. In agreement with experiment, we find that the field effect can be important at fields as low as 1 V/cm and that the effect can seriously decrease the sensitivity of the mobility to chain length. We also examine the fluctuation of the migration distance, the degree of orientation induced by the field, and the transient effects occurring when the field direction is rotated by a right angle.

INTRODUCTION

In an earlier note on the mobility of DNA in gel electrophoresis,1 two of us proposed a simple model and used it to obtain an expression for the electrophoresis mobility. Somewhat previously, Lerman and Frisch had obtained much the same result.2 In this sequel, we indicate how the dependence of the mobility on electric field, which is a serious complication in experiments with long chains, follows naturally from the model. Our overall aim is to understand accurately the behavior of this simple model so that the sources of deviation of real systems from this behavior can be better identified.

We assume that the polyion moves through a “tube” in the gel under the influence of the electric field. The tube is determined by constraints that the fixed fibers of the gel put on the sideways motion of the polyion; the tube is therefore equivalent to a section of the Doi-Edwards “primitive path” among the gel fibers, the primitive path being the path that would be taken by the polyion if it were pulled taut by a

* To whom correspondence should be addressed.
† Present address: Centre de Recherches sur les Macromolécules, 6, rue Boussingault, 67083 Strasbourg Cedex, France.

c© 1985 John Wiley & Sons, Inc. CCC 0006-3525/85/081573-21$04.00
stretching force on its ends. By simply taking the component of the electrostatic force on the axis of the tube, assuming that the only motion of the polyion is along the axis, and then taking the component of the motion on the field direction, we obtain the following expression for the mobility:

\[
\frac{\langle \dot{x}_{\text{cm}} \rangle}{E} = \frac{Q}{\zeta} \langle h_z^2/L^2 \rangle
\]  

(1)

where \( \langle \dot{x}_{\text{cm}} \rangle \) is the mean velocity of the polyion's center of mass in the field direction, \( E \) is the constant applied electric field, \( Q \) the total charge of the polyion, \( L \) the contour length of the tube, and \( \zeta \) the friction coefficient for translational motion of the whole chain along the tube. The symbols \( \langle \rangle \) denote averaging over the conformations of the tube.

The field dependence of \( \langle h_z^2/L^2 \rangle \), which involves the mean-square component of the polymer's end-to-end vector parallel to the field, is the subject of this article. In this work, we neglect any motion of the polyion that is not along the tube, since such motion is either limited in extent or leads to a state of very low entropy. In the simplest approximation, and in accord with the idea of the primitive path, we represent the tube by a series of straight segments, each with length equal to the distance between two neighboring gel fibers. The polyion moves along this tube by Brownian motion biased by the electric field. The configuration of the tube changes only when it is renewed by motion of the polyion along its axis, when the "head," or leading segment of the polyion, makes electric-field-biased, but otherwise random, steps among the stationary gel chains. After the polyion has moved its own length or more under the influence of the field, every segment of the tube (equals primitive path) has the bias that is acquired when it was laid down by the motion of the head. This last postulate, that the orientation of the head eventually determines the orientation of all the other segments, is of crucial importance in what follows.

To obtain tractable formulas, we first assert that the fluctuations of the tube length, \( L \), are relatively small and uncorrelated with those of \( h_z \), so that we can write

\[
\frac{\langle h_z^2 \rangle}{L^2} \approx \frac{\langle h_z^2 \rangle}{\langle L^2 \rangle}
\]  

(2)

First, we deal with the tube length, \( L \). (In our first article, we did not distinguish between the length of the tube and the length of the chain.) Let \( a_i \) be the length of the \( i \)th segment of the tube, \( L = \sum a_i \). Then,

\[
\langle L^2 \rangle = \left( \sum a_i^2 \right) + \left( \sum \sum a_i a_j \right)
\]  

(3)
We believe that fluctuations in the total number of segments per tube is rather small, at least with long chains, so that we may replace this number by its mean value, \( N \), to good approximation. Also, we assume that there is no correlation between the lengths of successive tube segments, so \( \langle a_i a_j \rangle = \langle a_i \rangle \langle a_j \rangle = \langle a \rangle^2 \). Then Eq. (3) becomes

\[
\langle L^2 \rangle = N(\langle a^2 \rangle - \langle a \rangle^2) + N^2 \langle a \rangle^2 = N^2 \langle a \rangle^2
\]  

Further, \( N \) is related to the contour length, \( L' \), of the chain, which is fixed by the chemical constitution of the chain, by

\[
N \langle l \rangle = L'
\]

where \( \langle l \rangle \) is the mean length of chain per segment. \( L' \) may be conveniently expressed in terms of the persistence length, \( p \), of the worm-like-chain model,\(^5\,^6\) \( L' = N' p \). These give

\[
N = N' p / \langle l \rangle
\]

as the relation between the number of tube segments, \( N \), and the number of persistence lengths, \( N' \), in the chain.

We now calculate \( \langle h_x^2 \rangle \) using this model. Let \( a_i \) again be the length of the \( i \)th segment of a particular tube and \( x_i \) be its component on the field direction. Then, we have

\[
h_x = \sum_i x_i \quad \text{(7a)}
\]

\[
h_x^2 = \sum_i x_i^2 + \sum_{i\neq j} x_i x_j \quad \text{(7b)}
\]

As a simplest approximation, we assume that the chain is flexible enough so that chain stiffness induces no correlation in the orientations of successive segments. Then, we have

\[
\langle h_x^2 \rangle = N \langle x^2 \rangle + N(N - 1) \langle x \rangle^2
\]

Equations (2), (4), and (6) inserted into (1) give

\[
\frac{\langle x_{cm} \rangle}{E} = \left[ \frac{Q}{\xi \langle a \rangle^2} \right] \frac{\langle l \rangle}{p N^2 (\langle x^2 \rangle - \langle x \rangle^2) + \langle x \rangle^2} \quad \text{(9)}
\]

The mobility consists of two terms. The first term produces the molecular-size dependence of mobility that makes gel electrophoresis so useful for separating polyions. We now need to assess the dependence of these two terms on electric field.

We now assume that \( x \) is governed by a Boltzmann distribution over the electrostatic energy of the leading segment in the field. The leading
segment is assumed to be able to bend freely at a hinge located at the point where the leading segment joins the next; by the primitive-path picture, the polion is constrained from translating sideways by a gel chain at this point. The electrostatic energy is \( q x E / 2 \), where \( q \) is the effective charge on the segment and the factor 2 arises because the center of the charge is at a distance, \( a / 2 \), from the hinge. With these assumptions, we can calculate \( \langle x \rangle \):

\[
\langle x \rangle = \langle a \mu \rangle
\]

\[
\mu = \frac{\int \cos \theta \exp(E' \cos \theta) \sin \theta \, d\theta}{\int \exp(E' \cos \theta) \sin \theta \, d\theta} = \Lambda(E')
\]

\[
\Lambda(E') = \coth(E') - 1 / E'
\]

where \( E' = aq E / 2 kT \) is a dimensionless reduced electric field, \( \theta \) is the angle between the segment and the field, \( \mu \) is the Boltzmann average of the cosine of this, and \( \Lambda(E') \) is the Langevin function. The Langevin function approaches \( E' / 3 \) at small \( E' \) and asymptotically approaches unity at large \( E' \). We also need \( \langle x^2 \rangle \), which, in the same way, is found to be

\[
\langle x^2 \rangle = \langle a^2 \nu \rangle
\]

\[
\nu = \frac{\int \cos^2 \theta \exp(E' \cos \theta) \sin \theta \, d\theta}{\int \exp(E' \cos \theta) \sin \theta \, d\theta} = \frac{d\mu}{dE'} + \mu^2
\]

with

\[
\frac{d\mu}{dE'} = \Lambda'(E') = \frac{1}{E'^2} - \frac{1}{\sinh^2(E')}
\]

\( \Lambda'(E') \) is \( 1 / 3 - E'^2 / 15 \) at small \( E' \) and declines asymptotically to 0 at large \( E' \). The mobility now is

\[
\frac{\langle x_{cm} \rangle}{E} = \left( \frac{Q}{\xi(a)^2} \right) \left[ \langle I \rangle \langle a^2 \Lambda(E') \rangle + \frac{\langle a^2 \Lambda^2(E') \rangle - \langle a \Lambda(E') \rangle^2}{pN} + \langle a \Lambda(E') \rangle^2 \right]
\]

\[
(13)
\]

**SEGMENTS OF CONSTANT SIZE**

If \( a \) is the same for all segments, we can disperse with the average signs. To simplify further, we can use the low-field limit of the Langevin function. The result is a simple formula
The mobility still consists of two terms. The first term, which varies as $1/N$, is independent of the field at low fields; the other term does not depend on $N$ but goes to zero at low fields. The reverse is true at high fields; the first term, as can be seen from Eq. (13), goes to zero, and the second, which is independent of $N$, becomes large. This behavior simulates the experimentally well-known decrease of molecular-size sensitivity at high fields.\(^7,8\)

**DISTRIBUTION OF SEGMENT SIZES**

Formula (14), while appealingly simple, is seriously limited in accuracy by the assumption of constant segment size. The space between fibers in actual gels, and hence the segment size of the moving molecule, is distributed over a wide range.\(^9\) (Waki et al.\(^10\) present pictorial evidence of this in electron micrographs of agarose gels.) Some idea of the nature of this distribution comes from the following preliminary considerations. We assume that the gel fibers are infinitely long and of vanishing thickness, and are randomly distributed in space. Hence, a contact point between the moving chains and a gel fiber can be thought of as a random point in the space. We can model a segment of the moving chain by starting at such a contact point and searching for the next contact point, which we take to be the nearest point on the nearest of the other fibers in any direction. (This will presumably give an underestimate of the segment size, since the next contact point of the primitive path will not always be the nearest point on the nearest fiber.)

The gel is characterized by having an average length, $\lambda$, per unit volume. If we start at a random point and move a distance, $a$, in a straight line and investigate all possible directions, we are effectively sweeping out a sphere of radius $a$. The probability that we will meet a fiber when we expand this sphere by an amount, $da$, is the same as the probability that the starting point lies within a cylindrical shell of radius, $a$, and thickness, $da$, around some fiber; this probability is the ratio of the volume of all such cylinders to the total volume, which ratio is $2\pi\lambda da$. But the probability that we have not encountered any fiber in previous increments, $da'$, is the product over previous steps of $1 - 2\pi\lambda da' = \exp(-2\pi\lambda da')$, where it is valid to use the exponential, since $da'$ is infinitesimal. This product over previous steps, $da'$, is

$$\exp(-2\pi\lambda \int_0^a da') = \exp(-\pi\lambda a^2)$$ (15)
The probability, \( f(a) \, da \), of a first encounter lying between \( a \) and \( a + da \) is the product of these two probabilities, which is

\[
f(a) \, da = 2\pi \lambda a \exp(-\pi \lambda a^2) \, da \tag{16}
\]

Coincidentally, this has the same functional form as Ogston's distribution of "pore sizes" in a gel\textsuperscript{11}

\[
f(r) = 2\pi \lambda r \exp(-\pi \lambda r^2) \tag{17}
\]

the size, \( r \), of a certain pore being defined as the radius of the largest sphere that can be placed on a random point without intersecting the surrounding fibers. However, this included-sphere radius is not the same thing as the interfiber spacing, so while the functional form is the same, the physical meaning of the two formulas is different.

Thus, we find that the mean and mean-square segment sizes are

\[
\langle a \rangle = \int af(a) \, da = \frac{1}{2\lambda^4} \tag{18}
\]

\[
\langle a^2 \rangle = \int a^2 f(a) \, da = \frac{1}{\pi \lambda} \tag{19}
\]

To get the contour length of chain segments in a space of size \( a \), we assume that the chain obeys the Kratky-Porod wormlike-chain formula,\textsuperscript{5,6} with persistence-length, \( p \), chain-contour length, \( l \), and mean-square end-to-end length, \( \langle h^2 \rangle \):

\[
2p^2[l/p - 1 + \exp(-l/p)] = \langle h^2 \rangle \approx a^2 \tag{20}
\]

Strictly speaking, we should have an equation for the mean contour length of chain that first passes the distance, \( a \), but, as far as we know, this first-passage problem has never been solved for the wormlike chain. However, the problem has been solved for a random-flights chain; Seshadri and Lindenburg have shown\textsuperscript{12} that the mean contour length (in their terms, the time that the chain has grown) at which the end-to-end distance first passes a given value, \( a \), is rigorously equal to the contour length at which the mean-square end-to-end length is equal to \( a^2 \), and the same thing is true, of course, for a rigid straight chain. Therefore, since the wormlike chain is intermediate between these two extreme cases, we hypothesize that replacing \( \langle h^2 \rangle \) by \( a^2 \) is a reasonable approximation in Eq. (20).

It is impossible to solve Eq. (20) for \( l \) in closed form, but we can write simple approximate solutions:

\[
l = a, \quad a < p \tag{21a}
\]

\[
l = p + a^2/(2p), \quad a > p \tag{21b}
\]
The mean length of chain in the segment, averaged over the segment size distribution with these formulas, is

$$\langle l \rangle = \int l f(a) da = \langle a \rangle \left[ \text{erf} \left( y + \frac{1}{\pi y} \right) \exp(-y^2) \right]$$  \hspace{1cm} (22a)

$$y = p(\pi \lambda)^{1/4} = p/\langle a^2 \rangle^{1/2}$$ \hspace{1cm} (22b)

**EFFECT OF THE INTERFIBER-SPACING DISTRIBUTION: LOW FIELDS**

Returning to the main problem, we introduce the linear charge density, $\rho$, of the moving chain; $\rho = Q/L'$, where $L'$, Eq. (5), is the total length of the chain. Then, if the length of chain in a segment is $l$, its charge, $q$, is $lp$. This gives us the expression

$$E' = a l p E/2kT$$ \hspace{1cm} (23)

which we can use in Eqs. (10)–(13). If we restrict ourselves to small values of $E'$, we get

$$\langle x \rangle = \langle a^2 l \rangle E^*/3\langle a^2 \rangle$$ \hspace{1cm} (24)

$$\langle x^2 \rangle = \langle a^2 \rangle/3 + 2 \langle a^4 l^2 \rangle E^*/45\langle a^2 \rangle^2$$ \hspace{1cm} (25)

where $E^* = \rho \langle a^2 \rangle E/2kT$ is an average dimensionless reduced electric field.

Equation (13) now becomes

$$\frac{\langle x \rangle_{cm}}{E} = \left( \frac{Q}{3\langle a \rangle^2} \right) \left( \frac{\langle a^2 \rangle \langle l \rangle}{N'p} + \frac{\langle a^2 l^2 \rangle E^*}{3\langle a^2 \rangle^2} \right)$$ \hspace{1cm} (26)

where we have kept only the leading term in each part of the expression. Using Eq. (21), we find the following useful approximation:

$$\langle a^2 l \rangle = \langle a^2 \rangle^{3/2} \left[ \frac{3\pi^{1/2}}{4} \text{erf}(y) + \left( \frac{1}{y} + \frac{y}{2} + \frac{y^3}{2} \right) \exp(-y^2) \right]$$ \hspace{1cm} (27)

where $y$ is given by Eq. (22b).

In this paper, we are primarily interested in the sensitivity of the electrophoretic mobility to the chain length and in the effect of field on this sensitivity. It seems reasonable to assume that the friction coefficient for axial motion, $\zeta$, is proportional to the chain length, and, of course, the same thing should be true for the total charge, $Q$. Thus, the ratio $Q/\zeta$ becomes independent of chain length, though of unknown value. The remaining part of, for example, Eq. (26) is a linear function
of $1/N'$. If we take the ratio of the intercept at $1/N' = 0$ to the slope, we have a measure of the relative importance of the $N'$-independent to the $N'$-dependent parts of the mobility. Calling this ratio $J$, we can rewrite Eqs. (14) and (26), respectively,

$$\frac{\langle \dot{x}_{cm} \rangle}{E} = \left( \frac{Q}{3p_1} \right) \left( \frac{1}{N'} + J \right) \tag{28a}$$

$$\frac{\langle \dot{x}_{cm} \rangle}{E} = \left( \frac{Q\langle l \rangle}{3\langle a^2 \rangle p_1} \right) \left( \frac{1}{N'} + J \right) \tag{28b}$$

where for segments of constant size, Eq. (28a),

$$J = \frac{pE'^2}{3\langle l \rangle} \tag{29}$$

while for Eq. (28b), the case of distributed segment size, we have

$$J = \frac{p\langle a^2 \rangle E'^2}{3\langle l \rangle \langle a^2 \rangle^3} \tag{30}$$

Although both formulas are qualitatively similar, there is a large numerical difference. For example, if we take the case where $\langle a^2 \rangle = p^2$ and $E^* = 0.1$, Eq. (30) with the preceding formulas gives $J = 0.01022$. If we now consider a gel of constant spacing $a = \langle a^2 \rangle^h$, so that $E' = E^*$, then $J = 0.00299$. The difference between these values of $J$ emphasizes the dependence of $J$ on the nature of the distribution of interfiber spaces in the gel.

The parameter $J$ determines, in a simple way, the sensitivity of the mobility to chain length. As a measure of this sensitivity, we can define a parameter, $S$, as minus the derivative of the logarithm of the mobility with respect to the logarithm of the chain length:

$$S = \frac{d \ln(\langle \dot{x} \rangle/E)}{d \ln N'} = \frac{1}{1 + JN'} \tag{31}$$

Thus, the sensitivity measured in this way is unity at small chain lengths but decreases as $JN'$ gets large, as happens for long chains at high electric fields.

**SATURATION AT HIGHER FIELDS**

At low fields, the effect of the field is to bias the orientation of the leading segment toward the field direction, but at high fields, the segment should approach complete parallelism with the field, and no further field effect would then be possible; in other words, the field effect should saturate. Putting the limiting values of $\Lambda$ and $\Lambda'$ into
Eq. (13), and using Eqs. (18)–(19), we find that the limiting value of $J$ at high fields is $\pi p / (4 - \pi \langle l \rangle) = 3.66 p / \langle l \rangle$. To treat fields of intermediate strength with our model, it is necessary to evaluate the averages in Eq. (13) by numerical integration, using the distribution function, Eq. (16), and using numerical solution of Eq. (20) for $l$.

Figure 1 shows curves obtained in this way for three chain lengths differing by factors of 10, for the case where the rms interfiber spacing of the gel is equal to the persistence length of the migrating chain; other cases should be similar. It is most noteworthy how the curves, which are separated by one decade from each other at low fields, approach each other at high fields, so that the sensitivity to chain length is all but lost at high fields, especially at the greater chain lengths.

The approximate low-field formulas, Eqs. (22)–(27), agree with these numerical results within 10% up to $E^* = 0.1$.

Figure 2 shows the results of numerical integration plotted in the usual way as chain length on a logarithmic scale versus mobility. The appearance of these curves is very similar to the data published by McDonell et al.7

**COMPARISON TO EXPERIMENT**

The observation that the field effect would take the form of Eqs. (13) or (28) was apparently first made by Southern,13 who found that with DNA in agarose at a low field, the reciprocal of the mobility was

![Fig. 1. Relative mobility, $z\xi/QE$, from Eq. (13) as a function of the dimensionless reduced electric field, $E^*$, for three chain lengths. The rms interfiber spacing, $a_{rms}$, and the chain persistence length, $p$, are assumed to be 67 nm.](image)
Fig. 2. Calculated relation between the chain length, $N'$, measured in persistence lengths, on a log scale and the relative electrophoretic mobility, $Q_k/Q_E$, on a linear scale at different strengths of the dimensionless reduced electric field, $E^*$. Conditions same as Fig. 1.

proportional to chain length; but at a higher field, it was necessary to subtract a constant from the mobility to get proportionality. This result is obviously consistent with Eqs. (13) or (28).

Extensive field-effect data on a series of DNA restriction fragments in agarose gels have recently been reported by Hervet and Bean (Refs. 14, 15, and personal communication) and N. C. Stellwagen, Serwer and Allen have reported data on the electrophoresis of spherical viruses of known size through a series of agarose gels that allow us to make some estimate of the interfiber spacing of agarose. If we follow Rodbard and Chrambach’s suggestion that the electrophoretic mobility of spheres through the gel is proportional to the volume fraction of those pores large enough to contain the spheres, then Ogston’s formula, Eq. (17), applies, and we find that

$$\lambda = 1/\pi R^2$$

where $R$ is the radius of the sphere whose mobility in the gel is $1/e$ times its mobility free in solution. Thus, by Eq. (19), $R^2$ of such a sphere is equal to $\langle a^2 \rangle$. (Note that $a_{rms}$ equals the radius of the sphere, not the diameter, as one might at first suppose; $a_{rms}$ is the average span across the pore, while $R$ is the radius of the largest sphere that fits within it.)

To take a specific example, Serwer and Allen’s data suggest that 1% agarose has a value of $\langle a^2 \rangle^{1/2}$ equal to about 40 nm. If $\lambda$ is
proportional to the concentration of agarose, as one would expect, then the rms interfiber spacing of a 0.6% gel would be 67 nm, which is approximately the persistence length, \( p \), of DNA in free solution at the electrophoresis salt concentration. In this range, \( J \) is almost independent of \( p \) at constant \( \alpha \), so the exact value does not matter much. From Hervet and Bean,\(^{15}\) we can get the ratio \( J \) at the various electric fields; the lowest that gives \( J \) with reasonable precision is 0.5 V/cm. We convert length units by assuming that one base pair is 0.34 nm, and we take \( \langle l \rangle \) to be equal to \( p = 67 \) nm; the result calculated from the slope and intercept of plots of the data is \( J = 5.0 \times 10^{-3} \). Then, using numerical integration of \( J \) and the formula below Eq. (25) for \( E^* \), we find that \( E^* = 0.0735 \), and \( p \) is 5.7 electron charges per base pair, not far from the formal value of 2. [If we had ignored the distribution of pore sizes and used Eq. (29) instead of (30), the result would have been larger by more than a factor of three.]

However, we know that in free electrophoresis the effective charge on DNA is about half of the formal charge.\(^{17}\) We could get closer to agreement with the free-electrophoresis charge by assuming a distribution of interfiber spacings that was broader than the Gaussian of Eq. (16), a quite reasonable possibility in our present state of limited knowledge.

The experiments seem to show more saturation of the field effect at high fields than we predict. For example, we can take the 0.6% agarose gel discussed above but assume a field of 5 V/cm; \( E^* \) would be 0.735. Using the complete Langevin function and numerical integration, as discussed above, we find that \( J \) is 0.26. On the other hand, calculating from Hervet and Bean’s data as before,\(^{15}\) we find that the experimental \( J \) is 0.13. Thus, the experimental \( J \) appears to be rising less rapidly with field than the theory predicts; in other words, it is saturating more rapidly. Agreement here would again be improved by assuming a larger interfiber spacing.

Inspection of the data shows that the saturation occurs mainly in the intercept of the plot against \( 1/N \), which is the second term of Eqs. (26) or (28); this arises from \( \langle x \rangle^2 \). The other term, which arises from \( \langle x^2 \rangle - \langle x \rangle^2 \), is less dependent on the field. This is what would be expected, since \( \langle x \rangle \) is entirely field-dependent and vanishes at zero field, while \( \langle x^2 \rangle \) has a large nonvanishing value even at zero field.

Figure 3 shows some of Stellwagen’s data (personal communication) plotted against reciprocal chain length, and Eq. (13), evaluated by numerical integration, is plotted on the same scale. The factor \( Q/\xi \) was assigned an arbitrary value to fit the point at the lowest experimental field (0.64 V/cm). \( J \) at 0.64 V/cm is 0.0236 and \( E^* \) is 0.0502, giving a value of 3.0 electrons per base pair for \( p \). Although this value of \( J \) is not accurately determined because of the long extrapolation to the intercept, the result is gratifyingly near that found from Hervet and Bean’s data.\(^{18}\)
Fig. 3. Equation (13) and data by Stellwagen for DNA fragments in 0.6% agarose (Ref. 15) plotted against reciprocal chain length. Mobility in units of cm²/V h. $Q/\zeta$ has been set to 18.3 cm²/V h to fit the data at 0.64 V/cm, and $a_{rm}$ and $p$ have been assumed to be 67 nm. Numbers on curves are $E^*$. Experimental fields in V/cm are: $\square$, 1.3; $\bigcirc$, 2.6; $\bullet$, 3.8.

The other lines in Fig. 3 are plots of Eq. (13) at higher fields, using the same value of $p$. While the fit to the data at 1.3 V/cm is reasonable, the fits at higher voltages are increasingly poor, and it is again obvious that the saturation of the field effect seen experimentally is much more pronounced than predicted by our theory.

The data for the smaller chains curve away from the limiting straight line. Qualitatively, this curvature is to be expected for two reasons: In the first place, when the contour length approaches the persistence length from above, correlations between the segments, which we have ignored in this work, become relatively more important, and $\langle h_2^2 \rangle$ becomes proportional to a power of $N$ higher than the first; as a result, by Eq. (1) the mobility becomes less dependent on $N$. In the second place, the mobility of long-chain polyelectrolytes in free solution without gel is well known to be independent of chain length. As the chain length becomes comparable to the gel interfiber spacing, the mobility must change from that characteristic of reptation in a tube to that characteristic of free electrophoresis; therefore, the curve must level off, in accordance with the trend seen in Fig. 3.

**FLUCTUATIONS IN THE MIGRATION RATE**

Since $\langle h_2^2 \rangle$ is a fluctuating quantity, the electrophoretic velocity must also fluctuate. We can make a rough estimate of the size of these
fluctuations as follows. Let $d$ be the distance migrated in a typical experiment at fields low enough so that no field effects are noticeable. Under these conditions, we can think of this distance as being the sum of separate, uncorrelated steps, each representing the traverse by the molecule of one tube length, whose component on the field direction is $h_x$. We then have

$$d = \sum_{k} |h_{x,k}|$$

(33a)

$$d^2 = \sum_{k,l} |h_{x,k}| |h_{x,l}|$$

(33b)

If we average these over many molecules, assuming that all go $P$ steps, we have

$$\langle d \rangle = P \langle h_x \rangle$$

(34a)

$$\langle d^2 \rangle = P \langle |h_x|^2 \rangle + P(P - 1) \langle h_x \rangle^2$$

(34b)

remembering that the individual steps are uncorrelated. The dispersion of $d$ is

$$\langle (d - \langle d \rangle)^2 \rangle = \langle d^2 \rangle - \langle d \rangle^2 = P \langle h_x^2 \rangle - \langle h_x \rangle^2$$

(35)

where we have replaced $\langle h_x^2 \rangle$ by the equivalent, $\langle h_x^2 \rangle$. For a Gaussianly distributed random variable, we have

$$\langle h_x^2 \rangle = (\pi/2) \langle |h_x|^2 \rangle$$

(36)

giving

$$\langle (d - \langle d \rangle)^2 \rangle = P(\pi/2 - 1) \langle |h_x|^2 \rangle = (\pi/2 - 1) \langle d \rangle^2 / P$$

(37)

Thus, since $\langle d \rangle$ is proportional to $P$, the standard deviation of the distance migrated, which is the square root of the dispersion, is proportionate to $P^2$, while the distance itself is proportional to $P$. To take a reasonable example, a DNA of 30,000 base pairs has a value of $\langle h_x \rangle$ of about $1 \mu$m, while $d$ in a typical experiment might be about 100 mm. Thus, $P = 10^5$, and the standard deviation of $d$, from the square root of the left-hand side of Eq. (37), is 0.24 mm. Such a small standard deviation in the distance migrated would not normally be noticeable in the presence of other complicating factors, such as gel heterogeneity.

**AVERAGE ORIENTATION**

The average orientation of the migrating chain is one of the potentially observable properties, for example, by linear dichroism (LD). In
this section, we make a preliminary estimate of this orientation, as measured by the anisotropy of the average optical polarizability, 

\[ A = \frac{(\alpha_{pl} - \alpha_{pr})}{(\alpha_{pl} + 2\alpha_{pr})} \]

where \( \alpha_{pl} \) and \( \alpha_{pr} \) are the polarizabilities for light polarized parallel and perpendicular to the field, respectively, and \( \alpha_{pl} + 2\alpha_{pr} \) is the trace of the tensor.

The average polarizabilities, \( \alpha_{pl} \) and \( \alpha_{pr} \), of the whole chain are the sums of the average polarizabilities of the separate segments. Assume that a segment has a diagonal polarizability tensor with principal values \( \beta_1 \) and \( \beta_2 \), where the first direction is parallel and the second is perpendicular to end-to-end vector of the segment, and that this vector makes an angle, \( \theta \), with the field direction. Also assume that the segment distribution is cylindrically symmetrical around the field direction. Then, the usual rules for transforming second-rank tensors give the following for the cylindrically averaged contribution of the segment to \( \alpha_{pl} - \alpha_{pr} \):

\[ (\beta_1 - \beta_2)(3\cos^2\theta - 1)/2 \]

The average of \( \cos^2\theta \) is \( \Lambda^2 + \Lambda' \), from Eqs. (11b)-(12). To get \( \beta_1 - \beta_2 \), we use a result from Kuhn and Grün,\(^{18}\) who calculated the polarizability of a chain of contour length, \( l \), with fixed rms end-to-end length, \( a \), but without any external orienting field. Their model, a chain of freely hinged rigid rods, has properties similar to the wormlike chain. Neglecting the effect of the external field on the internal conformation of the segment is, obviously, an approximation in our case, but one that is probably not serious for short segments, since we do take account of the effect of the field on the orientation of the segment end-to-end vector. Kuhn and Grün give a formula in terms of the inverse Langevin function, \( y = \Lambda^{-1}(x) \), such that \( x = \Lambda(y) \). Their formula is

\[ (\beta_1 - \beta_2)/(\beta_1 + 2\beta_2) = \Gamma F(a/l) \]

where

\[ \Gamma = \frac{\gamma_1 - \gamma_2}{\gamma_1 + 2\gamma_2} \quad \text{and} \quad F(a/l) = 1 - \frac{3a/l}{\Lambda^{-1}(a/l)} \]

\( \Gamma \) is the intrinsic anisotropy of a short piece of chain and \( \gamma_1 \) and \( \gamma_2 \) are the principal polarizabilities of the short piece. We have used the fact that the trace, \( \gamma_1 + 2\gamma_2 \), is invariant to rotation, so that \( \beta_1 + 2\beta_2 \) is the sum of \( \gamma_1 + 2\gamma_2 \) over all the subelements of the segment; likewise, \( \alpha_{pl} + 2\alpha_{pr} \) is the sum over all segments of \( \beta_1 + 2\beta_2 \). Two series representations are accurate enough for our purposes. When \( a/l \) is less than about 0.7, we have, from Kuhn and Grün,

\[ F(a/l) = \frac{3(a/l)^2}{5} + \frac{36(a/l)^4}{175} + \frac{108(a/l)^6}{875} + \cdots \]

(40a)
For larger $a/l$, we propose (from iterative inversion of the Langevin function):

$$F(a/l) = 1 - 3\left(\frac{2}{l}(1 - \frac{a}{l} + 2\exp\left(-\frac{2}{1 - a/l}\right) + \cdots\right)$$  \hspace{1cm} (40b)

Thus, we have

$$A = \frac{3\langle\cos^2\theta\rangle - 1}{2}F\left(\frac{a}{l}\right)$$  \hspace{1cm} (41a)

$$= \frac{3\langle\cos^2\theta'\rangle - 1}{2}$$  \hspace{1cm} (41b)

where $\theta'$ is the angle between any given short piece of the chain and the field direction. Equation (41b) arises by analogy with Eq. (38). This allows us to write

$$\langle\sin^2\theta'\rangle = 1 - \langle\cos^2\theta'\rangle = \frac{2}{3}\left(1 - \frac{A}{\Gamma}\right)$$  \hspace{1cm} (42)

The quantity $\langle\sin^2\theta'\rangle$ is a measure of the degree of orientation of the chain induced by the electric field; its value, when there is complete orientation, is zero; and when there is no orientation, $2/3$. From it we can calculate, for example, the LD, using formulas given by Weil and Sturm. Let $\epsilon$ be the absorbance of the DNA-containing gel in the absence of electric field, and $\delta\epsilon_{pl}$ and $\delta\epsilon_{pr}$ be the changes in absorbance induced by the field for light polarized in the parallel and perpendicular directions, respectively; then, from Weil and Sturm we can show that

$$\frac{\delta\epsilon_{pl}}{\epsilon} = -\frac{2\delta\epsilon_{pr}}{\epsilon} = (3\cos^2\chi - 1)\left(1 - \frac{3\langle\sin^2\theta'\rangle}{2}\right) = (3\cos^2\chi - 1)\frac{A}{\Gamma}$$  \hspace{1cm} (43)

Here, $\chi$ is the angle between the transition moment of the chromophore and the local axis of the DNA helix. Weil and Sturm also present formulas for the anisotropy of fluorescence in terms of $\langle\sin^2\theta'\rangle$, but since these also involve the direction of the emission transition moment, which changes with time, we shall not go into them (I. Hurley, to be submitted to this journal).

We can now substitute from Eqs. (41a) and (42) in (43) and average over segment sizes with the distribution function, $f(a)$. Some results from numerical integration are shown in Table I, where we have assumed that $\chi = \pi/2$, appropriate for the nucleotide bases of DNA; we
TABLE I  
Field-Induced Anisotropy and Linear Dichroism, $A/\Gamma = -\Delta \varepsilon_{pl}/\varepsilon$

<table>
<thead>
<tr>
<th>$E^*$</th>
<th>$a_{\mu\nu}/p = 0.5$</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.00220</td>
<td>0.00204</td>
<td>0.00196</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0404</td>
<td>0.0339</td>
<td>0.0255</td>
</tr>
<tr>
<td>1</td>
<td>0.106</td>
<td>0.0848</td>
<td>0.0571</td>
</tr>
<tr>
<td>10</td>
<td>0.495</td>
<td>0.365</td>
<td>0.207</td>
</tr>
<tr>
<td>100</td>
<td>0.716</td>
<td>0.530</td>
<td>0.298</td>
</tr>
</tbody>
</table>

have also assumed that $\langle a^2 \rangle = p^2$. For reference, we recall that the completely oriented chain gives $A/\Gamma = 1$ and $\delta \varepsilon_{pl}/\varepsilon = -1$ in this case, while the unoriented chain gives $A/\Gamma = 0$ and $\delta \varepsilon_{pl}/\varepsilon = 0$. It can be seen that the amount of orientation grows rapidly (initially quadratically) with increasing field and should be readily detectable at fields where a pronounced effect on the mobility is apparent, that is, where $E^*$ is in the neighborhood of unity or greater.

CROSSED FIELDS

Schwartz and Cantor have proposed a method in which the direction of the electric field is repeatedly switched from one direction to another orthogonal to it; they have shown that this procedure gives improved resolution of large DNA molecules. Without attempting a detailed analysis of this procedure, we can make some preliminary observations.

By Eq. (1), the important quantity is $\langle h^2 \rangle$. We can calculate what this becomes when a field in the $y$-direction has been applied for a time long enough to establish steady state; to do this, we need $\langle x \rangle$ and $\langle x^2 \rangle$. In place of Eq. (10b), we now have

$$p = \mu \times \text{proportional to } J \sin \theta \cos \phi + \exp(E' \cos \theta) \sin \theta$$

where the polar axis is now the $y$-axis and $\phi$ is the angle of rotation around it. In place of Eq. (11b), we now have

$$\nu = \frac{\int \sin^2 \theta \cos^2 \phi \exp(E' \cos \theta) \sin \theta \, d\theta \, d\phi}{\int \exp(E' \cos \theta) \sin \theta \, d\theta} = \frac{\Lambda(E')}{E'}$$

Now, with Eqs. (8), (2), and (1), we get

$$\langle x \rangle = \left(\frac{Q}{\xi} \right) \frac{\langle x^2 \rangle}{N \langle a \rangle^2} = \left(\frac{Q}{\xi} \right) \frac{\langle a^2 \Lambda(E')/E' \rangle}{N \langle a \rangle^2}$$

(46)
Thus, we find a strict dependence of mobility on $N^{-1}$, regardless of field strength. Obviously, the size resolution of the technique would be improved if one could operate under such conditions, as indeed Schwartz and Cantor found.\textsuperscript{20} We must remember, though, that our picture holds only during a short time interval after the field has been switched from the y-direction to the x-direction, before the molecule has moved far enough to become reoriented. We now consider the length of this interval.

By our model, the reorientation time is the time needed for the molecule to move axially its own length, so that all segments acquire the biased orientation created by the leading segment. By the theory described in our previous work,\textsuperscript{1} the axial velocity, $\dot{s}$, is given by

$$\dot{s} = EQ|\dot{h}_x|/\zeta L$$

where $x$ is the field direction. We assume that a low fields $h_x$ is Gaussianly distributed, as is usual for randomly coiling chains. Then, the probability distribution of $|h_x|$ is given by

$$P(|h_x|) = 2P(h_x) = 2\left(\frac{3}{2\pi Na^2}\right) \exp\left(-\frac{3h_x^2}{2Na^2}\right)$$

where the factor of 2 enters because two values of $h_x$ correspond to each value of $|h_x|$. Since to each value of $|h_x|$ there uniquely corresponds a value of $\dot{s}$, it follows that the probability distribution of $\dot{s}$ is given by

$$W(\dot{s}) \, d\dot{s} = P(|h_x|) d|h_x|$$

where $d\dot{s}$ and $d|h_x|$ are related by Eq. (47).

Now, the time needed for the molecule to form a new tube is

$$\tau = L/\dot{s}$$

Since there is thus a unique relation between $\tau$ and $\dot{s}$, we can write a probability distribution for $\tau$:

$$w(\tau) \, d\tau = W(\dot{s}) \, d\dot{s}$$

with $d\dot{s} = -Ld\tau/\tau^2$ by (50).

This distribution diverges as $\tau$ approaches zero, that is, in chain conformations such that the axial electrophoretic velocity vanishes. In these conformations, axial diffusion, that is, reptation, becomes the dominant form of motion.\textsuperscript{34} In diffusion, the time for the molecule to move its own length is

$$\tau_d = L^2/2D = L^2\zeta/2k_BT$$
where $D = k_B T/\zeta$ is the axial diffusion coefficient of the center of mass. We have assumed that $\zeta$ for diffusion is the same as $\zeta$ for electrophoresis, which is not exactly correct, but justifiable in view of the preliminary nature of this calculation. We now assume that the effective $\tau$ is the lesser of Eqs. (50) or (52). Then, we can calculate a mean value:

$$
\langle \tau \rangle = \int_0^{\tau_d} \tau w(\tau) \, d\tau + \int_{\tau_d}^\infty \tau_d w(\tau) \, d\tau
$$

(53)

A change of variable converts the first integral into the exponential integral and the second into the error function; when these are expanded in series, the result is

$$
\langle \tau \rangle = \left( \frac{2\langle h_z^2 \rangle}{\pi} \right)^{\frac{1}{2}} \frac{1}{\chi_{\text{cm}}} \left[ -0.2886 + \ln \left( \frac{\langle h_z^2 \rangle^{\frac{1}{2}} QE}{2 k_B T} \right) + \left( \frac{2}{\langle h_z^2 \rangle^{\frac{1}{2}}} \frac{k_B T}{Q E} + \cdots \right) \right]
$$

(54)

Consider a numerical example. In one of Schwartz and Cantor's experiments, T2 DNA was examined at a field of about 10 V/cm, giving a migration rate of about 1 cm/h.\textsuperscript{20} T2 DNA has about 180,000 base pairs, giving an effective electrophoretic charge\textsuperscript{17} (taken as half the formal charge) of $2.9 \times 10^{-14}$ Coulombs. For this DNA, $\langle h_z^2 \rangle^{\frac{1}{2}}$ is about 1.5 µm; we take the temperature of the gel to be about 300 K. Then, we calculate $\langle \tau \rangle$ to be 3.8 s.

If the time between field changes is much longer than $\langle \tau \rangle$, the molecule will be migrating for most of the time in a unidirectional, constant field. On the other hand, if this time is much shorter than $\langle \tau \rangle$, approximately half of the segments of the chain will be oriented along one field direction and half along the other, so that the field effect will be present at half strength in both directions; thus, the mobility will contain considerable field- and length-independent terms, and the separation of different-sized chains will not be much better than in the constant-field case. Therefore, we would expect that size resolution would be best when the time between field changes was about equal to $\langle \tau \rangle$.

Schwartz and Cantor tried switching the field at intervals of 2, 10, and 100 s, and they found the best separation of T2 DNA from the smaller T7 DNA at the 10-s interval.\textsuperscript{20} The agreement with our estimated $\langle \tau \rangle$ is as good as can be expected from this simple calculation.

**DISCUSSION**

The usefulness of gel electrophoresis with biochemical polyelectrolytes comes from its ability to resolve chains of slightly differing lengths, such as nucleic acids or polypeptide–detergent complexes. It
has been disappointing to find that this resolving power declines as larger and larger chains, such as DNA molecules above $10^4$ base pairs, are examined. In fact, in one of the earliest attempts to use the technique on DNA, by Flint and Harrington, almost no size dependence at all was observed from molecular weight $10^5$ to $10^6$ in polyacrylamide gels at 5–25 V/cm. In this paper, we have shown that at least part of the trouble can arise from distortion of the migration chain by the electric field, the distortion causing a molecular-weight-independent term to appear in the mobility, in addition to the low-field term that is inversely proportional to the molecular weight.

The qualitative agreement of our results with experiment is encouraging, even if the actual numerical values are not truly close to the experimental ones. Before too much can be learned from the disagreements, however, it will be necessary to know more about the structure of the gel. For one thing, the various experimenters did not all use the same kind of agarose. Second, we are not yet sure how to estimate pore sizes of the gel from electrophoresis of spheres of known size; the open-space formula proposed by Rodbard and Chrambach ignores hydrodynamic interactions between the moving particle and the gel, interactions that calculations of ours published elsewhere suggest should be important. Third, our assumption that the gel fibers are distributed at random in space is, at present, merely an assumption and is not supported by any other evidence.

The appearance of field-induced distortion at fields of less than 1 V/cm is astonishing when one considers that, for example, the fields needed to induce measurable electrooptical (Kerr) effects in free solution, even with large DNA molecules, are at least of the order of hundreds of volts per centimeter. This result easily follows from the chain-in-tube picture, as the above calculations have shown. In qualitative terms, the basic concept is the following. The chain lies in a random tube and continually undergoes random diffusional movements along its axis. Superimposed on these random movements is a bias produced by the electric field, the result of which is a continual drift toward that end of the tube that lies down field, the "leading segment." The component of this drift on the field direction is the macroscopically observed electrophoresis. When the leading segment moves forward, it forms a new segment of the tube. In this process, at sufficiently low fields, it makes a practically random choice of direction, so $\langle h_z^2 \rangle$ at low fields depends on $N$ to the first power only, as is usual for random walks. In contrast, in the high-field process of tube renewal, the orientation of the leading segment is biased by the field. The amount of bias can be very small, but since it is always in the same direction, its effects on $h_z$ accumulate in proportion to the number of segments, $N$, and hence its effects on $\langle h_z^2 \rangle$ accumulates as $N^2$. The result at large $N$ is a pronounced field effect at fields much lower than those usually considered to be capable of causing significant biasing effects on molecular conformations.
We have not discussed how the chain reverses its direction in the tube, a process that must happen sometimes when $h$, becomes small; in this situation, diffusional random motions must sometimes cause the leading segment to become the trailing segment and vice versa. The overall field-induced drift, however, is always toward the momentary leading segment.

If we accept the picture of DNA in the gel as a coil similar to its state in free solution, and if we accept the picture of the gel as a random network of long fibers, then it is hard to avoid the conclusion that translation of large DNA, at least, must take place by reptation along a tube among the gel fibers. The rms end-to-end distance of the larger molecules studied by Hervet and Bean (Refs. 14, 15, and personal communication) is about 1 $\mu$m in free solution, while the pore size of the agarose gels, from the experiments of Serwer and Allen, is of the order of 50 nm. Thus, the coil must either be compressed by a factor of about 20 to fit into a pore, or it must lie in a tube. Compression is not a completely unreasonable idea, considering that sufficiently concentrated solutions of neutral polymers do cause DNA molecules to collapse; however, the tube model seems simpler, so we have selected it for detailed analysis. It may also be that the chain moves selectively into the larger spaces in the gel, in that way maximizing its entropy, so that while the migrating chain lies in a tube, the tube is a selected tube. This is a refinement that we have not yet attempted to deal with.

One can easily think of other considerations that should be examined. We noted above that our estimate of the appropriate interfiber spacing was probably too low; this estimate could be refined. We have limited ourselves in this work to cases where the correlation in orientation between successive segments was weak, and this effectively limits us to dilute gels with large spacings; the case of tight gels needs a different analysis. The question of the ranges of chain length and interfiber spacing in which the tube picture is appropriate, and in what ranges another model might be better, is still unclear.

Thus, the present work appears to be only the beginning.

This research has been supported by NIH Grant GM-11916 to B.H.Z. P.D. benefited from a fellowship from NATO. We are indebted to C. P. Bean and H. Hervet for a copy of their data. Likewise, we are indebted to Nancy Stellwagen for permission to use her data in Fig. 3. Discussions with F. W. Studier, L. Lerman, H. Frisch, C. P. Bean, and P. Serwer were valuable in developing and clarifying ideas. We thank Ian Hurley for bringing the paper by Weil and Sturmlg to our attention. Helpful correspondence with E. M. Southern and Roger West is gratefully acknowledged.

References
GEL ELECTROPHORESIS OF DNA


Received September 20, 1984
Accepted January 16, 1985