uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

David Sean-Fortin
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.Sc. (Physics)
GRADE / DEGREE

Department of Physics
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

On the Mobility of Partially Denatured DNA in Gel Electrophoresis: A Theoretical Investigation

TITRE DE LA THÈSE / TITLE OF THESIS

Gary Slater
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

J. Harden

B. Joos

T. Xu

Gary W. Slater
Le Doyen de la Faculte des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

On the mobility of partially denatured DNA in gel electrophoresis: a theoretical investigation

David Sean
B.Sc. University of Ottawa 2002

THESIS SUBMITTED TO THE
FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE IN PHYSICS

Your file Votre référence
ISBN: 978-0-494-73823-8
Our file Notre référence
ISBN: 978-0-494-73823-8

SUMMARY

There are technologies which exploit a rapid reduction of the gel electrophoretic mobility of DNA arising from partial denaturation. The underlying phenomenon behind these experiments — the mechanisms which reduce the mobility — are not very well understood. Such is the purpose of my thesis.

The first chapter provides a brief introduction to the field of polymer physics. The subjects covered are carefully chosen to directly relate to the forthcoming research. There is a published semi-empirical formula used to model the rapid decrease of mobility which is largely considered to be consistent with experimental data. The second chapter of this thesis demonstrates that there is a fundamental confusion in the literature regarding the fitting parameter L_r, in the said formula. By going back to the original derivation, a physical interpretation can be given to L_r. This interpretation yields theoretical values which are consistent with what has been published. However, we find that an underlying assumption — that the effect of the denaturation does not depend on its position along the DNA fragment — may systematically overestimate experimental observations of L_r.

To measure the impact of this assumption, a simulation model of DNA is presented. The article presented in the third chapter reveals that indeed, the position of the denatured region affects the migration of the DNA fragment. A refined version of the formula which takes these factors into account is proposed. The simulations also reveal that, for certain fields, an unexpected conformation completely dominates during migration of the fragment. This surprising result: a squid-like conformation, is explored in chapter four.

SOMMAIRE

Certaines technologies exploitent une réduction rapide de la mobilité électrophorétique (sur un gel) de l'ADN suite à une dénaturation partielle. Le phénomène sous-ajacent derrière ces expériences — les mécanismes qui réduisent la mobilité — ne sont pas très bien compris. Voilà l'objet de ma thèse.

Le premier chapitre offre une brève introduction au domaine de la physique des polymères. Les sujets traités sont soigneusement choisis pour leurs liens directs à la recherche qui suit. Il existe une formule semi-empirique qui est utilisée depuis plus 25 années pour modeler la diminution rapide de la mobilité; on la considère en général comme étant cohérente avec les données expérimentales. Le deuxième chapitre de cette thèse démontre qu'il existe une confusion dans la littérature scientifique quant au paramètre L_r dans la dite formule. En retournant à la dérivation d'origine, une interprétation physique peut être attribuée à L_r. Cette interprétation produit des valeurs théoriques qui sont cohérentes avec ce qui a été publié. Cependant, on a découvert qu'une hypothèse sous-jacente – que l'effet de la dénaturation ne dépend pas de sa position le long du fragment d'ADN – peut systématiquement surestimer les observations expérimentales de L_r.

Un modèle de simulation numérique a été élaboré afin de mesurer l'impact de cette hypothèse. L'article présenté dans le troisième chapitre révèle qu'en effet, la position de la région dénaturée a un effet sur la migration du fragment d'ADN. Une version perfectionnée de la formule qui tient compte de ces facteurs est donc proposée. Les simulations révèlent également que pour certaines valeurs du champ éléctrique, une conformation inattendue domine complètement au cours de la migration du fragment. Ce résultat surprenant est étudié dans le chapitre quatre.

STATEMENT OF ORIGINALITY

I herby declare that the following thesis is the product from my original work under the supervision of Dr. Gary W Slater. The core of the simulations were performed using a version of the simulation package ESPResSo which was modified by Hendrick W. de Haan and I. I have written the scripts which were used to program the simulations. The vast majority of the analysis is performed using routines which I coded in C and Python. Unless otherwise noted, I have produced all of the schematics and plots. The second chapter of this thesis has been submitted for publication:

> **David Sean**, Gary W Slater, *Physical interpretation of the L_r parameter in the theory for the gel electrophoresis of partially denatured DNA*, Electrophoresis (Submitted April 24, 2010)
> elps.201000239.

Two manuscripts describing the results presented in the third and fourth chapters respectively are currently being processed and will shortly be submitted. As a partial requirement in my program, I presented my research at the Ottawa Carleton Institute of Physics symposium:

> **David Sean** Gary W Slater "Blocking mechanisms of partially melted DNA during gel electrophoresis", Ottawa Carleton Institute of Physics, May (2009)

During the course of my years as a graduate student, I have had the opportunity to work on a group review article. I was responsible for writing Section 2.4 *Langevin and Brownian dynamics.* With Mykyta V. Chubynsky and I wrote: Section 3.3 *Polymer-obstacle collisions.* Christine Kingsburry and I wrote Section 3.4 *Ogston Modeling sieving in hydrogels.* In the Outlook section, I also wrote a short segment on DNA denaturation. The preprint version of this article is included, with written permission from the editors, in the Appendix of this thesis.

> Gary W. Slater, Christian Holm, Mykyta V. Chubynsky, Hendrick W. de Haan, Antoine Dube, Kai Grass, Owen A. Hickey, Christine Kingsburry, **David Sean**, Tyler N. Shendruk, Lixin Zhan "Modeling the separation of macromolecules: A review of current computer simulation methods" *Electrophoresis* **30**, 792-818 (2009)
> DOI: 10.1002/elps.200800673

I had the immense opportunity to participate at the American Physical Society March meeting in 2009. I presented a poster:

> **David Sean**, Gary W Slater, "How does denatured DNA stop moving in a gel?", *American Physical Society*, March (2009)
> Abstract ID: BAPS.2009.MAR.S1.25
> http://meetings.aps.org/Meeting/MAR09/Event/93001

ACKNOWLEDGEMENTS

It is easy to loose track of the people that contributed to my research. This is especially true when their active role was played at the beginning. For this reason, I would like to start with Jean-François Mercier, Martin Kenward, and Frédérique Tessier, who initiated me to this wonderful field. Your patience and genuine interest in my challenges are greatly appreciated. Your presence in the group is greatly missed.

I have been fortunate to have equally great new members join the group: Henrick de Haan, Tyler Shendruck and Mykyta Chubynsky. Not to forget the permanent resident Owen Hickey and the office D.J. Antoine Dubé. Your scientific curiosity combined with sharp physical intuition truly brings our discussions to bear fruit.

I have had the immense opportunity to have Gary Slater as a supervisor. Thank you for tolerating and offering guidance towards to my endless tangents. Thank you for feeding my curiosity with your endless list of "cute" problems. You have the gift of being able to sustain scientific debate for hours without the slightest hint of hostility. Your curiosity-driven scientific approach is a true motivation. Gary you are a model scientist, orator, teacher, administrator, supervisor — and mentor.

The past years as a student has been some of the greatest of my life. During the span of my studies at the University of Ottawa, I have married the woman whom I regard as the most beautiful person possibly imaginable: Pascale. I have also had three incredibly intelligent and beautiful daughters: Loic, Juliette, and Soriya, who never cease to amaze me. I thank you for supporting me during this last stretch and I promise to be more present from now on.

« Courage les filles, c'est presque terminé! »

Contents

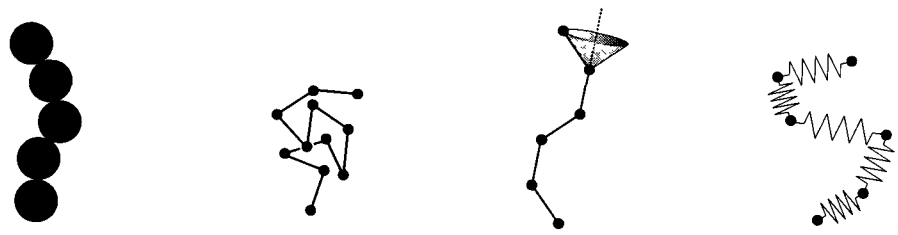
Chapter1

Introduction

1.1 Polymers

A polymer can be pictured as a molecular chain: a chain of repeating units (or "-mers") linked by chemical bonds. Perhaps the simplest physical picture is that of pearl-necklace—a sequence of beads each attached to the former in a linear fashion. However, polymers need not to be arranged in a linear manner. Sometimes there are branching points, which can yield a wide range of possible *configurations*. For example, the linear polymers in Fig. 1.1 have a different configurations than the branched examples of Fig. 1.2. The polymer configuration is not to be confused with the term *conformation*. The polymer conformation is a term used to describe instantaneous positions of the beads relative with one another. It is usually a qualitative description of the overall shape of the polymer. For instance, if a linear polymer is stretched out from end to end one may say it has a straight conformation. If it then bends near the center it may be said to have a "U"-like of "V"-like conformation. The conformation is thought to change with time whereas the configuration is usually static. All rules have exceptions, I will later present a case with DNA where this is not true.

For the moment, we can consider a generic linear polymer where the monomers, or repeating units, are hard-spheres connected to one another at the sphere surface. This is called the *pearl-necklace model*. Using this model, Fig. 1.1(a) depicts a polymer of $N = 5$ repeating units. By design, the distance between neighboring beads, or the *bond-length* b, is



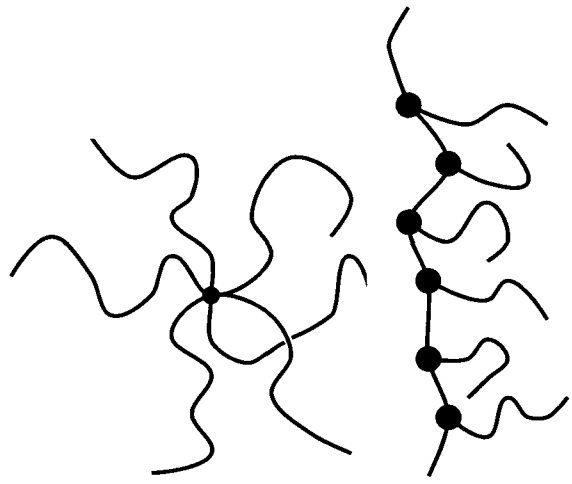
(a) (b) (c) (d)

Figure 1.1 Shown here lying on the same plane, are various theoretical polymer models. The bonds are usually thought to be massless. a) The pearl-necklace model of a polymer consists of modeling the repeat-units as spheres connected at the surface. b) The bead-stick model shown here has the sole constraint of a fixed length between subsequent beads. c) The chain shown has an additional constraint— a fixed bond angle of 45° between subsequent units. The cone illustrates other possible positions for the top unit if it wasn't restricted to the plane. d) The bead-spring model can have fluctuating bond lengths.

equal to the bead diameter. An extension of this model to include bonds longer than the bead diameter, called the *bead stick model*, is shown in Fig. 1.1(b). Similar to the pearl-necklace model, the beads are freely-joined and are only constrained by the distance between their immediate neighbors. We will continue this section using a bead-stick model with point-like beads (basically a chain of sticks). The only relevant parameters are the number of beads N and the bond length b.

Bond-crossings or interactions between beads other than direct neighbors will be ignored for the moment. This theoretical construct is, of course, highly idealized, but even with this ideal model, a single linear polymer of N beads with $N-1$ bonds has nonetheless $3+2^{N-1}$ spatial degrees of freedom. With increasing N, the number of variables needed for an exact description of a single polymer becomes too high to be useful. A computer can easily register the positions of a few hundred beads belonging to a polymer, but the human mind is incapable of making any sense of this raw information. Polymer physicists have devised a handful of statistics to describe a single polymer. Polymer physics being quite a vast field, we will only consider the ones pertinent for the research at hand.

Perhaps the most fundamental spatial measurement of an object is its position and size. The statistic for the position which is most commonly used in my research is the *center of*



(a) (b)

Figure 1.2 Monomers can be arranged in a variety of exotic ways in order to form polymers with different configurations. (a) Star polymers can be seen as multiple linear polymers branching at the same point. (b) Comb polymers can be seen as a main linear polymer with multiple branches attached on its contour. (Here the branching points have been emphasized with black dots to show a clearer picture.)

mass. It can be physically interpreted as the average position, weighted by the mass[1], over all beads. The polymers considered in my research have monomers of identical mass. Thus the center of mass is reduced to the average of the bead positions [1]

$$\vec{R}_{\mathrm{CM}} = \frac{\sum_{i=1}^{N} m_i \vec{R}_i}{N \sum_{i=1}^{N} m_i} = \frac{1}{N} \sum_{i=1}^{N} \vec{R}_i \tag{1.1}$$

where m_i and $\vec{R}_i$ are respectively the mass and position vector of monomer i. I could simply change terminology and call this statistic the *average of the bead positions*, but this term does not invoke the same intuitive picture as does *center of mass position*. Using $\vec{R}_{\mathrm{CM}}$ also eliminates a possible confusion with the concept of a time-average (have I called it the average of the bead positions).

Defining the characteristic size of a polymer is not an easy task. The *contour length* L_{C},i.e., the length along the path going through successive beads is perhaps the most intuitive

[1] No pun intended

choice for a linear polymer. It is obtained by multiplying the bond length with the number of bonds

$$L_{\mathrm{C}} = b(N-1). \tag{1.2}$$

It is a measure proper to the polymer and does not vary with different conformations. But unless the polymer is always in the same conformation, this measure is not very indicative of the size of the molecule. The polymer can change shape: it can bend in half, stretch out end to end, wrap itself up into a coil, *etc.* It is dynamic!

It is usually more practical to consider a statistic over the possible conformations. The easiest to calculate is R_{F}, the *root-mean-squared end to end distance.* This only depends on the position of the last bead relative to the first [1],

$$R_{\mathrm{F}}^2 = \langle(\vec{R}_N - \vec{R}_1)^2\rangle. \tag{1.3}$$

Although this value can change significantly between extreme conformations, they are few in number, so R_{F} is well-behaved. But why only use the two end-points? For a better statistic, one can extend this to all the beads! Let us anticipate a possible confusion and remove the averaging brackets in Eqn. 1.3. We shall first consider the statistic for a particular conformation. We will find its average over all conformations later on.

Instead of the square distance between the first and last bead, let us find the average square distance between all bead combinations, $\langle(R_i - R_j)^2\rangle$. This average comes from a sum of N^2 terms, but we do not need all of them. First, there are N terms $(R_i - R_i)^2 = 0$ which do not change the sum and are of no physical importance. Second, there is a series of superfluous terms $(R_i - R_j)^2 = (R_j - R_i)^2$, so an $i < j$ type of sum could be used. But I digress; this is a mere computational argument, not a physical one. For simpliy, let us use the naive form [1]

$$\langle(R_i - R_j)^2\rangle = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\vec{R}_i - \vec{R}_j)^2. \tag{1.4}$$

Let us remember that it would make more physical sense not to consider the $i = j$ terms. Choosing this would change the N^2 numerator to $N(N-1)$; we will get back to this later. This *extension* of R_{F} has the advantage to be applicable to a variety of configurations whereas R_{F} can only apply for a linear polymer.

We now have a measure of size which can be interpreted as *the average square-distance between all monomers* and a measure of position which is $\vec{R}_{\rm CM}$. Let us consider the distribution of monomers about the center of mass. The root of the average square distance from the center of mass is called the *radius of gyration* $R_{\rm G}$ and can be calculated using [1]

$$R_{\rm G}^2 = \frac{1}{N}\sum_{i=1}^{N}(\vec{R}_i - \vec{R}_{\rm CM})^2. \tag{1.5}$$

It turns out that Eqn. 1.4 and Eqn. 1.5 are mathematically equivalent measures, save a factor of two (in other words, $\langle(R_i - R_j)^2\rangle = 2R_{\rm G}^2$). In polymer physics, $R_{\rm G}$ is largely thought to be a geometrical measure of size. As a rule of thumb, a polymer is thought to have a roughly spherical shape of radius $R_{\rm G}$. It is interesting to note that what we have just defined as a geometric measure of position and size, could be found using another argument. Given the monomer positions $\vec{R}_i$, one could define $\vec{R}_{\rm CM}$ and $R_{\rm G}^2$ respectively as the mean and variance of the monomer distribution in space:

$$\vec{R}_{\rm CM} = mean(\vec{R}_i) = \frac{1}{N}\sum_{i=1}^{N}\vec{R}_i \tag{1.6}$$

$$R_{\rm G}^2 = var(\vec{R}_i) = \frac{1}{N}\sum_{i=1}^{N}(\vec{R}_i - \vec{R}_{\rm CM})^2 \tag{1.7}$$

In other words, $R_{\rm G}^2$ can be seen as the second moment of the monomer distribution. Statisticians may argue that the sum in Eqn. 1.7 should be divided by $(N-1)$. This has little practical importance since the difference is negligible at high N. It is nonetheless a very interesting coincidence that we have previously arrived at the same conclusion via a purely physical argument (see note after Eqn. 1.4).

An interpretation which is less common is related to the *moment of inertia* $MR_{\rm G}^2$, where M is the polymer mass ($M = \sum_{i=1}^{N} m_i$). Probably for the same reasons I prefer to use the phrase *center of mass* for the average monomer positions, the term *radius of gyration* is usualy used outside the context of the mechanical property called moment of inertia. Polymer physicists will usually talk about the radius of gyration as a measure of polymer size— or a measure of the monomer distribution, more than a measure of the mass distribution. In fact, many textbooks will omit the m_i terms which should appear in Eqn. 1.5. Naturally, this omission is not severe since in the models used, the masses are thought to be the same. In most cases, polymers in a solvent are assumed to be in the over-

damped regime so the mass plays a negligible role[2]. We can now define the average radius of gyration $\langle R_G^2 \rangle$ as an average over all conformations. Again, this definition varies in the literature, some authors choose to define the radius of gyration *already* as the average over all conformations. To avoid any confusion we chose R_G^2 to be an instantaneous value, and $\langle R_G^2 \rangle$ as its conformation average. Using the theoretical linear polymer models shown in Fig. 1.1 it is possible to derive the values for $\langle R_G^2 \rangle$ and $\langle R_F^2 \rangle$ by assuming the the conformations can be generated by a random walk. Unlike R_F, the statistical measure R_G, can be applied to polymer configurations other than a linear chain. For simplicity, we will only consider the results for a linear chain.

The important random walk result worth noting is that the average square displacement after $(N-1)$ steps of length b comes up to $b^2(N-1)$. We can apply this to a bead-stick polymer model with N monomers— and thus $(N-1)$ bonds, to obtain [1]

$$\langle R_F^2 \rangle = (N-1)b^2 \approx Nb^2. \tag{1.8}$$

Naturally, we are not limited to the end points. The average squared distance between any two monomers with positions $\vec{R}_i$ and $\vec{R}_j$ is $b^2|i-j|$. We can apply this to Eqn. 1.4 and obtain [1]

$$\langle R_G^2 \rangle = \frac{1}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \langle (\vec{R}_i - \vec{R}_j)^2 \rangle = \frac{b^2}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} |i-j| \approx \frac{1}{6} Nb^2, \tag{1.9}$$

where the last approximation is valid at large N. We can apply these two results to the bead-spring model of Fig. 1.1(d) by replacing b^2 with $\langle b^2 \rangle$ (assuming the bond lengths have a gaussian distribution). Theoreticians have found found how to generalize this for a variety of constraints. For example, if excluded volume interactions are considered, a different exponent on N is obtained in Eqn. 1.9. Let us focus on the effect a stiffness parameter has on the size of a linear polymer.

To consider an inherent stiffness, we can use the *freely rotating segment* model shown in Fig. 1.1(c). In the example shown, there is a fixed angle of $\theta_b = 45°$ between the successive segments. At best, it would take four consecutive segments to complete a 180° rotation. This chain would have the mean squared end-to-end distance $\langle R_F^2 \rangle = \frac{1+\cos\theta_b}{1-\cos\theta_b} Nb^2$ and a maximum end-to-end distance $L_C = \cos(\frac{\theta_b}{2})Nb$ [2].

---

[2] High-frequency regimes represent an exception.

If we were to redefine this chain — that is, define the repeat units as, say 8 consecutive monomers as shown in Fig. 1.3 — we would have a polymer with $N' = \frac{N}{8}$ units which would essentially be following the freely-joined-chain model (since it only takes four to rotate by 180°). This re-grouping / re-scaling concept was first brought forward by Werner Kuhn. The *Kuhn length* L_{K} is the critical (or minimum) length scale required to obtain N_{K} "freely-joined" units called *Kuhn segments* [3]. The physical size of this equivalent freely-rotating chain should not differ, thus the equations we obtained earlier should hold [1, 2]

$$\langle R_{\mathrm{F}}^2 \rangle = N_{\mathrm{K}} L_{\mathrm{K}}^2 = \frac{1 + \cos\theta_b}{1 - \cos\theta_b} N b^2 \tag{1.10}$$

$$\langle R_{\mathrm{G}}^2 \rangle = \frac{1}{6} N_{\mathrm{K}} L_{\mathrm{K}}^2 = \frac{1}{6} \frac{1 + \cos\theta_b}{1 - \cos\theta_b} N b^2 \tag{1.11}$$

$$L_{\mathrm{C}} = \cos\left(\frac{\theta_b}{2}\right) N b = N_{\mathrm{K}} L_{\mathrm{K}}. \tag{1.12}$$

Another length scale which characterises the flexibility is called the *persistence length* L_{P}. It is conceptually the length for which the orientation between two points on a linear polymer is lost, or equivalently, the length scale for which the orientation of one segment *persists* along the chain. If we take the previous example of a freely rotating segment model and consider $\theta_{\mathrm{b}} \approx 0$ in the limit $b \to 0$ (and $N \to \infty$, thus keeping L_{C} constant) we obtain a smooth semirigid chain which we will call a *wormlike chain* [1]. Let us consider a unit tangent vector $\vec{v}_i$ on some monomer i: one can imagine the tangent vector on $\vec{v}_{i+1}$ to be closely aligned ($\langle \vec{v}_i \cdot \vec{v}_{i+1} \rangle \approx 1$). If we consider n monomers away from i, the alignment (on avergage) will eventually be lost ($\langle \vec{v}_i \cdot \vec{v}_{i+n} \rangle \approx 0$). The "alignment" experiences an exponential decrease as we consider monomers further away. The persistence length L_{P} can more rigorously defined as some length scale which characterises the loss of alignment [1],

$$e^{-n/L_{\mathrm{P}}} = \langle \vec{v}_i \cdot \vec{v}_{i+n} \rangle. \tag{1.13}$$

Note that here L_{P} is expressed in units of monomers, but is usually expressed as a length (e.g., for ssDNA L_{P} is said to be $\approx$ 10nm, rather than $\approx$ 23 bases). Since we are now considering what is essentially a smooth polymer derived from the (discrete) freely rotating model, it is more appropriate to express a stiffness parameter as L_{P} instead of θ_{b}. Exact

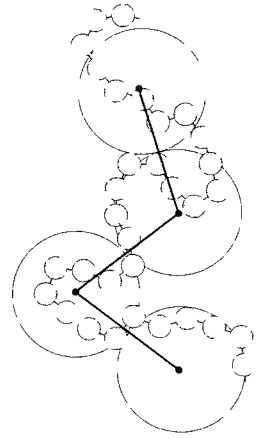


Figure 1.3 A polymer can be rescaled using spheres of diameter L_K such that these Kuhn beads (or equivalently segments) can be seen as being freely-joined— there is no significant potential energy associated with the longer length scale bending

values of the size of such a chain can be calculated [1] :

$$\langle R_F^2 \rangle = 2L_P[L_C + L_P(\exp(-L_C/L_P) - 1)] \tag{1.14}$$

$$\langle R_G^2 \rangle = \frac{1}{3}L_P L_C - L_P^2 + 2\frac{L_P^3}{L_C}\Big(1 - \frac{L_P}{L_C}[1 - \exp(-L_C/L_P)]\Big). \tag{1.15}$$

In the long chain limit $L_C \gg L_P$ one obtains

$$\langle R_F^2 \rangle \approx 2L_P L_C = 2L_P N_K L_K. \tag{1.16}$$

Comparing this to Eqn. 1.10 we obtain the relationship [1]

$$L_K = 2L_P. \tag{1.17}$$

The conceptual picture behind this relation can be explained as follows [3]: The persistence length is the characteristic length over which a particular segment's orientation persists along the chain, as described by Eqn. 1.13. This orientation must also persist on the other side of $\vec{v}_i$, thus at a length scale of $2L_P$, the orientations are not correlated— which is the requirement for L_K.

1.1.1 Deoxyribonucleic acid

Yes, DNA is a polymer. As a matter of fact, it is a worm-like chain.

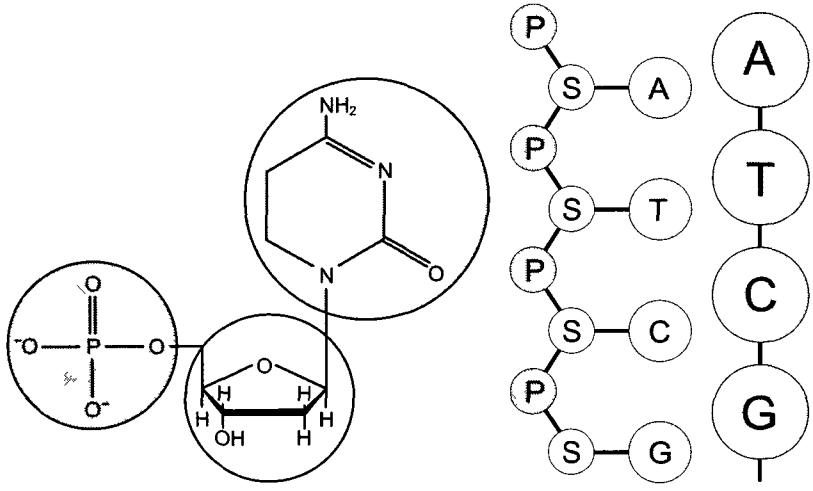


Figure 1.4 (a) A cartoon representation of the molecular structure of a single nucleotide (the repeating units of a ssDNA fragment) Circled into groups are (from left to right), a phosphate group, a sugar group, and coding base (cytosine is pictured here) (b) The sugar (S) and phosphate (P) groups make up the backbone unto which are attached the coding bases, adenine (A), thymine (T) cytosine (C) and guanine (G) (c) The bases are shown by a circular representation These figures were inspired by [4]

This section briefly covers deoxyribonucleic acid (DNA) from a structural point of view to understand the design motivations of our coarse-grained DNA model. One can consider double stranded DNA (dsDNA) as two side-by-side chains with hydrogen bonds linking every monomer of one chain to a complementary monomer on the other chain. A single stranded DNA (ssDNA) is one of the two polymer chains. Let us consider the monomers that make up these chains. Figure 1.4(a) shows the molecular structure of these units— the nucleotides. The nucleotide shown in the figure is more specifically a cytosine nucleobase. There are three more nucleobases: adenine, thymine and guanine. These differ from one another by the base attached to the sugar group. See Fig. 1.4(b) for a cartoon of these nucleotides constituting a small $N = 4$ ssDNA chain.

In order to form a dsDNA fragment, the nucleobases need to form hydrogen bonds with their complementary nucleobase, thus forming a *basepair*: adenine with thymine, and cytosine with guanine; the former with two hydrogen bonds and the latter with three. Thus AT (or TA) basepairs are less stable than CG (or GC) basepairs. Two complementary

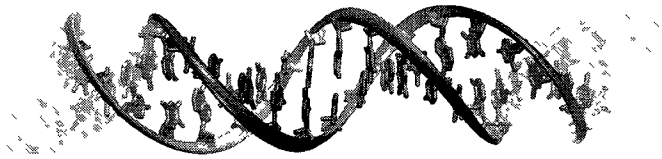


Figure 1.5 The famous double helix is formed when two complementary ss-DNA fragments are paired. This image has been generated using the standard coordinates obtained from the protein databank and visualised using VMD.

ssDNA fragments form the iconic double helix shown in Fig. 1.5 when they are paired. Close observation of Fig. 1.5 can reveal an important cooperativity mechanism between neighboring basepairs— *stacking interactions.*

The planes of the aromatic compounds of the nucleobases tend to lie parallel with one another through stacking interactions. This contributes to the stability and rigidity of the DNA structure— neighboring bases on the same strand stack together, but are also subject to aligning themselves with the complimentary strand via the pairing hydrogen bonds. The significant cooperativity between neighboring bases will make it difficult to break a single basepair into two unpaired bases, since it is surrounded by neighboring bases that (via the stacking interactions) keep them from changing positions. We will see later that these will usually break in groups, or domains. We can also take this opportunity to mention that naturally, unzipping at the extremities is easier since half of the stacking interactions (which contribute to the stability) do not exist.

The stacking interactions are, of course, dependent on the identity of the nucleaobase. This phenomenon can be observed by direct measurement of the stiffness along the chain. Sequence-dependent rigidity, as well as sharp bends, or *kinks*, along a DNA chain are mainly due to these interactions.

As far as the physics behind the structure of a DNA fragment is concerned, we have

covered what is needed to understand my research, but far more remains. One cannot talk about DNA without (even briefly) mentioning biological implications. One of my favorite metaphors has been attributed to Trevor Spencer Rines: "DNA: the web which spins the spider" [5]. I can maybe venture to give my own (biased) metaphor: "The output of a quine which doubles as its compiler"[3]. Of course, this cannot quite serve as a metaphor intended for the layperson. In short, DNA contains the genetic makeup of an organism— but it needs the organism to make anything out of it.

So the genetic makeup of an organism lies in combining the four bases into long (sometimes circular) strings. Due to the sheer size of a complete genome (for humans$\sim 10^9$ basepairs), sequencing the whole DNA can be an overkill for certain applications. Biologists use a variety of tricks to draw from this well of information. The DNA strings can be cut at very specific sites; for instance, the ECOR1 restriction enzyme will cut a DNA string at every occurrence of the sequence `gaattc`. If the resulting fragments can be separated by size, the distribution of fragment lengths can be used as a genetic fingerprint. There are many ways to achieve this; one that will be the focus of my research is gel electrophoresis. It consists of pushing the DNA fragments into a sieving gel... but what is a gel?

1.1.2 Polyacrylamide gels

Polymers can be used to make gels.

When I was a kid, I was not too happy to learn that my favorite desert was made from boiled animal skin and bones— the long triple helix of collagen being the magic ingredient in most gelatins. I like to picture a block of gel as a sponge with a very fine structure. It is not very exciting to see a sponge retain a given amount of water. But there is something intriguing (perhaps because we do not see the gel fibres) when we consider that a gelatin cube is mostly water— with the gel fibres holding its shape. Here we will solely focus on gels made from polyacrylamide. Polyacrylamide gels are linear polymers made from (you guessed it) acrylamide monomers (C_3H_5NO), crosslinked via N,N' ethylene bisacrylamide ($C_7H_{10}N_2O_2$). It can be pictured as a thatch-like structure of fine polymer fibers chemically cross-linked with one another. A cartoon of the gel structure is shown in Fig. 1.6(a). The black dots represent bisacrylamide, which permit a linear segment to branch out, like a 'Y'-shape, into two other linear segments. As the gel grows, or as *polymerization* progresses it

[3] A.I. software engineers look no further, here also lies the source code to consciousness.

(a) (b) (c)

Figure 1.6 The structure of polyacrylamide gels. (a) Cartoon depicting the crosslinked structure of a polyacrylamide gel at various stages of polymerisation (b) The gel (c) Transmission-Electron Microscopy image of a polyacrylamide gel with 10% T and 5% C [7]. This image is licensed under the Creative Commons Attribution 3.0 Unported and can be reproduced without written permission from the copyright owner [8].

begins to ressemble a sponge. Ends of linear acrylamide segments are considered to be short and very flexible, thus they can be neglected from our physical picture. A transmission-electron microscopy image is shown in Fig. 1.6(c). The gel can be characterised by its chemical makeup: %T and %C, where %T is the combined (w/v) concentration of acrylamide with crosslinker, and %C is the ratio (w/w) of cross-linker found in %T [6]. But for us, what matters are its physical traits. No, not bulk characteristics like Youngs' modulus, but rather micro-traits like an average pore size and effective fibre radius.

1.2 DNA electrophoresis in a Polyacrylamide gel

Size separation of DNA fragments can be achieved by pushing the fragments through a gel, which acts like a sieve. It generally permits smaller fragments to pass through faster than longer fragments. In solution, DNA fragments become charged. Thus an applied electric field can be used to drive the fragments through the medium; this process is called *gel electrophoresis*. The migration of a DNA fragment is characterised by its *mobility* μ. This is defined as the net velocity divided by the electric field. The stronger the mobility depends on the number of monomers N, the better the separation. There are three major

electrophoretic regimes which depends on the field intensity, DNA size $\langle R_G \rangle$, and average pore size $\langle R_P \rangle$. We shall present a brief qualitative overview of these regimes. The main qualitative differences between the regimes are the relative size of the DNA fragment with the gel pore size and the dominant conformations of the fragment.

The way a DNA fragment is confined in a gel depends on both the size of the fragment $\langle R_G \rangle$ and the size of the gel pores $\langle R_P \rangle$. Here, we will assume that the DNA fragment is flexible on a length scale comparable with the pore size, $L_K \ll \langle R_P \rangle$. Sometimes when the fragment is larger than the pores, it is forced to take up in multiple pores. The plot in Fig. 1.7(d) shows how this effect can lead to various migration regimes.

The so-called *Ogston* regime occurs when the size of the DNA fragment is smaller than the average pore size $\langle R_G \rangle < \langle R_P \rangle$, and when the field is too weak to significantly deform the DNA fragment. The cartoon in Fig. 1.7(a) shows a fragment in a gel under these conditions. The sieving picture here is probably accurate: larger fragments form bigger coils and have more collisions with the obstacles, and thus will migrate significantly slower. Looking at Fig. 1.7(d) there is a mobility dependence on the number of monomers, which makes size separation possible.

The *reptation* regime occurs when the DNA fragment needs to span multiple pores $\langle R_G \rangle > \langle R_P \rangle$. The fragment can still be seen as having a random walk conformation but since it is 'fused' in the gel, lateral movements are not possible. The only mechanism of migration available to the fragment is by moving head first like a snake in high grass, hence the term reptation. This is usually modeled as a polymer performing a 1D random walk inside a sinuous random 3D *tube*. Again, looking at Fig. 1.7(d) we can see that in this regime, size separation is possible.

At longer DNA lengths the fragment will orient itself with the field. The *oriented reptation* regime is depicted in Fig. 1.7(c). The right part of Fig. 1.7(d) shows that under these conditions, separation is not possible: the mobility becomes independent of N.

We just went over a qualitative description of the migration process. We did not go over the theoretical approaches used to model these regimes nor do we mention their results. We also did not talk about the effect of the field on these regimes. These effects are beyond the scope of this introduction. However, before we continue, we should mention that in a random gel, such as polyacrylamide, there is sometimes a sub-regime which occurs at the crossover

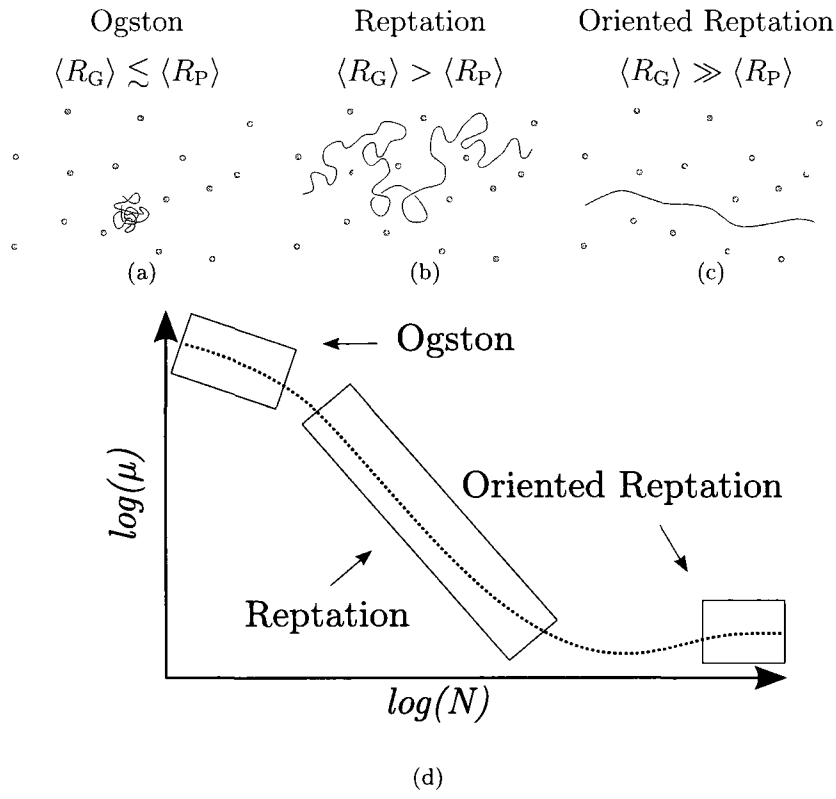
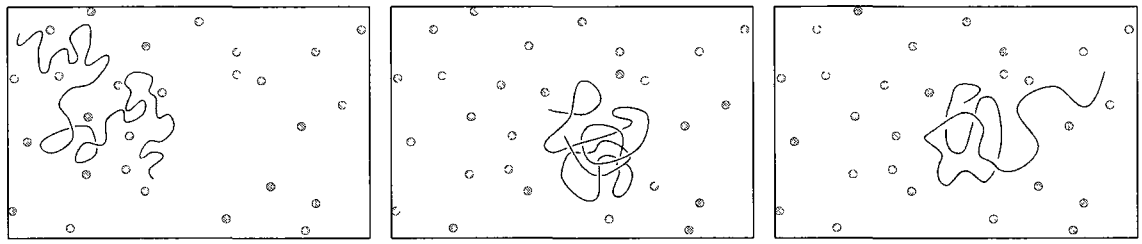


Figure 1.7 A two dimensional cartoon representation of the three main electrophoretic regimes. Cross sections of the gel fibres are shown as small grey circles. (a) In the Ogston regime a DNA fragment is roughly confined in a single pore. (b) In the reptation regime the DNA is too big to be confined in a single pore. It takes up multiple pores in a random walk-like conformation. (c) In the oriented reptation regime, a very long fragment orients itself in the direction of the field. (d) In this log-log plot of the DNA mobility *vs.* the number of monomers N, we can see the three dominant slopes corresponding to the three regimes. For the oriented reptation regime, the zero-slope means that the mobility is size-independent. This image has been inspired from [9]

between Ogston and reptation. This is due to the natural variance of the gel pore sizes. In this regime called *entropic trapping*, the fragment exhibits behaviours which were previously exclusive to the reptation and Ogston regime. The cartoon in Fig. 1.8(a) shows a fragment reptating in a gel. Sometimes the fragment will reptate in a volume of the gel for which the fibres are not as closely packed. In such a *depleted* region, showed in Fig. 1.8(b), the fragment is allowed form a coil. In order to continue its migration, the coil has to be stretched out because the neighboring pores are too small to accomodate this conformation. Due to



(a) (b) (c)

Figure 1.8 Entropic Trapping occurs between the Ogston and reptation regime $\langle R_G \rangle \approx \langle R_P \rangle$. The naturally occurring variance in the pore sizes causes the fragment to sometimes collapse as a coil inside a pore— but will need to reptate between these states. (a) The fragment reptates in a volume of the gel that happens to have small pores. (b) Naturally occurring large pores, or *pockets* enables the fragment to collapse into a random coil, maximizing its conformational entropy. (c) In order to continue its migration the fragment needs to, again, reptate in multiple pores, but this imposed conformation is not entropically favourable.

the entropic cost related to this effective stretching, the fragment will spend an additional amount of time in this trap before eventually leaving head first, as shown in Fig. 1.8(c).

If I can allow myself to foreshadow, it might be useful to note that these traps do not have to come from occurrences of bigger pores. If we go back to Eqn. 1.11, we can see that $\langle R_G^2 \rangle$ depends on the stiffness of a polymer. If we can have a DNA fragment which has a value for L_K which is allowed to be dynamic, perhaps entropic trapping can occur during the times when the fragment is very flexible.

It was previously mentioned that the hydrogen bonded basepairs in DNA can sometimes be broken. At high temperatures, the basepairs become unstable and the dsDNA fragment can begin to unzip into two ssDNA fragments. This is called *melting*, or *denaturation*. Since the stability of the basepairs are dependent on the nucleaobase identity, AT pairs being weaker than CG pairs, the denaturation profile of a DNA fragment is sequence-dependent. It has been experimentally observed that a dsDNA fragment exhibits a steep reduction in gel electrophoretic mobility when it is in an environment that favours denaturation.

There is an experimental technique called *temperature gradient gel electrophoresis* (TGGE) which exploits this phenomenon. It consists of two electrophoretic separation processes in different dimensions. First, a DNA is cut at specific sites using a restriction enzyme, then

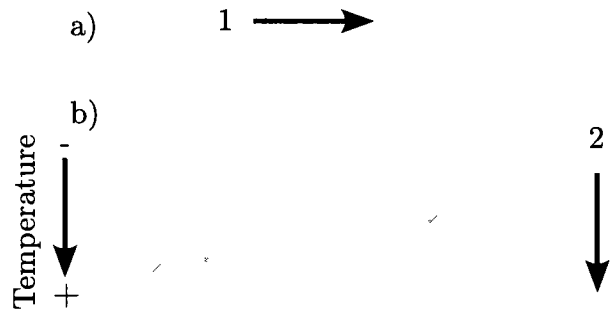


Figure 1.9 A cartoon of a typical result from temperature gradient gel electrophoresis (a) Standard gel electrophoresis of digested fragments is first conducted to separate the fragments by size (b) Sequence separation is then achieved by an electrophoretic run against an increasing temperature gradient The AT-rich fragments denature sooner and exhibit a sudden decrease in mobility which can be observed by earlier blocking in the gel

the *digested* fragments are size separated using standard gel electrophoresis. This constitutes the first dimension of the 2-dimensional profile. In the second dimension, the (now size-separated) fragments are again electrophoresed in a gel slab, but this time, against a temperature gradient (from cold to hot). As the fragments move to a hotter environment, the basepairs begin to denature and a steep reduction in mobility is observed. For all practical purposes, the fragment is *blocked.* Thus, fragments which had a similar size and could not be resolved during the first electrophoretic run, like the third band in Fig. 1.9(a), can perhaps be separated during a sequence-specific separation method— shown in Fig. 1.9(b).

There are many variants to this setup: the temperature gradient is sometimes perpendicular to the field or a chemical denaturant gradient is sometimes used. The main concept however, remains the same, so we will not go over them. It has been shown that sequences which differ by the identity of a single basepair can generally be differentiated using this technique [10]. One can use this technique as a litmus test for a particular strain, say H1N1[4] of the influenza virus. One could use the result obtained from this strain as a reference and compare it to an unknown sample Other uses include profiling the evolution of a certain species by mapping the mutations.

[4] Sorry, it was too tempting

1.3 Simulation Model

The articles in Chapter 3 and Chapter 4 describe the simulation model which I used as a primary investigation tool. Each chapter focuses on a specific result of the simulation and equally focuses on the model details which are relevant to these results. My simulation model basically consists of two bead-spring polymers, each representing a ssDNA chain, pair-bonded with one another using again more springs, thus creating a dsDNA chain. The individual beads have a repulsive potential between one another which is modeled by a truncated and shifted Lennard-Jones potential. The potential is truncated at the well minimum and shifted to zero at this point to prevent discontinuity issues. This is also called the Weeks-Chandler-Anderson (WCA) potential [11]:

$$U_{WCA}(r) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \epsilon & \text{for } r < r_c \\ 0 & \text{for } r \geq r_c \end{cases}, \tag{1.18}$$

where r is the separation between two beads, ϵ is the depth of the well— used here as our fundamental unit of energy, σ is the nominal bead diameter— used as our fundamental unit of distance, and r_c is the cutoff radius — chosen at the well minimum $r_c = 2^{\frac{1}{6}}\sigma$. This choice is a practical one. It is absolutely not meant to model Pauli repulsion combined with the attractive Van der Waals interactions. It is merely a convenient form for excluded-volume which is relatively simple to calculate. Since distances are calculated by $r^2 = r_x^2 + r_y^2 + r_z^2$, any mathematical form which has even powers of r can avoid the costly square-root function.

The springs used are called finitely extensible nonlinear elastic (FENE) springs. They too have a root in physics as first terms of the inverse Langevin function, but we won't go into that. Although it can be debatable, it is again chosen here due to its convenient form:

$$U_{\text{FENE}}(r) = -\frac{1}{2}\, k_{\text{FENE}}\, r_0^2 \ln\left[1 - \frac{r^2}{{r_0}^2}\right], \tag{1.19}$$

where k_{FENE} is the spring constant and r_0 is the maximum extension. The unphysical situation of bond crossing can be avoided by using the Kremer-Grest parameters [12]: $k_{\text{FENE}} = 30\epsilon/\sigma^2$ and $r_0 = 1.5\sigma$.

Once the beads are given an initial position and velocity, they can be integrated in time. Langevin dynamics (LD) is the principal time-integration scheme which is used. I

have written about it in a group review article which can be found in Appendix A (see section 2.4 of the review article for my description of LD). In short, it can simply be seen as an application of Newton's second law of motion:

$$m\vec{a}(t) = \vec{F}^{\mathrm{C}}(t) + \vec{F}^{\mathrm{D}}(t) + \vec{F}^{\vec{B}}(t), \tag{1.20}$$

where m is the mass of a bead, $\vec{F}^{\mathrm{C}}$ is the resulting force of the two potentials we discussed previously, $\vec{F}^{\mathrm{D}}$ is the hydrodynamic drag term and $\vec{F}^{\mathrm{B}}$ results from random kicks from the solvent molecules. Other details concerning the simulation model or LD will be discussed in the following Chapters and in the Appendix.

1.4 Bibliography

[1] Iwao Teraoka. *Polymer solutions. an introduction to physical properties.* John Wiley & Sons, New York, USA (2002).

[2] M Rubinstein, RH Colby. *Polymer physics.* Oxford University Press, USA (2003).

[3] JRC van der Maarel. *Introduction to biopolymer physics.* World Scientific Publishing Co. Pte. Ltd., Toh Tuck Link, Singapore (2007)

[4] TKnotts IV, N Rathore, D Schwartz, J de Pablo. *A coarse grain model for dna,* ***The Journal of Chemical Physics*** **126** (2007).

[5] S Strauss. *We need a satisfactory metaphor for dna,* ***New Scientist*** (2009).

[6] NancyC Stellwagen, Earle Stellwagen. *Effect of the matrix on dna electrophoretic mobility,* ***Journal of Chromatography A*** **1216**, 1917–1929 (2009).

[7] Reinhard Rüchel. *Tem image of a polyacrylamide gel.* http //en wikipedia.org/wiki/File:TEM_image_of_a_polyacrylamide_gel.jpg.

[8] http //creativecommons org/licenses/by/3 0/deed.en.

[9] G Slater, J Noolandi. *The biased reptation model of dna gel electrophoresis: Mobility vs molecular size and gel concentration,* ***Biopolymers*** **28**, 1781–1791 (1989).

[10] R Myers, S Fischer, L Lerman, T Maniatis. *Nearly all single base substitutions in dna fragments joined to a gc-clamp can be detected denaturing gradient gel electrophoresis,* ***Nucleic Acids Research*** **13**, 3131–3145 (1985).

[11] JD Weeks, D Chandler, HC Andersen. *Role of repulsive forces in determining the equilibrium structure of simple liquids,* ***J. Chem. Phys.*** **54**, 5237–5247 (1971).

[12] G Grest, K Kremer. *Molecular dynamics simulation for polymers in the presence of a heat bath,* ***Physical Review A*** **33** (1986).

Chapter2

Physical interpretation of the L_{r} parameter in the theory for the gel electrophoresis of partially denatured DNA

D Sean, GW Slater (2010) In press *Electrophoresis* (Accepted for publication on July 8, 2010.

Short Communication:

Physical interpretation of the L_r parameter in the theory for the gel electrophoresis of partially denatured DNA

David Sean, Gary W Slater

Department of Physics, University of Ottawa,
150 Louis-Pasteur, Ottawa, Ontario, Canada, K1N 6N5

September 3, 2010

Corresponding author:
Prof Gary W Slater
email gary slater@uOttawa ca
fax +1-613-562-5190

Keywords:
DNA denaturation, DNA gel electrophoresis, Denaturing gel electrophoresis, Electrophoretic mobility modeling

Running title:
Physical interpretation of the L_r parameter in the theory for the gel electrophoresis of partially denatured DNA

Abbreviations:

Abstract

Partial strand melting of dsDNA during gel electrophoresis typically results in an abrupt reduction of mobility. Several DNA analysis technologies are based on this phenomenon. Inspired by the de Gennes theory for the reptation of branched polymers in gels, Lerman *et al* (*Ann. Rev. Biophys. Bioeng.* 1984 *13* 399–423) proposed a mathematical expression to predict this reduced mobility. The latter contains only two parameters: the average number of denatured bases p (which can be obtained using a theory for DNA melting) and a constant L_r. However, there is confusion in the literature regarding the physical interpretation of L_r, and little is known about its dependence upon experimental parameters. The purpose of this short communication is to derive an explicit equation for the parameter L_r from the de Gennes theory of reptation. Our derivation sheds light on the meaning of L_r, clarifies the scope of the underlying approximations, and makes predictions about the dependence of L_r upon the gel pore size and the persistence length of ssDNA.

In a well known 1975 paper Pierre-Gilles de Gennes considered the reptation of branched polymers from a theoretical point of view [1] He showed that an exponential decrease of the diffusion coefficient with the length of the branches should be expected [1 2] Given this dependence it was argued that in a crosslinked gel the diffusion coefficient will rapidly drop for a polymer with a branch occupying only a few pores A few years later Lerman & Frisch [3] noticed the qualitative similarities between the de Gennes prediction and the steep decrease of electrophoretic mobility that they had been observing for partially denatured DNA These authors recognized the topological similarities between a polymer with a branch and a partially denatured DNA fragment and proposed that the mechanisms behind the electrophoretic blocking of partially denatured DNA could be similar to the ones described by de Gennes for the reptation of a branched polymer [3] Lerman *et al* [4] then suggested the following formula for the mobility of a partially melted DNA fragment in a gel

$$\mu = \mu_0 e^{-p(T)/L_r} \tag{1}$$

where μ_0 is the electrophoretic mobility of the native fragment T is the effective temperature and $p(T)$ is the sum over the base pairs along the fragment s contour length of the probability of being in a melted state (this can be obtained via the Poland algorithm [4 6]) It is conceptually simpler to think of $p(T)$ as the mean number of bases in the melted state at any given moment

Unfortunately the physical meaning of the remaining parameter L_r is unclear Its numerical value which is essentially always used as a fitting parameter (this may explain part of the success for the formula) varies between 45-200 bases in the literature [7 9] The reported values generally exceed both the Kuhn length of ssDNA and the typical pore size in a polyacrylamide gel When Lerman *et al* first published their theory they introduced L_r as being the length of the flexible unit [4] Unfortunately the paper does not clearly define which flexible unit is involved Furthermore the paper does not describe how the properties of the gel should be related to this parameter Interestingly the physical interpretation of L_r has not varied much in the literature since 1984 The very fact that the ill-defined expressions the length of the flexible unit' [4 8 10–14] or a constant [15 16] are used in most if not all of the literature indicates that the theory is still incomplete As far as

the present authors are aware there is no systematic experimental study of this parameter in the literature In the following we present a short version of the de Gennes calculation followed by our interpretation of the Lerman *et al* formula This will lead to an explicit equation for the parameter L_r (which includes the gel pore size a key parameter since it is the gel that blocks the molecule) and a number of associated testable predictions

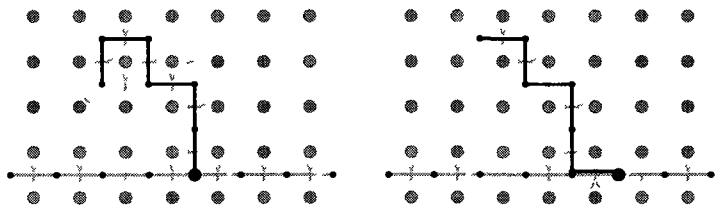


Figure 1 The system pictured by de Gennes (a) A branched polymer in a relaxed state (b) An elastic restoring force appears after displacing the polymer backbone one unit lattice length towards the right

The de Gennes model consists of an ideal linear polymer made of freely jointed segments with a single branch attached to the backbone at some point away from the ends [1] The polymer is placed in a gel modeled as a square lattice of point like obstacles Figure 1(a) shows the state of the polymer after it has had adequate time to reach equilibrium If the main polymer takes a steps of unit lattice length part of the branch will be forced to lie on top of the main polymer ($a = 1$ in Fig 1(b)) We can picture the effect of this folded section as having reduced the length of the branch by a segments The entropic cost of this retraction is $\Delta S \cong -a \ln z$ where z is the lattice coordination number ($z = 4$ for the square lattice pictured here) The resulting entropic restoring force generally dominates the reptation dynamics of the main polymer because the branch needs to relax in its new position between the reptation steps To consider the branch relaxation let us hold the main polymer in place and wait for the arm to relax to its new (translated) lateral position de Gennes calculated the fraction of states for which the end of the arm retreats back on itself (as a mechanism of relaxation) and obtained the fraction of retreated states $P(N_{\rm seg}) = \tilde{A}(N_{\rm seg})e^{-\alpha N_{\rm seg}}$ where $N_{\rm seg}$ is the number of segments in the arm $\tilde{A}(N_{\rm seg})$ is some weak function of $N_{\rm seg}$ which we

will neglect and $\alpha = \frac{1}{2}[\ln z + \ln(\ln 2)]$ is a constant of order unity related to the nature of the lattice. Neglecting the slowly varying pre-factors de Gennes argued that the diffusion coefficient should be reduced by $P(N_{seg})$ and thus obtained

$$\mu = \mu_0 e^{-\alpha N_{seg}} \tag{2}$$

where μ_0 is now the mobility of the unbranched polymer. This expression from de Gennes is the starting point of the model proposed by Lerman *et al* [4]. In fact, comparing the two expressions we obtain

$$L_r = \frac{\bar{p}}{N_{seg}\alpha} \tag{3}$$

Since $\bar{p}$ is the number of bases in the side branch, L_r is effectively the number of denatured bases per effective de Gennes lattice segment or blob. Therefore the expression "the length of the flexible unit [in bases]" [4] actually refers to the number of denatured bases required to fill one gel pore. Since the blobs in the de Gennes theory trace a freely-jointed random walk path emanating from the backbone, it is proper to call this *the size (instead of the length) of the flexible unit (or segment) in the corresponding reptation model*.

To estimate the number of denatured bases in one de Gennes segment, we consider branches arising from the denatured domains situated at the ends of the fragment since it is typically there that denaturation starts. dsDNA denaturation obviously gives rise to two melted arms (see Fig. 2(a) and Fig. 2(b)). We consider one of the arms as a branch and the other as belonging to the main polymer. Since ssDNA is more flexible than dsDNA, this process actually shortens the length of the backbone of the resulting branched polymer. It is also possible to treat a hernia formed by a bubble as a branch as depicted in Fig. 2(c), but these branches are typically shorter (for the same number of denatured bases) and will thus have less effect on the mobility. Since $p(T)$ gives the average number of denatured base pairs, and since they typically denature in domains, we can, as a first approximation, use it as the number of bases in a side branch. It is important to note that using $p(T)$ as a measure of the degree of denaturation does not capture the topological difference between bubbles and split ends, nor does it discriminate between single and multiple branches. Nonetheless it has been used as a measure of the effective total branch size and we also adopt it as such for our calculations.

(a) (b) (c)

Figure 2: (a) The native dsDNA fragment inside the gel. (b) Two ssDNA branches arrise from melting at one of the ends. We will consider only one branch. and the other as the continuation of native fragment. (c) When the denatured region is not placed at the end of the fragment. the effective branches is though to have a smaller impact on the mobility. These cases will not be treated.

In the context of the blob theory. a branch is a chain of N_{seg} blobs with radius b. the mean pore size of the gel. Thus the *primitive chain* of blobs has a *Kuhn length* of $2b$. As indicated in Eq. 3. the key parameter is the number of bases per blob. $\bar{p}/N_{\text{seg}}$. The square end-to-end distance of the primitive chain must be the same as the square end-to-end distance of the underlying ssDNA chain. In the flexible chain limit where $L_K < b$. we obtain:

$$N_{\text{seg}} b^2 = \bar{p} \ell L_K. \tag{4}$$

where ℓ is the contour length of a base and L_K is the Kuhn length of ssDNA. Re-arranging this and inserting it in Eq. 3 gives

$$L_r = \frac{b^2}{\alpha L_K \ell}. \tag{5}$$

This shows that L_r is actually proportional to the square of the mean pore size b and inversely proportional to the flexibility of ssDNA as measured by L_K. There are thus two ways to change the value of L_r. First. one can change the gel concentration C in order to affect the pore size $b(C)$. Typically. the mean pore size is related to the gel concentration through a power law relation $b \sim \frac{1}{C^x}$ [17]. where the exponent x is generally in the range $\cong \frac{1}{2} - \frac{3}{4}$. Second. one can change the flexibility of the ssDNA branch by changing the buffer's ionic strength I since the Kuhn length L_K contains a component related to the electrostatic repulsion between the bases; this component is $\sim 1/I$ [18]. In order to test our predictions. one could vary C and/or I and measure the effect on the experimental value of the stopping parameter $L_r(C.I)$.

Let us now examine the published numerical values of L_r given the equation that we have derived above. Using the standard values $\ell = 0.43$ nm, $L_K \approx 6$ nm [18] and $\alpha = 7$ (corresponding to a cubic lattice with $z = 6$) together with the published range of $L_r = 45 - 200$ [7, 9], we obtain pore sizes $b \cong 9 - 19$ nm which is consistent with what one could expect for polyacrylamide gels [17].

The fact that the Lerman *et al.* approximation is considered to be valid only for the initial stages of denaturation [8, 13, 14] supports the idea that blocking may primarily be due to denaturation from the ends (for entropic reasons, the bases at the ends are usually the first to melt). Therefore, the parameter $\bar{p}(T)$ (which takes into account the total amount of denaturation along the fragment) will most likely overestimate the number of denatured bases constituting branches located at the ends of the molecules; as a result, we expect that the experimental values of L_r obtained by fitting data will be systematically higher than the value predicted by theory. However, it is in principle possible to correct for this by redefining $p(T)$ as a weighed sum. More weight can be given to the melted bases that are located near the ends of the fragment (this can be computed using the available software for DNA melting).

For the same number of denatured bases, the effective length of a branch composed of a bubble is expected to be shorter than a branch arising from a denatured end (compare Fig. 2(b) with Fig. 2(c)). We thus predict that the value of L_r would be larger for fragments where the dominant denatured region is a bubble. In 1982 Lyamichev *et al.* [19] devised an experiment to compare the electrophoretic blocking of a DNA fragment containing a denatured bubble with one containing a split end. We believe the results they obtained (their Figure 8) to be inconclusive. It seems that both of these fragments are completely blocked in the gel— the reptation, as de Gennes would put it, is completely *quenched*. In order to appreciate any differences between these cases, it is imperative that the comparisons are not made in the $\mu \to 0$ limit. The experiment should thus be repeated with a smaller denatured region or with a gel characterized by larger pores. Experiments are sometimes performed with an attached CG-rich region to prevent complete fragment denaturation [11]. Another method to probe the effects of split ends *vs.* bubbles would be to attach such a clamp at each end, thus imposing a bubble-like domain.

In conclusion we have shown how the Lerman equation can be derived from the de Gennes theory for the diffusion of branched polymers in gels The parameter L_r can be physically interpreted as the number of denatured bases (constituting a branch) per gel pore this clarifies the fundamental nature of this parameter Based on Eq 5 we made a series of testable predictions that could be used to validate the theory Several methods use denaturing gradients along the direction of net migration to separate molecules with the same size but with different compositions In essence these methods make p increase during migration Eq 5 actually suggests that a similar result could in principle be obtained by decreasing the value of $L_r(C\ I)$ along the direction of migration This could be achieved by the use of a gel concentration gradient or an ionic strength gradient Furthermore in experiments using denaturant/temperature gradient gel electrophoresis (DGGE/TGGE) conditions could be chosen which would yield an optimal value for $L_r(C\ I)$ ($i\ e$ the resolution of a specific target fragment in a 2DGE experiment can be optimized by careful tuning of C and I)

The authors would like to thank Hendrick de Haan and Tyler Shendruk for helpful discussions and critical reading of the manuscript The work was supported by a Natural Science and Engineering Research Council (NSERC) of Canada Discovery Grant to GWS

References

[1] de Gennes P G *J Phys* 1975 *36* 1199 1203

[2] de Gennes P G *Scaling concepts in polymer physics* Cornell University Press Ithaca 1979

[3] Lerman L S Frisch H L *Peptide Science* 1982 *21* 995 997

[4] Lerman L S Fischer S G Hurley I Silverstein K Lumelsky N *Ann Rev Biophys Bioeng* 1984 *13* 399 423

[5] Fixman M Freire J J *Biopolymers* 1977 *16* 2693 2704

[6] Poland D *Biopolymers* 1974 *13* 1859 1871

[7] Mercier J Kingsbury C Slater G W Lafay B *Electrophoresis* 2008 *29* 1264 1272

[8] Zhu J Wartell R M *Biochemistry* 1997 *36* 15326 15335

[9] Steger G *Nucleic Acids Res* 1994 *22* 2760 2768

[10] Myers R Fischer S G Maniatis T Lerman L S *Nucleic Acids Res* 1985 *13* 3111–3129

[11] Myers R Fischer S G Lerman L S Maniatis T *Nucleic Acids Res* 1985 *13* 3131–3145

[12] Riesner D , Steger G Zimmat R Owens R *Electrophoresis* 1989 *10* 377–389

[13] Ke S H Wartell R M *Biochemistry* 1995 *34* 4593–4600

[14] Wartell R M Hosseini S Powell S Zhu J *J Chromatogr A* 1998 *806* 169–185

[15] Shiraishi, M Oates A J Li, X Chuu Y H Sekiya T *Biol Chem* 2004 *385* 967–973

[16] Shiraishi M Sekiya, T , *Nucleosides, Nucleotides and Nucleic Acids* 2006 *25*, 463–473

[17] Holmes D L Stellwagen N C *Electrophoresis* 1991, *12*, 612–619

[18] Desruisseaux C Long D Drouin G Slater G W *Macromolecules* 2001 *34* 44–52

[19] Lyamichev V Panyutin I Lyubchenko Y *Nucleic Acids Res* 1982 *10* 4813–4826

Chapter3

Low-field electrophoretic mobility of partially denatured DNA in a gel: Qualitative and semi-quantitative differences between bubbles and split ends

D Sean, GW Slater (2010) To be submitted shortly

Electrophoretic mobility of partially denatured DNA in a gel: Qualitative and semi-quantitative differences between bubbles and split ends

David Sean, Gary W Slater

Department of Physics, University of Ottawa,
150 Louis-Pasteur, Ottawa, Ontario, K1N 6N5, Canada

September 7, 2010

Corresponding author:
Prof Gary W Slater
email gary slater@uOttawa ca
fax +1-613-562-5190

Keywords:

Langevin Dynamics simulations, Computational modeling, DNA denaturation Gel electrophoresis

Running title:

Differences between bubbles and split ends during the electrophoresis of partially denatured DNA

Abbreviations:

FENE Finitely extensible nonlinear elastic

WCA Weeks-Chandler-Andersen

Abstract

Partially melted DNA is known to exhibit an abrupt decrease of electrophoretic mobility in a gel. Although this is the main phenomenon exploited in TGGE/DGGE not much is known about the physical process responsible for the blocking. While there is a commonly used formula for the reduced mobility based on the theory of branched polymers it does not discriminate between denatured domains bounded on one (split-end) or two sides (bubble). To better understand how the blocking occurs in both of these cases a coarse-grained model of DNA gel electrophoresis is simulated using Langevin Dynamics. The simulations reveal that the low-field mobility is much more sensitive to larger denatured domains located at the ends of the fragment. A denatured domain occurring at the center of the fragment reduces the mobility but at a lower rate.

1 Introduction

In their seminal 1979 paper Fischer and Lerman [1] successfully showed that two-dimensional gel electrophoresis (2DGE) can be used for the separation of DNA The method exploits the fact that upon partial denaturation the mobility of a DNA fragment is drastically reduced By introducing a denaturing gradient in a gel the electrophoretic mobility of DNA fragments with a higher AT content will typically drop first — thus sequence-specific separation was shown to be possible

Lerman and Frisch [2] proposed that the de Gennes model for the reptation of branched polymers, which predicted an exponential decrease of the diffusion coefficient as a function of the length of the branches [3] may also apply for the gel electrophoresis of partially denatured DNA and that the observed blocking could be attributed to the same mechanisms A sufficiently long denatured domain at the end of a fragment can be seen as being branched and a domain forming a bubble can perhaps form a lateral hernia which could effectively act like a branch The number of denatured bases can be accurately calculated from DNA melting theory Using the Poland-Sheraga algorithm [4] Lerman *et al* [5] calculated the thermodynamic average of the total number of denatured bases $\bar{p}(T)$ and argued that the de Gennes prediction can be adapted to predict the mobility μ of partially denatured DNA

$$\mu = \mu_0 e^{-\bar{p}(T)/L_r} \tag{1}$$

where μ_0 is the gel electrophoretic mobility of the native fragment and L_r can be seen as *ipso facto* the characteristic number of denatured bases needed to initiate the blocking Using L_r as a fitting parameter Eq 1 has been found to be in fair agreement with the results of several experiments [6 13]

We have recently proposed a physical interpretation of L_r as the average number of denatured bases needed to form a primitive segment acting like a branch i e the number of denatured bases required to fill one gel pore [20]

$$L_r = \frac{b^2}{L_K \ell} \tag{2}$$

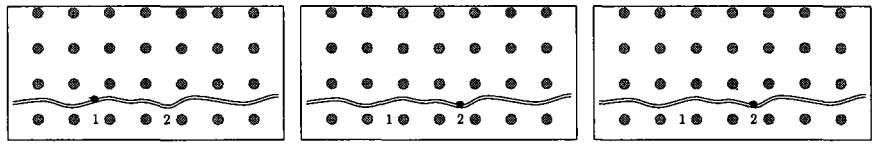
where b is the average gel pore size L_K is the Kuhn length of ssDNA and ℓ is the distance between ssDNA bases Since this form assumes a linear denatured branch it is only valid for split ends Based on the idea that a hernia from a bubble may need more denatured basepairs to form a branched segment of a given size compared to a branch formed by denaturation at the ends we predicted that the value of L_r would be larger for fragments which exhibit blocking solely from bubbles [20] It is not trivial to verify that a branch from a bubble is indeed shorter than a branch from a split-end since the topology of a bubble allows for possible multiple branched segments (clover-leaves) Clearly different mechanisms are at play between these two cases therefore L_r should also depend on the nature of the denatured domains (bubbles *vs* split-ends)

In the case where one end of the dsDNA fragment is partially split into two ssDNA sub chains only one of the sub-chains can be seen as a branch the other would simply the continuation of the main backbone (this would result in a factor of two in the denominator of Eq 2 but other factors of order unity have been dropped for simplicity) The maximum length of the branch constituting one of the ssDNA sub chains would be half the contour length of the total denatured region (which includes the other ssDNA sub-chain)

In the case where the denatured region is bound on both sides by native basepairs (bubble) the denatured bases which does not constitue a hernia cannot be seen as a branch These too would simply be the continuation of the main backbone If we consider the limiting case where the entire bubble forms a branched hernia the length of the two resulting bubbles branch cannot exceed one quarter of the contour length of the denatured bubble Thus for the same number of denatured bases a hernia from a bubble can *at best* form two branches half as long as one of the denatured split ends

Thus the effect of a branch composed of a hernia should have less of an impact on the mobility for two reasons 1) for the same number of denatured bases the hernia arising from a bubble should be smaller (i e it should fill fewer pores constituting a branch) and 2) the entropic cost of *threading* a bubble between pores (which is the underlying cause for the blocking) [3] should be lower than for a branch

If we consider de Gennes argument for a branch attached to a reptating polymer [3]

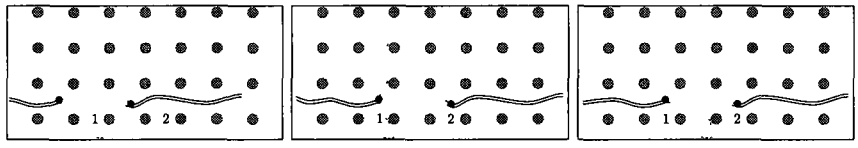


(a) (b) (c)

Figure 1 The de Gennes picture for the reptation of a branched polymer (a) The main polymer can reptate until the elastic restoring force of the branch dominates (b) Before reptation can continue the branch needs to relax in its new position (c) This this thought to occur by having the tip of the branch retract along its contour

we see that the polymer can reptate while pulling along the branch but only up to some point as shown in Fig 1 (a) and Fig 1 (b) After the branching point moves from (1) to (2) the main polymer needs to wait for the branch to relax before the reptation can resume (the elastic restoration force becomes too great for reptation to continue) As a mechanism of relaxation de Gennes proposed that the branch needed to retract upon itself as shown schematically in Fig 1(c) Neglecting weak pre-factors the time required for this method of relaxation has an exponential dependence on the length of the branch— thus the form of Eq 1

When this argument is considered for a bubble one can see that the limiting step should also be the bubble s relaxation But due to the different topology for the same number pores occupied this limiting time should be much shorter as we shall see Figure 2 (a) shows a hernia acting like a branch The standard reptation motion can occur as the main polymer moves forward (this is essentially pure reptation in a tube running along the main polymer and the hernia) Figure 2 (b) shows the state of the system after it can no longer reptate towards the right the fork is at the base of the hernia At this point it can reptate back or wait until the hernia relaxes / diffuses into other pores But by looking at Fig 2 (c) one can see that the bubble can relax in ways the branch couldn t full retraction of the bubble is not necessary since the bubble can slowly slide from one side of a fiber to the other The entropic barrier to the relaxation of a bubble is thus lower than that of the branch In effect the nucleation process for the relaxation mechanism described in Fig 2 (c) involves only



(a) (b) (c)

Figure 2 The limiting scenario for a hernia The opposite bubble is not shown for simplicity (a) the length of the denatured region needs to be at the very least twice as long to form a hernia with an equivalent length This number is increased when considering that parts of the denatured region do not occur in the hernia (b) Reptation can occur in the same manner up to the point where the fork meets the base of the hernia (c) The bubble need not relax in the way described by de Gennes for a branch Different relaxation modes exist which are not as restrictive

about half of the monomers in the hernia Qualitatively this is one of the reasons why the L_r parameter should be much larger for bubbles

Besides these topological considerations many other factors are believed to affect the blocking For example in DGGE the gel porosity the solvent viscosity the homogeneity of the electric field and even the net charge density of the DNA fragment are thought to be influenced by the presence of the chemical denaturant alone [5 14] The fact that similar results can be reproduced with TGGE (even though the solvent s viscosity decreases with increasing temperature) together with the qualitative agreement with Eq 1 highlights the importance of the underlying phenomenon partial denaturation The goal of this paper is to focus solely on physical blocking mechanisms arising from partial denaturation For this reason the side-effects of the denaturant (either chemical or by temperature) on the system as well as the attractive interactions between the gel and the nucleotides (which have been observed to affect the migration [14]) have been purposely omitted from our model

2 Model

In our coarse-grained Langevin Dynamics model a native dsDNA fragment is formed by pairing two ssDNA fragments modeled by two bead-spring chains of the same length They are paired together such that the first bead of one chain is paired to the first bead of the other chain This results in a blunt end dsDNA fragment with no leading ssDNA beads at the extremities (see Fig 3 and Fig 4)

A Lennard-Jones potential truncated at the well minimum is used for excluded-volume interactions between all beads This is also known as the Weeks-Chandler-Andersen (WCA) potential and has the form [15]

$$U_{\mathrm{WCA}}(r) = \begin{cases} 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] + \epsilon & \text{for } r < r_{\mathrm{c}} \\ 0 & \text{for } r \geq r_{\mathrm{c}} \end{cases} \tag{3}$$

where r is the distance between the beads $r_{\mathrm{c}} = 2^{1/6}\sigma$ is the cutoff radius $\epsilon = k_{\mathrm{B}}T$ is the well depth and σ is the nominal bead diameter Consecutive beads of a ssDNA chain are joined by FENE springs which have the potential

$$U_{\mathrm{FENE}}(r) = -\frac{1}{2}\, k_{\mathrm{FENE}}\, r_0^2 \ln\left[1 - \frac{r^2}{{r_0}^2}\right] \tag{4}$$

where k_{FENE} is the stiffness constant and r_0 is the maximum extension The Kremer-Grest values ($r_0 = 1\,5\sigma$ $k_{FENE} = 30\epsilon/\sigma^2$) are used to prevent bond crossing [16] The FENE potential is also used to form the pairing bonds between the monomers of two parallel ssDNA chains (same k_{FENE} and r_0 as before) A melted region of the fragment is created by deleting the pairing bond between two beads Due to the form of the FENE potential the resulting DNA fragment will not have pairing bonds which can break or be formed on their own—the denaturation is static and imposed For this study we refer to ds-beads as beads being in the native state (beads which have a pairing FENE bond with the other chain) and ss-beads as the beads without the pairing FENE bond (see Fig 3) A native segment (segment made up of two rows of ds-beads) will be referred to as a ds-subchain and similarly a ss-subchain will refer to a chain made up of ss-beads (see Fig 4)

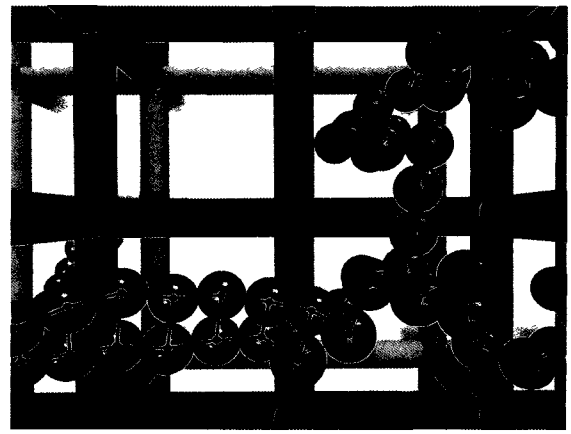


Figure 3: A simulation snapshot displaying the FENE bond structure for a fragment with a denatured region. The denaturation is imposed by permanently removing the pairing bonds from a group of ss-beads (colored blue). The ds-beads (orange) make up sections of the fragment which are in a native state. The gel is modeled by stiff WCA poles arranged in a periodic cubic lattice.

Due to the geometrical constraints arising from being pair-bonded to another chain, we expect the persistence length of the ds-subchain in free solution to be greater than that of the ss sub-chain. Our coarse-grain model does not explicitly capture the stiffness contribution arising from the cooperatively between neighboring base-pairs (stacking interactions). For this reason, an additional bending potential is needed for the ds-subchain [19]:

$$U_{\text{Bend}}(\phi) = k_{\text{Bend}} \left[1 - \cos(\phi - \phi_0)\right], \tag{5}$$

where ϕ is the angle between by three consecutive beads on a ds sub-chain, ϕ_0 is the equilibrium angle (chosen to be 0). The value for the stiffness constant k_{Bend} was obtained by simulating a completely native dsDNA fragment in free solution and calculating the persistence length L_{p} using the standard definition:

$$e^{-n/L_{\text{p}}} = \langle \hat{v}_i \cdot \hat{v}_{i+n} \rangle \tag{6}$$

where the unit tangent vector on bead i is $\hat{v}_i$. Bead separations of $n = [0, N - i]$ were

considered No bending potential is attributed to the ss sub-chains (we verified and found the expected value of $L_p^{ss} = 1\ 0$ from the simulation of a single pure ssDNA fragment) Different values of k_{Bend} were tested until we obtained ratio which corresponds with experimental observations of $3 - 7$ nm for ssDNA and 50 nm for dsDNA [17] We augmented the initial ($L_p^{ds} = 2\ 2$) free solution value to $L_p^{ds} = 7\ 6$ with a stiffness constant of $k_{\text{Bend}} = 3$ We thus obtain a ratio $L_p^{ss}/L_p^{ds} = 7\ 6$ which corresponds with the low-end of the experimental observations (50nm/7nm = 7 1)

The gel is modeled as a cubic network of rigid and immobile WCA-fibres (see Fig 3) These fibres are arranged to form a cubic lattice and extend to the extrema of the periodic simulation box The gel fibres have an effective diameter of σ and lattice spacing 5σ For this spacing long-range hydrodynamic interactions are assumed to be screened by the fibres and are thus not considered It is also assumed that the electric field lines are straight throughout the medium such that the effect of the field can be taken as a constant force applied on all beads For simplicity we will sometimes refer to this force as the field itself and it is given in units of $\frac{\epsilon}{\sigma}$ We have chosen a field which points in the (1 1 1) direction (with respect to the gel lattice) to reduce alignment effects of the fragment in the gel hallways For these results the field was chosen to be $F_i^{\text{ext}} = 0\ 01$ in units of $\frac{\epsilon}{\sigma}$ along each axes, the total magnitude is thus $F^{\text{ext}} = \sqrt{3}(0\ 01)$ This field intensity was chosen such that the migration of the native molecule was in the reptation without orientation regime In other words the field intensity was too low to either orient the reptating molecule or force the molecule to migrate via geometration [18] A study of the effect of high field intensities will be presented in a separate paper

Two separate types of domains are examined The first shown in Fig 4(a) considers denatured regions which are placed at the ends of the fragment This case reflects what typically happens when a fragment begins to denature since there is an added entropic gain when the denaturation occurs at the ends The other case considered shown in Fig 4(b) has the melted region inside of the fragment In all the simulations 100 bead long DNA fragments (with a total number of beads $N_{\text{Tot}} = 200$) are simulated The degree of denaturation defined as the number of ss-beads over the total number of beads ($N_{\text{ss}}/N_{\text{Tot}}$) varies from 0% (native fragment) to 48%

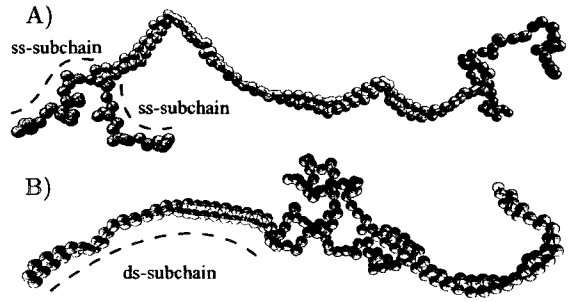


Figure 4: Two distinct denaturation profiles are considered; (a) when the denaturation occurs at the ends of the fragment and (b) when it is occurs at the middle. In all cases, the denatured domains are placed in a symmetrical fashion. In these two sample simulation snapshots $N_{ss} = 88$ and $N_{ds} = 112$ corresponding to 44 removed FENE pairing bonds and 56 pairs of beads with the FENE bond.

The fragments are first chosen to be symmetrical to remove complications arising from considering a different number of branches. For example, let us consider a fragment with a single bubble with $N_{ss} = 20$ denatured beads. At best, this can form two 5-bead long lateral branches. To compare with the split-end case, we denature both sides of the same native fragment by 10 beads, thus as before, N_{ss}=20 denatured beads. This case will also give rise to two branches with a maximum contour length of 5σ. Thus any observable difference between these cases will be from the topology.

The DNA model is integrated in time using the velocity-Verlet algorithm implemented by the ESPResSo simulation package [19] with an integration timestep of $\tau = 0.01$. Although denaturation is increased by removing bonds, the temperature is kept constant $\epsilon = k_B T = 1$. Coulombic interactions are considered to be screened at the relevant length scales and are thus not considered.

3 Results and discussion

3.1 Pure split ends *vs.* pure bubble

A plot of the scaled electrophoretic mobility μ/μ_0 vs the number of denatured beads is shown in Fig 5 From inspection it is clear that there is an important difference between the two types of molecules studied here In both cases the mobility decreases as the amount of denaturation increases but it clearly occurs at different rates The case where the denaturation is placed at the ends of the fragment shows a steep decrease similar to an exponential drop If p is taken as N_{ss}^{end} a fit with the form of Eq 1 can be used to describe the data An effective value of $L_r^{end} = 26 \pm 1$ is obtained However it is not evident that an exponential decease is observed for the bubble case Nevertheless a fit with an exponential form can be conducted and a value of $L_r^{bub} = 103 \pm 6$ is obtained

We have recently shown [20] that the denaturation variable $\bar{p}$ the total number of bases in the denatured state is actually related to the number of pores occupied by the branches that form during denaturation Our results shown in Fig 5 suggest that the *effective* number of denatured beads which contributes to a significant reduction in the mobility depends on the topology of the partially melted DNA Our results show that the mobility of the centered bubble case decays with a characteristic number of denatured beads of about four times the case with symmetrical split-ends The ratio $w = \frac{L_r^{end}}{L_r^{bub}} \approx 0\ 25$ could be used as a first order approximation of how the different denatured bases should be weighted when estimating the effective value of $\bar{p}$ For instance one could redefine the denaturation variable as a weighted sum

$$\bar{p}_w = w N_{ss}^{bub} + N_{ss}^{end} \tag{7}$$

Note that this is essentially equivalent to using

$$\mu = \mu_0 \exp - \left(\frac{N_{ss}^{bub}}{L_r^{bub}} + \frac{N_{ss}^{end}}{L_r^{end}} \right) \tag{8}$$

This first-order linear combination assumes that 1) the effect of multiple arms can simply be summed and 2) the mobility decrease due to bubbles is also exponential

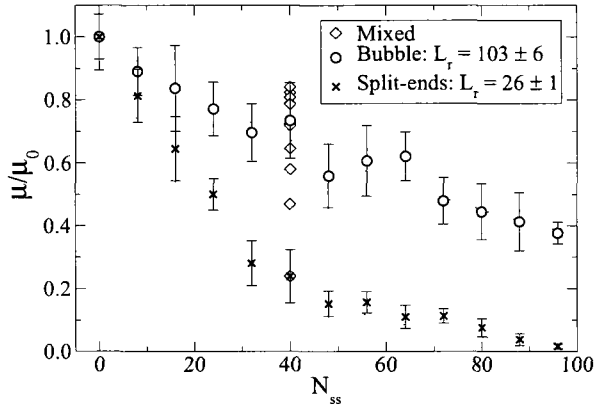


Figure 5: Reduced mobility vs degree of denaturation N_{ss} for same length molecules ($N_{\text{tot}} = 200$) with two identical split-ends (○) or one centered bubble (×). A value for the characteristic length L_{r} can be determined by fitting the data with an exponential function. as described in Eq. 1. The '◇' show simulations results with different combinations of both types of denaturations.

3.2 Counting occupied pores

The physical interpretation we proposed for L_{r} [20] suggested that this parameter is related to the number of beads per pore. Using the obtained simulation data. we can test whether there are in fact more beads per pore (thus less occupied pores for a given number of denatured beads) for the bubble case. The results are shown in Fig. 6.

The volume which constitues a given pore in a random gel is usualy difficult to define. However. this problem is trivial for a gel modeled as a cubic lattice of poles. as is the case here. The average number of pores occupied by the ss-beads can thus be simply counted by binning the simulation space (of course the bin borders are chosen to be aligned with the gel poles). Since the occurrence of a single ss-bead in an otherwise empty pore will

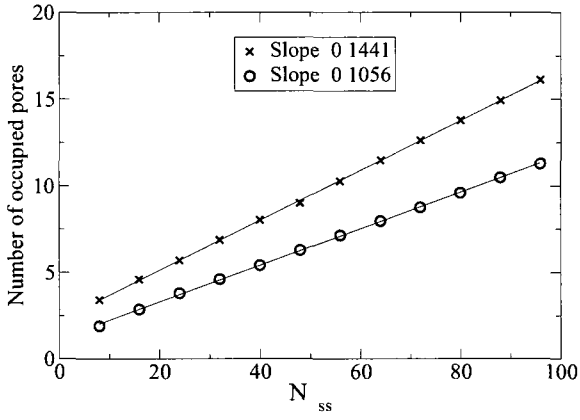


Figure 6 The average number of pores occupied by the ss-beads *vs* the number of ss-beads

skew the resulting *number of occupied pores* we chose to use an average weighted by the pore s population (similar to the concept of weight-averaged molecular weight in polymer chemistry) Our method can be explained as follows first we consider all the pores which contain ss-beads At this point we could choose to eliminate the pores which have a low bead count (using some cutoff) but instead we calculated the average pore population (weighted by their population) This value was used as the mean number density of ss beads in occupied pores To obtain the mean number occupied pores (the y-axis of Fig 6) we simply divided N_{ss} by the mean number density of ss beads in occupied pores

We can see in Fig 6 that a fragment with a bubble occupies fewer pores than a fragment with split ends This is not a surprise since segments forming a hernia need to 'come back in the same pores This means that for the case of a bubble there is a higher concentration of ss-beads in the occupied pores By this argument if there are hernias which form effective branches it is expected that for the same value of N_{ss} these branches will be shorter than the branches occurring from the split ends By taking the ratios obtained from the analysis

of Fig 6 we see that Eq 2 would predict $\frac{L_r^{end}}{L_r^{bub}} \approx 0.73$ This means that for the same N_{ss} branches formed by bubbles are about 73% shorter than branches formed by split ends This of course is a best case scenario where all of the ss-beads are part of a lateral hernia we actually expect that parts of the denatured domain close to the forking points will not play a role in the hernia

This value for w numerically differs from what we obtained from the mobility measurements but the qualitative results are the same bubbles should matter less The discrepancies can be attributed to the fact that the *average number of occupied pores* is not a good proxy for the length of a branch The former does not consider that one of the split arms (for the split-end case) or that ss-beads which are not in a hernia (for the bubble case) must be considered as the continuation of the main backbone

3.3 Hybrid case: centered bubble with symmetrical split-ends

A fragment with a bubble and the two ends denatured was simulated to test if Eq 7 (or equivalently Eq 8) gives consistent results when the two cases are present simultaneously We kept $N_{ds} = 160$ and the sum $N_{ss}^{bub} + N_{ss}^{end} = 40$ constant while varying $N_{ss}^{end} = 0 \; 40$ in steps of four Thus we have $N_{Tot} = 200$ as before since we are only changing the ratio between the number of beads belonging to bubbles and split-ends The scaled mobility for these cases are overlaid with the results from the unmixed denaturation types in Fig 5 As expected the fragment s mobility eventually drops to the pure split-ends case However we can see that some points they lie above the pure bubble case which means that the fragment actually migrate easier when a small fraction of the denatured beads in placed at the ends For better visualization the scaled mobility *vs* the number of ss-beads at the ends for this hybrid case is shown in Fig 7 Indeed a slight increase is observed (shoulder) before the decrease We attribute this to the incapacity of the rather small number of denatured beads at the end to act like a de Gennes primitive segment The first datapoint shown ($N_{ss}^{end} = 0$) represents a pure-bubble case The next datapoint shows the same case with four less beads in the bubble and one denatured bead at each of the four ends of the DNA fragment For this point we have effectively reduced the size of the bubble with negligible frictional impact

from the split-ends This is coherent with the de Gennes viewpoint until a full segment of the primitive chain can be formed the exponential decrease does not manifest itself The mobility eventually drops near $N_{ss}^{end} = 20$ this is when four arms have a contour length in the order of the pore size 5σ

The presence of the maximum in Fig 7 shows that a simple linear combination of the effective L_r s does not capture the complete physical picture The dotted line shows a fit with of the form

$$\frac{\mu}{\mu_0} = A \exp \frac{-\sqrt{(wN_{ss}^{bub})^2 + (N_{ss}^{end})^2}}{L_r} \tag{9}$$

with $w = 0\,66$ $L_r = 16\,9$ and $A = 3\,3$ Note that this is just an empirical fit chosen to capture the effect of the maximum A weighed geometric mean was chosen to give less weight to the small values of N_{ss}^{bub} and N_{ss}^{end} We are currently investigating a similar hybrid case with more denatured beads ($N_{ss} = 96$) to capture the effect of multiple primitive chain segments at the split ends

3.4 Asymmetrical split ends

The effect of asymmetrical split-ends is shown in Fig 8 The pure split-ends case where $N_{ss} = 20$ is investigated by varying the fraction on the left and right ends Again it can be noticed that even when solely considering the split-end case the total number of denatured beads N_{ss} is not adequate to capture the complete migration dynamics The effect on the mobility seems to be heightened when all of the denatured beads are distributed in one end of the fragment — one long arm stops the fragment more than two medium arms This observation puts into question the simultaneity requirement for the relaxation of the arms It seems that the two arms need not to retract at the same time Rather, the dynamics is dominated by the longest timescale Since the time for retraction increases exponentially with the size of the branch in the time window required for the relaxation of the longest branch all of the smaller branches were permitted to relax as well With the idea that the small branches should matter less than the long ones as before a fit with the form

$$\mu = A \exp \frac{-\sqrt{(N_{ss}^{left})^2 + (N_{ss}^{right})^2}}{L_r} \tag{10}$$

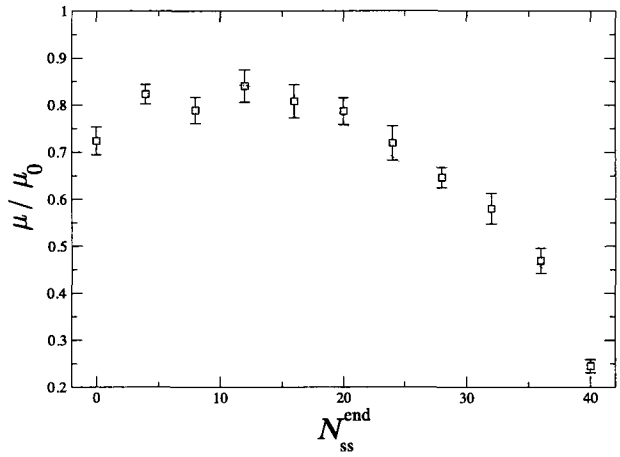


Figure 7: While holding $N_{ss} = 40$ constant. a hybrid case with a centered bubble and split-ends is considered. Going from a pure bubble case ($N_{ss}^{end} = 0$) to a pure split-ends case ($N_{ss}^{end} = 40$). the rapid decrease of mobility due to the ends is only felt when the ends are long enough to form a primitive branch segment.

is shown in Fig. 8 with $L_r = 10.3$ and $A = 0.83$. Again. these results support the claim that perhaps the requirement of simultaneity can be questioned. The dynamics seem to be dominated by the length of the longest arm.

4 Conclusion

We used Langevin Dynamics computer simulations to test whether the Lerman approximation for the mobility of partially denatured DNA needs to take into account the location of the denatured regions. Four distinct cases were studied: 1) a fragment with the denaturation at both ends. 2) a fragment where the denaturation is placed at the center. 3) a hybrid of both of these cases (for $N_{ss} = 40$) from a pure centered bubble to a pure split-end case, and 4) a split-end case where the beads were distributed in an asymmetrical fashion.

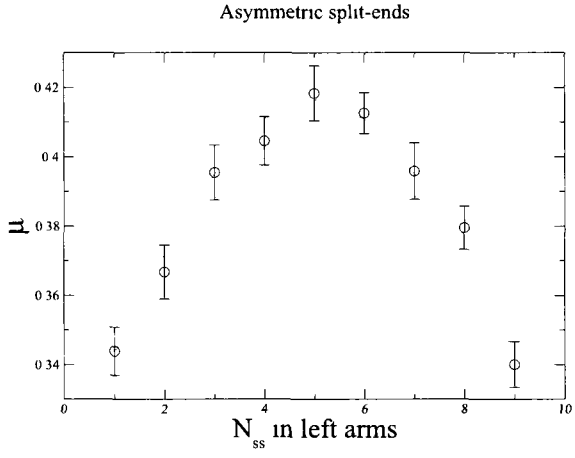


Figure 8 Keeping the total number of denatured beads constant at $N_{ss} = 20$ the distribution of denatured beads is varied from the left end to the right end Here we can see that the mobility is more sensitive to one denatured end with $N_{ss} = 10$ on two arms than denaturation at both ends with $N_{ss} = 5$ beads on each of the four arms

For the pure bubble and pure split-ends case it was found that a melted region has a higher impact on the mobility when it is positioned at the ends of the fragment Assuming that both of these cases give rise to an exponential decrease we suggest that two characteristic L_r values be used in the Lerman approximation We found that the split-end case contribute more to the blocking by a factor of roughly four Thus for a random sequence it is believed that the blocking is mostly dominated by the denaturation at the ends

If the two denaturation topologies are present in the fragment different weights (depending on which case the domain belongs to) can be assigned to the bases in the sum $p_w = wN_{ss}^{bub} + N_{ss}^{end}$ Our simulations yield a weighting factor of $w \approx 0\ 25$ However when this case was simulated our simulations revealed that short-arm effects have to be taken into account When a branch is too short to fill at least one pore it does not play a significant role in the reduction of mobility This finding is coherent with the de Gennes picture too

short a branch cannot really be seen as a primitive chain segment We suggested an empirical form for summing the two denaturation domains $\sqrt{(wN_{ss}^{bub})^2 + (N_{ss}^{end})^2}$ which can partly take this effect into account Simulations of higher degrees of denaturation are currently being conducted in order to test if this form holds when many primitive segments are present When we investigated the effect of asymmetrical split ends we found that a similar form can capture the reduction of mobility $\sqrt{(N_{ss}^{left})^2 + (N_{ss}^{right})^2}$ Again this form was chosen to give more weight to longer branches

The occurrence of bubbles can lead to an overestimation of either L_r or the final blocking position in the gel unless a different denaturation measurement is used By generalising our findings a denaturation measurement could perhaps have the form

$$\frac{\mu}{\mu_0} = \exp - \frac{\sqrt{(N_{ss}^{left})^2 + (wN_{ss}^{bub})^2 + (N_{ss}^{right})^2}}{L_r} \tag{11}$$

We have not tested a case which presents assymetry of the split ends with a centered bubble nor have we yet investigated the effect of off-center or multiple bubbles — we have carefully chosen the cases studied for their simplicity

Since bubbles will usually not form a single branch-like' hernia but rather have a complex structure with multiple leaves' it is difficult to evaluate the effective branch length from the simulation data Using as a naive proxy for the branch length the average number of occupied pores we can correct for the effect of the compacted ss-beads Even accounting for the difference in effective branch length important differences between the mobility of split-ends vs bubbles is observed These results indicate that an assessment of the total number of denatured basepairs is insufficient for a detailed prediction of the resulting mobility Rather our results indicate that details concerning the topology of the partially denatured DNA fragment are crucial to obtaining a complete physical picture and prediction of the sequence-dependent mobility In the derivation of Eq 1 the prefactors arising from the calculation of the relaxation of the ssDNA chain [3 5 20] are neglected It is possible that the observed differences between the bubble and split-end cases can be attributed to these factors

To our knowledge there is no experimental study that offers conclusive insight with regards to the effect of bubbles and split-ends on a fragment s mobility The parameter

$p(T)$ suggested by Lerman et al is a simple unweighted sum over all basepairs It does not account for the position of the melted domains nor the size of the individual domains We believe that the melted domain with the longest effective branch should be considered as the dominant blocking source For entropic reasons this will mostly be the case for domains at the end of a fragment Fortunately the Poland algorithm [4] gives the formulae to calculate the average length of the left and right denatured ends However the final formulae given in Equation (34) of reference [4] seems to have typographical errors We present the corrected formulae in the Appendix

5 Acknowledgements

All the simulations were performed using the ESPResSo [19] simulation package and were visualised using VMD [21] D S would like to thank Hendrick de Haan Mykyta Chubynsky and Tyler Shendruk for the insightful discussions and suggestions The work was supported by a Natural Science and Engineering Research Council (NSERC) of Canada Discovery Grant to GWS The present authors would like to thank Owen Hickey and Hendrick de Haan for the critical reading of the manuscript

A Corrected Poland Formulae

The Poland-Sheraga algorithm treats the DNA fragment like an Ising-like chain Each basepair can be in one of two states either native (1) or denatured (0) Ignoring total strand disassociation Poland first describes the net average of the number of denatured basespairs on the left $\langle N_0 \rangle_{\mathrm{L}}$ and right end $\langle N_0 \rangle_{\mathrm{R}}$ of the fragment as

$$\langle N_0 \rangle_{\mathrm{L}} = \sum_{m=1}^{N-1} m p_m(01) \tag{12}$$

$$\langle N_0 \rangle_{\mathrm{R}} = \sum_{m=1}^{N-1} (N-m) p_m(10) \tag{13}$$

where m is the basepair index N is the total number of basespairs and the probabilities $p_m(01)$ and $p_m(10)$ are the probabilities of having a string of m denatured bases on the left and right (respectively) of a native basepair This is basically a thermodynamic average weighted by the length of the denatured string

Poland later presents a convenient change of variables, here we only repeat the ones pertinent to Eq 12 and Eq 13

$$p_m(01) = p(1_1)\gamma_m \qquad (m = 1\ N-1) \tag{14}$$

$$p_m(10) = p(1_m)\beta_m \qquad (m = 1\ N-2) \tag{15}$$

$$p_{N\ 1}(10) = p(1_{N-1})\frac{r_N}{1+r_N} \tag{16}$$

where $p(1_m)$ is the probability that the m^{th} basepair is in the native (1) state The original paper by Poland [4] should be consulted for more details regarding the variables β_m γ_m and r_N If we take Eq 12 together with the change of variables defined in Eq 14 we obtain

$$\langle N_0\rangle_{\mathrm{L}} = p(1_1)\sum_{m=1}^{N-1} m\gamma_m \tag{17}$$

The same can be done using Eq 15 and Eq 16 into Eq 13

$$\langle N_0\rangle_{\mathrm{R}} = p(1_{N-1})\frac{r_N}{(1+r_N)} + \sum_{m=1}^{N-2}(N-m)p(1_m)\beta_m \tag{18}$$

These expressions differ from those in ref [4] For the details concerning the recursion relations and the algorithm please refer to the original paper by Poland [4] We recommend that the approximation suggested by Fixman and Freire [22] be considered for the actual implementation of the algorithm

References

[1] Fischer S Leiman L *Cell* 1979 *16* 191 200

[2] Lerman L Frisch H *Peptide Science* 1982 *21* 995 997

[3] de Gennes P *J Phys* 1975 *36* 1199 1203

[4] Poland D *Biopolymers* 1974 *13* 1859–1871

[5] Lerman L Fischer S Hurley I Silverstein K Lumelsky N , *Annual review of biophysics and bioengineering* 1984 *13* 399–423

[6] Myers R Fischer S Lerman L Maniatis T *Nucleic Acids Res* 1985 *13* 3131–3145

[7] Myers R Fischer S Maniatis T Lerman L *Nucleic Acids Res* 1985 *13* 3111–3129

[8] Riesner D , Steger G Zimmat R Owens R *Electrophoresis* 1989 *10* 377–389

[9] Steger G *Nucleic Acids Res* 1994 *22* 2760–2768

[10] Ke S Wartell R *Biochemistry* 1995 *34* 4593–4600

[11] Wartell R , Hossein, S Powell S Zhu J *J Chromatogr A* 1998 *806* 169 185

[12] Zhu J Wartell R *Biochemistry* 1997 *36* 15326–15335

[13] Mercier J Kingsbury, C Slater, G Lafay B *Electrophoresis* 2008 *29* 1264–1272

[14] Stellwagen, N C Stellwagen E *J Chromatogr A* 2009 *1216* 1917–1929

[15] Weeks, J D Chandler D Andersen H C *J Chem Phys* 1971 *54* 5237–5247

[16] Grest G Kremer K , *Phys Rev A* 1986 *33* 3628–3631

[17] Stellwagen E Lu Y , Stellwagen N *Biochemistry* 2003 *42* 11745–11750

[18] Viovy, J , *Rev Mod Phys* 2000, *72*, 813–872

[19] Limbach H -J , Arnold, A Mann B A Holm C , *Comput Phys Commun* 2006 *174*, 704–727

[20] Sean D Slater G W , *Submitted to Electrophoresis* 2010 *elps 201000239*

[21] Humphrey W Dalke A Schulten K *Journal of Molecular Graphics* 1996, *14* 33–38

[22] Fixman M Freire, J *Biopolymers* 1977 *16* 2693–2704

Chapter4

The dominant conformation of high field simulations of DNA partially denatured at the ends

D Sean, GW Slater. To be shortly submitted

The dominant conformation of high field simulations of DNA partially denatured at the ends

David Sean, Gary W. Slater

Department of Physics, University of Ottawa,
150 Louis-Pasteur, Ottawa, Ontario, K1N 6N5, Canada

September 8, 2010

Corresponding author:
Prof Gary W Slater
email gary slater@uOttawa ca
fax +1-613-562-5190

Keywords:

Langevin Dynamics simulations, Computational modeling, DNA denaturation, Gel electrophoresis

Running title:

The dominant conformation of high field simulations of DNA partially denatured at the ends

Abbreviations:

FENE Finitely extensible nonlinear elastic

WCA Weeks-Chandler-Andersen

Abstract

Gel electrophoresis of a partially denatured DNA fragment is simulated using Langevin Dynamics In our simulations the denaturation was placed at the ends of the fragment in a symmetrical fashion A squid-like conformation was found to sometimes cause the fragment to completely block in the gel This conformation is the principal cause of the steep reduction in mobility observed in the simulations As the field is increased it is found that the occurrence of this conformation dominates the migration dynamics Although the squid-conformation seems to be more stable at high fields the field can eventualy force the fragments to thread through the gel pores regardless We qualitatively explore the behavior of this squid-like conformation across a range of fields and denaturations

1 Introduction

Gel electrophoresis of partially denatured DNA can be used for sequence-specific separations of the fragments. Myers *et al.* demonstrated that the majority ($\approx$ 96%) of single base-pair substitutions can be detected by gel electrophoresis of DNA in a gel with a chemical denaturant gradient [1]. This approach exploits the fact that the mobility of a DNA fragment is considerably reduced when partial denaturation is introduced. In spite of the success of this method, to our knowledge, the physical mechanisms behind this phenomenon have never been carefully examined. Lerman and Frisch [2] suggested that the reptation model of branched polymers developed by de Gennes [3] can be used to predict the electrophoretic mobility of a partially denatured DNA fragment. However, high field intensities are known to affect molecular conformations during gel electrophoresis [4], therefore, it is unlikely that the de Gennes zero-field model can apply in such cases.

Considering the trapping of circular DNA in agarose gels, a physical mechanism for trapping is suggested by the impalement of the strands on the gel fibres [5, 6]. In this picture, high field values would reduce the number of de-trapping events by stabalizing the impalements. For linear DNA fragments however, it is unclear what mechanism could be responsible for the reduction of mobility at high fields, since linear DNA fragments with denaturation only at the ends cannot be trapped by the impalement mechanism. Furthermore, in polyacrylamide gels, the fiber ends are not thought to be stiff and long enough for this to occur.

To investigate this, we simulate a coarse-grained DNA model using Langevin dynamics. This model has previously been employed to explore the effects of a denatured region forming a bubble in the fragment, *vs.* a denaturation region occurring at the ends. In that work, we found that the denatured regions placed at the ends of the fragment had a much greater impact on the mobility than the denatured regions at the middle (see Chapter 3). In light of these results and noting that DNA fragments in the lab will typically first denature at the ends (in the absence of a pair CG clamps), we assume that the blocking can completely be attributed to the state of the bases placed at ends of the fragment. Hence, for this work, we focus solely on the effect of denatured segments at the ends of the fragment.

2 Model

In our bead-spring model excluded volume interactions are modeled by the Weeks-Chandler-Andersen (WCA) potential [7] It has the form of a truncated and shifted Lennard-Jones potential

$$U_{\mathrm{WCA}}(r) = \begin{cases} 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] + \epsilon & \text{for } r < r_{\mathrm{c}} \\ 0 & \text{for } r \geq r_{\mathrm{c}} \end{cases} \tag{1}$$

where r is the distance between two beads $r_{\mathrm{c}} = 2^{\frac{1}{6}}\sigma$ is the cutoff at which the potential is truncated ϵ is the well depth of the corresponding Lennard-Jones potential— it will be used as our fundamental unit of energy— and σ as our unit of length is the nominal bead diameter

A coarse-grain ssDNA chain is modeled using WCA beads bonded by finitely-extensible-nonlinear-elastic (FENE) springs with the Kremer-Grest parameters to prevent bond crossing [8] Two such chains are then paired side-by-side using another series of identical FENE springs to create a native dsDNA chain Since the FENE springs are unbreakable we are able to construct a static denaturation profile To do so the pairing FENE springs can be removed from a native fragment to impose the needed profile Note that this implies that re-hybridization of the ssDNA with the complementary strand or with other beads along its own chain is not considered For simplicity we only consider a DNA chain with has a native length of 100σ and thus the total number of simulation beads remain constant at $N_{\mathrm{Tot}} = 200$ Denaturation is introduced by deleting FENE bonds symmetrically from both ends of the fragment (see Fig 1) The total number of denatured beads is given by N_{ss} In other words for any denaturation profile (beside the trivial native case) there is a DNA fragment with 2 single ssDNA-chains of length $N_{\mathrm{ss}}/4$ at each end and a dsDNA sub-chain of length $N_{\mathrm{Tot}} - N_{\mathrm{ss}}$ at the center Although we do observe a higher rigidity of the ds sub chain due to the geometrical arrangements of the beads an additional bending potential is needed to acquire experimentally observed persistence length ratios between a dsDNA chain and an ssDNA chain (see Chapter 3 for more details)

For computational efficiency the gel fibres are modeled as immovable stiff poles of

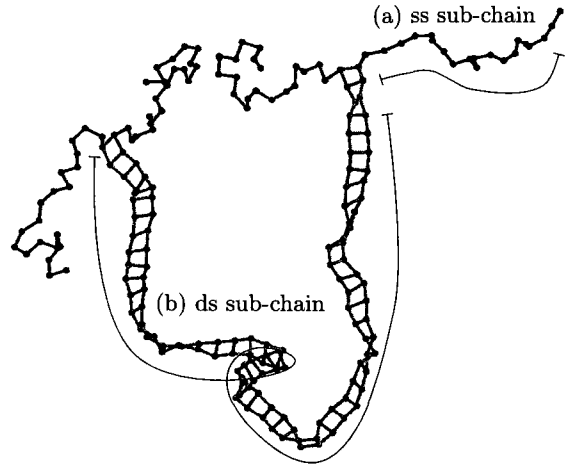


Figure 1: The bead-chain framework for our DNA model. All the bonds are modeld by FENE springs and all the beads have a repulsive WCA potential. We shall refer to a continuous run of beads which have had the pairing FENE spring removed as a ss sub-chain. Similarly, a continuous run of beads which have a pairing FENE spring will be called a ds sub-chain. In this paper, all of the denaturation occur at the ends. For this snapshot, $N_{ss} = 80$. Thus there are four ss sub-chains of 20 beads each, and a 60 bead long ds sub-chain of 120 beads. The gel has been removed for visualization.

nominal diameter σ defined by the repulsive WCA potential These fibres are arranged in a cage like cubic lattice such that there are no gel ends which stick out and thus the pore size is mono-dispersed For this study the gel lattice spacing is set to 5σ At this gel spacing long-range hydrodynamic interactions between beads are assumed to be screened by the gel fibers and thus are not explicitly calculated (pure Langevin Dynamics neglects these interactions)

Since we suspect that the DNA fragment will align itself in the hallways of the cubic gel lattice the external driving force was chosen to point in the 1 1 1 direction To model the electric field a constant force is simply applied on all the beads While this is an efficient approach it does assume that the field lines are not altered by the gel fibres (or the DNA fragment itself) and that the effective charge on every bead is the same as electrostatic interactions between the beads are assumed to be partially screened by the presence of counter-ions (which are not explicitly simulated) In this study we will refer to the electric field as the applied force along one of the axes $\vec{F}^{\text{ext}}$ (with $i = x\ y\ z$) in scaled units of $\frac{\epsilon}{\sigma}$ All of the simulations are performed at a constant temperature $k_{\text{B}}T = \epsilon$ The time axis in all the plots are in units of simulation frames where a frame is written every 250 integration time steps of size $\Delta\tau = 0\ 01\sqrt{\frac{m\sigma^2}{\epsilon}}$ where m is the mass of a bead

Since the total number of beads is $N_{\text{Tot}} = 200$ and the denaturation is symmetrically placed at the ends the denaturation profile can be deduced from the single parameter N_{ss} In this study we vary $N_{\text{ss}} = 0\ \ 96$ in steps of four At each N_{ss} the fragment is simulated across a range of electric field intensities $\vec{F}^{\text{ext}} = 0\ 01 - 0\ 04$ To determine if the range $\vec{F}^{\text{ext}} = 0\ 01 - 0\ 04$ crosses from the *low* to the *high* field regime of reptation theory we can compare this with the theoretical condition for a weak external force [4] of $|\frac{Q\vec{E}a}{k_{\text{B}}T}| < 1$ where Q is the charge per pore $\vec{E}$ is the electric field and a is the mean pore radius Here the value of $|Q\vec{E}|$ is simply $\sqrt{3}|\vec{F}^{\text{ext}}|$ multiplied by the number of beads per pore Using our lattice model of a gel we can determine the average population of beads in the occupied pores via the *weighted-average number of ds-beads per occupied pore* as discussed in our recent article [9] From this analysis there are roughly 10 beads in an occupied pore for a dsDNA fragment with a random walk conformation If we consider the pore size a as $\frac{5}{2}\sigma$ we obtain $|\frac{Q\vec{E}a}{k_{\text{B}}T}| = 25|\vec{F}^{\text{ext}}|$ We thus obtain a range of $0\ 433 - 1\ 73$ indicating that the $\vec{F}^{\text{ext}} = 0\ 01$

limit is *borderline* low field with $\vec{F}^{\text{ext}} = 0.04$ crossing over to a *high* field regime

To verify these regimes a qualitative assessment using these field values was conducted By manually visualising the simulations of a native DNA fragment for these two limiting cases the reptation picture was observed to be accurate for $\vec{F}^{\text{ext}} = 0.01$ Conversely for $\vec{F}^{\text{ext}} = 0.04$ hernias were observed to occur often As the occurrence of hernias is a signature behavior of the high field limit [4] the cross-over between these two regimes was validated

3 Results and Discussion

For all the plots the center-of mass vector for the DNA fragment is rotated to point in the direction of the field A plot of the center-of mass position *vs* time for a sample split-end denaturation profile ($N_{\text{ss}} = 80$) is shown in Fig 2 For this data the DNA fragment has a denaturation characterized by $N_{\text{ss}} = 80$ and results from two different field values of $\vec{F}^{\text{ext}} = 0.01$ and $\vec{F}^{\text{ext}} = 0.02$ are shown The low field (bottom curve) shows the fragment smoothly reptating in the direction of the field as time progresses forward A simulation snapshot of a typical conformation is shown on the curve (This is the field intensity that we used in Chapter 3) However when the $\vec{F}^{\text{ext}} = 0.02$ curve is considered (top curve) a radical change in migration behaviour is observed The dominant features of this plot are the long plateaus where the DNA fragment becomes trapped and does not migrate in the direction of the field Between the trapping events a rapid (steep slope) migration in the direction of the field is observed The inset (a) of Fig 2 shows a simulation snapshot taken when the fragment is in one of the plateaus It seems that the fragment becomes trapped by adopting this type of squid-like conformation The fragment is observed to reptate with a conformation similar to the inset (b) of Fig 2 between the trapping events This result is similar to a plot in a previous study [10] where a pulley-like conformation is responsible for the plateaus In this case the conformations are very similar but the U shape points in the opposite direction! The fragment completely bends in half with the ss sub-chains *up* the field and the hernia tip *down* the field

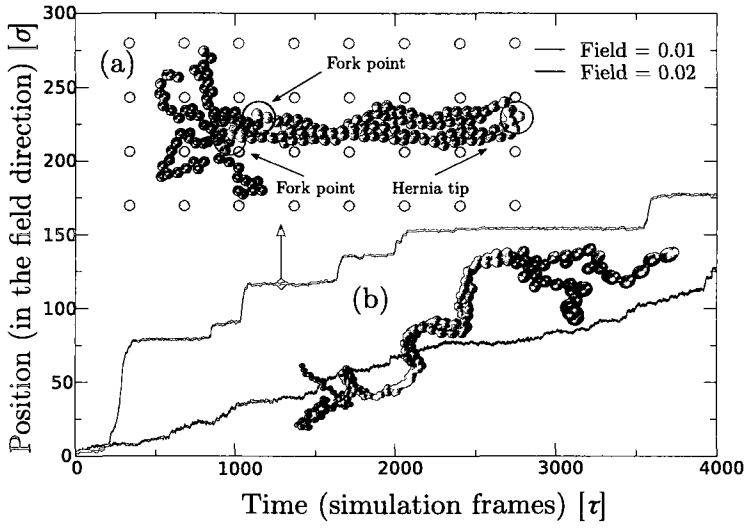


Figure 2: A plot of the center of mass position of the fragment *vs.* time (in units of frame numbers). When $\vec{F}^{\text{ext}} = 0.01$ (bottom curve) the fragment smoothly migrates with a fairly constant velocity (slope). When the field is increased to $\vec{F}^{\text{ext}} = 0.02$ (top curve), the time plot shows a step-like migration pattern. The fragment sometimes become stuck in the gel (long plateaus) for some time. Between these trapped states are short-lived events where the fragment migrates with high velocity (steep slopes). A simulation snapshot taken in the plateaus and steep-slope regions are shown in insets (a) and (b) respectively. The latter is also representative of the DNA's conformation when the field intensity $\vec{F}^{\text{ext}} = 0.01$. For this data, the denaturation profile is characterized by $N_{\text{ss}} = 80$.

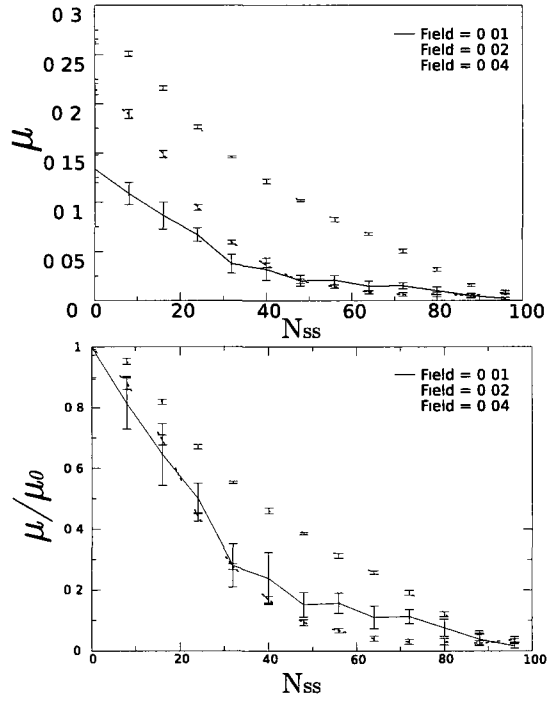


Figure 3 Plot of the scaled mobility *vs* the percent denatured (a) For each field the mobility is rescaled with the different μ_0 (b) The same curves without rescaling

For the physical picture of this blocking mechanism we can consider this process as a combination of steric and entropic trapping where the ss sub-chains (or the squid *tentacles*) *resist* threading through the pores That is there is an increased drag on the DNA fragment due to the ends of the chain The resulting conformation thus has the ss sub-chains dragging behind with the ds sub-chain portion bent almost in half yielding an overall symmetric conformation The hernia causes a reduction of the size of the pore opening due to the the presence of additional monomers

The interplay between the entropic and steric cost of threading through a pore *vs* the energetic gain by going down the field can be investigated by simulating different lengths of a ss-subchain across different fields The net mobility of a fragment can be calculated from the total simulated change in position over the total simulation time divided by the magnitude of the field

The resulting mobilities are shown in Fig 3 The calculated mobility μ is scaled by the mobility of the native fragment μ_0 for Fig 3 (a) Note that the value of μ_0 varies for different fields for the fields of $\vec{F}^{\text{ext}} = 0\ 01$ $\vec{F}^{\text{ext}} = 0\ 02$ and $\vec{F}^{\text{ext}} = 0\ 04$ the native mobility is respectively $\mu_0 = 2\ 9 \pm 0\ 1$ $\mu_0 = 10\ 5 \pm 0\ 1$ and $\mu_0 = 25\ 5 \pm 0\ 1$ In Fig 3 note the steep decrease as denaturation is introduced The expected exponential decrease is obtained for the low field of $\vec{F}^{\text{ext}} = 0\ 01$ (as seen in Chapter 3) As the field is increased the exponential decrease seems to become more like a sigmoidal curve with an inflection point

To determine if the squid conformation is responsible for the abrupt mobility decrease shown in Fig 3 discrimination between the squid conformation and a reptating conformation needs to be made By visualisation of the simulations we found that the fork-to-fork distance (distance between the end points of the ds-subchain shown in Fig 2) and the average number of native beads per pore are good independent indications of the squid conformation Although we do not show the results here we also found that using the cross-sectional area of the two secondary axes of the radius of gyration is also a good measure of the squid conformation

As a measure of the number of ds-beads in occupied pores the *weighted-average number of ds-beads per occupied pore* is again used (see Chapter 3) Due to the increased bead

population in a hernia this number should be increased roughly two-fold between the squid conformation and a reptating fragment

To verify if this measurement discriminates between reptation and a squid conformation the position *vs* time plots are color coded with red points when the number of ds-beads per pore surpasses a chosen cutoff of 13 The points are otherwise shown in black

The step-like curve in Fig 2 is re-plotted using this color-scheme Since the visualisation of the complete simulation validated the assumption that the squid was responsible for the plateaus we expect that the measurement will yield red plateaus Conversely we need to make sure that this measure does not give positive results where the squid conformation is not due As a test we can similarly use the bottom curve of Fig 2 since reptation was observed throughout the migration

The resulting two curves of Fig 2 are shown in Fig 4 (a) and Fig 4 (b) From these we can validate the squid measurement— indeed the squid is responsible for all the regions of low mobility and is not formed when the fragment reptates

The squid conformation is omnipresent when same denaturation profile is considered for yet a higher field shown in Fig 4 (c) It becomes nearly impossible to undo the hernia against the field thus the squid conformation is very stable Moreover visual inspection of the simulations reveal that the ss sub-chains are hooked around the gel fibres and point in the direction of the field This can be seen as an exotic version of the famous pulley

When the low denaturations are considered the low field case yields again a reptation regime but this time Fig 4 (d) the ss sub-chains are not long enough to quench reptation If we raise the field for this denaturation shown in Fig 4 (e) the reptation picture is completely lost to the squid But this time the squid is permitted to thread through the pores since the ss sub-chains are not long enough to provide significant drag But it seems that the drag they do provide is nevertheless enough to keep the DNA fragment folded at the center

A phase diagram has been sketched to separate the different behaviours Points for the plots which we have just considered are shown in the diagram This should be seen as a

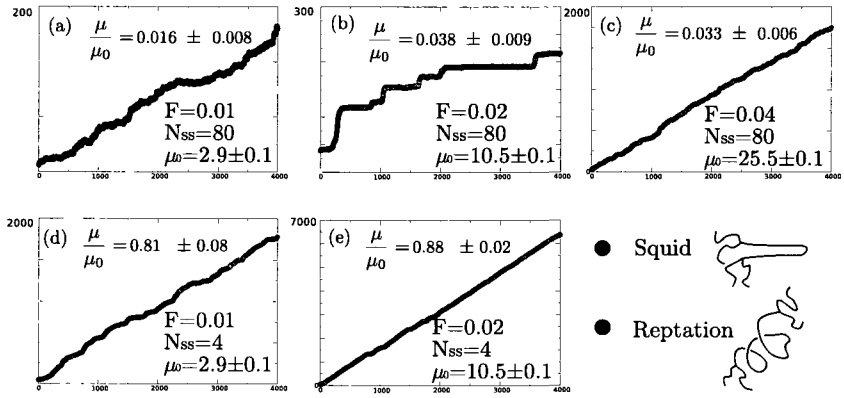


Figure 4: The *center-of-mass* position *vs.* time plots for points corresponding in Fig. 5. The curves are color-coded to indicate the instantaneous conformation: red when squid-like and black when it reptates. The value for the scaled mobility is given for proper comparison between the plots. (a) Low field and high denaturation: the fragment has a low mobility and has a dominant reptation conformation. (b) When the field is increased, the fragment mostly stays in a stuck state with a squid conformation. (c) When the field is increased even more, the fragment does not leave the squid conformation and stays stuck. (d) At a low denaturation and low fields, the fragment reptates most of the time. (e) when the field is increased, the squid conformation dominates, but it does not block migration.

rough qualitative sketch of the different phases The dotted lines do not represent definitive crossovers Points which are close to the dotted borders have behaviours which are proper to both regimes For example the step like plot of Fig 4 (b) represents an erratic oscillation between reptation and stuck squid

The phase diagram can be read as follows from left to right the crossover between the high and low field separates reptation from squids In other words we need a sufficient force to bend the dsDNA segment into two the necessary step to forming the squid From bottom to top starting at $\vec{F}^{\text{ext}} = 0\,01$ The fragment always reptates and an exponential decrease in mobility is experienced by the fragments From bottom to top starting at $\vec{F}^{\text{ext}} = 0\,02$ the mobility drops following a sigmoidal curve thus even with increasing denaturation the mobility is not significantly reduced until a critical value is obtained and then an exponential decrease follows (see middle curve of Fig 3) From bottom to top starting at $\vec{F}^{\text{ext}} = 0\,04$ the drop of mobility is again sigmoidal but this time the decrease is not as steep It thus takes a higher denaturation to reach a critical mobility decrease

At high fields the stabilizing effect of the denaturation on the DNA conformation can be investigated using the fork-to-fork distance The fork-to-fork distance is calculated from the fork-points defined in the inset of Fig 2 This is simply the end-to-end distance for the ds sub chain The simulation results from the high and low field limiting cases are shown in Fig 6 The error bars do not indicate statistical error but the standard deviation Inspection of the two native fragments $N_{\text{ss}} = 0$ confirm the existence of a crossover between the low-field (reptation) and the high field regime (geometration)

The stabilizing effect of the squid conformation can be noticed for the high field case shown in red As soon as the denaturation is introduced the squid conformation quickly dominates and the fork-points remain close to one another The last point of the low-field case shown in black has a low standard deviation because it is basically stuck in the given conformation for the entire simulation

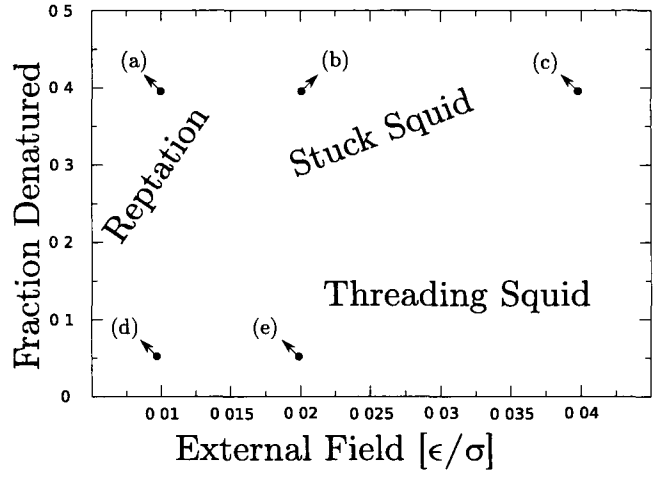


Figure 5 A free-hand phase diagram is drawn separating the fragment s dominant behaviours Depending on length of the ss sub-chains and the applied field the fragment can either have a random walk conformation or a squid In the squid case it can either stay stuck or thread through the pores Position *vs* time plots for the selected points in this diagram are shown in Fig 4

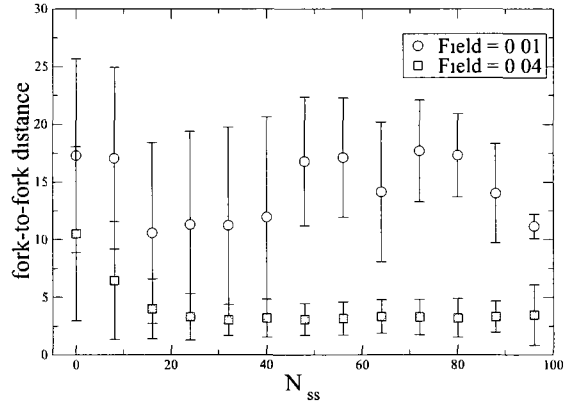


Figure 6 The average fork-to-fork distance *vs* the number of denatured beads N_{ss} Only the two limiting fields are shown here to alleviate the plot The error-bars represent the standard-deviation from the mean Notice that for the native fragments $N_{ss} = 0$ the mean end-to-end distance is lower for the high-field case This is a direct consequence of the hernias which are more common in this regime

4 Conclusion

Computer simulations across a range of denaturation profiles and field intensities which cross over from low to high fields regimes give new insight into the blocking mechanism of partially denatured DNA

For the low field case the reptation picture accurately describes the dynamics We obtain an exponential decrease in mobility as predicted by Lerman *et al*

When the field was increased we found an increase in the occurrence of hernias indicating that reptation becomes unstable versus the deformation of the chain by the electric force this is the onset of geometration [4]

When denaturation was introduced in this regime the denatured ends of the fragment seems to provide a stabilization effect on a hernia The dragging effect of the denatured ends causes the tip of the hernia to be placed at the center of the fragment

The ss sub-chains at the denatured ends are more flexible and can exhibit entropic-trapping-like behaviours by forming random coils inside the pores— even though the gel pores are mono-dispersed in our model Thus the ends provide more drag on the fragment and cause the chain for fold in half the ends at the *rear* and the hernia tip at the *front* Since the formation of a hernia is a necessary precursor to the squid conformation it is not clear whether the entropic trapping picture is accurate at these fields The steric crowding effects and the hooking interactions of the ss sub-chains need to be investigated

In some cases this conformation causes the fragment to completely stop migrating Further migration is only possible after the hernia is undone by completely pulling it upfield from the nucleation point The fragment is only allowed to migrate between trapping events

If the field is increased the fragment can be forced through the pores and it is no longer blocked However the squid conformation remains stable We present a sketch of a phase diagram for the possible states (reptation *vs* stuck-squid *vs* threading-squid) of the fragment *vs* the field intensity and the amount denatured

It is not clear whether the stuck squid is blocked from further migration due to the entropic cost of threading the ss sub-chains or due to the steric interactions of threading both denatured ends simultaneously in the pore It is possible that this is an artifact of our model since the gel is modeled as a stiff array of poles A flexible gel fiber may cede under influence of the fragment and a stuck squid is perhaps not possible Different lattice spacings is considered for future work since we think that this will minimize the steric blocking at the pore interface

5 Acknowledgements

D S would like to thank Hendrick de Haan and Tyler Shendruk for the insightful discussions and suggestions The authors would like to thank Hendrick de Haan for the critical reading of the manuscript All the simulations were performed using the ESPReSsO [11] simulation package and were visualised using VMD [12] The work was supported by a Natural Science and Engineering Research Council (NSERC) of Canada Discovery Grant to GWS

References

[1] Myers R Fischer S Lerman L Maniatis T *Nucleic Acids Res* 1985 *13* 3131 3145

[2] Lerman L Frisch H *Peptide Science* 1982 *21* 995-997

[3] de Gennes P *J Phys* 1975 *36* 1199-1203

[4] Viovy J *Rev Mod Phys* 2000 *72* 813 872

[5] Stellwagen N C Stellwagen E *Journal of Chromatography A* 2009 *1216* 1917 29

[6] Åkerman B *Biophys J* 1998 *74* 3140-3151

[7] Weeks J D Chandler D Andersen H C *J Chem Phys* 1971 *54* 5237 5247

[8] Grest G Kremer K *Physical Review A* 1986 *33*

[9] Sean D Slater G W *not submitted* 2010

[10] Noolandi J Rousseau J Slater G Turmel C *Phys Rev Let* 1987 *58* 2428-2431

[11] Limbach H J Arnold A Mann B A Holm C *Comput Phys Commun* 2006 *174* 704 727

[12] Humphrey W Dalke A Schulten K *Journal of Molecular Graphics* 1996 *14* 33 38

Chapter5

Conclusion

In this thesis, the effects of DNA denaturation on the gel electrophoretic mobility of DNA were studied. In the first chapter a short introduction to the basic concepts of polymer physics and gel electrophoresis of DNA was given. The concepts introduced are considered adequate to prepare the reader with a general physics background to the three main chapters. A review article which I co-authored, about computer simulation models for systems pertinent to my research is presented in the Appendix.

In the second Chapter we investigated the L_r parameter in the semi-empirical formula proposed by Lerman *et al.* [1]. This has been submitted as a Short Communication to the journal *Electrophoresis*. The formula proposed by Lerman *et al.* states that the DNA mobility is expected to decrease exponentially with the number of denatured basepairs:

$$\mu = \mu_0 e^{-\bar{p}(T)/L_r}, \tag{5.1}$$

where μ is the expected gel electrophoretic mobility of the partially denatured fragment, μ_0 is the gel electrophoretic mobility of the native fragment, $\bar{p}(T)$ is the thermodynamic average number of denatured basepairs as a function of the effective temperature given by T, and L_r is a fitting parameter which represents a characteristic number of denatured basepairs which reduces the mobility by a factor e^{-1}. Although this formula gives fitted results which appear to be in agreement with experimental observations, it is odd that there is no dependence on

the gel pore size or topology of the denatured domains. To my knowledge, there has been no rigorous investigations in this regard.

This formula was re-derived using the theoretical formalism introduced by the polymer physicist Pierre-Gilles de Gennes [2]. From this picture, a physical interpretation of the L_r parameter is given as *the average number of denatured basepairs per pore.* Using blob theory, the resulting formula for L_r is obtained:

$$L_r = \frac{\langle R_P \rangle^2}{\alpha L_K \ell} \tag{5.2}$$

where $\langle R_P \rangle$ is the average pore size, α is a numerical factor close to unity related to the gel, L_K is the *Kuhn-length* of the denatured ssDNA chain, and ℓ is the length of a ssDNA base. This formalism considers that only one ssDNA chain results from partial denaturation placed at one end of the DNA chain. Thus, this formula does not account for multiple denaturation domains, nor does it account for bubbles— domains which are not at the ends of the fragment.

Chapter 3 investigates the impact of the denatured region's topology on the gel electrophoretic mobility. Here, Langevin Dynamics computer simulations are performed on a model DNA fragment which is presented in detail. Two distinct cases were considered: *split-ends* which is a DNA fragment with both ends denatured *vs.* a *bubble* which is a DNA fragment with the denaturation at the center. The results indicate that there is an important difference between these cases and they should be treated differently. Since the value of $\bar{p}(T)$ pertains to the total number of denatured bases on the DNA fragment, we propose that a different measurement for denaturation be used. Melting theory can provide the necessary thermodynamic averages between these cases. We present an erratum for two of these necessary formulae in the Appendix of the presented article manuscript. We propose that a weighed linear combination between the number of bases belonging to a bubble and the number of basepairs belonging to a split-end could be used. This form is tested with simulations of cases where the DNA fragment has a combination of the two denaturation topologies. The current data shows a smooth transition between the two limiting cases, but it is yet too early to see if the linear interpolation formula as suggested is sufficient to capture the combined effect of split-ends and bubbles.

For simplicity, only symmetrical fragments were considered but future work in this direction needs to address situations where only one end is denatured — such is the case when experimental CG clamps are used — and/or the bubble is situated off-center.

Chapter 4 investigates the high-field effects only on DNA fragments with the ends denatured. A dominant conformation was found: a hernia in the ds sub-chain grows until the fragment is folded in half. Sometimes this conformation resulted in a completely blocked fragment, other times it was allowed to migrate. A phase diagram of this was produced which separates these two behaviours. High fields result in a stable squid conformation. In this regime the DNA is folded perfectly in half, and its mobility seems to exhibit a linear decrease with the amount denatured. There is perhaps a separation process which can exploit this stable novel conformation.

Future work needs to probe the effect of the rigidity of the gel fibres. It is believed that steric interactions may be partly responsible for the apparition of the squid conformation. Different gel pore sizes or the simulation of flexible gel fibers can probe this hypothesis. Furthermore, possible entropic trapping in the gel pores may occur when the fragment denatures. The more flexible ssDNA chains can form a random coil in the pores adding conformational entropy which was previously unaccessible. This effect can be probed in future work by changing the rigidity of the ssDNA chains.

Although it is the main subject of many experiments, the physical mechanisms behind the electrophoretic migration of partially denatured DNA in a gel has not received much attention. The simplistic scenarios considered in my thesis are merely a first peak into this field and the behaviours are rich and complex. They merit further theoretical investigations backed by simulations and experimental data. In most of the situations which were investigated, experimental analogues can be built to confirm the predictions — but that is another story.

5.1 Bibliography

[1] LS Lerman, SG Fischer, I Hurley, K Silverstein, N Lumelsky. *Sequence-determined dna separations*, ***Annual review of biophysics and bioengineering*** **13**, 399–423 (1984).

[2] PG de Gennes. *Scaling concepts in polymer physics.* Cornell University Press, Ithaca (1979).

AppendixA

Modeling separation processes: A review of current computer simulation methods

Gary W. Slater, Christian Holm, Mykyta V. Chubynsky, Hendrick W. de Haan, Antoine Dubé, Kai Grass, Owen A. Hickey, Christine Kingsburry, **David Sean**, Tyler N. Shendruk, Lixin Zhan. *Electrophoresis* **30**, 792-818 (2009)

Modeling separation processes: —A review of current computer simulation methods

Gary W. Slater[1], Christian Holm[2 3], Mykyta V. Chubynsky[1],
Hendrick W. de Haan[1], Antoine Dubé[1], Kai Grass[2], Owen A. Hickey[1],
Christine Kingsburry[1], David Sean[1], Tyler N. Shendruk[1], Lixin Zhan[1]

[1]*Department of Physics, University of Ottawa,*
150 Louis-Pasteur, Ottawa, Ontario, Canada, K1N 6N5

[2]*Frankfurt Institute for Advanced Studies, JW Goethe Universität,*
Ruth-Moufang-Strasse 1, D-60438 Frankfurt, Germany

[3]*Max Planck Institute for Polymer Research,*
Ackermannweg 10, D-55128 Mainz, Germany

August 19, 2009

Corresponding author:
Prof. Gary W Slater
email gary.slater@uOttawa.ca
fax +1-613-562-5190

Keywords:

Computer simulations, Electrophoresis, Microfluidics, Modeling, Separation methods

Running title:

Modeling Separation Processes

Abbreviations:

BD	Brownian Dynamics
BFA	Bond-Fluctuation Algorithm
BRM	Biased Reptation Model
DPD	Dissipative Particle Dynamics
ELFSE	End-Labeled Free Solution Electrophoresis
FENE	Finitely Extensible Nonlinear Elastic
HI	Hydrodynamic Interactions
LB	Lattice Boltzmann
LD	Langevin Dynamics
LJ	Lennard-Jones
LMC	Lattice Monte Carlo
Mbp	Megabase pairs
MC	Monte Carlo
MD	Molecular Dynamics
OMRC	Ogston-Morris-Rodbard-Chrambach
SRD	Stochastic Rotation Dynamics
VS	Verdier-Stockmayer
WCA	Weeks-Chandler-Andersen

Abstract

Theory and numerical simulations play a major role in the development of improved and novel separation methods In some cases computer simulations predict counterintuitive effects that must be taken into account in order to properly optimize a device In other cases simulations allow the scientist to focus on a subset of important system parameters Occasionally simulations even generate entirely new separation ideas! In this article we review the main simulation methods that are currently being used to model separation techniques of interest to the readers of *Electrophoresis* In the first part of the article we provide a brief description of the numerical models themselves starting with molecular methods and then moving toward more efficient coarse-grained approaches In the second part we briefly examine nine separation problems and some of the methods used to model them We conclude with a short discussion of some notoriously hard-to-model separation problems and a description of some of the available simulation software packages

Contents

1 Introduction

Computers are getting cheaper and more powerful every year At the same time separation systems are getting smaller and faster The convergence of these two trends has led to a situation where it is now possible to simulate the key parts of some separation systems at

the molecular level An example of this is the atomistic simulation of the translocation of a ssDNA molecule through a nanopore [1] a process that may lead to the $1000 genome sought by the NIH in the USA

Transport-based separation systems generally represent a compromise between the physical separation of several molecular species (*e g* due to their different velocities in the device) and their spatial spreading due to various diffusion-related processes Since most separation devices employ sieving liquids and electric forces modeling efforts must generally include long-range electrostatic forces long-range hydrodynamic forces frictional and diffu sion contributions conformational effects (for macromolecules) entropic factors gradients of various types and interactions with surfaces and obstacles It is the role of the theo- retician to reduce the number of factors to a bare minimum in order to design models that can be solved either analytically or numerically For example long-ranged hydrodynamic interactions are often neglected in the case of gel based separations

Simulating complex processes using numerical models that include various levels of detail is now widely seen as the third approach to scientific discovery complementing the well-established experimental and theoretical methods Computer simulations are more than mere attempts at reproducing experimental results Because we have full control over the simulation parameters and because we can measure every conceivable property (including correlations between properties) of a system during a *numerical experiment* simulations allow detailed autopsies and diagnostics not normally achievable in a laboratory They also allow us to explore numerous systems and geometries at relatively low cost

Historically several discoveries were first made on a computer i e *in silico* In the field of electrophoresis one of us (GWS) and his co-workers [2] first discovered the phenomenon of DNA band inversion using a computer simulation of the Biased Reptation Model (BRM) [3 4] Experiments later confirmed the existence of this counterintuitive phenomenon

In view of the growing importance of computer simulations in the field of separation methods we think that this review is timely Our hope is that it will be of equal interest to the experimentalist who wants to understand the relevant literature and use simulations

to guide his/her lab work to the theoretician who wants to explore novel systems using numerical methods and finally to the computational scientist who wants to use new tools

This review article has three main parts In Sec 2 below we describe several approaches that are currently being used to model separation processes We first examine methods that include molecular details then describe algorithms that simplify the description of the liquid phase and finish with several coarse-graining methods that generate even simpler numerical models In Sec 3 we review nine well-known problems and the simulation methods that are being used to study them In Sec 4 we look at the future and discuss a few problems that have proven hard to simulate and understand with the current methods An Appendix then lists some of the most well-known software packages in the field however it is important to note that many (perhaps most) simulation studies have actually been based on simulation programs written by the researchers themselves The choice of material is obviously that of the authors and choices had to be made in order to limit the length of this article we thus recognize that several important algorithms and problems are missing from this review Among those approaches based on the continuum description of the flow and the distribution of analytes [5 6] and reptation models of gel electrophoresis [4 7–12] are perhaps the most obvious

2 Simulation methods

In this section we describe some of the most popular simulation techniques used in computational studies of electrophoretic separation methods It is obviously impossible to review them all — as a matter of fact new methods are proposed every year In that light we present some of the main methods in a way that we believe is both logical (from the more microscopic to the most simplified) and compact In our descriptions we do not strive for the level of detail that would allow readers having no prior knowledge to do their own simulations Rather we present a bird s-eye view of the hierarchy of different computational methods Correspondingly the number of references is large as the interested reader will need to consult specialized texts in order to fully appreciate the intricacies of these simulation

methods

2.1 Molecular Dynamics

The Molecular Dynamics (MD) technique is used to follow the evolution of a large number of interacting particles by numerically integrating the classical equations of motion Although MD is often used to simulate systems at the molecular level it is also suitable for modeling larger scale systems by implementing coarse grained methodologies Hence it is a common approach to simulating polymer dynamics

When performing MD simulations the first step is to calculate the net force on each particle Newton s second law $\mathbf{F} = m\mathbf{a} = m\ddot{\mathbf{r}}$ then provides the particle s acceleration Numerically integrating by $\mathbf{v} = \mathbf{v_0} + \mathbf{a}\Delta t$ yields the particle velocity at a short time Δt later a second integration $\mathbf{x} = \mathbf{x_0} + \mathbf{v}\Delta t$ gives its position (in practice more accurate numerical integration techniques such as the velocity-Verlet algorithm [13 14] are employed) After performing these calculations for each particle the new configuration of the system at time $t + \Delta t$ is obtained This process is repeated to generate trajectories for each particle that together comprise a series of snapshots describing the evolution of the system For conceptual simplicity we begin by describing a fully atomistic MD simulation (each atom is represented by one particle) However as this is generally too computationally expensive we subsequently introduce various techniques to coarse-grain the models and make MD a viable tool for studying separation methods Although we will only discuss a few key topics there are a great number of excellent books [13 16] and review articles [17 19] which can be used to explore the rich field of MD simulations in greater depth

2.1.1 The force field

The interactions between the particles in the system obviously play a major role The mathematical forms and parameters dictating these interactions are known as the force field the key concepts used in this review are discussed below

Lennard-Jones: The interaction between two free uncharged atoms implies two primary effects First there is a short ranged repulsion preventing overlap Second there is a long-range attraction arising from weak but favourable interactions due to induced dipole effects (dispersion forces) The Lennard-Jones potential [20] is commonly used to model these effects

$$U_{LJ}(r_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \quad (1)$$

where r_{ij} is the separation between the particles i and j ϵ_{ij} is the depth of the potential well and σ is the effective size of the particle The numerical values used for ϵ and σ dictate the details of the interaction

Coulombic forces: The other primary non-bonded interaction arises from the electrostatic interaction between pairs of particles given by the Coulombic potential $U_{Coul}(r_{ij}) = \frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{\epsilon r_{ij}}$ Here q_i and q_j are the effective charges on each particle ϵ_0 is the permitivity of free space and ϵ is the dielectric constant of the medium In atomistic simulations with an explicit polar solvent model including ϵ is not necessary However for simulations with either implicit or non-polar solvent models (including the mesoscopic models common in coarse-grained simulations) ϵ is an effective dielectric constant that includes the screening effects due to the medium (*e g* ϵ=80 for water) Although the magnitude of the net charge is obvious for free ions in molecules where charges are shared via bonds the value of the effective partial charges is a vital component of the force-field Unlike the Lennard-Jones interaction which decays relatively quickly the Coulombic interaction is long ranged While a cut-off value (beyond which contributions are not considered) is appropriate in one dimension the long-ranged contributions are important in two or three dimensions For this reason inclusion of electrostatic effects can be computationally expensive and many sophisticated techniques have been developed to address this particular problem For a review of many of these methods see [21 22], further methods are discussed in [23 24]

Bonded interactions: In MD simulations elements of chemical bonds are captured by implementing potentials to maintain bond lengths and bond angles Although other forms are used a common choice for both is a harmonic potential such that the bond stretching

$U_{bond}(r_{ij})$ and bond bending $U_{bond}(\theta_{ijk})$ potentials are given by

$$U_{bond}(r_{ij}) = \frac{1}{2}k_{ij}(r_{ij} - r_0)^2 \quad (2)$$

$$U_{bond}(\theta_{ijk}) = \frac{1}{2}k_{ijk}(\theta_{ijk} - \theta_0)^2 \quad (3)$$

Here k_{ij} and k_{ijk} are force constants θ_{ijk} is the angle formed by the bonds joining atoms i j k and r_0 and θ_0 are the equilibrium separation and bond angle respectively Hence in this model atoms are bonded together via Hookean springs while the bond angles oscillate around the equilibrium value In atomistic simulations the various parameters are an essential part of the force field as they dictate the details of these interactions Note that while we discuss only bond stretching and bending here terms can be added to model effects such as restrictions to torsional angles and *cis vs trans* configurations [25]

2.1.2 Coarse graining

A fully atomistic MD simulation is quite detailed in the features it replicates and is appropriate to study dynamics that occur below the order of μsec and μm In general however they are prohibitively expensive computationally to probe the large systems and long time scale dynamics of interest for electrophoretic phenomena For this reason coarse-graining the system description is crucial for effective simulations

United Atoms: The first step in coarse graining a system is to lump groups of atoms together and simulate them as a single particle For example in atomistic MD simulations of proteins it is common to group the hydrogens of amino acid side-chain carbons in with the carbon to form a united atom Reducing the number of particles results in reduced computing time while preserving the dynamics of interest One can extend this idea and model entire side chains (or even entire monomers) as single particles From this we can then simulate any polymer as a string of bonded generic beads - each of which represents a monomer Of course the interaction between adjacent beads is vital such local effects can be incorporated by using *e g* a potential between adjacent bonds in the chain to give a finite stiffness to the backbone (for example the bond angle potential discussed above [26])

We can extend the level of coarse-graining even further by lumping polymer beads together and simulating n monomers as a single particle that now represents a subchain of the polymer. If n is large enough that the length of the subchain is equal to or exceeds the Kuhn length (a measure of the stiffness of the polymer chain), correlations between subchains are negligible and we can simulate the coarse-grained polymer as a freely jointed chain [27]. This is a standard model for studying a polymer via MD simulations as it allows for computational simplicity and efficiency while preserving the dynamics of interest. What is lost is the direct correspondence to the system in study and hence an extra layer of abstraction between simulation and reality is introduced.

The WCA Potential: In coarse-grained simulations it is common to use the purely repulsive Weeks-Chandler-Andersen (WCA) potential to model particle interactions [28]. This is a Lennard-Jones potential that is shifted and truncated at its minimum value so that it ends smoothly at a distance $r_c = 2^{\frac{1}{6}}\sigma$.

$$U_{WCA}(r_{ij}) = \begin{cases} 4\epsilon_{ij}\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] + \epsilon_{ij} & \text{for } r_{ij} < r_c \\ 0 & \text{for } r_{ij} \geq r_c \end{cases} \tag{4}$$

However, the standard Lennard-Jones interaction is often used to study systems with varying affinities between different types of particles (this is implemented by varying the ϵ_{ij} s).

Finitely extensible bonds: In a coarse-grained model, the use of harmonic potentials to connect adjacent beads can lead to unphysical bond stretching effects. For this reason, a finitely extensible force is needed such that there is a hard limit on the bond length. The most common choice is the Finitely Extensible Nonlinear Elastic (FENE) potential.

$$U_{\text{FENE}} = -\frac{1}{2}\, k\, r_0^2 \ln\left(1 - \frac{r^2}{{r_0}^2}\right) \tag{5}$$

Here k dictates the stiffness of the bond and r_0 gives the maximum bond length (the potential shoots to infinity as r approaches r_0). For appropriate values of k and r_0, the occurrence of bond crossing is extremely rare and hence behaviour appropriate for a self-avoiding polymer is produced [29]. It is interesting to note that this form is an approximate solution to the inverse Langevin function, in some cases it may thus account for the entropic restoring force which arises from stretching a subchain [27].

2.1.3 Simulating a polymer

Putting the above together the simplest model of a polymer in a coarse-grained simulation is a freely-jointed chain of beads (which can either carry a net charge or be neutral) interacting by the WCA potential (or the Lennard-Jones potential when short range attractions are desired) and with immediately neighbouring monomers bonded via the FENE potential Extension to branched polymers is trivial While we have sacrificed many details we have preserved effects such as excluded volume the entropic elasticity and the non-crossability of bonds While this model may seem very crude it does allow for a realistic scaling behaviour of polymeric properties such as the radius of gyration and the diffusion coefficient with respect to the degree of polymerization This approach also represents the most realistic model of polymers for which experimentally relevant simulations are feasible Note that although this bead-spring model is a common approach other models can be employed For example algorithms can be implemented to maintain the bond distance at a fixed length [30 31] resulting in a bead-rod model (or a pearl-necklace model for the case of the stick length being comparable to the diameter of the bead) [32]

Now that we have a model for the polymer we can add other components For example a common scenario involves simulation of polyelectrolytes with free ions To accomplish this ions can be defined as WCA particles carrying a net charge Additionally one can include features such as obstacles or boundaries (*e g* walls) by building them out of particles or defining them using mathematical constraints The final and critical component of the simulation system is the solvent As the model used for the fluid impacts both the dynamics which are observed and the computational time Sections 3 2 3 3 and 3 4 will focus on various ways of treating the fluid

At this point it is important to mention that choosing a method of controlling the temperature is often a crucial consideration in MD simulations, in fact it is also one that is often intimately connected to the choice of solvent model Implementing a thermostat is particularly critical when an external field is adding energy to the system – as in electrophoresis While there are schemes based on rescaling the velocities [33] or periodically assigning random velocities [34] such algorithms can lead to artifacts as local momentum

is not conserved (a particular problem when hydrodynamics are of interest [35]) In coarse-grained simulations the method chosen for treating the fluid often involves a particular temperature control scheme via the coupling of the analyte (*e g* the polyelectrolyte) to the fluid An important case is the thermostat developed for Dissipative Particle Dynamics (Section 3 3 1) which can be used independently of the fluid model and for several reasons has been a significant advancement in the field [36] Finally in the case of performing Langevin Dynamics or Brownian Dynamics the temperature is explicitly in the equation of motion (Section 3 4) While discussing the details of these algorithms (and others [37–40]) is beyond the scope of this review the interested reader is encouraged to investigate these various schemes starting with the citations given and the discussion of solvent models in the following sections

2.2 Explicit fluid

At the atomistic level many models exist which explicitly describe a water molecule [41] However in coarse-grained simulations we group the atoms and even molecules together and simulate them as a single bead Use of the WCA potential for this solution of beads is found to provide a good solvent regardless of temperature [42] The advantage of using an explicit fluid is that it is conceptually obvious and preserves the long-range hydrodynamic interactions that correlate the movement of objects in a fluid However in practice this results in much of the simulation time being dedicated to calculating the details of the fluid bead dynamics which are often not of interest For these reasons several techniques for modeling the fluid to maintain the hydrodynamic interactions while neglecting the computationally costly details of the fluid motion have been developed These models are discussed in the following section

2.3 Mesoscopic fluid

A number of computational approaches to the simulation of liquids are based on a coarse-grained representation of the underlying physics The description of the hydrodynamic effects

in these approaches lies in between the use of an explicit description of fluid molecules and the discretization of the continuous Navier-Stokes equations In particular a clever use of conservation laws allows these methods to employ local algorithms to recover the solution to the hydrodynamics equations in the large scale/long-time limit bypassing at the same time the molecular detail of the fluid and thus reducing dramatically the computational cost of the simulation In the following sections we will review three mesoscopic approaches namely Dissipative Particle Dynamics Stochastic Rotation Dynamics and Lattice Boltzmann

Following the description of these mesoscopic methods Section 2 4 will introduce Langevin and Brownian Dynamics A comparison of the different methods discussed ensues in Section 2 5 where a schematic is presented to summarize each of the methods

2.3.1 Dissipative Particle Dynamics

In Dissipative Particle Dynamics (DPD) the fluid is modeled by large particles interacting via a *soft* potential [43 45] This allows for a reduction in computing time in two ways First as each DPD fluid bead represents a cluster of fluid molecules moving together in a coherent fashion the simulation tracks a much lower number of interacting objects Second since with a soft potential the forces cannot be arbitrarily large we can reduce computing times by increasing the integration time step All particles interact by three forces a conservative force $\mathbf{F}^{\mathrm{C}}$ a dissipative force $\mathbf{F}^{\mathrm{D}}$ and a random force $\mathbf{F}^{\mathrm{R}}$

$$\mathbf{F}_i = \sum_{i \neq j} \left(\mathbf{F}_{ij}^{\mathrm{C}} + \mathbf{F}_{ij}^{\mathrm{D}} + \mathbf{F}_{ij}^{\mathrm{R}} \right) \tag{6}$$

where it is assumed that the interactions are negligible beyond a cutoff radius r_{c} Pairwise potentials ensure that momentum and angular momentum are conserved The force $\mathbf{F}_{ij}^{\mathrm{C}}$ represents and can include any conservative forces which act on the particle A common choice is a soft repulsion (see Fig 1) acting along the line of centres such as

$$\mathbf{F}_{ij}^{\mathrm{C}} = \begin{cases} a_{ij} \left(r_{\mathrm{c}} - r_{ij} \right) \mathbf{r}_{ij} & \text{for } r_{ij} < r_{\mathrm{c}} \\ 0 & \text{for } r_{ij} \geq r_{\mathrm{c}} \end{cases} \tag{7}$$

where a_{ij} parametrizes the maximum repulsion The dissipative force is an inter-drag force between a pair of soft fluid particles moving through each other opposing their relative motion $\mathbf{u}_{ij}$ and dissipating heat

$$\mathbf{F}_{ij}^{\mathrm{D}} = -\omega^{\mathrm{D}}(r_{ij})\zeta(\mathbf{r}_{ij} \ \mathbf{u}_{ij})\hat{\mathbf{r}}_{ij} \tag{8}$$

where ζ is the friction constant between the two clusters The random noise force is

$$\mathbf{F}_{ij}^{\mathrm{R}} = \omega^{\mathrm{R}}(r_{ij})\sqrt{2\zeta k_{\mathrm{B}}T}\frac{\gamma_{ij}}{\sqrt{\Delta t}}\hat{\mathbf{r}}_{ij} \tag{9}$$

where γ_{ij} is a random number of zero mean and unit standard deviation To satisfy the fluctuation-dissipation theorem the dissipative and random forces are interrelated through the weight functions as $\omega^{\mathrm{D}}(r) = \left[\omega^{\mathrm{R}}(r)\right]^2$

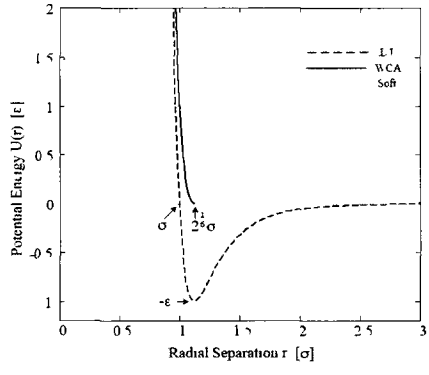


Figure 1 Comparison of the soft potential often used in DPD with the LJ and WCA potentials The positive soft potential extends far further and avoids the singularity of the LJ and WCA potentials

The random and dissipative forces act as a source and a sink for heat respectively Therefore DPD unlike BD or LD is an implicit thermostat [36 46] that conserves linear and angular momentum and thus recovers hydrodynamics in the macroscopic limit It should be noted that DPD does not conserve total energy only mass and momentum Solutes can be included as DPD beads and by the inclusion of a bead-spring type force in Eq 6 DPD

can simulate polymers as well [46 47] In fact DPD can be used as an effective thermostat that conserves hydrodynamic interactions independently of the fluid model [36]

2.3.2 Stochastic Rotation Dynamics

The DPD beads represent clusters of particles but all bead-bead interactions must still be evaluated In Stochastic Rotation Dynamics (SRD) [48-50] also called Multi-Particle Collision dynamics (MPC) [51 52] or Real-coded Lattice Gas (RLG) [53 54] collisions between fluid particles are replaced by multiparticle collision events that omit the molecular details and eliminate the need to calculate the forces between the fluid particles These events are defined to conserve mass momentum and energy such that the hydrodynamic equations of motion are obeyed on sufficiently long length and time scales [55] SRD simulations occur in two steps During the first or *streaming* step the particles move ballistically and their positions $\mathbf{r}_i(t)$ are updated in discrete time intervals δt

$$\mathbf{r}_i(t + \delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\delta t \tag{10}$$

The second or *collision* step transfers momentum between particles The simulation domain is partitioned into cells The number of particles in each cell may vary from one cell to another but the total number is conserved Each cell has a centre of mass velocity $\mathbf{v}_{\mathrm{CM}}$ which corresponds to the local macroscopic velocity The collision step is a simple non-physical scheme that is constructed to conserve momentum Multiparticle collisions within each cell are represented by the operation

$$\mathbf{v}_i(t + \delta t) = \mathbf{v}_{\mathrm{CM}}(t) + \mathbf{R}\,(\,\mathbf{v}_i\,(t) - \mathbf{v}_{\mathrm{CM}}\,(t)\,) \tag{11}$$

By making the collision operator $\mathbf{R}$ a rotation through an angle α about a randomly chosen axis conservation of energy isotropy and a Maxwell-Boltzmann velocity distribution are met in the continuum limit Other choices allow SRD to operate as a thermostat as well [56] Unfortunately Galilean invariance is broken by the discretization of space into cells However this can be completely remedied by performing the collision operation in a cell grid which is shifted each time step by a random vector [57 58]

Solute particles are mainly coupled to the fluid by inclusion in the collision step [51 52 59–64] although hybrid SRD/MD schemes where solvent-solute and solute-solute interactions are handled by an MD approach while solvent-solvent interaction is simulated by SRD are also common [49 65 66] To ensure that the momentum transferred to the polymer during the collision step is well distributed throughout the chain MD time steps must occur many times every collision step [63]

2.3.3 Lattice Boltzmann method

Lattice Boltzmann (LB) coarse-grained models [50 67 68] are based on a solution of the discretized Boltzmann transport equation The main quantity in the LB approach is the velocity field rather than fluid particles and it is an inherently statistical approach where discrete momentum distributions are represented on a spatial grid

The discretization of positions and momenta using finite sets of directions greatly simplifies the problem The most frequent mesh types for the LB simulations are the D2Q9 D3Q15 and D3Q19 lattices where DkQn refers to the number k of dimensions and to the discrete number n of velocity vectors $\mathbf{e}_i$

A set of distribution functions $\Gamma_i(\mathbf{r}\ t)$ is defined on each lattice site $\mathbf{r}$ Each of these can be interpreted as the fraction of fluid that will move with the i-th discretized velocity at time t The discretized Boltzmann equation provides a generic description for the time evolution of the probability density but there is freedom in the choice for the acutal form of the collision integral A common formulation is the Bhatnagar-Gross-Krook approximation where

$$\Gamma_i(\mathbf{r} + \mathbf{e}_i\Delta t\ t + \Delta t) = \Gamma_i(\mathbf{r}\ t) - \frac{\Delta t}{\tau}(\ \Gamma_i(\mathbf{r}\ t) - \Gamma_i^{\mathrm{EQ}}(\mathbf{r}\ t)\ ) \tag{12}$$

where τ is the phenomenological relaxation time which prescribes the timescale for the relaxation of the actual population Γ_i to the equilibrium particle distribution function Γ_i^{EQ} In the low velocity approximation Γ_i^{EQ} can be written as

$$\Gamma_i^{\mathrm{EQ}} = w_i\rho\left[1 + \frac{\mathbf{e}_i\ \mathbf{u}}{c_s^2} + \frac{(\mathbf{e}_i\ \mathbf{u})^2}{2c_s^4} - \frac{u^2}{2c_s^2}\right] \tag{13}$$

where ρ is the hydrodynamic density and c_s is the speed of sound which is determined by mesh properties The weights w_i must be suitably chosen to recover the macroscopic Navier-Stokes equations and are dependent on the mesh configuration [50] The viscosity of the LB fluid is determined by the choice of relaxation rates

LB can be coupled to small suspended spheres by treating them as point particles that interact with the fluid through a friction force proportional to the relative velocity obtained via linear interpolation from the surrounding lattice sites [69 70] By adding fluctuation terms to both the fluid and the embedded particles LB can operate as an adequate thermostat [69] The correct treatment of a fluctuating LB algorithm has been recently adressed in several papers [71 72]

2.4 Langevin and Brownian dynamics

Going beyond mesoscopic models further coarse graining can be achieved by avoiding direct simulation of the fluid altogether The timescale of individual collisions with solvent molecules (causing frictional drag and Brownian motion) is much smaller than the time scales relevant to electrophoresis It is computationally advantageous to coarse grain out the fine details of the collisions and simply include their statistical effect on the solute One may consider *implicitly* the two main effects of the fluid acting on the particle i) a frictional force opposing its motion and ii) random kicks arising from collisions with the solvent The frictional (or dissipative) force $\mathbf{F}^D(t)$ removes energy from the particle while the fluctuating Brownian force $\mathbf{F}^B(t)$ adds energy to the particle Hence at this coarse-grained level the fluid is included solely in a statistical manner governed by the fluctuation-dissipation theorem By replacing the explicit fluid with a drag and a Brownian force we lose the long-range particle-particle interactions mediated by the fluid This makes such an approach particularly tempting when the hydrodynamic interactions are negligible or not of primary concern Nevertheless we will see how they can still be included

2.4.1 Pure Langevin and Brownian Dynamics

In Langevin dynamics starting with Newton s 2^{nd} law as we did for pure MD we now add a dissipative drag force $\mathbf{F}^{\mathrm{D}}(t)$ and a Brownian force $\mathbf{F}^{\mathrm{B}}(t)$ (in addition to the conservative forces we had before $\mathbf{F}^{\mathrm{C}}(t)$) to end up with Langevin s equation

$$m\mathbf{a}(t) = m\ddot{\mathbf{r}}(t) = \mathbf{F}^{\mathrm{C}}(t) + \mathbf{F}^{\mathrm{D}}(t) + \mathbf{F}^{\mathrm{B}}(t) \tag{14}$$

where m is the mass of the particle For the dissipative term one usually assumes Stokesian drag on a spherical particle $\mathbf{F}^{D}(t) = -\zeta\mathbf{v}(t)$ ζ being the friction coefficient of the particle in the fluid The velocity $\mathbf{v}(t)$ is then the velocity of the particle with respect to the local solvent velocity This is an important detail if one wants to consider flow or long-range hydrodynamic interactions Although $\mathbf{F}^{\mathrm{B}}(t)$ is due to the solvent molecules colliding with the particle it can only model the net effect of a large number of collisions The Brownian force is taken as a centered Gaussian random variable [73] with zero mean and variance $\frac{2\zeta k_{\mathrm{B}}T}{\Delta t}$ where Δt is the integration time step The fact that the variance is related to the friction coefficient ζ is again a consequence of the fluctuation-dissipation theorem On the time scale of interest the values of the Brownian force are uncorrelated at different time steps

It can be shown that the energy transferred to the particle from a single collision with a solvent molecule decays on the viscous time scale $\frac{m}{\zeta}$ [74] If as is typically the case this is much smaller then the timescale over which $\mathbf{F}^{\mathrm{C}}(t)$ changes (overdamped limit) we may set $m\mathbf{a}(t) = 0$ in the Langevin equation and obtain the following discretized equation of motion

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \frac{\Delta t}{\zeta}\left[\mathbf{F}^{\mathrm{B}}(t) + \mathbf{F}^{\mathrm{C}}(t)\right] \tag{15}$$

which defines *Brownian Dynamics* For a more rigorous derivation see [75]

2.4.2 Incorporating long-range hydrodynamic interactions

Traditionally Langevin Dynamics does not include hydrodynamic interactions (HI) between particles This approximation is valid for certain systems (e g when HI are screened out) but in many cases HI have a significant impact on the dynamics [76]

To see how we can incorporate them let us consider a particle i As it is moving with a certain velocity it slows down due to the drag force Since the drag force exerted by the fluid on a particle must be equal and opposite in direction to the force exerted by the particle on the fluid this causes the fluid to move Hence the effect particle i has on particle j is that the latter experiences a modified drag as it is no longer surrounded by a stationary fluid The magnitude of this coupling depends on their relative separations On the time scale of interest this perturbation can generally be considered to be felt instantaneously In a many-particle system a common approximation is to ignore screening issues and to consider a superposition of all pairwise hydrodynamic interactions [77]

Consideration of the fluctuation-dissipation theorem implies that any modulation of the drag term must be accompanied by correctly modifying the magnitude of the Brownian term An interesting consequence is the correlation of the Brownian forces on different particles at the same time In practice the relationship between the drag and the Brownian term is determined via interaction tensors (mainly the Oseen or Rotne Prager-Yamakawa tensor) Application details can be found in these references [74 75 77–82]

2.5 Comparison of fluid models

We have presented a hierarchy of fluid models from the most detailed explicit solvent models to the mesoscopic and finally implicit (LD and BD) approaches

The advantage of explicit fluid models is their conceptual simplicity as the dynamics of each fluid particle simply follows from Newton s laws The obvious disadvantage is high computational cost Using such models (especially the atomistic variant) is essential when high accuracy is required and the chemical details (such as different hydrophobicities of different parts of the analyte) are important In most other cases simpler approaches should be used

All three mesoscopic methods (DPD SRD and LB) use a simple (but often sufficient) model to describe fluid dynamics and can also act as thermostats that define the local

temperature when coupled to MD particles As such they are well suited to be used for electrophoresis simulations as they can dissipate the energy transferred to the system by the applied electric field

Due to their conceptual differences these methods use different types of parameters to describe the fluid which results in a different suitability for specific problems For example while the fluid viscosity is directly accessible in LB methods it becomes a combination of different parameters and can only be controlled indirectly in DPD [83 84] Similarly while DPD and LB can only approximate the continuous-time dynamics of the fluid when the discrete time step is small SRD is proven to yield correct long-time hydrodynamics for any step size However SRD s transport properties depend explicitly on the chosen time step [55] More differences between the methods arise if confined fluids or interactions with large obstacles or particles are studied Here the ability to treat different boundary conditions becomes important which is covered in detail in the literature [49 59 85 89]

All three methods share a similar computational efficiency and computation times depend mainly on the implementation the computer system and also the investigated system However the speedup over explicit fluid simulations can be a factor 20 or higher [69 90]

These mesoscopic models describe compressible fluids in which hydrodynamic forces need time to propagate through the medium For this reason as well as considerations of computational speed performing Langevin dynamics with proper inclusion of hydrodynamic interactions can be preferable in some cases [91] Furthermore if the full treatment of hydrodynamic interactions is not necessary the use of pure Langevin or Brownian dynamics is advisable

We summarize the different ways of treating the solvent in MD simulations with a schematic that depicts the most important features of different fluid models (Fig 2)

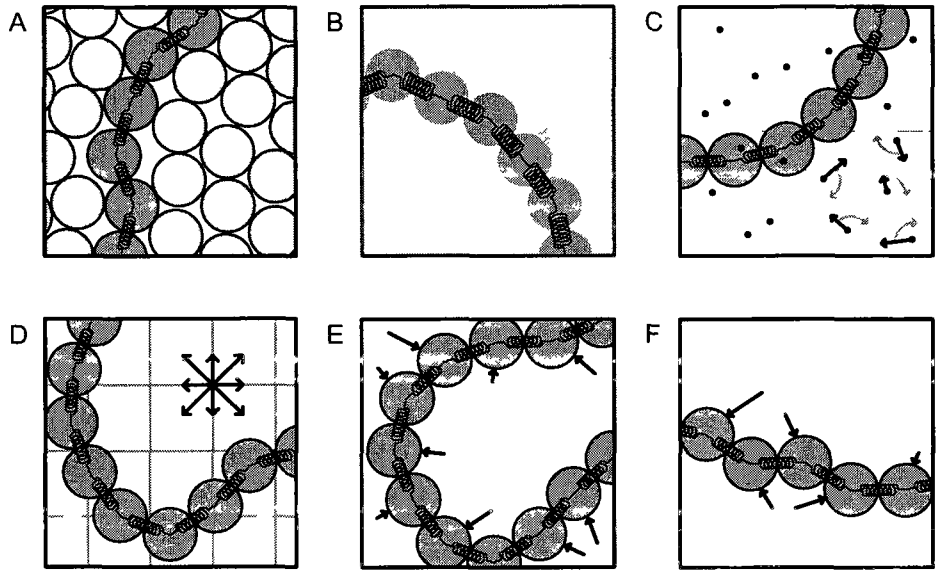


Figure 2 Schematic depicting different methods of modeling the solvent in MD simulations In each case a representative polymer is shown as a chain of beads connected by springs (A) The explicit fluid approach models the solvent as beads interacting through a short-range potential and simulated to the same level of detail as the polymer Fluid particles are often chosen to be the same size as the monomers (B) The DPD method typically models the solvent as large overlapping softly repulsive spheres (C) The SRD method replaces collisions between particles by a multiparticle collision operator which models the local collisions as a rotation of particle velocity through an angle about a randomly chosen axis (D) In LB the fluid is discretized to a grid Arrows symbolize the discrete set of allowed velocities The mesh shown is a D2Q9 (eight velocity vectors plus the zero velocity in two dimensions) (E) LD/BD represents the fluid implicitly by including a friction and a random term in the equation of motion of each monomer Hydrodynamic interactions can be included by introducing an additional long-range force between monomers (illustrated for one bead as a series of dashed lines) (F) Pure LD/BD neglects hydrodynamic interactions friction and Brownian forces are the only effects of the solvent that are included After Padding and Louis [92]

2.6 Monte Carlo simulations

In the previous section we discussed how atomistic models can be coarse-grained to make simulations faster The resulting models are simpler but they are designed to keep the essential features of the system (e g electrostatics and hydrodynamics) However once the coarse graining is done dynamics is simulated in all detail solving the equations of motion as carefully as possible It is reasonable to ask if it is possible to also coarse-grain the dynamics itself in order to achieve further computational speedup This is the motivation behind the development of the various *Monte Carlo (MC) methods*

In the broadest sense of the term any computational method involving randomness can be called a Monte Carlo method (the name says it all!) This broad definition would however include e g Brownian dynamics already considered under the banner of MD In the spirit of the preceding discussion we would therefore call a MC method any approach that sacrifices *at least some* of the dynamical details of the MD methods How much is sacrificed can vary At one extreme are methods designed to study *equilibrium* properties by quickly exploring large parts of the phase space In this case unphysical moves such as rotations of large parts of chains in the pivot algorithm [93] can be introduced *on purpose* to speed up the simulation Since in simulations of electrophoresis and related separation methods we are primarily interested in the *dynamics* of the analytes such MC approaches are less useful for our purposes although they can be applied to some auxiliary problems such as generating an equilibrated entangled sieving polymer solution in which the analytes would then migrate studying the properties of capillary coatings or finding the free energy of a DNA molecule as a function of its coordinate along a nanopore We do not discuss these methods here they are reviewed e g in Refs [94–96] Instead we concentrate on *dynamical MC algorithms*

MC methods can be divided into two groups In *off-lattice MC* methods particles and monomers of polymer chains can occupy any positions in space But one can also discretize the space allowing the particles to reside only at the nodes of a lattice Such methods are known as *Lattice Monte Carlo* (LMC) Since the sets of possible configurations and moves are then discrete they can be described using integer arithmetic and often special computational

techniques such as multispin coding ([97] Section 15) which allow significant savings in computer time and memory compared to off-lattice MC Of course the cost of increased efficiency in LMC is even less realism!

Our ideal goal when simulating electrophoresis problems is to make quantitative predictions for the mobility and diffusivity of the analyte Although this is possible in simple cases MC simulations are generally used to predict trends (including scaling laws) and qualitative features which in many cases is quite acceptable

We start our discussion of MC methods with approaches to simulating hard particles We then discuss some chain simulation methods As discussed in the Introduction we will not describe the MC methods used to simulate DNA reptation models [98]

2.6.1 Methods for particles

Consider a particle undergoing Brownian motion in the presence of an external force (e g an electric field) and some obstacles (e g gel fibers) In MC this becomes a simple biased random walk At each time step a move is randomly selected from a predefined set and a test is used to accept or reject it A simulation is simply a series of such moves Different MC algorithms are defined by a) the set of moves and the probability of selecting a particular move from that set b) the acceptance test for the selected move c) the time step per move Although the physics behind these three elements can be subtle (the main issue is generally the definition of the time scale) the simulation itself is often quite simple

The most popular MC approach is *Metropolis MC* [99] a method designed to study equilibrium configurations In the simplest lattice variant at each step the particle simply tries to move to any of the neighboring sites on the lattice and all possible moves can be selected with equal probability If the change in energy from the initial to the final configuration is negative ($\Delta U < 0$) the move is accepted otherwise it is accepted with probability $\exp(-\frac{\Delta U}{k_B T})$ (the Boltzmann factor) This is known as the Metropolis test The latter guarantees that the correct equilibrium properties are obtained However reproducing

the correct dynamics is problematic First of all how the MC unit of time (an attempted move) is related to actual time is usually left undefined which makes quantitative predictions difficult Moreover in a high field the method is even *qualitatively* wrong once $\frac{|\Delta U|}{k_B T} \gg 1$ all steps along the field are accepted and all those against the field are rejected, as a result the velocity saturates and the diffusion coefficient vanishes A different approach is then required

In the 2D lattice variant of such an approach [100] the system is modeled as a square lattice where each node is either free or is an impenetrable obstacle At each time step the particle can move in one of four directions ($\pm\hat{x}$ $\pm\hat{y}$) Unlike in the Metropolis algorithm the probabilities of selecting each of these moves are no longer equal instead if an external force $\vec{F} = F\hat{x}$ is applied to the particle they are given by [100]

$$\begin{aligned} p_{\pm x}(\varepsilon) &= \frac{1}{(1+e^{\mp 2\varepsilon})\,(1+\tanh(\varepsilon)/\varepsilon)} \\ p_{\perp}(\varepsilon) &= \frac{1}{2\,(1+\varepsilon\coth(\varepsilon))} \end{aligned} \tag{16}$$

where the scaled force ε is given by

$$\varepsilon = \frac{Fa}{2k_B T} \tag{17}$$

with a the mesh size of the lattice Note that the algorithm is *rejection-free* in the absence of obstacles, however moves leading to an overlap between the particle and an obstacle are rejected The time step is no longer arbitrary as it is given by the expression

$$\tau(\varepsilon) = \frac{\tau_B}{1+\varepsilon\,\coth(\varepsilon)} \tag{18}$$

where $\tau_B \equiv \tau(\epsilon = 0)$ is the mean duration of a (Brownian) jump in the absence of an external field The latter is directly related to the free-space diffusion coefficient D_0

$$\tau_B = \frac{a^2}{2D_0} \tag{19}$$

It is this relation that connects the MC and experimental times The above choice of transition probabilities and the time step can be shown [100] to give the correct average velocities for fields of arbitrary strength However the dispersion coefficient is only correct in the limit of a vanishingly small field [101] In this limit the dispersion coefficient D can actually be

obtained more efficiently by using the Nernst-Einstein relation

$$D = \lim_{F \to 0} k_B T \mu(F) \tag{20}$$

where $\mu(F) = \frac{v(F)}{F}$ is the mobility In a non-vanishing field the correct dispersion coefficient can be obtained by varying the time step (making it a random number) [101] If a constant time step is desired (as is the case for the numerically exact algorithm that we describe next) the MC moves themselves must be modified [101 102]

Exact Calculation Method

Standard MC methods require a large amount of simulation data in order to have a low statistical error In recent years our group has developed a numerical method that allows one to compute the exact mean velocity and diffusion coefficient of a particle moving on a lattice with impenetrable obstacles This method which basically gives the exact solution to the Monte Carlo simulation is both faster and more precise We now show the basics of this approach For further details the reader can refer to [103 104]

The first step is to obtain the transition matrix T whose elements T_{ij} are probabilities that a particle on site j jumps to site i in a single time step If at time t the probability of presence of the particle on site i is $n_i(t)$ then after a single time step

$$n_i(t + \tau) = \sum_j T_{ij} n_j(t) \tag{21}$$

Implicitly this assumes that the time step is unique since the latter depends on the field intensity the method only works in a uniform field [unless the field is so weak that the field dependence of $\tau(\varepsilon)$ can be neglected] The steady state defined by the equality $n_i(t + \tau) = n_i(t) \equiv n_i$ is thus the normalized eigenvector of T with the eigenvalue of unity (the normalization condition is $\sum_i n_i = 1$) This eigenvector which can be obtained with an arbitrary precision by a simple numerical calculation is the exact solution of the LMC algorithm

Once the steady-state occupation probabilities $n_i(\epsilon)$ are computed the mean velocity of the particle $v(\varepsilon)$ can be obtained by averaging over all sites $v(\varepsilon) = \sum_i n_i(\varepsilon)v_i(\varepsilon)$ where the average *local* velocity on site i is

$$v_i(\epsilon) = \frac{p_{+x}(\varepsilon)L_+(i) - p_{-x}(\varepsilon)L_-(i)}{\tau(\varepsilon)} \tag{22}$$

with the displacements $L_\pm = a$ if there is no obstacle in the given direction and zero otherwise

The dispersion coefficient D in the zero-field limit can be calculated using Eq (20) In a non-vanishing field D can be obtained with a numerically exact method based on the generalized Taylor-Aris dispersion theory [105]

2.6.2 Methods for chains

Simulating polymer chains presents additional challenges since connectivity and (often) non-crossability of chains have to be maintained during the simulation In MC simulations of polymer chain dynamics only one monomer (or at most a small local group of monomers) is moved at each step This choice is necessary in order to keep the frequency of moves leading to overlaps sufficiently low However a unit of time should now correspond to one attempted move *per monomer*

Off-lattice methods

In the most straightforward and commonly used off-lattice MC method for *bead-spring* chains [106] a single step consists in displacing a monomer chosen at random along each axis by an amount chosen from the uniform distribution on $[-\frac{\Delta}{2} + \frac{\Delta}{2}]$ where Δ is a predefined constant The event is accepted or rejected according to the Metropolis criterion but all attempts make the clock advance If the arbitrary parameter Δ is too small the evolution of the system may be too slow, if Δ is too large the rejection rate is too high The optimal Δ normally corresponds to an acceptance rate of about 50% although a lower Δ

can make the dynamics more realistic The same potentials as for MD (e g the FENE and WCA potentials) can be used Besides this one can use much simpler potentials such as a square well bond potential that is zero within a specified range and infinity outside with the range chosen (just as in FENE) to avoid chain crossings This improves the efficiency as the computation of complicated potential functions is avoided However in this case the chain will have no tension and so square-well potentials should be avoided in strong fields (see below)

In the *bead rod* model [107] the bond lengths have to be preserved explicitly The simplest moves are then rotations of a monomer around the axis connecting its two neighbors (for an end monomer the rotation is on the sphere with the center at its neighbor) Note though that in a strong electric field when the chain is stretched such moves are rather inefficient and this may lead to unphysical artifacts

Bond-length-preserving lattice methods

In lattice models of polymers monomers still hop between lattice sites like in single-particle models To avoid introducing a bond potential and still make sure that neighboring monomers remain close in space only those motions that keep all bond lengths within a certain range are allowed The first models were particularly restrictive in this respect keeping all bond lengths strictly fixed Verdier and Stockmayer (VS) [108] have used a simple cubic lattice requiring the polymer bonds to coincide with the lattice bonds and thus making them all of unit length In the first version of the model only single-monomer motions were allowed This was later found too restrictive and various two-monomer motions were added the most popular of which is the so called crankshaft motion [109] Unfortunately the dynamics in such models become very slow if we introduce excluded volume interactions by forbidding two monomers to reside on the same site Moreover Madras and Sokal [110] showed that for *any* model with *any* finite set of moves where neighbors along the chain remain neighbors on the lattice some configurations cannot be reached (i e the algorithm is nonergodic) Nevertheless in those cases where excluded volume interactions can be neglected the VS and other similar approaches can still be useful

Bond-Fluctuation Algorithm

The deficiencies of the fixed bond length MC models led to the development of lattice methods with fluctuating bond lengths. The most popular one is the Bond Fluctuation Algorithm (BFA) first proposed by Carmesin and Kremer for two-dimensional problems [111] involving both linear and branched polymers. While BFA is not strictly ergodic its nonergodicity problems can probably be neglected for all practical purposes. The algorithm was extended to 3D by Deutsch and Binder [112].

In the BFA the monomer is represented by a 2×2 square (four lattice sites) in 2D and a $2 \times 2 \times 2$ cube (eight lattice sites) in 3D. A lattice site can only be occupied by one particle at a time. In 2D the bond lengths between connected monomers must be less than 4 while in 3D bond lengths must be $\leq \sqrt{10}$ excluding $\sqrt{8}$. These simple conditions allow for a self-avoiding walk in which there is no crossing of bonds. Each attempted move consists of first picking a monomer at random and moving it by one lattice site along one of the lattice axes. As long as the new conformation does not create monomer overlaps or create a forbidden bond length the move is accepted. As before a unit time corresponds to one attempted move per monomer.

For simulations of electrophoresis obstacles can be placed at some lattice sites and the field is treated as in other models by using the Metropolis test. In principle other interactions such as intra- and interchain interactions between monomers can be included as well but in most cases this is unnecessary (an exception is for the proper modeling of polymer stiffness an important factor for DNA simulations).

Problems in strong fields

Lattice MC models considered here have severe problems in strong electric fields. For instance in gel electrophoresis simulations the speed of the polymer decays exponentially as a function of the field while experimentally the electrophoretic mobility is essentially

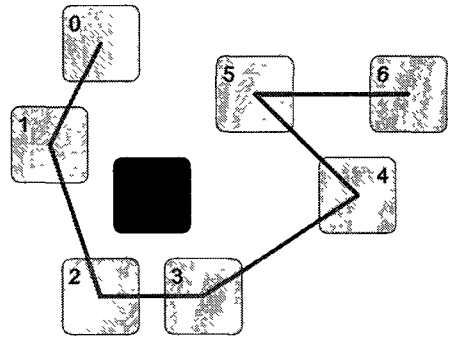


Figure 3: A polymer chain in the 2D Bond-Fluctuation Algorithm. Each gray square represents a monomer occupying 2×2 sites. All allowed bond lengths from 2 (between monomers 2 and 3) to $\sqrt{13}$ (between monomers 3 and 4) are present in this figure. Allowing only these bond lengths and forbidding the monomers to overlap with the obstacle (black) is sufficient to ensure that the chain does not pass through the obstacle and does not cross itself. After Ref. [113]

field-independent in the strong-field limit. To understand the reasons for this failure, consider a chain hooked upon a post with the two unequal arms pointing along the field and sliding off the post [114]. The sliding time is inversely proportional to the field intensity. But in lattice MC, since only local moves (e.g., single-monomer moves) are allowed, the only way for the chain to move is via chain slacks originating at the end of the short arm and propagating against the field. The probability for the slack to go all the way to the top decays exponentially with the potential energy difference between the tip of the arm and the post. If the short arm of the hooked conformation is of length L_s, the mean time between successful events will thus increase roughly like $\exp(+FL_s/kT)$, where F is the force on the monomer. This increase being an artifact of the model (the tension does not propagate along the backbone), the model can only work if the argument of this Boltzmann factor is much smaller than unity. Since L_S is proportional to the number of monomers in the chain, N, the maximum allowed external force scales like $F \propto N^{-1}$. This is too restrictive to be useful in practice. In order to solve this problem, one must modify the lattice models by adding *non-*

local moves an idea first implemented by Deutsch and Reger [114] and Duke and Viovy [11] for reptation models Azuma and Takayama [115] added such moves to the BFA While their approach is not very carefully justified it produces qualitatively reasonable results

Numerically exact methods for chains

Lattice MC algorithms for chains can sometimes be solved exactly similar to how this is done for single particles This was done by Boilcau and Slater [116] for the Bond-Fluctuation Algorithm The major complication is that each possible chain conformation and location should be considered a separate state and the number of such conformations grows exponentially with the chain length For this reason the approach is only practical for very short polymers (linear or branched) Of course the solutions are only exact for the dynamics of specific algorithms which are themselves approximate We should also mention a numerically exact solution of the MC approach to studying polymer translocation through nanopores [117] that reduced this problem to a 1D biased random walk by using a calculated dependence of the entropy of the chain on the number of monomers that have passed through the nanopore

3 Simulation examples

In this section we discuss nine current problems in the fields of separation science and electrophoresis We examine the systems by focusing on the type of simulation methods that are being used to study them The choice of a numerical model is directly related to the question being asked For example coarse-grained methods are ideal for generic investigations of the basic mechanisms leading to separation while microscopic methods may be required when more specific problems are to be solved To give the reader a broad view of the importance of simulations in our field we have selected a wide range of problems

3.1 Electro-osmotic flow

Electro-osmotic flow (EOF) is ubiquitous in electrophoresis EOF occurs when there is an electric field with a tangential component at a charged surface in contact with a liquid The mobile counterions which make up the Debye layer next to the surface viscously drag the rest of the fluid Although it is sometimes possible to solve for the fluid velocity profile analytically [118] computational modeling is often used to garner a better understanding of EOF Let us examine the simple example seen in Fig 4 where we can see the results of some MD simulations which are described in detail by Tessier and Slater [119] The solid line shows the net charge distribution which has a maximum next to the wall followed by an exponential decay as predicted by Debye-Huckel theory The slight bump in this curve is the result of packing of the water beads near a fixed corrugated wall This effect is even more evident if one looks at the individual ion density profiles The fluid velocity profile (dotted curve) changes slope in the region where there is a net charge but takes on a bulk value outside of the thin charged region near the wall

For microfluidic devices with complex geometries the EOF profile is often non-trivial to find For the entropic trapping device developed by Han and Craighead [120] simulations were done using DPD with a slip boundary condition to reproduce a realistic EOF profile in order to include its effects in more complex simulations [121] Similarly it has been shown that the LB method can also be used to effectively model EOF [122] The problem of EOF in complex geometries has also been investigated by solving the Navier-Stokes equation numerically [6] In fact the only fluid model presented in this review that has not been used for modeling EOF to our knowledge is the recently developed SRD algorithm

In the context of electrophoretic separation the presence of EOF often has a deleterious effect on the resolution For example the EOF increases dispersion in capillary electrophoresis because the EOF is non-uniform due to the non-uniform charge distribution on the wall [5 119 123] For this reason polymer coatings are often used to quench the EOF They have the additional benefit of preventing wall-analyte interactions which cause additional dispersion in the system [124] Simulations of polymer coatings tend to use a system size which is thicker than both the Debye layer and the polymer coating Beyond

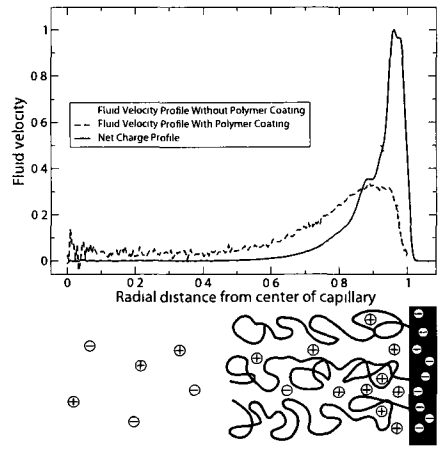


Figure 4 The radial profiles of the fluid velocity from MD simulations with and without a polymer coating as well as the net charge density profile The fluid velocities have been normalized by the bulk fluid speed in the coating-free case while the charge density has been normalized by its maximum value The radial distance is in units of the inner radius of the tube The fluid density is $0\,8\sigma^{-3}$ while the bulk and surface charge densities are $0\,02\sigma^{-3}$ and $0\,1\sigma^{-2}$ respectively The coating has a grafting density of $0\,05\sigma^{-2}$ and degree of polymerization $N = 20$ Further details can be found in [119 125] The schematic representation of the system is roughly aligned with the graph

the counter-ions and polymer coating the fluid velocity profile reaches a plateau The fluid speed in this plateau region (referred to as the bulk speed) is the same regardless of the system size and thus these miniature (often nano-scale) simulation systems provide realistic models for much larger experimentally relevant systems

Simulations of EOF in the presence of a polymer coating are fairly recent due to the high computational overhead involved in simulating them Recent numerical investigations [119 125] have been able to reproduce the scaling predictions of Harden *et al* [126] for grafted polymer coatings using coarse-grained Molecular Dynamics (MD) simulations These studies looked in particular at the case of when the polymer coating is thicker than the Debye length

and looked at two regimes the mushroom (isolated chains) and brush (high grafting density) In both regimes the MD simulations were able to confirm some of the predicted scaling behaviours of the bulk EOF with respect to the properties of the polymer layer such as the scaling with respect to the degree of polymerization N and grafting density The dashed line in Fig 4 shows a simulation where a polymer coating of length $N = 20$ beads and grafting density of $0\,05\sigma^{\,2}$ is used [here σ is the bead size in the WCA potential Eq (4)] Even these relatively short polymers clearly quench the majority of the EOF in the bulk of the fluid (note that in experiments the thin region near the wall where flow is generated makes up only a very small fraction of the total system size) Simulations by Qiao and He [127] using the DPD algorithm investigated the same situation showing interesting non-linearities in the EOF as due to dynamic coupling between the polymer s conformation and the fluid velocity profile The simulations also confirmed the fluid velocity profile as a function of the distance from the wall for a quenched polymer brush

More detailed atomistic simulations by Qiao [128] have shown that for cases where the Debye layer is on the same scale as the polymer layer more complex behaviour can result They found that at low grafting densities hydrophilic polymer coatings can actually increase the thickness of the Debye layer and thus the potential difference between the wall and the bulk fluid (termed the zeta potential) which increases the bulk EOF This effect was attributed to a reduction in the amount of water in the region of the polymer layer which caused the counterions to move further from the surface At higher grafting densities a larger suppression of EOF as the friction between the polymers and the fluid becomes larger was shown

3.2 Free-flow electrophoresis

Free-flow electrophoresis is widely used to separate and characterize biomolecules When a polyelectrolyte in gel-free solution is subject to a constant electric field its average drift velocity depends not only on the applied field but also on hydrodynamic interactions between the polyelectrolyte the counterions and the solvent Because the interrelation between the different forces is quite complex free-solution electrophoresis is not easily accessible to complete

analytical treatments Results do exist for long-chain limits where certain simplifications are applicable [129–131] but experimental evidence indicates that these models are not sufficient to explain short length-scale behaviour [132–134]

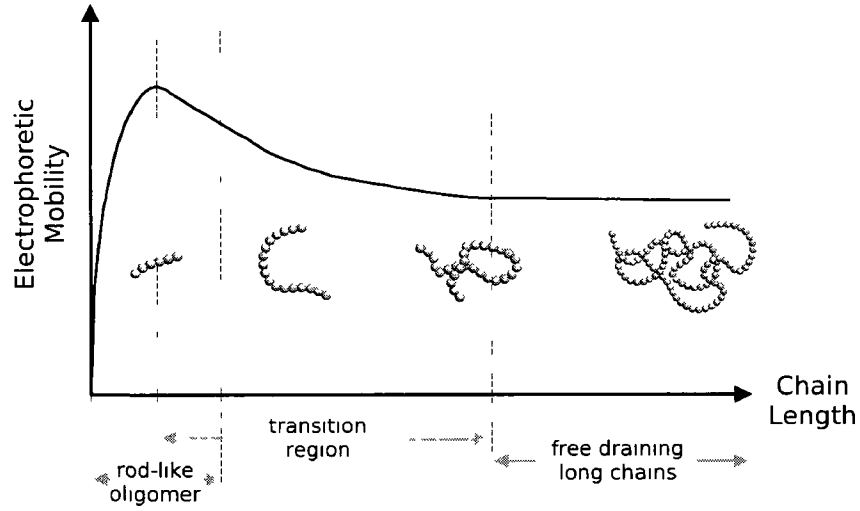


Figure 5 Schematic of the dependence of the electrophoretic mobility on the length of a flexible polyelectrolyte in free solution Short rod-like oligomers show a sublinear increase in electrophoretic mobility with chain length The mobility approaches a constant for long globular chains in the free draining regime Between these two behaviours lies a transition region in which both counterion and hydrodynamic effects must be accounted for

In the schematic of the free-solution electrophoresis of flexible polyelectrolytes (Fig 5) we see that the mobility is a function of length that approaches a constant with increased chain length so that separation of longer macromolecules by electrophoresis is not possible This limit called the free draining regime is well described by analytical methods [129–131]

The behaviour of short chains exhibiting not only length dependence but a nonmonotonic behaviour in the transition from oligomers to long flexible chains is not adequately

described by current theoretical approaches Here modern simulation methods introduced in Sec 2 of this paper provide much insight The ability to coarse grain certain interactions more than others facilitates probing different aspects of the behaviour one at a time In particular fully atomistic MD (Sec 2 1) can look at small oligomers thereby focusing on chemical details and has been employed to accurately describe the dynamic behaviour of short (3 and 6 units) fragments of ssRNA [135] The diffusion coefficient (corrected for finite-size effects [136] and solvent viscosity) and the increase in electrophoretic mobility from 3 to 6 nucleotides are consistent with experimental results The simulations show the importance of the counterions in reducing the effective charge by transiently binding to the polyelectrolyte resulting in the sublinear increase of the mobility with chain length that we see in Figure 5 The fact that the mobility follows the increase in effective charge demonstrates how the hydrodynamic friction of short rod-like polyelectrolytes depends only weakly on the length of the chain

Longer-chain behaviour can only be understood by investigating the hydrodynamics of polyelectrolytes undergoing conformational changes from a rod-like to a globular state As the number of counterions bound to the polyelectrolyte increases with length in this midlength regime their contribution remains essential Two recent studies using mesoscopic techniques (SRD (Sec 2 3 2) and LB (Sec 2 3 3)) have investigated the transition region between short fragments and long chains in detail [137 138] They show that without hydrodynamic interactions mobility would actually decrease with length and approach a constant value for large molecular weights However by including hydrodynamic interactions simulations accurately reproduce the experimentally observed nonmonotonic behaviour of the mobility By determining the effective charge due to counterion binding estimates of the effective friction s dependence on length were determined A transition from logarithmic to linear friction scaling was observed This change can be attributed to the correlated movement of the counterions in the vicinity of the polyelectrolyte effectively canceling long range hydrodynamic interaction this signals the transition to a free draining regime

When the chain length is increased further screening eliminates the need to explicitly account for hydrodynamic effects In this case implicit fluid techniques without hydrodynamics (Sec 2 4) but with explicitly included counterions have been applied to study

polyelectrolytes in electric fields [139 140] It should be noted here that in order to create at least microscale motion in the nanosecond timescales of simulations high electric field strengths that typically exceed the fields in experiments by several orders of magnitude must be used However below a critical field strength the static and dynamic properties of the polyelectrolyte remain unaffected and continue to agree with experimental data However when using even larger fields alignment of the polyelectrolyte with the field and an increased electrophoretic mobility due to dissociation of counterions is observed It is not clear if these effects are of any practical relevance

The above works emphasize the importance of correlations between counterions the polyelectrolyte and the resulting screening of hydrodynamic interactions Unlike the electrophoretic motion the diffusive motion of the polyelectrolyte is not correlated with counterion motion and hydrodynamic interactions remain unscreened [138 141] It has been shown that the diffusion of long polyelectrolytes can be correctly modeled when the counterions are neglected as long as hydrodynamic effects are included [142] In this simulation the Rotne-Prager-Yamakawa formulation (Sec 2 4 2) was used to quantitatively predict equilibrium and non-equilibrium diffusivity of DNA molecules up to 126μm in length

3.3 Polymer-obstacle collisions

In many electrophoretic methods size selectivity is due to the interaction between the analyte and obstacles of some sort To better understand this interaction Deutsch and Madden [143 144] have done pioneering 2D BD simulations of a polymer migrating through an ordered matrix of obstacles These authors found that in a very strong field the polymer migrated through the gel in an unexpected fashion very different from the conventional reptation picture [98] The chain goes periodically through a sequence of states a coil collides with a post extends its arms around it slides off the post leaving it in a fully extended state and then collapses into a coil again This process termed *geometration* was later observed in videomicroscopy experiments [145 146] Collision with a post is an important part of geometration obviously this process can be studied computationally in more detail if a system with just a single obstacle is considered

Nixon and Slater [147] did the first such computational study using the BD approach in 2D They considered a chain of beads (without excluded volume) connected by a FENE-like spring inside a narrow channel with an obstacle The field was assumed to be uniform i e the field lines penetrate the obstacle The chain started in the random coil state As the collision begins the coil gets deformed and becomes pancake-shaped The resulting behaviour of the polymer is similar to that observed by Deutsch and Madden but more clearly seen as other neighboring obstacles do not interfere Based on these simulations the authors developed an analytical theory predicting both the average retardation of the polymer due to the collision and the variance of this retardation

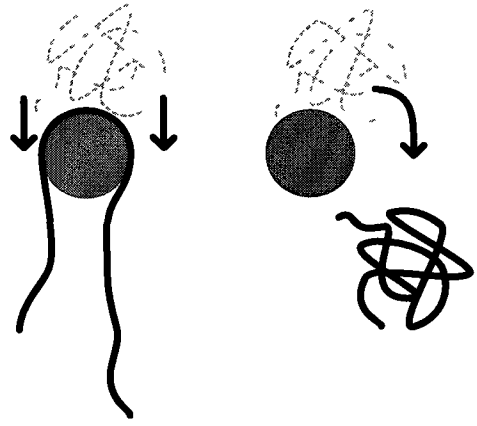


Figure 6 Different types of collisions of a chain with an obstacle a collision involving hooking of the chain upon the obstacle (left) and a rolling off collision during which the chain remains a coil (right)

More detailed simulation studies of collisions using a very similar 2D BD approach were carried out by Sevick and Williams [148] and Saville and Sevick [149] The authors studied the properties of collisions as a function of the *impact parameter b* the initial distance in the direction perpendicular to the field between the center of mass of the chain and the center of the obstacle Different obstacle sizes were also considered The authors found that besides the hooking collisions studied by Nixon and Slater another possibility is the rolling off collisions where the coil 'rolls over the surface of the obstacle and does not

deform much during the collision (Fig 6) For rolling off collisions that become dominant for large obstacles the duration of the collision depends mostly on the obstacle size rather than the chain length Among hooking collisions besides the conventional type (called U/J collisions because the hooked chain conformation resembles these letters) collisions with multiple hooking (later termed W collisions [150] as the conformation can resemble a double-U) were also found

More recently Randall and Doyle [151] found in their fluorescence microscopy experiments that besides U/J and W collisions yet another type of collision (called X for extending) is possible These collisions resemble the U/J type but the longer arm of the chain is not fully unwound at the beginning of the unhooking process containing a coil at the end which unwinds gradually as the unhooking proceeds To study this collision type in detail Kim and Doyle [152] carried out 3D BD simulations with excluded volume interactions Unlike previous work they did not make the assumption of a uniform electric field instead calculating the field assuming that the obstacle (cylindrical in shape) is a perfect insulator They indeed found X collisions along with previously known U/J and W types It turned out that surprisingly the X type is the most dominant one in a broad range of chain lengths and field strengths Given that this remains the case even in simpler models perhaps previous authors simply did not distinguish between completely and incompletely unwound chains

The work discussed so far deals with the case of strong fields when the thermal effects are negligible or at least secondary A very recent paper by Holleran and Larson [153] also considers the case when on the contrary the field is weak and thermal diffusion dominates In this case the arms are never extended and the polymer remains a coil that drifts slowly past the obstacle In fact in this regime the polymer at least semiquantitatively can be considered as a rigid particle and be studied using e g the exact MC methods described in Sec 2 6 1 The paper is also interesting because of its use of a novel computational model of the chain (developed by the authors and described in a separate publication [154]) where the springs connecting the beads and not the beads themselves are repelled by the obstacles which ensures that the chain cannot penetrate the obstacle even when the distance between the beads is much larger than the obstacle size This allows the authors to treat very long

chains (longer than 1000 Kuhn lengths)

All simulations discussed so far used Brownian Dynamics neglecting Coulomb interactions between monomers and hydrodynamic interactions Coulomb interactions are screened by counterions The justification for neglecting HI offered by Kim and Doyle [152] is that both their simulation and the experiment it models [151] were carried out in a slit in which case HI should be screened However even in the bulk the results are at least qualitatively correct as the comparison with recent simulations by Kenward and Slater [155] using *explicit solvent* shows Introducing a solvent can certainly produce some quantitative changes Neglecting HI assumes that the polymer is free-draining which is true in free solvent but not when the chain is slowed down by an obstacle The result is the modification of the friction force on the chain which also becomes conformation-dependent Kenward and Slater also studied collisions between two polymer chains of which only one is driven by an external force but both are mobile Such collisions are important in the case of electrophoresis in polymer solutions [156 157] In this case hydrodynamic effects influence the conformations of the chains Likewise studying situations where the colliding chain is driven by a flow rather than an external field is of course only possible when the solvent is included in some way, Kenward and Slater considered this case as well

Finally we mention other computational approaches applied to this problem Starkweather Muthukumar and Hoagland did off-lattice bead-rod MC simulations of a chain colliding with an immobile random coil [158] and a mobile chain [159] MC methods often have problems in strong fields and their use is especially dangerous when applied to this problem as it involves U-shaped chain configurations The authors had to restrict themselves to moderate field strengths when the chain is far from being fully stretched Andre *et al* [150] used a special algorithm tracking the evolution of different loops and arms of the chain after the impact

3.4 Ogston: Modeling sieving in hydrogels

A specific electrophoretic regime exists when the size of the analyte is smaller than or comparable to the mean pore size of the gel This regime is often called the Ogston regime for electrophoretic sieving Although the concept is technically restricted to rigid analytes. it is possible to extend its use to flexible polyelectrolytes such as DNA if one assumes that the chain takes on a spherical conformation with an effective radius R. By coarse-graining out the fine details of the individual monomers and considering the analyte as a solid sphere. we can discretize our system on a lattice (Fig. 7). and use simulations or exact calculation methods to study the electrophoretic mobility or diffusion coefficient of the analyte in this regime. A straightforward lattice approach to modeling the gel system would be to consider

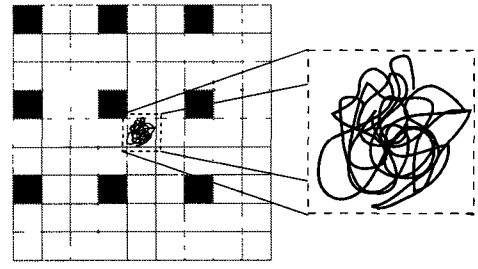


Figure 7: Modeling a polyelectrolyte in the Ogston regime with a Lattice Monte Carlo model

a gel fibre as an impenetrable obstacle. In Fig. 7. the analyte and the obstacles are of the same size (the lattice step size); although this is the case that we will consider below it is equally easy to study larger obstacles and/or larger analytes that occupy more than one lattice unit (more on this later).

The first such numerical model was used by Slater and Guo [160] to test the key hypothesis of the Ogston-Morris-Rodbard-Chrambach (OMRC) model According to the OMRC model. the mobility of the analyte in this regime is linearly proportional to the fractional gel volume that it can occupy. a purely geometric parameter that can be computed quite easily for the model shown in Fig. 7 All the other assumptions of the OMRC model (e g a low field intensity) being compatible with the numerical model. the results of the

study represented a direct test of the fractional volume hypothesis The Slater and Guo exact numerical calculations showed that the mobility of the analyte is higher in an ordered gel compared to a random one even if the fractional volume is the same This was the first demonstration that the OMRC model is incomplete, in fact these authors also showed that the OMRC model corresponds to a mean-field model valid for an annealed gel (a gel with rapidly moving obstacles) [160]

As mentioned earlier it is possible to extend the exact method to treat larger particle [161] In this case molecules are also viewed as rigid spherical particles but can be larger than the obstacles These results are valid in the zero-field limit since the interactions with the obstacles are assumed to be hard-core i e the particles do not deform when colliding with an obstacle It is also possible to extend this calculation method to treat attractive interactions between the analyte and the gel structure [162]

DNA molecules can also be modeled using the MC exact calculation method without making the hard sphere approximation [116] Indeed one can use multiple particles linked by bonds to represent a flexible chain For example such polymers can be described by self-avoiding walks and bond-fluctuation algorithms This extension opens the door to a fundamental study of the electrophoretic sieving of oligomers rod-like molecules vesicles star-shaped macromolecules etc One could argue that the MC exact calculation model is not a sufficiently good representation of a gel matrix since the field is assumed to be uniform throughout the gel In reality the field lines are affected by the gel structure (the obstacles) It has been shown [163] that the lattice MC exact method can also be extended to treat spatial variations of the electric field and that this has little impact on the results at low field intensity

It is possible to apply the exact method to treat high fields instead of vanishingly small external fields [100] For non-deformable analytes this can reproduce the trapping than sometimes occurs in real electrophoresis experiments Multiple obstacle geometries have been studied and it is possible to properly model experimental observations of trapping and pulsed field de-trapping A modified version of the initial algorithm [164] also allows the simultaneous calculation of the mobility and of the dispersion coefficient, this is obviously

needed in order to predict the resolution of a specific device

3.5 Microfluidic ratchets

The Brownian motion of particles is what gives rise to diffusion It is possible to exploit these natural thermodynamic fluctuations for the separation of particles by adding an external force which biases the dynamics Brownian ratchets are devices that use an asymmetry either temporal or spatial to drive the motion of a Brownian particle even when the net external force is zero For example a temporal asymmetry could be a zero-mean field alternating between a short high-intensity forward pulse and a longer low-intensity backward pulse A spatial asymmetry could take the form of asymmetrically shaped obstacles or walls Using any or all of these types of asymmetry a non-zero net velocity can observed in the presence of a field even if the net force is zero (hence the name *ratchet*) The random motion of the particle then plays a major role (this is why it is called a *Brownian ratchet*)

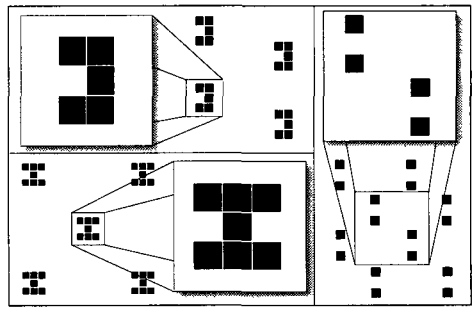


Figure 8 Three types of obstacles that can be used to design a Brownian ratchet separation system *Left, top* Trap-shaped obstacles with a left-right spatial asymmetry An unbiased AC electric field would lead to a net velocity pointing in the negative direction *Left, bottom* Symmetric obstacles with traps in both directions Here one would need an asymmetric pulsed field to drive the Brownian ratchet *Right* The distribution of obstacles proposed by Gauthier and Slater [165] misaligned rows of obstacles The lattices are not explicitly shown for clarity purpose

As an example consider the case of particles in a properly designed microfabricated array of obstacles (see Fig 8) Lattice Monte Carlo methods can be used to simulate Brownian ratchets of this kind (see Section 2 6) First the system is represented as an array of impenetrable obstacles A starting point is chosen and the particle is moved randomly with the probabilities defined in Eq (16) With this method long computational times would be required in order to obtain statistically precise results A faster and more precise computational method was described in Section 2 6 The increased accuracy is vital at very low fields where the statistical uncertainty associated with the normal approach is often larger than the studied behaviour

Gauthier and Slater [165] examined several ratchet systems using this simulation approach The system shown on the right of Fig 8 is interesting since it uses a symmetric array of obstacles together with an asymmetric pulsed field These authors observed several current reversals (changes in direction) for different parameters of the field For certain choices of parameter particle of different sizes but having the same charge were predicted to move in opposite directions a remarkable result (in some cases a particle can even move against the direction of the net field a phenomenon called absolute negative mobility) A Brownian ratchet of this type was later built experimentally by another group [166] An aqueous solution of charged spheres of radius 2 μm was placed in a periodic array of misaligned posts with alternating small and large gaps The applied asymmetric external field was a superposition of a constant (E_{DC}) and an alternating signal ($\pm E_{AC}$) The system behaved as expected and absolute negative mobility was observed a nice example of the kind of new ideas that simulations can suggest

Tessier *et al* [167] [168] studied the system proposed by Han *et al* [169] in the ratchet regime both spatial and temporal For the spatial asymmetry the system was modified to introduce a geometrical asymmetry The bond-fluctuation algorithm (see Sec 2 6) was used to model the polymer Tessier *et al* also simulated the system using a zero integrated pulsed field [170] In both cases separation was predicted These predictions have yet to be tested

3.6 Nanopore translocation

By threading ssDNA through a narrow pore and identifying the bases as they pass through, nanopores offer a promising avenue for the development of sequencing technologies [171]. In contrast to gel electrophoresis, nanopore sequencing could offer rapid (thousands of bases per second) sequencing of a single DNA molecule; a revolution which would have a great impact in fields associated with the life sciences. The same approach can also be used for other purposes, such as sizing molecular contour lengths. A significant advancement occurred in 1996 when Kasianowicz et al. demonstrated that RNA and DNA could be detected passing through a biological nanopore (α-hemolysin) by monitoring the disruption of ionic currents [172]. Subsequently, there have been a great number of theoretical [173–176] and experimental studies [177–182] focused on nanopore translocation (note that the number of nanopore related publications is staggering and in this very brief review, we are limited to providing only a few selected publications for each topic). Providing a bridge between theory and experiment, many computer simulations have also been performed and, in fact, most of the techniques discussed in section 2 of this paper have been used to study the translocation process.

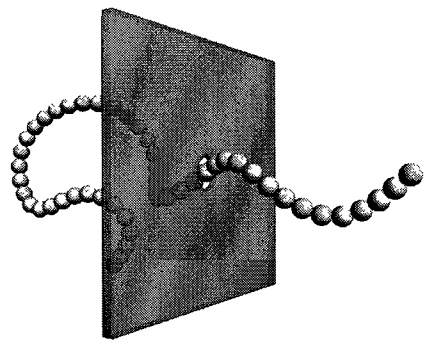


Figure 9: Schematic of a polymer translocating through a solid state nanopore.

At the coarsest level, lattice MC simulations (see Section 3.5.2) are often performed in conjunction with theoretical studies to test the resulting predictions [175, 176, 183–187]. Other studies have used off-lattice MC to study translocation driven by an external or

adsorption force [188–190] Finally one approach has mapped the translocation onto a one dimensional diffusion process and then employed an exact numerical technique to obtain results [191 192] While such studies are able to investigate very long polymers at low fields the details of the dynamics are not produced and additionally the model is coarse to the extent that effects such as hydrodynamic interactions can be included only in an approximate manner At the opposite end of the spectrum fully atomistic MD simulations of DNA inside a channel have also been performed These studies have revealed interesting details about the viability of distinguishing between bases of DNA translocating through a synthetic nanopore [1] and the orientation dependence of DNA inside the biological α-hemolysin channel [193] They are however limited in terms of time scale and are unable to simulate the full translocation process In between these extremes many coarse-grained *MD simulations using an implicit fluid* [194–196] *an explicit fluid* [197 199] *or a mesoscopic* fluid model [200 203] have been performed in an attempt to include the critical factors while requiring a reasonable level of computational power

Using one or more these approaches these studies have investigated the role of underlying physical mechanisms such as conformational entropy [192] hydrodynamic interactions [200] the solvent effect [204] counter ions [199] pore-polymer interactions [205] and the pore geometry [198 206] Additionally the dependence on the nature of the polymer has been studied by simulating charged polymers [207] and heteropolymers [208 209] Further considering application to sequencing it is not surprising that a great number of studies have also examined driven translocation by implementing a pulling force [210] an adsorption force [190] or an external field (discussed below) As is obvious from this list which is in itself incomplete there is not enough space in this brief review to cover all of these results Rather as a single example we will discuss some of the results for a key aspect of the general translocation problem the scaling of the translocation time τ with the degree of polymerization N of the polymer for driven translocation

In experiments on the biological α-hemolysin pore both Kasianowicz et al and Meller et al found a linear dependence of the translocation time on the polymer length ($\tau \sim N$) *for short ssDNA fragments* [172 182] *In contrast the experiments of Storm et al* found a scaling of $\tau \sim N^{1.27}$ when driving long DNA strands (6500 to 97000 base pairs)

through a solid-state nanopore [177] This result is in agreement with their prediction that $\tau \sim N^{2\nu}$ for long polymers when HI are taken into account ($\nu \cong \frac{3}{5}$ is the three-dimensional Flory exponent) From the analytical side considering translocation driven by a chemical potential gradient Muthukumar predicted linear scaling [183] while Kantor and Kardar predicted $\tau \sim N^{1+\nu}$ [185] The latter group also performed bond-fluctuation MC simulations but were unable to verify the prediction due to limited polymer lengths However additional simulation studies for long polymers have found a scaling of $\tau \sim N^{1+\nu}$ using 2D fluctuating bond MC [211] and using the exact numerical method [192] Finally Dubbeldam et al predicted a scaling of $\tau \sim N^{\frac{2}{2\nu+2-\gamma_1}}$ (where γ_1 is the surface exponent) [212] and found results consistent with this from off-lattice MC simulations Moving towards a more detailed simulation of the dynamics driven translocation has also been studied using Langevin Dynamics and Brownian Dynamics simulations Performing 2D LD simulations one study found a scaling of $\tau \sim N^{1+\nu}$ [196] - a results consistent with the Kantor and Kardar prediction and the MC simulations cited above Meanwhile others have found a linear dependence [194 195] The discrepancies between these results may be attributed to differences in the particular system setups such as the pore length polymer length magnitude of the external field and polymer model

A limitation of all the simulation results cited thus far (MC and LD/BD) is that they neglect hydrodynamic interactions As long ranged correlations through the fluid are conceivably important in the translocation process much of the current work focuses on MD simulations using mesoscopic fluid models Using the DPD approach He et al have found a linear dependence of τ on N [202] Meanwhile Izmith et al [200] and Fyta et al [201] each using a Lattice Boltzmann fluid model but with different polymer models have both found an exponent of 1 28 a result in good agreement with the experimental data of Storm et al corresponding to $\tau \sim N^{2\nu}$ Both of these studies also directly tested the impact of hydrodynamic interactions by performing the same simulations without the LB solvent Izmith et al found a negligible change as the exponent rose to 1 31 while Fyta et al observed a slightly larger effect with the exponent rising to 1 36 The key parameter here appears to be the molecular size is there a critical polymer size below which the hydrodynamic interactions are negligible because the translocation is then dominated by the polymer-pore

interaction and not by the polymer-fluid interaction? In our opinion a combination of careful simulations and experimental studies will be needed to answer this question

As demonstrated by this one example a wide range of simulation techniques have been employed to elucidate various details of the translocation process from MC simulations testing analytical results to coarse-grained MD simulations revealing details of the translocation dynamics all the way to atomic scale MD simulations probing the precise interaction between the DNA and the pore In fact given this wide array one must keep in mind the scope of a particular simulation approach and the limitations of a chosen system setup when considering the results of each study Ultimately, however this diverse amount of information is an advantage in fully characterizing the system especially since the scientist has full control of the variables in a computer simulation (which is not the case in a laboratory!) Hence computer simulations are not only proving to be an invaluable tool in the investigation of the translocation process but promise to be vital in the future design of sequencing devices

3.7 Entropic trapping

In Ogston-type models of electrophoresis it is assumed that analytes cannot pass through constrictions that are smaller than their diameter This is true for hard particles but polymer coils can deform and still pass through holes that are much smaller than their radius of gyration R_g However as deformed coils are no longer completely random this has an entropy cost and a polymer entering a narrow space has to overcome an *entropic barrier* In a porous medium such as a gel such barriers are especially important when the average pore size is comparable to R_g so there are pores both smaller and larger than R_g In this situation *entropic trapping* is possible in which case the polymer can stay trapped for a long time in a large pore since all ways out of it involve overcoming entropic barriers The consequences of this were first studied in off-lattice MC simulations (Sec 2 6 2) by Baumgartner and Muthukumar [213 215], one interesting result is a much stronger size dependence of the diffusion coefficient (and thus via the Nernst-Einstein relation [Sec 2 6 1] of the electrophoretic mobility) than predicted for larger polymers spanning many pores (the reptation regime) This was later confirmed experimentally [216]

Entropic trapping in gels, while an important issue, is still just one of the factors influencing the separation. On the other hand, Han *et al.* [120.169.217.218] have fabricated and studied a device where, by design, entropic trapping is the dominant contribution to separation. The device is an array of cavities separated by long and narrow *slits*. The size of the cavities is much larger than R_g for the typical DNA sizes whose separation is desired, whereas the width of the slit is much smaller than R_g in one direction, but, importantly, is still $\gg R_g$ in the other direction. Han *et al.* found that the mobility *increases* as the size of the polymer increases. This is rather counterintuitive, given that larger polymers should deform more passing through the slit and this should be more entropically costly. Han *et al.* explained this by suggesting that rather than entering the slit as a whole, the polymer stays around the slit and loops (or *hernias*) get inserted in the slit (Fig. 10). Such insertion has an entropy cost proportional to the insertion length, but it also causes the decrease of the electrostatic energy proportional to the square of the length. As a result, the free energy increases until the top of the free energy barrier is reached, but then starts decreasing. The escape rate, as always in transition-state theory, depends exponentially on the barrier height, but also depends on the prefactor (the attempt frequency). It was argued that the barrier height is inversely proportional to the field strength, but is independent of the polymer size, and so the separation is entirely determined by the prefactor. This prefactor should be proportional to the size of the part of the polymer exposed to the slit, as this determines the number of hernias that can form simultaneously; this size is proportional to R_g which increases with the polymer size, and therefore the escape rate indeed increases with the size, as observed experimentally.

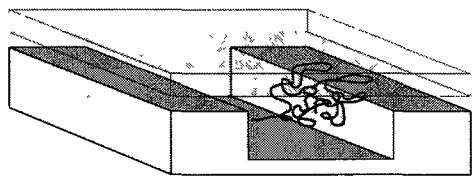


Figure 10: Schematic drawing of a part of the entropic trap device of Han *et al.* with a DNA chain near the entrance to a narrow slit. Several hernias are inserted in the slit

This simple theory, while appealing, was in need of verification by simulations, es-

pecially given that it relied on the independence of the free energy barrier height of the polymer size, since this barrier height enters in the exponent even a slight dependence may completely overwhelm the dependence contained in the prefactor With this in mind several simulations have been carried out Tessier *et al* [167] used the MC Bond-Fluctuation Algorithm (Sec 2 6 2) The field used in the simulation was computed by numerically solving the Laplace equation The results for the mobility obtained in the simulations are in qualitative agreement with the experiments Overall the simulations confirmed the theory by Han *et al* but also further refined it In particular it was found that for weaker fields the mean trapping time indeed depends exponentially on the inverse field with the slope on the semilogarithmic plot independent of the molecular size confirming that the activation energy is indeed size-independent On the other hand for longer chains there was a deviation from the perfect exponential at higher fields attributed to the change in the shape of the coil near the entrance to the slit At moderate fields the coil around the slit acquires a pancake shape and the radius of gyration behaves as that of a *2D* random walk, but as the field gets stronger the escape into the slit is so fast that the coil has no time to deform This has obvious consequences for the escape rate as the prefactor depends on the extent of the coil along the slit entrance The field dependence of the critical hernia length as estimated by Tessier *et al* is also largely in agreement with the theory

Chen and Escobedo [219] looked more directly at the free energy barrier associated with the entrance into the slit Since free energy calculations are done in equilibrium one of the equilibrium MC methods (in their case configuration-bias MC [220]) was used (see Sec 2 6) The advantage of this approach is that reliable results can be obtained even for very small fields when the escape is so slow that good statistics cannot be obtained in dynamical simulations The result of these free energy calculations is that over a large range of parameters the barrier height is indeed size-independent however deviations are observed at very low fields for very short chains (the regime that Tessier *et al* could not study very carefully) These deviations are not unexpected at very low fields the critical hernia length is very large and if the chain is short may actually become comparable to or even (in theory) exceed the chain length

Streek *et al* [221] used Brownian Dynamics simulations (Sec 2 4) to study essentially

the same system Based on their results these authors suggest another separation mechanism due to size-dependent trapping of chains in the corners of the cavity The authors claim that it is this mechanism rather than entropic trapping which determines the mobility This may well be true at the rather high fields that they consider when the entropic barriers are presumably essentially negligible On the other hand it is clear that at very low fields the entropic trapping mechanism should dominate as no mechanism based on diffusion alone without any activation barriers involved would be able to compete The intermediate case when the entropic barrier height is $\sim k_BT$ is the most interesting practically and deserves further study, the answer may depend on many details such as the size of the cavities

We should also mention newer work likewise using Brownian dynamics Panwar and Kumar [222] recognized that when the trapping barrier is not very high two other time scales besides the trapping time contribute the time it takes the DNA to approach the constriction and the crossing time They studied the field and chain length dependences of all three times These authors used a bead rod model of the DNA Lee and Joo [223] did a similar study for a bead-spring chain

All work described here uses methods that do not take hydrodynamic effects into account For this reason the DNA behaves as a free-draining chain when in reality a trapped chain is not free-draining As Tessier *et al* [167] point out this implies that the field intensity needed to overcome the entropic barrier is underestimated by a factor of $\sim N^{2/5}$ where N is the chain length As always at this level of modeling quantitative comparisons with experiment are difficult because the effective charge of the DNA is different from the actual charge due to counterion condensation These effects need to be taken into account in future work

3.8 Surface electrophoresis

A novel electrophoretic separation technique based on the DNA s interaction with a surface was first reported by Pernodet et al [224] in 2000 By adsorbing DNA to a surface length dependent separation on a flat surface without any restrictions or any sieving matrices was

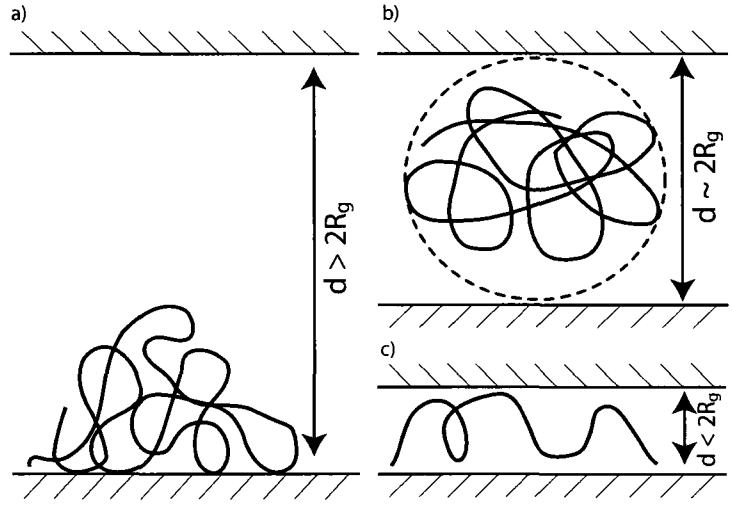


Figure 11 Different regimes of confinement used in some gel-free separation techniques a) weak confinement ($d > 2R_g$) in the presence of an attractive surface b) stronger confinement ($d \sim 2R_g$) that starts to influence the chain conformations and c) strong confinement ($d < 2R_g$) in which the chain conformations are determined by the walls of the channels

achieved (see Fig 11a) It was found that the interactions between the molecule and the substrate essentially act as a length dependent source of friction enabling electrophoretic separation The initial experimental observations were accompanied by Molecular Dynamics simulations [225 226] and have been followed up by further studies [227 228] under different conditions The results showed that the DNA surface interaction is a key parameter for the process a strong attraction leaves the molecules fully adsorbed and no separation is possible too weak an attraction lets molecules desorb and resume bulk behaviour where likewise no separation is possible Additionally the interaction can be noticeably influenced by choosing a specially patterned surface [229 230] an exciting and unique approach to designing optimized and custom-made separation systems

Since the exact nature of the interaction and the resulting separation mechanism remain elusive at this point there is a pressing need for more elaborate theoretical studies that include electrostatic and hydrodynamic effects alike as they are crucial when the molecules approach the surface This has been neglected so far

3.9 Confinement-driven separation

Recent progress in design and fabrication of microfluidic devices on a sub-micrometer length-scale [231 233] demands a good understanding of the statics and dynamics of the polyelectrolytes under steric confinement We can distinguish several regimes of confinement In a device that is much larger than the size of the polyelectrolytes given by its radius of gyration R_g the conformations are unperturbed and isotropic (weak confinement) Reducing the dimensions of the devices to the order of R_g the conformations of the polyelectrolytes start to become restricted by the walls and show deviations from the equilibrium (strong confinement see Fig 11b) With further reduction of the device size the polyelectrolyte becomes extremely restricted and the static and dynamic properties undergo significant changes [234 235] (Fig 11c)

The decrease in size of microfluidic devices used in actual experiments and the growth in size of the systems that are addressable by means of computer simulations - due to

advancement of simulation methods together with the increase of computer power - led to a cross-over creating systems that can be worked on from both sides Recent experiments in slit-like nano channels studied the static and dynamic properties of single molecules [236–242] and showed how confinement can be used as a tool to change polymer conformations as well as the dynamics through modulation of the hydrodynamic interactions Consequently the role of hydrodynamics in confinement has been the focus of several recent computer studies [70 91 242–248] The results indicate that under weak confinement the hydrodynamic interactions between polymer and wall drive the polymer to the center of the channel if an external flow is applied However in high confinement a migration towards the channel walls was observed Since this interesting physical phenomenon depends on the ratio between the size of the polymer and the width of the channel as well as on the strength of the driving force it seems to be an ideal candidate for free-solution separation of polyelectrolytes Recently reported measurements on the diffusion and the electrophoretic mobility of DNA in strongly confined systems [249 251] indicate a possible electrophoretic separation mechanism based on the modified dynamics in strong confinement A systematic simulation study of this subject has yet to be done in order to verify these results In particular electrostatic interactions and the influence of counterions on the hydrodynamic interactions have been neglected so far but they should be assumed to be of great importance if the length scales of the system become comparable to the Debye length below which electrostatic interactions are not fully shielded by the solvent [248 252]

4 Outlook

This review has hopefully convinced the reader that the computational approach has been useful in understanding a large variety of electrophoretic separation systems and relevant electrokinetic phenomena Compared to theory many situations that cannot be studied theoretically without gross simplifications can be treated computationally, in fact simulations provide valuable clues to theorists as to *exactly what* simplifications and assumptions they are allowed to make Compared to experiments simulations have an unmatched ability to look at the microscopic level, but perhaps the most important is the possibility to

switch on and off different effects (such as hydrodynamic interactions) at will something that is not available to experimentalists and yet is extremely helpful in developing a better understanding of the systems and phenomena of interest

Considering the future of both the simulation methods and their applications to separation phenomena the most straightforward approach is the brute-force one harnessing the inevitable increases in computational power to conduct more detailed simulations of larger systems for longer times However in striving for more realistic simulations in this manner it is important to recognize just how wide the gulf between experiments and detailed simulations is To see this consider atomistic MD simulations which are already being used to study DNA inside a nanopore (section 3 6) While these simulations are yielding interesting results significant increases in computational power would greatly enhance what they are able to study For example detailed atomistic simulations may be able to aid in the design of a probe to read the bases as they pass through the channel However current atomistic MD simulations are typically limited to trajectories of hundreds of nanoseconds for limited system sizes On the other hand in the lab the dynamics of interest can evolve on a time scale up to seconds or even longer Consequently while any increase in computational power is enthusiastically welcomed the gap between detailed simulation and experiment remains large and the development of efficient techniques to appropriately coarse-grain the simulations is equally vital to broadening the scope and relevancy of computational studies

Hence as a final thought to this review we examine a number of emerging or long-standing-but-difficult areas where new ideas and computer simulations will be needed over the next few years

Multiscale simulations In simulations often different regions of the system require different levels of detail For instance a molecular-level fluid model may be required in the vicinity of the analyte but at larger distances a faster mesoscopic approach may be sufficient Likewise the accuracy of a translocation simulation may be improved if the polymer is simulated atomistically in the vicinity of the nanopore To avoid simulating the whole system with the highest level of detail needed only in a small region multiscale simulations can be used The biggest challenge is the seamless coupling between regions with different

level of detail which is especially difficult if the region boundaries are required to move Several multiscale methods for fluids have appeared recently (see e g Refs [253–258]) Such methods should see increased use in electrophoresis simulations Besides more trivial *serial* multiscale studies where the outcome of a more detailed simulation is used to construct a coarse-grained model and vice versa [259] will be useful

Mesoscopic solvents SRD is a relatively new method of mesoscopically handling solvents and as such its full potential has yet to be captured Despite being a model that is conceptually and implementationally simple few of the electrophoretic simulation examples discussed above have taken advantage of SRD s ability to treat relatively large and multicomponent systems with complex and dynamic boundary conditions Furthermore careful choice of the multiparticle collision operator allows for hydrodynamics to be turned off and simply replaced with a Brownian heat bath in order to explore HI effects in the system

Hydrodynamic effects in gel electrophoresis The separation of small analytes by gel electrophoresis is often called sieving Small typically means that the size of the object is smaller than the mean pore size of the gel matrix The word sieving suggests that separation is related to steric interactions restricting the motion of the analyte Although detailed obstruction models have been proposed their validity is not clear because they ignore hydrodynamic interactions and do not properly treat electrostatic interactions Since we have three length scales here (the Debye length the particles radius and the mean pore size) many different regimes must exist Understanding the simultaneous screening of the electrostatic and hydrodynamic interactions in a sieving gel especially if the latter is an entangled polymer solution will require new ideas and probably extensive coarse-grained molecular dynamic simulations

Electrophoresis in polymer solutions Electrophoresis in capillaries and microchips using polymer solutions as sieving matrices is increasingly important Conventional wisdom is that well-entangled polymer solutions behave essentially like gels However the absence of crosslinking may be important especially in strong fields typically used for capillary electrophoresis Recent videomicroscopy experiments [260 261] seem to suggest that the DNA is able to drag matrix polymers even well above the entanglement concentration Simulations

are needed to help understand this process The consequences for the basic geometration mechanism of DNA motion can perhaps be understood even in rather simple models similar to that by Deutsch [143] but with draggable obstacles Electrophoretic motion in more dilute solutions around the critical entanglement concentration is of interest as well

Drag-tags for free-solution separation Attaching suitable molecules to DNA fragments can restore size-dependent mobility regardless of the free-draining property [262–264] The read length is optimized by choosing large but perfectly monodisperse drag-tags As it is an experimental challenge to produce such polymer labels two recently proposed alternatives seem promising Haynes et al [265] proposed to use branched polymers with well-defined architecture whereas Grosser et al [266 267] introduced nonionic surfactant micelles as drag-tags with very large hydrodynamic friction Computer simulations can be used to characterize these new labels and provide ideas on how to extend these approaches

Electro-osmotic flow As mentioned in Section 3 1 it has been shown that the mesoscopic fluid models such as SRD LB and DPD provide realistic EOF profiles It has also been shown that they provide realistic hydrodynamic interactions One would suspect that given that they do these two things accurately they will probably also be able to model EOF in the presence of polymer coatings If this is the case mesoscopic models could allow for the simulation of more complex coatings and a wider range of parameters due to the increase in speed of computation

DNA denaturation As a means of sequence dependent separation denaturing gradient gel electrophoresis (DGGE) exploits a rapid decrease in the mobility of a dsDNA fragment when sections of it denature It is currently unclear whether the experimentally observed blocking is an actual trapping or a steep reduction in mobility The bubble dynamics could play an important role in the blocking thus a static helical-coil configuration may or may not be enough to give the correct picture Unfortunately simulating accurate bubble dynamics sets a coarse graining length scale to the length of a single bp A MD model which could incorporate the correct breathing dynamics of a dsDNA in the presence of a gel while being able to treat dsDNA lengths of interest could be an important tool for investigating the blocking phenomenon

Separation of large DNA The idea of a Human Genome Project became realistic when pulsed field gel electrophoresis (PFGE) made it possible to separate DNA molecules as large as a few megabase pairs (Mbp) a necessary step for chromosome mapping and sample preparation After a few years of rapid progress PFGE saturated at about 5 Mbp Of course the physics of PFGE is expected to be complicated for molecules that are millimeters in contour length! Agarose gels with their sub-μm pores are probably not the best material to extend the usefulness of PFGE Recent ideas based on new separation concepts have yet to become commercial products Computer simulations played a key role in the development of PFGE and will again be needed in order to optimize the separation of such huge molecules and to design new and faster devices

Protein separations While the electrophoretic separation of nucleic acids small ions and spherical particles has been modeled extensively the same cannot be said of protein electrophoresis Proteins are complex molecules with non-trivial charge distributions on their surface They can be separated in their native state or denatured in free-flow electrophoresis or in gels under uniform conditions or in the presence of gradients of various kinds Although some models developed for particles or DNA can possibly be used for proteins as well this has never been carefully tested Computer simulations will be required to optimize the separation of proteins which remains a difficult issue in the laboratory However generic simulation methods are not likely to work well since the precise shape and charge distribution that characterize a specific protein must be taken into account

In conclusion as computers become more powerful and new algorithms are developed the future of computational studies of separation phenomena looks even brighter!

Acknowledgments

We are grateful to Dr Marcello Sega for useful discussions during the process of writing this review This research was supported by grants from AFMnet NSERC and the NIH (Grant No 2 R01 HG001970-07 through Stanford University) The findings opinions and

recommendations expressed in this letter are those of the authors and not necessarily those of Stanford University or the NIH Funds from the Volkswagen Foundation the DAAD and DFG under Grant No TR6 are also gratefully acknowledged OAH and TNS would like to thank the NSERC-CGS program and the University of Ottawa for financial support

A Molecular Dynamics Simulation Packages

While the theory behind an MD simulation is conceptually straightforward in practice it can be a large undertaking to code from scratch This is particularly true when one wishes to implement more involved algorithms in order to efficiently calculate long-ranged electrostatic interactions or incorporate mesoscopic fluid models For this reason there exists a multitude of simulation packages in which the routines necessary for performing the simulation have already been implemented Hence the user can simply supply input information pertaining to their system select simulation features and parameters and then use the packages to execute the simulation produce trajectory files and frequently perform analysis Thus to aid the interested reader we present in this appendix a table of the more prominent MD simulation packages (please note that the list given here is by no means exhaustive [268])

In addition to providing the simulation package name (and related references) we also include some details for each package The information given here is meant to convey what the package is most often used for and also to indicate any special features unique to this package These entries should not be interpreted as restrictive but rather as highlighting interesting or distinct features among a group of softwares which all accomplish similar goals

For example consider delineating the packages between atomistic or coarse-grained simulations In principle all the cited packages are capable of performing fully atomistic simulations But as such simulations require a fully atomistic forcefield it is easiest to start with a package that either comes with some forcefields implemented (the first four packages in the list) or at least is designed to be compatible with the forcefields from other packages (the next two entries) Likewise all of the packages could be used for coarse-grained simulations In fact due to it s efficiency in calculating non-bonded interactions GROMACS

a biomolecule oriented package has been used for coarse-grained polymer simulation work However in choosing a more coarse-grained oriented package one is more likely to find other useful elements such as the implementation of mesoscopic fluid models as found in the last two entries

Finally we include a Free column A check mark here indicates that the program is free for academic use (at a minimum) This column is included to encourage the interested reader to download and begin experimenting without any monetary investment

Name	Details	Free
GROMACS [269] www gromacs org	- Includes forcefields for fully atomistic - Efficient calculation of non-bonded interaction	✓
CHARMM [270] www charmm org	- Includes forcefields for fully atomistic - Pioneer for MD simulations	
NAMD [271] www ks uiuc edu/Research/namd/	- Includes forcefields for fully atomistic - Capable of steered and interactive MD	✓
AMBER [272] amber scripps edu	- Includes forcefields for fully atomistic - AMBER forcefield is compatible with and used in many of these MD packages	
DL_Poly [273] www ccp5 ac uk/DL_POLY/	- Compatible with GROMACS or AMBER forcefields - Includes potentials for non-biological material	✓
LAMMPS [274] lammps sandia gov	- Compatible with CHARMM AMBER and GROMACS forcefields - Includes DPD	✓
ESPResSo [275] www espresso mpg de	- Designed for coarse-grained - Includes many algrorithms for electrostatics - Includes LB DPD	✓

Note all these packages can be installed in parallel to take advantage of high performance computing clusters

References

[1] Aksimentiev A Heng, J B Timp G Schulten, K *Biophys J* 2004 *87* 2086–2097

[2] Noolandi J Rousseau J Slater G W Turmel, C Lalande M *Phys Rev Lett* 1987 *58* 2428–2431

[3] Slater G W Noolandi J *Phys Rev Lett* 1985 *55*, 1579–1582

[4] Slater G W Rousseau J Noolandi J , *Biopolymers* 1987 *26* 863–872

[5] Potocek B Gaš, B Kenndler E Stedry M *J Chromatogr A* 1995 *709* 51–62

[6] Patankar N A Hu H H *Anal Chem* 1998 *70* 1870–1881

[7] Noolandi J Slater G W Lim H A , Viovy J L *Science* 1989 *243* 1456–1458

[8] de la Cruz, M O Deutsch J M Edwards, S F *Phys Rev A* 1986 *33*, 2047–2055

[9] Duke T A J *Phys Rev Lett* 1989, *62* 2877–2880

[10] Zimm B H *J Chem Phys* 1991 *94* 2187–2206

[11] Duke T A J Viovy, J L *J Chem Phys* 1992 *96* 8552–8563

[12] Graham, R S Larson R G *Macromolecules* 2007 *40* 366–378

[13] Rapaport D C *The art of molecular dynamics simulation*, Cambridge University Press, Cambridge New York 1995

[14] Frenkel D Smit B *Understanding Molecular Simulation* Academic Press Inc , Orlando FL, USA 2001

[15] Allen M P Tildesley, D J *Computer Simulations of Liquids* Clarendon Press Oxford U K 2002

[16] Haile J M *Molecular dynamics simulation elementary methods* Wiley, New York 1992

[17] van Gunsteren W F Berendsen H J C *Angewandte Chemie International Edition in English* 1990 *29* 992-1023

[18] Tuckerman M E , Martyna G J *J Phys Chem B* 2000 *104* 159–178

[19] Glotzer S C Paul W , *Annu Rev Mater Res* 2002 *32* 401–436

[20] Lennard-Jones J E *Proc Phys Soc* 1931 *43* 461–482

[21] Arnold A Holm, C , *Adv Polym Sci* 2005, *185* 59–109

[22] Deserno M Holm C *J Chem Phys* 1998 *109* 7678–7693

[23] Greengard L Rokhlin, V *J Comput Phys* 1987 *73* 325–348

[24] Lee, F S Warshel A *J Chem Phys* 1992 *97* 3100–3107

[25] Leach. A R.. *Molecular Modelling: Principles and Applications.* Addison Wesley Longman. Essex. England 1996

[26] Binder. K . *Monte Carlo and Molecular Dynamics Simulations in Polymer Science.* Oxford University Press. Inc.. New York. NY. USA 1995.

[27] Underhill. P T , Doyle. P S . *J. Non-Newtonian Fluid Mech* 2003, *122.* 3–31

[28] Weeks. J D.. Chandler. D . Andersen. H. C . *J. Chem. Phys.* 1978. *54,* 5237–5247

[29] Grest. G S., Kremer. K , *Phys. Rev. A* 1986. *33.* 3628–3631.

[30] Ryckaert. J -P . Ciccotti, G . Berendsen, H J C . *J. Comput. Phys.* 1977. *23,* 327–341.

[31] Andersen. H C . *J. Comput. Phys* 1983. *52.* 24–34

[32] Teraoka. I.. *Polymer solutions : an introduction to physical properties.* Wiley. New York 2002

[33] Berendsen. H J. C.. Postma. J. P M.. DiNola. A.. Haak. J R . *J. Chem Phys.* 1984. *81.* 3684–3690.

[34] Andersen, H C , *J. Chem Phys.* 1980. *72* 2384–2393

[35] Stoyanov. S D . Groot. R D.. *J. Chem. Phys.* 2005. *122.* 114112

[36] Soddemann. T . Dunweg. B . Kremer. K , *Phys Rev. E* 2003. *68.* 046702

[37] Nosé. S , *J. Chem. Phys.* 1984. *81.* 511–519.

[38] Hoover. W G . *Phys. Rev. A* 1985. *31.* 1695–1697

[39] Lowe. C P , *Europhys Lett.* 1999, *47.* 145–151.

[40] Hunenberger. P. H . *Adv Polym. Sci* 2005. *173.* 105–147

[41] Guillot. B . *Journal of Molecular Liquids* 2002. *101.* 219–260.

[42] Dimitrov, D I . Milchev, A . Binder, K.. Heermann. D W.. *Macromolecular Theory and Simulations* 2006. *15.* 573–583.

[43] Hoogerbrugge. P. J . Koelman. J M. V A . *Europhys. Lett.* 1992. *19.* 155–160

[44] Groot, R. D . Warren. P., *J. Chem. Phys.* 1997. *107.* 4423–4435.

[45] Español. P . *Phys. Rev. E* 1995, *52.* 1734–1742

[46] Kong, Y.. Manke. C. W.. Madden. W G.. Schlijper, A G., *J. Chem. Phys.* 1997. *107.* 592–602.

[47] Goujon. F., Malfreyt. P , Tildesley. D J , *Molecular Physics* 2005. *103.* 2675–2685

[48] Malevanets. A . Kapral. R . *J. Chem. Phys.* 1999. *110.* 8605–8613

[49] Malevanets. A . Kapral. R , *J. Chem. Phys.* 2000. *112.* 7260–7269

[50] Yeomans. J. M.. *Physica A* 2006. *369.* 159–184.

[51] Ripoll, M Mussawisade K Winkler R G Gompper G *Europhys Lett* 2004 *68* 106–112

[52] Ripoll M Mussawisade K Winkler R G Gompper, G , *Phys Rev E* 2005 *72* 016701

[53] Inoue Y Chen, Y , Ohashi H *Computer Physics Communications* 2001 *142*, 114–116

[54] Inoue Y Chen Y Ohashi H *J Comput Phys* 2004 *201* 191–203

[55] Pooley C M Yeomans J M *J Phys Chem B* 2005 *109* 6505–6513

[56] Noguchi H Kikuchi N Gompper G , *Europhys Lett* 2007 *78* 10005

[57] Ihle, T Kroll D M *Phys Rev E* 2001 *63* 020201

[58] Ihle T Kroll D M *Phys Rev E* 2003 *67* 066705

[59] Malevanets A Yeomans J M *Europhys Lett* 2000 *52* 231–237

[60] Kikuchi, N , Gent A Yeomans J *Eur Phys J E* 2002 *9* 63–66

[61] Mussawisade K Ripoll M , Winkler R G Gompper, G , *J Chem Phys* 2005 *123* 144905

[62] Winkler R G Ripoll M Mussawisade K Gompper G *Computer Physics Communications* 2005 *169* 326–330

[63] Webster M A , Yeomans J *J Chem Phys* 2005 *122*, 164903

[64] Ripoll M Winkler R G Gompper G *Phys Rev Lett* 2006, *96* 188302

[65] Falck E Lahtinen, J M , Vattulainen I Ala-Nissila T *Eur Phys J E* 2004 *13*, 267–275

[66] Padding, J T Louis A A , *Phys Rev Lett* 2004 *93*, 220601

[67] McNamara, G , Zanetti G *Phys Rev Lett* 1988 *61* 2332

[68] Raabe D *Modelling Simul Mater Sci Eng* 2004 *12* R13

[69] Ahlrichs P , Dunweg B *J Chem Phys* 1999 *111*, 8225–8239

[70] Usta O Ladd A , Butler, J *Journal of Chemical Physics* 2005 *122* 094902

[71] Adhikari, R Stratford K Cates M Wagner A *Europhys Lett* 2005, *71* 473–479

[72] Dunweg, B Schiller, U D Ladd, A J C *Phys Rev E* 2007 *76* 036704

[73] Coffey, W Kalmykov, Y P Waldron J T *The Langevin equation* World Scientific River Edge, NJ 1996

[74] Russel, W B Saville D A , Schowalter W R *Colloidal Dispersions* Cambridge University Press, Cambridge 1992

[75] Ermak D , McCammon J *J Chem Phys* 1978 pp 1352–1360

[76] Doi M Edwards S F *The theory of polymer dynamics* Clarendon Press New York 1986

[77] Ottinger H C *Stochastic Processes in Polymeric Fluids Tools and Examples for Developing Simulation Algorithms*, Springer, New York 1996

[78] Graessley W W *Polymeric Liquids & Networks Dynamics and Rheology* Taylor & Francis Group New York 2008

[79] Rotne J Prager S *J Chem Phys* 1969 *50* 4831–4837

[80] Yamakawa H *Modern Theory of Polymer Solutions* Harper & Row, New York 1971

[81] Fixman M *Macromolecules* 1986, *19*, 1204–1207

[82] Jendrejack, R M Graham M D de Pablo, J J *J Chem Phys* 2000 *113* 2894–2900

[83] Junghans C Praprotnik M Kremer K *Soft Matter* 2008 *4* 156–161

[84] Noguchi H Gompper G *Phys Rev E* 2008, *78* 016706

[85] Inoue Y Chen Y Ohashi H *Journal of Statistical Physics* 2002 *107* 85–100

[86] Revenga M Zuniga I Espanol, P Pagonabarraga I *International Journal of Modern Physics C* 1998 *9* 1319 1328

[87] Smiatek J Allen M P Schmid F *Eur Phys J E* 2008 *26* 115–122

[88] Chen S Y , Martinez D , Mei R W , *Physics of Fluids* 1996 *8* 2527–2536

[89] Ladd, A J C , Verberg R , *Journal of Statistical Physics* 2001 *104* 1191–1251

[90] Horbach, J Succi, S , *Phys Rev Lett* 2006 *96*(22), 224503

[91] Chen Y-L , Ma H Graham M D , de Pablo J J *Macromolecules* 2007 *40* 5978–5984

[92] Padding J T Louis A A *Phys Rev E* 2006, *74* 031402

[93] Madras, N Sokal A D *J Stat Phys* 1988 *50* 109–186

[94] Leontidis, E Forrest B M Widmann, A H Suter U W *J Chem Soc Faraday Trans* 1995 *91* 2355–2368

[95] de Pablo J J Yan Q L Escobedo F A *Annu Rev Phys Chem* 1999 *50* 377–411

[96] Binder K Paul W *Macromolecules* 2008 *41* 4537–4550

[97] Newman M E J , Barkema G T *Monte Carlo Methods in Statistical Physics* Oxford University Press USA New York 1999

[98] Viovy J-L *Rev Mod Phys* 2000 *72* 813–872

[99] Landau R H Paez, M J *Computational Physics Problem Solving Using Computers* Wiley Interscience New York 1997

[100] Gauthier M G Slater G W *J Chem Phys* 2002 *117*(14) 6745–6756

[101] Gauthier M G Slater G W *Phys Rev E* 2004 *70*, 015103

[102] Gauthier M G Slater G W *Physica A* 2005 *355* 283–296

[103] Mercier J F Slater G W *J Chem Phys* 1999 *110*(12) 6050–6056

[104] Mercier J F Slater G W *J Chem Phys* 1999 *110*(12) 6057–6065

[105] Gauthier M G Slater G W Dorfman K D *Eur Phys J E* 2004 *15* 71–82

[106] Binder K Milchev, A , *J Comput Aided Mater Des* 2002 *9* 33–74

[107] Baumgartner A Binder K *J Chem Phys* 1979 *71* 2541–2545

[108] Verdier P H , Stockmayer W H *J Chem Phys* 1962, *36* 227–235

[109] Gurler M T Crabb C C Dahlin D M Kovac J *Macromolecules* 1983 *16* 398–403

[110] Madras N Sokal A *J Stat Phys* 1987 *47* 573–595

[111] Carmesin I Kremer K *Macromolecules* 1988 *21* 2819-2823

[112] Deutsch H P Binder K *J Chem Phys* 1991 *94* 2294–2304

[113] van Heukelum A Barkema G T *Electrophoresis* 2002 *23* 2562–2568

[114] Deutsch, J M Reger, J D *J Chem Phys* 1991 *95* 2065–2071

[115] Azuma, R , Takayama H *Phys Rev E* 1999, *59*, 650–655

[116] Boileau, J Slater G W *Electrophoresis* 2001 *22*, 673–683

[117] Gauthier M G Slater G W *J Chem Phys* 2008 *128*, 065103

[118] Keh H , Tseng H *J Colloid Interf Sci* 2001 *242* 450–459

[119] Tessier, F Slater G W *Macromolecules* 2006 *39* 1250–1260

[120] Han, J Craighead H *J Vac Sci Technol A* 1999 *17* 2142–2147

[121] Duong-Hong, D Wang J-S Liu G , Chen Y Z Han J Hadjiconstantinou N G , *Microfluid Nanofluid* 2007 *4*, 219–225

[122] Guo Z , Zhao, T Shi Y *J Chem Phys* 2005, *122* 144907–144917

[123] Gas B Kenndler E *Electrophoresis* 2000 *21* 3888–3897

[124] Štedrý, M , Gas, B Kenndler, E *Electrophoresis* 1995, *16* 2027–2033

[125] Tessier F Slater G W *Macromolecules* 2005, *38* 6752–6754

[126] Harden J L Long D Ajdari A *Langmuir* 2001 *17* 705–715

[127] Qiao R He P *Langmuir* 2007 *23* 5810–5816

[128] Qiao R *Langmuir* 2006 *22* 7096-7100

[129] Muthukumar M *Electrophoresis* 1996 *17* 1167–1172

[130] Volkel A Noolandi J , *Journal of Chemical Physics* 1995, *102* 5506–5511

[131] Mohanty U , Stellwagen N *Biopolymers* 1999 *49* 209–214

[132] Hoagland D Arvanitidou E Welch C *Macromolecules* 1999 *32* 6180–6190

[133] Cottet H Gareil, P Theodoly, O Williams C *Electrophoresis* 2000, *21*, 3529–3540

[134] Stellwagen E Lu, Y Stellwagen N *Biochemistry* 2003 *42* 11745–11750

[135] Yeh I Hummer, G *Biophysical Journal* 2004 *86*, 681–689

[136] Yeh I Hummer G *Journal of Physical Chemistry B* 2004 *108* 15873–15879

[137] Frank S Winkler R G *Europhys Lett* 2008 *83* 38004

[138] Grass K Boehme U Scheler U Cottet H Holm C *Physical Review Letters* 2008 *100* 096104

[139] Netz R *Journal of Physical Chemistry B* 2003 *107* 8208–8217

[140] Netz R *Physical Review Letters* 2003 *90* 128104

[141] Stellwagen E Stellwagen N *Electrophoresis* 2002 *23* 2794–2803

[142] Jendrejack R de Pablo, J Graham M *Journal of Chemical Physics* 2002 *116*, 7752–7759

[143] Deutsch, J M *Science* 1988 *240*, 922–924

[144] Deutsch J M , Madden, T L , *J Chem Phys* 1989, *90*, 2476–2485

[145] Smith S B , Aldridge P K Callis J B *Science* 1989, *243*, 203–206

[146] Schwartz D C Koval M *Nature* 1989, *338* 520–522

[147] Nixon G I Slater G W *Phys Rev E* 1994 *50*, 5033–5038

[148] Sevick E M Williams D R M *Phys Rev Lett* 1996 *76*, 2595–2598

[149] Saville P M Sevick E M *Macromolecules* 1999, *32* 892–899

[150] André, P Long D , Ajdari A *Eur Phys J B* 1998 *4* 307–312

[151] Randall G C Doyle P S , *Macromolecules* 2006 *39* 7734-7745

[152] Kim J M , Doyle, P S *Macromolecules* 2007 *40* 9151–9163

[153] Holleran S P Larson R G , *Macromolecules* 2008 *41* 5042–5054

[154] Holleran S P Larson R G *Rheol Acta* 2008 *47* 3–17

[155] Kenward M Slater G W , *Eur Phys J E* 2006, *20* 125–141

[156] Barron A E Blanch H W Soane D S *Electrophoresis* 1994 *15* 597–615

[157] Hubert S J , Slater G W Viovy J -L *Macromolecules* 1996 *29* 1006–1009

[158] Starkweather M E Muthukumar, M Hoagland D A *Macromolecules* 1998 *31* 5495–5501

[159] Starkweather M E Muthukumar M Hoagland D A *Macromolecules* 1999 *32* 6837–6840

[160] Slater G Guo H *Electrophoresis* 1996 *17* 977–988

[161] Slater G W Guo H L *Electrophoresis* 1996 *17* 1407–1415

[162] Labrie J Mercier, J Slater G *Electrophoresis* 2000 *21* 823–833

[163] Mercier J Tessier F Slater G *Electrophoresis* 2001, *22* 2631–2638

[164] Dorfman K Slater G Gauthier M *J Chem Phys* 2003 *119* 6979–6980

[165] Gauthier M G Slater G W *Electrophoresis* 2003, *24* 441–451

[166] Anselmetti D Duong, T T Eichhorn R Reimann P , Regtmeier, J Ros A *Nature* 2005 *436* 1138

[167] Tessier F Labrie J Slater G W *Macromolecules* 2002 *35* 4791–4800

[168] Tessier, F Slater G W *Applied Physics A* 2002 *75* 285–291

[169] Han J Turner S W Craighead H G *Phys Rev Lett* 1999 *83* 1688–1691

[170] Griess, G , Rogers, E , Serwer P *Electrophoresis* 2001 *22* 981–989

[171] Deamer D W Akeson M *Trends Biotechnol* 2000 *18* 147–151

[172] Kasianowicz J J Brandin, E , Branton D Deamer D , *Proc Natl Acad Sci* 1996 *93* 13770–13773

[173] Sung W Park P J , *Phys Rev Lett* 1996 *77* 783–786

[174] Muthukumar M *J Chem Phys* 2003, *118* 5174–5184

[175] Wolterink, J K Barkema G T , Panja D *Phys Rev Lett* 2006 *96* 208301

[176] Dubbeldam J L A Milchev A Rostiashvili V Vilgis T *Phys Rev E* 2007 *76* 010801

[177] Storm A J Storm, C , Chen J Zandbergen H Joanny J -F , Dekker, C , *Nano Lett* 2005, *5* 1193–1197

[178] Dekker C *Nature Nanotechnology* 2007 *2*, 209–215

[179] Zwolak M Ventra M D *Rev Mod Phys* 2008 *80*, 141–165

[180] Soni G Meller A *Clinical Chemistry* 2007 *53* 1996–2001

[181] Li J , Gershow M Stein D Brandin E Golovchenko J A *Nat Mater* 2003 *2*, 611–615

[182] Meller, A Nivon, L , Branton D *Phys Rev Lett* 2001 *86* 3435–3438

[183] Muthukumar M *J Chem Phys* 1999 *111* 10371–10374

[184] Chuang J Kantor, Y , Kardar M , *Phys Rev E* 2002, *65* 011802

[185] Kantor Y Kardar M *Phys Rev E* 2004 *69* 021806 1-021806 12

[186] Panja D Barkema G T *Biophys Jour* 2008 *94* 1630–1637

[187] Luo K Ala-Nissila T Ying S C , *J Chem Phys* 2006 *124* 034714

[188] Chern S Cardenas A Coalson R *J Chem Phys* 2001 *115* 7772–7782

[189] Chen C M , *Physica A* 2005 *350* 95–107

[190] Milchev A Binder K Bhattacharya A *J Chem Phys* 2004 *121* 6042–6051

[191] Gauthier M Slater G W *J Chem Phys* 2008 *128* 065103

[192] Gauthier, M Slater G W , *J Chem Phys* 2008 *128* 205103

[193] Mathe J , Aksimentiev A Nelson, D R Schulten K Meller, A *Proc Natl Acad Sci U S A* 2005 *102* 12377–12382

[194] Guo L Luijten E *Computer simulation studies in Condensed Matter Physics XVIII* 2007 *105* 159–164

[195] Tian P Smith G D *J Chem Phys* 2003 *119* 11475–11483

[196] Huopaniemi I Luo, K , Ala-Nissila T Ying S C *J Chem Phys* 2006 *125* 124901

[197] Guillouzic S Slater G W , *Physics Letters A* 2006, *359* 261–264

[198] Gauthier M G , Slater, G W , *Eur Phys J E* 2008 *25* 17–23

[199] Lansac, Y Maiti P K Glaser M A *Polymer* 2004 *45* 3099–3110

[200] Izmitli A Schwartz D C Graham M D de Pablo J J *J Chem Phys* 2008 *128* 085102

[201] Fyta, M , Kaxiras E Melchionna, S Succi, S *Computing in Science & Engineering* 2008 *10* 10–19

[202] He Y D Qian, H J Lu Z Y Li Z S *Polymer* 2007 *48* 3601–3606

[203] Ali I Yeomans J M , *J Chem Phys* 2005 *123*(23) 234903

[204] Wei D Yang, W Jin X Liao Q *J Chem Phys* 2007 *126* 204901

[205] Luo, K , Ala-Nissila T Ying S C Bhattacharya A *Phys Rev Lett* 2007 *99* 148102

[206] Kong C Muthukumar M *Electrophoresis* 2002 *23* 2697–2703

[207] Ambjornsson T Apell S , Konkoli Z Marzio E D Kasianowicz J *J Chem Phys* 2002 *117* 4063–4073

[208] Luo K Ala-Nissila T Ying S Bhattacharya A *J Chem Phys* 2007 *126*, 145101

[209] Gauthier M Slater G *J Chem Phys* 2008 *128* 175103

[210] Huopaniemi I Luo K , Ala-Nissila T Ying S *Phys Rev E* 2007, *75* 061912

[211] Luo K Huopaniemi I , Ala-Nissila T Ying S C *J Chem Phys* 2006 *124* 114704

[212] Dubbeldam J Milchev A Rostiashvili V Vilgis T *Europhys Lett* 2007, *79* 18002

[213] Baumgartner A Muthukumar M *J Chem Phys* 1987 *87* 3082-3088

[214] Muthukumar M Baumgartner A *Macromolecules* 1989 *22* 1941–1946

[215] Baumgartner, A Muthukumar M *Adv Chem Phys* 1996 *XCIV* 625

[216] Rousseau J , Droun G Slater, G W *Phys Rev Lett* 1997 *79* 1945–1948

[217] Han J Craighead H G *Science* 2000 *288* 1026–1029

[218] Han J Craighead, H G *Anal Chem* 2002, *74* 394–401

[219] Chen, Z Escobedo F A *Mol Simul* 2003 *29* 417–425

[220] de Pablo J J Laso M Suter U W *J Chem Phys* 1992 *96* 2395–2403

[221] Streek M Schmid F Duong T T Ros A *J Biotechnol* 2004 *112* 79–89

[222] Panwar A S Kumar S *Macromolecules* 2006 *39* 1279–1289

[223] Lee Y M , Joo, Y L *J Chem Phys* 2007 *127* 124902

[224] Pernodet N Samulov, V , Shin, K Sokolov J Rafailovich M , Gersappe, D Chu, B *Physical Review Letters* 2000 *85* 5651–5654

[225] Seo, Y , Samulov V Sokolov J Rafailovich M Tinland B Kim J , Chu B , *Electrophoresis* 2002 *23* 2618–2625

[226] Luo, H Gersappe D , *Electrophoresis* 2002 *23* 2690–2696

[227] Seo, Y Luo H Samulov V Rafailovich M Sokolov J Gersappe D Chu B *Nano Letters* 2004 *4*, 659–664

[228] Li B Fang X Luo H Petersen E Seo, Y Samulov V Rafailovich M Sokolov J Gersappe D Chu B *Electrophoresis* 2006 *27* 1312–1321

[229] Reichhardt C J O Reichhardt, C *Physical Review E* 2006, *74* 051908

[230] Hoda N Kumar S *Langmuir* 2007 *23* 11747–11760

[231] Odom T Thalladi V Love J Whitesides G , *Journal of the American Chemical Society* 2002 *124* 12112–12113

[232] Kan C Fredlake, C Doherty E Barron A *Electrophoresis* 2004 *25* 3564–3588

[233] Fredrickson C Fan Z *Lab On a Chip* 2004 *4* 526–533

[234] Reisner W Morton K Riehn R Wang Y Yu Z Rosen M Sturm J Chou S Frey E Austin R *Physical Review Letters* 2005 *94* 196101

[235] Krishnan M Moench I Schwille P *Nano Letters* 2007 *7* 1270 1275

[236] Pennathur S Santiago J *Analytical Chemistry* 2005 *77* 6782 6789

[237] Pennathur S Baldessari F Santiago J G Kattah M G Steinman J B Utz P J *Analytical Chemistry* 2007 *79* 8316–8322

[238] Balducci A Mao P Han J Doyle P S *Macromolecules* 2006 *39* 6273–6281

[239] Zheng J Yeung E *Analytical Chemistry* 2002 *74* 4536 4547

[240] Zheng J Yeung E *Analytical Chemistry* 2003 *75* 3675 3680

[241] Chen Y Graham M de Pablo J Randall G Gupta M Doyle P *Physical Review E* 2004 *70* 060901

[242] Chen Y Graham M de Pablo J Jo K Schwartz D *Macromolecules* 2005 *38* 6680 6687

[243] Jendrejack R Dimalanta E Schwartz D Graham M de Pablo J *Physical Review Letters* 2003 *91* 038102

[244] Jendrejack R Schwartz D Graham M de Pablo J *Journal of Chemical Physics* 2003 *119* 1165 1173

[245] Jendrejack R Schwartz D de Pablo J Graham M *Journal of Chemical Physics* 2004 *120* 2513–2529

[246] Hernandez-Ortiz J P Ma H de Pablo J J Graham M D *Physics of Fluids* 2006 *18*

[247] Usta O Butler J Ladd A *Physics of Fluids* 2006 *18* 031703

[248] Usta O B Butler J E Ladd A J C *Physical Review Letters* 2007 *98* 098301

[249] Mathe J Di Meglio J-M Tinland B *Journal of Colloid and Interface Science* 2007 *316* 831 835

[250] Mathe J Di Meglio J M Tinland B *Journal of Colloid and Interface Science* 2008 *322* 315–320

[251] Cross J D Strychalski E A Craighead H G *Journal of Applied Physics* 2007 *102* 024514

[252] Pennathur S Santiago J *Analytical Chemistry* 2005 *77* 6772–6781

[253] Delgado-Buscalioni R De Fabritiis G *Phys Rev E* 2007 *76* 036709

[254] Barsky S Delgado-Buscalioni R Coveney P V *J Chem Phys* 2004 *121* 2403–2411

[255] Dupuis A Kotsalis E M Koumoutsakos P *Phys Rev E* 2007 *75* 046704

[256] Williams S A Bell J B Garcia A L *Multiscale Model Simul* 2008 *6* 1256–1280

[257] Praprotnik M , Delle Site L Kremer K *Annu Rev Phys Chem* 2008 *59* 545–571

[258] Delgado-Buscalioni R Kremer K Praprotnik M *J Chem Phys* 2008 *128* 114110

[259] Muller-Plathe F *Soft Materials* 2002 *1* 1–31

[260] Chiesl T N Forster R E Root B E Larkin M Barron A E *Anal Chem* 2007, *79* 7740–7747

[261] Fredlake C P Hert D G Kan C W Chiesl T N Root B E Forster R E Barron A E *Proc Nat Acad Sci U S A* 2008 *105* 476–481

[262] Heller C Slater G Mayer P Dovichi N Pinto D Viovy J Droun G *Journal of Chromatography a* 1998 *806* 113–121

[263] Ren H Karger A Oaks F Menchen S Slater G Droun G *Electrophoresis* 1999 *20*, 2501–2509

[264] Desruisseaux C Long D Droun G Slater G *Macromolecules* 2001 *34* 44–52

[265] Haynes R Meagher R Won J Bogdan F Barron A *Bioconjugate Chemistry* 2005 *16* 929–938

[266] Grosser S T Savard J M Schneider J W *Analytical Chemistry* 2007 *79* 9513–9519

[267] Savard J M Grosser S T Schneider J W *Electrophoresis* 2008 *29* 2779–2789

[268] Wikipedia Molecular dynamics — wikipedia the free encyclopedia http //en wikipedia org/w/index php?title=Molecular_Dynamics&oldid=231786544 [Online accessed 23-August-2008] 2008

[269] Van Der Spoel D Lindahl, E , Hess B , Groenhof, G , Mark, A E Berendsen H J , *J Comput Chem* 2005 *26* 1701–1718

[270] Brooks B R Brucoleri R E Olafson, B D States D J Swaminathan S Karplus M *Journal of Computational Chemistry* 2005 *4* 187–217

[271] Phillips J C Braun R Wang W Gumbart J Tajkhorshid E Villa E Chipot, C Skeel R D Kale L Schulten K *Journal of Computational Chemistry* 2005 *26* 1781–1802

[272] Case D A Cheatham T E Darden T , Gohlke H Luo R , Merz K M Onufriev, A Simmerling, C , Wang B Woods R J *J Comput Chem* 2005, *26* 1668–1688

[273] Smith W Young C Rodger P *Molecular Simulation* 2002 *28* 385–471

[274] Plimpton S J *J Comput Phys* 1995 *117* 1–19

[275] Limbach H J Arnold A Mann B A Holm C *Comput Phys Commun* 2006 *174* 704–727