Perspective: AIChE 2012 Allan P. Colburn Award for Excellence in Publications by a Young Member of the Institute

The Fluid Mechanics of Genome Mapping

Kevin D. Dorfman
Dept. of Chemical Engineering and Materials Science, University of Minnesota, 421 Washington Ave. SE, Minneapolis, MN 55455

DOI 10.1002/aic.14002
Published online January 16, 2013 in Wiley Online Library (wileyonlinelibrary.com)

Keywords: optical mapping, DNA barcoding, low Reynolds number flow, complex fluids

Introduction

As chemical engineering research becomes increasingly interdisciplinary, it is important to identify areas where chemical engineers can make a unique contribution. This challenge is particularly manifest for problems that originate in biology. This Perspective makes the case that chemical engineers have something unique to contribute to the area of genome mapping through the discipline's core strength in fluid mechanics. At first glance, it may seem surprising that the dynamics of fluid flow are so closely related to genomic technologies, which are normally associated with biochemistry. However, chemical engineering research into this area is a natural extension of the discipline's long standing contributions to low Reynolds number hydrodynamics[1] and the rheology of complex fluids.[2]

To develop this case, this Perspective begins with an explanation of genome mapping and its connection to DNA sequencing. As this biological application may be unfamiliar to the fluid mechanics community, the problem is recast as a more familiar analogy with numerical methods. After outlining three of the newer approaches used to create genomic maps, this Perspective highlights some of the best examples of chemical engineering work on two of these methods: DNA in extensional flow and DNA in confinement. These examples hardly represent a thorough review of the literature, and readers interested in such a review should consult a recent article.[3] The end of this Perspective discusses some outstanding questions that are ideally suited to a chemical engineering mindset.

Genome Mapping and DNA Sequencing

The key tool for obtaining detailed genomic information is DNA sequencing. The last decade has witnessed a transformation in DNA sequencing.[4,5] These so-called "next generation sequencing" technologies are based on massively parallel sequencing of many short DNA fragments. For example, a single run of the Illumina HiSeq 2000 sequencer[6] used by researchers at the University of Minnesota produces up to 200 gigabases of information, an enormous yield. Unfortunately, the information comes out as approximately 10^9 short sequences of 100 bases. Moreover, structurally complex regions of the genome, such as telomeres, are difficult to sequence,[7] often leading to errors or gaps in the sequence. In order to produce useful information, the data onslaught produced by next generation sequencing needs to be assembled into a readable genome.

To understand the connection between genome mapping and DNA sequencing, it is useful to imagine how one would try to understand the genome if it was a computer program, a literal "code of life." Trying to decipher a genome directly from the data produced by next generation sequencing would be like trying to understand the computer program in Figure 1 from the 15 short, overlapping snippets of code in the left panel. This example is actually an ideal case, since the "sequences" in the figure completely cover the entire program and do not have any errors.

Assembling a genome from short sequencing data is similar to putting together a puzzle. In the relatively small example of Figure 1, you might be able to figure out all of the overlaps in the snippets and then string together the complete program. However, the program assembly is much easier with the "map" provided in the center panel of Figure 1. Here, the letter x is the "sequence-specific" label and the # signs indicate the number of characters between each label. With this map, one can quickly ascertain that the original program is the Newton-Raphson solution to the nonlinear system of equations $2x_1-x_1x_2=0$ and $2x_1^2-x_2=0$ with initial guesses $x_1^{(0)}=3$ and $x_2^{(0)}=5$.

Figure 1 also shows the analogy between this programming example and the corresponding nomenclature in genome assembly. In practice, a genome map consists of a number of short reference sequences and the genomic distance (number of bases) between repeats of these reference sequences along the genome. While the map in the center panel of Figure 1 has "single character" resolution, genomic maps do not require single base resolution to be useful. Rather, kilobase pair resolution[8] is sufficient to facilitate genome assembly.

Correspondence concerning this article should be addressed to K. D. Dorfman at dorfman@umn.edu.

Figure 1. Understanding the connection between genome mapping and DNA sequencing in an analogy with a computer program.

(left) Next generation sequencing provides short, overlapping sequences in a random order. The panel shows 15 randomly sorted, overlapping "reads" of the computer program that are 20 characters in length. (middle) A genome map provides a scaffold for assembly of the program. In this example, the map identifies all of the locations of the letter x in the program and the number of characters between each repeat of x. (right) The complete program is obtained by assembling the overlapping snippets of code in the left panel onto the scaffold provided by the middle panel. To make the assembly clear, the letter x in the snippets of code has the same color as the map. The code is a MatLab program for the Newton-Raphson solution to a system of two nonlinear equations.

It should be clear from this toy example that genome maps are critical for *de novo* genome assembly, where short sequencing reads are assembled without resorting to any outside information. It is tempting to regard mapping as an obsolete technology when there exists a reference genome, such as the human genome,[9] to guide this assembly. (Undergraduate students in numerical methods may recognize the corresponding analogy to Figure 1, as they sometimes try to "assemble" their homework solution onto the solution from the previous year.) Indeed, if there is a high-quality reference genome for a species, then assembling the sequencing data from another member of the species should be straightforward. However, this is not always the case[10] and mapping plays a key role in finishing genomes.[11] Even more important, mapping provides rapid access to large-scale genomic information. For example, many important genomic variants (e.g., in cancer) are phenotypically relevant rearrangements or repeat sequences in the range of tens of kilobases to megabases,[12] which are extremely difficult to detect from 100 base sequencing reads. Large-scale rearrangements in otherwise very large genomes can often be detected directly from the map,[13,14] while avoiding the data deluge inherent in DNA sequencing.[15]

How to Make a Genomic Map

Genomic maps are estimates for the number of bases between repeats of a particular sequence on the genome. Traditionally, these reference sequences are sites cut by enzymes known as restriction enzymes. In the classic approach, a solution of genomic DNA is incubated with a given restriction enzyme, producing a large number of fragments of different sizes that are often in the hundreds of kilobase pair range. These fragments are then separated by size using gel electrophoresis. Measuring the sizes of the fragments provides the genomic distance (i.e., the number of base pairs) between restriction sites in the original genomic DNA, although the analysis is complicated by end effects. While this tried and true technique works, it is slow and labor intensive. Gel electrophoresis in a dc electric field can only separate DNA up to several tens of kilobase pairs.[16,17] For DNA mapping, one normally needs to resort to pulsed field gel electrophoresis,[18,19] where periodic changes in the direction of the electric field impart an electrophoretic mobility that depends on the reorientation time for the DNA.[17] Pulsed field gel electrophoresis can separate DNA into the megabase range, permitting the sizing of whole chromosomes,[18] but it is a slow process requiring many hours to days.[17] The past decade has witnessed continued efforts to develop microfabricated separation devices that could replace pulsed field gel electrophoresis.[3,20] While such microfabricated "artificial gels" have reduced the separation time for long DNA into the minute range,[3] they remain at the prototype stage and have not translated into routine use.

Optical mapping is an alternate approach to obtain the same information using single molecules of DNA.[3,21] Figure 2 shows the output of such an experiment.[14] Each row in this figure is a single long piece of human genomic DNA in the hundreds of kilobase pair range that has been stretched, immobilized, cut with a restriction enzyme, and dyed with a fluorescent intercalator that sticks in the backbone of the DNA. The distance between restriction sites is obtained by

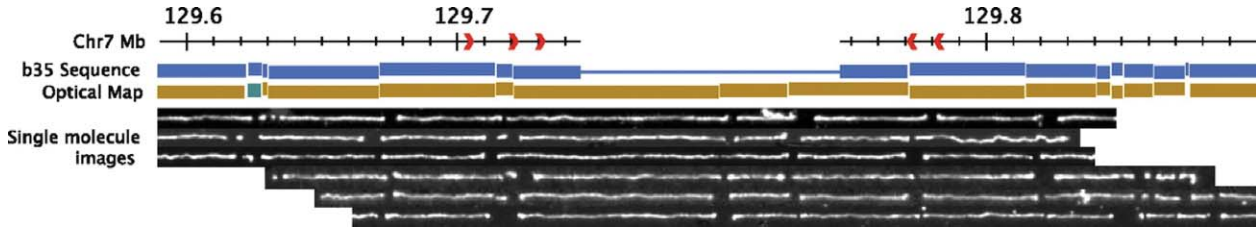

Figure 2. Principle behind optical mapping.

Single molecules of DNA are stretched in a flow, immobilized to a surface, and cut by restriction enzymes. These cuts produce dark spots in the optical image, and the genomic distance between cuts is obtained from the fluorescence intensity of the intact DNA segments. This particular example is human genomic DNA, where the optical mapping uncovered a 90 kilobase pair insertion on chromosome 7. (Reproduced with permission from Ref. 14. [Copyright 2010 National Academy of Sciences.])

measuring the fluorescence intensity between dark spots in the image. In addition to its aesthetic beauty and much higher throughput than pulsed field gel electrophoresis, optical mapping has an important bioinformatic advantage. In the electrophoresis experiment, the DNA are cut first and then separated by size. As a result, one obtains the distribution of distances between the restriction sites but no information about their relative locations along the genome. Consequently, assembling the pulsed field gel electrophoresis data into a map of the genome is challenging. In contrast, Figure 2 reveals that an optical map can yield many stretched DNA that have some overlap in their black/white pattern. The ability to stretch intact, long genomic DNA and then cut the fragments greatly simplifies the data analysis. Indeed, the high degree of automation now present in optical mapping is critical to the analysis of genomic problems in large, difficult genomes like maize[13] and humans.[14]

DNA barcoding is another way to obtain similar large-scale genomic information from single DNA molecules.[3,22] The moniker "DNA barcoding" is very appropriate, but for trademark reasons different companies have used other names for the same basic idea.[3,23–25] In any case, the idea behind DNA barcoding is to insert fluorescent, sequence-specific probes into the DNA. Figure 3 shows one approach by nick extension,[22] but there are many ways to insert the probes.[3] In addition to identifying particular sequences along the genome, there are also barcodes that can detect epigenetic markers such as DNA methylation.[26] In contrast to the optical mapping method in Figure 2, where the sequence-specific probes are the dark spots in the image, DNA barcoding is a two-color measurement. The backbone of the DNA is dyed with an intercalating dye that fluoresces at a different wavelength than the sequence-specific probes. Thus, as seen in Figure 3, the genomic distance between barcodes is obtained by integrating the green fluorescence intensity between the red markers on the stretched DNA.

Technologies for Stretching DNA

At equilibrium in free solution, polymers like DNA adopt a random coil configuration to maximize their configurational entropy. Both optical mapping and DNA barcoding require stretching DNA out of this conformation by driving the DNA out of equilibrium and/or shifting the equilibrium conformation away from a random coil. Figure 4 illustrates the three main approaches to accomplish this task (1) molecular combing,[27] (2) extensional flow,[25] or (3) confinement in a nanochannel.[28] All of these methods are currently being pursued as commercial technologies, and the key companies working on each technique are listed in Figure 4.

By far, the most advanced method for DNA stretching is the molecular combing technique illustrated in Figure 4a, which has already been commercialized by OpGen. Molecular combing uses a receding contact line to stretch DNA on a chemically treated surface that favors adsorption at the DNA ends.[27] We can think of molecular combing as a combination of a nonequilibrium mechanism and shifting the equilibrium configuration. When one end of the DNA is attached and the rest of the chain is stretched by the flow, the DNA is certainly not in its equilibrium configuration. However, once the other end of the chain is attached to the surface, the DNA is now in a stable conformation due to the large surface adsorption energy. While early molecular combing experiments[29] demonstrated successful stretching, the alignment and density of DNA on the surface were not well controlled, hindering high-throughput analysis. Subsequent advances in flow control using the coffee ring effect[30] and, eventually, the microchannel device[8] in Figure 4a led to

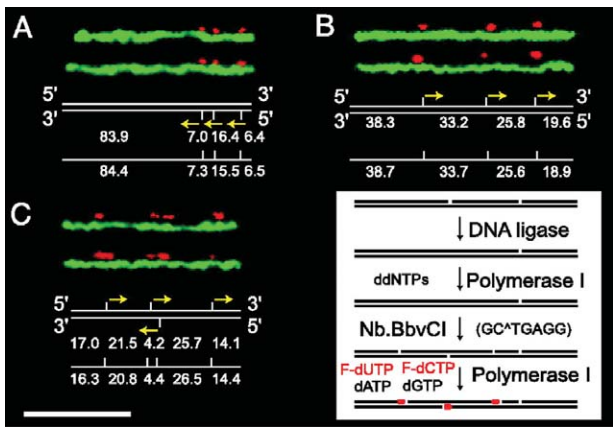

Figure 3. Principle behind DNA barcoding by nick extension.

The lower-right panel indicates the biochemical workflow (1) Repair any existing single-strand breaks (known as nicks), (2) enzymatically introduce sequence-specific nicks, and (3) fill in these nicks with fluorescent nucleotides. The other images show stretched, barcoded DNA, where the backbone is stained with a green dye and the fluorescent nucleotides are red. (Reproduced with permission from Ref. 22. [Copyright 2007 National Academy of Sciences.])

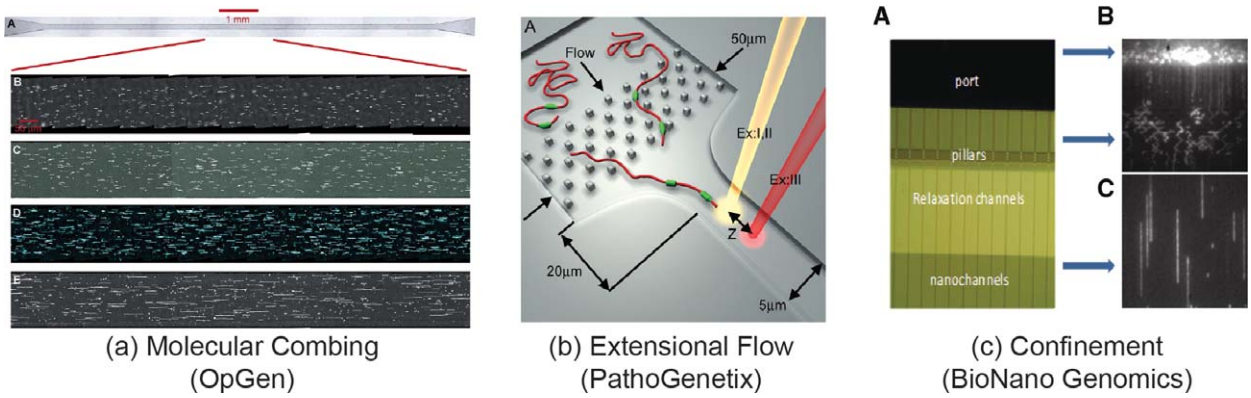

Figure 4. Different technologies for DNA stretching.

The companies associated with each approach are listed in parentheses. (a) In molecular combing, the DNA is stretched by a receding contact line on a treated surface that preferentially adsorbs the DNA at its end. The panels show different length DNA stretched by this method (adapted with permission from Ref. 8. Copyright 2004 American Chemical Society), (b) In extensional flow, the DNA is stretched by a velocity gradient produced by a hyperbolic channel. PathoGenetix was formerly known as US Genomics (reproduced with permission from Ref. 25. Copyright 2004 Creative Commons License). (c) In confinement, the DNA is electrokinetically injected into nanochannels with a size commensurate to the persistence length of the DNA. The key technology is an entropy gradient that facilitates the smooth injection of the DNA. BioNano Genomics was formerly known as Bionanomatrix. (Adapted with permission from Ref. 23. [Copyright 2010 by the authors. Published by Oxford University Press.])

well controlled DNA alignment that permits automated machine reading of the image data.

The second method for stretching in Figure 4, extensional flow, is clearly a nonequilibrium approach. Here, the genomic information is read from the DNA while it is being extended by the flow. The seminal work on stretching DNA in flow was published by Steven Chu's group in the late 1990s,[31–34] sandwiched between his Nobel Prize winning work in cold atom physics and his appointment as the Secretary of Energy. While these studies used fluorescently stained DNA as a model polymer to test theories of polymers in flow, it quickly became clear that the same method could be used to read DNA barcodes. The first article[25] demonstrating the genomic potential for the technology used the device design illustrated in Figure 4b. The relevant dimensionless parameter here is the Deborah number, quantifying the relative importance of extension in the flow to the relaxation of the DNA. The DNA stretching in the device is provided by the hyperbolic contraction, which creates a strong extensional component, and, thus, a large Deborah number. The technology, commercialized by PathoGenetix, has advanced considerably since its introduction in 2004, including a highly integrated "sample in-answer out" device[35] and applications to mapping problems.[36]

The third method in Figure 4, confinement in a nanochannel, involves shifting the equilibrium conformation from a random coil to a stretched configuration. The relevant dimensionless parameter for DNA stretching in confinement is the ratio of the channel size to the persistence length of the DNA, i.e., the characteristic length scale for bending DNA under thermal energy. Naturally, the best stretching occurs in the smallest channels; the emerging commercial technology[24] from BioNano Genomics uses 45 nm 45 nm channels. As illustrated in Figure 4c, the key to injecting long DNA molecules into nanochannels is a gradual change in the number of allowed configurations by using a gradient in the confinement[37] In addition to DNA barcodes, there are also proof-of-principle experiments from academic labs using nanochannels for epigenetic analyses based on methylation[26] or compacted DNA (chromatin).[38,39] Traditional restriction mapping is also possible (but not easy) in nanochannels.[40]

Examples of Chemical Engineering Insight

Many chemical engineering contributions toward elucidating the fluid mechanics of genome mapping methods appeared in a recent review.[3] In the interest of space, this section focuses on three areas that provide the best examples of applying a chemical engineering mindset to the problem (1) Ron Larson's work on the source of molecular individualism in extensional flow,[41,42] (2) Pat Doyle's connections between the kinematics of electrophoresis and hydrodynamic flow,[43–46] and (3) Mike Graham and Juan de Pablo's development of methods for efficiently simulating the hydrodynamics of confined polymers.[47–52] The first two examples connect genome mapping technologies to the mode of thinking in complex fluid rheology, while the third example highlights the importance of low Reynolds number hydrodynamics.

Molecular individualism

de Gennes[53] coined the term "molecular individualism" to describe the remarkable experimental observation[32] that two DNA molecules experiencing the same residence time in an extensional flow can display completely different behavior. (de Gennes colorfully referred to extensional flow as "molecular torture.") It is challenging to investigate the underpinnings of molecular individualism using a purely experimental approach, since (1) optical diffraction limits the resolution of the chain, and (2) there is no facile experimental approach to produce an ensemble of DNA trajectories starting from identical configurations. Simulation is a powerful alternative because the resolution of any measurement is set by the discretization of the model, and it is easy to produce chains with the same initial condition but different trajectories simply by repeating the simulation with a

different random number seed for the Brownian motion. Brownian dynamics simulations of a bead-spring model showed that the molecular individualism arises in part from the initial configuration of the chain when the extensional flow starts.[41] Although the simulations did not include any hydrodynamic interactions between the parts of the chain, the quantitative agreement between the simulations and the experiments is remarkable.[41] Importantly, these simulations also showed that two chains with the same initial conditions are not "predestined" to have the same trajectory. Rather, Brownian fluctuations play an important role in the unraveling process, even for cases where the rate of extension in the flow is hundreds of times faster than the relaxation rate of the chain.[41]

In a follow-up study, Larson[42] showed that preshearing the molecules can reduce molecular individualism in an extensional flow. This important insight is embodied in the post array in Figure 4b. Here, the collisions of the DNA molecule with the post array lead to unraveling of the chain,[20] which should homogenize the initial conditions of the chains as they enter the extensional flow. However, a subsequent publication from US Genomics[54] (remarkably, authored by a different Larson) showed that the posts are superfluous. Rather, because the depth of the microfluidic device is not much larger than the radius of gyration of the DNA, the shear flow between the floor and the ceiling already removes most of the molecular individualism. This is exactly the preshearing concept described by Ron Larson in his simulations.[42]

Analogy between electrophoresis and hydrodynamic flow

The stretching of DNA in a hydrodynamic flow arises from the inhomogeneous drag exerted on the DNA by the moving fluid, which creates tension in the chain. If one instead thinks about the dynamics from the perspective of the DNA molecule, the key to stretching is the kinematics of the external force acting on different parts of the chain. While hydrodynamic drag is a standard way to exert such a force, DNA is a polyelectrolyte and thus undergoes electrophoresis in an electric field. In a bead-spring model of the DNA dynamics, the action of the electric field is equivalent to exerting a force that produces an electrophoretic "velocity" $\mathbf{v}(\mathbf{r})=\mu\mathbf{E}(\mathbf{r})$ on one bead in the chain, where μ is the electrophoretic mobility of the DNA and $\mathbf{E}$ is the electric field vector at the bead position $\mathbf{r}$. Doyle and coworkers[43,44] recognized that this analogy provides a very easy way to stretch a DNA molecule by electrophoresis using a T-junction. The kinematics of the electrophoretic "flow" field in a T-junction includes a stagnation point at the intersection, with the electrophoretic flow pulling the DNA out toward the two arms of the T. While the throughput of this stretching method is much lower than the device in Figure 4b, it very nicely illustrates that flow kinematics govern the stretching, regardless of the origin of these kinematics.

Electrophoresis also provides a convenient mechanism for removing molecular individualism. When a DNA molecule moves through the interface between a gel and free solution, there is a step change in mobility that produces chain extension.[45] Doyle and coworkers[46] showed that a photopolymerized gel, used in place of the posts in the device in Figure 4b, enhances the stretching in a hyperbolic contraction. Unfortunately, the gel device is difficult to use in practice[55] and, as just mentioned, the preconditioning is not essential in the commercial device.[42,54] Although this clever work[46] on *in situ* gels did not yield a useful technology for genome mapping, it was the fountainhead for the Doyle group's recent (and wildly successful) work on flow lithography,[56] which has been spun-off in the startup company Firefly BioWorks.

Methods for confined hydrodynamics

All of the technologies in Figure 4 require accounting for the dynamics of DNA molecules proximate to a surface. In molecular combing and extensional flows, the problem is further complicated by the imposed fluid flow. To construct models of DNA dynamics in these devices, it is essential to have an efficient simulation method that accounts for hydrodynamic interactions. While the segment-segment hydrodynamic interactions between parts of a given DNA molecule are not particularly important, due to the extended state of the DNA, hydrodynamic interactions between the DNA and the walls are critical to producing an accurate model. Over the past decade, Mike Graham and Juan de Pablo led the development of a suite of increasingly sophisticated computational methods[47–52] to address the hydrodynamic problem, which are nicely summarized in Graham's recent review.[57] For relatively simple geometries such as a square nanochannel, recent experience[58] indicates that their grid-based method[48] is an excellent compromise between the complexity of the code and the speed of the calculation. However, for cases where there are many interacting particles[50,51] or a complicated geometry,[52] more sophisticated models are preferable.

An important complement to the hydrodynamic calculation is the model used to represent the DNA molecule. The coarse-grained, bead-spring DNA model[48] used in Graham and de Pablo's articles has found widespread application for modeling DNA in microfluidic devices, in particular for electrophoretic separations.[20] The model is well parameterized, yielding excellent agreement between the simulations and experiments for both electrophoresis in post arrays[59] and diffusion in slit-like confinement.[60] However, one should take caution when using such a bead-spring model to capture DNA dynamics in nanochannels similar to Figure 4c. The polymer coarse-graining in the bead-spring model increases the computational efficiency, but the tradeoff is a lower resolution of the conformation of the chain. Here, the best approach at present is to combine a fine-scale model for the DNA[61] with one of the hydrodynamic methods.[47–52]

Chemical Engineering Opportunities in Genome Mapping

Chemical engineers have already made numerous contributions, both direct and indirect, to genome mapping technologies over the past decade. There is good reason to be optimistic that chemical engineers will continue to have a strong impact in this field in the coming years. Some of the outstanding questions, discussed below, are ideally suited for a chemical engineering skill set and outlook. The discussion is partitioned in the context of the different technologies in Figure 4.

Molecular combing

While molecular combing is a well-developed technology, there is still no detailed model quantifying the stretching of the DNA on a surface in the state-of-the-art microchannel device[8] used in recent genomic studies.[14] The problem itself seems complicated but tractable, and an excellent model system for understanding the fluid mechanics of coating processes. In this context, the molecular combing experiment offers the opportunity to analyze not only the final morphology of the coating (i.e., the data in Figure 4a), but also the dynamics that lead to the coating in the first place. Any improvement in the uniformity and extent of the stretching will only aid in the subsequent analysis of the optical maps.[8]

From the modeling side, the key challenge surrounding molecular combing is accurately accounting for the contact line dynamics and the coupling between the fluid flow and the polymer. The problem is further complicated by the fact that the DNA spans from the bulk fluid to (presumably) a thin film of fluid left on the surface by the receding contact line. In the microchannel device,[8] the contact line is produced inside a relatively small conduit. As a result, one also needs to consider the wetting of the fluid on the channel walls.

While constructing such a model for molecular combing will certainly involve a number of assumptions, the veracity of these assumptions can be tested by experiments in the same geometry. The microchannel fabrication is straightforward, consisting of a number of parallel 10 mm long, 100 μm wide, 8 μm deep channels produced by replica molding.[8] Likewise, the experimental apparatus for visualizing DNA dynamics is widespread (although the cameras are expensive). Indeed, the ability to visualize the DNA dynamics has been driving this scientific field for almost 20 years.[31] Molecular combing experiments would benefit from total internal reflection microscopy, which is a proven technology for understanding DNA adsorption to surfaces.[62,63] The biggest experimental challenge may be the speed of the receding contact line, since the DNA is not especially bright.

Extensional flow

The technological situation is similar for the extensional flow devices in Figure 4b, where there are excellent opportunities for combining simulation and experiment. The design of latest generation of devices[64] used macroscopic mass balances to compute the average velocity as a function of distance down the funnel. This model is one-dimensional (1-D), whereupon the inhomogeneous drag on the DNA arises solely from the gradients in the mean velocity along the channel. Such a model neglects any details of the flow field or the hydrodynamic interactions in the confined geometry. The presence of the walls certainly leads to nontrivial effects.[48,49] At the very least, the DNA-wall friction should increase the relaxation time of the DNA, leading to a larger Deborah number (more stretching) than one would expect from a free solution analysis. At even higher degrees of confinement, the scaling laws for the diffusion coefficient will change.[58]

The time is ripe for a more detailed analysis of the hydrodynamics of DNA flowing through a hyperbolic contraction. In particular, the immersed boundary method[52] used this year to study DNA flow through nanopits is an ideal tool for designing microfluidic stretching funnels. The immersed boundary method provides considerable flexibility in setting the device geometry, which is essential to modeling the various types of funnels[54,64] proposed for DNA stretching. Moreover, since these devices are not overwhelmingly difficult to fabricate, there is ample opportunity to compare the predictions of the simulations to detailed experiments. We again have an excellent confluence of experiment and simulation that can be readily addressed using tools from the chemical engineering skill set.

Confined DNA

While there are engineering opportunities in both molecular combing and flow extension, the potential scientific rewards are even larger in the area of confined DNA. Well before the turn of the century, the description of a polymer confined to a tube was considered a solved problem, with tight confinement described by Odijk's theory[65] and weak confinement described by de Gennes' theory.[66] However, the seminal experiments for the extension of DNA in different size channels by Reisner et al.[67] which did not agree with either theory, led to a complete re-examination of the properties of a confined semiflexible chain.[68] The thermodynamic problems are starting to be resolved, especially understanding the DNA extension seen in nanochannel experiments.[3,61,68] However, almost all of the dynamic questions remain open and require a very careful combination of modeling and experiments. For example, while a surprising result exists for the Kirkwood diffusivity of a moderately confined chain,[58] there is still no consensus for the relaxation time of the chain or any experimental data to test the prediction. There are other dynamic phenomena, such as the increased extension of the chain under electrophoresis seen in experiments,[69] that remain mysterious. The problem only becomes more complicated by the finite length of the DNA used for genome mapping, and the beautiful theories developed for infinitely long chains may not apply to experimentally relevant situations.

While there are now a number of options for incorporating hydrodynamic interactions, simulations in confined geometries would benefit greatly from improvements in the coarse-grained wormlike chain model used to represent the DNA. In weak confinement,[49] where the channel size is only slightly smaller than the radius of gyration of the DNA, one does not require exquisite resolution of the DNA configuration. In this case, it is possible to use a coarse-grained bead-spring model where each spring represents many persistence lengths. In very tight confinement, where the channel size is near the persistence length of the chain, a touching bead model[70] provides the requisite resolution at the length scale of the DNA backbone width.[58] In such a model, many beads are used to represent one persistence length of the chain. Unfortunately, there are many situations of interest where the confinement lies between these two limits. On one hand, a bead-spring model with many persistence lengths per spring would lead to discretization errors in moderate confinement, since such a spring cannot capture the local deformation of the chain. On the other hand, while a touching bead model resolves the local curvature of the chain, it is computationally expensive to extend such a model out to the long chain lengths used in experiments.

A model that spans between the "many persistence lengths per spring" limit and the "many beads per persistence length" limit would be a tremendous asset. There is already an approach using an effective persistence length[71] in the spring force that moves the bead-spring model down to "a few persistence lengths per spring." As a result, this model permits a higher resolution of the chain than the conventional level of coarse-graining,[48] but there is still a large gap between "a few persistence lengths per spring" and "many beads per persistence length." At the 2012 AIChE meeting, Elena Koslovar and Andy Spakowitz presented a new approach to coarse-graining wormlike chains that may span this gap. Their work is still in progress, but it seems like a promising solution to the challenges in modeling DNA in moderate confinement.

While DNA mapping technologies can certainly continue to evolve through heuristic designs, the most pressing issues in this field are exactly the kind of messy, out-of-equilibrium problems that attract many people to chemical engineering in the first place. Chemical engineers are in a unique position to address these problems, and, thus, to have a broad impact on this interdisciplinary field.

Acknowledgments

I thank Mike Graham (University of Wisconsin - Madison) and Pat Doyle (MIT) for sharing insights into their contributions to this field, Han Cao (BioNano Genomics) for his input on the importance of genome mapping, and Andy Spakowitz (Stanford) for providing a preprint of their article on modeling wormlike chains. My group, in particular Sarit Dutta, provided useful input on earlier versions of this manuscript. My research in this area is currently supported by NIH (R01-HG005216), NSF (DMR-1206045), the David and Lucille Packard Foundation, and a Teacher-Scholar Award from the Dreyfus Foundation.

Literature Cited

1. Happel J, Brenner H. Low Reynolds Number Hydrodynamics. The Hague, Martinus Nijhof Publishers; 1983.
2. Bird RB, Armstrong RC, Hassager O. Dynamics of Polymeric Liquids, Vol 1: Fluid Mechanics. New York: Wiley; 1987.
3. Dorfman KD, King SB, Olson DW, Thomas JDP, Tree DR. Beyond gel electrophoresis: Microfluidic separations, fluorescence burst analysis, and DNA stretching. *Chem Rev*. 2013 (in press). doi:10.1021/cr3002142.
4. Schendure JH. Next-generation DNA sequencing. *Nat Biotechnol*. 2008;26:1135–1145.
5. Mardis ER. A decade's perspective on DNA sequencing technology. *Nature*. 2011;470:198–203.
6. Specification sheet for the Illumina HiSeq 2000 Sequencing System. http://www.illumina.com/documents/products/datasheets/datasheethiseq2000.pdf. Accessed Nov. 14, 2012.
7. Green ED, Guyer MS. National human genome research institute, charting a course for genomic medicine from base pairs to bedside. *Nature*. 2011;470:204–213.
8. Dimalanta ET, Lim A, Runnheim R, Lamers C, Chura C, Forrest DK, de Pablo JJ, Graham MD, Coppersmith SN, Goldstein S, Schwartz DC. A microfluidic system for large DNA molecule arrays. *Anal Chem*. 2004;76:5293–5301.
9. International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome. *Nature*. 2004;431:931–945.
10. Garber M, Zody MC, Arachchi HM, Berlin A, Gnerre S, Green LM, Lennon N, Nusbaum C. Closing gaps in the human genome using sequencing by synthesis. *Genome Biology*. 2009;10:R60.
11. Zhou S, Bechner MC, Place M, Churas CP, Pape L, Leong SA, Runnheim R, Forrest DK, Goldstein S, Livny M. Schwartz DC. Validation of rice genome sequence by optical mapping. *BMC Genomics*. 2007;8:278.
12. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet*. 2006;7:85–97.
13. Zhou S, Wei F, Nguyen J, Bechner M, Potamousis K, S. Goldstein S, Pape L, Mehan MR, Churas C, Pasternak S, Forrest DK, Wise R, Ware D, Wing RA, Waterman MS, Livny M, Schwartz DC. A single molecule scaffold for the maize genome. *PLoS Genet*. 2009;5:e1000711.
14. Teague B, Waterman MS, Goldstein S, Potamousis K, Zhou S, Reslewic S, Sarkar D, Valouev A, Churas C, Kidd JM, Kohn S, Runnheim R, Lamers C, Forrest D, Newton MA, Eichler EE, Kent-First M, Surti U, Livny M, Schwartz DC. High-resolution human genome structure by single-molecule analysis. *Proc Nat Acad Sci USA*. 2010;107:10848–10853.
15. Kodama Y, Shumway M, Leinonen R. International nucleotide sequence database collection, the sequence read archive: explosive growth of sequencing data. *Nucleic Acids Res*. 2012;30:D54–D56.
16. Heller C, Duke TAJ, Viovy J-L. Electrophoretic mobility of DNA in gels. II. Systematic study in agarose gels. *Biopolymers*. 1994;34:249–259.
17. Viovy J-L. Electrophoresis of DNA and other polyelectrolytes: Physical mechanisms. *Rev Mod Phys*. 2000;72:813–872.
18. Schwartz DC, Cantor CR. Separation of yeast chromosome-sized DNA by pulsed field gradient gel electrophoresis. *Cell*. 1984;37:67–75.
19. Clark SM, Lai E, Birren BW, Hood L. A novel instrument for separating large DNA molecules with pulsed homogeneous electric fields. *Science*. 1988;241:1203–1205.
20. Dorfman KD. DNA electrophoresis in microfabricated devices. *Rev Mod Phys*. 2010;82:2903–2947.
21. Schwartz DC, Li X, Hernandez LI, Ramnarain SP, Huff EJ, Wang Y-K. Ordered restriction maps of Saccharomyces cerevisiae chromosomes constructed by optical mapping. *Science*. 1993;262:10–114.
22. Jo K, Dhingra DM, Odijk T, de Pablo JJ, Graham MD, Runnheim R, Forrest D, Schwartz DC. A single-molecule barcoding system using nanoslits for DNA analysis, *Proc Nat Acad Sci USA*. 2007;104:2673–2678.
23. Das SK, Austin MD, Akana MC, Deshpande P, Cao H, Xiao M. Single molecule linear analysis of DNA in nano-channel labeled with sequence specific fluorescent probes. *Nucleic Acids Res*. 2010;38:e177.
24. Lam ET, Hastie A, Lin C, Ehrlich D, Das SK, Austin MD, Deshpande P, Cao H, Nagarajan N, Xiao M, Kwok PY. Genome mapping on nanochannel arrays for

structural variation analysis and sequence assembly. *Nat Biotechnol*. 2012;30:771–776.

25. Chan EY, Goncalves NM, Haeusler RA, Hatch AJ, Larson JW, Maletta AM, Yantz GR, Carstea ED, Fuchs M, Wong GG, Gullans SR, Gilmanshin R. DNA mapping using microfluidic stretching and single-molecule detection of fluorescent site-specific tags. *Genome Res*. 2004; 14:1137–1146.
26. Lim SF, Karpusenko AJ, Sakon J, Hook JA, Lamar TA, R. Riehn, DNA methylation profiling in nanochannels. *Biomicrofluidics*. 2011;5:034106.
27. Bensimon A, Simon A, Chiffaudel A, Croquette V, Heslot F, Bensimon D. Alignment and sensitive detection of DNA by a moving interface. *Science*. 1994;265: 2096–2098.
28. Tegenfeldt J, Prinz C, Cao H, Chou S, Reisner WW, Riehn R, Wang YM, Cox EC, Sturm JC, Silberzan P, Austin RH. The dynamics of genomic length DNA molecules in 100-nm channels. *Proc Nat Acad Sci USA*. 2004;101:10979–10983.
29. Michalet X, Ekong R, Fougerousse F, Rousseaux S, Schurra C, Hornigold N, van Slegtenhorst M, Wolfe J, Povey S, Beckmann JS, Bensimon A. Dynamic molecular combing: stretching the whole human genome for high-resolution studies. *Science*. 1997;277:1518–1523.
30. Jing J, Reed J, Huang J, Hu X, Clarke V, Edington J, Housman D, Anantharaman TS, Huff EJ, Mishra B, Porter B, Shenker A, Wolfson E, Hiort C, Kantor R, Aston C, Schwartz DC. Automated high resolution optical mapping using arrayed fluid-fixed DNA molecules. *Proc Nat Acad Sci USA*. 1998;95:8046–8051.
31. Perkins TT, Smith DE, Larson RG, Chu S. Stretching of a single tethered polymer in uniform-flow. *Science*. 1995;268:83–87.
32. Perkins TT, Smith DE, Chu S. Single polymer dynamics in an elongational flow, *Science*. 1997;276:2016–2021.
33. Smith DE, Chu S. Response of flexible polymers to a sudden elongational flow, *Science*. 1998;281:1335–1340.
34. Schroeder CM, Babcock HP, Shaqfeh ESG, Chu S. Observation of polymer conformation hysteresis in extensional flow. *Science*. 2003;301:1515–1519.
35. Meltzer RH, Krogmeier JR, Kwok LW, Allen R, Crane B, Griffis JW, Knaian L, Kojanian N, Malkin G, Nahas MK, Papkov V, Shaikh S, Vyavahare K, Zhong Q, Zhou Y, Larson JW, Gilmanshin R. A lab-on-chip for biothreat detection using single-molecule DNA mapping. *Lab Chip*. 2011;11:863–873.
36. Protozanova E, Zhang M, White EJ, Mollova ET, Ten Broeck D, Fridrikh SV, Cameron DB, Gilmanshin R. Fast high-resolution mapping of long fragments of genomic DNA based on single-molecule detection. *Anal Biochem*. 2010;402:83–90.
37. Cao H, Tegenfeldt JO, Austin RH, Chou SY. Gradient nanostructures for interfacing microfluidics and nanofluidics. *Appl. Phys Lett*. 2002;81:3058–3060.
38. Cipriany BR, Zhao RMurphy PJ, Levy SL, Tan CP, Craighead HG, Soloway PD. Single molecule epigenetic analysis in a nanofluidic channel. *Anal Chem*. 2010;82: 2480–2487.
39. Benitez JJ, Topolancik J, Tian HC, Wallin CB, Latulippe DR, Szeto K, Murphy PJ, Cipriany BR, Levy SL, Soloway PD, Craighead HG. Microfluidic extraction, stretching and analysis of human chromosomal DNA from single cells. *Lab Chip*. 2012;12: 4848–4854.
40. Riehn R, Lu M, Wang Y-MLim SF, Cox EC, Austin RH. Restriction mapping in nanofluidic devices. *Proc Nat Acad Sci USA*. 2005;102:10012–10016.
41. Larson RG, Hu H, Smith DE, Chu S. Brownian dynamics simulations of a DNA molecule in an extensional flow field. *J Rheol*. 1999;43:267–304.
42. Larson RG. The role of molecular folds and 'pre-conditioning' in the unraveling of polymer molecules during extensional flow. *J Non Newtonian Fluid Mech*. 2000;94: 37–45.
43. Tang J, Doyle PS. Electrophoretic stretching of DNA molecules using microscale T junctions. *Appl Phys Lett*. 2007;90:224103.
44. Balducci AG, Tang J, Doyle PS. Electrophoretic stretching of DNA molecules in cross-slot nanoslit channels. *Macromolecules*. 2008;41:9914–9918.
45. Underhill PT, Doyle PS. DNA stretch during electrophoresis due to a step change in mobility. *Phys Rev E*. 2007;76:011805.
46. Randall GC, Schultz KM, Doyle PS. Methods to electrophoretically stretch DNA: microcontractions, gels, and hybrid gel-microcontraction devices. *Lab Chip*. 2006;6: 516–525.
47. Jendrejack RM, Graham MD, de Pablo JJ. Hydrodynamic interactions in long chain polymers: Application of the Chebyshev polynomial approximation in stochastic simulations. *J Chem Phys*. 2000;113:2894–2900.
48. Jendrejack RM, de Pablo JJ, Graham MD. Stochastic simulations of DNA in flow: Dynamics and the effects of hydrodynamic interactions. *J Chem Phys*. 2002;116: 7752–7759.
49. Jendrejack RM, Schwartz DC, Graham MD, de Pablo JJ. Effect of confinement on DNA dynamics in microfluidic devices. *J Chem Phys*. 2003;119:1165–1173.
50. Hernandez-Ortiz JP, de Pablo JJ, Graham MD. N log N method for hydrodynamic interactions of confined polymer systems: Brownian dynamics. *J Chem Phys*. 2006; 125:164906.
51. Hernandez-Ortiz JP, de Pablo JJ, Graham MD. Fast computation of many-particle hydrodynamic and electrostatic interactions in a confined geometry. *Phys Rev Lett*. 2007;98:140602.
52. Zhang Y, De Pablo JJ, Graham MD. An immersed boundary method for Brownian dynamics simulation of polymers in complex geometries: Application to DNA flowing through a nanoslit with embedded nanopits. *J Chem Phys*. 2012;136:014901.
53. de Gennes PG. Molecular individualism. *Science*. 1997; 276:1999.
54. Larson JW, Yantz GR, Zhong Q, Charnas R, D'Antoni CM, Gallo MV, Gillis KA, Neely LA, Phillips KM, Wong GG, Gullans SR, Gilmanshin R. Single DNA molecule stretching in sudden mixed shear and elongational microflows. *Lab Chip*. 2006;6:1187–1199.

55. Balducci A, Doyle PS. Conformational preconditioning by electrophoresis of DNA through a finite obstacle array. *Macromolecules*. 2008;41:5485–5492.
56. Pregibon DC, Toner M, Doyle PS. Multifunctional encoded particles for high-throughput biomolecule analysis. *Science*. 2007;31:1393–1396.
57. Graham MD. Fluid dynamics of dissolved polymer molecules in confined geometries. *Annu Rev Fluid Mech*. 2011;43:273–298.
58. Tree DR, Wang Y, Dorfman KD. Mobility of a semiflexible chain confined in a nanochannel. *Phys Rev Lett*. 2012;108:228105.
59. Olson DW, Ou J, Tian M, Dorfman KD. Continuous-time random walk models of DNA electrophoresis in a post array: Part I. Evaluation of existing models. *Electrophoresis*. 2011;32:573–580.
60. Chen YL, Graham MD, de Pablo JJ, Randall GC, Gupta M, Doyle PS. Conformation and dynamics of single DNA molecules in parallel-plate slit microchannels. *Phys Rev. E*. 2004;70:060901.
61. Wang Y, Tree DR, Dorfman KD. Simulation of DNA extension in nanochannels, *Macromolecules*. 2011;44: 6594–6604.
62. Isailovic S, Li H-W, Yeung ES. Adsorption of single DNA molecules at the water/fused-silica interface. *J Chromatogr A*. 2007;1150:259–266.
63. Liu X, Wu Z, Nie H, Liu Z, He Y, Yeung ES. Single DNA molecules as probes for interrogating silica surfaces after various chemical treatments. *Anal Chim Acta*. 2007;602:229–235.
64. Griffis JW, Protozanova E, Cameron DB, Meltzer RH. High-throughput genome scanning in constant tension fluidic funnels. *Lab Chip*. 2013;13:240–251.
65. Odijk T. On the statistics and dynamics of confined or entangled stiff polymers. *Macromolecules*. 1983;16: 1340–1344.
66. Daoud M, de Gennes PG. Statistics of macromolecular solutions trapped in small pores. *J Phys Paris*. 1977;38: 85–93.
67. Reisner W, Morton KJ, Riehn R, Wang YM, Yu Z, Rosen M, Sturm JC, Chou SY, Frey E, Austin RH. Statics and dynamics of single DNA molecules confined in nanochannels. *Phys Rev Lett*. 2005;94:196101.
68. Reisner WW, Pederson JN, Austin RH. DNA confinement in nanochannels: Physics and biological applications. *Rep Prog Phys*. 2012;75:106601.
69. Reccius CH, Stavis SM, Mannion JT, Walker LP, Craighead HG. Conformation, length and speed measurements of electrodynamically stretched DNA in nanochannels. *Biophys J*. 2008;95:273–286.
70. Hagerman PJ, Zimm BH. Monte Carlo approach to the analysis of the rotational diffusion of wormlike chains. *Biopolymers*. 1981;20:1481–1502.
71. Underhill PT, Doyle PS. Alternative spring force law for bead-spring chain models of the worm-like chain. *J Rheol*. 2006;50:513–529.