Biophysical Journal
New and notable

Biophysical Society

Building a better bridge between models and experimental data for DNA

Oscar Gonzalez[1,*]
[1]Department of Mathematics, University of Texas, Austin, Texas

It has been known for decades that when all else is the same, the bending and flexibility of DNA depend on the sequence composition of its bases (1). This variability in material properties is believed to be both chemically and biologically important. For example, in addition to a direct readout of sequence information that contains the genetic code for amino acids, there can be an indirect readout of mechanical information that contains signals that may be important for various biological processes (2,3).

A multitude of studies have been aimed at quantifying the material properties of DNA, which has led to the development of various detailed models (4). Of particular interest here are those of the rigid base type (5), which can account for the position and orientation of each base, and even each phosphate group, both along and across the two backbone strands and provide a prediction of the free energy of a molecule, as a function of its configuration and sequence composition. Detailed variations in bending and flexibility can now be resolved, at the Angstrom level, for DNA in various forms.

Such models evolved from the classic, highly idealized chain models of polymer physics but are now much more detailed and refined and have been under continuous development over the last few decades. Their development was made possible in large part by the recent availability of extensive databases, terabytes in size, produced by atomistic-level computer simulation (6). These models are true products of the "Big Data" age and offer great promise as tools for the analysis of sequences.

But the validation of such models has been lacking. While different kinds of studies have been performed using simulated data, there is relatively little in the way of experimental data. A main reason is that there are few experimental methods that can probe, in a sufficiently direct and sensitive way, local variations in the bending and flexibility of DNA. For example, classic experimental methods based on electrophoresis, crystallography, and microscopy fall short because they may introduce significant perturbations due to a gel matrix or crystal packing forces, among other things.

In contrast, here we consider an experimental quantity known as the J-factor, which is a measure of the probability that a DNA with sticky ends will close up into a cyclized (circular) configuration, as illustrated in Fig. 1. This quantity is distinguished by the fact that it is highly sensitive to the bending and flexibility properties of a molecule, with fewer artifacts (7). While the concept of the J-factor is extremely interesting and attractive as an experimental probe, it is a rather intractable quantity to predict for all but the simplest, most idealized models (8). Indeed, for detailed models of DNA as considered here, the prediction of the J-factor is a major challenge and must necessarily involve numerical techniques. For instance, direct Monte Carlo can be used, but it is expensive and time-consuming.

In this issue of *Biophysical Journal*, Manning (9) presents a numerical technique for computing the J-factor for some detailed models of DNA of the rigid base and rigid basepair types. The technique draws upon a number of different mathematical and computational tools and is deterministic, as opposed to probabilistic, in form, with various enhancements aimed at efficiency.

The overall approach is based on an integral representation of the J-factor, where the integral is evaluated over a high-dimensional unbounded (infinite size) set in configuration space. The numerical evaluation of such an integral is a formidable task. However, for integrals with an exponential integrand, an analytical result known as the Laplace approximation can be applied. Other noteworthy approaches

Submitted October 23, 2024, and accepted for publication November 25, 2024.

*Correspondence: og@math.utexas.edu

Editor: Filip Lankas.
https://doi.org/10.1016/j.bpj.2024.11.3317

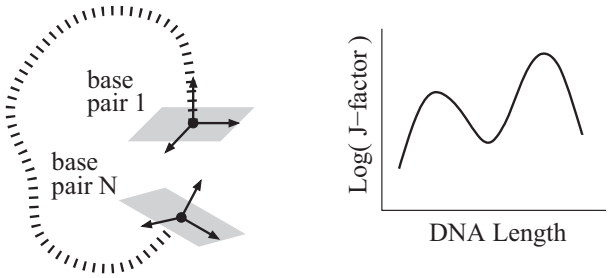

FIGURE 1 Definition of the J-factor and a typical profile curve for DNA. Left: for cyclization to occur in a thermally fluctuating fragment, the relative position and orientation of the two ends (basepairs 1 and N) must come within a specified range. The J-factor can be interpreted as the probability of this event. The dotted curve represents basepairs 2 to $N-1$, and the shaded planes represent basepairs 1 and N. The local frame in each plane represents the three-dimensional position and orientation of the basepair. Right: the J-factor is a sensitive function of length and material properties of DNA at scales of a few hundred basepairs and can vary by orders of magnitude. A plot of the J-factor versus length provides a characteristic profile curve for a given family of fragments. Differences in the location, period, and amplitude of undulations in this curve could potentially be used to probe detailed material properties related to the bending and torsional flexibility, intrinsic curvature, and helical repeat of DNA.

have been considered, such as those based on Fourier transforms (10) and thermodynamic integration (11), but within the context of more idealized models.

A unique feature of the treatment by Manning is the choice of configuration variables. Two types are considered: internal variables, which describe how neighboring bases are displaced and rotated relative to each other, and absolute variables, which describe the position and orientation of each base with respect to a given external frame. Whereas the (elastic) free energy for the model is conveniently defined in terms of internal variables, these variables lead to difficult expressions for the cyclization conditions, which are most naturally expressed in terms of absolute variables. The approach by Manning is to exploit the best of both worlds and employ both types of variables where appropriate. As a result, the cyclization conditions and the J-factor integral take simpler forms than in other approaches (10) and are more manageable.

In the Laplace approximation, the J-factor integral is made tractable by expanding the integrand about an energy-minimizing cyclized configuration, which then leads to Gaussian-type integrals, which can be evaluated in closed form. An important issue is that a given DNA fragment can have multiple local energy minimizers with similar energies, and care must be taken to include them when appropriate. Another core issue is the treatment of configuration variables that represent three-dimensional rotations. Mathematically, such rotations can be represented as points in a three-dimensional set, where the measurement of distances and volumes depends on a local metric. Care must be taken in the handling of the rotation variables, for instance when computing energy minimizers and volume integrals in configuration space. Manning uses a variety of tools to create a mathematically sound and nuanced method for these computations.

The work by Manning provides a bridge from a mathematical model of DNA to an experimentally accessible quantity that is highly sensitive to detailed features of the model, which promises to be a valuable tool for purposes of validation as well as parameter estimation. The current form of the method is based on only a leading-order approximation, with limited accuracy depending on the length of the DNA. Enhancements aimed at higher-order approximations and longer lengths would be natural areas for future research. This is an exciting time for this important and challenging problem. Interesting questions abound, much work remains to be done, and the advent of machine learning and AI may offer new opportunities.

DECLARATION OF INTERESTS

The author declares no competing interests.

REFERENCES

1. Hagerman, P. J. 1990. Sequence-directed curvature of DNA. *Annu. Rev. Biochem.* 59:755–781.
2. Basu, A., D. G. Bobrovnikov, …, T. Ha. 2021. Measuring DNA mechanics on the genome scale. *Nature*. 589:462–467.
3. Pedersen, A. G., L. J. Jensen, …, D. W. Ussery. 2000. A DNA structural atlas for *Escherichia coli*. *J. Mol. Biol.* 299:907–930.
4. Dans, P. D., J. Walther, …, M. Orozco. 2016. Multiscale simulation of DNA. *Curr. Opin. Struct. Biol.* 37:29–45.
5. Sharma, R., A. S. Patelli, …, J. H. Maddocks. 2023. *cgNA+web*: A visual interface to the *cgNA+* sequence-dependent statistical mechanics model of double-stranded nucleic acids. *J. Mol. Biol.* 435:167978.
6. Pasi, M., J. H. Maddocks, …, R. Lavery. 2014. μABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res.* 42:12272–12283.
7. Shore, D., J. Langowski, and R. L. Baldwin. 1981. DNA flexibility studied by covalent closure of short fragments into circles. *Proc. Natl. Acad. Sci. USA*. 78:4833–4837.
8. Shimada, J., and H. Yamakawa. 1984. Ring-closure probabilities for twisted wormlike chains: application to DNA. *Macromolecules*. 17:689–698.
9. Manning, R. S. 2025. Laplace approximation of J-factors for rigid base and rigid base pair models of DNA cyclization. *Biophys. J.* 124:40–61.
10. Zhang, Y., and D. M. Crothers. 2003. Statistical mechanics of sequence-dependent circular DNA and its application for DNA cyclization. *Biophys. J.* 84:136–153.
11. Giovan, S. M., A. Hanke, and S. D. Levene. 2015. DNA cyclization and looping in the wormlike limit: normal models and the validity of the harmonic approximation. *Biopolymers*. 103:528–538.