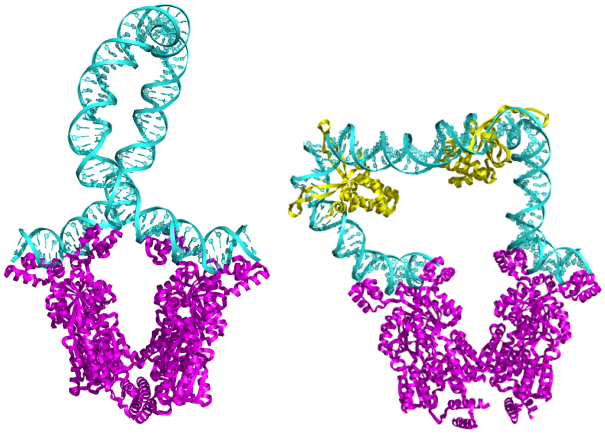
Investigating the Physics of DNA Looping By Simulating the *lac* Operon

Lac Operon, Proteins, DNA, Simulations, Polymers

Efficient gene expression is crucial for a living organism's survival. For example, in the absence of lactose, *Escherichia coli* can shut down production of its lactose-digesting enzymes. The manner in which *E. coli* achieves this is through an action of the *lac* operon, a well studied genomic region containing genes for lactose-digesting enzymes. Two distant sites of *lac* operon come close together to form a loop, blocking off the cell's DNA reading mechanism. Interestingly, these loops are much shorter than normally feasible by the random motions of DNA. For reasons not yet fully understood, DNA appears to gain flexibility when in a *lac* operon loop.

The *lac* operon contains three protein binding sites, or operators, for the *lac* repressor or LacR, a V-shaped protein with two DNA binding headpieces on the tips of its arms (See Figure 1). When LacR binds two operators, the two operators move close together preventing transcription of the three *lac* genes. Previous computational and physical treatments of the *lac* operon have neglected important biology. These models consider only idealized symmetric operator sequences and structures, omit key proteins, restrict the motion of LacR about a hinge joint at the base of the V, and neglect electric charges from LacR and DNA. In addition, physical theories which treat DNA as a continuous, elastic tube cannot account for the sharp bends induced within short loops [1].



(a) (b)

Figure 1: Shapshots of simulated 92 base pair DNA loops mediated by the V-shaped LacR assembly (in purple) in the (a) absence and (b) presence of HU (in yellow). The random binding of HU increases the J-factor by $\sim 10^7$ and precludes access to key recognition sites on DNA.

The looped DNA contains too many atoms for an atomic-level treatment. Instead, DNA can be described as a series of rigid rectangular planes representing base pairs with successive displacements characterized by six rigid-body parameters, termed step parameters. Base pair fluctuations are governed by an elastic potential function [2]. Only a small number of generated configurations will meet the end conditions imposed by proteins. The fraction of looped configurations out of all possible configurations is termed the J-factor and can be compared to J-factors obtained in gene repression studies and single molecule experiments.

Using computer simulation, I propose drastically improving upon previous *lac* operon models in order to better understand the mechanics of DNA looping. These simulations will take into account the complex biophysics of the system and provide a physical basis for future theories. I aim to increase the complexity of the modeled *lac* operon, add more binding proteins to the simulation, and construct a freely-available Internet database of simulated *lac* operon structures.

New research has shown that there are substantial differences in the geometry and sequence of operators [3]. Atomic-level simulations of LacR suggest that the headpieces should also be able to rotate with respect to each arm of *lac* [4]. I will approximate the system's electrostatic effects using Debye-Hückle theory, a framework for treating charges within ionic solution. Integrating

this information into modeling will increase realism of the simulations, and may even yield loop formations not yet seen in simulations [5].

I hypothesize that including architectural proteins, highly-abundant proteins which bind at random sites on the DNA and induce sharp bends, will greatly increase calculated J-factors. Experiments and my previous simulations show the absence of architectural protein HU significantly inhibits formation of loops [6]. Loops with bound HU, shown in Figure 1b, differ greatly from ideal loops, shown in Figure 1a. Including other architectural proteins present in *E. coli* such as Fis, H-NS, and Hfq may further change J-factors [7] and loop types.

There are two candidate algorithms to generate realistic DNA structures. I currently use a method devised by Czapla [8] where step parameters are directly sampled from a Gaussian probability distribution based on the elastic energy function. Another potential method is the Metropolis Monte Carlo algorithm [9]. It selects a random base pair, makes a small perturbation, then either accepts or rejects the move with probability derived from the elastic energy function. Czapla's method samples more configurations in a shorter time than Metropolis sampling, but the Metropolis method can also provide some semblance of the dynamics of the system. Further work will be required to determine the best algorithm to incorporate binding proteins and electrostatic forces.

My ongoing collaboration with Dr. Babis Kalodimos has shown there is demand for accurately simulated loops for use in facilitating structure determination. NMR experiments cannot directly image the looped DNA in *lac* loops, but researchers can make inferences using their imaged proteins and my simulated DNA. Constructing a database of looped DNA structures available online would enable Dr. Kalodimos' group and other research groups to visualize *lac* and answer spatial questions. By making this information publicly available along with the source code of my simulations, those who use this database could help in refining my algorithm. This database would also contain articles explaining the biological importance and physical basis for DNA looping geared towards other students.

My research is both biologically and physically important. DNA flexibility plays a critical role in the stability of the nucleosome, the basic compacting unit of eukaryotic DNA [10]. DNA loops are relevant in cancer biology through interactions of p53, DNA Topoisomerase I, chromatin remodeling factors, and RXR [11]. Applying modeling techniques learned from *lac* on these other loops is my ultimate goal. Understanding protein-mediated genetic switches such as the *lac* operon may allow future engineers to control gene expression or construct nano-technological devices.

My affiliation with Dr. Olson and Rutgers University places me in an unique position to carry out this research. Dr. Olson is an established researcher who has extensively studied the interactions and flexibility of DNA. Since Rutgers is also the home of the Protein Data Bank, I will be able to foster close ties to Rutgers' top structural biologists, like Dr. Kalodimos, who can assist in determining the structures needed for my project. Lastly, my background in physics and expertise in high-preformance computing will enable me to implement my ideas in a quantitative manner.

[1] P. A. Wiggins, *et al.*, *Nat Nano* **1**, 137 (2006).
[2] W. K. Olson, A. A. Gorin, X. J. Lu, L. M. Hock, V. B. Zhurkin, *Proc Natl Acad Sci U S A* **95**, 11163 (1998).
[3] J. Romanuka, *et al.*, *J Mol Biol* **390**, 478 (2009).
[4] E. Villa, A. Balaeff, K. Schulten, *Proc Natl Acad Sci U S A* **102**, 6783 (2005).
[5] O. V. Tsodikov, *et al.*, *J Mol Biol* **294**, 639 (1999).
[6] N. A. Becker, J. D. Kahn, L. J. Maher, 3rd, *Nucleic Acids Res* **35**, 3988 (2007).
[7] T. Ali Azam, A. Iwata, A. Nishimura, S. Ueda, A. Ishihama, *J Bacteriol* **181**, 6361 (1999).
[8] L. Czapla, D. Swigon, W. K. Olson, *J Mol Biol* **382**, 353 (2008).
[9] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
[10] J. Virstedt, T. Berge, R. M. Henderson, M. J. Waring, A. A. Travers, *J Str Bio* **148**, 66 (2004).
[11] H. G. Garcia, *et al.*, *Biopolymers* **85**, 115 (2007).