Single Molecule Spectroscopy:

With the advent of Next-Generation-Sequencing (NGS) technologies, an enormous volume of *DNA* sequencing data—in excess of one billion short reads per instrument per day—can be generated at low cost, placing genomic science within the grasp of everyday medicine. Mired in this voluminous data, a new problem has emerged: the assembly of the genome from the short reads. *De novo* assembly is an NP-hard problem and repetitive segments longer than the read length are the crux of the matter.[1,2] It becomes exponentially harder to assemble a genome as the number of repeats grows. We propose to develop a nanopore device for *de novo* sequencing of a single *DNA* molecule with very long (>1 kbp) reads. Nanopore sequencing has the potential for very long reads, reducing the computational burden posed by alignment and genome assembly, while at the same time eliminating logistically challenging and error-prone amplification and library formation due to its exquisite single molecule sensitivity.[3] Nanopore sequencing relies on the electrolytic current that develops when a *DNA* molecule, immersed in electrolyte, is forced by an electric field to translocate through a pore. Each nucleotide in the pore presents an energy barrier to the passage of ions, which blocks the current through the pore in a characteristic way. However, long, high fidelity reads demand stringent control over both the *DNA* configuration in the pore and the translocation kinetics. The configuration determines how the ions passing through the pore contact the nucleotides which affect the signal, while the kinetics affect the time allowed for data acquisition.

To sequence, we plan to force the propeller-like, helical structure of double-stranded *DNA* (*dsDNA*) through a nanometer-sized slit, with dimensions smaller than the double helix, in a solid-state membrane as thin as 2 nm. Molecular dynamics (MD) simulations of our measurements show that by controlling the electric field in such a pore, we can force the *dsDNA* to stretch, tilting the base-pairs (bps) inside the slit, while the B-form canonical structure is preserved outside it. In this configuration the *DNA* is effectively trapped so that the translocation velocity can be controlled, facilitating reads and allowing for discrimination of base-pairs by simply measuring the current. To sequence we intend to first trap a *dsDNA* in the nanoslit; then use low-noise, lock-in measurements of the blockade current to read a base-pair; and finally impel the *DNA* through the pore from one bp to the next by applying a high-speed voltage pulse to the membrane, continuing the cycle. Our preliminary data obtained using a 2.6x2.1nm pore in a 15nm thick membrane to trap a DNA molecule in a configuarion like that shown in Figure 2 indicates that it's possible to recognize the sequence with at least 3bp resolution, right now. With a nanoslit tailored to the *DNA* helical structure and a thinner membrane, we expect to improve on that performance.

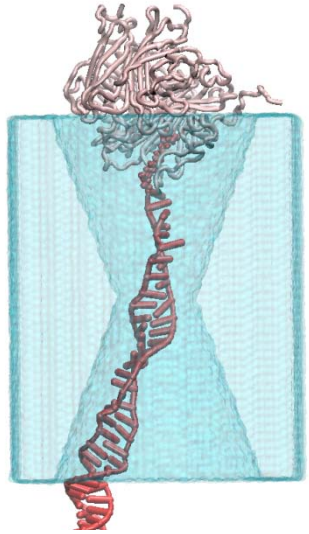

FIGURE 2. Streptavidin bound biotin-DNA duplex trapped by the electric field in a nanopore. An MD snapshot of biotinylated dsDNA bound to streptavidin in a 2.6nm x 2.1nm cross-section pore in a 23nm thick membrane showing *dsDNA*, and stretched *dsDNA*, and *ssDNA* respectively. The molecular conformation is stretched in the constriction beyond the 0.34nm about 8-20%, depending on the applied voltage.