Large Scale Genome Mapping on An Array of Nanochannels

by

YONG ZIXIN

Supervisor: A/P Johan R.C. van der Maarel

A Final Year Project Report

National University of Singapore

Department of Physics

6 April 2014

Content

Abstract

As the first step of genome mapping in nanochannels, quantum dot labeling of double stranded DNAs and single stranded DNAs were investigated. The double stranded DNAs were nick labeled and combed. The single stranded DNAs were prepared by Alkaline denaturation of double stranded DNAs and protein polymer coating. It was found by fluorescent microscopy and atomic force microscopy that the single stranded DNAs were indeed coated and could be labeled by oligonucleotides and quantum dots.

I. Introduction

Genetics

Since the turn of the twentieth century, genetics has been playing an increasingly important role in our understanding of life. The developments of recombinant DNA technology, polymerase chain reaction (PCR) and other powerful techniques have been driving discoveries in biological science constantly. These advancements not only changed the way we understand living organisms, but also provided valuable tools to enhance human welfare. Agriculture relied heavily on recombinant DNA technology to introduce new species with desired traits that can increase the productivity. Human insulin was made available by genetic engineering. Forensic science utilizes DNA finger printing to identify criminals. Medical experts are seeking clues of genetic diseases from the human genetic code (1). Genetics has become an indispensible source of innovations and discoveries. Although DNA is merely made of four types of bases: adenine (A), guanine (G), cytosine (C) and thymine (T), it took us 50 years to progress from the discovery of DNA double helix structure to the completion of Human Genome Project. The 3 billion base pairs on 23 chromosomes impose enormous challenge to researchers over decades and continue to be a core research subject.

Sanger sequencing

Sanger sequencing was published by Sanger's group in 1977 and remained the golden standard of gene sequencing till present. This method was groundbreaking since it provided a reliable and practical sequencing method for the first time and had largely

promoted modern biology's developments. Classical Sanger sequencing utilizes chain-terminating nucleotides dideoxynucleotides (ddNTP) in a polymerasing reaction of single stranded DNA, which is to be sequenced. Four types of ddNTP (ddATP, ddTTP, ddCTP, ddGTP) are used in four separate sequencing reactions. Polyacrylamide gel electrophoresis is then used to separate the reaction product –DNA mixture labeled with different ddNTP at different location. The visualization of electrophoresis bands is achieved by radiography. Following the innovative Sanger sequencing, important improvements such as four-color fluorescent labeling, capillary electrophoresis and novo fragment detection methods have rendered the Sanger sequencing a mature technique. Subsequently, it became the pillar of the Human Genome Project, which also relied on shotgun sequencing to assemble the genome (2).

Next generation sequencing

Though Sanger sequencing is reliable, it is expensive. This limitation has driven researchers to develop the Next-Generation Sequencing (NGS) technology. The first type of NGSs relies on emulsion PCR to produce ensembles of DNA fragments that will amplify the light signal emitted in the nucleotide adding events. The light signals are recorded and analyzed to produce sequencing reads of small pieces of DNA. Example of this type of NGS is Roche platform. The second type of NGS is single molecule detection based and does not require DNA amplification. For instance, the Single Molecule Real Time (SMRT) sequencing is based on single molecule synthesis. Nano-scale structures called zero-mode wave guides are used as the synthesis site. When the free flowing nucleotides are incorporated into the DNA, Fluorescent labels on them detach and the light signals are registered on CCD. NGS technology dramatically decreased the cost of gene sequencing to the level of 1000 dollars while high accuracy can be achieved. Technologies that accomplish sequencing without the aid of labeling are being developed as well. Oxford Nanopore technology makes use of biological or solid-state nanopores to detect changes in current amplitude when single nucleotide passes

through a nano-scale hole. Because different bases (A, T, C, G) will result in different current patterns, single nucleotide detection could be achieved (3).

Limitations

Both Sanger sequencing and Next Generation Sequencing are limited by the short reading length of typically a few hundred base pairs, and genetic information like haplotypes will also be lost in amplifying process (4). The structural variations such as repetitive sequences, inversion and translocations are biologically important, but are also hard to be mapped using either Sanger sequencing or NGS. In particular, the highly repetitive sequences in human genome impose great challenge for whole genome assembly. Efforts are put into developing more sophisticated computational algorithms and using multiple sequencing methods to compensate data loss. These methods are time consuming and costly. The ultimate solution will have to be much longer reading length in base pair sequencing process (5).

Solutions

To circumvent the problem arises from the intrinsic weakness of current sequencing technologies, strategies involving optical mapping were developed. One of the approaches is optical mapping based on chromosome combing. Fluorescent labeled DNA molecules, which contain thousands of base pairs, are stretched on treated glass surface and cut in situ, such that the fragment order is preserved. This method is essential for shotgun sequencing and whole genome assembly, but is restricted by its low throughput. Alternatively, fluidic micro channels are used as platform for large scale DNA mapping. Double stranded DNA molecules are labeled at specific sequence motifs and stretched in micro channels. They are then visualized using fluorescent microscopy (6). However, the resolution of the image is limited because the molecules are in

motion when they are driven through the channels. Low throughput is another limitation for this method (7).

In 2012, Lam et al proposed a technique to address the challenge of large scale genome mapping. They used lithography to fabricate a nanofluidic device that contained three sets of 4000 channels of 45nm in diameter and 0.4mm in length. A gradient region was constructed in front of the channel entrances so that DNA molecules were uncoiled before entering the channel. Since the channels' small diameter was comparable to the persistence length of DNA molecules, they were forced to linearize inside the channels. The fluorescent labeled DNAs were then visualized and imaged by an automatic imaging system. The huge amount of data was processed to give accurate information on the large scale variation of the DNA sequence and provided basis for de nove assembly (8). This approach had a high throughput and resolution. However, it was highly priced and not very accessible.

Our method

Because of the inaccessibility of the silicon based nanochannel array, our lab dedicates to develop a novo technique that does not require very sophisticated nano device fabrication. Instead, we aim to achieve linearization and large scale mapping of DNA in nanochannels of diameter on the level of a few hundred nanometers. Nanochannels of this dimension can be made from elastomer like polydimenthylsiloxane (PDMS) by stamping on a master chip. As a result, the fabrication process will be greatly simplified and the cost will be much lower than a small dimension silicon channel array. With a wider channel, experiments that involve in situ change of solution condition can be performed. Moreover, electrophoresis can be used to load the DNA molecules into channels of this dimension.

To further eliminate the restriction that only specific motifs can be labeled and used as mapping flags, we are researching on the feasibility of single stranded DNA (ssDNA) mapping in nanochannels. Because ssDNAs could be labeled at any desired sites by oligonucleotides synthesis, this approach could serve as a powerful gene locating tool, and large scale genome mapping could become more accurate.

The amplified stretch of dsDNA and ssDNA in a nanochannel is achieved with a protein diblock copolymers coating. The polymer coating increases the stiffness of the DNA and so its persistence length. The DNA molecules then elongates in the nanochannel accordingly. Our previous work has shown that dsDNAs (λ-phage, 48,502 base pair) coated with diblock polymer C_4K_{12} would achieve a uniform stretch of 85% the contour length inside a channel of a 200nm diameter. The coated dsDNAs were site specific labeled with fluorescent dye Alexa Fluor 546 and YOYO-1 and had the same labeling efficiency as labeled bare dsDNAs that are linearized by combing. The labels on the coated dsDNAs were shown to be at the right nicking sites. This proved that large scale genome mapping was feasible in a relatively large and thus easily fabricated nanochannel with the help of polymer coating (9).

II. Theory

Fluorescent Microscopy (10)

Fluorescence was first introduced in 1852 by Sir George G. Stokes when he observed the light emission caused by UV light. By the 1930s fluorescence has been used in biology. Today fluorescent microscopy is widely used in biology and biomedical imaging. The possibility of multiple labeling enables researchers to identify multiple targets with relatively high resolution at the same time. The development of various fluorophores also makes this technique more versatile.

A fluorescence microscope works by detecting the emitted light from the fluorescence dye on the sample, which is much weaker than the excitation light. Light from an arc-discharge lamp first passes through a selective excitation filter. Then the filtered light is reflected on a dichromatic mirror that is placed at 45 degree angle relative to the incoming light. The excitatory light is then reflected towards the specimen, while a small fraction of light that passes through the dichromatic mirror is absorbed by a coating of the mirror. Light directed to the specimen is condensed by the objective lens so that the light is intense enough to stimulate the fluorophore. Most of the excitatory light just passes by the specimen and a small fraction stimulates the dyes and induces emission light. The emitted light is then gathered again by the same objective lens and sent back to the dichromatic mirror. This time, however, the emitted light passes through the mirror rather than being reflected. Any residue of the excitation light is blocked. At last, the imaging system records the incoming light signals.

The emission spectra are functions of the absorption spectra. Because energy is lost during the electrons excitation in the dyes, the emission spectra show a shift towards longer wavelength and this shift is called Stoke's shift. An appropriate combination of fluorescence dyes and filters enables maximum emission intensity by putting the absorption wavelength at the peak. To increase the emission light intensity, a powerful light source such as high energy short arc discharge lamp is standard component. The efficiency of the fluorescence emission is also related to the absorption rate of the dye, the quantum yield and fluorescence lifetime. Photobleaching is the decomposition of the dye and often unavoidable. This mechanism imposes obstacle to long time observation, but can also be exploited to fulfill certain experiments' requirement.

Atomic Force Microscopy (11)

Atomic Force Microscopy (AFM) has a high resolution on the scale of nanometer, which is far beyond the visible light spectra and hence the optical microscopy's application ambit. This technology was developed in the 1980s and was an essential tool for imaging on the nanoscale. AFM uses a mechanical probe that consists of a cantilever and a tip of a few nanometers in radius to scan specimen surface. Due to the small distance between the tip and the surface being probed, electrical forces will repel or attract the cantilever. A laser is constantly pointed towards the top surface of the cantilever and reflected onto a photodiode panel. When the cantilever is deflected, the laser beam will deflect accordingly and the topography change of the surface can be mapped based on the data collected from the photodiodes. To achieve accurate movement on the surface so that the tip would not collide with the structure, piezoelectric elements that respond to mechanical stress are used to navigate the cantilever. If cantilever with special properties are incorporated, more functions such as probing the electrical conductivity of the surface can be realized.

Two imaging modes are generally used: the contact mode and tapping mode. Contact mode means that the tip and surface are in contact during measurement. The deflection of the cantilever is measured to obtain the structure information of the surface. In tapping mode, the cantilever oscillates up and down when probing through the surface. The oscillation amplitude changes when the tip is close to the surface due to electrical interaction and the height information can be inferred. This mode is less destructive to the surface and the tip, compared to the contact mode.

Except for the high resolution, AFM is able to provide 3D information of the surface probed and thus advantageous over other microscopy. It does not require much modification of the specimen either. The working environment for AFM is not stringent, so it is more versatile. However, there are also disadvantages largely due to its intrinsic property: the scanning area and speed is very limited. A 5um square surface takes about 10 minutes to scan, leading to long operation time.

Polymer conformation (12)

Worm-like chain

DNA can be modeled as a filament obeying Hooke's law. In this model, DNA is assumed to be of smooth continuous structure such that the step length (base pair) goes to 0. Suppose we have one such filament with length s and curvature θ/s, the bending energy obeys Hooke's law:

$$\Delta U = \frac{1}{2} s \kappa_b \left(\frac{\theta}{s}\right)^2$$

where κ_b is a constant representing characteristic bending rigidity. The mean square bending angle $\langle \theta^2 \rangle$ is thermodynamically determined:

$$\langle \theta^2 \rangle = 2\frac{\int exp\,(-\Delta U/kT)\theta^2 d\theta}{\int exp\,(-\Delta U/kT)d\theta} = 2\frac{s}{\kappa_b}kT$$

Here a quantity called persistence length can be defined:

$$\langle cos\theta(s)\rangle = exp\left(-\frac{s}{L_p}\right)$$

The left hand side of the definition represents the directional correlation of the chain. Therefore, the persistence length is a measure of the stiffness of the polymer. DNA shorter that the persistence length tends to behave like a rod and much longer ones would be more likely in a coil form. In the case that the polymer length is small, $\langle \cos\theta(s)\rangle$ can be expanded over the parameter $\theta(s)$:

$$\langle cos\theta(s)\rangle = 1 - \frac{1}{2}\langle \theta^2(s)\rangle + \cdots$$

The right hand side can also be expanded:

$$exp\left(-\frac{s}{L_p}\right) = 1 - \frac{s}{L_p} + \cdots$$

Equating both expansions we obtain:

$$\langle \theta^2(s)\rangle = 2\frac{s}{L_p}$$

With the value of $\langle \theta^2 \rangle$ determined previously, a relation between persistence length and Hooke's law emerge:

$$L_p = \frac{\kappa_b}{kT}$$

We would like to know the characteristic size of the worm-like chain as well. This requires us to calculate the end to end vector of the chain $\vec{h}$, since $\sqrt{\langle h^2 \rangle}$ is proportional to the radius of the polymer. The chain has been assumed to be smooth, so we have:

$$\vec{h} = \int_0^L \vec{\ell}(s)\, ds$$

where $\vec{\ell}(s)$ is the tangent unit vector at s.

With the definition of persistence length, $\langle h^2 \rangle$ can be calculated:

$$\langle h^2 \rangle = \int_0^L ds \int_0^L ds' \langle \vec{\ell}(s) \cdot \vec{\ell}(s') \rangle = 2{L_p}^2 \left[\frac{L}{L_p} - 1 + exp\,(-\frac{L}{L_p}) \right]$$

In the limiting case, we have:

$$\langle h^2 \rangle = L^2,\ L \ll L_p\ (rod)$$

$$\langle h^2 \rangle = 2LL_p,\ L \gg L_p\ (coil)$$

Confinement in tube

Scaling law is applied to investigate the polymer confinement in a tube. The basic idea is to define a length within which the dynamics of chain is unaffected. Consider a tube with diameter $D \gg L_p$, we can assume the chain is undisturbed within diameter D. Suppose the polymer chain inside the tube is a chain of blobs with diameter D, each consists of g links, then $D \cong \ell g^{\nu}$, the extension of the chain can be calculated:

$$R_{\parallel} = \frac{N}{g} D \cong N\ell (D/\ell)^{(\nu-1)/\nu}$$

Take excluded volume interaction into consideration, $\nu = 3/5$, thus:

$$R_{\parallel} \cong N\ell(D/\ell)^{-2/3}$$

In the case of a narrow tube with diameter much smaller than polymer persistence length, Odijk treated the chain as a worm-like chain. This means $\langle\theta^2(s)\rangle = 2\frac{s}{L_p}$ holds.

Geometrically, we also have the relation $\theta \approx \frac{D}{\lambda}$. These two relations give us:

$$\langle\theta^2\rangle = 2\frac{\lambda}{L_p} \cong \left(\frac{D}{\lambda}\right)^2$$

The extension of the case $D \ll L_p$ can then be determined:

$$R_{\parallel} = L\langle cos\theta\rangle \cong L\ （1 - \frac{1}{2}\langle\theta^2\rangle）= L\ （1 - (\frac{D}{L_p})^{2/3}）$$

Circular Dichroism (13)

A circularly polarized light can be decomposed into two orthogonal lights that are linearly polarized and out of phase. The electrical field of the circularly polarized light will trace out a helix. By combining two circularly polarized light with same magnitude but in opposite polarization, we obtain a plane polarized light. However, if one of the circularly polarized lights undergoes change in magnitude, an elliptically polarized light is obtained.

When a combination of two circularly polarized lights-- one left polarized and one right polarized--passes through an optically active medium, the absorptions of these two lights are different due to the intrinsic property and geometrical formation of the medium. The absorptions can be quantitatively represented by the molar extinction coefficient ε. The absorption difference A of the left and right circularly polarized lights

is the Circular Dichroism (CD). This quantity is related to the extinction coefficient difference $\Delta\epsilon = \epsilon_L - \epsilon_R$ by the relation:

$$CD = A_L - A_R = \Delta\epsilon c\ell$$

where c is the concentration and ℓ is the path length. Absorption of lights leads to the change in amplitude. The ellipticity of the elliptically polarized light is also a measurement of the amplitude change of the two circularly polarized lights. CD is thus proportional to the ellipticity θ. Specifically,

$$\theta = 33.0\,\Delta A = 33.0\,(A_L - A_R)$$

CD can be accurately measured with the aid of instrumentation and ellipticity can be calculated from above equation.

Biopolymers' 3D structure can be inferred from their CD spectrum. For example, the secondary structures (helix, sheet and coil) of a polypeptide have a characteristic CD spectrum. By comparing the experimental data to these spectra, we can infer the conformation of the polymers being studied. Calculation of the fraction of the polymer in helix, sheet and coil structure is also possible. This method is non-destructive and requires only tiny amount of sample.

Quantum Dots (14)

Quantum dots (QDs) are inorganic nanocrystals made from semiconductor material. QDs demonstrate excellent optical properties and are new options for fluorescence imaging. To qualify as an excellent label, a dye should be conveniently excitable, bright, stable and easily reproducible, etc. A comparison between QDs and traditional organic dyes is made by Ute Resch-Genger et al as follows.

QDs can be manufactured precisely according to the optical properties desired. QDs have a narrow emission band and by controlling the size of QDs, the absorption peak can be chosen to demonstrate a greater shift from the emission band. This enables easy separation of the absorption and emission lights, which is important for sensitive imaging of the molecules. Organic dyes, however, generally have a small Stokes shift and therefore large overlap between absorption and emission bands. QDs also have a wider absorption bands compared to organic dyes. This property combined with the narrow emission bands allows spectral multiplexing analysis with mono-excitation wavelength. Moreover, QDs' higher quantum yields in the NIR wavelength permits wider applications. A higher molar absorption coefficient of QDs is also favorable for effective light sensing. Another QDs' quality is the longer lifetime that can be advantageous in certain experiments. The excellent stability of QDs exceeds traditional dyes and makes it suitable for longtime exposure.

QDs are relatively new dyes compared to organic dyes and are hence less established in terms of standard labeling procedures and surface functionalization. How the microenvironment can affect QDs optical properties awaits more investigation. Furthermore, limited by the relatively large size of QDs, it is challenging to transport them into cells for intracellular targeting, although extracellular targeting can be achieved more easily. Another problem is the observed blinking of QDs that can influence the observation.

We have in one hand well studied, established and more accessible traditional organic dyes, the other hand QDs with better optical properties and stability but less researched. QDs are well suited for single molecule analysis that can benefit from the stability and brightness, but more exploration is needed before its wide application.

III. Materials and Methods

Materials

Linear lambda DNA (0.5g/L), Nb.BbvCl nicking endonuclease (10,000 U/ml), 10xNE Buffer 2, dNTP set (100 mM), VenR (exo-) DNA polymerase (2000 U/ml), T4 DNA ligase and T4 DNA ligase buffer were purchased from New England Biolanbs, Ipswich, MA. YOYO-1 (1μM), dUTP-biotin (1mM) and Qdot 565 (1μM) was purchased from Invitrogen, Carlsbad, CA. AMICON 100kDA centrifuge tubes were purchased from Merck Millipore, Ireland. 0.45 μm pore size filter was purchased from Whatman, NJ. 1xTE buffer was prepared by mixing 10mM Tris with 1mM EDTA and adjusted to pH 7.9 with HCl.

Double stranded lambda DNA Nicking and Polymerase

1. Dilute dATP, dGTP, dCTP to 0.5 mM with 1xTE buffer one day before (1 μl each + 197μl 1xTE)
2. Heat λ-DNA (0.5g/l) 2 μl at 65℃ then cool down to room temperature one day before
3. Mix the components in following order: Water (34.5μl) => 10xNE Buffer 2 (5μl) => DNA dNTP (1μl) => dUTP-biotin (0.5μl) => VentR (exo-) DNA polymerase (5μl) => Nb.BbvCl nicking endonuclease (2μl)
4. Gently stir the solution using pipette tip to mix it up
5. Incubate the solution at 37℃ for 30 mins to allow polymerase
6. Proceed to purification immediately

Purification with AMICON 100kDA centrifuge tube

1. Fill the tube with 1xTE buffer and centrifuge at 5500xg for 5 mins to wash the tube
2. Load 450 µl 1xTE buffer and the prepared DNA solution (50µl) in the tube
3. Balance the rotor down to 0.1mg and centrifuge at 2000xg for 10 mins, repeat 6 times
4. Collect the remaining solution by centrifuging at 1000xg for 1 min

Quantum Dot Labeling

1. NaCl solution (2M) 1.8µl was added to the above solution (70µl) to make NaCl concentration 50mM
2. 0.75 µl diluted Qdot 565 (0.1µM) was added into the solution, assuming 20% DNA loss in the purification process, the Qdot:DNA ratio is 3:1
3. Take 2 µl the above solution and add 6.92 µl diluted YOYO-1 (5µM) solution, the YOYO-1: DNA base pair ratio is 1:10. The obtained solution was wrapped with aluminum foil to avoid photobleaching and stored in room temperature for 24 hours before fluorescence imaging

Quantum dot labeling of single stranded λ-DNA

1. Mix 20 µl λ-DNA (0.05 g/L) with 180 µl NaOH (0.5M) and incubate at 37 ℃ for 10 minutes for Alkaline denaturing
2. Take 10 µl of the above solution and add in 1ml C_8S_7 polypeptide solution (0.2 g/L)
3. Incubate the solution in room temperature for 1 hour

4. Centrifuge 500 µl the above solution using 3KDa cutoff tube for 2 hours till 20 µl
5. Make 6 tubes of the solution in step 4, each contribute 20 µl
6. Prepare the following solution: 33.5 µl of DI water, 15 µl of 10xT4 DNA ligase buffer, 1.5 µl of oligonucleotide solution (100 µM), mix well
7. Take 100 µl λ-DNA solution from step 5 and add in the solution from step 6 (50 µl), incubate at room temperature overnight
8. Add 5 µl T4 DNA ligase into the solution in step 7 and incubate at room temperature for 2 hours
9. Heat the solution at 65℃ for 20 mins
10. Wash the 3KDa centrifuge tube with 1xTE buffer at 2000xg for 5 minutes at room temperature
11. Centrifuge the λ-DNA solution at 7000xg for 30 minutes, repeat 3 times
12. Assuming the final solution obtained was 0.05 g/L and 100 µL, add in Qdot 565 (1µM) 15.9 µl so that the ssDNA : Qdot ratio is 1:50

DNA combing (15)

The DNA molecules' extremities are more negatively charged than the middle sections. This property was exploited for linearization of the molecules on specially treated glass surface by J.F.Allemend et al. They developed the molecular combing technique to stretch DNA molecules on a glass surface in a non-destructive manner. The DNA itself was not modified. By adjusting the pH of the solution, which was used to coat the surface, highly efficient stretch of molecules could be achieved. DNA combing was conducted following the procedures listed below.

1. 0.5g of polystyrene was dissolved in 10ml toluene and filtered using 0.45 µm pore size filter
2. Glass cover slips were spin coated with the above solution at 2000rpm for 30 seconds
3. 2.5 µl droplet of the DNA solution was dragged along the treated glass cover slip using a pipette tip.

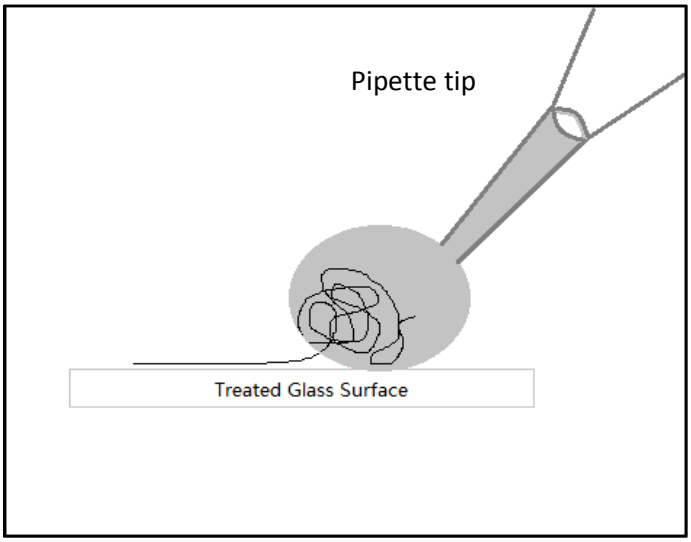


Figure 1. Schematic drawing of DNA combing on a treated glass surface.

DNA Coating (16)

DNAs' physical properties can be effectively modified by coating a protein diblock copolymer onto the molecules, as proposed by Armando H.G. et al. The synthesized protein copolymer C_4K_{12} (~37 KDa) consists of a large hydrophilic colloidal block (about 400 amino acids) and a short binding block (about 12 amino acids). This structure efficiently prevented intra- and intermolecular bridging when the polypeptides were coated around DNA. The brush-coated molecules gained a large persistence length increase subsequently. The aggregation of the polymers was avoided due to the steric repulsion exerted by the colloidal block. With genetic engineering, such protein can be

produced in high yields as well. The C_8S_7 protein polymer used in our experiments demonstrates similar properties, while the molecular weight is about twice that of C_4K_{12}. However, the research work regarding its properties has not been published by the research team that synthesis it. Therefore no reference is available.

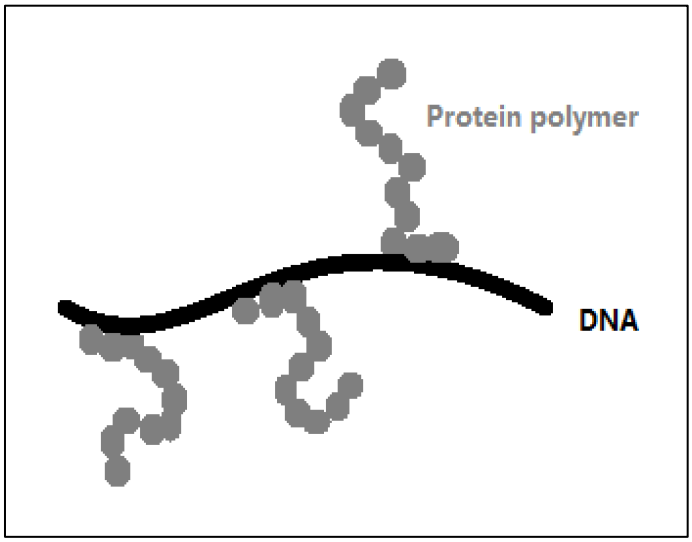


Figure 2. Schematic drawing of DNA coated by protein polymers. The image is not drawn to scale.

Fluorescence Imaging

The labeled ssDNA was visualized in bulk phase and combed dsDNA on the glass surface. A Nikon Eclipse Ti inverted fluorescence microscope with a 100x oil immersion objective lens was used to visualize the samples. A set of filters to distinguish YOYO-1 and Qdot 565 were used. Images were collected by a CCD camera and analyzed using software IMAGEJ.

Circular Dichroism

3ml solution was injected into a quartz cell of 1 cm path length. The sample was loaded into the cell compartment before the UV shutter was switched on. The wavelength band was set to range from 200 nm to 250 nm with data pitch of 1 nm. The same sample was scanned 5 times to generate an averaged value. A CD spectra distribution was auto-generated by a corresponding program.

Atomic Force Microscope

A Dimension 3000 atomic force microscope, Veeco, Woodbury, NY was used for imaging. The cantilever was set to tapping mode. The scanning was operated at 1Hz scanning rate at 512x512 pixel resolution, in room temperature. Images obtained were flattened. The AFM samples were prepared as following.

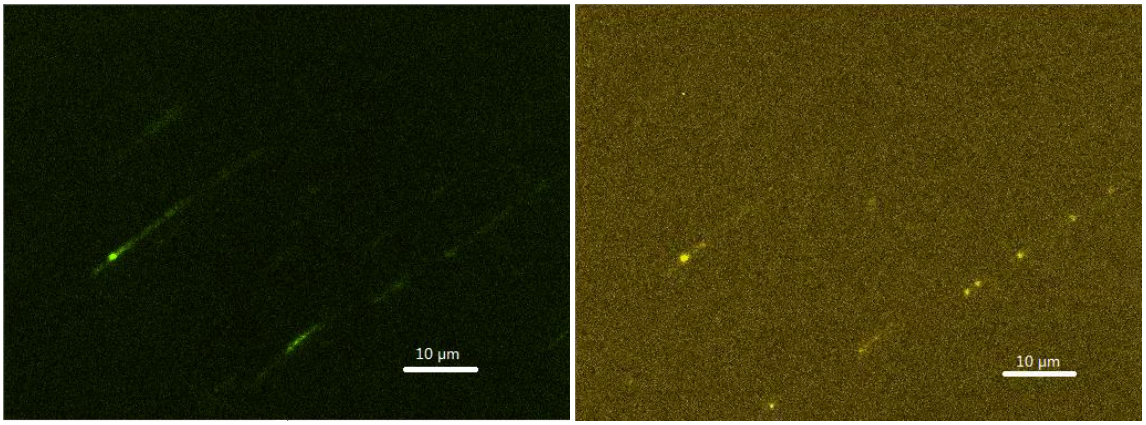
1. Newly cleaved silicon chips were immersed in 100% ethanol and treated with sonicate for 15 mins
2. After washing with DI water and ethanol, the silicon chips were plasma oxidized
3. 10 µl of sample was dropped on the silica surface and await 3 minutes before the surface was slowly washed with DI water (1000 µl)
4. The silica chips were dried with flowing Nitrogen gas before attaching to a metal plate used for holding the sample

The ssDNA smaples for AFM were prepared as following:

1. Mix 20 μl λ-DNA (0.05 g/L) with 180 μl NaOH (0.5M) and incubate at 37 ℃ for 10 mins for denaturing
2. Take 10 μl of the above solution and add in 1ml C_8S_7 polypeptide solution (0.2 g/L)
3. Incubate the solution in room temperature for 1 hour
4. Centrifuge 500 μl the above solution using 100KDa cutoff tube for 5min till 20 μl

IV. Results and Discussion

Quantum Dot Labeling of double stranded λ-DNA



(A) (B)

Figure 3. The nick labeled dsDNAs (48,502 base pairs) were stretched by combing and imaged under fluorescent microscope. The DNA molecules were labeled by both YOYO-1 and Qdot 565. (A) The DNAs were under filter for YOYO-1 (B) The DNA is under filter for Qdot 565.

As shown by the images, DNA molecules were not uniformly stretched due to the manual combing process. There were also some fragments that could be generated by the centrifuge and manipulation process. In the cases that the DNAs were fragmented or not linearized, mapping could not be done with accuracy and the molecules have to be discarded from analysis. The aggregation of Qdots was also observed in the experiments with a low chance. This, however, could lead to the insensitivity of CCD and failure in recognizing individual labels that were relatively dim. In such case, the data was discarded. Quantum dot blinking was constantly observed during the experiment. Therefore a longer recording time was required to confirm the labeling sites. Longer

exposure time also made quantum dots brighter and more being stimulated. However, the YOYO-1s on the molecules that were used to distinguish the whole molecule would undergo photobleach in relatively short time, which set a limitation on long time exposure.

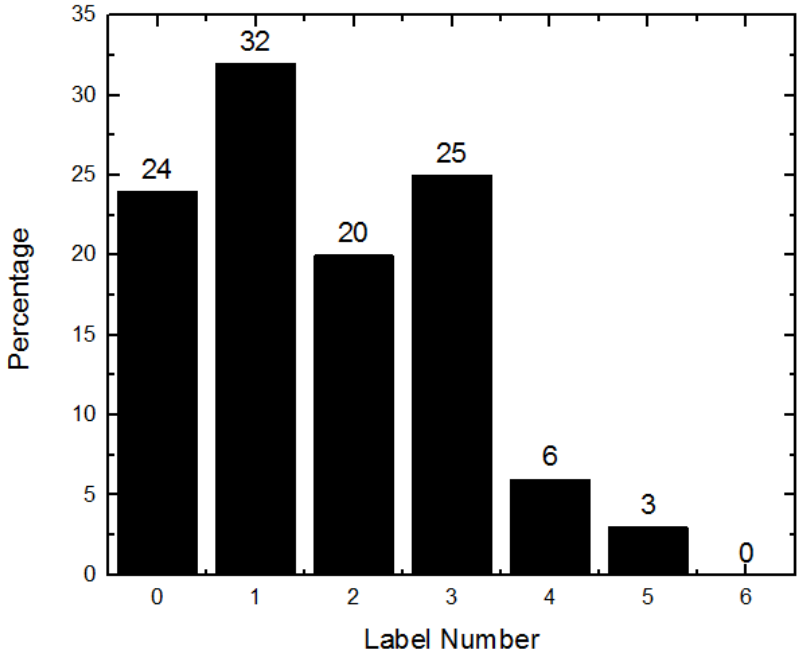


Figure 4. A pool of 100 DNA molecules with average extension 83% of the contour length was counted for their Quantum dot labels.

There were in total 6 discernible labeling sites on λ DNA molecule. From the figure, 9 out of 100 were labeled with more than 4 labels and 24 were not labeled at all. This labeling efficiency was not comparable to the traditional dye such as Alexa Fluor 546 (9) and was not sufficient for nanochannel experiments. One solution could be increasing the Qdots – DNA ratio and wash out the Qdot by centrifuging. This, however, would lead to greater loss of DNA and a brighter background because of the remained extra Qdots. Hence, mapping using the method combining combing and Qdot requires an automated combing system to increase the uniformity of the molecules. A more efficient quantum dots labeling protocol is of due importance.

Quantum dot labeled single stranded λ DNA

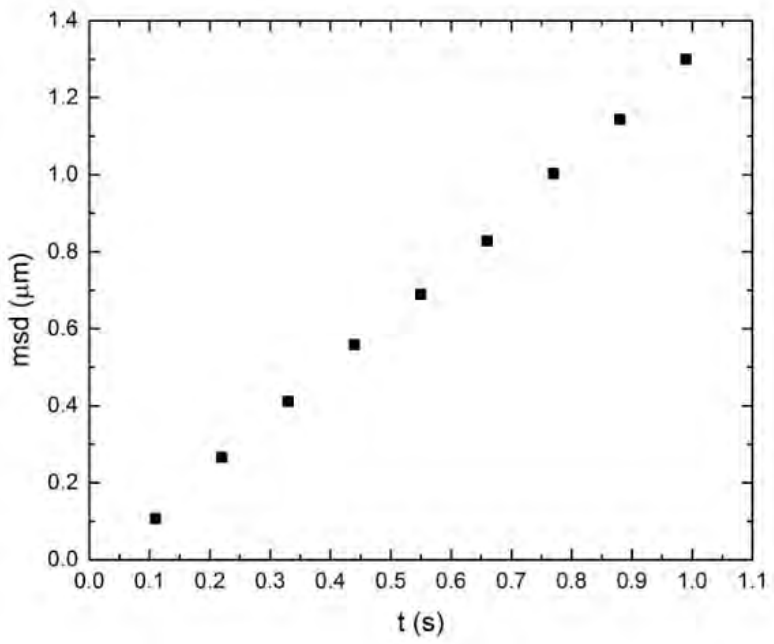


Figure 5. Polypeptide-coated and Qdot labeled Single stranded λ DNA molecules suspended in solution and moved freely. The positions of the molecules were tracked and the mean square displacements were calculated as a function of time.

The Brownian motion of the molecules was recorded and analyzed. The frame rate of the videos taken was 9.091 fps, corresponding to a minimum 0.11 time step. The hydrodynamic radius was calculated according to the relation:

$$R_H = \frac{k_B T}{6\pi\eta_s D}$$

where η_s was the solution viscosity and D the diffusion coefficient. The Diffusion coefficient was given by the relation formulated by Einstein:

$$\overline{\langle \Delta x^2 \rangle} = 2Dt$$

The results obtained were

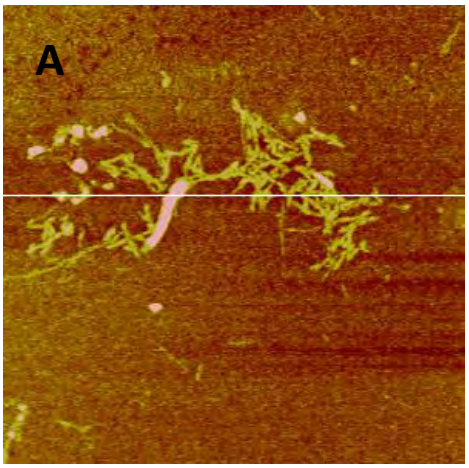
$$D = 6.17 \times 10^{-13} \pm 5.23 \times 10^{-14}\ m^2/s$$

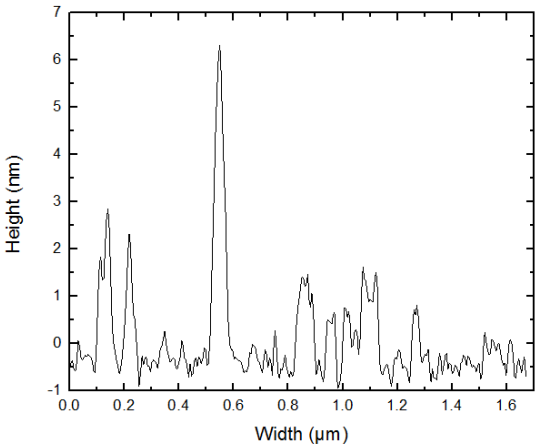
$$R_H = 395 \pm 33\ nm$$

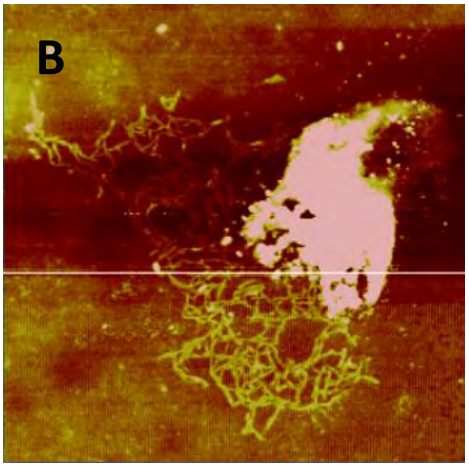
The calculated hydrodynamic radius of the molecules was in the same range as bare double stranded λ DNA (17). Considering the additional condition that the molecules were labeled by oligonucleotides and Qdots, which required the presence of their counterpart sequences, we were sure that the DNA molecules observed were either ssDNAs coated by polypeptides or dsDNAs that were partially denatured and labeled. If the Alkaline denaturation protocol led to perfect denaturation, it was more probable that the molecules were polymer-coated ssDNAs. Additionally, a larger hydrodynamic radius than the experimental one should be observed for a coated double stranded DNA. The problem, however, remained that to what extent have the DNA totally denatured and thus gave the desired sample of pure coated ssDNAs.

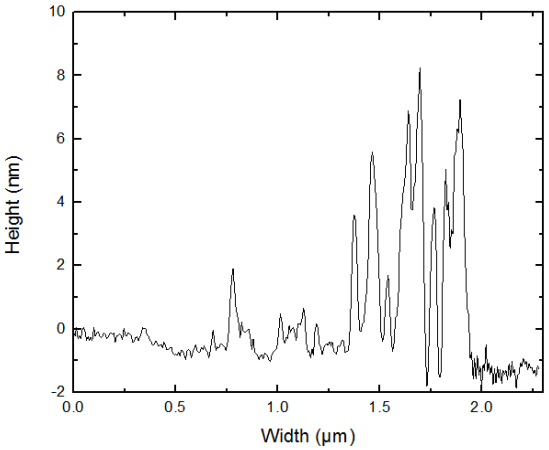
Combing by the same method as for dsDNA was tried for the coated ssDNAs. It was found that the solution was too hydrophilic to be dragged through the treated glass surface. This may due to the high concentration of protein polymers presented in the solution, since they were hydrophilic. To proceed further, the excess polymers left in the solution after the coating reaction were filtered by 100KDa filters instead of 3KDa filters in subsequent experiments.

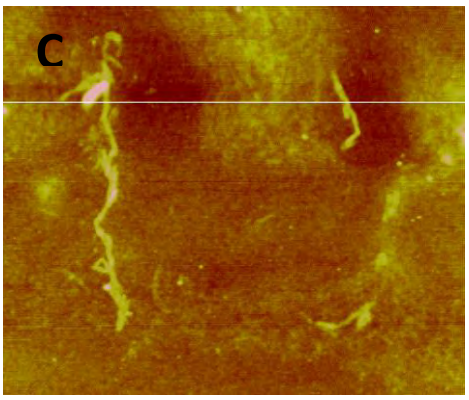
Atomic Force Microscopy

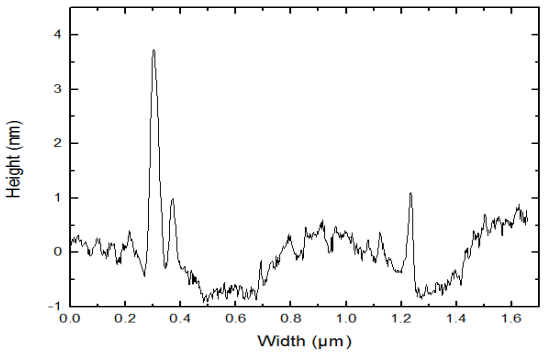












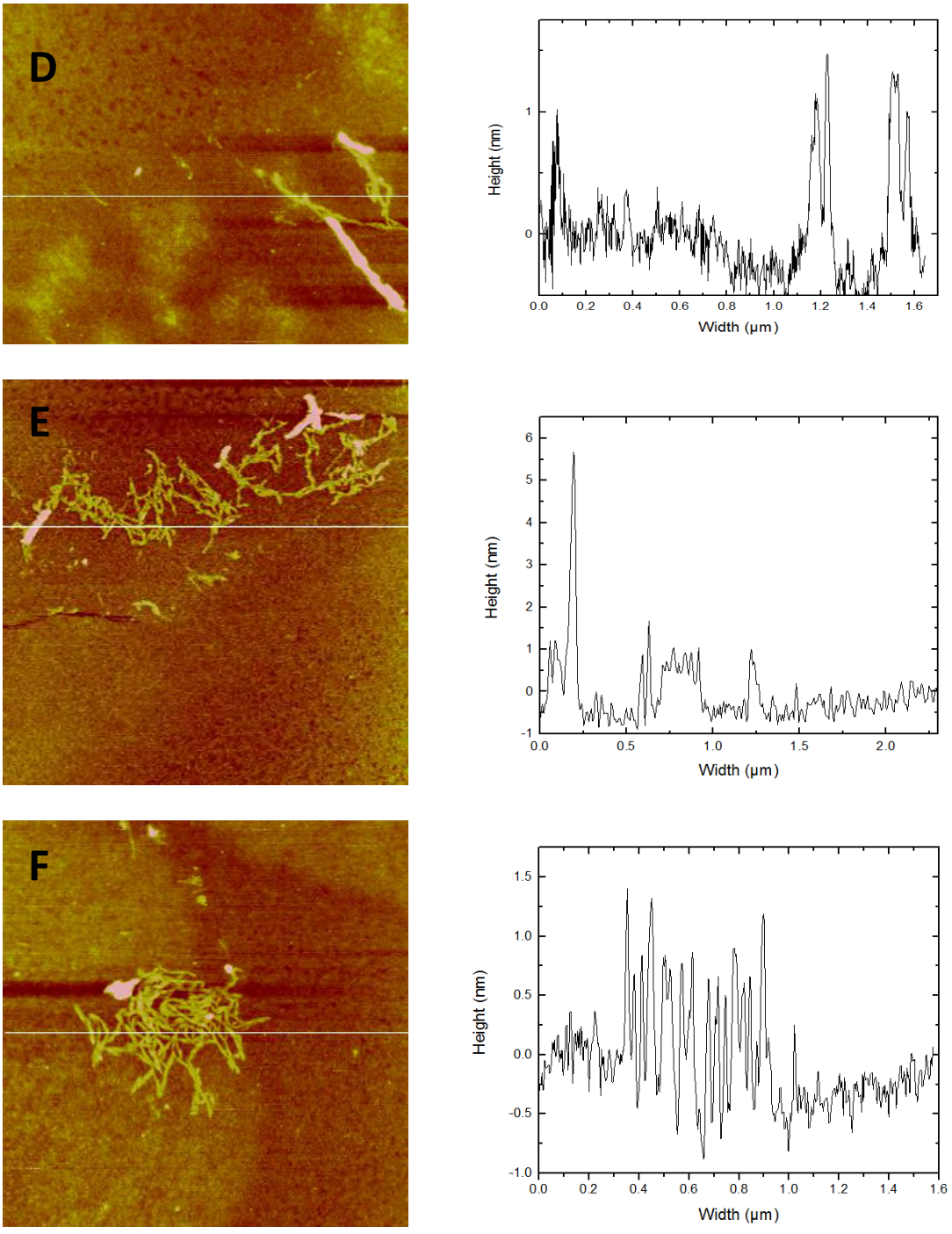


Figure 6. Left columns were AFM images taken in air of polypeptide-coated λ-DNA after denaturation on silica. Images on the right were corresponding height profiles of the line sections of the images on the right.

AFM images A and B showed that DNA coils with diameter on the order of 1 μm were formed. Molecule A had a contour length roughly equal to the contour length of λ DNA (16.5μm). The height, however, exceeded the expected height of double stranded DNA, which is about 0.68 nm. There was intramolecular bridging in the case of molecule A and intermolecular bridging in the case of molecule B, which has a contour length larger than one λ DNA molecule would have. The parts of the molecule with larger height (>6nm) should be protein polymer aggregate, other than DNA itself. This was concluded if we compared them to other parts with clear string like structures.

More linearized DNA molecules were observed as well (Image C, D). Judged by their small contour length and height, these molecules should be DNA fragments. The linearized parts extended to 0.5-1 μm, indicating that their persistence length was much longer than that of a single or double stranded DNA molecule. Therefore, the molecules must be coated. The parts without obvious overlaps were measured and had an average 0.6-1.2 nm height (Figure D, F). This could correspond to uncoated dsDNAs or coated ssDNAs and dsDNAs. Since we have already deduced that the molecules should be coated, the possibility was that the 0.6-1.2 nm parts were coated ssDNAs or dsDNAs.

According to our labs' preliminary work, the C_8S_7 coated single stranded λ DNAs had an extension of 30% the contour length in a 250-300 nm diameter nanochannel, compared to the 85% contour length extension of double stranded λ DNA coated by C_4K_{12} , which had a weight about half of the C_8S_7's (16). This result suggested that the DNA molecules were indeed denatured, but could have undergone intramolecular bridging that led to a smaller extension.

As the images shown, there were particularly high region along the molecules. They should be excess protein peptide aggregate. Accordingly, we needed to filter out excess polypeptides to achieve a uniform coating. A filter of 300KDa pores was then tested for its ability to allow the C_8S_7 polymers pass through.

Circular Dichroism

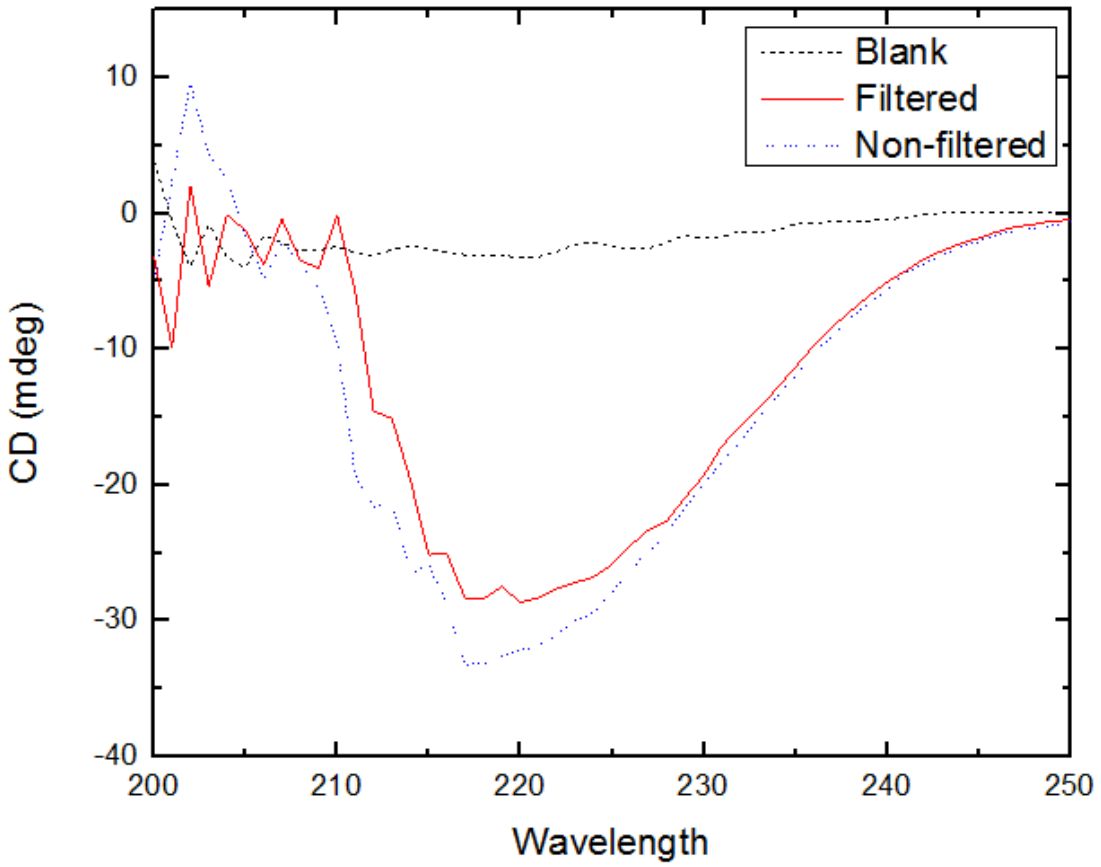


Figure 7. The Circular Dichroisms of three samples were measured. Solid line represented the sample filtered out by 300KDa tubes. Dotted line was unfiltered and the dashed line was blank TEx1 buffer.

The CDs of the filtered and unfiltered samples had a similar trend, while the unfiltered sample showed a larger CD in general. This result indicated that large parts of the C_8S_7 polymers could pass through the filter's 300KDa pores. Such filters could thus be used in future experiments in order to achieve a more uniform polymer coating. However, it was still possible that filters of this dimension cannot meet the requirement and further optimization was needed.

V. Conclusion

The feasibility of Quantum dots as labeling dyes for dsDNA site specific nick labeling and ssDNA oligonuleotide labeling was verified. However, to achieve genome mapping, a more efficient labeling protocal and imaging method was required. The hydrodynamic radius of the C_8S_7 polymer coated single stranded λ DNA in aqueous solution was determined to be $395 \pm 33\,\mathrm{nm}$. The AFM images confirmed the coating of the DNA molecules while partial excess coating was observed. The Circular Dichroism measurement suggested that the 300KDa filter could be an ideal candidate to eliminate the excess C_8S_7 polypeptides left in the solution after coating and result in a more uniform polymer coating.

References

1. Russell, P. J. (2010). *iGenetics A Molecular Approach 3rd ed.* pp.2-3 San Francisco : Pearson Education, Inc.

2. Artem, E. and Men, P. W. (2008). Sanger DNA Sequencing. In M. Janitz, *Nexet-Generation Sequenceing* (pp. 3-6). Wiley-VCH Verlag Gmbh & Co. kGaA.

3. Mardis, E. R. (2008). Next-Generation DNA Sequencing Methods. *Annual Review of Genomics and Human Genetics* , pp. 387-402.

4. Ming, X. and Pui-Yan, K. (2008). A Single DNA molecule Barcoding Method with Applications in DNA Mapping and Molecular Haplotyping. JanitzMichal, Next-Generation Genome Sequencing WILEY-VCH Verlag GmbH & Co.KGaA. pp. 117-120.

5. Treangen, T. J., and Salzberg, S. L. (2012). Repetitive DNA and next-generation sequencing: Computational challenges and solutions. *Nature Reviews.Genetics*, 13(1), 36-46.

6. Jing, J. et al (1998). Automated high resolution optical mapping using arrayed, fluid-fixed DNA molecules. *Proc Natl Acad Sci U S A.* 95(14): 8046–8051

7. Chan, E.Y. et al (2004). DNA mapping using microfluidic stretching and single-molecule detection of fluorescent site-specific tags. *Genome Res.* 14, 1137-1146

8. Lam, E. T., Hastie, A., Lin, C., Ehrlich, D., Das, S. K., Austin, M. D. … Kwok, P. (2012). Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nature Biotechnology*, 30(8), 771-6.

9. Zhang, C., Doyle, Patrick S., Vries, R., van der Maarel, Johan R C… and Dai, L. (2013). Amplified stretch of bottlebrush-coated DNA in nanofluidic channels. *Nucleic Acids Research* 41 (20), 1-8

10. Kenneth R. Spring, Michael W. Davidson. *Introduction to Fluorescence Microscopy.* Retrieved on March 26 2014 from http://www.microscopyu.com/print/articles/fluorescence/fluorescenceintro-print.html

11. G. Binnig, C. F. Quate, and Ch. Gerber. (1986) Atomic Force Microscope. *Physical Review Letters*, Volume 56, Issue 9, pp. 930 - 933

12. Johan R.C. van der Maarel (2008) *Introduction to Biopolymer Physics*. World Scientific. pp27-38.

13. Depeertment of physics, NUS (2007) *Circular Dichroism of Protein* Retrieved on March 26 2014 from http://www.physics.nus.edu.sg/~Biophysics/pc3267/CD-2007.pdf

14. Ute R.G., Markus G., Sara C., Roland N. and Thomas N. (2008) Quantum dots versus organic dyes as fluorescent labels *Nature Methods* 5, 763 – 775

15. Allemand, J.F., Bensimon, D., Jullien, L., Bensimon, A., and Croquette, V. (1997) pH-dependent specific binding and combing of DNA. *Biophys. J.,* 73, 2064-2070

16. Hernandez-Garcia, A., Werten, M. W. T., Stuart, M. C., de Wolf, F. A. and de Vries, R. (2012), Coating DNA: Coating of Single DNA Molecules by Genetically Engineered Protein Diblock Copolymers *Small*, 8: 3490. doi: 10.1002/smll.201290121

17. Hidehiro O., Kanta T., Yuko Y., and Kenichi Y. (2002) Folding transition of large DNA completely inhibits the action of a restriction endonuclease as revealed by single-chain observation, *FEBS Letters*, Volume 530, Issues 1–3, pp143-146