University *of Ljubljana*
Faculty *of Mathematics and Physics*

Department of Physics

Seminar

Pulling strands of DNA through tiny hole, can dramatically speed up the sequencing of human genome

Author: Tea Kokot Klančnik

Menthor: prof. dr. Rudolf Podgornik

Ljubljana, May 2012

Abstract

In the beginning, we are going to talk about human genome and sequencing of human genome. Then we are going to discuss about DNA sequencing, double-stranded DNA sequencing and single-stranded DNA sequencing through nanopore. On the other hand, we are going to talk about statistical physical model of polymer translocation through a pore in a membrane treated as the diffusion prosecc across free energy barrier and we determine the translocation time.

Contents

1 Introduction

The year 2007 marks the 30th anniversary of the introduction of modern dna sequencing method and the first complete sequence of a DNA molecule [1]. In genetics, sequencing means to determine the primary structure of an unbranched polymer. DNA sequencing is the process of determing the nucleotide order of a given DNA fragment. When we have defined sequencing in general, we can now focuse of the term nanopore sequencing, but first we have to explain what is nanopore; nanopore is simply a small hole, of order of 1 nanometer in internal diameter. Recent experiments show that a nanopore in a thin insulating membrane separating two ionic sulution filled chambers can serve as a single-molecule sensing device that provides insight into te distributin of static and dynamic molecular activities, properties or interactions. Most nanopore studies have relied on measurring ionic current blockages caused by dna translocating through a nanopore formed by the alpha-hemolysin(alpha HL) [2] protein in a lipid bilayer. There is a nanopore known as a solid-state nanopore, that is typically a nanometer sized hole formed in a sinthetic membrane. The pore is usually fabricated by focused ion beam or focused electron beam. Nanopore sequencing is now determined as a method for determining the order in which nucleotides occur on a strand of DNA.

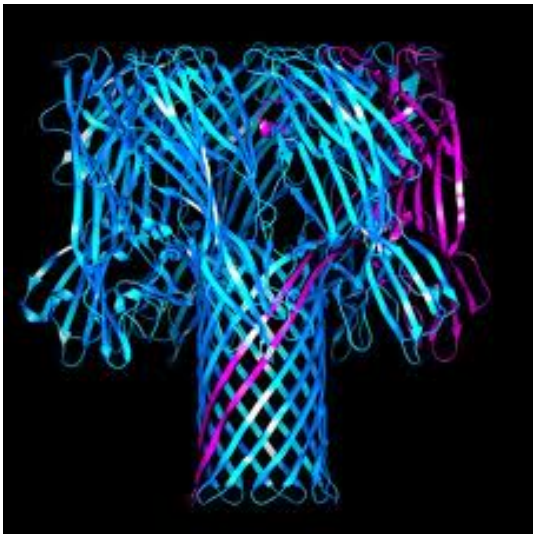

Figure 1: alpha-hemolisyn, a naturally occurring nanopore[3]

2 Human genome

Before we go more deeper in the DNA sequencing we are going to discuss about human genome and human genome sequencing. The human genome is stored on 23 chromosome pairs[4] and in the small mitochondrial dna(DNA located in organelles called mitochondria, structures with in eucaryotic cells that convert the chemicall energy from food into a form that cells can use, adenisine triphosphate).

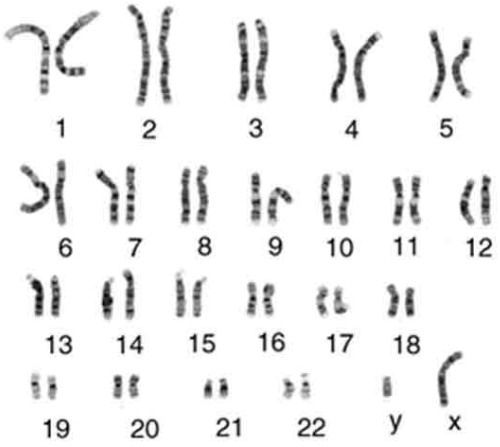

Figure 2: the chromosomes.[4]

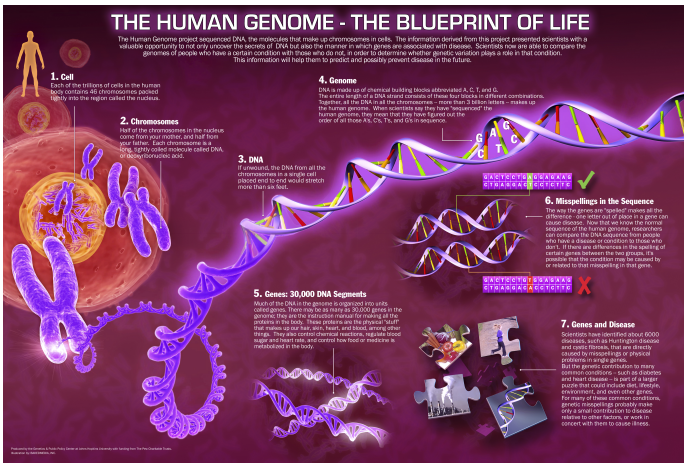

Figure 3: the human genome[5]

3 Human genome sequencing

First of all, full genome sequencing is a laboratory process that determines the complete DNA sequence of an organisms genome at a single time [6].

In principle, a genome of arbitrary size may be directly sequenced by the shotgun method [7], provided that it contains no repeated sequence and can be uniformly sampled at random. The genome can than be assembled using the simple computer science technique of hashing. Practical difficulties arise because of repeated sequences and cloning bias. Small amounts of repeated sequence pose little problem for shotgun sequencing. There are two approaches for sequencing large repeat-rich genomes; a whole genome shotgun sequencing and hierarchical shotgun sequencing. We are going to concentrate on the second approach. This approach involves generating a set of large-insert clones covering the genome and separately performing shotgun sequencing on appropriately chosen clones. In hierarchical sequencing, a low-resolution physical map of the genome is made prior to actual sequencing. from this map, a minimal number of fragments that cover the entire chromosome are selected for sequencing.

Now, when we know basic concept of human genome sequencing, we will extend our knowledge about DNA sequencing through nanopore.

4 Polymer translocation

Now we construct a new statistical physical model of polymer translocation through a pore in a membrane treated as the diffusion process across a free energy barrier. There have been a few efforts to investigate quantitatively the driving force of translocation on physical grounds [8]. There is a study via simulation the translocation of polymer directly through lipid bylayer, driven by the concentration in balance of lipids that exist at high-curvature regions in membrane. On the other hand, there are considered protein translocation through a channel or pore and postulated that its driving force is random thermal motion reflected by ratchets which give rise to directional diffusion. Now we consider a simple but tenable model for the membrane; a rigid wall of negligible thickness with a pore, which is assumed to be small enough to allow only a single segment passage. We describe the translocation dynamics as a stochastic process crossing the free energy barrier calculated from the chain configuration partition function. The conformation of a flexible polymer during its translocation is significantly effected by stering interaction with the membrane, leading to a reduction of the polymer entropy and increase of its free energy. We adopt as our model an ideal chain with N segments, each

with length b. First consider a chain with the initial segment anchored on a rigid wall introduced in yz plane. We define a boundary condition, that the other segments do not cross the surface. With that the probability of finding the end segment at r, given initial one at r_0 on surface $G(r, r_0; n)$ is given as the probability for all configurations in free space, the gaussian distribution

$$G(r, r_0; n) = (2\pi \cdot n \cdot \frac{b^2}{3})^{-\frac{3}{2}} \cdot exp(\frac{-3(r - r_0)^2}{2nb^2}) \tag{1}$$

minus the probability for the chain crossing the surface $G_0(r, r_0; n)$.

$$G(r, r_0; n) = G_0(r, r_0; n) - G_0(r, r_0; n) = (2\pi \cdot n \cdot \frac{b^2}{3})^{-\frac{3}{2}} \cdot \frac{Gxe}{nb^2} \cdot exp(\frac{-3r^2}{2nb^2})) \tag{2}$$

where $r_0 = (e.0.0)$ and e is an arbitrarily small distance of the anchored segment from the surface. The steric constraint factor of a chain, given as $Z_s(n) = \int_{x<0} G(r, r_0; n)\, \mathrm{d}r < 1$ scales as $n^{\frac{1}{2}}$ in the absence of the constraint, the partition function is given by

$$Z_b(n) = exp(-\beta \cdot \mu \cdot n); \beta = \frac{1}{k_b T} \tag{3}$$

.

$$\mu = (\frac{\partial F}{\partial n})_t \tag{4}$$

in the limit n-¿inf. $F(n)$ is the free energy given from the full partition function

$$F(n) = -k_{bt} \cdot ln(Z_s(n) Z_b(n)) = \frac{1}{2} \cdot k_b \cdot ln(n) + \mu \cdot n + const. \tag{5}$$

the whole chain during translocation can be decomposed into two independent end-anchored chains each in the oposite half spaces. For the decomposition into n and n-N segments[2], the total free energy is

$$F(n) = \mathcal{F}(n) + \mathcal{F}(N - n) = \frac{1}{2} \cdot k_b + ln(n(N - n)) - n\delta\mu + const. \tag{6}$$

where $\delta\mu$ is the excess chemical potential per segment of the trans side relative to that on the cis side. For the long-time scale behaviour of translocation we construct a coarse-grained description in terms of the translocated segment number n adopted as a relevant stochastic variable and in terms of the associated free energy barrier. It can be treated as a diffusive random process, which is described by a Fokker-Planck equation for P(n,t) [8]. The translocation time, defined by $\tau = \tau(N - 1; 1)$ is integrated to be

$$\tau = b^2 \cdot \int_{1n-1} \mathrm{d}n \cdot (\frac{1}{D(n)} \cdot exp(\beta \cdot F(n)) \cdot \int_{1n} \mathrm{d}n \tag{7}$$

we assume, that D (chain diffusivity) does not change in the course of translocation. In the case of the rigid chain without chemical potential defference $\delta\mu, F(n) = const.$ the translocation time is simply reduced to

$$\tau = \frac{L^2}{2D} = L^{2+\nu}; L = N \tag{8}$$

where L is the length of the whole chain. The exponent ν is 1 if the hydrodynamic interaction between the segments is neglected, and $\frac{1}{2}$ if it is included. To incorporate the chain flexibility effect, the free energy function should be included in equation for τ resulting in, for $\delta\nu = 0; \tau = (\frac{pi}{8} \cdot (L^2 2D)) = L^{2+\nu}$ the chain flexibility retards translocation by 23%, the translocation time is proportional to $N^{2\cdot\nu}$ if there in nonvanishing chemical potential difference, the translocation time can be calculated, having the analitycal expressions for limiting cases, where $|\mu^*| << 1; \mu^* << -1; \mu^* >> 1$

$\mu^* = N\beta\Delta\mu$

When the chemical potential per segment is reduced on the trans side, the translocation time is given by first two formulas. There is remarkable sensitivity of translocation to chemical potential asymmetry, and is a cooperative phenomenon arising from chain connectivity; the segment respond all hand in hand rather than as individuals to a driving asymmetry. The chain diffusivity can also change during translocation. The Brownian ratchet mechanism assumes fast chemical binding of chaperones on the chain entering the trans side of the membrane. The binding sites are assumed to be uniformly distributed with an interval of ρ along the chain. The whole space of translocation coordinate is divided into intervals of length $\rho = \frac{L}{M}$ where M is the number of binding sites. The range of the i-th interval is $(i-1)\alpha+1 < h < i\alpha+1; \alpha = \frac{N}{M}$. the number of polymer segments in each interval. The dynamics is now consecutive translocation of each interval subject to the free energy therein, as well as to the boundary conditions at both borders of the interval, reflecting boundary condition at the left border and absorbing boundary condition at the right. In general the translocation time can be written as $\tau = \frac{L^2}{2DM\omega(\mu',M)}$

5 DNA sequencing

The basic idea that DNA could be sequenced by running the strands through the tiny hole and reading off the bases by electrical detection, was suggested a couple of years ago [9]. Nanopores fabricated from graphene-sheets of carbon only one or a few atom thick-might have crucial advantages for this application. Graphene is a material with extraordinary electrical and mechanical properties.[9], [6] There is naturally occuring nanopore named alpha

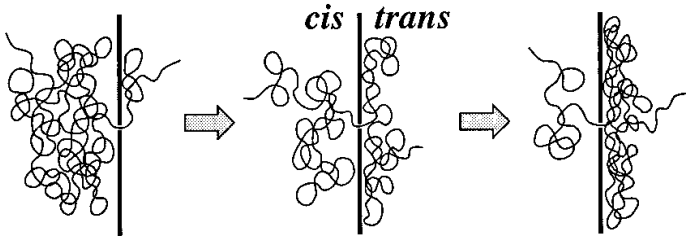

Figure 4: translocation of a polymer through a pore. polymer segments are absorbed on the trans side due to the attractive interaction at sufficiently low temperatures[2]

hemolisyn(alpha HL) and there is hypothesis, that we can use this nanopore to sequence DNA.

Single-molecule DNA sequencing using arrays of nanopores offers the possibility to obtain genome sequences in less time.

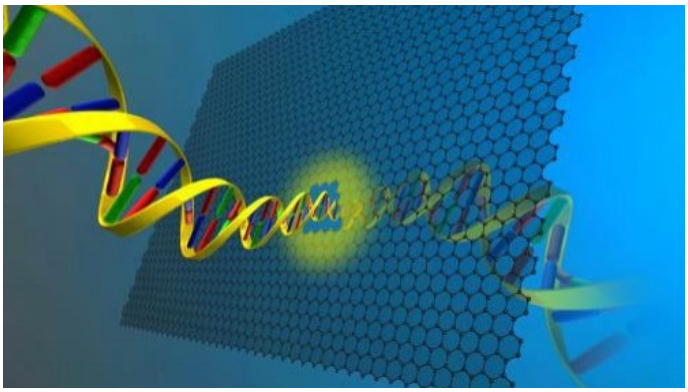

Figure 5: sequencing of DNA through a graphene[6]

6 Sequencing of ssDNA

Now we will theoretically show the possibility of a protocol for sequencing based on the distributions of the transverse electrical currents of single-stranded DNA while it translocates through a nanopore [10]. Estimates, based on the statistics of these distributions, reveal that sequencing of an entire human genome could be done with wery high accuray without parallelization. One of suggested method for sequencing DNA, is to measure blockade current. In this method, a longitudinal electric field is applied to

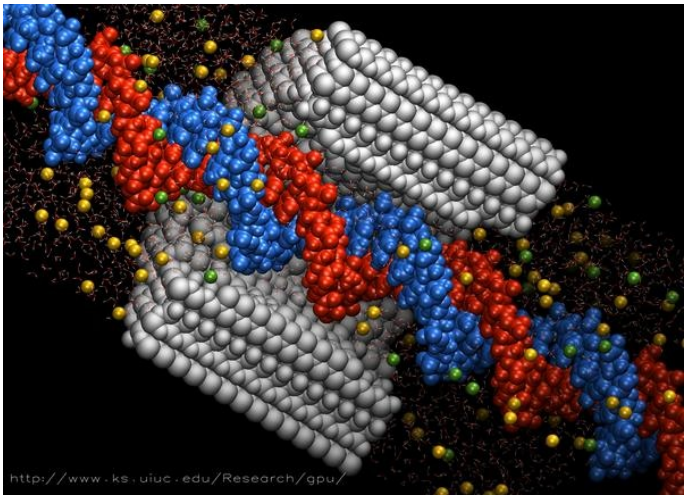

Figure 6: nanopore DNA sequencing[10]

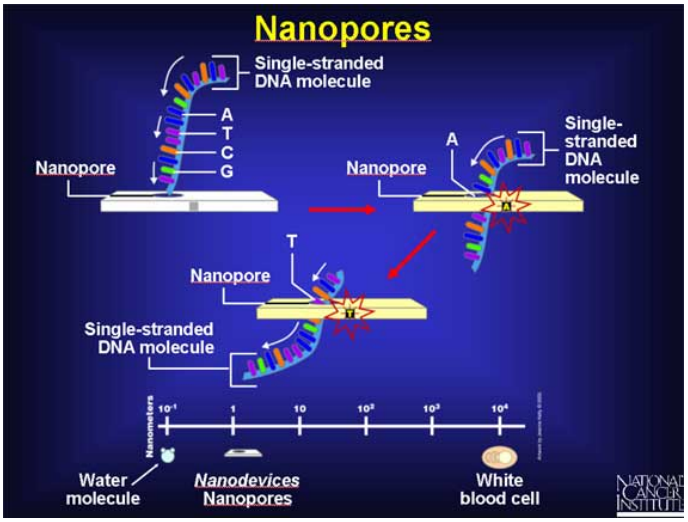

Figure 7: nanopore DNA sequencing[11]

pull DNA through a pore. As the DNA goes through a segnificant fraction of ions is blocked from entering the pore simultaneously. By measuring the ionic current continuously, one can detect single molecules of DNA. an alternative idea is one, that would allow single-base resolution by measuring the electrical current perpendicular to the DNA backbone, while a single strand immersed in a solution translocates through a pore. To do this, we envision embedding electrodes in the walls of a nanopore. The DNA is se-

quenced by using the measured current as an electronic signature of the bases as they pass through the pore. We use a Greens function method to calculate the current across the electrodes embedded in the nanopore. In that function we have $\sum_t, \sum_b$ which are the self-energy terms, describing the coupling between the electrodes and the DNA. Now we can express the total current. where we define $T(E)$ the transmission coefficient and is given by $T(E) = TL[\gamma_t \cdot \zeta_{DNA} \cdot \gamma_b \cdot \zeta_{DNA}]$. f_t, f_b is the Fermi-Dirac function of (top) bottom electrode, and $\gamma_{(t,b)} = i \cdot \sum_{t,b} - \sum_{t,b}$. The electrodes are comprised of 3x3 gold atoms arranged as a surface two layers deep and are biased at 1v. The electrode spacing is 12.5A. The results provide a good indication that DNA can be sequenced if its dinamicd through the pore can be controlled. Such control is provided by a transverse field of the same magnitude as that driving the current. Obviously in a real device there will allways be fluctuations of the current. These fluctuations are mainly due to two sources: structural fluctuations of the DNA, ions and water; noise associated with the electrical current itself, like thermal, shot and 1/f noise. The most significant source of noise is thus due the structural motion of the DNA and its environment. In the work,they have explored[ref] this structural noise by coupling molecular dynamics simulations with electronic transport calculations to obtain the real time transverse current of the ssDNA translocating through a $Si_{(3)}N_{(4)}$ nanopore.

The electric field generated by the electrodes is not included when the ssDNA translocates through the pore because the driving field is much larger in magnitude. Analyzing the current in two perpendicular directions gives us additional information on the orientation of a nucleotide inside the pore. Because of structural fluctuations and the irregular dynamics of the ssDNA, a single measurement of the current of each base is not enough to distinquish the different bases with high precision. We conclude that a distribution of electrical current values for each base need to be obtained. These can be done by slowing DNA translocation in the pore so that each base spends a larger ammount of time aligned with the electrodes. Most importantly, we find that when the field that drives the dna through the pore is smaller than the transverse field that generates the current on base that a time can allign with a pair of electrodes quite easily. This is due to the fact that dna backbone is charged in solution so that the position can be controlled by the transverse field. Main results are shown by the calculated distribution of transverse currents for each base in a realistic setting when the driving field is much smaller than the transverse field [12],[13]. We obtain these distributions by turning off the driving field and sampling the current while one base fluctuates between the electrodes.the distributions for each base are inded different.first, one needs to calibrate a given nanopore device by obtaining the

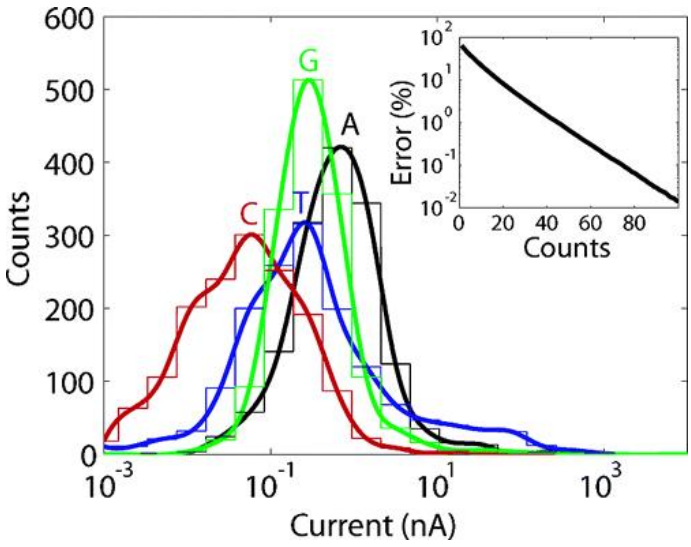

Figure 8: The calculated distribution of transverse current for each base in a realistic settings, when the driving field is much smaller than transverse field[13]

distributions of current with, for example, short homogenous polynucleotides, one for each base. Second, once this target distributions are obtained, a given sequence can be extracted with the same device by comparing the various currents with this target distributions and thus assigning a base to each measurement within a certain statistical accuracy. Finally we can show how many independent electrical current measurements one needs to do in order to sequence DNA within that accuracy. The number of current measurements will dictate how fast we can sequence. The average probability that we can correctly sequence a base after N measurements is given by

$$\langle P \rangle = \sum_{X=A,T,C,G} \frac{1}{4} \sum_{\{I_n\}} \frac{\prod_{n=1}^{N} P_X^n}{\prod_{n=1}^{N} P_A^n + \prod_{n=1}^{N} P_T^n + \prod_{n=1}^{N} P_C^n + \prod_{n=1}^{N} P_G^n} \tag{9}$$

where A,T,C,G are the distributions for the four bases. p_x Is the probability that a base is considering only the current for measurement h. The sum over I_n is the sum over all possible sets of measurements of size N. If we create a graph we see that 1-P the exponentially decaying ration of falsely identified bases versus the number of independent counts of the current averaged over the four bases, where the ensemble average is performed using Monte carlo methods.

7 Sequencing of dsDNA

Now we are going to report about fast dsDNA translocation through a solid-state nanopore-siliconoxide nanopore. Here we present a theoretical model, where hydrodynamic drag on the section of the polymer outside the pore is the dominant force counteracting the electrical driving force. First of all, solid-state nanopores were started to use, because they offer a range of obvious advantages such as tunable pore size and stability over a wide range of voltages, temperatures and buffers of varying salinity and PH. The translocation process consist of two separate stages. In the capture stage, a DNA molecule initially in solution in the negative reservoir has to come close enough to the pore to experience the electrostatic force and get pulled in. The driving force is only felt in the direct vicinity of the pore. The second stage is, where DNA passes the pore until it has reached the other side. We assume that one end of the DNA has entered the pore and calculate the time required for complete translocation. How fast the translocation is depends on the length of the polymer. We consider a linear polymer consisting of N nanomers, each of length b. This polymer is partially treated through a narrow pore. Time $t = 0$, sets the moment of initial capture. We will let $L(t)$ denote the contour length of the untranslocated part of the polymer, so that $L(0) = Nb = L_0$. A dwell time T is therefore determined by $L(\tau) = 0$. A second time scale in the problem is the charateristic relaxation time scale of the translocating polymer. The Zimm time, given by $t_z = 0.4\mu \cdot \frac{R^3}{k_b \cdot T}$ can be considered an upper bound on the time it takes the polymer to relax to an entropically and sterically favored configuration. μ Is the solvent viscosity and R is the radius of gyration of the polymer. That radius scales with the polymer length as $R = L_0^{\mu}$ where μ is the swelling exponent[**?**]. For translocation of DNA through $\alpha(HL)$ at room temperature, the measured velocity is about $0.8\mu s$ per base. When we compare this to the zimm time for the same polymer fragment, about $0.2\mu s$ we see that relaxation is much quicker than the translocation. If $t >> t_z$ we call that slow translocation, othervise that is a fast translocation. A full λ - phagegenome is found to take only arround 2ms to transverse a 10nm SiO_2 pore. There is zimm time much longer than translocation time. The important reason for fastness of our sistem is that we use dsDNA not ssDNA, because dsDNA has a much larger persistence length and consequently a longer relaxation time. First we consider driving force. Only the part of the polymer inside experiences the driving force. We can write it as $F_d = 2\frac{eV}{a}$ where e is the elementary charge, V is the potential difference and a=0.34nm is the spacing between nucleotides. The viscous drag per unit length in the pore can be estimated

as

$$2\pi \cdot \mu \cdot r \cdot \frac{\nu_{lin}}{(R-r)} \tag{10}$$

where R is the pore radius, r is the polymers cross-sectional radius, μ the solvent viscousity and ν_{lin} is the linear velocity of the polymer inside the pore. Finally, we estimate the hidrodinamic drag on the untranslocated part of the polymer outside the pore. We approximate the untranslocated part. As a sphere of radius Rg. As the polymer threads through the pore, the center of mass of this sphere moves toward the pore at a velocity $\frac{\mathrm{d}R_g}{\mathrm{d}t}$ assuming that the solvent inside the coil moves with the polymer, the coil experiences a stokes drag force of

$$6 \cdot \pi\mu R\nu_{blob} = 6 \cdot \pi\mu R_g \frac{\mathrm{d}R_g}{\mathrm{d}t}. \tag{11}$$

Clearly, in this case the hydrodynamics friction on the part of the polymer outside the pore is the dominant force counteracting the driving force. We propose that the principal effect of hydrodynamics is to resist motion with a hydrodynamic drag on the DNA coil. Such a drag force can be expressed as $\zeta_{eff}\nu_{blob}$ where ζ_{eff} is an effective friction coefficient proportional to the relavant length scale in the direction of motion and ν_{blob} is the velocity. The relavant hydrodynamic velocity is $\frac{R_g}{\tau}$ as the polymer radius must decrease from R_g to 0 during the τ. We write

$$F_{drag} = \zeta_{eff}\nu_{blob} = R_g \cdot \frac{R_g}{\tau} \tag{12}$$

the driving force must balance the hydrodynamic friction $F_d = -F_driv.$ as the driving force is constant during the whole process the same should hold for F_{drag}. this leads us to conclude that $R_g \cdot \frac{R_g}{\tau} = const.$ Results suggests a straightforward way of predicting the outcome of a wide range of translocation experiments. First, one determines the dominant contribution to the friction. In most cases, it suffices to compare the hydrodynamic friction inside the pore $F_{pore} = \zeta_{eff}\nu_{lin}$ to the stokes drag of the coil $6 \cdot \pi\mu R_g \frac{\mathrm{d}R_g}{\mathrm{d}t}$.

If the pore friction dominates, force balance with respect to the constant driving force implies that the translocation time scales linearly with the polymers length. A possible reason for a large pore friction could be the presence of specific interactions. When hydrodynamic drag dominates, we have shown that $\tau = R_g^2$. Depend on the length of the polymer, different regimes are thus obtained: when the polymer is short compared to its persistance length $R_g = L_0$ and we find that $\tau = L_0^2$. For long polymers we have shown that $\tau = L_0^{(2\nu)}$ [14].

8 Conclusion

In the last few years, there has been made a progress in creating a synthetic nanopore. when subjected to a soft beam of argon ions, an electron beam created hole in a polyamide film can be sculpted. The surface reorganizes in this sculpting process, slowly flowing material into the hole and closing it. By carefully monitoring the ion beam current, a hole of desired cross-section can be created. The focus initially is on measuring the time it takes for DNA to cross the pore. The final goal is to be able to call off the sequence of the bases as the strand flies through the pore at a very high speed. To summarize, we have investigated mechanism affecting polymer translocation through a pore in a membrane. polymer translocation is treated as the diffusion process across a free energy barrier. The conformation of a flexible polymer during its translocation is significantly affected by steric interaction with the membrane, leading to a reduction of the polymer entropy and increase of its free energy.

9 References

[1] Clyde A. Hutchinson. *DNA sequencing:bench to beside and beyond. Nucleic acid research*, 35(18):6227–6237, Sep 2007.

[2] W. Sung and P. J. Park. *Polymer translocation induced by adsorption. J.Chem.Phys.*, 108:3013–3018, feb 1998.

[3] T. A. van de Goor. *Nanopore Detection: Threading DNA Through a Tiny Hole*, Mar/Apr 2004.

[4] `http://www.biotechnologyonline.gov.au/popups/img_karyotype.html(26.4.2012)`.

[5] `http://www.dnapolicy.org/pub.graph.html(26.4.2012)`.

[6] `http://www.fastcompany.com/1665802/dna-nanopore-identification-gene-sequencing-medicine-roche-ibm(26.4.2012)`.

[7] institute for biomedical research center of genome research Stanford genome technology center Genome sequencing centres, human genome center. *Initial sequencing and analysis of the human genome. Nature*, 409:6227–6237, Feb 2001.

[8] W. Sung and P. J. Park. *Polymer translocation through a pore in a membrane. Phys. Rev. Lett.*, 77:783–786, Jul 1996.

[9] Hagan Bayley. *Holes with an edge. Nature*, 467(7312):164–165, Sep 2010.

[10] `http://www.web-books.com/eLibrary/ON/B0/B20/11MB20.html(26.4.2012)`.

[11] `http://textbookofbacteriology.net/themicrobialworld/chemoc.html(26.4.2012)`.

[12] Massimiliano Di Ventra Johan Lagerqvist, Michael Zwolak. *Fast DNA sequencing via transverse electronic transport. nano letters*, 6(4):779–782, mar 2006.

[13] Massimiliano Di Ventra Johan Lagerqvist, Michael Zwolak. *Fast DNA sequencing via transverse electronic transport*, mar 2006.

[14] Jianghua Chen Henny Zandbergen Jean-Francois Joanny Cees Dekker Arnold J.Storm, Cornells Storm. *Fast DNA translocation through a solid-state nanopore. Nano letters*, 5:1194–1197, jul 2005.