BIONANOTECHNOLOGY REPORT: DNA AND NANOPORE

Alicja Nowakowska, 234857, Bionanotechnology Lecture

SECTION 1. INTRODUCTION

Undubtedly, the DNA sequencing is one of the keys for the further progress in the life sciences. Its presence in both the research as well as in clinical applications is constantly increasing since the prominent Human Genome Project success. Therefore, development of sequencers that would be accurate, cheap, fast and capable of handling long and multiple sequences is crucial and occupies numerous researchers.

The DNA sequencing tools are often mentioned with regards to the generation they belong to. The first generation techniques date back to the 1970s when Frederick Sanger developed a foremost sequencing technology later called Sanger Sequencing Technology that spread throughout the scientific community. The method is based on a few steps that include the amplification of the DNA sequence of interest, then its denaturation in order to obtain a single strand that will serve as a template and usage of chemically modified nucleotides when the polymerase comes to action to create a complementary strand. These changed nucleotides are called dideoxynucleotides and terminate the chain at the specific point due to the lack of 3' OH group and deoxynucleotides – marked to be fluorescent or radioactive so that the sequence can be furtherly read. The four different reactions can take place depending on which ddNTP the reproduction was finished. This information is next used to apply the gel electrophoresis in four lanes and separate the chains by their size. Finally, from the generated bands we are capable of reading the sequence. This technology was used for the first read genome - the one of bacteriophage ϕX174 and later its modifications in Human Genome Project. Nowadays, it is considered to be a relatively slow and expensive method that does not allow multiple complex genomes sequencing though it happens to be used for small sequence clinical analysis [1][3].

Having in mind the obstacles connected to the first generation methods, the appearance of the second seems natural (happened around 2004). These new technologies primarily sped up the process and lowered the costs. It became possible to perform analysis on multiple short read (around 600 bases) in parallel without using the problematic electrophoresis. One of the significant examples is 454 sequencing that utilizes the pyrophosphate optical detection – the signal is given upon the new nucleotide binding. The DNA preparation is based on firstly the DNA fragmentation and ligation to a specific adapter molecules and further replication through PCR. The problems connected to this method are those of the applied detection method. If the signal is too high or too weak we underestimate or overestimate the number of identical nucleotides in the sequence. Although the method is like the Sanger's one "sequence-by-synthesis" type it had few crucial benefits: possibility of using natural nucleotides and real-time observation of the process instead of waiting for electrophoresis to happen. Some other second generation technologies include Ion torrent sequencing (detecting hydrogen ions when sequencing happens) or Illumina sequencing (synthesis and metal-oxide semiconductor based) [2][3].

Finally, we shall move to the third generation sequencers – the most modern ones that cure the problem of the replication that is time and money consuming. In the heart of the method we find

an idea of reading a very long sequences (dozen of kilobases) so that no cutting and amplification is needed anymore. The two technologies dominating this group are Pacific Biosiences' single-molecule real-time (SMRT) and Oxford Nanopore Technologies sequencing (ONT). SMRT devices "detect fluorescence events that correspond to the addition of one specific nucleotide by a polymerase tethered to the bottom of a tiny well" [1] what leads to the so-called polymerase read. The great advantage of SMRT apart from longer reads is its possibility of catching common DNA modifications such as methylation. Nevertheless, regarding the accuracy it is often criticized to be significantly lower comparing to some of the widely-used second generation methods such us Illumina (randomly distributed errors with a rate of about 13%) [1][4].

The nanopore technology that will be discussed in this report is based on a very different idea than that of SMRT. First of all, it shall be recalled that the protein nanopores, that are naturally spread across membranes, are very small holes that create entrances. The most common nanopore used is α-hemolysin, toxin from *Staphylococcus aureus*. It does have excellent sensitivity with a relatively low noise level though their main drawback is a fixed diameter. On the other hand, synthetical nanopores such as those constructed from silicon nitride seem very propitious due the higher stability and more controllability during the fabrication with regards to the thickness and diameter yet a higher background noise characteristic for them is an obstacle [5][8].

The ONT technology makes a clever use of the nanopores by inserting them on a membrane that is placed in an electrolyte solution. The membrane itself is highly resistant in terms of electricity though if the appropriate potential is applied, the ions can pass through the nanopore leading to the ionic power generation. The DNA sequence goes through the nanopore strand by strand while these remain linked to one another by a hairpin. Upon the strand translation through the pore, that happens with the help of a motor protein, the changes specific for each nucleotide appear in the observed ionic current and are recorded. Next the plot can be created and bases read according to the characteristic variations[5].

The ONT provides a significant advantages: small and portable devices for sequencing, low-cost, no need to mark nucleotides and long sequence reading in real-time. Nevertheless, similarly to SMRT it is also error-prone (around 12% [1]). Because of the relatively unstable translocation times the identification of how many repetitions appeared in a row results to be a remarkable challenge as in the second generation methods [6]. In order to fix these issues it has been proposed to combine the technique with Illumina and perform many such runs.

As a final remark from this section it should be mentioned that apart from the fact that the third generation technologies seem to be very promising their further development still is to be done. Like as not our traditional experimental approaches are not good enough to provide us with a fast breakthrough due to the scale they must deal with. In consequence, the call for the computational methodology is resonating in the air especially with regards to the nanopore technology that has already been developing for 30 years and still faces significant problems.

SECTION 2. COMPUTATIONAL APPROACHES

As it has been already mentioned nanopore sequencing faces many challenges. Although the smart mechanism of DNA translocating through the nanopore and measuring the current seems to be easy in fact is a very complicated process. It was shown that DNA strand moves very fast and adapts various hard to predict conformations that influence on the quality of the measurement. It can even happen that the due to this relatively flexible behaviour inside the pore, DNA bases do not pass in a proper order through the signalization point [7]. If we were to improve the situation and introduce some modifications to the system architecture the nano

scale imposes serious problems in terms of performing a reliable experimental investigation of the issues. Accordingly, the computational approach appears to be an attractive option.

The molecular dynamics studies of solid-state nanopores (e.g. Si_3N_4) have already demonstrated their meaningfulness by providing the atomic resolution studies of ssDNA and ions translocation phenomena. In [9] through an all-atom simulation model the investigators showed that the differentiation between the homo-oligonucleotides is possible because of the informative ion-nucleotide interactions (measuring the ionic current of K^+ and Cl^- through nanopore and nanochannels) though the signals were hidden in random thermal fluctuations and therefore hard to extract. In [10] it was observed that the increase in nanopore size that leads to the weak potential drop across the sizeable nanopore causes the DNA to move slower through the solid-state nanopore (what was experimentally confirmed). Nevertheless, the larger the size the lower the blockade ionic current amplitude and the lower the difference between the signal and noise and thus more difficult sequencing. In [11] the effect of the partial DNA charge reduction was investigated by means of MD. It was noticed that the decrease in the counterions size increases the DNA translocation time. Unquestionably, already these few cited works give us a strong idea on how powerful the in-silico methods can be.

SECTION 3. TASKS REALIZATION

This report was done by following the "Bionanotechnology Tutorial" of Alek Aksimentiev and Jeffrey Comer. For all the tasks both VMD and NAMD2 were used. The solid-state nanopore of Si_3N_4 along with DNA strands are analysed.

Task 1. Branching pore

In the first task the branching pore, that is the one that has an interior Y-shaped encounterable in a nanofluidic application, was to be constructed. The following coordinates rules were applied in the drillBranchedPore .tcl file inside the procedure insidePore{x,y,z}:

$$\text{if } z < 0 \text{ then } x^2 + y^2 < r_0^2$$

$$\text{if } z > 0 \text{ then } y^2 + \frac{1}{5}(z - 2x)^2 < r_1^2 \text{ OR } y^2 + \frac{1}{5}(x + 2x)^2 < r_1^2$$

The corresponding code fragment and the obtained result are presented in the Figure 1 and 2. The criteria were well applied and a desirable shape was found.

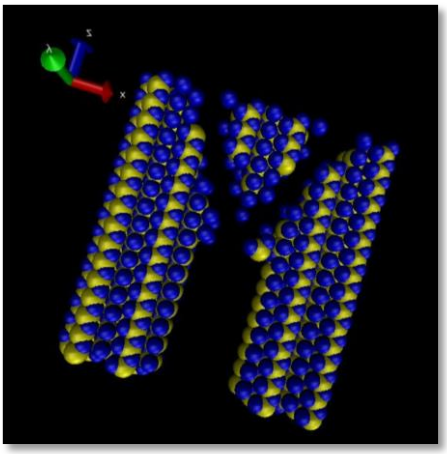


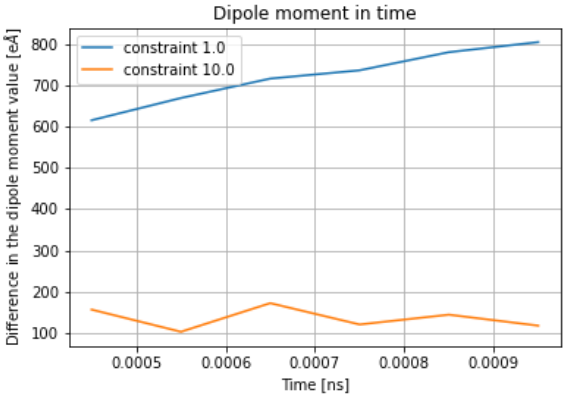
Figure 1. Y-shaped interior of the nanpore

```
Plik Edycja Format Widok Pomoc
# Cut a branched pore in a membrane.
# Use with: tclsh drillBranchedPore.tcl
# jcomer2@uiuc.edu

# Input:
set pdbIn block_hex.pdb
# Output:
set pdbOut branch.pdb

# Determine whether the position {x y z} is inside of pore and
# should be deleted.
proc insidePore {x y z} {
        set r0 8.
        set r1 5.

        if {$z < 0} {
                set isIn [expr $x*$x + $y*$y < $r0*$r0]
                return $isIn
        } else {
                # NOTE: Add your criteria to the expr commands.
                set isIn0 [expr $y*$y+0.2*($z-2*$x)*($z-2*$x)<=$r1*$r1]
                set isIn1 [expr $y*$y+0.2*($z+2*$x)*($z+2*$x)<=$r1*$r1]
                return [expr $isIn0 || $isIn1]
        }
}
```

Figure 2. Corresponding code modification

Next steps

After the realization of the first task the anaysis of the force field calibration in respect to the dielectric constant, that is of a great importance due to the electrostatic interactions dominance in the research, was performed. The electric field was applied to the nanopore with no free surfaces and the dielectric dipole moment was measured. The results for two different values of the constraints are presented in the figure and table below.



Constraint	Mean dipole	Std. error	Dielectric constant
1.0	619.89	97.79	2.29
10.0	116.22	19.81	7.85

Table 1. Plot and statistics for the dielectric constant calculation

It can be observed that the curves are rather stable in time and the great difference between them is apparent. The 10.0 constraint provides us with a good approximation of the real value which was experimentally calculated in [13] to be 7.2. This effect of the constraint value on the quality of the dielectric constant has its roots in the flexibility of the molecules that is imposed by this variable. The higher the constraint, the lower the rigidity and thus the change of the intermolecular interaction in response to the local environment is better reproduced.

Next, the system needed to be prepared for the MD simulation. The calibration was continued, water and salt (2 mol/kg KCl concentration) was added to mimic the physiological conditions (Figure 3.), the minimization was performed.

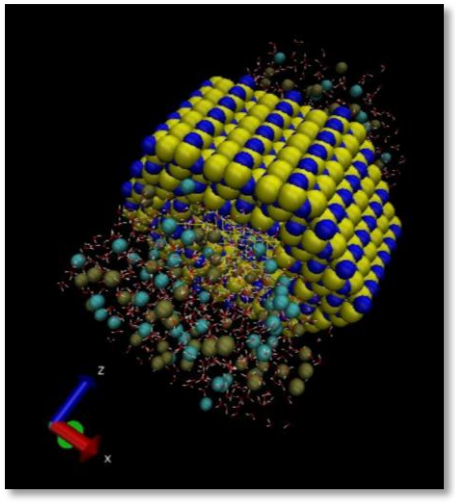


Figure 3. Nanopore, solvent and salt

The simulations with constant volume (raising temperature from 0 to 295 K, 500 steps) and with constant pressure (Langevin thermostat, 1000 steps) were performed for the equilibration goals. Nevertheless, as it might be seen in the Figure 4. such timing was not enough for the system since the volume still at the end of the run experiences down trends instead of fluctuations around some value what brings the conclusion that the system did not equilibrated. In further steps the ready to use files with an equilibrated system are utilized.

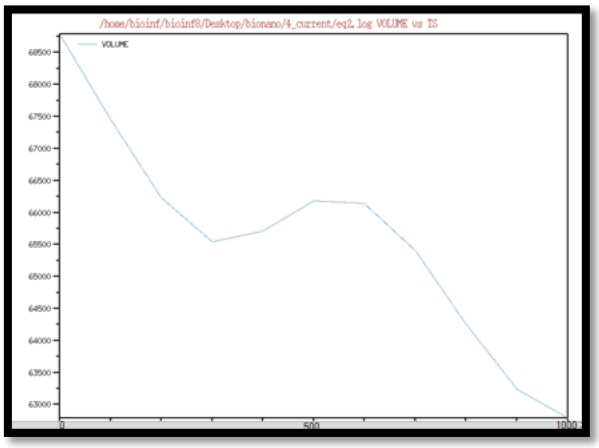


Figure 4. Equilibration run - volume in steps

The so constructed nanopore underwent the MD simulation with an applied electric field with the voltage of 20V. The temperature was plotted in steps (Figure 5.) and a significant raise in its value was observed as a consequence of applying a large ionic current. The current itself was also calculated for the aims of the Task 3. where the comparison of its value of the system with DNA and without it is performed.

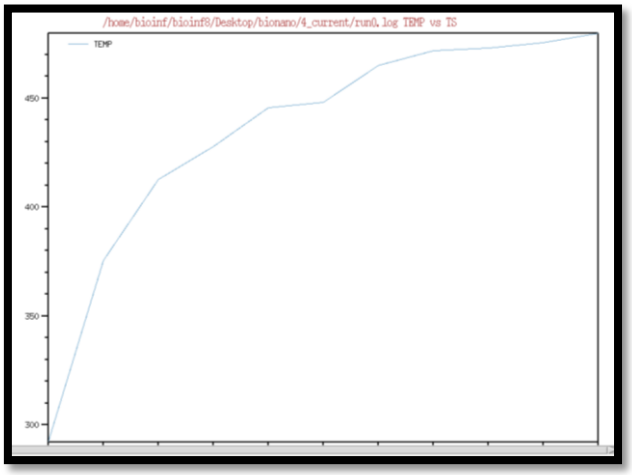


Figure 5. Temperature in simulation steps

Task 2. A different conformation

Before performing the nanopore-DNA strand simulation different DNA manipulations are done and for this purpose the 8-base pair double stranded DNA from the tutorial files is used (Figure 6). ADNA has the following base order: ADE, ADE, THY, THY, GUA, THY, GUA, ADE. No difference in reference to the AMBER and CHARMM model for the DNA was observed. The single strand DNA was also generated (Figure 7.)

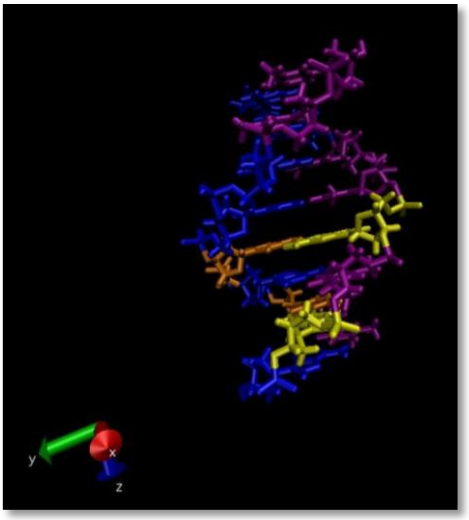


Figure 6. dsDNA (GUA yellow, ADE blue, THY violet, CYT orange)

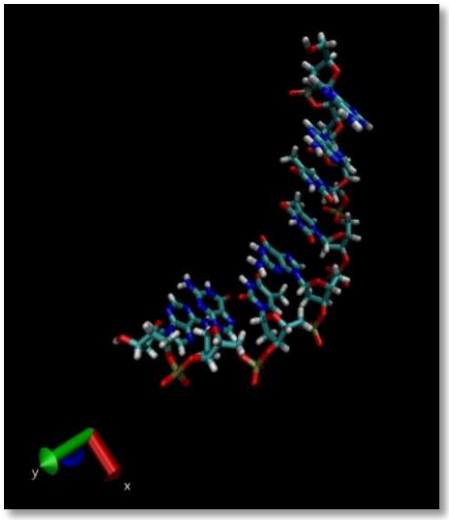


Figure 7. ssDNA

The ssDNA was modified to obtain different conformations using Sculpt and for one conformation a little bit similar to the U-shaped one but with some modifications applied the blocking positioning with regards to the nanopore was done what is shown in the Figure 8.

The great advantage of the in-silico modelling is the investigation of DNA conformations impact on its relation with the nanopore. The computational approach allows atomic resolution checking whether the specific conformation is adequate or maybe shall be somehow avoided. The taken space localization shall be rather avoided in the nanopore sequencing. It is evident that such blocking wouldn't lead to the correct read. It could even happen that the strand would not go through the pore or would go through it in an inappropriate order. Moreover, the great advantage of such visualizations may be noted in this example – observation of the space conformations from different perspectives. The first perspective does not clearly reveal the problem that we are facing while the second one does. It could be the case in classic experimental approach that the pore with the strand could be analysed only from the first perspective what would could lead to the false conclusions.

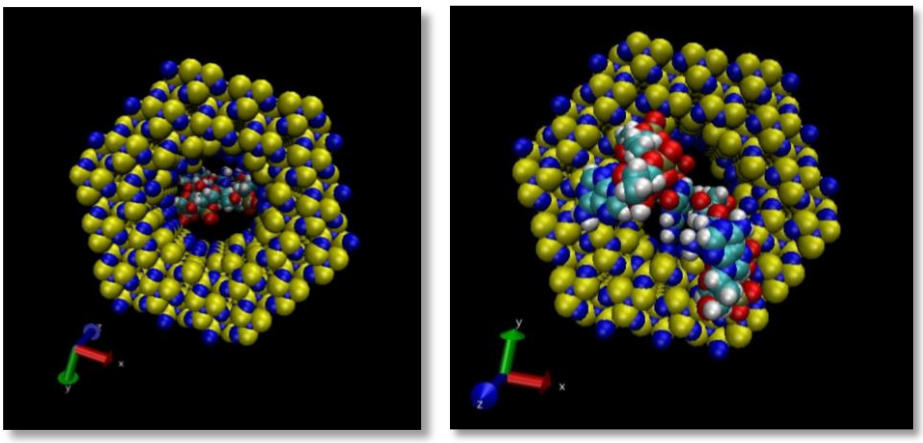


Figure 8. Blocking the nanopore with ssDNA – different perspectives

Another important aspect is that such a parallel localization still would cause the drop in the ionic current that would be of a comparable value to the perpendicular space location what was shown in [12]. Therefore one should remember that the decrease in the current does not need to

indicate the DNA translocation through the nanopore. In consequence, not only the current measuring matters but also knowledge about the DNA conformation is of great importance. Nevertheless, it was observed that the cure for this ambiguity is the usage of a multi-layered silicon wafer that provides the differentiation between these two signals in terms of the current values.

Task 3.

In the final part the MD simulation of length of 0.001 ns (1000 steps) with temperature equal 295K and usage of Langevin thermostat of the nanpore system with water ions and ssDNA was performed as in the part "next steps" when the system simulation was run without the ssDNA. In the Figure 9 the exemplary framework from the simulation is presented and in the Figure 10 the current values are plotted for both situations.

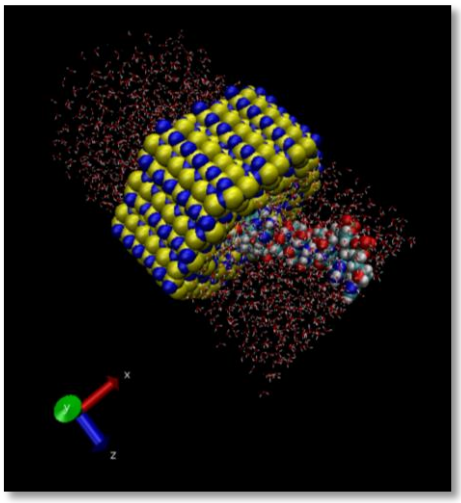


Figure 9. Frame from MD simulation with DNA

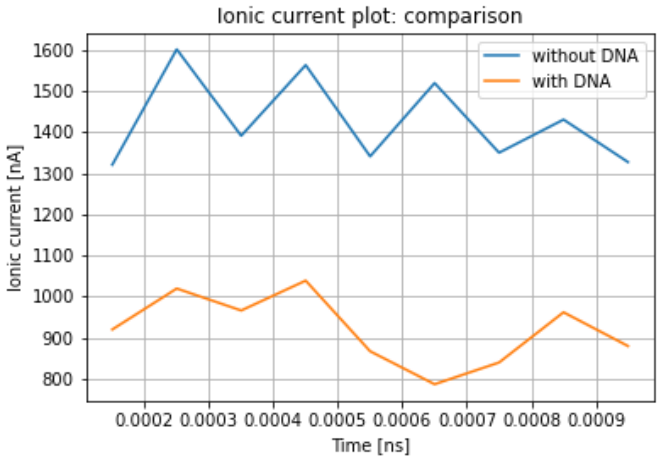


Figure 10. Current comparison

Mean without DNA : 1426.7491 nA

Mean with DNA: 920.1458 nA

When we apply the voltage across the membrane as it is done in the first simulation without the ssDNA, the system's ions flow according to the electric field. Under these circumstances the ion current is present and can be measured as in it is shown in the plot (blue curve). On the other hand when we analyse the system with the DNA that is placed closely to the nanopore the significant drop in the value of the ionic current is noted. Such situation can be explained by the fact that large charged molecules like DNA simply physically block the ionic flow what causes the reduction of the current. Furtherly if the ssDNA was to translocate through the nanopore, should the simulation be much longer (few hundred nanoseconds), even higher drops would happen and their value would depend on the base passing. In both simulations we observe the fluctuations – these are natural due to the dynamic nature of the atoms – e.g. DNA can take different conformations blocking less or more the ionic flow.

Finally, it should be mentioned that the current values themselves should be treated carefully. It is the case that the bigger the voltage applied in the simulation the more can be seen in the simulation – everything happens faster. Nevertheless, it is a double-edged sword – such big voltage numbers are not realistic with regards to the experiment in which they would be significantly smaller and thus the resulting average current could have a different value. However, the observation of the difference when the DNA is present and when it is not is indisputable.

SECTION 4. DISCUSSION

In this report the power of the computational approach in respect to the bionanotechnological techniques such as nanopore DNA sequencing has been shown. It turned out that using Molecular Dynamics we are capable of reproducing various system parameters and model different DNA conformations and the consequences of such. This not only provides us with an in-silico laboratory where we can check the experimental observations but also gives as an unique insights into new possibilities, challenges and in the further analysis hints for the solutions that such technology can encounter. Although, it is often indicated that the valuable simulation should be as long as possible with a time step as small as it can be, it was exposed that even very short and fast simulations can give us beneficial notion about the nanoscale process. In fact, it was possible to reproduce the critical impact that the DNA presence next to the nanopore has on the ionic current and analyse different DNA conformation.

The technical aspect of the Molecular Dynamics simulations itself was also considered. It was emphasized that in the bionanotechnological world we are constantly dealing with new combinations of materials what always brings an urgent need for the proper individual-approached calibration of the force field. Such a task is challenging and always shall be treated with caution and notion of which parameter has the greatest importance for the reproduction of the specific experimental variable. In the analysed case the reproduction of the experimental value of the dielectric constant was done and the particular influence of the constraints was shown. By the same token, the criticality of the careful investigation of the system after equilibration was underlined when it turned out that too short running times may cause its lack of appearance. Finally, it should be remarked that in order to perform relevant MD simulation one needs both the detailed knowledge about the system as well as accurate understanding of the algorithms applied.

Obviously, the topic of the nanopore and DNA system analysis with means of MD was not exhausted and the simulations of the translocation events would be an interesting option. Similarly, the incorporation of different salts instead of KCl and performing longer run in general would bring us more information about the system. The general choice of MD over the

experimental techniques is not a legitimate path. The computational approaches like this one has their own drawbacks connected mainly to the computer power they can operate in and therefore the results accuracy can vary and any obtained conclusion should be examined and checked in a wet lab. As such, the combination of techniques both seems to be the best option.

REFERENCES

[1] Generations of Sequencing Technologies: From First to Next Generation, Mehdi Kchouk, Jean-François Gibrat and Mourad Elloumi, Kchouk et al., Biol Med (Aligarh) 2017, 9:3, DOI: 10.4172/0974-8369.1000395

[2] 454 Life Sciences: Illuminating the future of genome sequencing and personalized medicine, Kristin L. Patrick, Yale J Biol Med. 2007 Dec; 80(4): 191–194.

[3] The sequence of sequencers: The history of sequencing DNA, James M.Heather, Benjamin Chain, Genomics, Volume 107, Issue 1, January 2016, Pages 1-8

[4] Opportunities and challenges in long-read sequencing data analysis, Shanika L. Amarasinghe, Shian Su, Xueyi Dong, Luke Zappia, Matthew E. Ritchie & Quentin Gouil, Genome Biology volume 21, Article number: 30 (2020)

[5] Chapter 5 - Sequencing in Precision Medicine, StephanVilgis Hans-Peter Deigner, Precision Medicine, Tools and Quantitative Approaches 2018, Pages 79-101

[6] Nanopore sequencing technology and tools for genome assembly: computational analysis of the current state, bottlenecks and future directions, Damla Senol Cali, Jeremie S Kim, Saugata Ghose, Can Alkan and Onur Mutlu, Brief Bioinform. 2019 Jul; 20(4): 1542–1559. Published online 2018 Apr 2. doi: 10.1093/bib/bby017

[7] Single-Stranded DNA within Nanopores: Conformational Dynamics and Implications for Sequencing; a Molecular Dynamics Simulation Study, Andrew T. Guy, Thomas J. Piggot, and Syma Khalid, Biophys J. 2012 Sep 5; 103(5): 1028–1036.

[8] A low-noise silicon nitride nanopore device on a polymer substrate, Wook Choi,Eun-Seok Jeon,Kyoung-Yong Chun,Young-Rok Kim,Kyeong-Beom Park,Ki-Bum Kim,Chang-Soo Han, PLOS ONE, July 20, 2018, https://doi.org/10.1371/journal.pone.0200831

[9] Molecular Dynamics Simulation Study of Transverse and Longitudinal Ionic Currents in Solid-State Nanopore DNA Sequencing, Mohsen Farshad, Jayendran C. Rasaiah , ACS Appl. Nano Mater. 2020, 3, 1438−1447

[10] Effect of nanopore size on poly(dT)30 translocation through silicon nitride membrane, SI Wei, SHA JingJie, LIU Lei, QIU YingHua & CHEN YunFei, SCIENCE CHINA Technological Sciences, October 2013 Vol.56 No.10: 2398–2402 doi: 10.1007/s11431-013-5330-2

[11] Slowing down DNA Translocation through a Nanopore in Lithium Chloride, Stefan W. Kowalczyk, David B. Wells, Aleksei Aksimentiev and Cees Dekker Nano Letters, 12(2), 1038–1044. doi:10.1021/nl204273h (2012)

[12] Aleksij Aksimentiev, Jiunn Benjamin Heng, Gregory Timp, and Klaus Schulten. Microscopic kinetics of DNA translocation through synthetic nanopores. Biophysical Journal, 87:2086-2097, 2004. (PMC: 1304610), http://www.ks.uiuc.edu/Research/nanopore/

[13] Dielectric properties of silicon nitride ceramics produced by free sintering, O.A. Lukianova, V.V. Sirota, Ceramics International 43 (2017) 8284-8288