DNA sequencing with nanopores

Grégory F Schneider & Cees Dekker

Major hurdles in the quest to sequence DNA with biological nanopores have now been overcome.

Sequencing DNA with nanopores offers exciting potential advantages over other sequencing technologies, but thus far the reading of the bases from a DNA molecule in a nanopore has been hampered by the fast translocation speed of DNA together with the fact that several nucleotides contribute to the recorded signal. In theory, the basic concept is straightforward: pass a DNA molecule through a nanoscale pore in a membrane from head to tail, and read off each base when it is located at the narrowest constriction of the pore, using the ion current passing through the pore to probe the identity of the base.

In this issue, two groups, Cherf *et al.*[1] led by Mark Akeson and Manrao *et al.*[2] led by Jens Gundlach, independently provide the first reports that sequence information can be obtained upon traversing a DNA molecule through a protein nanopore. Both groups use a polymerase (**Fig. 1a**) to slow DNA translocation through a protein nanopore to a speed that is conducive to reading nucleotide-specific current levels, which paves the way for sequencing. Together with UK-based Oxford Nanopore Technologies' recent announcement of the first commercial nanopore sequencing devices, these studies represent a clear step toward sequencing of single DNA molecules using nanopores.

DNA sequencing is a rapidly growing industry, making inroads in basic research, clinical medicine, pathogen epidemiology, forensic science and agricultural and industrial biotechnology. A variety of 'conventional' sequencing techniques exist, each with their own pros and cons[3]. Most techniques currently on the market require extensive biochemical labeling and sample preparation, and importantly do not allow long read lengths to be achieved.

Grégory F. Schneider and Cees Dekker are in the Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, Delft, The Netherlands.
e-mail: c.dekker@tudelft.nl

The idea of using nanopores for DNA sequencing was proposed more than 20 years ago[4]. The first experimental proof of translocation of DNA molecules was reported in 1996 using the biological protein pore α-hemolysin[5] (**Fig. 1b**), albeit without sequence information. It has taken significantly more time to establish that single nucleotides can be discriminated within the traversing DNA, most prominently because the speed of translocating DNA has been too fast (on the order of a microsecond per base) to identify individual bases[6].

Several strategies have been tried to address the fast translocation speeds. Single-stranded DNA molecules were statically captured in the α-hemolysin nanopore using a DNA hairpin or a protein attached to the DNA end, either of which are too large to enter the pore. This allowed the detection of single-nucleotide mutations in the immobilized DNA[7–9]. In other work, α-hemolysin was modified to include a

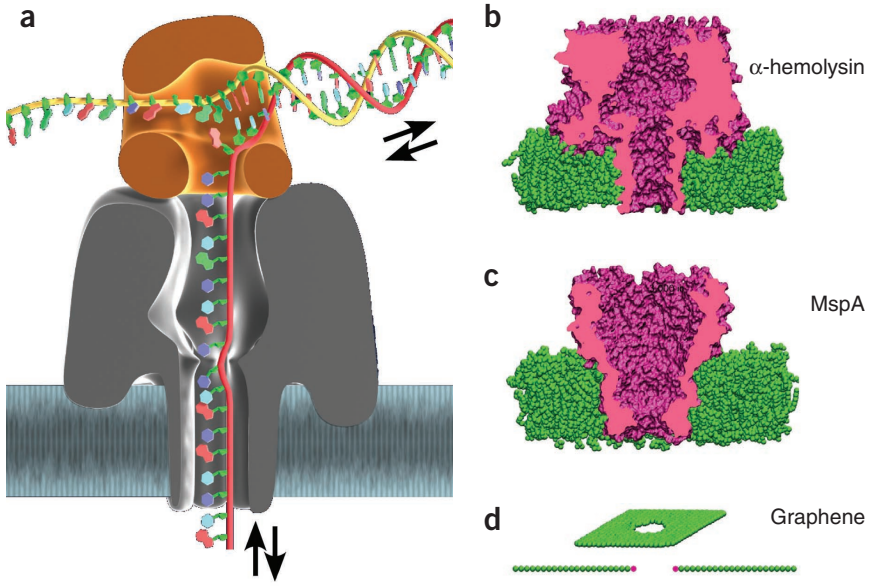

Figure 1 Nanopores for DNA sequencing. (**a**) DNA inserted in a nanopore, with speed control provided by a phi29 DNA polymerase (brown). An α-hemolysin nanopore (gray) is embedded in a lipid bilayer that separates two chambers containing buffered potassium chloride solution. The DNA substrate (red backbone) is inserted into the pore by an applied electric field. Its motion into or out of the pore (arrows) can be controlled by the applied electric field and the polymerase activity. DNA sequence information is obtained by changes in the ionic current running along the DNA through the nanopore, which occur as the DNA is ratcheted through the pore by the polymerase. This image has been extracted from Supplementary Video 1 from Cherf *et al.*[1]; the full movie shows the blocking strategy in better detail. (**b**) Structure of an α-hemolysin nanopore inserted into a lipid bilayer. (**c**) Structure of a MspA nanopore inserted into a lipid bilayer. (**d**) A graphene nanopore drawn on the same scale as the α-hemolysin and MspA protein pores in **b** and **c**; inset above represents a three-dimensional rendering of the graphene nanopore. Green spheres (**b**–**d**) represent individual carbon atoms in the supporting membrane. Purple color represents atoms in the close vicinity of the nanopore. Pink displays the cross-section of the nanopore at its center. Credit: Iemedia solutions (**a**) and Aleksei Aksimentiev at the University of Illinois (**b**–**d**).

cyclodextrin ring that binds free mononucleotides, which therefore reside in the pore for long enough (up to 10 ms) to be distinguished by different ionic current levels for each of the four bases[10].

More recently, the Akeson group has advanced the use of DNA polymerases that drive a DNA template through a nanopore in single-nucleotide steps as DNA is synthesized. They found that a DNA polymerase from the phage phi29 was potentially suitable to this approach because it remained bound to DNA, even against the force of an applied voltage needed to insert the DNA into the pore[11]. The polymerase, which processes DNA at a rate of about one nucleotide every ten milliseconds or slower, lowered the translocation speed of DNA by four orders of magnitude compared with freely translocating DNA. This study, however, relied on the transient chemical protection of the DNA primer to prevent elongation and excision in bulk phase, yielding only a short timeframe to capture and process DNA molecules from bulk (~20 minutes).

Now, Cherf *et al.*[1] describe a 'blocking oligomer' strategy in which up to 500 DNA molecules can be sequentially translocated through an individual nanopore over a period of many hours. To achieve this, they bound phi29 DNA polymerase to a single-stranded DNA template hybridized to a primer, and threaded the single-strand end through an α-hemolysin nanopore using an applied voltage. As the polymerase extends the primer to synthesize double-stranded DNA, it acts like a motor that ratchets the DNA through the pore in single-nucleotide steps.

To protect the 3′ end of the primer in solution and allow the template to be positioned in the pore, Cherf *et al.*[1] used a special oligonucleotide that binds downstream of the primer. The resulting complex is thermodynamically stable in bulk, but when the single-stranded end of the template is electrophoretically driven through the pore by an applied voltage, the force resulting from the applied voltage can mechanically dissociate the blocking oligomer in a process referred to as 'unzipping,' thereby allowing DNA synthesis to proceed (**Fig. 1a**).

Although Cherf *et al.*[1] succeeded in slowing the DNA translocation to a speed that should be compatible with sequencing, they were not able to directly associate current levels with individual nucleotides. They measured a variety of current levels and manufactured a reference map to correlate current amplitudes to known nucleotide sequences by carrying out four independent experiments using the same DNA template but with one of the four dNTP substrates present at lower concentrations in each experiment. Using this method, they were able to match current steps to the nucleotide sequence of the known DNA molecule with an error of 10–25%. It remains to be seen whether this approach can be applied to DNA having an a priori unknown sequence.

Manrao *et al.*[2] pursued the same strategy of combining a polymerase and a nanopore. However, instead of α-hemolysin they used a mutated version of the protein pore MspA, which has a slightly different geometry (**Fig. 1c**). Whereas many (~10) DNA bases within the long beta barrel of α-hemolysin contribute to the magnitude of the current level, MspA locally confines the DNA merely over a few bases at its smallest constriction (cf. **Fig. 1b**,**c**).

Even with this advantage, the current signatures obtained by Manrao *et al.*[2] were more complicated than anticipated. If the ionic current were strictly influenced by the single base located at the narrowest point of the constriction, four current levels corresponding to the four bases should be observed. Instead, Manrao *et al.*[2] observed a wide variety of distinct current levels. This likely can be attributed to the influence of nucleotides neighboring the base of interest, that is, a small series of adjacent bases, rather than a single base, determining the current signal.

Thus, although both groups have achieved the translocation of DNA through a nanopore in single-nucleotide steps at an appropriate rate, determining the identity of the nucleotides from the current levels calls for further advances in data analysis for base calling. Interestingly, both groups also show that the sequence information is present in the ionic current traces recorded during both the unzipping and synthesis modes of passing DNA through the pore, thereby providing double-pass readings of one molecule.

These two scientific reports coincided with an announcement by Oxford Nanopore at February's Advances in Genome Biology and Technology (AGBT) conference in Marco Island, Florida, that the company will launch the first commercial nanopore sequencing devices later this year (http://www.nanoporetech.com/news/press-releases/view/39). These nanopore-based DNA sequencing devices presumably use a layout similar to the one discussed above. The achievements reported are quite impressive, but no data have been released publicly yet. Most importantly, the 48,490-bp genome of lambda DNA was sequenced as a complete fragment—in one read. This read length is three orders of magnitude larger than the typical length of DNA fragments that are read in current sequencing approaches using fluorescent nucleotide labeling[3].

The long read lengths potentially achievable with nanopore sequencing present an important technological advantage, as the conventional short read lengths call for an elaborate process to computationally assemble the separate reads to deduce a genome sequence. Long reads also directly provide information on insertions, deletions and gene duplicates that are sometimes difficult to determine from assembling short reads. Multiple DNA repeats are found in regulatory DNA sequences, and long read lengths will thus help resolve haplotypes.

In principle, the head-to-tail sequential feed-through of DNA molecules allows for unlimited read length, although with the current approach, the processivity of the DNA polymerase will likely become the limiting factor. Other advantages of nanopore sequencing include the single-molecule read-out, the fact that multiple molecules can be processed by a single pore, the absence of expensive fluorophore labeling and the lack of need for amplification steps.

Looking ahead, several potential improvements are apparent. The error rate must be decreased (a 4% error rate was reported by Oxford Nanopore), the read length can likely be increased further and the stability of the protein pores over time may need to be improved.

An alternative approach to nanopore sequencing may involve solid-state nanopores[12]. Solid-state nanopores have tunable pore size, are more stable than biological membranes, can be reused after cleaning and allow for scaling and device integration. So far, however, DNA sequencing has not been demonstrated with these devices. Conventional silicon-based nanopore membranes are relatively thick, typically ~30 nm, which corresponds to ~60 bases along a single-stranded DNA molecule. Therefore, they are not directly applicable in DNA sequencing applications, although solid-state nanopores are excellent new tools for biophysical studies[12].

The recently introduced solid-state nanopores made of graphene, on the other hand, present an intriguing alternative[13–15]. Graphene is the ultimate nanopore membrane as it is a carbon sheet with a thickness of only a single carbon atom (**Fig. 1d**). It is also electrically conductive, which opens up new modalities for measuring the traversing nucleotides, for example by running a tunneling current through the DNA molecule that is passing through a graphene gap to directly probe the chemical nature of the bases[16]. DNA sequencing using graphene nanopores is currently still science fiction, but so was sequencing with biological pores two decades ago.

Even with the recent rapid progress with DNA sequencing using protein nanopores, the future commercial success of nanopore sequencing is hard to estimate, given the many competitors

and rapid pace of development in the field. It will be interesting to see how things progress in the coming years, both in sequencing applications and in fundamental science, where nanopores provide versatile single-molecule sensors.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Cherf *et al. Nat. Biotechnol.* **30**, 344–348 (2012).
2. Manrao *et al. Nat. Biotechnol.* **30**, 349–353 (2012).
3. Ansorge, W.J. *New Biotechnol.* **25**, 195–203 (2009).
4. Deamer, D.W. & Akeson, M. *Trends Biotechnol.* **18**, 147–151 (2000).
5. Kasianowicz, J.J., Brandin, E., Branton, D. & Deamer, D.W. *Proc. Natl. Acad. Sci.* **93**, 13770–13773 (1996).
6. Branton, D. *et al. Nat. Biotechnol.* **26**, 1146–1153 (2008).
7. Ashkenasy, N., Sanchez-Quesada, J., Bayley, H. & Ghadiri, M.R. *Angew. Chem. Int. Ed.* **44**, 1401–1404 (2005).
8. Stoddart, D., Heron, A.J., Mikhailova, E., Maglia, G. & Bayley, H. *Proc. Natl. Acad. Sci.* **106**, 7702–7707 (2009).
9. Vercoutere, W. *et al. Nat. Biotechnol.* **19**, 248–252 (2001).
10. Clarke, J. *et al. Nat. Nanotechnol.* **4**, 265–270 (2009).
11. Lieberman, K.R. *et al. J. Am. Chem. Soc.* **132**, 17961–17972 (2010).
12. Dekker, C. *Nat. Nanotechnol.* **2**, 209–215 (2007).
13. Schneider, G.F. *et al. Nano Lett.* **10**, 3163–3167 (2010).
14. Merchant, C.A. *et al. Nano Lett.* **10**, 2915–2921 (2010).
15. Garaj, S. *et al. Nature* **467**, 190–193 (2010).
16. Postma, H.W.C. *Nano Lett.* **10**, 420–425 (2010).

Selecting antigens for cancer vaccines

Francesca Avogadri & Jedd D Wolchok

A cancer vaccine is streamlined by identifying a small set of potent immunogens in a tumor cDNA library.

Vaccination as a therapeutic approach to cancer differs in important ways from classic vaccination against infectious agents. A cancer vaccine must induce an immune response to 'self' rather than to 'non-self' antigens without provoking harmful autoimmunity; it must work when administered after a tumor has formed and is exerting immunosuppressive effects; and it should be insensitive to heterogeneous tumor antigen expression among patients. Promising results have been achieved in animal models by vaccination with whole cells from autologous or allogeneic tumors[1] or with cDNA libraries[2], but it is debatable whether using a broad antigen repertoire is the safest and most effective strategy. In this issue, Vile and colleagues[3] show that melanoma in mice can be treated by vaccination with a human melanoma cDNA library delivered by a vesicular stomatitis virus (VSV) vector. Using a novel screening strategy, the authors identify a set of three antigens in the library that, when used in combination to vaccinate mice bearing established melanoma tumors, were as effective as the complete library. Selection of optimal antigen sets has the potential to improve cancer vaccines because it avoids the delivery of whole-cell or cDNA-library antigens that could stimulate autoimmunity or irrelevant immune responses.

Different types of whole-cell vaccines have been proposed and are being investigated for the treatment of various cancers, including prostate cancer, colon cancer, leukemia and melanoma[1]. Early efforts involved co-administration of mycobacterial adjuvants with irradiated or lysed cancer cells derived either from a patient's own surgically removed tumor (autologous) or from patients with the same type of cancer (allogeneic). A second generation of whole-cell vaccines comprised genetically modified cells that expressed immunostimulatory cytokines or surface receptors to fine-tune immunogenicity. These vaccines have not induced robust tumor regression, perhaps owing to the inherently low immunogenicity of mammalian cells. More recently, the Vile laboratory showed that a VSV-based cDNA library expressing 'self' epitopes from normal human prostate tissue induces rejection of prostate tumors in mice[2].

In the present report, Vile and colleagues[3] extend the VSV-expression library approach to melanoma. In the new study, the cDNA is derived from tumor cell lines, which, in contrast to normal tissue, contain numerous genetic mutations that are potentially immunogenic (a list of mutated epitopes is available at: http://cancerimmunity.org/peptidedatabase/mutation.htm). Both studies[2,3] show that a broad 'whole-cell library of antigens' approach overcomes immunological tolerance to cancer antigens. In addition, regardless of whether the cDNA library is derived from cancer cells or normal tissue, the anti-tumor effect is greater when the cDNA originates from a different species (xenogeneic). This supports the hypothesis that xenogeneic antigens are more immunogenic owing to small differences in the protein sequence that are capable of generating a cross-reactive immune response[4]. Interestingly, although the VSV melanoma cDNA library contained antigens that are known to elicit $CD8^+$ T-cell immunity, such as the melanoma differentiation antigen gp100, most of the anti-cancer effect did not involve cytotoxic $CD8^+$ T cells but instead involved $CD4^+$ T cells. This finding is in contrast to the generally accepted view that $CD8^+$ T cells have a central role in tumor immunity, but is in line with more recent data pointing to the importance of other arms of the immune response, including $CD4^+$ T cells and antibodies[5,6].

The main novelty of the current study[3] is the use of an innovative screening strategy that is based on the authors' finding that immunity generated by the VSV library involves production of interleukin (IL)-17. The cytokine IL-17 is secreted by T-helper cell 17 (T_H17) cells (a subset of $CD4^+$ T cells), which have recently emerged as having a major role in autoimmunity, inflammation, host defense and tumor immunity. Tumor-specific T_H17 cells from immunized mice were detected after only two rounds of stimulation with the viral library and a final pulse with tumor-cell lysates. However, the laborious culture conditions and the large amount of virus needed for stimulation could present a limitation when screening hundreds of pools of antigens. This could pose a real challenge if a similar strategy was used to treat individual patients. Addition of heat shock protein (hsp)70 to the culture medium allowed a reduction in the multiplicity of infection that was necessary to induce detectable amounts of IL-17 and represents a crucial technical advance for the scalability of this screening approach (**Fig. 1**).

The IL-17–based screening step led to the identification of three immunogenic proteins: TYRP1, NRAS and cytochrome C1. TYRP1 is a melanoma rejection antigen that activates $CD4^+$ and $CD8^+$ T-cell responses and antibodies. Recent studies in an adoptive T-cell transfer mouse model have also shown that T_H17-like $CD4^+$ T cells specific for TYRP1 have higher anti-tumor activity than T_H1-like cells[5]. Notably, only mild autoimmunity was induced by the VSV-based cDNA library relative to that of other vaccines targeting TYRP1. It would be

Francesca Avogadri and Jedd D. Wolchok are at Swim Across America Laboratory, Ludwig Center for Cancer Immunotherapy and Ludwig Institute for Cancer Research, Memorial Sloan-Kettering Cancer Center, New York, New York, USA.
e-mail: wolchokj@mskcc.org