

Chroniques génomiques

Séquençage d'ADN : l'offensive des nanopores

Bertrand Jordan



L'idée de séquencer des molécules d'ADN en leur faisant traverser des nanopores est apparue dès les années 1990, et la première démonstration de faisabilité (alors encore très préliminaire) a été publiée en 1996 [1]. De nombreux laboratoires travaillent à caractériser différents nanopores biologiques¹ (alpha-HL [alpha-hémolysine], MspA [*Mycobacterium smegmatis* porine A]), ou à mettre au point des pores dits *solid state* dans des membranes de silicium ou de carbone (voir [2] pour une revue récente). Plusieurs entreprises se sont intéressées au développement de systèmes de séquençage (notamment IBM, Illumina, Roche, etc.) mais un seul acteur, la firme britannique *Oxford Nanopore Technologies* (ONT) en est arrivée, aujourd'hui, au stade de la commercialisation. Fondée en 2005, l'entreprise a annoncé son premier produit, le Minlon, en 2012 [3] et a réellement débuté sa commercialisation (en mode *early access*²) en 2014. Cet appareil occupe une place à part dans le monde des séquenceurs par ses dimensions (la taille d'une grosse clef USB) et son coût (500 euros, ou moins) ; il a fait la preuve de performances assez convaincantes, et l'on peut penser qu'avec ses successeurs il va réellement changer le paysage du séquençage de nouvelle génération. Cette chronique va donc présenter le Minlon, ses spécificités, ses points forts et ses faiblesses ainsi que les évolutions prévues et leur place dans ce marché très mobile.

¹ Un nanopore est un pore de diamètre compris entre 1 et 100 nm. Il peut être constitué par une protéine transmembranaire (un canal ionique) ou correspondre à un trou dans un matériau (une membrane en polymère par exemple). Le diamètre du pore peut être ajusté pour correspondre aux dimensions d'une molécule biologique donnée. Quand la molécule traverse le pore, elle modifie ses propriétés électriques.

² L'accès anticipé (*early access*) est un mode de financement et de développement d'objets qui permet aux développeurs d'obtenir des retours des utilisateurs avant finalisation et commercialisation. Le même procédé est également utilisé en informatique pour de nouvelles applications.



UMR 7268 ADÉS, Aix-Marseille, Université/EFS/CNRS, Espace éthique méditerranéen, hôpital d'adultes la Timone, 264, rue Saint-Pierre, 13385 Marseille Cedex 05, France ; CoReBio PACA, case 901, parc scientifique de Luminy, 13288 Marseille Cedex 09, France.
bertrand.jordan@univ-amu.fr
brjordan@orange.fr

Un véritable mini-séquenceur

Le Minlon se présente comme une sorte de grosse clef USB (Figure 1) destinée à être branchée sur un ordinateur qui va l'alimenter en électricité et effectuer les calculs nécessaires pour interpréter les signaux en termes de séquence de bases.

La membrane porte 512 nanopores biologiques (leur nature exacte évolue au fil des mises au point et est confidentielle) ; la solution contenant l'ADN à séquencer (une petite fraction de microgramme) est déposée et conduite sur la membrane. Les circuits électroniques



Figure 1. Le séquenceur Minlon. La membrane portant les 512 pores se situe à droite de l'appareil, la micropipette est en train de charger la solution d'ADN à séquencer (200 µl). Longueur environ 10 cm, poids 90 g (site ONT).

contenus dans l'appareil gèrent l'application d'une différence de potentiel entre les deux côtés de la membrane, laquelle entraîne les molécules d'ADN à travers les nanopores. Au niveau de chaque nanopore, les variations de courant dues au passage de l'ADN sont relevées, mises en forme et transmises à l'ordinateur qui les interprète en termes de séquence des bases le long de l'ADN³. L'appareil est à usage unique (on le charge une seule fois), la session dure de 48 à 72 heures et produit plusieurs gigabases (Gb, 10⁹ bases) de séquence brute. Point intéressant, les séquences lues sont très longues, plusieurs dizaines de kilobases (Kb), le record actuel étant de 882 Kb ! Elles ne sont en fait limitées que par la taille de l'ADN employé.

Le point faible du Minlon est son taux d'erreurs, qui est actuellement de l'ordre de 5 % pour une seule lecture (séquence « brute »). Ceci est directement lié au mode de lecture, qui ne procède pas base par base comme on pourrait l'imaginer. En effet les pores biologiques actuellement utilisés ont une dimension telle que la longueur d'ADN présente à l'intérieur du pore correspond à cinq ou six bases : le signal observé est donc caractéristique d'une suite de bases et non d'une seule d'entre elles. La lecture se fait par comparaison informatique entre le courant observé et des courants de référence établis pour tous les hexamères possibles. Cela entraîne diverses difficultés et notamment l'impossibilité de déchiffrer correctement de longs homopolymères (comme GGGGGGGGGG, G étant une guanine). Ces problèmes sont en partie résolus grâce à la redondance des lectures ainsi qu'à des astuces techniques permettant de séquencer successivement les deux brins d'un ADN bicaténaire et d'améliorer ainsi la qualité du résultat. À plus long terme, l'amélioration des pores et des procédés informatiques devraient permettre de progresser – en attendant la mise au point de pores non biologiques (trous microscopiques dans une membrane de graphène ?) qui, grâce à leurs dimensions réduites, pourraient autoriser une réelle lecture base par base.

Des applications déjà nombreuses

Après des débuts qui furent souvent difficiles entre les mains des premiers utilisateurs, le Minlon a fait la preuve de ses capacités. À titre de banc d'essai, la séquence de l'ADN d'*Escherichia coli* a pu être ainsi assemblée à partir des seules données de cet appareil et avec un taux d'erreur de 0,5 % [4] ; des analyses métagénomiques ont pu être effectuées très rapidement, à partir d'échantillons d'urine [5]. Plus spectaculaire, l'emploi du Minlon sur le terrain a permis de caractériser les souches du virus Ébola lors de la récente épidémie en Afrique, en 2015, et ainsi de révéler les voies de propagation de l'agent pathogène [6, 7]. Au cours de ces travaux, 142 échantillons ont pu être séquencés en local, avec des résultats obtenus en quelques heures. La plus grosse difficulté technique était la connexion par internet au *Cloud*, nécessaire à l'époque pour l'interprétation des signaux en termes de séquence de bases... Plus récemment, l'emploi du Minlon a été rapporté pour séquencer des communautés microbiennes sur

le terrain, en Antarctique [8]. Notons que, dans ces applications cliniques, il est possible d'obtenir des résultats très rapidement, parfois en moins d'une demi-heure, la lecture se faisant en temps réel et sans nécessairement interrompre l'expérience. Dans ces applications qui visent la détection et l'identification de souches virales ou bactériennes, la facilité de mise en œuvre et la rapidité du système de nanopores sont des avantages déterminants, et le taux d'erreurs relativement élevé n'est généralement pas trop gênant.

Un travail qui s'apparente à un exercice de style a été récemment rapporté dans le serveur de *preprints* bioRxiv⁴ [9] : il s'agit d'une tentative de séquençage d'un génome humain (utilisant une lignée cellulaire de référence déjà séquencée). Ce travail, qui a directement utilisé l'ADN extrait de la lignée, sans amplification, a nécessité l'emploi de 39 Minlon qui ont fourni au total 91 gigabases de séquence (soit une couverture de 30X environ) ; la taille moyenne des lectures était de 10 kb, mais (avec des précautions spéciales pour la préparation d'ADN) des lectures très longues allant jusqu'à 882 kb ont été obtenues. Ces séquences ont pu être assemblées, sans faire appel aux données de référence, en moins de 3 000 contigs⁵, ce qui est déjà très bien pour le séquençage *de novo* d'un aussi grand génome. La comparaison avec la séquence de référence a montré que 90 % des polymorphismes (SNP, *single nucleotide polymorphism*) connus dans cette lignée étaient correctement détectés. On voit ainsi que la perspective d'utiliser le séquençage par nanopores pour caractériser un ADN humain par rapport à la séquence de référence n'est pas totalement chimérique – même si le coût de 39 Minlon reste prohibitif.

Des améliorations continuent à être apportées tant à l'appareil et au cœur de la technologie (nouveaux pores) qu'en amont (préparation de l'ADN) et en aval (analyse et interprétation des signaux). Entre les mains d'utilisateurs avertis, un Minlon peut maintenant fournir plus de 10 Gb de séquences, tandis qu'ONT revendique 20 Gb dans son laboratoire ; le séquençage direct de l'ARN ainsi que la détection de bases modifiées ont déjà été démontrés [10]. L'absence de barrière d'accès due au faible coût du Minlon⁶ favorise sa diffusion (plus de 4 000 exemplaires ont déjà été vendus) et

³ Jusqu'à début 2017, ce calcul nécessitait un accès au *Cloud* (nuage informatique, un système dématérialisé de stockage utilisant des serveurs internet) ; il est maintenant effectué en local.

⁴ Un site donnant accès aux versions préliminaires d'articles (non validées par une revue), mode de diffusion très utilisé en physique et qui commence à se répandre pour la biologie.

⁵ Blocs de séquence continue.

⁶ Un kit de démarrage comprenant un Minlon plus un ensemble de réactifs et d'accessoires est proposé pour 1 000 €, mais le prix d'un Minlon seul est nettement plus bas et varie en fonction des quantités commandées. En tout état de cause, cela n'a rien à voir avec le coût des séquenceurs concurrents (au moins 100 000 €).



encourage des expérimentations susceptibles de lui ouvrir de nouveaux domaines d'application.

Des évolutions en cours

ONT a annoncé depuis plus de deux ans une machine « haut de gamme », le Promethlon, qui doit regrouper 48 cellules (comportant chacune 3 000 nanopores et non plus 512) et avoir un débit total de plusieurs milliers de gigabases par session – dans la même catégorie que les plus grosses machines existantes. Seul ennui, le Promethlon est très en retard, quelques appareils ont été livrés à des utilisateurs-testeurs, mais les cellules ne sont toujours pas arrivées. Plus récemment, ONT a annoncé le lancement d'un système intermédiaire, le Gridlon, qui doit fonctionner avec cinq cellules Minlon et contenir l'informatique nécessaire au décodage des bases. Cette dernière machine est déjà en vente (environ 100 000 euros), mais ses avantages par rapport à l'emploi de cinq Minlons en parallèle sont discutables. Plus intéressant peut-être, l'entreprise développe aussi un système encore plus portable, le Smidglon, qui comporterait 128 nanopores et serait destiné à être branché sur un téléphone portable (Figure 2).

Le concept est séduisant dans la perspective d'une généralisation du séquençage clinique au chevet du malade (dit POC, *point of care*), reste à voir s'il sera effectivement commercialisé et quelles seront ses performances. On peut se demander notamment si l'ordinateur contenu dans un *smartphone* sera suffisamment puissant pour effectuer la complexe analyse du signal – et, plus prosaïquement, combien de temps la batterie pourra assurer le fonctionnement d'un accessoire gourmand en énergie.

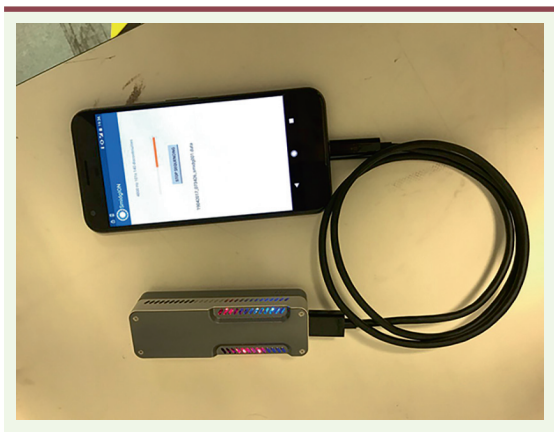


Figure 2. Le Smidglon. Dernier séquenceur de Oxford Nanopore Technology, destiné à fonctionner avec un téléphone portable (site ONT).

Il est clair en tous cas que ONT a de grandes ambitions et, qu'après une longue période de mise au point, d'importants investissements (plus de 350 millions de livres Sterling depuis 2005) et quelques annonces prématurées (notamment en ce qui concerne le Promethlon), l'entreprise dispose aujourd'hui d'un produit stable, le Minlon. Celui-ci présente des caractéristiques intéressantes (lecture longue, simplicité d'emploi, portabilité, absence d'investissement) qui lui donnent un net avantage dans certaines applications malgré la qualité perfectible de la séquence obtenue. Le paysage actuel du séquençage de nouvelle génération est marqué par la domination de l'entreprise *Illumina*, dont la technologie est déclinée en toute une gamme de machines et qui fournit des séquences courtes (150 bases) mais de bonne qualité et à un coût qui a fait du « génome à 1 000 dollars » une réalité⁷. Ses concurrents « historiques » ont disparu (*454*, *Solid*) ou sont en perte de vitesse (*Ion Torrent*). Une firme, *Pacific Biosciences*, reste dans la course grâce à un procédé qui autorise des lectures longues (plusieurs kilobases) avec un taux d'erreurs relativement faible, mais moyennant des coûts d'investissement et de fonctionnement élevés. *Oxford Nanopore Technology* apparaît donc comme un possible concurrent sérieux d'*Illumina*. L'entreprise a une niche évidente dans les applications où la facilité de mise en œuvre (portabilité) et l'absence d'investissement sont déterminantes. Elle peut aussi, comme *Pacific Biosciences*, être complémentaire d'*Illumina* dans des projets de séquençage *de novo* où le gros des données est fourni par le séquençage de petits fragments, mais où l'organisation générale du génome étudié repose sur l'obtention de longues séquences. On peut enfin imaginer que des progrès au niveau de la nature et de la structure des pores permettent à l'avenir une véritable lecture base par base, qui autoriserait sans doute une baisse spectaculaire du taux d'erreurs et mettrait alors ONT en concurrence directe avec *Illumina*. Quoi qu'il en soit, on commence à percevoir une évolution sensible des technologies de séquençage à grande échelle, qui pourrait aboutir à une accessibilité fortement accrue et sans doute à des applications cliniques encore insoupçonnées. ♦

SUMMARY

Nanopore DNA sequencing becomes a reality

A practical DNA sequencer based on nanopores has now demonstrated interesting performance and may be on the way to capturing a significant slice of the sequencing market while opening up new application fields thanks to its portability and low capital cost. ♦

LIENS D'INTÉRÊT

L'auteur déclare n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

RÉFÉRENCES

1. Kasianowicz JJ, Brandin E, Branton D, Deamer DW. Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci USA* 1996 ; 93 : 13770-3.

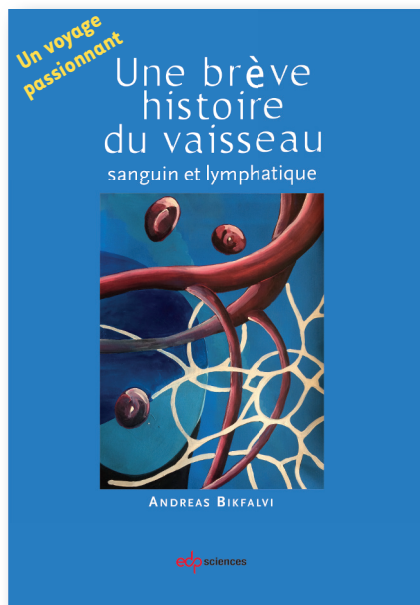
⁷ Du moins pour les machines les plus productives et avec une utilisation optimale

RÉFÉRENCES

2. Feng Y, Zhang Y, Ying C, *et al.* Nanopore-based fourth-generation DNA sequencing technology. *Genomics Proteomics Bioinformatics* 2015 ; 13 : 4-16.
3. Eisenstein M. Oxford nanopore announcement sets sequencing sector abuzz. *Nat Biotechnol* 2012 ; 30 : 295-6.
4. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat Methods* 2015 ; 12 : 733-5.
5. Schmidt K, Mwaigwisya S, Crossman LC, *et al.* Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J Antimicrob Chemother* 2017 ; 72 : 104-14.
6. Quick J, Loman NJ, Duraffour S, *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* 2016 ; 530 : 228-32.
7. Pennisi E. Genomics. Pocket DNA sequencers make real-time diagnostics a reality. *Science* 2016 ; 351 : 800-1.
8. Johnson SS, Zaikova E, Goerlitz DS, *et al.* Real-time DNA sequencing in the Antarctic Dry Valleys using the Oxford Nanopore sequencer. *J Biomol Tech* 2017 ; 28 : 2-7.
9. Jain M, Koren S, Quick J, *et al.* Nanopore sequencing and assembly of a human genome with ultra-long reads. *bioRxiv Preprint* first posted online April 20, 2017. doi: <https://doi.org/10.1101/128835>. <http://biorxiv.org/content/biorxiv/early/2017/04/20/128835.full.pdf>
10. Jain M, Olsen HE, Paten B, Akeson M. The Oxford nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* 2016 ; 17 : 239.

TIRÉS À PART

B. Jordan



ISBN : 978-2-7598-1863-1

202 pages

25 €

Ce livre, intéressant et lisible à la fois pour le spécialiste et le grand public, apporte des observations originales et nouvelles concernant l'angiogenèse, et notamment l'histoire des différentes découvertes, et discute les aspects et les concepts plus généraux en les plaçant dans le contexte de la philosophie des sciences.

Facile à lire, bien illustré, cet ouvrage cherche à comprendre et à faire comprendre les enjeux de la recherche sur l'arbre vasculaire en développement et en pathologie. Il intéressera non seulement les étudiants et post-doctorants en biologie, mais aussi les chercheurs actifs dans ce domaine de recherche ainsi que toute personne intéressée par la biologie et la médecine et par l'histoire des sciences. Un voyage passionnant à travers l'histoire et les concepts les plus actuels concernant les recherches sur le vaisseau sanguin.

Andreas Bikfalvi est Professeur à l'université de Bordeaux et Directeur d'une unité de recherche Inserm sur le cancer et la biologie vasculaire. Il est, par ailleurs, membre senior de l'Institut Universitaire de France (IUF) et reconnu internationalement pour ses recherches dans le domaine de l'angiogenèse tumorale.

BON DE COMMANDE

À retourner à EDP Sciences, 109, avenue Aristide Briand, 92541 Montrouge Cedex, France
Tél. : 01 41 17 74 05 - Fax : 01 49 85 03 45 - E-mail : françois.flori@edpsciences.org

NOM : Prénom :

Adresse :

Code postal : Ville :

Pays :

Fonction :

Je souhaite recevoir

Une brève histoire de vaisseau : 25 € + 3 € de port = 28 € TTC

en exemplaire, soit un total de €

Par chèque, à l'ordre de EDP Sciences

Par carte bancaire :

Visa

Eurocard/Mastercard

Carte n° | | | | | | | | | | | | | | | | | | | | | |

Date d'expiration : | | | | | | | | N° de contrôle au dos de la carte : | | | | | | | | Signature :