

Chroniques génomiques

Séquençage par nanopores : les protéines aussi ?

Bertrand Jordan

Des protéines à l'ADN – et vice-versa

Le séquençage de protéines, initié par Fred Sanger en 1951 [1], a constitué une part importante de l'activité des laboratoires de biologie moléculaire jusqu'à la fin des années 1970. Avec la révolution du génie génétique et de l'ADN recombinant, qui a permis un accès direct aux gènes, et la mise au point (toujours par Fred Sanger) du séquençage de l'ADN, il est devenu beaucoup plus rapide et commode de lire les gènes plutôt que leurs produits, et le séquençage de protéines a beaucoup perdu de son importance. Pourquoi donc s'intéresse-t-on encore à la mise au point de nouvelles approches ? En fait, pas vraiment pour le séquençage *de novo* : sauf cas exceptionnel, la voie de l'ADN est bien plus efficace. Mais la complexité de l'organisation exon/intron des gènes et la fréquence des phénomènes d'épissage alternatif (souvent fonctionnellement signifiants) imposent fréquemment une analyse au niveau des ARN, voire de la protéine, ce dernier permettant, en plus, de caractériser la grande variété des modifications post-transcriptionnelles qui, elles aussi, peuvent changer le rôle d'une protéine.

Un sujet difficile, des techniques lourdes

L'étude du protéome présente bien des difficultés. La variété des protéines, compte tenu des épissages alternatifs et des modifications post-transcriptionnelles, est supérieure d'un ou deux ordres de grandeur à celle des gènes. Et leur concentration dans les échantillons biologiques varie d'au moins dix ordres de grandeur [2], ce qui ne facilite pas l'étude des composants minoritaires ! De plus, les différences de charge entre acides aminés et la structure tridimensionnelle des protéines contrastent avec l'organisation très régulière des chaînes d'ADN dont les propriétés restent constantes le long de la chaîne nucléotidique quelle que soit la séquence. Le travail initial de Sanger pour la séquence de l'insuline



Biologiste, généticien et immunologiste, Président d'Aprogène (Association pour la promotion de la Génomique), 13007 Marseille, France. brjordan@orange.fr

reposait sur des séries de dégradations partielles par des protéases, suivies d'analyses chimiques et de la recherche de recoupements entre les peptides obtenus. Par la suite, on s'est rapidement tourné vers la dégradation récurrente de la protéine à partir de son extrémité N-terminale, selon la méthode d'Edman¹ [3], qui a été largement automatisée (Figure 1), et plus récemment vers l'analyse par spectrométrie de masse [4]. Mais ces approches restent lentes et imparfaites. Le boom des nanopores pour le séquençage d'ADN suggère d'envisager cette approche pour les protéines.

Une check-list pour les nanopores

On imagine, dans l'idéal, le passage d'une chaîne d'acides aminés à travers un nanopore qui va identifier chaque acide aminé lors de son transit dans la zone de détection (Figure 2) grâce aux variations qu'il induit dans le courant ionique qui passe à travers le nanopore. Pour en arriver là, il faut créer des milliers de nanopores sur un support adéquat, déplier la structure tertiaire de la protéine pour qu'elle se présente comme une chaîne linéaire, assurer son passage régulier et suffisamment lent à travers le pore, et être capable de différencier les vingt acides aminés lors de leur passage par la zone de détection. Des amorces de solutions existent même si elles ne sont pas encore totalement opérationnelles. En tous cas il ne manque pas d'équipes travaillant à les perfectionner : un article sur cette approche, publié dans la revue *Nature Biotechnology* [5], comporte pas

¹ Développée par le biochimiste suédois Pehr Edman (1916-1977) dans les années 1940, cette méthode biochimique permet de séquencer jusqu'à environ 50 acides aminés à partir de l'extrémité N-terminale de la protéine.



Figure 1. Séquenceur de protéines Beckman-Coulter Porton LF3000G, automatisant la dégradation séquentielle d'Edman, dans la limite d'une cinquantaine d'acides aminés (image Wikipedia commons).

moins de 89 références, en majorité très récentes. L'état de l'art est décrit en détail dans une autre revue récente assez technique mais très complète et bien illustrée [6].

Les nanopores employés sont quasiment tous de nature biologique (comme pour le séquençage d'acides nucléiques), l'un des plus fréquents étant le pore formé par la protéine MspA issue de la bactérie *Mycobacterium stegmatis* et bien sûr produite par génie génétique (Figure 2). Les pores envisageables se différencient par leur hauteur (en général de l'ordre de 10 nm) et surtout par la taille de leur région la plus étroite (la zone de détection, dite de « constriction ») qui va gouverner les variations du courant ionique lors du passage des acides aminés. Pour le nanopore MspA, cette dimension est de l'ordre de 0,6 nm, une des plus petites – mais qui reste très supérieure à la taille d'un acide aminé : le signal proviendra de la dizaine d'acides aminés présents dans la constriction et, comme pour le séquençage d'ADN, il faudra une analyse informatique performante pour en déduire la séquence.

Le déploiement de la protéine peut être obtenu par l'emploi de conditions dénaturantes douces (compatibles avec l'intégrité des pores). Contrairement aux acides nucléiques, les chaînes protéiques ne portent pas des charges électriques régulières, on ne peut donc pas compter sur le champ électrique existant entre les faces de la membrane pour pousser la protéine à travers le pore : il faut mobiliser la chaîne par un « moteur » assurant son déplacement, à une vitesse suffisamment lente (une fraction de seconde par acide aminé) pour autoriser la lecture. Dans l'exemple de la Figure 2 ce rôle est assuré par un oligonucléotide attaché à la protéine et sur lequel agit une hélicase bactérienne². Il existe d'autres solutions et notamment

² Enzyme impliquée dans l'ouverture du double brin d'ADN lors de la réplication.

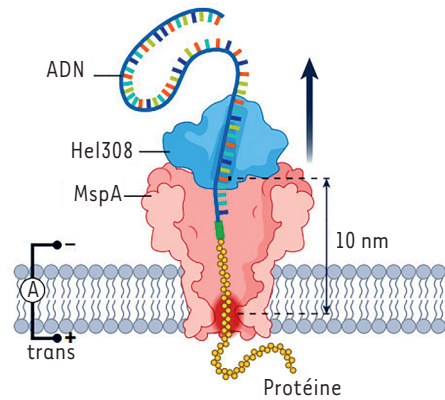


Figure 2. Lecture de la séquence d'une protéine par un nanopore MspA. Dans cette image la protéine est « tirée » à travers le nanopore (flèche) par une séquence d'ADN et une hélicase bactérienne (Hel 308). La zone de détection du pore est figurée en rouge foncé ; le passage de la protéine modifie le courant ionique (A) mesuré entre les deux faces de la membrane (Extrait partiel et modifié de la figure 4 de [6]).

des protéines capables de jouer ce rôle ; elles peuvent aussi aider à « dérouler » la structure secondaire et tertiaire de la protéine.

La reconnaissance des acides aminés est naturellement un des points critiques, d'autant plus qu'il s'agit là de distinguer entre vingt entités parfois très proches (et non quatre seulement comme pour l'ADN). Plusieurs publications très récentes revendiquent l'identification de tous les acides aminés par des nanopores adaptés [7] – mais il s'agit là de la reconnaissance d'acides aminés isolés ou encore portés à l'extrémité d'un homo-polypeptide³. On n'en est pas encore à la reconnaissance « au vol » des acides aminés lors de leur transit dans le nanopore, mais le fait que ces identifications soient possibles est un bon point de départ. En tout état de cause, la dimension des constriction entraîne la présence simultanée de plusieurs acides aminés dans ce site de reconnaissance, et donc (comme pour les ADN) la nécessité d'une interprétation très performante des signaux observés. À moins qu'on ne parvienne, comme dans la nouvelle technique de séquençage de l'ADN « SBX » [8] (→) (→ Voir m/s n° 6-7, 2025, page 611) annoncée par la firme Roche (et maintenant baptisée « Axelius »), à « étirer » la molécule à séquencer pour faciliter sa lecture – mais c'est peu probable. La détection des modifications post-transcriptionnelles semble, elle, très possible [9] et peut constituer un point fort de cette approche.

³ Polymère formé à partir d'un seul acide aminé.

Une technologie en devenir

Comme on le voit, le séquençage de protéines à l'aide de nanopores n'est pas une technique mature comme l'est devenu le séquençage d'ADN. La plupart des éléments nécessaires existent déjà, mais ils doivent encore être considérablement améliorés pour aboutir à un système fiable. Le but ne serait pas (ou guère) le séquençage *de novo* de protéines inconnues (il vaut mieux passer par l'ADN), mais l'analyse fine des modifications post-transcriptionnelles (au niveau de molécules individuelles), la détection et la quantification de protéines très peu abondantes (important pour le diagnostic) et bien sûr la caractérisation des variants d'épissage. En d'autres termes, de nombreuses applications susceptibles d'apporter des informations importantes du point de vue médical. Il n'est donc pas étonnant que des entreprises s'investissent dans ce domaine. En premier, bien sûr, le spécialiste des nanopores, Oxford Nanopore Technologies (ONT)⁴, qui a transformé, pour l'ADN, cette curiosité de laboratoire en un outil largement répandu et disposant de multiples atouts. ONT affiche un vif intérêt pour l'analyse des protéines, et annonce des projets grandioses – l'entreprise aura-t-elle les reins assez solides pour les mener à bien ? D'autres firmes se risquent aussi dans cette aventure, comme Portal Biotech (Pays-Bas) qui manifeste beaucoup d'assurance sur son site⁵, s'affirmant capable de séquencer de grandes protéines. On anticipe sans doute un peu, comme souvent

dans le monde de la biotech. Quoi qu'il en soit, on peut prévoir que les nanopores vont jouer un rôle important dans l'analyse des protéines, et ouvrir de nouvelles possibilités tant scientifiques que médicales. ♦

SUMMARY

Will nanopores also sequence proteins ?

DNA sequencing using nanopores is now well established. Could the approach be extended to proteins ? There are a number of difficulties but work by many groups has shown proof-of-principle solutions for most of them. Further progress could indeed lead to a workable protein sequencing approach that would have many applications in research and in the clinic. ♦

LIENS D'INTÉRÊT

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

RÉFÉRENCES

1. Sanger F, Tuppy H The amino-acid sequence in the phenylalanyl chain of insulin. *Biochem J* 1951 ; 49 : 463-81.
2. Zubarev RA. The challenge of the proteome dynamic range and its implications for in-depth proteomics. *Proteomics* 2013 ; 13 : 723-6
3. Edman P. A method for the determination of amino acid sequence in peptides. *Arch Biochem* 1949 ; 22 : 475.
4. Steen H, Mann M. The abc's (and xyz's) of peptide sequencing. *Nat Rev Mol Cell Biol* 2004 ; 5 : 699-711.
5. Lu C, Bonini A, Viel JH, Maglia G. Toward single-molecule protein sequencing using nanopores. *Nat Biotechnol.* 2025 ; 43 : 312-22.
6. Ritmejeris, J., Chen, X. & Dekker, C. Single-molecule protein sequencing with nanopores. *Nat Rev Bioeng* 2025 ; 3 : 303-16.
7. Wang K, Zhang S, Zhou X, Yang X, et al. Unambiguous discrimination of all 20 proteinogenic amino acids and their modifications by nanopore. *Nat Methods.* 2024 ; 21 : 92-101.
8. Jordan, B. Séquençage de l'ADN : la fin d'un quasi-monopole ? *Med Sci (Paris)* 2025 ; 41 : 611-4.
9. Restrepo-Pérez L, Wong CH, Maglia G, Dekker C, Joo C. Label-Free Detection of Post-translational Modifications with a Nanopore. *Nano Lett.* 2019 ;19 : 7957-64.

⁴ <https://nanoporetech.com/es>

⁵ portalbiotech.com

TIRÉS À PART

B. Jordan



Avec m/s, vivez en direct les progrès et débats de la biologie et de la médecine

CHAQUE MOIS / AVEC LES ARTICLES DE RÉFÉRENCE DE M/S
CHAQUE JOUR / SUR WWW.MEDECINESCIENCES.ORG

Abonnez-vous sur www.medecinesciences.org