infOGM

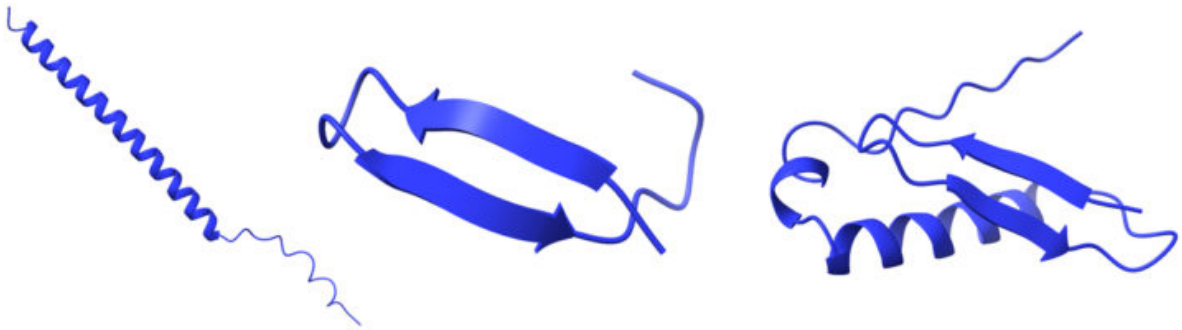
Veille citoyenne d'information sur les OGM et les semences

Une avancée majeure en biologie : le protéome obscur

Par

Publié le 16/01/2025, modifié le 20/01/2025

Des publications scientifiques récentes s'intéressent à une partie du génome de l'espèce humaine découverte tardivement, et jusqu'alors ignorée. C'est à partir de recherches sur le cancer que ces travaux sur ce qui est aujourd'hui nommé « *protéome obscur* » ont pris leur essor. Fondamentale d'un point de vue de la biologie, cette nouvelle approche des gènes et des protéines est d'ores et déjà vue comme une source d'opportunités pour les biotechnologies, notamment dans le domaine médical.



Leron Kok - Les structures probables de trois nouvelles mini-protéines putatives, prédites par l'intelligence artificielle AlphaFold3, montrent leur simplicité.

Une des plus grandes surprises du premier séquençage[i] du génome humain, il y a une vingtaine d'années, a été sa taille. Elle était plus petite qu'attendue : trois milliards de paires de nucléotides[ii]

avec un nombre de gènes considéré comme faible (20 379), soit moins d'un tiers du nombre prévu par certains scientifiques. Par gènes, il fallait entendre des « *séquences d'ADN* » qui codent pour des protéines[iii], une des familles de molécules qui construisent les organismes et les font vivre. Une grande partie de cet ADN était dit « *non codant* ». Depuis quelques décennies, on sait en fait que de l'ADN non codant peut... coder pour des ARN[iv]. De plus, on vient de découvrir qu'il peut aussi coder d'autres protéines beaucoup plus petites. Cet ensemble de petites protéines codées par de l'ADN, jusqu'alors méconnu, a été nommé « *protéome obscur* »[v]. Au début de cette découverte, peu de crédit avait été accordé à ces mini-protéines, car elles étaient rapidement dégradées. Mais, finalement, certains scientifiques s'y sont intéressés et c'est donc grâce au protéome obscur qu'a été découvert cette partie cachée de l'ADN. Déroulons un peu cette histoire.

Réviser les modèles de la biologie des gènes

Un gène, dans sa définition la plus répandue, consiste en une longue séquence d'ADN codant pour des protéines dont il est admis qu'elles ont plus de 100 acides aminés[vi]. Ce gène comporte des signaux indiquant aux enzymes cellulaires où commencer et où s'arrêter pour lire la séquence ADN et la transcrire en ARN messager (ARNm). Celui-ci, acheminé dans le cytoplasme, servira de modèle à d'autres structures : les ribosomes[vii]. Cela leur permettra d'ordonner les acides aminés disponibles dans la cellule en séquence protéique.

Cependant, une récente analyse systématique[viii], menée par un très grand nombre de chercheurs, suggère que les scientifiques ont manqué des milliers de gènes qui se cachent dans des parties du génome précédemment ignorées et qui produisent de petites protéines. Il y a quelques années, l'un des responsables de l'analyse, John Prensner, neuro-oncologue pédiatrique de l'Université du Michigan, a commencé à scruter le protéome obscur dans le but de découvrir les gènes qui en sont à l'origine, car ses recherches parmi les gènes connus pour trouver ceux qui sont liés à des cancers n'aboutissaient à rien. « *Je me suis intéressé à ce que le reste du génome avait à offrir* », résume John Prensner[ix].

Ces dernières années, les biologistes qui étudient tous les domaines, de la levure aux serpents en passant par les plantes et l'Homme, ont exploré l'ADN et découvert une pléthore de ces gènes de très courte séquence et dépourvus des signaux préliminaires classiques, souvent transcrits en ARN puis, le cas échéant, traduits en petits polypeptides (mini-protéines) de moins de 12 acides aminés[x]. Ces gènes sont dits « *non traditionnels* ». Ces découvertes ont été rendues possibles par les progrès de certaines techniques, comme la spectrométrie de masse[xi] appliquée aux protéines et le profilage des ribosomes[xii]. Les algorithmes utilisés comme « *intelligence artificielle* », à qui des bio-informaticiens ont demandé de rechercher ces nouvelles protéines, ont aussi contribué à leur découverte.

Ainsi la communauté scientifique est amenée à réviser collectivement les modèles de la biologie des gènes eucaryotes. C'est une évolution majeure en biologie moléculaire, basé sur l'étude d'une partie de l'ADN jusqu'alors ignorée. Une partie du travail ultérieur consistera à éclairer ce protéome obscur : faire le lien entre les séquences géniques et les mini-protéines, découvrir la nature et l'incidence de celles-ci. Dans un premier temps, les efforts se porteront surtout sur l'espèce humaine.

Un très grand champ de connaissances s'ouvre. Certes, ce champ de connaissances appartient au domaine de recherche qui s'intéresse essentiellement à l'ADN, domaine souvent qualifié de « *tout génétique* » car restreignant bien trop le vivant à cette seule molécule. Mais le point intéressant est que cette découverte montre justement que la complexité du vivant affecte aussi cette molécule, avec des conséquences largement méconnues à ce jour et qui interrogent sur les discours de

maîtrise technique portés par certaines multinationales.

Des applications en médecine déjà envisagées

Identifiées dans des cellules tumorales ou cardiaques, parfois en grand nombre (jusqu'à 550 dans les cellules tumorales), les mini-protéines doivent encore être caractérisées : « *tant que nous ne connaîtrons pas la nature des protéines du protéome sombre et leur contribution, notre capacité à traiter les maladies sera limitée* », déclare Alan Saghatelian, biologiste chimiste au Salk Institute for Biological Studies[xiii]. C'est surtout dans le domaine des thérapies du cancer que ce protéome obscur donne des espoirs. Il faut d'abord montrer l'importance de la mini-protéine dans les cellules de tel ou tel cancer, en créant artificiellement des mutations sur le gène qui code la protéine. On comprendra ainsi, petit à petit, sur quoi et comment elle intervient. Puis, en connaissant la nature de ces mini-protéines associées à tel ou tel cancer, il est attendu que des médicaments puissent être élaborés pour détruire ou empêcher d'agir ces mini-protéines oncogènes.

Les maladies métaboliques, telles le diabète ou l'obésité, et les maladies immunologiques sont aussi concernées. Thomas Martinez, biochimiste spécialiste des protéines à l'Université de Californie participant à l'analyse systématique, et son équipe recherchent de minuscules protéines impliquées dans le cancer du pancréas et les maladies métaboliques. « *Je suis très enthousiaste à l'idée de traduire les micro-protéines en efforts thérapeutiques, en espérant qu'elles serviront à la fois de biomarqueurs et de cibles médicamenteuses* », explique-t-il. « *Une fois cette barrière franchie, je pense que l'intérêt pour ce domaine augmentera* »[xiv]. John Prensner, un des responsable de l'analyse, conseille déjà une société du Massachusetts, ProFound Therapeutics, qui s'est associée au géant pharmaceutique Pfizer en promettant de mettre au point des traitements de l'obésité basés sur des cibles mini-protéiques.

A ce stade, des certitudes se profilent. Alors même que le protéome obscur n'est étudié que depuis peu de temps, que les connaissances concernant ces mini-protéines sont encore infimes et que les modalités de leur utilisation ne sont pas encore abouties, les promesses thérapeutiques vont vite faire affluer les capitaux dans les labos privés et publics pour des applications industrielles. D'autre part, la médecine servira probablement de cheval de Troie pour l'utilisation du protéome obscur dans les technologies végétales, pour « *améliorer* » et artificialiser encore plus les plantes.

i Le séquençage est l'opération qui consiste à donner la suite des nucléotides dans l'ADN d'un organisme. Les nucléotides sont au nombre de 4 se définissant par leur base azotée : adénine (A), thymine (T), guanine (G) et cytosine (C). Le génome humain est constitué de 3 milliards de paires de nucléotides.

ii Par comparaison, le plus grand génome connu est celui d'une fougère des îles du Pacifique. Découvert en 2024, il contient 160 milliards de paires de nucléotides, dont beaucoup d'ADN répétitif non codant.

iii Les protéines sont des chaînes d'acides aminés qui sont les constituants de base des protéines. Il y a 20 acides aminés différents qui entrent dans la constitution des protéines. En biologie, on parle de protéines lorsque celles-ci ont plus de 100 acides aminés. En dessous de ce chiffre, on parle de polypeptides.

iv Frédéric Jacquemart, « Un tournant théorique pour les OGM », *Inf'OGM, le journal*, n°113, novembre/décembre 2011.
Annick Bossu, « Biotechnologies médicales à ARN : nouveau Graal ? », *Inf'OGM*, 30 mars 2023.

v Protéome : ensemble de protéines d'un organisme.

vi Voir note iii.

vii Ribosomes : structures composées d'ARN et de protéines présentes dans le cytoplasme des cellules et responsables de la synthèse des protéines.

viii Elizabeth Pennisi, « ‘Dark proteome’ survey reveals thousands of new human genes », *Science* , Volume 386, Issue 6725, Novembre 2024.

ix *Ibid.*

x Bradley W. *et al.*, « The dark proteome: translation from noncanonical open reading frames Wright », *Trends in Cell Biology*, Volume 32, Issue 3, 243 - 258, mars 2022.

xi La spectrométrie de masse est une technique physique d'analyse permettant de détecter et d'identifier des molécules par mesure de leur masse.

xii Le profilage ribosomique permet de suivre très précisément la traduction de l’ensemble des ARNm cellulaires à un instant donné.

xiii Elizabeth Pennisi, « ‘Dark proteome’ survey reveals thousands of new human genes », *Science* , Volume 386, Issue 6725, Novembre 2024.

xiv *Ibid.*

Adresse de cet article : https://infogm.org/une-avancee-majeure-en-biologie-le-proteome-obscur/