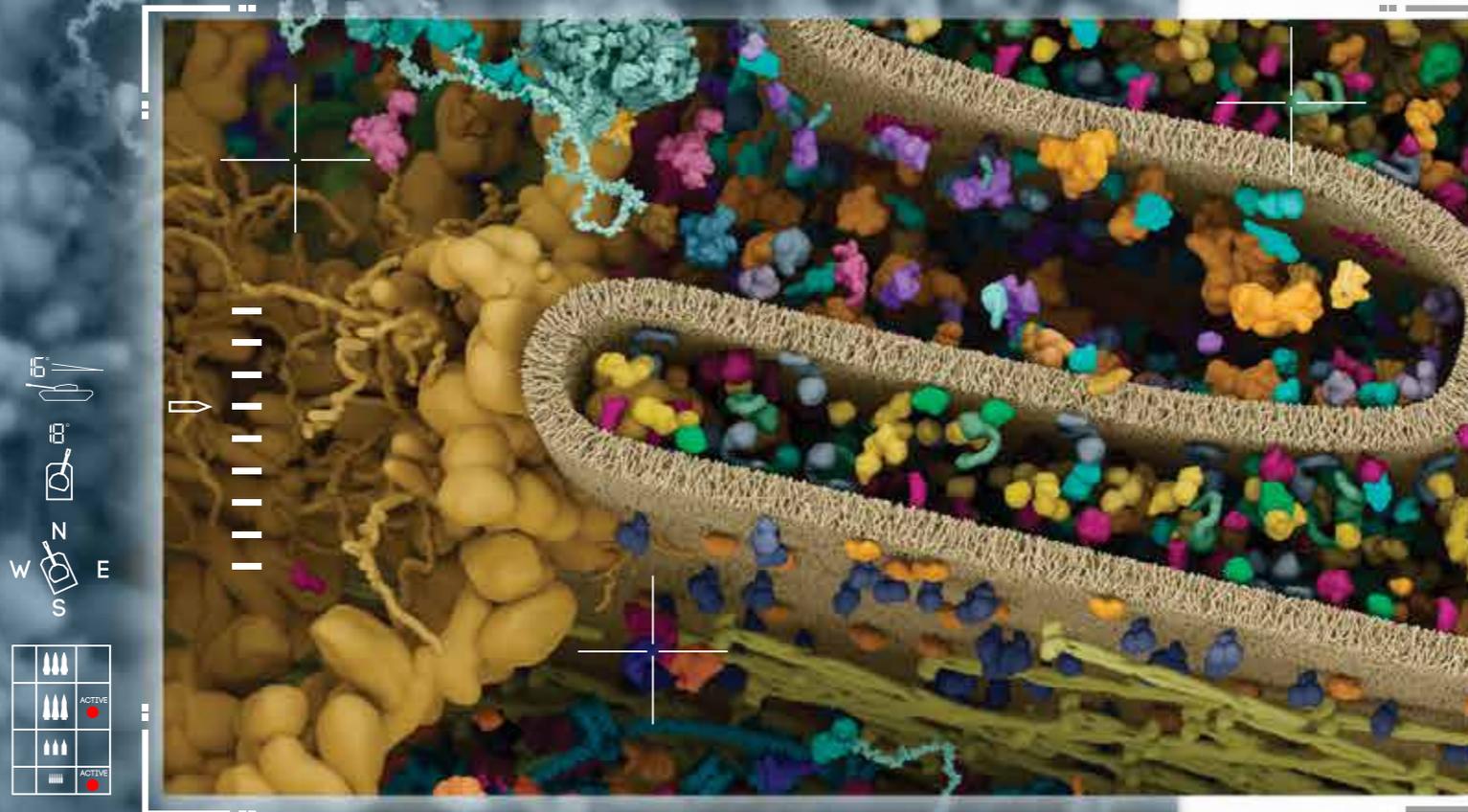


Le mystère des protéines fantômes

REC 3X TIME 05:08 DATE 20.09.2016



NW 16° 18' 8032 7206

Elles ont toujours été là, au cœur de nos cellules... mais incognito. La découverte de milliers de protéines que nul ne soupçonnait est un choc. Car elles pourraient jouer un rôle clé, analyse **Marine Corniou**.

C'est une bombe. Une découverte qui bouleverse nombre de certitudes et donne le vertige face à l'étendue insoupçonnée de notre ignorance.

Jugez plutôt : il y aurait, au cœur de nos cellules, des dizaines de milliers de protéines totalement inconnues. Officiellement, chez l'être humain, les biologistes en répertoriaient jusqu'ici 70 000 ; et voici qu'il y en aurait en fait deux, trois, voire quatre fois plus !

Information sidérante quand on sait que les protéines

sont les ouvrières de la vie, les rouages matériels élémentaires de toute activité. Construites à partir des instructions contenues dans le génome, ces molécules permettent à nos cellules de se structurer, de se réguler, de respirer, bref, d'être vivantes. Eh bien, il faut admettre qu'on ignore encore tout de la grande majorité de ces ouvrières, qui jouent potentiellement un rôle essentiel dans la grande machinerie cellulaire...

"L'impact pourrait être énorme, révolutionnaire", s'enthousiasme Xavier Roucou, biochimiste à l'université de Sherbrooke, au Canada, qui fut le premier, il y a six ans, à se mettre en chasse de ces fantômes, se heurtant longtemps au scepticisme de ses pairs.

Aujourd'hui, de plus en plus de biologistes lui donnent raison. *"Il y a clairement des centaines, voire des milliers de protéines qui sont passées sous le radar. C'est très*

Repères

L'ADN humain est constitué de longs filaments moléculaires comptant 3 milliards de nucléotides (A, T, C ou G). Officiellement, 2 % de ce texte génétique contient des séquences codantes, soit environ 20 000 gènes. Lesquels portent les instructions de construction pour environ le triple de protéines (certains gènes pouvant produire plusieurs versions d'une même protéine). Un nombre très sous-estimé, vient-il d'être découvert.

→ téines. Le reste, longtemps appelé "ADN poubelle", a été mis de côté, puisque non codant. "Une des grosses surprises des années 2000, c'est lorsqu'on s'est aperçu que la grande majorité du génome était en fait transcrit, c'est-à-dire copié en ARN", rappelle John Rinn, qui dirige un laboratoire de recherche sur l'ARN à l'université Harvard. Ces petits filaments mobiles issus de l'ADN poubelle sont soupçonnés de jouer un rôle dans l'expression de certains gènes ou dans l'activité des protéines dans la cellule. Mais, comme le souligne John Rinn, "ces ARN sont appelés 'longs ARN non codants', car ils ne servent pas, en théorie, à fabriquer des protéines".

Faux, indique le résultat de Judith Steen. Et ce n'est pas le seul.

300 000 FANTÔMES !

Depuis trois ou quatre ans, grâce au séquençage rapide et à des spectromètres de masse plus efficaces, les chercheurs ne cessent en fait de découvrir de petites protéines (ou peptides), codées justement par ces ARN considérés comme non codants. "Il y avait plusieurs exemples depuis la fin des années 1990, mais ils étaient considérés comme des exceptions", relate Serge Plaza, chercheur au Centre de biologie du développement de Toulouse, dont l'équipe a découvert en 2010 de tels peptides chez la drosophile. D'ailleurs, Jonathan Weissman a lui aussi constaté que de très nombreux ARN considérés comme non codants sont en fait associés aux ribosomes... et donc, potentiellement traduits !

"Tous les chercheurs ont l'air d'être surpris, mais je ne comprends pas pourquoi. Au sein des milliers de longs ARN non codants décrits, il y a forcément ce qu'on appelle de petits cadres de lecture, dont certains peuvent coder pour des petites protéines. Nous autres, spécialistes de l'ARN, nous le savons, nous l'avons dit, et personne ne nous a écoutés", ironise John Rinn.

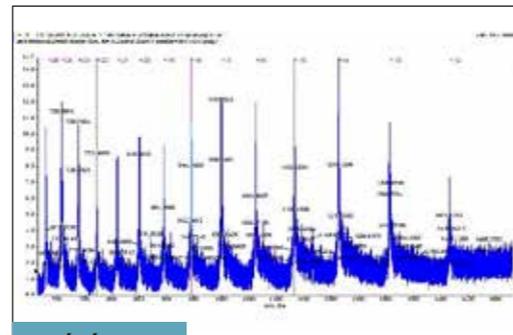
Toujours est-il que le bilan donne aujourd'hui le vertige. Dans une toute récente étude plus exhaustive, car portant sur l'intégralité du génome, Xavier Roucou vient de finalement revoir à la hausse le nombre de ces acteurs imprévus du grand jeu de la vie: "Selon nos estimations, basées sur des modèles informatiques, il pourrait y avoir jusqu'à 300 000 protéines fantômes", conclut-il, sans pouvoir donner un chiffre précis.

Mais enfin, comment a-t-on pu passer à côté de quelque chose d'aussi énorme ?

"Après le séquençage du génome humain, il y a quinze ans, il y avait tellement de données à analyser que les bio-informaticiens ont répertorié une seule séquence codante par gène, la plus longue", explique Xavier Roucou. De quoi obtenir de belles bases de données répertoriant tous les gènes et les protéines possible. "Ces algorithmes ont été mis en place pour discriminer ce qui était codant du bruit de fond", résume Serge Plaza. On a donc ignoré, volontairement et de façon arbitraire, beaucoup de majuscules, de points et de petites phrases potentielles. Puis on a oublié ces simplifications...

Protéines fantômes : trois techniques rivalisent pour les identifier

C'est grâce à de nouvelles techniques d'investigation cellulaire, bien plus fines et précises, que les protéines fantômes peuvent être détectées. Et cela, de trois façons différentes.

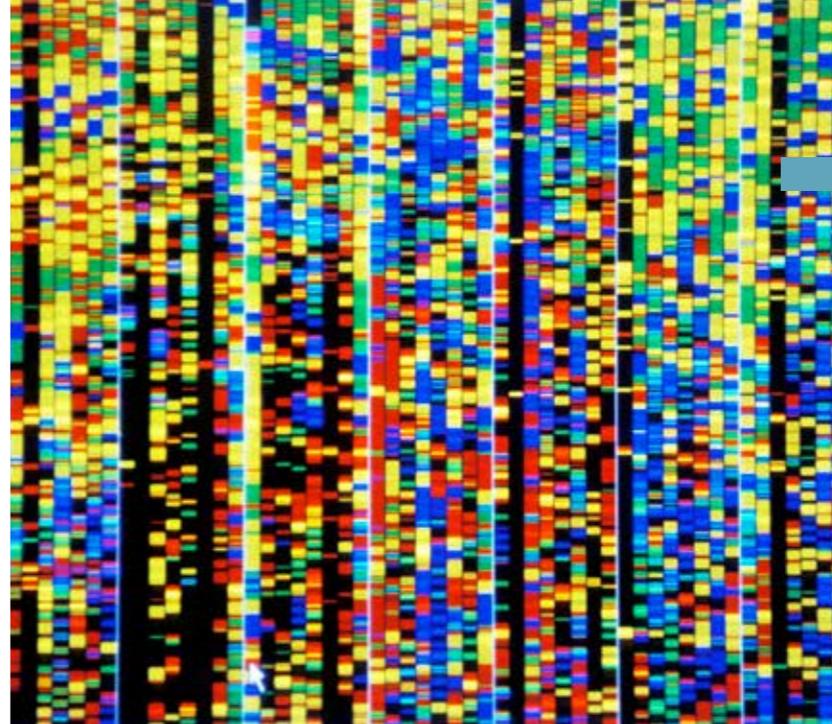


REPÉRÉES GRÂCE AU SPECTROMÈTRE

Les nouveaux spectromètres permettent de déceler plus facilement les petites protéines, qui se noyaient dans le bruit de fond. Ici, chaque pic représente une protéine.

Oui, mais pourquoi la présence matérielle de ces invités surprises n'a-t-elle pas été détectée plus tôt? Serge Plaza souligne les limites des techniques de spectrométrie de masse: "Les petits peptides ont toujours été difficiles à détecter, car ils sont noyés dans les produits de dégradation", explique-t-il. Xavier Roucou y voit aussi les conséquences d'une sorte d'aveuglement qui ne doit rien au hasard: "En fait, en spectrométrie de masse, on rejette habituellement 50% des données puisqu'elles ne correspondent à rien de connu." Selon lui, il aurait suffi de chercher les fantômes pour les trouver. "Si on est passé à côté, c'est surtout parce qu'on s'est enfermé dans des dogmes et qu'on ne s'est pas suffisamment interrogé dessus."

Reste la question centrale: à quoi servent ces protéines oubliées? Sont-elles toutes fonctionnelles? Jouent-elles des rôles cruciaux ou secondaires? Il est encore trop tôt



ANALYSÉES PAR SÉQUENÇAGE HAUT DÉBIT

Le séquençage à haut débit fournit un portrait détaillé de l'ADN et des ARN transcrits. "Scanner" ces données à l'aide de logiciels permet alors d'identifier tous les "cadres de lecture" du génome, représentant autant de protéines potentielles.

pour le dire. "Mais la porte est ouverte", s'enthousiasme Serge Plaza.

Ce qui est sûr, c'est qu'il ne s'agit pas que de "bruit", ou d'erreurs, comme certains ont pu l'avancer. A preuve, les petites protéines de drosophile découvertes par Serge Plaza, qui jouent un rôle dans le développement de l'insecte. Ou celles répertoriées par Judith Steen, dont la quantité varie en fonction de l'activité des neurones, ce qui suggère qu'elles ont une fonction biologique. Ou encore celle mise au jour en 2015 par Eric Olson, de l'université du Texas Southwestern, nommée myoréguline et située dans un ARN "non codant", qui joue un rôle majeur dans la régulation de la contraction musculaire.

STRESS DE LA CELLULE

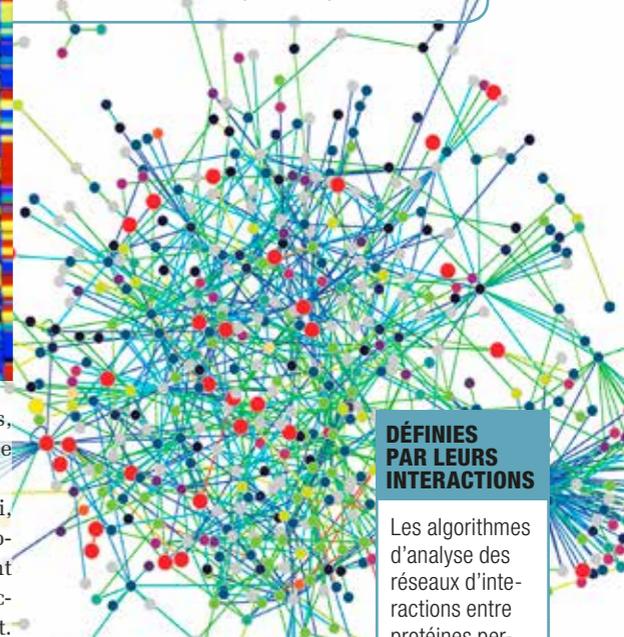
Judith Steen ne l'a pas encore prouvé, mais elle pense que ces protéines alternatives sont surtout produites lorsque la cellule est soumise à un stress. Comme si l'organisme puisait dans ses stocks de nou-

velles séquences codantes, pour avoir plus de chances de s'adapter.

Xavier Roucou, quant à lui, imagine que les protéines codées par le même ARN sont impliquées dans la même fonction biologique et s'entraident. "On commence à avoir des preuves d'interactions fonctionnelles", lâche-t-il, prudent.

Quoi qu'il en soit, on ne peut plus continuer à ignorer ce monde parallèle. "Lorsqu'on bloque un gène à des fins de recherche, ce sont potentiellement 3 ou 4 protéines qu'on supprime, indique-t-il. Et dans certains cas, les protéines alternatives sont même plus abondantes que la protéine de référence!"

Reste à convaincre les biologistes qui, dans leur majorité, ferment encore les yeux. "Nous avons simplifié la biologie à outrance. Pour chaque portion du génome, il faut s'intéresser aux rôles de l'ADN, de l'ARN et des éventuelles protéines de toute taille", affirme John Rinn, arguant que la spécialisation des biologistes dans



DÉFINIES PAR LEURS INTERACTIONS

Les algorithmes d'analyse des réseaux d'interactions entre protéines permettent de déterminer leur fonction, en décelant leurs partenaires de jeu biologiques.

l'un ou l'autre de ces éléments a nuï à leur lucidité, brouillant la vision d'ensemble.

Mais il ne faut pas voir tout cela comme un constat d'échec. Plutôt comme le début d'une nouvelle ère. "On découvre un degré supplémentaire de complexité du génome. C'est peut-être cette complexité qui permettra d'expliquer la diversité du vivant, que l'on n'arrive pas à appréhender en regardant uniquement le génome connu", avance Serge Plaza. Et d'ajouter que c'est aussi cela, la recherche: être confronté à des concepts qui s'effondrent le jour où l'on découvre que la réalité est plus complexe que prévu.

En attendant, la révolution a commencé. Des milliers de protéines, aujourd'hui, ne sont déjà plus des fantômes.



A lire : les publications mentionnées dans l'article.

EN SAVOIR PLUS

science-et-vie.com