

► protéines étaient impliquées, mais les données ont raconté une autre histoire.

Boucle après boucle, deux facteurs protéiques se sont distingués. L'un, appelé CTCF, a été découvert en 1990 par Victor Lobanenko, alors au Royal Cancer Hospital, à Londres, et ses collègues. Il contient onze structures dites «à doigt de zinc» qui lui permettent de se lier étroitement à certains endroits de l'ADN. Le second facteur, la cohésine, décrit en 1997 par Kim Nasmyth, aujourd'hui à l'université d'Oxford, est un complexe en forme d'anneau constitué de plusieurs protéines. À l'époque de nos travaux, les biologistes pensaient que deux anneaux de cohésine étaient capables de se lier et de fonctionner ensemble, chacun encerclant l'ADN et glissant librement sur lui comme un anneau sur un collier.

Retrouver ces protéines n'était pas si surprenant : de nombreuses études antérieures avaient suggéré qu'elles intervenaient dans le repliement du génome. Mais leur implication dans un mécanisme aussi omniprésent que la formation des boucles d'ADN était inattendue.

DES MOTIFS À BOUCLE

C'est alors que nous avons fait une découverte très étrange. Suhas Rao, Miriam Huntley et moi avons demandé à Ido Machol, un nouvel informaticien au laboratoire, d'étudier la distribution des histones – des protéines qui aident à compacter l'ADN dans le noyau – à proximité des facteurs CTCF. Ido Machol a remarqué que les histones étaient plus nombreuses juste à l'extérieur des boucles qu'à l'intérieur, comme si ces protéines savaient où une boucle était positionnée par rapport aux molécules CTCF. Ce résultat ressemblait à une erreur, mais au fil des semaines, Ido Machol n'a détecté aucun bug dans son programme.

Nous avons donc recherché une explication biologique. Victor Lobanenko avait montré que CTCF ne s'attache pas à des positions arbitraires sur l'ADN. Au contraire, le facteur se lie toujours à une séquence spécifique d'environ 20 bases, nommée «motif». L'ADN étant une double hélice, il comporte deux brins orientés en opposition. Chaque brin peut porter des motifs, par conséquent orientés dans un sens ou l'autre sur la longue molécule d'ADN. En général, l'orientation relative des motifs le long de l'ADN est aléatoire, comme un tirage au sort : il y a 50% de chances qu'un motif donné pointe dans un sens et 50% de chances qu'il pointe dans l'autre. Nous pensions donc qu'il en serait de même pour les motifs de liaison à CTCF.

Or, à notre grand étonnement, les deux minuscules motifs de liaison à CTCF qui intervenaient dans la formation d'une boucle – même s'ils étaient séparés par des millions de bases dans l'ADN déplié – pointaient ►

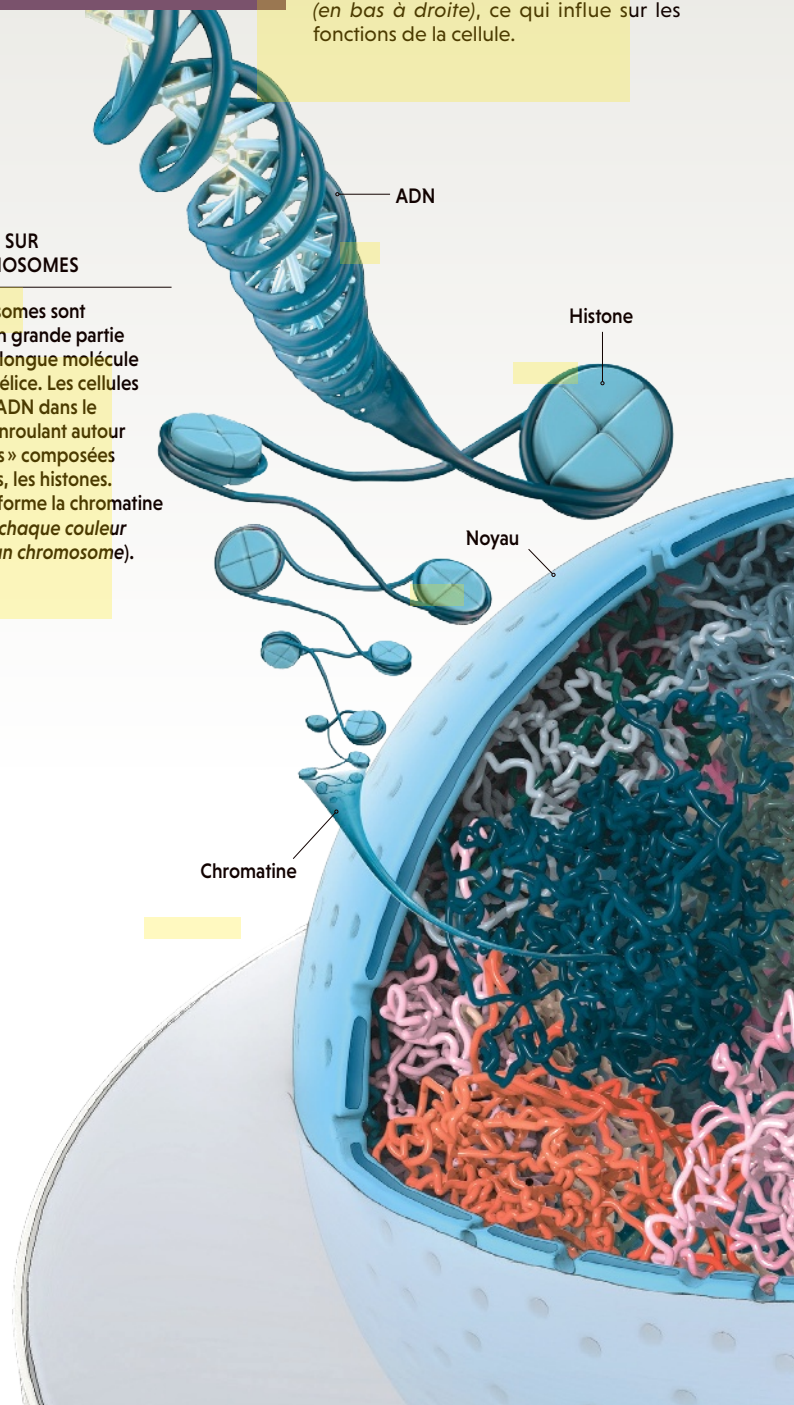
UN GÉNOME BOUCLÉ

Au microscope, le génome d'une cellule – l'ensemble de ses chromosomes – ressemble à un enchevêtrement de nouilles. Pourtant, la disposition est loin d'être aléatoire. Par exemple, avec une meilleure résolution que celle utilisée ici, on verrait que le génome se plie en environ 10 000 boucles qui ne s'entremêlent pas. Des travaux récents ont révélé un processus clé de formation des boucles, nommé «extrusion» (en haut à droite). Ces boucles favorisent l'expression ou l'activation de certains gènes dans différentes cellules (en bas à droite), ce qui influe sur les fonctions de la cellule.

GROS PLAN SUR LES CHROMOSOMES

Les chromosomes sont constitués en grande partie d'ADN, une longue molécule en double hélice. Les cellules emballent l'ADN dans le noyau en l'enroulant autour de «bobines» composées de protéines, les histones. L'ensemble forme la chromatine (ci-dessous, chaque couleur représente un chromosome).

© Falconieri Visuals



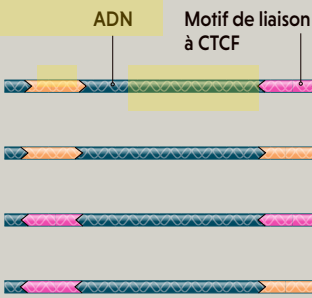
COMMENT LES BOUCLES SE FORMENT

L'extrusion d'une boucle commence quand un « complexe d'extrusion » rencontre l'ADN et s'y arrime. La boucle se forme lorsque deux sous-unités du complexe, comportant chacune une structure en anneau nommée « cohésine », se mettent à glisser dans des directions opposées

le long de l'ADN tout en restant attachées ensemble. Une protéine, CTCF, est capable d'arrêter la croissance de la boucle dans certaines conditions. Elle se lie à une séquence particulière de l'ADN, laquelle pointe dans un sens ou l'autre de la molécule (A). Lorsqu'un anneau de cohésine rencontre une protéine CTCF liée à un motif qui

pointe vers la boucle, l'anneau s'arrête. Si le motif est orienté dans l'autre sens, l'anneau continue d'avancer. Quand les deux anneaux rencontrent des protéines CTCF liées à des motifs pointant vers l'intérieur, ils s'arrêtent donc tous les deux, ancrant la boucle (B). Dans toutes les autres configurations, la boucle continue de croître (C).

A Quatre orientations possibles des motifs

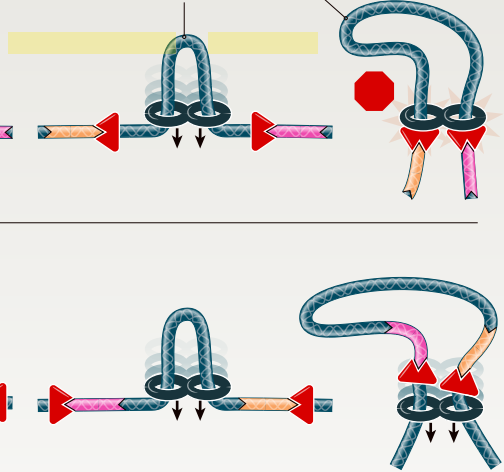


B Une boucle en formation se stabilise quand les anneaux de cohésine rencontrent deux motifs de liaison à CTCF qui pointent vers la boucle.

CTCF Anneau de cohésine

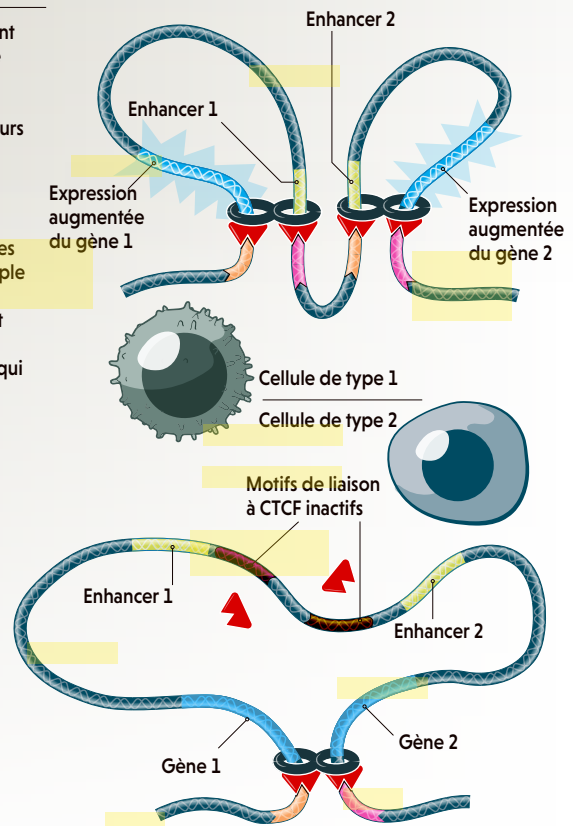
C Ici, les anneaux de cohésine dépassent les motifs de liaison à CTCF, car ils ne pointent pas vers la boucle naissante, laquelle continue à croître.

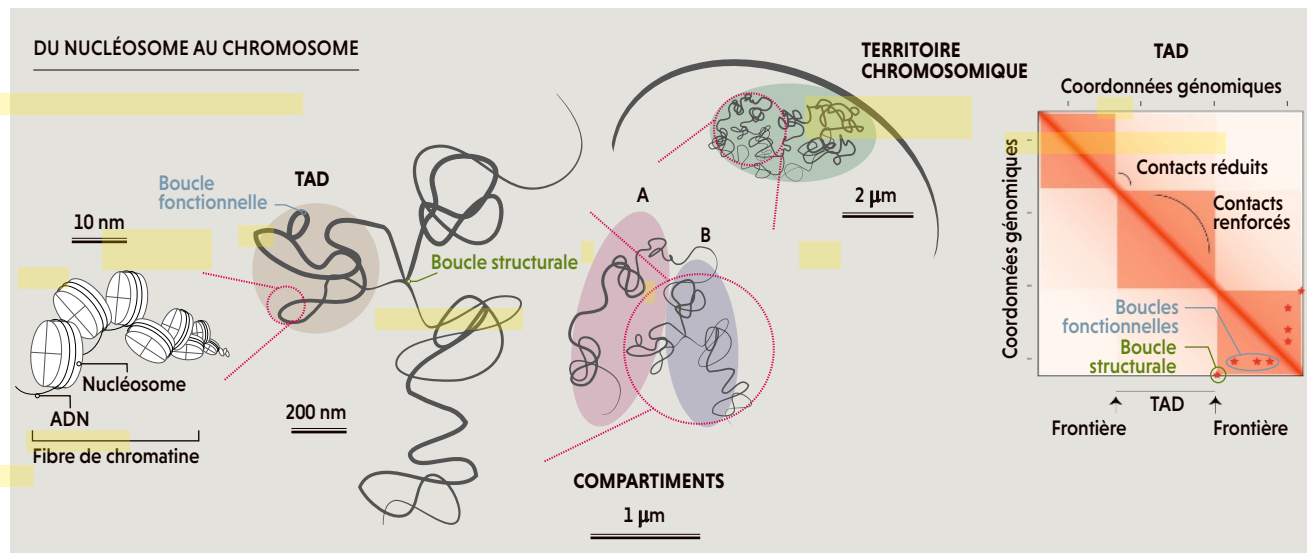
Boucle en formation Boucle stable



L'INTÉRÊT DES BOUCLES

Pour fonctionner correctement, certains gènes doivent entrer en contact avec un segment d'ADN nommé « enhancer ». Mais les enhancers se trouvent souvent loin des gènes qu'ils activent. Les boucles rapprochent les enhancers et leurs gènes (en haut). Toutes les cellules présentent les mêmes gènes et les mêmes motifs de liaison aux protéines CTCF, mais elles ne forment pas toutes les mêmes boucles. Il arrive par exemple que les motifs de liaison à la protéine CTCF soient désactivés à certains endroits de l'ADN, ce qui empêche la protéine de s'y lier (en bas).





UNE ORGANISATION À TOUTES LES ÉCHELLES

L'organisation spatiale du génome est loin d'être aléatoire au sein des noyaux de nos cellules. On connaissait le premier échelon de cette organisation, le nucléosome, dont la structure cristallographique avait été révélée en 1997. Le nucléosome permet d'enrouler la molécule d'ADN et représente donc un premier niveau de compaction de l'ADN pour former la fibre de chromatine (voir la figure ci-dessus à gauche). À l'autre bout de l'échelle, on savait, grâce à des approches cytologiques et de microscopie photonique, que les chromosomes s'organisent en territoires chromosomiques individuels, occupant des positions préférentielles dans les noyaux suivant leur activité génique, leur densité en gènes ou encore leur taille. En revanche, au milieu de l'échelle, les choses paraissaient beaucoup plus floues. On avait notamment observé des repliements à large échelle entre des régions génomiques éloignées, ainsi que des regroupements de gènes même lorsqu'ils étaient situés sur

des chromosomes différents. Bien sûr, on soupçonnait aussi l'existence de repliements de la fibre de chromatine et, par là même, de rapprochements physiques entre les promoteurs des gènes et leurs séquences stimulatrices ou « enhancers » – les boucles qui font l'objet de cet article. Les enhancers sont de véritables boutons d'activation, parfois situés à plusieurs dizaines de milliers de paires de bases des gènes qu'ils régulent. Des avancées majeures dans la compréhension de l'organisation spatiale du génome, de la centaine de kilobases (100 000 paires de bases) jusqu'à la mégabase (1 million de paires de bases), ont vu récemment le jour grâce à l'application de nouvelles techniques de biologie moléculaire, telles que la méthode Hi-C.

En 2009, on a vu apparaître les premières cartes de Hi-C, qui saisissent l'ensemble des contacts entre régions génomiques dans l'espace tridimensionnel du noyau. Plus précisément, ces représentations cartographient la fréquence de chacun des contacts possibles sous la forme de points plus ou moins intenses (voir ci-dessus à droite). Les premières cartes avaient une résolution relativement faible, mais on y distinguait déjà des organisations préférentielles de larges régions génomiques et, notamment, l'organisation en compartiments actif et réprimé, appelés respectivement compartiments A et B. En 2012, l'application de la méthode Hi-C ou de dérivées avec une résolution accrue a permis à trois équipes – la nôtre, celle d'Edith Heard, à l'institut Curie, et celle de Bing Ren, à l'institut Ludwig pour la recherche sur le cancer à La Jolla,

aux États-Unis – de découvrir de nouveaux domaines d'organisation de la chromatine, qui sont désormais appelés TAD (pour *topologically associating domains*).

Le long des chromosomes, les TAD représentent des zones du génome caractérisées par des interactions internes bien plus fortes que les interactions de TAD adjacents. Entre les TAD, on trouve de véritables frontières, qui conduisent à la partition du génome en une série de TAD. L'obtention de cartes de Hi-C à très haute résolution a permis de mettre en évidence de nombreuses boucles d'ADN, décrites dans cet article. Certaines, dites « structurales », séparent les TAD les uns des autres. D'autres, situées à l'intérieur des TAD, joueraient divers rôles fonctionnels. Certaines rapprocheraient des promoteurs et des enhancers, d'autres les deux extrémités d'un gène (ce qui renforcerait la direction selon laquelle le gène est transcrit), d'autres encore se formeraient entre des sites reconnus par des protéines impliquées dans la répression de nombreux gènes – les protéines du groupe Polycomb –, renforçant leur action. En d'autres termes, les TAD permettent de mieux confiner certaines régions du génome, ce qui facilite des repliements de la fibre chromatinienne et la régulation de l'expression des gènes.

La découverte de ces différents échelons de l'organisation des chromosomes a ouvert de nouvelles voies d'investigation sur la compréhension du fonctionnement du génome, ainsi que sur celle de maladies dont l'origine restait mystérieuse. Notamment, des réarrangements chromosomiques, du type