

# Génétique et épigénétique

Miguel Beato

*Directeur du Centre de régulation génomique de Barcelone*

Le fait que nous possédions maintenant la séquence entière du génome humain est un grand bond en avant dans les informations dont nous disposons, une avancée majeure qui nous ouvre un nombre incroyable de possibilités. Cependant, à l'heure actuelle, cette connaissance ne nous a pas encore réellement aidés à résoudre quoi que ce soit parce que nous n'avons considéré que les gènes et que nous disposions déjà de cette information génétique. Bien que nous connaissions maintenant tous les gènes en cause, nous n'avons pas encore appris grand-chose. Les questions biologiques les plus importantes n'ont pas encore été résolues et nous devons utiliser ces informations de la manière préconisée par Axel Kahn, c'est-à-dire en recherchant comment ces informations contextuelles et combinatoires – et la manière dont les combinaisons sont assemblées – sont codées dans le génome. Cependant, jusqu'à présent, nous ne savons pas comment cela se fait. Nous ne comprenons pas les mécanismes sous-jacents de sorte que, bien que la connaissance des informations génétiques soit un préalable indispensable, beaucoup d'autres informations manquent.

Je voudrais donc examiner quelles sont les informations qui sont codées dans le génome et, à partir de là, envisager comment il serait possible d'utiliser ces informations pour développer un organisme et contrôler sa fonction.

Parmi les divers types d'informations qui se trouvent dans l'ADN, nous avons évoqué principalement jusqu'ici celles qui permettent de traduire en protéines, grâce au code génétique, les séquences portées par les ARN messagers et par les gènes : un triplet de trois nucléotides pour un acide aminé. Elles ne représentent pourtant que moins de 3 % des informations fournies par le génome humain : en d'autres termes, 97 % des informations du génome ne sont pas directement liées aux gènes. Le génome est bien plus qu'un ensemble de gènes.

Vous connaissez déjà une partie de ces informations. Ce sont celles qui concernent les fonctions régulatrices : de courtes séquences de paires de bases lues par des protéines liant l'ADN et qui déterminent quels gènes sont exprimés, la force de cette expression et le tissu dans lequel elle se produit. Ces informations peuvent être lues de manière courante. Je vais vous donner des exemples. Les parois des cellules sont très souvent palindromiques, elles ont des axes symétriques et sont enroulées de la

même manière dans les deux directions, et c'est pourquoi les informations génétiques qui les concernent sont lues par des dimères protéiques. Un autre exemple est celui de l'information reconnue par des récepteurs des hormones stéroïdes : il existe deux molécules réceptrices capables de se lier chacune à la moitié d'une double hélice, c'est-à-dire à l'ADN.

Ici, il faut noter un fait important. L'ADN est souvent traité par les généticiens comme une simple séquence linéaire d'informations, dépourvue d'informations conformationnelles. En réalité, l'ADN est une molécule compliquée dont la structure tridimensionnelle est d'une importance vitale pour sa fonction. Il ne s'agit pas d'un texte linéaire, c'est une molécule qui possède une structure très précise. Cette structure est lue, est reconnue par des processus qui mettent en jeu la pénétration des chaînes latérales des acides aminés dans le sillon majeur de l'ADN – qui est la partie plus ouverte de la séquence – et la lecture des séquences de paires de bases. Ces informations régulatrices représentent environ 1 % du génome humain.

Nous savons maintenant de quoi se constitue 4 % du génome humain, ce qui laisse quand même 96 % d'inconnu... Un autre aspect que nous commençons à comprendre est lié à ce que l'on appelle les informations topologiques. Le premier pas vers la compréhension de ces informations complexes consiste à déterminer comment l'ADN est disposé dans les zones du noyau qui constituent les unités de base de la chromatine. Le génome humain n'est pas libre dans le noyau de la cellule, il est emballé dans la chromatine. La double hélice est enroulée autour d'unités de base appelées nucléosomes, qui ressemblent à des cylindres aplatis disposés le long d'un cordon d'histones. Cet ensemble est à son tour enroulé sur lui-même pour former une super hélice comportant environ six nucléosomes par tour d'hélice et un pas de 110 Å. C'est ainsi qu'est constituée la fibre de chromatine. Cette fibre est elle-même enroulée en boucles radiales. Il résulte de la superposition de ces différents niveaux d'organisation que le génome est compacté. Mesurant deux mètres de long, il est confiné dans un noyau cellulaire dont le diamètre n'est que de 2 micromètres environ. C'est dire à quel point il doit être compact.

Pendant longtemps, on a cru que la seule fonction de l'empaquetage de l'ADN dans la chromatine était de le rendre suffisamment compact pour qu'il tienne dans le noyau cellulaire. Maintenant, il devient très clair que l'ADN est emballé dans la chromatine d'une manière très précise, au nucléotide près, et que cet emballage influe sur l'expression du message génétique. Selon la manière dont il est emballé, il sera lu ou non par les protéines qui reconnaissent la séquence. Les séquences du sillon majeur de l'ADN sont exposées tandis que d'autres, dirigées vers l'intérieur du nucléosome, ne sont pas lisibles. Elles ne sont pas accessibles aux protéines qui recherchent des informations dans le sillon majeur de l'ADN. Si le positionnement est précis, cette information topologique domine toutes les autres dans la hiérarchie des informations régulatrices, parce qu'elle détermine si ces informations

régulatrices peuvent être lues ou non. Une protéine susceptible de reconnaître une séquence donnée ne sera pas capable de la voir que celle-ci est dirigée vers l'extérieur, est exposée. Il s'agit bien d'informations génétiques topologiques car la manière dont la molécule d'ADN est disposée dans l'espace est une propriété intrinsèque de la séquence.

Il existe ainsi, en quelque sorte, plusieurs alphabets dans l'ADN et ces alphabets se recouvrent :

- la double hélice est le support du code génétique, dont l'alphabet est constitué par ces triplets de bases qui permettent l'appariement à l'ARN de transfert, la transcription en ARN messager et finalement la synthèse des protéines. C'est ce qu'ont décrit Watson et Crick,

- le deuxième alphabet est le code régulateur, qui permet la lecture par les protéines de séquences situées dans le sillon majeur de l'ADN,

- vient enfin le code topologique, qui repose sur la séquence de nucléotides elle-même. C'est cette dernière qui détermine la manière dont l'ADN se replie pour s'enrouler autour des nucléosomes. C'est un repliement très impressionnant : si vous essayez d'en faire autant avec une maquette de la double hélice d'ADN, elle se casse... Vous devez déformer tous les angles pour former ces minuscules cercles d'ADN qui s'enroulent autour du nucléosome. Cette disposition spatiale est entièrement déterminée par la séquence des nucléotides : c'est une propriété intrinsèque de la séquence.

Il existe bien d'autres informations topologiques, dans la chromatine par exemple. Considérons deux protéines qui doivent se lier à deux séquences séparées par 80 paires de bases : elles ne seront pas capables d'interagir dans l'ADN libre – qui n'est d'ailleurs qu'une vue de l'esprit –, mais peuvent devenir très proches l'une de l'autre dans la chromatine, du fait des repliements. La chromatine prépare ainsi le terrain pour des interactions qui ne sont possibles que lorsque l'ADN est enroulé.

Le même type de rapprochements se produit aux sites d'entrée et de sortie du nucléosome, qui sont séparés dans la séquence linéaire par 200 paires de bases et qui sont pourtant très proches dans l'espace de la chromatine. Dans la fibre de chromatine elle-même, qui est formée de six nucléosomes, vous pouvez voir devenir très proches l'une de l'autre des séquences théoriquement séparées par 1 200 paires de bases.

Il existe donc dans la structure de la chromatine des informations qui peuvent influencer sur la manière dont le message est lu. La cellule comporte une quantité considérable d'informations sur les enzymes qui modifient la structure de la chromatine, ce que l'on appelle les machines à remodeler la chromatine. Ce sont des enzymes très complexes, à 10 ou 15 polypeptides, qui utilisent l'énergie de l'ATP pour ouvrir ou fermer la chromatine. Certaines protéines telles que le récepteur de la progestérone, par exemple, sont capables de recruter la machine qui déclenche le

remodelage des protéines. Grâce à l'énergie de l'ATP, elles ouvrent le nucléosome en dissociant des histones et permettent ainsi à d'autres protéines, telles que des facteurs de transcription, de reconnaître leur cible sur la chromatine.

Il est bien démontré que la mise en œuvre du message génétique est interactive. L'ADN n'est pas un simple alignement de séquences reconnues par des protéines, il contient des informations qui lui permettent de se positionner dans le nucléosome de sorte que seules certaines séquences soient visibles.

Les informations topologiques portées par l'ADN ne se manifestent que dans la structure de la chromatine. Vous avez donc une séquence qui n'a aucun sens intelligible – sur le plan topologique – dans laquelle la chromatine introduit un ordre plus élevé de signification, où sont inscrites des informations topologiques qu'il est impossible de lire dans la séquence d'ADN. Si vous ne disposez que de la séquence de l'ADN, vous n'avez aucun moyen de décoder les informations topologiques : vous devez pour cela y superposer la structure de la chromatine.

Il existe donc au moins trois types d'informations disponibles et je ne doute pas que bien d'autres restent à découvrir. Bien qu'il puisse y avoir dans l'ADN certaines informations inutiles, il ne faut pas céder à la facilité de considérer comme inutile ce que nous ne comprenons pas. Il existe des informations génétiques qui permettent créer des réseaux de gènes, des domaines de chromatine, sur lesquelles nous savons déjà beaucoup – je les ai évoquées – mais aussi des programmes morphogénétiques et métaboliques. Je suis sûr qu'il existe dans le génome une structure qui régule la manière dont les programmes sont mis en œuvre.

Outre ces facteurs génomiques – gènes à séquences de nucléotides, gènes à séquences régulatrices, codes topologiques, réseaux, boucles de régulation –, il existe quelque chose d'autre qui est maintenant réellement à la pointe de l'évolution en science fondamentale et en biologie moléculaire fondamentale. Ce qui retient maintenant l'attention ce n'est plus la génomique mais l'épigénomique. C'est la manière dont, sans changer la séquence d'ADN, une modification des histones peut changer la structure de la chromatine et la mise en œuvre du message génétique. Dans certains cas, ces processus sont associés à une méthylation de la séquence d'ADN. On a pu parler à propos de ces phénomènes, encore très mal connus, de « code des histones ».

Ce qu'il faut retenir, c'est que ces modifications influent sur la physiologie de la cellule. Par exemple, la phosphorylation d'une sérine en position 10 sur l'histone H3 peut, en fonction de la kinase mise en jeu, déclencher la transcription ou la mitose. La phosphorylation de l'histone H2A sur la sérine en position 139, ou de l'histone H2B sur la sérine en position 32, peut déclencher ou contribuer à déclencher les processus d'apoptose. Selon la manière dont elle agit, une enzyme donnée peut même produire deux résultats différents concernant le message. Par exemple, quand la cellule est en interphase, la phosphorylation de l'histone H3 en position 10

stimule l'acétylation d'autres résidus lysine, ce qui aboutit à la transcription. Au contraire, à la métaphase, il se produit les phosphorylations sont différentes et portent sur deux résidus sérines, non seulement en position 10 mais aussi 28, et alors la protéine se condense pour la métaphase.

Le comportement de la cellule est influencé de manière spectaculaire par de simples modifications de la manière dont la chromatine est emballée. C'est très important pour le traitement par cellules souches et aussi pour le clonage, car l'une des raisons pour lesquelles le clonage à partir de noyaux somatiques est tellement inefficace est que le programme épigénétique d'un noyau somatique est complètement différent de celui d'un blastocyte. De même, le programme épigénétique d'une cellule souche diffère radicalement de celui de la lignée germinale.

Pour résumer le code nucléosomique des histones, de multiples voies en série convergent sur les histones. Elles agissent en modifiant les sérines de manière sélective, aboutissant à des combinaisons mettant en jeu l'acétylation, la méthylation et la phosphorylation de résidus spécifiques. Ces modifications affectent la structure d'ordre supérieur de la chromatine et l'interaction avec les protéines régulatrices. Le code nucléosomique influence la lecture du message génétique de manière transmissible. Le message nucléosomique peut être transmis durant la division cellulaire : c'est pour cette raison que les hépatocytes continuent d'être des hépatocytes après la division cellulaire, c'est parce que les informations épigénétiques sont transmissibles.

En conclusion, l'épigénétique est le support d'un système de régulation fondamentalement nouveau, qui se situe au-delà de la séquence d'informations de notre code génétique. L'épigénétique est peut-être l'une des clefs futures de la compréhension des maladies héréditaires et de l'identification de nouvelles cibles thérapeutiques.