

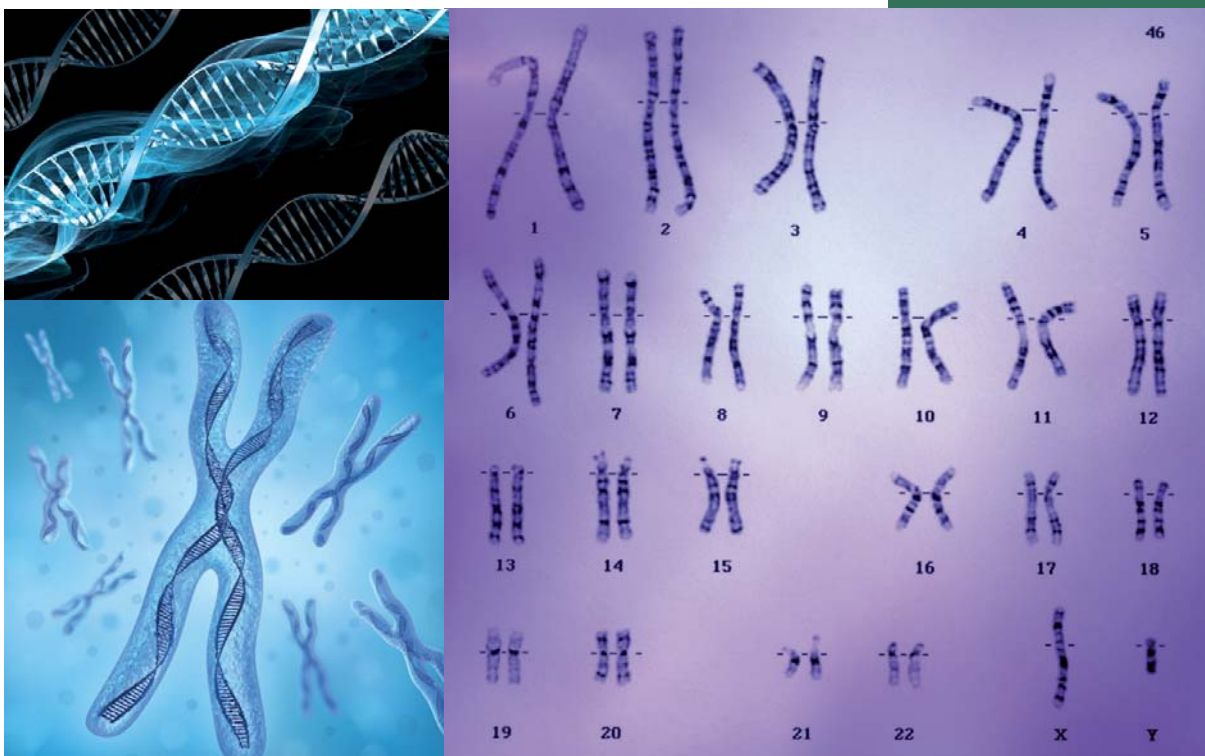
# Cibler l'ADN : pour la compréhension du vivant

Au sein de nos cellules se trouve l'ensemble de nos gènes – le génome –, contenus dans une grande molécule en forme de double hélice : l'ADN ou Acide DésoxyriboNucléique. L'ADN contient toutes les informations permettant à l'organisme de vivre et de se développer ; il est le support de notre information génétique, mais également celui de l'hérédité (**Figure 1**). Une molécule *a priori* relativement simple peut-elle commander à elle seule tout le fonctionnement d'un organisme ?

Si nous connaissons maintenant tout le génome humain depuis son séquençage en 2003 (**Encart « Un peu d'histoire sur l'ADN »**), nous ne connaissons néanmoins les fonctions que de 10 % de nos gènes... Il reste à découvrir quelles informations renferment les autres. Pour avancer dans la connaissance du génome, une méthode consiste à cibler et à agir sur ces gènes pour qu'ils ne s'expriment plus normalement ; c'est de cette manière que l'on peut parvenir à identifier leurs fonctions respectives.

Figure 1

L'ADN, double hélice constitutive de nos chromosomes, molécule support de l'information génétique et de l'hérédité.



UN PEU D'HISTOIRE SUR L'ADN

des gènes à l'ADN...

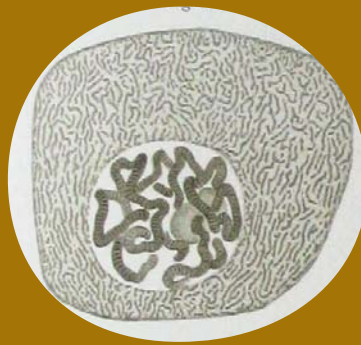
1865 : **Johann Gregor Mendel** établit les bases de l'hérédité en définissant la manière dont les gènes se transmettent de génération en génération : ce sont les **lois de Mendel**.



Johann Gregor Mendel (1822-1884), le père fondateur de la génétique.

1869 : **Johann Friedrich Miescher** découvre dans le noyau des cellules vivantes une substance riche en phosphate – la nucléine –, qui sera nommée au xx<sup>e</sup> siècle « ADN » ou Acide DésoxyriboNucléique.

1882 : **Walther Flemming** met en évidence les chromosomes, constitués de molécules d'ADN, qui regroupent plusieurs gènes. Il décrit pour la première fois la **mitose**, phénomène par lequel les cellules se divisent et permettent la croissance et le renouvellement cellulaires.



Dessin de chromosomes dans un noyau de cellule, par J.F. Flemming.

1928 : **Phoebus Levene** puis **Erwin Chargaff** déterminent la structure chimique de l'ADN avec sa composition en bases azotées : adénine A, thymine T, guanine G et cytosine C (**Figure 2**).

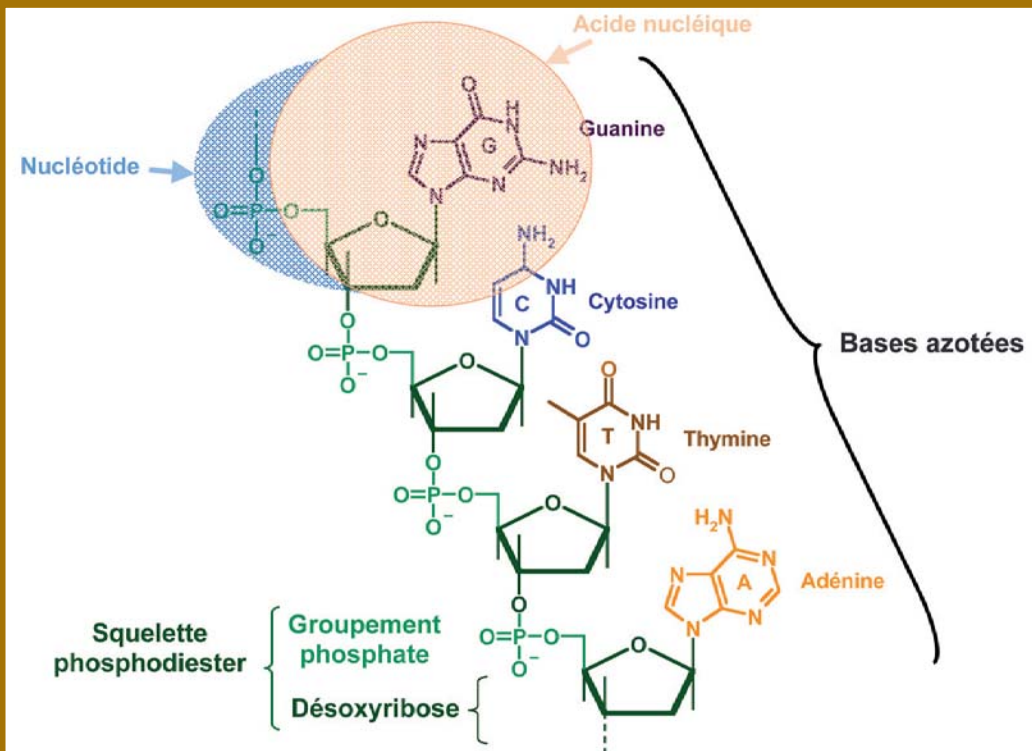


Figure 2

Formule chimique d'un fragment de brin d'ADN. L'ADN est une grande chaîne dont les maillons sont des bases azotées (A, T, G ou C) qui s'enchaînent dans des ordres différents. L'ensemble {désoxyribose + groupement phosphate + base azotée} forme un nucléotide. L'ADN est donc un polymère de nucléotides.

**1944 : Oswald Avery** établit que l'ADN est le transporteur de l'information génétique.

**1952 : James Watson et Francis Crick** établissent la **structure en double hélice de l'ADN (Figure 3)**, ce qui leur valut le prix Nobel de physiologie et de médecine en 1962.

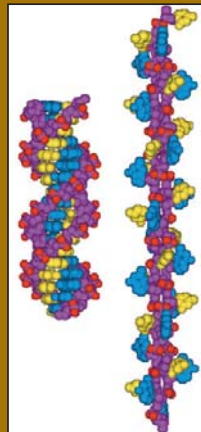


Figure 3

L'ADN en double hélice, la figure emblématique de la biologie moléculaire.

### La nouvelle ère : en route pour l'épopée du séquençage du génome humain...

En 1995, les chercheurs ont pu « lire » pour la première fois tout l'ADN contenu dans le génome d'un organisme unicellulaire. Depuis cette date, des pas de géant ont été franchis en génétique, et les chercheurs ont ainsi séquencé des génomes de plus en plus longs, aboutissant en avril 2003 au **séquençage complet du génome humain**.

C'est le résultat de nombreuses années de travail impliquant plusieurs pays à travers le monde, associés dans ce « *projet génome humain* », parfois appelé « *Apollo de la biologie* » (Figure 4). Et pour cause, ce fut un travail titanesque de plus de dix ans, à l'issue duquel les chercheurs ont décrypté les quelque 3,5 milliards de bases de notre ADN. L'idée admise jusque-là était que le génome humain contenait au moins 100 000 gènes... ! Puis l'équipe française du Génoscope, dirigée par Jean Weissenbach\* suggéra le nombre de 30 000 gènes, ce qui fut jugé un peu iconoclaste. L'évaluation actuelle a encore réduit ce chiffre pour le ramener à environ 25 000... Devant les 37 000 gènes du simple grain de riz, cela incite à la réflexion.



Figure 4

Le projet « *Apollo de la biologie* » : son ambition était à l'image de celle de marcher sur la Lune. Centre national du séquençage – Génoscope d'Evry.

Aujourd'hui, plus de 5 000 génomes ont été séquencés, ce qui aurait été impossible sans les sauts technologiques énormes réalisés ces dernières années. Il faut citer l'invention de la carte génétique de haute précision par Jean Weissenbach\*, qui a été décisive pour le diagnostic précoce de pathologies génétiques.

La recherche est maintenant armée pour séquencer les gènes plus rapidement et à un moindre coût. Ce domaine de recherche, connu pour représenter un travail colossal et un véritable « puzzle », évolue encore très vite et prévoit de plus en plus de gènes et de fonctions à étudier, de plus en plus d'éléments à comprendre sur le fonctionnement du vivant.

\* Jean Weissenbach, responsable du puissant génotype au Centre national de séquençage d'Evry (centre de séquençage français du « projet Apollo »), a reçu pour ses travaux pionniers la prestigieuse médaille d'or du CNRS en 2008.

Où en sommes-nous dans le développement de cette approche toute récente qu'est le **ciblage de l'ADN** ? Comment peut-elle être appliquée en thérapie génique et de manière générale dans les manipulations génétiques ?

## 1 Prise de connaissance avec la cible : l'ADN

### 1.1. L'ADN : découverte de sa structure et de son fonctionnement

Depuis le XIX<sup>e</sup> siècle, la nature et le rôle de cette grande molécule en forme d'hélice se sont révélés par étapes successives aux scientifiques puis au grand public (*Encart « Un peu d'histoire sur l'ADN »*).

C'est ainsi que dès le milieu du XIX<sup>e</sup> siècle, nous avons déjà compris les bases de notre fonctionnement.

C'est le long de la chaîne d'ADN qu'est inscrit tout notre patrimoine génétique ; et c'est l'enchaînement des bases azotées A, T, G, et C – encore appelé séquence d'ADN –, différent d'un individu à l'autre, qui détermine nos caractéristiques physiologiques, et donc notre identité.

*Faisons un zoom sur les pages de notre hérédité, dont les mots sont A, T, G et C...*

L'ADN a une structure très simple et répétitive, qui repose sur un squelette phosphodiester (groupements phosphates et désoxyriboses, *Figure 2*). Sur ce squelette sont attachées des briques élémentaires qui sont les bases azotées, parmi les quatre possibles : adénine A, thymine T, guanine G et cytosine C (*Figure 2*). Squelette phosphodiester et bases forment un brin. L'ADN est constitué de deux brins qui s'enroulent l'un autour de l'autre de manière très précise en une double hélice. Ils sont maintenus solidaires grâce à la formation de paires de bases : l'adénine se lie avec la thymine – et seulement avec elle – en établissant des liaisons de faible intensité (liaisons hydrogène), et la cytosine se lie avec la guanine (*Figure 5*).

C'est cette structure très simple dans son principe et à la fois très riche qui a été à la base du « **dogme de la biologie moléculaire** » : la molécule d'ADN est le support de l'information génétique et, à travers plusieurs étapes, elle conduit à la synthèse de protéines, lesquelles sont



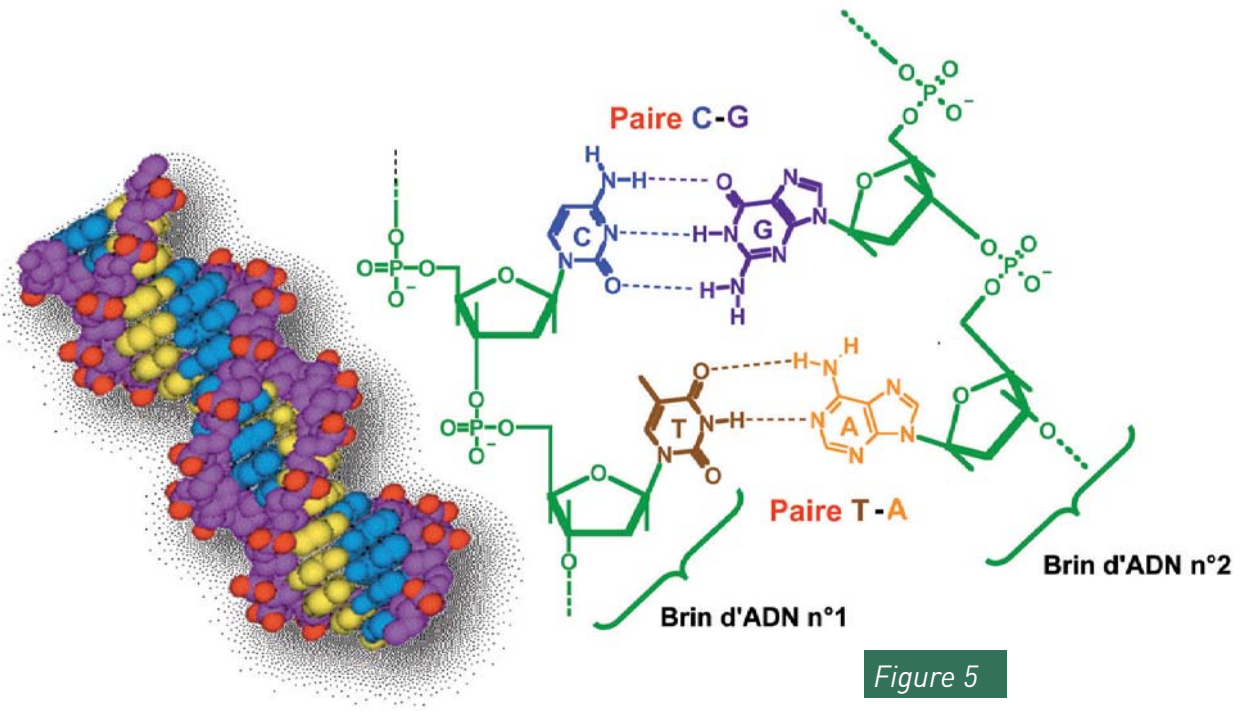


Figure 5

Les paires de bases qui permettent aux deux brins d'ADN de s'associer. L'adénine et la thymine se lient entre elles par deux liaisons hydrogène, tandis que trois liaisons hydrogène lient la cytosine et la guanine.

associées à des fonctions bien précises dans la cellule. Ces étapes sont les suivantes (Figure 6) :

1) la **transcription** : c'est la production d'une copie de l'ADN en ARN messager (ARN = Acide RiboNucléique), qui a une structure chimique très proche ;

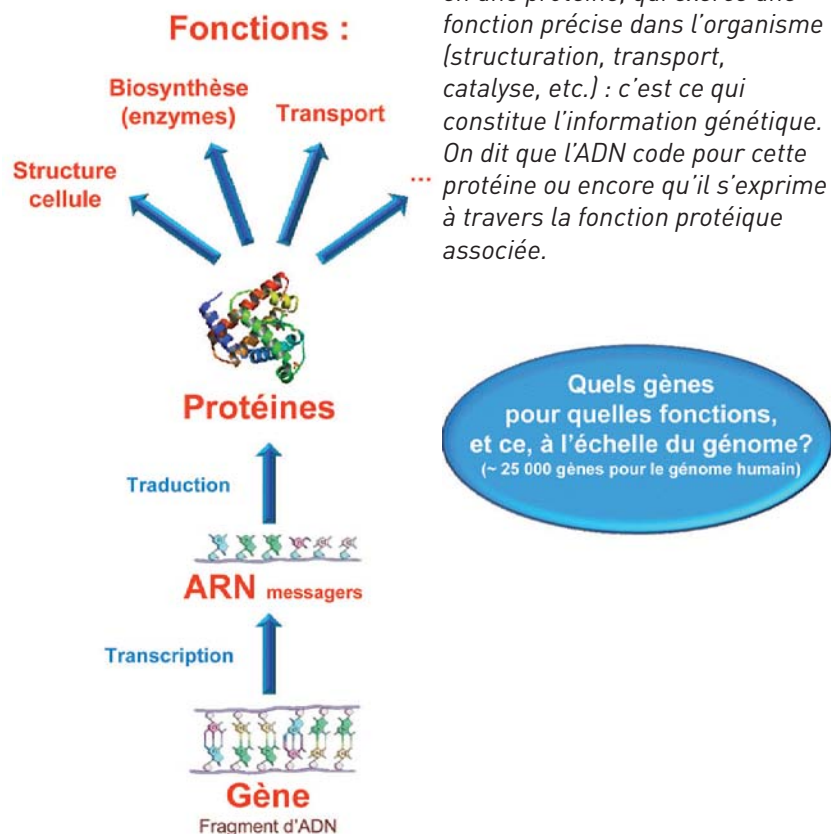
2) la **traduction** : l'ARN est le messenger de l'information génétique, et son message se traduit par la synthèse d'une protéine. Les protéines ainsi produites par l'organisme jouent des rôles physiologiques divers : rôle structural (membranes des cellules, etc.), biosynthèse de molécules indispensables à la vie (glucides, lipides, acides nucléiques), rôle de messagers (hormones), etc. Les protéines sont donc des constituants majeurs et les principaux bâtisseurs cellulaires.

Un corolaire découlant du dogme de la biologie moléculaire est que l'ADN peut être copié. Chaque brin d'ADN pouvant servir de copie, une molécule d'ADN peut donc en donner deux au moment

de la division cellulaire (la mitose), permettant ainsi le transfert de l'information génétique de la cellule-mère aux cellules-filles. C'est ainsi qu'est transmise l'information génétique d'une cellule à l'autre au cours du développement cellulaire et de son renouvellement, et c'est ainsi que nous transmettons notre

Figure 6

Le dogme de la biologie moléculaire : une séquence d'ADN (un gène) est transcrite en un ARN messager, lequel est traduit en une protéine, qui exerce une fonction précise dans l'organisme (structuration, transport, catalyse, etc.) : c'est ce qui constitue l'information génétique. On dit que l'ADN code pour cette protéine ou encore qu'il s'exprime à travers la fonction protéique associée.



information génétique à nos descendants.

### 1.2. Le génome est très complexe et les gènes ne sont pas seuls dans la partie

Comment une molécule relativement simple comme l'ADN pourrait-elle contenir toute l'information génétique qui nous est nécessaire ? En fait, le dogme de la biologie moléculaire a été récemment révisité, et l'on sait maintenant que pour les quelque 25 000 gènes que nous avons, il existe dix fois plus d'ARN, et trois mille fois plus de protéines (soit 10 000 000 protéines) ! On voit donc qu'un gène peut être associé à une multitude de fonctions différentes, ce qui rend très complexe l'étude du génome.

Quant aux ARN, on sait maintenant que 5 % d'entre eux ne donnent pas de protéines : ce sont des ARN non codants appelés ARN régulateurs, à l'instar de nombreuses protéines connues pour jouer un rôle de régulation dans l'organisme. Ils représentent une quantité importante par rapport à l'ensemble de l'information génétique.

Pour compliquer le tout, il se trouve que le génome n'est pas constitué que de gènes. Les régions codantes n'en représentent que 5 % ! On a montré qu'il existe des régions sans gène, mais qui ont des fonctions biologiques très particulières. Mais quelles sont donc ces fonctions ?

Seuls 5 % de notre génome correspondent à des séquences codantes.

En fait, nous verrons que ces 5 % ne résument pas à eux seuls tout notre fonctionnement, et que l'ADN n'est pas un élément autonome. Il existe en effet des protéines dans l'organisme qui interagissent avec lui et jouent le rôle de facteurs régulateurs, c'est-à-dire qu'ils activent ou inhibent la fonction associée. Il a été prouvé récemment qu'au cours des divisions cellulaires, certaines informations sont transférées non pas directement par le séquençage de l'ADN, comme résumé par le dogme de la biologie moléculaire, mais par des modifications de facteurs régulateurs appelées **modifications épigénétiques** (déjà évoqués dans le *chapitre de J.-F. Bach*). En particulier, les nucléosomes sont des structures jouant un rôle important dans ces régulations épigénétiques (*Encart « L'ADN, une pelote d'information génétique dans nos cellules »*).

### 1.3. Identifier les fonctions des gènes

Les scientifiques se sont donné l'objectif d'identifier les fonctions de tous nos gènes, pour aboutir à connaître toute l'information génétique renfermée dans notre génome. Pour cela, l'approche du **ciblage de l'ADN** a été envisagée pour modifier son expression.

Par ailleurs, en ciblant l'ADN, on peut interférer avec les facteurs régulateurs associés, et donc identifier leur rôle dans la régulation de son expression.

Quelles molécules peut-on donc utiliser pour cibler

## L'ADN, UNE PELOTE D'INFORMATION GÉNÉTIQUE DANS NOS CELLULES

Le génome humain comporte plus de trois milliards de paires de base et, selon J. Watson et F. Crick, la distance entre deux paires de bases est de 3,4 Å, ce qui donne deux mètres linéaires d'ADN dans nos cellules, mis bout à bout ! Un noyau cellulaire ayant un diamètre de quelques microns, il faut donc compacter l'ADN pour le contenir dans ce noyau. La solution trouvée par la nature est d'enrouler l'ADN autour de protéines appelées histones, pour former des sortes de pelotes appelées nucléosomes (Figure 7). Ces pelotes de nucléosomes ne sont pas anodines, car elles ont leur rôle à jouer dans les régulations épigénétiques !

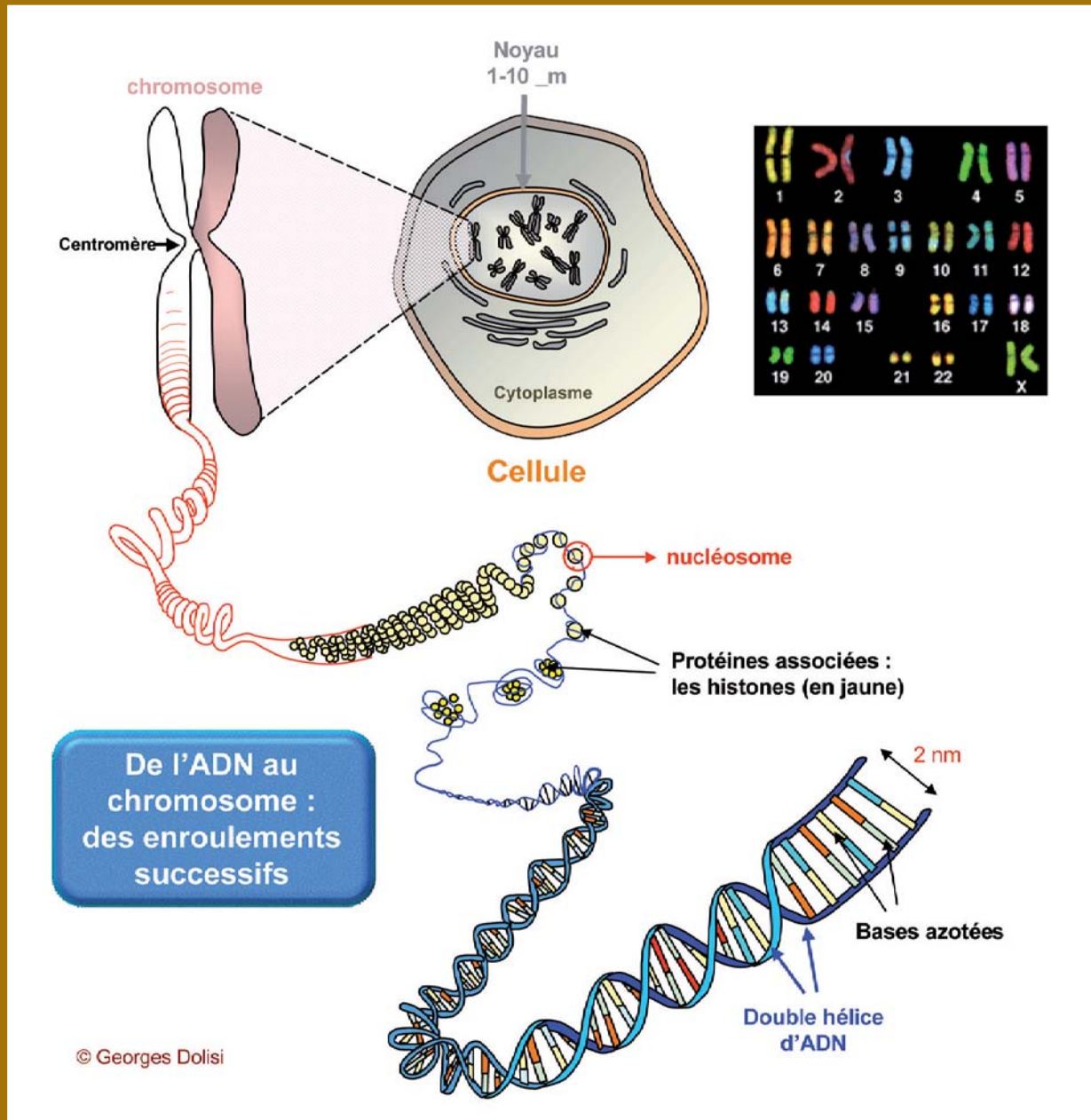


Figure 7

Les différentes échelles du transfert de l'information génétique : dans chacune de nos cellules se trouve un noyau (1 à 10 microns) ; dans chaque noyau se trouve notre génome, c'est-à-dire l'ensemble de nos chromosomes (22 paires de chromosomes et nos chromosomes sexuels) ; chaque chromosome possède deux molécules d'ADN reliées au niveau du centromère. Chaque molécule d'ADN, dont les différents fragments constituent des gènes, est enroulée en nucléosomes (autour de protéines appelées histones).



## LES ARN INTERFÉRENTS : UNE GRANDE INNOVATION DANS LA BIOLOGIE MOLÉCULAIRE

L'approche **antisens** consiste en un ciblage de l'ARN messenger, ce qui revient en fait à moduler l'expression du gène qui l'engendre. Au début des années 2000, Tom Tuschl a montré qu'une famille de molécules appelées **ARN interférents** est capable de reconnaître des fragments précis d'ARN et de s'y fixer en formant des mini-duplex. Cela a pour effet de bloquer l'étape de traduction et donc d'empêcher la production des protéines correspondantes.

Les oligonucléotides antisens et les ARN interférents sont de petits enchaînements de moins de vingt-cinq nucléotides, ce sont donc de petits bouts d'ADN ou d'ARN, respectivement. Les chimistes savent en synthétiser rapidement et en grande quantité (avec des robots synthétiseurs). Ce n'est pas par hasard si cette découverte a été faite en 2001, au même moment que le séquençage du génome. En effet, c'est un outil largement utilisé pour inactiver des gènes et comprendre à quoi ils servent. Par ailleurs, ces molécules ont également un avenir prometteur en tant qu'agents thérapeutiques (anti-infectieux, anti-inflammatoires).

l'ADN ? L'intervention et l'imagination des chimistes se révèlent indispensables dans cette recherche.

## 2 Cibler l'ADN : comment ? Avec quelles molécule ?

L'approche du ciblage de l'ADN consiste à concevoir des molécules capables de reconnaître certaines séquences bien précises d'ADN et d'en modifier la séquence en bases azotées. En résumé, on touche à un seul gène sans toucher aux autres. Une fois touché, ce gène ne pourra plus donner la (ou les) protéine(s) dont il permet la synthèse habituellement. En étudiant les conséquences de l'élimination de cette protéine, il sera ensuite possible de déterminer son rôle, et en particulier de voir si elle est impliquée dans une pathologie (maladie génétique ou développement d'une tumeur).

Contrairement à une autre approche bien connue des

biologistes appelée « **antisens** », qui cible l'ARN (**Encart « Les ARN interférents : une grande innovation dans la biologie moléculaire »**), le ciblage de l'ADN est une approche dite permanente car la modification des séquences de bases qu'il vise a un impact direct et permanent sur la fonction associée.

Afin de cibler l'ADN, les chimistes savent aujourd'hui synthétiser deux types de molécules capables de reconnaître spécifiquement des fragments d'ADN, ce qui permet ensuite d'en modifier la séquence de bases avec précision (**Figure 8**) :

– **méthode 1** : on utilise des oligonucléotides pouvant reconnaître des séquences de paires de bases et se fixant dans le grand sillon pour former des triplex (triples hélices) ;

– **méthode 2** : on utilise des molécules de la famille des pyrroles et imidazoles, qui se fixent dans le petit sillon de l'ADN.

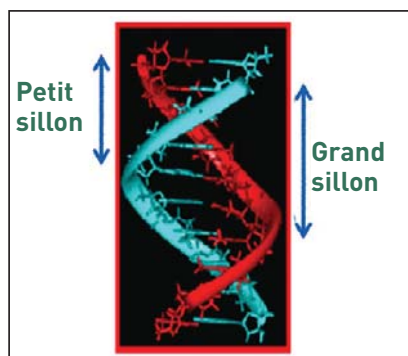
### 2.1. Méthode 1 : des oligonucléotides fonctionnalisés. Vers une application dans la thérapie génique et la transgénèse

#### Étape 1 : reconnaître des fragments précis d'ADN

Les deux brins d'ADN s'associent entre eux grâce à la formation de paires de bases selon quatre possibilités : T-A, A-T, C-G et G-C. Il est possible de les distinguer en utilisant des molécules qui vont interagir de manière différente avec l'une ou l'autre, selon les liaisons qu'elles vont pouvoir établir avec certains sites

Figure 8

L'ADN peut être ciblé par des molécules de synthèse à deux niveaux possibles : au niveau du grand sillon ou au niveau du petit sillon.





situés sur ces paires de bases (en particulier des liaisons hydrogène). Ces sites sont : « donneurs (D) », « accepteurs (A) » ou des méthyles, et leur enchaînement est différent selon la paire de bases, mais aussi selon que l'on se trouve au niveau du grand sillon ou du petit sillon. C'est de cette manière que l'on peut différencier les quatre paires de bases possibles (Figure 9).

En théorie, il est donc possible d'arriver à reconnaître n'importe quel enchaînement de paires de bases sur le génome, et l'enjeu pour les chimistes est alors de trouver la molécule qui y parviendrait ! Une famille de molécules peut accomplir cet exploit : les oligonucléotides.

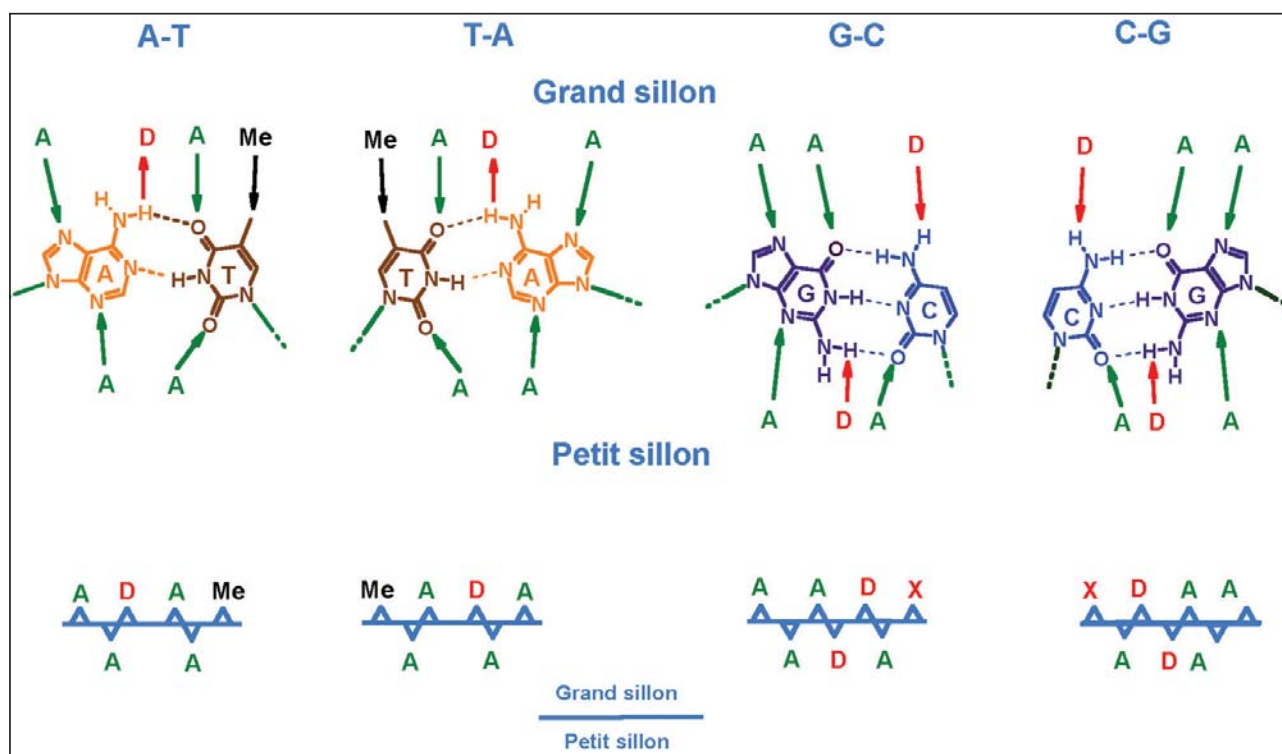
Revenons sur l'historique de cette découverte : en 1957, peu de temps après la découverte de la structure en double hélice de l'ADN, le chercheur Gary Felsenfeld découvre sur une figure de diffraction

de polymères une structure à trois brins d'ADN, qu'il interpréta comme une triple hélice ! Pendant trente ans, cette découverte demeura une curiosité de laboratoire. Il fallut attendre 1987 pour que Peter Dervan et Claude Hélène montrent simultanément qu'il est possible de former ce type de structure à trois brins avec de courts oligonucléotides (Encart « Le miracle de la triple hélice »). Cette démonstration a été le départ de ce qu'on a appelé la stratégie « anti-gène » pour « anti-ADN ». Et en tant que grands pionniers dans ce domaine, Peter Dervan et Claude Hélène ont obtenu, en 1996, le grand prix de la Fondation de la Maison de la Chimie.

Un travail considérable a été réalisé par les chimistes dans la recherche d'oligonucléotides qui soient capables de reconnaître sélectivement des séquences en double

Figure 9

Pour chaque paire de bases possible, et selon que l'on se place au niveau du grand ou du petit sillon de la double hélice d'ADN, on trouve des sites d'interaction différents (accepteur A, donneur D ou méthyle Me), avec des ordres d'enchaînements bien définis. Cela permet de distinguer chacune des quatre bases possibles et au final donne accès à toutes les séquences d'ADN possibles.



## LE MIRACLE DE LA TRIPLE HÉLICE

Comment une triple hélice peut-elle se former ? Dans le grand sillon de l'ADN, il est en fait possible de fixer un troisième brin qui va interagir avec l'un des deux brins de la double hélice. Il va reconnaître une purine (adénine ou guanine) de la double hélice, laquelle est déjà engagée dans une paire, et va former deux liaisons hydrogène avec cette purine. De la même manière, l'adénine d'une paire adénine/thymine est reconnue par une thymine, et la guanine de la paire guanine/cytosine peut être reconnue soit par une cytosine soit par une guanine. On forme donc des triplets de bases et le troisième brin va ainsi s'enrouler autour de la double hélice (*Figure 10*).

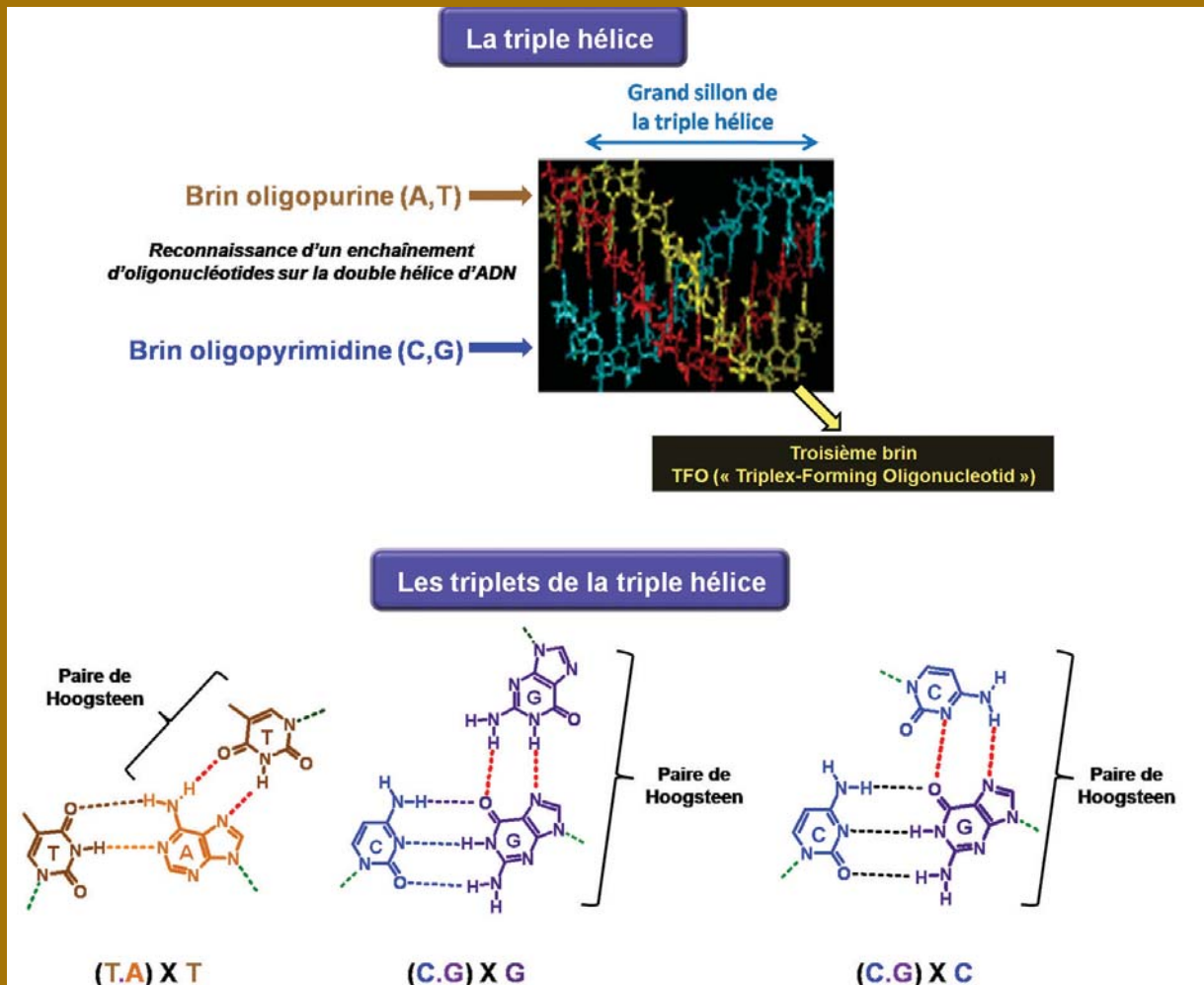
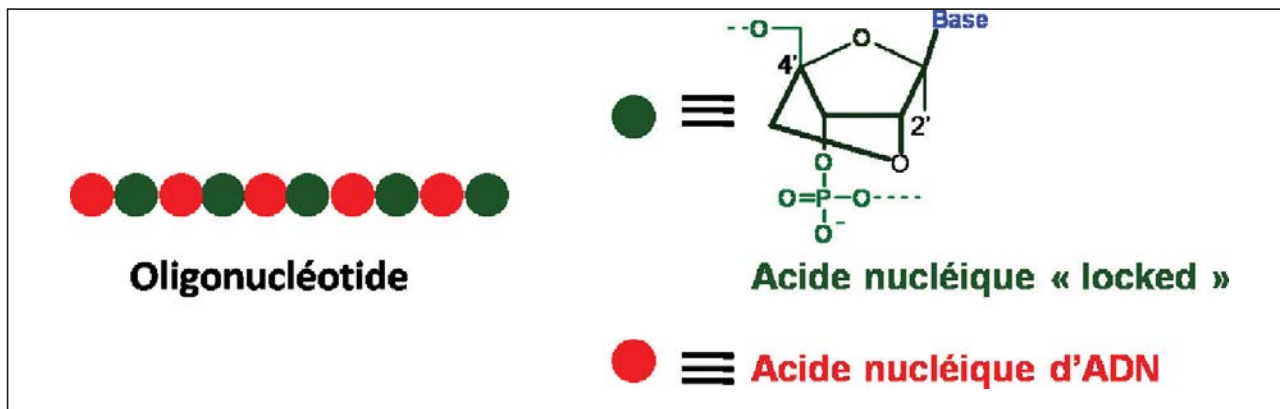


Figure 10

Formation de la triple hélice : le troisième brin (oligopurine ou oligopyrimidine) va reconnaître l'un des deux brins de l'ADN et s'enrouler avec la double hélice.



hélice de l'ADN, en formant une triple hélice stable. Et pour cause, il a fallu apporter plusieurs modifications chimiques aux nucléotides pour que le ciblage d'ADN soit efficace. En effet, un oligonucléotide standard mis dans une cellule sera immédiatement dégradé, car l'environnement cellulaire regorge de nucléases, enzymes chargées de dégrader les oligonucléotides. D'autre part, les oligonucléotides standards ont généralement des affinités moyennes avec l'ADN que l'on veut cibler.

Ce n'est qu'après de nombreuses études de modifications chimiques (squelette, désoxyribose, bases) que les oligonucléotides adéquats ont pu être mis au point pour cette application en milieu cellulaire (Figure 11).

## Étape 2 : modifier la séquence d'ADN

Une fois que l'on s'est positionné sur un fragment bien précis de l'ADN, comment y apporter les modifications

souhaitées sur sa séquence en bases ? Il suffit d'attacher à un court oligonucléotide une molécule capable de couper l'ADN. Lorsque le brin court va se fixer sélectivement sur une région précise de l'ADN pour former un triplex, la coupure de l'ADN s'effectuera à un endroit précis (Figure 12).

On peut aller plus loin avec la même méthode : on peut couper, mais on peut aussi couper puis insérer un fragment contenant une séquence correcte qui peut alors corriger la séquence, comme une sorte de pansement.

On dispose donc de deux stratégies de modification de l'ADN : « couper » ou « couper, coller ».

### 2.1.1. Modifier un gène selon la méthode « couper »

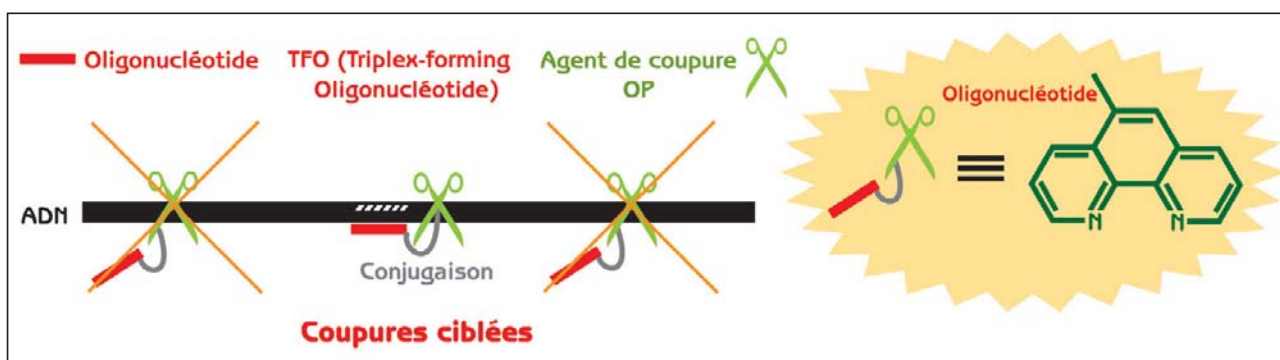
Des expériences ont été menées sur différents types de conjugués [oligonucléotide TFO (Triplex-forming Oligonucléotide) + agent de coupure OP] ayant un nombre de sites de coupure différents. Les

Figure 11

Les oligonucléotides avec des acides nucléiques dits « locked » sont parmi les plus efficaces en cellules : le désoxyribose y est bloqué par un pont méthylène entre l'oxygène en 2' et le carbone en 4'.

Figure 12

Un oligonucléotide (en rouge) conjugué à un agent de coupure reconnaît une séquence précise d'ADN et l'agent de coupure (en vert) coupe la chaîne en un endroit précis. L'agent de coupure peut être l'orthophénanthroline (OP), qui, en présence de métal et de réducteur, est capable de générer des espèces radicalaires et ainsi de casser le squelette phosphodiester [1].



résultats montrent qu'avec ces conjugués, il est possible de cibler des coupures en des sites choisis du génome.

Une fois la coupure effectuée, une cellule contenant des brins d'ADN coupés a naturellement envie de réparer cette coupure. Mais de cette autoréparation il restera des « séquelles » : des mutations vont se produire dans la région de la coupure, c'est-à-dire des changements de paires de bases, et la séquence de l'ADN s'en trouvera modifiée ; par conséquent, son évolution jusqu'à la production d'une protéine sera également modifiée. Cette approche permet ainsi d'éteindre un gène et a été récemment utilisée *in vivo* avec succès pour obtenir des organismes génétiquement modifiés, KO pour un gène d'intérêt [3].

Figure 13

L'approche « couper-coller », ou réparation par recombinaison homologue. Sur un gène muté que l'on veut réparer, on effectue une coupure au niveau du site de mutation, on ajoute un ADN correcteur (le fragment dit « homologue »), qui vient se coller à l'endroit que l'on veut réparer. On récupère alors un gène réparé.

### 2.1.2. Modifier un gène selon la méthode « couper-coller »

Prenons par exemple un gène qui a subi une mutation et qui est responsable d'une pathologie. On souhaite réparer ce gène en coupant la partie mutée et en la remplaçant par

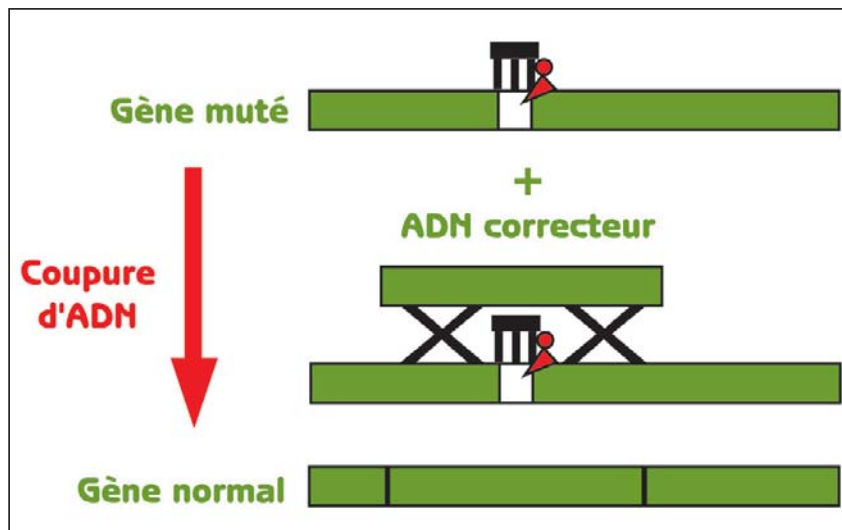
un fragment de gène normal, le « pansement ».

Pour procéder à cette opération de « couper-coller », les scientifiques utilisent une machinerie cellulaire : la « **recombinaison homologue** ». C'est une méthode développée depuis longtemps et encore actuellement.

Si l'on utilise un fragment d'ADN extérieur, homologue à la séquence d'ADN que l'on veut réparer, mais sans les mutations indésirables, il peut s'intégrer dans le génome. Le problème rencontré est le manque d'efficacité de la méthode pour les cellules de mammifères (mais elle s'avère très efficace dans des cellules de levures). Néanmoins, on a trouvé depuis 1994 le moyen d'augmenter l'efficacité de l'intégration de l'ADN, et donc l'efficacité de correction : il s'agit d'effectuer la coupure à proximité du site que l'on souhaite modifier, d'où l'importance de bien cibler les coupures (Figure 13) [2].

Par rapport à la méthode précédente, où l'étape de recollage n'était pas contrôlée, on recolle dans ce cas exactement comme on le désire en intégrant la séquence voulue dans le milieu cellulaire.

Mais on peut aller encore plus loin dans cette approche, en intégrant un gène entier ; et l'on s'est rendu compte qu'il est possible de faire exprimer dans une cellule des **transgènes**, rien qu'en les intégrant avec des fragments homologues... De là à envisager de faire des organismes génétiquement modifiés ?





### Applications du « couper-coller » : des organismes génétiquement modifiés à la thérapie génique

Effectivement, cette méthode de recombinaison homologue a été utilisée dans le cas de cellules souches embryonnaires de souris. Le prix Nobel de médecine 2007 a été décerné à Mario Capecchi, Oliver Smithies et Martin Evans qui ont réalisé l'exploit de générer des souris **transgéniques** grâce à des recombinaisons homologues réalisées dans des cellules souches embryonnaires (**Figure 14**). La population de cellules modifiées est enrichie, pour générer au final l'animal génétiquement modifié, le mutant.

Le problème est que, pour l'instant, on sait faire des cellules souches pour la souris mais pas encore pour beaucoup d'autres organismes, tels que la drosophile ou le poisson zèbre, qui sont pourtant des modèles importants en biologie. On n'est donc pas encore capable de faire facilement des mutants ciblés.

Mais il a été montré récemment qu'avec des ciseaux non synthétiques, des « nucléases à doigts de zinc » [3], il est possible de cibler des coupures en utilisant ce type d'approche. En plein développement, cette méthode permet d'espérer faire de la **transgénèse** dans d'autres organismes.

Par ailleurs, les chercheurs suspectent que la réparation de l'ADN change selon l'endroit où l'on se situe sur le génome. Ici, la chimie a un rôle important à jouer. En effet, il est possible de cibler

des coupures avec une chimie donnée, bien contrôlée, ce que ne font pas les nucléases de nature protéique (qui font des réactions enzymatiques). Cela permettrait de mimer des coupures qui se produisent naturellement et d'en étudier les mécanismes de réparation.

À plus long terme, quand on saura parfaitement corriger un gène, on pourra penser à utiliser cette méthode pour la **thérapie génique**. En effet, avec une **maladie monogénique** (voir le chapitre de **J.-F. Bach, paragraphe 1**) et un gène à corriger, on peut non pas introduire un gène mais simplement utiliser cette approche de modification de séquence.

Beaucoup de questions ne sont pas encore résolues. En effet, les résultats décrits sont obtenus avec une première génération de nucléases, mais la spécificité à l'échelle du génome est encore à préciser car, pour des applications de thérapie génique, il ne faut pas se tromper dans la correction du gène. Il faut aussi réussir à reconnaître de manière fine

**Figure 14**

*Les souris génétiquement modifiées constituent un outil précieux pour la recherche thérapeutique.*



n'importe quelle séquence sur le génome. Il faut enfin savoir contrôler la chimie des coupures. Les chimistes ont donc encore beaucoup de travail en perspective !

### 2.2. Méthode 2 : cibler des gènes pour comprendre le fonctionnement des régions non codantes du génome

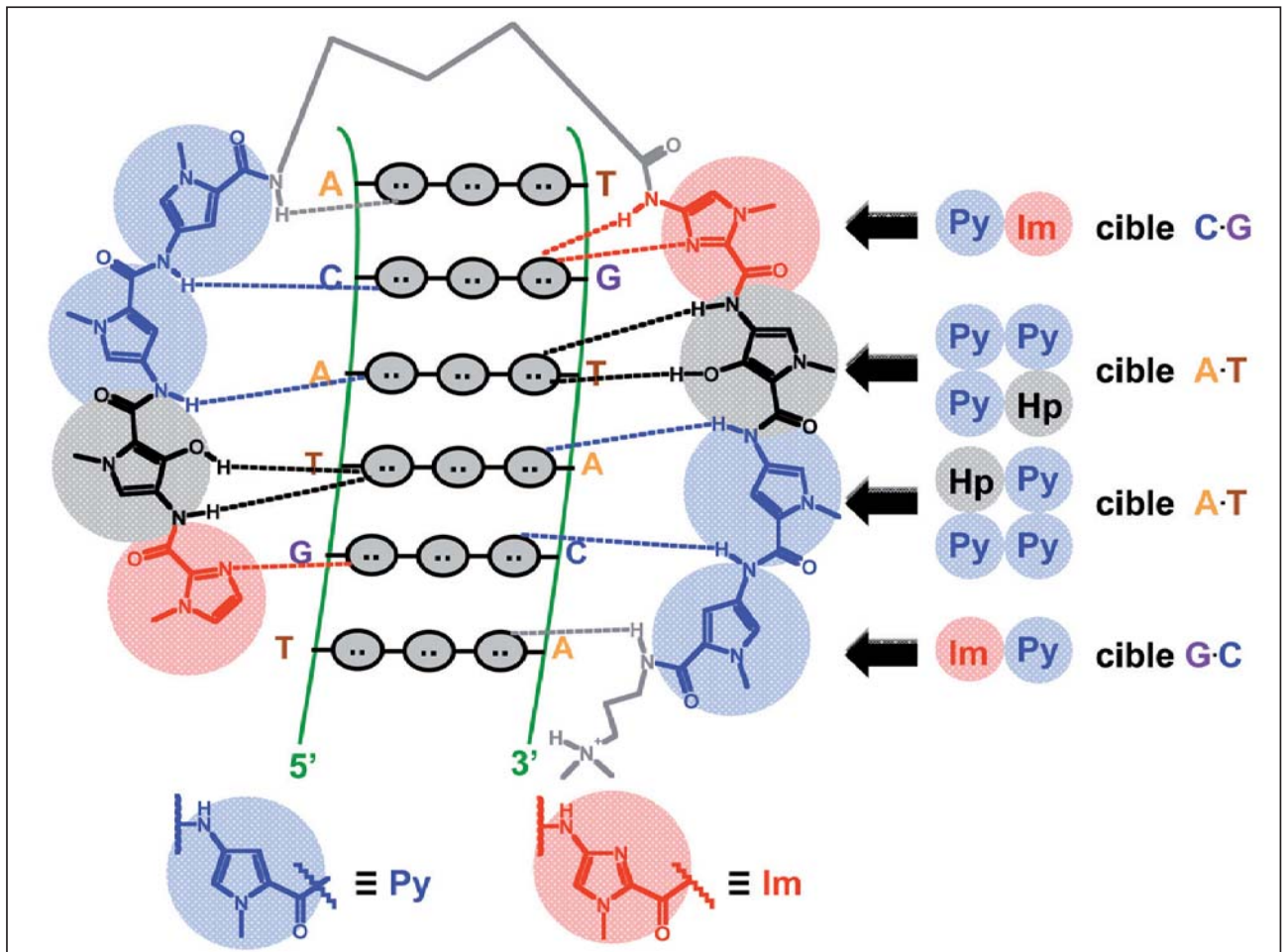
Nous avons vu que seulement 5 % du génome correspondent à des régions codantes, c'est-à-dire qui codent pour la synthèse de protéines nécessaires au fonctionnement cellulaire. Qu'en est-il du reste ? C'est l'objet de toute une recherche qui utilise justement le ciblage de l'ADN.

### À la découverte de régions non codantes du génome et de la régulation de l'expression des gènes

Pour étudier les régions non codantes, les chercheurs ont utilisé un type de molécules composées de plusieurs modules : pyrroles (Py) et imidazoles (Im). Elles se fixent dans le petit sillon de l'ADN et des paires Im-Im et Im-Py vont reconnaître les paires de bases de cet ADN. Le code de reconnaissance a été identifié par le chercheur Peter Dervan : une paire Py-Im reconnaît un couple C-G, une paire Py-Py reconnaît un couple A-T et une paire Im-Py reconnaît un couple G-C. On peut donc théoriquement reconnaître les quatre paires de bases (**Figure 15**).

Figure 15

Reconnaissance des quatre paires de bases.



Ulrich Laemmli à Genève a étudié le rôle de certaines séquences non codantes appelées « séquences satellites » chez la drosophile [4] ; et Emmanuel Kas à Toulouse a essayé d'expliquer [5] les résultats observés, qui sont les suivants : près du centromère du chromosome X de la drosophile, on trouve les régions **satellites** (appelées hétérochromatine ou régions silencieuses), dont le « satellite III » est une région très riche en paires A-T. Il pourra donc être repéré par les paires de molécules Py-Py. À côté de ces séquences satellites, on est surpris de découvrir des gènes codant pour des ARN ribosomiaux, que l'on sait être nécessaires à la prolifération cellulaire, et très exprimés (**Figure 16**).

La question que se sont posée les chercheurs était la suivante : quel est le rôle des régions satellites ?

Pour y répondre, l'expérience suivante a été réalisée : chez la drosophile sauvage, il existe un gène (le gène « *white* ») qui est un transporteur de pigment et qui donne

une couleur rouge à l'œil de la drosophile. Mais chez des mutants naturels, l'œil est blanc. Or on voit que pour ces derniers, ce gène « *white* » est situé dans la région d'hétérochromatine, près du satellite III. Les régions satellites de ces mutants perturberaient-elles ce gène censé donner l'œil rouge ?

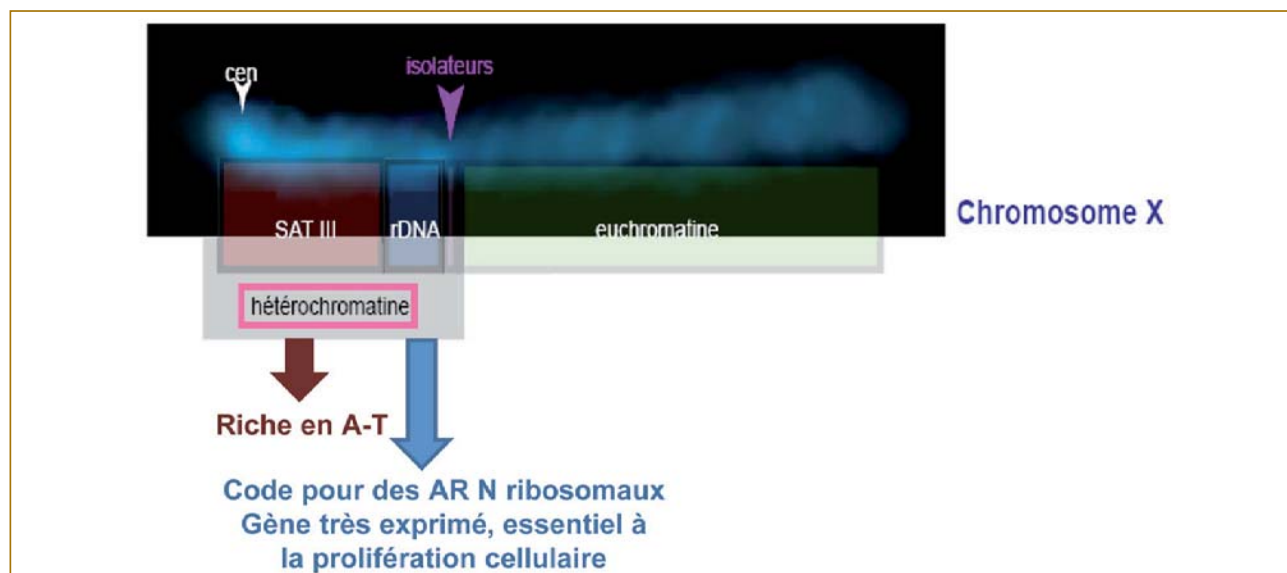
Après avoir traité des larves de drosophiles mutantes (œil blanc) avec un oligopyrrole (P9), on a observé que leur œil devenait rouge. Cette molécule capable de se fixer sur le satellite III a donc une influence sur le gène « *white* », qui se trouve juste à côté. Ce gène n'était donc pas exprimé chez le mutant, mais le devient dès que le P9 s'est fixé sur le satellite III (**Figure 17**).

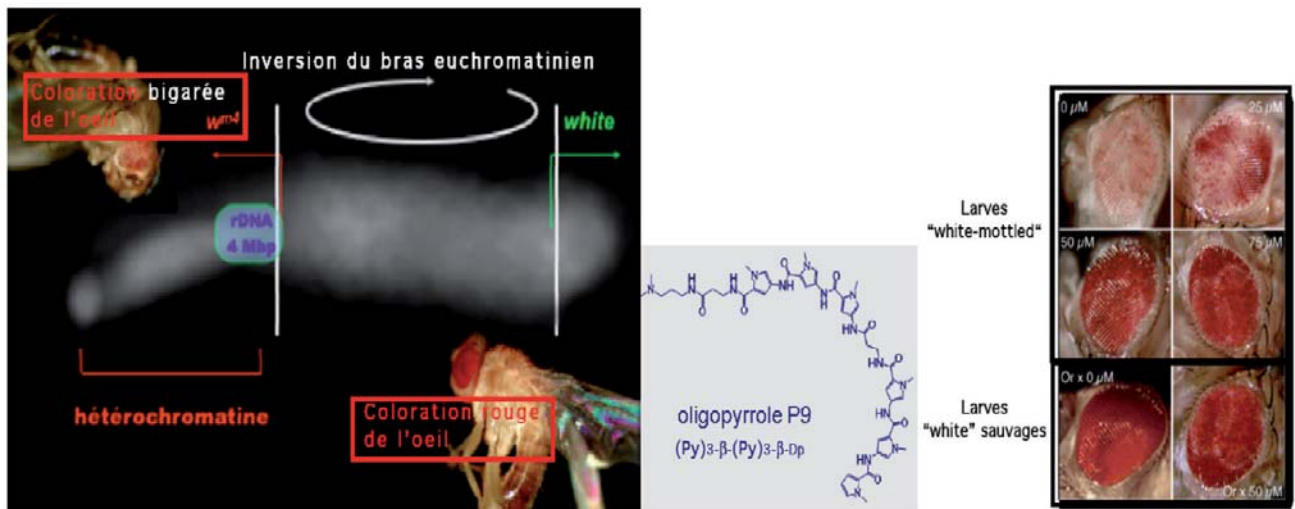
La réponse est donc trouvée : les régions satellites régulent l'expression des gènes avoisinants, lesquels contribuent à donner la couleur rouge aux yeux.

Cet exemple parmi d'autres montre bien que l'on dispose de molécules chimiques capables de repérer les séquences

**Figure 16**

Les séquences satellites dans le modèle drosophile. Emmanuel Kas (Toulouse) et Ulrich Laemmli (Genève).





**Figure 17**

L'oligopyrrole (P9) a été injecté chez la lignée mutante « white-mottled » de la drosophile. Étant capable de reconnaître des séquences A-T, il se fixe sélectivement sur le satellite III de l'hétérochromatine, ce qui a pour effet de lever l'inhibition du gène « white », et ce qui redonne à l'œil sa couleur rouge.

non codantes de l'ADN pour pouvoir les étudier, et avancer dans l'élucidation de leur rôle à l'échelle du génome.

On découvre ici une illustration du rôle essentiel joué par des régions non codantes du génome, ou régions satellites : réguler l'expression des gènes.

### Vers des ciblage plus spécifiques pour la médecine du futur ?

La chimie peut donc concevoir et synthétiser des molécules capables de reconnaître une séquence choisie d'ADN. Néanmoins, il reste encore du travail pour parvenir à reconnaître n'importe quelle séquence d'ADN et réussir à caractériser de manière très précise le niveau de spécificité de ces molécules. En effet, seule une attaque totalement spécifique du gène ciblé fera la puissance de la méthode et de ses applications, surtout si celles-ci sont d'ordre thérapeutique.

Les nombreux exemples développés dans ce chapitre ont néanmoins permis de montrer la richesse potentielle des applications, notamment en biologie mais aussi pour des applications thérapeutiques qui seront à développer dans le futur.



## Bibliographie

**[1]** Simon P., Cannata F., Perrouault L., Halby L., Concordet J.-P., Boutorine A., Ryabinin V., Sinyakov A., Giovannangeli C. (2008). Sequence-specific targeted DNA cleavage mediated by bipyridine polyamide conjugates. *Nucleic Acids Res.*, **36**: 3531-3538.

**[2]** Pingoud A., Silva G.H. (2007). Precision genome surgery. *Nat Biotechnol.*, **25**: 743-744.

**[3]** Doyon Y., McCammon J.M., Miller J.C., Faraji F., Ngo C., Katibah G.E., Amora R., Hocking T.D., Zhang L., Rebar E.J. *et al.* (2008). Heritable targeted gene

disruption in zebrafish using designed zinc-finger nucleases. *Nat. Biotechnol.*, **26**: 702-708.

**[4]** Janssen S., Cuvier O., Müller M., Laemmli U.K. (2000). Specific gain- and loss-of-function phenotypes induced by satellite-specific DNA-binding drugs fed to *Drosophila melanogaster*. *Mol. Cell.*, **6**: 1013-1024.

**[5]** Blattes R., Monod C., Susbielle G., Cuvier O., Wu J.H., Hsieh T.S., Laemmli U.K., Käs E. (2006). Displacement of D1, HP1 and topoisomerase II from satellite heterochromatin by a specific polyamide. *EMBO J.*, **25**: 2397-2408.

# Crédits photographiques

- Fig. 3 et 5 (gauche) : CNRS Photothèque/Université de Montpellier 2, UPR 9080  
- Laboratoire de Biochimie Théorique - Paris.
- Fig. 4 : CNRS Photothèque/Lamoureux Richard, GIP-GENOSCOPE - Centre national de séquençage.
- Fig. 7 : Georges Dolisi.
- Fig. 14 : CNRS Photothèque/Raguet Hubert, UMR 8612  
- Physico-Chimie, Pharmacotechnie, Biopharmacie  
- Châtenay-Malabry.
- Fig. 16 et 17 : E. Kas.