

## 3. Architecture et structure de l'ADN

*Les 98,7% d'ADN "non-codants" ont un rôle fonctionnel*

*Si tous ces ARN qui ne codent pas pour des protéines sont fonctionnels, comme les faits le suggèrent de façon croissante, alors une grande partie et peut-être la presque totalité du génome humain est fonctionnel. Si oui, la programmation génétique des organismes supérieurs a été mal comprise pendant les 50 dernières années, à cause de la supposition [...] que l'information génétique est transmise par les protéines. (J. Mattick, 2007)*

### Alain Boudet

Dr en Sciences Physiques  
www.spirit-science.fr

2012

#### Autres articles de la série :

1. La molécule d'ADN et le code génétique
2. L'ADN et ses modes d'expression
4. La musique de l'ADN et des protéines
5. L'ADN électrique
6. L'ADN électromagnétique

**Résumé:** Les parties codantes des gènes de l'ADN, qui détiennent les codes de fabrication des protéines qui régulent notre corps, n'occupent que 1,3% environ de la totalité de l'ADN. Les zones non-codantes dans et entre les gènes intriguent les scientifiques par leur présence énigmatique. Ayant abandonné l'idée que ces zones sont inutiles, ils commencent à mettre en évidence leurs fonctions possibles. Ces zones ont des rôles de régulation et de contrôle. Les zones intergènes comportent des séquences caractéristiques pour chaque individu, au point qu'elles ont été retenues par la législation comme base de l'empreinte génétique. Il existe donc une architecture significative dans l'ADN. Par des méthodes d'analyse statistique, des mathématiciens ont mis en évidence **un ordre fractal qui varie selon le type d'ADN**.

#### Contenu de l'article

- Les zones qui portent le code génétique et les autres
- Les introns: des zones sans code génétique incluses dans les gènes
- A quoi servent les introns?
- A la fois intron et exon - l'épissage alternatif
- Des séquences répétées dans l'ADN non-codant
- Variabilité individuelle des satellites
- Les séquences non-codantes plus nombreuses dans les organismes supérieurs
- Le réseau de régulation des gènes
- Des séquences mobiles, les transposons

- La communication par les ARN non-codants
- L'information itérative
- Un ordre global dans l'ADN
- L'ordre fractal
- Le code génétique électromagnétique

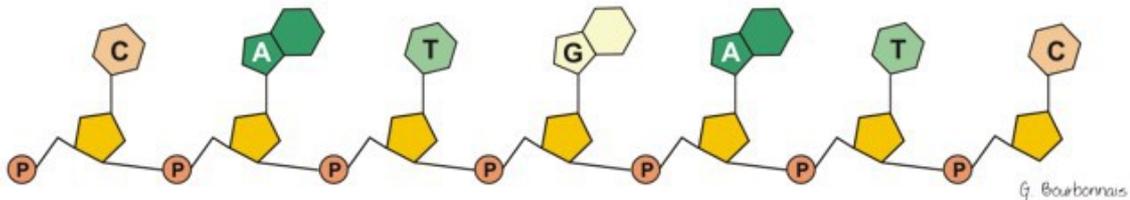
#### Annexes

Empreinte génétique et test ADN  
En savoir plus

Dans la première partie de cette étude sur l'ADN, nous avons fait connaissance avec la molécule d'ADN au sein des noyaux des cellules, et nous avons découvert que certains fragments de cette molécule appelés gènes portent des instructions pour le fonctionnement et le développement physique du corps. L'ensemble de ces instructions est appelé le code génétique. Il contrôle la fabrication des protéines, grosses molécules qui sont les ouvrières du fonctionnement et de la croissance du corps. Mais nous avons remarqué que les gènes humains n'occupent que 10% de la molécule d'ADN. Pourquoi alors **la présence des autres 90%? Quelle est leur fonction?** Dans cette troisième partie, nous allons mettre en évidence que la chaîne est structurée par une sorte d'architecture, aussi bien dans les gènes qu'en-dehors d'eux et que cette architecture est le signe d'une fonctionnalité.

Comme indiqué dans la première partie de cette étude (*La molécule d'ADN et le code génétique*), nos cellules renferment dans leur noyau 23 paires de **chromosomes**, qui sont chacun constitués d'un filament d'ADN, décoré d'autres éléments tels que les grosses protéines appelées histones. Un **filament** ou **molécule** d'ADN est formé de l'enroulement en hélice de 2 **brins** qui se font face. La totalité des filaments d'ADN de l'ensemble des chromosomes s'appelle le **génome**.

Chaque **brin d'ADN** est constitué de l'enchaînement en grand nombre de nucléotides. Un **nucléotide** désigne l'assemblage d'une base azotée, du "pentagone" sur lequel elle est fixée (voir figure 1) qui représente le *désoxyribose*, et du groupement *phosphate* de la chaîne centrale. Les bases azotées sont prises parmi 4 types qu'on représente par les lettres **C** (cytosine), **A** (adénine), **T** (thymine), **G** (guanine). Les deux brins du filament sont associés chimiquement par les bases face à face. Celles-ci se révèlent être complémentaires l'une de l'autre, car la base C s'associe à G et la base A s'associe à T.



*Fig.1- Schéma d'un brin d'ADN. Il est composé d'une chaîne centrale, faite de l'assemblage linéaire de motifs identiques (les pentagones oranges représentant le désoxyribose, articulés par le groupement phosphate P), sur laquelle sont fixés des groupements C, A, T et G  
Merci à G. Bourbonnais*

Si l'on considère les 2 brins assemblés, la molécule d'ADN est constituée d'un enchaînement de **paires de bases**. Si donc on décrit l'enchaînement des bases de l'un des brins, il est inutile de préciser l'autre, car il s'en déduit par complémentarité. On a donc coutume d'évaluer la longueur d'un segment d'ADN en nombre de paires de bases (ou bp).

## Les zones qui portent le code génétique et les autres

Lorsque nous observons la molécule d'ADN et l'enchaînement des nucléotides, aucune structure particulière ne se distingue. Par exemple, on ne constate aucune discontinuité dans la succession des nucléotides, qui pourrait mettre en évidence le début d'un gène. Celui-ci est seulement signalé par un code qu'il est nécessaire de connaître pour pouvoir le repérer. Les bases se suivent dans un ordre qui peut sembler aléatoire, mais un comptage plus fin employant des méthodes statistiques se révélera instructeur.

### Gènes

C'est seulement par leur rôle génétique que les gènes se distinguent dans la molécule, non par leur aspect. En effet, un **gène** est une portion du filament d'ADN qui commande la production d'une **protéine** déterminée, protéine qui joue un rôle indispensable au fonctionnement et à la croissance des cellules. On dit que le gène *code pour la protéine*.

**Le code d'un gène pour la fabrication d'une protéine est donné par l'ordre de succession de ses nucléotides.** Sachant qu'une protéine est faite de l'enchaînement d'acides aminés, le code commande la fabrication successive de chacun des acides aminés. Le code pour un acide aminé est fait d'un groupe de 3 nucléotides successifs (un **codon**).

### Zones intergènes

Si l'on comptabilise la totalité des gènes reconnus dans les chromosomes, ils n'occupent qu'une faible partie de l'ADN,

environ 10% du génome chez l'homme (l'estimation varie selon les chercheurs). Entre les gènes, se trouvent de longues **zones non-codantes**, ce qui signifie qu'on ne leur a pas trouvé de fonction de fabrication d'une protéine. On les appelle **zones intergènes** ou encore **séquences espaceurs**. Elles représentent la plus grande partie du génome (environ 90 % chez l'homme).

## Introns

Mais ce n'est pas tout. A l'intérieur d'un gène, il existe aussi des zones non-codantes. Si l'on repère les suites de codons, il y a ceux qui codent pour les acides aminés, éléments de la protéine. Il y en a aussi quelques-uns qui codent pour des contrôles: par exemple, un codon en tête pour repérer le début de fabrication et un autre en fin pour en marquer l'arrêt. Il y en a quelques autres. Mais tous ces codons n'occupent qu'une faible partie du gène.

Insérés entre les codons, d'autres fragments d'ADN ne codent pas pour des acides aminés. On les nomme des **introns**.

## Zones non-codantes

Nous avons repéré des régions non-codantes à l'intérieur des gènes (les introns) et de bien plus grandes entre les gènes (les espaceurs). Au total, elles représentent **98,7% de régions non-codantes**. Il reste donc seulement une faible proportion de régions codantes dans un filament d'ADN, 1,3% environ. C'est très intrigant. Pourquoi la nature a-t-elle conçu ces régions non-codantes?

D'ailleurs sont-elles réellement non-codantes? Même si elles ne codent pas la production d'acides aminés, il est possible qu'elles portent des codes qui contrôlent d'autres fonctions. Le mot **code** a un sens très général et très profond, beaucoup plus étendu que le simple fait de *coder pour une protéine* (voir [article Code, information et mémoire](#)). Lorsque nous employons un mot, il est important de sentir qu'il peut enfermer notre pensée dans une habitude.

Depuis quelques années, les scientifiques entrevoient des **fonctions possibles des zones "non-codantes"**. Les termes de *ADN déchet*, *ADN poubelle*, ou *junk DNA*, encore employés il y a quelques années, ont été abandonnés par la communauté scientifique au profit de *séquences non-codantes pour des protéines*. Puis-je proposer le terme *non-protéinocodantes* ou *npc*? Prenons connaissance de ces nouvelles pistes de réflexion.

## Les introns: des zones sans code génétique incluses dans les gènes

Les séquences de nucléotides dans les gènes ont tout d'abord été observées et analysées sur des bactéries, organismes composés d'une seule cellule, sans noyau (**procaryotes** - ce qui signifie en gros *pas encore de noyau, mais ça viendra*). Ce sont ces observations qui ont conduit à démontrer que chaque groupe de 3, le codon, contrôle la fabrication d'un acide aminé, et que l'enchaînement des acides aminés de la protéine est calqué sur l'enchaînement des codons du gène.

Or lorsqu'on s'est penché sur des organismes plus élaborés tel que l'être humain, dont les cellules comportent un noyau (organismes appelés **eucaryotes** - ce qui signifie *avec un vrai noyau*), il est apparu que seules certaines parties du gène sont des séquences codantes. **Le gène est constitué d'une suite alternée de séquences codantes (les exons) et de séquences non codantes (les introns)**.

### Élimination des introns dans la molécule ARN messagère

Comment sait-on que les introns ne sont pas impliqués dans le codage des acides aminés? En suivant le procédé par lequel une protéine est élaborée à partir du gène.

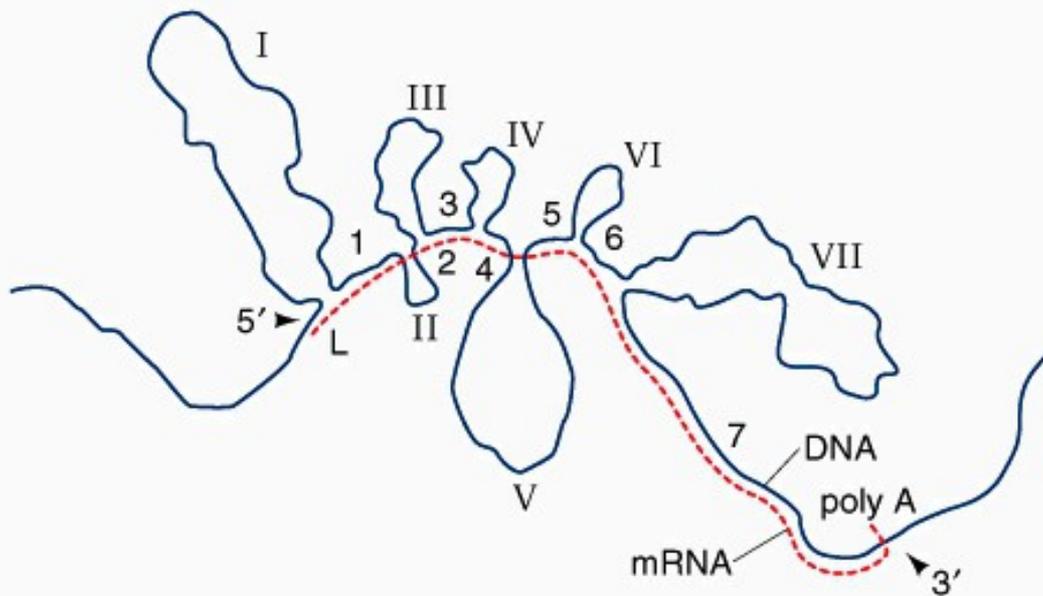
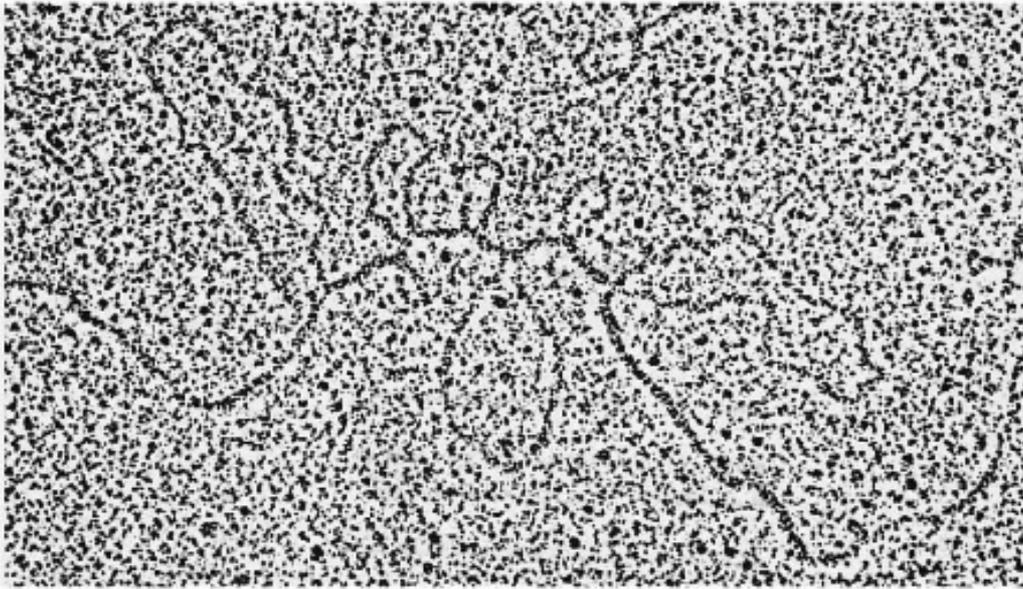
Tout d'abord, **le gène est transcrit en une copie moléculaire**, à la façon d'une empreinte complémentaire. La molécule de cette empreinte est de l'ARN (**Acide RiboNucléique**). Elle diffère de l'ADN, d'une part par une légère différence chimique dans le pentagone de la chaîne centrale (fig.1) qui est ici un ribose au lieu d'un désoxyribose (d'où le **R** d'ARN au lieu du **D** d'ADN), et d'autre part par le remplacement de la thymine (**T**) par l'uracyle (**U**).

Observer cette empreinte en ARN est un moyen de repérer le gène, ses extrémités et sa longueur. L'empreinte ARN (appelée **pré-ARN**) est élaborée à partir du gène par complémentarité des bases, elle est moulée dessus. Elle a donc exactement la longueur du gène. La base **C** du gène est reproduite en **G** et la base **A** est reproduite en **U**, et ainsi de suite. Ainsi le codon CGA de l'ADN qui code pour l'alanine est reproduit en GCU dans l'ARN. Dans une deuxième étape, le pré-ARN est remanié. **Des fragments d'ARN sont découpés et éliminés** (excision), puis les extrémités des fragments restants sont ressoudées. Ce mécanisme est nommé **épissage**.

L'ARN restant qui en résulte est envoyé en mission hors du noyau, dans l'usine de production de la cellule (les **ribosomes**) où il est décodé en acides aminés pour produire la protéine. On le nomme **ARN messenger**.

C'est le modèle classique en vigueur depuis les années 1950, dans lequel les instructions circulent dans un seul sens, de l'ADN vers la protéine. Ce modèle a été nommé **dogme central de la génétique moléculaire** par Francis Crick

lui-même, le découvreur de la structure en double hélice, dans un article publié en 1958 par le journal *Nature*.



From Chambon, P., *Sci. Am.* 244(5), 61 (1981).  
Copyright 1999 John Wiley and Sons, Inc. All rights reserved.

*Fig.2. En haut, le gène d'ovalbumine (protéine du blanc d'uf) vu en microscopie électronique. En bas, sa représentation schématique: En bleu, l'ADN du gène tel qu'il est vu en microscopie électronique. En rouge, l'ARN messager (mRNA) après excision des introns. Les boucles I à VII sont des introns. Les fragments L, et 1 à 7 sont les exons. L est le début de lecture du gène.*

Les parties du pré-ARN éliminées au cours de l'épissage ne servent donc pas au codage de la protéine. Par définition, **les introns sont les parties du gène qui sont éliminées lors de la formation de l'ARN messager** (nommés ainsi par le biologiste américain *Walter Gilbert* en 1978).

En définitive, bien qu'il forme un ensemble compact avec deux extrémités, **un gène est morcelé**: des séquences qui sont traduites en protéines sont incorporées dans une matrice de régions silencieuses.

### **Nombres et longueurs des introns**

En examinant des organismes variés, des plus simples comme les bactéries, jusqu'aux plus élaborés comme l'être humain, les chercheurs se sont aperçus que les mêmes protéines avaient des rôles semblables d'un organisme à l'autre. De plus une même protéine est codée par un même ARN messager, qui est issu de gènes analogues dans l'ADN de chacun des organismes.

Mais alors que les gènes analogues comportent les mêmes exons, leur teneur en introns est différente. **Plus les**

**organismes sont élaborés et plus leurs gènes contiennent d'introns.** Par exemple, un gène analogue peut contenir 0 intron dans une levure, et 5 à 50 chez l'homme.

Prenons l'exemple du gène de l'*ovalbumine* chez la poule. Il contient 8 exons, séparés par 7 introns (voir fig.2). La longueur totale du gène est d'environ 7700 paires de bases, réparties en seulement 1870 environ dans les exons et 5830, soit **75%, pour les introns. La longueur totale des introns est donc environ 3 fois celle des exons.**

Les introns constituent souvent une partie importante du gène. Ce ne sont pas seulement quelques petits "blancs".

La **longueur** d'un intron est typiquement de 65 à 20'000 nucléotides. Elle descend à 31 dans un des gènes du virus SV40. Le gène humain de la dystrophine qui a une longueur de 2,5 millions de nucléotides, est composé de seulement 11000 nucléotides environ pour les exons, et de plus de 75 introns dont le plus long a 210'000 nucléotides. Cela fait **0,5% d'exons et 99,5% d'introns.**

**Les gènes humains sont composés en majorité d'introns.** Au total les zones non-codantes de l'être humain couvrent environ 90% de l'ADN en zones intergènes, + environ 90% des gènes, le total étant estimé à **98,7%** de zones non-codantes.

**Sources:** Pierre Chambon, les gènes en mosaïque. Des gènes aux protéines; dossier Pour la Science, éd. E. Belin, 1987.  
Jerry Bergman, The Functions of Introns: From Junk DNA to Designed DNA

## A quoi servent les introns?

Puisque les introns sont éliminés au cours du processus d'épissage du pré-ARN, pourquoi existent-ils? Ont-ils une fonction?

Certains chercheurs, n'entrevoiant aucun rôle apparent aux introns, en sont restés à cette apparence. Ils ont émis l'idée que ces introns ne servent à rien. En 1972, S. Ohno a qualifié les zones non codantes d'**ADN rebut ou déchet** (*So Much ? Junk DNA in our Genome*). Pour justifier leur présence, ils ont imaginé que c'étaient des résidus d'ADN de nos ancêtres qui se sont incorporés dans le génome à une lointaine époque. Ils y vivent par eux-mêmes, pour eux-mêmes, comme un parasite inoffensif. Ces chercheurs l'ont donc qualifié d'**ADN égoïste** (suivant *Richard Dawkins*). Mais si c'était le cas, répondent d'autres, les exons qui sont séparés par ces introns colonisateurs du gène devraient correspondre à des sous-structures de la protéine codée par ce gène. Or c'est rarement le cas.

Quand les introns ont été découverts il y a 30 ans, on a immédiatement et universellement supposé que ce vaste amoncellement de séquences non-codantes pour des protéines insérées dans les gènes était non-fonctionnel, en dépit du fait qu'elles sont transcrites. Leur présence a été rationalisée en tant que restes d'une évolution ancienne des gènes. En même temps, la découverte qu'une part importante du génome des mammifères (45% chez les humains) dérive des transposons qui sont censés être principalement des nomades parasites, a conduit au concept d'ADN égoïste. Cela a renforcé la vue conventionnelle dominante que les génomes des eucaryotes complexes est largement constitué de débris évolutifs accumulés. (J. Mattick, 2007)

Pourtant, il est bien connu que **certaines séquences d'un gène ont une fonction autre que celle de coder pour un acide aminé.** Par exemple, la **séquence de tête d'un gène** est faite de quelques dizaines de nucléotides décomposés en deux codes: une zone activatrice ou inhibitrice de la transcription en ARN, et un promoteur qui indique le début de la transcription. On dit que c'est le début du **cadre de lecture**. Quant à la **séquence de fin**, c'est un codon (3 nucléotides) qui dit "stoppez la lecture". Ce sont des exemples qui montrent que des séquences peuvent ne pas coder pour un acide aminé et avoir cependant des rôles d'instructeurs.

Des zones de régulation ont été mises en évidence dès les années 1960 par les biologistes français *François Jacob* (né en 1920) et *Jacques Monod* (1910 - 1976). Ils ont montré que certains gènes des bactéries, appelés **opérons**, ne codent pas des protéines, mais déclenchent l'activation de la fabrication de l'enzyme de digestion du lactose normalement inactif, lorsque l'environnement nutritif est anormal (*voir ADN et modes d'expression*). Le modèle de régulation par l'opéron démontre clairement qu'il existe une information rétroactive vers l'ADN, là où se trouve le code d'activation et d'inhibition.

Depuis quelques années, et surtout depuis que le programme **Génome humain** a livré ses conclusions en 2004, les chercheurs découvrent, non seulement de nouvelles fonctions des introns, mais aussi la part que jouent les ARN dans leur fonctionnement. Ils ont été amenés à repenser leur vision des gènes et des introns.

*Ne pas avoir reconnu l'importance des introns pourrait bien se révéler comme la plus grande erreur de l'histoire de la biologie moléculaire.* (J. Mattick dans *The Gems of "Junk" DNA*, 2003)

Examinons cela.

**Sources:** J. Mattick, *The Human Genome: RNA Machine* - The Scientist, 2007, 21, 10, 61

A. Stoltzfus, D. Spencer, M. Zuker, J. Logsdon, Jr. W. Doolittle, Testing the Exon Theory of Genes: The Evidence from Protein Structure, *Science*, 1994, 265, 202

Jacob F, Perrin D, Sánchez C, Monod J, Edelman S. The operon: a group of genes with expression coordinated by an operator. *C.R.Acad. Sci. Paris*, 1960, 250



Il existe aussi un certain nombre de **répétitions géantes**, dont le nombre de motifs peut être très grand. La taille de ces séquences peut aller de quelques dizaines à quelques centaines de milliers de nucléotides.

En-dehors des satellites, les chercheurs ont également répertorié chez les vertébrés des **séquences dispersées** répétées à plusieurs endroits différents sur les chromosomes: nommées petits éléments nucléaires intercalés ou **SINE** et longs éléments nucléaires intercalés ou **LINE**.

*Remarque: Pour la suite de l'article, il est inutile de retenir ces distinctions. Je les ai données ici brièvement pour ceux qui désirent approfondir et lire des articles scientifiques. Ils vont se trouver en face de textes qui emploient ces mots sans les définir en supposant qu'ils sont connus du lecteur. Les articles scientifiques sont écrits pour la communauté scientifique spécialisée. Il est difficile de trouver des articles pour les non-spécialistes, sauf... modestement dans ce site.*

## Variabilité individuelle des satellites

Nous savons que la molécule d'ADN est constituée de l'enchaînement de nucléotides à la manière d'une suite de lettres prises parmi 4 (A, T, C, G). Mais nous savons aussi (voir *L'ADN et le code génétique*) que cette suite est structurée en une sorte d'**architecture** comportant des emplacements ou cadres ou, en terme de biologie, un **locus**. C'est par exemple l'emplacement d'un gène donné. C'est une partie du chromosome qui est repérable de la même façon sur les chromosomes de tous les individus de la même espèce, disons l'être humain.

Ce cadre est rempli de nucléotides qui peuvent être identiques chez tous les individus, ou différents en partie. Chacune des versions possibles du locus s'appelle un **allèle**. Les allèles qui sont identiques chez tous les êtres humains portent les caractères propres à l'espèce humaine. Les allèles variables portent les caractéristiques propres à un individu, comme la couleur de ses cheveux.

Les zones non-codantes sont elles-aussi structurées en loci, avec leur début et leur fin. On peut repérer un même satellite chez des individus différents par leur emplacement ou locus. On constate que **les satellites d'un même locus sont très variables d'un individu à l'autre. Ce qui varie, c'est le nombre de répétitions de son motif.**

De telles modifications se produisent pendant la formation des chromosomes au cours de la division des cellules, provoquant le raccourcissement ou l'allongement d'un satellite. Cela affecte également les cellules sexuelles, entraînant

## Empreinte génétique et test ADN

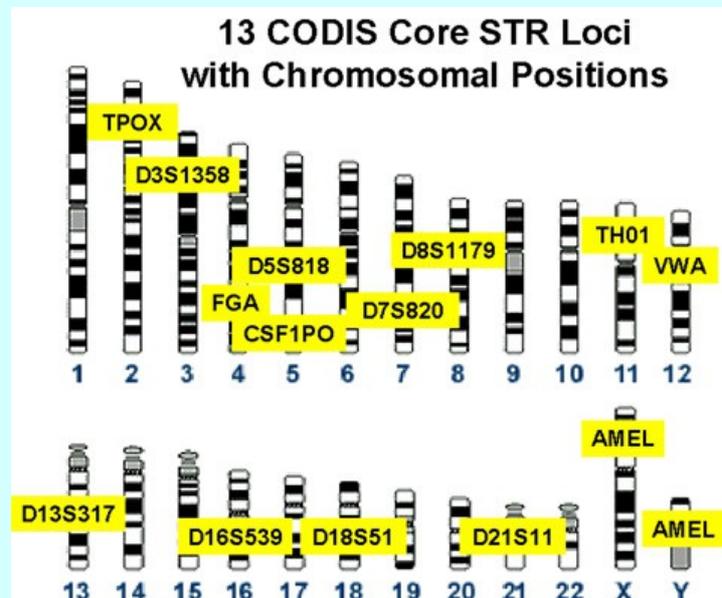
Notre génome comporte 3,2 milliards de paires de nucléotides dont 3,2 millions diffèrent de celles des autres, c'est-à-dire 0,1 %. Ce sont ces différences dans notre génome qui constituent notre **empreinte génétique**.

Lorsque la police scientifique cherche à identifier un individu à partir de son ADN, les laboratoires d'analyse ne déterminent pas l'ensemble des parties variables de son génome, car ce serait bien trop long et trop coûteux. Pour simplifier la tâche, on ne retient que quelques **satellites**, ceux qui peuvent être aisément amplifiés pour l'analyse, qui sont suffisamment simples et qui présentent une grande variabilité d'un individu à l'autre.

Les **minisatellites** présentent cette grande variabilité et sont les meilleurs candidats pour caractériser un individu. Les **microsatellites** présentent moins de variations, ont donc plus de chance d'être semblables chez 2 individus. Mais ils sont plus aisément amplifiables par les techniques routinières de laboratoire et sont donc plus souvent employés dans les tests ADN.

Dans une parenté, la probabilité des microsatellites (ou STR) d'être identiques chez 2 individus n'est pas nulle. On a plus de chance de faire la distinction entre deux individus si on examine plusieurs loci. C'est pourquoi la législation a retenu un nombre minimum de microsatellites (10 ou plus), situés à des loci déterminés et normalisés, dont le choix est variable selon les normes des pays.

En France, les empreintes génétiques sont rassemblées dans le **Fichier national automatisé des empreintes génétiques** (FNAEG), destiné à l'origine à recueillir les empreintes des personnes condamnées pour infractions à caractère sexuel, mais dont **l'usage s'est rapidement étendu à toutes sortes de délit**. En Amérique du Nord (USA, Canada), la norme CODIS est la plus utilisée.



*Localisation des 13 microsatellites dans les 22+2 chromosomes humains, qui constituent l'empreinte génétique selon la législation Codis  
Merci à Wikipedia - Source: Short Tandem Repeat DNA Internet DataBase*

des changements de longueur de l'allèle d'une génération à l'autre. De ce fait, **il est possible de définir une sorte de profil génétique d'un individu par la carte des nucléotides de ses satellites** (voir encadré).

**Sources:** Guillaume Achaz, Etude de la dynamique des génomes; les répétitions intrachromosomiques.

Les empreintes génétiques en pratique judiciaire, Christian Doutremepuich, Bull. Acad. Natle Méd., 2012, 196, 6, 1117-1130

Matt Ridley, Chromosome 8, l'intérêt personnel

## Les séquences non-codantes plus nombreuses dans les organismes supérieurs

Depuis longtemps, on pressent que les zones non-codantes pour des protéines ont un rôle indispensable.

Dans son ouvrage *La Statue intérieure* (1987), F. Jacob se rappelle *l'un des plus vieux problèmes en biologie: dans des organismes faits de millions, voire de milliards de cellules, chacune des cellules possède l'ensemble complet de gènes. Comment se fait-il alors que tous les gènes ne fonctionnent pas de la même façon dans tous les tissus?* Pourquoi **les cellules d'un organisme, qui détiennent exactement le même génome, se développent-elles pourtant de façon différenciée selon leur fonction:** cellules nerveuses, cellules du foie, du muscle, etc?

Une autre question est devenue cruciale depuis le séquençage du génome humain. Selon les biologistes, la croissance et le fonctionnement du corps sont régis par un nombre considérable de protéines (dont de nombreux enzymes, qui sont des protéines), elles-mêmes contrôlées par les gènes. On s'attendait donc à découvrir un grand nombre de gènes chez l'humain. Or il est apparu qu'il en a beaucoup moins que ce qu'on supposait. **Seulement 25000 gènes couvrant 1,3% de l'ADN codent pour des protéines. C'est insuffisant pour expliquer la complexité de l'organisme** avec ses nombreuses fonctions.

Par comparaison, le ver *Caenorhabditis elegans* le plus simple des organismes évolués (eucaryotes) possède 19300 gènes environ, soit presque autant que les humains. Mais ils couvrent 24% de l'ADN, le maximum pour les eucaryotes, car les zones non-codantes y sont bien moins abondantes.

Chez les procaryotes, la quantité de zones codantes monte à 88% (bactérie E. Coli). Les zones non-codantes y sont rares et le génome est court.

Nous savons maintenant que les séquences codantes couvrent 80% ou plus du génome des procaryotes, mais par exemple moins de 2% chez les humains. Une bactérie telle que E. Coli a environ 4000 gènes, dans un génome de 4 millions de nucléotides, tandis que les génomes humains ont seulement 2300 gènes selon la dernière estimation, alors que la quantité totale de nucléotides de l'ADN est de 3 milliards. Donc il y a dans ce cas **une relation inverse entre la complexité des organismes et le rapport ADN codant/ ADN non codant.** (M. et M. Buiatti, 2008)

Plus l'organisme est évolué et complexe, plus nombreuses sont les zones non-protéinocodantes. Cela implique que **les séquences non-protéinocodantes sont en rapport avec des fonctions élaborées de gestion et d'instruction dans les organismes complexes.** Les chercheurs à l'heure actuelle font preuve d'humilité et n'emploient plus le terme d'ADN rebut. Ils admettent qu'il y a encore beaucoup de choses à découvrir et à comprendre dans le fonctionnement de l'ADN.

## Le réseau de régulation des gènes

La découverte de la structure de l'ADN a profondément transformé la biologie, catalysant le séquençage du génome humain et a engendré une nouvelle vision de la biologie en tant que science de l'information... L'ADN a 2 types d'information numérique - les gènes qui codent pour les protéines, qui sont les machines moléculaires de la vie, et **le réseau de régulation des gènes qui spécifient le comportement des gènes** (Hood L, Galas D., 2003).

John S. Mattick, biologiste à l'Institut pour la Bioscience moléculaire (université de Queensland, Australie), écrit: *Le dogme central de la biologie affirme que l'information génétique passe normalement de l'ADN à l'ARN, puis à la protéine. En conséquence, on a généralement supposé que les gènes codaient généralement pour des protéines, et que les protéines remplissaient des fonctions non seulement structurales et catalytiques, mais aussi régulatrices, dans toutes les cellules, des microbes aux mammifères. Toutefois, cela pourrait ne pas être le cas dans les organismes complexes* (John S. Mattick, 2003).

La preuve des fonctions régulatrices des zones non-codantes se trouve dans le fait que **certaines maladies proviennent de déviations des informations issues de séquences non-codantes.** Des millions, pour ne pas dire des centaines de millions de personnes mourraient de maladie de l'ADN "rebut" tandis qu'officiellement 98,7% de



*Le ver Caenorhabditis elegans a presque autant de gènes que nous!  
Merci à AJC1*

l'ADN humain était considéré comme intouchable (A. Pellionisz). Sont ainsi concernées des maladies du système nerveux, cardiovasculaires, cancer, et bien d'autres (voir une [liste ici](#))

On commence à comprendre avec quels outils est effectuée cette régulation. On découvre la réalité et la complexité de toute une nouvelle gamme d'acteurs: pour l'essentiel, les **transposons** et diverses catégories d'**ARN**. Voyons de quoi il s'agit.

**Sources:** Hood L, Galas D., [The digital code of DNA](#), Nature, 2003, 23, 42

John S. Mattick, [Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms](#), BioEssays 2003, 25, 10; [The Human Genome: RNA Machine](#), The Scientist, 2007, 21, 10, 61

A. Pellionisz, [Hereditary Diseases: « Not all is in the genes »](#), compilation par International HoloGenomics Society

Buiatti M. et M., [Chance vs. Necessity in Living Systems: A False Antinomy](#), Biology Forum, 2008, 101, 29; Menconi G., Bencia V., Buiatti M., [Data compression and genomes: a two dimensional life domain map](#), J Theor Biol. 2008, 21, 253

## Des séquences mobiles, les transposons

Dans l'article précédent ([ADN et expression](#)), j'ai déjà exposé ce que sont les transposons et comment ils ont été découverts par B. McClintock, puis oubliés. Dès 1965, elle a suggéré que les transposons pourraient jouer un rôle de régulation en indiquant à quel moment un gène doit devenir actif. **Les transposons sont des séquences d'ADN qui sont copiées et insérées dans une autre partie non-codante du génome.** En 1969, pour expliquer la différenciation des cellules d'un organisme, Britten et Davidson ont suggéré que le transposon pourrait déclencher un signal d'activation ou de désactivation selon l'endroit où il se localise dans l'ADN (non-codant). Actuellement, on estime que les transposons occupent **plus de 40% du génome humain.**

Les séquences dispersées répétées SINE décrites plus haut, dont la longueur varie de plusieurs centaines à plusieurs milliers de nucléotides, sont de tels transposons. En 2002, des chercheurs japonais (Ogiwara, 2002) ont étudié une nouvelle famille de séquences répétitives **SINE** chez les vertébrés. En comparant les divers vertébrés, ils ont constaté combien ces SINE ont été bien conservés d'une espèce à l'autre pendant des années, ce qui est hautement significatif **de leur rôle indispensable.**

En 2003, [Gill Bejerano](#) et ses collaborateurs (Université de Standford, USA) ont découvert une autre famille de SINE exceptionnellement stable depuis des millions d'années chez les coelacanthes, poissons indonésiens. Ils ont confirmé le **rôle de régulation** de certains transposons, en constatant la façon dont la couleur du poisson est transmise. Le transposon active à distance le gène de la couleur.

**Sources:** Leslie Pray, [Transposons, or Jumping Genes: Not Junk DNA?](#) Nature Education 2008, 1, 1

J. Mattick, [The Human Genome: RNA Machine](#), The Scientist, 2007, 21, 10, 61

Ogiwara I, Miya M, Ohshima K, Okada N., [V-SINEs: a new superfamily of vertebrate SINEs that are widespread in vertebrate genomes and retain a strongly conserved segment within each repetitive unit](#); Genome Res. 2002, 12, 2

G. Bejerano, C.B. Lowe, N. Ahituv, B. King, A. Siepel, S.R. Salama, E.M. Rubin, W.J. Kent and D. Haussler. [A Distal Enhancer and an Ultraconserved Exon are Derived From a Novel Retroposon](#). Nature, 2006, 441, 7089

C. Wills, "Exons, Introns, and Talking Genes: The Science Behind the Human Genome Project", ed. Basic Books, 1991, 239, cité dans [The Functions of Introns: From Junk DNA to Designed DNA](#), par Jerry Bergman, 2001

[Introns: a Mystery](#)

Steven Zimmerly, Huatao Guo, Phillip S. Perlman and Alan M. Lambowitz, [Group II Intron Mobility Occurs by Target DNA-Primed Reverse Transcription](#), Cell, 1995, 82, 545

Jian Yang, Steven Zimmerly, Peter Perlman and Alan M. Lambowitz, [Efficient integration of an intron RNA into double-stranded DNA by reverse splicing](#), Nature, 1996, 381, 332.

## La communication par les ARN non-codants

Les questionnements évoqués plus haut - origine de la différenciation des cellules d'un même organisme, rôle des zones non-protéinocodantes et des transposons - trouvent des réponses dans les nombreuses avancées effectuées dans l'analyse de la transcription de séquences d'ADN non-protéinocodantes en ARN (npcARN).

**Un nombre impressionnant de nouveaux ARN non-protéinocodants qui ont un rôle génétique ont été découverts** dans les organismes complexes tels que les mammifères et les humains.

*Les ARN non-protéinocodants semblent être particulièrement abondants dans des rôles qui demandent une reconnaissance hautement spécifique des acides nucléiques sans catalyse complexe, telle que diriger la régulation post-transcriptionnelle de l'expression des gènes ou guider les modifications de l'ARN (Eddy, 2001).* **Les npcARN dominant l'expression génomique des organismes supérieurs et contrôlent l'architecture chromosomique, le flot des ARN messagers, et le timing de l'expression de protéines, et pourraient aussi réguler la transcription et l'épissage alternatif.** (J. Mattick, 2003).

Les npcARN sont impliqués dans les maladies évoquées plus haut relatives aux zones non-protéinocodantes.

Le fonctionnement correct des npcARN est également important pour la santé humaine: des changements dans les npcARN ont été impliqués dans **des attaques cardiaques, et des maladies comme le cancer.** Beaucoup sont

visibles dans le cerveau et au moins un est impliqué dans des attitudes comportementales [chez la souris]. (Mattick 2007)

## Les micro-ARN

Vers 2000, personne n'avait encore entendu parler des micro-ARN. Puis on en a découvert dans la biologie des vers, et cela a été considéré comme étrange. Ce sont de petits npcARN. En fait, ils sont très abondants et contrôlent l'expression des gènes dans tous les animaux et les plantes, et même quelques virus. Les scientifiques les ont repérés dans tous les systèmes organiques majeurs.



John Mattick

*En 3 ans, des chercheurs ont découvert des centaines de microARN, dont plus de 150 chez les humains. Ils semblent un moyen bien établi pour les organismes de moduler les gènes. Plusieurs sortes d'ARN travaillent: catalyse, signalement, activation comme n'importe quelle protéine. (Wayt 2003)*

Certains micro-ARN agissent sur les ARN messagers qui vont porter les codes génétiques des protéines dans les ribosomes de la cellule. Ils sont capables de les inhiber. Ils ont aussi bien d'autres fonctions: *Presque tous les gènes importants et les voies de synthèse sont régulés à de multiples niveaux par une variété de micro-ARN. C'est vraiment une strate entièrement nouvelle de la biologie* (Deepak Srivastava, University of California, San Francisco, USA). (Phillips 2007)

*Il y a de bonnes raisons de penser que ces micro-ARN représentent une strate de régulation jusqu'à maintenant cachée qui encode le programme de développement des eucaryotes et partant, s'est largement répandue dans les organismes complexes. Il est de plus en plus reconnu que les npcARN, y compris ceux qui proviennent des introns, sont impliqués dans tous les aspects des processus de régulation cellulaire, y compris le changement de conformation de la chromatine et la mémoire épigénétique (voir article précédent), le transport des enzymes du noyau, et l'activation ou répression de la transcription... Les micro-ARN connus et autres petites séquences ARN sont seulement un faible pourcentage de ce que comprend probablement le transcriptome fonctionnel des npcARN. (Mattick, 2007)*

Il est clairement établi que **toutes les parties de l'ADN ont un rôle dans la machinerie génétique**, une machinerie qui se révèle très complexe. Les parties qui ne codent pas pour les protéines exercent un contrôle sur le développement et les traits distinctifs de tous les organismes, des bactéries jusqu'aux humains. Les micro-ARN en sont les acteurs.

*Si tous ces npcARN sont fonctionnels, comme les faits le suggèrent de façon croissante, alors **une grande partie et peut-être la presque totalité du génome humain est fonctionnel**. Si oui, **la programmation génétique des organismes supérieurs a été mal comprise pendant les dernières 50 années**, à cause de la supposition - largement vraie pour les procaryotes, mais pas dans les eucaryotes complexes - que l'information génétique se manifeste et est transmise par les protéines. (Mattick, 2007)*

Le sens donné habituellement aux termes **gène** et **code génétique** n'est peut-être plus adéquat. Ils devraient être redéfinis. *Le terme **gène** a toujours été quelque chose de mal défini; ces gènes pour ARN élargissent sa signification. Pour éviter toute confusion, dit Claes Wahlestedt (Karolinska Institute, Suède), "nous tendons à ne plus parler de "gènes"; nous disons d'un segment d'ADN transcrit en ARN qu'il est une **unité de transcription**". (G. Wayt, 2003)*

**Sources:** Eddy SR., *Non-coding RNA genes and the modern RNA world*, Nat Rev Genet. 2001, 2, 919

Mattick John S., *The Human Genome: RNA Machine*, The Scientist, 2007, 21, 10, 61

Mattick John S., *Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms*, BioEssays, 2003, 25, 10

Phillips Melissa Lee, *MicroRNAs: An emerging portrait*, The Scientist, 2007, 21, 55

Gibbs W. Wayt, *The Gems of "Junk" DNA*, Scientific American, 2003, November

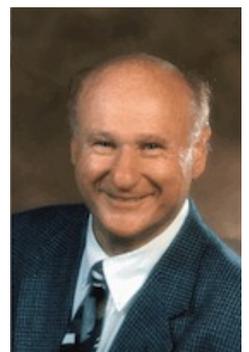
Buiatti M et M, *Towards a statistical characterisation of the living state of matter*, Chaos, Solitons and Fractals 2004, 20, 55

## L'information itérative

Pour Andrés Pellionisz, un biophysicien anciennement à l'université de New York, les 100% de l'ADN fournissent des informations codées qui entrent en jeu à certaines phases des processus biologiques en échangeant des informations **de façon itérative**.

Lorsqu'une cellule est en construction, par exemple une cellule neuronale du cerveau ou cellule de Purkinje, un squelette de protéine est d'abord construit selon le modèle classique de la transcription: gène > ARN messager > protéine. Mais avant la transcription en ARN, il y a un signal, une instruction qui est donnée: *Activer le gène*.

Lorsque cette instruction a été lue et effectuée, un marqueur sous forme de protéine se lie à l'ADN pour signifier *Éteindre le gène*. Pour le désactiver, le marqueur accroche un groupe méthyle sur la partie adéquate du gène, comme il a été vu dans l'article [Expression de l'ADN et](#)



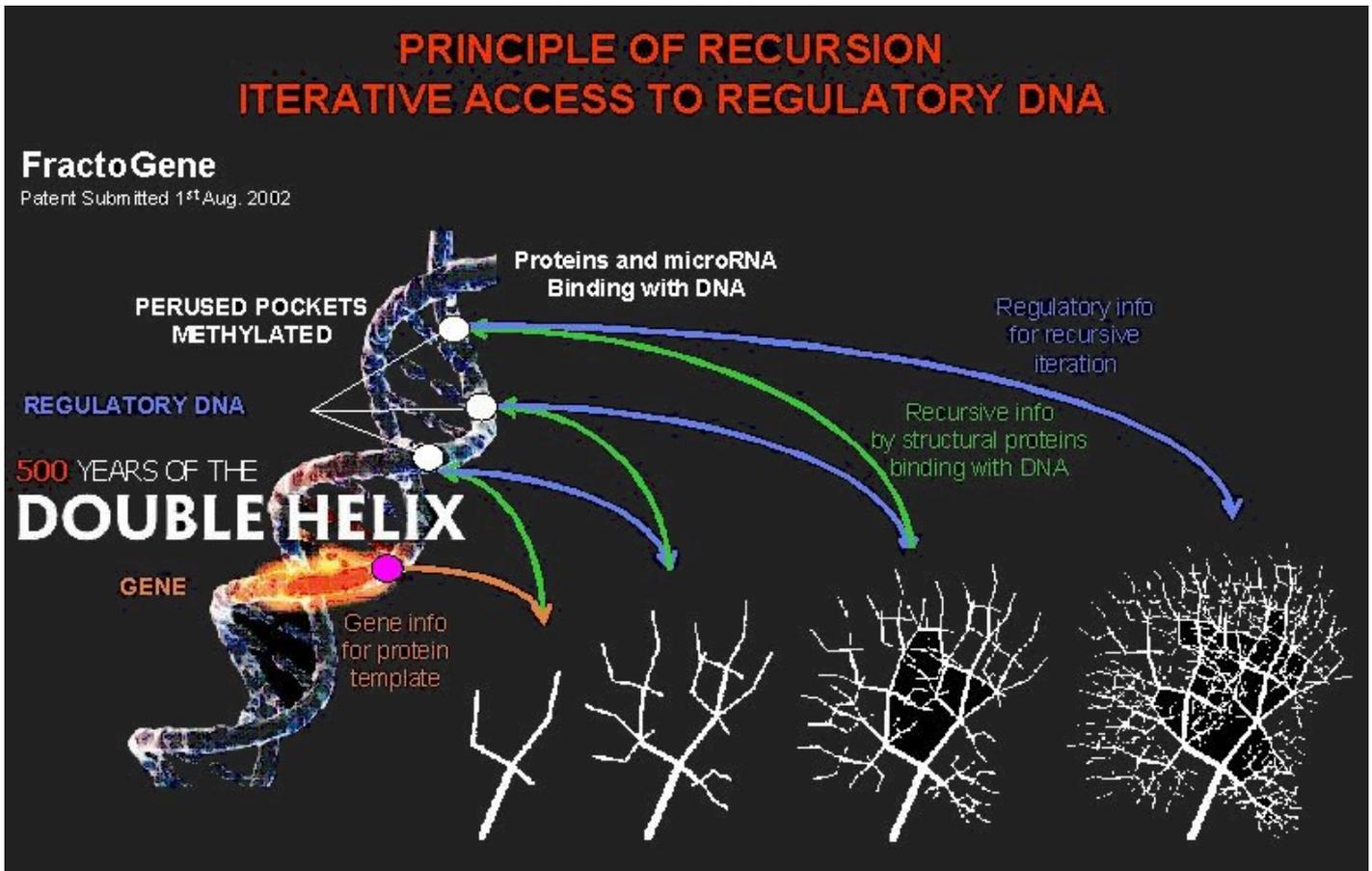
Andreas Pellionisz

épigénétique. L'**épigénétique** est ainsi replacée dans une perspective plus large, l'**hologénétique**. Ajoutons que les **microARN** sont des acteurs de ces instructions.

Ceci est la première étape, dont le résultat est un embryon de protéine.

Puis cet embryon va chercher un deuxième lot d'instructions pour continuer son assemblage. **Ces instructions auxiliaires sont fournies par une séquence de l'ADN située dans une partie non-protéinocodante, une de ces séquences quasi-répétitives** dont on a parlé plus haut. Justement, les variantes du motif de répétition sont hautement significatives et font partie des instructions.

A la troisième étape, la protéine en construction va chercher un autre lot d'instructions complémentaires dans une autre séquence non-codante, et ainsi de suite. L'entité se construit par étapes, et à chacune d'elle, elle se développe avec une hiérarchie fractale, de la protéine au corps entier en passant par la cellule et le tissu.



*Schéma du processus itératif de construction fractale d'une cellule selon A. Pellionisz ©*

Plus loin, nous allons explorer la structure fractale de l'ADN non-codant. La hiérarchie fractale de l'entité biologique en est le reflet. Ce processus itératif est appelé l'**hologénétique**. Ainsi, tout l'ADN est sollicité dans la construction du corps, y compris les 98,7% d'ADN non-codant. Il fournit le plan d'assemblage.

Les maladies hologénétiques seraient dues à des erreurs des constructions fractales. Selon Pellionisz, on peut diagnostiquer ces défauts en mesurant la structure fractale des séquences génomiques non-codantes. Il a breveté cette méthode.

**Sources:** Andras J. Pellionisz, [The Principle of Recursive Genome Function](#), The Cerebellum, 2008, 7

Hal Plotkin, [Junk DNA: Revisited Silicon Valley startup claims to have unlocked a key to its hidden language](#). San Francisco Chronicle, 2002, 21

## Un ordre global dans l'ADN

Toutes les avancées dans la recherche des fonctions de l'ADN que j'ai présenté jusqu'à présent reposent sur le même type d'investigation, à savoir qu'on porte son regard sur le comportement des séquences. Il s'agit d'un regard local, morceau par morceau. Or un autre type d'approche a été entrepris depuis 1992 environ. Il envisage l'architecture globale du brin d'ADN, dont les propriétés statistiques ont été intensément étudiées. Le regard est porté sur l'ensemble. S'en dégage-t-il des formes, des ordonnancements?

Les recherches ont été stimulées par la publication progressive du séquençage des gènes, dont les résultats ont été mis à disposition dans des banques de données. L'effort le plus important pour fournir ces données a été fait dans le cadre du [Projet génome humain](#) international, qui a été terminé en 2004.

## Le décompte statistique

Certains ont utilisé des statistiques en comptant les différents éléments de l'ADN et en calculant comment ils se distribuent. Quels éléments?

Si on ignore la distinction entre zones codantes et non-codantes, on peut considérer l'ADN comme une chaîne continue de nucléotides. Si l'on reconstruit mentalement cette chaîne, on place d'abord un élément, puis un deuxième et se pose alors la question: le choix du prochain élément à poser est-il régi par une règle? Y a-t-il un ordre, ou bien les éléments sont-ils posés au hasard?

En science, le **hasard** est défini de la façon suivante. Nommons les éléments constitutifs sous la forme de 4 lettres, A, T, C, G, et imaginons qu'elles sont inscrites sur des carrés de scrabble en nombre illimité, avec une répartition égale des 4 lettres. Tirons au sort la première lettre, nous avons autant de chance d'avoir une lettre qu'une autre. Pour la seconde et toutes les suivantes, même chose. En définitive, chacune des lettres a 1 chance sur 4 d'apparaître dans le filament **quelle que soit sa place**. Si le filament est long, il y aura en moyenne autant de lettres de chaque espèce, 1/4 de A, 1/4 de T, etc. Donc, **dans le cas du hasard, la fréquence d'apparition des lettres est la même et égale à 25%**. Ou, autre façon de le présenter, la fréquence de A par rapport à TCG est de 1/3 (**0,33**).

Ainsi, une méthode statistique d'analyse rudimentaire est de dénombrer la **fréquence d'occurrence des lettres**. Si cette fréquence s'écarte notablement de la valeur 0,33, cela montre qu'un processus autre que le hasard est en oeuvre. Il crée les enchaînements selon une autre règle, une certaine logique, et par là il installe un ordre.

Au lieu de considérer les éléments constitutifs, les briques élémentaires, comme étant une seule lettre à la fois, on peut les définir comme une groupe de lettres. Ainsi, comme nous savons que les codons sont des groupes de 3, on peut définir l'ADN comme une chaîne de groupes de 3, au moins dans les parties codantes. Ou même des groupes de 2, 4 ou 10. Les résultats peuvent être différents selon le groupement envisagé.

Une autre façon consiste à compter ensemble la somme des C et G d'une part et la somme des A + T d'autre part. En effet, les chercheurs ont remarqué que **les nombres de C et de G étaient sensiblement les mêmes**. De même pour A et T. C'est une sorte de symétrie étonnante de l'ADN.

## A la recherche du nombre d'or

Imaginons que la base T se trouve systématiquement tous les 10 emplacements. Ce serait un ordre très régulier qui nous interpellerait. Il existe de telles sortes d'ordres, mais plus flous. Par exemple, au 10e emplacement, on a une chance plus élevée de trouver T que les autres bases. On dit que les emplacements des T sont **corrélés** entre eux. On mesure ainsi la probabilité de trouver telle base à telle distance d'un point de départ.

C'est une analyse de ce type qu'a entreprise le mathématicien Jean-Claude Pérez pour mesurer la fréquence d'apparition des bases sur des segments de différentes longueurs. Il a trouvé que **les fréquences de répartition des nucléotides ne sont pas dues au hasard**.

La méthode semble toutefois curieuse, car au lieu de rechercher la valeur de ces fréquences, il en fixe une a priori et il recherche si elle se produit. La valeur de fréquence recherchée d'une base par rapport aux 3 autres est 1,618. Cette valeur est traditionnellement nommée le **nombre d'or**. Il est indéniable que lorsque cette valeur, qui est presque le double de la valeur du hasard 0,33, apparaît dans les résultats du calcul, cela signifie que la distribution des nucléotides ne tient pas du hasard.

Chaque fois que cela se produit, le mathématicien considère cette occurrence comme une résonance positive à sa quête, et il dénombre ces résonances. Le total constitue une mesure de l'ordre. Les résultats révèlent des résonances abondantes. Pérez constate alors que l'ordre est différent selon les segments. Il décroît fortement quand des parties y sont insérées artificiellement comme c'est le cas dans les OGM (voir *L'ADN décrypté*, J.C. Pérez, Ed. Résurgence).

## Des corrélations dans les parties non-codantes de l'ADN

L'examen de milliers et millions de bases dans un segment pour trouver des corrélations est fastidieux. Mais les mathématiques offrent des outils qui permettent d'extraire les corrélations d'un ensemble d'éléments. Ce sont au fond des façons de présenter les ensembles d'objets d'une façon telle que le cerveau peut mieux appréhender les structures globales.

Ces analyses ont été effectuées sur des séquences de longueurs variées, de gènes courts à des filaments entiers, sur des zones codantes et non-codantes.

*Pour les connaisseurs, je citerai quelques-unes de ces méthodes. Les **transformées de Fourier** permettent de mettre en évidence les répétitions ou les périodicités d'une courbe, un procédé qui est couramment employé pour détecter les harmoniques d'un son (voir exemple dans *Nature du son et Timbre et Harmoniques*). Les **fonctions de corrélation** permettent de mesurer si des corrélations existent entre les éléments de la structure analysée. On emploie aussi la transformée en ondelettes, et des méthodes de calcul d'entropie.*

Le résultat qui en émerge est que **l'ADN est caractérisé par des corrélations à courte portée et des**

## corrélations à longue portée.

Précisons un peu ce que signifie une corrélation dans ce cas. Si on trouve une base T à un certain emplacement, cela influe sur la probabilité d'avoir un G (ou une autre base) à un autre emplacement, que ce dernier soit situé 100 nucléotides plus loin sur la chaîne, ou 1000, 10'000 ou plus. Comme le dit I. Amato (*Revue Science*, 1992), c'est comme si, sachant qu'il fait beau le 14 janvier de chaque année, on savait mieux prévoir s'il y aura du brouillard le 8 septembre. Une telle corrélation ne donne pas d'explication sur le type de lien entre les emplacements. On se contente de constater qu'il y a un lien de probabilité, mais on ne sait pas quelle en est la cause.

Ces corrélations sont très différentes dans les zones codantes et dans les zones non-codantes. **Dans les exons (séquences codantes), on ne trouve pas ou peu de corrélations.** Les nucléotides sont disposés selon un ordre, certes, qui est celui du code génétique, mais il est localisé précisément. Par contre, **dans les zones non-codantes, on trouve des corrélations à longue portée.**

La répartition des bases le long du brin d'ADN n'est donc pas le fruit du hasard, y compris et surtout dans les zones non-codantes.

### Sources:

- Li, W. and Kaneko, K. 1992. Long-range correlation and partial 1/f spectrum in a noncoding DNA sequence. *Europhys. Lett.*, 17, 655-660; Li W., Marr T.G., Kaneko K., [Understanding long-range correlations in DNA sequences](#), 1994, *Physica D* 75, 392; Li W., [The Study of Correlation Structures of DNA Sequences - A Critical Review](#), 1997, *Computers & Chemistry* 21, 4, 257; Li W. [Universal 1/f noise, crossovers of scaling exponents, and chromosome-specific patterns of guanine-cytosine content in DNA sequences of the human genome](#), 2005, *Phys. Rev. E* 71
- Voss R., [Evolution of long-range fractal correlations and 1/f noise in DNA base sequences](#), 1992, *Phys. Rev. Letters*, 68, 3805
- Amato I., [DNA shows unexplained patterns writ large](#), 1992, *Science* 7, 257, 5071, 747
- [Noisy Nucleotides, DNA sequences show fractal correlations](#), 1992, *Scientific American* Sept 92
- Peng C. -K., S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, 1994, *Phys. Rev. E*, 49, 1685; Peng C. -K., S. Havlin, H. E. Stanley, and A. L. Goldberger, [Long-range correlations in nucleotide sequences](#), 1995, *Chaos*, 5, 82; Havlin S., S. V. Buldyrev, A. L. Goldberger, R. N. Mantegna, C.K. Peng, M. Simons and H. E. Stanley, [Statistical Properties of DNA Sequences](#), 1999
- Buiatti M. et M., [Towards a statistical characterisation of the living state of matter](#), 2004, *Chaos, Solitons and Fractals* 20, 55
- Selvam A.M., [Universal spectrum for DNA base C+G frequency distribution in Human chromosomes 1 to 24](#)

## L'ordre fractal

Dans la recherche de corrélations, les chercheurs ont constaté que le degré de corrélation reste le même pour des portions courtes de 1000 nucléotides, ou longues de 10'000 ou 100'000, donc quelle que soit l'échelle d'observation. C'est aussi ce qu'a constaté J.C. Pérez pour l'existence des résonances. Autrement dit l'architecture de l'ADN a un visage semblable quelque soit l'échelle à laquelle on l'analyse.

Par définition, une structure qui présente la même forme à différentes échelles est nommée **structure fractale**. Le degré de corrélation est sa dimension fractale (*voir Les images fractales*).

En comparant la dimension fractale de gènes de différents organismes, R. Voss (*Revue Science*, 1992) a trouvé que chaque catégorie génétique d'êtres vivants (primates, invertébrés, organelles, etc.) avait chacune sa dimension fractale particulière. Les organismes les moins évolués tels que les bactéries sont peu corrélés, la corrélation est de plus en plus forte en montant sur l'échelle de complexité. Cela provient du fait que la quantité de zones non-codantes augmente.

## Des similitudes avec le langage humain

Certaines équipes ont comparé la structure fractale de l'ADN avec celle du langage humain. La **fractalité du langage humain** repose sur des études faites par [Georges Zipf](#) (1902-1950) qui a compté la fréquence avec laquelle chaque mot apparaît dans un texte littéraire. Il en a déduit que cette fréquence est inversement proportionnelle au rang du classement de ce mot. Par exemple, la fréquence du mot qui vient au 3e rang a sa fréquence d'apparition 3 fois moindre que celle du mot en première position. Par la suite, Mandelbrot a corrigé cette loi sans la remettre fondamentalement en cause.

Elle a été vérifiée dans plusieurs langues: anglais, turc (Luděk Hřebíček 1992, Hacinliyan), etc. La structure est présente parce que cette loi d'occurrence est vraie quelle que soit la longueur du texte, du moins au-dessus d'un nombre minimum de mots qui permet de faire une statistique.

Selon différentes équipes, les lois de corrélation des bases de l'ADN sont semblables à la fractalité du langage. Il faut seulement définir ce qu'est un mot dans le texte de l'ADN. Dans les parties codantes, il est clair que c'est le codon, comprenant 3 lettres, les 3 nucléotides qui codent pour un acide aminé. Pour les parties non-codantes, les chercheurs de l'équipe de Stanley et Mantegna (1992) ont testé des mots de 3 à 8 paires de nucléotides. Dans tous les cas, ils ont trouvé que la fréquence d'apparition des mots dans les zones non-codantes suit effectivement la loi de Zipf-

Mendelbrot.

Une autre caractéristique commune au langage humain et à l'ADN, c'est la **redondance**. Le caractère de redondance dans le langage humain signifie qu'une phrase peut être comprise même si certains mots sont déformés ou manquants. Le nombre de lettres et de mots contenus dans une phrase est plus grand que le minimum requis pour en percevoir le sens. En appliquant une analyse statistique selon la théorie de l'information de Shanon (1950), la même équipe de chercheurs a trouvé que les zones non-codantes contiennent 3 à 4 fois plus de redondance que les zones codantes.

**Sources:**

Mantegna RN, Buldyrev SV, Goldberger AL, Havlin S, Peng CK, Simons M, Stanley, [Linguistic features of noncoding DNA sequences](#), Phys Rev Lett. 1994, 5, 73, 23, 3169

[Noisy Nucleotides, DNA sequences show fractal correlations](#), 1992, Scientific American, Sept

Luděk Hřebíček, Text as a Self-Similar Structure, 1992, Text in Communication, Quantitative Linguistics Vol. 48, Universitätsverlag, Bochum, 1992; Luděk Hřebíček, Fractals in language, 1994, Journal of Quantitative Linguistics, 1, 1, 82;

Avadis Hacınlıyan, Murat Erentürk, and Gökhan Şahin, [Possible Chaotic Structures in the Turkish Language with Time Series Analysis](#), 2008, Unifying Themes in Complex Systems, Part III.:, Pages 618-625

## Le code génétique électromagnétique

La structure fractale démontre que **l'ADN est construit selon un ordre**, autrement dit une organisation. Toute organisation porte en elle une information, car enregistrer une information dans la matière suppose une certaine forme d'arrangement codé de cette matière (*voir article [Codes et information](#)*).

Des chercheurs russes sous la direction de P. Gariaev ont étudié l'information génétique que cette organisation pouvait enregistrer. Et ils ont trouvé des pistes très intéressantes. Ils n'ont pas traqué des molécules éventuellement capables de lire ces informations, comme l'ont fait toutes les autres investigations présentées jusqu'ici dans cet article.

Ils ont trouvé que le brin d'ADN pouvait être lu avec une onde électromagnétique, comme un rayon laser lit des signaux enregistrés sur un CD. On envoie une onde de caractéristiques adéquates sur l'ADN. **En parcourant l'ADN, cette onde électromagnétique est modulée par la structure et transporte l'information codée ailleurs dans l'organisme.**

Les 98,5% [restant] de la totalité de l'ADN contiennent des programmes de niveau significativement supérieur (*P. Gariaev et coll, 2002*)

Pour étudier cela, nous devons auparavant faire connaissance avec le champ électromagnétique de l'ADN, ce que nous faisons dans la [partie 6](#).

## En savoir plus

### Autres articles de cette série

1. [La molécule d'ADN et le code génétique](#). De la cellule aux gènes, en passant par les chromosomes et l'ADN, vous êtes invités à visiter **les rouages du programme génétique qui commande notre développement physique**. Comment fonctionne-t-il? Jusqu'à quel point nous contrôle-t-il? Quel est son langage? Vous pourrez le découvrir sans notion de biologie ou de chimie en observant le paysage, tel un voyageur qui s'aventure dans le monde des molécules.

2. [L'ADN et ses modes d'expression](#). Contrairement à l'idée répandue selon laquelle nous sommes programmés par notre code génétique, des scientifiques ont montré que celui-ci est en réalité un stock de données qui peuvent être activées ou non selon nos conditions de vie (nutritionnelles et psychiques). La science de l'**épigénétique** a montré que cette activation était due à des modifications chimiques réversibles du gène. Chacun de nous est donc dans un état épigénétique qui lui est propre et qui se modifie avec l'âge. Dans certaines circonstances, cet état est transmissible à la **descendance**, et cela bouleverse les idées figées des scientifiques sur l'évolution des espèces par la sélection naturelle. D'autres observations nous démontrent que l'ADN et les gènes ne sont pas des assemblages constitués de façon fixe et définitive. L'ADN se recompose en partie lorsque certains fragments (les **transposons**) changent de place. La **plasticité des cellules nerveuses** est un autre exemple qui montre combien nos cellules ne sont pas constituées une fois pour toutes, mais possèdent la capacité étonnante de s'adapter au changement et d'inventer de nouvelles formes.

4. [La musique de l'ADN et des protéines](#). La structure de l'ADN et des gènes sous-tend une harmonie que certains artistes et compositeurs ont transcrite en musique. Au-delà de ces visions d'artiste, la physique quantique montre, grâce à Joël Sternheimer, qu'**à chaque acide aminé composant une protéine est associée une onde d'échelle, qui peut être transcrite en note de musique**. Par la musique des protéines ou **protéodies**, il est possible d'entrer en dialogue intime avec l'organisme, ce qui ouvre des perspectives passionnantes et nouvelles en agriculture et en médecine.

5. [L'ADN électrique](#). On représente habituellement la molécule d'ADN sous forme de volumes géométriques: hélices, rubans et segments. Au-delà de son occupation dans l'espace, une vie électronique intense se manifeste dans les molécules, responsable de leurs attirances, associations et assemblages. De nombreuses recherches ont été conduites sur la conductivité électrique de l'ADN nu, donc en-dehors du corps. Récemment, il a été démontré que l'ADN est électro-conducteur et peut être considéré comme un minuscule fil électrique. Ces recherches sont motivées par la possibilité d'utiliser l'ADN comme constituant de nano-circuits électroniques (à l'échelle du nanomètre). Des ordinateurs à base d'ADN ont été construits et testés. L'ADN participe ainsi à la grande course des nanotechnologies qui permettent de fabriquer des puces et autres dispositifs de taille très inférieure à celles élaborées avec le silicium. Une technologie qui se répand pour le meilleur et pour le pire.

6. [L'ADN électromagnétique et la communication entre molécules](#). Depuis une centaine d'années, des scientifiques de plusieurs pays (Gurwitsch, Kaznacheev, Gariaev, Inaba, Popp, et d'autres) ont montré que les organismes vivants émettent de la lumière (biophotons) à très faible intensité. Tel un laser, l'ADN est à la fois la source et le lieu de stockage de ces photons. L'ensemble des biophotons de l'organisme constitue un champ cohérent porteur d'information, sous forme d'hologrammes, qui dirige les processus vitaux de l'organisme et maintient son intégrité. Grâce à ces rayonnements, les cellules communiquent entre elles et envoient des informations sur leur état énergétique et sanitaire. D'autres informations constituent un code génétique électromagnétique holographique qui assure et coordonne le développement de l'organisme. Cela explique des phénomènes inexplicables par la génétique moléculaire comme la différenciation des cellules. Des applications pratiques de ce phénomène ont été conçues pour évaluer la qualité des aliments et améliorer l'état de santé des êtres vivants par des techniques non destructrices.

### Documents sur internet

- [Il était une fois l'ADN](#), Dolan DNA Learning Center, Cold Spring Harbor Laboratory, traduction Université de Genève. En particulier, 32 [Certains fragments d'ADN peuvent sauter](#), B. McClintock; 33 [Les gènes peuvent être induits et réprimés](#), J. Monod et F. Jacob
- [Projet génome humain](#), Génoscope - Centre National de Séquençage
- [Biologie végétale et génétique](#), Gilles Bourbonnais, Collège de Sainte-Foy, Québec. Voir [Chromosome 8](#), l'intérêt personnel
- Encyclopédie Wikipedia: [Épissage](#); [Gène](#); [Empreinte génétique](#); [Fichier national des empreintes génétiques](#); [Génome](#); [Projet génome humain](#)
- [Empreintes génétiques: FNAEG: Fichier national des empreintes génétiques](#),
- [En quête de l'identité](#), les empreintes génétiques et l'identification judiciaire, Eric Heilmann

- [Les empreintes génétiques en pratique judiciaire](#), Christian Doutremepuich, Bull. Acad. Natle Méd., 2012, 196, 6, 1117-1130
- [L'architecture brisée de l'ADN](#), J.C. Perez

### Autres sites en anglais

- [Introns: a Mystery](#)
- [Introns first](#) par Donald R. Forsdyke
- [The Functions of Introns: From Junk DNA to Designed DNA](#), par Jerry Bergman 11/18/2001 Reprinted from Perspectives on Science and Christian Faith Volume 53, Number 3, September 2001
- [The Human Genome: RNA Machine](#), J. Mattick, The Scientist, 2007, 21, 10, 61
- [The Principle of Recursive Genome Function](#), Andras J. Pellionisz, The Cerebellum, 2008, 7
- [Transposons, or Jumping Genes](#), Leslie Pray, Nature Education 2008, 1, 1
- [The Gems of "Junk" DNA](#), Gibbs W. Wayt, Scientific American, 2003, November

Texte conforme à la [nouvelle orthographe française](#) (1990)

13 novembre 2009 - Ajouts et révision 03 juillet 2012

© Copyright 2009-2015 - Alain Boudet  
[www.spirit-science.fr](http://www.spirit-science.fr) - France

Tous les documents présents sur ce site sont protégés par les lois sur les droits d'auteur.

[Les publications de ce site sont identifiées par le numéro international ISSN 2430-5626](#)

**Cet article est l'aboutissement d'études, d'investigations, de compréhensions, de synthèse, de réflexions, de clarifications et de reformulation en langage simple, qui ont demandé une somme importante de travail.**

**Si vous deviez en tirer parti devant un public de lecteurs ou de spectateurs ou pour quoi que ce soit, merci de le citer.**