

# INTRODUCTION A LA GENETIQUE MOLECULAIRE

Jean-Marc Lobaccaro, Serge Lumbroso, Charles Sultan

Unité de Biochimie Endocrinienne du Développement et de la Reproduction  
Hôpital Lapeyronie et INSERM U 58, 34059 Montpellier, Cedex

**INTRODUCTION TO MOLECULAR GENETICS.** Following overviews of the structure of DNA, the organization of the human genome, and a description of the process of gene transcription, the second part of this review considers the principal molecular biological tools used to study genetic anomalies. Significant progress in molecular genetics during recent years and has led to the discovery of many new gene loci, greatly expanding the human gene map. Restriction fragment length polymorphism and enzymatic amplification studies are now routinely applied in the diagnosis of an ever increasing number of genetic diseases. Finally, elucidation of the molecular mechanisms of gene transcription facilitate the study of regulatory factors of gene expression. **Key Words :** molecular genetics, gene transcription, polymorphisms, gene expression, screening for mutations. **Andrologie, 1992, 2, 104-107.**

Les outils performants de la biologie moléculaire ont bouleversé la génétique durant ces 10 dernières années: carte génétique humaine enrichie, recherche de marqueurs polymorphes de l'ADN associés à la maladie, reconnaissance directe d'une mutation. Des techniques directes ou indirectes sont désormais appliquées au diagnostic d'un nombre de plus en plus grand de maladies héréditaires. Deux ouvrages de références rapportent l'état actuel de la carte génétique de l'homme (4) et les avancées de la biologie moléculaire dans le diagnostic génétique (5).

## 1. STRUCTURE GÉNOMIQUE

A l'intérieur des cellules, dont le corps humain est constitué d'environ 40 mille milliards d'unités, le noyau renferme l'ensemble du patrimoine génétique de l'individu qui est porté sur les chromosomes. Toutes les informations nécessaires au développement, à la vie et même au renouvellement des cellules et de l'organisme sont contenues dans un long ruban d'ADN qui, chez l'homme, mesure à peu près 2 mètres.

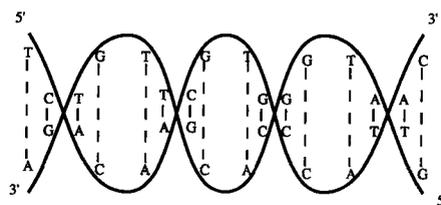


Fig. 1 : Structure dans l'espace de la double hélice de l'ADN. Le squelette phospho-déoxyribose est schématisé en gras. A, T, G, C sont les quatre bases. Les liaisons hydrogènes entre les deux chaînes sont schématisées en pointillés.

L'ADN est classiquement représenté par une double hélice constituée de deux chaînes complémentaires de nucléotides (Fig. 1) qui s'ouvrent comme une fermeture éclair quand la cellule a besoin de se diviser, et ce phénomène est indispensable à la duplication de l'ADN avant le début de la mitose cellulaire.

Le message héréditaire contenu dans la molécule d'ADN peut être transmis grâce à des petits fragments de la double hélice -les gènes- qui contrôlent spécifiquement la synthèse d'une chaîne polypeptidique particulière.

Le génome humain comporte  $3 \times 10^9$  paires de bases et il y a dans ce génome à peu près 100 000 gènes qui s'expriment et dont la taille est très diverse: de quelques centaines à plus d'un million de paires de bases.

Les gènes se localisent sur les chromosomes dans des sites qui s'appellent les locus et chacun des deux gènes situés dans le même locus de

chromosomes est appelé allèle. Il y aura homozygotie si l'on trouve deux allèles identiques d'un locus particulier sur une paire de chromosomes homologues, et hétérozygotie s'il y a deux allèles différents dans un locus particulier sur une paire de chromosomes homologues.

La transmission du message héréditaire de cellule à cellule est faite grâce à la réplication de l'ADN, phénomène qui a lieu pendant le cycle cellulaire à l'approche de la mitose (en phase S de l'interphase). Les deux chaînes d'ADN s'ouvrent, à ce moment, comme une fermeture à glissière par l'action d'une hélicase et au contact de chacun des brins ainsi dédoublés vient se former, à partir des nucléotides libres disséminés dans la cellule et sous l'action d'une polymérase d'ADN, nouvelle chaîne respectant scrupuleusement le principe de complémentarité des bases (Adénine-Thymine; Cytosine-Guanine). A la fin de ce processus, la cellule dispose de deux chaînes exactement identiques à la double chaîne initiale. Par ce mécanisme, chacune des deux cellules-filles aura la même information génétique que la cellule mère originale. Par ailleurs, dans la cellule elle-même, l'ADN nucléaire impose aux organites cellulaires responsables les instructions nécessaires à la synthèse des protéines. On peut dire, au sens figuré, qu'une cellule connaît deux langages pour exprimer son discours fonctionnel: le langage nucléaire qui s'exprime par mots de 3

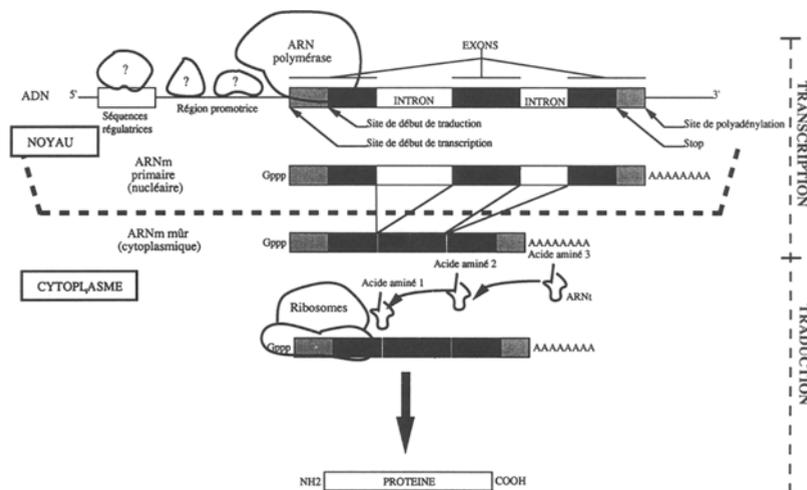


Fig. 2 : Schéma de la transcription et de la traduction.

lettres -les triplets ou codons, composés à partir d'un alphabet qui n'a besoin que de 4 lettres-; et le langage protéique qui s'exprime en termes d'acides aminés. La cellule passe d'un langage à l'autre par une traduction précédée d'une transcription (6).

Au moment de la transcription, la double hélice de l'ADN s'ouvre sur toute la partie correspondant à un gène. Sur l'un des deux brins viennent se placer des nucléotides libres d'ARN selon les principes de la complémentarité et grâce à l'action d'une polymérase de l'ARN. La séquence d'ARNm ainsi constituée est donc la réplique complémentaire d'une séquence d'ADN (telle une copie-miroir).

Les mécanismes de transcription, fort complexes, sont par ailleurs soumis à des règles de régulation et contrôle très particuliers.

## 2. TRANSCRIPTION GÉNÉRIQUE (Fig. 2)

Il y a dans chaque gène une région spécifique qui code la synthèse protéique et qui s'appelle gène de structure. Il existe, par ailleurs, d'autres régions adjacentes au gène qui ne sont pas directement concernées par la synthèse de protéines spécifiques mais qui jouent un rôle capital.

Leur mission est justement de réguler l'activité des gènes de structure et de contrôler leur propre expression (5, 6).

Juste en amont des gènes qui seront transcrits, il existe des séquences appelées promoteurs. Ces promoteurs comportent le site de fixation de l'ARN polymérase ainsi que les sites de fixation de protéines régulatrices de la transcription. En effet, un gène ne se limite pas à sa partie transcrite et encore moins à sa partie codante.

Le gène commence en 5' par une séquence non transcrite dont la présence est nécessaire pour que la transcription s'effectue quantitativement et qualitativement de manière normale. Ces séquences qui peuvent être très éloignées (jusqu'à quelques kb en amont) sont très difficiles à mettre en évidence et à délimiter de manière précise. On peut schématiquement distinguer plusieurs parties rencontrées successivement de 5' en 3': le promoteur où se fixe l'ARN polymérase, des séquences où se fixent un ou plusieurs facteurs protéiques de transcription ou certains récepteurs stéroïdiens activés par la liaison de l'hormone, une séquence TATA où se fixe l'ARN polymérase II. Vient ensuite le site d'initiation de la transcription, correspondant le plus souvent à une base purique et la séquence ATG, codon méthionine, qui signale le lieu d'ini-

tiation de la transcription. Suivent ensuite une succession d'exons et d'introns. Le signal d'arrêt de la traduction est donné par l'un des codons stop. Enfin, 10 à 20 paires de bases avant la fin du dernier exon, est retrouvée une séquence de reconnaissance pour la coupure du transcrit primaire. La terminaison du transcrit est bien plus en aval, dans une région riche en G et C, au niveau de signaux dont la nature n'est pas connue. Le gène se termine par une région 3' adjacente très mal connue où l'on a parfois caractérisé des séquences régulatrices (5, 6).

Il existe également des séquences d'ADN appelées amplificateurs qui stimulent en amont la transcription de certains gènes. Leur mécanisme d'action n'est pas connu, mais il est vraisemblable qu'ils agissent par l'intermédiaire de protéines. Ces amplificateurs pourraient également agir en modifiant la structure spatiale de l'ADN ou son taux de torsion.

Dans le mécanisme de formation de l'ARNm à partir d'ADN, il existe une étape de maturation. Cette étape comprend la fixation d'une guanine méthylée sur l'extrémité 5' avec méthylation des 2 premiers nucléotides, la fixation d'une séquence polyadénylée en 3', et l'épissage qui, après séparation des introns, va réunir les exons bout à bout. Cette molécule, ainsi réduite, transite alors vers le cytoplasme cellulaire où elle va être traduite en protéine (traduction).

La synthèse protéique est donc réalisée dans le cytoplasme grâce au rôle des ribosomes qui feront la lecture séquentielle des triplets et rassembleront les acides aminés correspondant aux codons. Ainsi, progressivement, les acides aminés iront s'attacher les uns à la suite des autres dans l'ordre exact déterminé précisément par la séquence d'ARN et en fait selon l'ordre inscrit sur le gène. C'est la traduction du message sous forme de synthèse d'une protéine spécifique (5, 6).

Un mécanisme particulier de transmission de l'information génétique qui explique certains traits héréditaires souvent difficiles à comprendre a été décrit sous la désignation anglaise d'*imprinting* ou empreinte génomique. Il s'agit d'une forme différente de régulation d'expression d'allèles homologues qui est due à un effet épigénétique. Par ce mécanisme, l'expression des allèles dépend de son origine paternelle ou maternelle, au contraire de ce qui se passe dans les conditions définies par l'hérédité classique ou le sexe des progéniteurs n'a rien à voir avec l'expression des 2 allèles d'un locus.

## 3. ANOMALIES DE L'INFORMATION

Les défauts de la transmission du message génétique peuvent être la conséquence d'anomalies touchant différents domaines du complexe système héréditaire.

Les anomalies chromosomiques numériques sont caractérisées par une altération du nombre des autosomes ou des gonosomes soit par excès (polysomies) soit par défaut (monosomies).

Les anomalies chromosomiques de structures sont la conséquence d'un changement au niveau chromosomique de la séquence normale des gènes, ce qui va perturber, de façon plus ou moins évidente, la transmission de l'information génétique. Plusieurs variétés d'anomalies de la structure des chromosomes sont bien connues depuis longtemps : délétions (terminales ou intercalaires), isochromosomes, chromosomes en anneau, inversions, insertions, translocations, etc...

Il y a encore des anomalies de l'information qui résultent de mutations géniques déterminées par une modification permanente de la structure de l'ADN, soit ponctuelle (changement d'une base par une autre), soit par mécanismes de délétion, insertion, conversion génique ou duplication de l'information transmise par les gènes.

## 4. OUTILS D'ÉTUDE DES ANOMALIES GÉNÉRIQUES

### 4.1. Polymorphismes de longueur des fragments de restriction (RFLP).

L'ADN diffère entre les individus d'une même espèce. Ces variations normales sont surtout localisées dans les zones intergéniques ou les zones non codantes des gènes où elles peuvent s'accumuler sans conséquence phénotypique. De telles variations peuvent supprimer ou faire apparaître un site pour une enzyme de restriction. Ces enzymes sont de véritables ciseaux biochimiques qui coupent la chaîne d'ADN en des sites spécifiques déterminés par une séquence. Par un tel mécanisme, ces endonucléases bactériennes vont cliver des séquences géniques connues et sectionnent la molécule d'ADN exactement aux endroits voulus. Une électrophorèse va permettre de trier les fragments d'ADN selon leur taille. Après transfert sur une membrane solide (méthode de Southern, 10), les fragments sont hybridés avec une sonde ADN spécifique du gène étudié. Les sondes utilisées pour détecter ces polymorphismes peuvent être intragéniques (dans ce cas, le risque de recombinaison entre un RFLP et un allèle particulier est considéré comme nul) justagéniques, ou extragéniques. Le Southern blot ne

peut mettre en évidence que les remaniements majeurs d'un gène comme les délétions (5 à 10 %) ou les conversions géniques. Les mutations ponctuelles ne peuvent être détectées que si elles créent ou abolissent un site de restriction.

Les RFLPs peuvent également être utilisés pour caractériser l'allèle porteur, suivre sa ségrégation au sein d'une famille et ainsi dépister les hétérozygotes ou réaliser un diagnostic anténatal.

#### 4.2. Amplification enzymatique par la réaction en chaîne de la polymérase (PCR).

Une approche plus facile est l'amplification élective d'une séquence d'ADN double brin, effectuée *in vitro* par l'extension itérative de 2 amorces, situées de part et d'autres de la région considérée, grâce à une ADN polymérase (8). L'amplification est effectuée par la répétition de cycles de dénaturation de l'ADN - hybridation des amorces - polymérisation, qui assure une amplification exponentielle de chaque brin (Fig. 3). Par cette méthode rapide, il est possible d'étudier spécifiquement tout ou partie d'un gène et d'obtenir de l'ADN en grande quantité. Des délétions d'exon peuvent ainsi être mises en évidence.

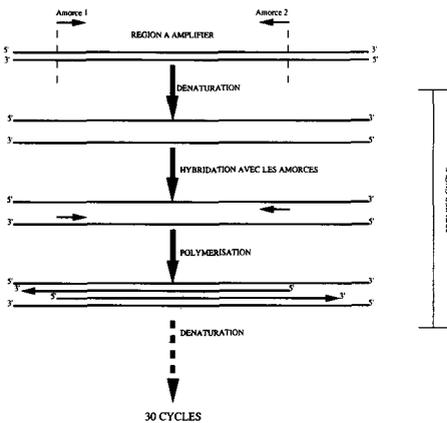


Fig. 3 : Schéma de l'amplification en chaîne par la polymérase (PCR). Au bout de 30 cycles il y aura théoriquement  $2^{30}$  copies de la région à amplifier.

Lorsque le gène est présent et la protéine non exprimée ou biochimiquement inactive, il est possible de coupler la PCR à des méthodes de recherches de mutations ponctuelles.

#### 4.3. PCR couplée à la DGGE.

L'électrophorèse sur gel d'acrylamide avec gradient de dénaturation (DGGE) est basée sur la dénaturation de la molécule d'ADN dans un gel avec un gradient croissant de dénaturation: dans une électrophorèse sur acrylamide, l'ADN double brin migre plus vite que l'ADN simple brin; de plus, les duplex où il existe un mauvais

appariement sont plus facilement dénaturés (par diminution du point de fusion) (3). La migration électrophorétique dans un gel où l'on crée un gradient croissant de dénaturation permet ainsi de discriminer un duplex plus stable d'un duplex moins stable (fig. 4): le premier est dénaturé plus tard et migre plus loin. Cependant, cette méthode ne peut être utilisée que s'il existe au moins deux domaines de fusion dans la molécule étudiée. En ajoutant un nombre variable de GC à l'une des extrémités des amorces, il est possible de créer artificiellement des domaines de fusion.

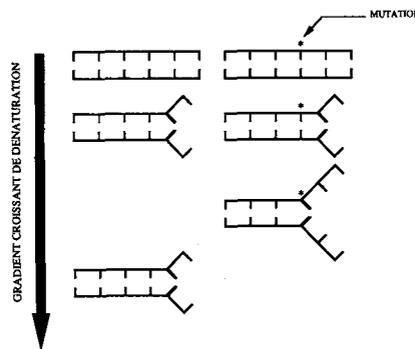


Fig. 4 : Principe de la détection d'une mutation par DGGE. La mutation (dans l'exemple, schématisée par\*) rend la molécule moins stable. Celle-ci, dénaturée plus tôt, migrera moins vite.

#### 4.4. PCR couplée à la SSCP

Le polymorphisme de conformation des simples brins d'ADN (SSCP) est basée sur la comparaison de migration relative de molécules d'ADN dénaturées dans un gel non dénaturant (7). Les simples brins d'ADN vont prendre une conformation tridimensionnelle dépendant de leur séquence. La migration dans le gel dépend de cette conformation dans l'espace (Fig. 5). Ainsi,

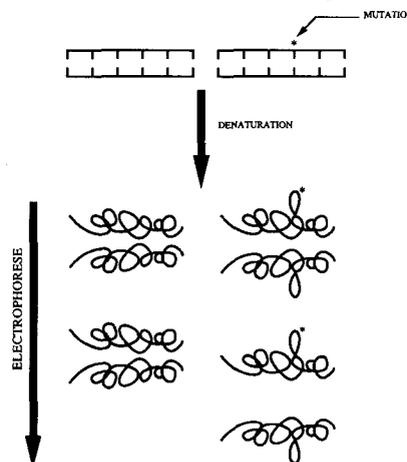


Fig. 5 : Principe de la détection d'une mutation par SSCP. Après dénaturation, la molécule d'ADN simple brin prend une conformation tridimensionnelle en fonction de sa séquence. L'anomalie de migration observée suggère la présence d'une mutation ponctuelle.

une anomalie de migration d'un échantillon suggère la présence d'une mutation ponctuelle.

#### 4.5. Séquençage

Lorsqu'une anomalie est détectée, le séquençage constitue la dernière étape pour mettre en évidence la mutation. La méthode la plus couramment utilisée est celle des di-désoxynucléotides (méthode de Sanger, 9). Le fragment à séquencer peut être obtenu après clonage ou après PCR.

#### 4.6. Expression

La découverte d'une mutation ponctuelle, si elle n'entraîne pas l'apparition d'un codon stop responsable d'une protéine tronquée, nécessite l'utilisation de vecteurs d'expression pour confirmer le rôle de la mutation. Ainsi, il est possible de déterminer si la protéine mutée *in vitro* a les mêmes caractéristiques que la protéine anormale *in vivo* (2). Après insertion du gène muté par mutagenèse dirigée dans un vecteur, le plasmide infecte une culture bactérienne (*E. coli*) résistante à certains antibiotiques. La colonie bactérienne va alors produire le gène muté en grande quantité. Après extraction, le gène est introduit dans des cellules encystées qui produisent la protéine mutée en grande quantité. A partir de là, il est possible de réaliser des immunoblots, western-ligand-blots, courbes de saturation. Par des expériences de co-transfection, il est possible de mettre en évidence une anomalie de transactivation d'une construction dans laquelle se trouve un gène rapporteur mis sous le contrôle de séquences régulatrices (2).

#### 4.7. Etude de l'ARNm

Les ARNm extraits sont séparés par électrophorèse en gel d'agarose dénaturant puis transférés sur filtre et enfin hybridés avec une sonde marquée (northern blot, 1). Sur l'autoradiogramme final, on observe une ou plusieurs bandes dont la taille, le nombre et l'intensité sont des indications qualitatives précieuses. C'est la technique la plus couramment utilisée.

La cartographie à la nucléase S 1 ou à la RNase A permet de rechercher des mutations ponctuelles sur l'ARNm sans passer par l'association transcriptase reverse-PCR.

#### REFERENCES

1. Alwine J, Kemp D, Stark G. Method for detection of specific RNAs in agarose gels by transfert to diazobenzylmethyl paper and hybridization with DNA probes. Proc. Natl. Acad. Sci., USA, 1977, 74:5350-5354.

## LEXIQUE (d'après 5 et 6)

2. Dalbadie-McFarlane G, Cohen L, Riggs A. Oligonucleotide-directed mutagenesis as a general and powerful method for studies of protein function. Proc. Natl. Acad. Sci., USA, 1982, 79:6409-6413.
3. Fischer S, Lerman L. DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. Proc. Natl. Acad. Sci., USA, 1983, 80:1579-1583.
4. Frezal J, Baule M-S, De Fougerolle T, Kaplan J, Le Merrer M. Genatlas. Montrouge, John Libbey Eurotext, 1989.
5. Kaplan J, Delpuch M. Biologie moléculaire et médecine. Paris, Flammarion Médecine Sciences, 1989.
6. Lewin B. Genes IV. New York, John Wiley and Sons Inc, 1989.
7. Orita M, Suzuki Y, Sekiya T, Hayashi K. Rapid and sensitive detection of point mutation and DNA polymorphisms using the polymerase chain reaction. Genomics, 1989, 5:87-879.
8. Saiki RK, Scharf S, Faloona F et al. Enzymatic amplification of  $\beta$  globin genomic sequences and restriction sites analyses for diagnosis of sickle cell anemia. Science, 1985, 230: 1350-1354.
9. Sanger F, Nicklen S, Coulson A. DNA sequencing with chain-termination inhibitors. Proc. Natl. Acad. Sci. USA, 1977, 74:5463-5467.
10. Southern E. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol, 1975, 98:503-507.

**RESUME :** Après un rappel succinct de la structure de l'ADN et de l'organisation du génome humain, les principales étapes de la transcription génique sont décrites. Dans une seconde partie, sont abordés les principaux outils de la biologie moléculaire utilisés pour l'étude des anomalies de l'information. L'utilisation des polymorphismes de restriction et l'amplification de régions spécifiques des gènes permettent l'enrichissement constant de la carte génétique chez l'homme par la découverte et la localisation de nouveaux gènes, et sont désormais appliquées au diagnostic génétique d'un nombre toujours plus grand de maladies. Enfin, l'élucidation des mécanismes moléculaires de la transcription génique ont ouvert la voie à l'étude des facteurs régulant l'expression des gènes.  
**Mots-clés :** génétique moléculaire - transcription génique - polymorphismes - recherche de mutations - expression génique. **Andrologie, 1992, 2, 104-107.**

**Acides nucléiques :** au nombre de deux, l'ADN (acide désoxyribonucléique) et l'ARN (acide ribonucléique). Ce sont des chaînes polynucléotidiques dont les éléments de bases, les nucléotides, sont constitués du groupement d'une base (adénine, guanine, cytosine, thymidine pour l'ADN et uracile remplaçant la thymidine pour l'ARN), d'un sucre (désoxyribose pour l'ADN et ribose pour l'ARN) et d'un groupement phosphate et sont liés les uns aux autres par des liaisons 5'-3' phosphodiester.

**ADNc :** séquence d'ADN complémentaire d'un ARNm obtenue par l'action d'une transcriptase reverse. Cette séquence correspond au gène sans les introns.

**Allèles :** versions alternatives d'un même gène différant par leur séquence nucléotidique. Les deux chromosomes homologues portent chacun un allèle de la séquence.

**Amplificateurs :** séquences activatrices particulières par leur possibilité de garder leur caractère activateur si elles sont inversées ou situées à distance du gène.

**Amorce :** courte séquence d'ADN ou d'ARN complémentaire du début d'une matrice servant de point de départ à son amplification par une ADN ou ARN polymérase.

**Carte de restriction :** alignement des sites de restriction existant sur une région précise de l'ADN.

**Cadre de lecture :** une des trois possibilités pour traduire une séquence nucléotidique en acide aminé. Lorsqu'il ne renferme pas de codon stop, on dit qu'il est ouvert.

**Cis-régulateur :** élément qui a un effet régulateur sur le taux de transcription d'un gène et qui est situé à proximité sur le même chromosome. Il peut s'agir d'un promoteur sur lequel viennent se fixer des protéines régulatrices (facteurs trans) ou des amplificateurs.

**Code génétique :** enchaînement des bases dans l'ADN qui contient l'information permettant de définir l'ordre des acides aminés sur la chaîne polypeptidique. Le code est lu par groupe de trois nucléotides ou codon.

**Crossing-over :** échange réciproque de matériel génétique entre chromosomes homologues, survenant en général au moment de méiose au niveau des chiasmata.

**Délétion :** perte d'une ou plusieurs paires de bases consécutives dans la molécule d'ADN.

**Enzymes de restriction :** endonucléases bactériennes coupant l'ADN au niveau de séquences parfaitement définies.

**Épissage :** mécanisme par lequel sont éliminés les introns au cours de la maturation des ARNm.

**Exons :** séquences d'ADN, codantes ou non, dispersées dans le gène et persistant sur l'ARNm mûr.

**Expression :** processus en plusieurs étapes par lequel l'ADN est transcrit en ARNm puis traduit en protéine.

**Facteur en trans :** facteur diffusible capable de moduler l'activité d'un ou plusieurs gènes.

**Hétérozygote :** sujet possédant deux allèles différents d'un même gène sur un locus déterminé.

**Homozygote :** sujet possédant deux allèles identiques d'un même gène sur un locus déterminé.

**Intron :** séquence d'ADN transcrite et éliminée par épissage au cours de la maturation des ARNm.

**Liaison génétique :** deux loci sont liés s'ils ségrègent ensemble à la méiose avec une fréquence plus élevée que celle attendue.

**Locus :** site d'une séquence d'ADN déterminée sur le chromosome.

**Méthylation :** au niveau des cystéines des motifs GC, associée à une diminution de la transcription des gènes.

**Polymorphisme :** variations individuelles de la séquence de l'ADN au niveau d'un locus pouvant donner la présence de deux allèles différents pour un même locus.

**Promoteur :** séquence d'ADN située en 5' d'un gène en amont des régions codantes et contenant des sites de fixation pour l'ARN polymérase et pour les facteurs de régulation de la transcription.

**Traduction :** mécanisme par lequel une molécule d'ARNm est traduite en protéine. Elle est effectuée dans les ribosomes.

**Transcriptase reverse :** enzyme permettant la rétrotranscription de l'ARN en ADN.

**Transcription :** mécanisme par lequel une molécule d'ARN est synthétisée à partir d'une matrice d'ADN.