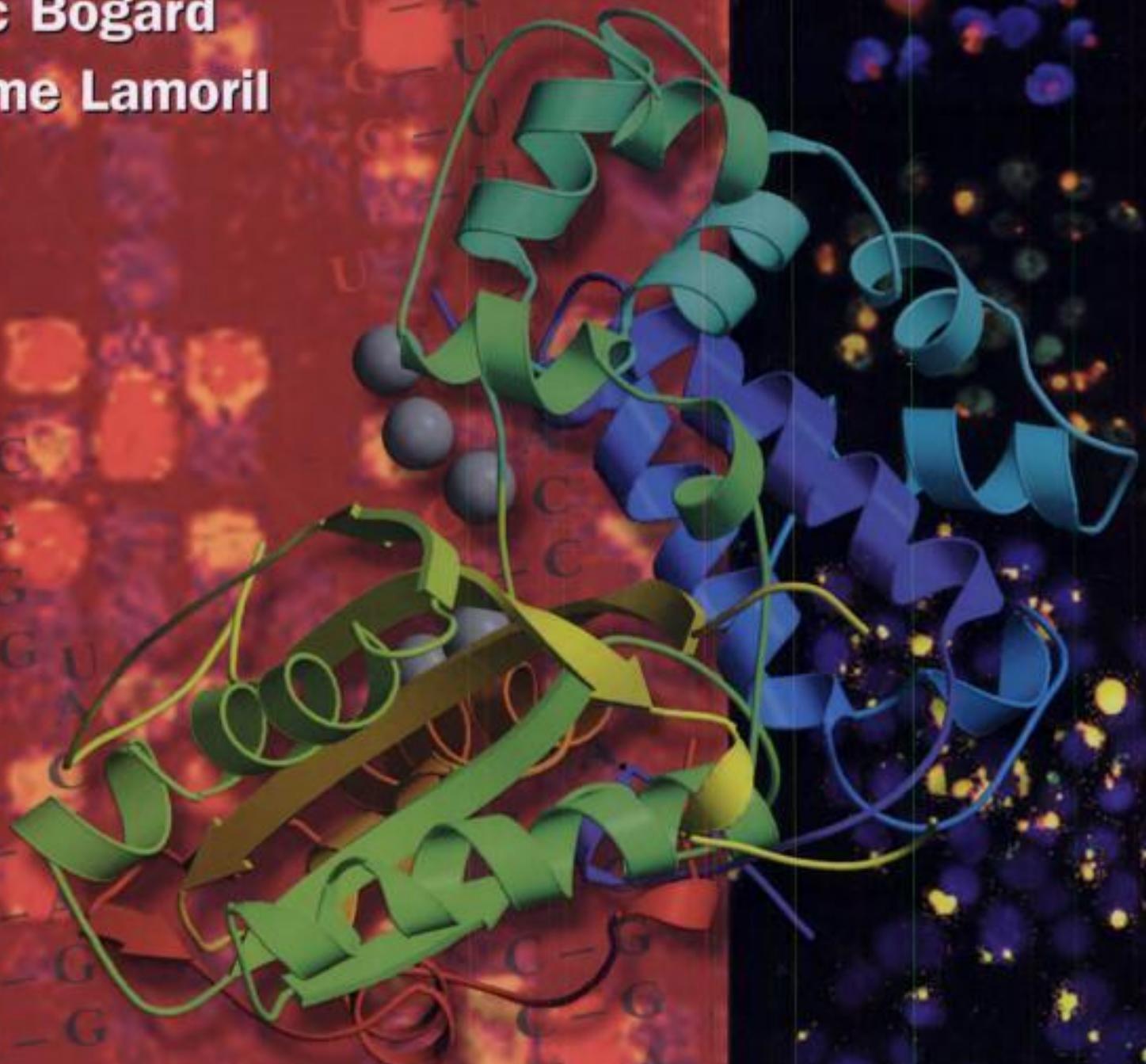


Campus  
référence

# Principes de biologie moléculaire en biologie clinique

Nedjma Ameziane  
Marc Bogard  
Jérôme Lamoril



**Nedjma Ameziane**, praticienne hospitalière, centre hospitalier de Sens, laboratoire de biologie et biologie moléculaire, ex-Inserm U479, 1, avenue Pierre-de-Coubertin, 89100 Sens

**Marc Bogard**, praticien hospitalier, centre hospitalier de Meaux, laboratoire de biochimie et biologie moléculaire, 6 et 8, rue Saint-Fiacre, 77100 Meaux

**Jérôme Lamoril**, praticien hospitalier, hôpital Louis-Mourier, laboratoire de biochimie et génétique moléculaire, Inserm U656, 178, rue des Renouilllets, 92700 Colombes

*Principes de biologie moléculaire en biologie clinique*

Responsable éditoriale : Marie-José Rouquette

Éditeur : Dragos Bobu

Chef de projet : Nathalie Morellato

Conception graphique et maquette de couverture : Véronique Lentaigne

Illustration de couverture : Transférase terminale. Figure extraite de Protein Data Bank IKDH, reproduite avec l'aimable autorisation de Protein Data Bank.

---

© 2006 Elsevier SAS. Tous droits réservés

23, rue Linois, 75724 Paris cedex 15

<http://france.elsevier.com>

---

L'éditeur ne pourra être tenu pour responsable de tout incident ou accident, tant aux personnes qu'aux biens, qui pourrait résulter soit de sa négligence, soit de l'utilisation de tous produits, méthodes, instructions ou idées décrits dans la publication. En raison de l'évolution rapide de la science médicale, l'éditeur recommande qu'une vérification extérieure intervienne pour les diagnostics et la posologie.

Tous droits de traduction, d'adaptation et de reproduction par tous procédés réservés pour tous pays. En application de la loi du 1<sup>er</sup> juillet 1992, il est interdit de reproduire, même partiellement, la présente publication sans l'autorisation de l'éditeur ou du Centre français d'exploitation du droit de copie (20, rue des Grands-Augustins, 75006 Paris).

*All rights reserved. No part of this publication may be translated, reproduced, stored in a retrieval system or transmitted in any form or by any other electronic means, mechanical, photocopying, recording or otherwise, without prior permission of the publisher.*

---

Photocomposition : SPI Publisher Services, Pondichéry, Inde

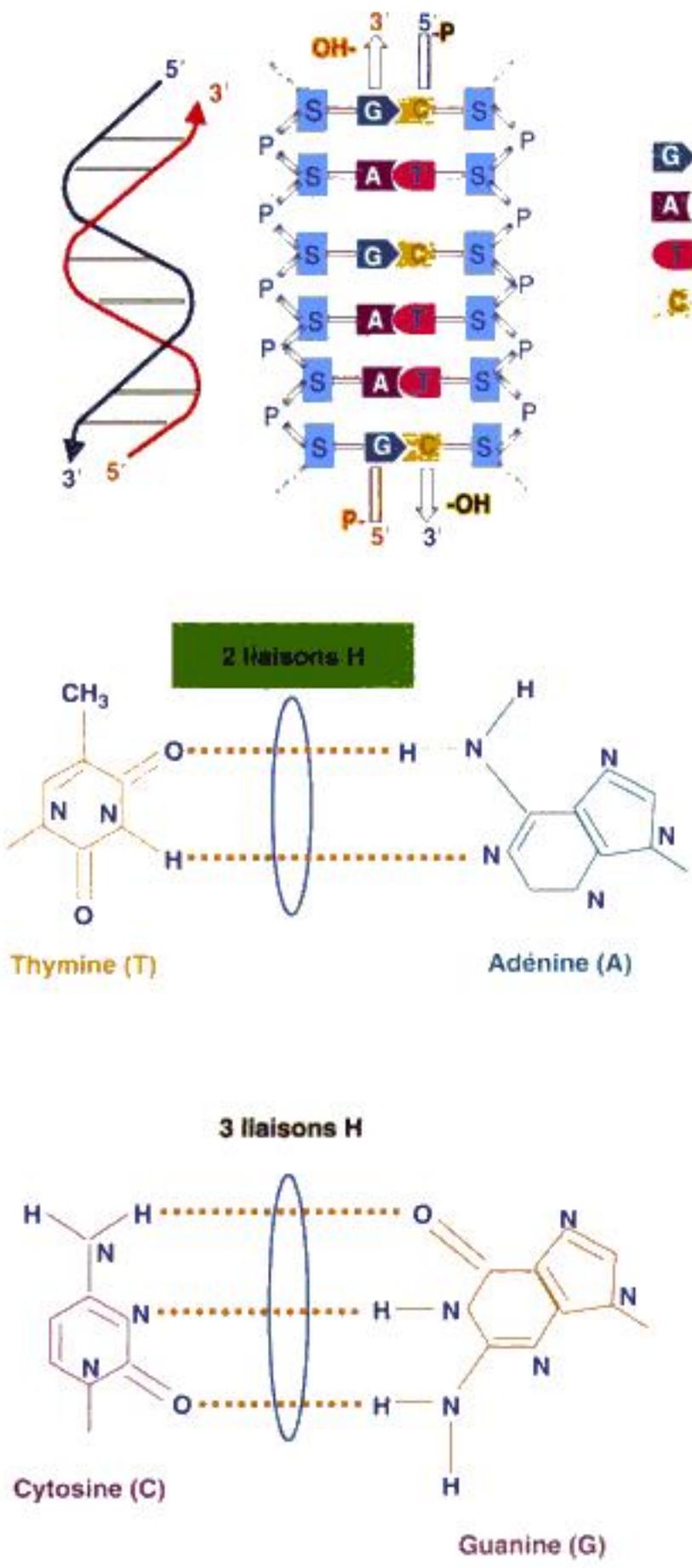
Imprimé en Italie par Priater Trento, 38100 Trento

Dépôt légal : Décembre 2005

ISBN : 2-84299-685-2

ISSN : 1768-2304

<b>Partie II. Les outils utilisés en biologie moléculaire</b>	<b>81</b>	Détection des hybrides marqués par des sondes	210
<b>3 Enzymes de restriction-modification</b>	<b>83</b>	Applications	211
Enzymes de restriction	84	Conclusion	211
Méthyltransférases	99	<b>9 Extraction-purification de l'ADN et de l'ARN</b>	<b>215</b>
Conclusion	105	Recueil du spécimen	216
<b>4 Polymérase</b>	<b>107</b>	Extraction des acides nucléiques	216
ADN polymérase	107	Automatisation de l'extraction-purification	225
ARN polymérase	123	Conclusion	226
Transcriptases inverses	125	<b>Partie III. Amplification génique</b>	<b>229</b>
Conclusion	126	<b>10 Polymerase chain reaction et autres systèmes d'amplification</b>	<b>231</b>
ADN polymérase et réparation de l'ADN	126	Limites de l'identification de séquences nucléiques par hybridation moléculaire	231
<b>5 Enzymes de modification</b>	<b>131</b>	Définition de l'amplification génique	232
Nucléases	131	<b>Amplification de cibles : la polymerase chain reaction</b>	<b>232</b>
Phosphatases alcalines	137	Amplification par PCR « classique »	232
Kinases	138	Amplification par PCR en temps réel	251
Transférase terminale	138	PCR quantitative	265
Enzymes intervenant dans la réparation de l'ADN	138	<b>Techniques alternatives d'amplification</b>	<b>285</b>
Enzymes et produits divers	150	Amplification par processus transcriptionnel : TMA, NASBA et 3SR	285
Conclusion	153	Amplification par déplacement de brin (SDA)	287
<b>6 Séparation des acides nucléiques</b>	<b>155</b>	Techniques d'amplification basées sur la rolling circle amplification	289
Électrophorèse	155	<i>Ligase chain reaction</i>	290
Méthodes de capture	167	<i>Cycling probe reaction</i>	292
Conclusion	170	ADN branché ou <i>branched DNA</i> (bDNA)	293
<b>7 Synthèse des oligonucléotides</b>	<b>173</b>	Système Hybrid capture®	294
Éléments constitutifs de la synthèse d'oligonucléotides	173	<b>Systèmes anticontamination</b>	<b>296</b>
Synthèse proprement dite	174	Méthodes physiques d'isolement	297
Caractéristiques physicochimiques des oligonucléotides	188	Destruction de l'ADN contaminant avant introduction de la cible	297
<b>8 Hybridation moléculaire</b>	<b>195</b>	Modification des amplicons afin de les différencier de la cible	298
Fusion de l'ADN	195		
Obtention et marquage des sondes	199		
Hybridation proprement dite	209		



**Figure 1.1** Schéma de l'acide désoxyribonucléique (ADN). **A.** ADN en structure double hélice orientée sens et antisens en 5' phosphorylé et 3' hydroxylé. G : guanine ; A : adénine ; T : thymine ; C : cytosine ; S : désoxyribose (sucre) ; P : acide phosphorique. **B.** Deux liaisons hydrogène relient A à T (A=T) et trois liaisons hydrogène relient C à G (C≡G).

Les deux chaînes d'ADN forment une hélice en établissant au niveau de leurs bases des liaisons de faible énergie, les liaisons hydrogènes. Par ailleurs, les deux chaînes d'ADN sont complémentaires. En effet, les liaisons hydrogènes se forment selon les règles d'appariement suivantes :

- la base « A » est associée avec la base « T » par deux liaisons hydrogènes (A=T) ;
- la base « G » est associée avec la base « C » par trois liaisons hydrogènes (G≡C).

Cette règle fondamentale est à la base des principales techniques de biologie moléculaire.

Chaque brin d'ADN possède une extrémité 5'-phosphate et une extrémité 3'-hydroxyle. Par convention, une orientation a été définie de l'extrémité 5' vers l'extrémité 3' (5'→3'). Les deux brins sont orientés en sens opposé : on dit qu'ils sont antiparallèles. Les liaisons hydrogènes étant de faible énergie, il est possible de séparer les deux brins d'ADN par chauffage ou par traitement chimique (par la soude, par exemple) : on parle alors de dénaturation.

Cette séparation est réversible et la réassociation s'appelle la renaturation ou hybridation (elle peut se faire ADN/ADN ou ADN/ARN).

Sur le plan structural, l'ADN existe sous trois principales formes hélicoïdales différentes : A, B (la plus fréquente) et Z (la plus rare), selon la séquence et la concentration ionique du milieu. Ces trois structures se différencient schématiquement par une torsion, un enroulement et une orientation différente et donc une accessibilité plus ou moins importante des bases.

**Structure de l'acide ribonucléique (ARN)**

L'acide ribonucléique (ARN) se distingue de l'ADN par son sucre, le D-ribose, et diffère d'une base pyrimidique, l'uracile (U) au lieu de la thymine (T) de structure proche. L'uracile s'apparie à l'adénine et forme deux liaisons hydrogènes avec elle (A=U).

Il existe trois grands types d'ARN :

- l'ARN messager (ARNm) : il provient de la transcription de l'ADN. Il sert de matrice pour la traduction en protéine ;
- l'ARN de transfert (ARNt) : il véhicule les acides aminés vers les ribosomes au cours de la traduction (on parle alors d'ARNt-aminoacyl). Chaque acide aminé fixé sur l'ARNt est associé à un anticodon complémentaire du codon correspondant (un acide aminé pouvant avoir plusieurs anticodons – voir le code génétique) ;
- l'ARN ribosomal (ARNr) : il s'associe à des complexes protéiques pour former le ribosome qui servira de support pour la synthèse des protéines lors de la traduction.

D'autres ARN jouent un rôle important. Nous les citerons brièvement :

- les petits ARN nucléaires de 100–190 pb (*small nuclear RNA*, snRNA) impliqués dans l'épissage du pré-ARNm et la maturation de l'ARNr ;

- les petits ARN nucléolaires d'environ 100 pb (*small nucleolar RNA*, snoRNA) impliqués dans les modifications et la maturation des ARNr ;
- les introns du groupe I (*group I intron*), ARN aux capacités d'auto-épissage présents dans les ARN 28S ;
- les microARN (*microRNA*, miRNA), ARN non codant d'environ 22 pb. Ces ARN interviennent dans la régulation post-transcriptionnelle et de la traduction ;
- les petits ARN d'interférence (*small interfering RNA*, siRNA), ARN issus d'ARN double brin et jouant un rôle dans la dégradation d'ARN dont ils sont complémentaires (voir le chapitre sur l'interférence d'ARN).

Les structures différentielles de l'ADN et de l'ARN sont résumées dans le tableau 1.1.

## Code génétique

La structure primaire de l'ADN (succession des bases le long de brin de l'hélice, appelée encore séquence de l'ADN) porte l'information génétique et constitue le génome.

L'information génétique des parties dites codantes est décryptée puis traduite en protéine dont le squelette est constitué d'acides aminés. Il n'existe que 20 acides aminés pouvant être intégrés au sein des protéines. Ceci impose que l'enchaînement des bases constituant la séquence soit lu par triplet, une lecture en doublet ne donnant que 16 combinaisons possibles ( $4^2$ ). Le triplet de base (appelé aussi codon) code donc pour un acide aminé, molécule de base composant les protéines. La relation triplet/acide aminé porte le nom de code génétique (tableau 1.2). Le nombre de codons possibles  $4^3$  (soit 64) étant plus important que le nombre d'acides aminés existant (20 acides aminés), le code génétique permettant d'interpréter la relation entre un triplet et l'acide aminé correspondant est dit dégénéré.

Ce terme signifie que plusieurs codons codent pour le même acide aminé. Deux acides aminés seulement possèdent un seul codon, la méthionine et le tryptophane.

En sus des triplets codant pour les acides aminés, il existe :

- un codon initiateur ATG à partir duquel commence la traduction (signal d'initiation de la traduction) et qui code aussi pour un acide aminé, la méthionine (appelée encore méthionine initiatrice quand il s'agit du codon correspondant au signal de début de la traduction) ;
- trois codons particuliers, les « codons-stop » (TAA, TAG, TGA), signaux de fin de traduction.

La succession de plusieurs codons sur l'ADN définit la partie codante d'un gène. Les codons seront traduits en acides aminés dont l'association formera le « squelette » d'une protéine. Le tableau 1.2 rappelle le code génétique chez l'être humain.

## Biais de codon – Variantes du code génétique

Le code génétique est universel. Néanmoins, selon les espèces, on observe un biais d'utilisation des codons, un ou plusieurs codons pour un même acide aminé (codons synonymes) étant plus utilisés que d'autres (tableau 1.3). Cette sélection des codons est le plus souvent corrélée à l'efficacité de l'expression des gènes.

## Description générale du génome humain

### Génome humain

Le génome humain est constitué de deux génomes distincts, le génome nucléaire situé dans le noyau de la cellule (souvent appelé par défaut le génome humain) et le génome mitochondrial localisé dans les mitochondries. Dans le noyau, le génome humain est constitué d'environ 3,2 milliards de paires de nucléotides (on dit aussi paires de bases, pb). Le nombre de gènes codant pour des protéines n'est pas connu avec précision. Il est estimé actuellement entre 20 000 et 25 000. Ce nombre n'inclut pas les gènes codant pour des ARN non codant (tels que les ARN de transfert, ribosomiaux, les microARN et les petits ARN nucléolaires, snoARN). Événement majeur dans l'histoire de l'humanité, la publication de la quasi-totalité du génome humain date de février 2001. À ce jour, on connaît la succession des nucléotides sur 99 % du génome humain. La quasi-totalité du génome a été identifiée en 2003 (annonce du 14 avril 2003). Mais le décryptage du génome humain est encore loin d'être réalisé.

De manière simplifiée, sur le génome, on peut distinguer la partie codante de l'ADN constituée des gènes puis les pseudogènes et les séquences non codantes (figure 1.2). Seule une faible partie du génome (environ 5 %) code pour des protéines. Une partie minime (environ 0,5 %) est constituée de pseudogènes. Le reste du génome est constitué d'introns et d'ADN intergénique. Par ailleurs, l'étude de la répartition des gènes a montré que leur densité est très variable d'un chromosome à

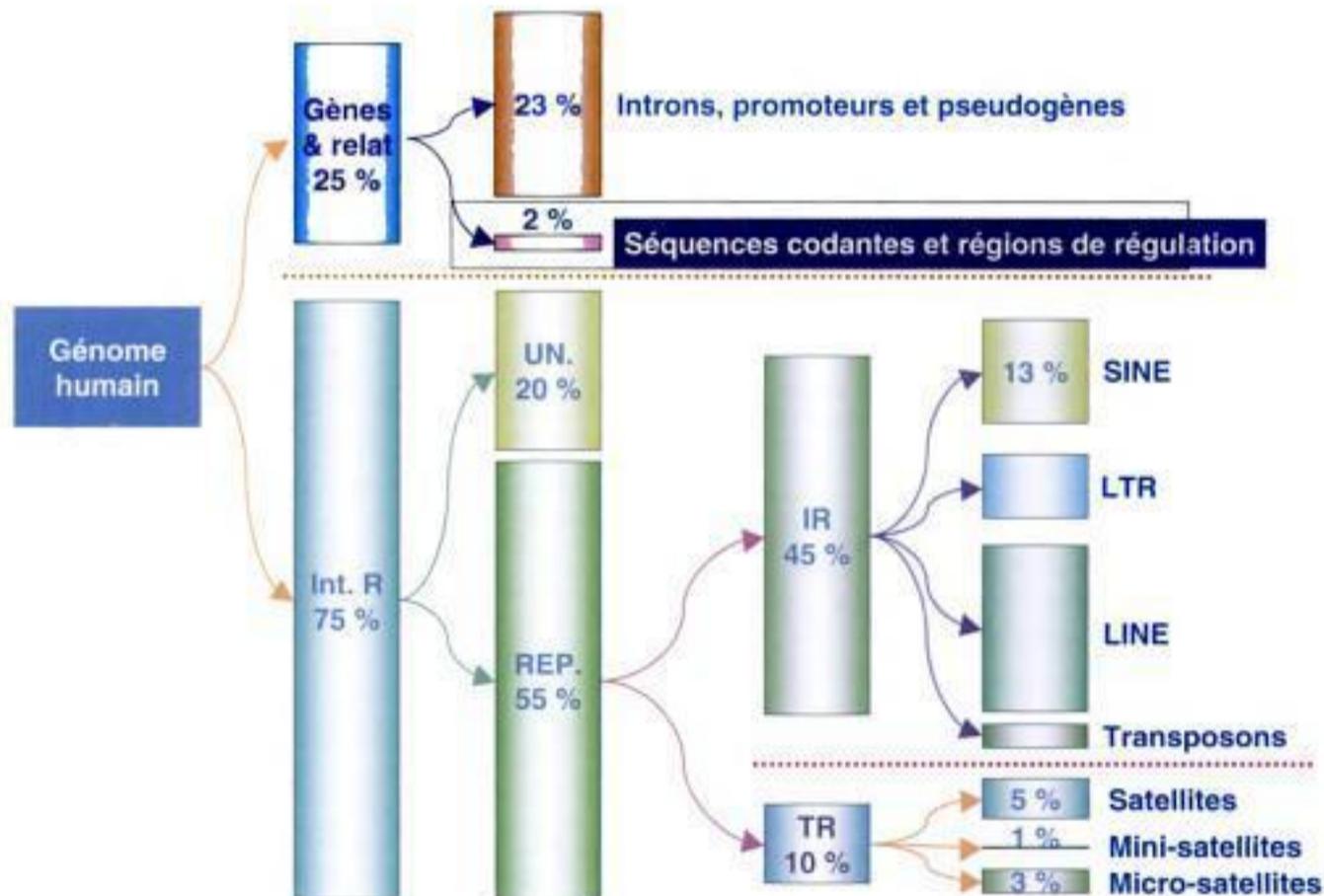
**Tableau 1.1. Différence structurale entre l'ARN et l'ADN**

Constituants	ARN (simple brin)	ADN (double brin)
Sucre	Ribose	Désoxyribose
Acide	Phosphorique	
	Bases	
Puriques	Adénine (A) Guanine (G)	
Pyrimidiques	Cytosine (C) Uracile (U)	Thymine (T)

Tableau 1.2. Code génétique

		U			C			A			G		
<b>U</b>	UUU	<b>Phe</b>	UCU	<b>Ser</b>	UAU	<b>Tyr</b>	UGU	<b>Cys</b>	<b>U</b>				
	UUC		UCC		UAC		UGC			<b>C</b>			
	UUA	<b>Leu</b>	UCA		UAA	<b>Stop</b>	UGA				<b>Stop</b>	<b>A</b>	
	UUG		UCG		UAG		UGG						<b>Trp</b>
<b>C</b>	CUU	<b>Leu</b>	CCU	<b>Pro</b>	CAU	<b>His</b>	CGU	<b>Arg</b>	<b>U</b>				
	CUC		CCC		CAC		CGC			<b>C</b>			
	CUA		CCA		CAA	<b>Gln</b>	CGA				<b>A</b>		
	CUG		CCG		CAG		CGG					<b>G</b>	
<b>A</b>	AUU	<b>Ile</b>	ACU	<b>Thr</b>	AAU	<b>Asn</b>	AGU	<b>Ser</b>	<b>U</b>				
	AUC		ACC		AAC		AGC			<b>C</b>			
	AUA		ACA		AAA	<b>Lys</b>	AGA				<b>Arg</b>		
	AUG	<b>Met</b>	ACG		AAG		AGG					<b>G</b>	
<b>G</b>	GUU	<b>Val</b>	GCU	<b>Ala</b>	GAU	<b>Asp</b>	GGU	<b>Gly</b>	<b>U</b>				
	GUC		GCC		GAC		GGC			<b>C</b>			
	GUA		GCA		GAA	<b>Glu</b>	GGA				<b>A</b>		
	GUG		GCG		GAG		GGG					<b>G</b>	

Mode d'emploi du tableau : on désire par exemple connaître l'acide aminé correspondant au codon ACU (ACT sur l'ADN). On procède comme suit : on prend la colonne 1 à gauche (exemple A), puis la ligne 2 du milieu (exemple C) puis la colonne 3 (exemple U). L'intersection des trois donne ACU soit Thr (thréonine).



**Figure 1.2** Composition générale du génome humain. Le pourcentage représente la quantité de séquences par rapport à la séquence totale connue du génome. Gènes & relat : gènes et séquences associées ; Int. R : régions intergénomiques ; Un. : séquences intergénomiques uniques ; Rep. : séquences intergénomiques répétitives ; IR : séquences intergénomiques répétitives dispersées ; TR : séquences intergénomiques répétées en tandem.

**Tableau 1.3. Codon(s) « préféré(s) » selon l'organisme**

Acide aminé	Nombre de codons possibles	Codon(s) préféré(s) chez				
		Mammifères	Plantes	Levures	Bactéries Gram +	Bactéries Gram -
Alanine	4	GCC	GCU	GCU	GCG	GCA
Arginine	6	AGG, AGA	AGA	AGA	CGC, CGU	AGA
Asparagine	2	AAC	AAC	AAU	AAC	AAU
Acide aspartique	2	GAC	GAU	GAU	GAU	GAU
Cystéine	2	UGC	UGC	UGU	UGC	UGC
Acide glutamique	2	GAG	GAA	GAA	GAA	GAA
Glutamine	2	CAG	CAA	CAA	CAG	CAA
Glycine	4	GGC	GGA, GGU	GGU	GGC	GGA
Histidine	2	CAC	CAU	CAU	CAC, CAU	CAU
Isoleucine	3	AUC	AUU	AUU	AUC	AUU
Leucine	6	CUG	UUG, CUU	UUG	CUG	UUA
Lysine	2	AAG	AAG	AAA	AAA	AAA
Méthionine	1	AUG	AUG	AUG	AUG	AUG
Phénylalanine	2	UUC	UUC	UUU	UUC	UUU
Proline	4	CCC	CCA	CCCA	CCG	CCU
Sérine	6	AGC	UCU	UCU	AGC	UCU, UCA
Thréonine	4	ACC	ACU	ACU	ACC	ACA
Tryptophane	1	UGG	UGG	UGG	UGG	UGG
Tyrosine	2	UAC	UAC	UAC	UAU	UAU
Valine	4	GUG	GUU	GUU	GUG	GUU
Codon-stop	3	UGA	UGA	UAA	UAA	UAA

Dans la mitochondrie, le code est modifié : UGA : tryptophane ; AUA : méthionine ; AGA et AGG : codons-stop.

Récemment, deux nouveaux acides aminés atypiques ont été décrits :

– la sélénocystéine : cet acide aminé a été décrit chez les eucaryotes et les procaryotes. Dans certains gènes, le codon UGA peut coder pour cet acide aminé en présence d'un complexe traductionnel spécifique. Certaines enzymes par exemple possèdent cet acide aminé telles que la formate déhydrogénase ou la glutathione peroxydase ;

– la pyrrolysine : cet acide aminé a été décrit chez certaines bactéries. Le codon UGA peut dans certains cas et à l'aide d'un mécanisme spécifique donner naissance à cet acide aminé.

l'autre. Par exemple, les chromosomes 17, 19 et 22 sont relativement riches en gènes contrairement aux chromosomes 4, 8, 13, 18 et Y.

## Chromosomes

L'ADN est compacté dans le noyau en ensembles, les chromosomes. On distingue 23 paires de chromosomes dans la cellule humaine, chacun ayant une composition en séquence caractéristique et donc unique. Vingt-deux paires sont aussi appelées autosomes et la vingt-troisième paire est constituée par les chromosomes sexuels X et Y. Les chromosomes ne présentent pas la même taille, le plus petit étant le chromosome 21 et le plus grand le chromosome 1 pratiquement cinq fois plus gros (tableau 1.4).

## Télomères

Dans une cellule eucaryote normale, après environ 40 à 60 cycles de réplication, la cellule meurt. Un des mécanismes expliquant cette mort naturelle est le raccourcissement des télomères. Les télomères sont les extrémités terminales des chromosomes indispensables pour préserver l'intégrité du matériel génétique au cours des divisions cellulaires. Leur rôle est multiple tel que :

- la protection de l'extrémité des chromosomes de la dégradation exonucléolytique ;
- la prévention d'une recombinaison chromosomique aberrante ;
- la ségrégation des chromosomes durant la mitose et la méiose ;
- l'attachement des chromosomes à la matrice nucléaire.

**Tableau 1.4. Taille des chromosomes humains**

Chromosome	Taille (Mpb)
1	249
2	237
3	192
4	183
5	174
6	165
7	153
8	135
9	132
10	132
11	132
12	123
13	108
14	105
15	99
16	84
17	81
18	75
19	69
20	63
21	54
22	57
X	141
Y	60

L'ADN télomérique est riche en séquences répétées en tandem (environ 1000 séquences) 5'-TTAGGG. Le télomère se termine par une extrémité 3' simple brin riche en G. Cette extrémité est associée à des complexes protéiques formant un chapeau « protecteur ». Par sa fonction de protection du chromosome, le télomère empêche la dégradation de l'extrémité chromosomique et la fusion des extrémités. L'érosion progressive des extrémités à chaque division cellulaire provoque la perte de protection des télomères avec instabilité, fusion des extrémités, l'ensemble aboutissant le plus souvent à l'apoptose. Les ADN polymérases habituelles ne pouvant répliquer efficacement les extrémités 3' des télomères, il en résulte une diminution de la longueur des télomères. À titre d'exemple, en culture cellulaire, on observe une diminution d'environ 100 pb par division cellulaire jusqu'à perte de viabilité des cellules. En effet, de manière générale, les cellules somatiques normales ne possèdent pas d'activité télomérase après la naissance. Cette activité est cependant retrouvée dans les cellules à renouvellement rapide telles que les cellules de la moelle osseuse et les leucocytes par exemple.

Deux classes cellulaires, les cellules de la lignée germinale et les cellules embryonnaires à un stade précoce présentent cependant un mécanisme protecteur évitant le raccourcissement des télomères. L'activité télomérase (voir infra) persiste durant la vie intra-utérine dans la majorité des tissus et est rapidement perdue à la naissance dans les tissus somatiques (sauf les leucocytes et la moelle osseuse). Une enzyme spécifique appelée télomérase est dédiée à la répllication des télomères. Cette télomérase est en fait composée d'une sous-unité catalytique, d'une transcriptase inverse et d'un ARN servant d'amorce. Elle participe en association avec un complexe protéique à la régulation et la réaction de répllication. L'enzyme catalyse l'addition de séquences TTAGGG aux extrémités 3'. La télomérase compense ainsi les pertes de séquences qui se produiraient si cette répllication s'effectuait à partir de la machinerie cellulaire « conventionnelle ». Son dysfonctionnement peut avoir des conséquences importantes. Par exemple, dans les cellules somatiques où l'activité télomérase est absente, la maintenance anormale de cette activité aboutit à des cellules rendues quasiment immortelles et participe au phénomène de cancérogenèse.

### Centromère

En cytologie, sur le chromosome, on reconnaît physiquement une zone de constriction se traduisant par un raccourcissement de l'épaisseur de la chromatine. Cette région est appelée centromère. Ce centromère joue un rôle fondamental puisqu'il est le site de fixation des microtubules de tubuline formés au cours de la division de la cellule. Cela aboutit à l'alignement correct des chromosomes en métaphase et à la ségrégation correcte des chromosomes dans les cellules filles au cours de l'anaphase. Une perte de centromère aboutit à une instabilité chromosomique. Plutôt que centromère, terme utilisé par les cytogénéticiens et cytologistes, nous parlerons de domaine centromérique qui englobe le centromère proprement dit et la région adjacente. Ce domaine est constitué d'ADN satellites, ensemble de séquences répétées comme, par exemple, la famille ADN alphasatellites, unités répétées d'environ 170 pb spécifique, de chromosomes. Dans cette région, on trouve aussi des rétrotransposons comme, par exemple, des éléments L1 (rétrotransposons de type LINE, voir infra).

### Variants non pathologiques du génome : les polymorphismes

Chez l'homme, entre deux individus non apparentés, il existe de petites variations de séquences, estimées entre 1 et 2 %, considérées comme non pathologiques, appelées polymorphismes. Ces polymorphismes génotypiques peuvent se trouver dans des régions codantes ou non codantes

d'un gène ou dans d'autres régions du génome. Au cours de l'étude de pathologies diverses (maladies génétiques, cancers) ou dans d'autres cas (épidémiologie bactérienne), ils peuvent servir de marqueurs. Bien que considérés comme non pathologiques, dans certains cas, ces polymorphismes peuvent avoir un rôle fonctionnel et participer aux mécanismes physiopathologiques (voir le chapitre « Principes de pathologie moléculaire humaine »).

On distingue schématiquement deux classes de polymorphismes.

### Polymorphismes bi-alléliques

Un polymorphisme peut en effet se caractériser par la présence d'un nucléotide différent sur chacun des deux chromosomes d'une même paire (par exemple, A sur un allèle et G sur l'autre allèle). On parle de polymorphisme bi-allélique (ou *single nucleotide polymorphism*, SNP).

### Polymorphismes multi-alléliques

Certains polymorphismes sont de courtes séquences répétées de quelques bases à plusieurs centaines de bases qui peuvent varier sur chacun des deux chromosomes. En effet, sur chaque chromosome, le nombre de répétitions peut ne pas être le même. On dit qu'il y a hétérozygotie pour le nombre de séquences répétées. Le nombre de variations de la séquence pouvant être différent entre deux individus, on parle alors de polymorphisme multi-allélique. Schématiquement, parmi les polymorphismes multi-alléliques, suivant leur taille, on peut distinguer :

- les séquences répétées « simples » : ce sont les mono-, di-, tri-, tétra- et pentanucléotides (par exemple :  $[AC]_n = 5'-AC AC AC... \text{ répétées } n \text{ fois}$ ). On les appelle encore STR (*short tandem repeats*). On estime leur nombre entre 50 000 et 100 000 copies dans le génome ;
- les séquences en tandem (*variable number of tandem repeats* ou VNTR) : ce sont des unités répétées en tandem, de nombre variable. Parmi ces VNTR, on distingue schématiquement :
  - les SINE (*short intersped repetitive nucleotides elements*), courtes séquences répétées de moins de 500 bp ; une séquence particulière parmi les SINE mérite une brève et schématique description, la séquence Alu,
  - les séquences Alu sont réparties dans le génome au niveau des introns le plus souvent. En moyenne, on considère qu'une séquence Alu est présente tous les 4 à 5 kb. En fait, les séquences Alu sont réparties en groupe, séparées par quelques centaines de bases d'ADN « non-Alu ». Ces séquences sont composées typiquement de deux parties homologues « droite et gauche » distinctes, dérivées d'un gène, le gène ARN 7SL,

- les LINE (*long intersped repetitive nucleotides elements*), séquences plus longues de quelques milliers de bp (environ 6 à 7 kb),
- les ADN « satellites » : on distingue classiquement les satellites  $\alpha$  et  $\beta$ . Les satellites  $\alpha$  sont situés au niveau des centromères. Ce sont des séquences répétées en tandem, de 171 bp constituant entre 1 et 3 % des chromosomes. Les satellites  $\beta$  sont des séquences répétées en tandem de 68 bp situées dans l'hétérochromatine des chromosomes acrocentriques et du chromosome 9. Ces séquences sont encore mal connues.

### Définition d'un gène

De manière générale, un gène est un segment d'ADN contribuant à une fonction ou un phénotype. Il constitue un ensemble de séquences nucléiques contenant l'information nécessaire pour la production d'un ARN (transcription) et/ou d'un polypeptide ou d'une protéine (traduction). Un gène peut cependant coder pour plusieurs protéines (par exemple, par épissage alternatif) et peut, dans certains cas, produire des transcrits antisens. La taille d'un gène est très variable, de 2000 pb à plus de 2 millions de pb.

### Nomenclature des gènes humains

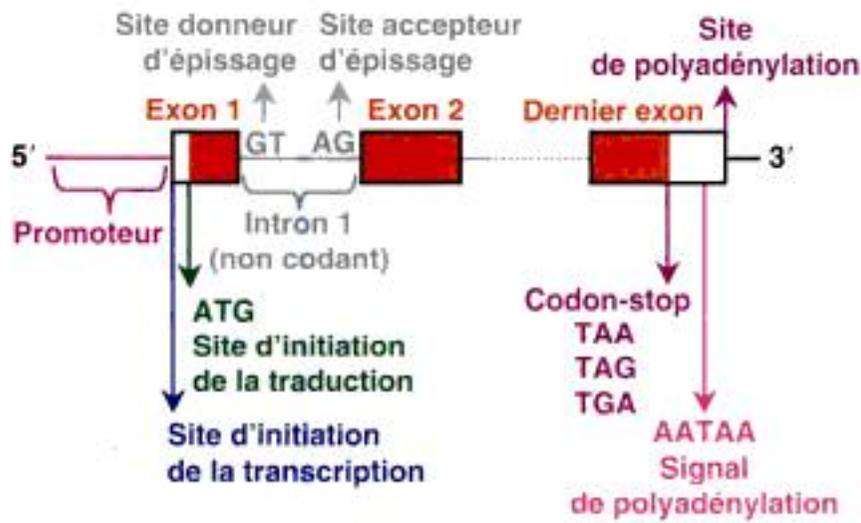
La nomenclature des gènes est en évolution permanente. Elle est évaluée par un comité au sein du projet HUGO (HUMAN Genome Organisation), organisme international aidant à la recherche sur le génome humain (voir le chapitre « Sites internet de biologie moléculaire »).

### Description générale d'un gène

Le gène d'un eucaryote est morcelé en fragments codants, les exons (dont la taille varie en moyenne entre 50 et 200 pb), séparés en général par des séquences non codantes, les introns. En amont du gène, se trouve une séquence régulatrice, le promoteur (figure 1.3).

### Promoteur

En amont du gène en 5', se trouve la région promotrice ou promoteur, séquence régulatrice de la transcription du gène, d'environ 800 à 1 kb. Cette séquence d'ADN, adjacente au site d'initiation de la transcription, est position- et orientation-dépendante. En effet, le promoteur définit le site d'initiation de la transcription et sa direction. De nombreuses séquences régulatrices de l'expression sont contenues dans cette région. Elles peuvent être soit tissu-spécifiques, soit ubiquitaires dans l'organisme. Dans ce dernier cas, on parle encore de gènes de ménage (*house-keeping*). Comme exemples de sites régulateurs, on



**Figure 1.3** Organisation générale d'un gène eucaryote. Le gène est orienté de 5' vers 3'. Les exons sont représentés par des rectangles et les parties foncées sont les parties codantes au sein des exons. En amont des exons, se trouve le promoteur et entre les exons se trouvent les introns.

peut citer la « TATA-box », 5'-TATAAA, et le site Sp1, 5'-GGCGG. Des facteurs protéiques, dits facteurs de transcription, interagissent entre eux et avec les séquences régulatrices, modulant de manière le plus souvent complexe l'expression du gène.

## Gène

Un gène est une entité discontinue dans laquelle les parties codantes (trouvées dans les exons) sont en général séparées entre elles par des parties non codantes (appelées introns) éliminées au cours de la maturation de l'ARNm.

Un gène commence en 5' de l'ADN par une séquence qui ne sera pas transformée en protéine (partie 5' non codante du gène, 5'NC). Le début du gène (situé sur l'exon 1) n'est pas traduit : la partie 5'NC intervient dans l'initiation de la transcription et la fixation d'un « chapeau » (*capping*) 5'-m<sup>7</sup>GpppX sur l'ARN messenger, indispensable pour sa traduction ultérieure.

En 3' de ce site d'initiation (de quelques dizaines à quelques centaines de bases après), le codon ATG signale le site d'initiation de la traduction, c'est-à-dire le début de la partie codante du gène. Ce site d'initiation est souvent présent sur l'exon 1. Il peut cependant être présent sur un autre exon. Par ailleurs, certains gènes contiennent plusieurs sites d'initiation de la traduction (sites alternatifs de traduction). Ces sites alternatifs peuvent être par exemple tissu-spécifiques. Puis vient une succession de séquences, soit codantes et conservées dans l'ARN (les exons), soit non codantes et éliminées lors de la transcription (les introns).

En 3' du gène, sur le dernier exon, un des trois codons-stop TAA, TAG, TGA signale l'arrêt de la traduction.

Après le codon-stop, se trouve la partie non codante du dernier exon (3' non codant, 3'NC). Pour un grand

nombre de gènes, cette région joue un rôle important, plus particulièrement dans la régulation post-transcriptionnelle du gène en contrôlant par exemple soit sa localisation dans la cellule, soit sa stabilité ou dans l'initiation de la traduction. Par ailleurs, la stabilité des ARNm dans la partie 3' non codante est régulée par certaines séquences telles des zones riches en éléments 5'-AU (*AU-rich elements*, ARE, par exemple la séquence 5'-AUUUA).

Dans la région 3'NC, se trouve aussi une séquence, signal de polyadénylation : 5'-AATAAA. Il peut cependant en exister plusieurs correspondant à des ARNm de taille différente. Ils peuvent être « tissu-spécifiques ». La régulation de ces sites alternatifs est encore mal connue. Ce signal de polyadénylation représente un site de reconnaissance pour la coupure de l'ARNm primaire. Elle est indispensable pour la stabilité de l'ARNm. Voir infra pour une description plus précise de la région 3'NC.

La taille des exons et leur nombre sont variables selon les gènes. La taille varie de 12 pb à environ 17 000 pb. Enfin, certains gènes ne présentent qu'un seul exon et aucun intron (gènes *intronless*).

## Introns

Certains introns jouent un rôle important dans la régulation de l'expression d'un gène (s'il y a présence d'un *enhancer* qui stimule l'expression d'un gène ou d'un *silencer* qui diminue voire inhibe l'expression d'un gène, par exemple, voir infra). Dans certains cas, un intron peut aussi contenir un gène (on parle alors de gène « niché », *nested gene*, « un gène dans un gène »). C'est le cas par exemple d'un gène appelé F8A [Omim 305423], *FVIII associated gene 1*, situé dans le vingt-deuxième intron du gène codant pour le facteur VIII de la coagulation. Ce gène possède aussi une autre caractéristique, il est dans un sens de transcription opposé à celui du facteur VIII. De façon générale, les introns n'existent pas chez les virus et les bactéries. Ils ne sont présents que chez les eucaryotes.

## Familles de gènes

Le décryptage progressif du génome humain montre que son organisation et sa régulation sont complexes. Ainsi, bien qu'un grand nombre de gènes soient transcrits dans le sens 5'→3', certains gènes sont transcrits à partir du brin opposé (le brin antisens). Certains gènes se trouvent en tandem et d'autres en sens opposé. Enfin, certains gènes se chevauchent (un gène commence « dans » un autre gène).

Par ailleurs, les scientifiques, au cours du décryptage (non achevé actuellement), constatent qu'un grand nombre de gènes peuvent être regroupés selon la similitude de séquence en un amas de gènes (on parle de

cluster). Cette parenté peut correspondre à de courtes séquences appelées motifs, de fonction semblable, ou à de longues séquences se rapportant à des gènes dont les produits possèdent des propriétés analogues (par exemple, les gènes codant pour certaines histones localisées sur le chromosome 6). Ces similitudes, outre leur classification à tel ou tel groupe de gènes (famille de gènes), permettent aussi, par comparaison entre les espèces, de comprendre, regrouper ou déduire le fonctionnement de telle ou telle classe de gènes. Des logiciels ont été développés pour aider au regroupement par famille, classe ou fonctionnement de gènes.

**Gènes chevauchant**

On parle de gènes chevauchant quand les exons d'un gène sont contenus dans les introns d'un autre gène (*overlapping gene groups*, OGG). Souvent, ces gènes sont codés sur les brins opposés (sens et antisens) de l'ADN. Il semblerait que ces gènes soient plus sensibles aux réarrangements géniques et que ces régions soient des points chauds de mutation.

**Pseudogènes**

Le pseudogène de structure proche du gène auquel il est rattaché ressemble structurellement à un gène fonctionnel. Mais ce n'est pas un gène, car il n'est pas traduit en protéine. On estime actuellement à environ 20 000 le nombre de pseudogènes dans le génome humain. Chez les procaryotes, les pseudogènes sont en nombre variable. En moyenne, il semblerait que le nombre de pseudogènes constitue environ 1 à 5 % du génome total (avec cependant des cas particuliers comme, par exemple, *Mycobacterium leprae* chez qui on dénombre 1100 pseudogènes pour 1600 gènes). Le pseudogène présente quelques caractéristiques générales :

- la séquence du pseudogène est proche de la séquence de l'ARN messager codant pour le gène normal (pouvant aller jusqu'à 98 % d'homologie de séquence). Il ne possède pas d'introns et présente des variations de séquence. En effet, la séquence contient des mutations ponctuelles, des insertions ou des délétions. La plupart du temps, ces mutations entraînent la formation de codons-stop ne permettant pas la transcription ;
- dans certains cas, cependant, le pseudogène peut être transcrit en ARN (mais en faible quantité). Dans ce cas, il ne peut être traduit. Aucune protéine fonctionnelle n'est synthétisée. À ce jour, un seul cas de pseudogène fonctionnel a été décrit. Il s'agit d'un pseudogène codant pour un ARN à fonction régulatrice. La possibilité de transcription d'un pseudogène en ARN peut poser parfois des problèmes au cours de l'étude de certains gènes présentant un pseudogène transcrit. La transcriptase inverse PCR (RT-PCR) peut dans ce cas donner des

résultats erronés dans la recherche de mutations sur les gènes correspondants (exemple du gène codant pour la protéine cytokératine 19 [KRT19, Omim 148020] dont le pseudogène est aussi transcrit) ;

- sa localisation peut être située sur un autre chromosome que le gène correspondant.

Certains auteurs distinguent quatre grandes classes de pseudogènes :

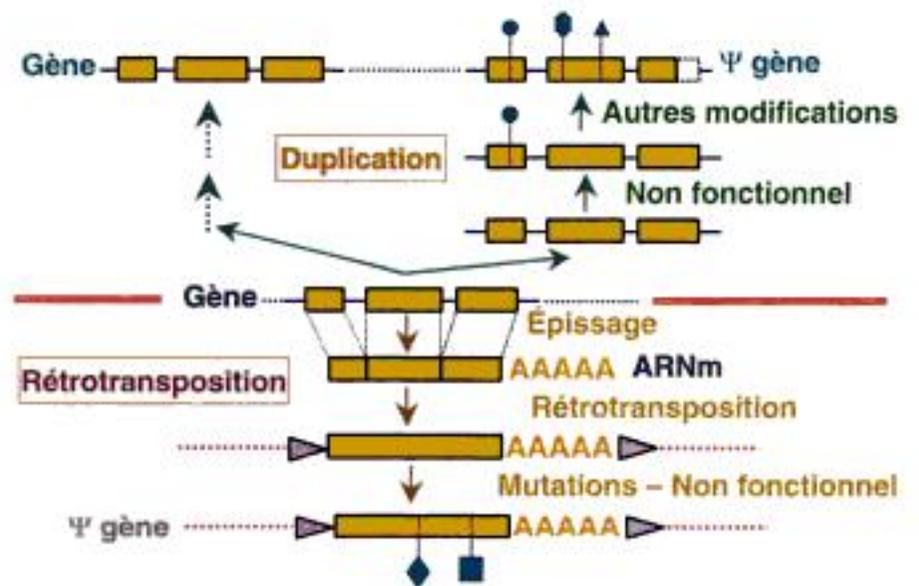
- les pseudogènes complets ou quasi-complets ;
- les exons ou groupes d'exons isolés, en dehors d'un cluster de gènes ;
- les exons « échappés » proche d'un gène parent ou d'un cluster de gènes ;
- les exons internes, dupliqués ou partiels, situés dans des gènes.

Cette classification n'est cependant pas encore acceptée sur le plan international.

Sur le plan conventionnel, un pseudogène est représenté par la lettre grecque psi ( $\Psi$ ), par exemple  $\Psi$ PGK-1 (pseudogène dérivé du gène PGK-1 [phosphoglycérate kinase 1]) ou par le suffixe P, par exemple CYP21P (pseudogène dérivé du gène CYP21). Les pseudogènes sont observés aussi bien chez les eucaryotes que chez les procaryotes. De manière générale, un pseudogène est la conséquence de la duplication d'un gène soit par rétrotransposition, soit par duplication d'ADN génomique (figure 1.4).

**Rétrotransposition**

On parle aussi de rétropseudogène. Il est caractérisé par l'absence de promoteur et d'introns, la présence de



**Figure 1.4** Exemples de mécanisme de formation d'un pseudogène. Deux mécanismes essentiels peuvent aboutir à la formation d'un pseudogène ( $\Psi$  gène), la duplication et la rétrotransposition. La rétrotransposition : un gène transcrit en ARNm est inséré (« rétrotransposé ») dans une zone chromosomique contenant des séquences répétées (triangle). Après insertion, cet ARNm intégré subit des mutations (losange et carré) le rendant non fonctionnel. Un pseudogène est créé. La duplication : au cours de l'évolution, certains gènes se sont dupliqués, intégrés et ont subi des mutations (carré, triangle et losange) et d'autres modifications les rendant non fonctionnels, aboutissant à la formation de pseudogène.

séquences répétées l'encadrant et d'une queue poly(A) en 3'. Il résulte de la transposition par insertion d'un ARNm normalement reverse-transcrit dans une zone non homologe d'un chromosome contenant de part et d'autre de l'insertion des séquences répétées. Dans un deuxième temps, de nombreuses mutations sur l'ADNc apparaissent dans cette région chromosomique et rendent le gène inactif. Il est devenu un pseudogène. Il s'agit par conséquent de rétrotransposons insérés dans le génome sous forme d'ADN double brin issu d'un ARN simple brin généré le plus souvent par une ARN polymérase II. La majorité de ces rétrotransposons sont donc inactivés en rétropseudogènes. Le mécanisme de rétrotransposition peut cependant ne pas aboutir dans certains cas à la formation d'un pseudogène. En effet, un gène fonctionnel peut être maintenu (ce n'est donc plus un pseudogène) sans intron (par exemple, le gène PGK-2 fonctionnel et le pseudogène  $\psi$ PGK-1 sont tous les deux issus par rétrotransposition du gène PGK-1). Par ailleurs, certaines séquences mitochondriales sont retrouvées dans le génome nucléaire. Elles ont ainsi pu migrer dans le génome nucléaire et se transformer en pseudogène. Bien que le mécanisme ne soit pas connu, certains auteurs ont suggéré la possibilité d'un phénomène de rétrotransposition à partir d'ARN mitochondrial converti dans un deuxième temps en ADN intégré.

#### Duplication d'ADN génomique

Au cours de l'évolution du génome, un certain nombre de gènes ont évolué à partir d'un ancêtre commun. La duplication est un des mécanismes expliquant la présence de famille de gènes au sein du génome. Certaines duplications ont évolué vers la formation de pseudogène (par exemple,  $\psi$ BRCA1 provenant de la duplication de l'extrémité 5' du gène codant pour BRCA1 et qui ne comprend que les exons 1A, 1B et 2).

Tous les gènes n'ont pas de pseudogènes. Cependant, l'analyse du génome humain semble indiquer que les pseudogènes sont assez fréquents. Quant aux procaryotes, le mécanisme de formation des pseudogènes est encore mal compris.

#### Séquences répétées

De nombreuses séquences répétées sont réparties sur le génome. Leur taille et leur nombre sont très variables. Selon leur origine, leur fonction ou leur constitution, plusieurs types de séquences répétées existent. Par exemple, les ADN satellites sont des séquences hautement répétitives, de fonction inconnue. Les séquences répétées parsemées (*intersped repeat sequences*) sont le produit de l'intégration d'éléments transposables (comme les rétrogènes et les rétropseudogènes), d'un gène fonctionnel. Schématiquement, on distingue trois classes de séquences répétées.

#### Microsatellites

Ce sont de courtes séquences répétées en tandem. Chaque répétition est constituée de 1 à 10 pb au maximum. Les plus fréquentes sont les séquences répétées [CA]. Leur fonction n'est pas connue. Leur nombre peut varier. Certaines maladies sont la conséquence d'une augmentation importante de leur nombre (expansion) comme, par exemple, l'expansion de trinuécléotides associée au syndrome de l'X fragile (Omim 309550).

#### Minisatellites

La taille des séquences répétées d'ADN varie de 1 à 15 kb. C'est le cas par exemple de l'ADN des extrémités des chromosomes (ADN télomérique) contenant des répétitions d'hexanucléotides, 5'-TTAGGG sur 10–15 kb, ajouté par une télomérase pour permettre une réplication complète des chromosomes.

#### Macrosatellites

Les répétitions peuvent aller jusqu'à une centaine de kb de séquences répétées en tandem. C'est le cas par exemple de l'ADN alpha-satellites qui constitue l'hétérochromatine centromérique des chromosomes.

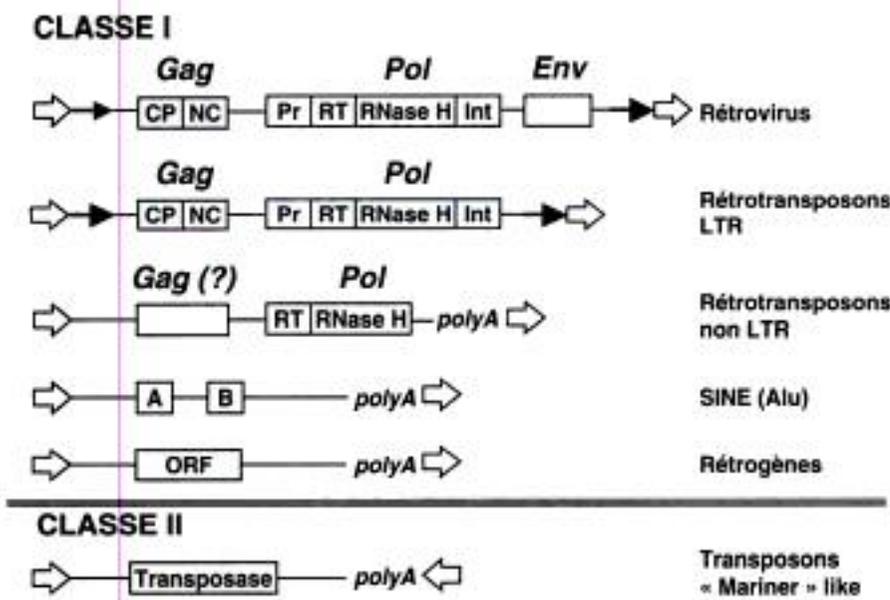
#### Éléments transposables

Schématiquement, selon le mode de transposition de ces éléments, on distingue (figure 1.5) :

- la classe I : les rétroéléments. Ils se transposent par réplication et nécessitent un intermédiaire ARN qui se rétrotranscrit avant d'être intégré dans l'ADN. C'est le cas, par exemple, des transposons LTR (*long terminal repeat* proches des rétrovirus), des éléments non LTR (*long intersped nuclear elements 1*, L1, LINE, et *short intersped nuclear elements*, SINE), des séquences Alu et des rétrogènes (pseudogènes) ;
- la classe II : ce sont des séquences s'intégrant par un mécanisme de « copier-coller ». Après excision de la séquence, celle-ci est réinsérée dans un autre lieu du génome.

#### Rétroéléments

Les rétroéléments se caractérisent par la nécessité de passer par un intermédiaire ARN transcrit à partir du rétrotransposon ADN pour pouvoir intégrer le génome. Les deux principales et plus abondantes classes de séquences répétées sont des rétrotransposons, les *long intersped repetitive elements* (LINE) et les *short intersped repetitive elements* (SINE). On en distingue cependant d'autres. Classiquement, on les sépare en deux classes, les rétrotransposons autonomes et non autonomes.



**Figure 1.5** Schéma des structures des principaux éléments transposables chez l'homme. Les flèches claires correspondent aux sites dupliqués d'insertion dans le génome. Les flèches sombres indiquent des LTR. CP : capsid ; NC : nucléocapsid ; Pr : protéinase ; RT : transcriptase inverse ; Int : intégrase ; ORF : cadre de lecture ouverte ; A et B : promoteurs de l'ARN polymérase III. On distingue deux classes : les classes I et II. Classe I : les rétrotransposons autonomes, séquences d'environ 6–7 kb, contenant des LTR (similaires au rétrovirus) ou n'en contenant pas (éléments LINE 1 ou L1). Ils rétrotransposent par l'intermédiaire d'un ARN soit comme les rétrovirus, soit par transcription inverse médiée par une amorce (*target-primed reverse transcription*, TPRT). Les rétrotransposons non autonomes comme les séquences Alu (environ 300 pb) et les pseudogènes (rétrogènes), copies d'ARNm sans introns et contenant des mutations non-sens ou des décalages de phase. Classe II : les transposons « mariner », séquences d'environ 1,4 kb qui s'intègrent par un mécanisme « couper/coller » ou « copie/coller » par une transposase.

**Rétrotransposons autonomes**

Les rétrotransposons contenant des séquences *long terminal repeat* (LTR) ont une structure similaire à l'ADN des rétrovirus. Dans le génome humain, la classe dominante est constituée par les rétrovirus endogènes (*human endogenous retroviruses*, HERV, dont on peut consulter la base de données à l'adresse HERVd, voir le chapitre « Sites internet de biologie moléculaire »). Leur taille varie entre 6 et 10 kb. On estime qu'ils représentent environ 0,1 à 1 % du génome. Ces HERV sont le plus souvent défectifs même si on pense que certains d'entre eux puissent rétrotransposer. Enfin, un certain nombre de ces HERV possèdent une activité transcriptionnelle (promoteur ou *enhancer*).

Pour les rétrotransposons ne contenant pas de LTR, les deux principales et plus abondantes classes de séquences répétées sont des rétrotransposons d'origine non virale, les LINE et les SINE. Sur le plan structural, ils présentent tous deux une région 3' riche en A et l'absence de longues séquences répétées terminales (LTR), ce qui les distingue des rétrovirus et des rétroéléments en dérivant. Les LINE (appelé aussi élément L1) sont disséminés dans le génome et sont surtout concentrés dans les régions riches en AT. Sans rentrer dans les détails, un LINE complet mesure environ 6 à 7 kb. Certains d'en-

tre eux sont potentiellement actifs. On estime que les LINE constituent environ 17 % du génome. Un grand nombre d'entre eux sont tronqués et constituent des pseudogènes aux extrémités 5' diverses par transcription inverse incomplète.

Le nombre de LINE est estimé à environ 100 000 dans le génome humain dont environ 3500 complets. Un pour cent environ de ces LINE complets ont une séquence promoteur d'ARN polymérase II et deux séquences codantes ou *open reading frame* (ORF) intactes permettant de générer de nouvelles copies d'éléments L1. Un LINE contient une queue poly(A) et à chaque extrémité des séquences répétées. L'activité LINE a pu être mise en évidence dans les lignées germinales et somatiques.

**Rétrotransposons non autonomes**

Le nombre des séquences Alu est estimé à environ 500 000–1 000 000 de copies ; elles constituent le « chef de file » des SINE. La séquence Alu est le transposon le plus efficace du génome et constituerait environ 10 % du génome. Réparties dans tout le génome, on estime qu'il y a en moyenne une séquence Alu tous les 5 kb. Dans certaines régions cependant, des groupes (*clusters*) de séquences Alu peuvent être observés.

De nombreuses études ont montré que les séquences Alu dérivent du gène codant pour l'ARN 7SL. Chaque élément Alu est constitué d'un fragment de 280 pb de structure dimérique et contient des séquences promoteurs pour l'ARN polymérase III. Elle possède une queue poly(A) et à ses extrémités des séquences répétées. Les séquences Alu de manière générale affectent la composition, l'organisation et l'expression du génome. Elles peuvent avoir un rôle pathogène (insertion de séquences Alu) mais aussi fonctionnel comme *enhancer* ou *silencer*, jouer un rôle dans le mécanisme d'empreinte ou être le composant d'un gène par exemple. Certains auteurs ont utilisé les séquences Alu comme marqueur génétique par exemple dans des études d'épidémiologie moléculaire.

**Éléments de classe II**

Ce sont des transposons proches de ceux observés dans les bactéries. Ils contiennent des séquences répétées inversées de 10 à 500 pb à leurs extrémités et codent pour une transposase catalysant la transposition. Après excision du site donneur, ils se réinsèrent ailleurs dans le génome.

**Autre « génome humain » : l'ADN mitochondrial**

La mitochondrie fournit l'énergie à la cellule en synthétisant le carburant cellulaire (l'ATP) grâce à un ensemble de

réactions chimiques complexes, la phosphorylation oxydative. Ces mitochondries contiennent leur propre ADN et diffèrent de l'ADN contenu dans le noyau de la cellule. Cet ADN est donc extrachromosomique. Il s'agit d'un ADN double brin, d'environ 16,5 kb, circulaire, codant pour 13 sous-unités protéiques constituant quatre complexes biochimiques et 24 ARN « structuraux » (2 ARNr et 22 ARNt) indispensables pour la traduction intramitochondriale des protéines (soit 37 gènes au total).

L'ADN mitochondrial est indépendant de l'ADN du noyau : la réplication, la transcription, la traduction sont donc indépendantes de celles se déroulant dans le noyau. L'ADN nucléaire code pour des protéines participant à la phosphorylation oxydative et pour les protéines nécessaires aux fonctions et structures mitochondriales. Ces protéines (codées par des gènes du noyau) sont sécrétées dans le cytoplasme avant d'aller dans la mitochondrie. Il existe donc une inter-relation ADN nucléaire/ADN mitochondrial.

### Caractéristiques de l'ADN mitochondrial

L'ADN mitochondrial ne possède pas d'introns.

- L'ADN mitochondrial a un taux de mutations environ dix fois plus élevé que l'ADN du noyau. Il n'a pas de système de réparation, ni d'histones. Par ailleurs, l'ADN polymérase gamma, responsable de la réplication de l'ADN mitochondrial, aurait un taux d'erreur de recopiage important.
- Il existe un grand polymorphisme génétique au sein de l'ADN mitochondrial.
- L'ADN mitochondrial est transmis de manière générale par la mère. En effet, les mitochondries dérivent de l'ovocyte. Néanmoins, cette règle n'est pas absolue. De très rares cas de transmission paternelle ont été décrits.
- L'ADN mitochondrial ne recombine pas.
- Chaque mitochondrie contient deux à dix molécules d'ADN.
- Chaque cellule contient de très nombreuses mitochondries (jusqu'à plusieurs milliers).
- Dans une même mitochondrie, peuvent coexister de l'ADN normal et de l'ADN muté. On parle alors d'hétéroplasmie.
- En cas de présence uniquement d'ADN normal ou uniquement d'ADN muté, on parle d'homoplasmie.
- Au cours de la réplication, la proportion d'ADN normal et mutant (en cas d'hétéroplasmie) peut être modifiée dans les cellules filles, à la suite d'une répartition au hasard des mitochondries.

Enfin, le code génétique dans la mitochondrie présente certaines modifications par rapport au code universel applicable dans le génome nucléaire. Ainsi, le codon UGA code pour le tryptophane (au lieu du codon-stop), le codon AUA pour la méthionine (au lieu

de l'isoleucine) et les codons AGA et AGG pour le codon-stop (au lieu de l'arginine).

### Réplication

La réplication constitue l'ensemble des événements aboutissant à la duplication de l'ADN au cours de la division cellulaire. L'information génétique est transmise aux deux cellules filles. Sans rentrer dans les détails des mécanismes fort complexes de la réplication, nous rappellerons quelques bases indispensables.

Tandis que chez les procaryotes de manière générale, un site de réplication existe (origine de réplication), chez les eucaryotes, la réplication se produit en de nombreux sites (environ  $10^4$ – $10^6$ ). Les unités de réplication ainsi composées s'appellent les réplicons.

Que ce soit les eucaryotes ou les procaryotes, les règles de base de la réplication sont universelles. Il s'agit d'un processus semi-conservatif et bidirectionnel. La vitesse de la réplication dépend de la fréquence de son initiation et de la vitesse de l'élongation ou de la progression de la fourche de réplication. De manière schématique, les enzymes-clés de la réplication sont :

- l'hélicase qui déroule les deux brins d'ADN et définit le point de la fourche de réplication ;
- la topoisomérase qui facilite la progression de la fourche ;
- l'ADN polymérase qui permet le recopiage de l'ADN par addition des nucléotides dans la direction  $5' \rightarrow 3'$ , les nucléotides étant complémentaires de l'ADN source et ajouté à l'extrémité  $3'$ -OH des nucléotides déjà polymérisés.

En plus de ces enzymes, des protéines jouent un rôle-clé, les protéines *single strand binding proteins* (SSB). Les SSB stabilisent l'ADN source simple brin.

La réplication étant bidirectionnelle et dans la direction  $5' \rightarrow 3'$ , à partir de chaque brin ADN source (sens et antisens), le mode de réplication est différent :

- le brin leader (*leading strand*) déroulé dans le sens  $3' \rightarrow 5'$  sert directement de source pour l'élongation  $5' \rightarrow 3'$  du brin complémentaire naissant dans la direction de la progression de la fourche ;
- le brin retardé (*lagging strand*) de direction antiparallèle doit être régulièrement ré-initialisé par une primase (ARN polymérase permettant la synthèse des petits brins d'ARN) à partir de petits brins d'ARN servant d'amorces pour l'extension par la polymérase. Il s'agit en effet d'un processus discontinu. L'extension depuis ces amorces ARN (*initiator ARN*) avec l'ADN polymérase aboutit à la production de petits fragments, les fragments d'Okazaki, de taille variable (en moyenne, 2 à 3 kb chez les procaryotes et 100 à 300 pb chez les eucaryotes). Ces fragments sont ensuite collés entre eux par une enzyme, l'ADN ligase, après remplacement des amorces ARN par de l'ADN.

Dans la mitochondrie, l'enzyme responsable de la réplication est une ADN polymérase spécifique, l'ADN polymérase gamma de structure proche de l'ADN polymérase I de *Escherichia coli*.

### Origine de réplication

La réplication est définie sur le plan de la séquence ADN mais aussi en fonction de déterminants structuraux que nous ne détaillerons pas. La séquence d'ADN qui initie la réplication s'appelle l'origine de réplication. Celle-ci interagit avec un grand nombre de protéines constituant un complexe (à titre d'exemple, les protéines Cdc6, Cdk2, les protéines lamines chez les mammifères). Chez les eucaryotes, l'origine de réplication se dénomme aussi *origins of bidirectional replication* (OBR). Des séquences en *cis* (appelées réplicateurs) définissent sa localisation. Les réplicateurs sont reconnus par des protéines initiateuses (*initiator protein*, IP, telles que les *origin recognition complex*, ORC, impliquées dans la réplication dans la levure, les *multiprotein replication complex*, MRC, chez les mammifères) qui interagissent sur ceux-ci. Le déroulement de l'ADN et le début de la synthèse se fait alors au niveau de l'origine. Certaines séquences de réplication sont transposables. C'est le cas par exemple des réplicateurs trouvés chez *Saccharomyces cerevisiae* qui peuvent aussi servir de réplicateurs dans des plasmides. Ces séquences sont appelées séquences répliquantes autonomes (*autonomously replicating sequence*, ARS).

### Réplication chez les eucaryotes

Celle-ci n'a lieu qu'une fois au cours d'un cycle cellulaire. À partir d'une unité de réplication (réplicon) qui contient une origine de réplication, deux fourches de réplication divergentes prennent naissance. Chaque réplikon a une taille variant entre 50 et 250 kb. La réplication d'un réplikon s'achève quand une fourche rencontre une autre fourche. Les deux fourches fusionnent alors. Les réplikons sont coordonnés en groupes de 20 à 100 unités, chaque groupe étant activé à différents moments de la phase S du cycle cellulaire. Par ailleurs, il existe plus de sites potentiels d'initiation de la réplication dans une cellule quiescente que ceux observés dans une cellule en phase S dans une cellule normale. La réplication est réalisée dans un ordre défini spatial et temporel. En fait, la sélection des sites d'initiation de la réplication varie selon le stade cellulaire (par exemple, embryogénèse, développement tissulaire). La régulation de la réplication au cours de la phase S est un phénomène complexe. Par exemple, au niveau des centromères, les alpha-satellites riches en séquences A et T sont répliqués tardivement au cours de la phase S. Le plus souvent, les gènes les plus activement transcrits sont les premiers répliqués et inversement pour les gènes inactifs.

Le déclenchement de la réplication est encore mal connu. De nombreuses protéines interviennent dans celle-ci, y compris, semble-t-il, des facteurs de transcription. La structure chromatinienne joue aussi un rôle important dans la réplication, un certain nombre de protéines liées aux nucléosomes étant impliquées.

À noter qu'au cours de la réplication, la méthylation de l'ADN est maintenue grâce à des ADN méthylases qui reconnaissent les duplex ADN hémiméthylés. Cette action se fait en présence de complexes protéiques et d'histones, ces derniers jouant un rôle important dans la reconnaissance de l'état de méthylation de l'ADN et permet à la fois la transmission de l'état de méthylation dans les cellules en division et permet le maintien de la mémoire épigénétique dans les cellules qui ne sont pas en division.

### Réplication chez les procaryotes

Comme pour les eucaryotes, la réplication est bidirectionnelle à partir de l'origine de réplication (*Ori*). Après fixation d'un complexe protéique sur *Ori* et en présence d'une enzyme, l'hélicase, la double hélice d'ADN s'ouvre dans les deux directions. D'autres enzymes interviennent, la gyrase (son activité est proche de celle de la topoisomérase), la topoisomérase, la primase et l'ADN polymérase. À noter à propos des amorces utilisées avec la primase que, chez certains procaryotes, pour les plasmides et pour certains bactériophages, une amorce ADN (au lieu d'une amorce ARN) est utilisée. Certains virus (par exemple, l'adénovirus) utilisent une protéine et non un acide nucléique. L'ADN polymérase (par exemple, polymérase III chez *Escherichia coli*) présente une capacité à polymériser (« la processivité ») importante et une activité correctrice des erreurs (activité exonucléasique 3'→5', *proofreading*). Comme chez les eucaryotes, la fourche de réplication est asymétrique avec un brin synthétisé de façon continue et l'autre de façon discontinue avec synthèse des fragments d'Okazaki. À noter que les ARN synthétisés par la primase et servant d'amorces sont ensuite détruits par une RNase H (ou par la polymérase I grâce à son activité exonucléasique chez *Escherichia coli*). La réplication se termine lorsque deux fourches se rencontrent. Au cours de la réplication, des erreurs de recopiage apparaissent. Celles-ci sont corrigées par l'activité exonucléasique 3'→5' de l'ADN polymérase mais aussi par un système de correction des erreurs de plusieurs protéines (chez *Escherichia coli*, le taux d'erreurs résiduel est faible : environ  $10^{-10}$  erreurs d'incorporation). Par ailleurs, le système de méthylation de l'ADN (chez *Escherichia coli*) permet de différencier le brin parental du brin néosynthétisé (qui n'est pas transitoirement méthylé). Enfin, outre les erreurs de recopiage, il peut se produire dans certains cas (par exemple au cours de l'arrêt de la progression de la réplication), des réarrangements de grandes séquences d'ADN.

### Réparation de l'ADN

Au cours d'une atteinte de l'ADN (par exemple, par exposition aux ultraviolets provoquant une dimérisation covalente de pyrimidines adjacentes ou après incorporation erronée de bases par erreur au cours de la réplication), la machine répliquative permet la réparation de l'ADN. Les mécanismes en sont complexes et impliquent de très nombreuses protéines. Dans certains cas, il n'y a pas réparation mais ignorance de l'anomalie (*bypass*). Ce phénomène de tolérance permet la réplication mais aboutit à la formation de mutation sur la nouvelle séquence. Enfin, parfois, il n'y a pas de réparation et l'atteinte de l'ADN provoque la mort cellulaire par apoptose.

### Du génotype au phénotype : l'expression d'un gène

L'expression d'un gène démarre au moment de l'activation de la transcription et se termine à la synthèse d'une protéine fonctionnelle (figures 1.6 et 1.7).

Le niveau d'expression de la plupart des gènes est régulé par des facteurs de transcription se fixant sur des séquences spécifiques au niveau de l'ADN, en amont du site d'initiation de la transcription. L'action des facteurs de la transcription est l'aboutissement d'un ensemble d'interactions multiples et complexes entre plusieurs voies de régulation.

L'ensemble des facteurs de transcription en association avec la machine transcriptionnelle favorise l'accès à l'ADN

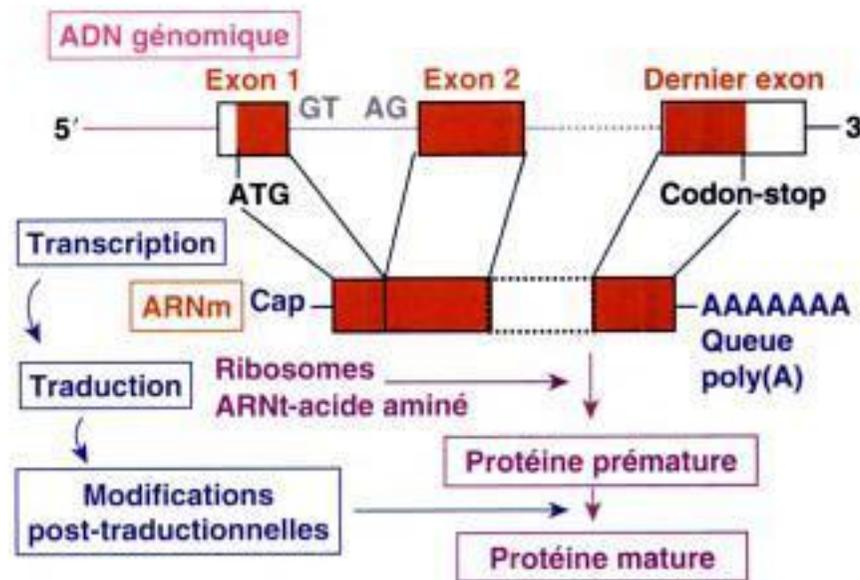


Figure 1.6 Relation gène/protéine.

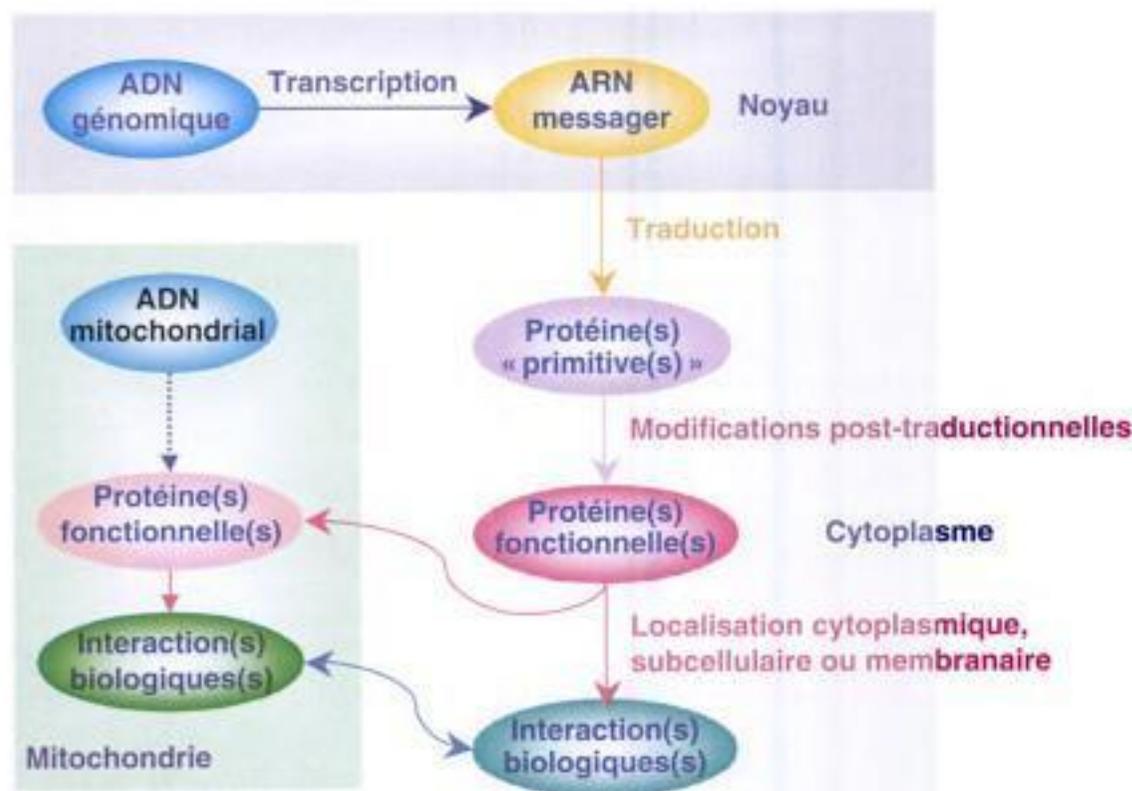


Figure 1.7 Relation génomes/protéine. Au sein de la cellule, les relations entre les génomes des mitochondries et du noyau (communément appelé génome humain) et l'ensemble de la machinerie cellulaire sont complexes.

et le recrutement des ARN polymérase au site d'initiation de la transcription. De manière générale, on distingue trois ARN polymérase (I, II et III) dont le fonctionnement est similaire, l'ARN polymérase II étant la plus importante.

## Transcription

La séquence de l'ADN contenue dans le noyau est lue et traduite en une séquence d'acides aminés qui constituera la protéine. Cette traduction passe par une étape intermédiaire qui transforme l'information contenue dans l'ADN en une copie sous forme d'ARN messager (ARNm) : cette étape s'appelle la transcription.

Cette étape est plus complexe chez les eucaryotes où plusieurs étapes de maturation sont nécessaires afin notamment d'éliminer les introns. En effet, en général, les procaryotes ne présentent pas d'introns.

De manière générale, la régulation de la transcription est la conséquence de la fixation de protéines agissant en *trans* (facteurs de transcription) sur des séquences d'ADN situées à proximité du gène à réguler (sites de fixation en « *cis* »). La fixation d'éléments *trans* sur des sites *cis* permet alors l'interaction avec le complexe d'initiation de la transcription.

De nombreuses séquences agissent sur l'ADN pour en réguler l'expression. Nous donnerons quelques exemples de ces éléments de régulation.

### Séquences d'ADN agissant en cis (*cis-acting factors*)

#### Locus control elements (LCR)

Ces régions sont définies par leur capacité à stimuler l'expression de certains gènes et maintenir un niveau physiologique d'expression des gènes sous leur dépendance. Ces séquences d'ADN sont très éloignées des gènes qu'elles contrôlent (plusieurs dizaines de kb en moyenne). Le premier LCR décrit mais aussi le plus étudié est celui du locus bêtaglobine. Elles sont tissu-spécifiques et [nombre de copies]-dépendant. Ces LCR sont localisés au niveau de sites d'hypersensibilité aux DNases (reflet indirect de la structure de la chromatine, cette propriété est associée à un état ouvert de la chromatine et donc à une accessibilité de l'ADN). Ils maintiennent la chromatine dans un état ouvert. Cependant, le mécanisme expliquant son action à distance est encore mal connu. Pour certains, elle pourrait jouer un rôle dans la dynamique d'acétylation des histones en stimulant l'activité des acétyltransférases au moins à certains loci. Le LCR pourrait aussi jouer un rôle dans la déméthylation tissu-spécifique de l'ADN. Ce qui est clair, c'est que le LCR joue un rôle sur la configuration et la conformation de la chromatine. Cet élément empêche ainsi l'évolution de la chromatine vers de l'hétérochro-

matine, propriété fondamentale qui la différencie des *enhancers*. D'autres séquences spécifiques telles que des activateurs ou des isolateurs (*insulators*) interviennent probablement pour moduler la fonction des LCR. Par ailleurs, le LCR possède plusieurs propriétés :

- il stimule de manière importante l'activité de transcription d'un gène. Ainsi, dans les souris transgéniques pour le locus bêtaglobine humaine, l'absence de LCR se traduit par une expression résiduelle d'environ 1 % par rapport à l'homologue murin. L'activité du locus d'origine humaine retrouve un niveau similaire à celle de son homologue murin, quand le LCR est rajouté ;
- l'activité LCR est tissu-spécifique. Cette spécificité est liée aussi à son interaction avec le promoteur du gène. Sur ce LCR, se trouvent des séquences spécifiques de fixation pour des facteurs de transcription ;
- sur le gène auquel il est lié, le LCR est capable de stimuler l'expression indépendamment de la position mais fonction du nombre de copies du gène ou locus lié à celui-ci.

Un certain nombre de LCR ont été décrits dans le génome. À titre d'exemple, on peut citer les LCR du gène codant pour l'adénosine déaminase (Omim 102700), du gène de l'hormone de croissance GH (Omim 139250) et du gène codant pour l'apolipoprotéine B (Omim 107730).

#### Promoteur

Cette séquence d'ADN est adjacente au site d'initiation de la transcription. La fonction du promoteur (ou région promotrice) est position- et orientation-dépendante. En effet, le promoteur établit le site d'initiation de la transcription et la direction dans laquelle la transcription va se dérouler. Chaque promoteur contient non seulement des séquences définissant le site d'initiation de la transcription mais aussi des séquences consensus pour les sites de fixation de l'ARN polymérase et des complexes de transcription.

Par exemple, certains gènes « tissu-spécifiques » contiennent, 20 à 30 bp en amont du site d'initiation de la transcription, une boîte TATA (5'-TATAAA-3') sur laquelle se fixent certains complexes protéiques alors que certains gènes s'exprimant dans tous les tissus (*house-keeping gene*) contiennent des séquences riches en GC, par exemple, 5'-GGGCGG-3', séquence sur laquelle se fixe le facteur de transcription Sp1.

Un gène peut avoir plusieurs promoteurs (certains d'entre eux pouvant être tissu-spécifiques).

#### Enhancers

Ces séquences d'ADN sont localisées en 5' ou en 3' du gène, dans un exon ou dans un intron. La fonction de l'*enhancer* est position- et orientation-indépendante. Chaque *enhancer* contient au moins un site de fixation pour un facteur de transcription (en général, plusieurs

sites sont présents qui agissent en synergie). Ces séquences stimulent l'expression d'un gène.

### Silencers

Ces séquences d'ADN diminuent ou inhibent l'expression d'un gène. Ils sont souvent localisés entre le promoteur et les *enhancers*. Ce sont des « anti-*enhancers* ».

### Insulateurs

Les insulateurs sont des séquences d'ADN parfois courtes (par exemple, 42 pb) situées entre un (et/ou plusieurs) *enhancers* ou un (et/ou plusieurs) *silencers* et un promoteur de gènes ou de groupes de gènes adjacents. L'insulateur empêche l'influence d'un gène (par activation ou répression) sur son voisin. À titre d'exemple, le promoteur du gène codant pour la chaîne delta du récepteur gamma/delta de la cellule T (*T cell receptor*, TCR) est juste à côté du promoteur du gène codant pour la chaîne alpha du récepteur alpha/bêta de la cellule T. La cellule T doit choisir une des deux chaînes. Un insulateur sépare le promoteur du gène alpha de celui du gène delta et évite ainsi que l'activation de l'un ne s'étende à l'autre. Ces insulateurs fonctionnent en association avec une protéine se fixant sur leur séquence, le CTCF (*CCCTC-binding factor*).

### Facteurs agissant en trans (*trans-acting factors*)

Ce sont les protéines qui agissent sur l'ADN en activant (activateurs) ou en inhibant partiellement ou totalement (répresseurs) la transcription.

### Rôle de la chromatine

L'ADN n'est pas nu dans le noyau. Il est empaqueté dans un complexe nucléoprotéique appelé chromatine dont on distingue deux formes, l'euchromatine et l'hétérochromatine. La chromatine est en effet un ensemble complexe constitué de l'association de l'ADN et de protéines histones et « non histones ».

L'organisation de base de la chromatine est appelée nucléosome : il est constitué d'ADN de taille 146 pb enroulé pratiquement deux fois autour d'un noyau protéique contenant deux copies de chacune des quatre protéines appelées histones (H2A, H2B, H3 et H4). La compaction de l'ensemble du génome est ensuite réalisée par un arrangement de nucléosomes par un mécanisme mal connu actuellement. Les modifications de la chromatine jouent un rôle important dans la régulation transcriptionnelle des gènes. Elle modifie l'accès de la machinerie transcriptionnelle aux gènes. Cependant, la chromatine n'est pas homogène. Les régions non transcrites du génome sont empaquetées dans des régions hautement condensées appelées hétérochromatine,

riches en séquences répétées et contenant peu de gènes alors que les gènes accessibles pour la transcription se trouvent dans les zones accessibles appelées euchromatine. L'hétérochromatine présente un certain nombre de modifications caractéristiques telles que l'hypoacétylation des histones sur des résidus lysine ou leur méthylation ou encore la méthylation des cytosines sur l'ADN. Cette subdivision euchromatine/hétérochromatine est différente selon le type de cellule et détermine les gènes qui seront activés en fonction du type de cellule. Elle se transmet aux cellules filles. Un exemple de ces modifications est l'X inactivé chez la femme, inactivation aléatoire provenant de l'X du père ou de la mère. Pour activer un gène, les protéines activatrices de la transcription doivent donc franchir la structure chromatinienne « fermée » (voir infra, régulation de la transcription).

### Mécanismes généraux de la transcription

La transcription a lieu dans le noyau. On distingue schématiquement trois étapes importantes :

- la fixation d'un chapeau en 5' du pré-ARNm ;
- l'élimination des introns (appelé aussi épissage ou *splicing*) ;
- l'addition en 3'-terminal d'une queue poly(A).

Ces trois phénomènes sont en fait interdépendants, l'ensemble [protection en 5', épissage et polyadénylation] étant couplé.

#### Mise en place d'un chapeau en 5' du pré-ARNm

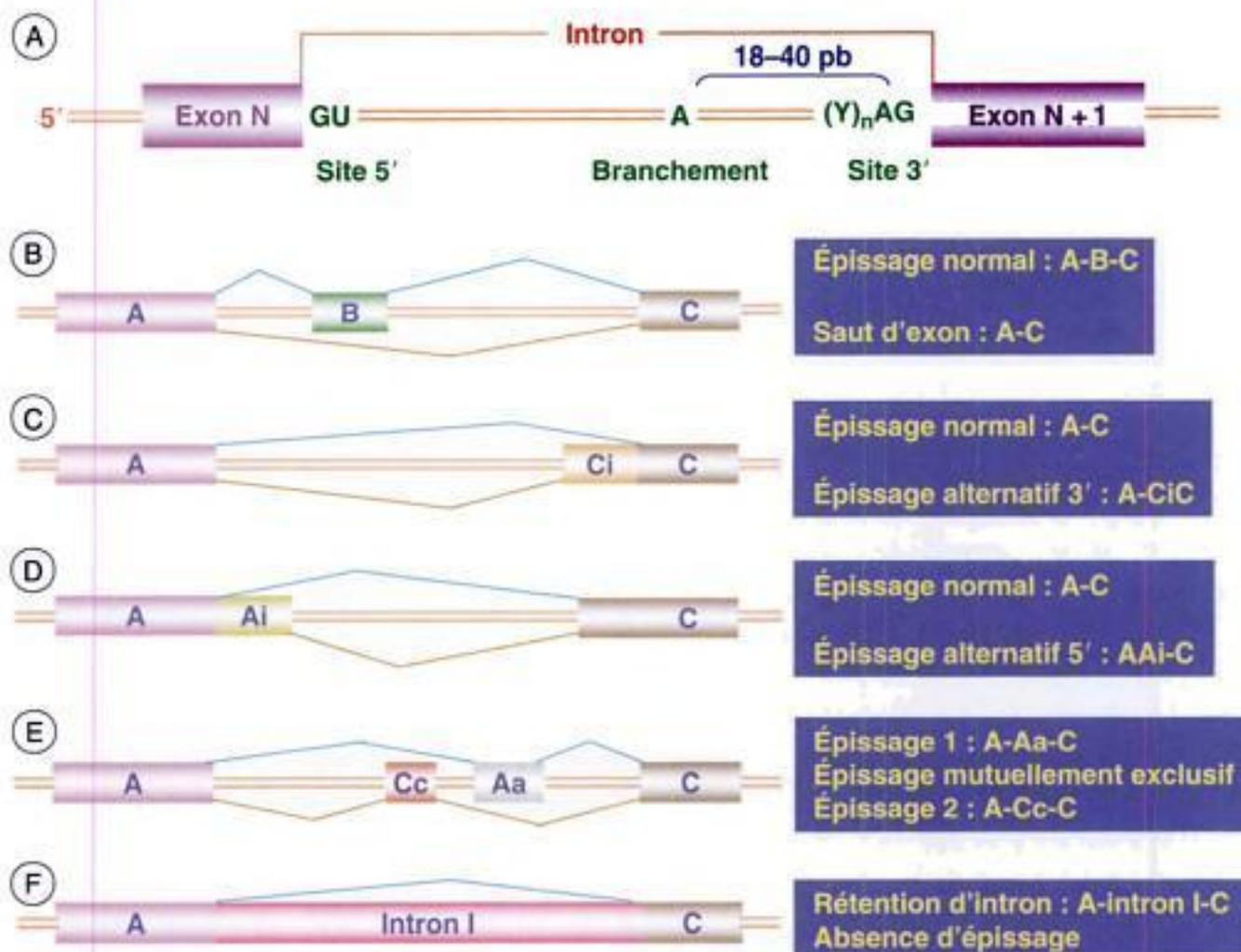
Dès l'initiation de la transcription par l'ARN polymérase (nous prendrons l'exemple de l'ARN polymérase II, les autres agissent de façon similaire), après la synthèse d'environ 20 à 30 nucléotides en général, l'ARNm naissant va être modifié par l'addition d'un chapeau (*cap*) en 5' par un processus complexe faisant intervenir plusieurs protéines. Au moment de l'addition de ce chapeau, l'ARN polymérase effectue une pause. Il semblerait que cette étape soit fondamentale : l'enzyme continue uniquement si le chapeau est ajouté. Celui-ci, stabilisant l'ARNm, le protège d'une digestion par les exonucléases 5'→3'. Il permet aussi la fixation de protéines indispensables à l'exportation de l'ARNm mature dans le cytoplasme et à sa traduction en protéine. Au début de la traduction, il participe à l'engagement des sous-unités ribosomales avec l'ARNm. L'élongation de l'ARNm se poursuit ensuite à partir de l'extrémité 5' de l'ADN vers l'extrémité 3'. Un certain nombre de protéines appelées facteurs d'élongation interviennent au cours de cette phase en régulant cette étape.

#### Épissage (*splicing*)

Cette étape a lieu après la transcription et avant la traduction. Au cours de la transcription, le gène est trans-

crit en pré-ARNm, copie de l'ADN génomique contenant les introns. Ces derniers sont éliminés au cours de l'étape d'épissage pour aboutir à l'ARNm mature. Au niveau du génome, les séquences codantes du gène (contenues dans les exons) sont séparées entre elles dans la majorité des gènes par de longues séquences non codantes, les introns. Ces derniers sont caractérisés par au minimum trois éléments, le site d'épissage 5', le site de branchement et le site d'épissage 3'. Ces introns sont éliminés par un phénomène complexe appelé épissage (épissage des pré-ARNm ou *splicing*). Le pré-ARNm possède en *cis* des séquences consensus essentielles pour l'épissage. La jonction 5' exon-intron (représenté par une barre verticale, |), appelée aussi site d'épissage, est constituée par la séquence 5'-AG|GURAGU (R, purine ; Y, pyrimidine). La fin de l'intron, le site d'épissage 3', est définie par la séquence YAG|RNN (N, n'importe quelle base) : 20 à 40 pb environ en amont du site 3' d'épissage, se situe une séquence appelée site de branchement dont le rôle est fondamental dans la réaction d'épissage. Ce site présente la séquence consensus : YURAY, l'adénosine est quasi constante (très conservée), suivie d'une séquence riche en pyrimidine. Nous ne décrivons pas le phénomène d'épissage qui sort du cadre de cet ouvrage. Cette réaction complexe fait intervenir

notamment des petits ARN nucléaires (*small nuclear RNA*, snRNA) et des ribonucléoprotéines, snRNP. Associées à d'autres protéines, elles aboutissent à la formation d'un ensemble comprenant l'intron qui sera éliminé. Ce complexe constitue le spliceosome dont le rôle principal est la reconnaissance des jonctions exon/intron et la catalyse de la réaction « couper/coller » éliminant les introns et joignant les exons. L'identification des exons, introns et de leurs jonctions est complexe. Des études ont montré l'existence de séquences soit dans les introns soit dans les exons stimulant l'épissage (*exonic splicing enhancer*, ESE, et *intronic splicing enhancer*, ISE) ou inhibant celle-ci (*exonic splicing silencer*, ESS, et *intronic splicing silencer*, ISS), ces séquences interagissant avec le spliceosome. L'ARNm peut aussi subir des épissages alternatifs aboutissant à partir de la même unité de transcription à plusieurs isoformes (figure 1.8). Par exemple, l'analyse du chromosome 22 a montré que 60 % des gènes au minimum possédaient au moins deux transcrits. Ainsi, dans certains cas, des exons pourront être excisés de façon différente suivant les tissus (tels que des isoenzymes dont l'expression peut être spécifique d'organe). L'existence d'épissage alternatif est en partie à l'origine de la diversité protéique observée dans les études de protéomique. Le nombre de gènes soumis à un épissage



**Figure 1.8** Épissage normal et exemples d'autres formes d'épissage. **A.** Les motifs conservés aux extrémités de l'intron sont les dinucléotides GU en 5' et AG en 3', ce dernier étant précédé d'une séquence polypyrimidique (Y) n. Le résidu A est le nucléotide conservé au niveau du site de branchement à 30-50 pb de l'extrémité 3' de l'intron. En **B, C, D, E, F**, cinq modes d'épissage alternatif sont schématiquement décrits. En **F**, il n'existe pas d'épissage.

alternatif est encore largement spéculatif actuellement. Un certain nombre d'auteurs pensent qu'au minimum 35 % des gènes humains présentent un ou plusieurs épissages alternatifs (le nombre d'épissage alternatif est en effet variable selon les gènes). Par exemple, le gène *KCNMA1* (*potassium channel, calcium-activated, large conductance, subfamily m, alpha member 1*; Omim 600150), codant pour un canal potassium activé par le calcium exprimé au niveau de la cochlée (oreille interne) et dont le rôle est important dans la détection des gammes de fréquence, comprend huit sites indépendants d'épissage alternatif aboutissant à 500 ARNm différents et autant de protéines fonctionnelles.

Les mécanismes à l'origine de l'épissage alternatif sont complexes. Certaines études ont montré par exemple un rôle possible du promoteur dans la régulation des épissages alternatifs en agissant sur l'ARN polymérase II par l'intermédiaire d'un complexe protéique (facteurs de transcription) et d'une séquence régulatrice d'épissage située sur l'intron en amont du gène épissé, l'ISE. Ceci a été mis en évidence par exemple dans le cas de la modulation d'épissage de l'exon 9 du gène codant pour le CFTR (Omim 602421). Cet exemple montre bien l'existence d'un couplage entre la transcription d'une part et l'épissage d'autre part.

Enfin, chez les eucaryotes, il faut noter que certains gènes n'ont pas d'introns et que les procaryotes ne possèdent pas d'introns en général.

#### Formation de l'extrémité 3' de l'ARNm

À la fin du gène, l'ARN polymérase arrête la transcription. L'ARNm synthétisé est clivé et une queue polyadénosine (poly(A), résidus adénine) est ajoutée en son extrémité 3'. Cette étape est réalisée dans le noyau chez les eucaryotes. Cette queue est indispensable pour la stabilité de l'ARNm chez les eucaryotes alors que chez les procaryotes (qui ne possèdent pas de noyau) elle semble jouer aussi un rôle dans la dégradation de l'ARNm (effectuée par une « déanylation » réversible). Chez les procaryotes, cette queue poly(A) paraît donc plus labile. Cette queue est constituée d'environ 200 nucléotides. La machinerie protéique impliquée dans ce mécanisme doit d'abord couper le pré-ARNm. Le site de clivage dans la plupart des cas est situé entre le site hautement conservé AAUAAA (appelé aussi signal de polyadénylation) et une séquence en aval, un motif riche en U ou en GU. Le clivage a lieu le plus souvent au niveau d'un dinucléotide CA. L'addition de la queue poly(A) est effectuée par une enzyme, la polymérase poly(A) [PAP] en présence de protéines spécifiques. Il faut noter que, dans un même gène, plusieurs signaux de polyadénylation peuvent être présents (sites alternatifs) et provoquer l'expression différentielle ou différente de gènes (par exemple, expression tissu spécifique ou après traduction, production d'une protéine différente).

Chez les eucaryotes, les gènes d'histone ne possèdent pas de queue poly(A).

La partie 3' non codante (3'NC) de l'ARNm joue également un rôle important dans la régulation de l'expression d'un gène. Par exemple, certaines séquences, les *cytoplasmic polyadenylation elements* (CPE) ou les *adenylation control elements* (ACE) interviennent dans la répression initiale puis dans l'activation de la traduction de nombreux ARNm maternels au cours de la maturation et du développement précoce des oocytes de souris et de l'activation concomitante [déanylation/re-anylation] qui modulent la longueur de la queue poly(A). Ces régulations se font par interaction avec des complexes protéiques spécifiques.

De manière générale, la partie 3' d'un ARNm est importante pour l'expression des gènes. Elle participe au contrôle de l'exportation hors du noyau, le statut de polyadénylation, la localisation cellulaire, les taux de traduction et de dégradation de l'ARNm.

#### Couplage chapeau 5', épissage et queue poly(A)

L'étude de la transcription a rapidement montré que ces trois réactions étaient liées. Il a pu être ainsi montré que la mise en place du chapeau en 5' stimulait l'épissage et la formation de l'ARNm. Les modalités exactes de ces interactions chez les eucaryotes sont encore mal connues. De la même façon, une influence entre le chapeau 5' et l'addition de la queue poly(A) a pu être mise en évidence. Cet effet semble néanmoins plus modeste.

La polyadénylation est fortement influencée par le phénomène d'épissage, plus particulièrement avec celui de l'exon terminal.

#### Facteurs de transcription

Après décompaction au niveau du promoteur (voir infra), la chromatine doit aussi être en relâchement au niveau des régions codantes. En fait, cette décompaction au cours de la transcription est synchronisée avec l'élongation en présence de l'ARN polymérase. Au moins deux protéines jouent un rôle important dans ce processus, le facteur d'élongation de la transcription spécifique de la chromatine, FACT, et la protéine *elongator*, l'ensemble aboutissant à une régulation fine de l'expression du gène.

De nombreux facteurs de transcription exercent leur activité en fonction du contexte et peuvent souvent être modulés eux-mêmes par d'autres facteurs de telle sorte qu'un même facteur de transcription peut activer un gène et réprimer l'expression d'un autre gène. Ainsi, la transcription d'un gène est-elle sous le contrôle d'une combinatoire de protéines. L'activité des facteurs de transcription est donc soumise à de multiples régulations. Parmi celles-ci, on peut citer le transport de ces facteurs protéiques entre le cytoplasme et le noyau à travers les pores nucléaires des protéines, canaux contrôlant

le passage des macromolécules. Ce transport se fait à l'aide de molécules spécialisées reconnaissant des motifs d'acides aminés.

On distingue deux types de motifs, les signaux d'export nucléaires (*nuclear export signal*, NES) retrouvés dans les protéines exportées du noyau vers le cytoplasme et des signaux de localisation nucléaire (*nuclear localisation signal*, NLS), pour l'importation de protéines dans le noyau.

En réponse à un stimulus, une cellule doit pouvoir rapidement réagir pour inactiver un ou plusieurs facteurs de transcription. Il semble que cela soit réalisé par le système ubiquitine/protéasome permettant la dégradation de ce(s) facteur(s). Après ubiquitination de la protéine par le système ubiquitine ligase, celle-ci est détruite par le protéasome. En cas d'activation prolongée du signal, un nouveau facteur de transcription est synthétisé et la transcription continue. Dans le cas contraire, la transcription s'arrête.

D'autres modifications post-traductionnelles semblent aussi jouer un rôle dans la régulation de la transcription.

#### Promoteur et transcription

Le(s) promoteur(s) d'un gène a(ont) un rôle très important dans la transcription. Il a pu être montré, par exemple, l'importance de celui-ci sur l'épissage alternatif observé dans certains gènes. Le mécanisme exact n'est pas connu.

#### Régulation de la transcription

Elle est complexe. Un ensemble de signaux de transduction dans la cellule va aboutir à l'activation de protéines régulant la transcription. Ces dernières viennent alors se fixer sur des séquences spécifiques du promoteur ou des *enhancers* du gène. Pour arriver au gène, ces facteurs doivent d'abord décompacter la chromatine. Pour cela, ces protéines nécessitent la coopération d'un ensemble de protéines corégulatrices et d'enzymes. Certains facteurs de transcription peuvent cependant accéder à l'ADN même en état condensé dans les nucléosomes (par exemple, le récepteur aux glucocorticoïdes)

#### Enzymes agissant sur les histones

En général, sur chaque gène, de nombreux sites de fixation spécifiques de multiples facteurs de transcription sont présents. Ils agissent de concert avec des corégulateurs dont on distingue deux types, les coactivateurs et les corépresseurs. Ces corégulateurs jouent un rôle fondamental dans la genèse de réponse promoteur-spécifique et tissu-spécifique. Une autre forme de corégulateur, appelé médiateur, complexe multiprotéique, joue aussi un rôle important dans la régulation d'un grand nombre de gènes. De manière générale, de nombreux corégulateurs sont des enzymes modulant la

structure de la chromatine, ce qui souligne l'importance de l'empaquetage de l'ADN pour son expression.

Ces enzymes sont de deux classes : celles remodelant les complexes nucléosomes (modification de conformation par exemple) et celles catalysant les modifications post-traductionnelles des histones.

Quatre classes de modificateurs d'histone ont été décrites :

- les histones acétyltransférases (HAT) : elles initient la première modification conformationnelle de la chromatine. Par exemple, l'acétylation des histones H3 et H4 à certains résidus lysine sont associés à une forte activité transcriptionnelle. De manière générale, les histones acétylés sont associés à des gènes activés et, inversement, les histones déacétylés à la répression des gènes au sein de l'hétérochromatine. Par conséquent, en général, les HAT activent donc la transcription et les HDAC (voir infra) inhibent la transcription. Bien que ce mécanisme soit le plus souvent observé comme précédemment décrit, le cas inverse est aussi possible montrant bien la complexité du système ;
- les histones déacétylases (HDAC) : recrutées par les répresseurs de transcription, elles aboutissent à la déacétylation des histones. Ces enzymes sont indispensables pour la répression ;
- les histones méthyltransférases (HMT) : comme les HAT, elles agissent dans la queue N-terminale des histones. Elles jouent un rôle fondamental dans l'activation ou la répression d'un grand nombre de gènes. La conséquence de leur action est fonction du résidu lysine ou arginine méthylé. Le mécanisme précis est encore mal connu ;
- les histones kinases (phosphorylation).

En fait, les conséquences des modifications d'histones semblent bien plus complexes. L'interdépendance des modifications sur les histones et la combinatoire qui en résulte joue un rôle fondamental dans les conséquences fonctionnelles et démontre la complexité du système.

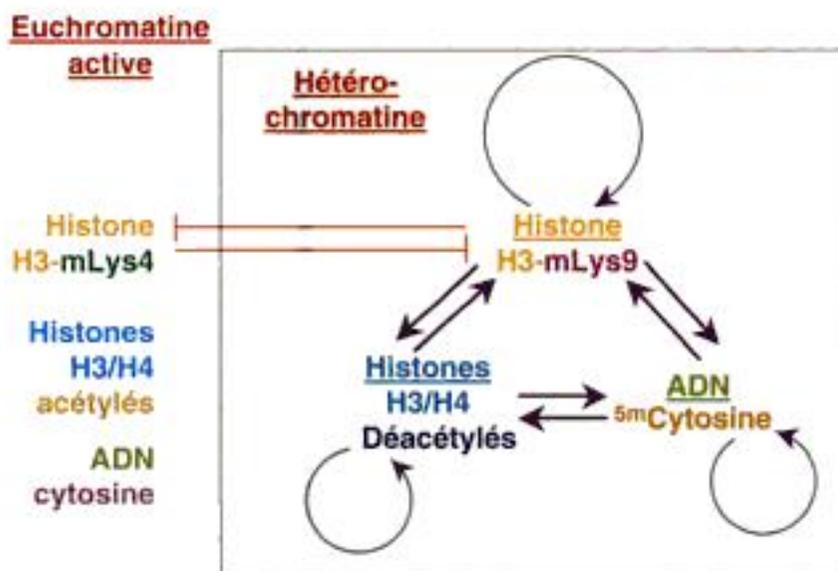
#### Code histone

Cette théorie s'applique aux eucaryotes, les procaryotes ne possédant pas d'histones. L'hypothèse du code histone a été proposée à partir des différentes observations sur les modifications des histones, leur nombre et leur variété ainsi que sur l'interdépendance de celles-ci sur la fonction des histones. De manière schématique, le code histone est l'ensemble des modifications des histones qui provoquent par leurs effets combinatoires et/ou séquentiels sur la ou les queues histones des modifications fonctionnelles uniques.

Les modifications des histones aboutissant à une réorganisation de la structure chromatinienne et jouant un rôle dans la régulation transcriptionnelle constituent un phénomène combinatoire fondamental. Certains auteurs ont suggéré un code histone résultant de la

combinaison des modifications des queues histones. Ce code est considéré comme un état de reconnaissance aboutissant à la réorganisation de la chromatine localement. Contrairement au code génétique qui est permanent sur le génome, le code histone est temporel. Ces modifications réversibles post-traductionnelles telles que l'acétylation (par l'intermédiaire d'enzymes comme les acétylases et les déacétylases), la méthylation, la phosphorylation, l'ubiquitination et l'ADP-ribosylation des histones nécessitent l'action d'enzymes spécifiques modifiant l'extrémité N-terminale des histones. Par exemple, les extrémités N-terminales des histones acétylées interagissent spécifiquement avec des domaines protéiques (appelés bromodomains) présents au niveau de complexes protéiques associés à la chromatine. La lysine (« lysine 9 ») méthylée des queues histone H3 est un site de reconnaissance spécifique pour un complexe protéique au niveau d'un domaine appelé chromodomaine (ce dernier étant en général nécessaire pour la formation de l'hétérochromatine et l'expression des gènes). L'action des protéines reconnaissant des sites de fixation sur ces histones modifiées (on parle de bromodomains et de chromodomains) et lisant ce code aboutirait à des transitions entre les différents états conformationnels. Par exemple, l'hypoacétylation des histones (ainsi que l'état de méthylation de certaines histones en N-terminal) a été associée à la formation d'hétérochromatine et l'inhibition de l'expression des gènes alors que l'hyperacétylation est associée à l'euchromatine (figure 1.9).

Il faut noter que ce code histone dépend en partie du code génétique, ce dernier contenant la majorité des informations déterminant l'état des histones (comme les *enhancers* par exemple). Par ailleurs, l'interaction avec des facteurs de transcription interagissant avec les nucléosomes provoque des modifications conformationnelles impliquées dans ce code. Cette théorie aux contours encore mal définis est l'objet d'intenses études.



**Figure 1.9** Marqueurs épigénétiques et leur influence sur l'état de la chromatine (hétérochromatine ou euchromatine). mLys9 : lysine méthylée ; <sup>5m</sup>cytosine : cytosine méthylée.

L'étude de la pathologie moléculaire humaine démontre sa réalité.

### Code épigénétique

De grands fragments du génome humain (dont un grand nombre de séquences répétées) sont sous forme de chromatine inactive, c'est l'hétérochromatine. Ces séquences inactives sont sous l'influence de la modification d'histones, la présence de complexes protéiques et d'un état spécifique de l'ADN. Cet état silencieux peut être transmis au cours des mitoses et des méioses. Outre les modifications d'acétylation et de méthylation des histones précédemment décrites, une autre modification joue un rôle fondamental dans le maintien de l'hétérochromatine et l'expression des gènes, c'est la méthylation des cytosines en 5' (5mC). Cette modification des cytosines est assurée par un groupe d'enzymes, les cytosine-ADN méthyltransférases (*Dnmt*). Ce phénomène mal connu est un processus dynamique observé par exemple au cours du développement précoce de l'embryon où on constate un effacement des méthylations et une reprogrammation de celles-ci sans que l'on sache quels processus gouvernent ces modifications. Certaines études proposent un lien entre la méthylation des cytosines sur l'ADN d'une part et la méthylation et acétylation des histones d'autre part.

Enfin, un mécanisme décrit chez les plantes est la méthylation de l'ADN dépendant de l'ARN (RdDM). Une hypothèse (qui reste à prouver chez les eucaryotes) est que l'ARN pourrait dans certains cas être un cofacteur des *Dnmt* pour les méthylations de cytosine de novo (mécanisme proche de l'interférence d'ARN, voir chapitre correspondant).

La méthylation des cytosines est transmise aux cellules filles au cours de la mitose et probablement aussi de la méiose. À l'inverse, la déméthylation des cytosines est possible grâce à l'action de déméthylases. De manière générale, les phénomènes de méthylation différentiels du génome humain appartiennent au domaine de l'épigénétique, domaine encore mal connu dans lequel la séquence nucléique de l'ADN n'est pas altérée mais son état de méthylation est modifié. Ce domaine est en cours d'exploration dans le cadre du projet humain d'épigénome (*human epigenome project*, HEP).

### Méthylation et transcription

Certaines zones du génome humain contiennent des régions riches en CpG : on les appelle les îlots CpG. Ces îlots sont définis par une taille minimum de 200 pb et un contenu en [G+C] supérieur à 50 %. Ces îlots sont essentiellement trouvés en 5' de nombreux gènes au niveau du promoteur, autour du site d'initiation de la transcription et se continuent dans un ou plusieurs exons. Dans le cas des gènes de ménage par exemple (*house-keeping genes*), les îlots CpG ne sont générale-

ment pas méthylés quel que soit le niveau d'expression du gène. Par ailleurs, dans un certain nombre de cas, il existe une relation inverse entre la méthylation des îlots CpG associés au promoteur et l'expression d'un gène, ce qui signifie de manière simplifiée que la méthylation inhibe l'expression d'un gène en inhibant la transcription. Ceci est essentiellement observé au niveau des promoteurs. Il est rare en effet que la méthylation des parties codantes ou en aval du promoteur inhibe la transcription : dans ce cas, on observe le plus souvent une augmentation de l'expression d'un gène. Dans certains cas par ailleurs, l'état de méthylation peut être tissu-spécifique. Enfin, dans l'ADN normal, les cellules différenciées ont un profil de méthylation spécifique transmis aux cellules filles au cours de la réplication.

### Méthylation et séquences répétées

Le génome humain est constitué d'un grand nombre de séquences répétées (au moins 35 % du génome) telles que les éléments transposables, les LINE, les séquences Alu. La méthylation de ces séquences jouerait un rôle important pour empêcher leur activation. En effet, les conséquences de leur activation sur le génome pourraient être dramatiques. À titre d'exemple, en cas de perte de méthylation d'une séquence Alu, un élément transposable pourrait être activé et s'insérer par rétrotransposition dans un gène (mutation par insertion) provoquant ainsi son inactivation (voir infra, interférence de transcription). La méthylation jouerait donc un rôle protecteur en inhibant l'activation de la séquence.

### Méthylation et développement

La méthylation du génome est un phénomène complexe et dynamique. Elle peut être ou non modulée, comme dans certains cas développés plus bas. Nous citerons :

- l'inactivation de l'X ;
- l'empreinte parentale ;
- le développement embryonnaire.

### Interférence de transcription

L'interférence de transcription est la perturbation d'une unité transcriptionnelle par l'activation d'une autre unité transcriptionnelle soit située à proximité (gènes en tandem par exemple), soit située à distance. Ce phénomène a été décrit par exemple chez les plantes où des transposons ayant perdu leur état de méthylation ont été activés et ont réprimé l'expression d'autres unités transcriptionnelles. De nombreux mécanismes expliquent l'interférence de transcription. On peut citer, outre l'état de méthylation, les modifications des histones, les ARN antisens (interférence d'ARN par exemple), la compétition en *cis* entre facteurs de transcription.

### Stabilité de l'ARNm dans la cellule

Certains auteurs ont étudié le niveau global des ARNm et leur durée de vie dans la cellule. De manière générale, la cellule répond rapidement et de manière coordonnée aux variations physiologiques qu'elle subit. Par exemple, dans les cellules de la lignée érythroïde, l'induction de la transcription du gène globine est couplée à la déstabilisation d'un grand nombre d'autres ARNm permettant ainsi la redirection des ressources cellulaires vers la production d'hémoglobine. De même, les hormones stéroïdes agissent sur la stabilité des ARNm ainsi que la transcription des gènes cibles. Les mécanismes de régulation, d'interaction et de coordination des synthèses et dégradation des ARNm et des protéines sont néanmoins encore mal connus.

### Exportation de l'ARNm du noyau vers le cytoplasme

La compartementalisation de la cellule en noyau et cytoplasme est une propriété caractéristique des eucaryotes. Chez ces derniers, l'ARNm mature doit ensuite passer dans le cytoplasme pour être transformé (« traduit ») en protéine. L'exportation de l'ARNm est effectuée par un processus complexe faisant intervenir des protéines reconnaissant et se fixant à l'ARNm et lui permettant de passer à travers des canaux, les pores du noyau (constitués d'un grand nombre de protéines dont les nucléoporines). Il s'agit d'un processus actif faisant intervenir des protéines de transport transmembranaires (membrane du noyau), les karyophérines (comprenant les importines et les exportines). Même si ces molécules jouent un rôle fondamental dans l'import/export des protéines, elles interviennent aussi (essentiellement les exportines) dans le passage des ARN (notamment ARNt, ARNr et ARNsn). Pour les ARNm en général, d'autres protéines interviennent. Il a en effet pu être démontré qu'il existait un couplage entre l'épissage et l'exportation (à travers les nucléoporines) des ARNm en présence d'un complexe protéique (appelé hétérodimère d'export de l'ARNm). Dès que l'ARNm pénètre dans le cytoplasme, les ribosomes se fixent sur ce dernier.

### Traduction

Elle a lieu dans le cytoplasme pour les eucaryotes. La séquence des codons sur l'ARNm dirige la synthèse du polypeptide qui aboutira dans un deuxième temps à la protéine fonctionnelle. La traduction de l'ARNm en protéines est réalisée à partir d'un complexe de ribonucléoprotéines appelé ribosomes. Les ribosomes cheminent le long de l'ARNm et, en association avec l'ARNt combiné

à un acide aminé (aminoacyl-ARNt) et un complexe protéique, permettent la synthèse du polypeptide.

### Éléments de structure générale du ribosome et traduction

Le ribosome est un large complexe de ribonucléoprotéines constitué de deux sous-unités 20S et 60S. Chaque sous-unité est constituée de trois sites de fixation désignés A (*aminoacyl*) qui accepte l'aminoacyl-ARNt, P (*peptidyl*) sur lequel vient se fixer l'ARNt contenant le polypeptide naissant et E (*exit*) qui fixe l'ARNt déacylé avant que celui-ci ne quitte le ribosome.

La petite sous-unité se fixe à l'ARNm et à l'anti-codon de l'ARNt. Elle vérifie la complémentarité des bases entre le codon sur l'ARNm et l'anticodon sur l'aminoacyl-ARNt correspondant.

La grande sous-unité fixe le bras accepteur de l'aminoacyl-ARNt et catalyse la formation de la liaison peptidique entre l'acide aminé présent sur l'ARNt au site A et au polypeptide naissant fixé à l'ARNt du site P.

Les deux sous-unités sont impliquées dans la translocation, mécanisme par lequel l'ARNt et l'ARNm se déplacent dans le ribosome codon par codon (un codon à la fois).

Après reconnaissance du site d'initiation de la traduction par les ribosomes en présence de facteurs d'initiation de la traduction et des sous-unités ribosomales, l'élongation en polypeptide puis la fin de la traduction se poursuit (au cours de la traduction, le ribosome peut faire des erreurs dont le taux est estimé à 1/10 000 codons).

Le polypeptide néoformé subit ensuite des modifications post-traductionnelles puis prend sa conformation définitive donnant une protéine fonctionnelle.

### Initiation de la traduction

Dans le cas des bactéries, l'initiation de la traduction fait intervenir dans la grande majorité des cas une séquence spécifique, la séquence de Shine-Dalgarno (située sur l'ARNm), complémentaire de l'extrémité 3' de l'ARN 16S de la sous-unité ribosomale 30S, et impliquée dans la traduction chez les procaryotes (on l'appelle aussi site de fixation du ribosome, RBS, *ribosome binding site*). Cette séquence est située 10–13 pb en amont du codon d'initiation de la traduction (AUG) dont la séquence optimale chez *Escherichia coli* (elle varie selon les procaryotes) est 5'-AGGAGG (le motif le plus conservé étant AGGA). Par exemple, 5'-AGGAGGACAGCUAUG (la séquence soulignée correspond à la séquence Shine-Dalgarno et le double soulignement au codon initiateur).

Chez les eucaryotes, la plupart des gènes présentent une séquence impliquée dans la traduction (RBS), la

séquence de Kozak (5'-A[/G]CCACCAUGG, AUG souligné correspond au site d'initiation de la traduction).

### Élongation au cours de la traduction

Chez les eucaryotes, après la sélection de l'ARNt initiateur en présence de facteurs d'initiation, l'ensemble se fixe à la sous-unité ribosomale 40S pour former le complexe de pré-initiation 43S. La fixation de 43S à l'ARNm se fait au niveau du chapeau m<sup>7</sup>G situé en 5' de l'ARNm. Parfois, comme c'est le cas par exemple pour certains virus (exemple, virus de l'hépatite C) ou certains gènes cellulaires (par exemple, *vascular endothelial growth factor*, VEGF, Omim 192240), l'initiation de la traduction peut commencer en des sites différents indépendants du chapeau, ce sont les sites d'entrée des ribosomes (*internal ribosomal entry site*, IRES). L'ensemble du complexe ribosomal va se déplacer de la partie 5' non codante de l'ARNm à la recherche du codon d'initiation (AUG, *scanning*) et former un complexe d'initiation 48S dans lequel l'anticodon présent sur l'ARNt-Met se fixe au codon initiateur (AUG, codon initiateur principal). Dans de rares cas (certains ARNm viraux ou cellulaires), les codons initiateurs peuvent être CUG, GUG, ACG et AUU. Quand le complexe d'initiation migrant le long de la partie 5'NC de l'ARNm rencontre le site AUG, le complexe 48S est remplacé par les sous-unités 40S et 60S. La traduction de l'ARNm démarre après l'association de l'ARNt initiateur (ARNt-Met), de l'ARNm et des sous-unités 40S et 60S (ces deux dernières constituant le ribosome 80S). L'ensemble de ce processus fait appel à plusieurs protéines, les facteurs d'initiation (*eukaryotic initiation factors*, eIF). À la fin du processus d'initiation, l'ARNt-Met initiateur se trouve dans le site P. Le site A du ribosome est alors vide. L'élongation va démarrer. De manière simplifiée, un aminoacyl-ARNt est amené au site A en présence d'un complexe protéique. Ces molécules vont lire les codons sur l'ARNm (c'est-à-dire les triplets de bases, par exemple : GAU → acide aspartique). L'association codon de l'ARNm et de l'anticodon complémentaire sur l'aminoacyl-ARNt provoque une modification de conformation du ribosome et une stabilisation de l'association ARNm/aminoacyl-ARNt. Cette réaction provoque la libération de l'extrémité aminoacyl de l'ARNt au site A et le passage de l'ARNt vers le site peptidyl transférase par un processus complexe appelé accommodation. Au niveau du site P, la formation de la liaison peptidique nécessite la déacylation de l'ARNt et provoque le transfert du peptide en cours de formation à l'ARNt au niveau du site A. Après transfert du peptide, on a donc un ARNt déacylé au site P et un ARNt-peptidyl au site A. L'ARNt déacylé est ensuite transféré au site E du ribosome, l'ARNt-peptidyl au site P. Le site A redevient prêt à recevoir un nouveau aminoacyl-ARNt. L'élongation peut se poursuivre.

## Terminaison

Celui-ci démarre quand le site A rencontre le codon-stop. La reconnaissance du codon stop est effectuée en présence de protéines, notamment les facteurs de libération de classe I (*release factors*, eRF1) qui reconnaissent les codons-stop. Cette reconnaissance au site A provoque une lyse puis la libération de la chaîne peptidique fixée sur l'ARNt fixé au site P. Après libération du polypeptide, le ribosome reste fixé sur l'ARNm avec un ARNt déacylé au site P. Pour que le ribosome se dissocie de l'ensemble et puisse être de nouveau utilisé pour une autre synthèse, un facteur protéique, le facteur de recyclage du ribosome (*ribosome recycling factor*, RRF) est nécessaire.

## Modifications post-traductionnelles

Une fois la protéine formée, celle-ci va encore subir des transformations dites post-traductionnelles : glycosylation, acétylation, phosphorylation, etc. L'ensemble de ces transformations est complexe et est réalisé au cours des passages dans le réticulum endoplasmique et l'appareil de Golgi.

Ces mécanismes ne seront pas développés car ils sortent du cadre de cet ouvrage.

## Contrôle général de la traduction

Un certain nombre de facteurs extérieurs interviennent dans la traduction. De manière générale, le contrôle global de la traduction est la conséquence de modifications post-traductionnelles des facteurs d'initiation et des protéines ribosomales alors que le contrôle spécifique de la traduction est assuré par des motifs nucléotidiques dans les régions non traduites de l'ARNm. Certains contrôles sont assurés par régulation du site d'initiation en 5'NC, d'autres par régulation de la stabilité de l'ARNm (en 3'NC par exemple). Ainsi, certaines hormones ou des facteurs de croissance peuvent stimuler la traduction. À l'opposé, le stress oxydatif, par exemple, diminue la traduction.

Certaines molécules peuvent se fixer sur l'ARNm au niveau de séquences comme la séquence *iron responsive element* (IRE) et modifier la traduction d'un gène en agissant sur sa stabilité (par exemple l'ARNm codant pour la ferritine par fixation de la protéine IRP1 [*iron regulatory protein*] sur la séquence IRE, cette interaction étant modulée par le fer). Par ailleurs, des effets directs de régulation de la partie 3'NC sur l'initiation de la traduction sont aussi décrits.

L'inhibition de l'initiation de la traduction est assurée par des éléments de régulation négative séquence-spécifique située en 5'NC. Ces régulations négatives sont complexes. Schématiquement, on en distingue deux types, des structures secondaires spécifiques de l'ARNm et des cadres de lecture particuliers en amont du cadre de

lecture physiologique (*upstream open reading frame*, uORF) situé en 5'NC.

- Les structures secondaires inhibitrices de l'ARNm sont constituées de séquences autoccomplémentaires stables repliées en boucle (*stem-loop*). Quand ces structures sont proches du chapeau 5', elles peuvent interférer avec l'assemblage du complexe de pré-initiation (par exemple, pour l'ARNm codant pour l'ornithine décarboxylase ; Omim 165640). Si elles sont situées plus en aval, elles peuvent être suffisamment stables pour résister au « déroulement » de l'ARNm par l'hélicase associée aux ribosomes et empêcher la vérification (*scanning*) du complexe par les ribosomes. Cette « résistance » peut être stabilisée par la fixation de protéines et provoquer une inhibition de la traduction (par exemple, les IRE avec les IRP sur le gène codant pour la ferritine).
- L'inhibition de la traduction peut aussi être la conséquence de la reconnaissance de codon AUG (uAUG) en amont du codon initiateur physiologique au cours du criblage (*scanning*) par le ribosome 40S. Ces uAUG sont suivis de codons-stop prématurés provoquant l'arrêt de la traduction (par exemple, sur l'ARNm codant pour l'interleukine 15). La traduction peut cependant être réinitiée et le ribosome 40S peut continuer le *scanning* jusqu'à l'AUG suivant (par exemple, pour l'ARNm de la thrombopoïétine ; Omim 600044). Il a pu être démontré que la réinitiation de la traduction pouvait être réalisée si une distance minimum de 16 nucléotides sépare les deux AUG.
- L'uAUG peut aussi être responsable de la formation d'un peptide qui inhibera lui-même le début de l'initiation (par exemple, ARNm codant pour le récepteur adrénergique bêta-2 ; Omim 109690).
- Ce n'est cependant pas toujours le cas. Certains promoteurs alternatifs utilisent ces uAUG pour donner des protéines aux extrémités 5' variables. Dans ce cas, les uAUG sont en phase de lecture avec le produit de traduction principal (par exemple c-myc ; Omim 190080). L'utilisation de tels sites de traduction alternatifs peut aboutir à des protéines aux extrémités N-terminales différentes et pouvant être actives, inactives ou aux propriétés différentes (par exemple, le facteur de transcription C/EBPβ activateur en cas d'utilisation d'un AUG en amont et inhibiteur en cas d'utilisation d'un AUG en aval).
- Enfin, dans certains cas, l'inhibition de la traduction peut être « surpassée » par l'initiation au niveau d'un IRES situé en aval d'une structure secondaire en boucle ou d'un uORF (par exemple, ARNm du VEGF, Omim 192240).

### Initiation alternative de la traduction

Outre la traduction telle que décrite précédemment, dans certains cas, l'initiation ne dépend pas du chapeau

mais d'un site d'initiation interne de la traduction encore appelé site interne d'entrée du ribosome (IRES). Ce mécanisme alternatif est observé au cours de la traduction de certains virus notamment (exemple, virus de l'hépatite C). Cela a aussi été rapporté dans certaines cellules.

### Régulation de l'expression d'un gène

Cette régulation est fort complexe. Nous soulignerons certains points importants.

L'ensemble des étapes depuis la transcription jusqu'à la synthèse protéique est en fait étroitement interdépendant. Il s'agit d'un processus continu. Par exemple, les facteurs de transcription jouent un rôle important dans le recrutement des protéines impliquées dans la formation du chapeau et la naissance de l'ARNm. L'épissage du pré-ARNm stimule l'élongation au cours de la transcription. Ainsi, des nombreuses études réalisées, il ressort que les facteurs protéiques responsables de chacune des étapes sont fonctionnellement et même parfois physiquement connectés entre eux. La régulation des différentes étapes est contrôlée à de nombreux niveaux et dépend du (ou des) gène(s) concerné(s). Il n'existe pas de règle générale permettant de décrire les différents modes de régulation de l'expression d'un gène. Ainsi, certains gènes dont l'expression doit être rapidement et précisément contrôlée sont souvent rapidement transcrits et traduits et la demi-vie de la protéine correspondante courte. Alors que d'autres gènes de rôle plus général (par exemple, les gènes de ménage, *house-keeping genes*) sont souvent plus lentement induits et possèdent une demi-vie plus longue. La régulation peut donc se faire à plusieurs stades de l'expression d'un gène et avoir différentes conséquences sur la protéine synthétisée. Brièvement, on peut citer :

- la régulation de l'activité de la protéine : en modifiant la quantité de protéines synthétisées. Cela peut être réalisé au niveau transcriptionnel et/ou traductionnel et/ou post-traductionnel ;
- la génération d'une diversité protéique : la combinaison des exons (épissage alternatif) amène dans certains cas à l'obtention de plusieurs protéines (isoformes par exemple). On obtient la même chose avec l'*editing* de l'ARN qui modifie la séquence codante de l'ARNm (voir infra). Il en est de même au cours des modifications post-traductionnelles ;
- la régulation par compartimentalisation : on obtient un contrôle de la concentration locale de la protéine. Après formation de la protéine, celle-ci est adressée vers un compartiment défini (mitochondrie, peroxy-some par exemple). Certaines séquences d'acides aminés permettent cet adressage (par exemple, dans la mitochondrie).

### Phénomène particulier : l'édition de l'ARN

De manière générale, l'édition (*editing*) de l'ARN est une modification post-transcriptionnelle se traduisant par la modification de la séquence nucléotidique de l'ARN à une ou plusieurs positions (figure 1.10).

#### Modification de la cytosine en uracile

L'édition a initialement été mise en évidence en observant l'addition ou la délétion d'un ou de plusieurs résidus uridyl sur l'ARN. Ce phénomène a été initialement constaté dans les mitochondries des trypanosomes. Par la suite, chez le même organisme, l'édition a été observée pour la transformation d'une cytosine en uracile. En présence d'un complexe protéique comprenant une déaminase, *ApoBec-1* (dont une séquence spécifique, UUUN (A/U) U, située à 3 pb du site édité, a été identifiée comme site de fixation), et appelé éditosome, l'édition permet de modifier la séquence d'un ARN simple brin par rapport à la séquence d'ADN génomique. Les conséquences fonctionnelles qui en découlent sont importantes (changement d'acide aminé, nouveau codon d'initiation, etc.). Ce mécanisme a été retrouvé chez les virus (par exemple, le virus de la rougeole) et les eucaryotes, l'exemple le plus connu étant celui de l'apolipoprotéine B (ApoB, Omim 107730) pour lequel existent deux formes (ApoB48 et ApoB100) de métabolisme totalement différent et codées par le même gène ApoB. L'une est synthétisée dans le foie (ApoB100) et l'autre dans l'intestin (ApoB48). Dans ce cas particulier, l'édition du gène au niveau d'une séquence CAA est effectuée par déamination de la cytosine (ce qui la transforme en uridine : C→U) et produit un codon-stop (UAA) et la formation d'une protéine tronquée, ApoB48. Cette édition survient après l'épissage, le spliceosome inhibant cette forme d'édition (contrairement à l'édition A→I qui se produit avant l'épissage, voir infra).

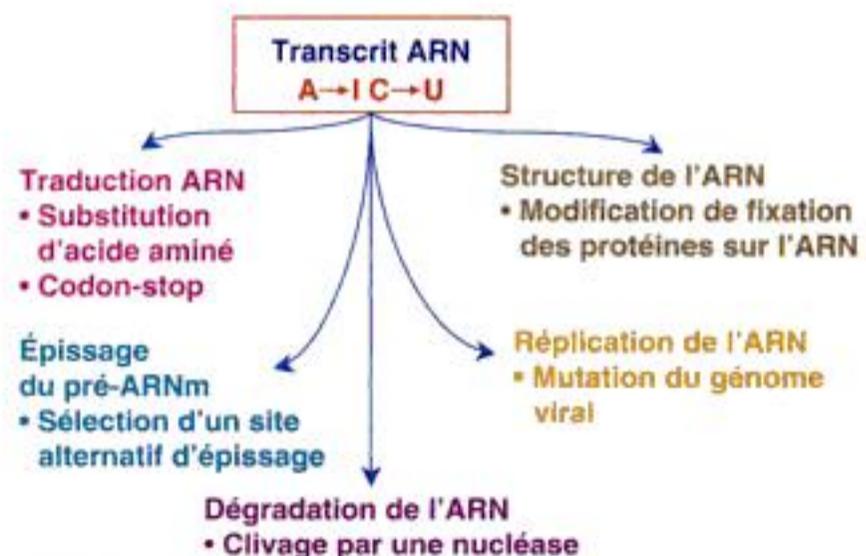


Figure 1.10 Mécanismes d'édition (*editing*).

### Modification de l'adénine en inosine

Un autre type d'édition, plus fréquent chez les eucaryotes, a aussi été décrit : la modification d'une (ou de plusieurs) adénosine en inosine (A→I, l'inosine étant reconnue comme un G par les ribosomes au cours de la traduction) sur le pré-ARNm en présence d'un complexe protéique et d'un groupe d'enzymes spécifiques des ARN double brin, la famille ADAR (*adenosine deaminase acting on RNA*). Cette réaction de déamination a été décrite chez les métazoaires (par exemple, *Xenopus* spp.). Ce mécanisme a aussi été retrouvé chez les eucaryotes (par exemple, sur les transcrits du gène codant pour les récepteurs de glutamate) et les virus (par exemple, virus de l'hépatite delta). Certaines études évoquent une interaction entre les mécanismes d'épissage d'une part et l'édition d'autre part.

De manière générale, le phénomène d'édition de l'ARN semble un processus fondamental chez les mammifères. Beaucoup d'inconnues demeurent cependant, notamment sur la régulation de ce phénomène et le nombre de gènes soumis à ce processus. D'autres mécanismes de modification des ARN restent probablement à découvrir.

### Empreinte génomique – Bases

La majorité des gènes présentent une expression bi-allélique avec un allèle d'origine paternelle et un autre d'origine maternelle. Dans de rares cas, l'expression d'un gène est dépendant de l'origine parentale de celui-ci : c'est l'empreinte génomique, conséquence de modifications épigénétiques (c'est-à-dire pour lesquelles aucune modification de la séquence d'ADN n'est observée). L'empreinte génomique est donc caractérisée par l'expression mono-allélique d'un gène. Un gène soumis à ce phénomène sera exprimé exclusivement soit à partir de l'allèle d'origine maternelle, soit exclusivement à partir de l'allèle d'origine paternelle. À ce jour, une trentaine de gènes soumis à empreinte génomique ont été identifiés (voir le chapitre « Sites internet de biologie moléculaire »). Leur rôle est fondamental dans le développement, la croissance fœtale ou la régulation cellulaire. Une anomalie de l'empreinte génomique a été décrite dans certaines maladies héréditaires et au cours de la carcinogenèse.

Par exemple, au cours de la gamétogenèse, certains gènes (environ une centaine répertoriés) sont soumis à empreinte génomique. De manière générale, ces gènes jouent un rôle important dans la croissance, le développement et la suppression des tumeurs. Cette caractéristique sera maintenue au cours des divisions et au cours de la différenciation cellulaire. L'expression mono-allélique est établie au niveau moléculaire. Parmi les mécanismes en cause, ont été décrits la méthylation différentielle, la structure chromatinienne et des élé-

ments de régulation. Ces mécanismes sont encore mal connus.

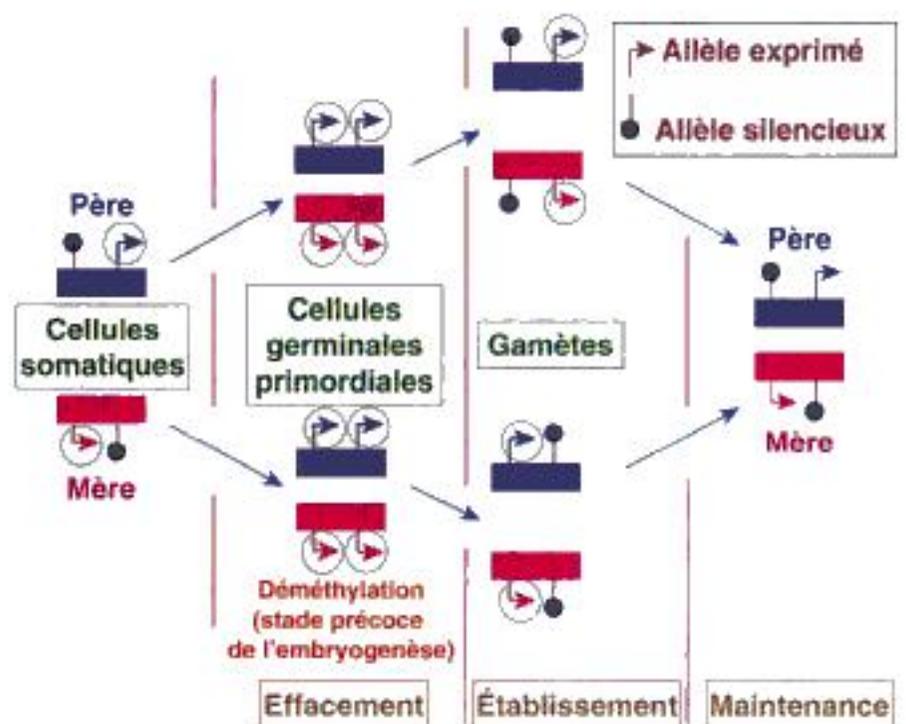
### Empreinte génomique et développement embryonnaire

Au cours des premiers stades du développement dans les cellules primordiales germinales, on observe une remise à zéro (*reset*) de la méthylation de l'ADN. L'effacement de l'empreinte génomique à ce stade semble fondamental pour son ré-établissement de novo qui se poursuit au cours de l'embryogenèse (effectué notamment par des méthyltransférases). L'empreinte génomique est transmise aux cellules filles au cours des mitoses successives (figure 1.11). Il a pu être démontré que la méthylation de l'ADN est un processus fondamental (d'autres mécanismes existent aussi) de cette empreinte génomique. On observe donc pour certains gènes une méthylation différentielle de l'ADN.

Les gènes soumis à empreinte génomique semblent organisés en groupes (*cluster*) dans de larges domaines chromatinien. La structure chromatinienne joue par ailleurs un rôle important au niveau de ces domaines pour assurer l'empreinte génomique.

### Mécanisme de l'empreinte génomique

Le mécanisme précis régissant l'empreinte génomique est encore mal compris. De nombreuses études ont été réalisées notamment au niveau du locus *Igf2/H19*, domaine d'environ 100 kb situé sur le chromosome 11p15.5. Dans cette région, deux gènes ont été particulièrement étudiés, *insulin-like growth factor 2* (*Igf2*, Omim 147470, jouant un rôle important notamment au cours du développement) et *H19* (Omim 103280, de



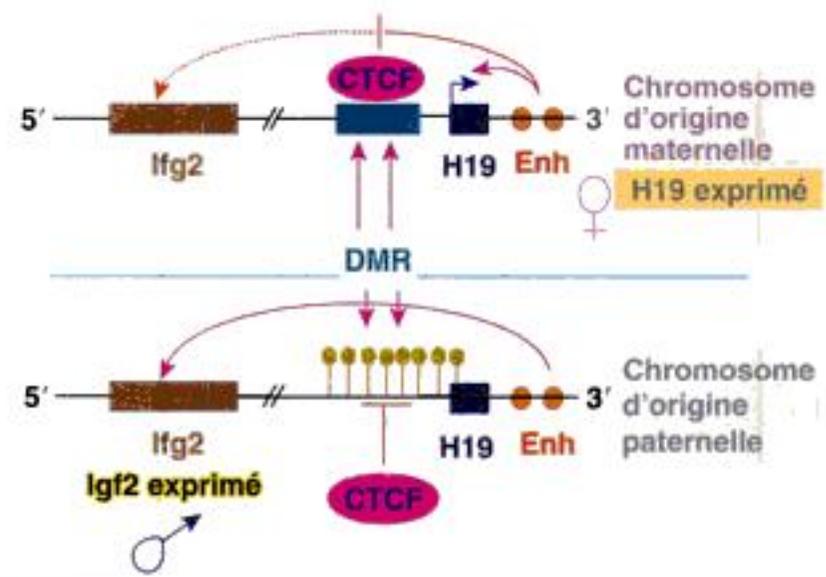
**Figure 1.11** Empreinte génomique et développement embryonnaire. Au cours du développement, les gènes soumis à empreinte génomique sont remis à zéro très précocement avant d'être reprogrammés. L'empreinte est ensuite transmise aux cellules filles.

fonction inconnue et dont le transcrit est un ARN non codant). Nous prendrons ce locus comme exemple pour illustrer de manière simplifiée la complexité du mécanisme de l'empreinte génomique. À partir de ces études, plusieurs hypothèses ont été évoquées.

Une localisation chromatinienne, appelée région méthylée différemment (*differentially methylated region*, DMR) et caractérisée par une sensibilité particulière aux nucléases, semble jouer un rôle important pour le contrôle de l'empreinte génomique du locus *Igf2/H19*. À noter que d'autres noms ont été aussi donnés au DMR tels que domaine méthylé différemment (*differentially methylated domain*, DMD) ou région de contrôle d'empreinte (*imprinting control region*, ICR). Cette zone d'environ 2 kb située en amont du gène *H19* contient des zones d'hyperméthylation aux DNases sur l'allèle non méthylé maternel, ce qui n'est pas le cas sur l'allèle méthylé paternel. On observe un remodelage chromatinien différentiel entre l'allèle maternel et paternel. L'acétylation différentielle des histones joue probablement un rôle. Ainsi, la déacétylation des histones participe à la répression de l'expression notamment au niveau de *H19* d'origine paternelle. Par ailleurs, la méthylation des histones jouerait aussi un rôle dans la répression de l'expression du gène *H19* d'origine maternelle. Ces modifications épigénétiques associées aux modifications conformationnelles de la chromatine sont des mécanismes importants mais pas uniques dans l'empreinte génomique.

### DMR

Il a pu être démontré qu'en général l'activité du DMR est héritée de la mère et est associée à une structure chromatinienne ouverte et déméthylée. Elle interagit avec certains complexes protéiques. Par opposition, le DMR hérité du père est méthylé et localisé dans une structure chromatinienne fermée. Ce DMR fait partie des isolateurs ou éléments frontières, séquences de régulation délimitant des domaines génomiques indépendants. Ces éléments peuvent bloquer l'action de promoteurs et/ou d'*enhancers* en fonction de leur position. Cette action est dépendante de l'interaction avec un facteur protéique nucléaire, le CTCF (*CCCTC-binding factor*, Omim 604167, protéine décrite comme activateur/répresseur d'un certain nombre de promoteurs). Cette protéine se fixe sur le chromosome maternel non méthylé (mais ne peut se fixer sur le DMR paternel méthylé). Dans le cas du locus *Igf2*, l'activité « isolateur » de DMR est limitée à ses propriétés de blocage d'*enhancer* méthylation-dépendant (figure 1.12). À noter que d'autres DMR ont été décrits et peuvent avoir des propriétés différentes de celles illustrées dans notre exemple (certains DMR par exemple sont méthylés sur le locus inactif et méthylés sur le locus actif).



**Figure 1.12** Schéma simplifié de l'action du DMR dans la région *Igf2/H19*. Exemple d'empreinte génomique et de sa régulation. Un *enhancer* spécifique est partagé par le gène *Igf2* exprimé chez le père et par le gène *H19* (exprimant un ARN non codant) exprimé chez la mère. La région DMR est une « frontière » fermée quand il est non méthylé (*Igf2* n'est pas exprimé) et ouverte quand il est méthylé (*Igf2* exprimé). Un facteur protéique nucléaire, le CTCF (*CCCTC-binding factor*), interagit sur la séquence DMR non méthylée de la mère (et non sur celle méthylée du père) et joue le rôle d'isolateur entre deux régions du génome. La fixation de la protéine CTCF permet en effet l'expression du gène *H19* et inhibe l'expression du gène *Igf2*. La méthylation chez le père du DMR ne lui permet plus de jouer son rôle d'isolateur et l'*enhancer* (*Enh*) peut stimuler le gène *Igf2*.

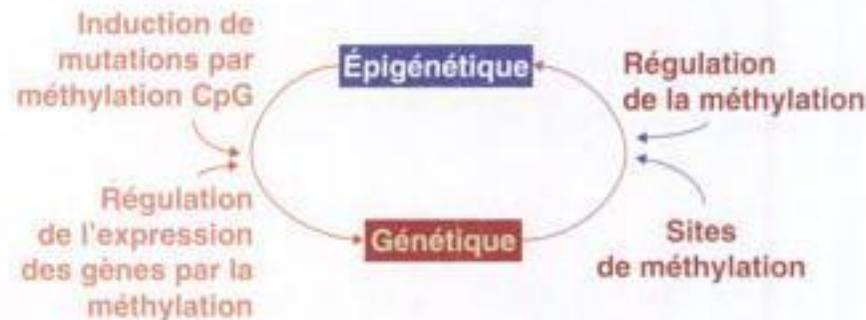
### Inactivation du chromosome X

Chez la femme (XX), un des deux chromosomes X est inactivé au hasard. C'est le phénomène de lyonisation. La femme est donc une mosaïque pour l'inactivation du chromosome X. En moyenne, environ 50 % des cellules ont l'X d'origine paternelle activé et 50 % des autres cellules l'X d'origine maternelle activé. Néanmoins, chez certaines femmes normales, l'inactivation n'est pas 50/50 mais déséquilibrée pour un des deux X (jusqu'à 90 % d'un des deux chromosomes X inactivé). Le mécanisme de ce biais d'inactivation n'est pas clairement identifié. Cette inactivation démarre tôt au cours du développement embryonnaire et s'accompagne d'une modification de la chromatine, caractéristique de l'hétérochromatine. L'inactivation de l'X est un des événements de méthylation de novo d'îlots CpG les plus précoces se produisant avant l'implantation de l'œuf. Un des deux chromosomes X d'origine paternelle ou maternelle est inactivé au hasard. Dans le chromosome inactivé, le processus démarre au niveau d'une séquence d'ADN appelée centre d'inactivation du chromosome X (*X-chromosome inactivation center*, *Xic*). À ce niveau, se trouve un gène codant pour un ARN non traduit, *X inactivation specific transcript* (*Xist*, Omim 314670), transcrit à partir de l'X inactivé. *Xist* joue un rôle fondamental dans l'initiation et la propagation de l'inactivation de l'X. En revanche, il ne joue aucun rôle dans le maintien de l'inactivation de l'X. *Xist* facilite la méthyla-

tion de novo des îlots CpG sur le chromosome X inactif, ce qui rend la majorité des gènes inactifs (et donc non exprimés). Cependant, certains gènes méthylés sur l'X inactif sont quand même exprimés (malgré l'inactivation de l'X correspondant). Chez les humains, la régulation et la maintenance de l'inactivation de l'X sont des mécanismes complexes dont on commence seulement à comprendre les différentes étapes.

## Interactions génétiques et épigénétiques

L'expression d'un gène est complexe et sous l'influence de nombreux facteurs génétiques et épigénétiques (figure 1.13). Bien évidemment, d'autres facteurs jouent un rôle très important. On peut citer les facteurs environnementaux généraux sur lesquels nous reviendrons dans la pathologie moléculaire.



**Figure 1.13** Interactions génétique/épigénétique. La méthylation de l'ADN se fait sur les îlots CpG dans le génome. L'expression des gènes dépend aussi de l'état de méthylation. La méthylation agit aussi indirectement sur le code génétique en favorisant la mutagenèse. Génétique et épigénétique sont donc étroitement liées.

## Points forts

### Les acides nucléiques

Les acides nucléiques sont des macromolécules constituées de chaînes de nucléotides. On distingue deux acides nucléiques, l'ADN (acide désoxyribonucléique) et l'ARN (acide ribonucléique). L'ADN forme une double hélice constituée d'unités de base, les nucléotides (union d'une base avec un sucre, le D-désoxyribose et un phosphate). Les bases composant un ADN normal sont au nombre de quatre : deux bases puriques (adénine [A], guanine [G]) et deux bases pyrimidiques (cytosine [C] et thymine [T]). La base « A » est associée avec la base « T » par deux liaisons hydrogènes (A=T). La base « G » est associée avec la base « C » par trois liaisons hydrogènes (G=C).

L'acide ribonucléique (ARN) se distingue de l'ADN par son sucre, le D-ribose, et diffère d'une base pyrimidique, l'uracile (U) au lieu de la thymine (T) de structure proche. L'uracile s'apparie à l'adénine et forme deux liaisons hydrogènes avec elle (A=U).

Il existe trois grands types d'ARN : l'ARN messager (ARNm), l'ARN de transfert (ARNt) et l'ARN ribosomal (ARNr).

### Le génome du noyau

La structure primaire de l'ADN (succession des bases le long de brin de l'hélice, appelée encore séquence de l'ADN) porte l'information génétique et constitue le génome.

L'information génétique des parties dites codantes est décryptée puis traduite en protéine dont le squelette est

constitué d'acides aminés. Il n'existe que 20 acides aminés codés chacun par un ou plusieurs triplets de base (appelé aussi codon). La relation triplet/acide aminé porte le nom de code génétique. En sus des triplets codant pour les acides aminés, il existe un codon initiateur ATG à partir duquel commence la traduction (signal d'initiation de la traduction) et trois codons particuliers, les « codons-stop » (TAA, TAG, TGA), signaux de fin de traduction.

Le génome humain est constitué de deux génomes distincts, le génome nucléaire situé dans le noyau de la cellule (souvent appelé par défaut le génome humain) et le génome mitochondrial localisé dans les mitochondries. Le nombre de gènes codant pour des protéines n'est pas connu avec précision. Il est estimé actuellement entre 20 000 et 25 000.

L'ADN est compacté dans le noyau en ensembles, les chromosomes. On distingue 23 paires de chromosomes dans la cellule humaine, chacun ayant une composition en séquence caractéristique et donc unique. Vingt-deux paires sont aussi appelées autosomes et la vingt-troisième paire est constituée par les chromosomes sexuels X et Y. Le chromosome comprend en ses extrémités les télomères indispensables pour préserver l'intégrité du matériel génétique au cours des divisions cellulaires. En cytologie, sur le chromosome, on reconnaît physiquement une zone de constriction se traduisant par un raccourcissement de l'épaisseur de la chromatine. Cette région est appelée centromère, site de fixation des microtubules de tubuline formés au cours de la division de la cellule.

(Suite)

## Points forts (suite)

Chez l'homme, entre deux individus non apparentés, il existe de petites variations de séquences, estimées entre 1 et 2 %, considérées comme non pathologiques, appelées polymorphismes. Ces polymorphismes génotypiques peuvent se trouver dans des régions codantes ou non codantes d'un gène ou dans d'autres régions du génome. On distingue schématiquement deux classes de polymorphismes, les polymorphismes bi-alléliques et les polymorphismes multi-alléliques.

De manière générale, un gène est un segment d'ADN contribuant à une fonction ou un phénotype. Il constitue un ensemble de séquences nucléiques contenant l'information nécessaire pour la production d'un ARN (transcription) et/ou d'un polypeptide ou d'une protéine (traduction). Le gène d'un eucaryote est morcelé en fragments codants, les exons séparés en général par des séquences non codantes, les introns. En amont du gène se trouve une séquence régulatrice, le promoteur. Dans certains cas, on peut trouver un pseudogène de structure proche du gène auquel il est rattaché. Il ressemble structurellement à un gène fonctionnel. Mais ce n'est pas un gène, car il n'est pas traduit en protéine.

De nombreuses séquences répétées sont réparties sur le génome (par exemple, les microsatellites et les minisatellites).

### Le génome mitochondrial

L'autre génome humain est l'ADN mitochondrial. Les mitochondries contiennent leur propre ADN, différent de l'ADN contenu dans le noyau de la cellule. Cet ADN est extrachromosomique, double brin et circulaire. Il est indépendant de l'ADN du noyau. Très polymorphe, il ne possède pas d'intron et est transmis par la mère.

### De l'ADN à la protéine

L'expression d'un gène démarre au moment de l'activation de la transcription et se termine à la synthèse d'une protéine fonctionnelle. Elle est régulée par des facteurs de transcription se fixant sur des séquences spécifiques au niveau de l'ADN, en amont du site d'initiation de la transcription. De nombreuses séquences agissent sur l'ADN pour en réguler l'expression par exemple, le *locus control région* (LCR), les *enhancers*, les *silencers*, les insulateurs, la conformation de la chromatine. La transcription d'un gène est donc sous le contrôle d'une combinatoire de protéines.

La transcription qui a lieu dans le noyau est constituée de trois étapes interdépendantes : la fixation d'un chapeau en 5' du pré-ARNm, l'élimination des introns (appelé aussi épissage ou *splicing*) et l'addition en 3'-terminal d'une queue poly(A).

Le code histone est basé sur les modifications des histones (acétylation et méthylation différentielles) aboutissant par leurs effets combinatoires et/ou séquentiels sur la ou les queues histones à des modifications fonctionnelles uniques.

Le code épigénétique est constitué du code histone et des modifications épigénétiques du génome, essentiellement la méthylation des cytosines en 5' (5mC).

La traduction a lieu dans le cytoplasme pour les eucaryotes. La séquence des codons sur l'ARNm dirige la synthèse du polypeptide qui aboutira dans un deuxième temps à la protéine fonctionnelle. La traduction de l'ARNm en protéines est réalisée à partir d'un complexe de ribonucléoprotéines appelé ribosomes. Les ribosomes cheminent le long de l'ARNm et en association avec l'ARNt combiné à un acide aminé (aminoacyl-ARNt) et un complexe protéique permettent la synthèse du polypeptide. Une fois la protéine formée, celle-ci va encore subir des transformations dites post-traductionnelles : glycosylation, acétylation, phosphorylation, etc. L'ensemble de ces transformations sont complexes et sont réalisées au cours des passages dans le réticulum endoplasmique et l'appareil de Golgi.

La régulation de l'expression d'un gène est fort complexe. L'ensemble des étapes depuis la transcription jusqu'à la synthèse protéique est en fait étroitement interdépendant. Il s'agit d'un processus continu.

L'édition (*editing*) de l'ARN est une modification post-transcriptionnelle se traduisant par la modification de la séquence nucléotidique de l'ARN à une ou plusieurs positions comme la modification de la cytosine en uracile ou la modification de l'adénine en inosine.

L'empreinte génomique est l'expression d'un gène dépendant de l'origine parentale de celui-ci. Conséquence de modifications épigénétiques, elle est caractérisée par l'expression mono-allélique d'un gène soit exclusivement à partir de l'allèle d'origine maternelle, soit exclusivement à partir de l'allèle d'origine paternelle.

L'inactivation du chromosome X : chez la femme (XX), un des deux chromosomes X est inactivé au hasard. C'est le phénomène de lyonisation. La femme est donc une mosaïque pour l'inactivation du chromosome X. En moyenne, environ 50 % des cellules ont l'X d'origine paternelle activé et 50 % des autres cellules l'X d'origine maternelle activé. Néanmoins, chez certaines femmes normales, l'inactivation n'est pas 50/50 mais déséquilibrée pour un des deux X (jusqu'à 90 % d'un des deux chromosomes X inactivé). On parle alors de biais d'inactivation.

L'expression d'un gène est complexe et sous l'influence de nombreux facteurs génétiques et épigénétiques.

## Références

- Agalioti T, Che G, Thanos D. Deciphering the transcriptional histone acetylation code for a human gene. *Cell* 2002 ; 111 : 381-92.
- Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. *Nature* 1981 ; 290 : 457-65.
- Blanc V, Davidson NO. C-to-U RNA editing : mechanisms leading to genetic diversity. *J Biol Chem* 2003 ; 278 : 1395-8.
- Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, et al, FANTOM Consortium ; RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group). The transcriptional landscape of the mammalian genome. *Science* 2005 ; 309 : 1559-63.
- Cazzola M, Skoda RC. Translational pathophysiology : a novel molecular mechanism of human disease. *Blood* 2000 ; 95 : 3280-8.
- Celera Genomics. The human genome sequence. *Science* 2001 ; 291 : 1304-51.
- Cervantes RB, Lundblad V. Mechanisms of chromosome-end protection. *Curr Opin Cell Biol* 2002 ; 14 : 351-356.
- Chan MF, Liang G, Jones PA. Relationship between transcription and DNA methylation. *Curr Top Microbiol Immunol* 2000 ; 249 : 75-86.
- Chiaromonte F, Miller W, Bouhassira EE. Gene length and proximity to neighbors affect genome-wide expression levels. *Genome Res* 2003 ; 13 : 2602-8.
- Clayton DA. Vertebrate mitochondrial DNA-a circle of surprises. *Exp Cell Res* 2000 ; 255 : 4-9.
- Conne B, Stutz A, Vassalli JD. The 3' untranslated region of messenger RNA : a molecular « hotspot » for pathology ? *Nat Med* 2000 ; 6 : 637-41.
- Dennis C, Campbell P, Watson JD, Crick FHC, Wilkins MHF, Stokes AR, et al. The double helix – 50 years. *Nature* 2003 ; 421 : 396-453.
- Dhaene K, Van Marck E, Parwaresch R. Telomeres, telomerase and cancer : an up-date. *Virchows Arch* 2000 ; 437 : 1-16.
- Dreyfus M, Regnier P. The poly(A) tail of mRNAs : bodyguard in eukaryotes, scavenger in bacteria. *Cell* 2002 ; 111 : 611-3.
- Esnault C, Maestre J, Heidmann T. Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* 2000 ; 24 : 363-7.
- Fairbrother WG, Chasin LA. Human genomic sequences that inhibit splicing. *Mol Cell Biol* 2000 ; 20 : 6816-25.
- Freitag M, Selker EU. Controlling DNA methylation : many roads to one modification. *Curr Opin Genet Dev* 2005 ; 15 : 191-9.
- Graveley BR. Alternative splicing : increasing diversity in the proteomic world. *Trends Genet* 2001 ; 17 : 100-7.
- Hirotsune S, Yoshida N, Chen A, Garrett L, Sugiyama F, Takahashi S, Yagami K, et al. An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* 2003 ; 423 : 91-6.
- International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001 ; 409 : 860-921.
- International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 2004 ; 431 : 931-45.
- John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. Human MicroRNA targets. *PLoS Biol* 2004 ; 2 : 1862-79.
- Karlin S, Chen C, Gentles AJ, Cleary M. Associations between human disease genes and overlapping gene groups and multiple amino acid runs. *Proc Natl Acad Sci USA* 2002 ; 99 : 17008-13.
- Kazazian HH Jr, Moran JV. The impact of L1 retrotransposons on the human genome. *Nat Genet* 1998 ; 19 : 19-24.
- Khrapko K, Coller HA, Andre PC, Li XC, Hanekamp JS, Thilly WG. Mitochondrial mutational spectra in human cells and tissues. *Proc Natl Acad Sci USA*. 1997 ; 94 : 13798-803.
- Kohlmaier A, Savarese F, Lachner M, Martens J, Jenuwein T, Wutz A. A chromosomal memory triggered by xist regulates histone methylation in X inactivation. *PLoS Biol* 2004 ; 2 : E171.
- Levinson B, Kenwick S, Lakich D, Hammonds G, Gitschier J. A transcribed gene in an intron of the human factor VIII gene. *Genomics* 1990 ; 7 : 1-11.
- Li Q, Peterson KR, Fang X, Stamatoyannopoulos G. Locus control regions. *Blood* 2002 ; 100 : 3077-86.
- Liu Y, Harrison PM, Kunin V, Gerstein M. Comprehensive analysis of pseudogenes in prokaryotes : widespread gene decay and failure of putative horizontally transferred genes. *Genome Biol* 2004 ; 5 : R64.
- Maas S, Rich A, Nishikura K. A-to-I RNA editing : recent news and residual mysteries. *J Biol Chem* 2003 ; 278 : 1391-4.
- Margueron R, Trojer P, Reinberg D. The key to development : interpreting the histone code ? *Curr Opin Genet Dev* 2005 ; 15 : 163-76.
- Mighell AJ, Smith NR, Robinson PA, Markham AF. Vertebrate pseudogenes. *FEBS Lett* 2000 ; 468 : 109-14.
- Nelson DR, Zeldin DC, Hoffman MGS, Maltais LJ, Wain HM, Nebert DW. Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. *Pharmacogenetics* 2004 ; 14 : 1-18.
- Numéro spécial de la revue *Cell*. Reviews on gene expression. *Cell* 2002 ; 108 : 431-582.
- Numéro spécial de la revue *Nature*. The double helix – 50 years. *Nature* 2003 ; 421 : 395-453.
- Ohlmann T, Derrington E, Lopez-Lastra M, Deffaud C, Bouchardon A, Darlix JL. L'initiation de la synthèse des protéines chez les eucaryotes. *M/S* 2000 ; 16 : 77-86.
- Ostertag EM, Kazazian HH Jr. Biology of mammalian L1 retrotransposons. *Annu Rev Genet* 2001 ; 35 : 501-38.
- Pagani F, Stuardi C, Zuccato E, Kornblihtt AR, Baralle FE. Promoter architecture modulates CFTR exon 9 skipping. *J Biol Chem* 2003 ; 278 : 1511-7.
- Pestova TV, Kolupaeva VG, Lomakin IB, Pilipenko EV, Shatsky IN, Agol VI, Hellen CU. Molecular mechanisms of translation initiation in eukaryotes. *Proc Natl Acad Sci USA* 2001 ; 98 : 7029-36.
- Plath K, Mlynarczyk-Evans S, Nusinow DA, Panning B. Xist RNA and the mechanism of X chromosome inactivation. *Annu Rev Genet* 2002 ; 36 : 233-78.
- Rakyan VK, Hildmann T, Novik KL, Lewin J, Tost J, Cox AV, et al. DNA methylation profiling of the human major histocompatibility complex : a pilot study for the human epigenome project. *PLoS Biol* 2004 ; 2 : 2170-82.
- Recillas-Targa F. DNA methylation, chromatin boundaries, and mechanisms of genomic imprinting. *Arch Med Res* 2002 ; 33 : 428-38.
- Reik W, Walter J. Genomic imprinting : parental influence on the genome. *Nature Rev* 2001 ; 2 : 21-32.
- Riggs AD, Xiong Z. Methylation and epigenetic fidelity. *Proc Natl Acad Sci USA* 2004 ; 101 : 4-5.
- Ross J. mRNA stability in mammalian cells. *Microbiol Rev* 1995 ; 59 : 423-50.
- Rowold DJ, Herrera RJ. Alu elements and the human genome. *Genetica* 2000 ; 108 : 57-72.
- Samuel CE. RNA editing minireview series. *J Biol Chem* 2003 ; 278 : 1389-90.
- Sandovici I, Naumova AK, Leppert M, Linares Y, Sapienza C. A longitudinal study of X-inactivation ratio in human females. *Hum Genet* 2004 ; 115 : 387-92.
- She X, Jiang Z, Clark RA, Liu G, Cheng Z, Tuzun E, et al. Shotgun sequence assembly and recent segmental duplications within the human genome. *Nature* 2004 ; 431 : 927-30.
- Shiota K, Yanagimachi R. Epigenetics by DNA methylation for development of normal and cloned animals. *Differentiation* 2002 ; 69 : 162-6.
- Spotswood HT, Turner BM. An increasingly complex code. *J Clin Invest* 2002 ; 110 : 577-82.
- Szmulewicz MN, Novick GE, Herrera RJ. Effects of Alu insertions on gene function. *Electrophoresis* 1998 ; 19 : 1260-4.
- Taanman JW. The mitochondrial genome : structure, transcription, translation and replication. *Biochim Biophys Acta* 1999 ; 1410 : 103-23.

- The genome international sequencing consortium. Initial sequencing and analysis of the human genome. *Nature* 2001 ; 409 : 860-921.
- Venter JC and The Celera Genomics Team. The sequence of the human genome. *Science* 2001 ; 291 : 1304-51.
- Wang Y, Liu CL, Storey JD, Tibshirani RJ, Herschlag D, Brown PO. Precision and functional specificity in mRNA decay. *Proc Natl Acad Sci USA* 2002 ; 99 : 5860-5.
- Waterston RH, Lander ES, Sulston JE. On the sequencing of the human genome. *Proc Natl Acad Sci USA* 2002 ; 99 : 3712-6.
- White JA, McAlpine PJ, Antonarakis S, Cann H, Eppig JT, Frazer K, et al. Guidelines for human gene nomenclature (1997). HUGO Nomenclature Committee. *Genomics* 1997 ; 45 : 468-71.
- Whitelaw E, Martin DI. Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nat Genet* 2001 ; 27 : 361-5.
- Woischnik M, Moares CT. Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. *Genome Res* 2002 ; 12 : 885-93.

## Annexe

**Tableau A. Contenu en acides nucléiques d'une cellule humaine**

ADN	6 µg
ARN total	10–30 µg
Taille du génome humain	3 10 <sup>9</sup> pb
Séquences codantes/ADN génomique	3 %
Nombre de gènes	0,5–1 10 <sup>5</sup>
Nombre de gènes actifs	1,5 10 <sup>4</sup>
Nombre de molécules d'ARNm	2 10 <sup>5</sup> –10 <sup>6</sup>
Taille moyenne d'un ARNm	1930 pb
Distribution des ARN dans la cellule	ARNr (5S, 18S, 28S) : 80–85 % ARNt : 15–20 % ARNm : 1–5 %

**Tableau B. Abréviations des acides aminés et symboles correspondants**

Acide aminé	Abréviation (en trois lettres)	Symbole
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Acide aspartique	Asp	D
Cystéine	Cys	C
Glutamine	Gln	Q
Acide glutamique	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Méthionine	Met	M
Phénylalanine	Phe	F
Proline	Pro	P
Sérine	Ser	S
Thréonine	Thr	T
Tryptophane	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

**Tableau C. Exemples de différences entre une cellule eucaryote et procaryote**

	Eucaryote	Procaryote
Appareil nucléaire	Présence d'un noyau Membrane nucléaire Histones Plusieurs chromosomes dont le nombre et la forme sont caractéristiques de l'espèce diploïdie	Pas de véritable noyau Pas de membrane nucléaire Chromosome unique, circulaire, pelotonné haploïdie
Structure membranaire	Cytoplasme structuré de façon complexe par le réticulum endoplasmique (RER) Organites : – mitochondries – Golgi – ribosomes 80S libres et sur le RER	Ribosomes 70S libres Les mêmes fonctions sont assurées par la membrane plasmique
Paroi	Pas chez tous les protistes Pas de glycopeptides	Presque toujours présente Polymère caractéristique : le peptidoglycane Constitutants spécifiques
Reproduction	Asexuée : mitose Sexuée : méiose	Asexuée (amitotique)
Exemples de modifications génétiques	<i>Crossing-over</i>	Conjugaison ou transduction

Pendant l'interphase, la cellule eucaryote contient un noyau, organe limité par une enveloppe, renfermant le matériel génétique sous forme d'acide désoxyribonucléique (ADN), molécule majeure des chromosomes. Nous distinguons :

- les protozoaires, formés par une seule cellule libre, souvent capables de se mouvoir (amibes, paramécies, Euglena) ;
- les métazoaires (animaux, plantes, levures) dont les cellules sont groupées en tissus (épithéliaux, musculaires, conjonctifs, de soutien [cartilagineux, osseux], nerveux).

Pendant l'interphase, la cellule procaryote ne possède pas de noyau. Un seul chromosome formé aussi d'ADN représente le matériel génétique. Aucune enveloppe ne le sépare du cytoplasme. En microscopie optique, après coloration, le chromosome apparaît comme un fin granule dénommé nucléoïde.

Les virus :

- ils échappent à cette classification : lorsqu'ils sont isolés, ils ne manifestent aucune activité vitale ;
- en revanche, dans les cellules qu'ils infestent (eucaryotes ou procaryotes), leur matériel génétique (ADN ou ARN) se réplique et commande les synthèses de protéines spécifiquement virales.

**Tableau D. ADN mitochondrial et nucléaire**

ADN nucléaire (ADNg)	ADN mitochondrial (ADNmt)
Un ADN dans le noyau	Environ 1000 ADNmt/cellule
Transmission par les deux parents	Transmission maternelle
≈ 3 10 <sup>9</sup> kb	16,5 kb
Organisation promoteur, région 5' non codante, exons/introns, région 3' non codante	Pas d'intron
≈ 25 000 gènes	37 gènes codés par l'ADNmt : – si pathologie : hétéroplasmie mélange d'ADNmt normal et pathologique – effet seuil : dans une cellule, le rapport ADNmt/ADNg doit dépasser un seuil pour que le phénotype se déclare (notamment défaut de la chaîne respiratoire) – variabilité du nombre d'ADNmt muté intracellule et/ou intra-tissu et/ou intra-organe et/ou intrafamiliale Interaction ADNg et ADNmt

## Bases de pathologie moléculaire

Génétique mendélienne  
 Mutations mitochondriales  
 Anomalies d'empreintes génomiques  
 Polymorphismes  
 Gènes de susceptibilité – Modulation polygénique  
 des maladies monogéniques  
 Gènes modificateurs  
 Maladies polygéniques et multifactorielles  
 Pathologie moléculaire : les mutations  
 Généralités sur les effets des mutations  
 Relation génotype-phénotype  
 Mutation(s) et pathologie  
 Mutation(s) et pathologie  
 Nomenclature des variations de séquence

Afin de comprendre l'intérêt des techniques de biologie moléculaire en médecine, nous décrirons dans ce chapitre les bases de la pathologie moléculaire chez l'être humain. Quelques exemples serviront à illustrer nos propos. Les lecteurs sont invités à se reporter au site internet *On line mendelian inheritance in man* (Omim) pour approfondir les sujets brièvement décrits. Après avoir rappelé quelques principes de base en génétique moléculaire, nous évoquerons les principales anomalies moléculaires responsables de pathologie médicale. Le domaine de la cytogénétique ne sera pas évoqué car il sort du cadre de cet ouvrage. Par ailleurs, la pathologie moléculaire appliquée à la microbiologie étant similaire à celle observée en pathologie humaine, nous ne l'aborderons pas.

### Génétique mendélienne

Ce chapitre de la génétique est abordé dans de nombreux ouvrages. Nous rappellerons quelques principes de base. De manière schématique, en génétique moléculaire, on distingue :

- les maladies monogéniques ;
- les maladies polygéniques.

Cette distinction est cependant artificielle. En effet, dans le cadre des maladies monogéniques, bien qu'un certain nombre de pathologies soit clairement associé au déficit d'un gène, la pénétrance d'un grand nombre de ces maladies monogéniques est variable (la pénétrance étant la proportion de sujets ayant une mutation dans le gène causal et développant la maladie par rapport aux

sujets ayant la même mutation et ne déclarant pas la maladie). Outre les facteurs environnementaux (non génétiques), on découvre de plus en plus de gènes influençant le développement et/ou protégeant un sujet d'une maladie. On parle de gènes modificateurs et de gènes de susceptibilité. Les réseaux de gènes interagissant entre eux soit directement, soit indirectement, amènent à délimiter une frontière de plus en plus floue entre ce qui est considéré comme monogénique et polygénique (figure 2.1).

La génétique mendélienne décrit les modes de transmission des parents à leur descendance selon des règles initialement décrites par Gregor Mendel en 1865. Le recensement des maladies héréditaires monogéniques est regroupé dans le catalogue Omim accessible sur internet (*On line Mendelian Inheritance in Man*). Dans les exemples que nous décrivons, le numéro omim est indiqué. Il s'agit d'un numéro conventionnellement donné à chaque pathologie sur le site internet Omim. Les maladies héréditaires monogéniques sont nombreuses (> 4000) mais rares. À titre d'exemple, les plus fréquentes d'entre elles ont une incidence faible comme la mucoviscidose (incidence 1/3000), le déficit en alpha 1-antitrypsine (incidence 1/1 700), l'hypercholestérolémie familiale (1/500) ou les neurofibromatoses (incidence 1/3000). Même si ces maladies sont rares, elles posent évidemment un réel problème à la fois humain et de santé publique. Classiquement, on décrit plusieurs modes de transmission :

- récessif lié à l'X : par exemple, l'hémophilie A (Omim 306700) ;
- dominant lié à l'X : par exemple, certaines formes de maladie de Charcot-Marie-Tooth (Omim 302800) ou le syndrome de Coffin-Lowry (Omim 303600). Cette forme de transmission est très rare ;
- autosomique récessif : par exemple, la mucoviscidose (Omim 219700). Pour que la maladie se déclenche, il

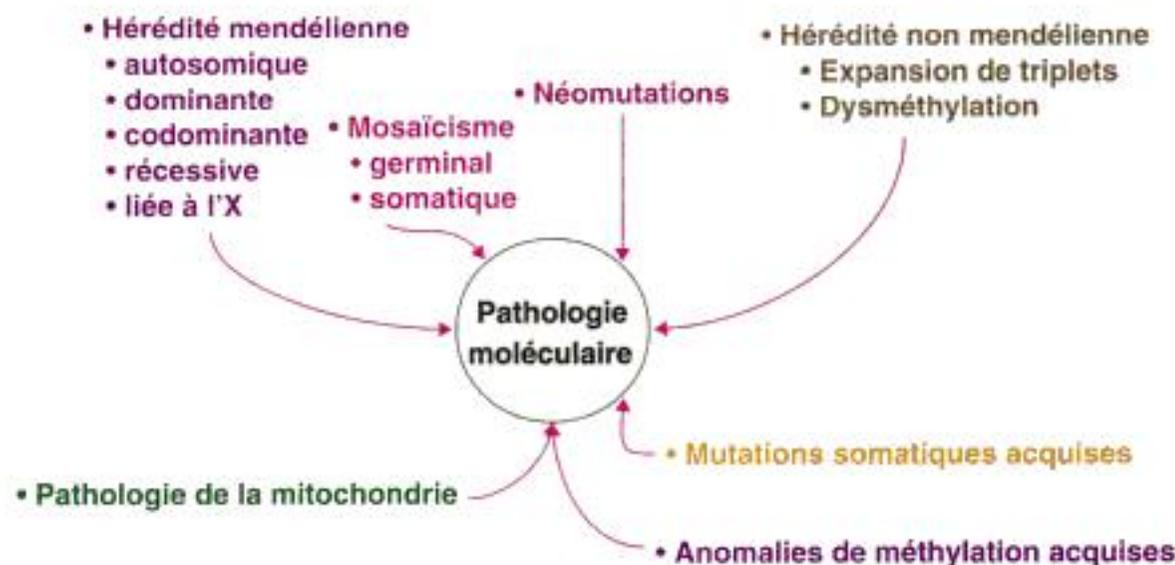
est nécessaire qu'une mutation soit présente sur chaque allèle de la paire d'autosomes portant le gène. La mutation sur un seul des deux allèles ne permet pas de développer la maladie (le sujet est porteur du trait héréditaire et est par conséquent normal) ;

- autosomique dominant : par exemple, la porphyrie aiguë intermittente (Omim 176000). Une mutation sur le gène d'un chromosome non sexuel (un allèle) suffit pour provoquer la maladie ;
- codominant : on parle de codominance quand deux allèles d'un même chromosome expriment chacun leur caractère de manière indépendante. Par exemple, les groupes sanguins ABO (Omim 110300).

### Transmission liée à l'X

Les gènes mutés sont localisés sur le chromosome X. Le risque clinique et la sévérité sont donc différents pour les deux sexes. La femme possède deux chromosomes X et, par conséquent, peut être homo-ou hétérozygote pour une mutation.

L'expression chez la femme est variable et est largement influencé par l'inactivation au hasard de l'X (phénomène de lyonisation). Au cours des maladies récessives liées à l'X, les hommes sont atteints et les femmes porteuses de la maladie. Chez la femme, cela est la conséquence de l'inactivation aléatoire de l'X (en général, 50 % de l'X d'origine maternelle inactivé et 50 % de l'X d'origine paternel inactivé). Cependant, dans certains cas, la femme peut présenter des manifestations cliniques liées à l'anomalie de l'X. En effet, dans de rares cas, on peut observer un biais d'inactivation de l'X chez la femme. Dans ce cas, un X d'origine soit paternel soit maternel sera majoritairement exprimé (au lieu d'une inactivation 50/50 la plus fréquemment observée). Ce biais peut être la conséquence d'un avantage sélectif de l'X normal sur



**Figure 2.1** Pathologie moléculaire héréditaire et acquise. La pathologie moléculaire, quel que soit son mécanisme, peut être d'origine héréditaire ou acquise. De nombreuses inconnues demeurent pour expliquer la physiopathologie des maladies acquises et héréditaires. La pathologie moléculaire ne constitue qu'une brique de ce complexe physiopathologique.

l'X muté (par exemple, l'X normal assure une meilleure croissance cellulaire ou améliore la durée de vie de la cellule) ou peut provoquer un avantage sélectif de l'X muté (par exemple, par défaut d'inactivation de l'X, conséquence d'une mutation sur le gène *Xist*, Omim 314670). Un biais d'inactivation en faveur de l'X muté peut donner l'apparence d'une transmission dominante liée à l'X.

On ne parle de maladie dominante ou récessive liée à l'X que chez la femme. Dans ce cas, les femmes atteintes peuvent avoir des fils ou des filles atteints mais les hommes atteints n'ont que des filles atteintes. Cette différence est difficile à préciser du fait de l'inactivation au hasard de l'X. Ces expressions sont donc rarement utilisées. C'est la raison pour laquelle on parle en général de maladie liée à l'X. Cependant, le plus souvent, la maladie est récessive. Les garçons présentent la maladie, mais les filles sont conductrices et peuvent avoir des fils atteints. Toutes les filles d'un homme atteint sont conductrices. En effet, les hommes transmettent à toutes leurs filles le chromosome anormal. Dans les maladies liées à l'X, il n'y a jamais de transmission d'homme à homme.

### Cas particulier : les gènes pseudo-autosomaux

Il s'agit de gènes présents sur les chromosomes X et Y. Leur mode de transmission n'est pas différent de celui des gènes autosomaux. À ce jour, aucune maladie n'a été formellement démontrée par anomalie dans la région pseudo-autosomale des chromosomes X et Y.

### Transmission autosomique dominante

La maladie se manifeste à l'état hétérozygote. Un seul allèle (l'allèle muté) est anormal. La mutation est présente sur un des 22 autosomes (chromosomes non sexuels).

### Caractéristiques générales

- Un des parents est atteint ou porteur sain.
- Un sujet atteint possède une chance sur deux (50 %) de transmettre l'allèle anormal à sa descendance.
- La transmission est verticale.
- Les enfants normaux d'un sujet atteint n'auront que des sujets normaux.
- Les hommes et les femmes sont atteints également.

Bien qu'encore imparfaitement compris, deux mécanismes expliquent au moins en partie le mode de transmission autosomique dominant d'un grand nombre de pathologies.

Ce sont l'haplo-insuffisance (HI) et l'effet dominant négatif (EDN) (figure 2.2). Dans les deux cas, la manifestation phénotypique est la conséquence de mutation(s) à l'état hétérozygote.

### Haplo-insuffisance (HI)

De manière générale, les copies d'un même gène autosomique sur chaque chromosome issu des deux parents s'expriment en proportion identique (codominance). L'haplo-insuffisance est une manifestation phénotypique consécutive à une mutation à l'état hétérozygote. Quand un seul allèle d'une cellule diploïde est présent ou fonctionnel (un seul des deux allèles est exprimé), la protéine correspondante est traduite en quantité insuffisante pour l'expression d'un phénotype normal (par exemple, une mutation sur un promoteur et modifiant l'expression d'un gène, la délétion complète d'un gène ou une mutation intragénique responsable de l'abolition de l'activité d'une protéine [allèle nul] ou de la diminution de son activité). En dehors du cas de la transmission autosomique dominante, on peut aussi parler d'haplo-insuffisance dans les cas où certaines mutations homozygotes ou double hétérozygotes (hétérozygotes composites) sont responsables de la forte diminution d'activité d'une protéine. Cette haplo-insuffisance peut aussi être sous la dépendance d'autres facteurs, que ce soit d'autres protéines (soit par action synergistique, soit par action de protéines sous influence de gènes modificateurs, voir infra) ou de facteurs environnementaux (toxique, médicament, tabac, aliment, infection, etc.) (figure 2.3).

### Effet dominant négatif (EDN)

L'effet EDN est le résultat de l'interaction d'une protéine anormale (produit d'un gène muté) avec la protéine normale (synthétisée par le gène présent sur l'autre allèle). Elle interfère avec l'activité de la protéine normale en diminuant son activité globale, voire en l'inhibant.

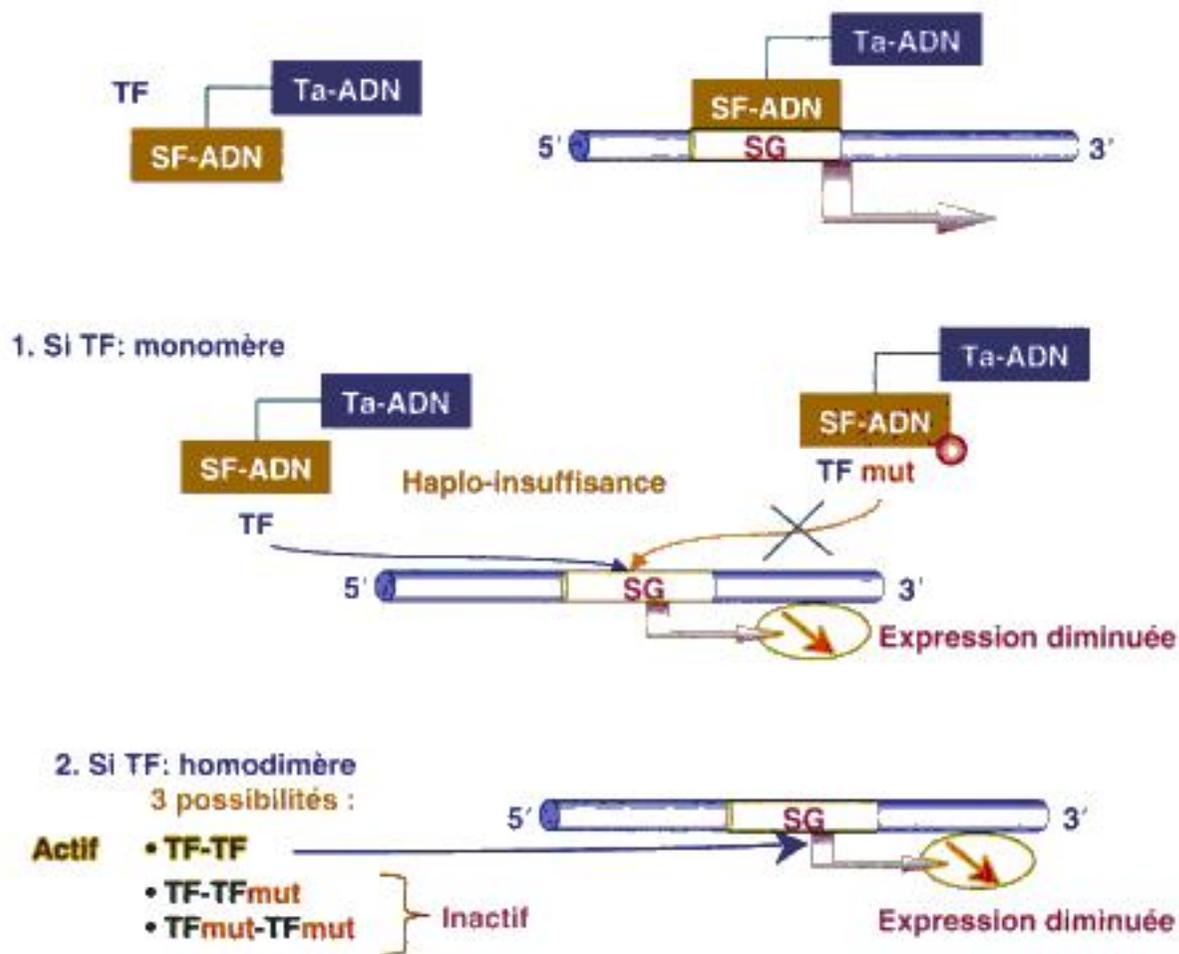
En pathologie moléculaire, selon leur localisation sur un même gène, certaines mutations peuvent être responsables d'un effet HI et d'autre d'un effet EDN.

### Un cas particulier : l'effet dominant positif

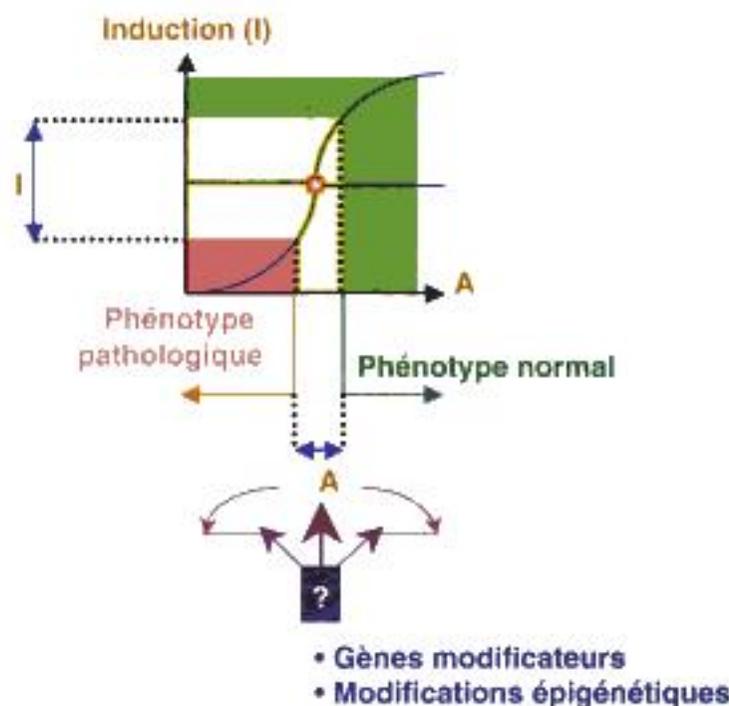
L'effet dominant positif est le résultat de l'interaction d'une protéine anormale (produit d'un gène muté) avec la protéine normale (synthétisée par le gène présent sur l'autre allèle). Elle est l'opposé de l'EDN. En interférant avec l'activité de la protéine normale, la protéine mutée va acquérir un gain de fonction.

### Un cas rare : la dominance autosomique spécifique de l'hétérozygote

Comme déjà indiqué, une maladie est de transmission autosomique dominante lorsque l'allèle muté détermine le phénotype à l'état hétérozygote. En règle générale, dans les maladies autosomiques, l'allèle muté agit en fait sur



**Figure 2.2** Notions d'haplo-insuffisance et d'effet dominant négatif. Prenons l'exemple d'un facteur de transcription (TF) se fixant sur une séquence spécifique du promoteur d'un gène (SG). Le TF contient deux domaines, le domaine de fixation à l'ADN (SF-ADN) et le domaine de transactivation (Ta-ADN). Le TF, en se fixant par son domaine de fixation à l'ADN, active le domaine de transactivation et permet la transcription du gène. 1. Haplo-insuffisance : supposons que le TF soit un monomère. En cas de mutation sur le SF-ADN, le TF muté (TFmut) ne pourra plus se fixer sur l'ADN cible et activer son expression. Seul, le TF normal pourra le faire. La mutation étant présente à l'état hétérozygote, la quantité de TF présente est diminuée de moitié (TF muté inefficace et TF normal actif) et l'expression du gène cible aussi. On aura un phénotype d'haplo-insuffisance. 2. Effet dominant négatif : supposons un TF agissant sous forme d'homodimère (deux TF associés pour obtenir l'action du TF sur le gène cible). Si un des deux allèles est muté, on aura un monomère TF normal et un monomère TF muté (TFmut). On obtiendra donc trois formes homodimères, TF-TF (normal/normal), TF-TFmut (normal/muté) et TFmut-TFmut (muté/muté). En cas d'effet dominant négatif, TFmut empêche l'action normale de TF. Seul l'homodimère TF-TF sera actif. L'action globale du TF (sous forme d'homodimère) sera diminuée.



**Figure 2.3** Haplo-insuffisance et influence multifactorielle. À noter que dans cet exemple, plusieurs protéines (groupe de FT) sont impliquées. Il pourrait n'y en avoir qu'une seule. Représentation schématique d'une réponse transcriptionnelle synergistique (par exemple, action de facteurs de transcription [FT] dont l'effet global, à savoir l'induction [I] de l'expression d'un gène est supérieure à un simple effet additif [la somme des actions] de ces FT) fonction de la concentration d'un activateur global A, somme de plusieurs FT (par exemple, par fixation sur le promoteur et/ou un enhancer). On observera une courbe sigmoïde résultant d'un effet coopératif de l'ensemble des FT. Autour du point d'inflexion (point rouge), une petite variation de l'activateur global provoque de grandes variations d'induction. Dans ce modèle schématique, au-delà d'un certain seuil, le phénotype est normal (sujet sain). En dessous d'un autre seuil, le phénotype est anormal (sujet malade). Entre les deux (deltaA), le phénotype est incertain. Il dépend d'influences extérieures (génétiques et environnementales) qui déclencheront l'évolution soit vers un phénotype normal, soit vers un phénotype pathologique. Ce basculement peut varier dans le temps, être réversible ou non par exemple.

un mode semidominant (ou codominant). En effet, lorsqu'un sujet est homozygote pour une maladie de transmission autosomique dominante, les effets des deux allèles mutés s'additionnent et donnent une maladie beaucoup plus sévère qu'à l'état hétérozygote. Ce n'est cependant pas toujours le cas. Certaines maladies sont strictement autosomiques dominantes (ou dominant complet) comme, par exemple, la chorée de Huntington (Omim 143100) et la néoplasie endocrinienne de type 1 (MEN1, *multiple endocrine neoplasia type 1*, Omim 131100). Dans ces deux maladies, aucune différence phénotypique n'est observée entre les sujets homo- et hétérozygotes.

La dominance autosomique spécifique de l'hétérozygote est un phénomène différent. Celle-ci a été observée par exemple dans une forme héréditaire de glaucome à angle ouvert résultant d'une mutation (K423E) dans le gène codant pour la myociline (Omim 601652). Dans cette pathologie héréditaire, les sujets hétérozygotes présentent la maladie alors que les sujets homozygotes pour la mutation sont asymptomatiques. La pénétrance de la maladie est donc largement supérieure chez les sujets hétérozygotes par rapport aux homozygotes. L'explication réside dans l'effet dominant négatif de la protéine présentant la mutation sur la protéine normale (le complexe devient non fonctionnel), effet neutralisé quand la protéine (agissant sous forme de multimère) est présente uniquement sous forme mutée. La protéine retrouve alors sa fonctionnalité.

### Transmission autosomique récessive

La maladie s'exprime cliniquement quand le sujet est homozygote (la même mutation sur les deux allèles) ou hétérozygote composite (une mutation différente sur chacun des deux allèles). Les deux allèles d'un même locus sont donc mutés. Le gène étant situé sur un des 22 autosomes, les deux sexes peuvent donc être affectés (nous excluons de ce cadre les disomies uniparentales).

#### Caractéristiques

- Les parents sont cliniquement normaux.
- Seule la descendance est atteinte. Il n'y a pas de transmission verticale. La probabilité d'avoir un enfant atteint est de 1/4 (25 %) si les parents indemnes sont porteurs. Les enfants indemnes sont porteurs avec une probabilité de 2/3.
- Les hommes et les femmes sont atteints également.

Ces maladies sont rares et la probabilité de consanguinité est importante.

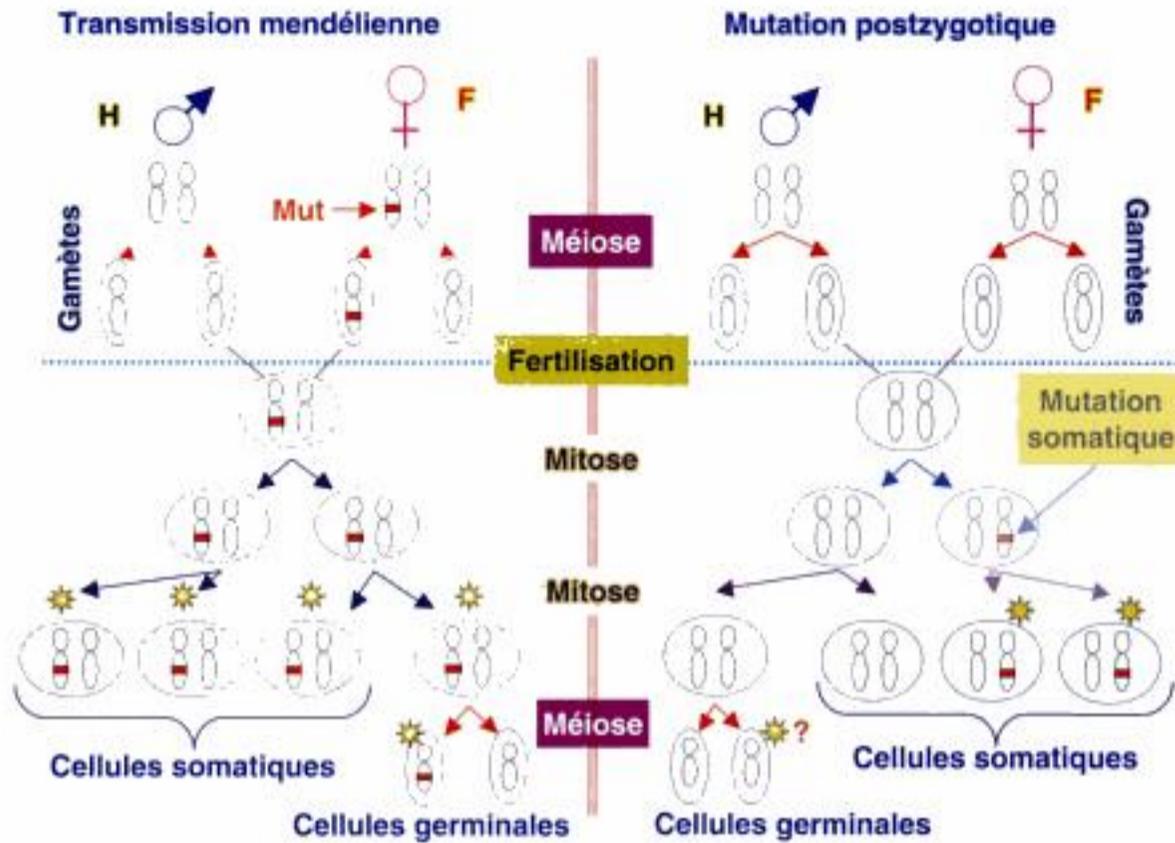
De manière générale, la survenue des maladies récessives apparaît tôt dans la vie, par opposition aux maladies dominantes.

### Néomutation (mutation de novo)

De manière générale, il s'agit de mutations apparaissant nouvellement chez un individu normal (nous excluons de ce cadre les mutations somatiques acquises, dans les cancers par exemple). La fréquence de mutation est approximativement de  $5 \cdot 10^{-6}$  mutation/gène/génération (ou encore  $1-2 \cdot 10^{-8}$  mutation/nucléotide/génération). La fréquence des néomutations serait alors de 1/100 000 nouveau-nés à chaque locus. Un grand nombre de ces mutations doit être probablement silencieuse ou récessive. La plupart, de fait, semblent silencieuses. Certaines cependant peuvent s'exprimer de manière dominante. Le parent chez qui la mutation est apparue dans la cellule germinale (et donc transmissible) est cliniquement normal. Les autres enfants seront normaux, la mutation germinale s'exprimant en effet dans peu de cellules germinales. Le nombre de gamètes atteints étant limité, la probabilité que les autres enfants soient atteints est faible. On peut cependant la retrouver chez d'autres enfants, dans ce cas, un mosaïcisme (au minimum gonadal) existe probablement. Bien évidemment, avant d'envisager ces possibilités, il est nécessaire de penser à la non-paternité (estimée à environ 5 % des cas) ou à la pénétrance faible d'une mutation. En général, cependant, les néomutations sont facilement suspectées dans les maladies liées à l'X et les maladies de transmission autosomique dominantes. Les néomutations récessives sont beaucoup plus rares (par exemple, un parent porteur hétérozygote d'une mutation et l'autre parent ayant une néomutation).

### Mosaïcisme

De manière générale, le mosaïcisme est une condition dans laquelle coexiste un mélange de population de cellules génétiquement distinctes (figure 2.4). Le mosaïcisme peut être la conséquence de mutations dans les cellules précurseurs de la lignée germinale. On a alors un mosaïcisme germinale. Le parent transmetteur normal (pas de mutation somatique) présentera ainsi un risque de transmettre une maladie liée à la mutation apparue dans la lignée germinale (par exemple, le père pourra avoir une mutation présente sur 10 % des spermatozoïdes pour un gène donné). Le mosaïcisme peut être la conséquence de mutations apparues après la fertilisation : on parle de mutations post-zygotiques. On observe alors chez un même sujet une prolifération clonale de deux groupes de cellules génétiquement distinctes comme, par exemple, dans certaines formes de neurofibromatoses de type 1 (Omim 162200). Il s'agit d'un mosaïcisme somatique. Un mosaïcisme somatique peut donc s'observer par suite d'une mutation apparue dans une cellule au cours de la gamétogenèse. Il peut s'observer plus tardivement par exemple en pathologie, au cours de la cancérogenèse. Le mosaïcisme est observé aussi dans les



**Figure 2.4** Mosaïcisme génétique et génétique mendélienne. Transmission mendélienne : un gène muté sur un chromosome d'origine maternel (rectangle rouge) est transmis de génération en génération (mitose) aux cellules somatiques (elles sont toutes atteintes : cellules possédant l'étoile orange) et germinales (transmission au cours de la méiose, étoile orange). Transmission postzygotique : une mutation somatique apparaît dans une cellule somatique après la fertilisation. Il y a un mosaïcisme. La mosaïque est constituée par des cellules sans mutation et d'autres avec mutation (chromosome possédant la barre rouge). Selon la nature de la mutation et la période d'apparition de celle-ci, la mutation reste somatique (cellules possédant l'étoile orange) ou peut être transmise aux cellules germinales.

mitochondries (hétéroplasmie), en cas d'altérations épigénétiques ou d'anomalies chromosomiques. Par ailleurs, un mosaïcisme germlinal peut coexister avec un mosaïcisme somatique.

L'inactivation au hasard de l'X (d'origine paternelle ou maternelle) dans les cellules XX est une forme de mosaïcisme du moins pour certains gènes exprimés à partir du chromosome X.

Enfin, certains auteurs ont calculé que nous étions tous probablement des « mosaïques ». En effet, le nombre de cellules d'un organisme humain adulte est estimé à environ  $10^{14}$  et la fréquence d'une mutation par division cellulaire est estimée à  $10^{-6}$  par gène.

Le diagnostic d'un mosaïcisme n'est pas toujours simple. Cette anomalie peut simuler d'autres formes de transmission héréditaire (par exemple, mosaïcisme germlinal et transmission autosomique dominante simulant une transmission récessive).

**Conséquences possibles du mosaïcisme**

Les conséquences d'une mutation somatique dépendent à la fois de sa nature et de sa date d'apparition. Il est fort probable que la majorité des mutations sont

phénotypiquement silencieuses, probablement par redondances soit de gènes similaires, soit de gènes aux fonctions moins importantes. Cela n'est pas toujours le cas. Voici quelques principes volontairement simplifiés et simplistes :

- si une mutation est provoquée dans un gène au rôle majeur (par exemple, une mutation provoquant une perte de fonction de la protéine correspondante), si le gène en question joue un rôle primordial dans le contrôle de la croissance cellulaire et si cette mutation est dominante, on peut s'attendre à l'apparition d'un cancer (où à un mécanisme favorisant l'apparition d'un cancer) ;
- si ce même gène n'est pas dominant, il représente un premier « choc » génétique dans un certain nombre de cellules. Ces cellules sont alors prédisposées pour un deuxième « choc » (mutation somatique sur l'autre allèle) inactivant complètement le gène ;
- une mutation post-zygotique peut dans certains cas être transmise à la génération suivante. Il s'agit dans ce cas d'une mutation présente dans les cellules donnant naissance aux gamètes. En fait, il semblerait que cela soit rare, la séparation entre les cellules germinales et somatiques se faisant tôt chez l'embryon. Il faudrait

alors une mutation précoce dans les premières étapes post-zygotiques du développement ;

- enfin, dans certains cas, le mosaïcisme n'est présent que dans les cellules de la lignée germinale. Ce cas rare peut être envisagé par exemple lorsque les parents sont tous les deux normaux et présentent plusieurs enfants malades.

## Notion de pénétrance et d'expressivité

### Pénétrance

La pénétrance est la proportion de sujets présentant un phénotype pathologique (sujets malades) par rapport au génotype responsable de la pathologie. Par exemple, une pénétrance de 10 % signifie que 10 % des sujets possédant l'anomalie moléculaire (la mutation par exemple) présente au moins un signe phénotypique de la maladie. Quand on parle de phénotype pathologique, il s'agit de n'importe lequel des symptômes de la maladie (depuis une anomalie biologique jusqu'à la présence d'un ou de plusieurs signes cliniques). Cette dernière notion est importante, la notion de phénotype pouvant varier selon les études (et, par conséquence, la fréquence de la pénétrance aussi), certains auteurs n'accordant pas toujours la même signification au terme de phénotype.

La pénétrance est complète quand tout porteur du gène délétère est atteint. La pénétrance d'une maladie est très variable selon les gènes en cause. Elle peut aller de 1 à 100 %. Cette pénétrance peut se traduire aussi en fonction de l'âge, certaines maladies se déclarant dès la naissance et d'autres plus tardivement au cours de la vie. Dans un grand nombre de pathologies, la pénétrance est incomplète.

### Expressivité ou variabilité de l'expression clinique

Il s'agit de l'ensemble des effets phénotypiques observés pour une mutation donnée. Les manifestations cliniques peuvent être différentes d'un sujet à l'autre pouvant aller d'une seule manifestation clinique à l'ensemble du phénotype clinique décrit pour cette maladie, par exemple, type et sévérité des symptômes, âge de survenue de la maladie.

L'expressivité et la pénétrance variables d'une maladie peuvent être dues à plusieurs facteurs :

- des gènes modificateurs situés sur d'autres loci ou d'autres variations génotypiques sur le même locus ;
- des facteurs environnementaux et exogènes.

### Transmission dominante ou récessive

Dans la génétique mendélienne classique, les maladies héréditaires sont classées selon le mode de transmission de

manière cartésienne. En fait, ce concept n'est pas toujours aussi tranché. Par exemple, le déficit héréditaire en facteur V de la coagulation (facteur V Leiden, Omim 227400) est considéré comme une maladie de transmission autosomique récessive. La mutation la plus fréquente (R506Q) est associée au risque élevé de thrombose veineuse (risque relatif compris entre 80 et 100 fois par rapport au sujet normal). Le sujet présentant la mutation à l'état hétérozygote présente aussi un risque de thrombose veineuse (risque relatif 5 à 10 fois supérieur à la population normale). La notion de récessivité doit donc être interprétée avec une certaine relativité dans certaines pathologies.

### Phénotype et diversité des modes de transmission

Certaines entités cliniques présentent plusieurs modes de transmission héréditaire selon le gène muté. C'est le cas par exemple dans la cardiomyopathie dilatée responsable d'une insuffisance cardiaque congestive chez le sujet jeune. Sans rentrer dans les détails de cette pathologie dans laquelle de nombreux gènes sont incriminés, selon le gène en cause, la transmission peut être autosomique dominante (par exemple, mutations sur le gène codant pour la troponine T, Omim 191045), mitochondrial (par exemple, mutation sur le gène codant pour l'ARNt-isoleucine, Omim 590045) ou liée à l'X (mutation sur le gène codant pour la tafazzine, Omim 300394).

### Génétique non mendélienne

La génétique mendélienne n'explique pas toutes les maladies génétiques. Ces vingt dernières années, de nouveaux concepts sont apparus et ont amené à des modes de transmission héréditaires différents classiquement appelés « génétique non mendélienne ». Cette classification relative ne doit pas faire oublier que l'ensemble des mécanismes physiopathologiques est plus complexe que ne laisse penser ce mode de classification.

### Instabilité du génome

#### Séquences répétées

On distingue schématiquement dans le génome :

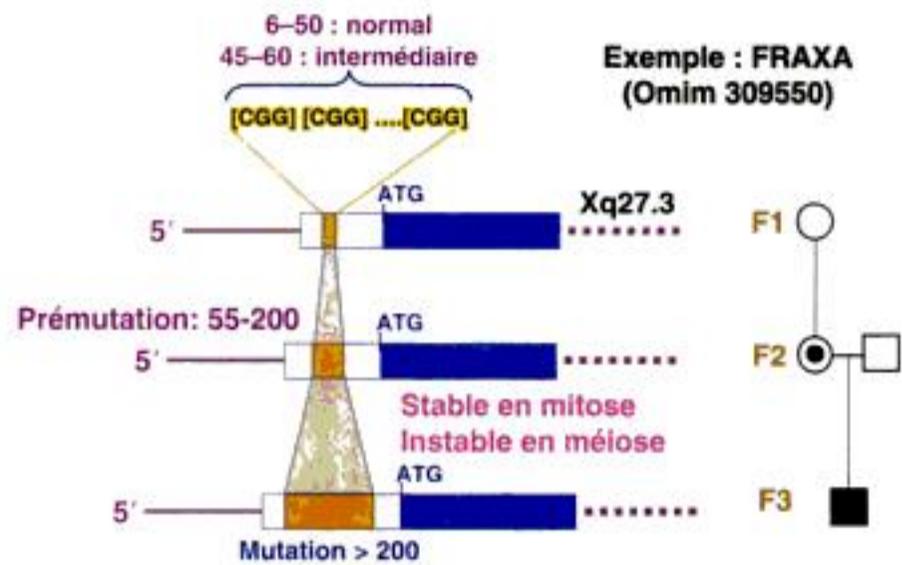
- les minisatellites, séquences répétées de 10 pb à plus de 100 pb. Ces minisatellites répétées en tandem varient en longueur, on parle de *variable number of tandem repeat* (VNTR). Ces VNTR peuvent être intragéniques (par exemple, dans le gène codant pour le récepteur à la dopamine D4, DRD4, Omim 126452) ou, le plus souvent, extragéniques. Ces VNTR peuvent jouer un rôle régulateur de la transcription : par exemple, le VNTR situé 600 pb en amont du site d'initiation de la transcription du gène codant pour l'insuline (INS,

Omim 176730). Dans ce cas, le VNTR contiendrait des sites de fixation pour des facteurs de transcription. Ces séquences peuvent parfois devenir instables soit au niveau de la lignée germinale, soit de la lignée somatique. D'autres séquences telles les séquences Alu peuvent induire des pathologies par insertion ;

- les microsatellites de taille inférieure à 10 pb (en moyenne, 1 à 4 pb) répétées en tandem. Ces séquences répétées en tandem (par exemple, CACACA... dinucléotide répété N fois) sont présentes dans tout le génome. Ces mono-, di-, tri- ou tétranucléotides (on parle encore de microsatellites) sont très polymorphiques chez l'être humain. Dans un grand nombre de cancers notamment, la répétition de ces séquences devient instable au cours des mitoses dans les cellules cancéreuses. Cette instabilité des microsatellites (MIN, *microsatellite instability*, voir infra) se traduit par une variation du nombre de séquences répétées (augmentation ou diminution) différentes de celles des cellules normales et peut être facilement détectée par les méthodes de biologie moléculaire. Un des mécanismes considérés comme essentiel dans les tumeurs et expliquant cette instabilité est une erreur de réparation de l'ADN au cours de la réplication (d'autres mécanismes ont aussi été décrits tel que la recombinaison homologue par exemple).

### Phénomène d'anticipation

Dans certaines pathologies humaines résultant de l'extension de séquences répétées en nombre anormal (par exemple, le syndrome de l'X fragile [Omim 309550] ou la chorée de Huntington [Omim 143100]), on observe de génération en génération une augmentation du nombre de séquences trinuécléotidiques répétées en tandem traduisant une instabilité de ces microsatellites (figure 2.5). Bien que le mécanisme moléculaire soit encore mal compris, cette instabilité serait le résultat de la structure spatiale secondaire instable de ces répétitions (formation d'épingles à cheveux [*hairpin*] notamment). Pour expliquer cette instabilité, plusieurs mécanismes (isolés ou associés) ont été décrits tels que les recombinaisons au cours de la méiose et/ou de la mitose (dans ce dernier cas, cela expliquerait les variations de longueur des séquences observées au cours de l'analyse moléculaire), les erreurs de réplication de l'ADN (par glissement de l'ADN polymérase), la conversion (par recombinaison homologue non réciproque) et les anomalies de réparation de l'ADN. Il existe une relation génotype/phénotype entre le nombre de séquences répétées et les symptômes (par exemple, sévérité de la maladie, âge de début). À partir d'un certain nombre de répétitions, la pathologie se déclare. Un état intermédiaire entre un nombre de répétition normale et pathologique existe, on parle alors de prémutation. Le nombre de répétitions au stade de prémutation est relatif



**Figure 2.5** Le phénomène d'anticipation. Dans le syndrome de l'X fragile (Omim 309550), la trinuécléotide CCGG est répétée entre 6 et 50 fois chez le sujet normal. En cas d'anticipation, d'une génération à l'autre, cette répétition de microsatellite augmente. Lorsque le nombre de répétitions du microsatellite est au-dessous du seuil de 200, la maladie n'est pas déclarée encore, c'est le stade de prémutation. Au stade de prémutation, il peut y avoir une instabilité de microsatellite aboutissant à une expansion du triplet (phénomène dynamique). Et au-delà d'un certain nombre de répétitions (> 200 en moyenne), la maladie apparaît et s'aggrave au fil des générations, apparaissant de plus en plus précocement : c'est le phénomène d'anticipation.

vement stable dans la lignée germinale de l'homme alors qu'elle est plus instable dans la lignée germinale de la femme. C'est le cas dans le syndrome de l'X fragile. Néanmoins, cela n'est pas toujours observé. Ces amplifications de séquences répétées se produisent au cours de la méiose avec parfois une amplification préférentielle d'origine maternelle ou paternelle. Ainsi, de génération en génération, la maladie s'aggrave et apparaît à un âge plus jeune : ce phénomène est appelé anticipation. Ces mutations sont ainsi dynamiques (tendance à l'allongement de la longueur des séquences répétées). Enfin, il a été décrit dans de rares cas non pas une expansion des séquences répétées mais une contraction (réduction des séquences répétées) dans la lignée germinale. Cet événement exceptionnel est encore appelé réversion (voir infra). Cette expansion est bien évidemment aussi présente dans les cellules somatiques où on observe souvent une instabilité des séquences répétées (variabilité du nombre de séquences répétées) plus particulièrement pour les allèles de grande taille. Cette expansion anormale aboutit le plus souvent à une méthylation anormale des séquences répétées (par exemple dans le syndrome de l'X fragile au niveau du promoteur) et l'absence de transcription du gène correspondant. Il peut donc y avoir une anomalie de méthylation associée à une expansion de triplets. Dans d'autres cas, elle aboutit à la formation d'un polyaminoacide dans la partie codante du gène (par exemple, polyglutamine dans la maladie de Huntington, Omim 143100) responsable de la formation d'agrégats et de mort cellulaire par apoptose.

Pour mémoire, d'autres pathologies développées plus loin, résultant d'interactions complexes, sont rappelées :

- les maladies polygéniques ;
- les maladies mitochondriales par atteinte du génome de la mitochondrie ;
- les maladies épigénétiques.

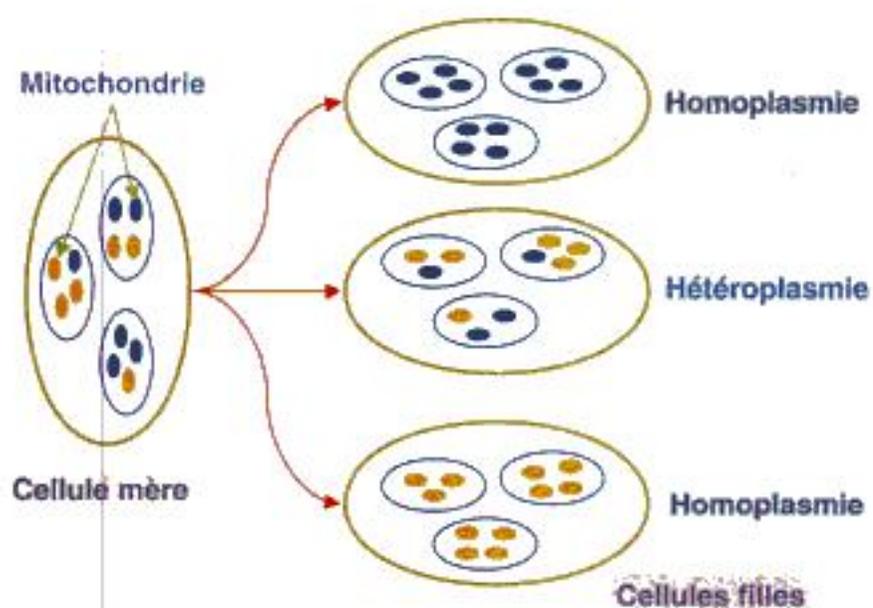
## Mutations mitochondriales

La transmission de l'ADN mitochondrial est maternelle en général (dans certains cas exceptionnels, la transmission d'ADN mitochondrial muté d'origine paternelle a pu être démontrée).

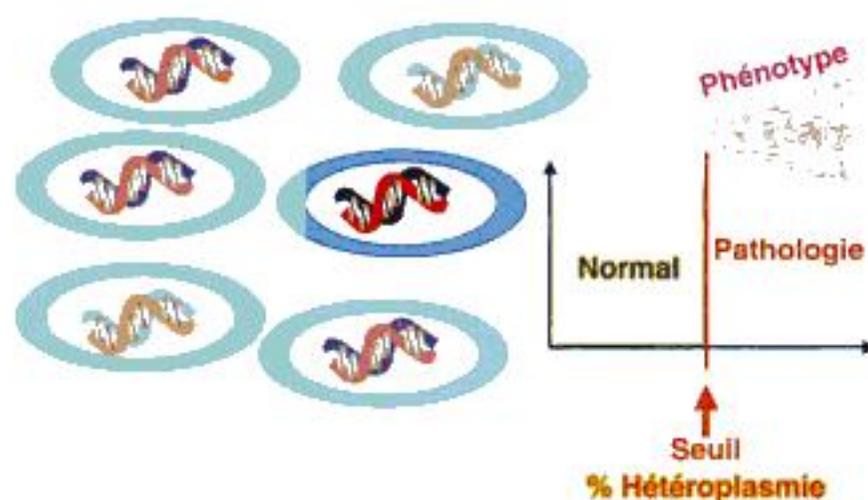
Les mutations dans le génome mitochondrial sont transmises par la mère et sont responsables de nombreuses pathologies, surtout neurologiques et neuromusculaires.

La ségrégation des mitochondries au cours de la répliation se fait au hasard entre les cellules filles (figure 2.6). Ce phénomène explique l'existence d'hétéroplasmie et permet d'aborder la notion de seuil (figure 2.7). Les manifestations cliniques dépendent notamment du nombre de mitochondries présentant la mutation. On peut ainsi définir :

- l'hétéroplasmie : les ADN des mitochondries ne possèdent pas tous la mutation causale. Ainsi, dans la cellule, le tissu et les mitochondries, il y aura différentes populations d'ADN mitochondrial normal et muté ;
- l'homoplasmie : l'ADN de chaque mitochondrie présente la mutation causale.



**Figure 2.6** Notions d'homoplasmie et d'hétéroplasmie. À partir d'une cellule mère, la transmission des mitochondries se fait au hasard dans les cellules filles aboutissant soit à l'hétéroplasmie, soit plus rarement à l'homoplasmie. Prenons l'exemple de deux populations de mitochondries (ADN mitochondrial différent). Soit la population mitochondriale transmise possède le même ADN de l'un ou l'autre variant ADN, c'est l'homoplasmie. Soit la répartition dans la cellule est un mélange aléatoire des deux populations, c'est l'hétéroplasmie.



**Figure 2.7** Notion d'effet seuil dans la pathologie mitochondriale. La répartition aléatoire entre le nombre de mitochondries normales et pathologiques est variable. Lorsque le nombre de mitochondries pathologiques atteint une certaine quantité (seuil), les phénomènes pathologiques se déclarent notamment dans les tissus dépendant en grande partie du métabolisme mitochondrial (par exemple, le cerveau).

Tout type de mutation décrite pour l'ADN nucléaire peut se voir dans la mitochondrie. On estime leur fréquence au niveau mitochondrial à 1/8000. Le pourcentage de mitochondries atteintes détermine la présence ou non de la maladie. Par ailleurs, le seuil est variable selon la pathologie en cause. Les conséquences des mutations ne seront pas les mêmes selon que le tissu en cause présente plutôt un métabolisme à dominante glycolytique (par exemple, la peau, les fibroblastes) ou aérobie (par exemple, le cerveau, le cœur, les muscles squelettiques, le foie, le pancréas, les yeux). Selon les exigences métaboliques en ATP notamment (effet seuil), l'impact sera supérieur dans les organes à exigence aérobie prédominante.

La distribution des mitochondries au cours de la mitose étant effectuée aléatoirement aux cellules filles, si le seuil pathologique est dépassé dans un tissu auparavant non atteint, le phénotype peut être modifié. Ce phénomène explique en partie la variabilité phénotypique observée en fonction de l'âge et du tissu.

Dans la mitochondrie, outre les maladies héréditaires (transmission maternelle), les mutations de l'ADN peuvent aussi être sporadiques.

## Mutations responsables de déficit d'importation de protéines dans la mitochondrie

Les protéines cytoplasmiques destinées aux mitochondries présentent en général une séquence signal d'importation (appelée aussi préséquence, peptide signal ou

séquence d'adressage). Cette séquence sera ensuite éliminée par des peptidases pour que la protéine puisse assurer sa fonction (protéine mature). Une mutation dans la séquence signal sera responsable de l'absence/diminution d'importation dans la mitochondrie.

Le peptide signal d'environ 20 à 60 acides aminés est le plus souvent chargé positivement, contient des résidus hydrophobes et hydroxylés, sans résidu acide et capables de former une structure secondaire amphiphile. Bien qu'il existe des protéines importées dans la mitochondrie et ne possédant pas ce peptide caractéristique, ce dernier est souvent indispensable à l'importation mitochondriale des protéines. Cette importation (dont le mécanisme sort du cadre de cet ouvrage) est sous la dépendance de complexes protéiques permettant le formatage tridimensionnel et la translocation des protéines vers les compartiments mitochondriaux (tels que les translocases de la membrane externe [*translocase of outer membrane*, TOM] et interne [*translocase of inner membrane*, TIM]). Des pathologies spécifiquement mitochondriales peuvent donc avoir pour origine une mutation dans un gène situé dans le génome nucléaire. Dans ce cas, le mode de transmission héréditaire peut être autosomique dominant, récessif ou lié à l'X. D'autres pathologies non spécifiquement mitochondriales sont aussi la conséquence de mutations dans le peptide signal.

### Anomalies d'empreintes génomiques

L'empreinte génomique représente l'expression différentielle d'un gène selon son origine maternelle ou paternelle. Certains gènes seront donc actifs sur un chromosome parental et inactif sur l'autre chromosome parental. Les anomalies de l'empreinte génomique peuvent être la conséquence d'anomalies chromosomiques et/ou épigénétiques.

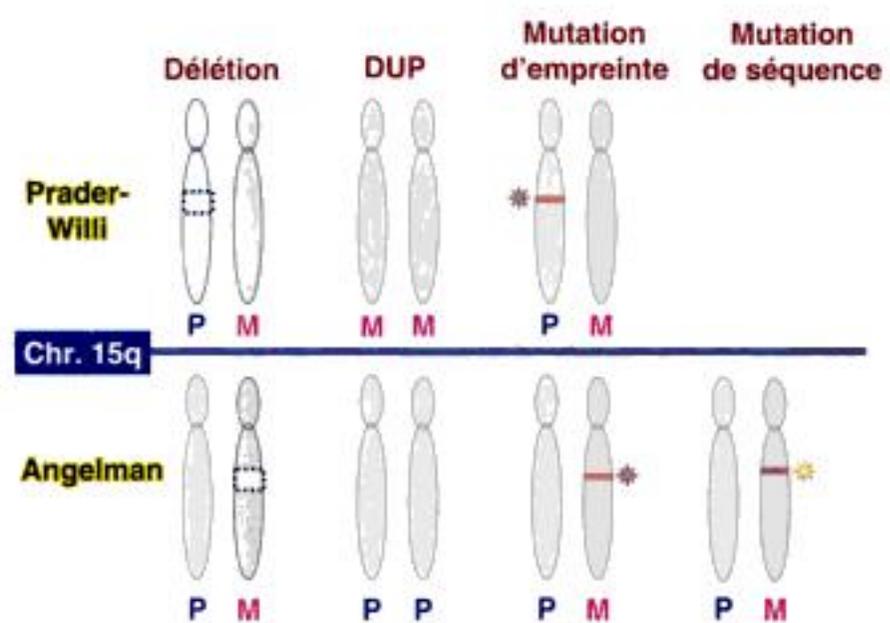
À ce jour, environ une trentaine de gènes soumis à empreinte parentale sont répertoriés.

### Pathologie de l'empreinte génomique

Les anomalies de l'empreinte génomique sont responsables de maladies dont le phénotype est différent selon que le locus malade est d'origine maternel ou paternel. À titre d'exemple, les rares syndromes de Prader-Willi (Omim 176270) et d'Angelman (Omim 105830) sont des pathologies résultant d'anomalies de l'empreinte génomique (figure 2.8). Sans rentrer dans les détails complexes de ces deux maladies, nous exposerons quelques éléments pour illustrer cette pathologie.

#### Délétion chromosomique

Dans chaque cas, l'absence du (des) gène(s) actif(s) est la cause de la maladie :



**Figure 2.8** Pathologie de l'empreinte génomique. Exemple des syndromes d'Angelman et de Prader-Willi. Ces deux syndromes résultent d'anomalies au même locus. Le syndrome de Prader-Willi (SPW) est la conséquence d'anomalie sur le chromosome d'origine paternelle et le syndrome d'Angelman (SA) sur le chromosome d'origine maternelle. Plusieurs anomalies sont possibles : une délétion chromosomique dans la même zone du chromosome 15 (15q11–q13) ; une disomie uniparentale (DUP) : les deux chromosomes normaux sont issus soit du père (SA), soit de la mère (SPW) ; une atteinte du centre de l'empreinte (mutation par exemple) peut être responsable d'une empreinte anormale sur les gènes cibles ; enfin, certains gènes peuvent aussi être mutés (par exemple, le gène UBE3A, Omim 601623 dans le syndrome d'Angelman).

- le syndrome de Prader-Willi (PWS) est la conséquence d'anomalies d'une région sur le bras court du chromosome 15 (dans 70 % des cas environ, une délétion, del [15q11–q13], le plus souvent de novo). La délétion est issue du chromosome d'origine paternel ;
- le syndrome d'Angelman (AS) résulte de la même délétion sur le chromosome d'origine maternel (dans 70 % des cas).

#### Disomie uniparentale

La disomie uniparentale (DUP) est une anomalie chromosomique dans laquelle une paire de chromosomes est transmise par le même parent (environ 20 à 25 % des cas pour PWS et 2 à 3 % des cas pour AS). L'autre parent ne transmet aucune copie. La DUP peut être partielle ou totale. Cette anomalie est principalement consécutive à une non-disjonction chromosomique. Dans ce cas, PWS est la conséquence de la présence de deux chromosomes 15 normaux provenant de la mère (aucun du père) et AS la conséquence de la présence de deux chromosomes 15 normaux provenant du père (aucun de la mère).

#### Délétion du centre d'empreinte

Dans la région délétée, une portion du génome appelée centre d'empreinte (*imprinting centre*, IC) a été identifiée. Elle contrôle les gènes environnants en établissant la

bonne empreinte au cours de la gamétogenèse. L'IC en fait régule la méthylation des gènes environnants, l'expression des gènes et la structure chromatinienne locale. En effet, les gènes soumis à empreinte sont essentiellement regroupés en « amas » (*cluster*) sur une portion de l'ADN. Le mécanisme d'action précis du IC n'est pas encore connu. On sait néanmoins qu'une partie de son action se fait par des modifications épigénétiques différentielles (telles que la méthylation de l'ADN au niveau des dinucléotides CpG, voir infra).

Dans certains cas (3 à 5 % environ), on ne retrouve ni délétion ni disomie uniparentale. On observe alors des microdélétions, des mutations ou, plus rarement, d'autres anomalies chromosomiques (par exemple, translocation) au niveau de l'IC. Le centre d'empreinte n'assure plus son rôle d'activateur/répresseur dans la région. Toutes ces anomalies vont être responsables d'anomalies épigénétiques modifiant régionalement l'expression des gènes cibles. Ainsi, dans l'AS, il peut s'agir de mutations sur un gène UBE3A (Omim 601623, environ 8 % des cas).

Enfin, environ 5 % des PWS et 15 % des AS n'ont pas de cause identifiée (aucune anomalie moléculaire ou cytogénétique).

### Effet de l'origine parentale sur la transmission des maladies héréditaires

Dans certaines maladies héréditaires, on observe une influence de l'origine parentale en l'absence de toute empreinte génomique. C'est le cas dans certaines maladies par expansion de répétitions de nucléotides (phénomène d'anticipation). Par exemple, dans la maladie de Huntington (Omim 143100), maladie autosomique dominante par expansion des triplets CAG sur le gène codant pour la huntingtine, l'âge plus précoce de survenue de cette forme de démence est lié à la transmission paternelle. La transmission paternelle est associée à une forte tendance à l'expansion de triplets (alors que l'expansion est très faible si la transmission est maternelle). Une des explications possibles de ce phénomène est l'instabilité plus importante du nombre de triplets au cours de la gamétogenèse chez l'homme. Des explications différentes ont été avancées dans d'autres maladies comme la maladie de Steinert (ou dystrophie myotonique, Omim 160900) de transmission autosomique dominante associée à une transmission maternelle par expansion anormale de trinucleotides CTG. On observe chez les filles transmettrices une augmentation du nombre de séquences répétées. Un avantage sélectif de l'expansion de triplets pourrait influencer sur la survie des spermatozoïdes. Dans ce dernier cas, la forme congénitale sévère de la maladie est le plus souvent associée à une transmission maternelle. En fait, ce phénomène non lié à l'empreinte génomique est encore mal connu.

### Empreinte génomique et hérédité multifactorielle

Certains auteurs ont suggéré l'influence de l'empreinte sur certains traits phénotypiques tels que la sévérité de l'atteinte, l'âge précoce du début de la maladie et la pénétrance. Cette hypothèse a notamment été évoquée par exemple dans certaines formes de diabète.

### Polymorphismes

Les polymorphismes de l'ADN sont des séquences intra- ou extragéniques considérées comme non pathologiques pour le sujet (un exemple de polymorphisme simple est la couleur des yeux, polymorphisme phénotypique). Dans le génome humain, on estime à environ 20 000 à 25 000 le nombre de gènes, la plus grande partie étant constituée d'une quantité importante d'ADN non codant. Malgré la taille considérable du génome humain, on estime que la variation entre deux sujets normaux dans la population mondiale est au maximum de 0,1 %.

### Diversité des polymorphismes

Les polymorphismes peuvent être bi-alléliques (exemple A et G, A/G) ou multi-alléliques (séquences répétées de type mini- ou microsatellites).

#### Polymorphismes multi-alléliques

On distingue classiquement deux classes :

- les microsatellites : on les appelle souvent *short tandem repeat* (STR). Ces séquences sont nombreuses (estimées à 50 000–100 000 copies par génome) et sont espacées assez régulièrement (environ tous les 10 kb). La plupart de ces microsatellites sont des séquences dinucléotidiques  $[CA/GT]_N$ . D'autres microsatellites sont des séquences répétées de 1 à 10 bp, et dont la taille totale est d'environ 100 à 300 bp. Ces marqueurs sont très informatifs et sont utilisés par exemple pour la cartographie, les études de liaison dans le génome humain ou encore en médecine légale ;
- les minisatellites comme les *variable number of tandem repeats* (VNTR) sont également utilisés. Ces derniers sont constitués d'une série de 15 à 70 bp répétées en tandem. D'autres minisatellites sont disséminés dans le génome comme les séquences Alu par exemple.

#### Polymorphismes bi-alléliques

Les polymorphismes bi-alléliques (*single nucleotide polymorphism*, SNP) sont les variants les plus fréquents dans le génome humain. De manière générale, un SNP est défini par les caractères suivants :

- la fréquence dans la population est supérieure à 1 % ;

- la transmission héréditaire est stable ;
- elle permet d'établir des haplotypes ;
- la densité moyenne est estimée à un polymorphisme toutes les 1,2 à 1,9 kb. Il existe cependant de grandes variations selon les chromosomes et selon les gènes. Cette densité s'explique schématiquement par la pression de sélection sur les gènes ainsi que les taux de mutation et de recombinaison, ces derniers variant à travers le génome. On estime en général que les mutations aléatoires sont de l'ordre de  $2 \cdot 10^{-8}$  (ce taux intervient dans les calculs appréciant l'évolution de l'être humain, en anthropologie par exemple).

Les SNP peuvent être classés selon les conséquences observées ou leur localisation dans le génome :

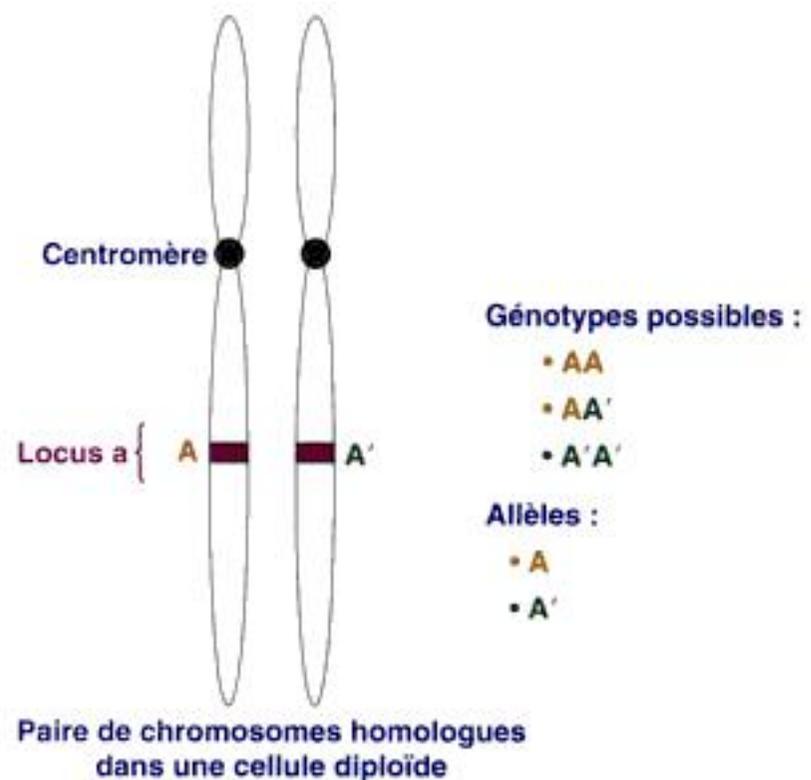
- changement d'acide aminé ;
- absence de changement d'acide aminé ;
- présent dans la région 5' non codante (5'NC) ;
- présent dans la région 3' non codante (3'NC) ;
- présent dans d'autres régions du génome (introns, régions intergéniques).

Chez un sujet normal, on estime le nombre de SNP à environ 11 millions. Les SNP aboutissent à la présence de variations génétiques (variants). Les différentes formes ainsi obtenues du même gène s'appellent des allèles (figure 2.9). Un grand nombre de ces SNP, malgré les variations qu'elles provoquent au niveau des parties codantes et/ou régulatrices d'un gène, n'affectent pas le plus souvent ni la fonction du gène ni celle de la protéine codée par ce gène. Dans certains cas, cependant, ces polymorphismes peuvent prédisposer ou être associés à certaines pathologies en modifiant par exemple la fonction de la protéine codée par ce gène. On parle alors de polymorphisme fonctionnel. Leur rôle dans les maladies multifactorielles telles que l'hypertension artérielle, l'athérosclérose ou certains diabètes a été démontré. L'intérêt des SNP réside aussi dans l'aide qu'ils peuvent apporter à la localisation des gènes de susceptibilité, dans l'étude de l'expression des gènes (notamment en cas de polymorphisme fonctionnel, par exemple en pharmacogénétique), dans les études de population et dans l'étude moléculaire de certaines maladies (étude d'association par exemple, tableau 2.1).

### Intérêts des polymorphismes

Les polymorphismes, que ce soit les STR ou les SNP, présentent un intérêt fondamental dans l'étude du génome humain, sa physiologie et sa physiopathologie. Parmi les nombreuses informations qu'ils apportent, on peut citer leur intérêt dans les domaines suivants :

- localisation et identification de gènes (*mapping*) ;
- mise en évidence d'anomalies chromosomiques (en cancérologie et en génétique par exemple) ;
- prédisposition à une maladie génétique (par exemple, maladie d'Alzheimer) ou non (par exemple, une maladie



**Figure 2.9** Notion d'allèle et de génotype. Le locus a est localisé sur une paire de chromosomes non sexuels (autosomes) dans une cellule diploïde. Ce locus représente une position d'une séquence d'ADN sur le chromosome. De manière générale, elle peut représenter un simple SNP ou une grande région d'ADN (par exemple, le locus HLA). La présence de variants au locus définit des allèles (A et A'). Un allèle peut représenter un variant de gène ou de polymorphisme. Le génotype est l'association au sein du locus des deux allèles d'un même gène (par exemple, AA').

infectieuse comme le paludisme) ou aux conséquences (positives ou négatives) d'un médicament (domaine de la pharmacogénétique) ;

- enquêtes familiales ;
- étude de la biodiversité humaine (anthropologie par exemple) ;
- médecine légale ;
- étude de paternité ;
- étude des maladies multifactorielles ;
- médecine prédictive.

De manière générale, l'étude des SNP est actuellement privilégiée pour plusieurs raisons :

- les SNP sont plus nombreux dans le génome que les STR ;
- les SNP sont plus stables que les STR ;
- certains SNP ont un retentissement fonctionnel.

### Polymorphismes épigénétiques

De nombreuses études réalisées sur des jumeaux monozygotes ont montré des variations dans l'expression de certaines maladies (par exemple, la schizophrénie). Ces résultats (concordance de 50 % entre jumeaux monozygotes dans la schizophrénie) permettent de déduire la responsabilité au moins en partie de facteurs environnementaux dans la physiopathologie de cette maladie (même si à ce jour ces facteurs n'ont pas été clairement mis en évidence). On devrait s'attendre en effet à 100 % de concordance entre les jumeaux monozygotes étant

**Tableau 2.1. Exemples de polymorphismes bi-alléliques (SNP) associés à des maladies**

Gène	Variant(s)	Pathologie associée
Gène de susceptibilité :		
- apolipoprotéine E	Allèle ε4	Maladie d'Alzheimer
- NRAMP1 ( <i>natural resistance-associated macrophage protein 1</i> )	Asp543Asn	Tuberculose
- PPARγ ( <i>peroxisome proliferator-activated receptor γ</i> )	G>C IVS4 Pro12Ala	
Pronostic :		
- MBL2 ( <i>mannose binding lectin</i> )	Variants structuraux B, C et D	Mucoviscidose
- TNF ( <i>tumor necrosis factor</i> )	-376G>A -308G>A dans le promoteur	Neuropaludisme
Pharmacogénétique :		
ADRB2 ( <i>β2 adrenergic receptor</i> )	Gly16Arg	Catécholamines (albuterol)
CYP2C9 (cytochrome P450 IIC, polypeptide 9)	Arg144Cys	Dosage de la warfarine

donné leur identité génétique (ce sont des « clones » naturels). Or, comme on l'a précisé ci-dessus, ce n'est pas le cas. En reprenant l'exemple de la schizophrénie, maladie multifactorielle (et multigénique), des auteurs ont suggéré qu'au cours du développement, il existait une différence d'expression de certains gènes probablement par modification épigénétique différentielle (de mécanisme actuellement inconnu). Parmi les mécanismes physiopathologiques impliqués dans la schizophrénie, l'hypothèse de polymorphismes épigénétiques fonctionnels a été proposée. Cette hypothèse reste à prouver. D'autres études ont montré la variation inter-individuelle des profils de méthylation de l'ADN sans les avoir encore bien caractérisés. On parle alors d'épigénotype. Ces études démontrent également que le profil épigénétique d'un individu intervient dans son phénotype et dans la susceptibilité à certaines pathologies (par exemple, les cancers). L'ensemble de ces observations permet donc de penser qu'au sein du génome, outre les polymorphismes de la séquence nucléotidique, des polymorphismes épigénétiques fonctionnels existent (en dehors des phénomènes d'empreinte génomique déjà évoqués). Il est probable par ailleurs que certaines mutations altèrent le profil de méthylation d'un individu, d'une cellule ou d'un tissu (par exemple, dans des zones de régulation de l'expression d'un gène) et donc altèrent le phénotype. Des interactions polymorphisme épigénétique/mutation existent probablement.

De même que pour les polymorphismes nucléotidiques, les variations épigénétiques (ou polymorphismes épigénétiques), indépendantes de la séquence d'ADN, sont transmises entre générations.

### Polymorphismes de l'ADN mitochondrial

La diversité nucléotidique de l'ADN mitochondrial est supérieure à celle de l'ADN génomique (présent dans le noyau). C'est la raison pour laquelle l'ADN mitochondrial (ADNmt) est souvent utilisé en médecine légale, la phylogénétique et la génétique des populations. Cette diversité est d'ailleurs inégalement répartie dans l'ADNmt. Du fait de cette propriété, en cas de découverte d'une nouvelle variation de séquence, la question se pose souvent de savoir s'il s'agit d'un polymorphisme ou d'une mutation responsable d'une pathologie d'autant plus qu'un polymorphisme peut aussi coségrégier avec la maladie et, s'il est suffisamment rare, peut être pris pour la mutation causale. Sans être absolus, les critères suivants orientent soit vers une mutation, soit vers un polymorphisme :

- la présence d'une hétéroplasmie : on peut imaginer en effet qu'une mutation à l'état homoplasmique soit létale. Malheureusement, ce critère est souvent absent. Les mutations peuvent aussi être homoplasmiques ;
- le pourcentage d'hétéroplasmie dans les tissus atteints et leur comparaison par rapport aux tissus sains (notion de seuil). Le seuil est cependant difficile à évaluer ;
- la comparaison de la séquence avec des séquences contrôles. Le choix des « bons contrôles » est souvent difficile. La comparaison de la séquence étudiée avec celle d'autres espèces peut aider à trancher (sites conservés au cours de l'évolution). Si, au niveau de la séquence étudiée, l'acide aminé est le même à travers les espèces, la probabilité que la mutation soit responsable est grande.

Dans la pathologie mitochondriale, la frontière polymorphisme/mutation n'est donc pas toujours évidente. Les nombreuses études démontrent l'importance de l'environnement et de l'influence du génome nucléaire dans les conséquences fonctionnelles de ces variations de séquence (polymorphisme/mutation). Deux exemples illustrent notre propos :

- le génome de la mitochondrie est relativement instable, par exemple par suite de la formation de radicaux libres produits par la chaîne respiratoire mais aussi du fait de l'absence d'un processus efficace de réparation de l'ADNmt et de l'absence d'histones. Des mutations somatiques sont donc acquises au cours de la vie cellulaire ;
- certains polymorphismes sont sensibles à l'environnement. C'est le cas par exemple de la mutation/polymorphisme A1555G sur le gène codant pour l'ARN 12S, cible des aminosides (Omim 580000). Dans ce cas, en cas d'exposition aux aminosides, les sujets peuvent être atteints d'hypoacousie, voire de surdité (toxicité cochléaire).

### Gènes de susceptibilité – Modulation polygénique des maladies monogéniques

De manière générale, les maladies monogéniques (c'est-à-dire, un seul gène anormal est responsable d'une maladie, par exemple, la drépanocytose) présentent une variabilité importante dans leur expression clinique (expressivité variable). Cette variabilité est bien entendu fonction de facteurs environnementaux multiples mais aussi sous l'influence de gènes modulant cette expressivité ou de polymorphismes au sein du gène ou de ses éléments régulateurs. Les gènes de susceptibilité (il serait plus approprié de parler des allèles de ces gènes associés à une susceptibilité) ne suffisent pas pour induire une maladie. Les allèles de susceptibilité confèrent un risque accru, mais le plus souvent minime, de développer la maladie.

### Polymorphismes fonctionnels

De nombreux polymorphismes n'ont aucune conséquence fonctionnelle (polymorphismes neutres). Certains d'entre eux cependant sont impliqués dans le mécanisme physiopathologique de nombreuses maladies, ce sont les polymorphismes fonctionnels (figure 2.10). Deux grandes classes de polymorphismes fonctionnels peuvent être distinguées schématiquement :

- les polymorphismes modulant la régulation d'un ou de plusieurs gènes. Par exemple, il a été montré le rôle important de certains polymorphismes fonctionnels qui peuvent moduler la fixation de facteurs de transcription sur le promoteur, sur le *locus control region* (LCR) ou sur des *enhancers* ;

- les polymorphismes modifiant la structure (et modulant la fonction) d'une protéine. Dans des pathologies telles que les maladies infectieuses (VIH par exemple), certains polymorphismes modifient l'évolution naturelle de la maladie. C'est par exemple le cas du polymorphisme delta32 sur le gène codant pour la chémokine CCR5 (Omim 601373) associé à une évolution plus lente du sida. D'autres polymorphismes sont associés à des pathologies et constituent des facteurs de risques pour le développement de ces maladies sans être pour autant une condition nécessaire ni suffisante. Des polymorphismes ainsi banaux en apparence peuvent, combinés, jouer un rôle important par exemple dans le développement et/ou l'évolution et/ou le pronostic et/ou la réponse thérapeutique au cours d'une maladie. C'est le cas du polymorphisme variant E4 de l'apolipoprotéine E (Omim 107741 ; gène pour lequel existent deux autres polymorphismes E2 et E3). Celui-ci est associé à la maladie d'Alzheimer sporadique. En termes clairs, cela signifie qu'un sujet présentant ce polymorphisme a un risque plus élevé de développer la maladie (selon les études et les ethnies, ce risque varie de 3 à 7 fois chez un sujet homozygote pour le polymorphisme E4). Cependant, et cela est fondamental, ce n'est ni une condition nécessaire ni suffisante pour développer cette pathologie. Enfin, pour citer un autre exemple, les polymorphismes peuvent jouer un rôle dans la réponse thérapeutique à certains médicaments (ces polymorphismes ne sont d'ailleurs pas toujours fonctionnels). Ce peut être par exemple par modification de la cible, du transport ou du métabolisme du médicament (domaine de la pharmacogénomique).

Ainsi, la mise en évidence de ces polymorphismes fonctionnels est importante. En effet, même si ces effets sont potentiellement mineurs à l'échelle de la protéine, dans le cas d'un réseau métabolique, on peut imaginer les conséquences importantes que de « petits effets » résultant de l'action de ces polymorphismes peuvent avoir sur la résultante globale de ce métabolisme.

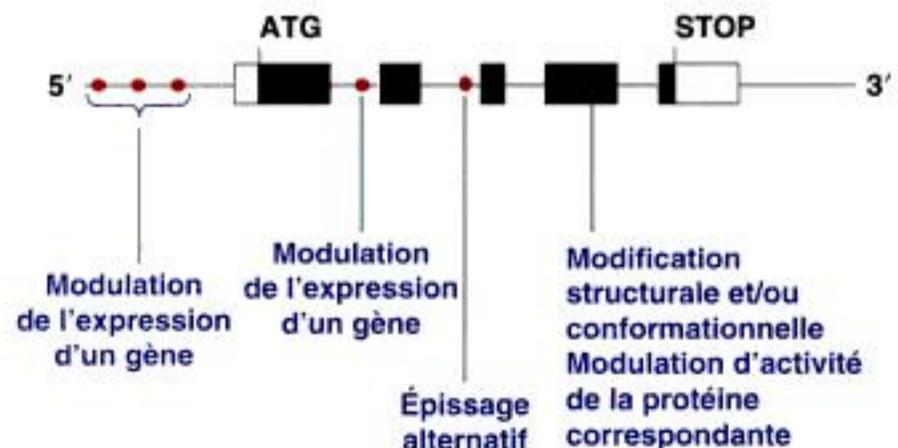


Figure 2.10 Conséquences fonctionnelles et/ou structurales potentielles des polymorphismes fonctionnels.

### Un cas particulier, l'hétérosis

Dans certains cas, un polymorphisme de type SNP à l'état hétérozygote peut présenter une association significativement différente à une pathologie par rapport à chacune des deux classes d'homozygote : c'est la notion d'hétérosis. Ceci est observé par exemple dans le cas d'un polymorphisme du promoteur du gène codant pour une sous-unité de l'interleukine 12 (IL12), IL12B (Omim 161561) dont les sujets hétérozygotes présentent un risque accru d'asthme sévère dans l'enfance par rapport aux sujets homozygotes pour ce polymorphisme.

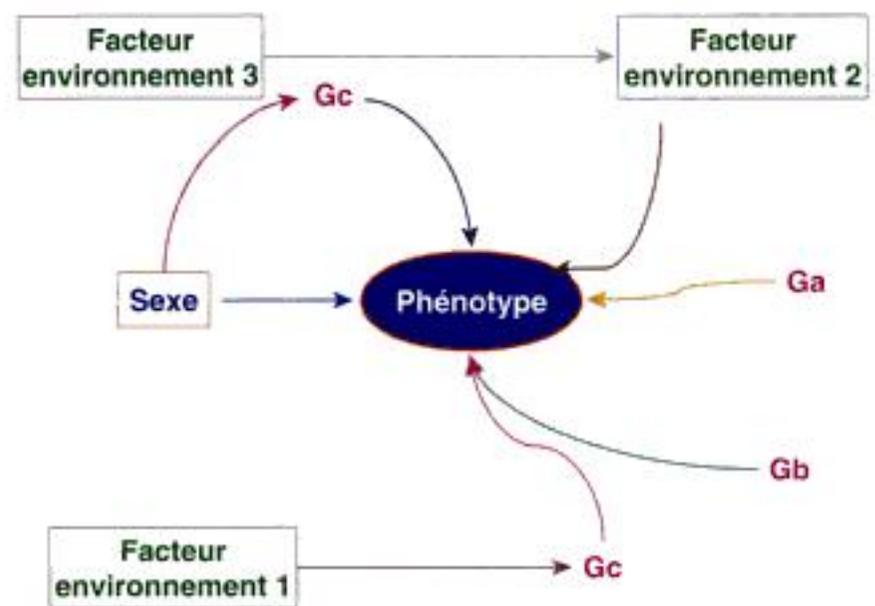
### Gènes modificateurs

Dans un grand nombre de maladies héréditaires monogéniques de transmission mendélienne, on ne retrouve pas de relation génotype/phénotype évidente. Certains gènes jouent un rôle pour moduler (« modifier ») l'expression d'une maladie (par exemple, âge de début de la maladie, symptômes, sévérité). Ce sont les gènes modificateurs. C'est ainsi par exemple que dans le cas de la mucoviscidose, l'iléus méconial est retrouvé dans environ 15 à 20 % des cas avec environ 30 % de récurrence familiale. Il a été démontré que le gène CFM1 (*cystic fibrosis modifier 1*, Omim 603855) était un gène modificateur impliqué dans la présence ou l'absence d'iléus méconial chez les nouveau-nés atteints de mucoviscidose.

Selon le contexte et/ou la situation, un même gène, quand il est muté, peut avoir des conséquences phénotypiques différentes. Cette variabilité phénotypique peut aussi bien être intrafamiliale (oncles, cousins, grands-parents) qu'intranucléaire (même fratrie).

### Comment définir un gène modificateur ?

En dehors de certains cas particuliers, l'identification des maladies monogéniques a montré qu'une mutation seule ne suffit pas pour le développement d'une maladie. Ainsi, dans une même famille, des individus présentant la même mutation peuvent présenter des phénotypes différents ou être sains (pénétrance variable). D'autres facteurs physiopathologiques tels que les facteurs environnementaux (alcool, tabac, état nutritionnel, infections, toxiques) sont nécessaires pour le développement de la maladie. Parmi ces facteurs, les gènes modificateurs jouent aussi un rôle important (figure 2.11). On peut considérer comme gène modificateur d'une maladie héréditaire tout gène dont les différents allèles ne sont pas à l'origine de la maladie mais peuvent modifier/moduler son phénotype. Cette définition schématique n'exclut pas qu'un gène modificateur d'une maladie puisse être lui-même à l'origine d'une autre maladie héréditaire. Par exemple, la mutation C282Y sur le gène HFE (Omim 235200) est associée à l'hémochro-



**Figure 2.11** Interactions multifactorielles. Le phénotype est la résultante d'interactions complexes individuelles (le sexe, l'âge, les génotypes modulateurs) et environnementales multiples. Un certain nombre de gènes (les gènes modificateurs et/ou de susceptibilité, Ga, Gb, Gc...) vont moduler l'expression d'une maladie en interaction entre eux et avec les facteurs environnementaux (1, 2, 3...). L'ensemble de ces interactions aboutit à un phénotype qui peut être variable d'un sujet à l'autre.

matose héréditaire. Cette mutation est aussi un facteur de susceptibilité pour une autre maladie, la porphyrie cutanée (Omim 176100).

Dans le cas des cancers et de la carcinogenèse en général, l'anti-oncogène p53 muté (Omim 191170) est un responsable d'une maladie héréditaire, le syndrome de Li-Fraumeni (Omim 151623, syndrome constitué d'un grand nombre de cancers d'apparition précoce et de localisation multiple tels que leucémies, ostéosarcomes, rhabdomyosarcomes...). Le gène p53 est aussi muté dans un grand nombre de cancers sporadiques (et donc non héréditaires) et peut être considéré alors comme un gène modificateur.

En résumé, Il est difficile de mettre en évidence les gènes modificateurs, ceux-ci participant à un réseau métabolique complexe en cours de déchiffrement.

### Maladies polygéniques et multifactorielles

Un grand nombre de pathologies sont en fait des maladies multifactorielles pour lesquelles l'influence polygénique a été abondamment prouvée. On peut citer, par exemple, l'obésité, les maladies cardiovasculaires, les schizophrénies. Ces maladies complexes résultent de très nombreuses interactions alléliques, constituant un réseau combinatoire complexe dont la résultante fonctionnelle aboutit ou non au développement de la maladie (figure 2.11). Ces variations peuvent être la conséquence d'une combinatoire à effet additif ou synergique de gènes ou de l'association au sein d'un cluster de gènes (gènes transmis d'un bloc [sans recombinaison] au niveau d'un locus) dont l'expression basale est modulée par d'autres gènes

**Tableau 2.2. Mutations rapportées dans la littérature (2093 gènes analysés)**  
(référence : <http://www.hgmd.cf.ac.uk/docs/hohoho.html>)

Mutation	Nombre	%
<b>Microlésions :</b>		
- mutations ponctuelles (non-sens, faux-sens)	31 251	57,3
- épissage	5188	9,5
- zones de régulation de l'expression	700	1,3
- petites délétions	9014	16,5
- petites insertions	3626	6,6
- petites insertions/délétions	533	1
<b>Macrolésions :</b>		
- variation de séquences répétées	143	0,2
- grandes insertions ou duplications	470	0,9
- réarrangements complexes (incluant les inversions)	639	1,2
- grandes délétions	2998	5,5
<b>Total</b>	<b>54 562</b>	<b>100</b>

(hors du cluster). Les facteurs environnementaux additionnels sont multiples. À titre d'exemple, nous pouvons citer les produits chimiques, le tabac, les médicaments, les agents infectieux (virus, bactérie, champignon...), l'alcool.

## Pathologie moléculaire : les mutations

Une mutation est une variation allélique anormale pouvant aboutir au développement d'une maladie. De nombreuses formes de mutations ont été décrites (tableau 2.2). Selon le type de lésion moléculaire observée, on distingue les mutations de grande taille (les macrolésions) et celles de petites tailles (les microlésions).

### Macrolésions

Elles constituent des réarrangements géniques dont la taille peut aller de quelques kilobases à plusieurs mégabases. On distingue plusieurs formes (figure 2.12).

#### Délétion

C'est l'excision d'un segment d'ADN pouvant aller de quelques bases à plusieurs millions de bases, voire un

chromosome. Elles sont souvent la conséquence de mécanismes de recombinaison. Dans les cas de délétion de gène ou de grande partie de gène, les mécanismes de recombinaison sont souvent observés au sein de famille de gènes regroupés en groupe (clusters), comme par exemple les gènes de la famille des globines (par exemple, les gènes alpha situés sur le chromosome 16p13.3 et dont les anomalies aboutissent à des alpha-thalassémies, Omim 141800 et 141850).

#### Duplication

C'est la « répétition » d'un fragment plus ou moins long de l'ADN (figure 2.13). La duplication de tout ou partie d'un gène peut se faire dans la même orientation ou en orientation inverse (inversion, voir infra) du gène dupliqué. Les mécanismes de recombinaison sont principalement impliqués dans leur formation. Dans les cas de duplication de gène ou de grande partie de gène, parmi les mécanismes de recombinaison observés, on peut citer :

- les recombinaisons observées au sein de famille de gènes regroupés en groupe (clusters), comme les gènes de la famille des cytochromes P450 ;
- celles observées au sein de séquences répétées comme dans la maladie de Charcot-Marie-Tooth de type IA

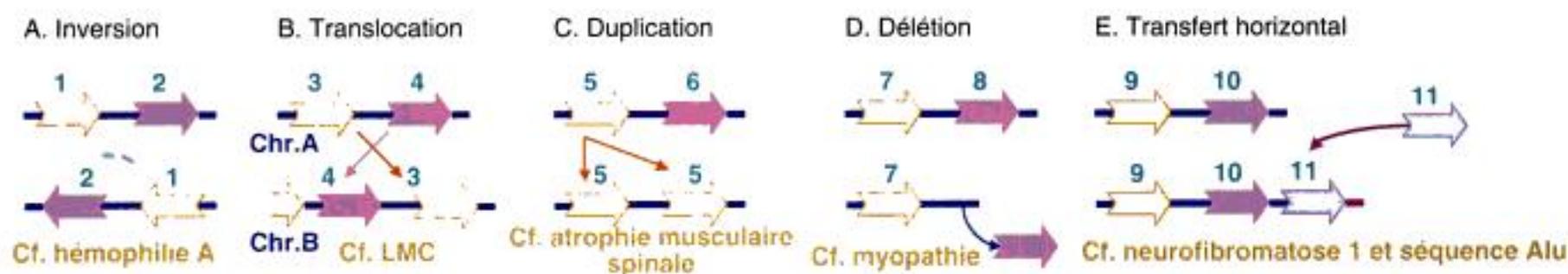
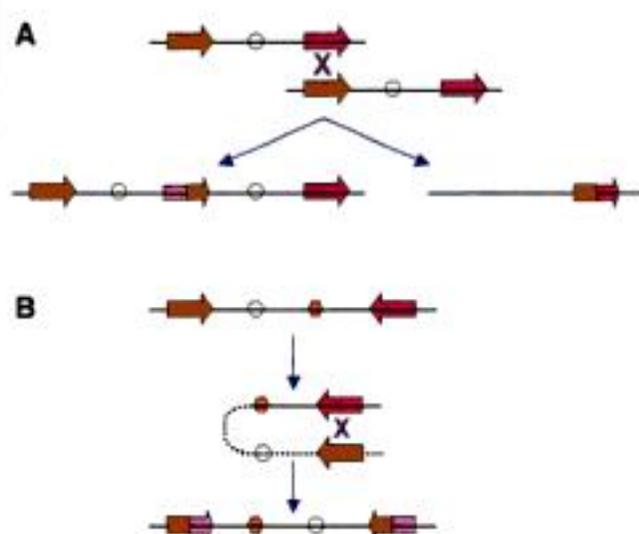


Figure 2.12 Réarrangements géniques -Exemples.



**Figure 2.13** Les mécanismes des réarrangements sont complexes. Dans ce schéma, deux exemples simples de réarrangements pendant la méiose sont décrits. Il s'agit d'une recombinaison non homologe entre deux séquences répétées (*low copy repeats*, LCR) représentées par des flèches de couleur différente et encadrant une séquence génomique (cercle ou hexagone). A. Crossover entre deux LCR de séquence proche et de même direction. Le résultat est soit une duplication, soit une délétion d'un segment génomique unique (cercle). Une recombinaison entre deux chromosomes ou intrachromosomes (inter- ou intrachromatide) peut aboutir à un tel réarrangement. B. En cas de recombinaison de deux LCR orientés en sens opposé, on obtient une inversion des segments génomiques (cercle et hexagone) sans perte ni gain de segment génomique. La recombinaison interchromosomique ou interchromatide peut provoquer un changement conformationnel important des chromosomes (chromosomes dicentrique et acentrique).

(CMT1A, Omim 118220) dans laquelle, parmi les mécanismes pathomoléculaires, on retrouve une duplication sur le chromosome 17 en p12-p11.2 par recombinaison non homologe ;

- celles à partir de séquences Alu comme dans le syndrome d'Ehlers-Danlos variant VI (Omim 225400).

### Inversion

C'est un changement d'orientation d'un segment d'ADN (figure 2.13). Cette mutation a été retrouvée par exemple dans la maladie de Charcot-Marie-Tooth de type 1, CMT1A, Omim 118220). Elles résultent en général de recombinaisons génétiques complexes dont les mécanismes précis sortent du cadre de cet ouvrage.

### Amplification

C'est la répétition le plus souvent en tandem de séquences normalement uniques (la taille du « tandem » peut être considérable). On peut le voir par exemple dans les cancers du sein où l'amplification du gène ERBB2 (*v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2*, Omim 164870) est associé à une résistance aux agents anticancéreux placicaxel et au tamoxifène.

### Fusion de gènes

Elle est le résultat d'un réarrangement avec double cassure de deux gènes différents et « transposition » de l'un dans l'autre. Elle peut aboutir à la formation de protéines chimériques aux propriétés délétères pour l'organisme.

### Conversion

Il s'agit d'une recombinaison non réciproque d'un fragment d'ADN aboutissant à la conversion apparente d'un fragment d'ADN en un autre (figure 2.14). Le mécanisme est complexe, souvent mal connu et ne sera pas décrit. On retrouve ce type de réarrangement dans la maladie de Gaucher de type I (Omim 606463 et 230800) ou le syndrome de Shwachman-Diamond (Omim 260400).

### Insertion

Il s'agit de l'introduction d'une séquence (par exemple transposon ou séquence virale) dans un gène. La mutation par insertion résulte de plusieurs mécanismes complexes (par exemple, recombinaison inégale, translocation, conversion génique, « boucles chromatiniennes »).

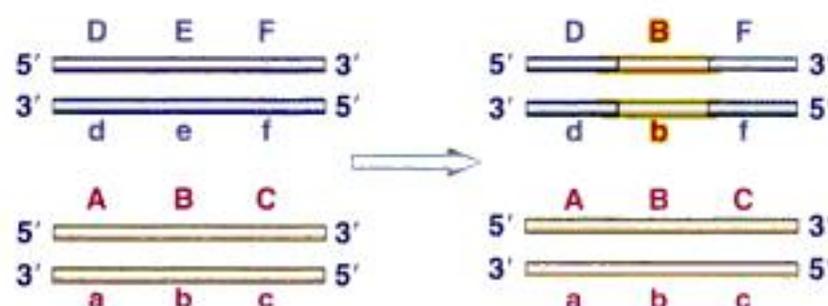
### Mécanismes complexes

Dans certains cas, un grand réarrangement est la résultante de mécanismes complexes et pouvant associer par exemple une délétion et une duplication (délétion/duplication) ou une délétion et une inversion (délétion/inversion). Ce genre de mécanisme complexe et rare a été décrit par exemple dans les myopathies de Duchenne et de Becker (Omim 310200 et 300376).

Le tableau 2.3 donne quelques exemples de pathologie héréditaire humaine résultant de réarrangements géniques.

### Microlésions

Ce sont des substitutions, suppressions ou additions d'un ou de quelques nucléotides. Lorsqu'une base est transformée en une autre base, on parle de mutation ponctuelle.



**Figure 2.14** Conversion génique.

Tableau 2.3. Exemples de gènes soumis à un effet de dose

Gène	Réarrangement causal	Pathologie – Omim
PMP22 (protéine myéline périphérique 22)	Duplication Délétion	Maladie de Charcot-Marie-Tooth (CMT1A) – 118220 Neuropathie héréditaire (HNPP) – 162500
PLP1 (protéolipide 1)	Duplication Délétion	Maladie de Pelizaeus-Mezbacher (PMD) – 312080
ELN (élastine)	Délétion	Maladie de Williams-Beuren (WBS) – 194050
JAG1 (jagged 1)	Délétion	Syndrome d'Alagille (AGS) – 118450
WT1	Délétion	Syndrome WAGR – 194072

### Généralités sur les mutations ponctuelles

De nombreux mécanismes sont responsables de ces mutations. Par exemple :

- dépurination ;
- désamination notamment des cytosines méthylées :  $^{mC} \rightarrow T$  sur le brin sens (et  $G \rightarrow A$  sur le brin antisens) ;
- erreurs de réplication ou de réparation.

Les termes ci-dessous sont parfois employés. Ils sont définis de la manière suivante :

- transition : mutation ponctuelle dans laquelle une base purique est transformée en une autre base purique (par exemple,  $G \rightarrow A$ ) ou une base pyrimidique en une autre base pyrimidique (par exemple,  $C \rightarrow T$ ) ;
- transversion : mutation ponctuelle dans laquelle une base purique est transformée en une base pyrimidique (par exemple,  $G \rightarrow T$ ) ou une base pyrimidique en une base purique (par exemple,  $C \rightarrow A$ ).

Le tableau 2.4 donne quelques exemples de mutations ponctuelles.

### Mutations dans les régions codantes

#### Mutations faux-sens

Elles entraînent le changement d'un acide aminé sur la protéine. Leur conséquence sur le fonctionnement de la protéine est variable. Dans certains cas, elles peuvent aussi être responsables d'une anomalie d'épissage.

#### Mutations silencieuses

Le plus souvent, elles modifient le codon mais ne modifient pas l'acide aminé. Cependant, dans certains cas, elles peuvent être responsables d'un épissage anormal de l'exon notamment lorsque cette mutation est située sur une séquence *exonic splicing enhancer* (ESE), séquence d'environ 6–8 nucléotides située dans des exons et nécessaire pour un épissage efficace. Il existe aussi d'autres séquences spécifiques intervenant dans l'épissage

normal, les *exonic splicing silencers* (ESS), favorisant un saut d'exon ou au contraire notamment sur le dernier exon, inhibant un saut d'exon. Une mutation sur ces séquences peut aussi aboutir à des anomalies d'épissage.

#### Mutations non-sens

Elles aboutissent à la formation d'un des trois codons-stop UAA, UAG ou UGA. Elles sont responsables d'une protéine tronquée, le plus souvent non fonctionnelle (parfois, une activité résiduelle peut persister) ou de la dégradation de l'ARNm correspondant (et absence de synthèse d'une protéine anormale). Dans certains cas, les mutations non-sens sont responsables d'épissage anormal d'exon (« saut d'exon »).

Certaines mutations ne modifient pas le phénotype (fonction de la protéine) : ces mutations sont dites conservatives. Les mutations qui affectent probablement/sûrement le phénotype d'une protéine sont dites non conservatives.

### Mutations perturbant le cadre de lecture

L'insertion ou la délétion d'une ou de plusieurs bases dans la région codante décale le cadre de lecture (*open reading frame*, ORF) et provoque l'apparition d'un codon-stop prématuré plusieurs bases (une dizaine à une centaine en général) en aval de la mutation. Les principales conséquences sont la formation d'une protéine tronquée (plus petite) ou la dégradation de l'ARNm correspondant (et absence de synthèse de la protéine normale).

### Mutations perturbant l'épissage

En dehors des anomalies d'épissage consécutives aux atteintes des protéines du spliceosome ou des signaux cellulaires agissant sur le phénomène d'épissage, de nombreuses

**Tableau 2.4A. Exemples de mutations ponctuelles.**  
**1) Sur un codon-sens**

Codon normal	Codon muté (mutation ponctuelle)	Conséquence théorique au niveau protéique	Conséquence fonctionnelle
GAA (acide glutamique)	GAG (acide glutamique) (A→G)	Pas de modification d'acide aminé	Mutation silencieuse : pas de conséquence Saut d'exon (mutation dans un ESE) : – isoforme protéique (polymorphisme) – mutation pouvant induire une maladie (absence de protéine, protéine tronquée fonctionnelle ou non)
GAG (acide glutamique)	GTG (valine) A→T	Faux-sens	Polymorphisme fonctionnel ou non Mutation ponctuelle pouvant induire une maladie Saut d'exon (mutation dans un ESE)
CAG (glutamine)	TAG (codon-stop) C→T	Non-sens	Arrêt prématuré de la traduction avec ARNm instable (absence de protéine) ou protéine tronquée (fonctionnelle, hypo- ou non fonctionnelle) Saut d'exon (mutation dans un ESE) responsable soit d'une protéine tronquée (plus ou moins fonctionnelle ou non), soit de l'absence de protéine

ESE : *exonic splicing enhancer*.

**2) Sur un codon-stop**

Codon normal	Codon muté (mutation ponctuelle)	Conséquence théorique au niveau traduction	Conséquence fonctionnelle
TAA	TAG (codon-stop) A→G	Non-sens	– Protéine tronquée – Saut d'exon – ARNm instable (absence de protéine)
TAA	CAA (glutamine) T→C	Faux-sens	Mutation induisant une maladie

**Tableau 2.4B. Exemples de mutations non ponctuelles**

Mutation	Conséquence possible
Délétion (1 à plusieurs pb)	Protéine tronquée ARNm instable
Insertion (1 à plusieurs pb)	Protéine tronquée ARNm instable
Insertion/délétion	Protéine tronquée ARNm instable

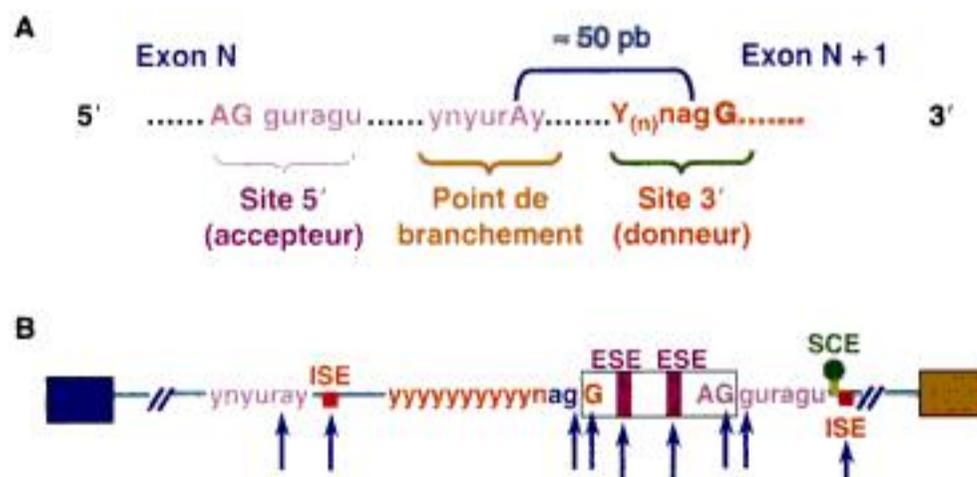
mutations ponctuelles ou non peuvent modifier l'épissage normal (*splicing mutation*).

L'épissage, pour être correct, nécessite certains sites situés aux jonctions exon/intron et caractérisés par des séquences consensus (figure 2.15). Ainsi, au niveau des introns, ces séquences sont GU pour le site 5' (site donneur), Y<sub>n</sub>AG pour le site 3' de l'intron (site accepteur) et le résidu A pour le site de branchement. En dehors de ces séquences consensus, d'autres séquences moins consensuelles adjacentes à celles-ci sont aussi nécessaires. Par ailleurs, certaines séquences différentes de celles-ci et moins consensuelles jouent un rôle important comme les *intronic splicing enhancers* (ISE) et les ESE situés respectivement dans certains introns et certains exons. Il est important de retenir que le phénomène d'épissage résulte de l'interaction de nombreux signaux présents au sein des introns et des exons. Ainsi, le processus d'épissage est un phénomène à la fois complexe (plus d'une centaine de protéines interviennent dans cette réaction) et d'une extrême précision. En pathologie moléculaire, les mutations d'épissage représentent environ 15 % des anomalies responsables de maladies génétiques. Les conséquences de ces mutations sont nombreuses (figures 2.16 et 2.17) :

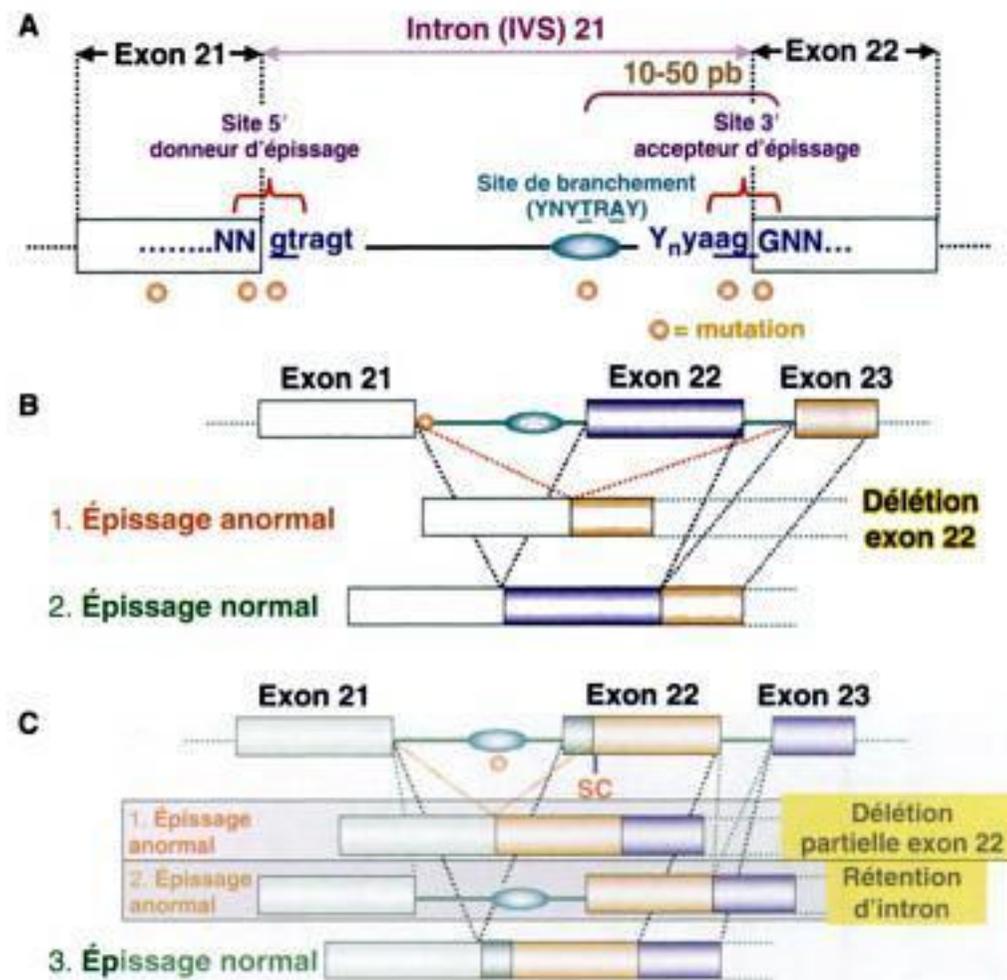
- saut d'exon (délétion) : la conséquence la plus fréquente. Ce saut d'exon aboutit dans la moitié des cas environ à la formation d'une protéine tronquée ;
- activation d'un site cryptique d'épissage dans un exon ou dans un intron. En dehors des zones d'épissage normales, des sites « atypiques » d'épissage, dits sites « cryptiques », peuvent être activés. Ils peuvent être situés dans un exon (l'épissage anormal raccourcira l'exon) ou dans un intron (une portion d'intron sera retrouvée sur le transcrite mature). Le plus souvent, un codon-stop apparaît rapidement en aval ;
- création d'un « pseudo-exon » dans un intron ;
- rétention d'un intron.

De part et d'autre de l'exon, au niveau des jonctions exon/intron se trouve une zone de consensus d'environ une dizaine de pb avec notamment un site donneur 5'-gt [bases +1 et +2 de l'intron suivant un exon] et un site accepteur 5'-ag [dernières bases -1 et -2 de l'intron précédant un exon]. Les sites donneurs et accepteurs de l'intron sont les plus fréquemment mutés. Une mutation dans cette zone de consensus provoque « l'omission » de l'exon, le saut d'exon (appelé aussi *exon skipping*, figure 2.16). La mutation au niveau du site accepteur ou donneur (séquences consensus) est la cause la plus fréquente des mutations aboutissant à un saut d'exon (environ 60 % des causes de saut d'exon). Les mutations au niveau du site donneur 5' sont plus fréquentes que celles au niveau du site accepteur 3' (62 % versus 26 %). Dans l'intron, au site donneur 5', les mutations au résidu G en +1 et celles au résidu en +5 sont les plus fréquemment décrites. Au site accepteur 3', les mutations en -1, -2 et plus rarement -3 sont les plus fréquentes. La mutation peut perturber le cadre de lecture (*frame-shift*) ou ne pas le modifier. On obtient alors une protéine « tronquée ». La protéine résultante peut ou non avoir une activité résiduelle. On peut aussi obtenir des conséquences à la fois différentes et simultanées d'une même mutation. Par exemple, une mutation (1592 + 5G > A) dans l'intron 10 du gène codant pour le récepteur au LDL (LDLR, responsable d'une hypercholestérolémie familiale, Omim 143890) induit le saut de l'exon 10 et la création d'un site cryptique d'épissage responsables d'un ARNm anormal provoquant la pathologie.

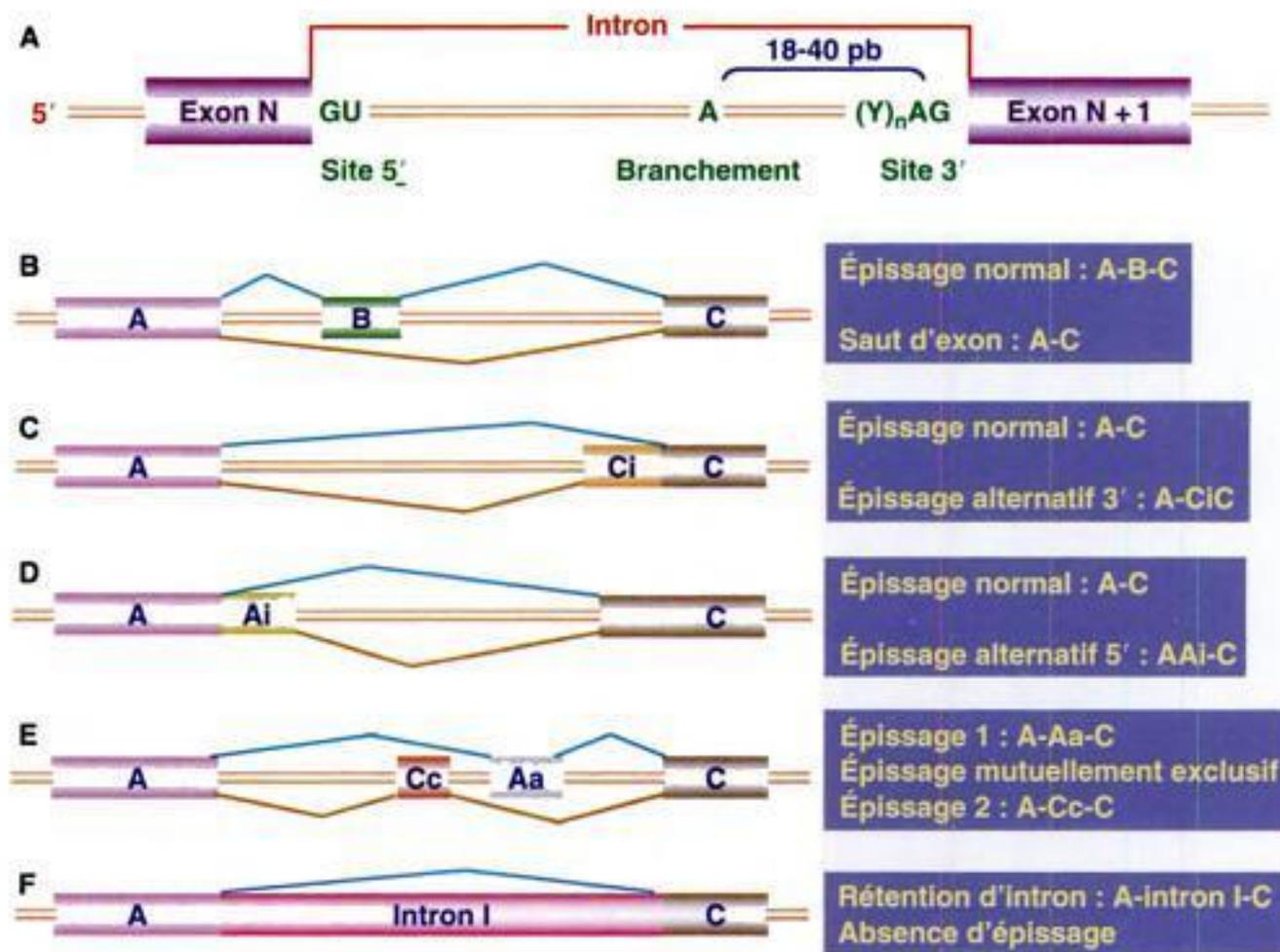
Un saut d'exon peut aussi ne pas avoir de conséquence pathologique et être un polymorphisme (responsable alors d'isoformes telles que des isoenzymes par exemple, figure 2.17). L'épissage alternatif peut être tissu-spécifique ou variable selon le stade cellulaire ou de développement.



**Figure 2.15** Sites consensus d'épissage. **A.** Séquence consensus général des sites 5' et 3' d'épissage et du point de branchement. r : base pyrimidique ; u : base purique ; n : n'importe quelle base. **B.** Exemples de sites de mutations pouvant être responsables de saut d'exon, de rétention d'intron ou de création d'un site cryptique d'épissage (SCE). On peut avoir ces mutations au niveau du site de branchement, des sites donneurs ou accepteurs, des ISE (dans les introns) ou des ESE (dans les exons). Ces mutations sont le plus souvent des mutations ponctuelles. Le plus souvent, ces anomalies aboutissent à la formation d'un codon-stop et la dégradation de l'ARNm par *nonsense mediated decay* (NMD).



**Figure 2.16** Épissage anormal - Exemples. **A.** Sites consensus d'épissage. Ces séquences interagissent avec le spliceosome pour permettre un épissage normal. Les ovales correspondent à des positions de mutations pouvant induire un épissage anormal. **B.** Exemple d'une mutation au site donneur d'épissage de l'intron 22. On observe le saut (absence) de l'exon 22 (délétion de l'exon 22). **C.** Exemple de mutation au point de branchement. Deux conséquences peuvent être observées : un site d'épissage caché (cryptique) situé dans l'exon 22, c'est-à-dire non utilisé par le spliceosome dans les conditions physiologiques habituelles, est activé. Ce site cryptique (SC) devient le nouveau site accepteur et une partie de l'exon 22 est épissée ; le spliceosome ne va pas épisser normalement l'intron 21, il y a rétention de cet intron.



**Figure 2.17** Épissages alternatifs. Cette figure donne quelques exemples de diversité d'épissage soit normal (aboutissant alors à des isoformes de protéines), soit pathologique. **A.** Les motifs conservés aux extrémités de l'intron sont les dinucléotides GU en 5' et AG en 3', ce dernier étant précédé d'une séquence polypyrimidique (Y) n. Le résidu A est le nucléotide conservé au niveau du site de branchement à 30-50 pb de l'extrémité 3' de l'intron. En B, C, D, E, F, cinq modes d'épissage alternatif sont schématiquement décrits. En F, il n'existe pas d'épissage.

Comme il a été dit plus haut, certaines mutations non-sens, faux-sens ou silencieuses peuvent être responsables d'un saut d'exon. Dans ce cas, une des hypothèses évoquées est l'abolition fonctionnelle par la mutation d'un site stimulant physiologiquement l'épissage, le site ESE, cette séquence intervenant dans l'épissage normal. Une autre hypothèse proposée est la modification de la structure secondaire, essentielle à l'épissage normal. L'épissage qui en résulte peut alors soit respecter le cadre de lecture, soit provoquer un décalage de lecture (*frame-shift*) et l'apparition d'un codon-stop prématuré. Ainsi, une mutation dans la partie codante d'un exon peut provoquer une anomalie d'épissage (saut d'exon par exemple) au lieu d'une modification d'acide aminé (ou être silencieuse s'il s'agit d'une mutation silencieuse).

Une mutation au niveau du site de branchement, site intervenant dans la réaction d'épissage, peut entraîner l'absence d'épissage. Dans ce cas, l'intron n'est pas éliminé et, par conséquent, est conservé dans l'ARNm et dans la protéine quand elle est synthétisée, l'ARNm pouvant aussi être instable et dégradé.

En résumé, les sauts d'exons peuvent être la conséquence de mutations à plusieurs sites (figure 2.18) :

- mutation au niveau du site donneur (SD) 5' (saut d'exon complet) ;
- mutation au niveau du site accepteur (SA) 3' (saut d'exon complet) ;
- mutation au niveau du site de branchement ;
- mutation au niveau des séquences moins consensuelles adjacentes au SD ou au SA. On peut alors observer une modulation du saut d'exon. Par exemple, sur le gène CFTR, il existe un polymorphisme du nombre de polypyrimidines [T] T<sub>5</sub>, T<sub>7</sub>, T<sub>9</sub> au niveau de l'intron 8 (IVS8)

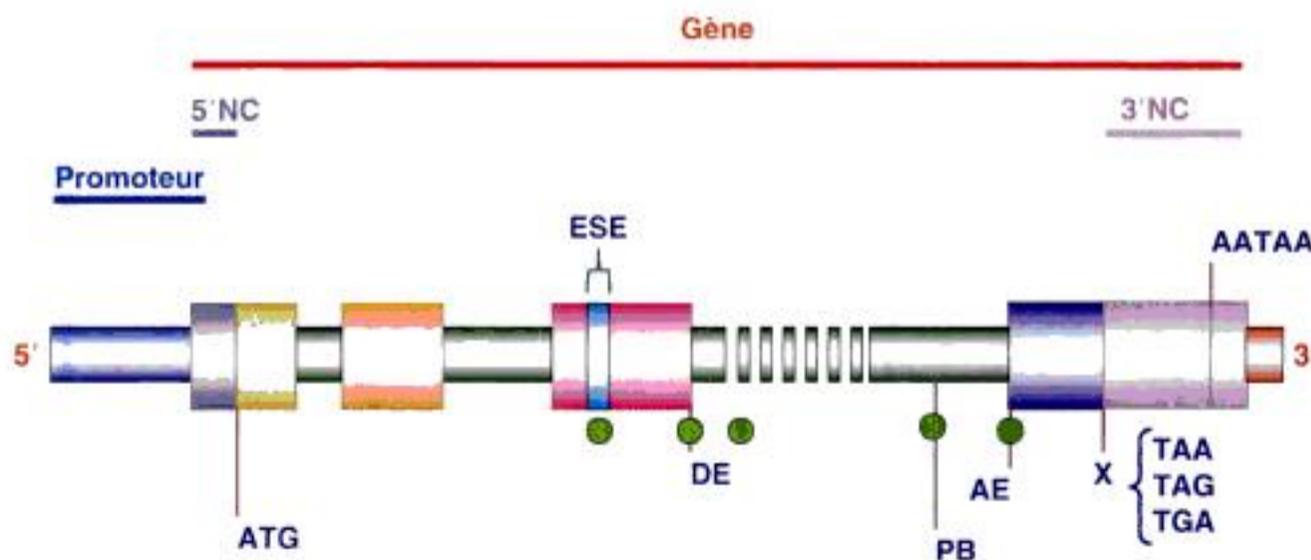
proche du SA en 3' de l'intron. Ce polymorphisme est responsable d'une diminution de l'épissage normal de l'exon 9 (modulation d'épissage), aboutissant à un saut d'exon partiel de ce dernier. Il existe probablement une variation de la régulation du spliceosome ;

- mutation (non-sens, faux-sens, silencieuse) dans l'exon, altérant un ESS ou un ESE. Selon les cas, il a été rapporté des sauts d'exon complet ou des modulations de saut d'exon.

Ces mutations peuvent abolir les sites alternatifs d'épissage aboutissant dans certains cas soit à des interactions anormales entre les différentes isoformes, soit à des pertes de fonction si les deux isoformes ne possèdent pas les mêmes propriétés.

### ARNm et codon-stop – Dégradation de l'ARNm consécutive à la création d'un codon-stop

Un système de surveillance complexe et encore mal connu de l'ARNm entraîne la dégradation de celui-ci (*nonsense-mediated mRNA decay*, NMD) lorsqu'il y a création d'un codon-stop prématuré. Les mécanismes aboutissant à la formation d'un codon-stop prématuré sont nombreux. On peut citer, par exemple, les mutations ponctuelles faux-sens, le décalage de lecture, le réarrangement génique, l'épissage anormal, l'erreur de transcription, l'activation d'un site cryptique, l'utilisation de sites d'initiation mineurs de la traduction (AUG), la rétention d'un intron ou le saut d'un exon. En effet, un codon-stop anormal peut amener à la formation d'une protéine non fonctionnelle ou aux conséquences délétères. Le système NMD est donc un système fondamental pour empêcher la formation de protéines tronquées pouvant avoir un effet dominant négatif.



**Figure 2.18** Mutations d'épissage. Les cercles représentent la localisation de mutations induisant une anomalie d'épissage. Les rectangles représentent les exons. ESE : *exonic splicing enhancer* ; DE : site donneur d'épissage ; AE : site accepteur d'épissage ; PB : point de branchement ; X : codon-stop ; AATAA : signal de polyadénylation ; ATG : site d'initiation de la traduction ; 5'NC : région 5' non codante du gène ; 3'NC : région 3' non codante du gène.

Ce mécanisme est même parfois activé au cours de la synthèse d'un ARNm normal lorsque des erreurs d'épissage se produisent et joue aussi un rôle dans la régulation de l'épissage alternatif normal. En général, ce système est activé si le codon-stop anormal est créé au moins à 50 pb en amont en 5' de la jonction de deux exons. En effet, un complexe protéique important vient se fixer au niveau du codon-stop constituant un espace volumineux. Certains auteurs ont aussi rapporté l'échappement au NMD pour des mutations non-sens proches de l'AUG initiateur. De même, certains gènes au cadre de lecture (ORF) court résisteraient à l'action du NMD. Son action aboutit à la destruction de l'ARNm pathologique (figure 2.19). Certaines mutations non-sens échappent donc à ce système de surveillance. Il en est de même lorsqu'il existe un site d'entrée interne de ribosome (*internal ribosomal entry site*, IRES), autre possibilité d'échappement à ce système de surveillance. Comme on peut s'en rendre compte, le système NMD est complexe et de nombreuses inconnues demeurent sur son mécanisme d'action précis.

En pratique, une mutation d'épissage peut être le résultat de nombreuses formes différentes de mutation. Toute suspicion de mutation d'épissage devrait donc être confirmée par l'étude de l'ARNm soit après RT-PCR,

soit après création d'un minigène artificiel dans un plasmide d'expression transfecté dans une cellule.

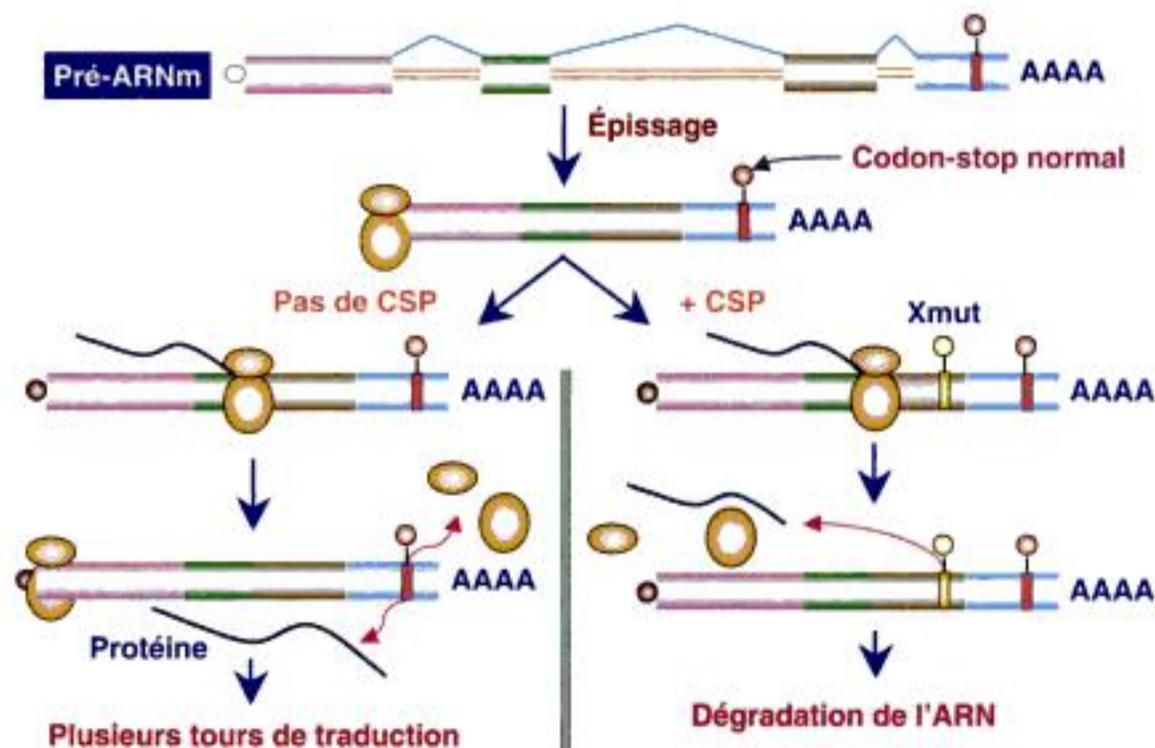
### Mutations non ponctuelles

Ce sont les délétions, insertions ou association des deux (tableau 2.4B). Les conséquences sont identiques à celles observées pour des mutations ponctuelles.

### Position des mutations sur un gène et leurs conséquences

En pathologie moléculaire, une mutation peut être localisée en n'importe quel point d'un gène ou des régions extragéniques interagissant avec lui. La figure 2.20 résume les conséquences possibles des mutations selon leur localisation. Selon les pathologies, plusieurs cas de figures sont possibles. Par exemple :

- une mutation peut être responsable de la quasi-totalité des cas (par exemple la drépanocytose, Omim 603903) ;
- le nombre de mutations identifiées est très important sans prédominance pour telle ou telle mutation (par exemple, la porphyrie aiguë intermittente, Omim 176000). Dans ce cas, on dit qu'il n'y a pas de point chaud (*hotspot*) de mutation ;



**Figure 2.19** Dégradation de l'ARNm par l'intermédiaire d'un codon-stop (*nonsense-mediated mRNA decay*, NMD). Une mutation, de quelque nature qu'elle soit, responsable de la formation d'un codon-stop prématuré, peut aboutir à la dégradation de l'ARNm. Au cours de la traduction, un certain nombre de complexes protéiques jouent un rôle déterminant en présence des ribosomes (comme le complexe *exon-junction complex* [EJC] et les protéines *up-frameshift suppressor proteins* [UPF]). Lors de la rencontre d'un codon-stop prématuré (CSP), le ribosome ne peut déplacer certains complexes tels que EJC provoquant alors la dégradation de l'ARNm. Ce mécanisme évite la formation de protéines à effet dominant négatif. Ce mécanisme joue aussi un rôle de régulateur dans l'épissage alternatif. Il est aussi observé lors des anomalies d'épissage pouvant se dérouler sur un ARNm normal.

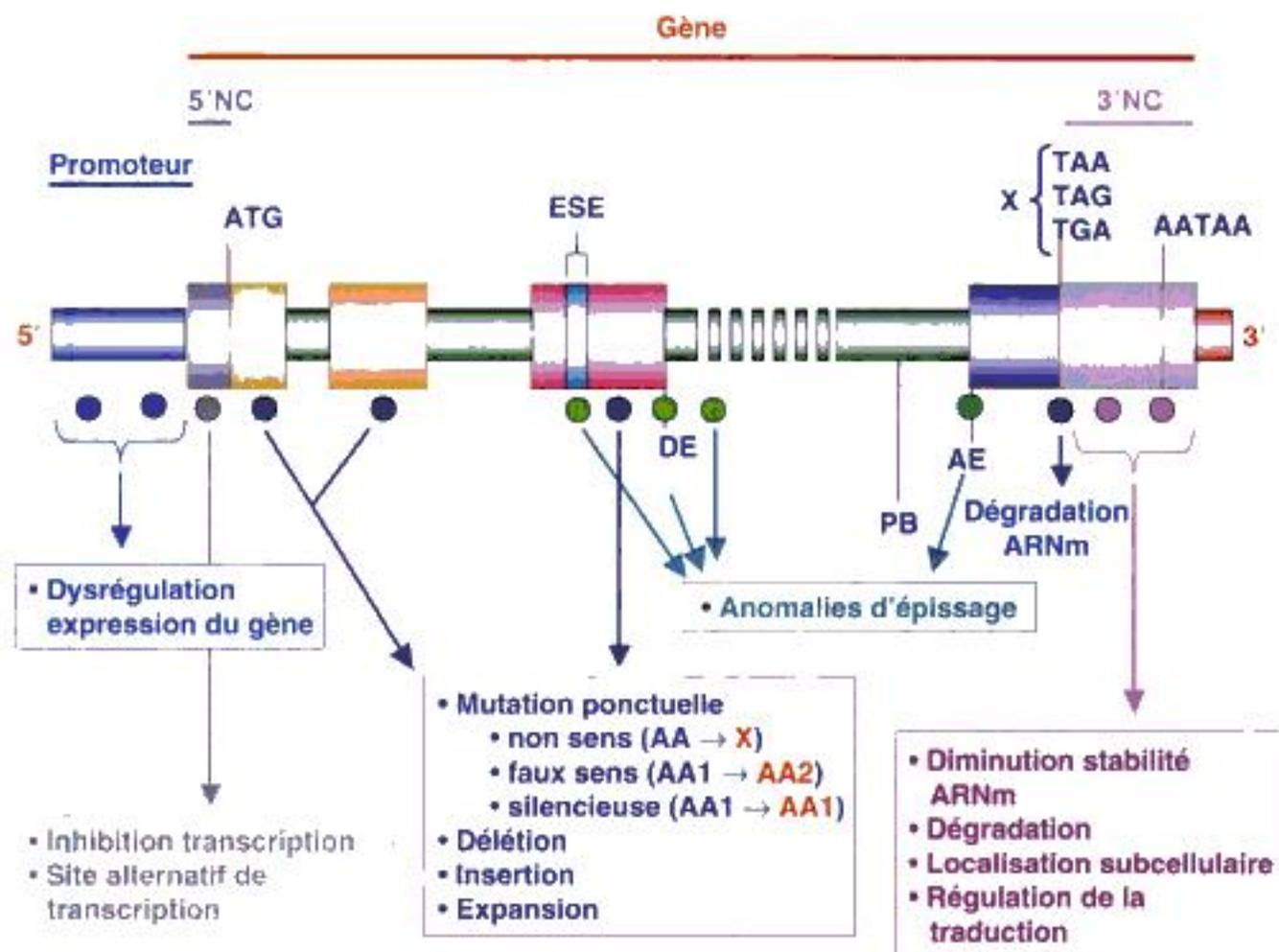


Figure 2.20 Principales localisations des mutations sur un gène et les conséquences de ces mutations.

- le nombre de mutations est important avec prédominance d'une mutation sur les autres (par exemple, la mucoviscidose dans laquelle plus de 1200 mutations sont répertoriées et une seule d'entre elles, la deltaF508, représente environ 70 % des cas, Omim 219700) ;
- la majorité des mutations sont situées dans une ou plusieurs régions du gène. Ces régions constituent des points chauds de mutations (*hotspot* ; par exemple, le gène suppresseur de tumeur TP53, Omim 191170). Deux régions ont des conséquences fonctionnelles que nous allons développer.

### Mutation dans la région 5' non codante (5'NC) et traduction

Parmi les conséquences possibles des mutations en 5'NC, deux mécanismes originaux peuvent être observés (figures 2.21 et 2.22) :

- certains ARNm présentent des séquences régulatrices sur leur extrémité 5' non codante. Ces séquences en forme de boucle (*hairpin*) permettent la régulation de leur traduction après fixation de protéines. C'est le cas par exemple des ferritines H et L (respectivement, Omim 134790 et 134770) contenant des séquences *iron responsive element* (IRE). Sur ces dernières peuvent se fixer des protéines *iron regulatory protein* (IRP). Ce mécanisme intervient dans la régulation du métabolisme du fer. La fixation des IRP sur les IRE des ferritines L et H en cas de déplétion

en fer aboutit à l'inhibition de leur traduction. Une mutation ou une délétion sur cette IRE aboutit à la perte de régulation de la traduction par ce mécanisme et à une hyperferritinémie (syndrome hyperferritinémie/cataracte, Omim 600886) ;

- certains ARNm présentent des uAUG (*upstream AUG*) en 5' non codant réprimant ou limitant la traduction de l'AUG physiologique (site d'initiation de la traduction). C'est le cas par exemple pour la thrombopoïétine (Omim 600044) qui présente plusieurs uAUG. Dans certains cas, la thrombocytémie héréditaire (Omim 187950), liée à une surproduction de thrombopoïétine (TPO), est la conséquence d'une mutation sur un uAUG aboutissant à l'augmentation de synthèse de la TPO par perte de répression de la traduction de l'uAUG.

### Mutations dans la région 3' non codante d'un gène

La région 3' non-codante (3'NC) joue un rôle important dans la régulation de l'expression des gènes. Une mutation dans cette région peut donc en modifier l'expression (figure 2.23). Elle peut aussi avoir un effet dominant négatif en séquestrant des protéines régulatrices interagissant en *trans* avec elle. À titre d'exemple, la dystrophie myotonique de Steinert (Omim 160900) de transmission autosomique dominant est la conséquence d'expansion anormalement élevée du triplet CTG dans la région 3'NC

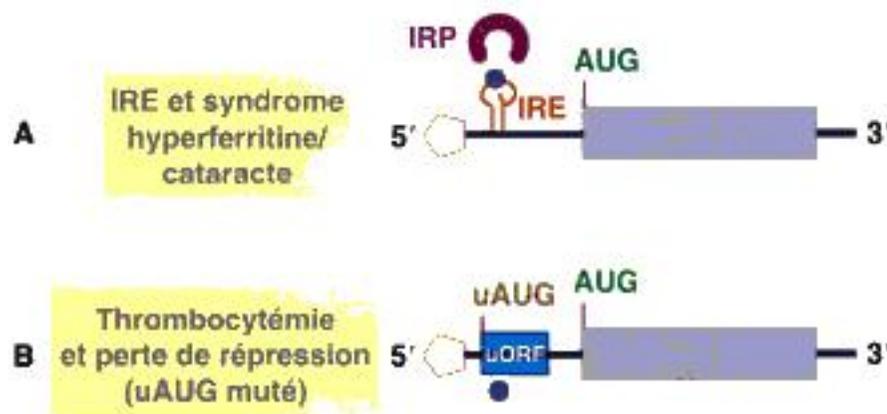


Figure 2.21 Mutations de la région 5' non codante. Exemples de mutations pouvant aboutir à des modifications d'expression de gènes. Voir le texte du paragraphe pour les explications. Le cercle correspond à une mutation.

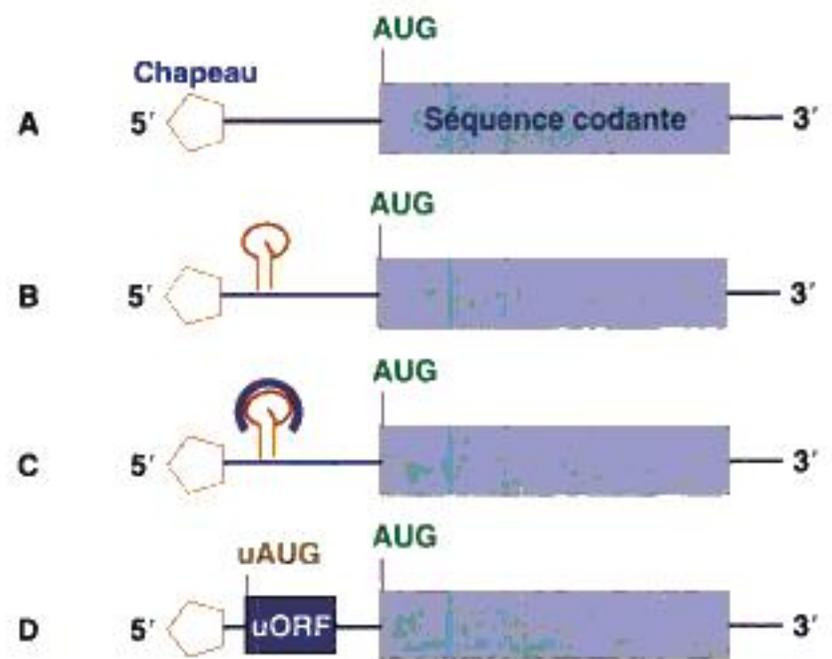


Figure 2.22 Régulation négative de la traduction. Exemples de mécanismes dans la région 5' non codante de l'ARNm. **A.** ARNm sans régulation négative. **B.** Séquence en boucle stable (*hairpin*) située entre l'extrémité 5' (chapeau) et le codon AUG. Cette boucle interfère dans l'assemblage du complexe de pré-initiation et/ou le scanning du ribosome à la recherche de l'AUG. Cette séquence en boucle peut être suffisamment stable pour empêcher l'activité de déroulement de l'hélicase et inhiber le scanning ribosomal. **C.** Sur certains ARNm, la séquence en boucle peut être stabilisée par des protéines. **D.** La partie 5'NC peut posséder des codons AUG alternatifs (uAUG) en amont du codon AUG principal ou physiologique.

du gène *dystrophia myotonica protein kinase* (DMPK, Omim 605377). Normalement, ces triplets, présents en nombre variable chez un sujet normal (environ 5 à 30), peuvent être présents jusqu'à 1000 fois chez un sujet malade. Une protéine physiologiquement se fixe sur les triplets CUG de cette partie 3'NC de l'ARNm. Parmi les hypothèses physiopathologiques, il a été montré que l'expansion importante de triplets aboutit à une modification de structure de l'ARNm, un effet négatif sur l'expression de la protéine DMPK et sur l'exportation de la protéine hors du noyau.

Dans cette même région, des mutations peuvent aussi survenir au niveau des séquences riches en 5'-AU telle que la séquence 5'-AUUUA (*AU-rich elements*, ARE). Ces séquences jouent un rôle important dans la demi-vie cytoplasmique des ARNm. Par exemple, dans certaines

formes de lymphomes, la protéine CCND1 (Omim 168461) est surexprimée. Parmi les mécanismes impliqués, la surexpression de la protéine peut être la conséquence d'une délétion de plusieurs ARE allongeant ainsi la demi-vie des ARNm en les stabilisant.

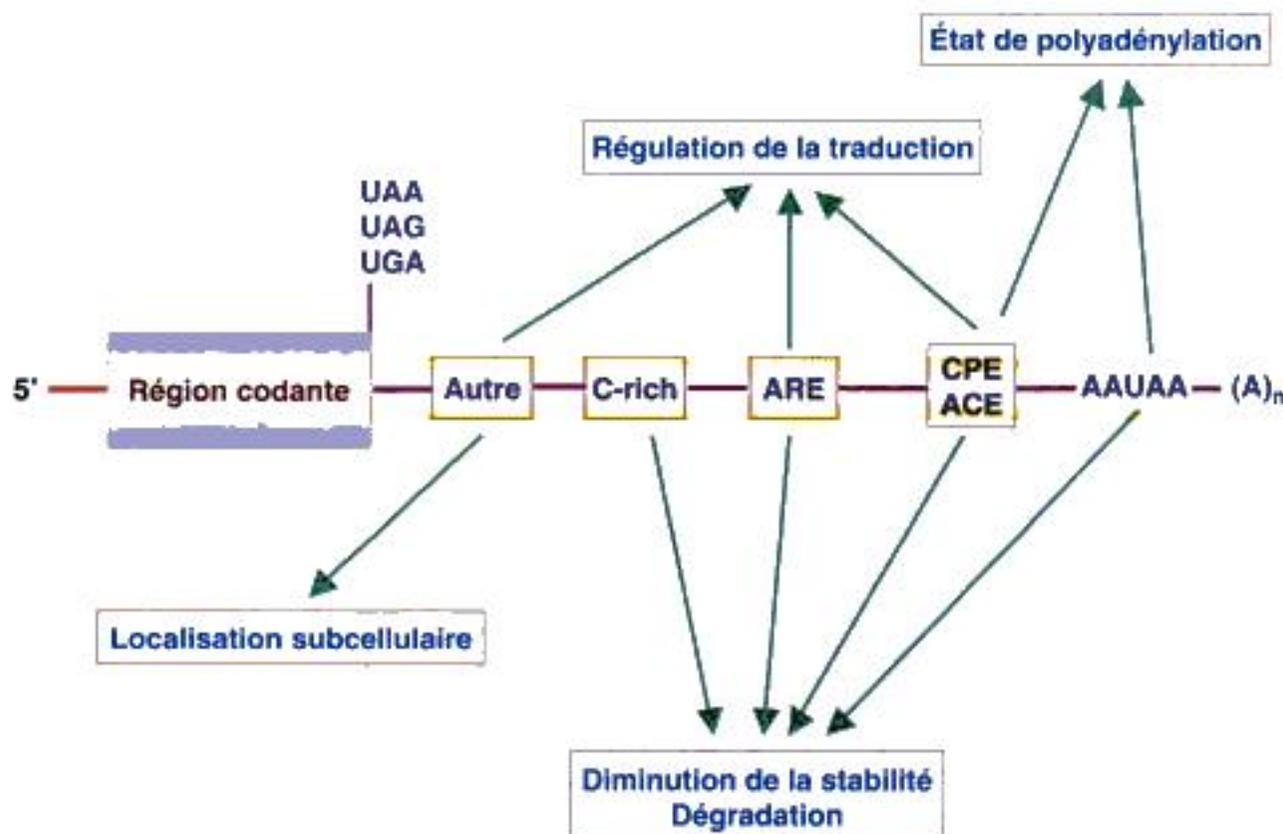


Figure 2.23 Mutations de la région 3' non codante. Voir le texte du paragraphe pour les explications.

Dans la région 3'NC de certains gènes tels que le gène alphaglobine dont l'ARNm est très stable, une mutation du codon-stop UAA en CAA aboutit à un ARNm instable (demi-vie diminuée). Le complexe ribosomal continue « sa route » dans ce cas et masque des déterminants de stabilité de l'ARNm. En fait, dans certains ARNm, se trouvent des zones riches en C interagissant notamment avec des protéines spécifiques (complexe ribonucléoprotidique alpha) et dont les mutations vont aboutir à une déstabilisation de l'ARNm. Dans notre exemple volontairement simplifié, ces zones sont masquées par la poursuite de la lecture du complexe ribosomal aboutissant à la diminution de la demi-vie de l'ARNm.

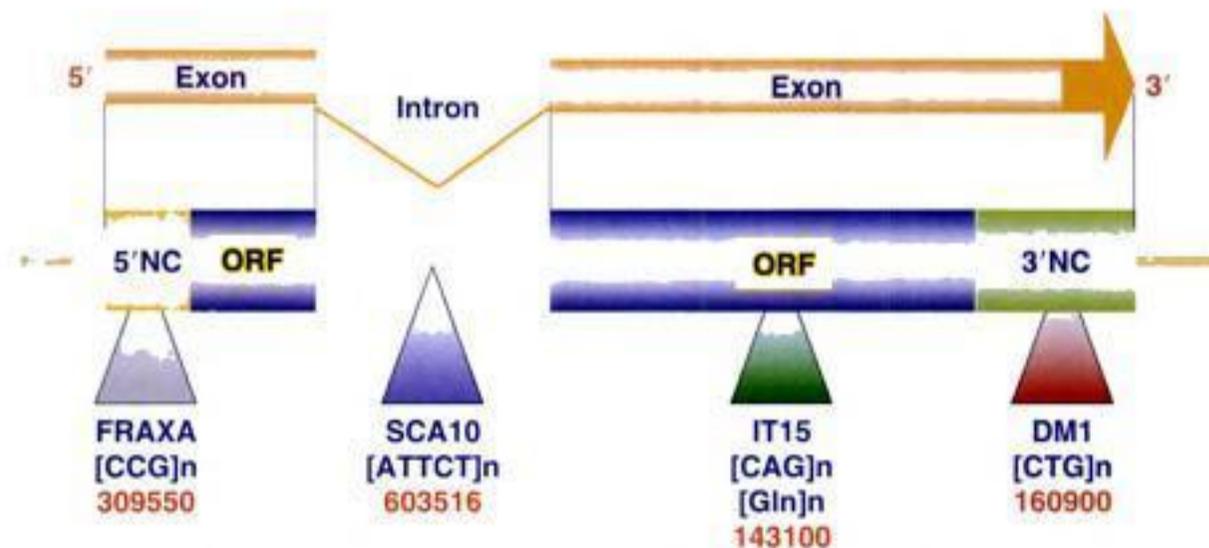
### Mutations de réplication ou dynamiques

La description précise de ces mutations sort du cadre de cet ouvrage. Nous décrivons quelques principes généraux. Les mutations dynamiques sont en général des erreurs de réplication mitotique ou préméiotique : elles résultent d'un dérapage (*slipped strand mispairing*) dans certaines zones instables du génome. Elles provoquent une amplification (mutation dynamique) ou une réduction des zones répétées. La modification observée se voit sur plusieurs générations. Le plus souvent, elle augmente au cours des générations successives. On parle alors de phénomène d'anticipation. Ce mécanisme est retrouvé dans des maladies génétiques au cours desquelles le phénomène d'anticipation observé est le résultat de l'instabilité de la répétition de trinuécléotides ou de dinuécléotides (figure 2.24). Si nous prenons l'exemple des trinuécléotides, en général, le nombre de répétitions de trinuécléotides est très variable dans le génome d'un sujet à l'autre (polymorphisme multi-allélique). L'existence de la variabilité de séquences répétées est intéressante. En effet, elle per-

met d'utiliser ces trinuécléotides comme marqueur génétique. Cependant, dans certains cas, quand leur nombre dépasse un seuil (un certain nombre de répétitions) à un locus donné, elles peuvent induire une pathologie. Ces mutations dynamiques sont responsables de l'expansion de séquences courtes, par exemple (CGG)<sub>n</sub> dans la maladie de l'X fragile (> 50 répétitions, Omim 309550). Dans un premier temps, dans une génération, une prémutation est observée. La « prémutation » contient un nombre de répétitions supérieur à la normale. Au cours de la génération suivante, la prémutation augmente de taille pour devenir une mutation. Ces répétitions présentent donc la particularité de changer de taille (essentiellement augmentation) au cours de la transmission des allèles des parents aux enfants. Ces mutations sont donc « dynamiques ». L'aspect « dynamique » de ces mutations explique en grande partie la variabilité du phénotype, de la pénétrance et de la gravité des maladies génétiques dans lesquelles ce phénomène est observé.

### Mutations par insertion d'éléments mobiles

Des insertions de type transposon, séquence LINE (Omim 151626) ou séquence Alu, peuvent se produire dans un gène. Cette insertion se fait en général par un mécanisme de rétrotransposition. Un certain nombre de pathologies ont été décrites par insertion de transposon. À titre d'exemple, on peut citer l'insertion d'un LINE dans l'exon 14 du facteur VIII, responsable d'hémophilie A (Omim 306700) ou l'insertion de séquences Alu dans la neurofibromatose de type 1 (maladie de Von Recklinghausen, de transmission autosomique dominante, Omim 162200). Par ailleurs, un grand nombre de LINE sont polymorphiques dans la population générale et donc sans conséquence apparente. Pour certains



**Figure 2.24** Exemples de localisation d'expansion de séquences répétées dans le génome. Le nombre indiqué en dessous de chaque abréviation de gène est le numéro Omim. 5'NC : région 5' non codante ; 3'NC : région 3' non codante ; ORF (*open reading frame*) : partie codante d'un gène ; FRAXA (*Fragile site mental retardation 1*) : syndrome de l'X fragile (le gène muté est FMR1) ; IT15 (appelé aussi huntingtine) : gène muté dans la chorée de Huntington ; SCA10 (spinocerebellar ataxia) : gène muté dans l'ataxie spinocérébelleuse ; DM1 (*dystrophia myotonica 1*) : gène muté dans la maladie de Steinert (dystrophie myotonique).

auteurs, cette hétérogénéité allélique pourrait jouer un rôle dans leur potentiel mutagène.

### Épimutations – Pathologie de l'épigénétique

L'épigénétique est un phénomène physiologique fondamental dans la cellule (par exemple, la méthylation au niveau des îlots CpG de certains promoteurs) et intervient dans de nombreuses régulations. De manière générale, au niveau des promoteurs notamment, l'hyperméthylation est associée à la répression de l'expression des gènes et participe à la régulation de l'expression de gène. Néanmoins, dans certains cas, des anomalies de méthylation peuvent être observées. On parle alors d'épimutations. Ces épimutations sont observées notamment dans un grand nombre de cancers (domaine de l'épigénétique) où elles modifient l'expression des gènes. Par exemple, quand une méthylation anormale d'ADN se produit au niveau d'un gène suppresseur de tumeur (anti-oncogène), cette épimutation participera au processus d'oncogénèse, l'anti-oncogène étant réprimé et ne pouvant plus assurer sa fonction physiologique. En cancérologie par exemple, de nombreux gènes peuvent être réprimés par une épimutation et provoquer le développement progressif de la cellule saine vers une cellule cancéreuse. C'est le cas par exemple du gène codant pour MLH1 (*Mut L Escherichia coli homolog of 1*, Omim 120436), gène impliqué dans le complexe de réparation de misappariements au cours de la réplication. Ce gène est hyperméthylé (et réprimé) dans un certain nombre de cancers sporadiques du côlon présentant une instabilité de microsatellites. Par contraste, une absence/diminution de méthylation normale peut aussi aboutir à une dysrégulation de l'expression des gènes. Les méthylation d'îlots CpG peuvent donc aboutir selon les cas à une surexpression ou une sous-expression d'un ou de plusieurs gènes. Certaines hypothèses ont été proposées pour expliquer ces mécanismes comme l'activation anormale de certaines méthylases, des erreurs de méthylation, des erreurs de réparation de méthylation ou encore des anomalies de remodelage chromatinien (par modification du code histone). Les mécanismes régulant la méthylation ne sont pas encore bien connus.

#### Relation mutation/épimutation

Les épimutations sont aussi observées dans des maladies héréditaires et sont la conséquence de mutations sur des gènes impliqués dans la méthylation. C'est le cas par exemple d'un syndrome rare de transmission autosomique récessif, le syndrome *immunodeficiency, centromeric instability and facial anomalies* (ICF, Omim 242860) caractérisé par un immunodéficit général associé à des anomalies du développement. Schématiquement, dans les

lymphocytes, des anomalies chromosomiques sont observées (hétérochromatine péricentromérique allongée sur le chromosome 1 ou 16 au niveau de régions riches en satellites alpha et en séquences répétées). Ces régions sont normalement fortement méthylées. Or, dans le syndrome ICF, elles sont hypométhylées. Ces anomalies de méthylation sont la conséquence d'une mutation sur le gène codant une ADN méthylase, DNMT3B. D'autres pathologies, conséquences de mutations sur des gènes impliqués dans le remodelage chromatinien (comme les histones acétyltransférases, les ATPases remodelant les nucléosomes par exemple), ont démontré le rôle fondamental de la chromatine et les conséquences pathologiques de sa dysrégulation (cancers, malformations, retards mentaux par exemple). Cet exemple illustre la forte inter-relation entre l'épigénétique et la génétique (figure 2.25).

### Anomalies de l'empreinte génomique

Certaines anomalies de méthylation provoquent un dérèglement de l'empreinte génomique. C'est le cas par exemple dans le syndrome de Beckwith-Wiedemann (BWS, Omim 130650) responsable d'anomalies du développement et de risque élevé de tumeurs dans l'enfance. L'expression bi-allélique du gène *insulin-like growth factor 2* (IGF2) est une des anomalies responsables de ce syndrome. Le gène codant pour IGF2 est normalement soumis à empreinte génomique et seul l'allèle d'origine paternel est exprimé (par méthylation d'une zone proche, la région du contrôle d'empreinte, *imprinting control region* [ICR]). Dans certains cas de BWS, la région ICR est méthylée sur les deux allèles au niveau de l'ICR provoquant alors une surexpression du gène codant pour IGF2. Ce descriptif est évidemment simplifié, la réalité physiopathologique étant plus complexe. Cet exemple illustre l'importance de maintenir une méthylation différentielle pour certains gènes.

### Répression de l'expression des gènes – Autres mécanismes

#### Répression par certains ARN

Des mécanismes différents des mécanismes d'épigénétique et aboutissant à la répression de l'expression des

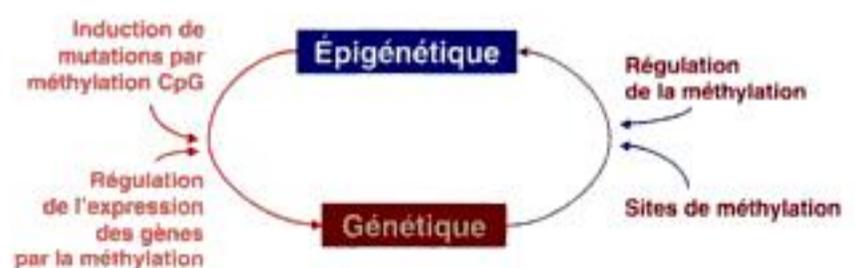


Figure 2.25 Inter-relation génétique/épigénétique.

gènes (*gene silencing*) ont été mis en évidence d'abord dans d'autres organismes avant d'être identifiés chez l'être humain. Nous les citerons brièvement. Ces mécanismes ont été observés d'abord dans la résistance des plantes à un certain nombre de virus. Il s'agit essentiellement de répression d'expression d'un gène par de l'ARN (*RNA silencing*). On distingue :

- L'inhibition de la transcription par l'intermédiaire d'ARN double brin ciblant le promoteur du gène. La séquence reconnue subit alors une méthylation de novo ce qui provoque un blocage de la transcription ;
- la répression post-transcriptionnelle des ARN (interférence d'ARN, pour les siARN par exemple et les mi-ARN (mi ARN)) – voir chapitres 24 et 25. Elle se fait par l'intermédiaire d'un ARN double brin et après un processus complexe aboutit à la dégradation de l'ARN cible. Bien que décrite dans le cytoplasme, il est probable que certains événements mal connus soient aussi présents dans le noyau. Certains auteurs ont en effet décrit dans certains cas une méthylation de novo des gènes réprimés dans le noyau concomitamment à l'action dans le cytoplasme.

L'action d'ARN dans la régulation de gènes peut être responsable de méthylation normale/anormale de gènes. On peut donc considérer que ce mécanisme est un phénomène épigénétique. Et, bien qu'à ce jour, aucune maladie présentant un tel mécanisme n'ait été démontrée chez l'homme, on peut supposer que de tels méca-

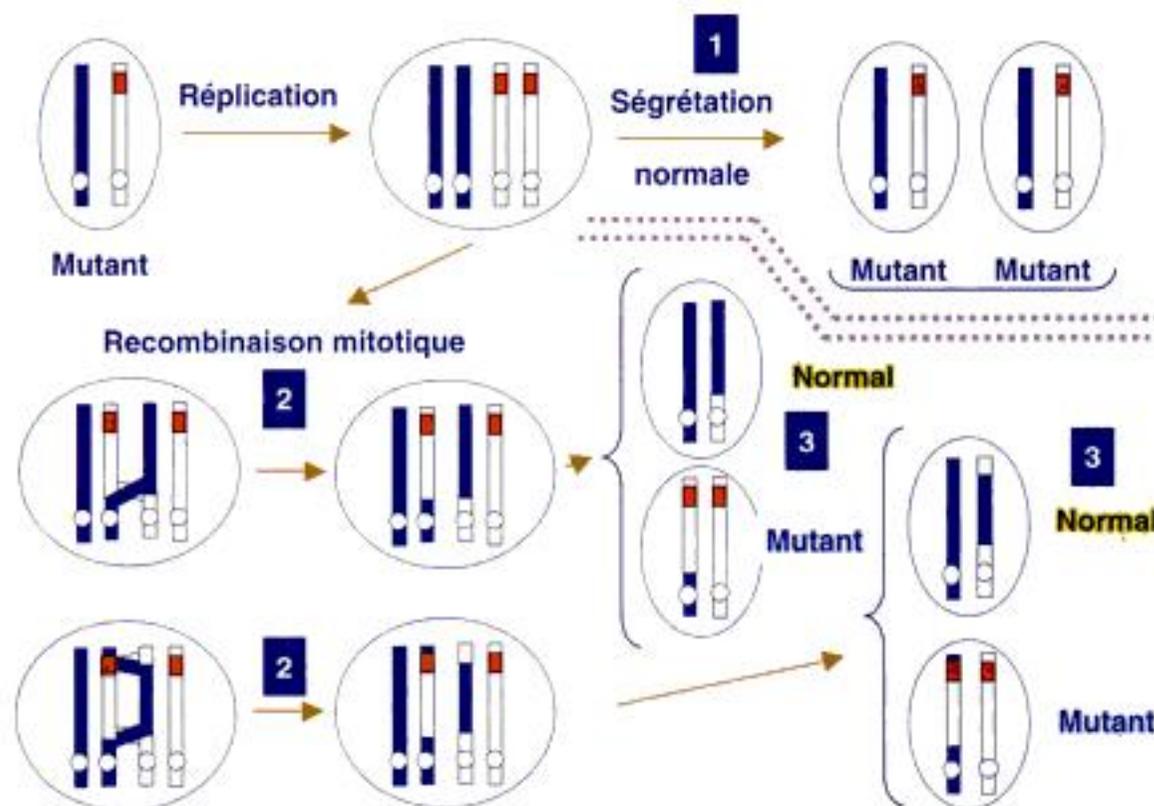
nismes physiopathologiques pourraient un jour être observés.

### Édition (*editing*) de l'ARN

Les modifications de l'ARN par *editing* (A→I→G et C→U) sont des phénomènes physiologiques rares (par exemple, pour l'ARN transcrit du gène codant pour l'apolipoprotéine B, Omim 107730). Sans avoir été confirmée à ce jour, l'hypothèse d'une anomalie de l'édition a été proposée dans certains cas (par exemple, le lupus érythémateux disséminé, Omim 152700). Elle pourrait être la conséquence de l'augmentation d'expression d'une déaminase spécifique de l'édition.

### Mutations de mutations ou mutations de réversion

Les maladies héréditaires dans leur grande majorité sont la conséquence de mutation(s) dans certains gènes indispensables au fonctionnement normal d'une enzyme, d'un récepteur cellulaire, d'un facteur de transcription ou de protéine d'autre fonction. En règle générale, ces mutations sont stables et transmises à la descendance. Néanmoins, dans de rares cas, il a pu être montré que ces mutations pouvaient être réversibles et donner ainsi lieu à une véritable thérapie génique naturelle (figure 2.26). Ces mutations sont dites de réversion.



**Figure 2.26** Exemple de mécanisme de mutation inverse. Prenons l'exemple d'une mutation hétérozygote (par exemple, une délétion ; carré rouge) dans une maladie de transmission autosomique dominante. L'allèle du père est en bleu et celui de la mère en jaune. 1. Au cours d'une ségrégation normale, la mutation est transmise aux cellules filles. 2 et 3. En cas de recombinaison (*crossing-over*) au cours de la mitose (réversion post-zygotique), on peut obtenir des révertants normaux au sein des cellules pathologiques (mosaïcisme révertant).

Le mutant de réversion devient un révertant caractérisé par une correction totale ou partielle du phénotype normal au moyen d'un mécanisme génétique ou non. Ce phénomène a été suspecté dans certaines maladies graves (comme le déficit en ADA [adénosine déaminase, Omim 102700] où, dans quelques cas, le phénotype observé restait étonnamment modéré alors que cette maladie est habituellement rapidement mortelle dans les deux premières années de la vie.

Ces mutations de réversion sont de deux catégories :

- les révertants par mutation inverse (mutation de mutation) de la mutation originelle. Ce peut être une mutation restaurant la séquence d'ADN normale ou aboutissant au codon normal (un codon pouvant avoir plusieurs triplets possibles). On observe alors la transformation d'une mutation aboutissant à la restauration de la séquence normale, par exemple, mutation sur un site cryptique d'épissage abolissant ce dernier et permettant le fonctionnement normal du site d'épissage physiologique ;
- les révertants par mutation sur un autre endroit du gène causal (deuxième site). Elle ne modifie donc pas la mutation originelle toujours présente mais permet une restauration totale ou le plus souvent partielle de la protéine défectueuse (phénomène de compensation).

Ces phénomènes aboutissent à la coexistence de deux populations cellulaires (mosaïque), la pathologique et la « corrigée ». Ainsi, si la mutation apparaît au cours d'une mitose et induit chez la cellule en réversion un avantage sélectif de croissance, on observera progressivement au remplacement des cellules pathologiques par celles contenant les révertants (cellules ayant subi une mutation de réversion). Cet événement exceptionnel peut alors modifier le phénotype de la maladie qui passe d'une maladie sévère à une maladie au phénotype modéré. Cela a par exemple été rapporté dans le déficit immunitaire combiné sévère lié à l'X (SCID, Omim 300400).

Si cette mutation de réversion apparaît dans une cellule germinale, elle peut être transmise à la descendance.

Les mécanismes expliquant ces réversions sont encore mal connus. Dans certains cas, une recombinaison intragénique par *crossing over* (décrit par exemple dans le syndrome de Bloom, Omim 210900), un glissement de l'ADN polymérase (erreur de recopiage de séquences répétées), une mutation ponctuelle ou une conversion génique ont pu être démontrés.

## Généralités sur les effets des mutations

### Déterminants pouvant expliquer les conséquences phénotypiques d'une mutation

Il est difficile de répondre à une telle question avec précision, de nombreuses études ayant montré qu'il n'existait

pas de règle générale. On peut néanmoins donner quelques orientations relatives.

### Type de mutation en cause

Une mutation est le plus souvent responsable d'une anomalie de synthèse d'une protéine (diminution/absence), d'une modification structurale (protéine instable), d'une anomalie de ciblage (membrane, organelle) ou d'une protéine non (peu) fonctionnelle. Les mutations en cause sont essentiellement des mutations non-sens, faux-sens, celles induisant des décalages du cadre de lecture (*frameshift*), une anomalie d'épissage (*splicing defect*), une délétion ou une insertion de grande taille. Si la mutation induit des modifications mineures au niveau de la protéine, on peut s'attendre à des conséquences modérées. Cela peut se voir dans certaines mutations ponctuelles (faux-sens, voire des petites délétions ou insertions).

### Mécanisme moléculaire

Outre ce qui a été décrit au paragraphe précédent, une mutation peut aussi induire une modification de la chromatine (par exemple, mutation sur un gène codant pour une protéine du noyau) en modifiant la conformation spatiale de la membrane nucléaire et/ou du noyau. Cette modification de chromatine provoquera des changements d'expression d'un certain nombre de gènes.

### Position de la mutation sur le gène

Certaines parties du gène contiennent des régions critiques pour la structure et/ou la fonction d'un gène. Une mutation dans ces régions aura des conséquences fonctionnelles importantes.

Par ailleurs, à travers les espèces, certaines régions d'un gène sont très conservées traduisant leur importance pour la protéine correspondante. Une mutation dans ces zones conservées aura probablement des conséquences négatives pour la protéine.

Dans le cas d'isoformes de protéine, selon la localisation de la mutation, toutes les isoformes d'une enzyme par exemple peuvent être sous-exprimés ou absentes si la mutation est présente sur un exon commun aux isoformes, alors qu'une mutation sur un/des exon(s) spécifique(s) d'une isoforme n'altérera que l'expression d'une seule des isoformes possibles (isoforme tissu-spécifique par exemple). Par conséquent, en fonction de la prédominance de telle ou telle isoforme dans un tissu, les conséquences peuvent être variables.

### Quantité de protéines fonctionnelles

Certaines mutations provoqueront une diminution de la quantité de protéine fonctionnelle active. À partir d'un

seuil critique, la quantité de protéine synthétisée sera probablement insuffisante.

### Modulateurs intragéniques

L'effet phénotypique d'une mutation peut être modulé par l'existence d'un variant allélique (polymorphisme fonctionnel). Par exemple, dans la pathologie liée au CFTR (Omim 602421), la mutation R117H associée au variant allélique 5T (séquence poly-[T] localisée au niveau du site accepteur de l'intron 8) est associée à une mucoviscidose sans insuffisance pancréatique alors que la même mutation associée au variant 7T est responsable d'une infertilité masculine uniquement.

### Conséquence d'une deuxième mutation

En reprenant l'exemple de la mucoviscidose (Omim 219700) de transmission récessive autosomique, selon le type d'association entre les deux mutations présentes sur chaque allèle, la conséquence phénotypique varie d'une forme sévère (avec insuffisance pancréatique par exemple) à une forme modérée (tableau 2.5).

Enfin, selon le génotype causal (l'association des mutations), les conséquences pourront être différentes sur les organes. Dans le cas de la mucoviscidose par exemple, certaines mutations sont associées à une atteinte pancréatique, d'autres ont un retentissement pulmonaire sans atteinte pancréatique.

### Interaction polymorphisme et mutation

Certains polymorphismes peuvent agir en synergie avec des mutations et aboutir à une pathologie. Par exemple, dans l'hyperoxalurie primitive de type 1 (*primary hyperoxaluria type 1*, PH1, Omim 259900), un polymorphisme (P11L) sur le gène codant pour l'alanine : glyoxylate aminotransférase (AGXT, Omim 604285) agit en synergie avec une mutation (I244T). La résultante est une modification conformationnelle de la protéine aboutissant à sa précipitation et la formation d'agrégats insolubles. En l'absence du polymorphisme, la mutation ne possède aucun effet délétère. Le polymorphisme, fréquent dans la population générale, ne possède aucune conséquence phénotypique lorsqu'il est isolé.

### Mutation et fonction d'un gène

Les conséquences d'une mutation sur la protéine codée par le gène délétère sont le plus souvent une diminution, voire une absence totale de fonction ou une fonction anormale. On dit que la mutation est responsable d'une perte de fonction. Dans certains cas, cependant, une mutation peut provoquer une conséquence inverse. On

parle de gain de fonction. Les conséquences d'un gain de fonction peuvent être :

- une activité augmentée de la fonction normale de la protéine (par exemple, résistance à l'apoptose, réplication illimitée, perte de régulation par certaines protéines) ;
- l'acquisition d'une fonction toxique ;
- l'acquisition d'une nouvelle fonction différente de celle de la protéine normale.

Par exemple, le gène SOD1 (superoxyde dismutase 1, Omim 147450), dont le produit naturel possède une activité anti-oxydante, présente une activité pro-oxydante responsable d'une forme héréditaire de sclérose latérale amyotrophique (Omim 105400) lorsqu'il est muté.

En fait, la conséquence d'une mutation peut être variable selon sa localisation dans le gène. Elle peut par exemple provoquer une modification de la structure primaire ou tertiaire de la protéine correspondante, modifier les liaisons avec des ligands et/ou des récepteurs spécifiques, changer/moduler le site actif d'une enzyme. Les conséquences phénotypiques des mutations ont donc des conséquences importantes. La détermination des mutations permet alors d'aider au suivi et/ou diagnostic des patients.

Ainsi, les mutations sur l'exon 11 du gène codant pour BRCA2 (Omim 600185) et impliqué dans les cancers du sein héréditaires (Omim 114480) sont associées au cancer de l'ovaire. Cette région a été dénommée par les Anglo-Saxons OCCR (*ovarian cancer cluster region*).

Les conséquences sont donc variables. De manière générale, on distingue :

- aucune conséquence. Il s'agit d'une mutation sans conséquences (variant allélique). Elle correspond à la définition de base du polymorphisme non fonctionnel ;
- la mutation aboutissant à une perte de fonction (*loss of function*) par diminution/abolition de l'activité fonctionnelle d'une protéine et/ou de sa quantité. Par exemple, un déficit en glucose-6-phosphate déshydrogénase (GP6PD, Omim 305900) dont le gène est situé sur le chromosome X diminue l'activité enzymatique et provoque chez le garçon une anémie hémolytique lorsque ce dernier est exposé à certains médicaments ou aliments ;
- la mutation aboutissant à l'agrégation de la protéine. On parle alors de mutation conformationnelle comme dans le cas de certaines formes du déficit en alpha-1-antitrypsine (Omim 107400) provoquant l'agrégation intrahépatocyttaire d'alpha-1-antitrypsine muté (mutant Z [E342K] homozygote) et le développement possible d'une cirrhose ;
- dans d'autres pathologies, les anomalies conformationnelles de la protéine aboutissent à une dégradation de celle-ci comme, par exemple, la protéine CFTR mutée (mutation delta F508) ;

**Tableau 2.5. Conséquences possibles des mutations et relation génotype/phénotype – Exemple de mutations sur le gène CFTR (Omim 602421) responsable de la mucoviscidose (Omim 219700), maladie de transmission autosomique récessive**

Mutations (> 1000)	Exemples de mécanismes	Conséquence(s) biologique(s)	Relation génotype/phénotype
Classe I Anomalies de la synthèse protéique	Mutations non-sens (14 %) Décalage de phase de lecture (22 %) Épissage anormal (8 %) Épissage alternatif aberrant (exon 9)	Synthèse anormale de CFTR Protéine tronquée ARNm détruit Anomalies d'épissage partiel	Phénotype sévère  Phénotype corrélé au niveau d'exon 9 non épissé
Classe II Anomalies de la maturation protéique	Mutations dans le promoteur La plus fréquente : ΔF508	Diminution de la quantité de CFTR synthétisé Réduction de synthèse du CFTR Localisation anormale de la protéine	Sévère
Classe III Régulation/Ouverture du canal chlore	Mutations ponctuelles localisées dans le domaine de fixation de l'ATP	Réponse diminuée ou absente après stimulation par l'AMPc (localisation correcte)	Phénotype modéré
Classe IV Anomalies de conductance du chlore	Mutations ponctuelles localisées dans le domaine transmembranaire	Variations anormales de l'activité globale du canal chlore	Phénotype modéré
Classe V Instabilité de la protéine	Mutations non-sens avec perte des 70–98 résidus C-terminaux du CFTR Décalage du phase de lecture en C-terminal (70–98 résidus)	Protéine labile – Demi-vie diminuée	Phénotype modéré
En général	Grande variabilité y compris intrafamiliale Phénotype corrélé au taux d'épissage anormal (pour un même génotype) Absence d'insuffisance pancréatique (environ 8 % des patients) : au minimum un allèle modérément pathologique (classe IV, classe I avec épissage alternatif variable, classe V) Insuffisance pancréatique : deux allèles sévères (classes I, II et III) Atteinte pulmonaire variable même pour une même mutation		
Cas particuliers	Deux mutations en cis (sur le même chromosome) associées à une mutation sur l'autre chromosome		Phénotype variable, une mutation pouvant moduler l'effet de l'autre mutation
Polymorphismes fonctionnels	Modulation d'épissage Diminution relative de la transcription	Effet synergique avec la ou les mutations présentes	
Gènes modificateurs	Influence du terrain génétique (* fond génétique *) Gènes modificateurs à identifier		
Autres pathologies associées (exemples)	Absence bilatérale congénitale du vas deferens (CABVD)	70 % : une mutation trouvée 10 % : deux mutations trouvées Association fréquente d'une mutation à l'épissage alternatif de l'exon 9 (hétérozygote composite)	Pas de mucoviscidose
	Pancréatite	Mutation CFTR et/ou épissage alternatif de l'exon 9	

- la mutation apportant/renforçant la fonction de la protéine mutée (par exemple, effet toxique, *gain of function*). C'est le cas par exemple dans la maladie de Huntington (Omim 143100) dans laquelle la protéine huntingtine mutante (conséquence d'une expansion anormale de triplets CAG codant pour la glutamine) séquestre l'huntingtine normale et provoque sa précipitation et des effets toxiques dans les neurones. Ce sont souvent des maladies de transmission autosomique dominante. Dans certains cas, une mutation peut modifier la propriété d'une protéine et lui donner une nouvelle fonction. C'est le cas, par exemple, pour la protéine DMT1/DCT1/Nramp2 (Omim 600523), molécule impliquée dans le transport transmembranaire du  $Fe^{2+}$  au niveau de l'intestin proximal. Une mutation dans le gène codant pour cette protéine G185R responsable d'une carence en fer et d'une anémie importante chez la souris modifie chez celle-ci la fonction de cette protéine de transporteur du fer en canal calcium (mutation *gain of function*) ;
- la modulation de l'expression des gènes : 2 à 5 % du génome humain contient des gènes. Le reste du génome est constitué de séquences pouvant modifier l'expression d'un ou de plusieurs gènes. Par exemple, un polymorphisme dans le gène codant pour le collagène (COL1A1, Omim 120150) dans l'intron 1 sur un site de fixation pour le facteur de transcription Sp1 est associé à un risque accru d'ostéoporose ;
- la mutation peut diminuer un risque de maladie non génétique (comme une maladie infectieuse par exemple). Ainsi, la délétion de 32 pb dans le gène codant pour un récepteur de chémokines CCR5 (*cc chemokine receptor 5*, Omim 601373). Il s'agit d'un des récepteurs pour l'entrée du virus VIH dans la cellule cible. À l'état homozygote, cette délétion est un facteur de protection majeur contre l'infection par le VIH, les sujets hétérozygotes ayant une progression plus lente de la maladie et les sujets homozygotes une résistance relativement importante à l'infection par le virus.

### Mutation et traduction

Nous le citons pour mémoire, ce paragraphe ayant été déjà développé dans ce chapitre.

### Relation génotype-phénotype

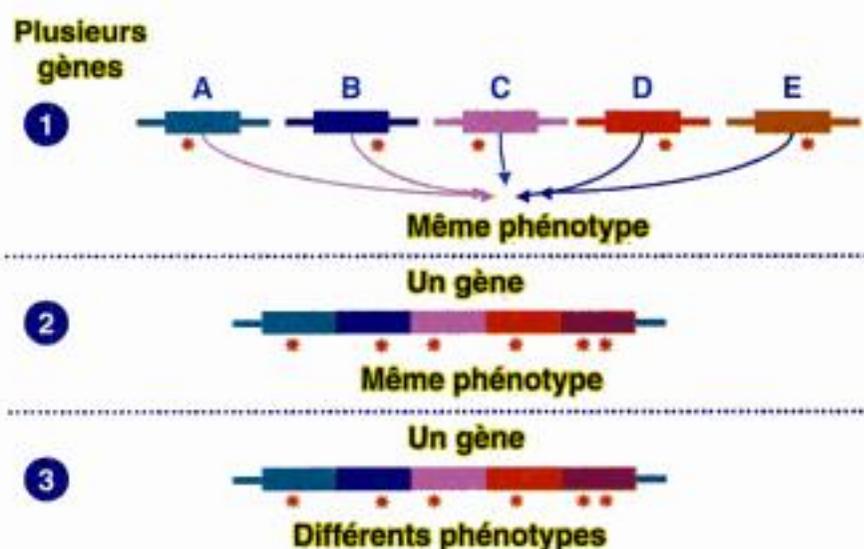
La génétique mendélienne est caractéristique des maladies monogéniques, par exemple, la mucoviscidose (Omim 219700), la maladie de Gaucher (Omim 230800, 230900, 231000) ou la maladie de Marfan (Omim 154700). Rapidement, l'étude des maladies monogéniques a montré que l'association « un génotype/un phénotype » était erronée. En effet, au début des études de génétique moléculaire, le postulat de base était

qu'à chaque génotype correspondait un phénotype. Les analyses moléculaires successives ont amené à réévaluer cette assertion et, à l'évidence, à constater que pour un grand nombre de maladies, ce postulat était inexact ou faux.

Parmi les différentes observations ayant abouti à remettre en cause et à éliminer ce postulat, on peut citer :

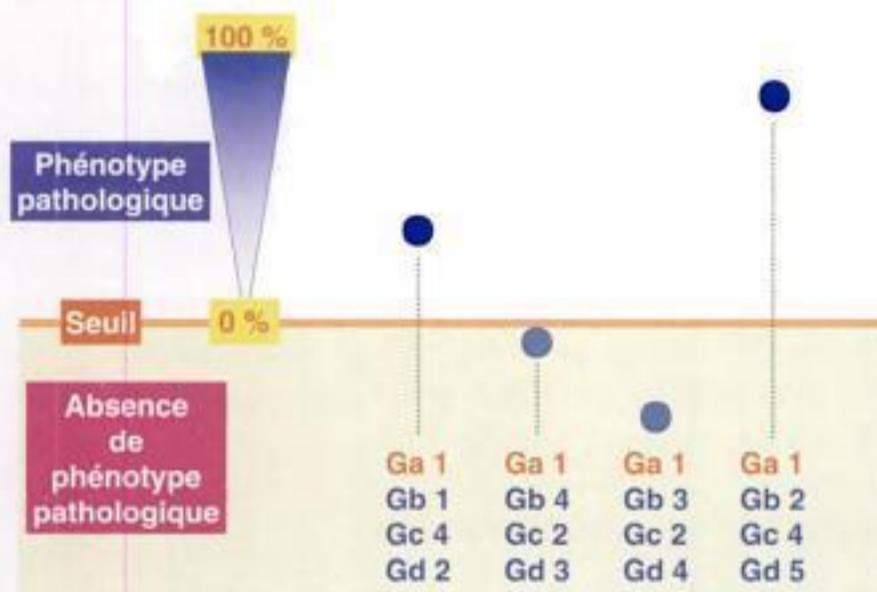
- différents génotypes sur des gènes distincts peuvent donner un phénotype différentiel pour une même maladie (hétérogénéité génétique) ;
- l'étude des mutations et de leur localisation sur le gène et leur(s) conséquence(s) sur le phénotype (hétérogénéité phénotypique ou génotypique, figure 2.27) ;
- l'étude de la relation mutations (notamment ponctuelles) et haplo-insuffisance ;
- l'étude de corrélations mutation/phénotype (notamment sur l'âge d'apparition, la sévérité, la diversité) pour une même maladie.

Une même mutation sur un même gène peut s'exprimer différemment dans une même famille et chez des sujets non apparentés (influence multifactorielle dont l'environnement, les gènes modificateurs, les modifications épigénétiques, l'interaction intragénique avec des polymorphismes en *cis* ou en *trans*, par exemple). L'analyse des génotypes et des phénotypes a montré la complexité des mécanismes génétiques. Cette relation est loin d'être toujours évidente. La plupart des maladies héréditaires sont phénotypiquement variables (expressivité variable). Pour une même maladie, entre différentes familles et au sein d'une même famille, le phénotype



**Figure 2.27** Hétérogénéités génotypique et phénotypique. Selon les gènes et les mutations, les conséquences phénotypiques peuvent être différentes. 1. Une même maladie peut être la conséquence de différents gènes : par exemple, syndrome du long QT (Omim 192500, 152427, 603830, 600163, 176261, 600681, 600919, 603796). 2. Plusieurs mutations sur un même gène sont responsables d'une même maladie (par exemple, porphyrie aiguë intermittente, Omim 176000). 3. Selon la localisation de la mutation, la maladie n'est pas la même. Par exemple, les mutations dans le proto-oncogène RET (Omim 164761). Chaque rectangle représente un gène en 1, un exon en 2 et en 3. Chaque étoile représente une mutation différente.

peut donc être différent (que ce soit sur le plan clinique, biologique, évolutif ou pronostique par exemple). En dehors des facteurs environnementaux, l'influence d'autres gènes (par exemple, gènes modificateurs ou de susceptibilité mais on devrait plutôt parler de gènes modulateurs) sur le(s) gène(s) muté(s) est évidente. La frontière entre maladies monogéniques et polygéniques est donc artificielle. En effet, on sait maintenant qu'aucun produit codé par un gène n'est vraiment limitant dans un métabolisme et que la régulation et la modulation d'un système est sous l'influence de plus d'un produit codé par un gène (il s'agit plus ou moins d'une redondance de nombreux mécanismes physiologiques jouant un rôle protecteur sur le métabolisme normal de la cellule et assurant une résistance relative aux agressions de toute forme). Les nombreuses interactions gène-gène(s) et gène-environnement commencent à être décryptées. Certains auteurs ont proposé la théorie de l'effet seuil déjà évoquée. Schématiquement, cette théorie propose qu'un phénotype se déclare si la combinaison d'un certain nombre de génotypes s'accumule chez un sujet donné (figure 2.28). En fait, l'étude de l'interaction de nombreux génotypes est encore à ses débuts. L'essor de la génomique et de la protéomique permettent progressivement d'élucider ces interrelations complexes. La bioinformatique apporte une aide majeure et les modèles mathématiques commencent à être développés pour



**Figure 2.28** Effet seuil. Des groupes de génotypes modulent l'expression phénotypique d'une maladie héréditaire. Dans certains cas, la maladie se déclare (au-delà d'un seuil) résultant de la combinaison d'un certain nombre de génotypes (au moins deux) chez un individu. Supposons une maladie héréditaire monogénique résultant d'une mutation sur un gène et dont le génotype est Ga1 (1 représentant la mutation causale et Ga, le gène responsable). Un certain nombre de gènes modulent l'expression de Ga (dénommés Gb, Gc, Gd). En fait, cette modulation est effective pour certains génotypes. La conséquence de l'action de ces génotypes (représentés sous forme d'haplotypes) est schématiquement représentée sur la figure. Les phénotypes observés sont alors fonction des haplotypes de chaque individu. Le phénotype est donc une combinatoire résultant de l'action/modulation de plusieurs génotypes. Dans le cas présent, les effets environnementaux n'ont pas été représentés.

comprendre l'ensemble de ces mécanismes et élaborer des modèles de métabolome, de physiome (voir définitions dans le glossaire).

### Un gène - Plusieurs maladies

Selon la localisation de la/les mutation(s) sur le gène, le phénotype observé est différent. Parmi les nombreux exemples, on peut citer :

- certaines myopathies comme celles liées aux mutations sur le gène codant pour la dystrophine (Omim 300377). Selon la mutation en cause sur ce gène, on observe de manière schématique deux types de dystrophies musculaires, la myopathie de Duchenne (Omim 310200) de pronostic sévère et la maladie de Becker (Omim 300376), d'apparition plus tardive, de meilleur pronostic et cliniquement moins invalidante ;
- le gène *rearranged during transfection* (RET, Omim 164761) est impliqué dans la maladie de Hirschsprung (ou mégacolon congénital, Omim 142623), le cancer médullaire de la thyroïde et les cancers endocrines multiples de type 2 (*multiple endocrine neoplasia*, MEN 2, Omim 171400). Dans le cas du gène RET, selon la localisation de la mutation dans le gène (plus exactement au niveau du domaine fonctionnel de la protéine traduite), l'une ou l'autre maladie se déclarera (Hirschsprung ou MEN2) ;
- dans certaines cardiomyopathies, une même mutation dans le gène codant pour la troponine I (TNNI3, Omim 191044) peut être responsable soit de cardiomyopathie hypertrophique (HCM, Omim 600858), soit de cardiomyopathie restrictive (RCM, Omim 115210). Il a même été rapporté le cas d'une famille dans laquelle les deux pathologies (HCM et RCM), conséquences de la même mutation, étaient observées chez deux apparentés.

### Plusieurs gènes - Une seule maladie

La grande hétérogénéité génétique a été démontrée pour de nombreuses pathologies, c'est le cas par exemple de la maladie de Hirschsprung (mégacolon congénital, Omim 142623). Parmi les gènes impliqués, on peut citer le gène RET, le gène *endothelin receptor type B gene* (EDNRB), le gène endothéline 3 (ED3), le *glial cell line-derived neurotrophic factor* (GDNF) et l'*endothelin converting enzyme-1* (ECE1). Par ailleurs, comme cela a pu être démontré dans cette maladie, des polymorphismes (par exemple dans le gène RET) et même certains haplotypes peuvent aussi moduler le phénotype de la maladie.

### Pathologie mitochondriale

La génétique mitochondriale est complexe. Nous donnons un aperçu simplifié de la relation génotype/phénotype afin de montrer la place de la biologie moléculaire

dans le diagnostic et la compréhension de ce groupe de pathologies. De manière générale, le phénotype des maladies mitochondriales résultent de l'interaction entre le niveau d'hétéroplasmie (ration ADN mitochondrial muté/normal) et la distribution des génomes mutants dans l'espace (dans les mitochondries, les cellules, les tissus et les organes). Il existe donc une grande diversité théorique. Un exemple caractéristique de cette diversité est la spécificité de tissu. Cette dernière est le plus souvent expliquée par l'effet seuil. Le cœur, le cerveau et les muscles sont les organes dépendant le plus de la phosphorylation oxydative. Ils seront donc plus rapidement atteints et le phénotype s'exprimera de manière prédominante dans ces organes. D'autres explications ont aussi été avancées. Par exemple, la quantité totale de mutation dans un tissu peut être moins importante que sa concentration locale et sa distribution cellulaire. Ainsi, la distribution topographique des mutations dans différents tissus peut-elle être expliquée par une expansion clonale d'une mutation sporadique (par exemple, une délétion) survenue tôt au cours de l'oogenèse ou de l'embryogenèse. La réplication de l'ADN mitochondrial ne se faisant pas avant la différenciation des couches germinales, quelques mutations sporadiques présentes au cours ou avant le stage blastocyte peuvent ségréger parmi les couches germinales ou les cellules filles. Il en résulte une charge mutationnelle différentielle importante entre les différents tissus. Un exemple de ces différences est observé par exemple dans l'ophtalmoplégie externe progressive (Omim 157640), forme de myopathie, et le syndrome de Kearns-Sayre (Omim 530000), maladie polysystémique. Ces deux maladies sont la conséquence d'une même délétion importante dans l'ADN mitochondrial (l'expansion clonale d'une telle mutation par rapport à l'ADN normal n'est pas clairement expliquée actuellement).

Un autre exemple est la distribution topographique différentielle des mutations dans un même tissu, comme par exemple dans le cas de deux syndromes, MELAS (Omim 540000) et l'ophtalmoplégie externe progressive (PEO, Omim 157640). Ces deux pathologies peuvent être la conséquence d'une mutation sur l'ARNt-Leu<sup>(UUR)</sup>. Chez les sujets atteints du syndrome PEO, 60 % des muscles possèdent la mutation et ceux-ci développent une myopathie et une ophtalmoplégie, alors que chez les sujets atteints du syndrome MELAS, 80 % des muscles possèdent la mutation. L'analyse fine des muscles montrent que la plupart des fibres musculaires dans les deux pathologies possèdent environ 50 % de mutations mais que les cellules *ragged red fibers* (RFF, marqueur de prolifération mitochondrial anormal dans les muscles) contiennent 95 % de mutations chacune dans le MELAS alors qu'elles en contiennent 95 à 98 % dans le PEO. Ces quelques pourcentages de différences suffisent à expliquer la différence de phénotype.

## Transfert d'ADN noyau/mitochondrie – Pathologie exceptionnelle

Dans l'ADN génomique se trouvent de nombreux fragments d'ADN mitochondrial ayant migré depuis la mitochondrie durant l'évolution de l'être humain. Il y a eu transfert de gènes de la mitochondrie vers le noyau. Un grand nombre de ces fragments sont des pseudogènes. Néanmoins, une des conséquences de ce phénomène est l'importation de protéines traduites à partir des gènes du noyau vers la mitochondrie. Le mécanisme est inconnu actuellement, même si des hypothèses ont été formulées (ARN mitochondrial ayant servi d'intermédiaire et converti en ADN avant d'être intégré dans le génome, ADN mitochondrial issu de mitochondries altérées et intégrées dans le génome nucléaire). Il a été décrit des cas exceptionnels de translocation mitochondrie/génome responsable de pathologie humaine. Ce mécanisme a été rapporté dans une maladie rare de transmission autosomique dominante, constituant un syndrome polymalformatif le syndrome de Pallister-Hall (Omim 146510).

## Mutation(s) et pathologie

### Mutation et maladies « communes »

Parmi les maladies de grande fréquence (dites encore « communes »), on retrouve aussi parfois des formes héréditaires. À titre d'exemple, on peut citer un des gènes impliqués dans les formes héréditaires de la maladie de Parkinson, l'alpha-synucléine (Omim 1633890), le gène codant pour BRCA1 (Omim 113705) impliqué dans les cancers du sein, les gènes MODY 1, 2 et 3 (*maturity-onset diabetes of the young*, respectivement Omim 125850, 125851, 600496) impliqués dans le diabète.

Une mutation dans un de ces gènes est associée à un risque important de développer une de ces pathologies. Malheureusement, les mutations décrites dans ces gènes ont une pénétrance élevée mais une prévalence faible dans la population générale.

### Mutation et mitochondries

Les mutations de l'ADN mitochondrial sont retrouvées dans un grand nombre de maladies multisystémiques ou localisées à certains organes. Plus de 200 maladies sont associées à des mutations dans les mitochondries ([www.mitomap.org](http://www.mitomap.org)). D'autres pathologies liées à des mutations dans le noyau de transmission mendélienne sont aussi associées à des mutations dans la mitochondrie. Les cellules contiennent environ 300 à 2000 mitochondries. Leur fonction est importante dans les tissus nécessitant beaucoup d'énergie comme les muscles squelettiques, le cœur, le cerveau et, à un degré moindre, le foie et le rein.

Les mutations dans la mitochondrie (délétion[s], mutation[s] ponctuelle[s]) sont le plus souvent hétéroplasmiques, c'est-à-dire qu'au sein d'une même cellule coexiste de l'ADN mitochondrial (ADNmt) normal et muté en proportion variable. Comme déjà souligné, les conséquences de cette hétéroplasmie sont variables (définie par le rapport ADNmt muté/ADNmt normal). Par exemple, de nombreuses études ont montré qu'il était nécessaire d'avoir plus de 60 % d'ADN mitochondrial muté ou délété pour observer un déficit de la chaîne respiratoire mitochondriale dans un organe spécifique. Par ailleurs, le pourcentage d'hétéroplasmie est souvent corrélé à la sévérité du phénotype et au développement de la maladie. Par exemple, la mutation T8993G peut être responsable d'une maladie létale, le syndrome de Leigh (Omim 256000) si l'hétéroplasmie est de 90 % ou du syndrome NARP (neuropathie, ataxie et rétinite pigmentaire, Omim 551500) si l'hétéroplasmie est de 60 à 80 %. Par ailleurs, certaines pathologies mitochondriales peuvent aussi être des anomalies quantitatives d'ADNmt soit une diminution (syndrome de déplétion), soit une augmentation (sur-réplication, *overreplication*). Ces différentes anomalies qualitatives et/ou quantitatives peuvent être observées spécifiquement dans un organe. Ainsi selon le(s) tissu(s) affecté(s), le phénotype observé peut être différent.

### Mutation et cancers

La tumorigenèse est une transformation complexe, multifactorielle en plusieurs étapes. Un grand nombre de modification de gènes de régulation sont impliqués dans ce processus, par exemple des mutations inactivant des

gènes suppresseurs de tumeurs (anti-oncogènes) ou des mutations activant des gènes impliqués dans la croissance (oncogènes) (tableau 2.6).

Sans rentrer dans les détails de la tumorigenèse, nous rappellerons quelques principes de base :

- l'hypothèse de Knudson : deux événements génétiques sur les allèles d'un même gène sont nécessaires pour l'inactivation complète de ce gène. Au cours du premier événement, une mutation est présente sur l'allèle d'un gène (mutation héréditaire ou acquise). Le deuxième allèle étant dominant, la cellule ne présente pas d'anomalie. Dans un deuxième temps, succède un deuxième événement acquis (mutation ou perte partielle voire totale d'un chromosome) responsable de l'inactivation des deux allèles du même gène. Ce phénomène a initialement été décrit dans le rétinoblastome (Omim 180200) ;
- la délétion partielle ou totale d'un chromosome appelée aussi perte d'hétérozygotie (*loss of heterozygosity*, LOH) (voir infra) ;
- dans certains cas, on observe une perte d'empreinte (*loss of imprinting*, LOI). Un gène normalement soumis à empreinte (un seul allèle actif) s'exprime alors sur les deux allèles (expression bi-allélique). On assiste alors à une dérégulation de l'expression de ce gène.

### Perte d'hétérozygotie

Le terme de perte d'hétérozygotie (LOH) a été défini à partir de l'hypothèse de Knudson. Supposons un allèle codant pour un gène suppresseur de tumeur présentant une mutation germinale (quelle que soit sa nature) et donc

**Tableau 2.6. Origine génétique des cancers : mutations germinales et somatiques**

	Mutation germinale	Mutation somatique et épigénétiques
Fréquence	Faible (≈ 5 %)	Importante
Mode de transmission	Héréditaire	Sporadique
Exemples	Cancer héréditaire du sein (Omim 114480)  Polypose colique familiale (Omim 175100) Cancer colorectal familial (Omim 114500) Syndrome de Van Hippel Lindau (Omim 193300)	Cancer du sein Cancer colorectal Cancer du rein
Localisation des mutations	Toutes les cellules de l'organisme	Cellules tumorales uniquement
Intérêt clinique	Prédire le risque de cancer	Détecter/suivre le cancer
Indications actuelles pour le patient	Important Recherche des porteurs de(s) mutation(s) Traitement préventif/curatif	Faible Détection précoce Intérêt pronostique Intérêt thérapeutique Adaptation thérapeutique

fonctionnellement inactif (première atteinte) et l'autre allèle normal, le sujet est donc hétérozygote. Si l'allèle normal subit dans un deuxième temps une mutation somatique (deuxième atteinte), la protéine correspondante est donc fonctionnellement inactive/absente. Il y a perte d'hétérozygotie. Dans le cas du produit d'un gène suppresseur de tumeur, son absence aura des conséquences sur le cycle cellulaire et/ou la communication cellulaire et/ou la réparation de l'ADN, par exemple, et donc une possible évolution vers une tumeur.

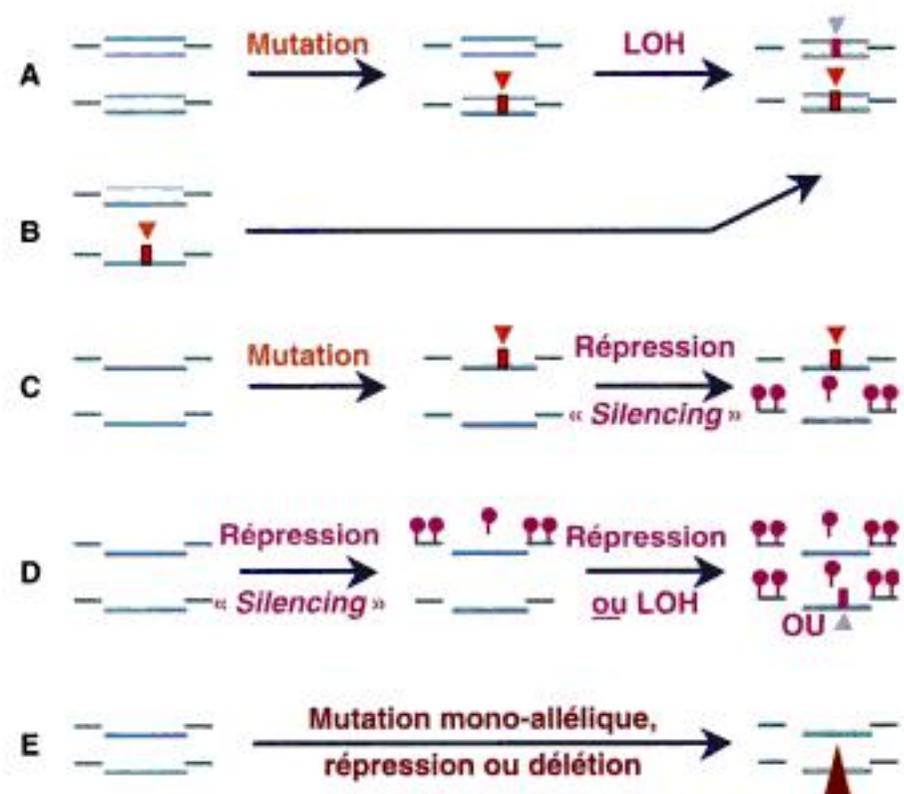
Plusieurs mécanismes peuvent expliquer la LOH (figure 2.29). Schématiquement, on peut citer :

- la délétion ;
- la recombinaison mitotique ;
- la perte chromosomique ;
- la conversion génique ;
- la mutation ponctuelle.

### Instabilité des microsatellites

L'instabilité des microsatellites (*microsatellite instability*, MIN, ou *replication error*, RER) constitue un ensemble d'altérations réparties dans une plus ou moins grande partie du génome au niveau de courtes séquences répétées (1 à 10 pb). Il s'agit d'une situation dans laquelle le nombre de séquences répétées a changé (augmentation

ou diminution). On les retrouve donc dans un grand nombre de cancers soit héréditaires comme le cancer du côlon non polyposique (HNPCC, Omim 114500), pathologie dans laquelle au moins 100 000 séquences répétées sont altérées, soit dans certains cancers sporadiques comme le cancer du côlon. Cette instabilité est la conséquence d'anomalies des complexes cellulaires engagés dans la réparation et/ou la réplication de l'ADN. Un certain nombre de gènes sont impliqués (par exemple, hMSH2 [Omim 120435], hMLH1 [Omim 120436], hMSH6 [Omim 600678], hPMS1 [Omim 600258] et hPMS2 [Omim 600259]). Le défaut de réparation de l'ADN est un phénomène impliqué dans la carcinogénèse par augmentation du taux de mutation sur l'ADN (encore appelé hypothèse du phénotype mutateur, *mutator phenotype hypothesis*). De manière générale, l'instabilité progressive liée à ce phénotype mutateur provoque un ensemble séquentiel ou non d'autres mutations et déstabilise un certain nombre de gènes (dans certains cas d'ailleurs, cette instabilité a été retrouvée dans les états précancéreux). Sans rentrer dans les détails, une cascade d'événements mutationnels apparaissent donc (de manière aléatoire ou non) et participent à la progression vers la tumeur maligne (avec d'autres événements tels que les anomalies de méthylation, les mutations dans les oncogènes et les anti-oncogènes par exemple). L'hypothèse du phénotype mutateur en cancérologie est cependant discutée par certains auteurs.



**Figure 2.29** Inactivation d'un gène suppresseur de tumeur. Exemples de mécanismes. **A et B.** Hypothèse de Knudson. En A, une mutation initiale déclenche l'inactivation du gène sur un allèle. L'autre allèle du gène est inactivé par perte d'hétérozygotie (*loss of heterozygosity*, LOH) au cours d'un deuxième événement (par exemple, délétion ou recombinaison mitotique par non-disjonction). La conséquence est donc l'absence de fonction de la protéine codée par le gène suppresseur de tumeur. **C.** Un événement mutationnel apparaît sur un allèle suivi de l'inactivation épigénétique du deuxième allèle (par exemple, par méthylation du promoteur). **D.** Répression épigénétique des deux allèles en deux étapes sans mutation ni LOH. **E.** Certains gènes suppresseurs de tumeur n'ont pas besoin de perdre une activité fonctionnelle par atteinte des deux allèles. L'atteinte d'un allèle suffit (haplo-insuffisance). Plusieurs mécanismes sont possibles tels que mutation, délétion ou répression épigénétique.

### Instabilité des chromosomes

Au cours des cancers notamment, l'instabilité chromosomique (*chromosomal instability*, CIN) est souvent importante et multiple, allant du réarrangement à la perte partielle ou totale d'un (ou de plusieurs) chromosome(s). L'aneuploïdie qui en résulte se traduit soit par une perte de gène(s) suppresseur(s) de tumeur, soit par une amplification d'oncogènes. CIN serait la conséquence d'une perte de fonction du point de vérification de la mitose (*mitotic checkpoint*, MCP) dans lequel sont impliqués plusieurs gènes comme, par exemple, *budding uninhibited by benzimidazoles 1* (hBUB1, Omim 602452).

### Méthylation anormale de l'ADN

Celle-ci a déjà été évoquée dans le cadre des modifications épigénétiques (avec la déacétylation des histones et la modification de structure de la chromatine). De manière générale, la méthylation de l'ADN est un phénomène de régulation transcriptionnelle importante d'un grand nombre de gènes. Son dérèglement aboutit à un certain nombre de pathologie, par exemple dans les cancers. De manière globale, dans les cancers, il semble y avoir une hypométhylation du génome par rapport au tissu sain. Cette hypométhylation aboutirait à la surexpression d'oncogènes. Une hyperméthylation a aussi été décrite dans un certain nombre de cancers aboutissant à l'abolition d'expression de gènes suppresseurs de tumeurs et l'apparition de mutations au niveau d'îlots CpG. Les anomalies de la méthylation sont donc fréquentes dans les cancers. Parmi les origines encore mal connues de ces troubles de la méthylation, on peut citer les produits toxiques et probablement certains régimes (déficit en acide folique par exemple). Par ailleurs, certains auteurs ont proposé un lien entre les troubles de la méthylation et la présence de MIN, les deux phénomènes étant probablement liés. Il en serait de même pour CIN. La connaissance de ce phénomène n'en est qu'à ses débuts.

### Virus

À titre de rappel, un certain nombre de virus sont impliqués dans les cancers, traduisant l'implication de nombreux processus biologiques, génétiques et environnementaux dans la genèse d'un cancer. Parmi les virus impliqués, on peut citer HTLV-I, le virus de l'hépatite B, les papillomavirus, le virus de l'hépatite C, le virus d'Epstein-Barr (EBV).

### Conséquences générales des mutations

Nous l'avons déjà évoqué dans ce chapitre de nombreuses fois. Nous le rappelons brièvement pour mémoire. Le phénotype d'une mutation peut être la résultante d'une protéine ayant acquis une nouvelle fonction (*gain of function*) à effet dominant négatif (*dominant-negative*) ou ayant perdu ses propriétés (*loss of function*)

Certaines pathologies résultant de délétions dans des zones non codantes peuvent être liées à un dysfonctionnement de la régulation de l'expression d'un gène.

On peut aussi observer :

- l'haplo-insuffisance : la perte ou l'inactivation d'un gène résulte dans la formation en quantité insuffisante de la protéine correspondante et son insuffisance pour un métabolisme normal. En général, les deux allèles sont donc indispensables pour une fonction normale ;
- à l'opposé, on peut avoir la surexpression d'un ou plusieurs gènes (par exemple, la trisomie 21) ;
- modification de la structure d'un chromosome : par exemple, duplication, délétion, cassures, translocation et inversions) ;
- un effet dominant négatif (gain « toxique ») ou une perte de fonction.

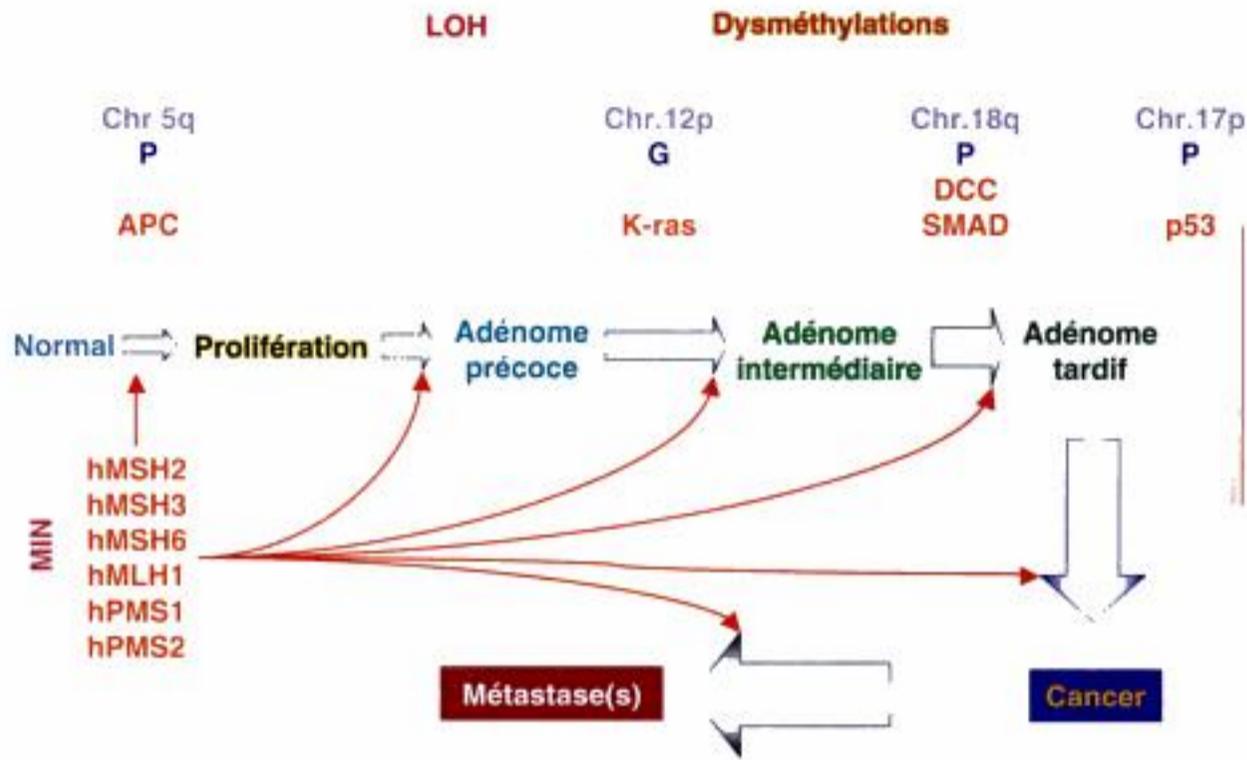
### Maladie génétique et pathogenèse moléculaire

Sans détailler les conséquences des mutations en pathologie humaine générale (hors maladies héréditaires monogéniques) qui sortent du cadre de cet ouvrage, nous donnerons un exemple simplifié des conséquences possibles d'un ensemble de mutations aboutissant à une maladie générale. Nous prendrons l'exemple du cancer colorectal sporadique. En dehors des cancers colorectaux d'origine héréditaire (par exemple, la polypose adénomateuse familiale, *familial adenomatous polyposis*, FAP, ou *adenomatous polyposis of the colon*, APC, Omim 175100), la plupart de ces cancers sont liés à l'acquisition séquentielle (au minimum) de mutations au cours de la vie, l'ensemble des mutations aboutissant au cancer proprement dit.

Ces mutations associent plusieurs types d'anomalies :

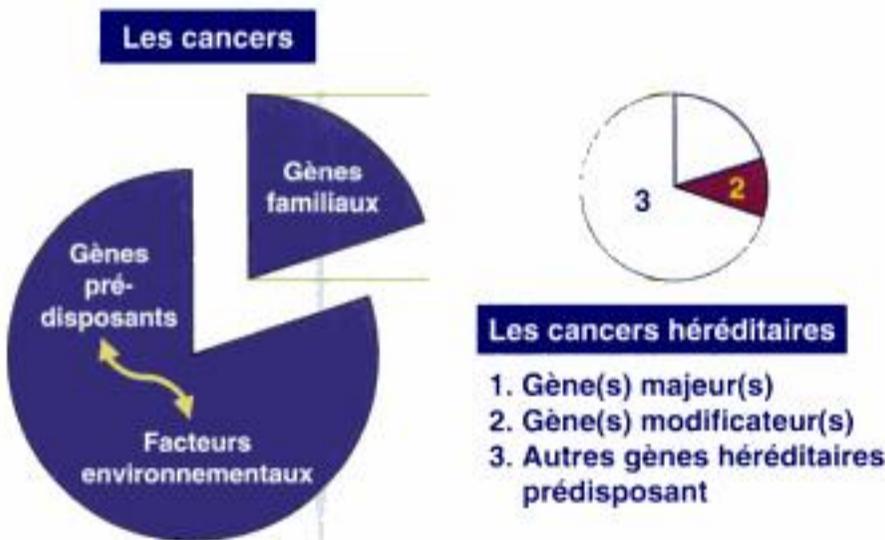
- anomalies chromosomiques ;
- mutations intragéniques ;
- altérations épigénétiques.

Elles provoquent des modifications d'expression de nombreux gènes par exemple, de proto-oncogènes, de gènes suppresseurs de tumeur et de gènes de réparation de l'ADN. Certains modèles théoriques de carcinogenèse contestés par certains ont été proposés (figures 2.30, 2.31). Ces différents événements sont bien évidemment associés à des facteurs environnementaux multiples (par exemple, régime alimentaire, agents infectieux, toxiques, tabac). Dans les cancers, il est probable qu'un grand nombre de mutations apparaissent de manière aléatoire et qu'un certain nombre d'entre elles aboutissent à un avantage sélectif de croissance. Par ailleurs, certaines mutations (par exemple, celles sur les gènes de réparation de l'ADN) peuvent augmenter le taux de mutations sur



**Figure 2.30** Exemples de modifications génétiques associées à la tumorigenèse colorectale. Cette séquence de modifications génétiques a été proposée par certains auteurs. Elle semble cependant insuffisante pour expliquer la plupart des cancers. D'après cette théorie, un certain nombre de mutations se développent et provoquent successivement un avantage sélectif de prolifération. De nombreux gènes sont impliqués. Les mutations (depuis la perte totale ou partielle d'un chromosome à la mutation ponctuelle) aboutissent à un processus progressif de prolifération puis de cancérisation des cellules. Chr : localisation chromosomique ; P : perte de fonction du gène par mutation ; G : gain de fonction du gène par mutation ; APC : *adenomatous polyposis of the colon* (Omim 175100) ; K-ras : *Kirsten rat sarcoma* (Omim 190070) ; DCC : *deleted in colorectal cancer* (Omim 120470) ; SMADs/MADH4 : *mothers against decapentaplegic homolog* (*Drosophila*), (Omim 601366, 600993, 603109) ; p53 (Omim 191170) ; les gènes de réparation de l'ADN, hMSH (*human mutS E. coli homolog 2, 3 et 6*, Omim 120435, 600887, 600678) ; hMLH1 (*human mutL E. coli homolog 1*, Omim 120436) ; hPMS1 et 2 (*postmeiotic segregation increased, S. cerevisiae, 1*, Omim 600258 ; *postmeiotic segregation increased, S. cerevisiae, 2*, Omim 600259) ; MIN : instabilité de microsatellites ; LOH : perte d'hétérozygotie. Le trait vert signifie qu'à de nombreuses étapes de cancérogenèse des dysméthylations peuvent se produire et le trait orange signifie que des LOH peuvent survenir au cours de nombreuses étapes de cancérogenèse.

d'autre gènes (phénotype mutateur précédemment décrit). Ces deux phénomènes (avantage sélectif et phénotype mutateur) sont probablement interdépendants. Ces différentes hypothèses sont actuellement sources de nombreuses discussions et l'objet d'intensives recherches.



**Figure 2.31** Héritéité, interactions géniques et environnement. Exemple des cancers. De manière générale, les cancers héréditaires représentent une part faible des cancers dans la population générale (par exemple, cancer du sein et gènes BRCA1 [Omim 113705] et BRCA2 [Omim 600185]). Des gènes modulateurs (par mutation et/ou polymorphismes) interagissent entre eux et avec l'environnement jouent une part importante dans la pénétrance et/ou l'expressivité de la maladie.

### Maladies « génomiques » et effet de dosage génique

Les maladies génomiques, conséquences de réarrangements souvent complexes, présentent une grande diversité phénotypique dépendant de plusieurs paramètres (essentiellement localisation, taille et type de recombinaison). La conséquence d'une délétion, duplication ou inversion peut être soit une modification de la structure du gène, soit une modification du « dosage génique ». Dans ce dernier cas, le réarrangement provoque en général des modifications sur une grande région génomique (en général, > 1 Mb) et comprenant plusieurs gènes et modifiant le nombre de copie de ce(s) gène(s). Les conséquences phénotypiques qui en découlent sont alors soit un « excès » du nombre de copies (duplication par exemple), soit un « déficit » du nombre de copies (haplo-insuffisance par délétion par exemple). Certaines études ont montré qu'une petite fraction de gènes était sensible à cet effet de dose (tableau 2.3).

## Mutation(s) et pathologie

Une mutation est un changement de séquence sur l'ADN (nous excluons les modifications épigénétiques de ce paragraphe). Soit elle est pathogène et induit une modification du phénotype, soit elle est sans importance et le phénotype n'est pas modifié.

Il est nécessaire de rappeler quelques notions importantes :

- après avoir éliminé les problèmes techniques, l'étude complète du gène doit s'assurer qu'il n'existe pas d'autres mutations susceptibles d'induire la pathologie (afin d'éliminer par exemple une variation de séquence en déséquilibre de liaison avec la vraie mutation) ;
- une mutation silencieuse (pas de variation d'acide aminé) peut être responsable de pathologie (saut d'exon, par exemple). L'étude de l'ARNm est donc importante ;
- une mutation peut se trouver sur un acide aminé non conservé entre les espèces et être vraiment responsable de la pathologie et, inversement, une mutation sur un acide aminé conservé à travers les espèces peut être sans effet phénotypique ;
- une mutation non-sens peut être sans effet si elle affecte le polypeptide en dehors de domaines fonctionnels importants

Les principaux arguments en faveur d'une mutation sont :

- une mutation dans la séquence codante génère un codon-stop ou un changement de phase de lecture (*frame-shift*, avec formation d'un codon-stop en aval) est probablement responsable de modification du phénotype ;
- l'analyse d'un gène doit être complète (promoteur, exons, jonctions exon/intron). Selon la méthode employée, la sensibilité et la spécificité varient. Il est important de préciser la technique employée et les zones explorées ;
- l'analyse de la famille : dans le cas d'une maladie dominante, la mutation doit ségréger (être transmise) avec le phénotype pathologique. En l'absence de mutation, le phénotype est normal. Si la mutation ségrége, la probabilité qu'il s'agisse de la mutation causale est grande. Dans le cas d'une maladie récessive, le même principe de raisonnement s'applique ;
- on considère souvent qu'un acide aminé conservé à travers les espèces possède une importance fonctionnelle importante. En cas de mutation modifiant cet acide aminé, les conséquences fonctionnelles et/ou structurales sont probables ;
- l'analyse de contrôles (sujets normaux de la population générale) est importante. La prévalence des mutations est normalement rare. Si un variant allélique a une fréquence supérieure à 1 % (par définition), il est considéré fréquent (polymorphisme). La difficulté d'interprétation en considérant ces arguments de fré-

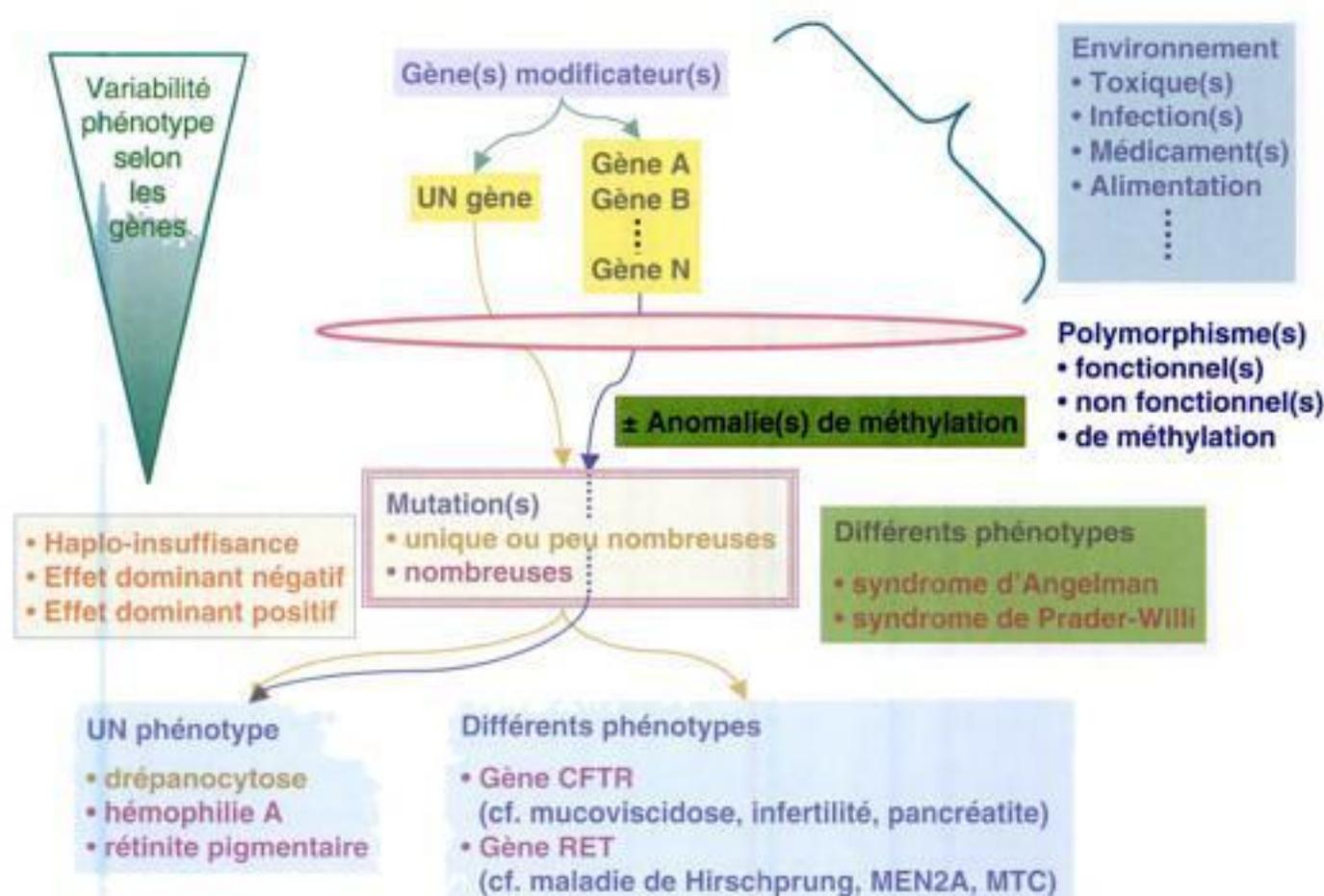
quence tient au fait qu'un polymorphisme peut avoir des conséquences fonctionnelles sur une mutation (polymorphisme fonctionnel) notamment (mais pas uniquement) dans les maladies multifactorielles. Une interaction est aussi possible entre un polymorphisme et une mutation rendant l'interprétation parfois délicate. En général, dans les études de mutations, la plupart des études analysent 100 chromosomes au minimum (50 sujets normaux) pour comparer la fréquence du variant allélique chez les sujets normaux et malades (à différencier des études épidémiologiques telles que les études de gènes candidats où le nombre de sujets contrôles et malades doit être beaucoup plus important [l'épidémiologie moléculaire sort du cadre de cet ouvrage et ne sera pas développé]) ;

- l'expression de la mutation (à partir de l'ADNc) devrait être effectuée dans un système eucaryote ou procaryote. L'activité biologique de la protéine mutée et éventuellement ses propriétés devraient être étudiées et comparées à celles de la protéine normale ;
- des logiciels ont été développés pour aider à la prédiction des conséquences fonctionnelles et/ou structurales d'une mutation sur une protéine. Ces logiciels ne sont cependant que des aides relatives et la confiance vis-à-vis de ceux-ci doit être limitée. Certaines études ont en effet montré leurs imperfections et les erreurs possibles de leur prédiction. Ils doivent donc être utilisés avec prudence et ne permettent jamais seuls de conclure sur la conséquence pathologique ou non d'une mutation. Seuls les tests fonctionnels des protéines mutées ou susceptibles de l'être peuvent conclure à la conséquence pathologique ou non d'une mutation.

En conclusion, la pathologie moléculaire est complexe et beaucoup d'inconnues (mécanismes en cause, interprétation, pénétrance...) demeurent. Le rôle des polymorphismes fonctionnels semble majeur et leur connaissance émerge progressivement. De nombreuses interactions multigéniques sont progressivement mises en évidence grâce aux outils de la génomique, de la protéomique et de la bioinformatique. La figure 2.32 représente quelques-unes des interactions complexes mises en évidence grâce aux outils de la biologie moléculaire. Les conséquences de ces découvertes sont majeures avec, par exemple, outre la compréhension de la physiologie humaine, des implications thérapeutiques considérables, l'apparition à court terme d'une médecine prédictive précise, des aides précieuses au diagnostic et au pronostic des maladies.

## Nomenclature des variations de séquence

Les variations de séquence (polymorphismes et mutations) peuvent être exprimées au niveau de l'ADN ou de la protéine. Les auteurs de la nomenclature utilisent le



**Figure 2.32** Physiopathologie générale en génétique moléculaire. Un ou plusieurs gènes peuvent être impliqués et interagir entre eux. Selon la pathologie, plusieurs possibilités : un gène est atteint et une seule mutation causale (par exemple, la drépanocytose) ; un gène est atteint et de nombreuses mutations sont responsables de la même pathologie (voir hémophilie A) ; plusieurs gènes donnent la même maladie (par exemple, la rétinite pigmentaire) ; un gène selon la localisation de la mutation peut donner différentes maladies. C'est le cas du gène CFTR dont les nombreuses mutations sont essentiellement responsables de la mucoviscidose. Mais certaines mutations peuvent donner une infertilité ou une pancréatite par exemple. Le principe est le même pour le gène RET impliqué dans la maladie de Hirschprung, le carcinome médullaire de la thyroïde (MTC) ou dans le syndrome MEN2A (cancer endocrine multiple). Enfin, selon le(s) gène(s) en cause, une relation génotype/phénotype peut ou non exister. La gravité de la maladie peut être variable de modérée à grave (par exemple, pathologie du gène dystrophine responsable de la myopathie de Duchenne et de la maladie de Becker). Des interactions complexes entre l'environnement, des gènes modificateurs et des polymorphismes fonctionnels peuvent moduler l'expressivité de la maladie. Soit celle-ci est absente, soit à des degrés variables le sujet peut être malade. Ces différents facteurs interviennent probablement dans la pénétrance d'une maladie. Enfin, il ne faut pas oublier les anomalies de méthylation responsables de pathologies spécifiques (anomalies d'empreinte génomique) ou participant à des maladies multifactorielles (cancers).

terme « variation de séquence » pour éviter toute confusion avec les termes mutation et polymorphisme. En effet, selon les disciplines, une mutation peut signifier un changement de séquence quelle que soit la conséquence ou un changement induisant une pathologie. Enfin, pour les polymorphismes, il peut s'agir d'un changement sans conséquence pathologique pour certains ou un changement dont la fréquence est supérieure à 1 % dans la population générale pour d'autres. Nous décrirons les grandes lignes de cette nomenclature, la totalité du document étant accessible sur internet (nomenclature des variations de séquence : <http://www.hgvs.org/mutnomen/>).

La nomenclature permet de différencier la variation au niveau de :

- l'ADN génomique (lettre g ; ADNg). Certains préfèrent utiliser l'ADN génomique du fait de cas difficiles de numérotation, par exemple en cas d'épissage alter-

natif, de sites alternatifs d'initiation de la transcription ou de la traduction ;

- l'ADN complémentaire (lettre c ; ADNc) ;
- la mitochondrie (lettre m ; ADNm) ;
- l'ARN (lettre r) ;
- la protéine (lettre p). Dans ce cas, on utilise le code des acides aminés à une lettre (par exemple, P pour proline).

La variation de séquence est décrite par rapport à une séquence référence possédant un numéro d'accèsion (les banques de données les plus utilisées étant Genbank, EMBL, SWISS-PROT, DDJB, toutes accessibles sur internet).

Les principales modifications pour la description des variations de séquence (que ce soit au niveau ADN, ARN ou protéine) concernent :

- les mutations ponctuelles ;
- les insertions ;

## Principes de biologie moléculaire en biologie clinique

Nedjma Ameziane, Marc Bogard, Jérôme Lamoril

Depuis ses débuts, il y a 50 ans, la biologie moléculaire ne cesse de révolutionner le monde de la biologie par ses nombreuses découvertes. Les avancées de cette discipline et leurs multiples implications bouleversent notre vision de l'humain et du vivant et s'accompagnent de débats cruciaux sur le plan de l'éthique médicale et scientifique.

À l'heure où le séquençage du génome humain est achevé avec succès, la PCR en temps réel, l'interférence ARN, la génomique, les puces d'expression, de diagnostic et de séquençage, la pharmacogénétique, la protéomique, s'imposent comme de nouveaux outils au service de la médecine et de la biologie.

Ce livre présente le vaste panorama des techniques moléculaires disponibles, leur puissance d'analyse et la complexité des phénomènes étudiés. Novateur à plus d'un titre, *Principes de biologie moléculaire en biologie clinique* fait apparaître l'importance et le statut particulier que la biologie moléculaire tient actuellement dans la pratique de la médecine. En effet, il semble dorénavant difficile de n'y voir qu'un groupe de techniques, mais plutôt une spécialité au sein de la biologie médicale.

Les auteurs exposent les procédés et les clés d'interprétation de manière transversale car les technologies moléculaires reposent sur les mêmes principes, quel que soit le domaine particulier de leur application dans la pratique médicale.

Extrêmement pédagogique et clair, cet ouvrage s'adresse à un large public : étudiants en médecine, en pharmacie et en sciences de la vie, biologistes et cliniciens.

<http://france.elsevier.com>