

# mini Manuel

de

## biologie moléculaire

Cours + QCM/QROC

**Abderrahman Maftah**

Professeur à l'université de Limoges

**Jean-Michel Petit**

Professeur à l'université de Limoges

**Raymond Julien**

Professeur émérite à l'université de Limoges

4<sup>e</sup> édition

DUNOD

Les illustrations de cet ouvrage ont été réalisées par Sébastien ARICO.

### Directeur d'ouvrage

Raymond JULIEN

Le pictogramme qui figure ci-contre mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du photocopillage.

Le Code de la propriété intellectuelle du 1<sup>er</sup> juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique

s'est généralisée dans les établissements

d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour

les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée.

Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation de l'auteur, de son éditeur ou du Centre français d'exploitation du

droit de copie (CFC, 20, rue des Grands-Augustins, 75006 Paris).



© Dunod, Paris, 2007, 2011, 2015, 2018

11, rue Paul Bert, 92210 Malakoff

[www.dunod.com](http://www.dunod.com)

ISBN 978-2-10-077368-8

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2° et 3° a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

# Table des matières

<b>1</b>	<b>Structure de l'ADN et de l'ARN</b>	<b>1</b>
<b>1.1</b>	<b>Les composants des acides nucléiques</b>	<b>1</b>
	La structure des nucléotides	3
	La structure des polynucléotides	5
<b>1.2</b>	<b>La structure en double hélice de l'ADN</b>	<b>5</b>
	La règle de Chargaff et les appariements complémentaires	5
	Les différentes formes d'ADN	8
	Dissociation et réassociation des brins d'ADN	9
	Les surenroulements de l'ADN	11
<b>1.3</b>	<b>Le nucléosome, la chromatine et les chromosomes</b>	<b>13</b>
	La structure du nucléosome	13
	La structure et le remodelage de la chromatine	15
	La structure des chromosomes et le cycle cellulaire	16
<b>1.4</b>	<b>La structure des génomes</b>	<b>19</b>
	Qu'est-ce qu'un génome ?	19
	La taille des génomes	19
	Les génomes viraux	21
	Les génomes procaryotes	21
	Les génomes eucaryotes	21
	Les génomes d'organites	22
<b>1.5</b>	<b>Les différents types d'ARN</b>	<b>23</b>
	Structure et fonction	23
	Qu'est-ce qu'un ARN non codant ?	25
	Qu'est-ce que l'ARN interférence ?	26
	<b>Points clefs</b>	<b>28</b>
	<b>QCM - QROC</b>	<b>29</b>
	<b>Réponses</b>	<b>31</b>

<b>2</b>	<b>Réplication, réparation, recombinaison et transposition de l'ADN</b>	<b>33</b>
	<b>2.1 Les mécanismes de réplication de l'ADN</b>	<b>33</b>
	La chimie de synthèse cellulaire des polydésoxyribonucléotides	34
	L'action de l'ADN polymérase	35
	La fourche de réplication	36
	Les autres enzymes et protéines de la réplication	37
	Les différentes ADN polymérases	40
	Les différentes étapes de la réplication	41
	<b>2.2 Les erreurs de réplication de l'ADN et leur réparation</b>	<b>46</b>
	Les altérations de la structure de l'ADN	47
	Les mécanismes de réparation	48
	<b>2.3 Les détériorations environnementales de l'ADN et leur réparation</b>	<b>50</b>
	L'hydrolyse spontanée et les détériorations physico-chimiques	50
	Les agents intercalants	51
	La réparation des détériorations	51
	<b>2.4 La recombinaison et la transposition de l'ADN</b>	<b>54</b>
	Les mécanismes de recombinaison homologue	54
	La recombinaison en des sites spécifiques et la transposition	61
	<b>Points clefs</b>	69
	<b>QCM - QROC</b>	70
	<b>Réponses</b>	72
<b>3</b>	<b>La transcription de l'ADN</b>	<b>75</b>
	<b>3.1 Les mécanismes de la transcription</b>	<b>75</b>
	Les ARN polymérases	75
	Les différentes étapes de la transcription	77
	<b>3.2 La transcription chez les bactéries</b>	<b>78</b>
	Les promoteurs bactériens	78
	Le démarrage de la transcription	79
	La phase d'allongement	79
	L'arrêt de la transcription	81

<b>3.3 La transcription chez les eucaryotes</b>	<b>82</b>
Les promoteurs eucaryotes et la polymérase II	82
Le démarrage : facteurs de transcription et complexe médiateur	83
Les phases d'allongement et d'arrêt	86
Les modifications des transcrits	86
Les deux autres polymérases eucaryotes	89
<b>3.4 L'épissage de l'ARN</b>	<b>90</b>
Le mécanisme général	90
Le splicéosome	93
L'épissage alternatif et sa régulation	93
Exemples de rôles biologiques de l'épissage alternatif	95
<b>3.5 L'« editing des transcrits »</b>	<b>96</b>
<b>Points clefs</b>	99
<b>QCM - QROC</b>	101
<b>Réponses</b>	102
<b>4 La traduction des ARN messagers</b>	<b>105</b>
<b>4.1 Le code génétique</b>	<b>106</b>
Le code génétique est dégénéré	106
Le code a été établi expérimentalement	108
Le code est lu sur l'ARN messager dans le sens 5'-3'	108
Les codons ne sont pas chevauchants	109
Les mutations modifiant le sens des codons	109
Le code génétique est universel	111
<b>4.2 Les principaux acteurs de la traduction</b>	<b>111</b>
Les ARN messagers	112
Les ARN de transfert	112
Le ribosome	116
<b>4.3 La traduction des ARN messagers bactériens</b>	<b>120</b>
Le démarrage (initiation) de la traduction	120
L'étape d'allongement (élongation) de la chaîne polypeptidique	122
L'arrêt de la synthèse (terminaison)	125

<b>4.4 La traduction des ARN messagers eucaryotes</b>	<b>125</b>
Le démarrage de la traduction eucaryote	125
Les étapes d'allongement et d'arrêt de la traduction eucaryote	128
Traduction et demi-vie des ARN messagers eucaryotes	128
<b>Points clefs</b>	131
<b>QCM - QROC</b>	132
<b>Réponses</b>	135
<b>5 Régulation de l'expression des gènes</b>	<b>139</b>
<b>5.1 Principes généraux</b>	<b>139</b>
Les protéines régulatrices : activateurs et répresseurs	139
Le recrutement des ARN polymérases	140
Autres exemples de facteurs de régulation	141
<b>5.2 Régulation chez les procaryotes</b>	<b>142</b>
L'exemple historique : l'opéron lactose	142
Autres exemples	147
La régulation complexe du cycle vital du bactériophage $\lambda$	151
<b>5.3 Régulation chez les eucaryotes</b>	<b>156</b>
Les régulateurs transcriptionnels	157
Le contrôle des régulateurs transcriptionnels	161
Le contrôle de l'épissage alternatif des transcrits ARN	164
<b>5.4 Régulation traductionnelle de l'expression des gènes eucaryotes</b>	<b>165</b>
Éléments de structure des ARN messagers influençant la traduction	166
Le contrôle général par la phosphorylation des facteurs de démarrage	167
Les mécanismes spécifiques régulant l'attachement du ribosome à l'ARN messenger	169
Les mécanismes de régulation plus tardifs	170
Les mécanismes de régulation de la traduction par les micro-ARN	170

L'irrésistible ascension de l'ARN comme régulateur de l'expression des gènes	174
<b>Points clefs</b>	177
<b>QCM-QROC</b>	179
<b>Réponses</b>	182
<b>6 Techniques de biologie moléculaire</b>	<b>185</b>
<b>6.1 La création de molécules d'ADN recombinant</b>	<b>185</b>
Couper l'ADN : les enzymes de restriction	185
Ligaturer l'ADN	187
<b>6.2 Les vecteurs de clonage</b>	<b>189</b>
Les plasmides	190
Les vecteurs viraux	193
Les cosmides	193
Les chromosomes artificiels bactériens	194
Les vecteurs pour levures	194
Les vecteurs pour les eucaryotes supérieurs	196
<b>6.3 Les banques d'ADN</b>	<b>196</b>
Les banques d'ADN génomique	196
Les banques d'ADN complémentaire	197
<b>6.4 Les techniques d'analyse de l'ADN</b>	<b>197</b>
Le séquençage des acides nucléiques	197
Les nouvelles techniques de séquençage	201
Application au séquençage de l'ARN	203
Application à la métagénomique	205
Une nouvelle révolution dans le séquençage	207
La réaction de polymérisation en chaîne (PCR)	210
La PCR quantitative en temps réel	210
Les techniques d'hybridation des acides nucléiques	213
Les techniques de localisation des sites de liaison à l'ADN	215
<b>6.5 Le criblage de cellules recombinées</b>	<b>217</b>
Criblage par PCR	217
Criblage à l'aide de sites de restriction	218

**6.6 Les applications de la technologie de l'ADN recombinant 219**

La mutagenèse 219

Le système double-hybride 223

Transfert et expression de gènes eucaryotes  
chez les procaryotes 224

Transfert et expression de gènes dans les levures 228

Génie génétique et cellules eucaryotes supérieures 229

Utilisation de vecteurs rétroviraux pour la transfection  
des cellules 231

Les gènes rapporteurs et études des séquences promotrices 232

Un nouvel outil remarquable d'ingénierie des génomes :  
le système CRISPR-Cas 9 235

Les techniques d'analyse de l'épigénome 239

**Points clefs 241****QCM-QROC 243****Réponses 246****Glossaire 249****Index 259**



# Structure de l'ADN et de l'ARN

## PLAN

- 1.1 Les composants des acides nucléiques
- 1.2 La structure en double hélice de l'ADN
- 1.3 Le nucléosome, la chromatine et les chromosomes
- 1.4 La structure des génomes
- 1.5 Les différents types d'ARN

## OBJECTIFS

- Définir les structures chimiques des acides nucléiques, notamment celle de la double hélice de l'ADN
- Étudier les assemblages complexes auxquels participe l'ADN (nucléosome, chromatine)
- Répertoire d'un point de vue structural et fonctionnel les différents types d'ARN

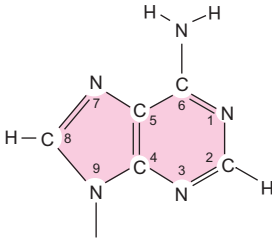
## 1.1 LES COMPOSANTS DES ACIDES NUCLÉIQUES

Les acides nucléiques (ADN et ARN) sont des macromolécules composées d'un enchaînement d'unités structurales appelées **nucléotides**. Ce sont donc des **polynucléotides**.

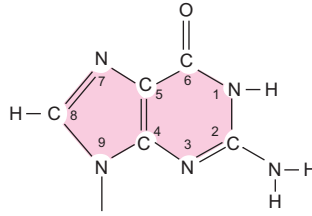
À l'état libre, chaque nucléotide est constitué :

- d'une base (Fig. 1-1), qui peut être une purine (composée de deux hétérocycles azotés) ou une pyrimidine (un seul hétérocycle azoté) ;
- d'un **pentose** (ose à cinq atomes de carbone, Fig. 1-2) ;
- d'un à trois groupes phosphate (Fig. 1-3).

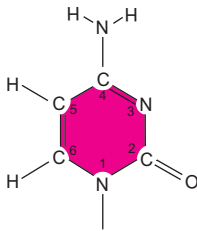
Les atomes des bases (C et N) portent les chiffres 1 à 6 pour les pyrimidines et 1 à 9 pour les purines, alors que les atomes C du pentose portent les chiffres 1' à 5'. Le signe prime ajouté à ces derniers les distingue de ceux des bases. L'ose, par son carbone 1', établit une **liaison N-glycosidique** avec l'azote 1 des pyrimidines ou l'azote 9 des purines. Par son carbone 5', il forme une liaison ester avec un premier groupe phosphate (désigné pour cette raison phosphate  $\alpha$ ). Ce dernier peut former des liaisons anhydride d'acide (Fig. 1-3) avec des groupes phosphate  $\beta$  et  $\gamma$  ou des liaisons ester avec un groupe OH porté par le carbone 3' du pentose précédent (Fig. 1-4).



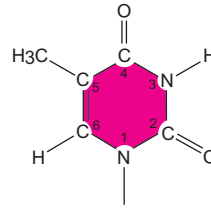
Adénine  
(6 amino**purine**)



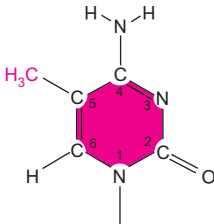
Guanine  
(2 amino, 6 oxo**purine**)



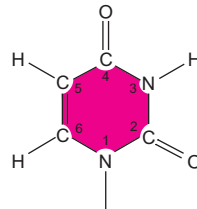
Cytosine  
(2 oxo, 4 amino**pyrimidine**)



Thymine  
(5méthyl-uracile)



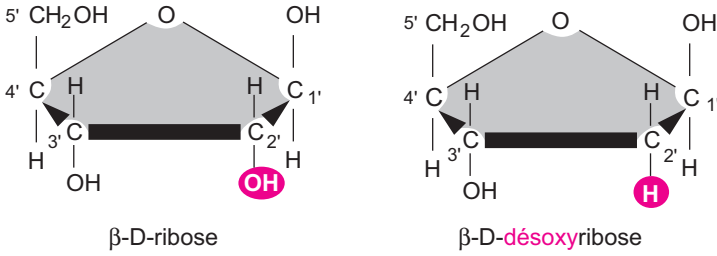
5 méthyl-cytosine



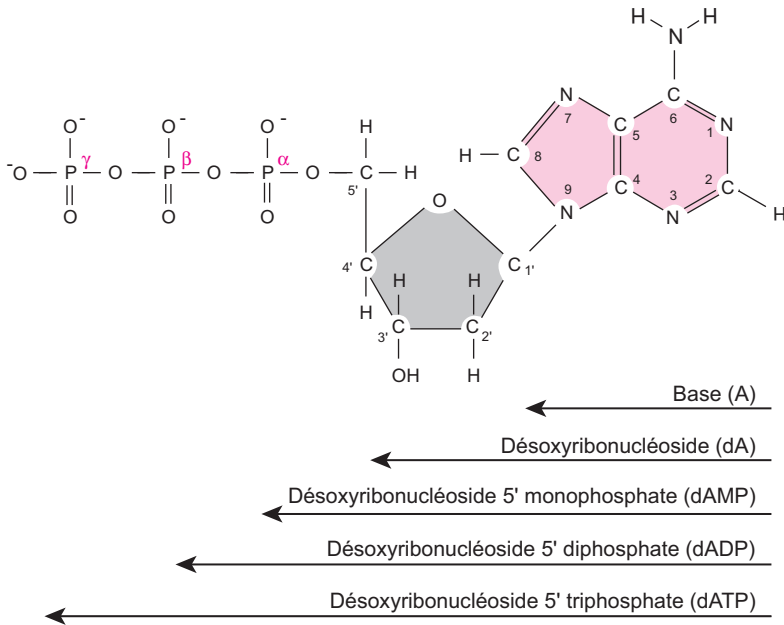
Uracile  
(2,4 dioxo**pyrimidine**)

**Figure 1-1** Structures chimiques des bases puriques (Adénine, Guanine) et des bases pyrimidiques (Cytosine, Thymine, Uracile) des acides nucléiques ADN et ARN. Des formes méthylées peuvent aussi être rencontrées (5-méthyl cytosine).

L'union d'une base et d'un pentose s'appelle un **nucléoside**. Dès qu'un groupe phosphate est présent, l'ensemble est désigné par le terme **nucléotide**. Il en résulte toute une nomenclature dont un exemple est donné (Fig. 1-3).



**Figure 1-2** Les deux pentoses respectifs de l'ARN et de l'ADN.

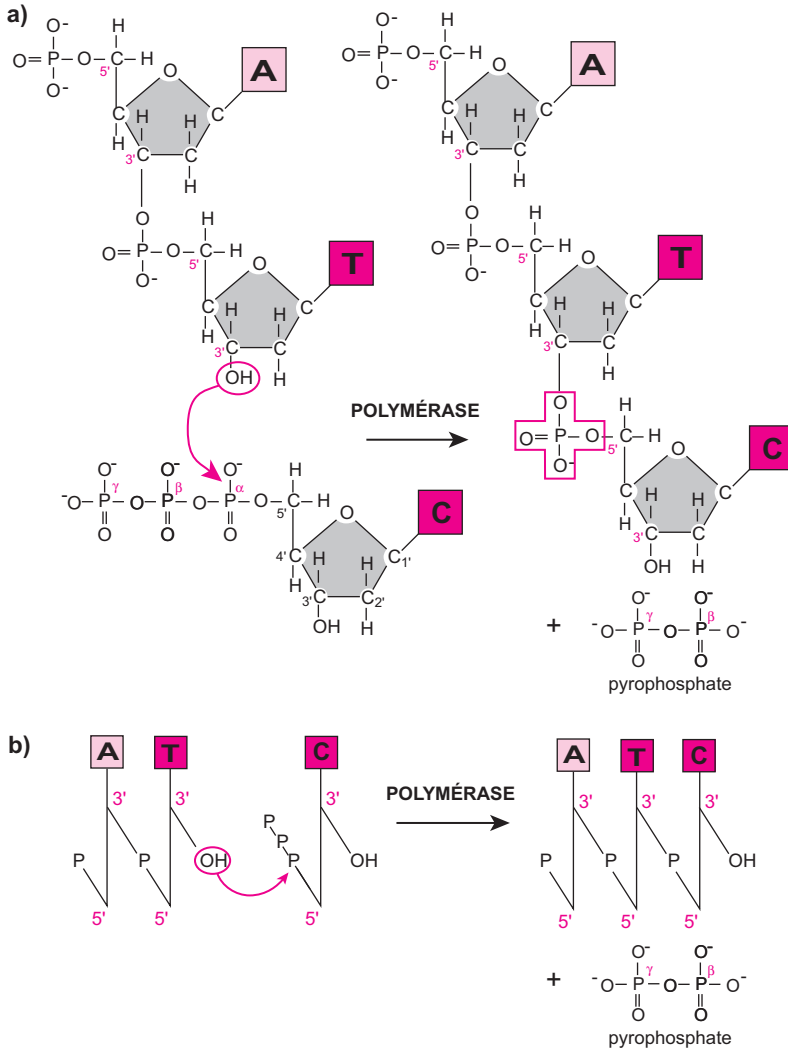


**Figure 1-3** Structure détaillée d'un déoxyribonucléoside triphosphate. En remplaçant la lettre A (Adénine) par les lettres G, C, T/U et le déoxyribose par le ribose, la nomenclature est généralisable à tous les nucléotides de l'ADN et de l'ARN.

## La structure des nucléotides

Dans l'ADN, l'ose est le **2'-déoxyribose** (Fig. 1-2) et les nucléotides sont des déoxyribonucléosides monophosphates (Fig. 1-4). Ils ne comportent qu'un seul groupe phosphate (le phosphate  $\alpha$ ) ;

- Les bases puriques de l'ADN sont l'**adénine** et la **guanine**.
- Les bases pyrimidiques de l'ADN sont la **cytosine** et la **thymine** (Fig. 1-1).



Dans l'ARN, l'ose étant un **ribose** (Fig. 1-2), les nucléotides sont des ribonucléosides monophosphates. Dans l'ARN, l'**uracile** remplace la thymine. Dans l'ADN, la cytosine peut être méthylée sur le carbone 5 pour donner la **5-méthyl-cytosine** (Fig. 1-1). Les cytosines méthylées sont souvent présentes dans des régions de séquences ADN

riches en répétition de nucléotides à guanine et cytosine (les îlots à dinucléotides CpG).

## La structure des polynucléotides

La structure d'un polynucléotide (polymère de nucléotides) repose sur la formation d'un nombre parfois élevé (plusieurs millions) de liaisons **phosphodiester** entre les nucléotides constitutifs (Fig. 1-4). Chaque liaison phosphodiester s'établit entre le groupe OH du carbone 3' d'un pentose et le groupe OH porté par le carbone 5' du pentose suivant. C'est la liaison phosphodiester 3'-5'. Il en résulte que l'orientation du polynucléotide se trouve définie par la position du groupe phosphate libre du premier nucléotide, généralement le phosphate porté par le carbone 5', et par le groupe OH libre porté par le carbone 3' du dernier nucléotide. On dit ainsi que le polynucléotide est orienté dans le sens 5' vers 3' (Fig. 1-4, voir aussi Fig. 1-6).

## 1.2 LA STRUCTURE EN DOUBLE HÉLICE DE L'ADN

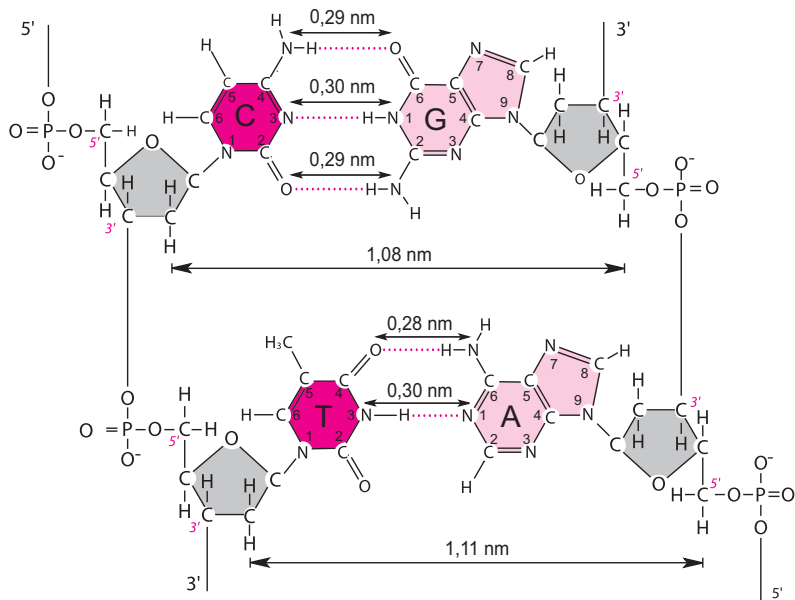
### La règle de Chargaff et les appariements complémentaires

L'ADN est un acide nucléique **bicaténaire**, c'est-à-dire constitué de deux brins associés par des **liaisons hydrogène** entre les bases. Les liaisons hydrogène s'établissent toujours entre une purine de l'un des brins et une pyrimidine de l'autre brin. Dans cet appariement complémentaire, l'adénine (A) est toujours associée à la thymine (T) par deux liaisons hydrogène et la guanine (G) interagit avec la cytosine (C) grâce à trois liaisons hydrogène (Fig. 1-5).

**Règle d'Erwin Chargaff.** Le nombre de thymines est égal au nombre d'adénines et celui des cytosines au nombre de guanines.

Ainsi, dans l'ADN, la quantité des purines (A+G) est égale à celle des pyrimidines (C+T). Par contre le rapport A+T sur G+C varie selon l'origine de l'ADN (Tableau 1-1). Le principe d'**appariement complémentaire** des bases a une grande valeur pratique : il permet au biologiste de déterminer avec exactitude la séquence nucléotidique d'un brin d'ADN dès lors qu'il connaît celle du brin complémentaire, qu'il s'agisse d'un brin ADN ou d'un transcrit ARN (voir chap. 3).

L'orientation d'un brin d'ADN est définie par la présence des groupes chimiques 5'-phosphate et 3'-OH portés par les carbones 5' et 3' des deux désoxyriboses extrêmes (Fig. 1-6 a). Dans la **double hélice**, les

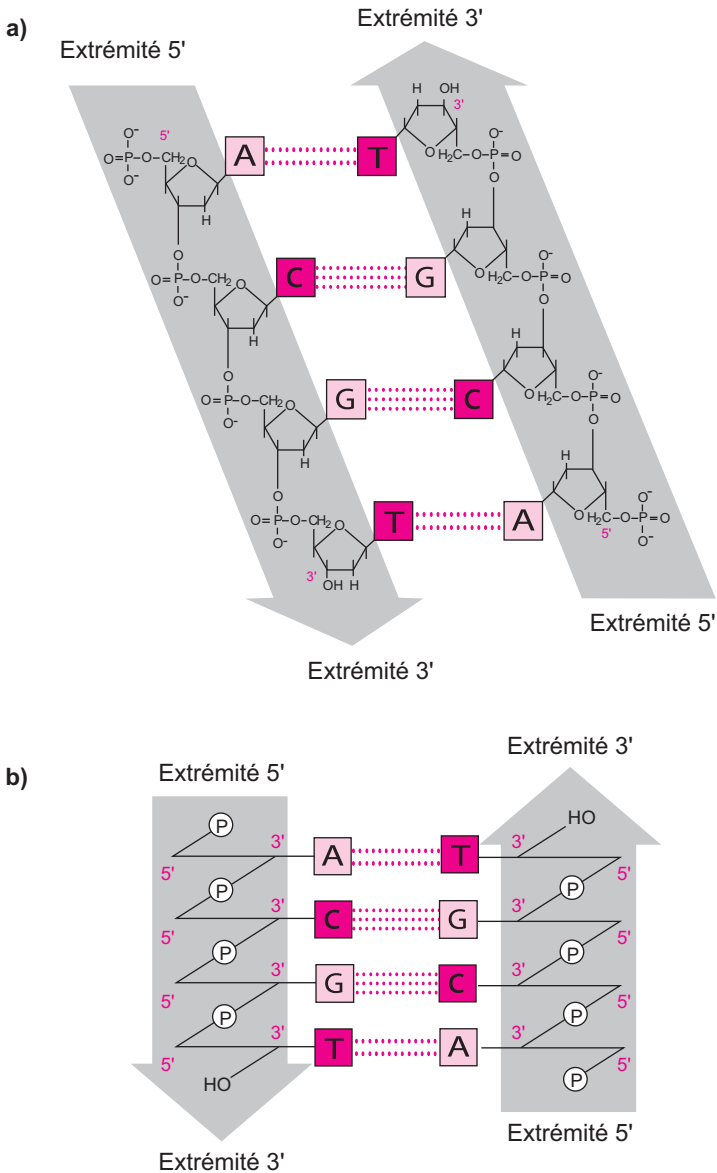


**Figure 1-5** Appariements complémentaires des bases G-C et A-T.

**TABEAU 1-1** COMPOSITION EN BASES DE L'ADN DE DIFFÉRENTES ORIGINES

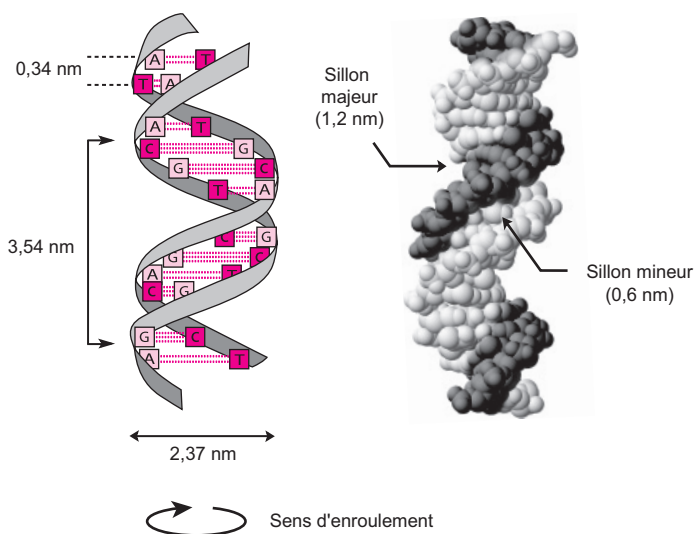
Origine de l'ADN	Bases (%)				(A+T)/(G+C)	(A+G)/(T+C)	%GC
	A	G	C	T			
Bactériophage T7	26,0	23,8	23,6	26,6	1,11	0,99	47,4
<i>Escherichia coli</i> B	23,8	26,8	26,6	23,1	0,88	1,01	53,2
Neurospora	23,0	27,1	26,6	23,3	0,86	1,00	53,8
Drosophile	30,7	19,6	20,2	29,5	1,51	1,01	39,8
Saumon	28,0	22,0	20,0	27,8	1,33	1,05	42,0
Poule	28,0	22,0	21,6	28,4	1,29	1,00	43,6
Rat	28,6	21,4	21,6	28,4	1,33	1,00	42,9
Vache	27,3	22,5	22,5	27,7	1,26	0,99	43,0
Homme	29,3	20,7	20,0	30,0	1,46	1,00	40,7

deux brins complémentaires d'ADN sont orientés en sens inverse (Fig. 1-6). On dit qu'ils sont **antiparallèles**.



**Figure 1-6** Structure des appariements complémentaires de deux brins antiparallèles d'ADN. **a)** Représentation développée ; **b)** représentation simplifiée.

L'association des deux brins par des liaisons hydrogène génère une contrainte physique qui impose à l'ensemble l'adoption d'une structure en double hélice (Fig. 1-7). Dans le modèle original, proposé par James Watson et Francis Crick en 1953, la double hélice a un diamètre de 2 nm et présente deux périodicités. La première de 0,34 nm, correspond à la distance séparant deux plateaux de bases appariées. La seconde de 3,4 nm correspond au pas de l'hélice qui couvre une longueur de 10,5 plateaux de bases (Fig. 1-7). À la surface de la molécule d'ADN, les deux brins torsadés ménagent deux **sillons hélicoïdaux** de largeur différente, un sillon majeur ou grand sillon et un sillon mineur ou petit sillon (Fig. 1-7). Ces deux sillons, tout particulièrement le grand sillon, exposent de nombreux groupes chimiques avec lesquels interagissent dans les conditions physiologiques de la cellule des protéines dites de liaison à l'ADN. Ces protéines jouent des rôles de régulation de l'expression de l'information génétique portée par la séquence en bases proprement dite (voir chap. 5).



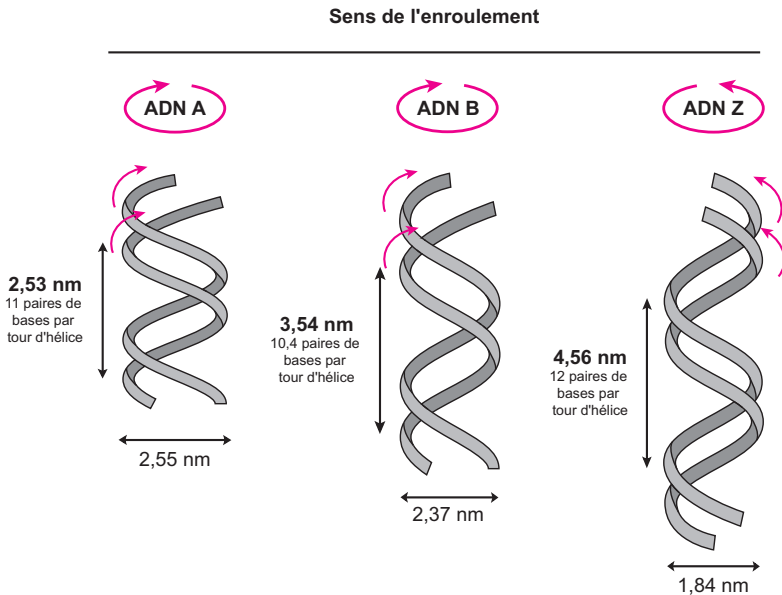
**Figure 1-7** Représentation d'un fragment d'ADN en double hélice. À gauche, avec les plateaux de bases ; à droite, les atomes sont représentés avec leurs rayons de Van der Waals.

## Les différentes formes d'ADN

En principe, les deux brins polynucléotidiques peuvent former une hélice de pas droit (forme dextre) ou une hélice de pas gauche (forme senestre). La première forme est la plus fréquente et existe en deux



variantes, l'ADN-A et l'ADN-B, qui diffèrent, entre autres, par la taille de leurs hélices (Fig. 1-8). L'ADN-A a un diamètre plus large (2,6 nm) que l'ADN-B (2 nm). Il est plus compact avec une distance de 0,23 nm entre deux plateaux de bases successifs (11 plateaux par tour). La forme A est observée notamment dans les hélices hybrides ADN-ARN. En solution et dans la cellule, c'est la forme B de l'ADN qui est prédominante. Une forme senestre de l'ADN étirée en zigzag (ADN-Z), avec un diamètre de 1,8 nm, est parfois observée (Fig. 1-8), mais sa signification physiologique reste incertaine.

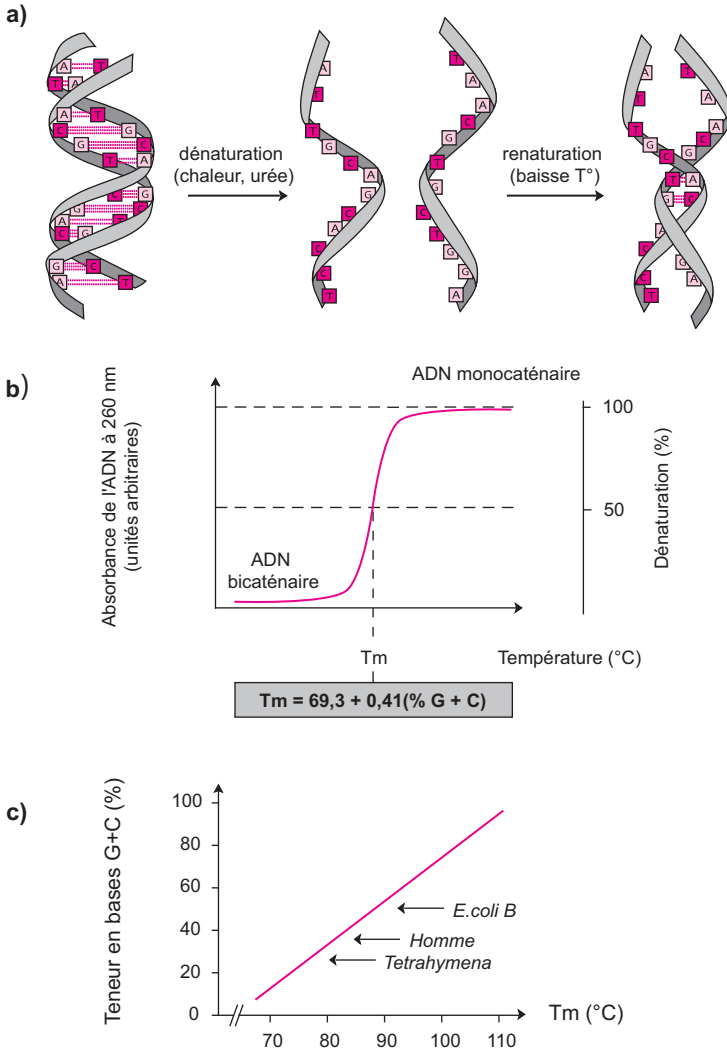


**Figure 1-8** Représentation simplifiée des formes A, B et Z de l'ADN.

## Dissociation et réassociation des brins d'ADN

En chauffant une solution d'ADN, l'agitation thermique provoque la rupture des liaisons hydrogène et la séparation des deux brins (Fig. 1-9 a). Cette dissociation est mise en évidence par la mesure de l'absorption de la lumière ultraviolette à 260 nm (longueur d'onde d'absorption maximale des bases puriques et pyrimidiques, due aux systèmes des doubles liaisons conjuguées liant les atomes des hétérocycles). Sous sa forme bicaténaire (en double hélice) l'ADN absorbe modérément la lumière UV. Après dissociation des deux brins, sous une forme monocaténaire, le démasquage des bases provoque une

absorption plus marquée de la lumière UV. L'enregistrement du phénomène à 260 nm fournit une courbe en forme de sigmoïde, typique de la dissociation de l'ADN (Fig. 1-9 b).



**Figure 1-9** La dissociation de l'ADN et sa réassociation dépend de sa teneur en G+C. **a)** Représentation simplifiée de la dissociation-réassociation de l'ADN sous l'action d'agents dénaturants ; **b)** représentation simplifiée de la courbe sigmoïde témoin du processus de dissociation. La formule établit la relation entre la teneur en G+C de l'ADN et la température de fusion de l'ADN ; **c)** droite reliant teneur en G+C et  $T_m$  de divers organismes.

La température pour laquelle on observe la dissociation de la moitié des molécules d'ADN est dénommée **température de fusion** ou **Tm** (« melting Temperature »).

Plus un ADN est riche en bases G et C qui sont appariées par des triples liaisons hydrogène, plus grande sera l'énergie nécessaire pour les rompre et plus élevée sera la température de fusion. À l'inverse, un ADN riche en bases A et T liées par seulement deux liaisons hydrogène, aura une température de fusion plus basse. La présence d'ions ou d'agents déstabilisants (urée ou formamide) peut modifier la Tm.



La teneur globale en bases G+C est une caractéristique variable suivant les grands groupes d'organismes (Fig. 1-9 c). Cela ne préjuge pas cependant de la distribution des paires AT et GC, souvent irrégulière, tout au long de la séquence.

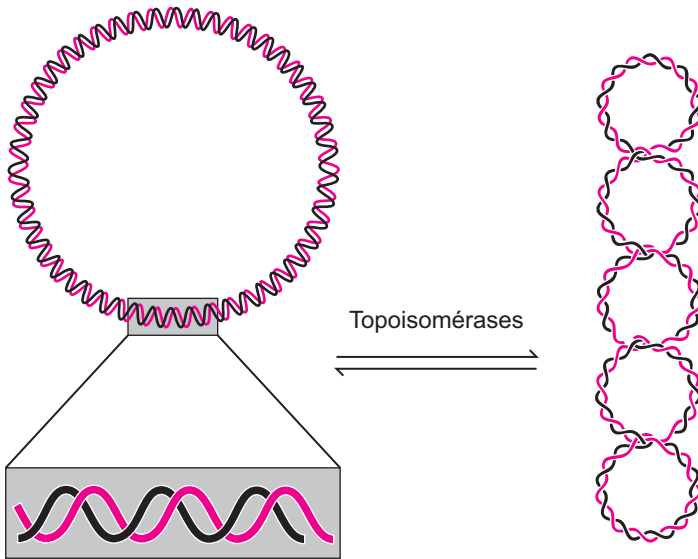
Dans certaines conditions expérimentales, les deux brins de l'ADN dissocié peuvent se réassocier et former à nouveau une double hélice. Le phénomène s'observe lors d'un refroidissement lent de la solution d'ADN préalablement chauffée (Fig. 1-9 a). Il est exploité pour réassocier des brins d'ADN d'origine différente. On parle d'**hybridation** moléculaire, et cette propriété est largement exploitée dans la recherche de séquences particulières dans l'ADN d'espèces différentes ou dans le génome d'une même espèce. Des duplex hybrides peuvent aussi se former de cette manière avec l'ARN et cette propriété est largement exploitée expérimentalement. La dissociation thermique de l'ADN suivie de sa réassociation par refroidissement est une étape de base utilisée dans la réaction de polymérisation en chaîne (voir chap. 6).

## Les surenroulements de l'ADN

Du fait même que l'ADN est un polymère flexible, sa structure dépend largement de son environnement salin et aussi des protéines avec lesquelles il peut se complexer.

Sous la forme linéaire souvent observée dans la plupart des cellules, les extrémités de l'ADN sont libres et les contraintes physiques introduites par ses divers ligands sont, pour cette raison, facilement ajustées. L'ADN peut cependant exister sous une forme circulaire, c'est-à-dire sans extrémités libres. C'est le cas de l'ADN bactérien, de beaucoup d'ADN viraux et de l'ADN mitochondrial et chloroplastique ou de levure (Tableau 1-2). C'est aussi le cas des **plasmides**, petits chromosomes bactériens surnuméraires capables de s'auto-répliquer de façon autonome. L'ADN circulaire subit des contraintes physiques

qui sont amoindries par un jeu de surenroulements parfois nombreux afin de stabiliser la molécule d'ADN dans son ensemble (Fig. 1-10).



**Figure 1-10** Exemple du surenroulement de l'ADN circulaire d'un plasmide bactérien et sa résolution par les topoisomérases.

**TABLEAU 1-2** TOPOLOGIE DE L'ADN DE DIVERSES ORIGINES

Source	Simple brin (S) ou double brin (D)	Circulaire ou linéaire	Taille*	
Virus Simien 40	D	circulaire	5243	pb
Virus X-φ174	S	circulaire	5386	b
Bactériophage M13	S	circulaire	6407	b
<i>Adénovirus AD-2</i>	D	linéaire	35 937	pb
<i>Virus Epstein-Barr</i>	D	circulaire	172 282	pb
Bactériophage T2	D	linéaire	$1,7 \times 10^5$	pb
<i>Escherichia coli</i>	D	circulaire	$4,7 \times 10^6$	pb
Drosophile	D	« linéaire »	$6,5 \times 10^7$	pb

(\*) b = base ; pb = paire de bases.

Dans l'ADN de très grande longueur des cellules eucaryotes, malgré la linéarité de la molécule, les **contraintes topologiques** demeurent au sein de la chromatine (voir plus loin). Il est donc important que ces contraintes soient réduites afin que les processus fondamentaux que sont la réplication et la transcription de l'ADN soient rendus possibles. Une classe spéciale d'enzymes, nommées **topoisomérases**, peut introduire des cassures spécifiques dans l'un ou l'autre brin ou même les deux. Ces cassures réduisent les contraintes, relâchent les surenroulements et rendent possibles les processus physiologiques.



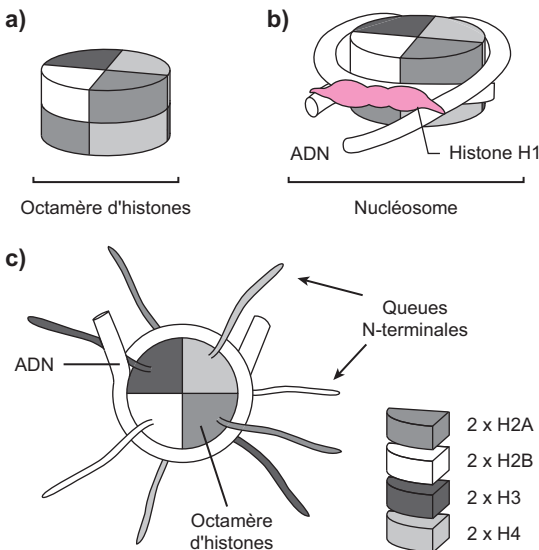
Ceci est particulièrement important lors du processus nécessaire à la séparation des deux molécules d'ADN résultant de la réplication et qui sont encore liées l'une à l'autre avant leur adressage aux deux cellules résultant de la division cellulaire.

## 1.3 LE NUCLÉOSOME, LA CHROMATINE ET LES CHROMOSOMES

### La structure du nucléosome

Dans les cellules eucaryotes, à la différence des procaryotes, l'ADN est très fréquemment associé à des protéines basiques, les **histones**.

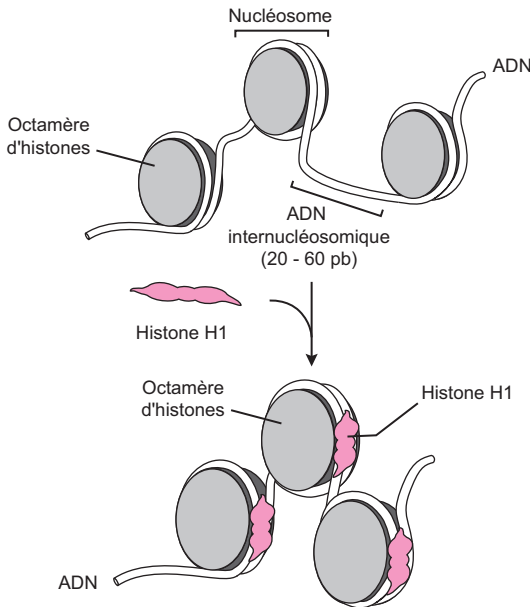
Le résultat de l'assemblage régulier entre ADN et histones est désigné par le terme de **nucléosome** (Fig. 1-11).



**Figure 1-11** Structure simplifiée d'un nucléosome. **a)** L'octamère d'histones ; **b)** l'enroulement de l'ADN autour de l'octamère et sa consolidation par l'histone H1 ; **c)** protubérance vers l'extérieur des queues N-terminales d'histones de l'octamère.

La formation des nucléosomes est la première étape d'un processus qui assure l'empaquetage de l'ADN en structures compactes réduisant énormément le volume occupé par la molécule. L'ensemble est désigné sous le terme de **chromatine**.

Chacune des classes d'histones est constituée de petites protéines riches en amino-acides chargés positivement (lysine et arginine). Leurs séquences en amino-acides sont restées remarquablement identiques au cours de l'évolution. Les **histones H2A, H2B, H3 et H4** (11 à 15 kDa) sont en quantité égale et chacune en double exemplaire, elles forment un **octamère** d'histones (Fig. 1-11 a) constituant un « cœur » protéique discoïde autour duquel l'**ADN nucléosomal** d'une taille de 147 paires de bases est enroulé (Fig. 1-11 b). L'histone **H1** (20 kDa) ne fait pas partie du « cœur ». Elle est attachée d'une part, à l'ADN de liaison situé entre les nucléosomes et, d'autre part, à l'ADN nucléosomal ce qui a pour effet de consolider l'ensemble (Fig. 1-11 et 1-12).



**Figure 1-12** Représentation simplifiée d'un enchaînement de trois nucléosomes montrant le rôle de l'histone H1 dans le resserrement de l'ADN autour de l'octamère d'histones.

De nombreux contacts existent entre les histones de l'octamère et l'ADN nucléosomal. Ils sont indépendants de la séquence en bases de l'ADN, ce qui explique que la presque totalité de l'ADN des cellules euca-

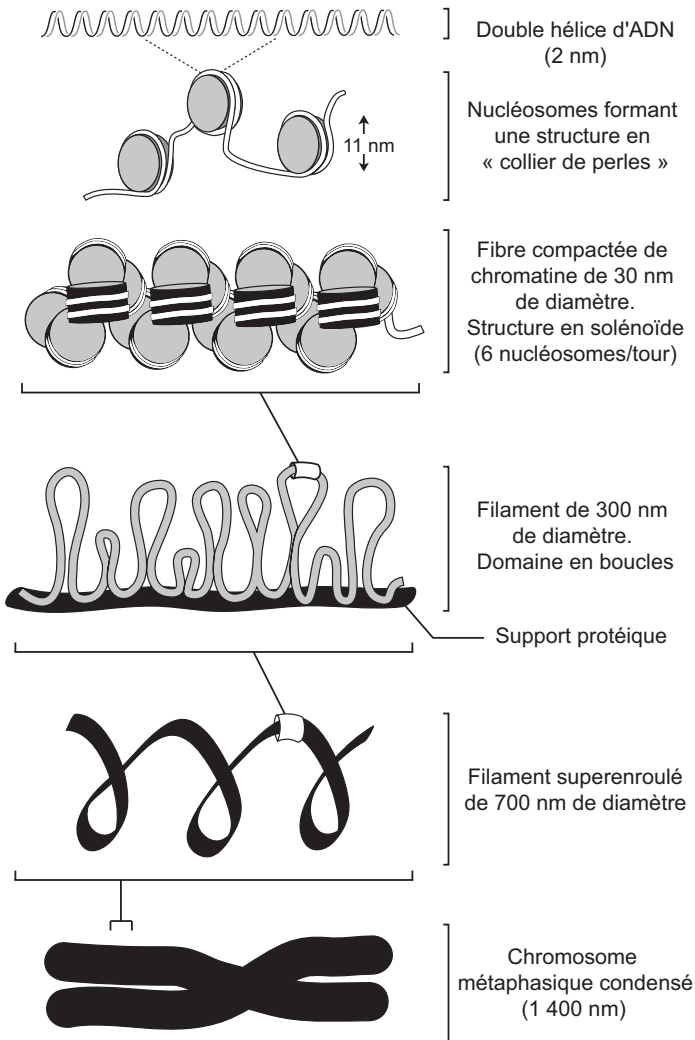
ryotes est organisée en nucléosomes. Les histones du cœur possèdent des extrémités N-terminales libres, désignées par le terme de « queues ». Ces queues émergent du nucléosome en des positions très précises (Fig. 1-11c). Elles facilitent la formation des structures d'ordre supérieur résultant des surenroulements de nucléosomes donnant la structure finale de la chromatine (Fig. 1-13). Certains amino-acides des **queues d'histones** sont parfois modifiés par des groupes acétyle et méthyle (lysine, arginine) ou phosphate (sérine). Ces modifications facilitent ou altèrent l'accessibilité de la chromatine. Elles constitueraient un « code » lu par les protéines impliquées dans la régulation de l'expression des gènes et dans les diverses autres fonctions de l'ADN. Les nucléosomes sont des structures dynamiques, constamment modifiées par de nombreux complexes protéiques de remodelage capables de les déplacer le long de la molécule d'ADN, de les transférer d'une molécule d'ADN à une autre ou de les modifier afin d'accroître l'accessibilité de l'ADN.

### La structure et le remodelage de la chromatine

La présence de l'histone H1 solidifie et resserre la structure du nucléosome. L'ADN nucléosomal représente alors 160 pb. L'histone H1 favorise aussi la formation de structure d'ordre supérieur. Les nucléosomes s'organisent alors en une **fibre de 30 nm** de diamètre qui est en fait une **super-hélice** comportant 6 nucléosomes par tour. Ceci ne suffit pas cependant pour empaqueter les 1 à 2 mètres d'ADN dans un noyau cellulaire de  $10^{-5}$  mètre de diamètre. Des repliements en forme de **boucles** de la fibre de 30 nm sont donc nécessaires. Les boucles seraient maintenues dans un ensemble compact par leur association à un support protéique, aujourd'hui remis en question (Fig. 1-13).



Le modèle présenté à la Figure 1-13 malgré ses vertus pédagogiques donne une vision trop simpliste de l'organisation tridimensionnelle de la chromatine et des chromosomes. En effet, de récentes analyses biophysiques très performantes (tomographie par microscopie électronique, cryo-microscopie électronique, diffraction des rayons X aux petits angles) montrent que les chromosomes sont plus élastiques que ne laisse supposer le schéma présenté et surtout que l'ADN y contribue autant que les protéines. Ces analyses excluent la possibilité que les chromosomes soient construits à partir d'un improbable échafaudage protéique ; ils posséderaient plutôt les propriétés d'une tresse d'ADN et de protéines. Les observations indiquent que les fibres de chromatine s'entrecroisent et sont à de fréquents intervalles liées entre elles par un complexe protéique de condensine jouant le rôle de liant. Par ailleurs, ni la fibre de 30 nm ni aucune autre organisation hiérarchique n'est observée avec ces techniques modernes. La structure la plus probable serait donc celle d'une chaîne de nucléosomes dont l'auto-organisation serait modulée grâce aux seules contraintes imposées par le complexe protéique de la condensine.



**Figure 1-13** Modèle hypothétique de l'organisation et de la compaction de la chromatine.

## La structure des chromosomes et le cycle cellulaire

Dans les cellules eucaryotes, le **chromosome** est le stade supérieur et ultime d'organisation de l'ADN (voir Fig. 1-13).

*Le nombre de chromosomes et leur composition sont des caractéristiques de chaque organisme.*



Le séquençage complet d'un certain nombre de génomes a montré que la quantité d'ADN effectivement utilisé pour coder les protéines ou les ARN varie énormément d'un organisme à un autre. Chez les bactéries, la presque totalité de l'ADN code des protéines et des ARN. Chez les mammifères, seule une petite portion (moins de 10 %) code réellement les protéines ou les divers ARN, le reste ou **ADN intergénique**, est constitué largement de séquences répétées (souvent des dinucléotides), appelées **microsatellites** (3 %) ou composées de grands éléments de **séquences transposables** (plus de 50 %).



La nécessaire et exacte duplication de la totalité de l'ADN à chaque division cellulaire et le fait que 80 % serait effectivement transcrit en ARN (voir *Mini-manuel de Génétique*, projet ENCODE) militent en faveur d'un rôle biologique majeur de l'ADN intergénique, même s'il est encore largement incompris.

Pour être maintenu comme tels au cours de la division cellulaire, chaque chromosome chez les eucaryotes doit posséder un **centromère**, des **télomères** et plusieurs **origines de réplication**.

Les origines de réplication sont les nombreux sites (3 000 par chromosome en moyenne) où s'assemblent les divers composants protéiques, dont les **ADN polymérases** (voir chap. 2), nécessaires au démarrage de la réplication (duplication de l'ADN) opération qui précède la division cellulaire.

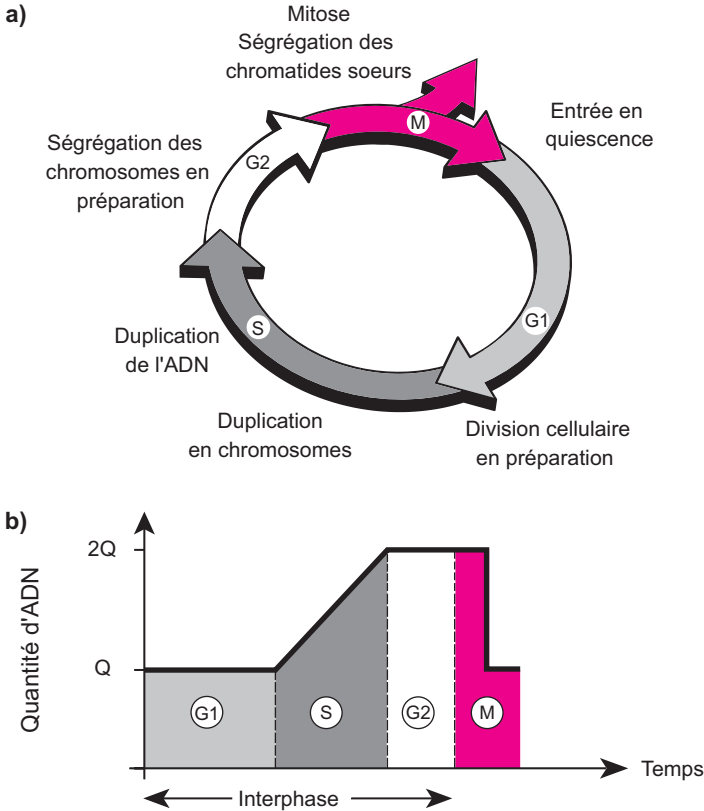
Les centromères sont des séquences d'une taille supérieure à 40 kb au contact desquelles se construit le **kinétochore**, un complexe protéique nécessaire à la séparation et l'adressage corrects des deux chromosomes-fils aux deux cellules filles.



Il est essentiel qu'un seul centromère soit présent par chromosome. Dans le cas contraire, il en résulte des cassures, des pertes ou des duplications de chromosomes dont les conséquences sont généralement graves.

Les télomères sont des séquences particulières localisées aux deux extrémités du chromosome. Elles sont reconnues par des protéines spécifiques qui protègent le chromosome des événements de recombinaison ou de dégradation, typiques des extrémités ADN libres. À l'endroit des télomères se trouve une origine de réplication particulière reconnue par une ADN polymérase spéciale, la **télomérase**, qui réplique un certain nombre de fois la séquence correspondante (voir chap. 2).

La duplication du chromosome puis sa ségrégation s'opèrent dans des phases distinctes du **cycle cellulaire**. On distingue dans l'ordre 4 phases : **G1**, **S**, **G2** et **M** (Fig. 1-14).



**Figure 1-14** Chromosomes et cycle cellulaire. **a)** Duplication et ségrégation des chromosomes au cours du cycle cellulaire. **b)** Variation de la quantité d'ADN «  $Q$  » au niveau des chromosomes à un instant  $t$  dans la cellule. Par simplification, l'ADN est représenté sous forme de chromosomes à toutes les phases du cycle.

Au cours de la phase S s'opère la synthèse de l'ADN conduisant à la duplication des chromosomes en deux **chromatides** soeurs maintenues ensemble par un complexe protéique appelé **cohésine**, jusqu'à la ségrégation.

Celle-ci s'opère durant la phase M (pour mitose). Chaque chromatide s'attache grâce au kinétochore au fuseau mitotique fait de protéines fibreuses appelées **microtubules** réunis aux **centrosomes** situés aux deux pôles de la cellule. La désagrégation de la cohésine favorise la séparation des deux chromatides soeurs. L'ensemble des opérations est soigneusement coordonné afin que la séparation assure un transfert complet et efficace des informations génétiques portées par les deux chromosomes-fils.

Les phases G1 et G2 (G pour gap) sont des phases de préparation où des mécanismes de contrôle du cycle opèrent qui autoriseront ou non sa poursuite jusqu'à son terme.

Au cours du processus complet de la division cellulaire, l'état de condensation des chromosomes varie profondément. Durant les phases G1, S et G2 (groupées sous le terme d'**interphase**) il est plus faible que durant la phase M bien que des changements discrets se produisent constamment.

## 1.4 LA STRUCTURE DES GÉNOMES

### Qu'est-ce qu'un génome ?

On désigne par le terme de **génome** (contraction des mots **gène** et **chromosome**) l'ensemble des informations génétiques, codées par l'ADN chez la plupart des organismes et parfois par l'ARN pour certains virus. De façon plus précise, le terme désigne la séquence de l'ADN correspondant à un jeu **haploïde** de chromosomes ( $n$  chromosomes). Une cellule somatique humaine, **diploïde** ( $2n$  chromosomes), contient donc deux génomes (un génome paternel et un génome maternel) alors qu'une cellule germinale (sexuelle), haploïde (ovule ou spermatozoïde), en contient un seul. Le terme génome s'applique aussi bien à l'ADN du noyau cellulaire (**génome nucléaire**) qu'à l'ADN des organites (**génome mitochondrial**, **génome chloroplastique**). On parle de **génomique** (contraction des mots *génome* et *informatique*) pour désigner l'étude à grande échelle des propriétés générales des génomes grâce à des moyens informatiques de plus en plus puissants. Les premières questions intéressant l'étude des génomes portent sur leur taille, le nombre de gènes qu'ils contiennent (au sens classique du concept), le nombre de chromosomes constitutifs et la manière dont les gènes sont distribués ou organisés.

### La taille des génomes

Pour désigner la taille d'un génome on utilise habituellement une unité exprimant le nombre de paires de nucléotides. Pour le millier de paires de nucléotides ou de bases soit  $10^3$ , l'unité est la kilopaire de bases ou **Kpb**, pour le million soit  $10^6$ , l'unité est la mégapaire de bases ou **Mpb**. On pensait généralement qu'il existait une relation progressive entre la taille des génomes, le nombre de gènes qu'ils contiennent et la complexité des groupes d'organismes vivants (Fig. 1-15, Tableau 1-3). En vérité, plus la connaissance intime de la structure et du fonctionnement des génomes progresse, plus cette relation s'avère inexacte. En effet, l'organisation des instructions génétiques au sein des génomes et surtout leur mode d'expression semblent à la fois très diversifiés et variables selon les organismes.

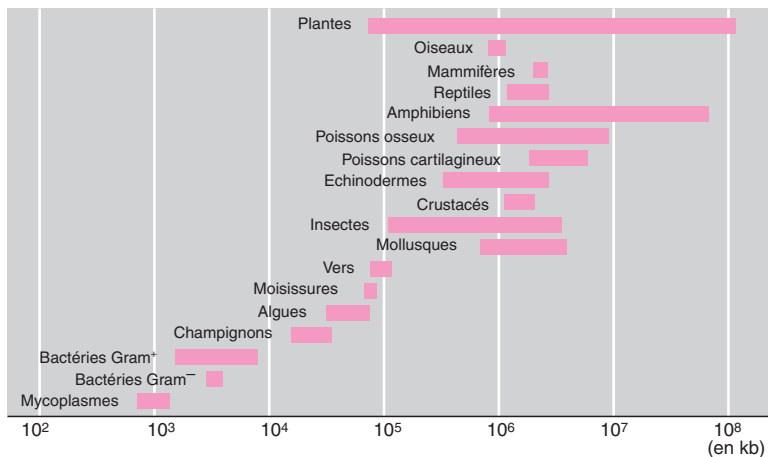


Figure 1-15 Quantité d'ADN présent dans les génomes de différents groupes d'organismes.

Génome	Groupe	Structure du génome	Taille (kb)	Nombre de gènes
<b>Eucaryotes :</b>				
<i>Saccharomyces cerevisiae</i>	Levures		13 500 (L)	6 000
<i>Caenorhabditis elegans</i>	Nématode		100 000 (L)	13 500
<i>Arabidopsis thaliana</i>	Plante		120 000 (L)	25 000
<i>Homo sapiens</i>	Humain		3 000 000 (L)	22 000
<b>Procaréotes :</b>				
<i>Escherichia coli</i>	Bactérie		4 700 (C)	4 000
<i>Haemophilus influenza</i>	Bactérie		1 830 (L)	1 703
<i>Methanococcus jannaschii</i>	Bactérie		1 660 (L)	1 738
<b>Virus :</b>				
Virus T4	Virus bactérien		172 (L/C)	300
HCMV (virus de l'herpès)	Virus eucaryote		229 (L)	200
<b>Organites</b>				
Mitochondrie de <i>S. cerevisiae</i>	Levure		78 (C)	34
Mitochondrie de <i>H. sapiens</i>	Humain		17 (C)	37
Chloroplaste	Plante		121 (C)	136
			L : Linéaire C : Circulaire	

TABLEAU 1-3 GÉNOMES DE DIVERS ORGANISMES : STRUCTURE, TAILLE ET ESTIMATION DU NOMBRE DE GÈNES.

## Les génomes viraux

Bien que les virus ne soient pas considérés comme des organismes vivants mais comme des parasites moléculaires, ils se reproduisent en infectant des cellules vivantes dans lesquelles ils injectent leur génome constitué par de l'ADN tels les adénovirus, protégé au sein d'une particule ou capsid de nature essentiellement protéique. Certains virus, comme les rétrovirus auxquels se rattache par exemple le VIH, ont un génome constitué d'ARN. Quel que soit l'acide nucléique, il est pour certains virus, sous une forme linéaire et pour d'autres, sous une forme circulaire. Dans tous les cas, le génome viral existe au moins dans une des phases de son cycle sous la forme d'un ADN double brins.



*Les génomes viraux géants.* Découverts récemment, les Mimivirus et Pandoravirus parasites d'eucaryotes unicellulaires (amibes) possèdent des génomes géants, parfois plus grands (plus de  $10^6$  Kpb) et plus complexes (plus de 1 000 gènes) que des génomes bactériens. Avec la découverte de l'existence de virus de virus (des « virophages » en quelque sorte) ces nouveautés biologiques n'obligent-elles pas à réviser la place des virus dans l'échelle du vivant ? À la conception du virus « parasite moléculaire non vivant » ne faudrait-il pas substituer l'idée que les virus posséderaient les propriétés du vivant à l'exclusion de la cellularité ?

## Les génomes procaryotes

Le génome de la plupart des organismes procaryotes correspond à un seul chromosome composé d'un ADN souvent circulaire et de très peu de protéines associées. Une bactérie en cours de division peut cependant contenir simultanément plusieurs copies de ce chromosome unique. Entre les gènes bactériens, il y a peu d'**espaces intergéniques**. Les gènes sont souvent organisés en unités, les opérons, transcrits en un seul ARN messager (voir Fig. 4-3). L'ADN associé à quelques protéines forme une masse dense, le **nucléotide** qui se désorganise en un écheveau filamenteux lorsque la cellule est broyée.

Un grand nombre de bactéries possède en supplément du chromosome bactérien des **plasmides**, éléments d'ADN circulaires et double brins. Ces molécules d'ADN extrachromosomique peuvent se répliquer indépendamment et bien qu'elles ne portent pas de gènes essentiels au métabolisme de la bactérie, elles peuvent lui apporter des avantages évolutifs comme la résistance à un antibiotique. En général, les bactéries possèdent un ou plusieurs plasmides dont le nombre de copies varie.

## Les génomes eucaryotes

Les espèces eucaryotes sont soit diploïdes (les cellules somatiques des animaux et plantes à fleurs possèdent dans leur noyau, deux jeux –  $2n$  – de chromosomes) ou haploïdes (un seul jeu –  $n$  – de chromosomes par noyau cellulaire chez la plupart des champignons et les

algues). Les organismes diploïdes produisent généralement des cellules spécialisées, les gamètes, haploïdes ( $n$  chromosomes) pour la reproduction (ovule et spermatozoïde chez les animaux) alors que les organismes haploïdes peuvent présenter au cours de la phase sexuée de leur cycle vital des cellules spécialisées à  $2n$  chromosomes. Le nombre  $n$  de chromosomes présents dans les cellules d'un organisme définit le niveau de ploïdie. Certains organismes (les végétaux souvent) présentent des niveaux de **ploïdie** élevés (polyploïdie).

Dans une cellule diploïde, les chromosomes sont présents par paires homologues qui possèdent une taille et un nombre de gènes identiques aux mêmes positions. Le terme de **caryotype** est utilisé pour désigner le nombre de chromosomes, établi par l'observation microscopique d'une cellule en métaphase. Au sein d'un génome, les chromosomes peuvent différer fortement en taille et en nombre de gènes portés. Chez l'homme, les chromosomes 1 (247 Mpb ; 2 200 gènes) et 2 (243 Mpb ; 1 400 gènes) ont une grande taille alors que les chromosomes 21 et 22 sont cinq fois plus petits (environ 48 Mpb) avec respectivement 300 et 540 gènes. On constate qu'il n'y a pas de lien direct entre la taille d'un chromosome et le nombre de gènes présents. Le chromosome 19 est celui qui possède la plus forte densité de gènes, 1 500 gènes qui se répartissent sur 64 Mpb.

## Les génomes d'organites

À côté des chromosomes nucléaires qui représentent l'essentiel du génome d'un organisme eucaryote, certains organites (les mitochondries et les chloroplastes, présents parfois en grand nombre dans une cellule) possèdent des chromosomes circulaires, d'une taille moyenne avoisinant  $10^4$  à  $10^5$  paires de bases, souvent en de nombreuses copies identiques. Pour ces raisons, il arrive que selon les organismes, le nombre de génomes d'organites par cellule s'avère très élevé jusqu'à plusieurs centaines parfois. En général, les génomes d'organites possèdent de l'ordre de 50 à 100 gènes dont certains avec des introns. Ces gènes sont spécifiques des fonctions assurées par l'organite lui-même. Mais beaucoup des fonctions de l'organite sont spécifiées par des gènes nucléaires sans qu'il y ait de redondance entre les gènes des deux compartiments, mais plutôt une complémentarité.



*Les épigénomes.* La notion d'épigénome est une réponse, partielle sans doute, à la question fondamentale posée en particulier chez l'homme : pourquoi les cellules d'un individu diffèrent-elles d'un tissu à l'autre (et parfois au sein d'un même tissu), alors qu'elles ont toutes le même génome nucléaire ?

Concrètement le terme « épigénome » désigne l'ensemble des marques (ou modifications chimiques) apposées sur les principaux constituants du génome (ADN et histones) sans modification de la séquence en bases de l'ADN. Ces marques dites