

mini **manuel**

Biologie moléculaire

Cours + Exos + QCM/QROC

4^e édition

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

DUNOD

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

Directeur d'ouvrage
Raymond JULIEN

NOUS NOUS ENGAGEONS EN FAVEUR DE L'ENVIRONNEMENT :



Nos livres sont imprimés sur des papiers certifiés pour réduire notre impact sur l'environnement.



Le format de nos ouvrages est pensé afin d'optimiser l'utilisation du papier.



Depuis plus de 30 ans, nous imprimons 70 % de nos livres en France et 25 % en Europe et nous mettons tout en œuvre pour augmenter cet engagement auprès des imprimeurs français.



Nous limitons l'utilisation du plastique sur nos ouvrages (film sur les couvertures et les livres).

© Dunod, 2007, 2011, 2015, 2018, 2023
pour la nouvelle présentation
11, rue Paul Bert, 92210 Malakoff
www.dunod.com
ISBN 978-2-10-086789-9

Table des matières

1	Structure de l'ADN et de l'ARN	1
1.1	Les composants des acides nucléiques	1
	La structure des nucléotides	3
	La structure des polynucléotides	5
1.2	La structure en double hélice de l'ADN	5
	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
1.3	Le nucléosome, la chromatine et les chromosomes	13
	La structure du nucléosome	13
	La structure et le remodelage de la chromatine	15
	La structure des chromosomes et le cycle cellulaire	16
1.4	La structure des génomes	19
	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
1.5	Les différents types d'ARN	23
	Structure et fonction	23
	Qu'est-ce qu'un ARN non codant ?	25
	Qu'est-ce que l'ARN interférence ?	26
	Points clefs	28
	QCM - QROC	29
	Réponses	31

2	Réplication, réparation, recombinaison et transposition de l'ADN	33
	2.1 Les mécanismes de réplication de l'ADN	33
	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
	2.2 Les erreurs de réplication de l'ADN et leur réparation	46
	Les altérations de la structure de l'ADN	47
	Les mécanismes de réparation	48
	2.3 Les détériorations environnementales de l'ADN et leur réparation	50
	L'hydrolyse spontanée et les détériorations physico-chimiques	50
	Les agents intercalants	51
	La réparation des détériorations	51
	2.4 La recombinaison et la transposition de l'ADN	54
	Les mécanismes de recombinaison homologue	54
	La recombinaison en des sites spécifiques et la transposition	61
	Points clefs	69
	QCM - QROC	70
	Réponses	72
3	La transcription de l'ADN	75
	3.1 Les mécanismes de la transcription	75
	Les ARN polymérases	75
	Les différentes étapes de la transcription	77
	3.2 La transcription chez les bactéries	78
	Les promoteurs bactériens	78
	Le démarrage de la transcription	79
	La phase d'allongement	79
	L'arrêt de la transcription	81

3.3 La transcription chez les eucaryotes	82
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
3.4 L'épissage de l'ARN	90
Le mécanisme général	90
Le splicéosome	93
L'épissage alternatif et sa régulation	93
Exemples de rôles biologiques de l'épissage alternatif	95
3.5 L'« editing des transcrits »	96
Points clefs	99
QCM - QROC	101
Réponses	102
4 La traduction des ARN messagers	105
4.1 Le code génétique	106
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
4.2 Les principaux acteurs de la traduction	111
Les ARN messagers	112
Les ARN de transfert	112
Le ribosome	116
4.3 La traduction des ARN messagers bactériens	120
Le démarrage (initiation) de la traduction	120
L'étape d'allongement (élongation) de la chaîne polypeptidique	122
L'arrêt de la synthèse (terminaison)	125

4.4 La traduction des ARN messagers eucaryotes	125
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
Points clefs	131
QCM - QROC	132
Réponses	135
5 Régulation de l'expression des gènes	139
5.1 Principes généraux	139
Les protéines régulatrices : activateurs et répresseurs	139
Le recrutement des ARN polymérases	140
Autres exemples de facteurs de régulation	141
5.2 Régulation chez les procaryotes	142
L'exemple historique : l'opéron lactose	142
Autres exemples	147
La régulation complexe du cycle vital du bactériophage λ	151
5.3 Régulation chez les eucaryotes	156
Les régulateurs transcriptionnels	157
Le contrôle des régulateurs transcriptionnels	161
Le contrôle de l'épissage alternatif des transcrits ARN	164
5.4 Régulation traductionnelle de l'expression des gènes eucaryotes	165
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
Points clefs	177
QCM-QROC	179
Réponses	182
6 Techniques de biologie moléculaire	185
6.1 La création de molécules d'ADN recombinant	185
Couper l'ADN : les enzymes de restriction	185
Ligaturer l'ADN	187
6.2 Les vecteurs de clonage	189
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
6.3 Les banques d'ADN	196
Les banques d'ADN génomique	196
Les banques d'ADN complémentaire	197
6.4 Les techniques d'analyse de l'ADN	197
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
6.5 Le criblage de cellules recombinées	217
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

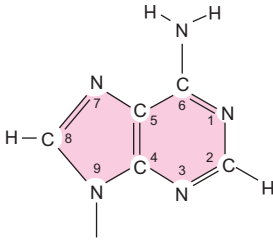
1.1 LES COMPOSANTS DES ACIDES NUCLÉIQUES

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

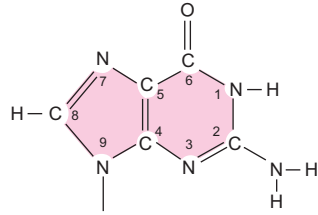
À l'état libre, chaque nucléotide est constitué :

- 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).

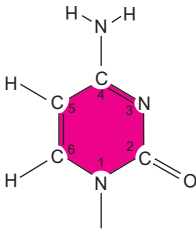
Les atomes des bases (C et N) portent les chiffres 1 à 6 pour les pyrimidines et 1 à 9 pour les purines, alors que les atomes C du pentose portent les chiffres 1' à 5'. Le signe prime ajouté à ces derniers les distingue de ceux des bases. L'ose, par son carbone 1', établit une **liaison N-glycosidique**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



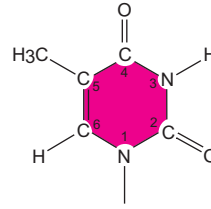
Adénine
(6 aminopurine)



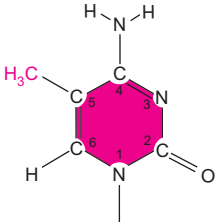
Guanine
(2 amino, 6 oxopurine)



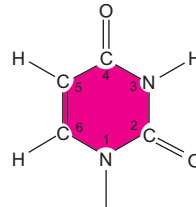
Cytosine
(2 oxo, 4 aminopyrimidine)



Thymine
(5méthyl-uracile)



5 méthyl-cytosine



Uracile
(2,4 dioxypyrimidine)

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).

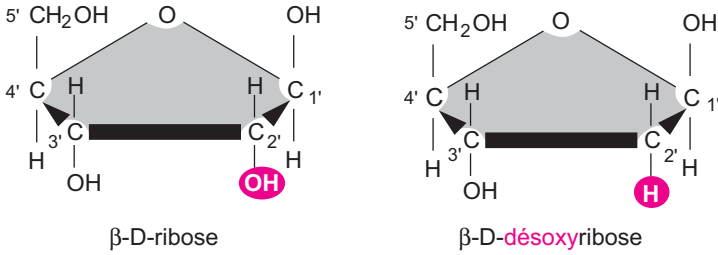


Figure 1-2 Les deux pentoses respectifs de l'ARN et de l'ADN.

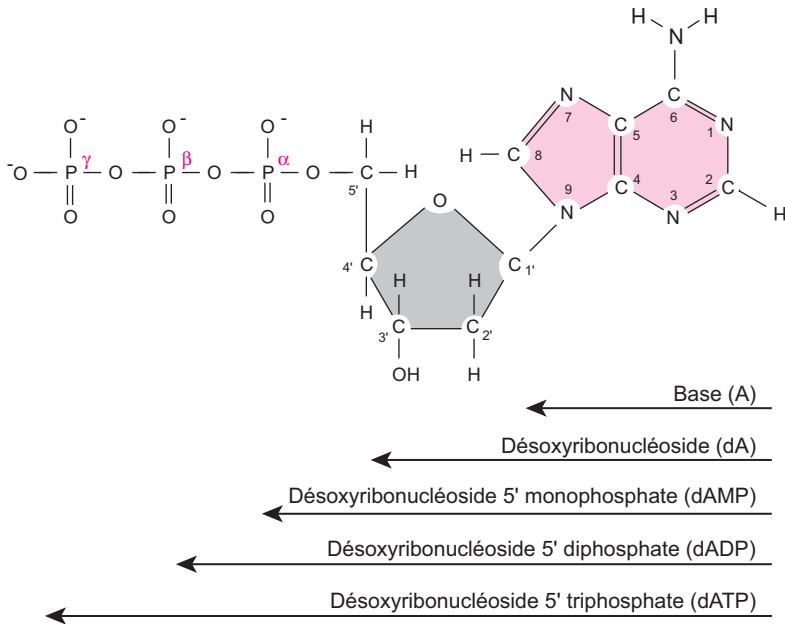


Figure 1-3 Structure détaillée d'un déoxyribonucléoside triphosphate. En remplaçant la lettre A (Adénine) par les lettres G, C, T/U et le déoxyribose par le ribose, la nomenclature est généralisable à tous les nucléotides de l'ADN et de l'ARN.

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 α) ;

- 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).

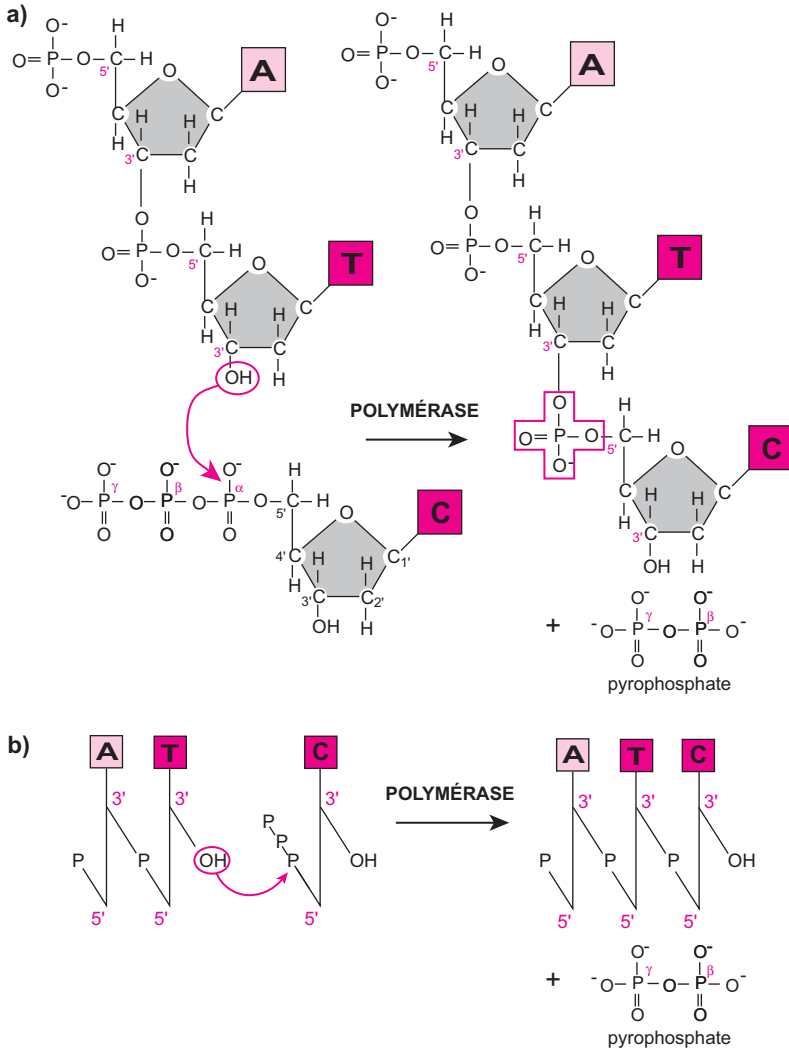


Figure 1-4 Principe de la formation d'un brin d'ADN par la polymérisation des nucléotides. Noter que le brin matrice qui est normalement copié n'est pas ici représenté. **a)** représentation développée ; **b)** représentation simplifiée.

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

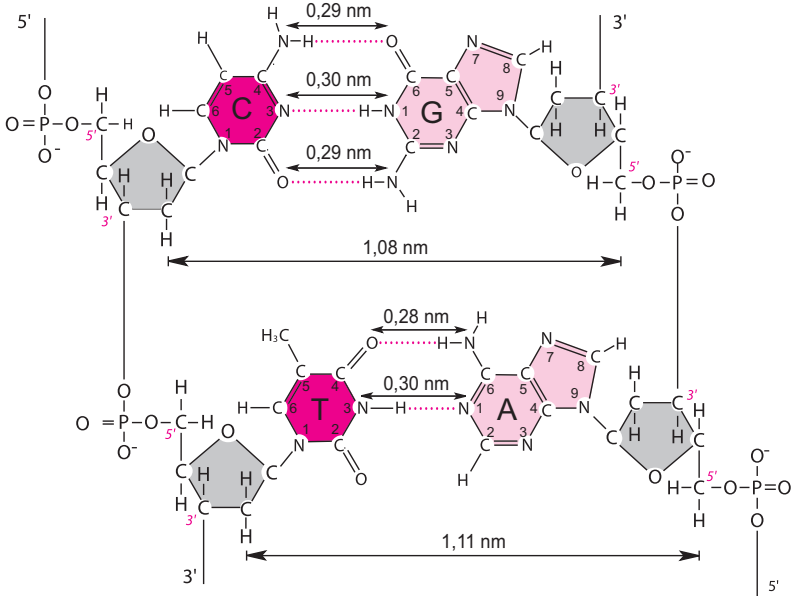


Figure 1-5 Appariements complémentaires des bases G-C et A-T.

TABLEAU 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**.

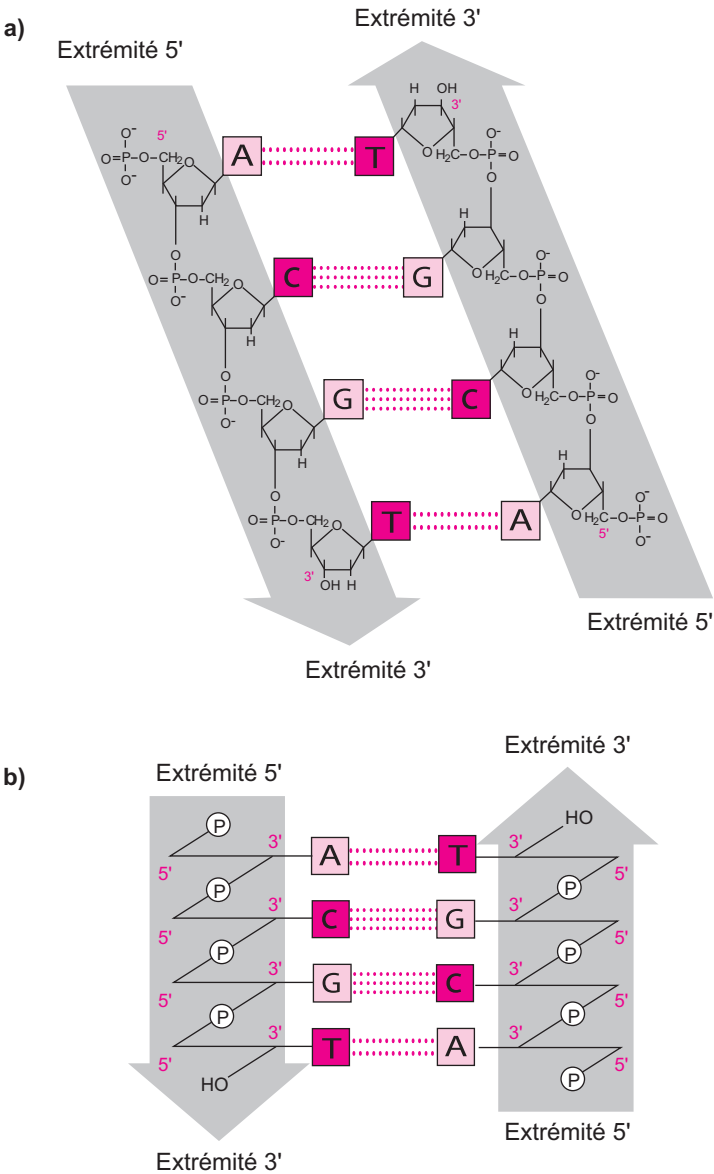


Figure 1-6 Structure des appariements complémentaires de deux brins antiparallèles d'ADN. **a)** Représentation développée ; **b)** représentation simplifiée.

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).

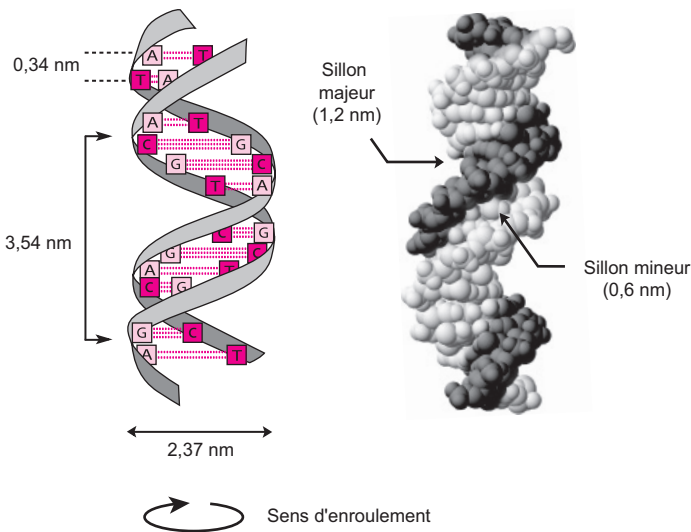


Figure 1-7 Représentation d'un fragment d'ADN en double hélice. À gauche, avec les plateaux de bases ; à droite, les atomes sont représentés avec leurs rayons de Van der Waals.

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.

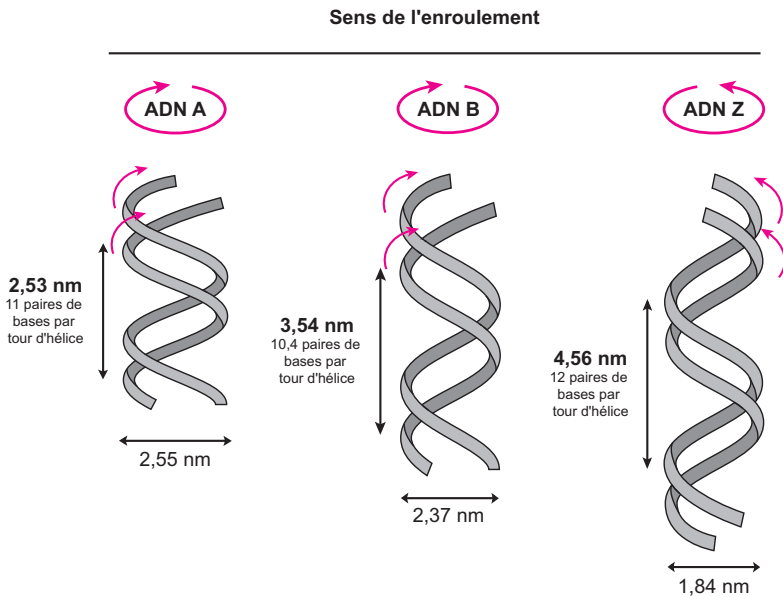


Figure 1-8 Représentation simplifiée des formes A, B et Z de l'ADN.

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).

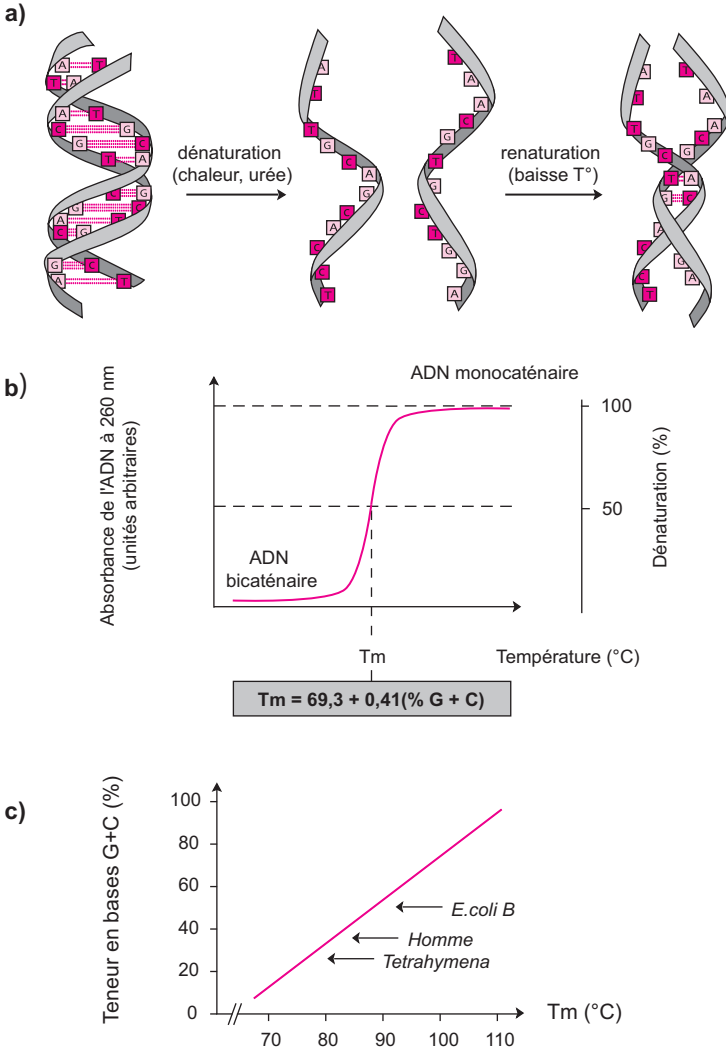


Figure 1-9 La dissociation de l'ADN et sa réassociation dépend de sa teneur en G+C. **a)** Représentation simplifiée de la dissociation-réassociation de l'ADN sous l'action d'agents dénaturants ; **b)** représentation simplifiée de la courbe sigmoïde témoin du processus de dissociation. La formule établit 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 **T_m** (« 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 T_m.



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).

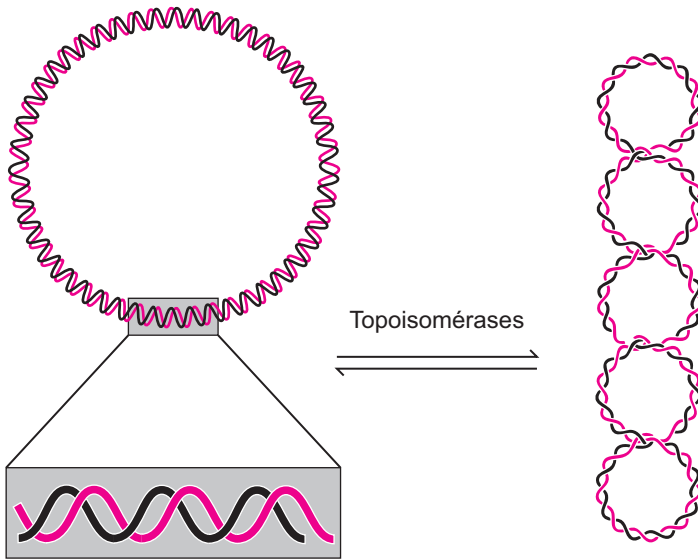


Figure 1-10 Exemple du surenroulement de l'ADN circulaire d'un plasmide bactérien et sa résolution par les topoisomérases.

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- ϕ 174	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.