

BIN702

INTRODUCTION AU CODE GÉNÉTIQUE

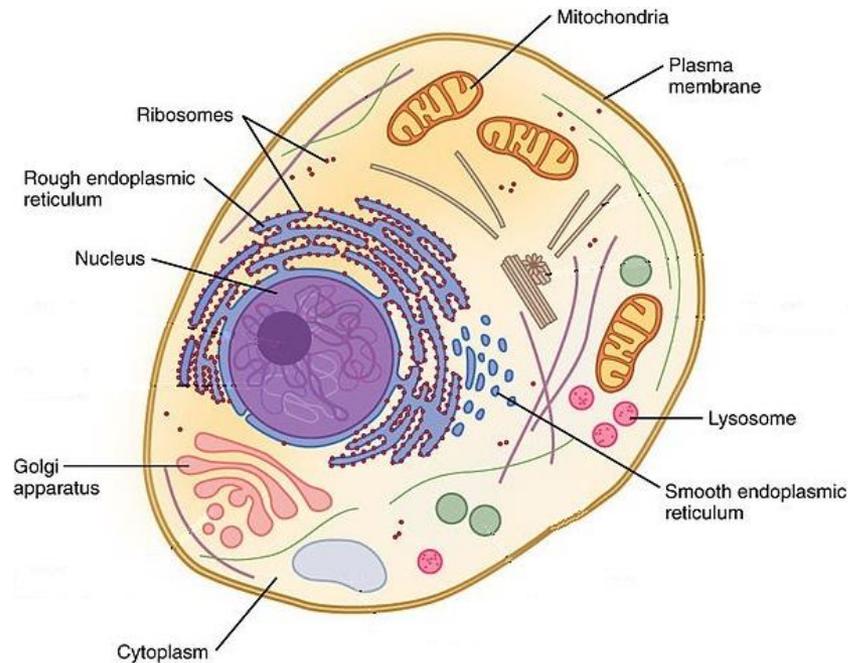
Manuel Lafond



UNIVERSITÉ DE
SHERBROOKE

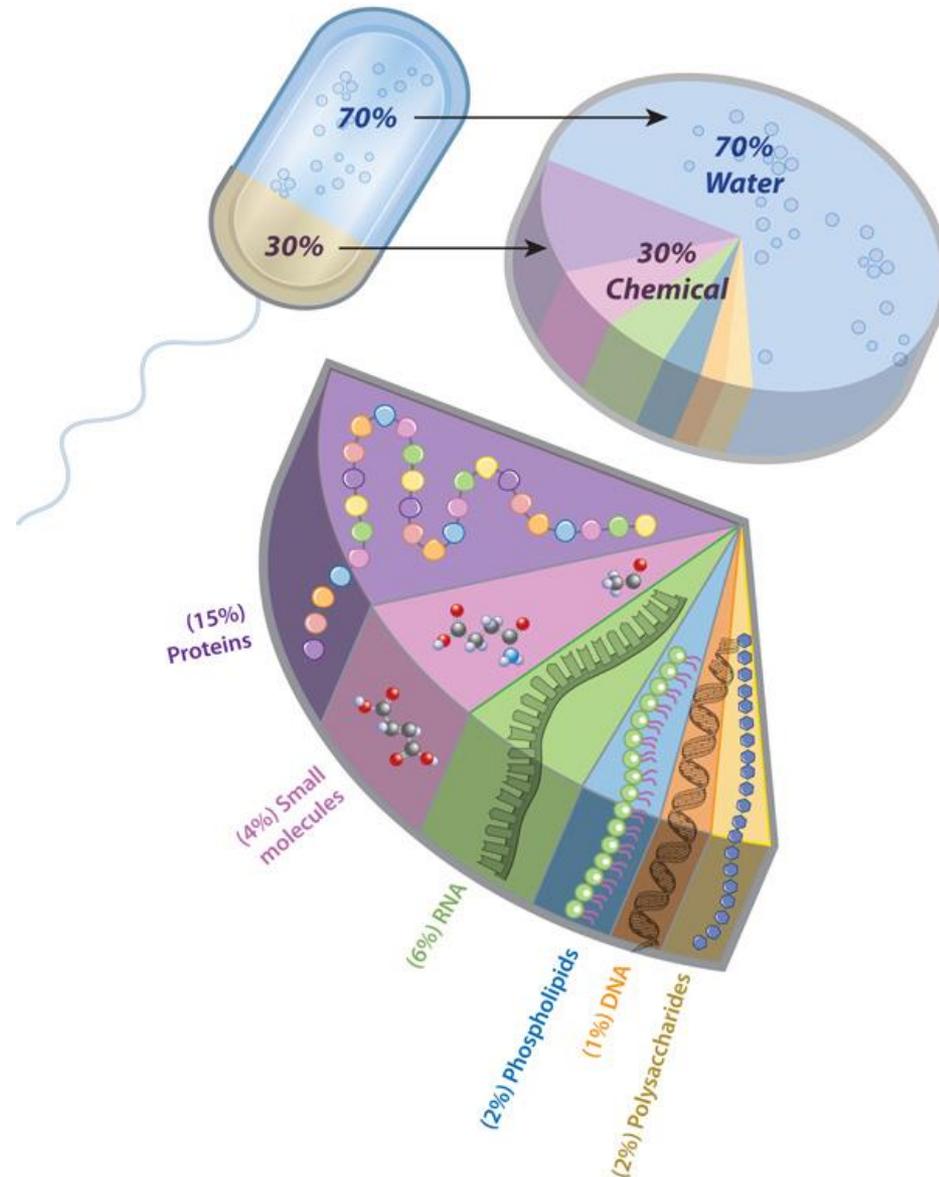
La cellule

- Tous les organismes vivants sont fait de cellules.
- Une cellule naît, se nourrit, se réplique et meurt.

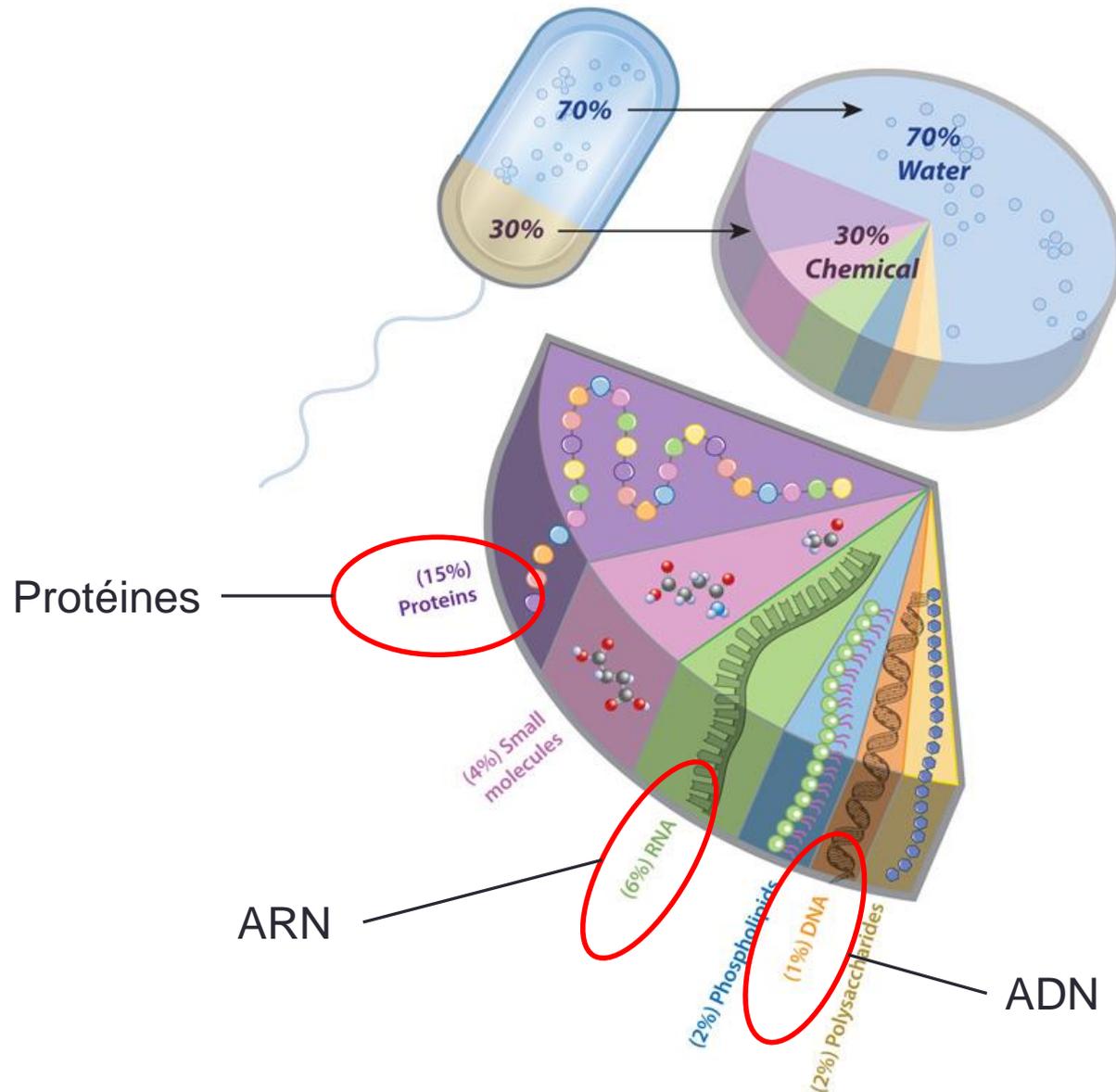


<http://physiologyplus.com/eukaryotic-cells/>

Composition d'une cellule

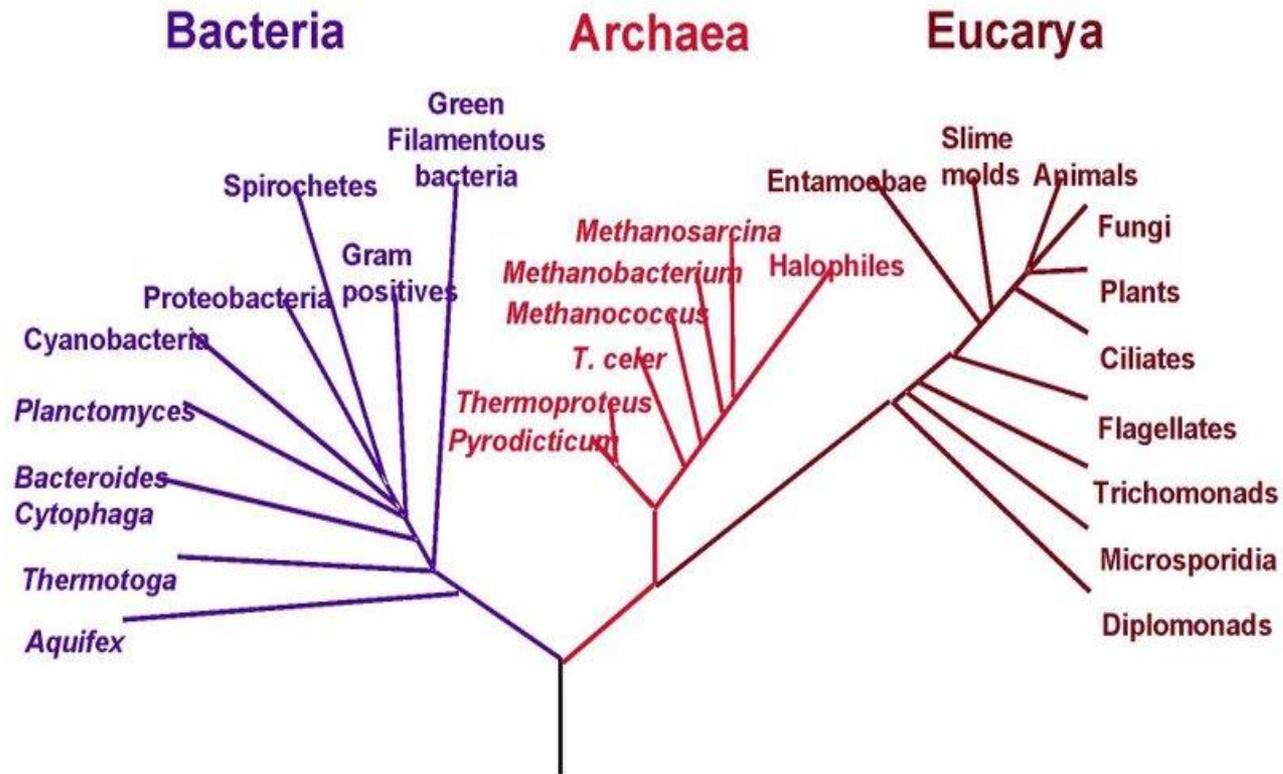


Composition d'une cellule



Les trois domaines du vivant

Phylogenetic Tree of Life

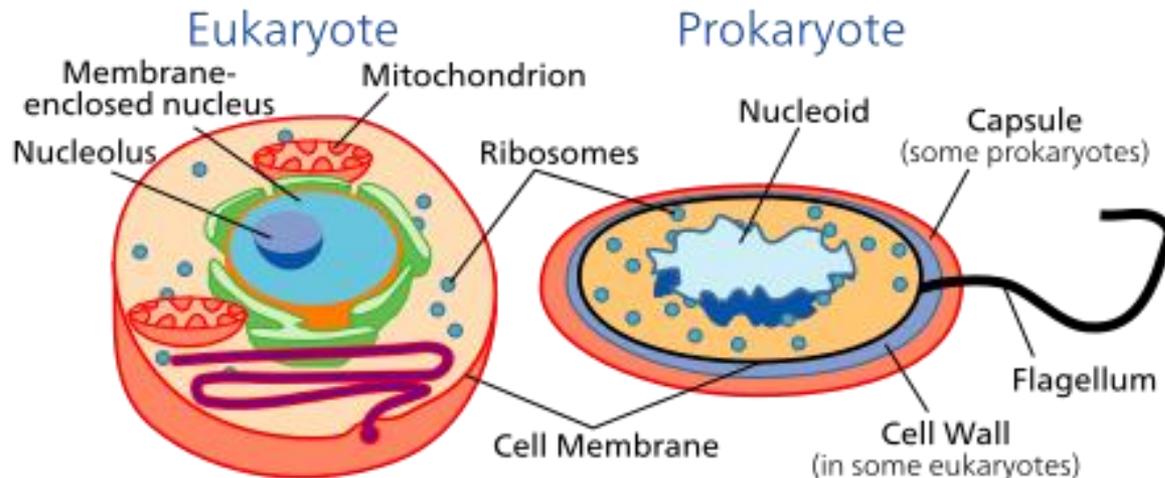


Les 3 domaines du vivant

- **Procaryotes = Bactéries+ Archaea.** Organismes unicellulaires.
- Les bactéries représentent ~ 60% de la biomasse terrestre. Il y a 2-3 milliards d'espèces différentes de bactéries
- **Virus:** Ne fait pas partie des 3 domaines du vivant (après plusieurs débats). Entité biologique nécessitant un hôte pour se répliquer (parasite). Formé d'ADN ou d'ARN, double ou simple brin.

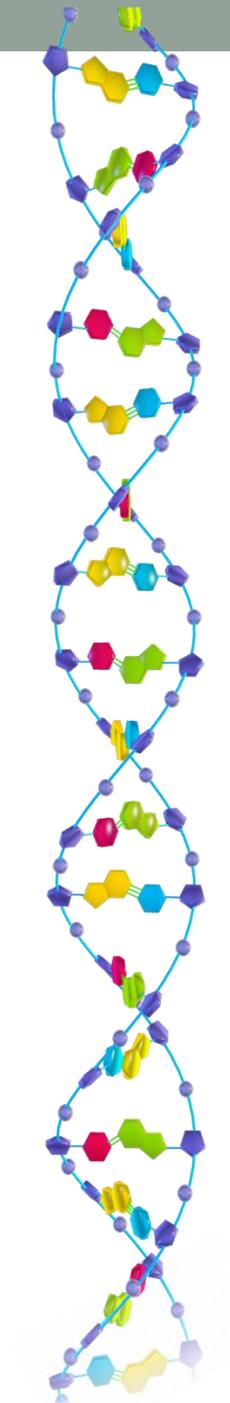
Les 3 domaines du vivant

- **Eucaryote:**
 - Noyau cellulaire protégé par une membrane
 - ADN nucléaire contenu dans le noyau
- **Procaryote**
 - Pas de noyau, pas d'organelles. ADN libre dans la cellule



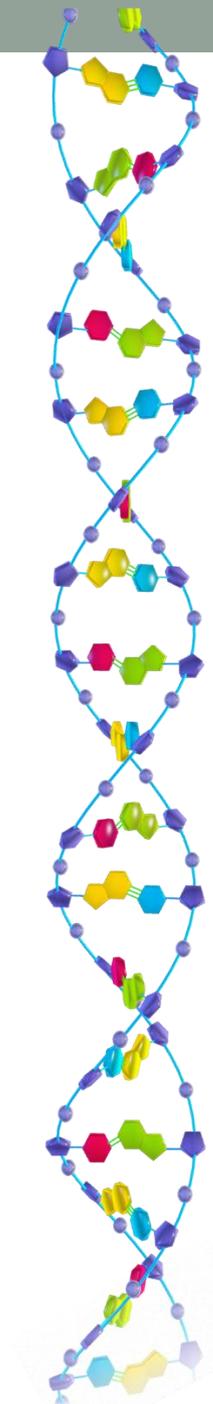
L'ADN

- Toutes les cellules d'un organisme contiennent le même bagage génétique (sauf exceptions).
- L'ADN (**acide désoxyribonucléique**) est le type de molécule utilisée pour **encoder** et **stocker** ce bagage génétique.



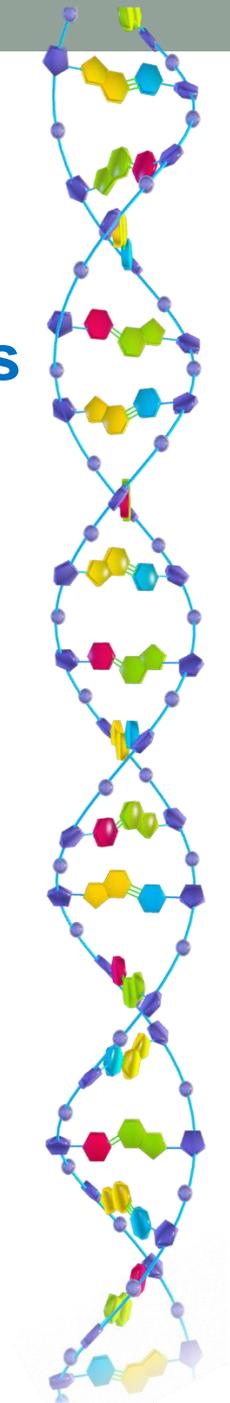
L'ADN

- Toutes les cellules d'un organisme contiennent le même bagage génétique (sauf exceptions).
- L'ADN (**acide désoxyribonucléique**) est le type de molécule utilisée pour **encoder** et **stocker** ce bagage génétique.
- L'ADN est transmis de façon héréditaire.
- Quand l'organisme a besoin d'effectuer une fonction spécifique, il utilise le "segment" d'ADN encodant la fonction voulue.



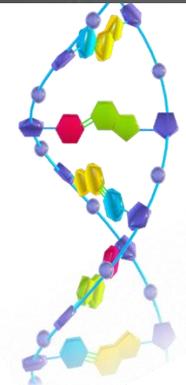
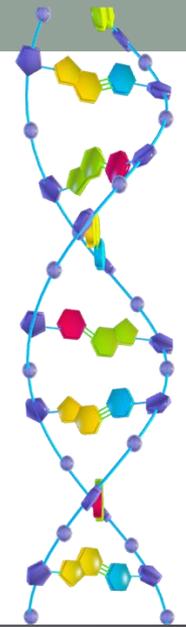
L'ADN

- ADN = chaînes de molécules appelées **nucléotides**
- Chaque molécule est aussi appelée une base.
- ADN de l'homme: ~3.1 milliards de bases
- ADN de la souris: ~2.7 milliards de bases

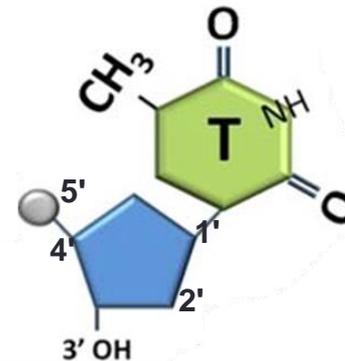
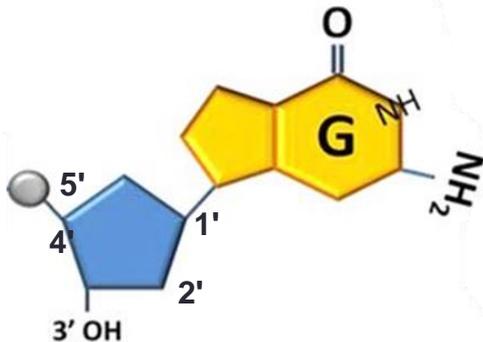
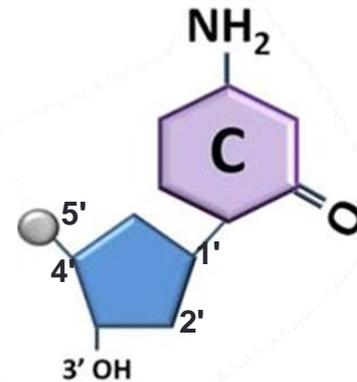
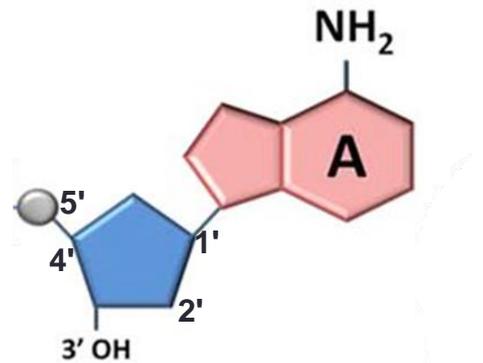


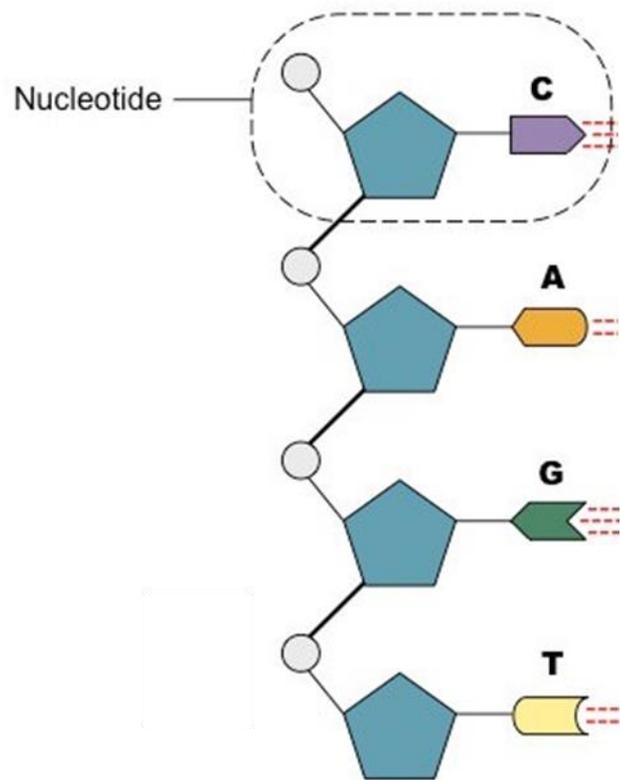
L'ADN

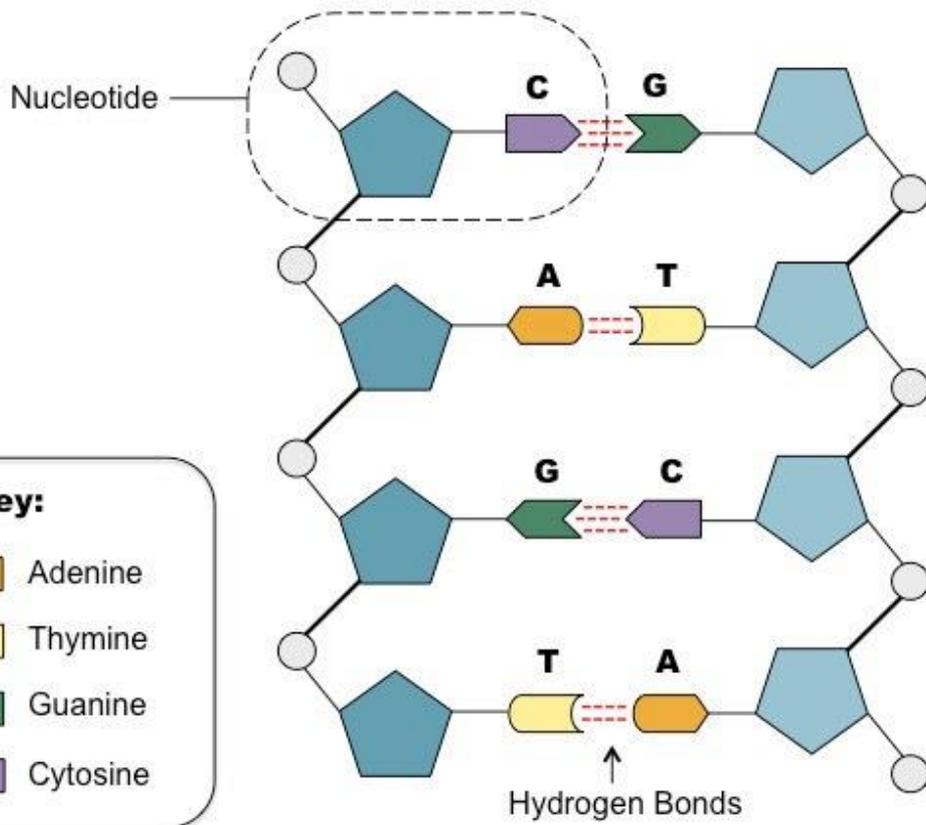
- ADN = chaînes de molécules appelées **nucléotides**
- Chaque molécule est aussi appelée une base.
- ADN de l'homme: ~3.1 milliards de bases
- ADN de la souris: ~2.7 milliards de bases
- Polychaos dubium: ~670 milliards de bases



- On distingue quatre nucléotides faisant partie de l'ADN.
- A (adénine) C (cytosine)
- G (guanine) T (thymine)





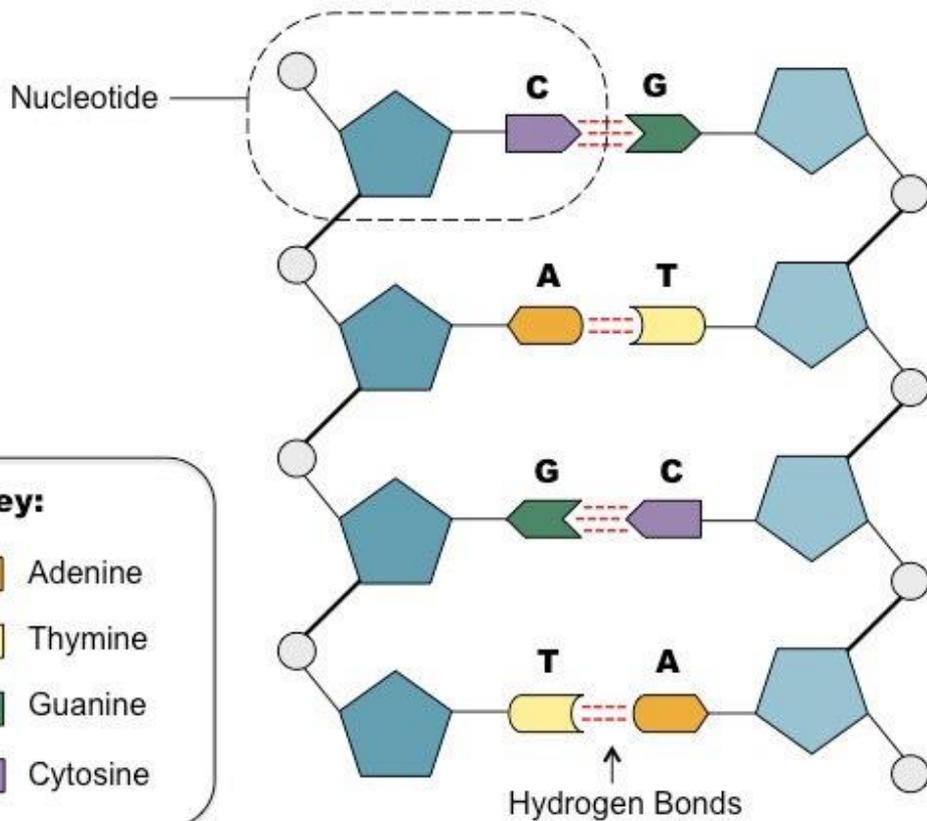


Antiparallel DNA Strands

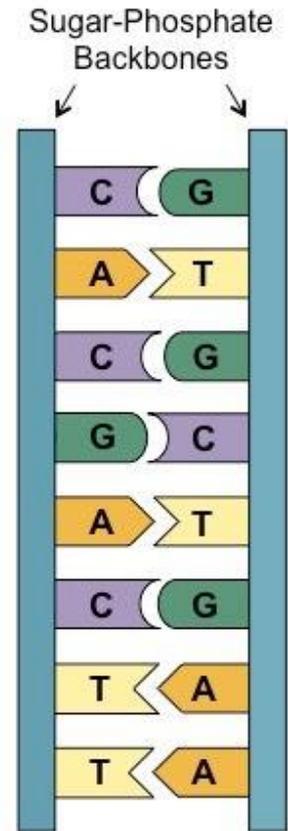
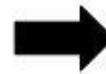
Les bases **s'apparient** avec d'autres.

A avec T
C avec G

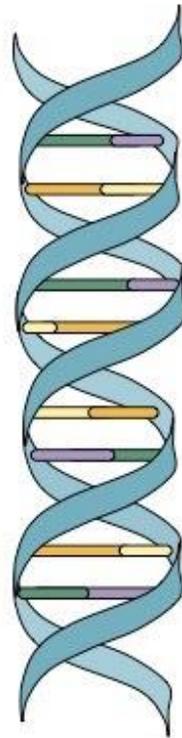
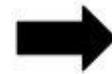
Toujours!



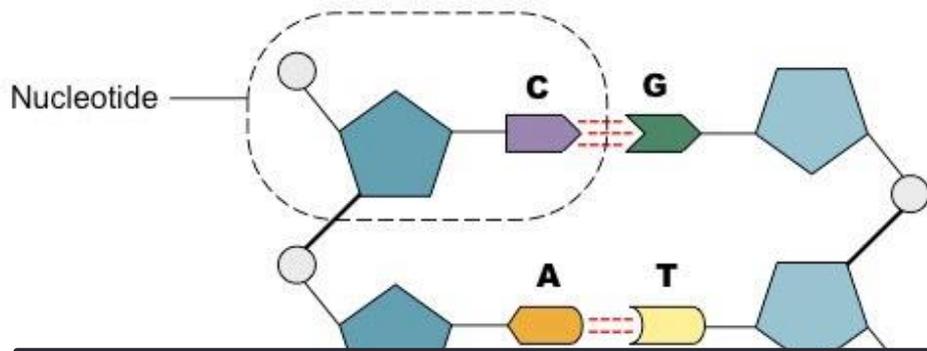
Antiparallel DNA Strands



DNA Ladder

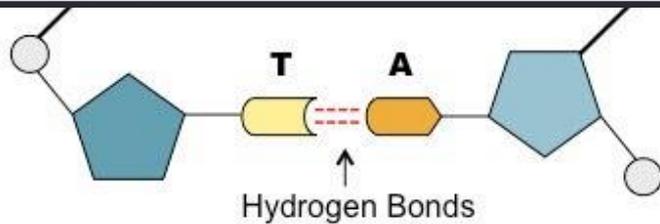


Double Helix

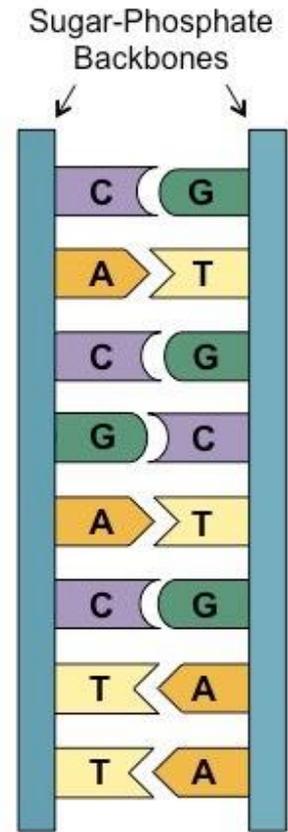


Pour nous: ceci est CACGACTT

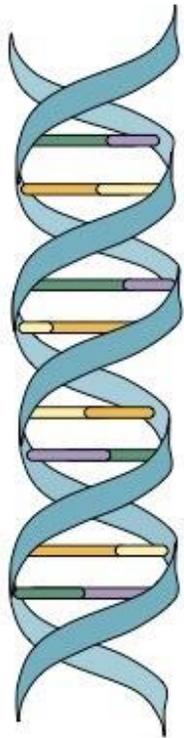
- Ke**
- Thymine
 - Guanine
 - Adenine
 - Cytosine



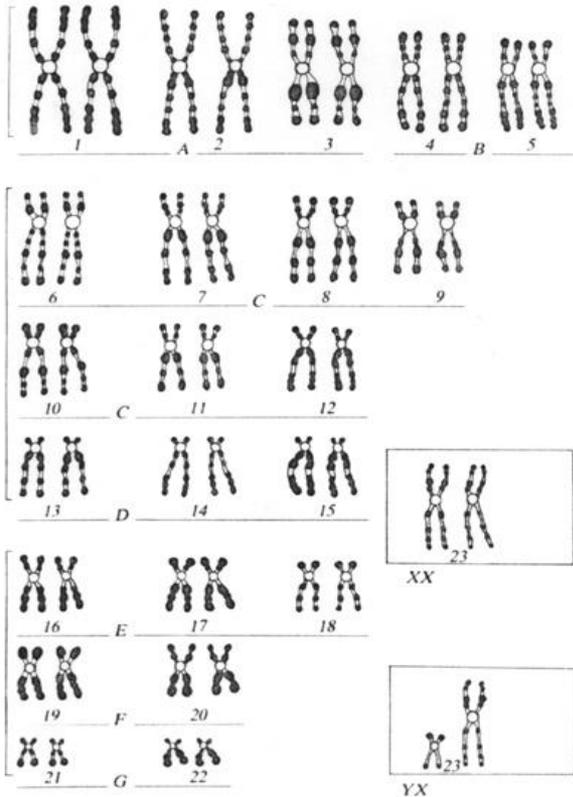
Antiparallel DNA Strands



DNA Ladder

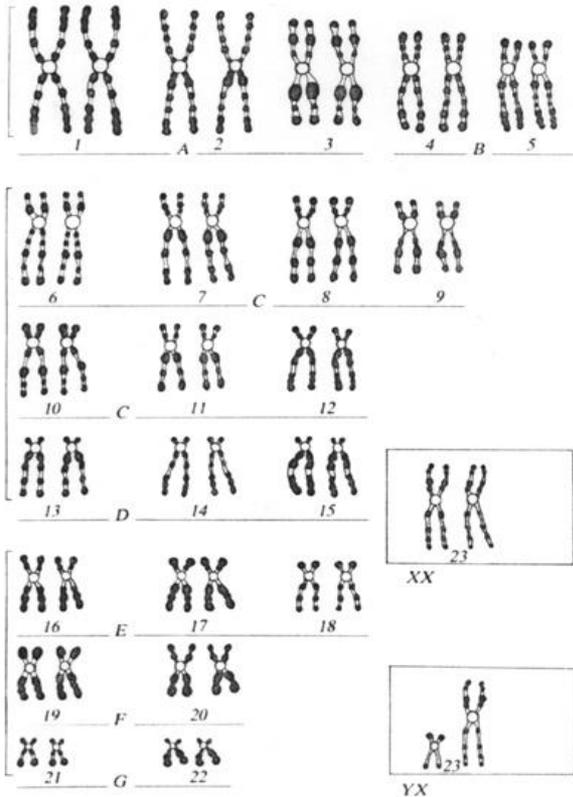


Double Helix



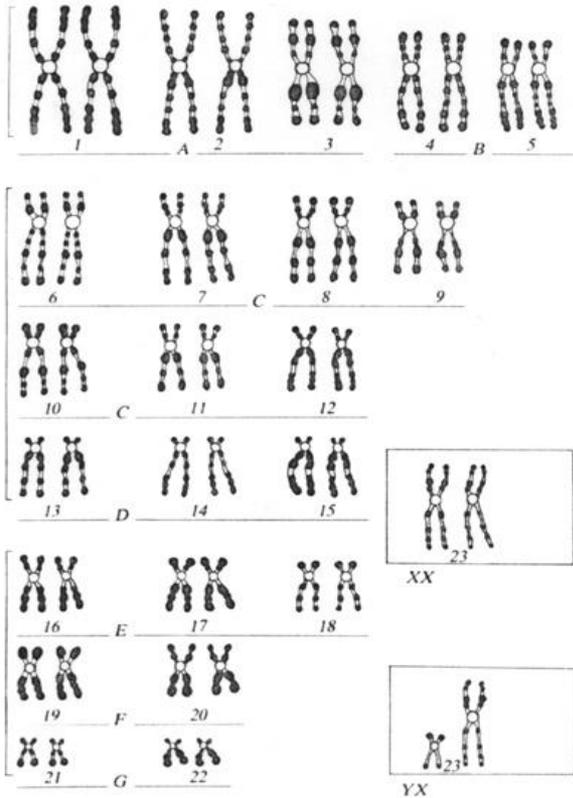
- L'ADN est partitionné en un ou plusieurs **chromosomes**.
- Le **génom**e est l'ensemble des chromosomes.

Karyotype humain: 23 paires de chromosomes homologues. La 23ème paire: XX chez la femme, XY chez l'homme.



Karyotype humain: 23 paires de chromosomes homologues. La 23ème paire: XX chez la femme, XY chez l'homme.

- L'ADN est partitionné en un ou plusieurs **chromosomes**.
- Le **génom**e est l'ensemble des chromosomes.
- Chaque chromosome peut-être *linéaire* ou *circulaire*.
 - Procaryote = circulaire
 - Eucaryote = linéaire
 - (sauf exception)

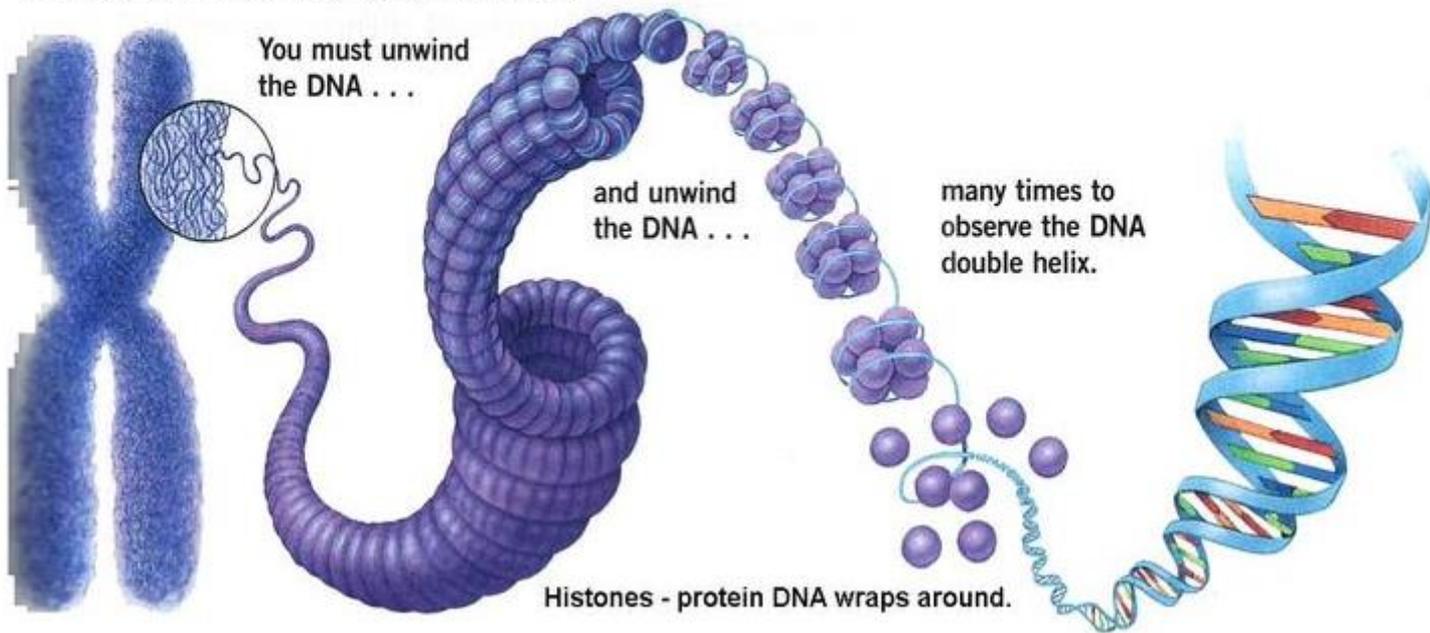


Karyotype humain: 23 paires de chromosomes homologues. La 23ème paire: XX chez la femme, XY chez l'homme.

- L'ADN est partitionné en un ou plusieurs **chromosomes**.
- Le **génom**e est l'ensemble des chromosomes.
- Chaque chromosome peut-être *linéaire* ou *circulaire*.
 - Procaryote = circulaire
 - Eucaryote = linéaire
 - (sauf exception)
- **Cellule diploïde**: contient deux copies « homologues » de chaque chromosome: une copie provenant de chaque parent.

Chromosome Structure

Chromosomes contain very tightly wound DNA



Passage de l'ADN chez les Eucaryotes

- L'ADN est hérité des parents (ou du parent chez les asexués).
- Chez les diploïdes, un organisme reçoit deux copies de chaque chromosome: une du père, une de la mère.
- La **meiose** est l'étape de processus de reproduction qui "choisit" le contenu génétique à envoyer chez l'enfant.

Erreurs de réplication

- Pendant la copie de l'ADN, des erreurs peuvent survenir, introduisant ainsi des **mutations**.

Erreurs de réplication

- Pendant la copie de l'ADN, des erreurs peuvent survenir, introduisant ainsi des **mutations**.

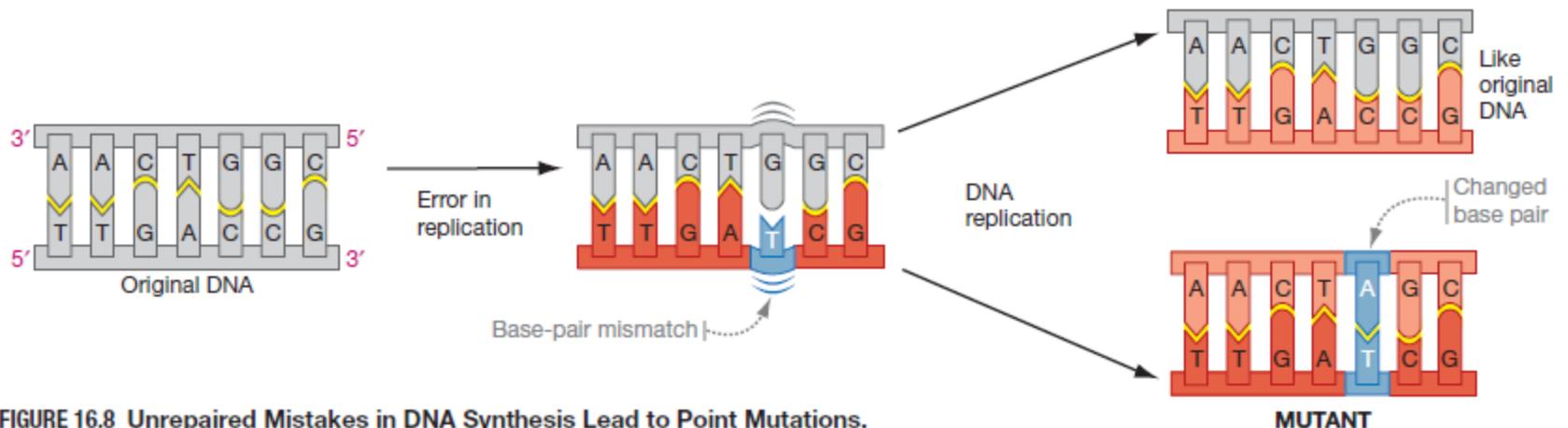
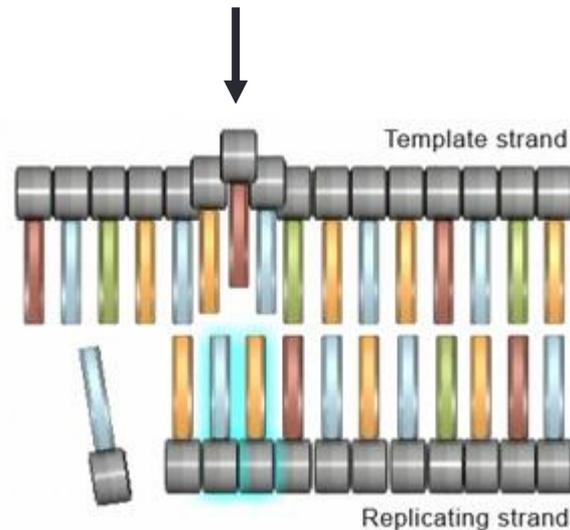


FIGURE 16.8 Unrepaired Mistakes in DNA Synthesis Lead to Point Mutations.

Erreurs de réplication

- Pendant la copie de l'ADN, des erreurs peuvent survenir, introduisant ainsi des **mutations**.
- Ici: une **deletion** (suppression en bon français) de nucléotide dans le brin de DNA copié.



À retenir

- Lorsque l'ADN est répliqué, il peut être modifié de plusieurs façons, ce qui mène à la diversité biologique.
- Recombinaison génétique
- Erreurs de réplication
 - Mutation
 - Insertion
 - Deletion
- Les erreurs peuvent parfois affecter de grand blocs d'ADN
 - Duplication segmentale, inversion de bloc, ...

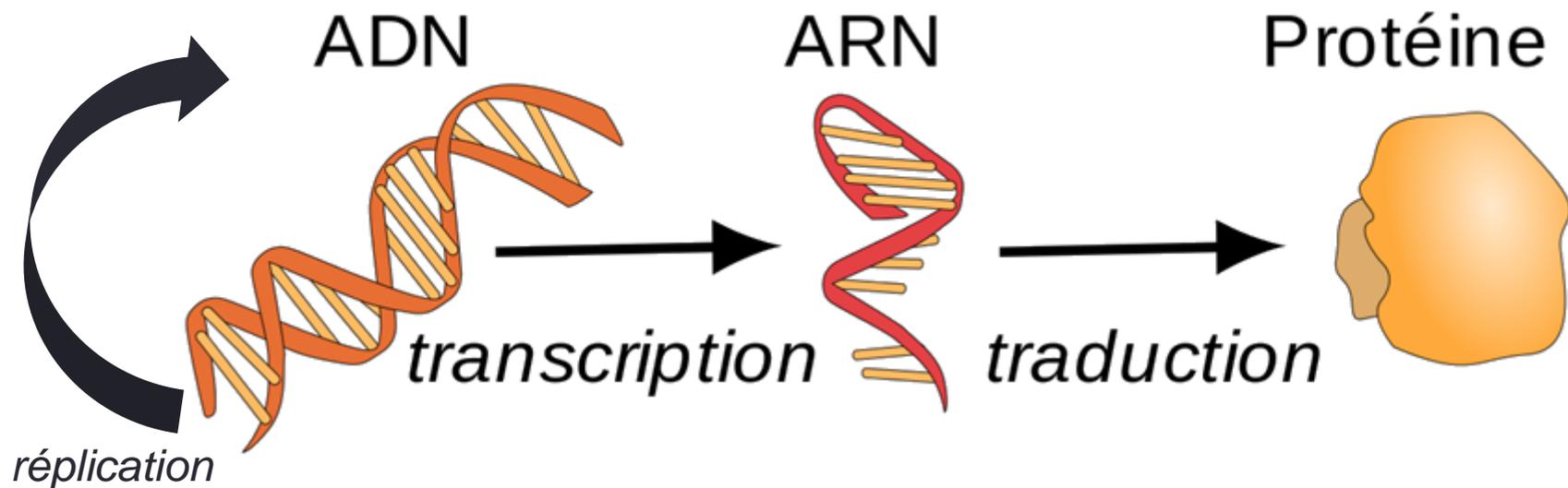
Mais à quoi sert l'ADN?

- Jusqu'à ici, on a vu comment l'ADN est stocké et transmis.
- Mais en quoi permet-il à l'organisme de fonctionner?

Mais à quoi sert l'ADN?

- Jusqu'à ici, on a vu comment l'ADN est stocké et transmis.
- Mais en quoi permet-il à l'organisme de fonctionner?
 - Il permet de "générer" de l'**ARN** et des **protéines**.

Dogme central de la biologie moléculaire



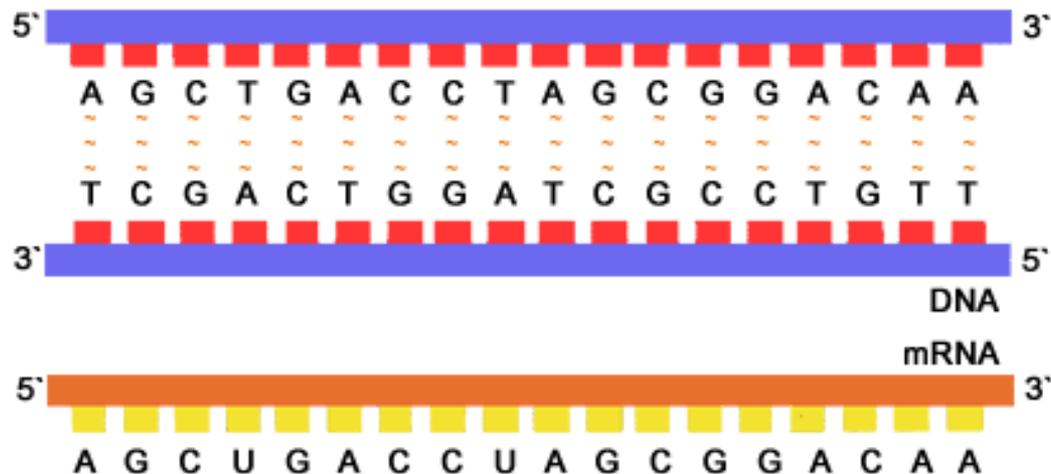
L'ARN

(RNA en anglais)

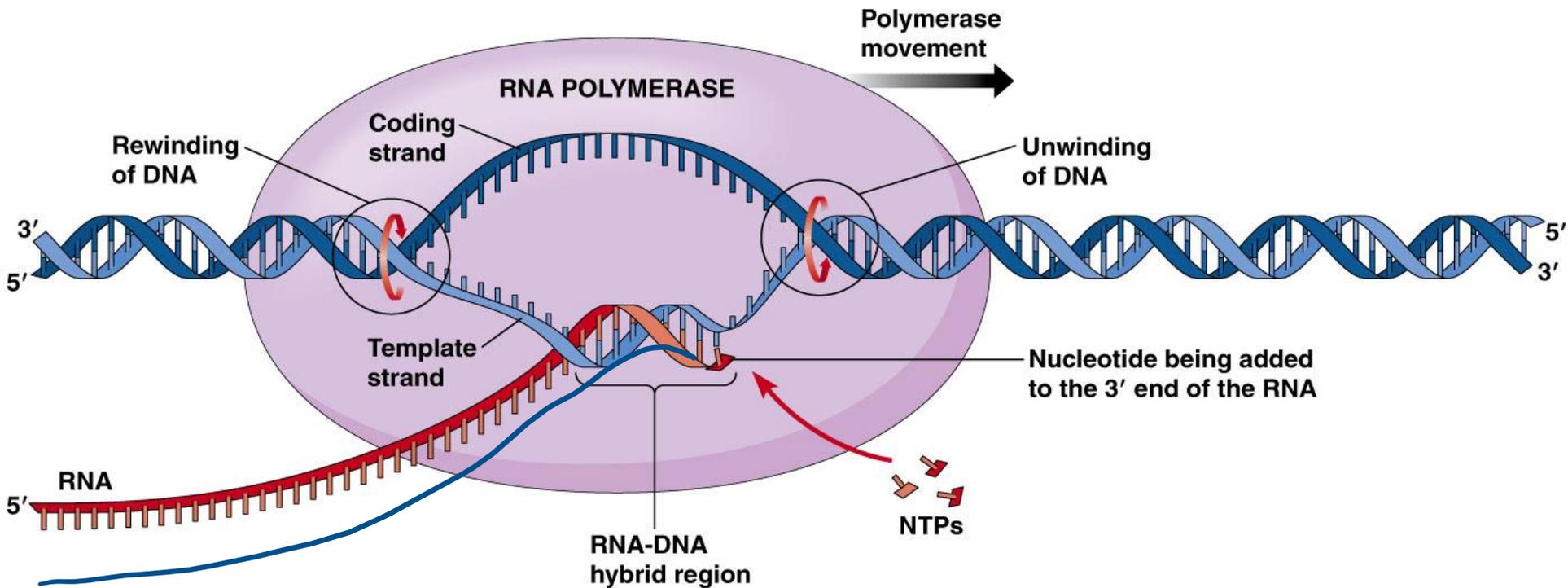
- Acide ribonucléique
- Rôle plus **actif** que l'ADN
 - (synthèse de protéines, régulation de l'expression des gènes, épissage, ...)
- 3 types principaux
 - ARNm (ARN messenger): résultat d'une "lecture" d'un segment d'ADN
 - ARNt (ARN de transfert): convertit un triplet d'ARNm en acide aminé
 - ARNr (ARN ribosomique): machinerie de traduction

Structure de l'ARN

- Structure primaire: séquence linéaire de 4 nucléotides possibles:
- A(dénine), C(ytosine), G(uanine), U(racile)
- Un gène est **transcrit** en une séquence primaire d'ARN par une copie effectuée dans le sens 5' - 3'.
- Pas de complémentarité dans l'ARN, et T devient U.



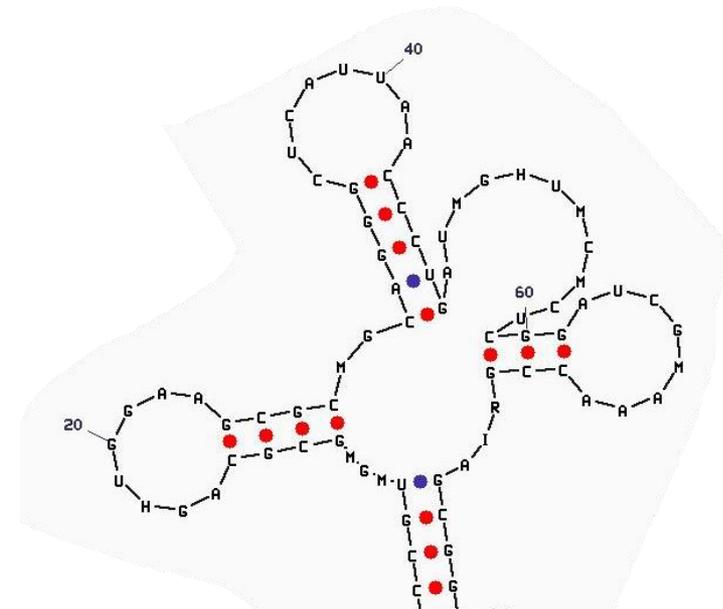
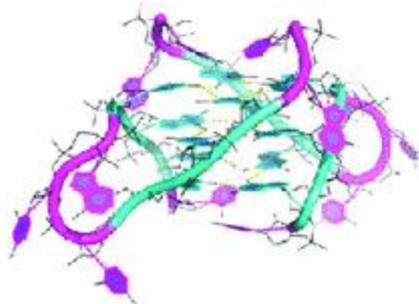
Transcription (ADN => ARN)



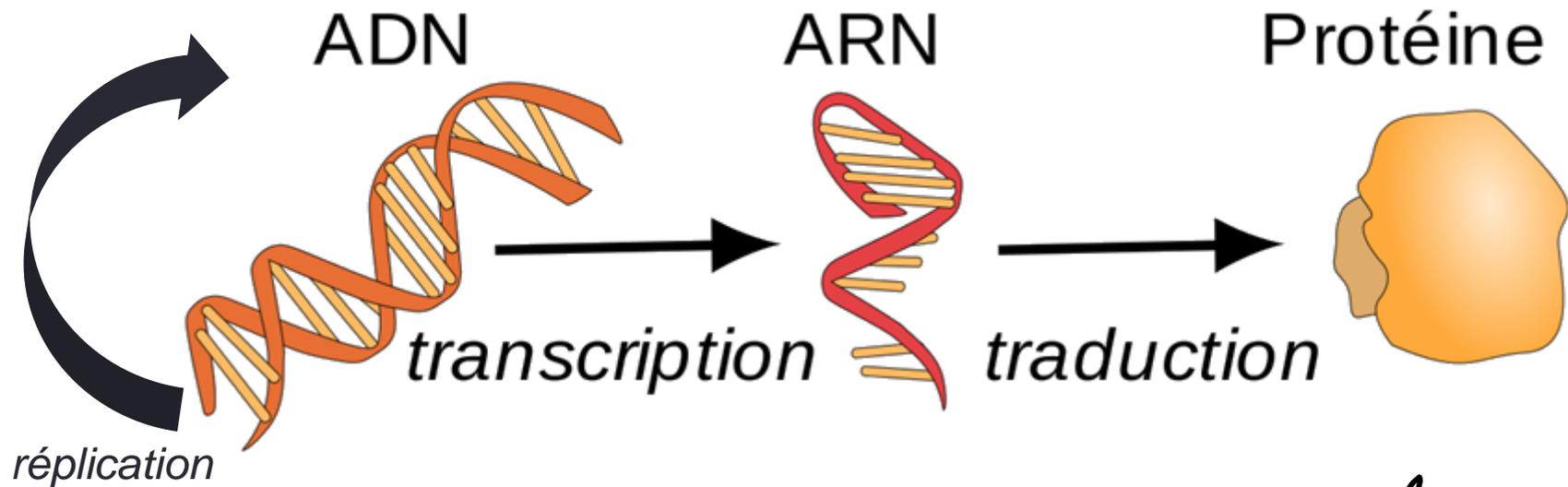
© 2012 Pearson Education, Inc.

Structure de l'ARN

- Pas de double-hélice.
- Repliement selon les appariements qui dépendent des nucléotides.
- Simplification en représentation 2D appelée la structure secondaire (plus tard dans le cours, nous allons prédire cette structure).



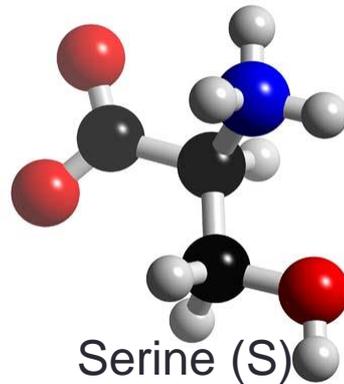
Dogme central de la biologie moléculaire



Voir vidéo

ARN => protéine

- L'ARNm (message) est le type d'ARN qui sera **traduit** en protéine.
- Chaque triplet de nucléotides est converti en ce qu'on appelle un **acide aminé** (AA).
- L'organisme utilise 20 acides aminés différents.



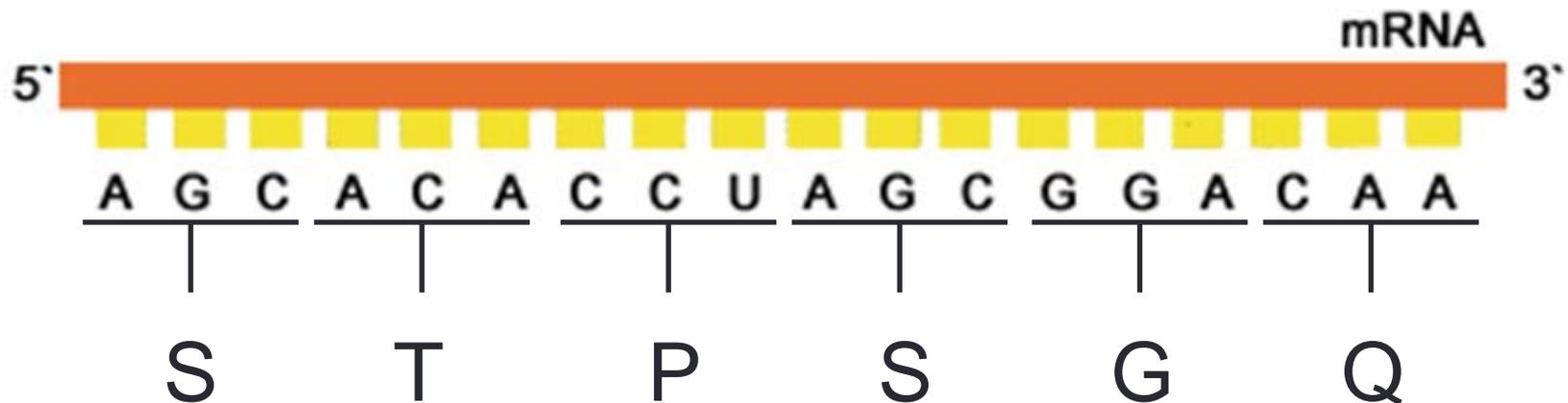
ARN => protéine

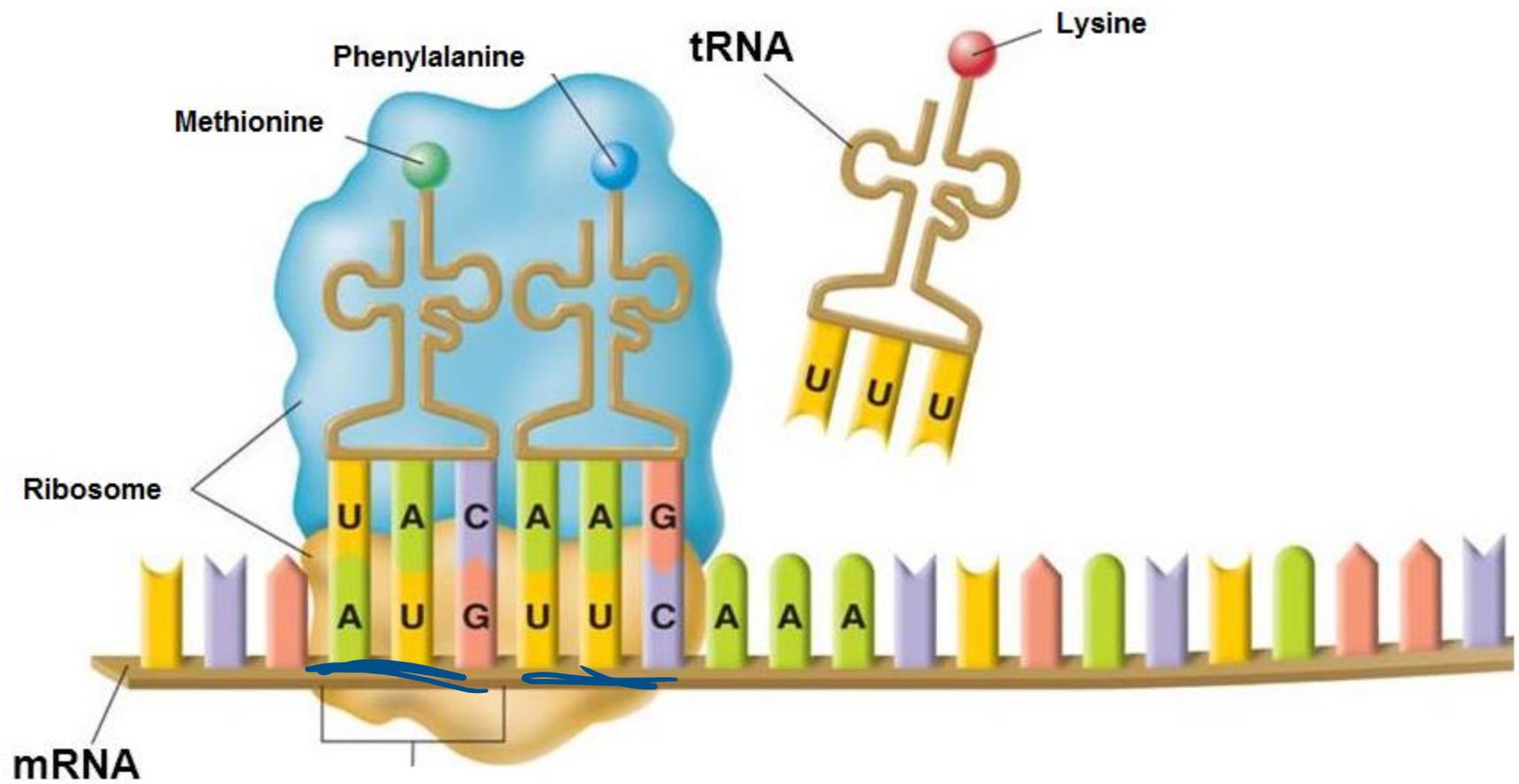
- L'ARNm (message) est le type d'ARN qui sera **traduit** en protéine.
- Chaque triplet de nucléotides est converti en ce qu'on appelle un **acide aminé** (AA).
- L'organisme utilise 20 acides aminés différents.



ARN => protéine

- L'ARNm (message) est le type d'ARN qui sera **traduit** en protéine.
- Chaque triplet de nucléotides est converti en ce qu'on appelle un **acide aminé** (AA).
- L'organisme utilise 20 acides aminés différents.





ARN => protéine

- Chaque triplet de nucléotides est **traduit** en ce qu'on appelle un **acide aminé** (AA).
- L'organisme utilise 20 acides aminés différents.
- Un triplet est appelé un **codon**.
- Les ARNt (de transfert) "apportent" les AA correspondant aux codons.
- Le **ribosome** est le nom de la "machine" effectuant la conversion.

Code génétique

- 3 nucléotides par codon, 4 nucléotides possibles
- $4^3 = 64$ acides aminés possibles
- Il y a 20 AA => plusieurs codons produisent le même AA
 - (principe du pigeonnier)

le code génétique

		Deuxième lettre								ijk
		U		C		A		G		
Première lettre (côté 5')	U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
		UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C
		UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop	A
		UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp	G
	C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
		CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	C
		CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
		CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
	A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
		AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
		AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
		AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
	G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
		GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
		GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
		GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G
		codon d'initiation				codon de terminaison				

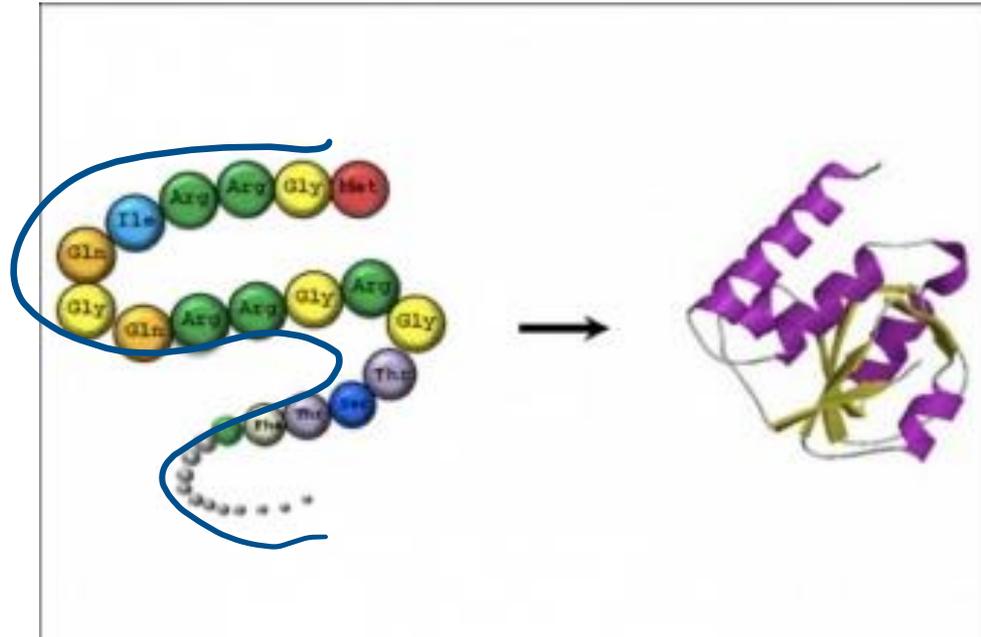
Troisième lettre (côté 3')

ARN => protéine

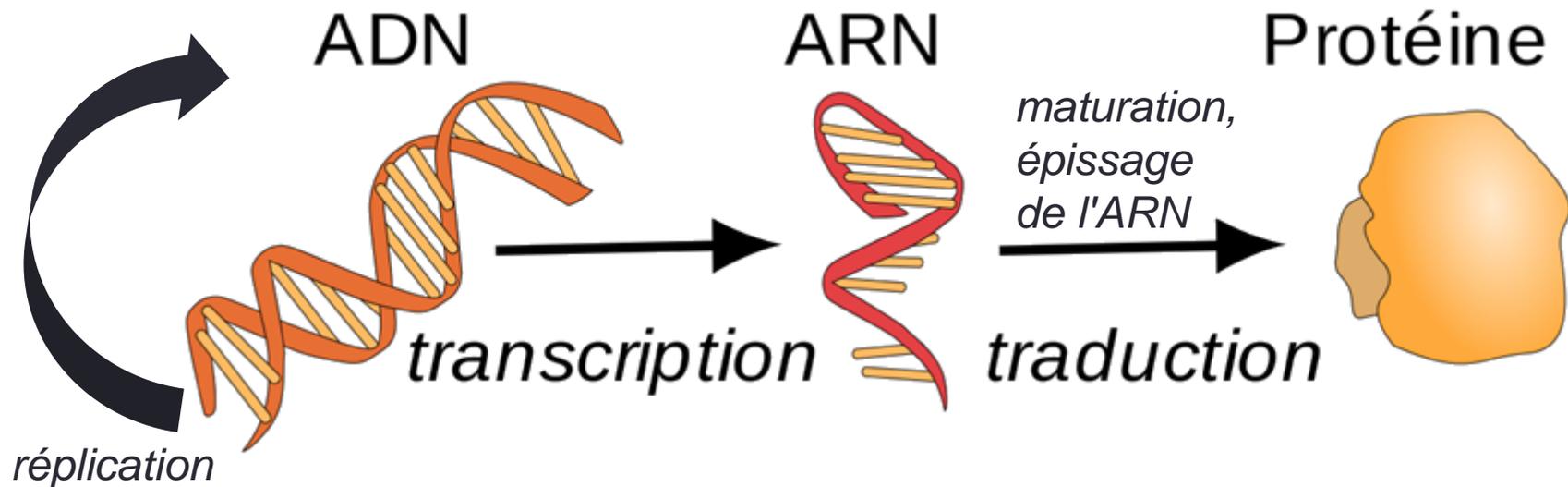
- **Codon start** : codon spécial indiquant au ribosome de débuter la traduction.
 - Presque toujours AUG (Méthionine)
- **Codon stop** : codon spécial indiquant l'arrêt de la traduction.
 - UAA, UAG, UGA

Les protéines

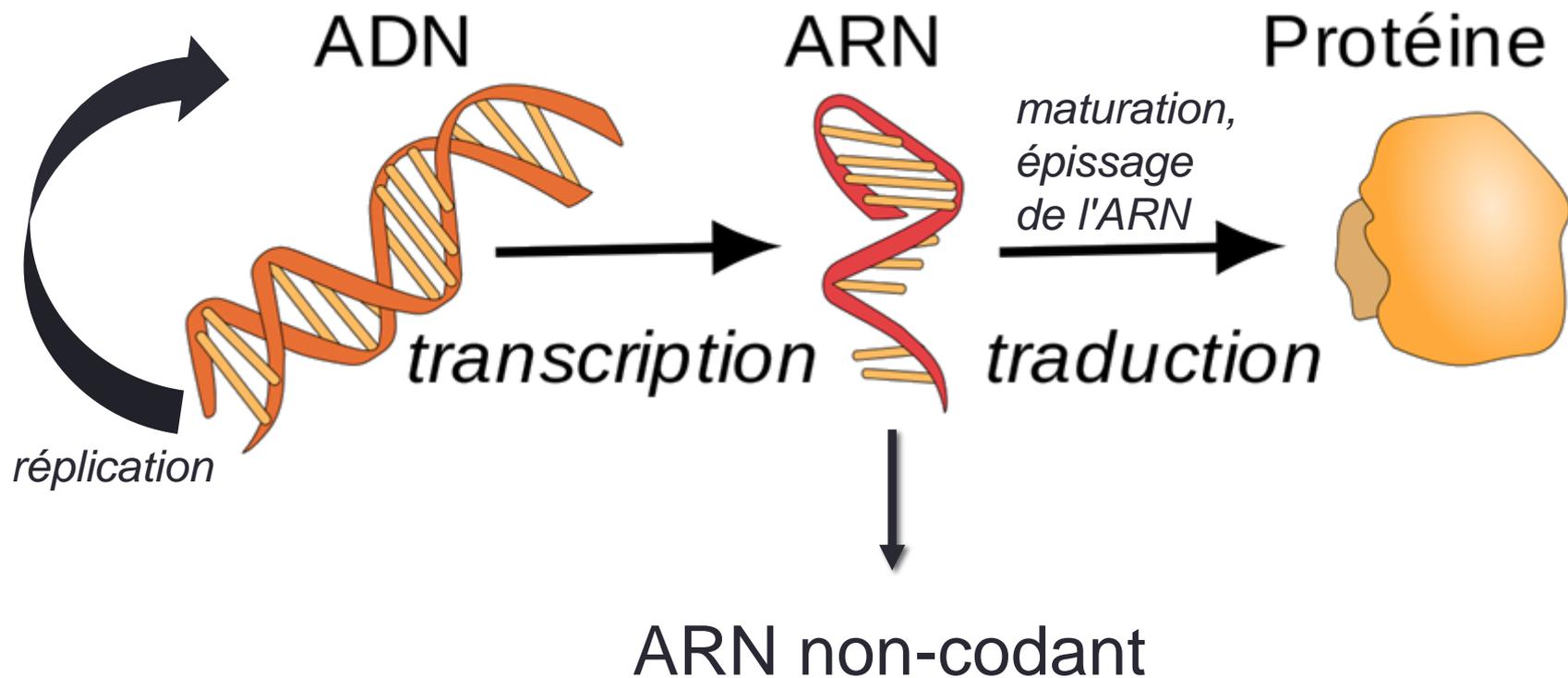
- Responsables de la plupart des fonctions d'une cellule
 - ex: digestion des nutriments, anticorps, hormones, ...
- Pour nous, ce sont des chaînes de caractères sur un alphabet de 20 lettres (les acides aminés).
- Longueur typique au alentour de 300 AA, étendue: 100-5000 AA



Dogme central de la biologie moléculaire



Dogme central de la biologie moléculaire

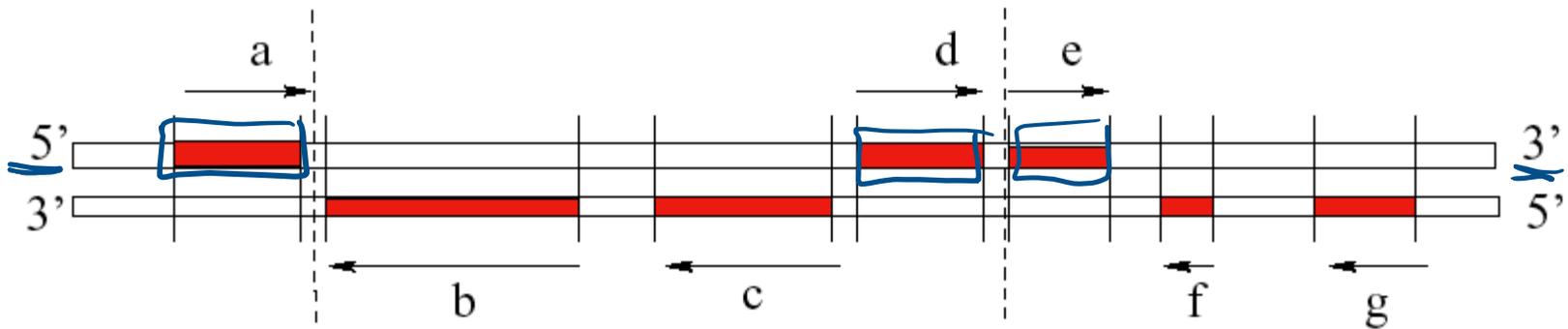


Quelques concepts restant à voir

- Qu'est-ce qu'un gène?
- Introns et exons
- Épissage et régions non-traduites (UTR)
- Cadres de lecture (open reading frame, ORF)

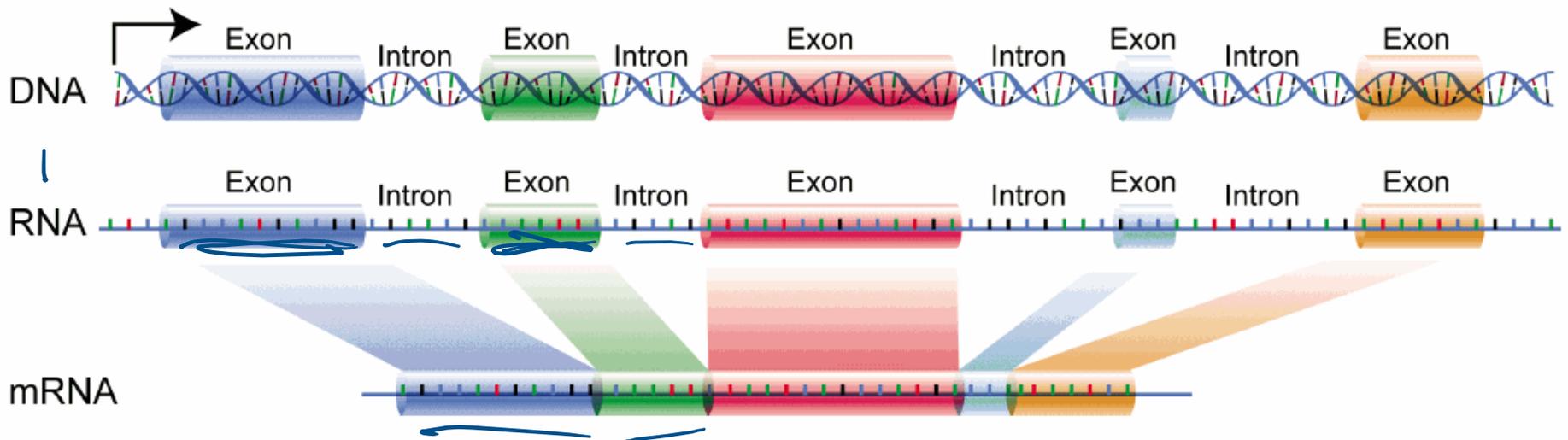
Qu'est-ce qu'un gène

- Segment d'ADN qui "sert à quelque chose"
 - Souvent: qui peut être exprimé en ARN.
 - Gènes les plus étudiés: ceux qui deviennent éventuellement des protéines.
- Pour nous: un segment d'ADN sur le génome.



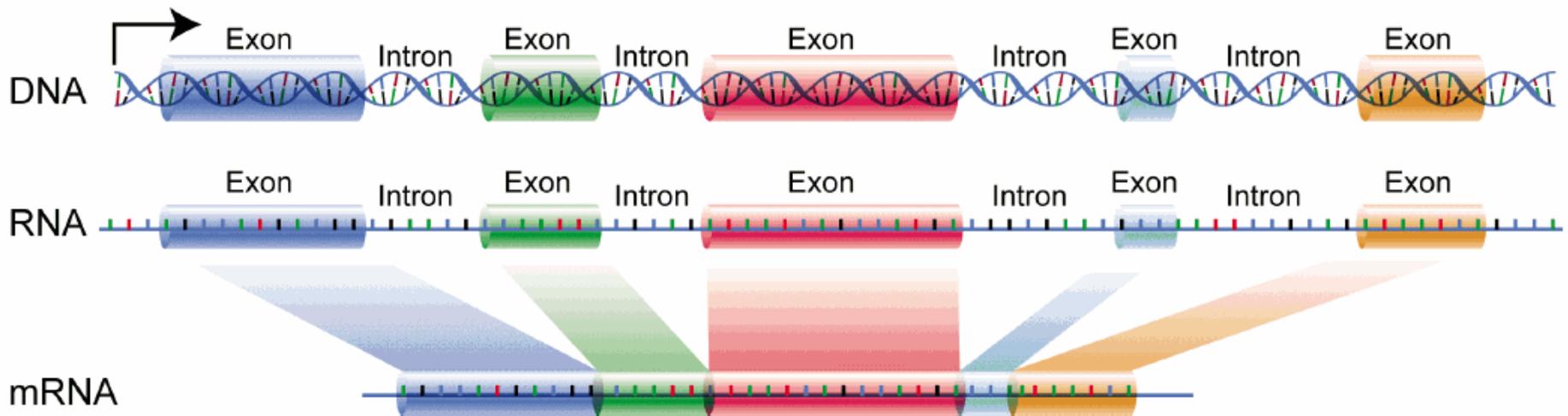
Épissage

- Entre la transcription et la traduction, certains segments peuvent être "coupés". On dit qu'ils sont **épissés**.



Épissage

- Entre la transcription et la traduction, certains segments peuvent être "coupés". On dit qu'ils sont **épissés**.
- **Exons**: segments retenus
 - (ex pour "exit" le noyau, ou encore ex pour "exprimé")
- **Introns**: segments épissés.
 - (*in* pour "inside" le noyau, *intragénique*, ou *intrus*)



Cadres de lecture

- La suite de codons dépend du point de départ de la transcription.
- 3 **cadres de lecture** possibles (en anglais: *reading frames*).
- **Cadre de lecture ouvert**: portion d'un cadre de lecture susceptible d'être traduit. Début START et fin STOP. (en anglais: *open reading frame, ou ORF*)
- Ne considère pas l'épissage ou les autres événements induisant un *décalage* de lecture.

Cadres de lecture (ouverts) pour

ATGCAATGGGGAAATGTTACCAGGTCCGAACTTATTGAGGTAAGACAGATTTAA

1. **ATG** CAA TGG GGA AAT GTT ACC AGG TCC GAA CTT ATT GAG GTA AGA CAG ATT **TAA**
2. ~~A~~ TGC AAT GGG GAA **ATG** TTA CCA GGT CCG AAC TTA TTG AGG **TAA** GAC AGA TTT AA
3. ~~A~~ GCA **ATG** GGG AAA TGT TAC CAG GTC CGA ACT TAT **TGA** GGT AAG ACA GAT TTA A

- Super vidéo qui explique la transcription:
- https://www.youtube.com/watch?v=8M198uHJd_8

- Autre vidéo
- <https://www.youtube.com/watch?v=gG7uCskUOrA>

Quelques lectures complémentaires

- Problème de l'héritabilité manquante
- https://en.wikipedia.org/wiki/Missing_heritability_problem

- Les ARN ne font pas que coder pour des protéines. Les ARN non-codants ont d'autres rôles.
- https://en.wikipedia.org/wiki/Non-coding_RNA
 - ex: les Riboswitches, <https://en.wikipedia.org/wiki/Riboswitch>

- Traduction non-AUG
- <http://genesdev.cshlp.org/content/31/17/1717.full>

- La vie des introns après l'épissage
- <https://onlinelibrary.wiley.com/doi/abs/10.1002/wrna.1187>