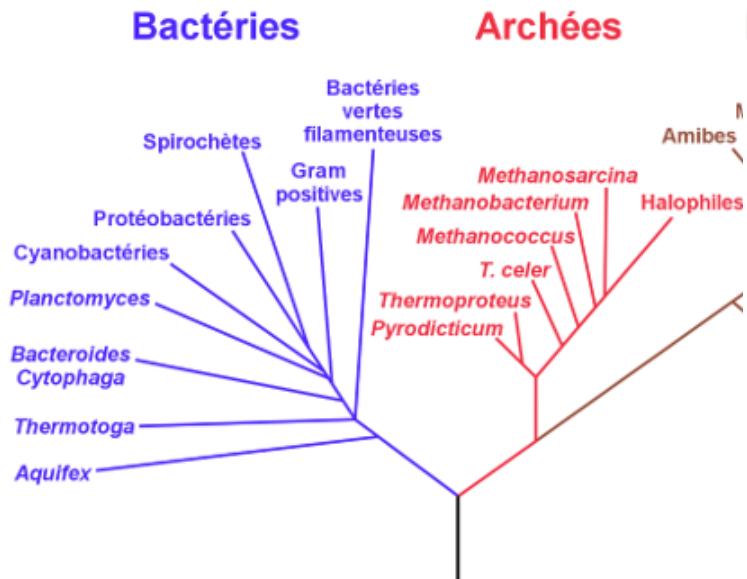


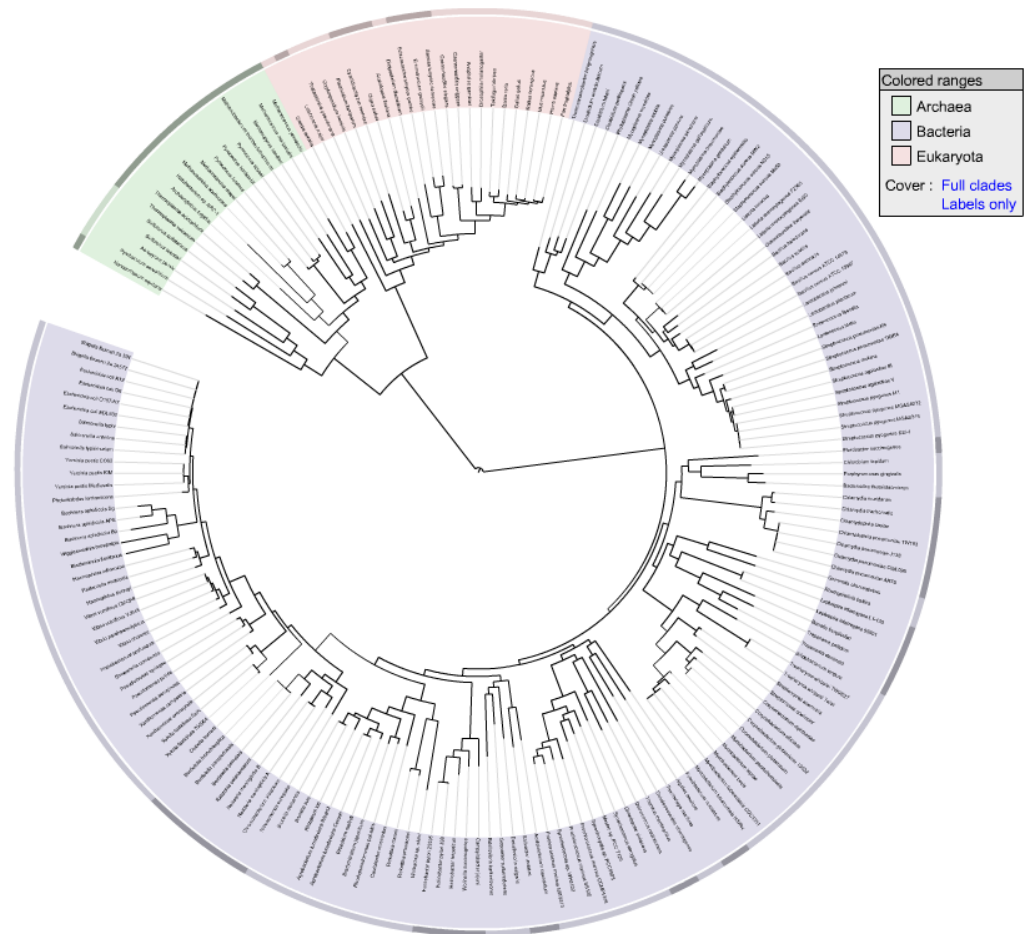
# Les trois domaines du vivant

## Arbre phylogénétique de la vie

Basé sur l'analyse des séquences d'ARN ribosomiques (C. Woese)



Extrait de L'évolution du vivant expliquée à ma boulangère  
Nepoux  
([http://www.ilvbibliotheca.net/librairie/levolution\\_du\\_ma\\_boulangere.html](http://www.ilvbibliotheca.net/librairie/levolution_du_ma_boulangere.html)).



# Les procaryotes

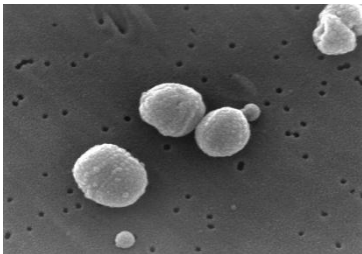
## Les bactéries et les archaeabactéries

Organismes unicellulaires dont la structure cellulaire ne comporte pas de noyau ni d'organites contrairement aux eucaryotes.

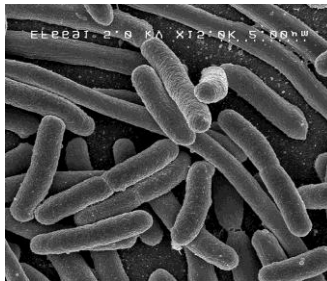
Le matériel génétique chez les procaryotes est regroupé dans une zone appelée nucléoïde qui n'est pas séparée physiquement du reste de la cellule. Chez les eucaryotes, ce matériel est contenu dans le noyau.

Cellule bactérienne typique : taille de 0,5 à 5  $\mu\text{m}$  (mais des exception comme l'espèce *Thiomargarita namibiensis* qui peut mesurer jusqu'à 0,5mm).

La plupart des bactéries sont soit sphériques (coques) ou en forme de bâtonnets (bacilles) mais d'autres formes existent.



*Streptococcus pneumoniae*

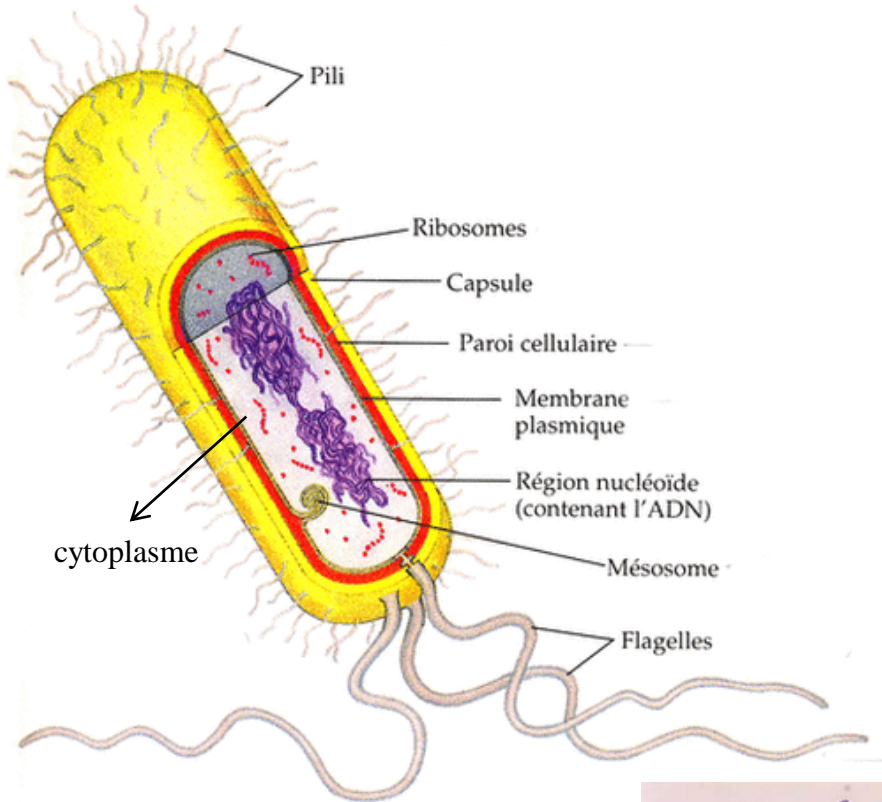


*Escherichia coli*



*Treponema pallidum*

# La cellule procaryote

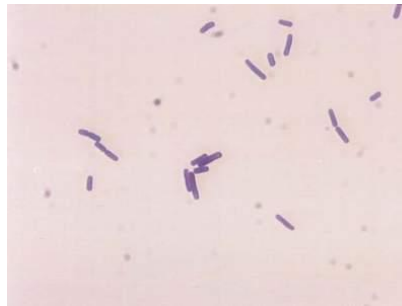


Plasmide : élément autonome pour la réplication. Code pour des fonctions d'adaptation à l'environnement.

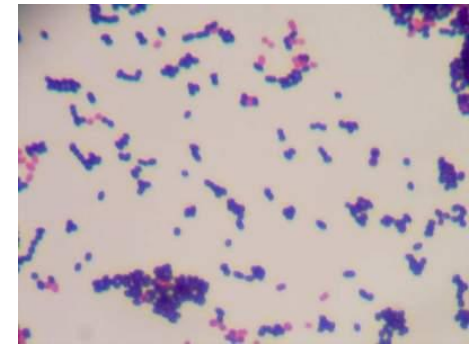
Mésosome : invagination de la membrane plasmique.

La nature de la paroi cellulaire permet de diviser les bactéries en deux groupes (mise en évidence par la coloration gram) :

Bactérie à Gram positif (Gram<sup>+</sup>)  
Bactérie à Gram négatif (Gram<sup>-</sup>)



Bactérie Gram<sup>+</sup> : *Bacillus subtilis*



Mélange :  
bactérie Gram<sup>+</sup> : enterocoques  
Bactéries Gram<sup>-</sup> : Acinetobacter

# Le génome

Ensemble du matériel génétique d'une espèce codé par dans la majorité des cas par son ADN (exception de certains virus dont le génome est porté par des molécules d'ARN).

Chez les bactéries ou archaeae :

- génome en général contenu dans une molécule d'ADN circulaire -> un seul chromosome circulaire
- présence également de génomes extrachromosomiques contenu dans les plasmides
- des exceptions : existence de génomes linéaires (ex : *Streptomyces* pour les bactéries et *Methanobacterium thermoautotrophicum* pour les archaea) et existence de procaryotes possédant plusieurs chromosomes (2 ou 3, ex: les *Burkholderia* (3 chromosomes))

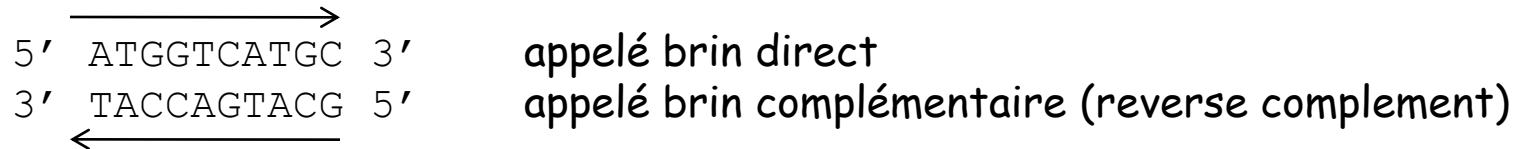
Chez les eucaryotes :

- génome nucléaire contenu dans le noyau et formé de plusieurs chromosomes (46 chez l'homme).
- génomes non nucléaires :
  - le génome mitochondrial contenu dans les mitochondries (chez la quasi-totalité des eucaryotes)
  - le génome chloroplastique contenu dans les chloroplastes (chez les algues et plantes supérieures)

# L'acide desoxyribonucléique : l'ADN

Les brins d'ADN sont orientés dans le sens 5'→ 3' (notation liée à la géométrie du sucre). Les deux brins sont donc complémentaires et antiparallèles

Ex : molécule de 10 nucléotides :



C'est grâce à l'alternance ordonné des 4 bases que l'on va pouvoir coder l'information génétique.

La taille du génome est mesuré en nucléotides ou en bases. En général on l'exprime en paires de bases (pb). On utilise souvent les multiples Kb (Kilo-base = 1000 bases) et Mb (Méga-base = 10<sup>6</sup> bases)

# Taille des génomes

## Exemple génomes procaryotes

Species	Genome size (Mb)
<b>Bacteria</b>	
<i>Mycoplasma genitalium</i>	0,58
<i>Haemophilus influenzae</i>	1,8
<i>Escherichia coli K12</i>	4,6
<i>Bacillus subtilis</i>	4,2
<i>Pseudomonas aeruginosa PAO1</i>	6,3
<b>Archaea</b>	
<i>Nanoarchaeum equitans</i>	0,49
<i>Aeropyrum pernix</i>	1,7
<i>Natronomonas pharaonis</i>	2,9
<i>Sulfolobus solfataricus P2</i>	3
<i>Methanosarcina mazei</i>	4,1

## Exemple génomes eucaryotes

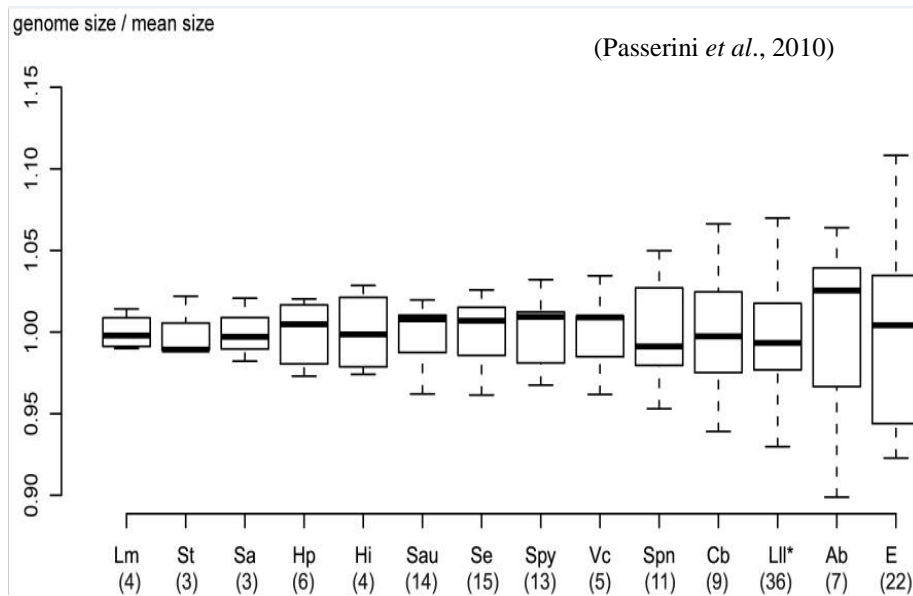
Species	Genome size (Mb)
<b>Fungi</b>	
<i>Saccharomyces cerevisiae</i>	12.1
<i>Aspergillus nidulans</i>	25.4
<b>Protozoa</b>	
<i>Tetrahymena pyriformis</i>	190
<b>Invertebrates</b>	
<i>Caenorhabditis elegans</i>	97
<i>Drosophila melanogaster</i>	180
<i>Bombyx mori</i> (silkworm)	490
<i>Strongylocentrotus purpuratus</i> (sea urchin)	845
<i>Locusta migratoria</i> (locust)	5000
<b>Vertebrates</b>	
<i>Takifugu rubripes</i> (pufferfish)	400
<i>Homo sapiens</i>	3200
<i>Mus musculus</i> (mouse)	3300
<b>Plants</b>	
<i>Arabidopsis thaliana</i> (vetch)	125
<i>Oryza sativa</i> (rice)	430
<i>Zea mays</i> (maize)	2500
<i>Pisum sativum</i> (pea)	4800
<i>Triticum aestivum</i> (wheat)	16 000
<i>Fritillaria assyriaca</i> (fritillary)	120 000



# Taille des génomes

Cependant, au sein d'une même espèce bactérienne, la taille du génome peut varier considérablement :

Exemple parmi 42 souches d'*Escherichia coli*, on observe une variation de la taille du génome allant de 2,6 Mb à 5,7 Mb



Lm, *L. monocytogenes*; St, *S. thermophilus*; Sa, *S. agalactiae*; Hp, *H. pylori*; Hi, *H. influenzae*; Sau, *S. aureus*; Se, *S. enterica*; Spy, *S. pyogenes*; Vc, *V. cholerae*; Spn, *S. pneumoniae*; Cb, *C. botulinum*; Lll, *L. lactis* subsp. *lactis*; Ab, *A. baumannii*; Ec, *E. coli*.

# L'ADN : support de l'information génétique

Le génome est composé de régions codantes, les gènes, et de régions non codantes.

Les gènes codent :

- des protéines
- des ARN ribosomiques
- des ARN de transfert

Le nombre de gènes dans un génome varie moins que sa taille, cependant la corrélation avec la complexité de l'organisme n'est pas parfaite.

Species	Genome size (Mb)	Nombre approximatif de gènes
<i>Mycoplasma genitalium</i> (bactérie)	0,58	500
<i>Haemophilus influenzae</i> (bactérie)	1,8	1800
<i>Escherichia coli K12</i> (bactérie)	4,6	4400
<i>Saccharomyces cerevisiae</i> (levure)	12	5800
<i>Drosophila melanogaster</i> (insecte)	180	14700
<i>Caenorhabditis elegans</i> (nematode)	103	20000

Species	Genome size (Mb)	Nombre approximatif de gènes
<i>Tetrahymena thermophila</i> (protozoaire)	125	2700
<i>Arabidopsis thaliana</i> (plante à fleur)	120	26500
<i>Oryza sativa</i> (riz)	430	~45000
<i>Zea mays</i> (maïs)	2200	> 45000
<i>Mus musculus</i> (souris)	2600	22000
<i>Homo sapiens</i>	3200	22000
Paramécie	72	40000



# Le « dogme central » de la biologie moléculaire



Les flèches indiquent le sens pour le transfert de l'information génétique. Terme « dogme central » employé en 1956 par Francis Crick.

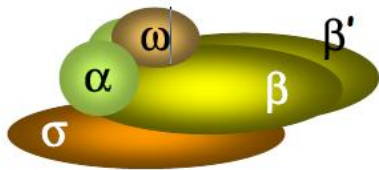
L'ARN peut parfois servir de matrice pour la synthèse d'un brin d'ADN complémentaire. Cependant, évènement rare comparé à la grande quantité de molécules d'ARN synthétisées à partir d'une matrice ADN.

Donc ce flux orienté de l'information reste d'actualité.

# La transcription

Elle est réalisée par une enzyme : l'ARN polymérase qui réalise essentiellement les mêmes réactions de la bactérie à l'homme.

Chez les bactéries, une seule ARN polymérase constituée de 5 sous-unités, deux exemplaires de la sous-unité  $\alpha$ , un exemplaire de chacune des sous-unités  $\beta$ ,  $\beta'$  et  $\omega$ . Le facteur  $\sigma$  permet la reconnaissance des séquences spécifiques sur l'ADN, appelée **promoteur**.



ARN polymérase procaryote

# La transcription

3 étapes :

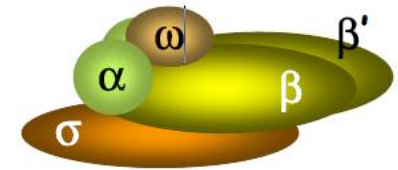
- l'initiation : reconnaissance de séquences spécifiques sur l'ADN : le promoteur
- l'élongation
- la terminaison

## Transcription des gènes codant pour des protéines

Chez les bactéries :

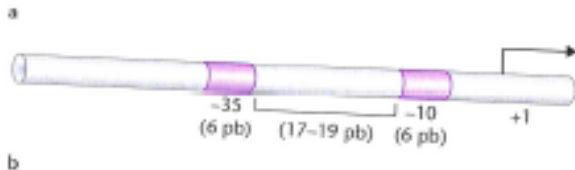
L'initiation :

Le facteur  $\sigma$  de la RNA polymérase reconnaît deux régions constituant le promoteur : la région -35 et la région -10 (TATAAT box) séparées par une région de taille variable (17 à 19 pb). Le promoteur se trouve en amont du début du (des gènes).

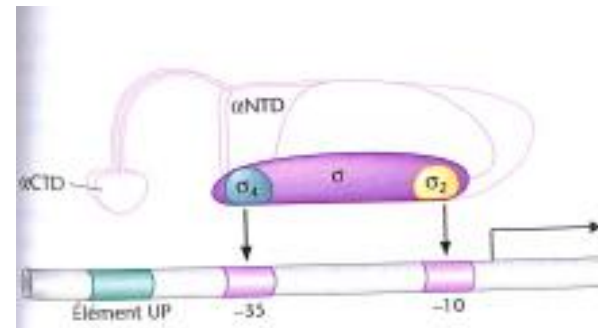


ARN polymérase procaryote

Promoteurs bactériens



Recrutement du cœur de l'ARN polymérase au promoteur

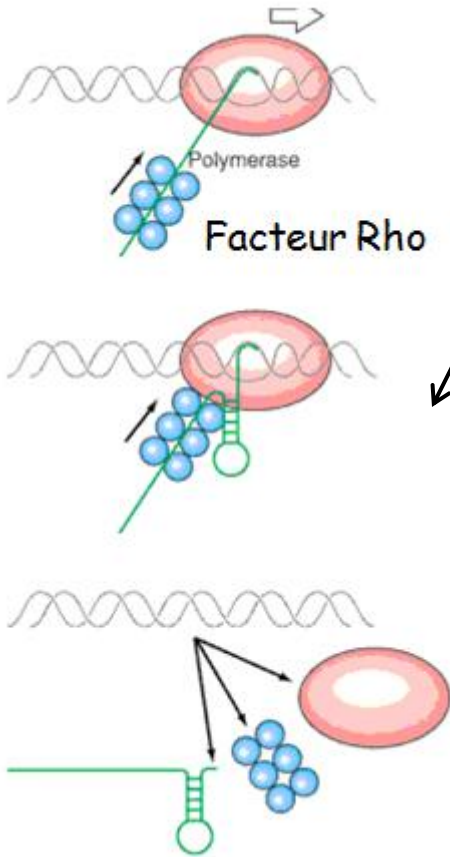


# La transcription

## Transcription des gènes codant pour des protéines

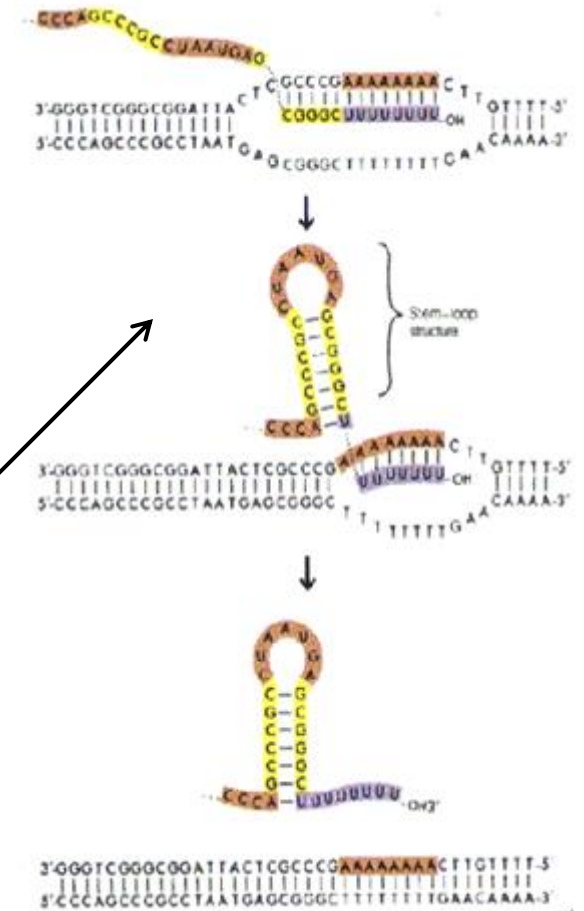
Chez les bactéries :

La terminaison : deux types de terminateurs Rho-dépendants et Rho-indépendants



Rho-dépendant : L'ATPase Rho, une protéine qui se déplace le long du transcrit naissant jusqu'à rattraper la polymérase, stimule la terminaison de la transcription

Rho-indépendant : deux éléments une courte séquence répétée inversée suivie d'une série d'environ 8 paires de bases A-T. Le transcrit est capable de former une structure secondaire tige-boucle qui provoque la désorganisation de l'ARN polymérase. La suite du transcrit est faiblement apparié au brin ADN ce qui conduit à sa libération.

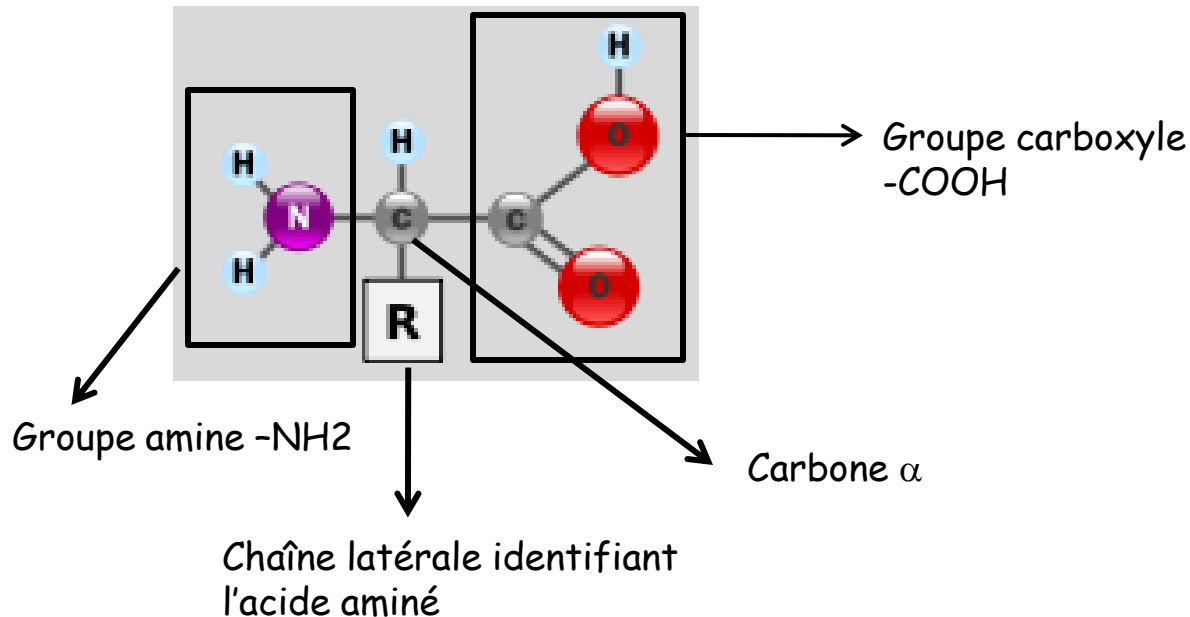


# La traduction

Etape qui permet la synthèse des protéines à partir de l'ARN messenger.

Une protéine est une macromolécule biologique composée d'acides aminés. Il existe 20 acides aminés qui sont rencontrés dans tous les organismes.

Structure générale d'un acide aminé

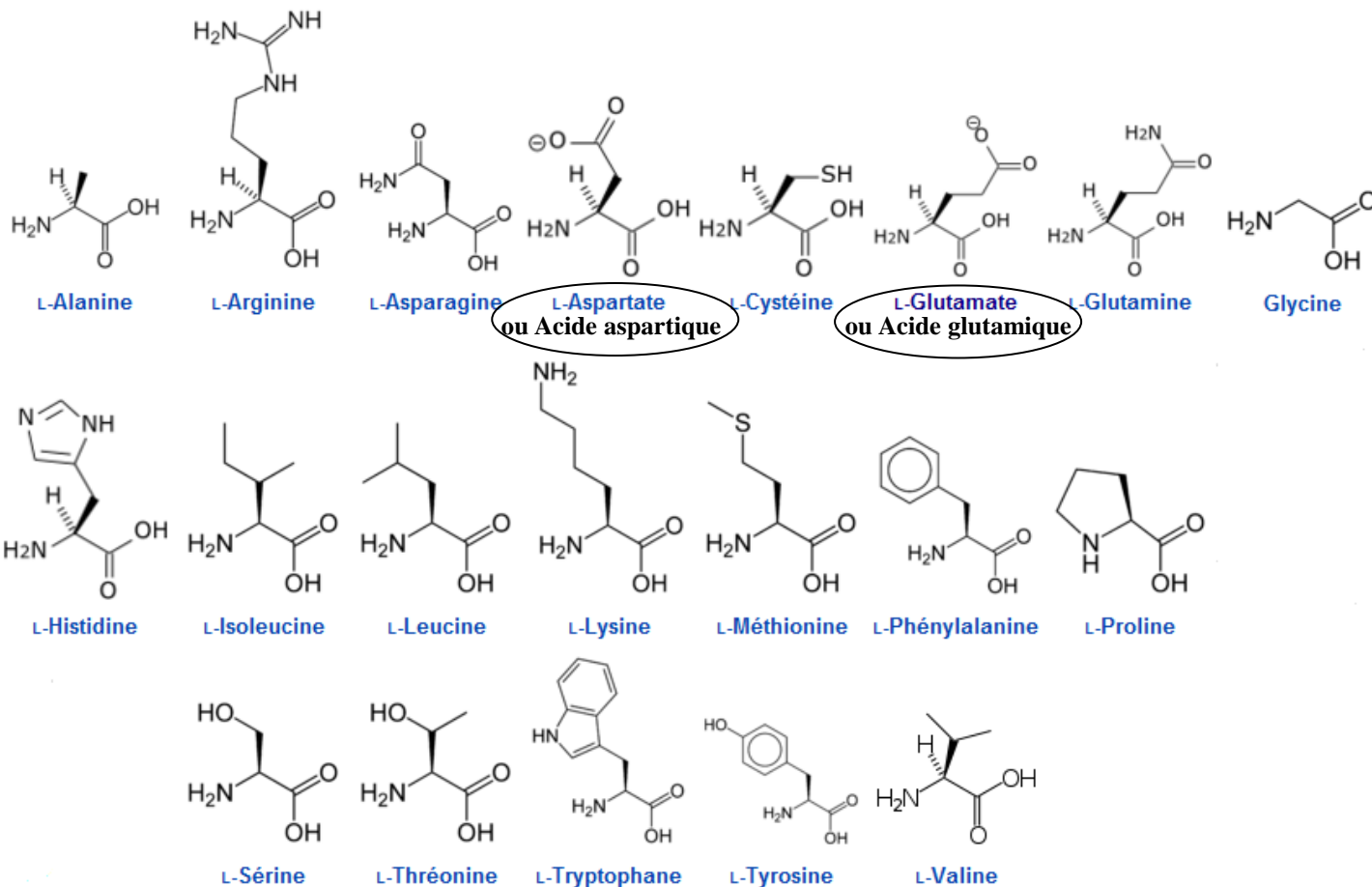


Les acides aminés polymérisent en formant des liaisons peptidiques

# Les acides aminés

codes et abréviations sont ceux spécifiés par le comité de nomenclature commun IUAPC - IUBMB

## Structure chimique des 20 acides aminés



Code	Abrév.	Acide aminé
A	Ala	Alanine
C	Cys	Cystéine
D	Asp	Acide aspartique
E	Glu	Acide glutamique
F	Phe	Phénylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Méthionine
N	Asn	Asparagine
O	Pyl	Pyrrolysine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Sérine
T	Thr	Thréonine
U	Sec	Sélénocystéine
V	Val	Valine
W	Trp	Tryptophane
Y	Tyr	Tyrosine

Les acides aminés sélénocystéine et pyrrolysine ne sont pas présents dans toutes les espèces (pyrrolysine que chez certaines archaea)

# La traduction

Comment passer d'une suite de bases de l'ARNm à une suite d'acides aminés de la protéine ?

La partie codante des gènes codant les protéines sont structurées en une suite de regroupements de 3 bases appelée codons. A chaque codon correspond un acide aminé. Cette correspondance est donnée par le code génétique. Un acide aminé peut être codé par plusieurs codons, on dit que le code est dégénéré.

3 codons particuliers - codons stop : UAG, UAA, UGA. Ils servent de terminaison à la traduction. Permettent d'identifier la fin de la protéine.

Le premier codon est appelé codon start, il correspond soit à AUG, UUG ou GUG. En face sera toujours incorporé une méthionine modifiée.

**Le code présenté est le code génétique standard utilisé par la majorité des organismes mais il existe quelques variantes**

Acide aminé ↕	↕	↕	Codons ↕
Alanine	Ala	A	GCU, GCC, GCA, GCG.
Arginine	Arg	R	CGU, CGC, CGA, CGG ; AGA, AGG.
Asparagine	Asn	N	AAU, AAC.
Acide aspartique	Asp	D	GAU, GAC.
Cystéine	Cys	C	UGU, UGC.
Glutamine	Gln	Q	CAA, CAG.
Acide glutamique	Glu	E	GAA, GAG.
Glycine	Gly	G	GGU, GGC, GGA, GGG.
Histidine	His	H	CAU, CAC.
Isoleucine	Ile	I	AUU, AUC, AUA.
Leucine	Leu	L	UUA, UUG ; CUU, CUC, CUA, CUG.
Lysine	Lys	K	AAA, AAG.
Méthionine	Met	M	AUG.
Phénylalanine	Phe	F	UUU, UUC.
Proline	Pro	P	CCU, CCC, CCA, CCG.
Pyrolysine	Pyl	O	UAG, après séquence PylIS.
Sélénocystéine	Sec	U	UGA, après séquence SecIS.
Sérine	Ser	S	UCU, UCC, UCA, UCG ; AGU, AGC.
Thréonine	Thr	T	ACU, ACC, ACA, ACG.
Tryptophane	Trp	W	UGG. (UGA)
Tyrosine	Tyr	Y	UAU, UAC.
Valine	Val	V	GUU, GUC, GUA, GUG.
START			AUG. (UUG, GUG)
STOP <i>Ambre</i>			UAG.
STOP <i>Ocre</i>			UAA.
STOP <i>Opale</i>			UGA.



# La traduction

		Second Letter								
		T	C	A	G					
First Letter	T	TTT } Phe TTC } TTA } Leu TTG }	TCT } TCC } Ser TCA } TCG }	TAT } Tyr TAC } TAA } Stop TAG } Stop	TGT } Cys TGC } TGA } Stop TGG } Trp	Third Letter	T	G		
	C	CTT } CTC } Leu CTA } CTG }	CCT } CCC } Pro CCA } CCG }	CAT } His CAC } CAA } Gln CAG }	CGT } CGC } Arg CGA } CGG }		T		G	
	A	ATT } ATC } Ile ATA } ATG } Met	ACT } ACC } Thr ACA } ACG }	AAT } Asn AAC } AAA } Lys AAG }	AGT } Ser AGC } AGA } Arg AGG }		T			G
	G	GTT } GTC } Val GTA } GTG }	GCT } GCC } Ala GCA } GCG }	GAT } Asp GAC } GAA } Glu GAG }	GGT } GGC } Gly GGA } GGG }		T			

La dégénérescence du code génétique touche essentiellement la 3<sup>ème</sup> base du codon et dans une moindre mesure la 1<sup>ère</sup> base

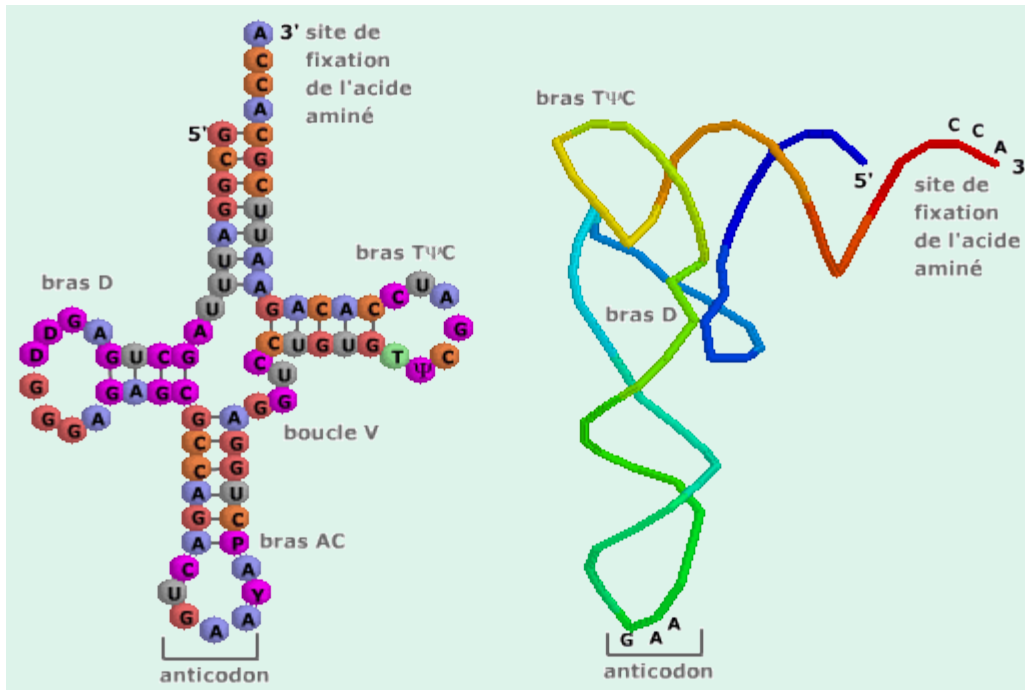
# La traduction

4 composants principaux font partie de la machinerie de traduction :

- l'ARN messenger (ARNm) contenant la séquence codant la protéine ainsi que des régions de reconnaissance pour l'initiation et la terminaison de la traduction.
- les ARN de transfert (ARNt) qui une fois chargé de leur acide aminé (processus réalisé par les aminocyl-ARNt- synthétases) vont mettre en relation un codon et un acide aminé.
- les aminocyl-ARNt- synthétases permettant donc de charger l'acide aminé sur l'ARNt
- le ribosome : gros complexe formé de deux sous-unités (une grande et une petite) chacune constituées par des protéines ribosomiques et un ou plusieurs ARN ribosomique (ARNr).

# La traduction : les ARNt

Ce sont de petites molécules d'ARN, d'une taille variant entre 70 et 100 nucléotides. Ils se replient en une structure secondaire caractéristique appelée feuille de trèfle et présente une structure tridimensionnelle en forme de L.



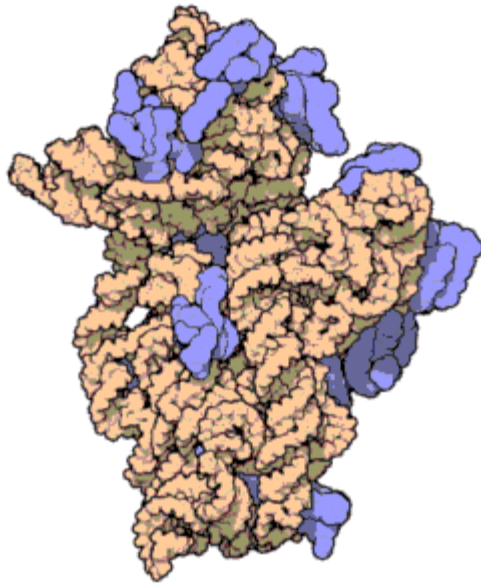
L'anticodon s'apparie avec le codon de l'ARNm. L'ARNt chargé portant à son extrémité 3' l'acide aminé correspondant permet donc de faire la correspondance codon-acide aminé. Après reconnaissance du codon, l'acide aminé est transféré à la chaîne peptidique en croissance. Ce processus est réalisé à l'intérieur du ribosome.

Structures secondaire et tertiaire d'un ARNt

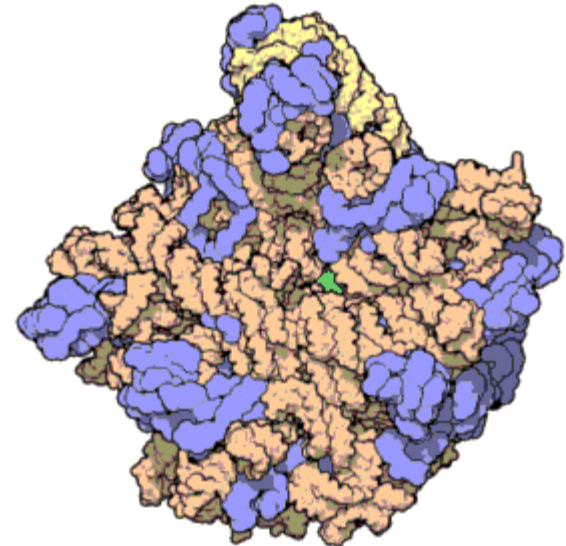
## La traduction : le ribosome

Le ribosome est constitué de deux sous-unités : la grande et la petite sous-unité.

Les procaryotes ont des ribosomes 70S constitué d'une petite sous-unité 30S et d'une grande sous-unité 50S. La petite sous-unité est composée de l'ARNr 16S (~1500 nt) et de 21 protéines. La grande sous-unité possède deux molécules d'ARNr, l'ARNr 5S (~120 nt) et l'ARNr 23S (~2900 nt), et 31 protéines.



Petite sous unité 30S de *Thermus thermophilus*.  
Bleue protéines, orange ARNr 16S



Grande sous unité 50S de *Haloarcula marismortui*. Bleue protéines, orange et jaune les deux ARNr. En vert au centre de la sous-unité, le site actif

# La traduction : le ribosome

Le ribosome : une usine de synthèse des protéines.

Il contient 3 sites de fixation pour les ARNt :

- le site A lie les ARNt chargés de leur acide aminé (ARNt aminoacylés)
- le site P lie les ARNt liés à la chaîne peptidique en cours de synthèse (peptidyl ARNt)
- le site E lie les ARNt libres (déchargé et décroché la chaîne peptidique) avant leur sortie du ribosome (E pour « exit »).

La traduction comprend également 3 étapes :

- l'initiation qui charge le ribosome sur l'ARNm
- l'élongation qui conduit à la synthèse de la protéine
- la terminaison conduisant à la dissociation du ribosome de l'ARNm

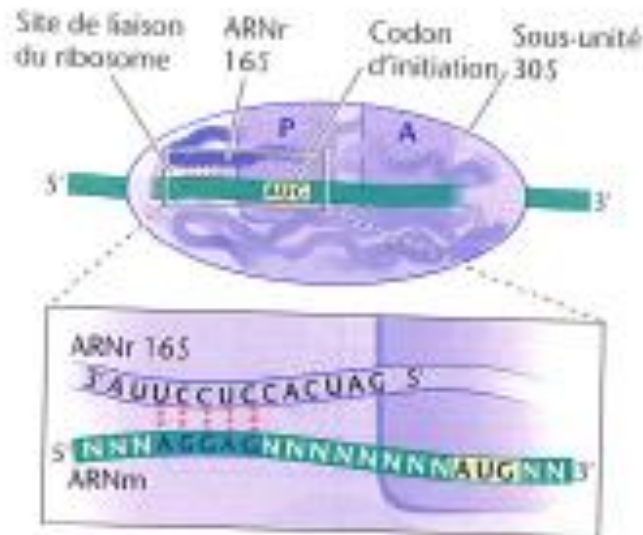
# La traduction

## L'initiation :

### ➤ chez les bactéries :

La petite sous-unité est chargée en premier sur l'ARNm, ceci grâce à un appariement des bases d'une région de l'ARNr 16S avec le RBS. Pour un RBS idéalement situé, le codon initiateur (AUG, GUG ou UUG) se trouve situé au site P du ribosome (et non au A comme pour l'élongation). Ceci requiert un ARNt particulier, appelé ARNt initiateur. Cet ARNt ne porte ni la méthionine, ni la valine, ni la leucine comme acide aminé, mais une méthionine modifiée (N-formylméthionine) d'où son nom ARNt fMet.

Quand l'ARNt fMet s'apparie au site P, il y a un changement de conformation de la petite sous-unité qui fait que la grande sous-unité peut se lier à elle pour former le ribosome 70S.



Interaction RBS et ARNr 16S pour positionnement du codon AUG au site P

RBS = Ribosome Binding Site

# La traduction

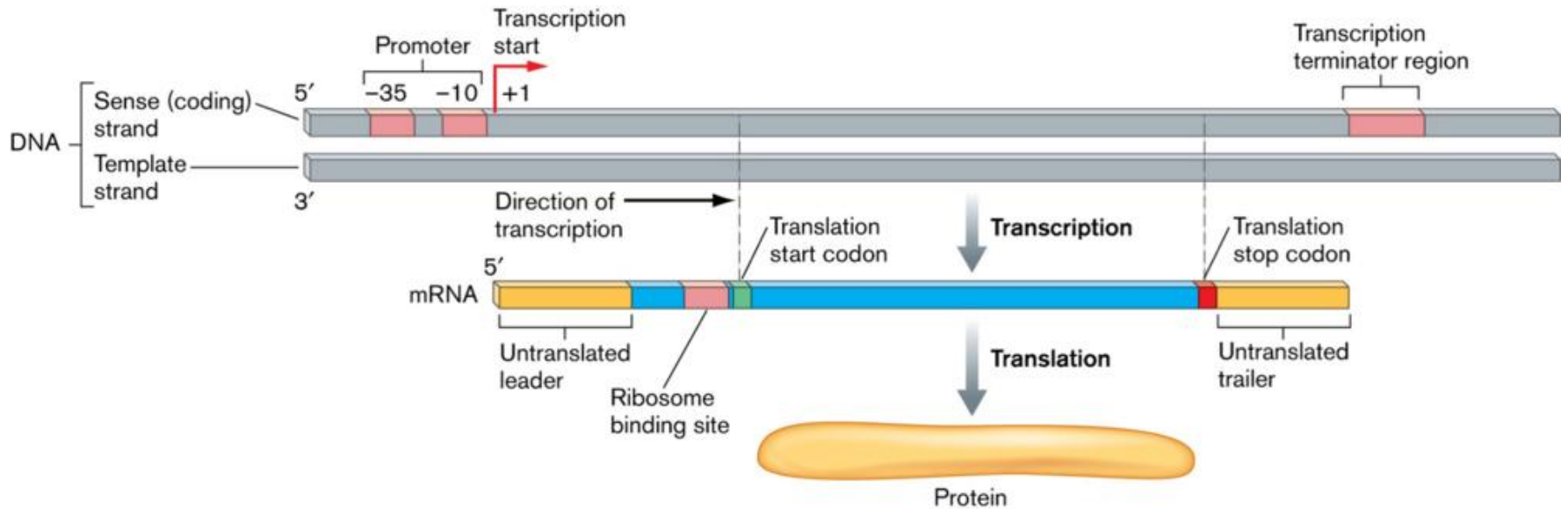
## La terminaison :

La traduction s'arrête quand le ribosome rencontre un codon stop au site A.

Ce codon stop est reconnu par un des deux facteurs de terminaison de classe 1 chez les procaryotes. Ce facteur stimule l'hydrolyse du polypeptide et du peptidyl-ARNt, libérant ainsi le peptide complet.



# Structure d'un gène bactérien



Les gènes peuvent être co-transcrits. Ils sont organisés en unité de transcription, appelée opéron entre un promoteur et un terminateur de transcription.

