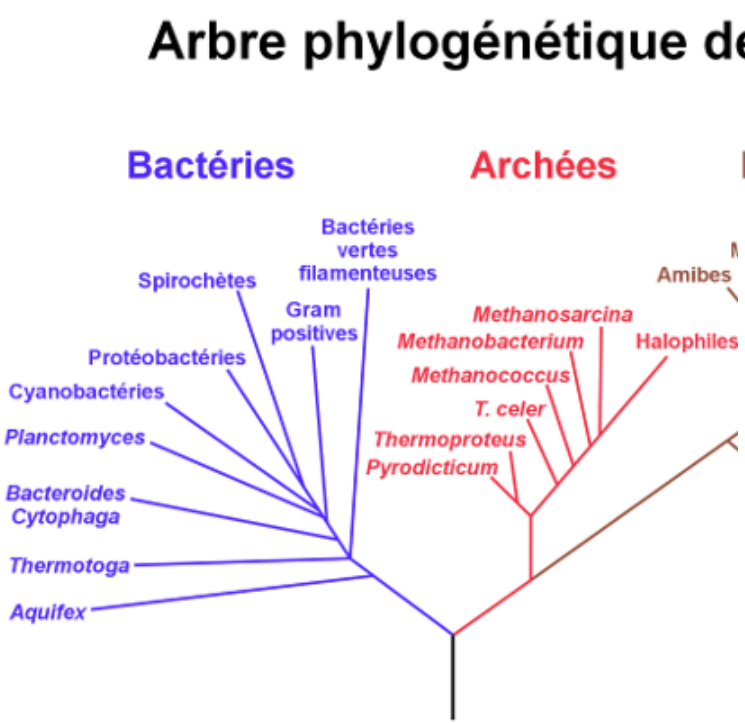
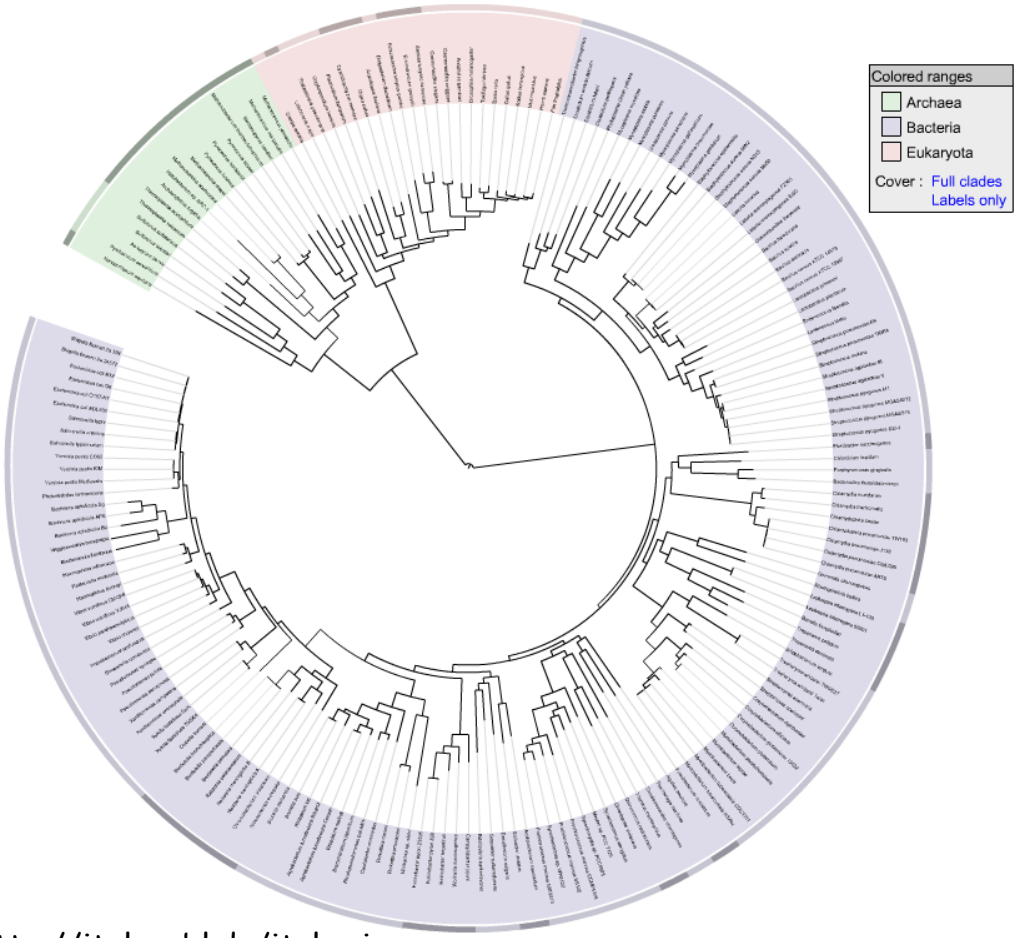
Les trois domaines du vivant

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



Extrait de L'évolution du vivant expliquée à ma boulan
Nepoux
(http://www.ilvbibliotheca.net/librairie/levolution_du
_ma_boulangere.html).



http://itol.embl.de/itol.cgi

Les procaryotes

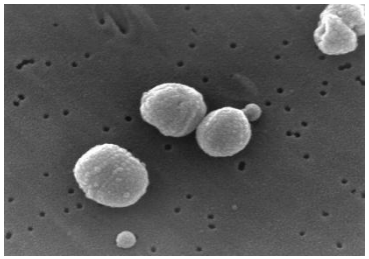
Les bactéries et les archaeabactéries

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

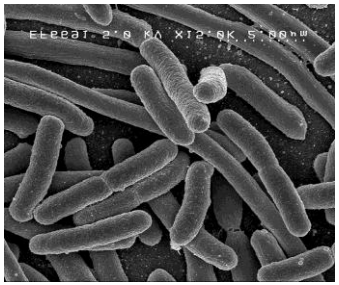
Le matériel génétique chez les procaryotes est regroupé dans une zone appelée nucléoide qui n'est pas séparée physiquement du reste de la cellule. Chez les eucaryotes, ce matériel est contenu dans le noyau.

Cellule bactérienne typique : taille de 0,5 à 5 μm (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.



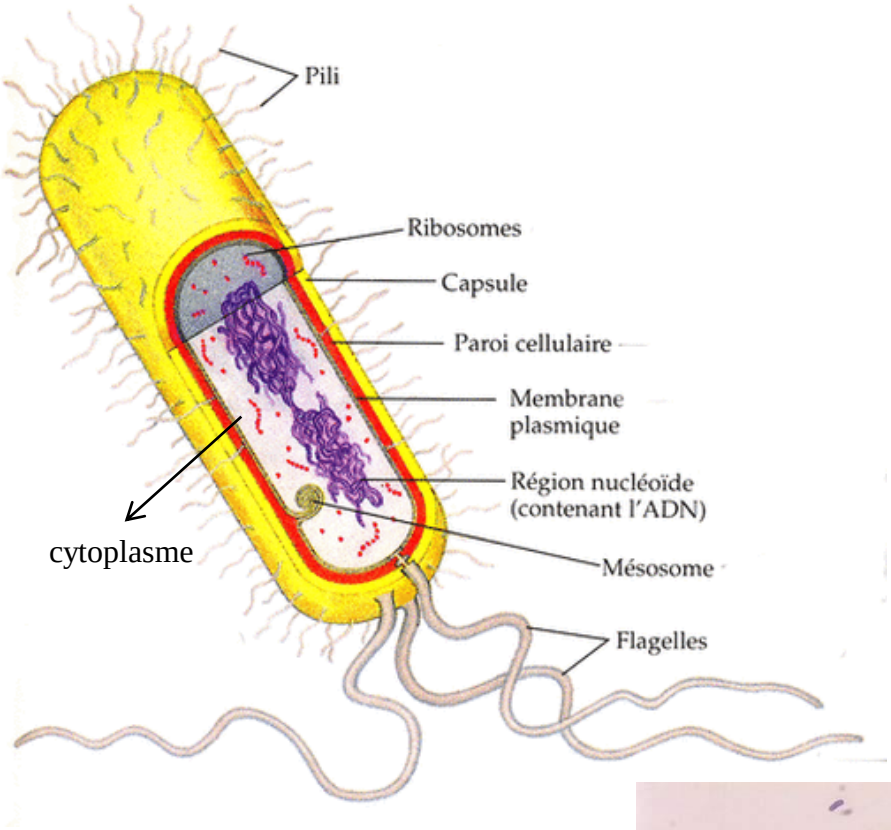
Streptococcus pneumonia



Escherichia coli

Treponema pallidum

La cellule procaryote

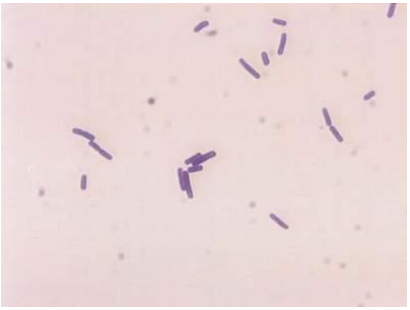


Mésosome : invagination de la membrane plasmique.

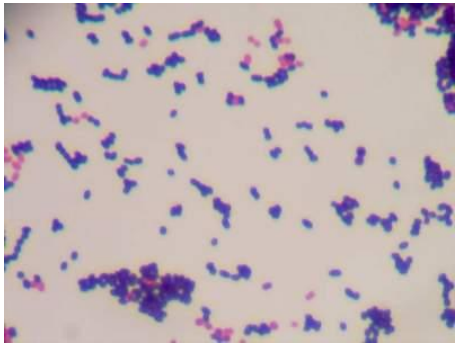
La nature de la paroi cellulaire permet de diviser les bactéries en deux groupes (mise en évidence par la coloration gram) :

Bactérie à Gram positif (Gram$^+$)
Bactérie à Gram négatif (Gram$^-$)

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



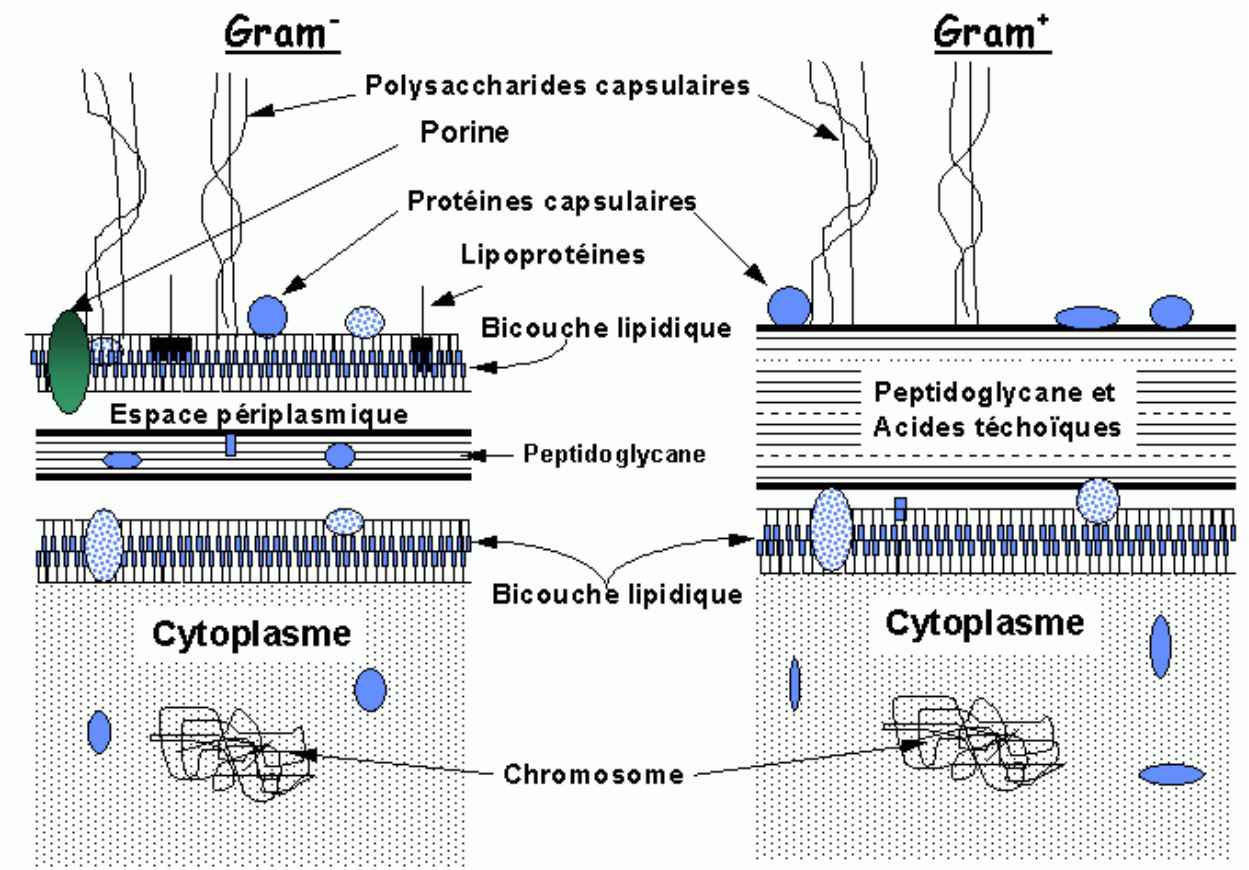
Bactérie Gram$^+$: *Bacillus subtilis*



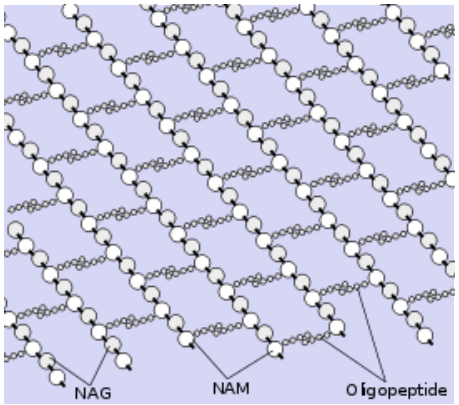
Mélange :
bactérie Gram$^+$: enterocoques
Bactéries Gram$^-$: Acinetobacter

La cellule procaryote

Bactérie Gram+ possède une paroi cellulaire formée d'une couche épaisse de peptidoglycane et d'acides téchoïques.
Bactérie Gram- possède une mince couche de peptidoglycane entouré d'une membrane interne et d'une membrane externe.



Structures de la membrane et de la paroi de peptidoglycane chez les bactéries Gram⁺/Gram⁻



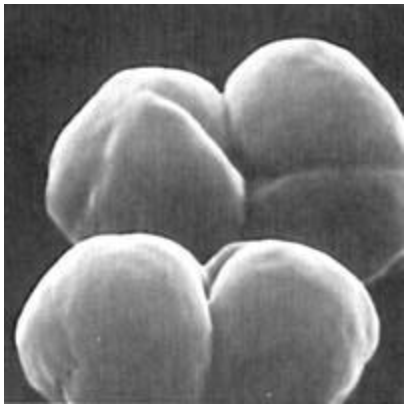
Structure du peptidoglycane

La cellule procaryote

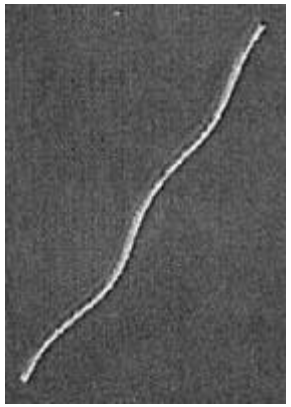
Les archaea :

- composition chimique de la paroi différente de celle des bactéries (plusieurs types de composition existent)

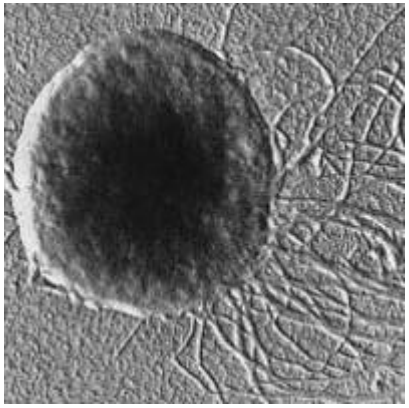
- phospholipides membranaires ont des caractéristiques spécifiques ne se retrouvant ni chez les bactéries, ni chez les eucaryotes



Methanosarcina barkeri



Methanobacterium thermoautotrophicum

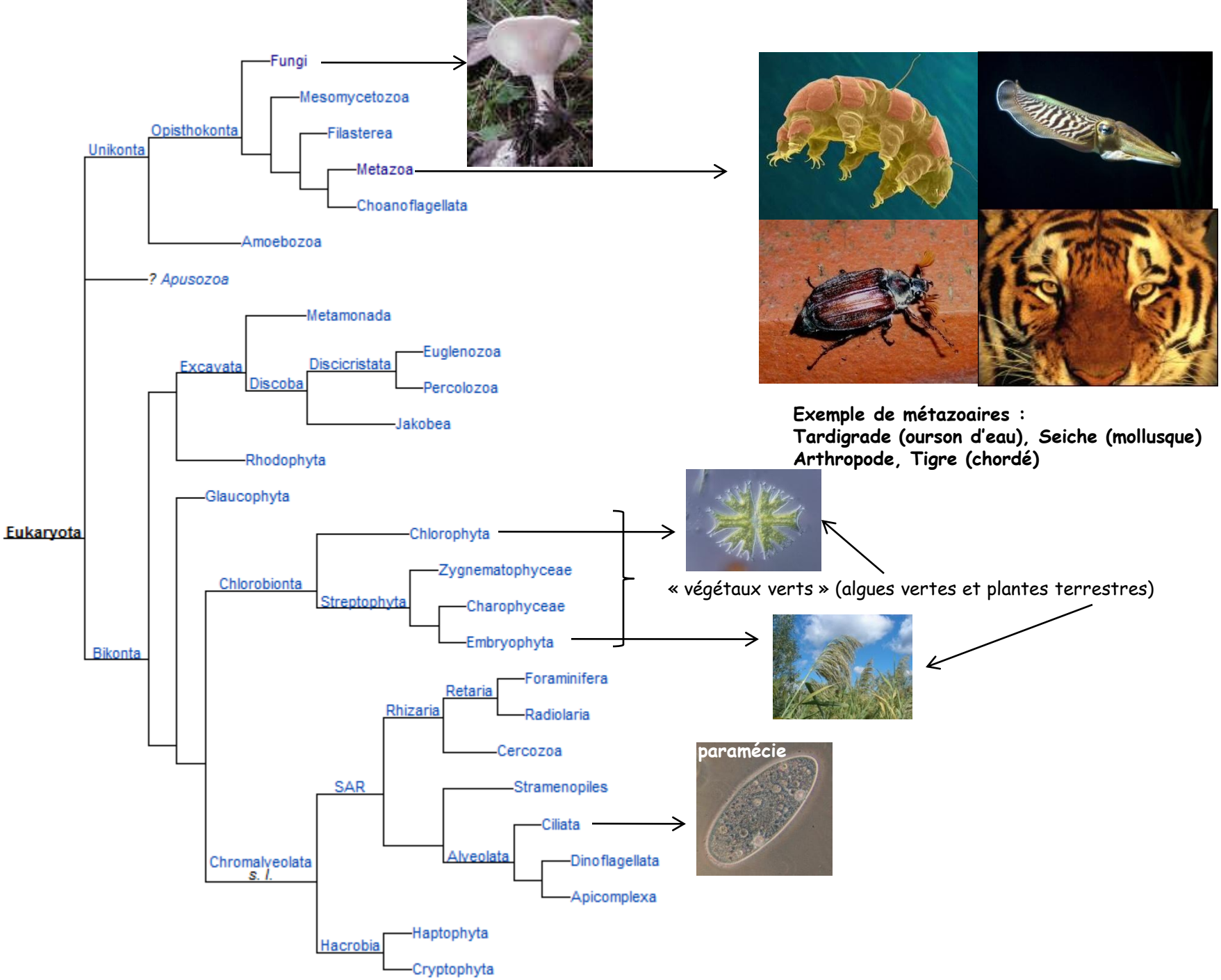


Methanococcus janaschii

Les eucaryotes

Domaine regroupant tous les organismes uni- ou pluricellulaire se caractérisant par la présence d'un noyau et de mitochondries dans les cellules.

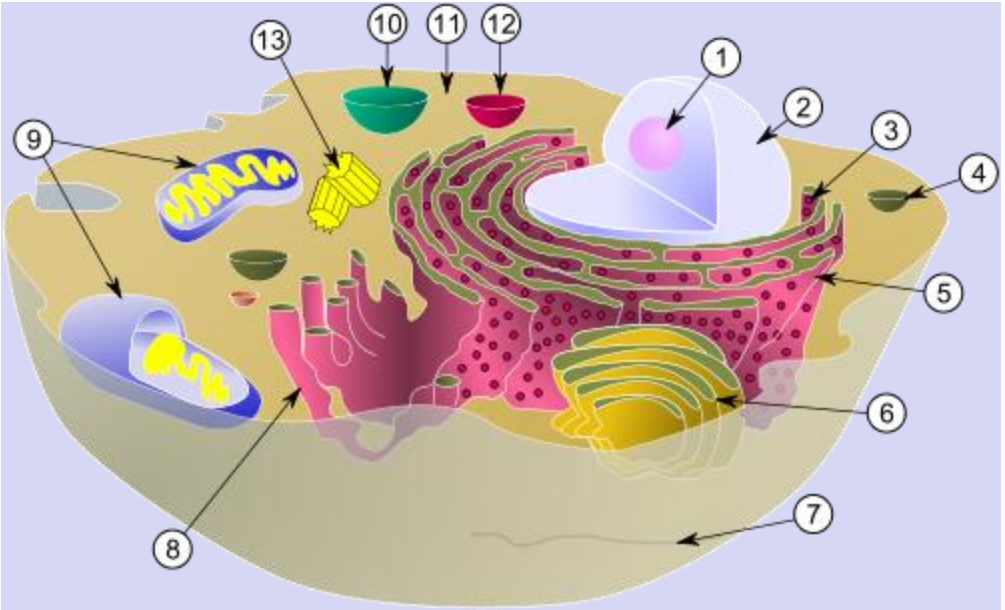
Reproduction sexuée


Eukaryota
Unikonta
Opisthokonta
Fungi
Mesomycetozoa
Filasterea
Metazoa
Choanoflagellata
Amoebozoa
? Apusozoa
Bikonta
Excavata
Metamonada
Discoba
Discicristata
Euglenozoa
Percolozoa
Jakobea
Rhodophyta
Glaucophyta
Chlorobionta
Chlorophyta
Streptophyta
Zygnematophyceae
Charophyceae
Embryophyta
Chromalveolata s. l.
SAR
Rhizaria
Retaria
Foraminifera
Radiolaria
Cercozoa
Stramenopiles
Alveolata
Ciliata
Dinoflagellata
Apicomplexa
Hacrobia
Haptophyta
Cryptophyta
paramécie
Exemple de métazoaires :
Tardigrade (ourson d'eau), Seiche (mollusque)
Arthropode, Tigre (chordé)
« végétaux verts » (algues vertes et plantes terrestres)

La cellule eucaryote

Elle se distingue de la cellule procaryotes notamment par l'existence de compartiments appelés organites :

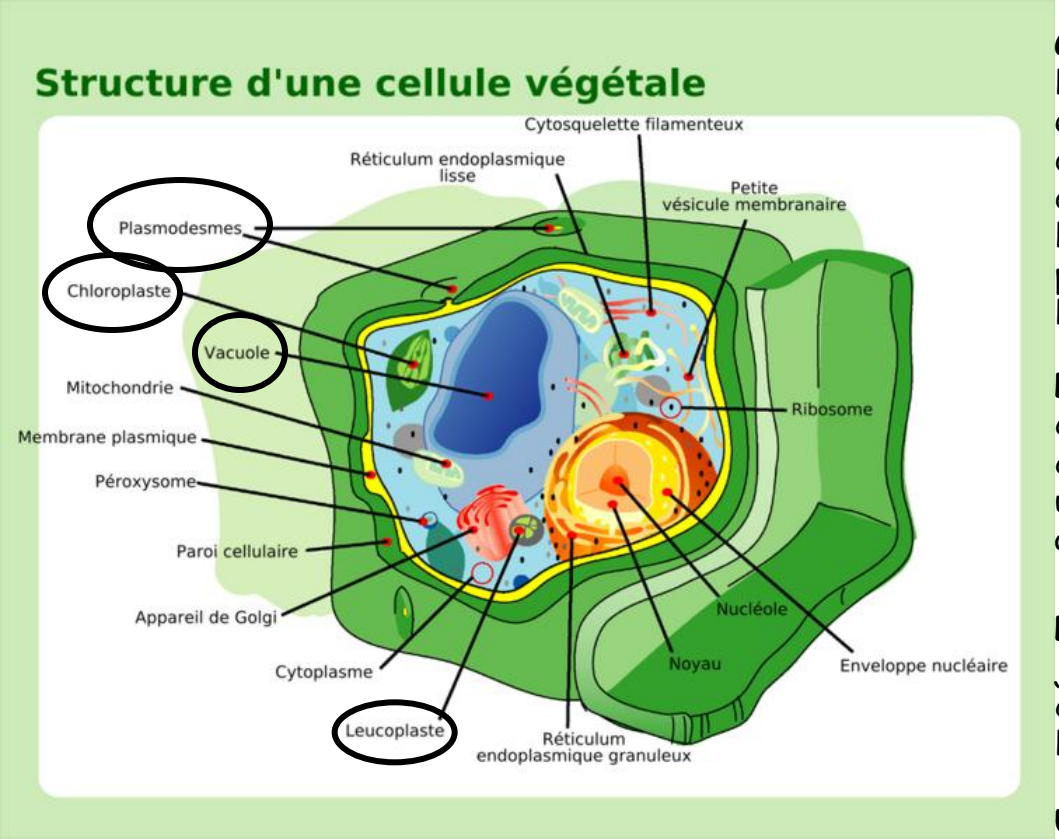
Cellule eucaryote typique



- 1 : nucléole : sous compartiment du noyau lieu de la transcription des ARN ribosomiques.
- 2 : noyau : organite contenant la majorité du matériel génétique
- 3 : ribosomes
- 4 : vésicule : Circulant dans le cytoplasme, elle peut stocker, transporter ou encore digérer des produits et déchets cellulaires.
- 5 : réticulum endoplasmique granuleux : site de la synthèse et du repliement des protéines
- 6 : appareil de Golgi : lieu où certaines protéines sont modifiées et rôle dans la sécrétion des protéines
- 7 : microtubule
- 8 : réticulum endoplasmique lisse : intervient dans plusieurs processus métaboliques (synthèse des lipides, métabolisme du calcium (stokage)), rôle dans la détoxification de la cellule…
- 9 : mitochondrie : « centrale énergétique », siège des deux dernières étapes de la respiration. Possède son propre matériel génétique (ADN mitochondrial).
- 10 : lysosome : digestion intracellulaire
- 11 : cytoplasme : désigne contenu de la cellule, émulsion colloïdale constituée par le cytosol, phase liquide où baigne les organites
- 12 : peroxysome : détoxification de la cellule
- 13 : centrosome (centre organisateur des microtubules)

La cellule eucaryote

Caractéristiques de la cellule végétale :



Grande vacuole centrale : maintien la turgescence de la cellule (eau entre dans la cellule, la vacuole se remplit et grossit. Ses membranes se tendent (turgescence des cellules). Cette eau exerce une pression sur la paroi des cellules et donne la rigidité aux parties soupes de la plantes (tiges, feuilles …). La vacuole sert aussi de poubelle à la cellule végétale. Afin de la vie de la cellule, la vacuole occupe 90% de l'espace cellulaire.

Les plasmodesmes : canaux traversant la paroi cellulaire. Ils relient les membranes plasmiques et les cytoplasmes des cellules adjacentes et permettent un passage contrôlé de molécules entre cellules adjacentes

Les plastes, en particulier les **chloroplastes** qui joue un rôle dans la photosynthèse et qui contiennent la chlorophylle. Les plastes possèdent leur propre matériel génétique.

Une paroi pectocellulosique faite de cellulose et de protéines, ainsi que de lignine dans de nombreux cas. Elle protège la cellule végétale. Elle diffère de la paroi cellulaire des champignons, faite de chitine, et des procaryotes, faite de peptidoglycanes.

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 moléculaire 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

Molécule constituant le génome

- l'ADN renferme les informations nécessaires au développement et fonctionnement d'un organisme -> support de l'information génétique.

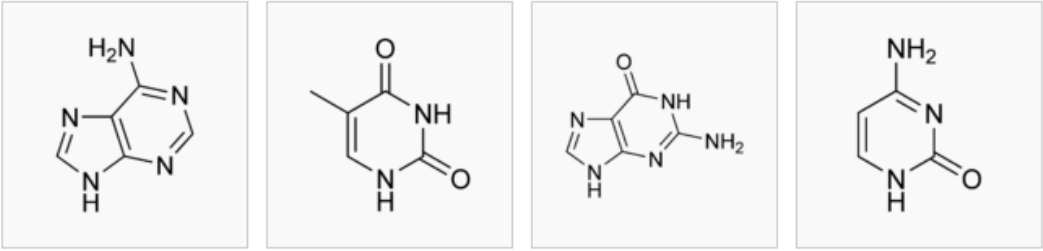
- l'ADN est le support de l'hérédité car il est transmis de génération en génération lors de la reproduction (sexuée ou non).

L'ADN est composé d'une suite de **nucléotides**. Chaque nucléotide est formé de trois éléments :

- un groupe phosphate lié à
- un sucre, le désoxyribose lié à
- une base azotée

Il existe 4 bases différentes :

- l'adénine (A)
- la thymine (T)
- la cytosine C
- la guanine (G)



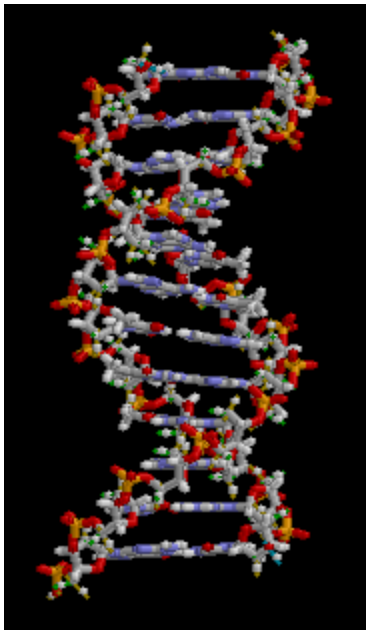
La thymine et la cytosine sont des pyrimidines
L'adénine et la guanine sont des purines

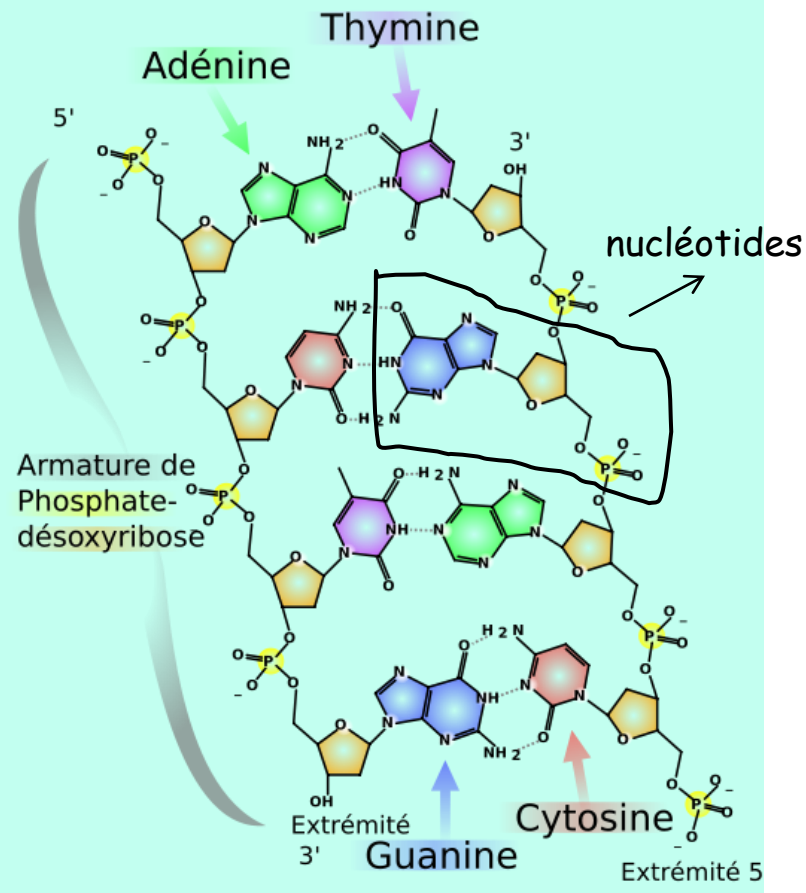
L'acide desoxyribonucléique : l'ADN

L'ADN est constitué de deux brins formant une double hélice. Ceci est possible car les bases sont complémentaires deux à deux et peuvent interagir par des liaisons hydrogènes.

On peut avoir les interactions suivantes appelées Watson-Crick

A-T ou T-A avec deux liaisons hydrogènes
C-G ou G-C avec trois liaisons hydrogènes





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 :

```
   ──────────────→
5'  ATGGTCATGC  3'      appelé brin direct
3'  TACCAGTACG  5'      appelé brin complémentaire (reverse complement)
   ←──────────────
```

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 = 10^6 bases)

Taille des génomes

Exemple génomes procaryotes

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

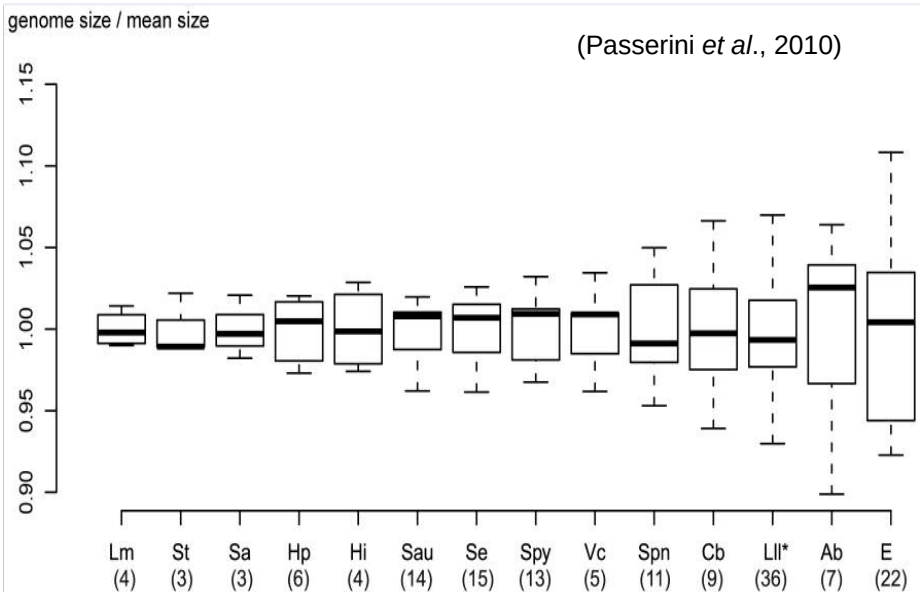
Exemple génomes eucaryotes

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

Taille des génomes

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

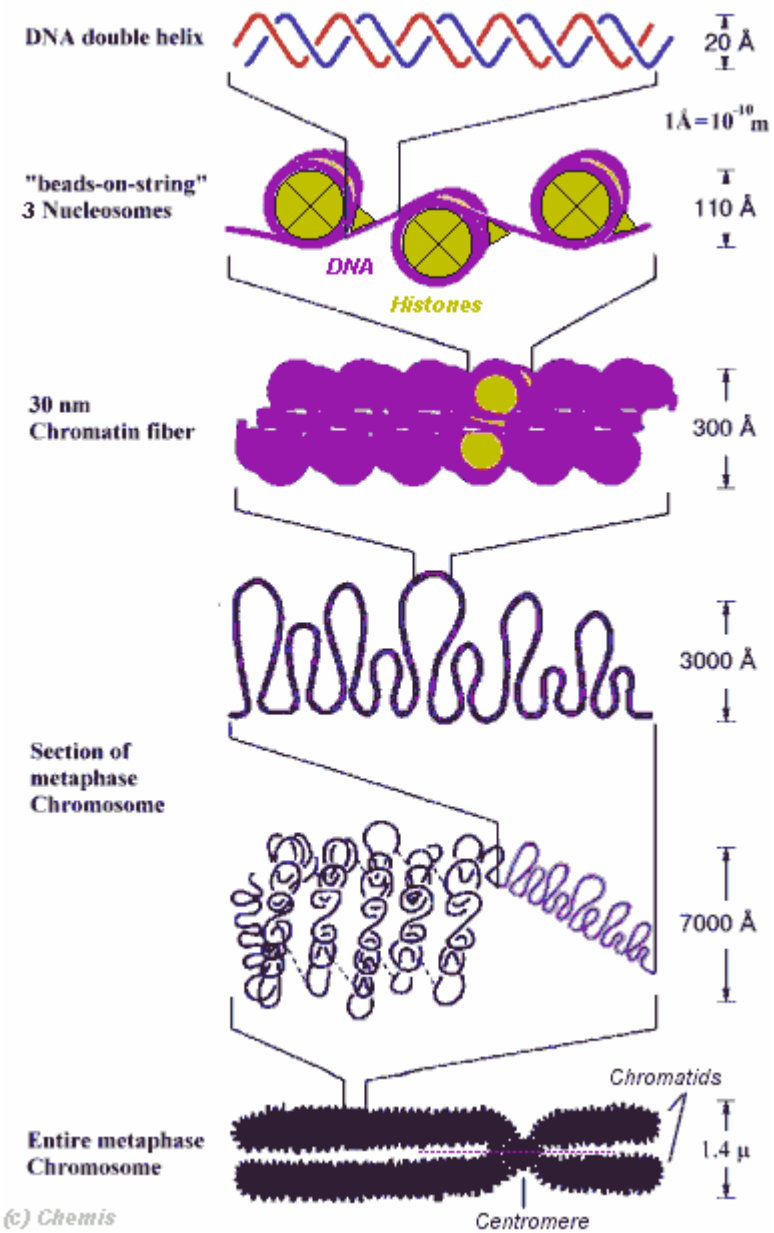
Exemple parmi 42 souches d'*Escherichia coli*, on observe une variation de la taille du génome allant de 2,6 Mb à 5,7 Mb



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*.

Structure tertiaire de l'ADN
= interaction protéine / ADN
Eucaryotes

Pour être contenu dans la cellule, l'ADN doit être compacté. Il existe plusieurs degré de compaction.



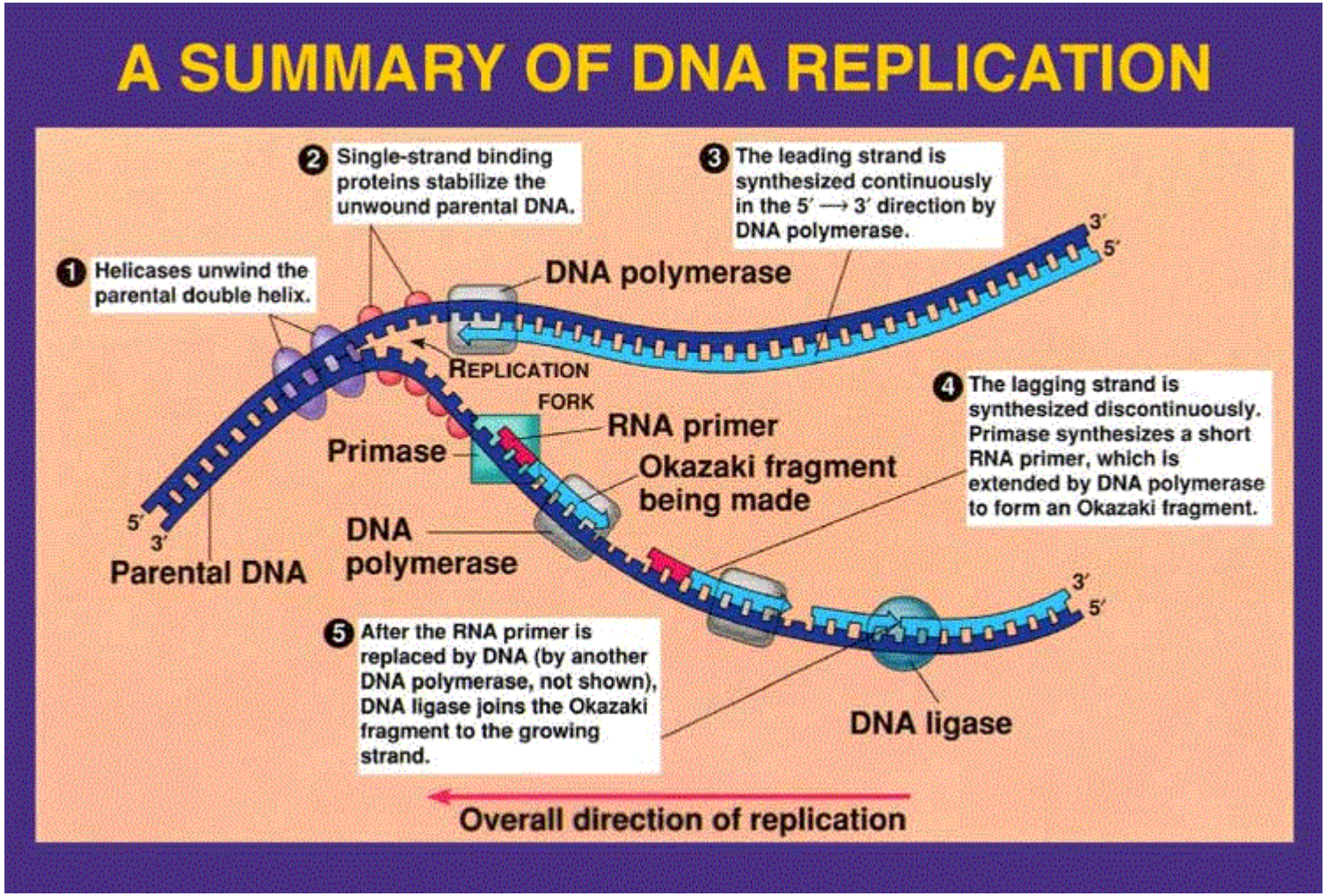
L'ADN : support de l'hérédité

A chaque division cellulaire, la molécule d'ADN double brin doit être dupliquée en deux molécules filles similaires. Ce processus est appelé **la réplication**.

Chacun des deux brins de l'ADN va servir de matrice pour la synthèse du nouveau brin à partir de nucléotides libres. A la suite de la réplication, chacune des deux molécules d'ADN sera constitué d'un brin parental (le brin matrice) et d'un brin néoformé. La réplication est donc dite semi-conservatrice.

La réplication est un processus complexe, faisant intervenir un nombre d'enzymes plus important que ce que l'on pensait au départ.

La réplication



L'ADN est synthétisé par l'ADN polymérase qui progresse toujours dans le sens 5' -> 3' pour le brin en création. Ceci engendre donc une asymétrie du processus pour les deux brins. Un brin est synthétisé de façon continue (le brin précoce ou leading strand) et l'autre de façon discontinue (le brin tardif ou lagging strand).

La réplication

La réplication débute à des sites spécifiques appelés origine de réplication. Suivant les organismes ce nombre peut varier de un à plusieurs milliers. Il n'y a qu'une seule origine de réplication chez les bactéries, une à trois chez les archaea suivant les espèces.

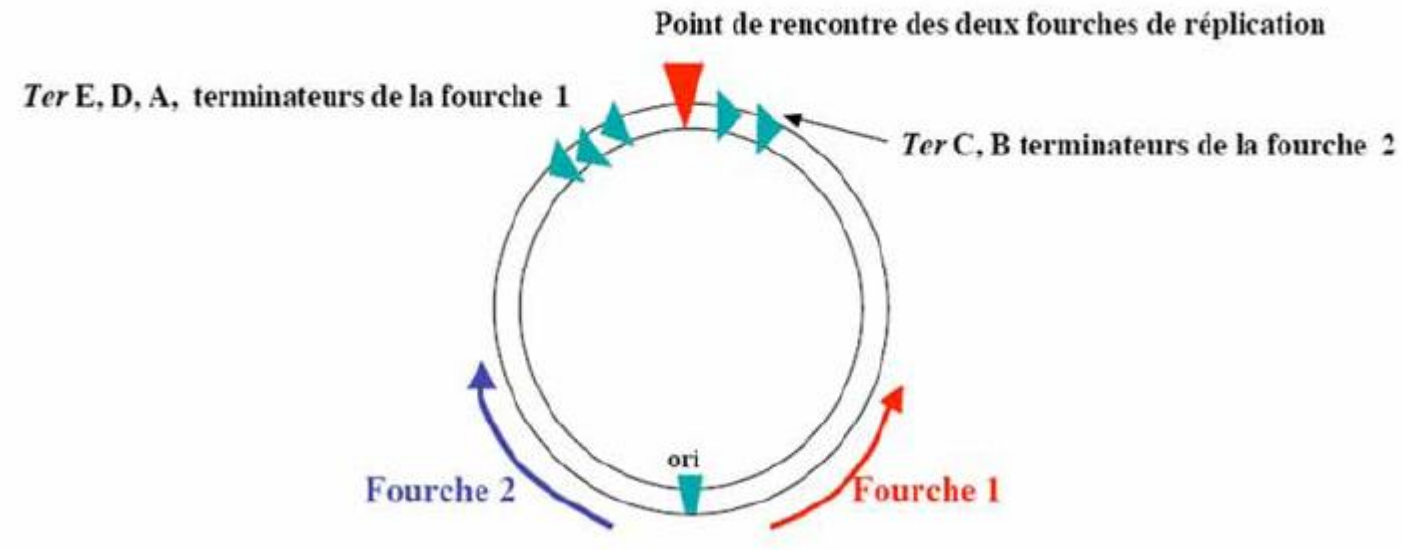
La réplication s'achève différemment selon que le chromosome est circulaire ou linéaire.

Chromosome circulaire : les deux fourches de réplication vont nécessairement se rencontrer et les deux chromosomes devront se séparer.

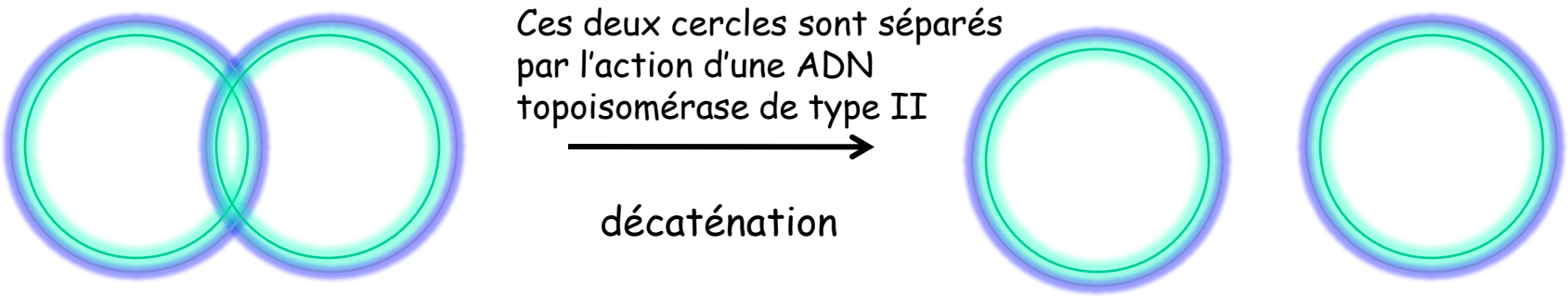
Chez *E. coli,* cette terminaison a lieu dans une région appelée *ter*, constituée de 2 régions ter longues chacune de 22 nucléotides. Ces deux sites sont dans des directions opposées sur le chromosome et ne peuvent être traversés par la fourche de réplication que dans un sens.
Si l'une des fourches de réplication a progressé plus rapidement que l'autre, elle traversera le premier mais pas le second.

La réplication

Terminaison



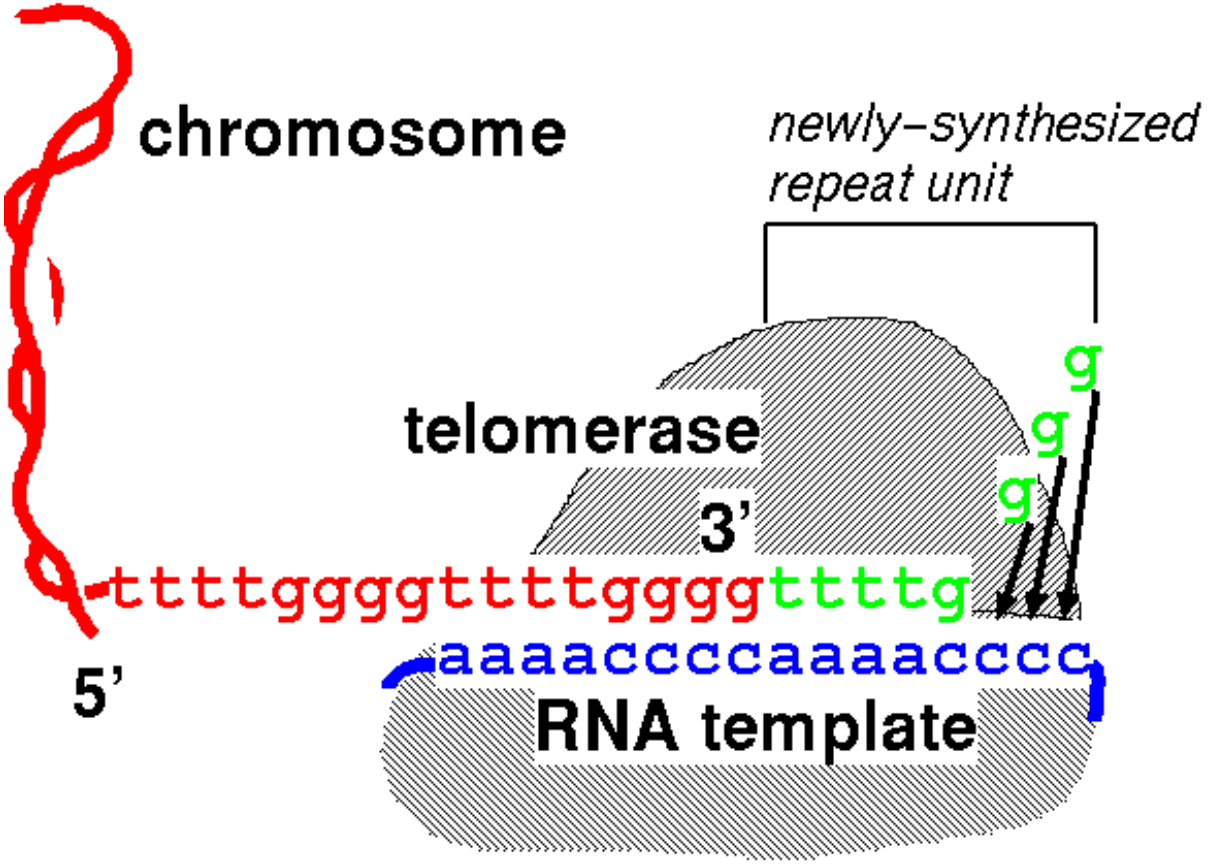
Quand les deux fourches se sont rencontrées, dans certains cas, les deux chromosomes restent liées entre eux.



Caténat

Terminaison chromosome linéaire

Problème sur le brin discontinu :



La réplication

Taux extraordinaire de fidélité : 1 erreur pour 10^{10} bases ajoutées

Si un nucléotide est incorporé par erreur, l'ADN polymérase utilise son activité 3'-5' exonucléasique qui retire le nucléotide en question.

Les ADN polymérases,

activité de polymérisation 5’ ➔ 3’, à partir d’une extrémité 3’-OH
utilise l’énergie des dNTP (3 phosphates) pour crée la liaison covalente
activité exonucléase 3 ’➔ 5’ éventuelle

Bactériennes

Pol I : réparation de l’ADN. Activités 5'→3' et 3'→5' exonucléase, synthèse des fragments d’Okazaki. élimination des amorces d'ARN.

Pol II : réplication de l'ADN endommagée, activité 5'→3' et activité a 3'→5' exonucléase.

Pol III : principale polymérase bactérienne de l'élongation lors de la réplication (brin avancé et fragments d’Okazaki). Constituée de dix sous-unités.

Eucaryotes

ADN Pol α-primase: synthétise les amorces ARN. Pas de fonction exonucléasique.

ADN Pol β: Processus de réparation de l'ADN. Pas de fonction exonucléasique.

ADN Pol δ: Polymérase principale de la réplication de l'ADN chez les Eucaryotes. Activité exonucléasique 3' vers 5' (correction des erreurs et dans des processus de réparation).

ADN Pol ε: Activité polymérase 5' vers 3' et une activité exonucléase 3' vers 5' et intervient dans la réplication et la réparation de l'ADN.

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
Mycoplasma genitalium (bactérie)	0,58	500
Heamophilus influenzae (bactérie)	1,8	1800
Escherichia coli K12 (bactérie)	4,6	4400
Saccharomyces cerevisaie (levure)	12	5800
Drosophila melanogaster (insecte)	180	14700
Caenorhabditis elegans (nematode)	103	20000

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

Différence de taille des génomes

Due bien sur à la différence du nombre de gènes codés par le génome mais pas seulement. En effet, il existe aussi dans le génome des régions non codantes. Cet ADN non codant peut être localisé soit :

- entre les gènes -> régions intergéniques (procaryotes et eucaryotes)
- à l'intérieur des gènes -> introns (eucaryotes)

ADN non codant comprend de multiples structures:

Pseudogènes

Eléments mobiles

Répétitions en tandem

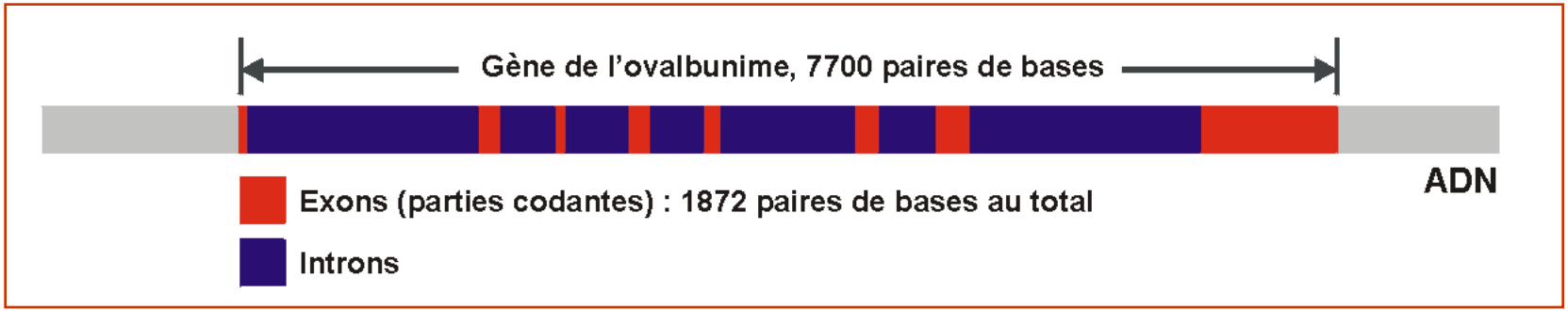
« microARNs »

ADN non codant : les introns

Introns = Séquences non codantes situées à l'intérieur des gènes

Exons = Parties codantes

Un exemple : le gène de l'ovalbumine : 8 exons et 7 introns

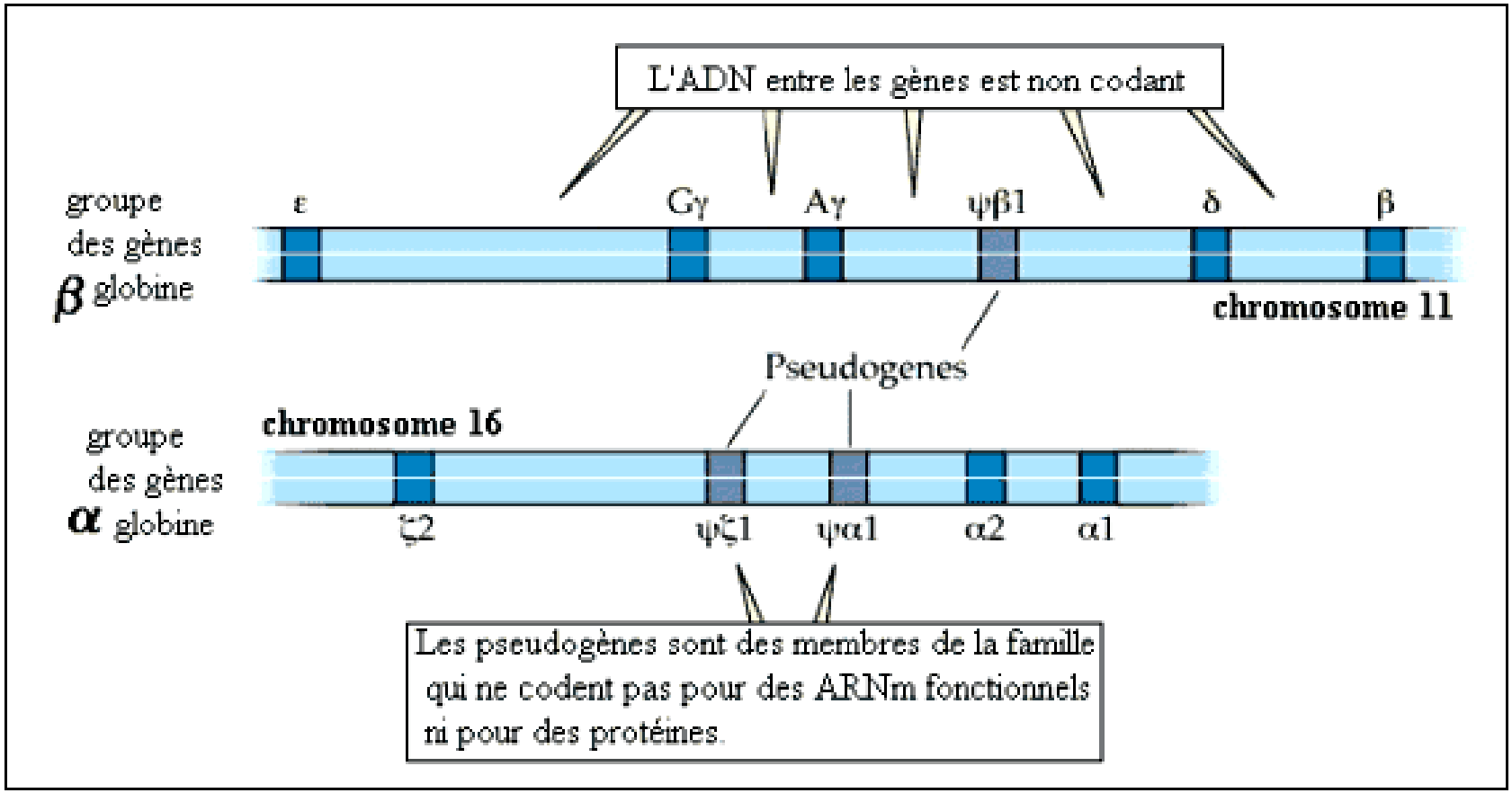


Taille des exons généralement plus petites que celles des introns. Taille moyenne d'un exon interne ~150 pb. Celle des introns varie de 31 à 210 000 pb.

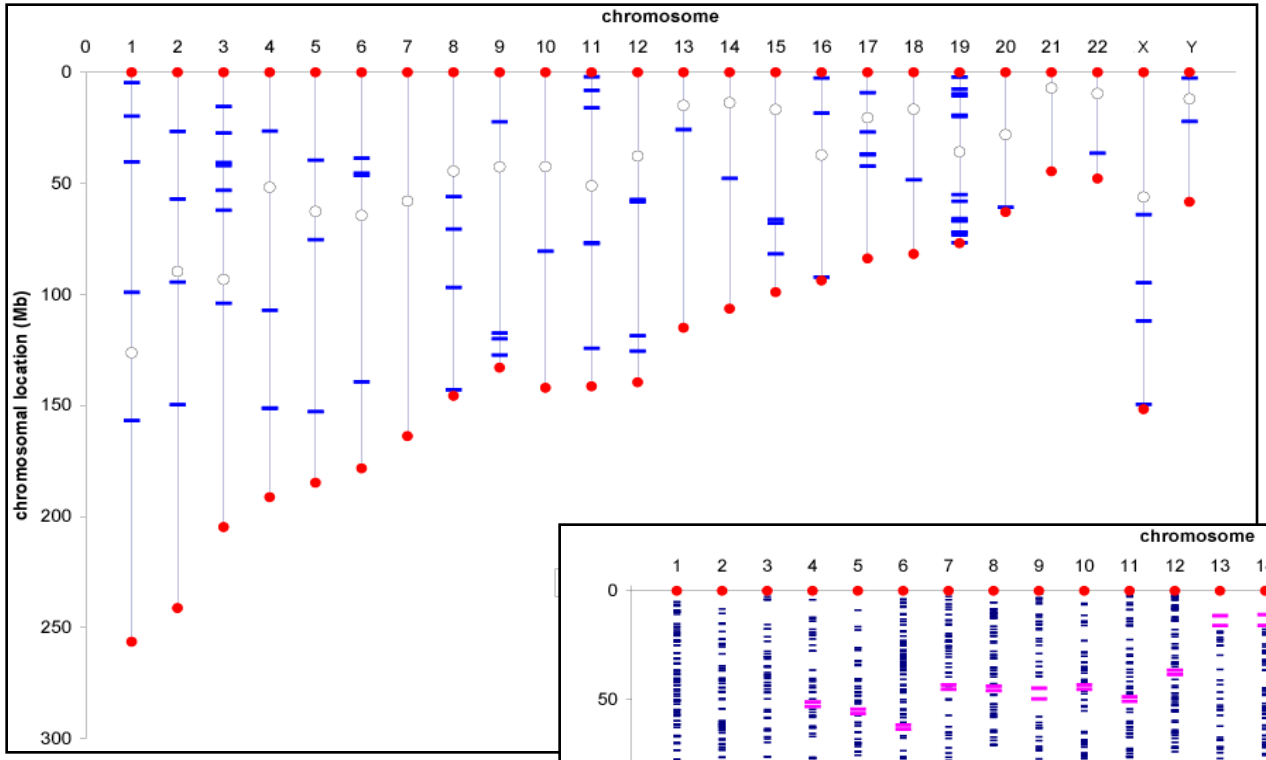
Certains gènes ont leurs séquences constituées d'environ 90% d'introns.
Le nombre d'introns peut être très élevé : exemple le gène du collagène contient 50 introns!

ADN non codant : les pseudogénes

Séquence d'un gène ayant subi des modifications (mutations) conduisant à un ARNm et/ou une protéine non fonctionnels (exemple : présence d'un codon stop, perte ou gain de bases : rupture du cadre de lecture, insertion d'un élément mobile)

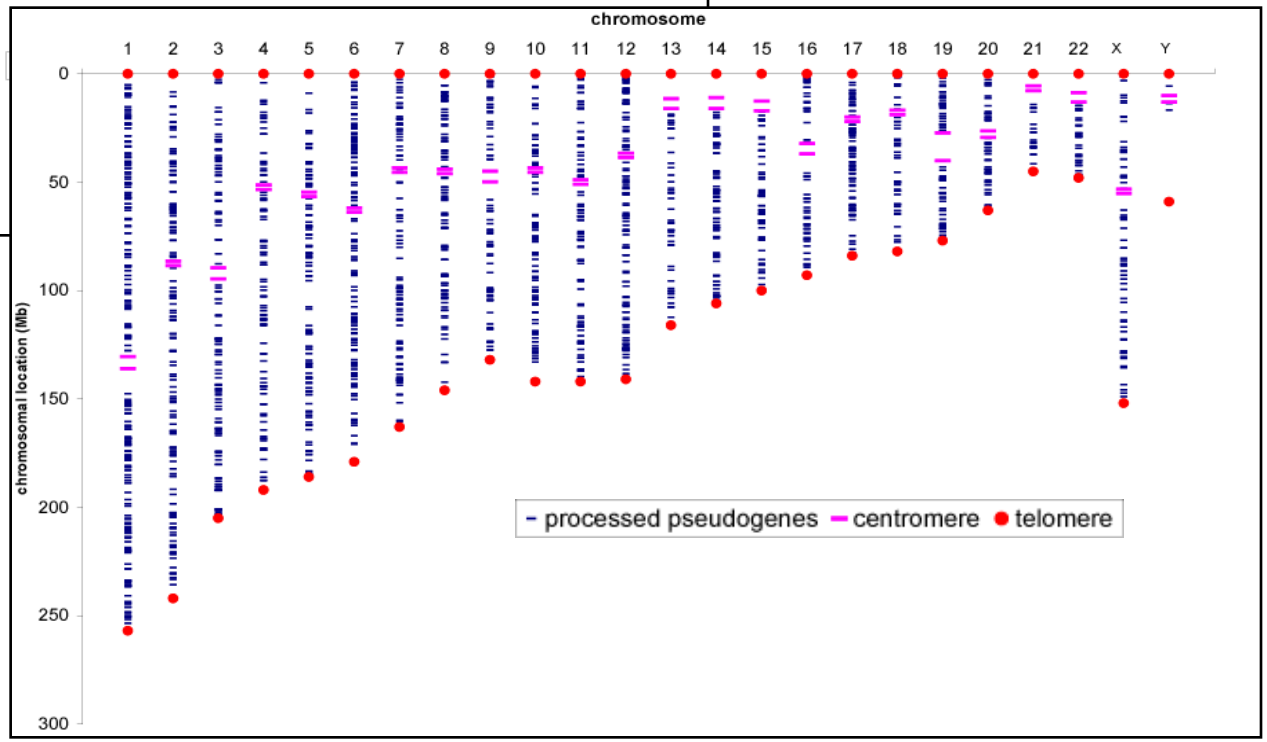


ADN non codant : les pseudogénes



Human Ribosomal Proteins (RP)

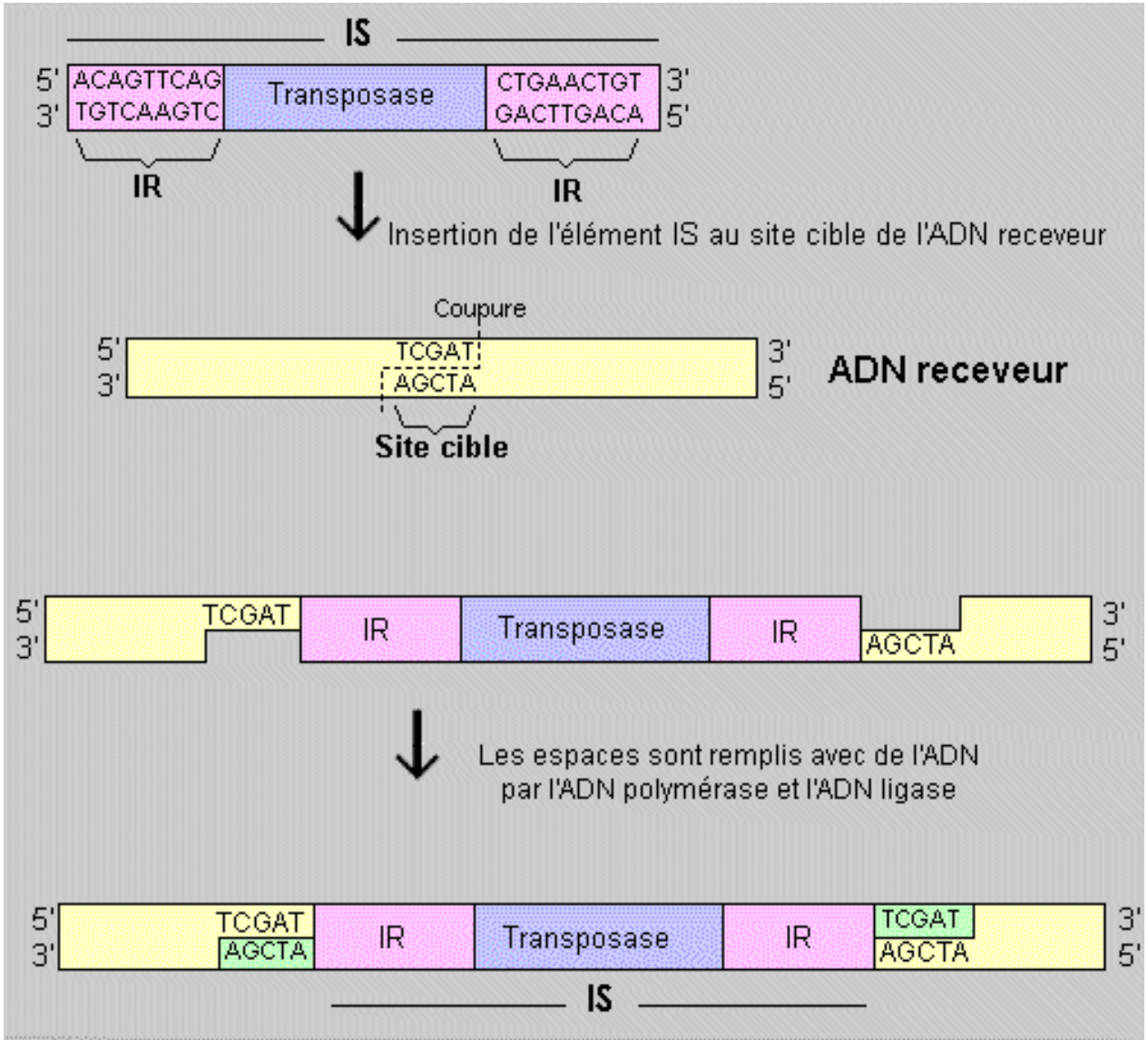
79 Functional RP genes



~ 2000 RP Ψgenes

ADN non codant : les transposons

Ex: variations de l'expression des protéines de surface du trypanosome impliqué dans la malaria.



Transposition d'un élément IS

Légende : - IR : Séquences répétées inversées
- IS : Élément IS (*insertion sequence*)

ADN non codant : les transposons

Les différents types de transposons chez l'homme

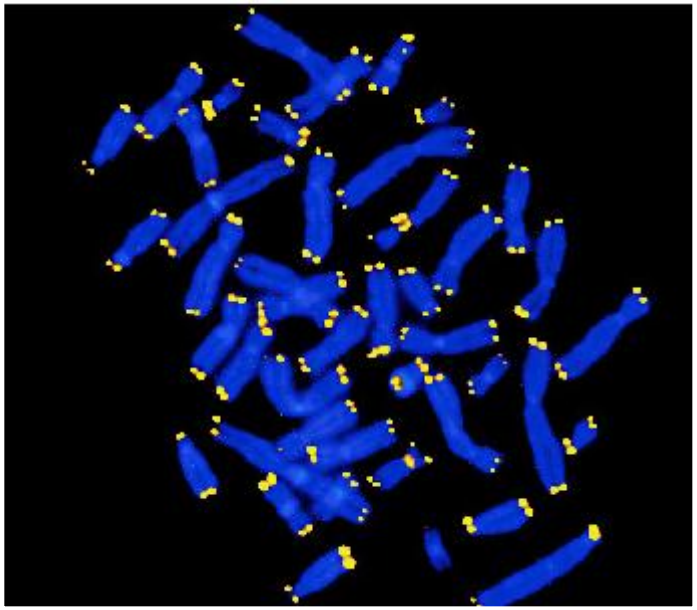
	longueur	Nombre de copies	Fraction du génome
LINEs SINEs	6-8 Kb 100-300pb	850000 1500000	21% 13%
Éléments qui Ressemblent à des rétrovirus	1,5-11Kb	450000	8%
Trace d'anciens transposons	De 80 pb à 3kb	300000	3%

Existence de plusieurs membres des ces différentes familles : Création de répétitions appelées : répétitions dispersées.

ADN non codant

Des séquences très répétées…

Séquences centromériques



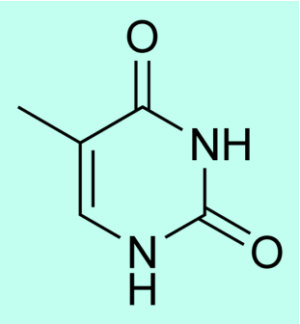
Séquences télomériques

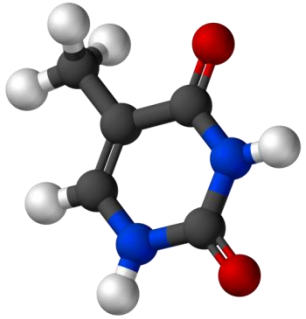
L'acide ribonucléique : l'ARN

L'ARN ou acide ribonucléique est une molécule très proche de l'ADN. Il est aussi formé de nucléotides qui contiennent donc un groupement phosphate, un sucre et une base azotée.

Différence par rapport à l'ADN :

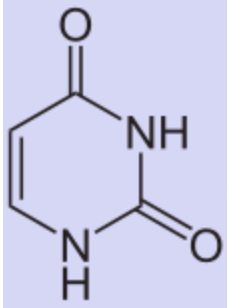
- le sucre = un ribose à la place d'un désoxyribose rendant l'ARN plus instable
- une des 4 bases : la thymine remplacée par l'uracile possédant les mêmes propriétés d'appariements de base avec l'adénine

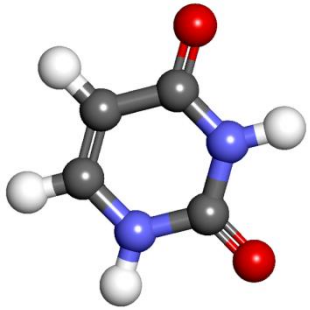




Thymine

rouge : oxygène
bleu : azote
gris : carbone
blanc : hydrogène



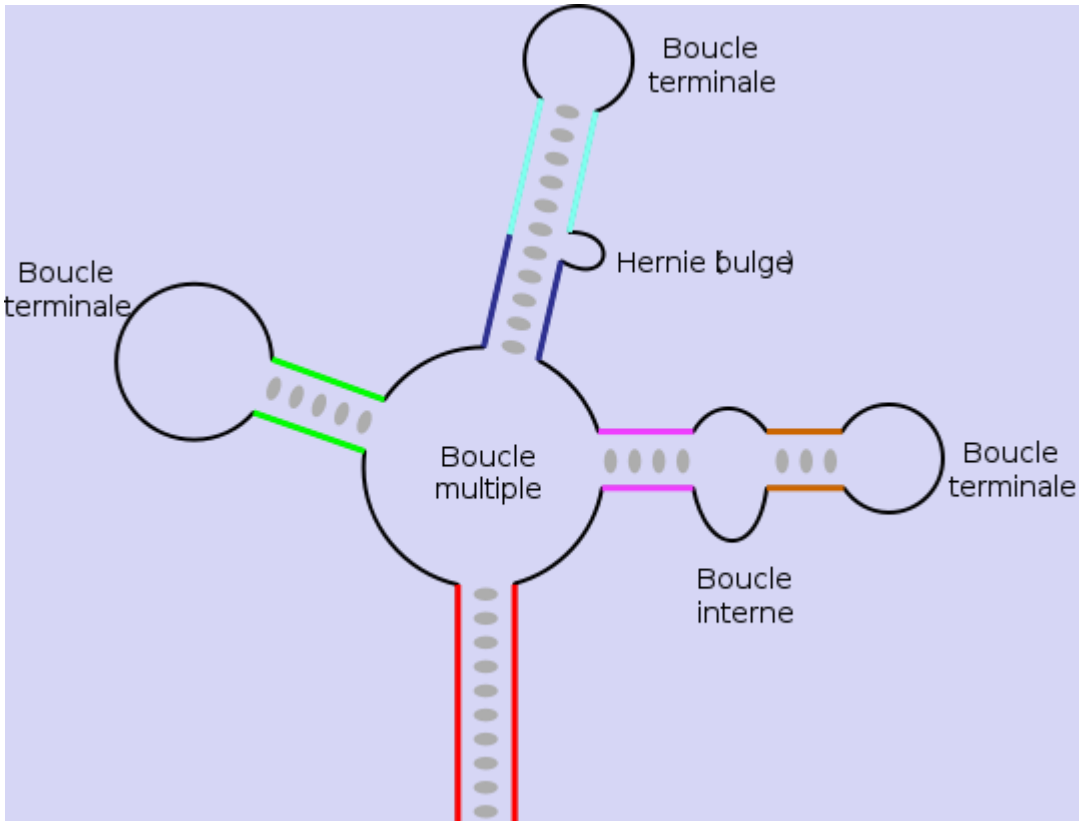


Uracile

La plupart des ARN *in vivo* sont présents sous forme simple brin. Ils se replient par contre souvent sur eux-mêmes formant des appariements internes entre bases. Ces appariements internes forment **la structure secondaire de l'ARN.** Cette structure peut former des appariements à longue distance, formant la **structure tertiaire de l'ARN.**

L'acide ribonucléique : l'ARN

Différents types de structures secondaires de l'ARN :



L'acide ribonucléique : l'ARN

Dans la cellule, l'ARN est produit à partir de la copie de certaines régions d'un brin de l'ADN par un processus appelé **transcription**.

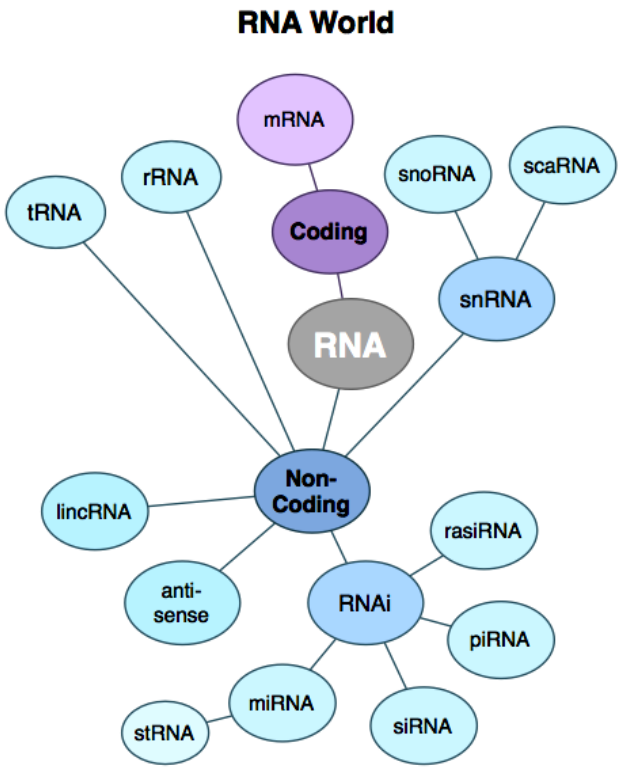
Ces ARN ainsi produits peuvent :

- porter l'information génétique d'un ou plusieurs gènes codant des protéines : **ARN messager (ARNm)** -> support temporaire de l'information génétique.

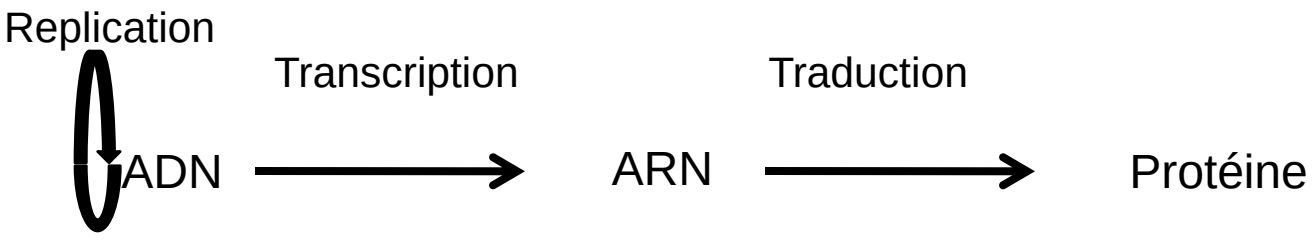
- avoir des fonctions catalytiques **(ARN ribosomique (ARNr), Ribonuclease P,..)**.

- intervenir dans la régulation de l'expression des gènes (ex ARN antisense).

- servir de guide pour des enzymes. Ils sont utilisé comme cofacteur par des protéines pour permettre leur ciblage vers des séquences spécifiques (ex : les snoARN, qui guident les enzymes de modification de l'ARN ribosomique).



Le « dogme central » de la biologie moléculaire

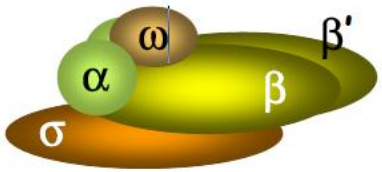


Les flèches indiquent le sens pour le transfert de l'information génétique. Terme « dogme central » employé en 1956 par Francis Crick.
L'ARN peut parfois servir de matrice pour la synthèse d'un brin d'ADN complémentaire. Cependant, évènement rare comparé à la grande quantité de molécules d'ARN synthétisées à partir d'une matrice ADN.
Donc ce flux orienté de l'information reste d'actualité.

La transcription

Elle est réalisé 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 α, un exemplaire de chacune des sous-unité β, β' et ω. le facteur σ permet la reconnaissance des séquences spécifiques sur l'ADN, appelée **promoteur**.



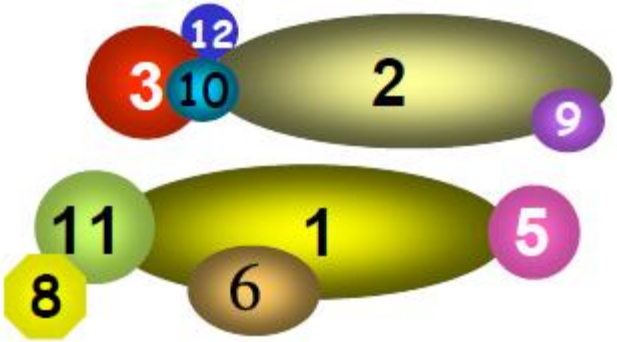
ARN polymérase procaryote

Chez les eucaryotes, présence de trois ARN polymérases :

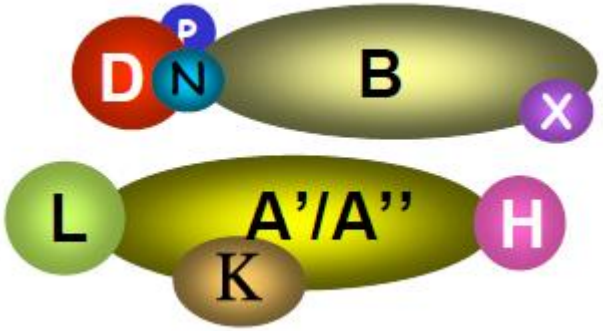
ARN polymérase	Synthèse	Constituants
I	- ARN pré-ribosomique	Ensemble de12 sous-unités plus 2 autres polypeptides
II	- ARN messagers - snoRNA*	Ensemble de12 sous-unités
III	-Les ARN de transfert -ARN ribosomique 5S -Petits ARN	Ensemble de12 sous-unités plus 5 autres polypeptides

La transcription

Chez les archaea, présence d'une seule ARN polymérase similaire à l'ARN polymérase II des eucaryotes. Elles est constituée de 8 à 10 sous-unités



ARN polymérase eucaryote



ARN polymérase archaea

La transcription

3 étapes :

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

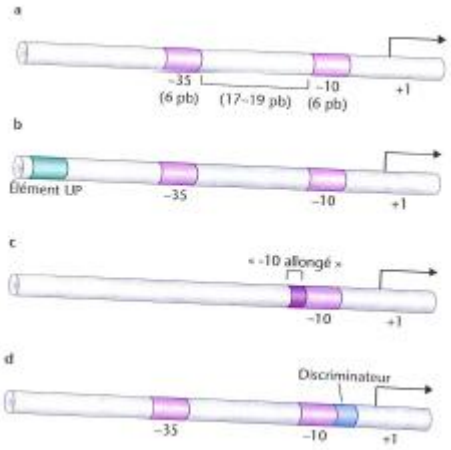
Transcription des gènes codant pour des protéines

Chez les bactéries :

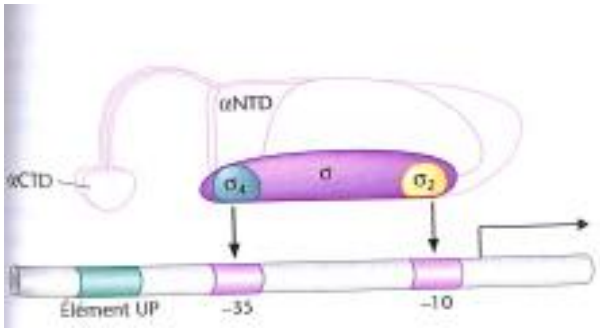
L'initiation :

Le facteur σ de la RNA polymérase reconnait 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).

Promoteurs bactériens



Recrutement du cœur de l'ARN polymérase au promoteur

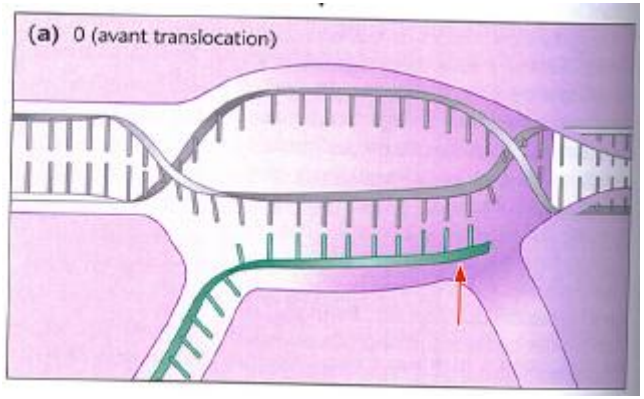


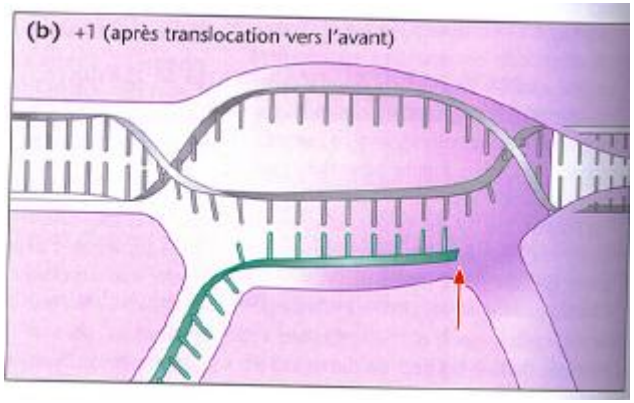
La transcription

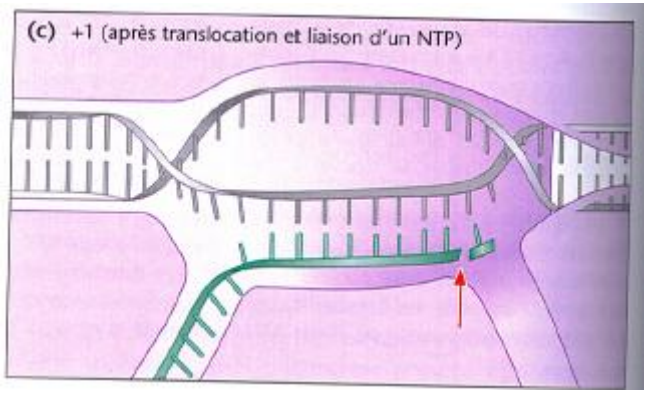
Transcription des gènes codant pour des protéines

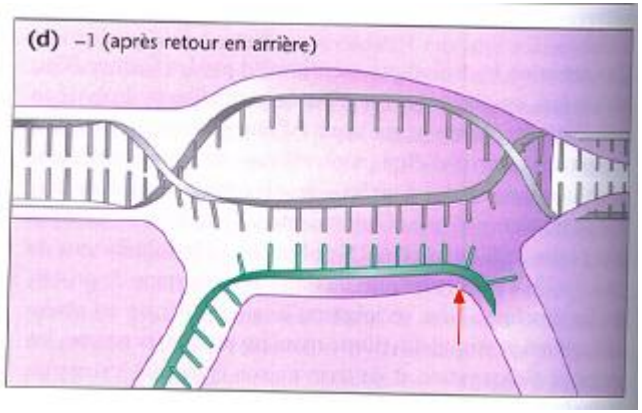
Chez les bactéries :

L'élongation : construction du brin ARN par copie du brin matrice ADN. La RNA polymérase peut revenir en arrière pour corriger une erreur.







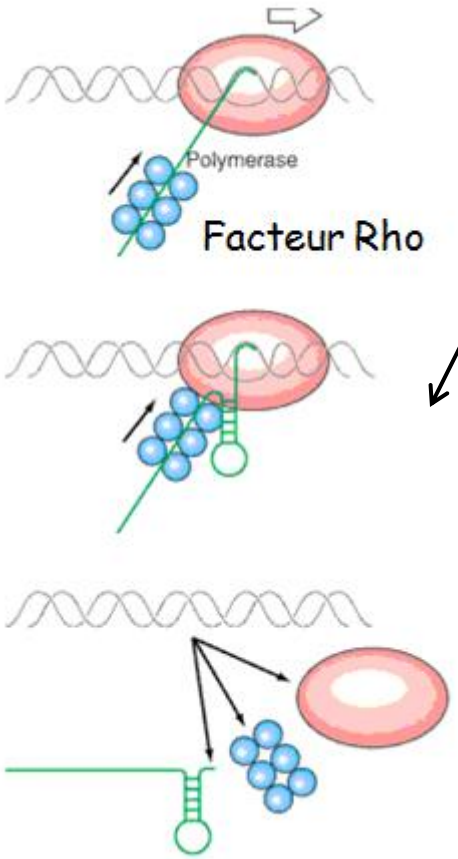


La transcription

Transcription des gènes codant pour des protéines

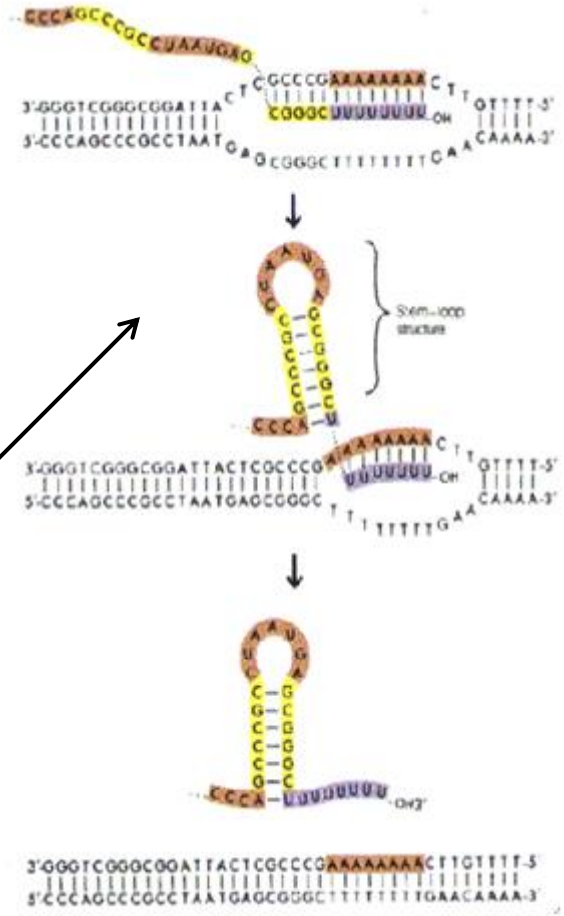
Chez les bactéries :

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.



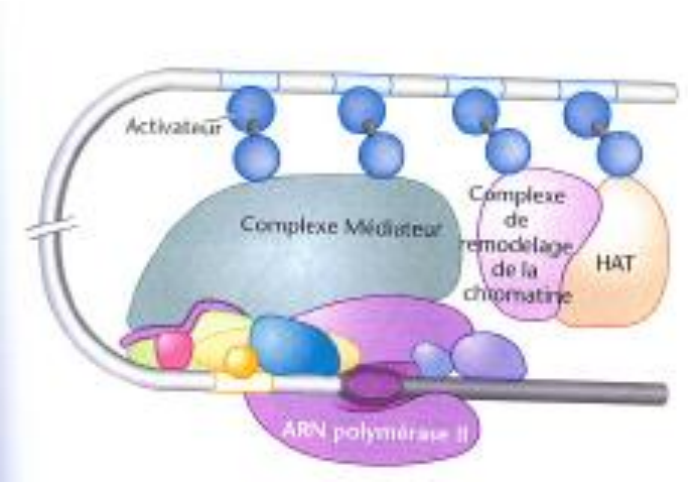
La transcription

Transcription des gènes codant pour des protéines

Chez les eucaryotes réalisée par l'ARN polymérase II :

L'initiation :

Elle nécessite un plus grand nombre de facteurs que dans la cas des procaryotes où il n'y en a qu'un seul. Ces facteurs sont appelés **facteurs généraux de transcription**. Cependant, l'ADN étant empaqueté sous forme de chromatine, des facteurs supplémentaires sont requis pour que les facteurs généraux puissent avoir accès au promoteur. Des protéine régulatrices de la transcription, appelées activateurs, aident au recrutement de l'ARN polymérase au promoteur. Le recrutement est réalisé grâce à des interactions entre ces activateurs, des facteurs modifiant la chromatine et des parties de la machinerie de transcription.
Une de ces interactions implique le **Médiateur**.
C'est un complexe formé de nombreuses protéines (plus de 20). Il est associé par une de ces faces avec la queue CTD de la grande sous-unité de l'ARN polymérase (absente chez les procaryotes) et ces autres faces sont impliquées dans des interactions avec les activateurs.



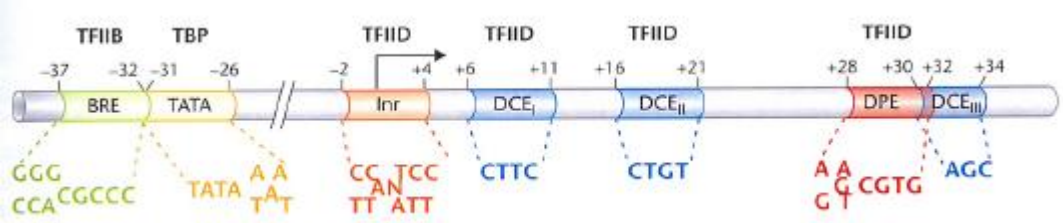
La transcription

Transcription des gènes codant pour des protéines

Chez les eucaryotes réalisée par l'ARN polymérase II :

L'initiation :

Le promoteur minimum eucaryote (plus petit ensemble d'éléments de séquence nécessaire pour assurer l'initiation de la transcription) a une longueur de 40 à 60 nucléotide et est localisé en amont ou en aval du site d'initiation.



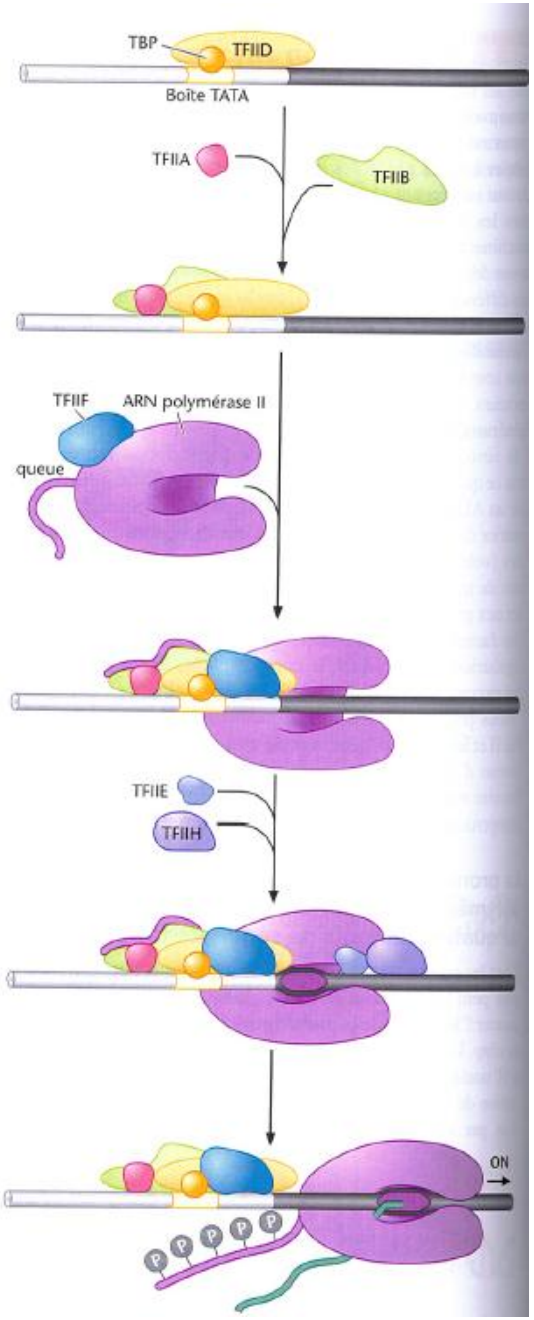
L'élément BRE est reconnu par le facteur TFIIB
La boîte TATA si présente pas d'élément DPE
Inr : élément initiateur est le plus courant, présent avec la TATA ou le DPE
DCE : Downstream core element (élément en aval de la polymérase)
DPE : Downstream promoter element (élément en aval du promoteur) quand présent pas la boîte TATA

La transcription

Transcription des gènes codant pour des protéines

Chez les eucaryotes réalisée par l'ARN polymérase II :

L'initiation : assemblage progressif du complexe



La polymérase a quitté le promoteur. Pour cela, il faut que sa « queue » soit phosphorylée. Elle a entamé sa phase d'élongation

La transcription

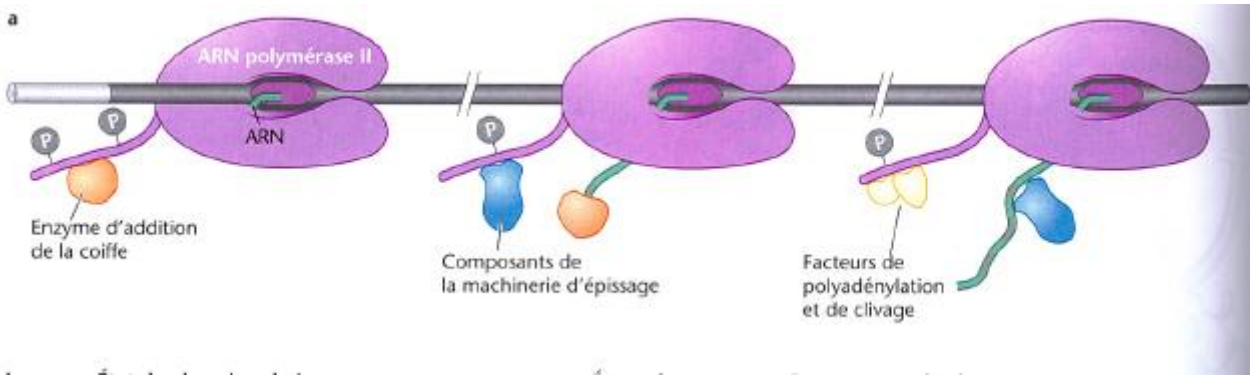
Transcription des gènes codant pour des protéines

Chez les eucaryotes réalisée par l'ARN polymérase II :

La maturation de l'ARN :

Une fois transcrit, l'ARN eucaryotes doit subir des modifications incluant :

- l'ajout de la coiffe à l'extrémité 5'
- l'excision des introns (régions non codantes séparant les régions codantes dans les gènes eucaryotes) pour produire l'ARNm mature. Ce processus est appelé épissage
- l'ajout d'une queue polyA à l'extrémité 3' : polyadénylation



La transcription

Transcription des gènes codant pour des protéines

Chez les eucaryotes réalisée par l'ARN polymérase II :

Terminaison:

Liée à l'étape de polyadénylation. Quand la polymérase atteind la fin d'un gène, elle rencontre des séquences spécifiques qui, une fois traduites en ARN, sont reconnues par les facteurs de polyadénylation. Ceci engendre les 4 évènements suivants :

- le clivage de l'ARNm
- l'ajout d'un grand nombre de résidus adénine à l'extrémité 3'
- la dégradation de l'ARN encore associé à la polymérase
- la terminaison de la transcription.

La transcription

Chez les eucaryotes, deux autres ARN polymérases I et III.

Elles reconnaissent des promoteurs différents en utilisant un groupe de facteurs de transcription spécifique.

L'ARN PolI est nécessaire pour la transcription d'un seul gène, celui codant pour le précurseur de l'ARNr. Plusieurs copies de ce gène dans le génome et il est exprimé à un niveau bien supérieur à celui des autres gènes. D'où une ARN polymérase dédiée.

Structure du promoteur

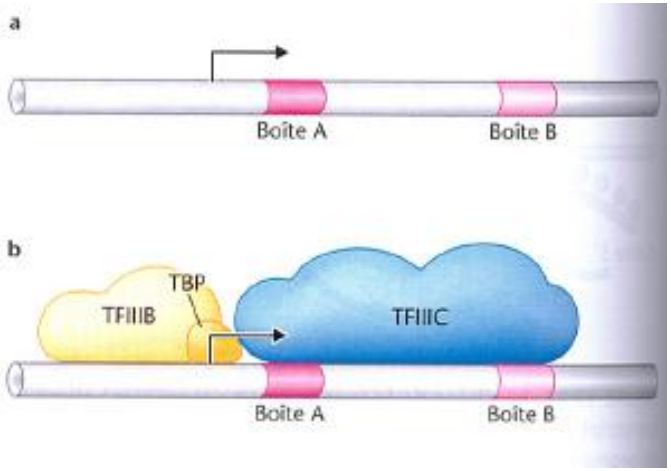
Deux élements :
Le promoteur central localisé autour du site de début de transcription
L'élément de contrôle (UCE) localisé en amont de 100 à 150 pb (du moins chez l'homme)

Facteurs de transcription de l'ARN PolI chez l'homme (différent chez la levure).

Le facteur SL1 comprend la TBP et 3 autres unités spécifiques. Ce facteur se fixe sur le promoteur central mais il ne fixe l'ADN qu'en présence du facteur UBF lié à l'UCE.

La transcription

Les promoteurs de l'ARN Pol III présentent différentes structures mais ont comme caractéristique d'être localisés après le site d'initiation de la transcription.
Certains promoteurs comme ceux des gènes codant les ARNt sont composés de deux boîtes A et B ; d'autres contiennent la boîte A et une boîte C (ex, le gène codant l'ARNr 5S); d'autres contiennent une boîte TATA comme ceux de la Pol II.
Ces promoteurs sont reconnus par des facteurs spécifiques : TFIIIB et TFIIIC pour les gènes d'ARNt plus le facteur TFIIIA pour le gène d'ARNr 5S.



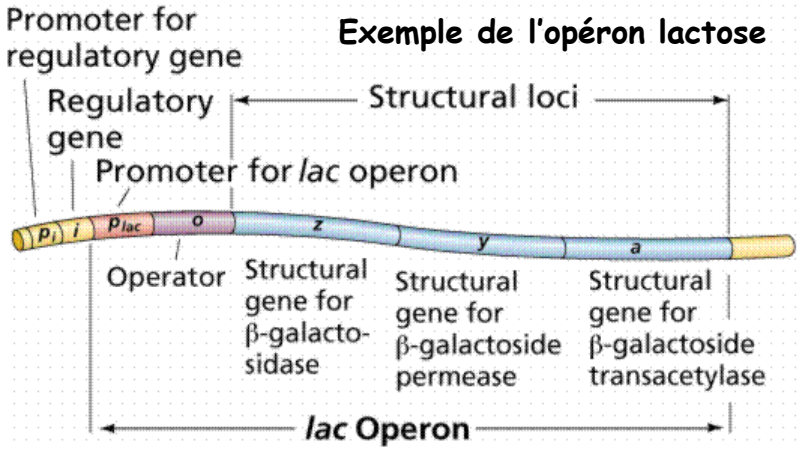
Promoteur schématisé d'un gène de levure codant un ARNt

La transcription

Différence entre procaryotes et eucaryotes :

• localisation : la transcription a lieu dans le cytoplasme chez les bactéries car pas de compartiment. Comme l'étape suivante de traduction a lieu au même endroit, elle peut commencer avant que la transcription soit terminée. Chez les eucaryotes, la transcription se déroule dans le noyau ainsi que sa maturation. L'ARNm mature est ensuite exporté dans le cytoplasme où il est traduit.

• Dans la majorité des génomes procaryotes, les gènes sont organisés en opérons. L'opéron comprend : i) une région régulatrice (opérateur), région ADN sur laquelle se fixe un activateur ou un répresseur, ii) un promoteur commun et iii) un ou plusieurs gènes. Ces gènes sont donc contrôlés de façon coordonnées et ils sont co-transcrits formant un seul ARNm dit polycistronique.
Chez les eucaryotes, les gènes sont transcrits de façon individuelle donnant un ARNm appelé monocistronique.

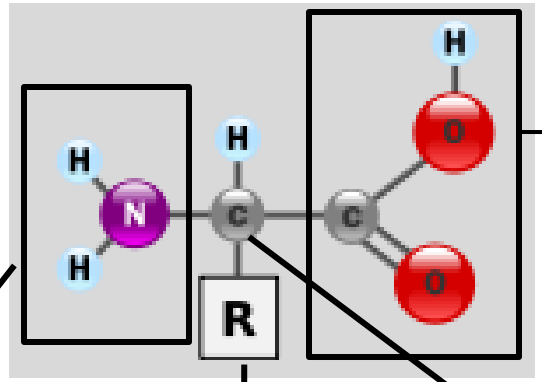


La traduction

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

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

Structure générique d'un acide aminé



Groupe carboxyle
-COOH

Groupe amine –NH2

Carbone α

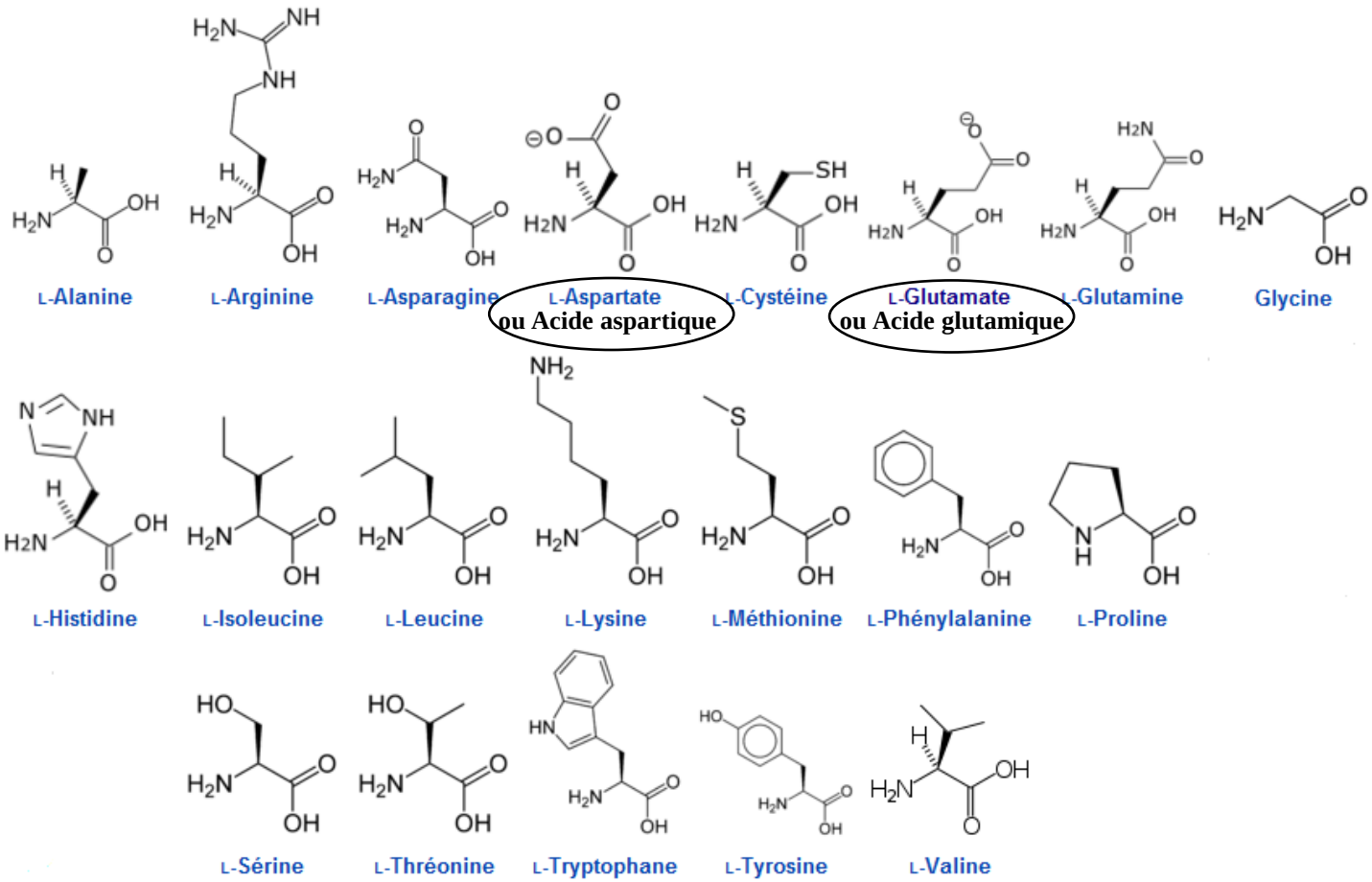
Chaîne latérale identifiant
l'acide aminé

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)

Les acides aminés

Les acides aminés possèdent des propriétés physicochimiques :

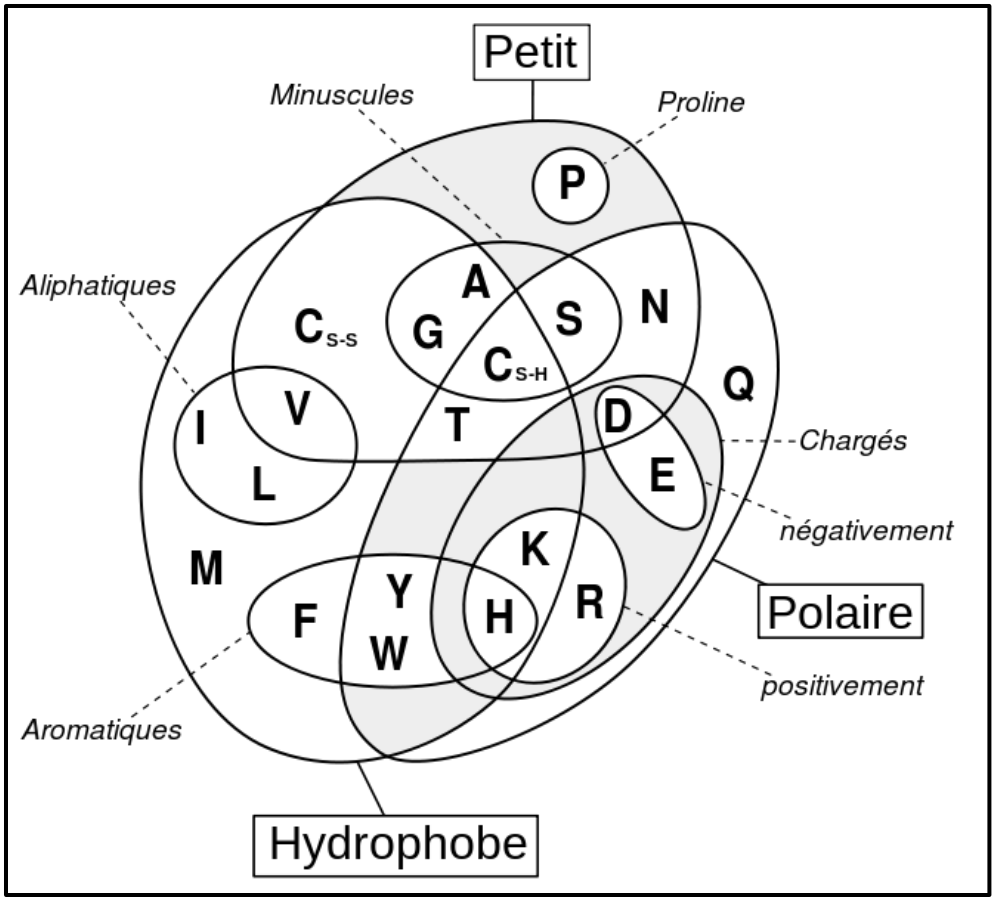


Diagramme de Venn des propriétés des acides aminés

Acide aminé polaire ou hydrophile, localisé le plus souvent à la surface des protéines (certains sont chargés)

Acide aminé hydrophobe ou apolaire ou aliphatique tendent à occuper le cœur de la protéine.

Les acides aminés aromatiques possèdent un cycle dans leur chaîne latérale.

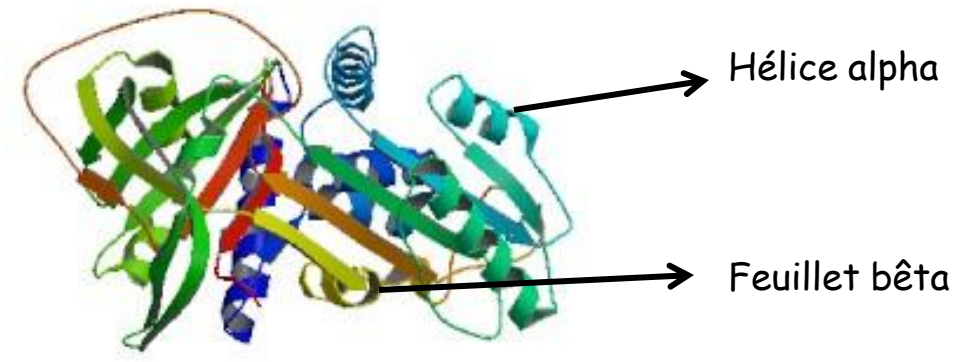
La cystéine permet d'établir des ponts disulfures

Les protéines

Les différents niveaux de structures :

- la suite linéaire des acides aminés, appelée aussi séquence protéique, constitue la **structure primaire.**
- la protéine peut se replier sur elle-même pour former des structures secondaires dont les plus connues sont les hélices alpha et les feuillets bêta.
- L'agencement des différentes structures secondaires forme la structure tertiaire ou structure 3D.

C'est sous sa forme repliée (structure 3D) que la protéine est active. Des protéines appelées chaperonnes facilitent ou sont nécessaires pour le repliement correct de la protéine. Si la protéine est mal repliée, elle sera dégradée.



α-1-antitrypsin

La traduction

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

La partie codantes 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.
Pyrrolysine	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 *Ambre*			UAG.
STOP *Ocre*			UAA.
STOP *Opale*			UGA.

La traduction

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

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

La traduction

4 composants principaux font partie de la machinerie de traduction :

- l'ARN messager (ARNm) contenant la séquence codant la protéine ainsi que des régions de reconnaissance pour l'initiation et la terminaison de la traduction.

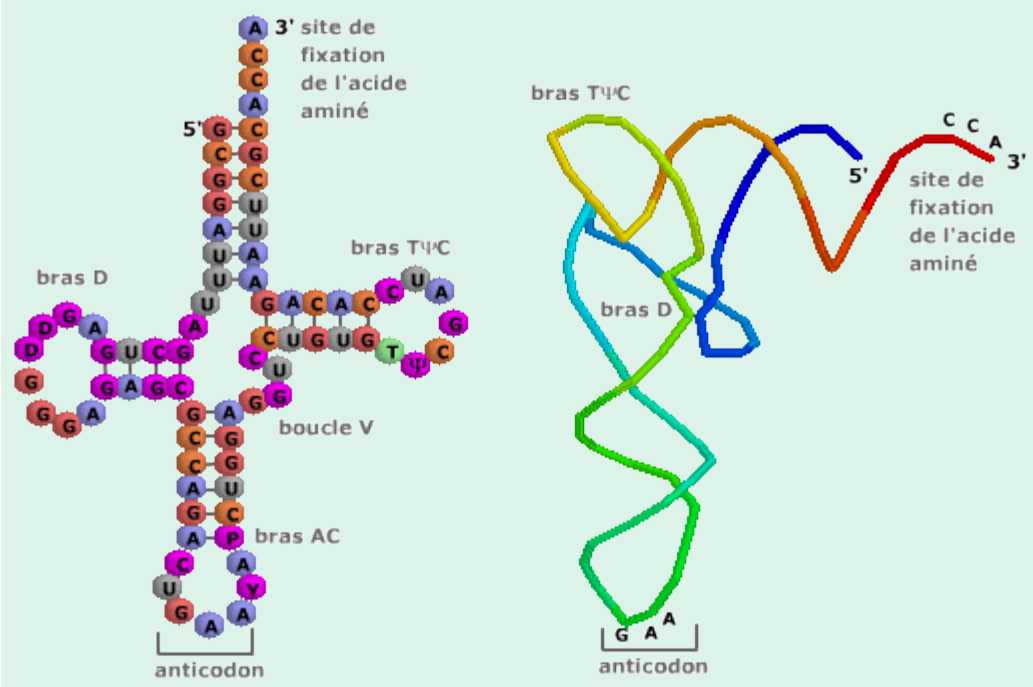
- les ARN de transfert (ARNt) qui une fois chargé de leur acide aminé (processus réalisé par les aminocyl-ARNt- synthétases) vont mettre en relation un codon et un acide aminé.

- 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.



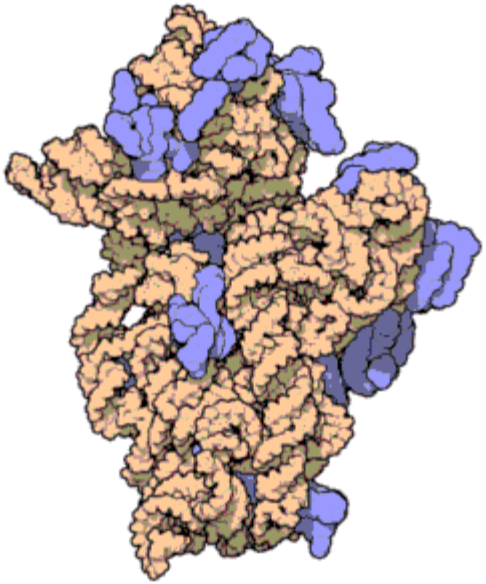
Structures secondaire et tertiaire d'un ARNt

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.

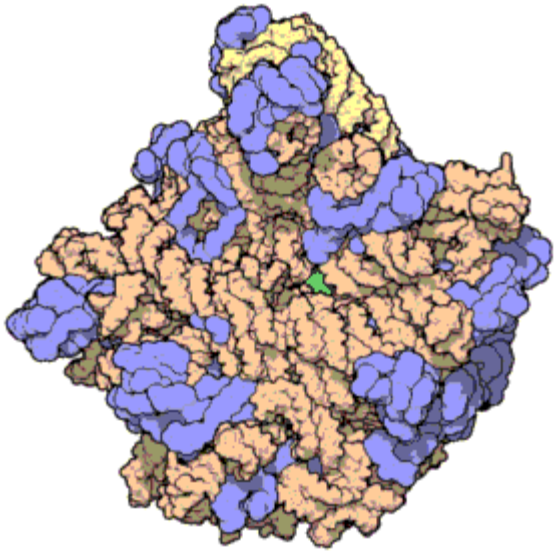
La traduction : le ribosome

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 proteines, orange ARNr 16S



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

La traduction : le ribosome

Les eucaryotes ont des ribosomes 80S constitué d'une petite sous-unité 40S et d'une grande sous-unité 60S. La petite sous-unité est composée de l'ARNr 18S (~1900 nt) et de 33 protéines. La grande sous-unité possède trois molécules d'ARNr, l'ARNr 5S (~120 nt), l'ARNr 28S (~4700 nt), et l'ARNr 5.8S (~160 nt) et de ~49 protéines.

Les ribosomes trouvés dans les mitochondries et chloroplastes forment aussi une particule 70S. Ces organelles sont supposées descendre de bactéries et auraient été acquises par endosymbiose.

La petite sous-unité contient le centre de décodage dans lequel les ARNt chargés décodent les codons de l'ARNm

La grande sous-unité contient le centre responsable de la formation des liaisons peptidiques entre acides aminés et donc de la croissance de la chaîne peptique.

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 :

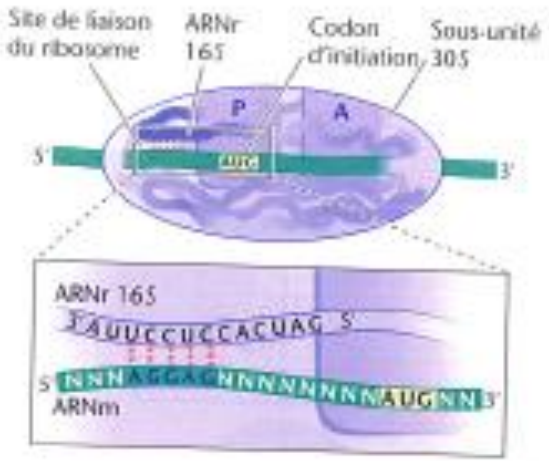
- l'initiation qui charge le ribosome sur l'ARNm
- l'élongation qui conduit à la synthèse de la protéine
- la terminaison conduisant à la dissociation du ribosome de l'ARNm

La traduction

L'initiation :

➢ chez les bactéries :

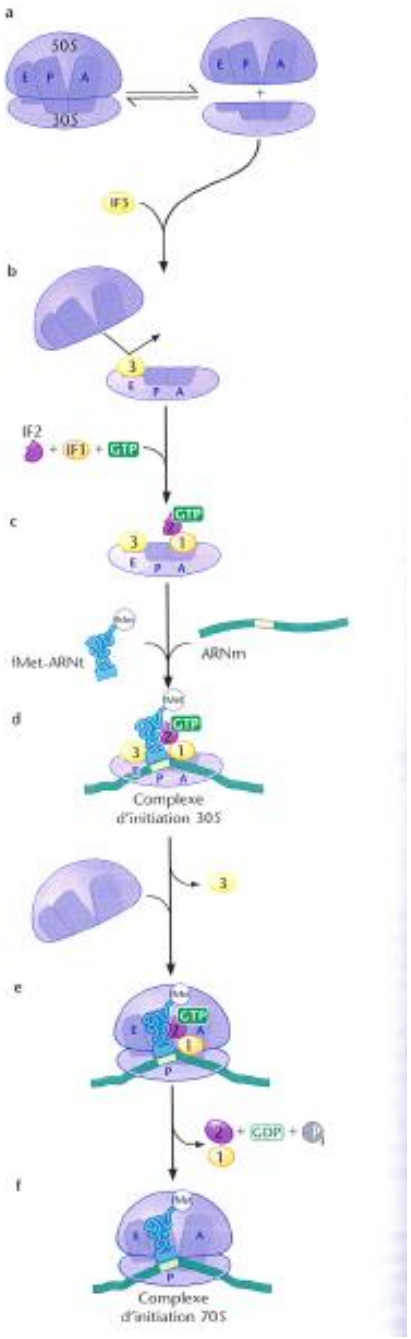
La petite sous-unité est chargée en premier sur l'ARNm, ceci grâce à un appariement des bases d'une région de l'ARNr 16S avec le RBS. Pour un RBS idéalement situé, le codon initiateur (AUG, GUG ou UUG) se trouve situé au site P du ribosome (et non au A comme pour l'élongation). Ceci requiert un ARNt 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'apparient 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

La traduction

Résumé de l'étape d'initiation de la traduction chez les bactéries

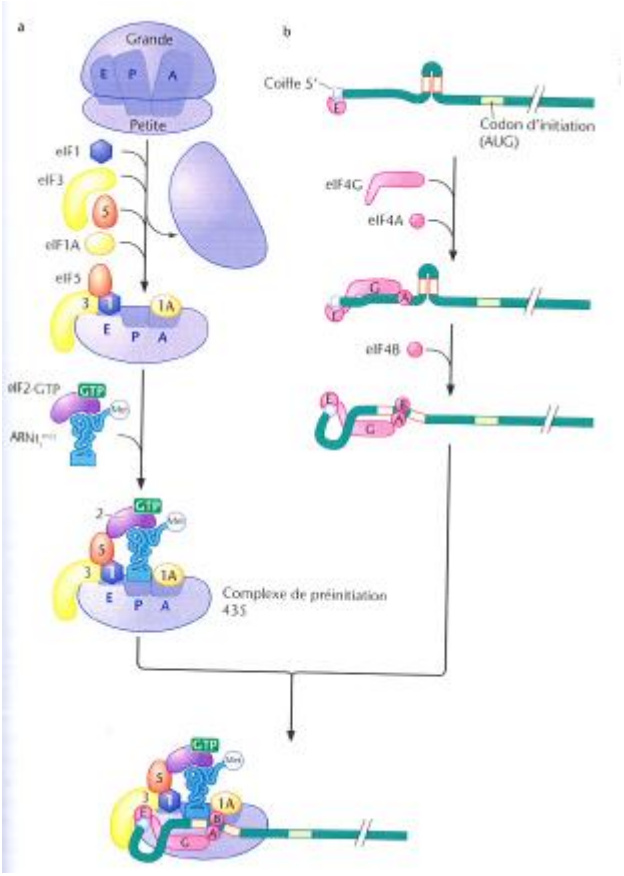


La traduction

L'initiation :

➢ chez eucaryotes:

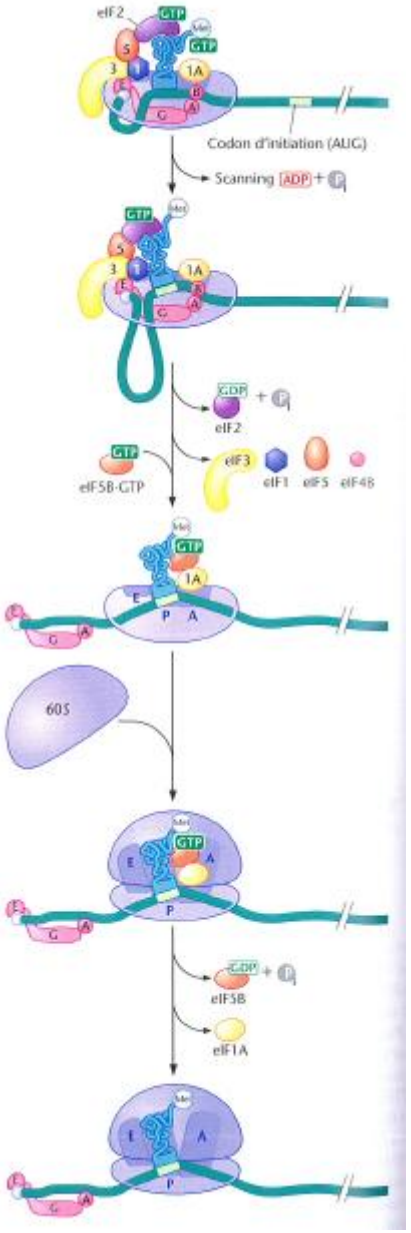
La petite sous-unité est également chargée en premier sur l'ARNm. Cependant, elle est déjà associée avec l'ARNt initiateur puis est recrutée au niveau de la coiffe de l'ARNm. Ensuite, elle « scanne » l'ARNm dans le sens 5'-3' pour atteindre le premier AUG qui est reconnu comme codon initiateur. La présence d'une séquence de Kozak (G/ANNAUGGNN) augmente l'efficacité de la traduction (pas toujours présente).
La grande sous-unité est recrutée après l'appariement de l'ARNt initiateur avec le codon initiateur conduisant à la formation du complexe 80S.



Recrutement de la petite sous-unité sur l'ARNm eucaryote

La traduction

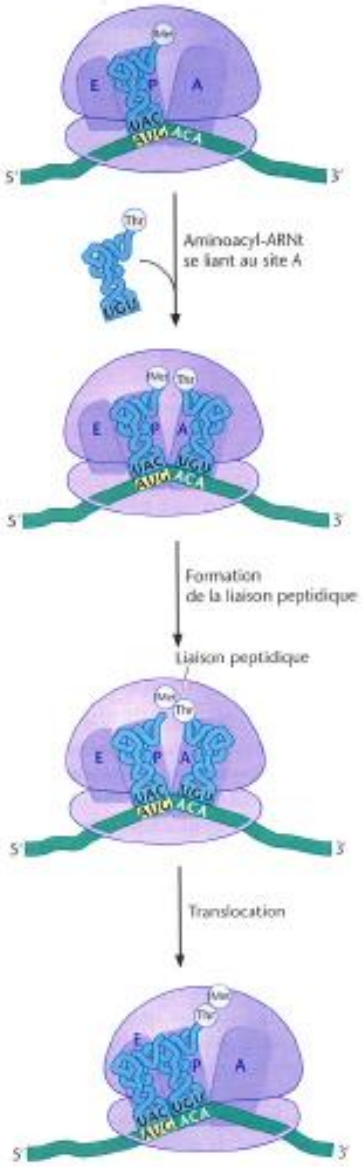
Recrutement de la grande sous unité pour former le complexe 80S sur l'ARNm eucaryote



La traduction

Résumé de l'étape des étapes de l'élongation.

Mécanisme très conservé entre cellules procaryotes et eucaryotes.



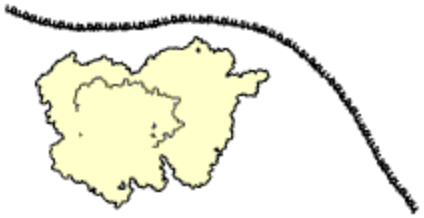
La traduction

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 et par le seul facteur de cette classe chez les eucaryotes. Ce facteur stimule l'hydrolyse du polypeptide et du peptidyl-ARNt, libérant ainsi le peptide complet.
D'autres facteurs achèvent la terminaison en libérant l'ARNm et en dissociant les deux sous-unités du ribosome.
La petite sous-unité est alors prête à recommencer un nouveau cycle de synthèse polypeptidique.

La traduction

Animation : étape d'élongation et d'adressage de la protéine au réticulum endoplasmique. Ribosome en vert, ARNt en bleu foncé, autres protéines impliquées en bleu clair



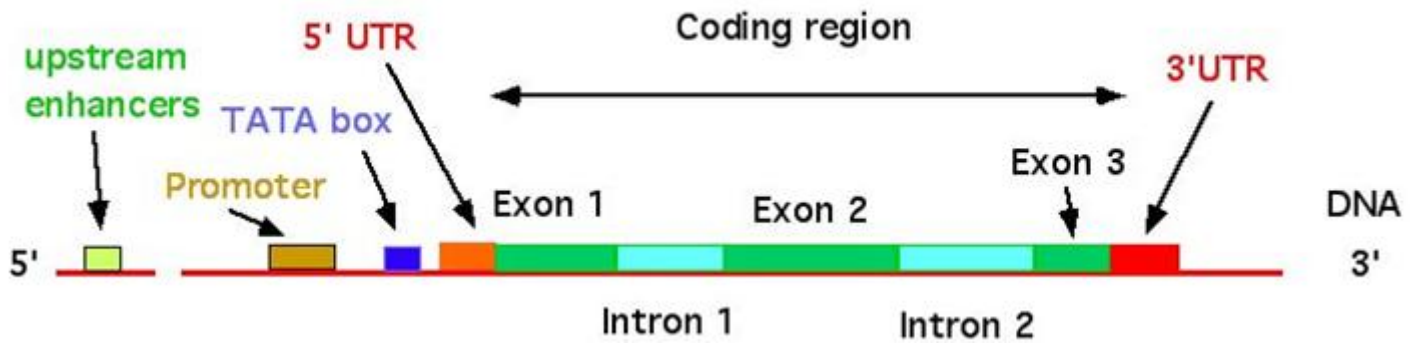
Chez les eucaryotes, un complexe ribonucléoprotéique [ARN 7S / 6 polypeptides] appelé "particule de reconnaissance du signal" ("signal recognition particle" - SRP) se fixe au peptide signal qui émerge du ribosome (séquence en début de protéine qui indique à la cellule le compartiment vers lequel l'adresser) , ralentissant ainsi l'élongation de la chaîne polypeptidique en cours de biosynthèse. Le SRP se fixe au récepteur du SRP. La chaîne polypeptidique en cours d'élongation emprunte le canal de translocation (pore dans la membrane du RE.

Différences procaryote - eucaryotes

Structure du gène :

Procaryote

Eucaryote

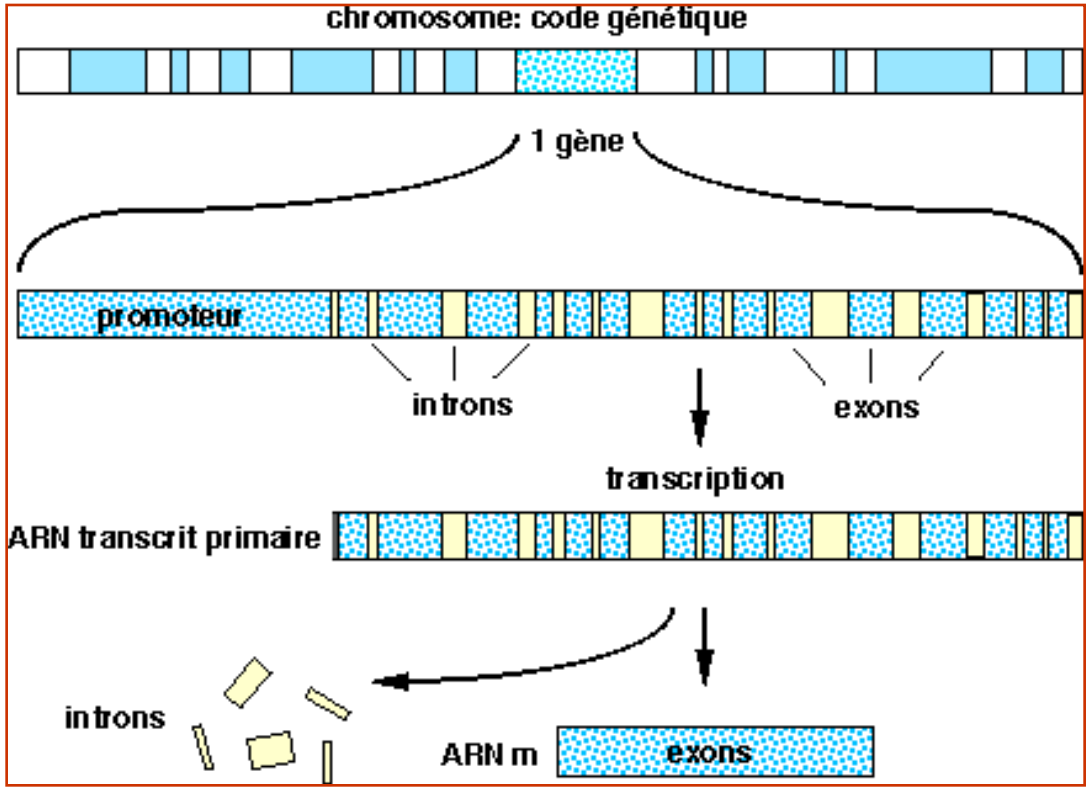


Différences procaryote - eucaryote

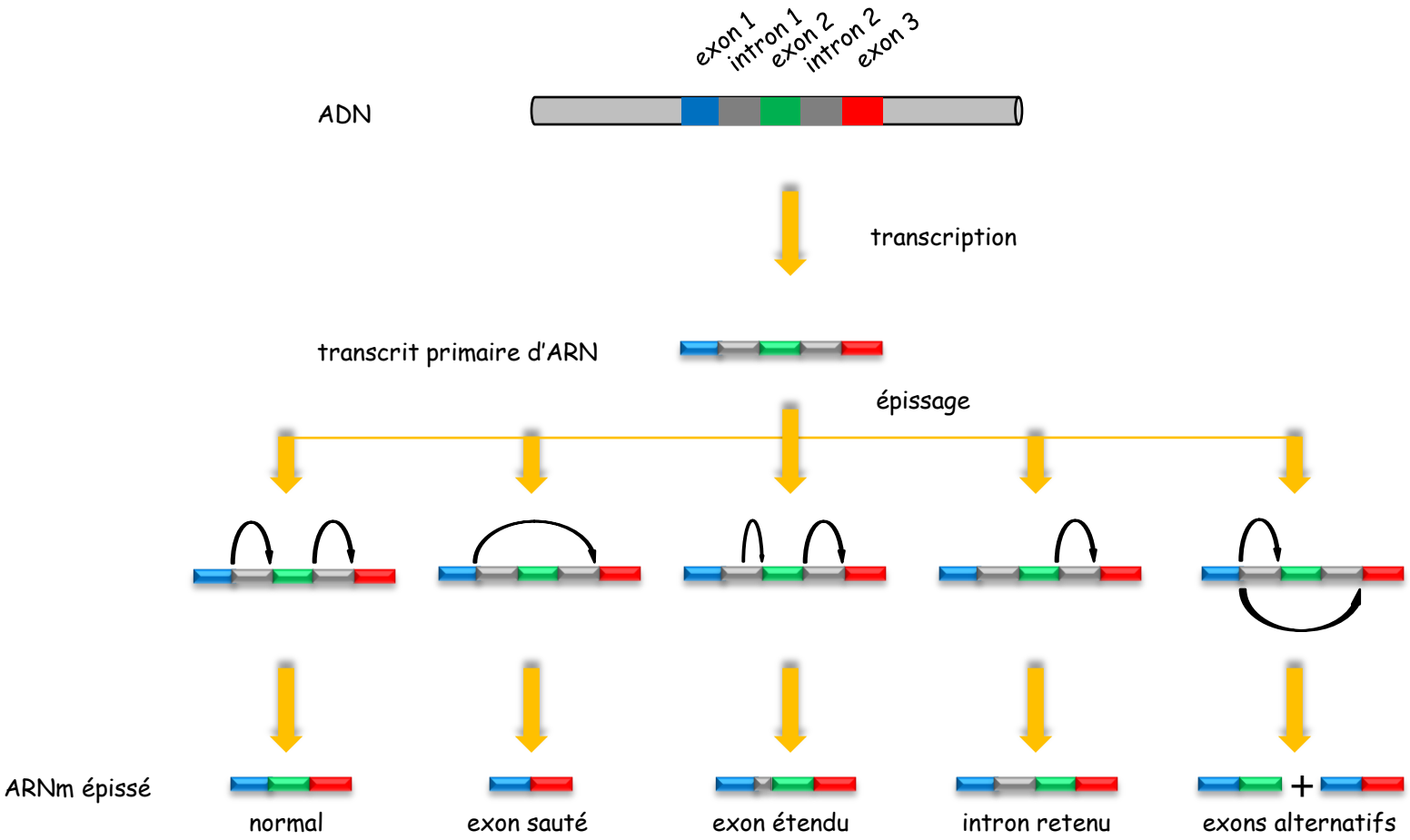
Maturation de l'ARNm

Lors de la transcription, tout le gène (introns + exons) est copié en ARN appelé transcrit primaire ou pré-messager.

L'ARN pré-messager doit ensuite être modifié pour, entre autres, en enlever les introns = épissage de l'ARN.



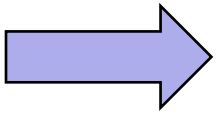
Cinq manières d'épisser un ARN



ADN non codant : les introns

L'épissage alternatif : les introns peuvent être épissés de différentes façons. Une partie de l'intron voir tout l'intron peut tantôt être non codant, tantôt être codant.

Combinatoire



A partir d'une séquence ADN on peut faire plusieurs séquences protéiques