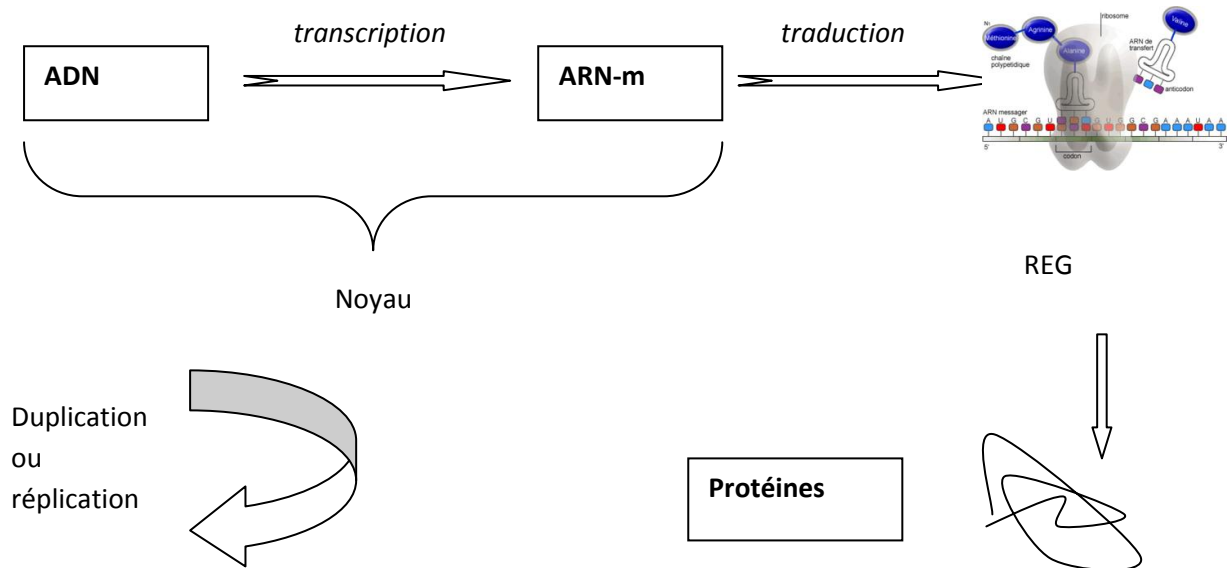


L'ADN : Réplication, synthèse des protéines

1) Principe général

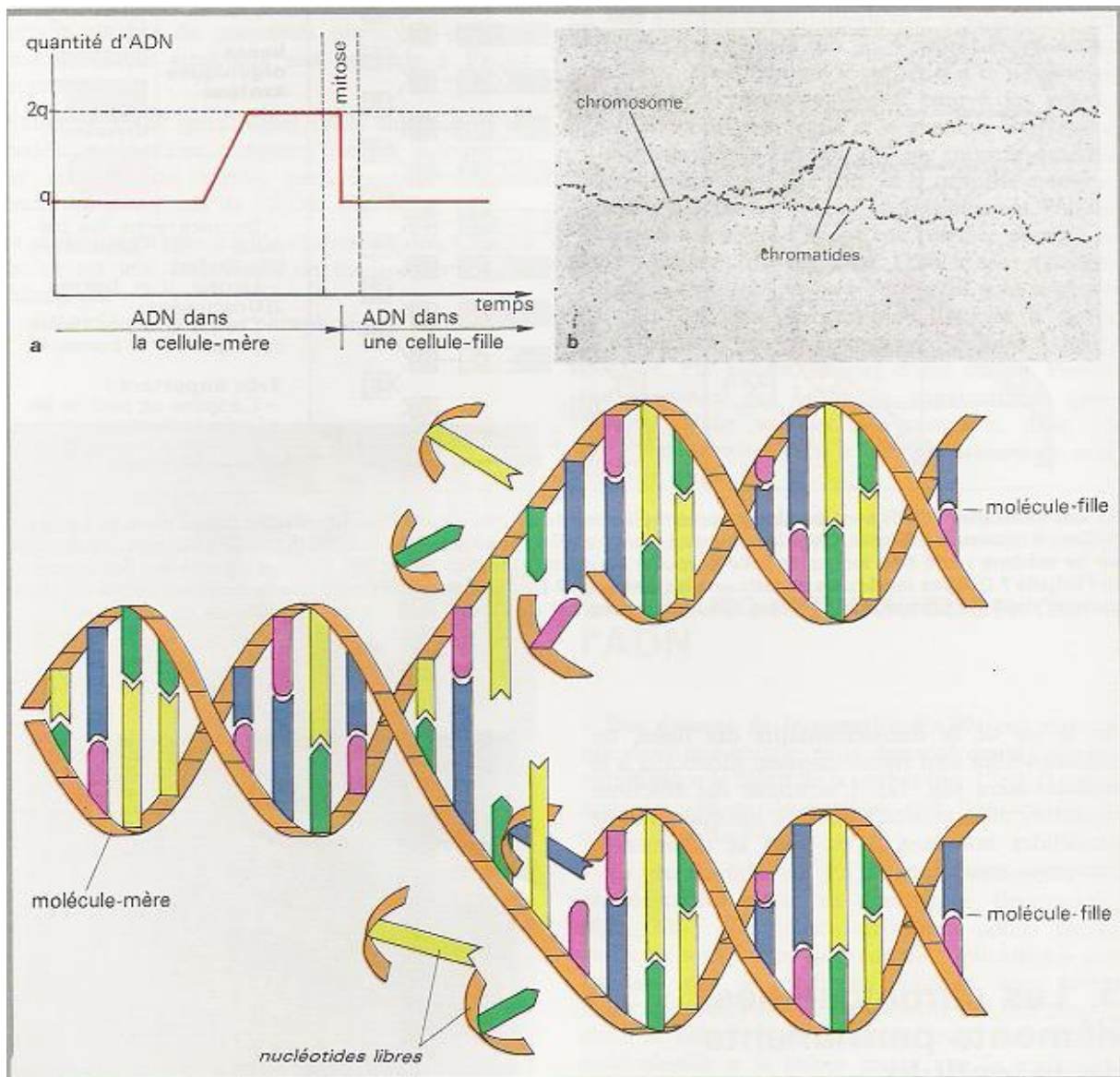
Le principe général est le suivant :



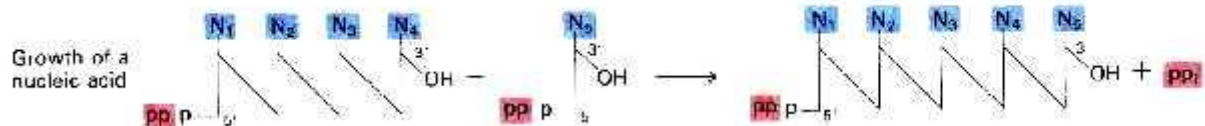
2) La réplication ou la duplication de l'ADN

Le premier postulat du dogme central, est que l'ADN est capable d'autoréplication. Ceci avait été immédiatement déduit par Watson et Crick du fait que chaque brin d'ADN spécifie univoquement son complément, mais il a fallu longtemps pour que les détails du mécanisme soient élucidés.

La première question fut celle de la conservation du mécanisme de réplication. Les expériences maintenant classiques de Meselson et Stahl prouvèrent que la réplication de l'ADN est **semi-conservative**, en ce que chacune des molécules filles hérite d'un brin de l'ADN parental. Il aurait été parfaitement possible d'imaginer un mécanisme conservatif, où les deux brins de la molécule fille auraient été synthétisés *de novo*, en utilisant les brins parentaux comme matrice. Il y a des avantages à une réplication semi-conservative, puisqu'elle permet de réparer des erreurs commises pendant la réplication, dans la mesure où le brin parental (et donc correct) peut être identifié.



Il fallut ensuite identifier et purifier des enzymes capables de polymériser des nucléotides sur une matrice d'ADN. Un grand nombre de **DNA polymérases** sont maintenant connues, qui toutes ont besoin d'une **amorce** à laquelle ajouter des nucléotides, et toujours à son extrémité 3'.

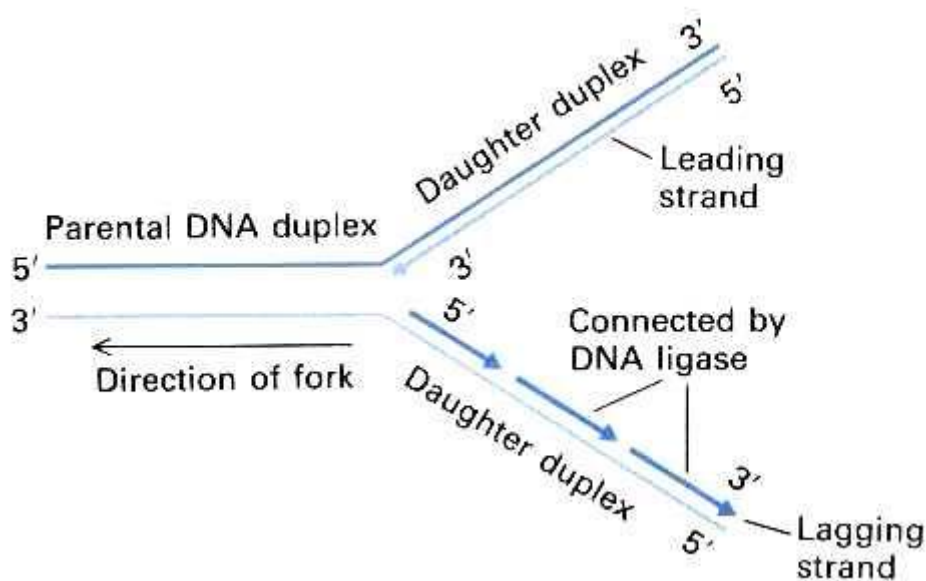


Elongation des acides nucléiques: polarité 5' -> 3'

La nécessité d'une amorce et la polarité 5' -> 3' obligatoire posent un sérieux problème: comment peut-on synthétiser de façon coordonnée deux brins à la polarité antiparallèle? La solution a été d'utiliser des mécanismes **différents** pour les deux brins:

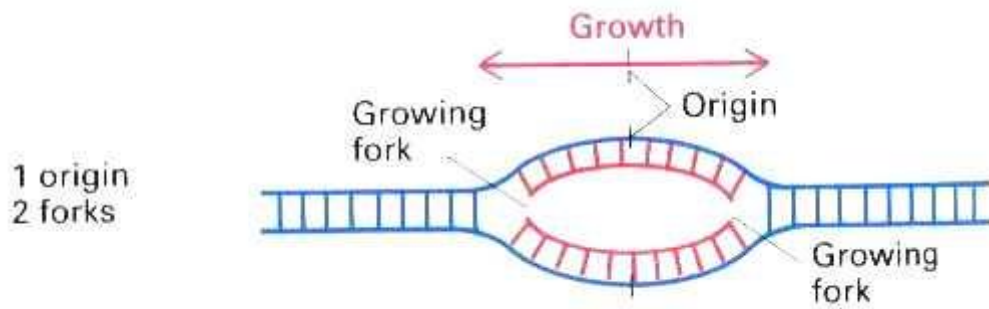
- 1) le brin primaire ("leading strand") où la polymérisation se fait en continu,
- 2) le brin secondaire ("lagging strand") où elle est discontinue.

De courtes amorces sont d'abord préparées par une RNA polymérase (qui peut démarrer sans amorce), puis allongées par une DNA polymérase, puis enlevées, les trous restants sont réparés, et les brins sont recollés par une "DNA ligase". Complicé, mais ça marche!

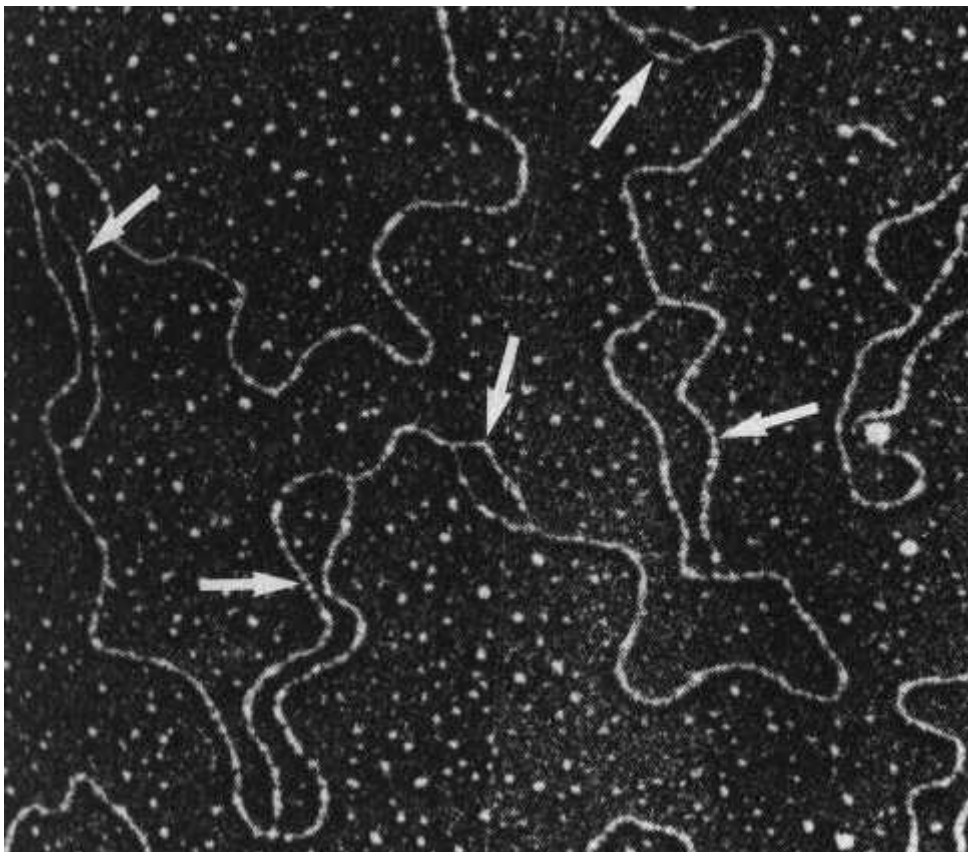


Anatomie d'une fourche de réplication de l'ADN

La plupart des molécules d'ADN sont soit très longues, soit circulaires. Il faut donc prévoir des points de démarrage ou **origines de réplication**. Un processus complexe sépare les deux brins d'ADN à ces origines, met en place les premières amorces, et démarre des fourches de réplication dans les deux sens. On obtient donc des **bulles de réplication** qui sont caractéristiques de l'ADN en phase répliative.

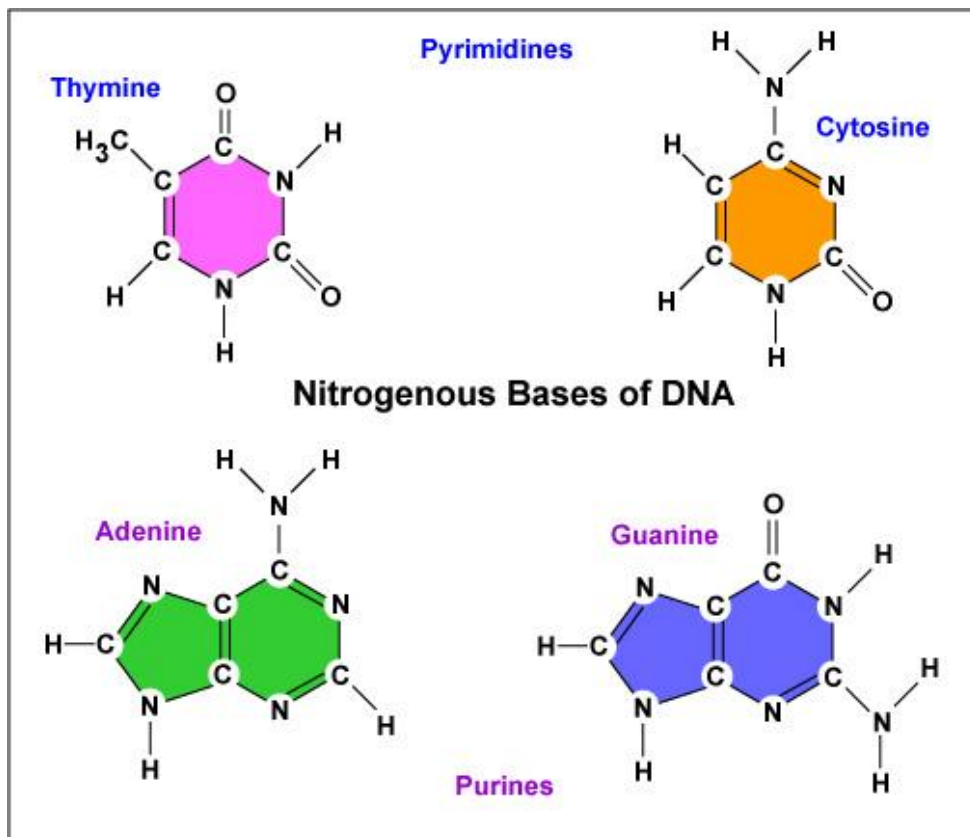


Une bulle de réplication s'étendant dans les deux directions

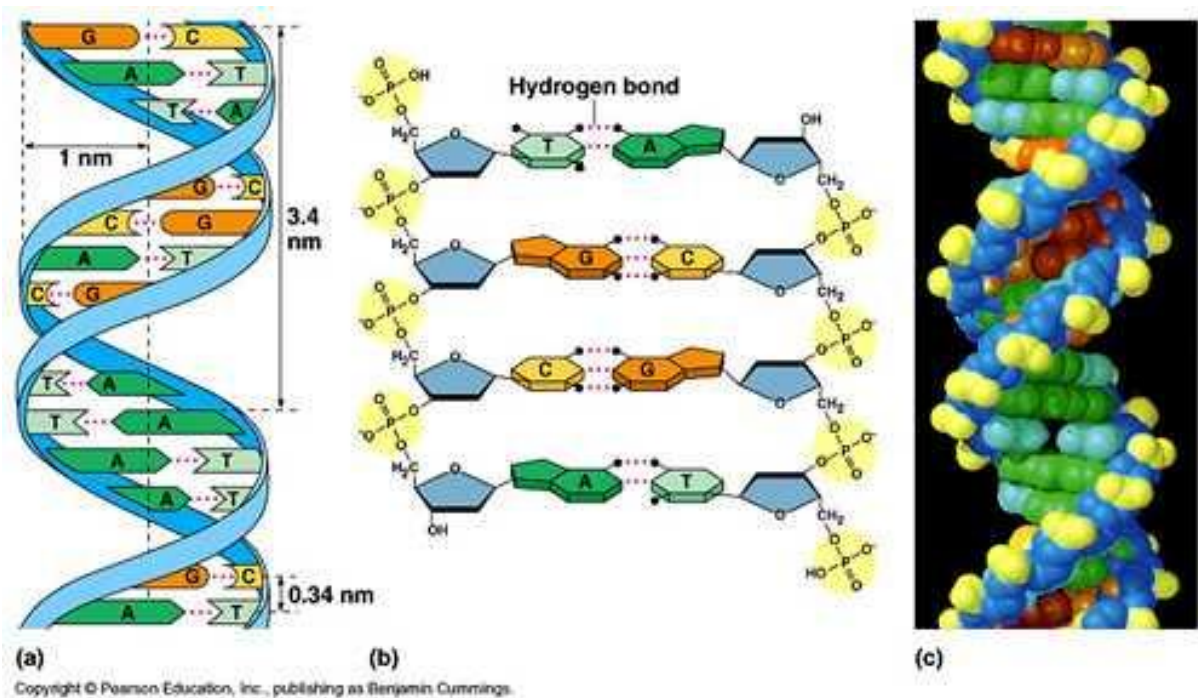


Photographie au microscope électronique d'ADN en réplication

Pour rappel , les bases présentes dans l'ADN sont :



L' Adénine étant toujours couplée à la Thymine et vice versa , la Cytosine (C) est couplée à la Guanine (G) , les liaisons entre bases sont des ponts hydrogène suivant le modèle ci-après

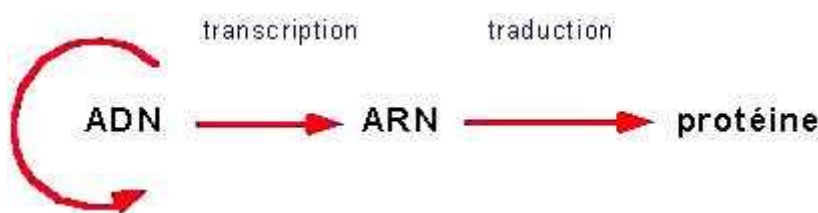


3) Transcription - types d'ARN

Le dogme central veut que l'ARN soit l'intermédiaire entre l'information stockée dans l'ADN et les agents d'exécution, de structure,... que sont les protéines.

Ceci est vrai seulement en partie, certains types d'ARN ont aussi, également, des rôles directement fonctionnels (c'est-à-dire qu'ils entrent dans la structure de certains organites).

Seul l'ARN messager (mRNA), qui ne représente que 3-5% de l'ARN total de la cellule, remplit strictement la définition donnée par le dogme central.



Néanmoins, il est vrai que tous les types d'ARN sont encodés dans le génome, et doivent être transcrits depuis l'ADN. Les principaux types sont les suivants:

- ARNr-r ou r-RNA = ARN du ribosome, 80% du total
- ARN-t ou t-RNA = ARN de transfert, 15% du total
- ARN- m ou mRNA = RNA messager, env. 3% du total

Il existe d'autres types d'ARN que nous citons ici bas)

hnRNA = précurseurs des messagers, trouvés dans le noyau

snRNA = RNA structuraux du noyau, impliqués dans la maturation du mRNA

scRNA = RNA structuraux du cytoplasme

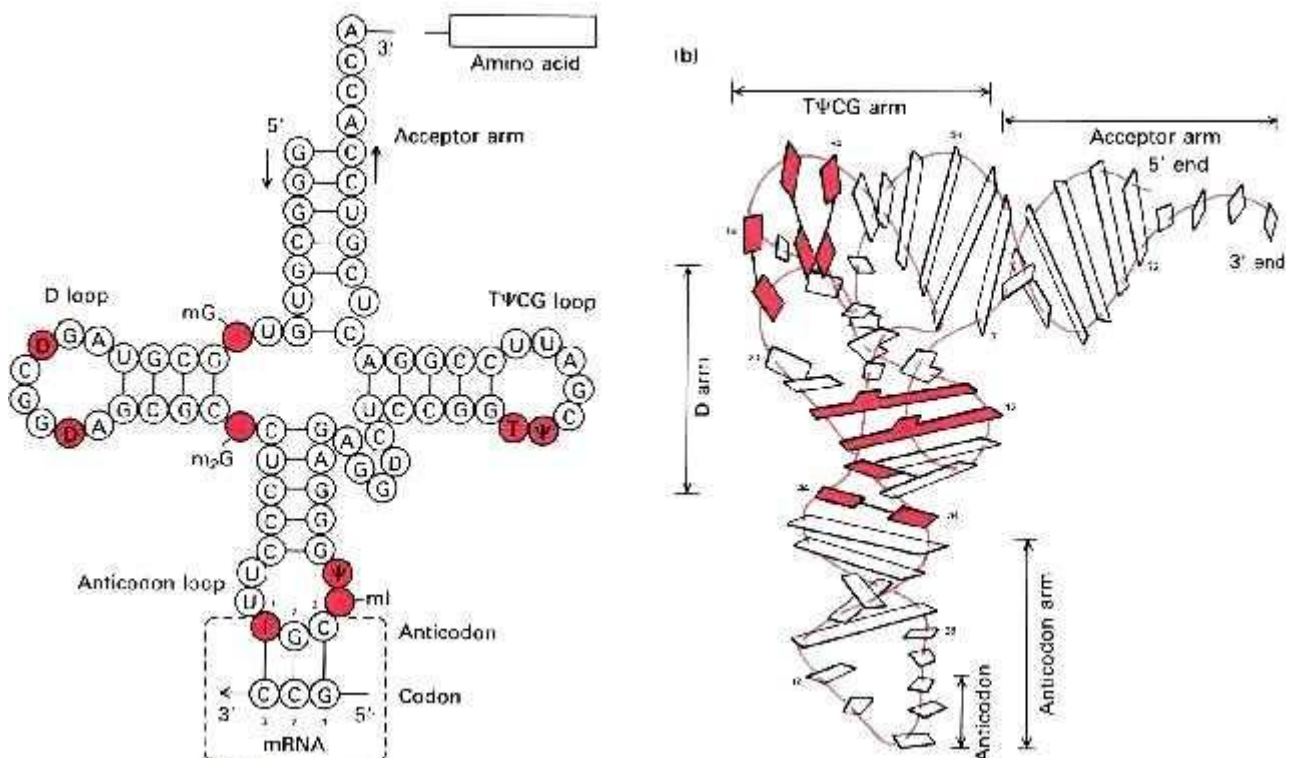
Alors que l'ADN a, dans la vaste majorité des cas, une structure en double hélice, les ARN ne contiennent normalement qu'un seul brin, et ne forment donc pas de structure stable avec leur complément.

Par contre, les bases de l'ARN peuvent s'apparier avec d'autres régions du même brin, et donc former des structures dites secondaires, entièrement dictées par la séquence du brin d'ARN. Ces appariements intramoléculaires jouent un rôle central dans la fonction de tous les ARN, en leur imposant une structure tridimensionnelle unique.

La structure des ARN de transfert, les adaptateurs qui permettent de décoder les ARN messagers, reflète parfaitement leur fonction.

La figure ci-après , représente un ARN-t : qui sert à transférer les AA , comme vous pouvez le voir , ce dernier possède une structure en « en croix », avec des connexions entre les bases , à côté de sa structure « plane » , vous trouverez sa structure « réelle » en 3 D .

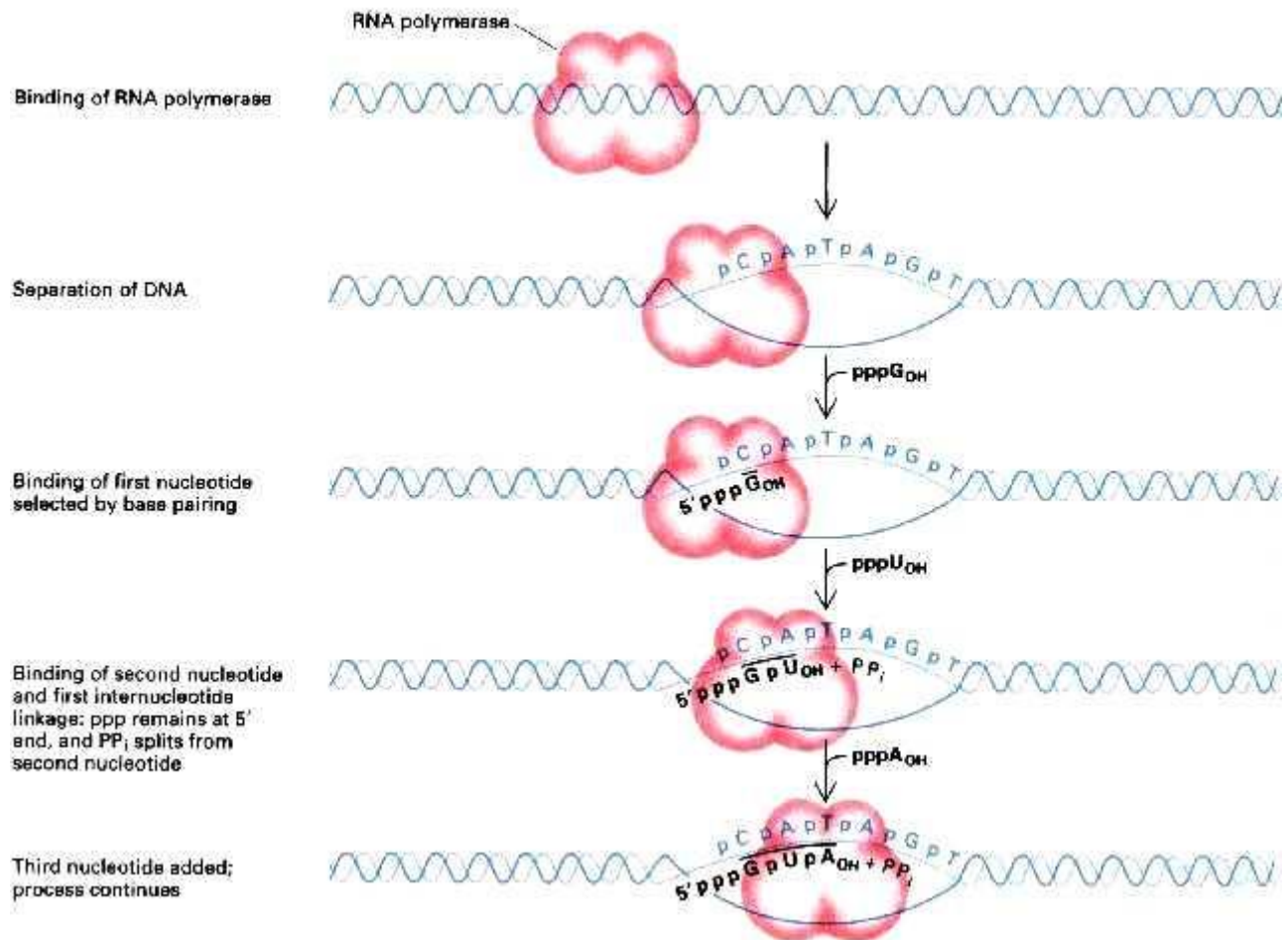
Cette structure lui permet de s'adapter à la « forme des cavités du ribosome dans lesquelles il va s'insérer



4) Transcription - mécanismes généraux

La transcription de l'ADN en ARN est accomplie par les **RNA polymérases**. Contrairement aux DNA polymérases responsables de la réplication et de la réparation de l'ADN, les RNA polymérases sont capables de synthèse *de novo*, sans avoir recours à des amorces.

Pour ce faire, elles doivent être capables de séparer les deux brins d'ADN, d'abord à leur site d'initiation, puis au fur et à mesure de leur progression le long de la matrice d'ADN.



Mécanisme d'initiation de la transcription

L'initiation de la transcription ne peut se faire qu'à des endroits précis sur la double hélice d'ADN, que l'on appelle les **promoteurs** (primers)

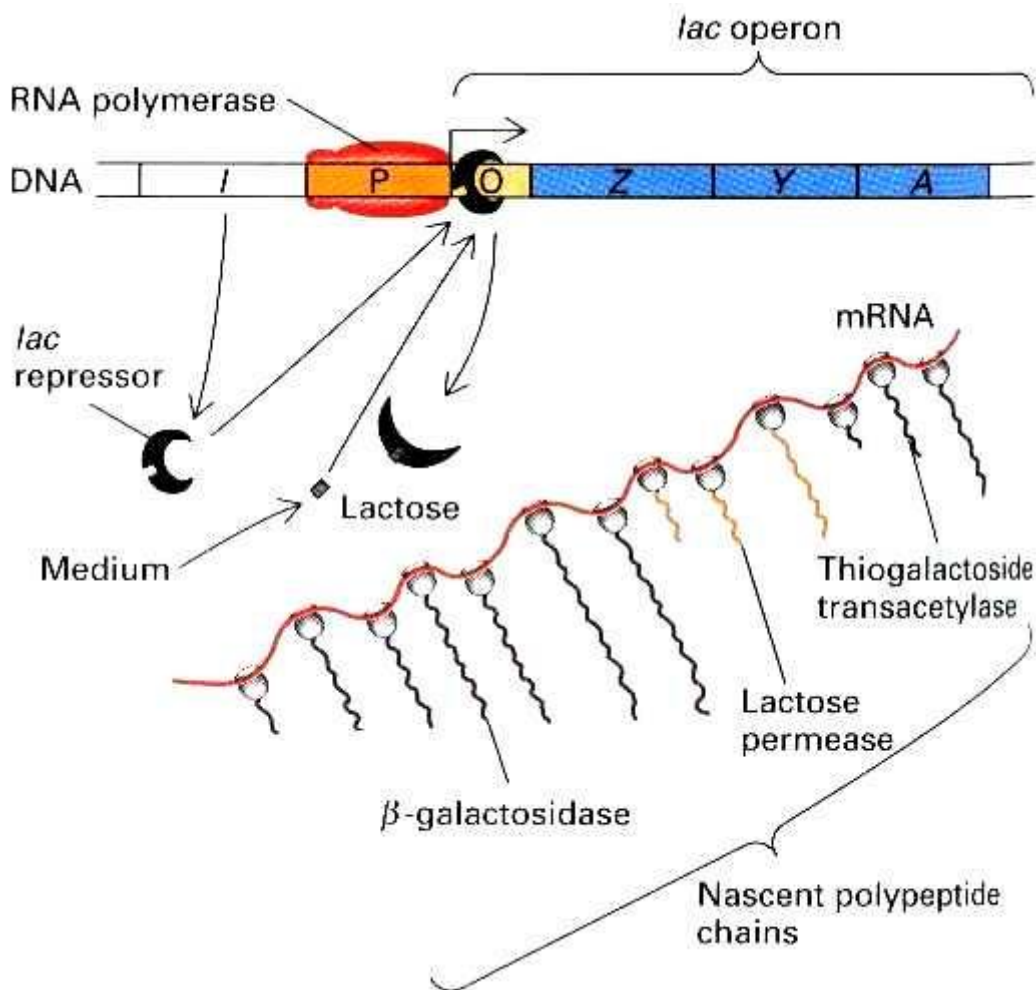
Chez la plupart des organismes (à l'exception de quelques virus infectant des bactéries), l'initiation de la transcription demande l'assemblage d'un complexe multiprotéique sur le promoteur, et ouvre donc la porte à des mécanismes régulateurs sophistiqués, qui font l'objet du chapitre suivant.

5) Transcription - mécanismes régulateurs

Considérant que le génome de chaque cellule porte le plan directeur complet pour former un nouvel organisme, il faut que l'expression des gènes soit strictement régulée pour permettre la différenciation et l'adaptation aux conditions externes de chaque cellule individuelle. Une partie importante de cette régulation se fait au niveau de la transcription des gènes. Des mécanismes à effet négatif et positif seront brièvement évoqués ici, mais il faut garder à l'esprit le fait que la plupart des gènes sont soumis à une régulation bien plus complexe.

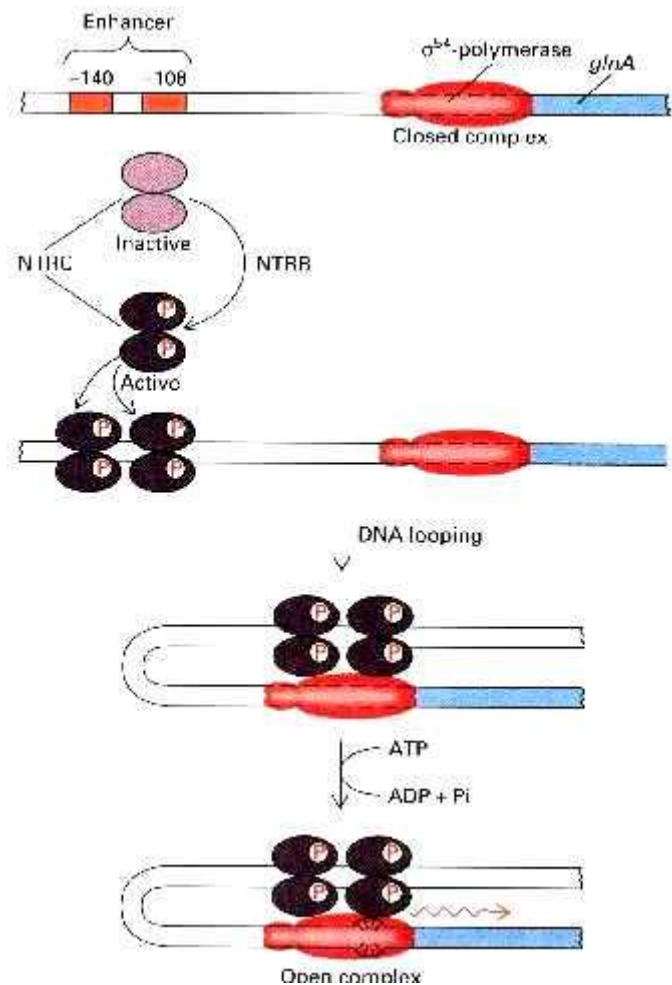
L'exemple classique d'une boucle régulatrice est celle dite de de l'opéron lactose *lac*, étudié dans les années 60 par Jacob, Monod et leur collègues.

En l'absence de lactose, un **répresseur** se lie à une séquence d'ADN appelée **l'opérateur**, et empêche ainsi la transcription de tout l'opéron, qui contient les gènes nécessaires à l'utilisation du lactose comme source d'énergie. Le lactose, quand il pénètre la cellule, se lie directement au répresseur, dont il empêche la liaison à l'opérateur. L'opéron peut donc être transcrit, et le mRNA qui en résulte être traduit en protéines.



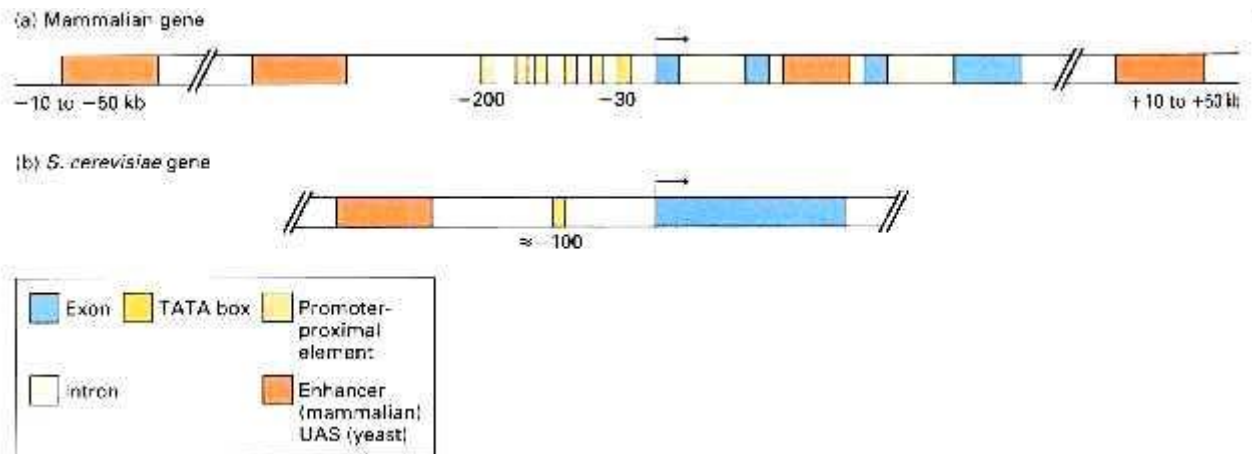
Mécanisme de régulation de l'opéron *lac*

Un mode de régulation très fréquent chez les eucaryotes implique des séquences à une certaine distance du promoteur, les "**enhancers**", qui sont reconnus par des protéines régulatrices (**facteurs de transcription**) dans certaines conditions mais pas dans d'autres. La reconnaissance peut dépendre de la migration du facteur vers le noyau, de son association avec une autre protéine, ou de modifications de sa structure (phosphorylation, glycosylation etc.) La liaison du facteur de transcription sur le "enhancer" provoque un changement de structure de l'ADN, et le recrutement d'autres facteurs menant finalement à l'assemblage d'un complexe d'initiation actif.



Activation de la transcription par action à distance

Les éléments régulateurs de la transcription peuvent se trouver à de multiples endroits par rapport au promoteur du gène. Chez les mammifères, on a trouvé des "enhancers" entre -50'000 et +50'000 nt par rapport au promoteur, souvent en synergie avec des éléments plus proches. Chez la levure, qui a un génome plus compact, ils sont normalement plus proches. Chez les bactéries, les éléments régulateurs se trouvent presque toujours à moins de 100 nt du promoteur.



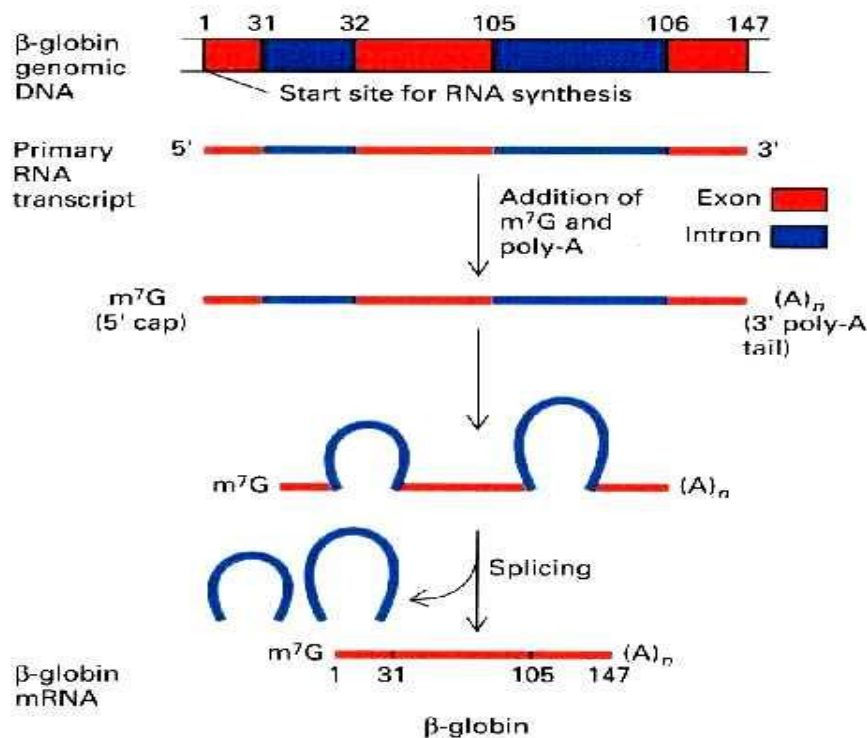
Eléments régulateurs chez les mammifères et les levures.
TATA box: élément caractéristique du promoteur
UAS: équivalent du "enhancer" chez la levure.

6) Modifications post-transcriptionnelles de l'ARN

L'ARN, copie fidèle du génome à sa synthèse, va encore subir une série de modifications avant de devenir une molécule fonctionnelle.

Chez les procaryotes (bactéries), ces modifications sont relativement mineures. Chez les eucaryotes par contre, pratiquement tous les types d'ARN sont modifiés après leur transcription. Nous nous concentrerons ici sur les ARN messagers. Les modifications principales des m-RNA sont:

- 1) L'ajout en 3' d'un polymère d'adénosine, ou **polyadénylation**
- 2) L'ajout en 5' d'un nucléotide modifié, le "**cap**"
- 3) La déletion des introns, ou **épissage**



Modifications post-transcriptionnelles de l'ARN messager

Le "capping" est une modification qui permet aux mécanismes de transport et de traduction de reconnaître le m RNA des autres formes d'ARN. Il est particulièrement important pour la liaison des ribosomes assurant le processus de traduction.

La polyadénylation, qui est aussi spécifique des ARN messagers, marque de façon précise l'extrémité 3' de ceux-ci, vu que la terminaison de la transcription est peu spécifique chez les eucaryotes. La place où la "queue" de poly(A) doit être ajoutée est marquée par une séquence spécifique (AAUAAA). Il semble aussi que la dégradation progressive du poly(A) dans le cytoplasme soit une horloge permettant de déterminer l'âge d'un m-RNA et de le détruire en temps voulu.

La modification la plus spectaculaire est l'épissage ou "splicing", qui élimine une grande partie de l'ARN initialement transcrit et reconstitue sa partie codante à partir de séquences discontinues. Des ARN nucléaires spéciaux, les snRNA, reconnaissent les extrémités 5' et 3' des **introns**, les parties à éliminer, et participent activement à une réaction de transestérification qui échange les liaisons phosphate de façon à laisser l'intron sous forme de "lasso" alors que les **exons** (constituant le produit final) sont reliés entre eux. Le nombre et la taille des introns augmentent avec la complexité de l'espèce et sa position dans l'arbre de l'évolution. Le "saut" des gènes en exons séparés semble être un élément important dans l'évolution, dans la mesure où il permet la formation de nouveaux gènes à partir d'exons issus de gènes différents.

7) Le code génétique

L'ARN messenger transcrit de l'ADN a une fonction principale, **celle de spécifier une protéine à synthétiser**.

Une grande partie des autres ARN cellulaires participent aussi à cette tâche, mais en tant que « machines à décoder » (bref les boîtes à musique) plutôt que de porteurs du code.

Le code à déchiffrer comporte quatre lettres, correspondant aux nucléotides (A, G, C, ou U), alors que sa traduction en compte 20 soit le nombre des acides aminés qui composent toutes les protéines.

Le nombre minimum de nucléotides nécessaires pour spécifier un acide aminé est donc de trois, vu que deux lettres tirées d'un alphabet de quatre ne peuvent coder que 16 combinaisons différentes.

Avec trois nucléotides, le nombre est de 64, ce qui permet d'incorporer une **punctuation**, nécessaire à la bonne lecture du code, et implique une **redondance** (plusieurs codons peuvent « coder pour le même acide aminé »). Ces deux prédictions ont été vérifiées expérimentalement, et l'un des grands succès de la biologie moléculaire des années 60 fut de déchiffrer complètement le code génétique.

TABLE 4-3 The Genetic Code (RNA to Amino Acids)*					
First Position (5' end)	Second Position				Third Position (3' end)
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop (och)	Stop	A
	Leu	Ser	Stop (umb)	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met (start)	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val (Met)	Ala	Glu	Gly	G

* Stop (och) stands for the ochre termination triplet, and Stop (umb) for the amber, named after the bacterial strains in which they were identified. AUG is the most common initiator codon; GUG usually codes for valine, but, rarely, it can also code for methionine to initiate an mRNA chain.

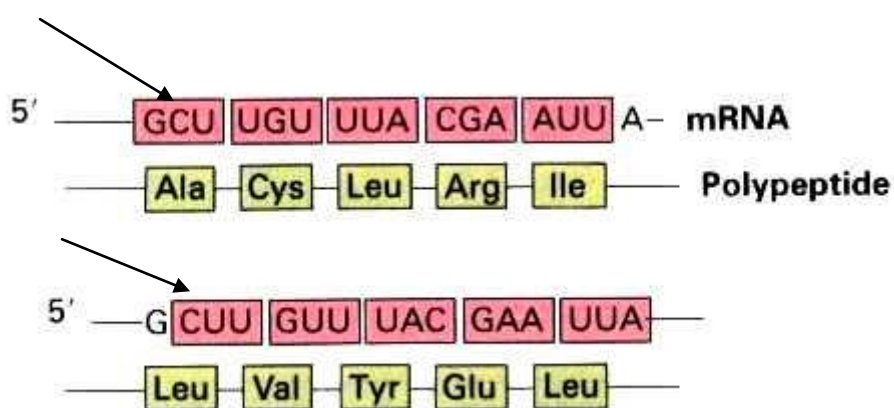
TABLE 4-4 The Degeneracy of the Genetic Code		
Number of Synonymous Codons	Amino Acid	Total Number of Codons
6	Leu, Ser, Arg	18
4	Gly, Pro, Ala, Val, Thr	20
3	Ile	3
2	Phe, Tyr, Cys, His, Gln, Glu, Asn, Asp, Lys	18
1	Met, Trp	2
Total number of codons for amino acids		61
Number of codons for termination		3
Total number of codons in genetic code		64

Le code génétique et ses éléments de redondance

Pour que le code soit déchiffré correctement, il faut aussi spécifier un **cadre de lecture** modulo 3 (ou un codon « start »).

Ceci est accompli par l'incorporation d'une méthionine (code AUG) en première position, dans un environnement favorable à l'attachement d'un ribosome. La première apparition du code AUG dans un m-RNA spécifie normalement le cadre de lecture, et il est souvent précédé de plusieurs codons "stop" (UAA, UAG, UGA) pour éviter toute ambiguïté.

Des mutations qui introduisent ou enlèvent 1 ou 2 nucléotides de la partie codante d'un gène vont décaler le cadre de lecture, et donc provoquer la synthèse de protéines non fonctionnelles. Ceci est une cause fréquente de mutations naturelles.



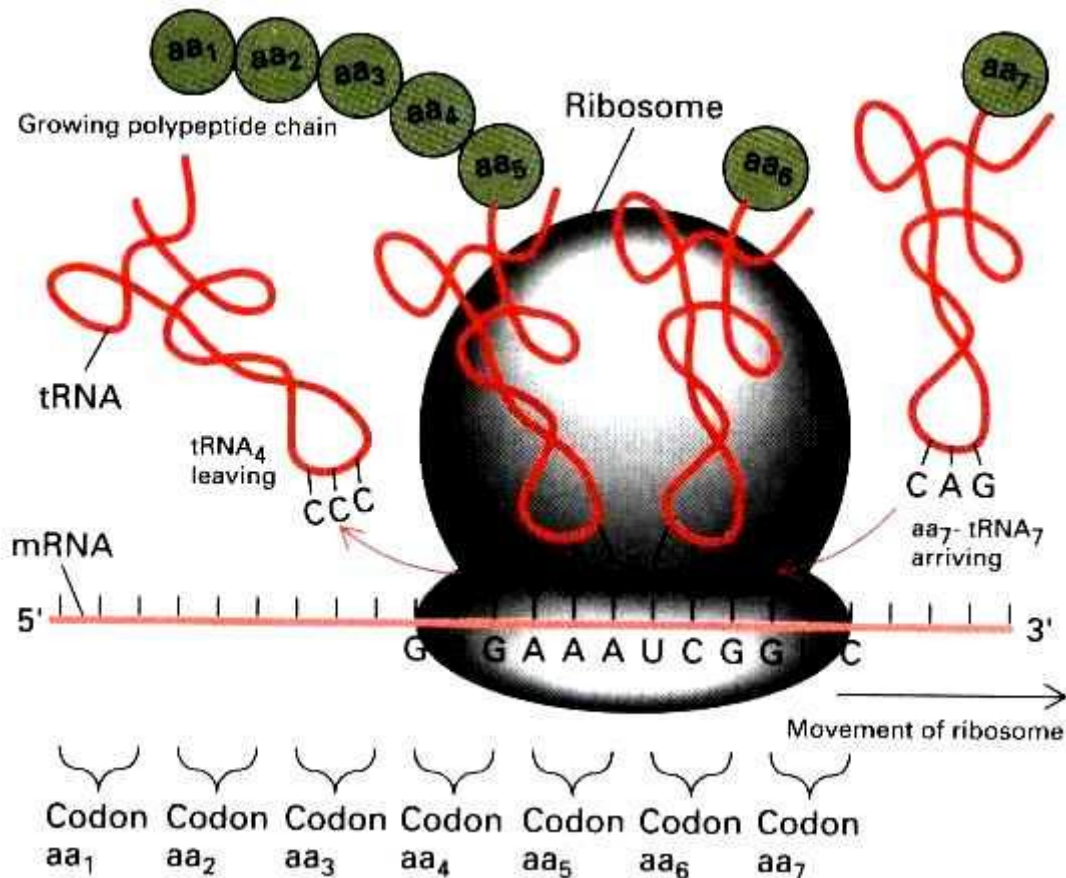
Deux cadres de lecture pour le même m- RNA donnant deux polypeptides différents

8) Traduction - mécanismes généraux

La traduction du code génétique en protéine utilise les trois types principaux d'ARN

- 1) l'ARN messager comme porteur du code,
- 2) les ARN ribosomiques comme machines à fabriquer les protéines,
- 3) les ARN de transfert comme clefs du code et qui servent au transport des AA vers le ribosome

Les trois ARN utilisent et sont les parties fonctionnelles de la machine à décoder : le **ribosome**.

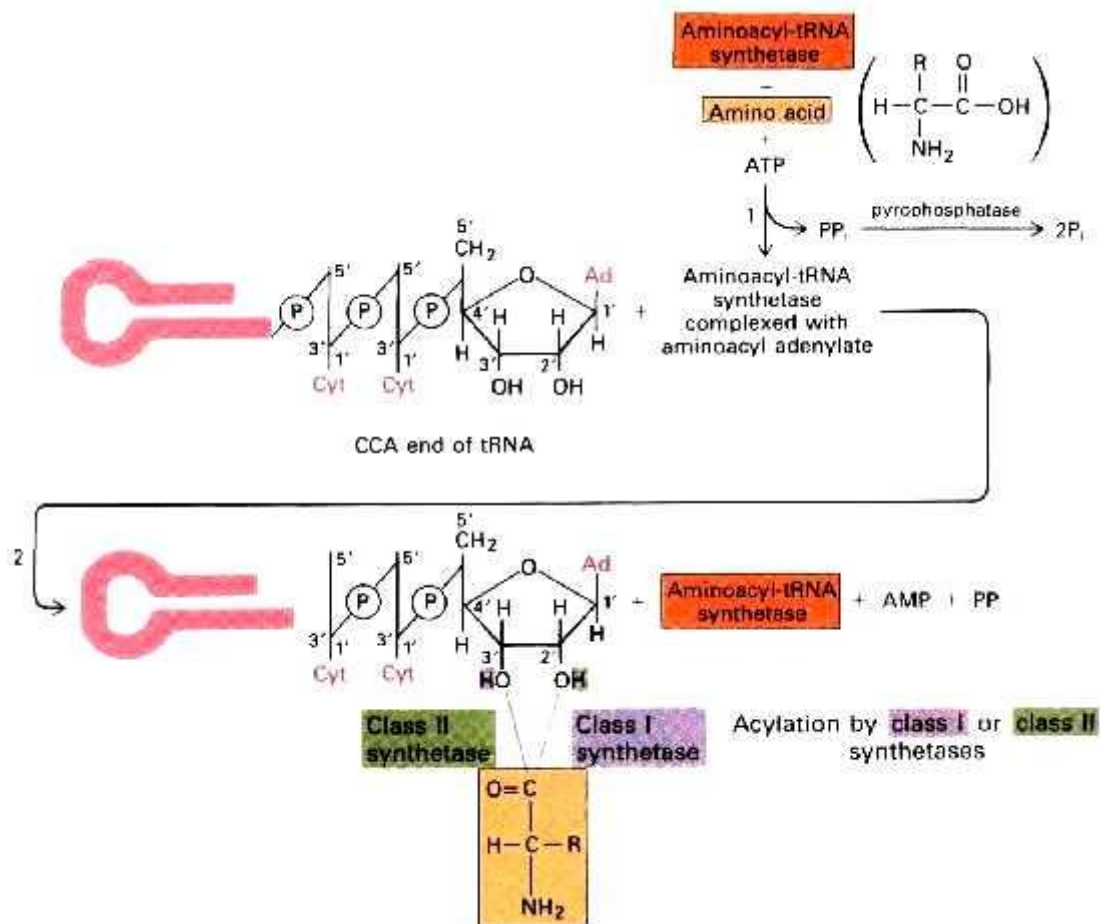


Fonction des mRNA, tRNA et rRNA dans la traduction du code

Le mécanisme de décodage, sur lequel nous reviendrons en plus de détail dans le chapitre suivant, implique un appariement complémentaire entre **les codons** se trouvant sur le m-RNA, et les **anticodons** se trouvant à une extrémité du t-RNA. Cet appariement de trois paires de

bases est guidé par le ribosome. A l'autre extrémité du tRNA se trouve un **acide aminé** attaché de façon covalente à son adénosine 5'-terminale. Une réaction de transestérification transfère l'acide aminé amené par le dernier tRNA à celui amené par l'avant-dernier, allongeant ainsi la chaîne peptidique d'une unité.

La traduction correcte du code dépend donc directement de l'attachement spécifique d'acides aminés précis à leurs tRNA respectifs, qui doivent être reconnus par des enzymes spécialisées, les **aminoacyl tRNA synthétases**. Ces enzymes sont toujours spécifiques d'un acide aminé donné, et reconnaissent un certain nombre d'éléments, dont l'anticodon, dans les tRNA. Il suffit d'une mutation dans un tRNA qui altère cette reconnaissance, ou l'anticodon, pour que le code ne soit plus interprété correctement. Ce type de mutation est normalement létal, mais existe néanmoins: il y a des mutants bactériens chez lesquels l'un des codons "stop" n'est pas reconnu correctement. Il y a aussi quelques embranchements dans l'arbre de l'évolution où l'interprétation d'un codon a changé, présumément à partir d'une mutation non létale.

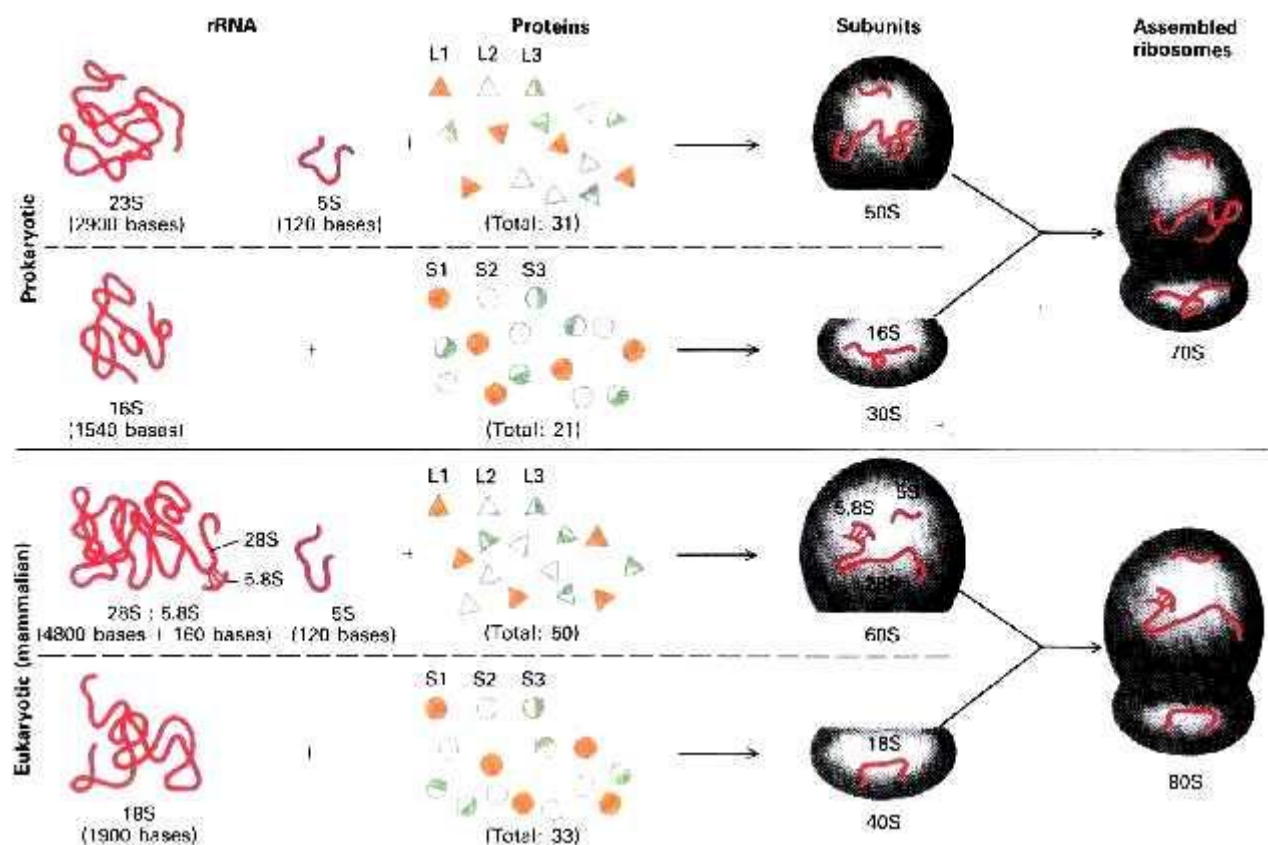


Mécanisme d'action des aminoacyl-tRNA synthétases

Traduction - structure du ribosome

Le ribosome, la machine à décoder, est une structure extrêmement complexe. En fait, nous ne connaissons toujours pas son architecture précise, seulement sa forme générale et l'identité de ses composants. Plusieurs ARN différents, connus sous des noms correspondant à leurs constantes de sédimentation, en forment l'échafaudage, et jouent aussi un rôle essentiel dans les propriétés associatives et catalytiques du ribosome.

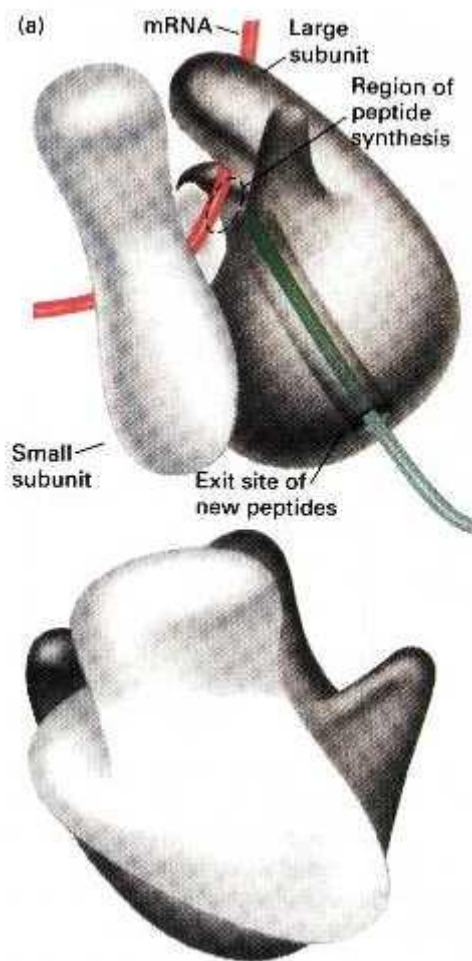
Le gros de la masse des ribosomes est formée de protéines, qui avec les ARN correspondants forment deux sous-unités, séparées au moment de l'initiation de la traduction, mais toutes deux nécessaires à sa progression.



Composants des ribosomes: ARN, protéines, sous-unités

Une vue plus exacte de la forme tridimensionnelle des ribosomes a pu être déduite par analyse d'images en microscopie électronique et par diffraction de neutrons.

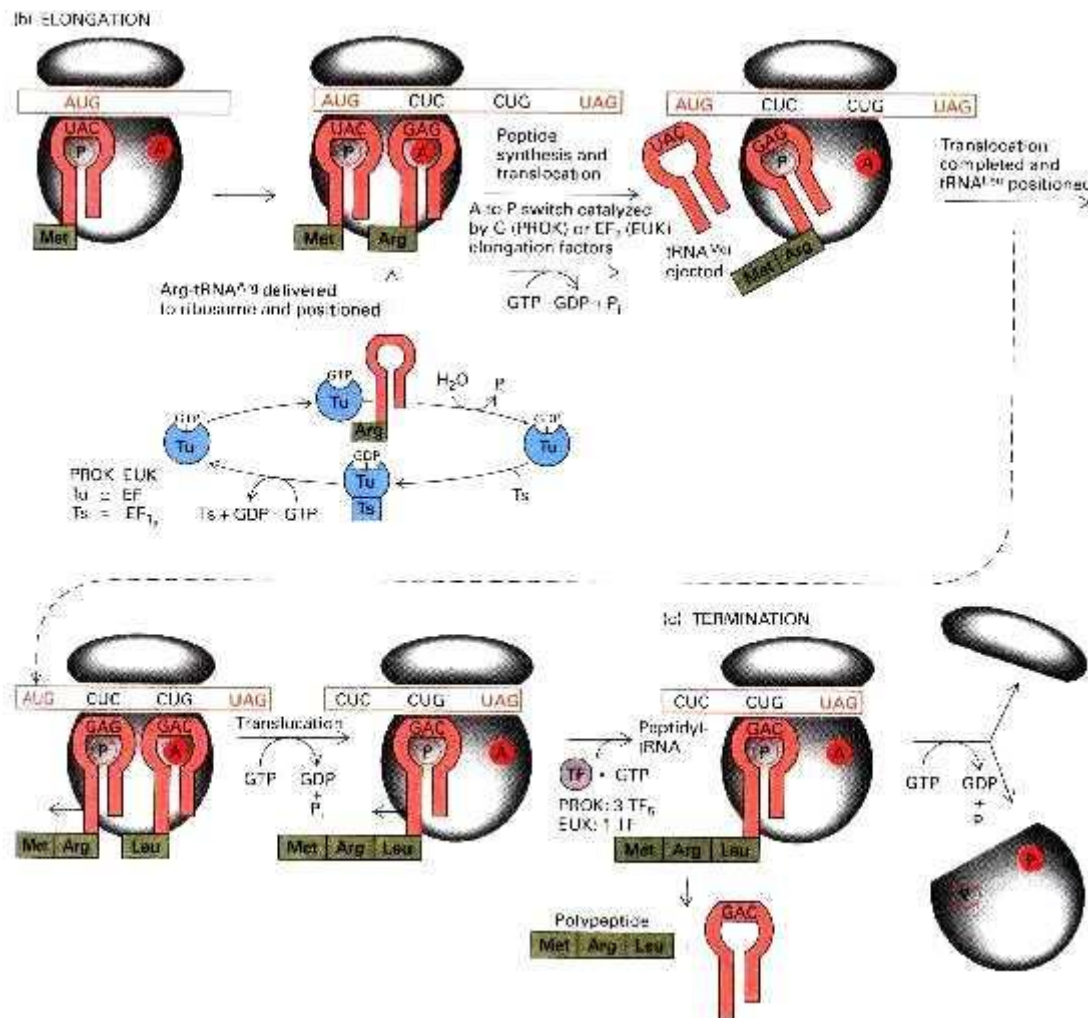
Elle montre plusieurs protubérances associées avec la région où coulisser le mRNA pendant la traduction, et un "tunnel" à travers lequel passe la chaîne peptidique nouvellement synthétisée. La structure et la position des tRNA pendant le processus de traduction ne sont pas encore connus. On sait pourtant qu'il existe deux sites de liaison du tRNA, occupés l'un après l'autre pendant la traduction.



Structure tridimensionnelle du ribosome à basse résolution

Traduction - quelques détails

Nous ne considérerons ici que le mécanisme d'élongation, qui opère après la mise en place du tRNA d'initiation, portant une méthionine. La phase d'initiation requiert une série de protéines ou **facteurs d'initiation**, pour associer les deux sous-unités ribosomiques avec le mRNA et le tRNA-Met, et mettre celui-ci dans le site de liaison P, prêt à recevoir le tRNA suivant (tRNA-Arg dans l'illustration). Chaque nouveau tRNA (avec son acide aminé attaché) doit d'abord être préparé par association avec un **facteur d'élongation**, EF1. Une fois le tRNA en place dans le site A, le facteur se dissocie de façon irréversible, dans une réaction hydrolysant du GTP, et donc consommant de l'énergie. Une deuxième réaction requérant l'hydrolyse de GTP et la présence d'un autre facteur d'élongation (EF2) effectue la liaison peptidique proprement dite et la translocation du tRNA du site A au site P. Ce cycle se répète jusqu'à ce que le ribosome rencontre un codon stop, empêchant l'occupation du site A. Un **facteur de terminaison** permet alors la dissociation du dernier tRNA, du mRNA, et des deux sous-unités ribosomiques.



Cycle d'élongation et terminaison de la traduction

Le dogme central - thème et variations

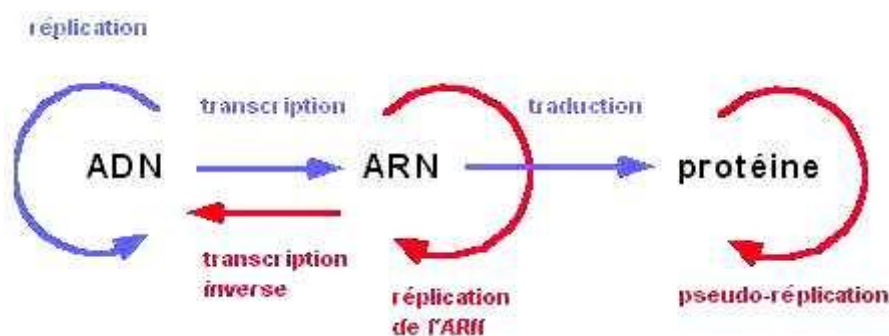
Le [premier chapitre](#) de ce cours a présenté le dogme central dans sa version la plus dogmatique (ADN → ARN → protéine), qui correspond à sa formulation originale par Francis Crick. Depuis lors, un certain nombre de phénomènes qui contredisent ce formalisme ont été décrits. Chacun ou presque a été salué comme une nouveauté sensationnelle, remettant en cause notre vision du flux d'information dans les êtres vivants. Avec un peu de recul, les contradictions au dogme central apparaissent comme des épiphénomènes qui ne sauraient exister si le dogme lui-même n'était pas respecté dans la vaste majorité des manifestations de la vie sur terre.

Les principales transgressions au code ont été:

1. **La transcription inverse.** Elle a été découverte comme faisant partie du cycle réplcatif des **rétrovirus**, dont le plus connu est le virus du SIDA, le HIV. Le génome des rétrovirus est constitué d'ARN, qui est transcrit en ADN après infection et va s'intégrer dans le génome de

l'hôte, d'où il est à nouveau transcrit pour former de l'ARN et des protéines virales. Il existe dans le génome de la plupart des eucaryotes des génomes rétroviraux, les **rétrotransposons**, défectifs en tant que virus mais encore capables de se multiplier dans le génome par transcription, transcription inverse, et réintégration. Les rétrotransposons sont devenus des facteurs significatifs dans l'évolution des espèces. Finalement, la transcription inverse réintroduit parfois des mRNA dans le génome.

2. **La réplication de l'ARN.** Ceci est une caractéristique de la plupart des virus à ARN, qui produisent leurs propres enzymes pour répliquer leurs génomes, dans des structures pratiquement indépendantes de la machinerie normale de l'hôte. Les virus étant en gros des petits groupes de gènes parasites ayant trouvé des stratégies diverses pour se propager, la réplication de leur ARN n'est pas une entorse majeure au dogme.
3. **La pseudo-réplication des prions.** Les prions sont des agents infectieux avec des caractéristiques similaires aux virus, mais qui ne contiennent pas d'acides nucléiques. Le dogme voudrait qu'ils ne puissent pas se répliquer, et ne puissent donc pas être infectieux. Il semble maintenant que les prions sont en effet des protéines porteuses d'information, mais que cette information n'est que structurelle: les prions, dérivés de protéines du cerveau, peuvent induire chez des protéines normales un changement de conformation qui se propage graduellement, et peut être transmis d'individu à individu. Les maladies à prions sont toutes des maladies dégénératives du système nerveux central (en particulier la maladie des "vaches folles"). La différence majeure par rapport aux acides nucléiques est que l'information contenue dans les prions ne peut se propager qu'à des protéines pré-existantes, et n'est donc qu'inductive, non créatrice.



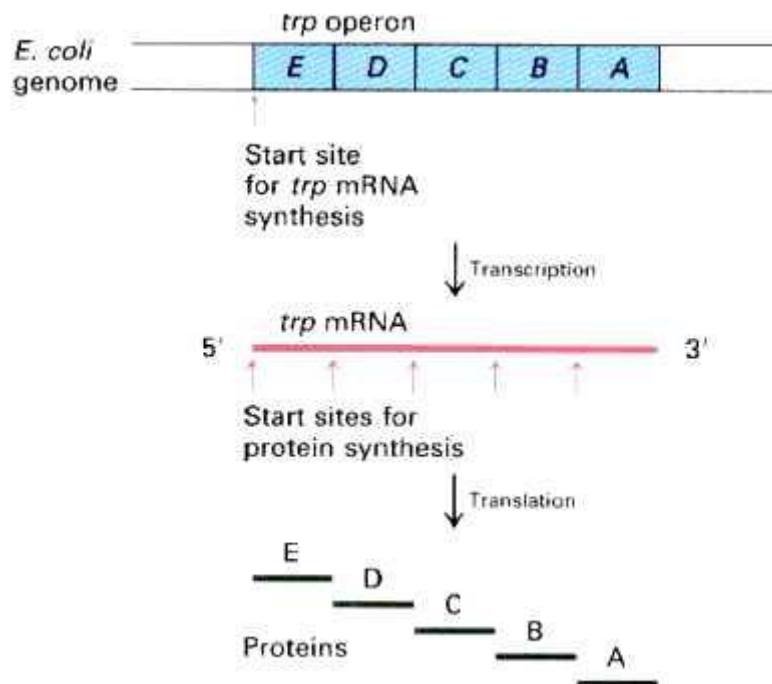
Le dogme central (bleu) et certaines de ses exceptions (rouge)

Il existe aussi des variations assez importantes entre procaryotes et eucaryotes dans la façon dont le code est stocké et interprété, même si sa substantifique moëlle, la traduction des codons, reste la même. Chez les procaryotes, une grande partie des gènes sont organisés en **opérons**, où un seul mRNA code pour une série de protéines avec des fonctions interdépendantes (p.ex. le métabolisme d'un composant nutritif). C'est une économie de taille génomique, de complexité, et d'énergie. D'autre part, les structures trop efficaces ont de la peine à évoluer.

Chez les eucaryotes multicellulaires, un mRNA ne code jamais pour plus d'une protéine, et on peut même argumenter que l'unité d'évolution est l'exon, qui représente souvent un domaine fonctionnel de la protéine. De plus, le phénomène de transcription inverse peut réintroduire dans le génome des copies sans introns des gènes, qui sont parfois transcrites mais ne sont pas fonctionnelles à cause de la disparition des introns. Le génome accumule aussi des

pseudogènes, non illustrés ici, qui résultent de duplication génique (l'un des moteurs de l'évolution) suivie de mutation laissant le gène non fonctionnel.

(a) Prokaryotes



(b) Eukaryotes

