

Chapitre 10 : Structure et fonction du gène : génétique biochimique

1- Notion de gène

La molécule d'ADN est organisée en unités appelées gènes. Un gène est fait d'une succession de nucléotides. C'est un facteur transmissible qui détermine un caractère.

Sur le plan fonctionnel, un gène est une séquence d'ADN avec une structure nécessaire à la synthèse d'un produit fonctionnel qui peut être sous la forme d'ARN ou de polypeptide.

2- Notions de locus et d'allèles

Chaque gène occupe un emplacement particulier le long du chromosome. Cet emplacement est appelé locus (loci au pluriel).

Les allèles sont les différentes formes que peut prendre un même gène à un locus donné.

Exemple : pour le caractère « forme des grains chez le petit pois », lisse et ridé sont les deux allèles possibles du gène responsable de ce caractère.

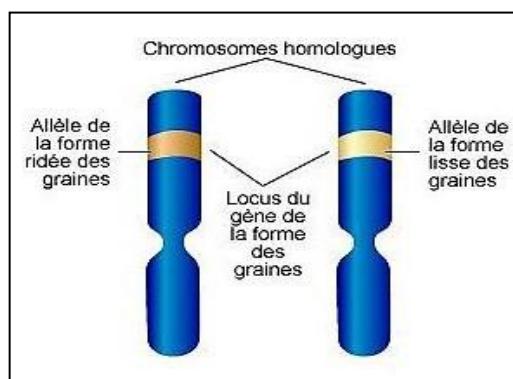


Figure 1 : Emplacement des allèles sur les chromosomes homologues

3- L'organisation du gène

La plupart des gènes eucaryotes contiennent une alternance de régions codantes appelées **exons** et de régions non codantes appelées **introns**. Chez les procaryotes, toutes les régions d'un gène sont codantes.

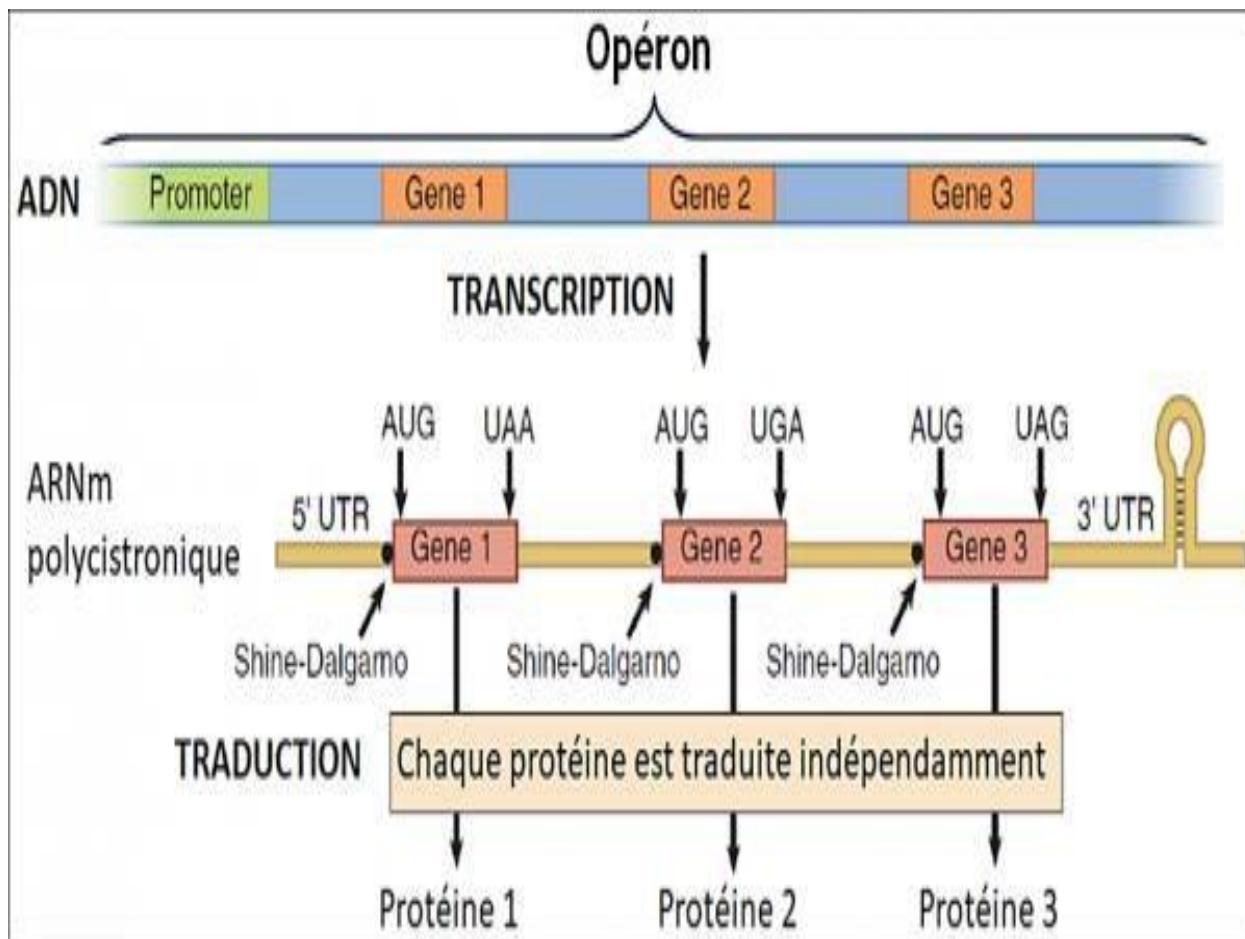


Figure 2: information génétique continue chez les procaryotes

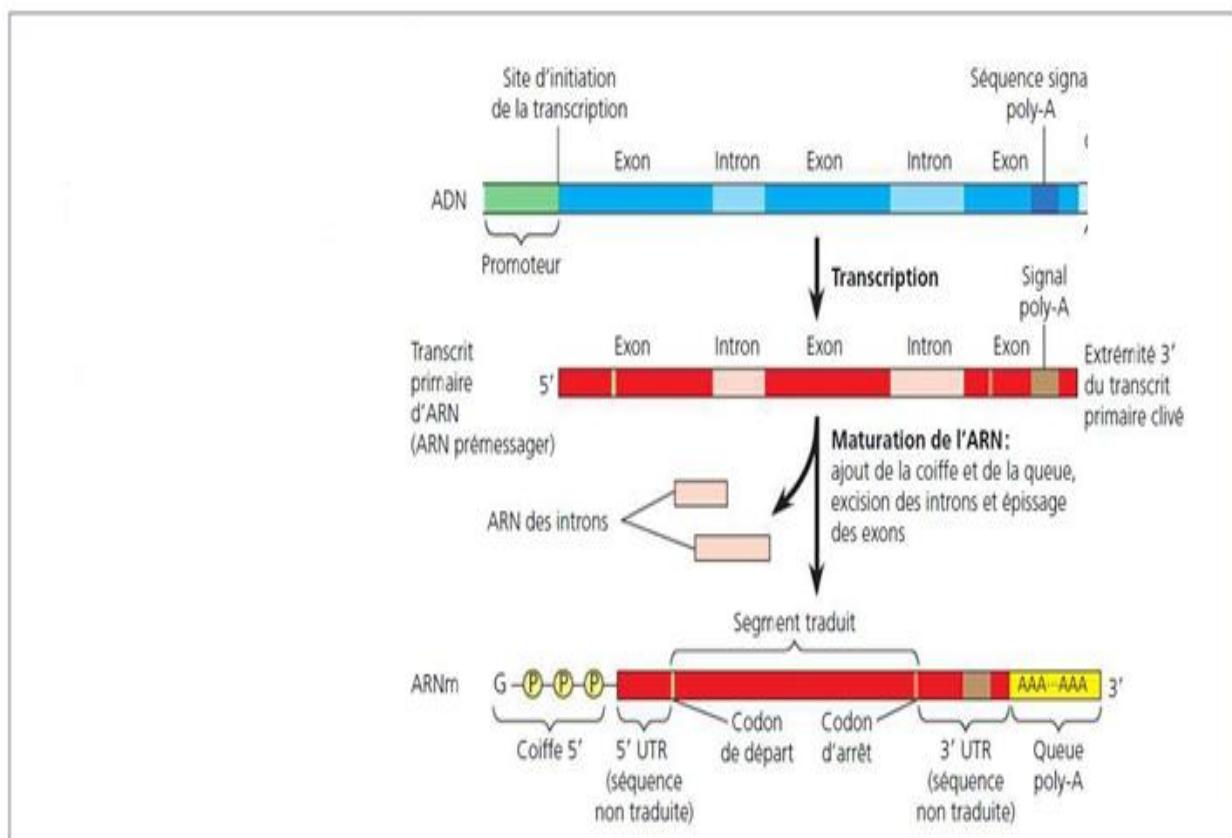


Figure 3 : exons et introns chez les eucaryotes

4-Structure des gènes eucaryotiques :

Un gène codant pour une protéine commence au site de départ de la transcription et termine par des séquences qui seront transcrtes mais pas traduites : sont appelées **UTR-5'** et **UTR-3'** (Untranslated Regions ; Régions non-traduites en 5' et 3' respectivement).

A quelques dizaines ou quelques centaines de paires de bases en aval, le codon d'initiation ATG est le site de départ de la traduction. Puis vient une succession de séquences tantôt codantes, les exons qui seront transcrits et traduits, tantôt non codantes, les introns (qui séparent deux exons) qui seront transcrits mais pas traduits.

Les exons et les introns sont numérotés dans la direction 5 'à 3' du brin codant. Les exons et les introns sont transcrits en un ARN précurseur (transcription primaire). Les segments non codants (introns) sont retirés du transcript primaire et les exons de chaque côté sont connectés par un processus appelé épissage. L'épissage doit être très précis pour éviter une modification indésirable du cadre de lecture correct. Les introns commencent presque toujours par les nucléotides GT dans le brin 5 'à 3' (GU dans l'ARN) et se terminent par AG .

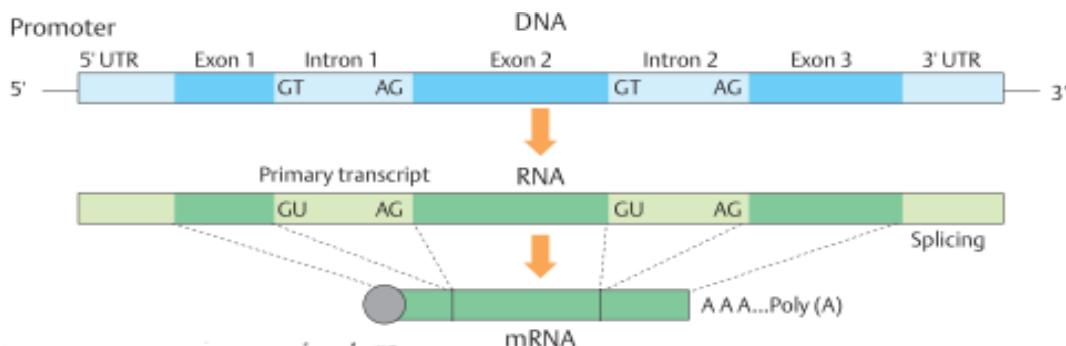


Figure 4 : Structure générale d'un gène eucaryotique

Les séquences à l'extrémité 5 ' de l'intron commençant par GT sont appelées site donneur d'épissage et à l'extrémité 3', se terminant par AG, sont appelées site accepteur d'épissage. A la fin du dernier exon, un codon stop (TAA, TAG ou TGA) est le site de fin de la traduction. Enfin 10 a 20 bases avant la fin du dernier exon est retrouée la séquence AATAAA improprement

appelée sequence de polyadenylation, sequence de reconnaissance pour la coupure du transcript primaire ,site de fixation de la poly A polymerase.

En amont, coté 5' du gène, se trouve **le promoteur** : cette séquence, d'une centaine de paires de bases, définit le site de départ de la transcription et sa direction, plus exactement le promoteur est un segment d'ADN en amont de la séquence codante comportant le site de fixation de l'ARN polymérase ainsi que les sites de fixation des protéines régulatrice de la transcription. Il comporte des séquences très conservées (Figure 4) :

- **La boîte TATA** : Elle est située à environ -25 paires de bases de l'origine de la transcription. C'est une séquence de six nucléotides riches en A et T. La séquence dite consensus (statistiquement la plus

rencontrée) est TATAAA.

- **La boîte GC** : (située le plus souvent dans la région entre -110 et -40). Elle peut se présenter sous forme d'hexanucléotides: 5'-GGGCGG-3'. Le motif riche en bases G et C peut être répété plusieurs fois.

- **La boîte CCAAT** : (souvent située dans la région entre -120 et -80). Cette boîte peut être située avant ou après une boîte GC ou même entre deux boîtes GC.

Par ailleurs, des séquences régulatrices sont présentes à des distances très variables du promoteur. Ces séquences constituent des éléments fondamentaux de la régulation de l'expression du génome des eucaryotes, leur importance fonctionnelle venant de leur propriété d'être régulées par des facteurs spécifiques de la différenciation tissulaire, voire par des hormones. Il s'agit des séquences extinctrices (ou inhibitrices) appelées « silencers » ou des séquences stimulatrices les dits « enhancers » (Figure 5).

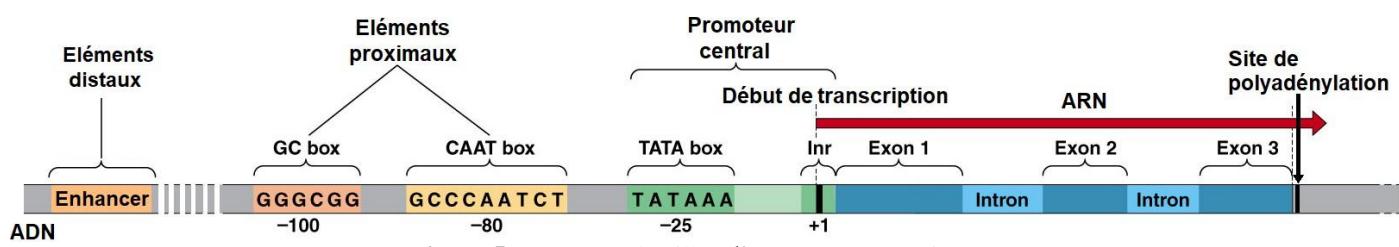


Figure 5 : Structure détaillée d'un gène eucaryotique.

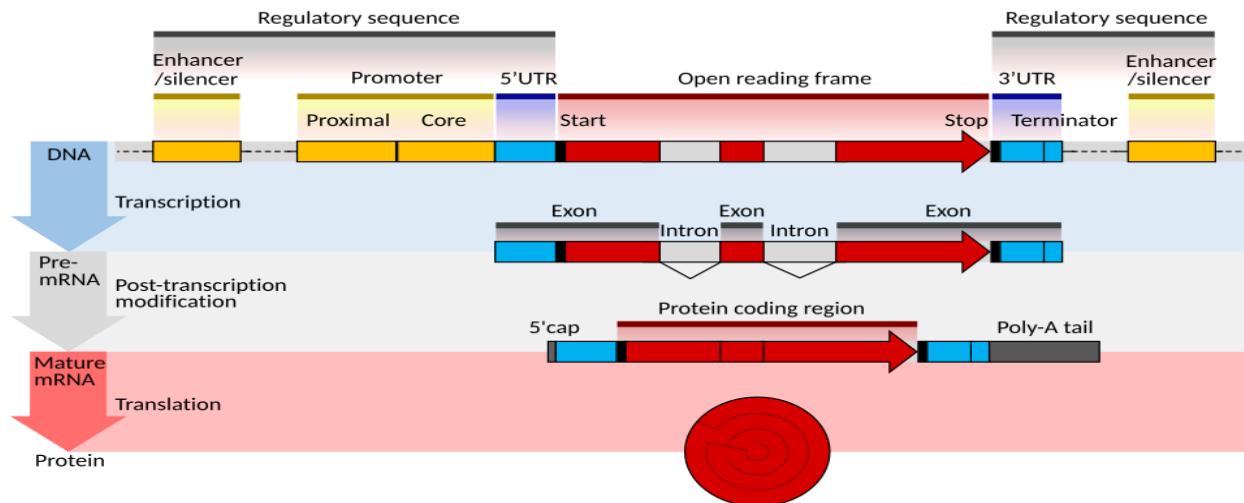


Figure 6 : Structure détaillée d'un gène eucaryotique

NB. Les régions non traduites de l'ARNm (en bleu) régulent la traduction en produits protéiques finaux

5-Classification des gènes : Les gènes sont classés selon

- **La complexité :**
 - Gène unique
 - Famille de gènes
 - Super famille de gènes
- **Le tissu d'expression :**
 - Gène spécifique

- Gène domestique qualifie un gène qui s'exprime dans tous les types cellulaires et dont les produits assurent les fonctions indispensables à la survie des cellules. Ils ne subissent donc pas de régulation.

• **La fonction :**

- Gène actif

- Pseudogènes Sont des gènes non fonctionnels (par absence de région promotrice, de codon d'initiation, d'un cadre de lecture suffisant). Il existe 2 sortes de pseudogènes :

-Les gènes dupliqués (mal dupliqués ou mutés au cours de l'évolution : ils sont pourvus d'introns).
-les gènes rétrotranscrits à partir d'ARNm (les rétropseudogènes n'ont ni promoteur, ni introns, mais une séquence poly (A)).

✓ Comme ils ne sont pas fonctionnels, les pseudogènes ne subissent aucune pression de sélection et les mutations s'y accumulent en grand nombre.

• **Le type de protéine :**

- Gène de structure : est un gène qui détermine la séquence d'acides aminés d'une protéine

- Gène de régulation : ou gène de contrôle est un gène qui possède la capacité de contrôler le fonctionnement des gènes de structure. C'est ainsi qu'il élabore selon les cas, un répresseur ou un inducteur de façon à freiner au contraire déclencher voir accélérer la transcription.

• **La transcription :**

- Gènes de classe I : ARNr

- Gènes de classe II : ARNm

- Gènes de classe III : ARNt

6-Les gènes procaryotiques

Chez les procaryotes, le matériel génétique est libre dans la cellule, le génome de la plupart des organismes procaryotes correspond à un seul chromosome composé d'un ADN souvent circulaire et de très peu de protéines associées. L'ADN associé à quelques protéines (non-histones) forme une masse dense, le nucléoïde. Les gènes bactériens sont souvent organisés en unités appelées opérons, transcrits en un seul ARN messager, le cytoplasme contient également des structures facultatives constituées d'ADN, appelées plasmides.

Chez les procaryotes, un gène est défini structurellement comme une séquence d'ADN comprenant **un promoteur, un site d'initiation et un site de terminaison**.

A) Le promoteur

Le promoteur correspond à une région non transcrive de l'ADN, généralement juste en amont du début de la région transcrive, dont la séquence permet le recrutement de l'ARN polymérase.

Chez les procaryotes les promoteurs font environ 40pb (région couverte par l'enzyme) et qui contiennent 2 séquences conservées:

- Une séquence consensus de 6 nucléotides, placée en -35 (-30à-35) du +1 de transcription.

Exp: TTGACA. Le rôle de cette séquence est de donner un signal pour la reconnaissance du promoteur par l'ARN polymérase.

- Une séquence de 6 nucléotides en -10 ou -12 du +1 de transcription = **boîte TATA box (Pribnow)**.

Cette dernière facilite la dissociation des deux brins d'ADN, car riche en A et T.

EXP: TATATT.

B) Site d'initiation

Par convention on appelle **+1** le premier nucléotide à partir duquel la transcription démarre et -1 qui précède. Le premier nucléotide est très souvent A ou G.

C) Sites de terminaison

Ce sont des sites qui indiquent la terminaison de la transcription. Les terminateurs font généralement partie de la séquence codante.

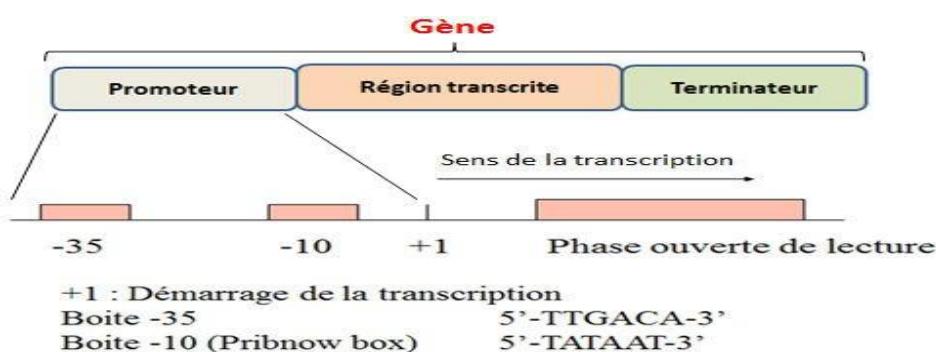


Figure 7 : représentation schématique d'un gène chez les procaryotes

Les opérons :

Chez les procaryotes, il existe des gènes organisés en opérons (voies métaboliques). Chaque opéron comporte un nombre variable de gènes de structure contigus qui possèdent un même promoteur et donnant des ARN **polycistroniques** (**donnant naissance à plusieurs protéines en même temps**). Mais on trouve également des gènes de structure plus simple ne contenant, comme chez les eucaryotes, qu'une seule unité de traduction. Chez les eucaryotes, les ARNm sont **monocistroniques**.

Opéron: Unité d'expression de gènes, codant plusieurs enzymes apparentées ou des ARNr, sous le contrôle d'un même promoteur: co-transcrits générant un long ARNm.

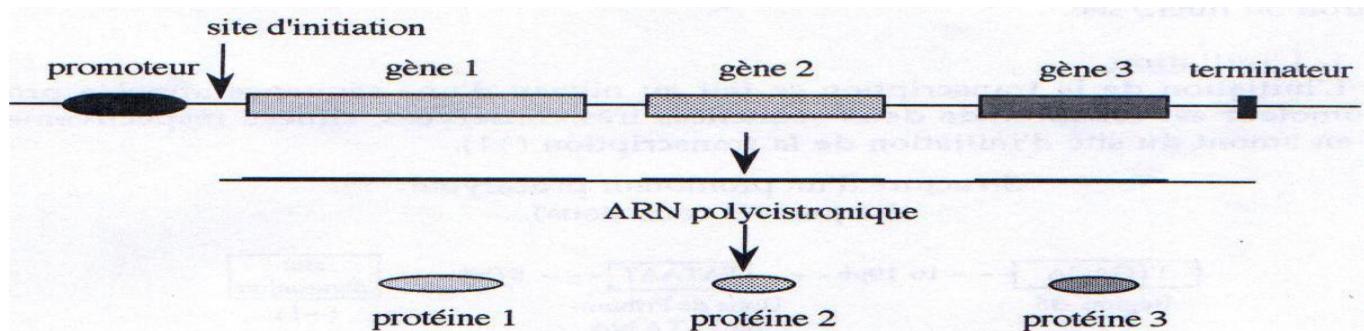


Figure 8 : Structure détaillée d'un gène procaryotique.

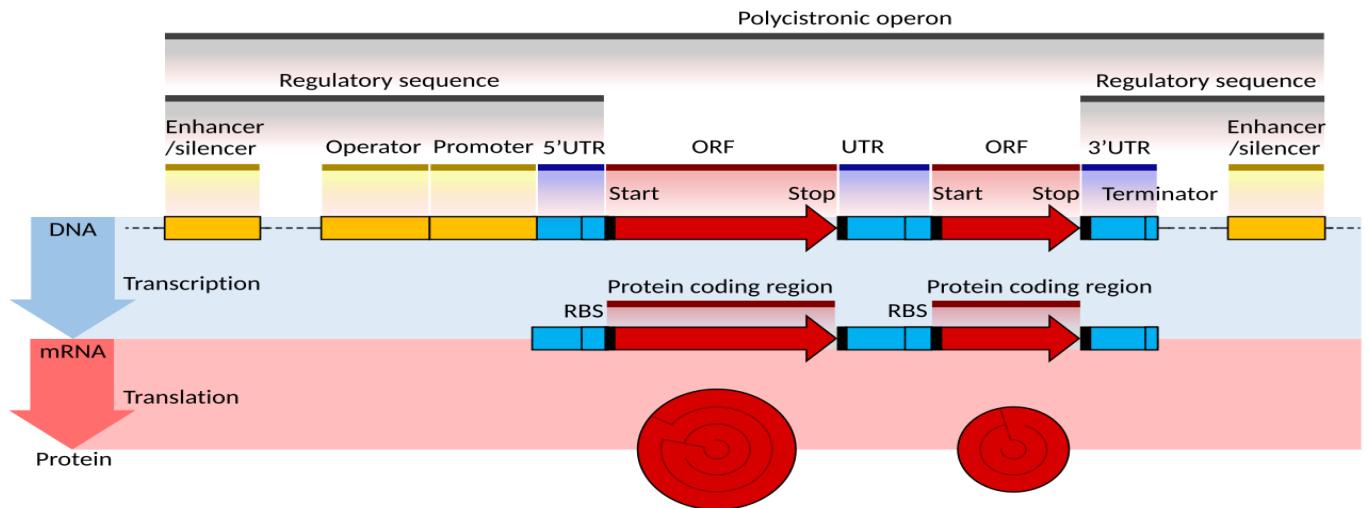


Figure 9 : Structure détaillée d'un gène procaryotique

NB. La structure d'un opéron procaryote de gènes codant pour des protéines. La séquence régulatrice contrôle le moment où l'expression se produit pour les multiples régions codantes pour des protéines (en rouge). Les régions promotrices, opératrices et *enhancers* (en jaune) régulent la transcription du gène en ARNm. Les régions non traduites de l'ARNm (en bleu) régulent la traduction en produits protéiques finaux.