

Chp 4 : Caractérisation des gènes

A – définition d'un gène

début et fin?

séquence d'initiation?

séquence de régulation?

**séquence de terminaison non
traduite?**

Gène = unité d'information génétique

**région du génome situé entre les bases
5' et 3' terminales de l'ARNm**

Gène = Unité de transcription



définition élargie:

→ + régions de régulation

→ unité de transcription =

unité d'information génétique

+ promoteur

+ régions régulatrices

+ régions de terminaison

a) Les séquences promotrices et activatrices

↪ Région promotrice = promoteur

nécessaire à la transcription d'un
gène

* situé sur le brin sens

* site de fixation de l'enzyme

* site de dissociation des 2 brins

* site d'initiation :

site d'incorporation du
1er nucléotide de l'ARN

= site « n + 1 »

chez les bactéries : *E. coli*

taille minimale : 12bp

Mais pas toutes juxtaposées

Parfois répétées

**Séparées par des fragments d'ADN
informatif par la longueur
non par la séquence**

**Souvent 2 séquences conservées de 6 bp
= hexamères**

- position « -10 » : TATA box

séquence consensus : TATAAT

**domaine d'activation du
complexe protéique de l'ARN polymérase**

- position « - 35 »

séquence consensus : TTGACA

**domaine de reconnaissance de
l'enzyme**



**Localisation très fréquente : séparation par
une distance appropriée pour la géométrie
de l'enzyme**

Mutation des séquences promotrices

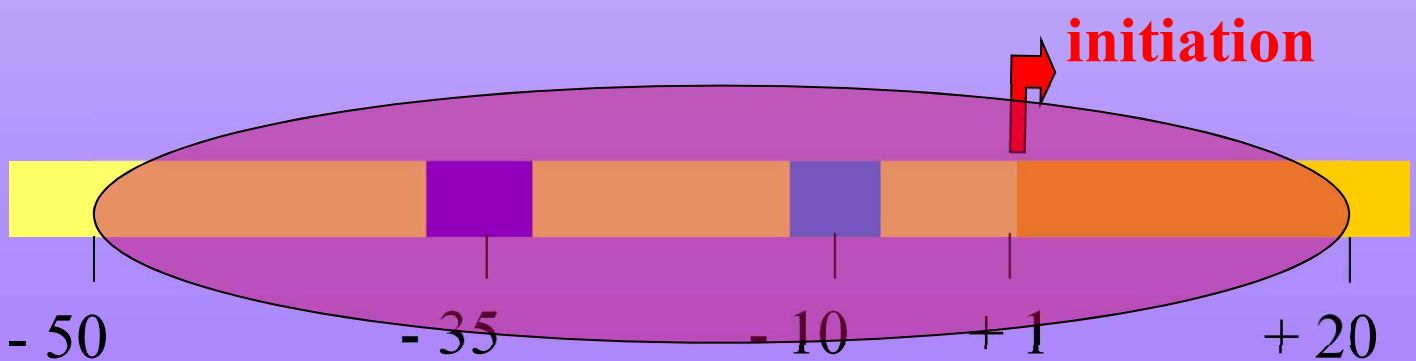
→ affecte le niveau d'expression de gène

→ n'affecte pas la nature du produit d'expression

Site de fixation de l'ARN polymérase :
beaucoup plus large

-50 à + 20

de part et d'autre du site d'initiation




Chez les Eucaryotes :

grande diversité des promoteurs



liée aux différents types d'ARN polymérase

*** pol I :  synthèse des ARNr
très peu variables**

2 séquences :

*** promoteur central**

de «- 45 » à « + 20 » bp

 suffisante

*** élément de contrôle**

position : « -180 » à « - 107 »

Pas indispensable mais

augmente l'efficacité

Curieusement : 2 régions riches en CG

identité : 85%

*** pol II :  synthèse des ARNm**

**homologies de séquences
très courtes**

proche du site d'initiation

- initiateur (Inr)

entre « - 3 » et « + 5 »

séquence consensus : Py₂C/APy₅

- « TATA » box


position : « -25 »

taille : 8 bp  T/A

**entourée de régions riches en
GC**

**site de fixation d'un facteur
protéique**

 indispensable à la transcription

*** pol III :  synthèse d'ARN de plusieurs classes**


très variables

2 grandes classes


• Situées en amont du site d'initiation

proches des promoteurs de pol II

 « TATA » box

 responsable de la spécificité de la polymérase

efficacité accentuée par la présence d'autres séquences

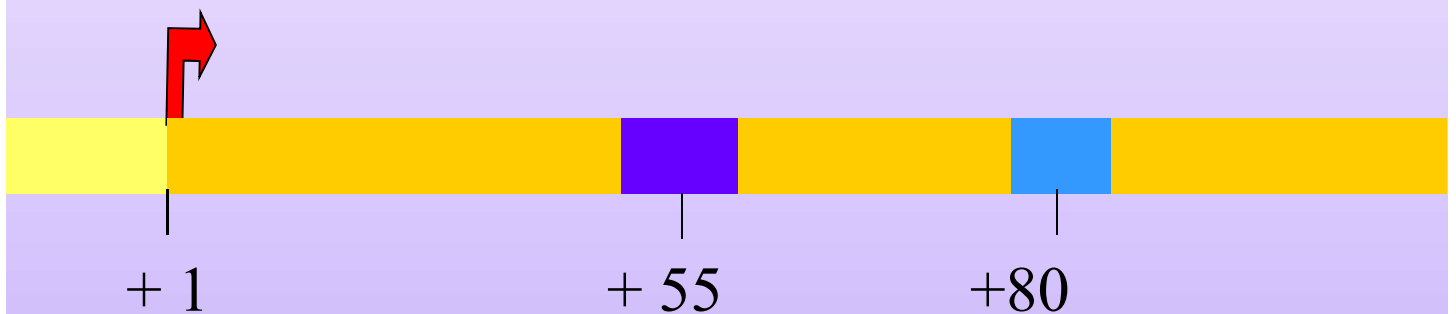
 fixation de facteurs influençant la transcription

• **Situées en aval du site d'initiation**

ex : ARN 5S

position : +55 à +80

 **promoteurs internes**



2 séquences consensus de distances définies

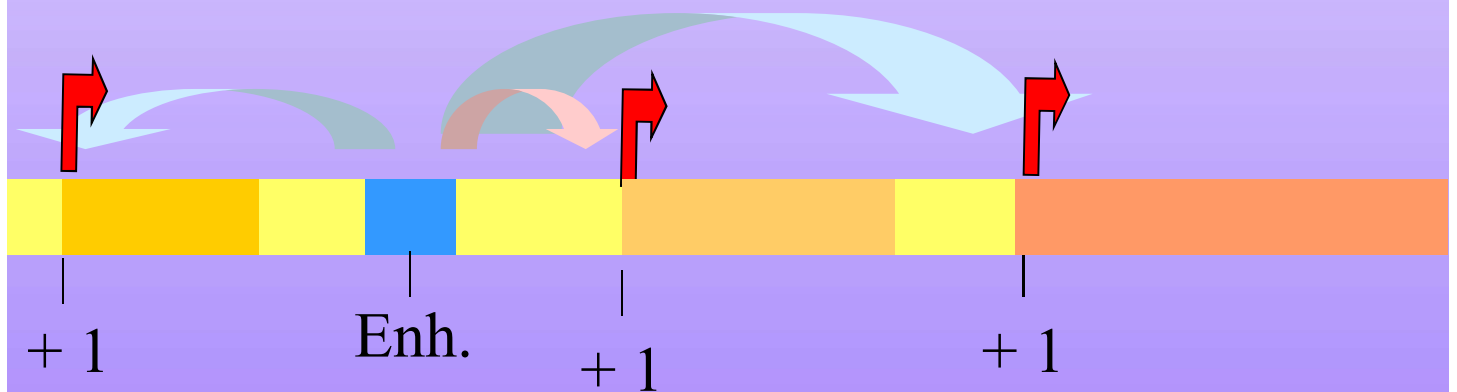
fonction - du gène

- des facteurs impliqués dans l'initiation de la transcription

↪ **Séquences activatrices (enhancer)**
active l'expression
mais pas indispensable
Position variable/ gène (amont/aval)
Action simultanée possible sur plusieurs gènes

Impliqués dans les régulations

tissus -spécifique



b) Les séquences terminatrices



**fin d'un gène moins bien définie que
le début**

**car transcription possible au delà de
l'extrémité 3' clivée**

**Ensemble de mécanismes d'arrêt de la
transcription**

Séquence nécessaire : terminateur

**→ chez les procaryotes et beaucoup
d'eucaryotes : formation d'une épingle à
cheveux**

**= structure secondaire de l'ARN
transcrit**

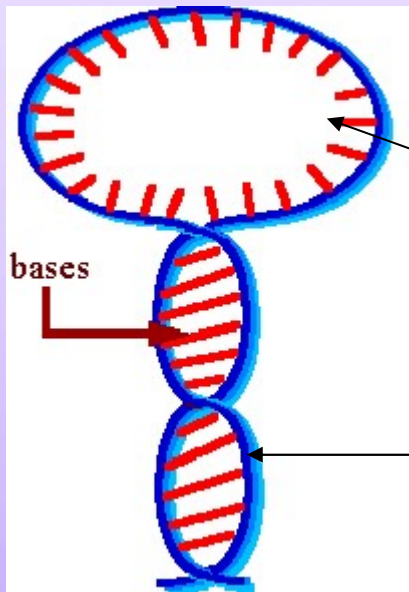
Liée à une séquence riche en GC

position intrinsèque

→ clivage en aval de l'épingle

Épingle à cheveux :

↪ séquences palindromiques
séparées par une courte
distance = séquence unique



Séquence
unique

Séquences
palindromiques

Structure secondaire de l'ARN

Pas d'arrêt de la transcription

↪ ralentissement de l'ARN
polymérase

séquences voisines riches en A

souvent 6 A successives sur le
l'ARN transcrit

↪ Arrêt

Liaison A-T faible énergétiquement

→ libération de l'ARN

Chez les Eucaryotes

↳ moins de choses connues

Analyse de l'ARN peu informative

car extrémité 3' =

- fin de transcription

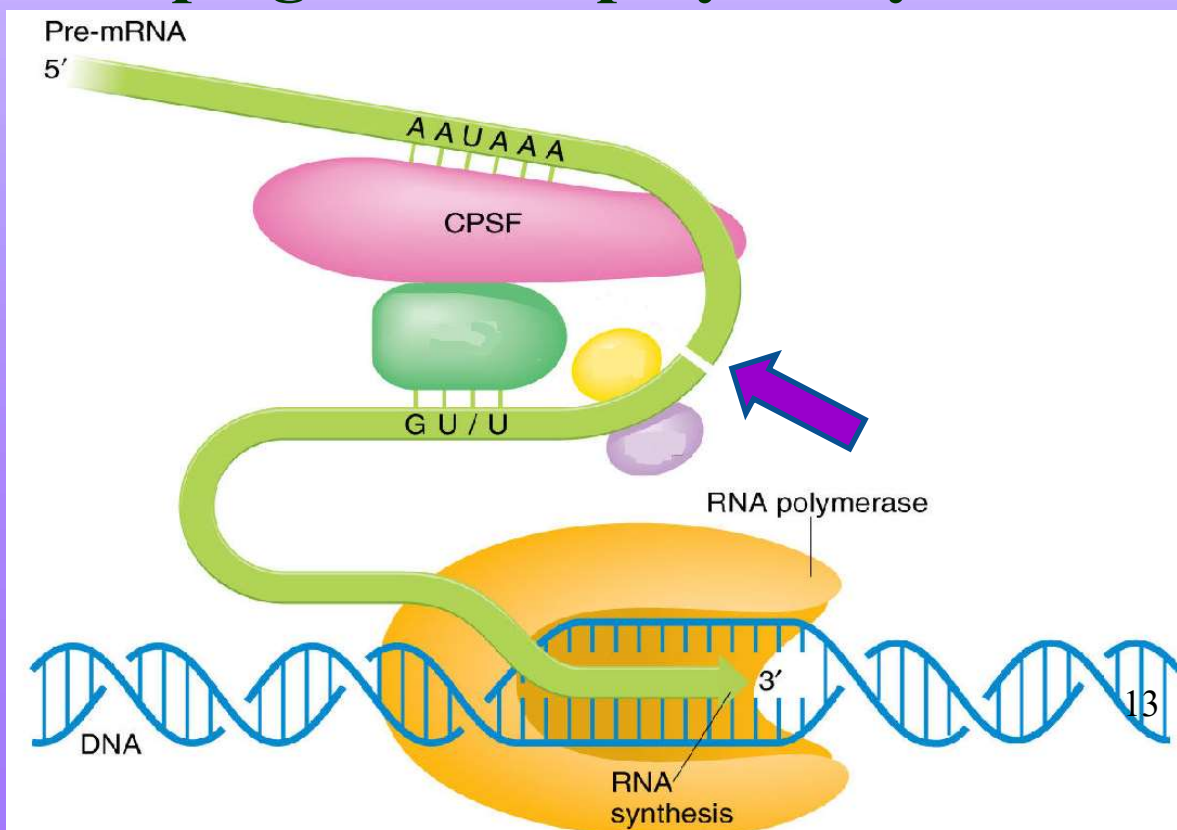
Ou

- clivage secondaire de la molécule
d'ARN

structure en épingle à cheveux

pas toujours observées

couplage avec la polyadénylation



II– Les gènes mosaïques

Faible partie du génome

↳ production de polypeptides

Taille du gène plus élevée que taille de
l'ARN traduit

existence d'interruptions

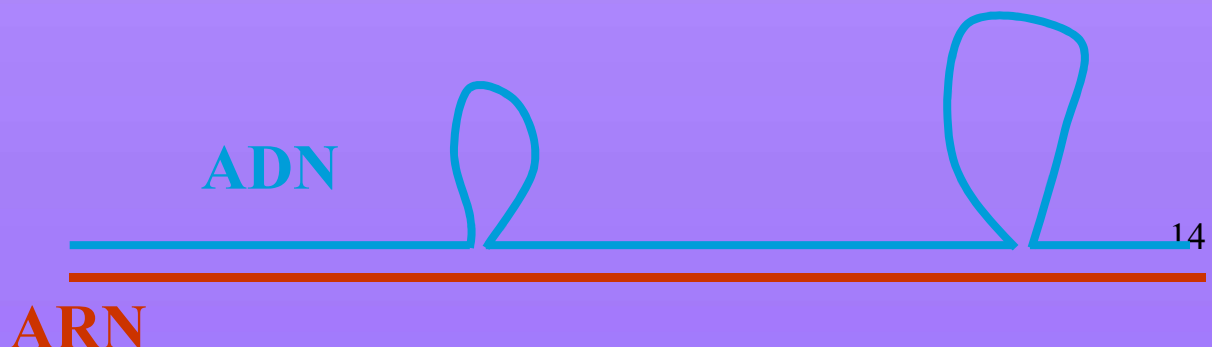
↳ séparation de séquences codantes
dans l'ADN

gènes morcelés

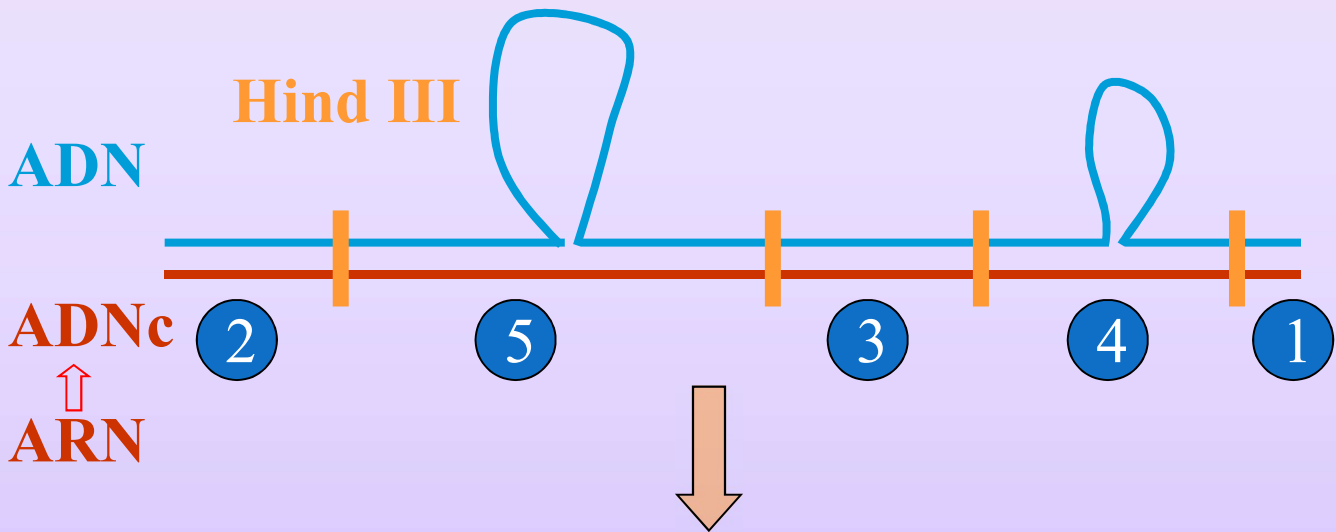
= gènes mosaïques

Microscopie électronique : comparaison

Taille ADN/ taille transcrit



Comparaison des cartes de restriction



Existence de séquences sans fonction

codante

= introns

= séquences introniques

Séquences codantes : exons

Par définition:

**Un gène commence et se termine par un
exon**



signification d'une région d'ADN

variable

a) Localisation des gènes mosaïques

Apparition chez les Eucaryotes inférieurs

*** aucun groupe d'Eucaryotes n'est
dépourvus**

*** toutes les catégories de gènes
peuvent être morcelés:**

- codant les protéines

- codant les ARNr

- codant les ARNt

- gènes mitochondriaux

- gènes chloroplastiques

**Mais existence de catégories de gènes non
mosaïque: ex : histone**

= gènes sans intron

**Présence de gènes mosaïques chez les
archées**

MAIS absence chez les eubactéries

b) Organisation des gènes mosaïques

*** alternance d'exons/d'introns**

**nombre d'introns variable d'un gène
à l'autre**

Mais caractéristiques communes :

- ordre identique sur l'ADN et l'ARN

 **gène morcelé**

non dispersé

- structure identique dans tous les tissus

*** lignées germinale et somatique
incluses**

*** gène exprimé ou non**

 **présence des introns invariable pour
un gène donné**

Caractéristiques communes :

- présence fréquente de codons de terminaison dans tous les cadre de lecture dans les introns

* taille des exons généralement faible :
codant 100 AA

* taille des introns beaucoup plus élevée

* nombre d'exons tend à augmenter avec la taille de la protéine

→ nombre d'introns aussi

Ex : Gène de la globuline : 125 bp codantes

⇒ 3 exons

Gène codant la pro α 2 collagène:

5 000 pb codantes

⇒ 50 exons : taille de 45 à 249 bp

introns parfois courts (cf exons)

Parfois longs

5 000 pb codantes sur 40 000 pb

 **Introns : proportion plus importante
que les exons**



Pas de relation directe

taille du gène/ taille de l'ARN

Pas de relation :

**taille du génome/ complexité de
l'organisme**

Cependant

D'une façon générale :

**Eucaryotes inférieurs : introns moins
fréquents**

 **gènes moins morcelés**

c) Mécanisme d'élimination des introns

Différence de taille ADN/ARN

élimination des introns sur l'ARN

avant la traduction

= excision-épissage

1ère étape de transcription



produit primaire

complémentaire du brin codant de l'ADN



ARN précurseur

Pas d'utilisation possible en l'état



Mécanismes de maturation



obtenir la matrice pour la traduction

Maturation : 2 étapes

Excision de l'intron



puis épissage des 2 extrémités

des exons de l'ARN

→ très grande précision

liée à la présence de séquences

spécifiques et bien conservées

incluse dans l'intron



éliminées

Intron : commence par G -T

se termine par A-G



Règle GT-AG

« GU-AG » car ARN

Jonctions exons/introns différentes



orientation possible de l'intron

Définition directionnelle des extrémités:

→ **site d'épissage gauche ou 5'**

= site donneur

→ **site d'épissage droit ou 3'**

= site accepteur

+ séquence consensuelle à l'intérieur de l'intron

UACUUAUCC

→ **permet la formation d'une boucle**

Ensemble de 3 séquences consensus chez les eucaryotes supérieurs



mécanismes communs

Séquences impliquées



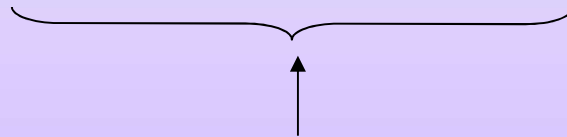
structures secondaires



rapprochement physique

des extrémités à épisser

UACUUAUCC



Ex1 ----- **intron** ----- **Ex2**

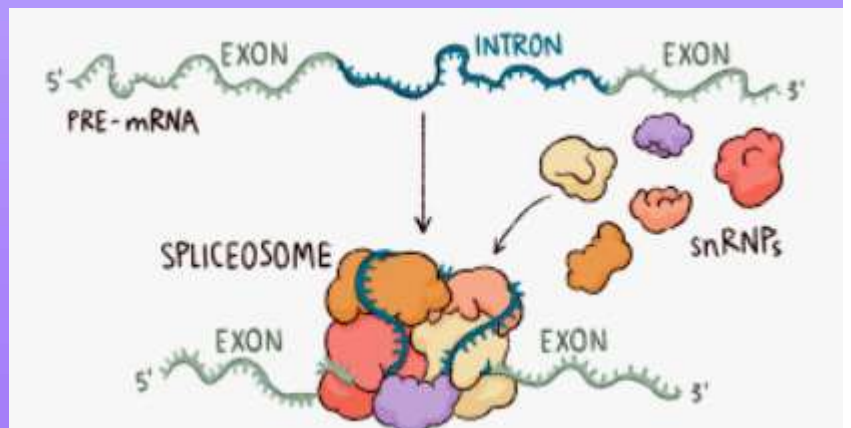


GGGUGAGUA

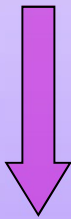
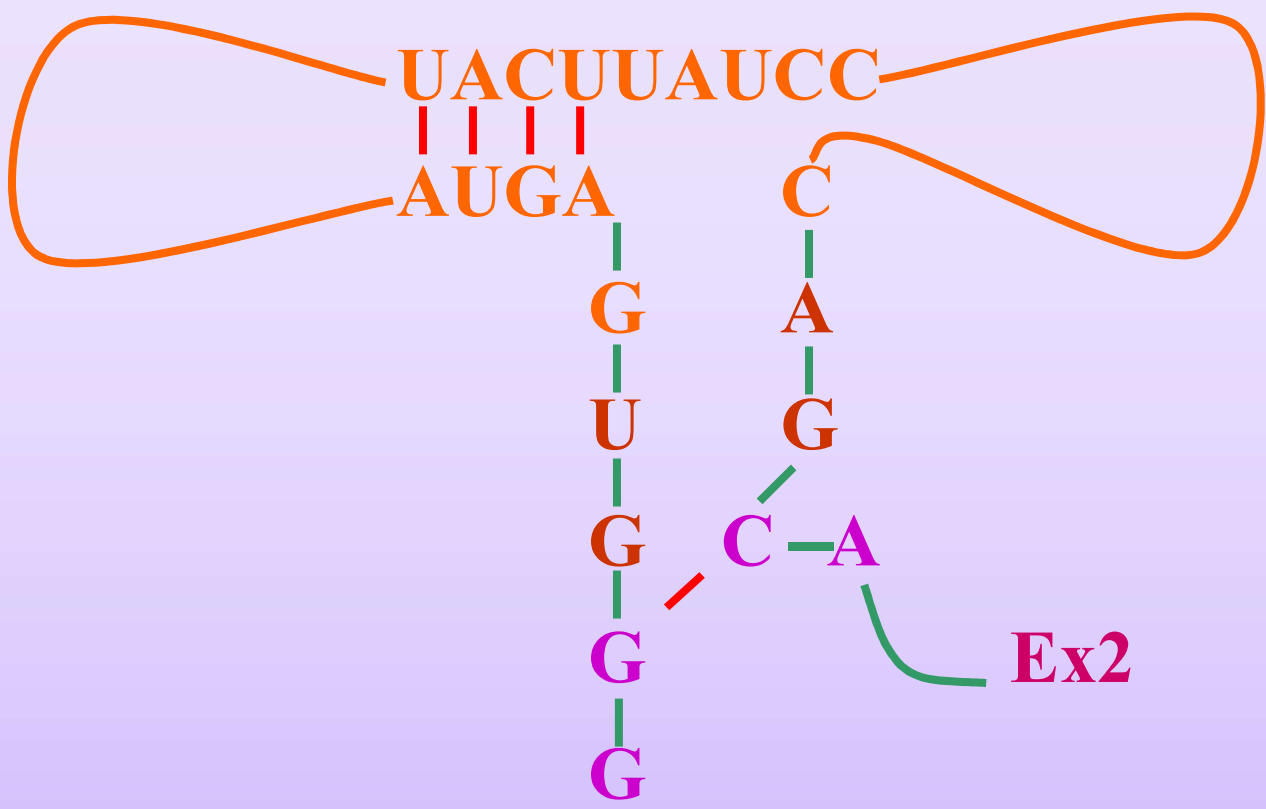
CAGCA

Action de complexes ribonucléoprotéiques

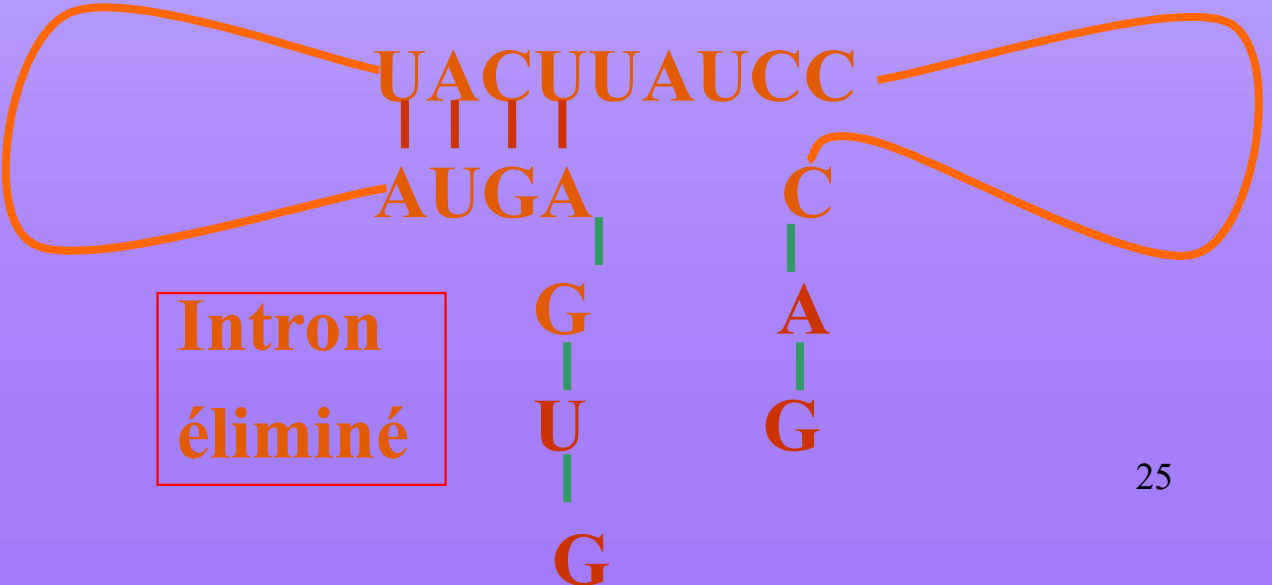
→ spliceosome



Structure conservée chez tous les Eucaryotes



+



d) Relations entre introns et exons de différents gènes

exons de gènes souvent

→ **apparentés dans un même groupe phylétique**

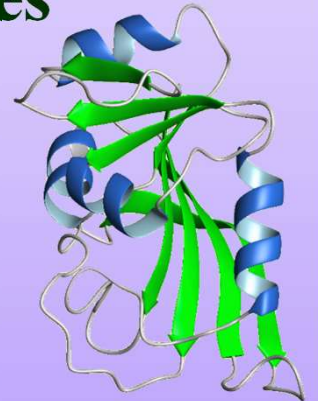
Mais introns peu ou pas apparentés



plus grande diversité

en nombre

en taille



Ex : Dihydrofolate réductase (DHFR)

chez les mammifères

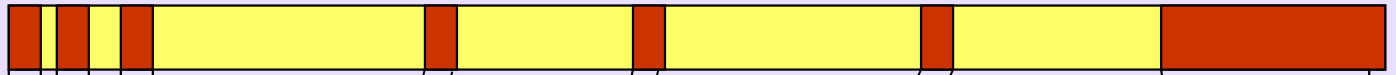
exons : +/- identiques

introns :

position relative conservée

mais longueur très variable

Humain



Souris



Hamster



Variation de taille

des introns et des exons de la DHFR

Relation de parenté entre 2 gènes

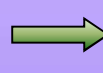
 **visible au niveau des exons**

 **degré de similitude plus élevé**

Exons : codent des acides aminés

 **potentiel de modifications de séquences limité**

 **modifications de séquences protéiques**

 **Divergence des exons +/- corrélée à la divergence entre les protéines**

 **substitutions de bases**

 **pas de délétion/insertion**

Introns : pas de rôle codant

 **peu de contrainte fonctionnelle**

 **accumulation possible de mutations:**

 **substitutions de bases**

Modifications  **changements de taille**

 **insertions**

 **délétions**



**Degré de divergence des introns plus élevé
que degré de divergence des exons**



**Mise en évidence de la
divergence génétique**

**Même taux de mutations au niveau des
exons et des introns**

MAIS

**Élimination plus efficace
au niveau des exons**

 **Phénomènes de contre-sélection**



**élimination des mutations des exons
plus rapide que pour les introns**



Evolution plus rapide des introns

III– Les gènes polycistroniques

Chez les procaryotes eubactériens

cistron : unité de transcription

= séquence codante

+ séquences promotrices

+ séquences de terminaison

**+ séquences de régulation
(si présentes)**

**Ensemble de séquences sous l'action d'un
même promoteur**

codent des protéines différentes



cistrons distincts



gène polycistronique

Eucaryotes :

**ensemble de séquences codantes sous
contrôle d'un même promoteur**

→ 1 seule protéine

→ 1 gène mosaïque

Eubactéries :

**ensemble de séquences codantes sous
contrôle d'un même promoteur**

→ plusieurs protéines

→ plusieurs cistrons


→ 1 gène polycistronique

Souvent :

code des protéines d'une même

voie métabolique

= opérons

 **grande souplesse de régulation**

Non utilisation d'une chaîne métabolique

 **pas de besoin protéique**

répression génique

de l'ensemble de la chaîne

Utilisation de la voie métabolique

 **Besoins protéiques**

induction de l'expression de tous les

gènes simultanément

 **par action sur 1 seule séquence
promotrice**

Synthèse de l'ensemble des protéines en quantités relatives toujours identiques

Organisation en groupe d'enzymes

↻ 1 seul transcrit primaire

= ARNm polycistronique

↻ traduction séquentielle
en différentes protéines

