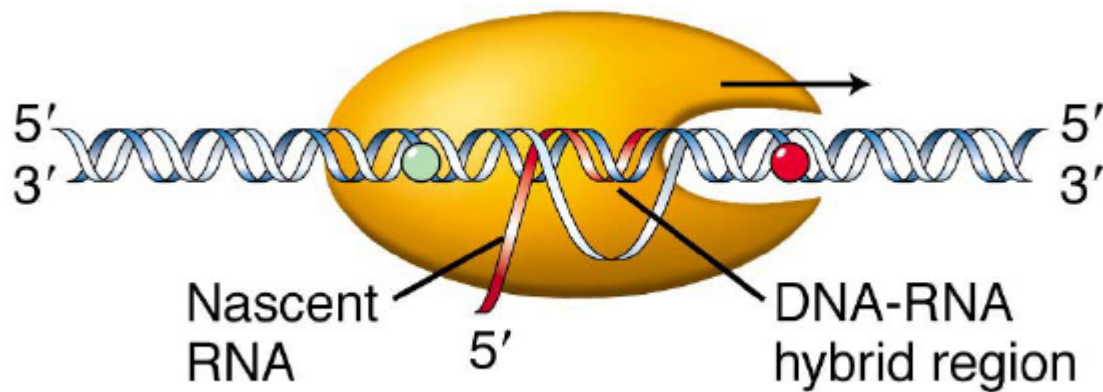
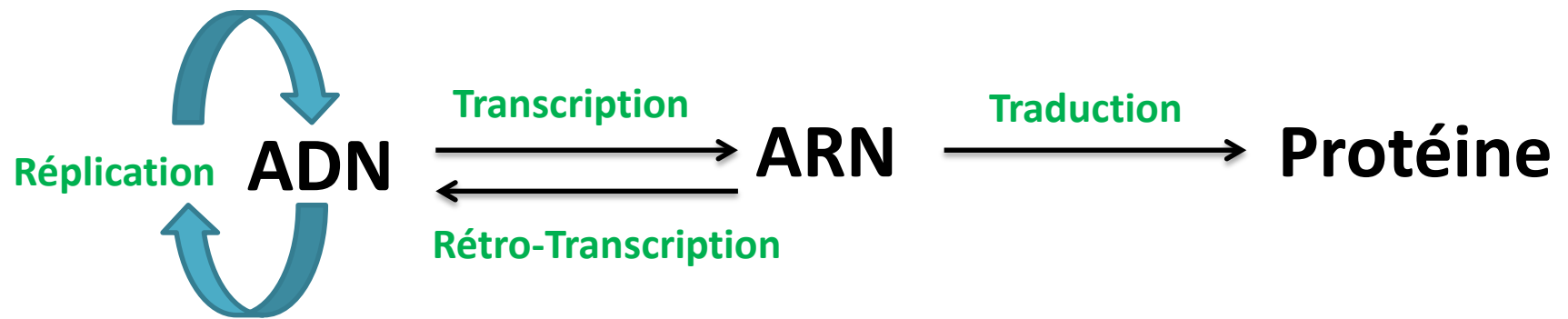


# La Transcription

## Chapitre II



# Le «Dogme Central Simplifié» de la Biologie Moléculaire



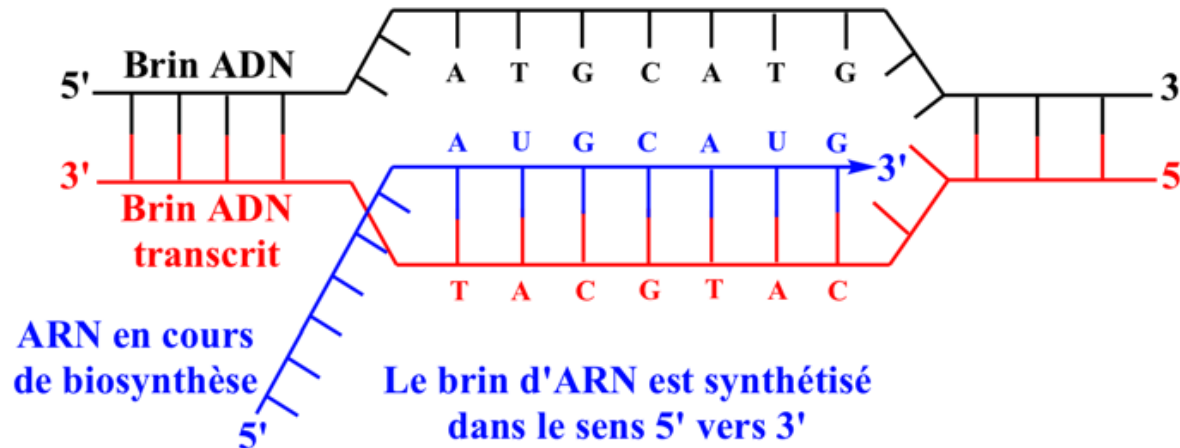
# Objectifs du cours

- Comprendre la transcription chez les procaryotes et les eucaryotes
- Comprendre les étapes clés de la transcription de l'ADN en ARN
- Comprendre la structure des gènes des procaryotes et eucaryotes
- Comprendre la structure et la fonction des ARN polymérases
- Comprendre la structure du promoteur et du terminateur
- Comparer la transcription entre procaryotes et eucaryotes
- Décrire les techniques moléculaires les plus couramment utilisées pour étudier la transcription.

# Définition: Transcription

- Synthèse d'un acide ribonucléique (ARN) par une ARN polymérase à partir d'une molécule d'ADN
- La structure primaire de l'ARN nouvellement formée est la copie identique du brin codant de l'ADN
- La transcription est une étape clé de l'expression d'un gène: synthèse de macromolécules (ARN et protéines) à partir d'une séquence d'ADN.

# Transcription: Synthèse de l'ARN

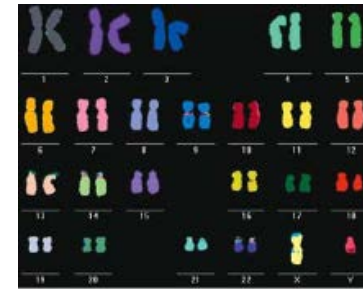


- La molécule d'ADN est formée de deux brins:
  - **Brin transcrit** (3'-5'): brin anti-codant: brin matriciel qui servira de matrice/modèle à l'ARN-polymérase
  - **Brin codant** (5'-3'): a une séquence identique à l'ARN néoformé (à l'exception que la Thymine est changée par l'Uracile dans l'ARN)
- Les ribonucleotides (rNTPs) sont ajoutés à l'extrémité 3' de la chaîne en cours de biosynthèse selon la complémentarité avec la matrice d'ADN
- **L'ARN messenger (ARNm)** synthétisé est monocaténaire (un seul brin) et est dirigé dans le même sens que le brin codant de l'ADN (5'-3').

# Généralités: Gène

- Un gène est un **segment d'ADN** qui contient l'information nécessaire pour la synthèse d'un acide ribonucléique (ARN) ou d'une protéine

- Un gène se caractérise par sa:
  - position sur le chromosome
  - série de succession des séquences de bases



ATGCTAGGTAT  
TUTCCGAATGA  
CCATTTAGCCG  
ATTAGGA.....

- Les gènes sont **Polycistroniques** chez les procaryotes. Ils forment une seule unité de transcription (plusieurs gènes sont transcrits en une seule molécule d'ARN).
- Les gènes sont **Monocistroniques** chez les eucaryotes: la transcription est indépendante entre les gènes et aboutie à la formation de plusieurs unités de transcription (chaque gène produit son ARN messenger primaire).

# Nombre de gènes par espèces

	<b>Espèce</b>	<b>Taille du génome</b>	<b>Nombre de gènes</b>
	Humains	2.9 milliard de paires de bases	20,000-25,000
	Drosophile ( <i>Drosophila melanogaster</i> )	120 million de paires de bases	13,601
	Levure ( <i>Saccharomyces cerevisiae</i> )	12 million de paires de bases	6,275
	Nématode ( <i>Caenorhabditis elegans</i> )	97 million de paires de bases	19,000
	<i>E. coli</i>	4.1 million de paires de bases	4,800
	Arabidopsis ( <i>Arabidopsis thaliana</i> )	125 million de paires de bases	25,000

# Le projet du genome humain

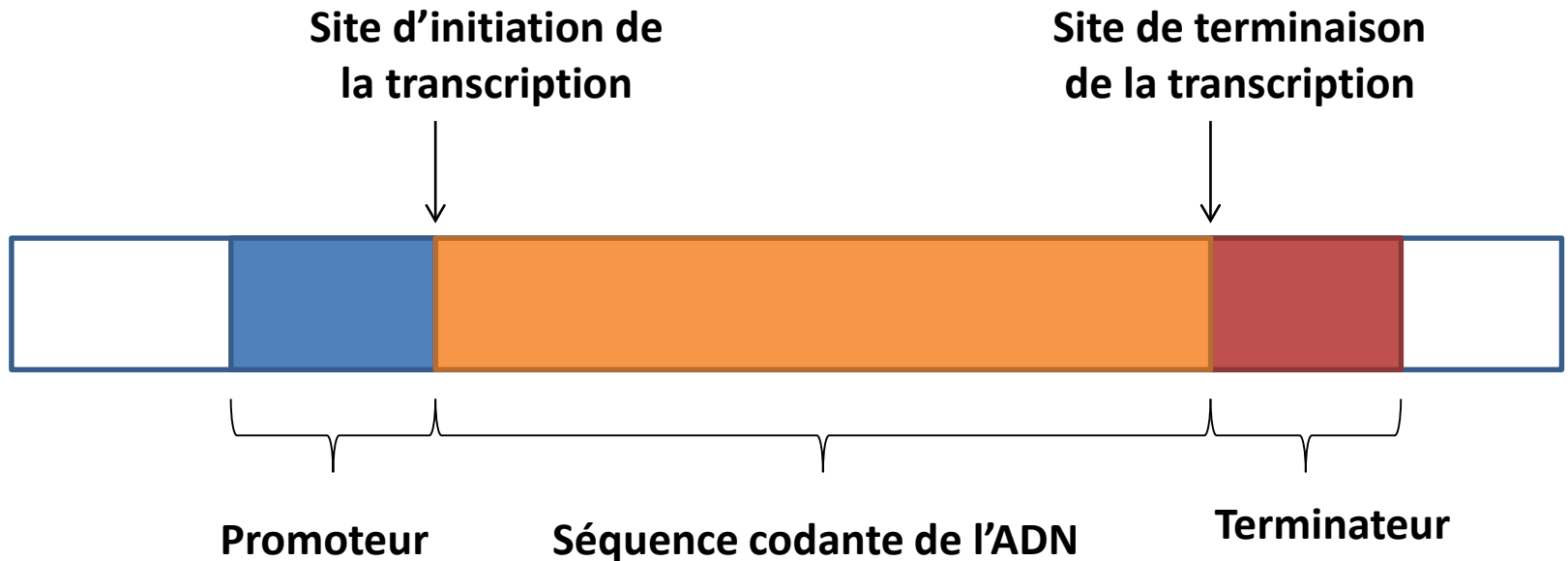


**Nature; February 15, 2001**

- Projet du gouvernement américain lancé en 1986 et terminé en Juin 2000 (annoncé à la Télévision par Bill Clinton)
- **Objectifs du projet:**
  - Identifier les gènes dans l'ADN humain
  - Déterminer les séquences des bases qui constituent le génome
  - Conserver les informations dans une base de données
  - Développer des outils d'analyses de ces données
  - Adresser les aspects éthiques, légaux, et sociaux qui peuvent survenir au cours de la recherche sur le génome
- Le premier génome de référence est un génome composite de différentes personnes (10-20 de races et d'ethnies différentes).



# Structure d'un gène procaryote



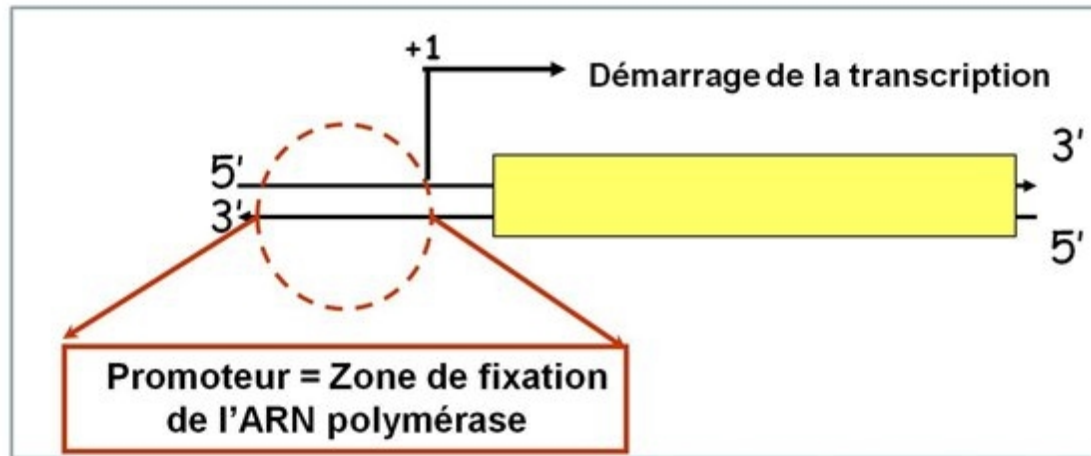
Le gène procaryote est composé de 3 régions:

- Promoteur
- Séquence codante de l'ADN
- Termineur

# Transcription chez les procaryotes

- La transcription a lieu dans le cytoplasme de la cellule bactérienne
- Les bactéries ont une seule ARN-polymérase qui catalyse la synthèse de tous les ARNs
- La transcription se déroule en 3 étapes distinctes:
  - Initiation
  - Elongation
  - Terminaison
- La réaction nécessite la présence
  - ADN bicaténaire (double brin) contenant un promoteur
  - ARN-polymérase
  - Magnésium (agent de protection des groupements SH)
  - Les 4 ribonucleotides (rNTPs).

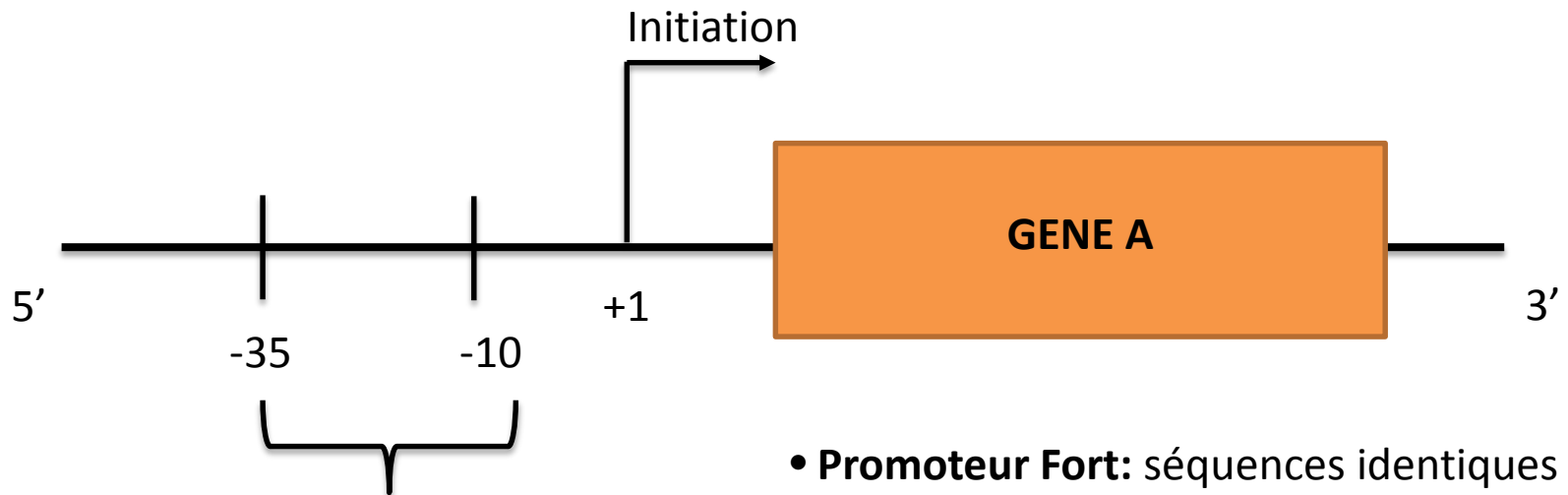
# Le promoteur



- Le promoteur est la région de l'ADN sur laquelle se fixe initialement l'ARN polymérase, avant de commencer la transcription de l'ARN.
- Elle est habituellement située en amont du site d'initiation de la transcription
- Elle contient des séquences reconnues par l'ARN polymérase
- Le promoteur est constitué de courtes séquences conservées appelées séquences consensus
- Promoteur est actif en **Cis**: Il agit sur la transcription du gène qui lui est adjacent sur le même chromosome
- Promoteur est actif en **Trans**: Il n'agit pas sur les gènes situés sur le même chromosome.

# Le promoteur des procaryotes (1)

- Contient deux boites de séquences consensus:
  1. <<TTGACA>>: -35 pair de bases (pb) du site d'initiation
  2. <<TATAAT>>: -10 pb du site d'initiation appelé **boite de Pribnow**

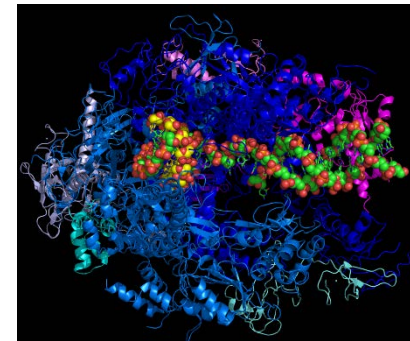
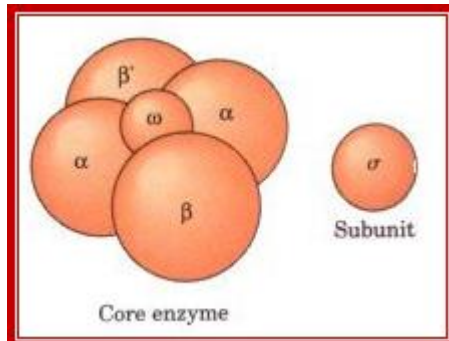


Site de fixation de  
l'ARN-polymerase

- **Promoteur Fort:** séquences identiques aux séquences consensus
- **Promoteur Faible:** séquences divergent des séquences consensus

# L'ARN polymérase bactérienne

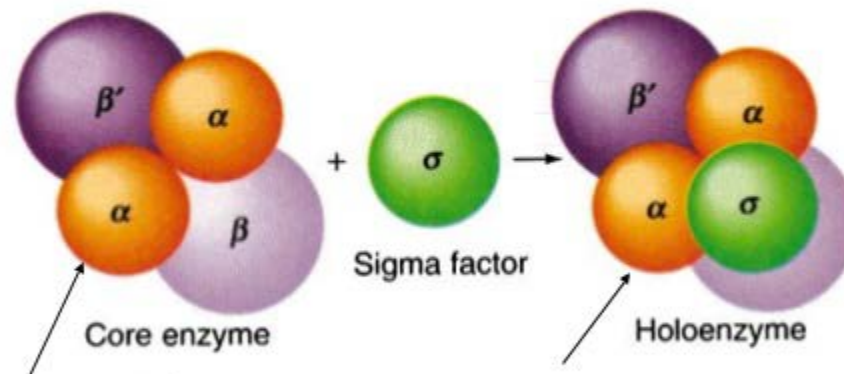
- Complexe enzymatique ADN-dépendante responsable de la synthèse de l'ARN à partir d'une molécule d'ADN
- Elle n'a pas besoin d'une amorce pour initier la synthèse de l'ARN
- Elle ne possède pas d'activité exo-nucléasique
  - Pas de fonction d'édition pour corriger les erreurs
  - Taux d'erreur plus élevé que pour les ADN-polymérases
- Elle est formée de sous-unités  $\alpha$ ,  $\omega$ ,  $\beta$ ,  $\beta'$  et  $\sigma$
- Elle se présente sous deux formes:
  - Le cœur de l'enzyme =  $\alpha_2\beta\beta'\omega$
  - Holoenzyme =  $\alpha_2\beta\beta'\omega\sigma$



# Structure de l'ARN polymérase

## *E. coli* RNA polymerase

2 $\alpha$ , 1 $\beta$ , 1 $\beta'$ , 1 $\omega$  and  $\sigma$  factor

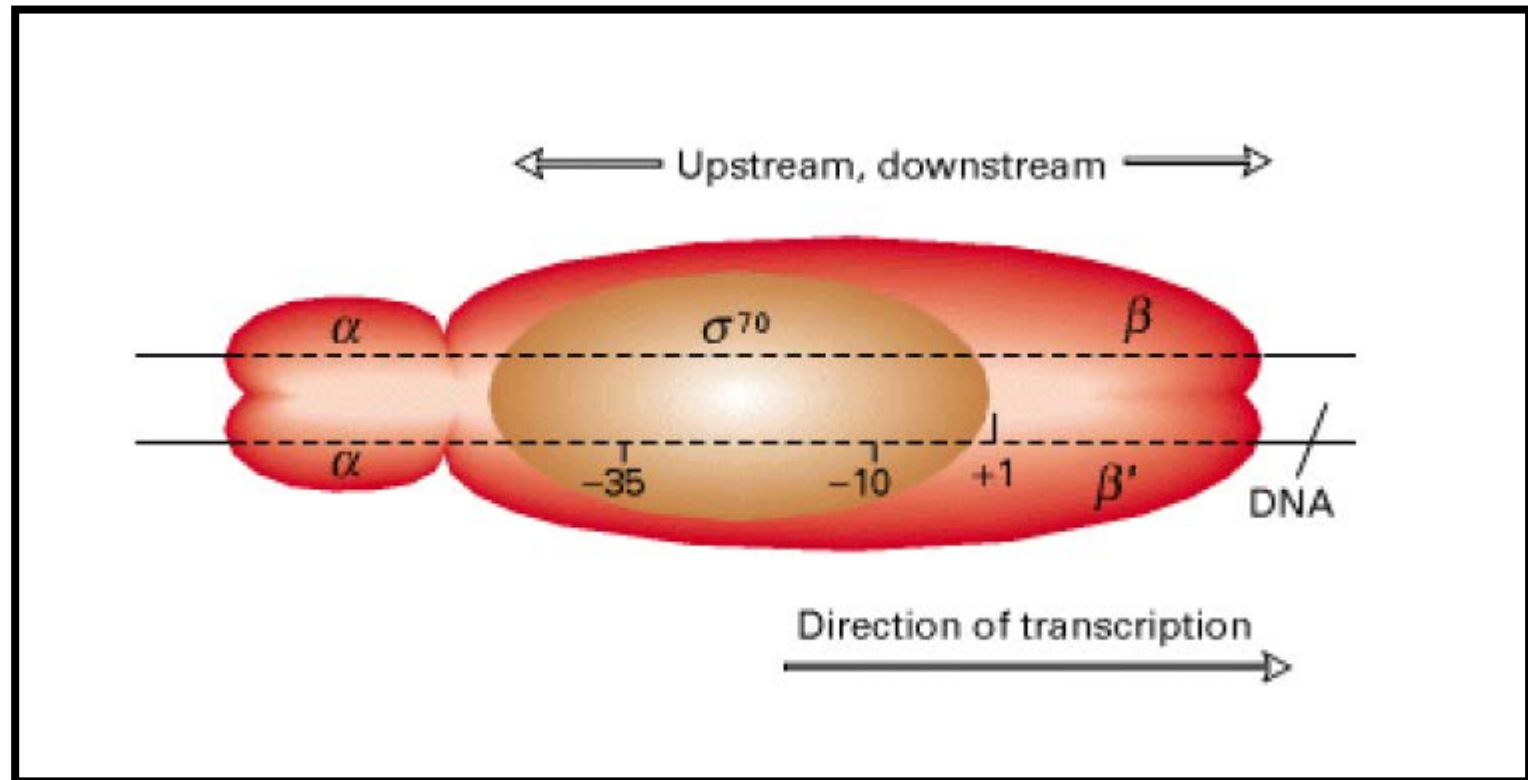


**Le Cœur de l'enzyme: nécessaire pour l'activité de polymérisation de l'ARN naissant. Elle a une affinité faible et non spécifique**

**L'holoenzyme: nécessaire pour l'initiation correcte de la transcription en reconnaissant le promoteur. Elle a une affinité très forte et spécifique**

- $\sigma$  permet une reconnaissance spécifique du promoteur par l'ARN-polymérase (donc elle réduit l'affinité de l'enzyme pour les séquences non promotrices)
- $\sigma$  a elle seule ne peut pas se fixer sur l'ADN. L'interaction avec le promoteur se fait grâce à l'aide de la sous-unité  $\beta$  qui est basique (l'ADN étant acide)

# Interaction entre l'ARN polymérase et le promoteur

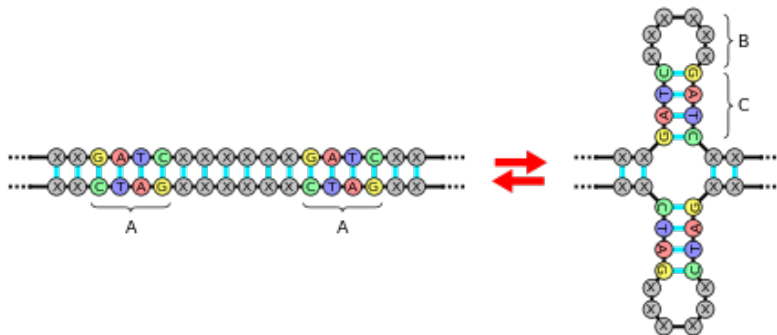


# Le Terminateur

- Le terminateur se présente sous la forme d'un **palindrome**
- Un palindrome est une séquence d'ADN ou d'ARN identique lorsqu'elle est lue dans le sens 5' → 3' sur un brin matriciel ou dans le sens 5' → 3' sur le brin complémentaire.

Exemple: 5'-GAATTC-3'  
3'-CTTAAG-5'

- Le palindrome favorise la formation d'une structure secondaire en **épingle à cheveux** (ou tige-boucle) due à la complémentarité des bases au sein de la molécule d'ARN.



Séquence palindromique d'ADN.

**A** : palindrome ; **B** : boucle ; **C** : tige:

- Il y a deux mécanismes de Terminaison selon l'usage ou pas du facteur Rho:
  - Rho-indépendant
  - Rho-dépendant

- Le facteur Rho est une hélicase ATP-dépendante.



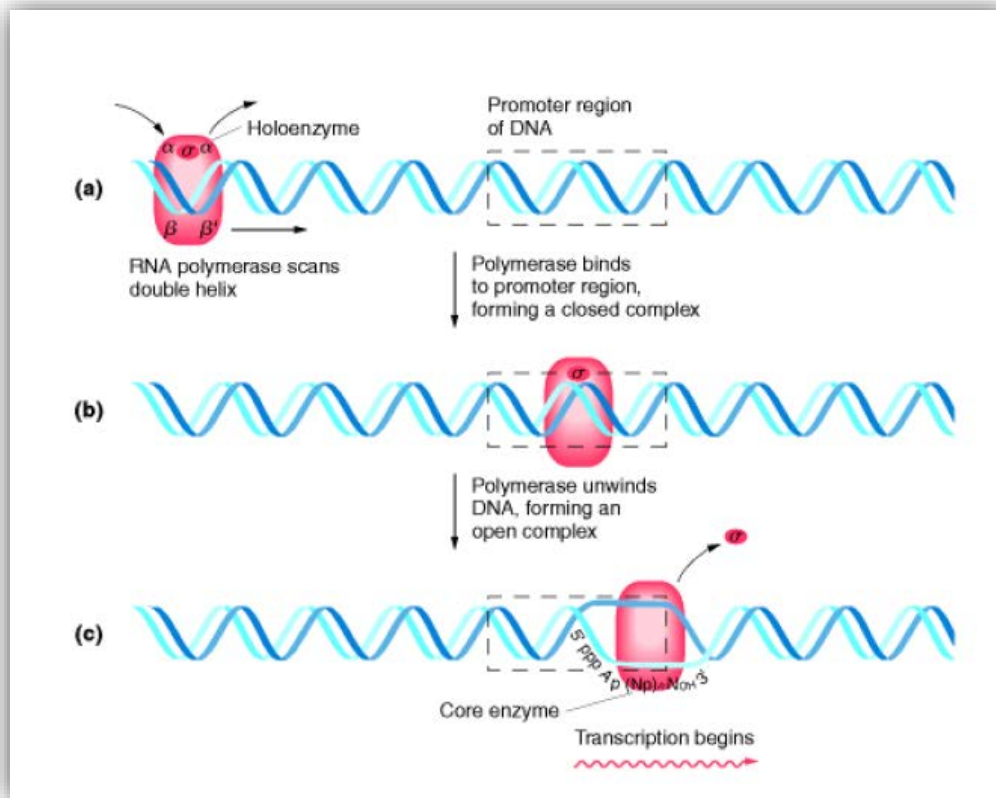
# Les étapes de la transcription

La transcription s'effectue en 3 étapes distinctes:

- Initiation
- Elongation
- Terminaison

# Transcription chez les procaryotes (1)

**A. Initiation:** Elle consiste à la formation de la première liaison phosphodiester qui est réalisée par la sous-unité  $\beta$  qui correspond à la région catalytique de l'ARN polymérase



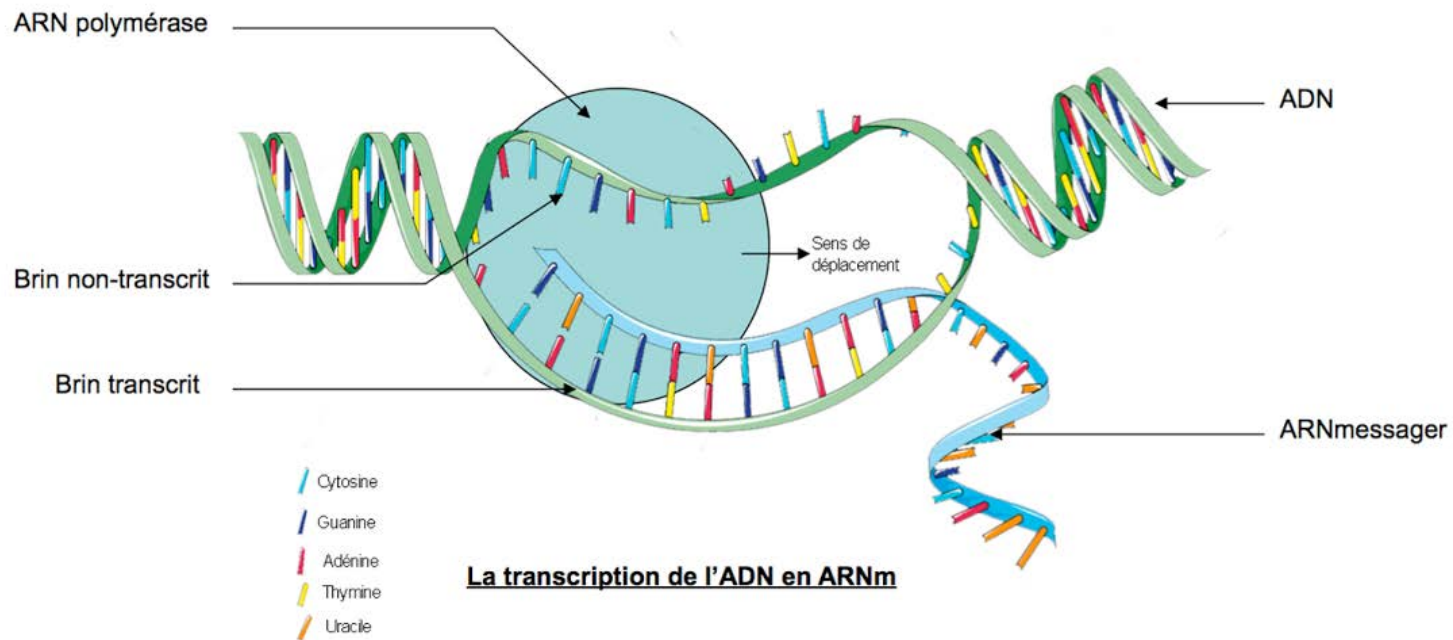
(a) **Formation d'un complexe fermé:**  
Attachement spécifique de l'holoenzyme au promoteur se fait avec l'aide du facteur sigma

(b) **Formation du complexe ouvert:** l'ARN polymérase déroule la double hélice de l'ADN sur 14-17 nucléotides

(c) **Formation de l'amorce:** mise en place du premier nucléotide (A ou G) puis addition de 4 à 5 nucléotides suivi du détachement du facteur sigma ( $\delta$ ), après la transcription des 4-5 premiers nucléotides.

# Transcription chez les procaryotes (2)

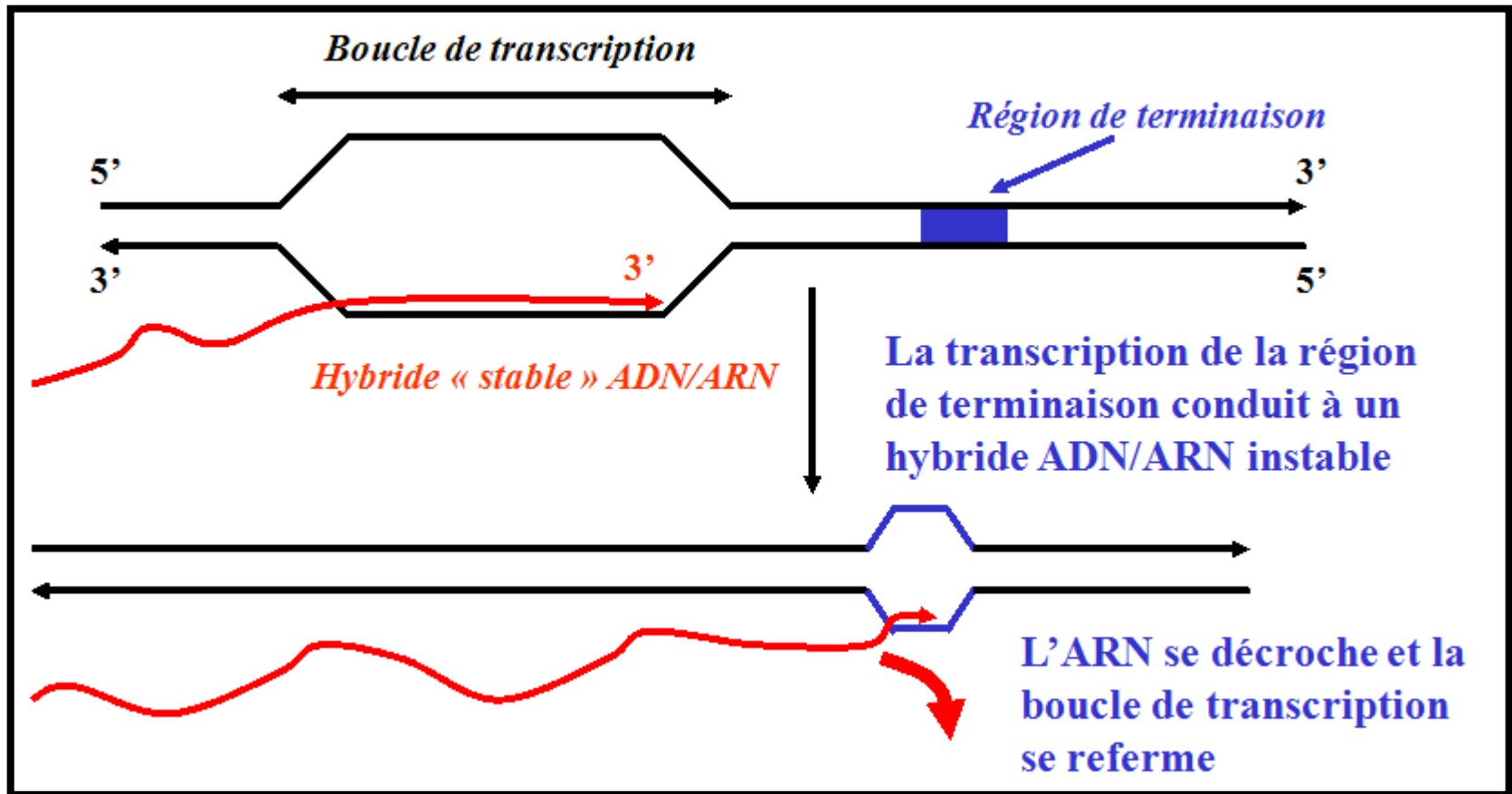
**B. Elongation:** Elle correspond à l'extension de l'amorce ARN formée lors du déplacement de la bulle de transcription le long de la molécule d'ADN. Elle est réalisée par l'enzyme  $\sigma$  (sans  $\delta$ )



- Déroulement de la double hélice au fur et à mesure que Pol avance sur le brin transcrit
- Addition de nucléotides à l'extrémité 3' du brin ARN naissant
- Formation d'une molécule hybride ADN-ARN sur une dizaine de paires de bases
- Les topo-isomérases suivent la bulle de transcription pour enlever les surenroulements.

# Transcription chez les procaryotes (3)

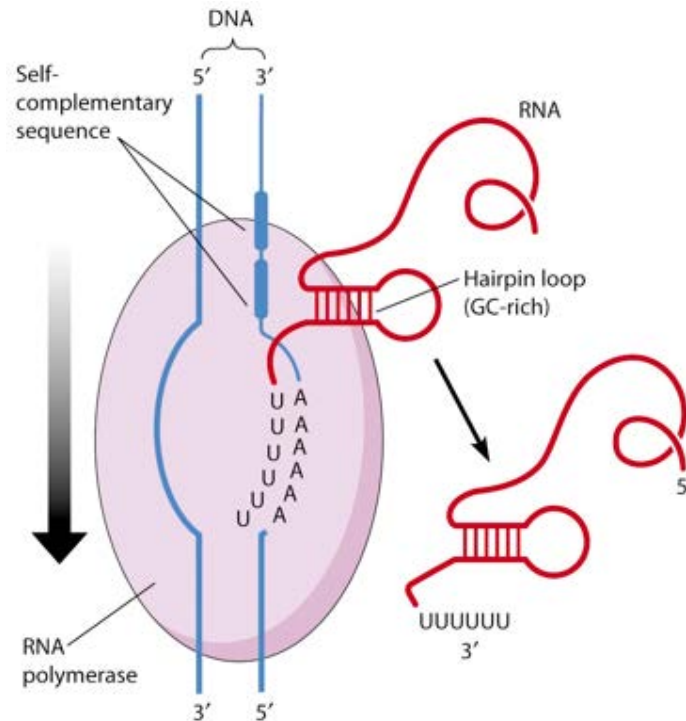
**C. La terminaison:** Elle se fait lorsque l'ARN polymérase arrive au niveau d'une séquence spécifique (région de terminaison) appelé **Termineur**. A ce niveau, la dissociation de l'ARN polymérase survient à l'extrémité 3' du brin d'ARN néosynthétisé.



# Terminaison Rho-indépendante

Structure en épingle à cheveux riche en G-C qui est suivie d'une séquence Poly-U d'environ 6 nucléotides permettant une dissociation plus facile de l'hybride ADN-ARN

ARN polymérase  
fait une pause  
et se dissocie du  
complexe



# Terminaison Rho-dépendante

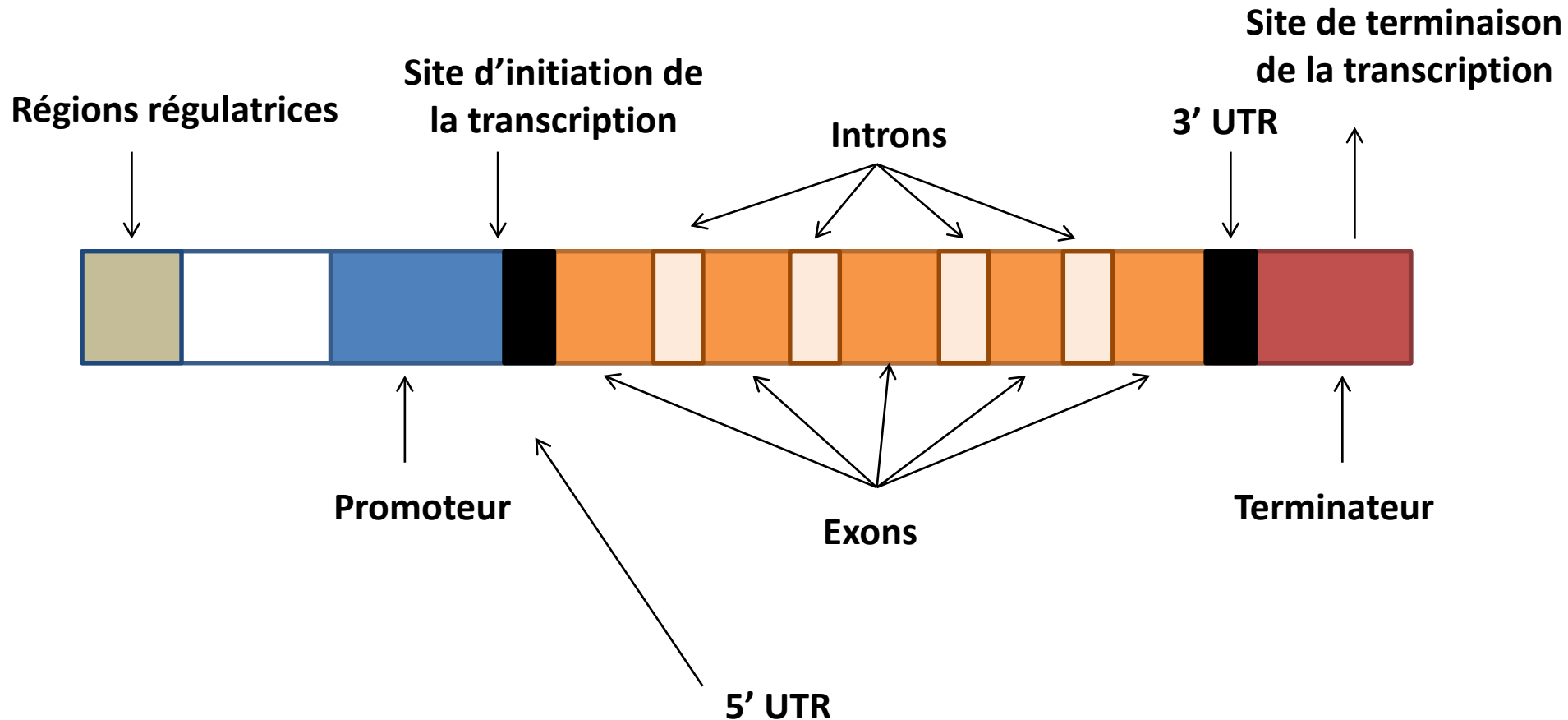
- Le facteur Rho est une **hélicase ATP-dépendante**
- Structure en épingle à cheveux est plus courte, n'est pas riche en G-C et n'est pas suivie d'une séquence Poly-U d'environ 6 nucléotides
- Rho a une affinité pour l'ARN en cours de synthèse
- Rho migre le long de l'ARN dans le sens 5' vers 3', localise le complexe, le déroule et libère l'ARN.

## Inhibiteurs de la Transcription

On peut pharmacologiquement arrêter la transcription en utilisant les antibiotiques:

- **Rifampicine** (inhibe l'initiation en bloquant l'action de la sous-unité  $\beta$  de ARN Pol)
- **Actinomycine** (inhibe l'élongation en s'intercalant entre les bases de l'ADN)

# Structure d'un gène eucaryote



**Exon:** région de la séquence d'un gène qui est transcrite et conservée dans la structure de l'ARN messager jusqu'à la traduction

**Intron:** région de la séquence d'un gène qui est transcrite mais n'est pas conservée dans la structure de l'ARN messager mature

**UTR (Untranslated région):** région non-traduite en protéines

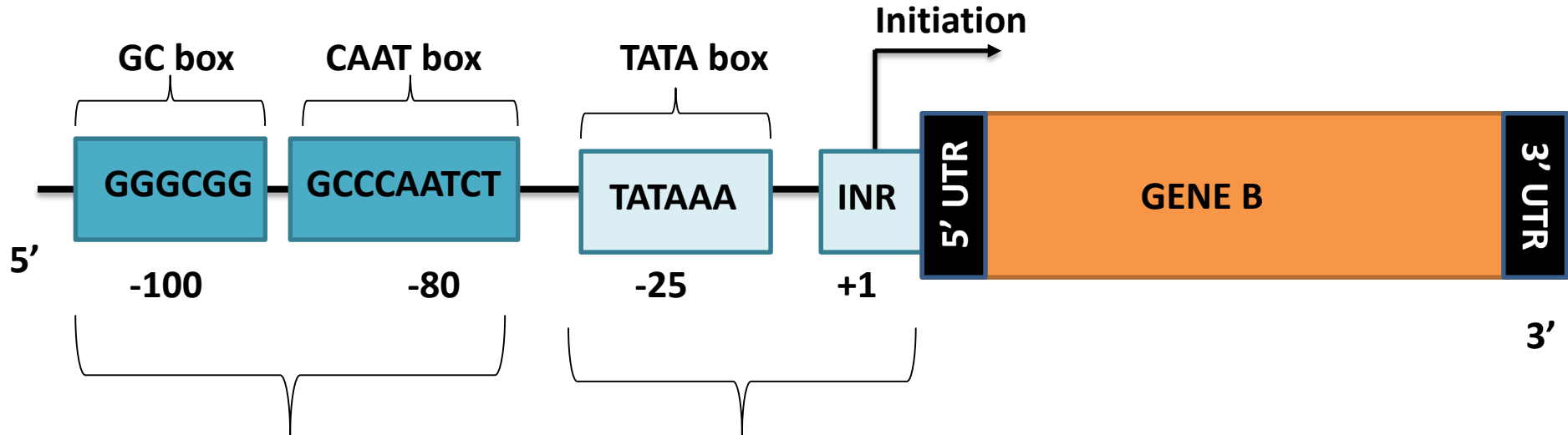
# Classes des gènes eucaryotes

Gènes Eucaryotes	Produits	Structure
<b>Classe I</b>	ARN ribosomiques (ARNr)	Séquences répétées en tandem et séparées par des espaces inter-géniques
<b>Classe II</b>	ARN messagers (ARNm) qui sont traduits en protéines	Exons (parties codantes) Introns (parties non-codantes)
<b>Classe III</b>	ARN de Transfer (ARNt) et autres ARNs de petites tailles	Séquence répétées en tandem (similarité avec class I)



# Structure du promoteur eucaryote

Le promoteur des eucaryotes contient plusieurs séquences consensus lui conférant une structure modulaire.



## Promoteur proximal

Site de fixation des facteurs de transcription et permettent ainsi la régulation de l'activité du promoteur minimum

## Promoteur minimum

Site de fixation de l'ARN polymérase II via des facteurs généraux de transcription

# Les ARN-polymérase eucaryotes

- Trois ARN-polymérase ont été identifiés chez les eucaryotes
- Elles sont classées selon trois critères:
  - Localisation dans le noyau cellulaire
  - Le type d'ARN synthétisé
  - Sensibilité aux inhibiteurs de transcription tel que l' $\alpha$ -amanitine.

Enzyme	Localisation	ARN synthétisé	Sensibilité aux inhibiteurs
ARN-polymérase I	Nucléole	ARN ribosomiques	Insensible (-)
ARN-polymérase II	Nucléoplasme	ARN messenger	Sensible (+)
ARN-polymérase III	Nucléoplasme	ARN de transfert et snARNs	Sensible (+)

# Transcription chez les eucaryotes: Généralités (1)

## A. Facteurs de transcriptions

ARN Pol II a elle seule ne peut pas initier la transcription. Elle a besoin de protéines qui peuvent se fixer sur le promoteur. Ces protéines sont des facteurs de transcription appelés TFII (Transcription Factor II: facteurs de transcription qui coopèrent avec Pol II)

Les **facteurs généraux de transcription** sont:

- **TFII D** constitue la protéine de liaison TBP (TATA binding protéine) qui se fixe sur la boîte TATA au niveau du promoteur minimum et les facteurs associés à TBP (TAFs)
- **TFII A** interagit avec la séquence en amont du TATA box
- **TFII B** interagit avec la séquence INR en aval du TATA box (site d'initiation)
- **TFII F** agit durant l'élongation
- **TFII H** possède une activité hélicase et de réparation de l'ADN. Elle a aussi une activité kinase qui permet de phosphoryler Pol II au niveau de sa queue CTD (C-terminal Domain)

# Transcription chez les eucaryotes: Généralités (2)

## C. Les régions cis-régulatrices

Les séquences consensus du promoteur sont des séquences cis-régulatrices qui sont reconnues par des protéines spécifiques appelées facteur Trans-régulateurs.

Ces séquences peuvent être de trois types:

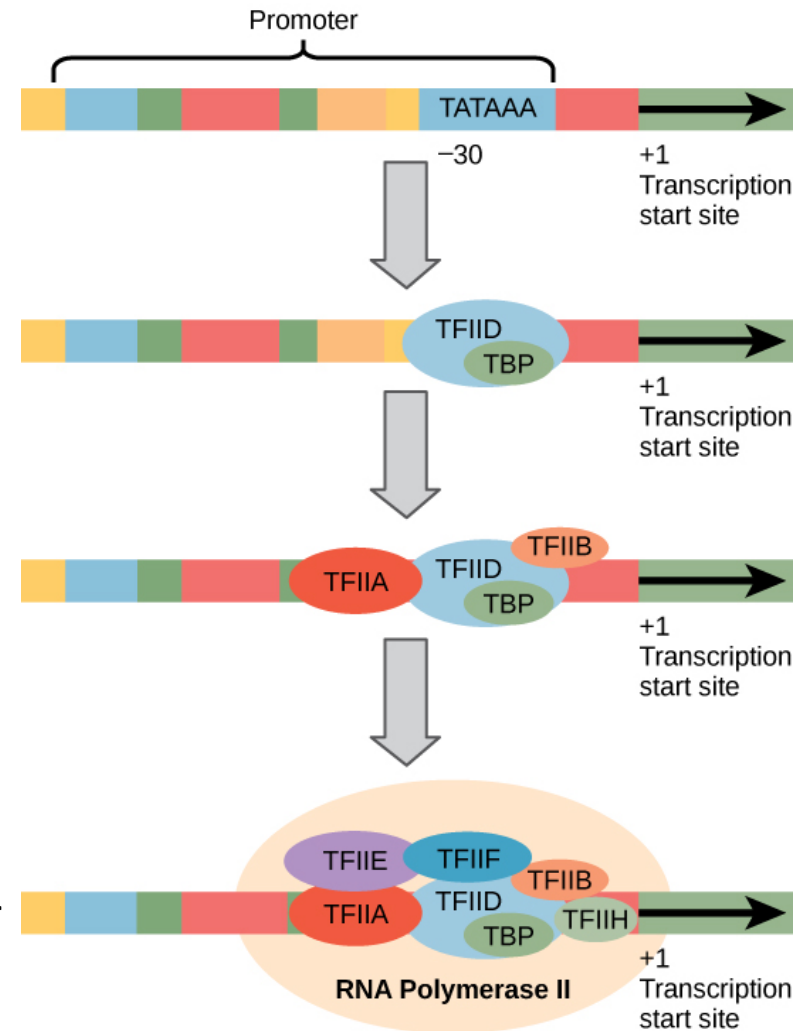
- Les enhancers:** sites de fixation des protéines **amplificatrices** de l'expression des gènes. Ils peuvent se trouver en amont et en aval du site d'initiation
- Les silencers:** sites de fixation des protéines **inhibitrices** de l'expression des gènes
- Les insulators:** séquences isolantes qui permettent d'isoler certaines régions du génome (par exemple ils bloquent l'interaction entre un promoteur et un enhancer)

NB: Les facteurs Trans-régulateurs agissent sur la vitesse de transcription des gènes. Leur fixation sur l'ADN dépend de leur structure qui peut être un motif hélice-boucle hélice, motif en doigt de zinc, et leucines zipper

# Transcription eucaryote: Initiation

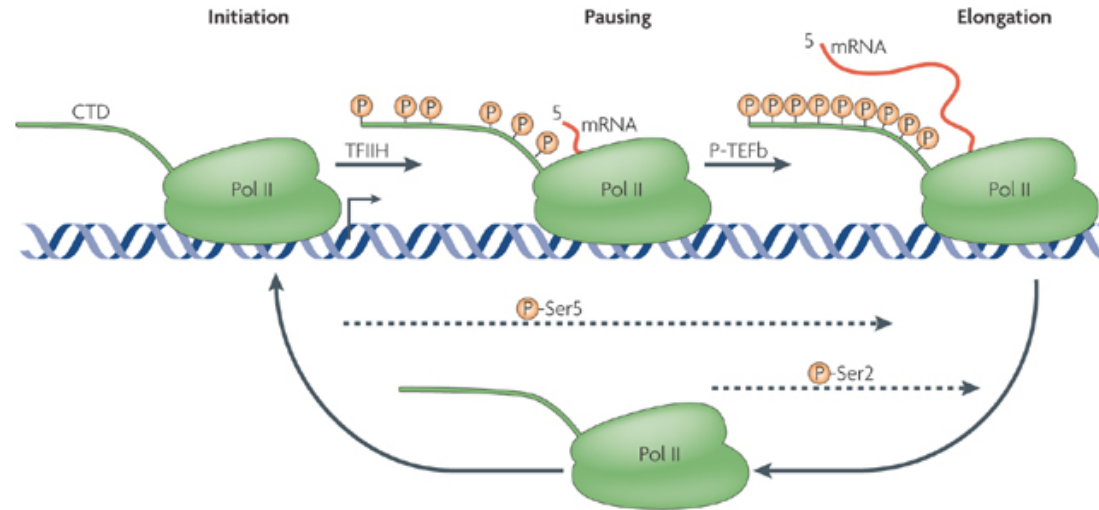
**Initiation:** correspond à l'assemblage des facteurs généraux de transcription sur le promoteur pour former le complexe d'initiation de la transcription

- TFIID se fixe sur TATA box grâce à TBP
- TFIIA et TFIIB se lient à TFIID pour stabiliser le complexe
- ARN Pol II et TFIIF se lient au complexe TFIID-TFIIA-TFIIB
- TFII E se fixe sur ce complexe
- TFII H (Hélicase ATP-dépendante) se lie et déroule le promoteur
- La transcription commence mais elle n'est pas optimale
- Les Trans-activateurs vont ensuite intervenir pour obtenir une transcription optimale
  - CTF se lie a la CAAT box
  - Sp1 se lie a la GC box



# Transcription eucaryote: Elongation

**Elongation:** correspond à l'addition séquentielle de nucléotides sur le brin d'ARN en cours de synthèse

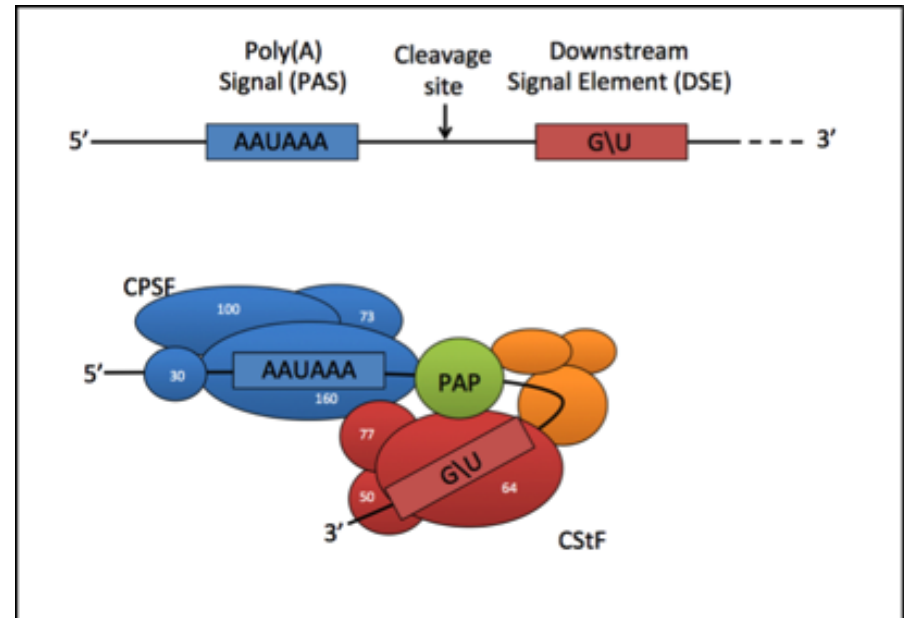


Nature Reviews | Molecular Cell Biology

- L'élongation est régulée par la **phosphorylation du Domain CTD** de l'ARN Pol II.
- CTD (queue) est riche en Serine et Thréonine qui sont des acides aminés pouvant être phosphorylés sur leur groupement OH.
- La phosphorylation du CTD est catalysée par la fonction kinase du TFIID H en présence d'ATP
- ARN Pol II se déplace sur le brin en ajoutant le nucléotide complémentaire au brin transcrit avec une vitesse de transcription d'environ 40 nucléotides par seconde.

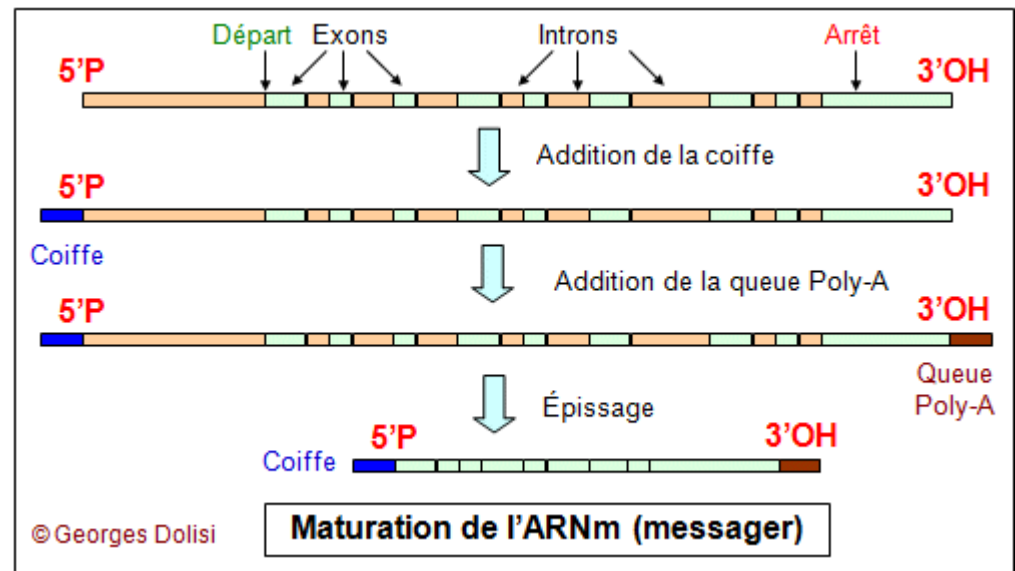
# Transcription eucaryote: Terminaison

- La terminaison est assurée par des séquences spécifiques
- La terminaison commence lorsque ARN Pol II arrive sur un site de polyadenylation 5'-AAUAAA-3'
- Le brin d'ARN néoformé est libéré par diverses facteurs
  - La protéine CPSF (cleavage and polyadenylation specificity factor protein) se fixe sur le site de polyadenylation
  - La protéine CSTF (cleavage stimulating factor) se fixe sur une séquence riche en GU et s'apparie avec CPSF pour former une structure en tige-boucle
  - CFI and CFII (cleavage factor protein).



# Maturation des transcrits primaires

- L'ARN formé lors de la transcription est un transcrit primaire appelé précurseur de l'ARN messenger (pré-ARNm ou transcrit primaire)
- Le pré-ARNm doit subir 3 étapes de modifications dans le noyau cellulaire pour être mature:
  - **Addition de la coiffe en 5'**
  - **Poly-adenilation en 3'**
  - **Excision des introns et épissage des exons**

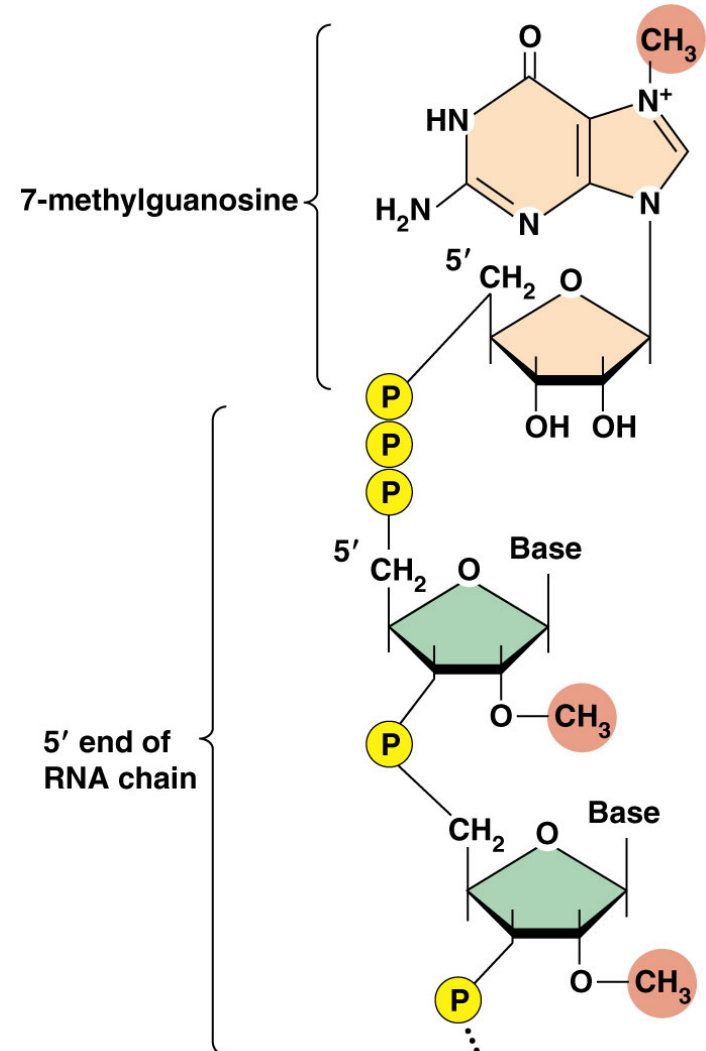




# Maturation des transcrits primaires (1)

## A. Addition de la coiffe en 5' (capping)

- Elle correspond à l'addition du **7-méthyl guanine (m7G)** à l'extrémité 5' du préARNm . Ce ajout se fait par une liaison 5'-5' triphosphate
- Le groupement m7G est composé de
  - 3 groupements phosphates
  - une molécule GTP qui est méthylée sur le carbone 7 (C7)
- Cette coiffe est ajoutée suite à l'action du Cap-binding Complex
- La coiffe est importante car elle:
  - Protège contre l'action des exonucléases
  - Sert de signature moléculaire lors de la translation par le ribosome

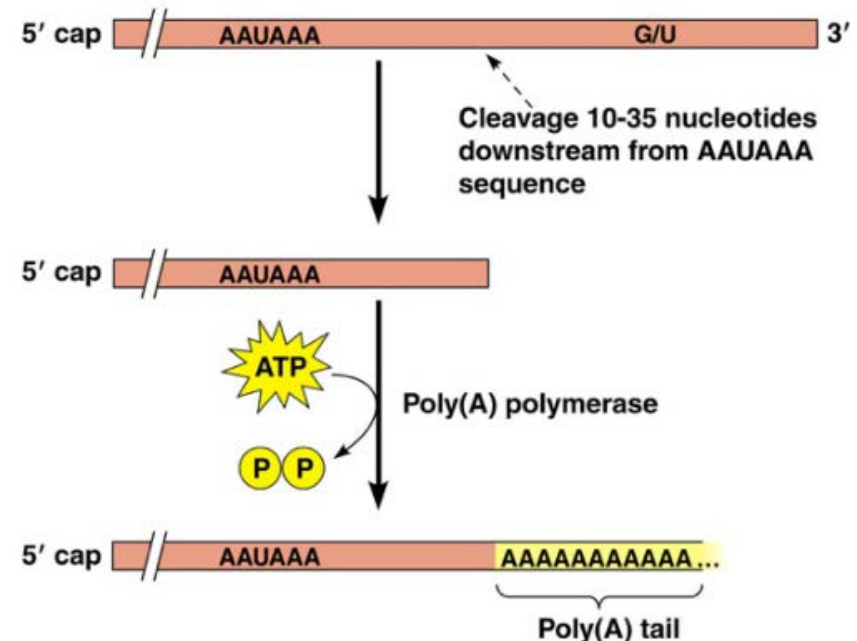


© 2012 Pearson Education, Inc.

# Maturation des transcrits primaires (2)

## B. Poly-adenilation 3'

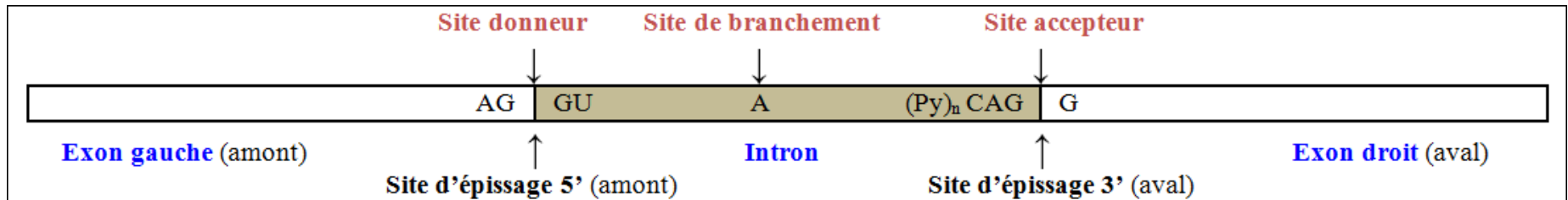
- Elle correspond à l'addition d'environ 200 nucléotides à l'extrémité 3' du préARNm .
- Ce ajout se fait sous l'action de la **Poly-A-polymérase (PAP)** qui reconnaît un autre signal de poly-adenilation différent de celui du CPSF (AAUAAA) sur le préARNm
- La modification est appelée la **queue Poly (A)**
- La queue Poly (A) est importante car elle:
  - Permet d'exporter l'ARNm du nucleus
  - Protège contre l'activité des exonucléases
  - Participe à la terminaison de la transcription
  - Participe à l'initiation et la régulation de la translation



# Maturation des transcrits primaires (3)

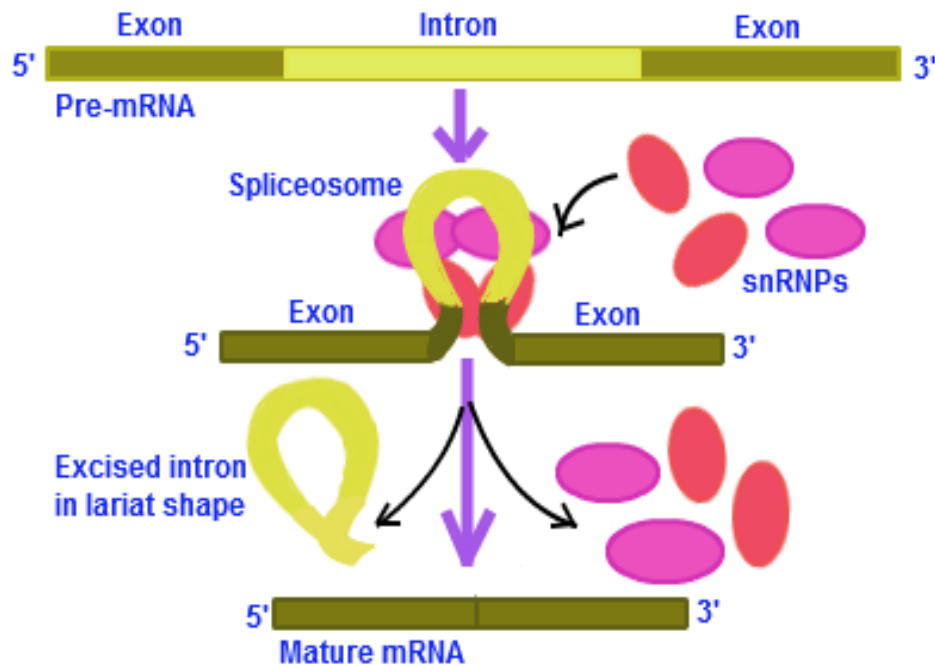
## C. Excision des introns et épissage des exons (splicing)

Elle correspond à l'élimination des introns de la structure primaire du pré-ARNm et à la liaison des exons les uns aux autres



- L' excision des introns se fait par la présence du:
  - **Site Donneur GU** à l'extrémité 5'
  - **Site Accepteur AG** à l'extrémité 3'
- Ces sites sont reconnues par les snRNPs (Small-Nuclear-Ribonucleo-protein-Particules) qui sont constitués de:
  - snRNA U1, U2, U3, U4, U5, U6
  - Protéines
- L'ensemble des snRNPS qui participent à la maturation du préARNm s'appelle le **Spliceosome**

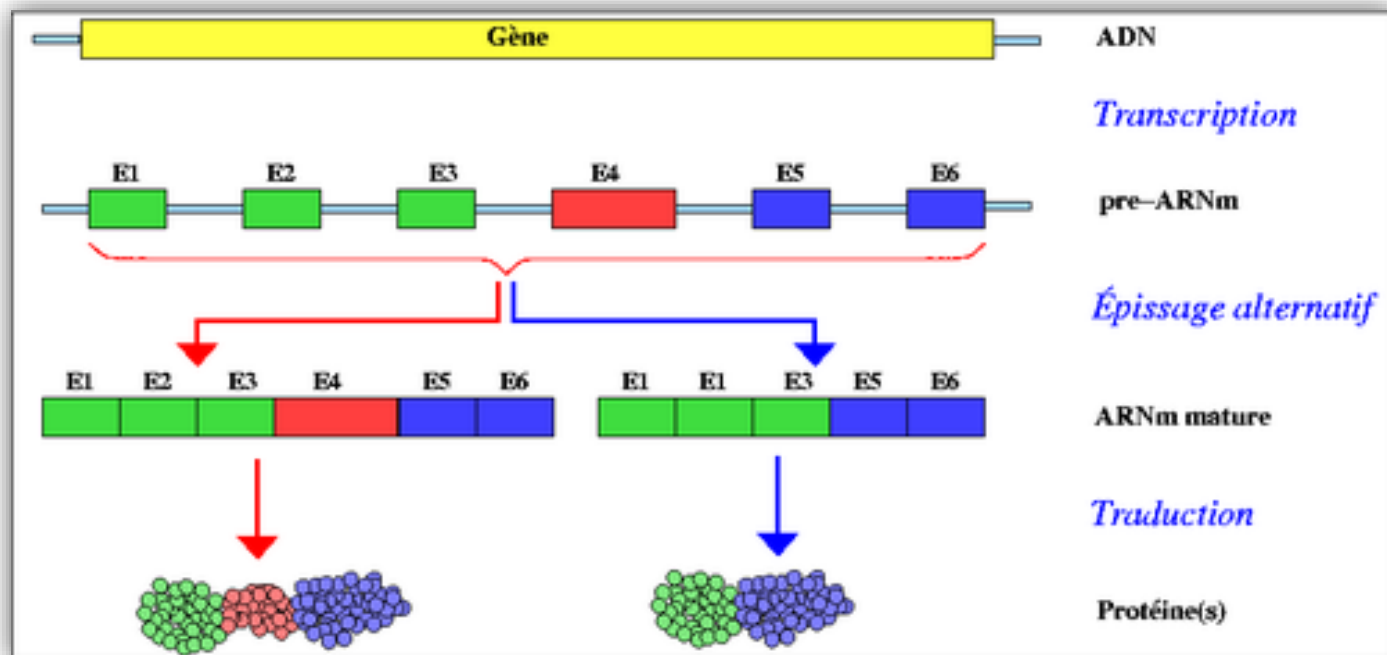
# L'excision des introns et l'épissage des exons en image



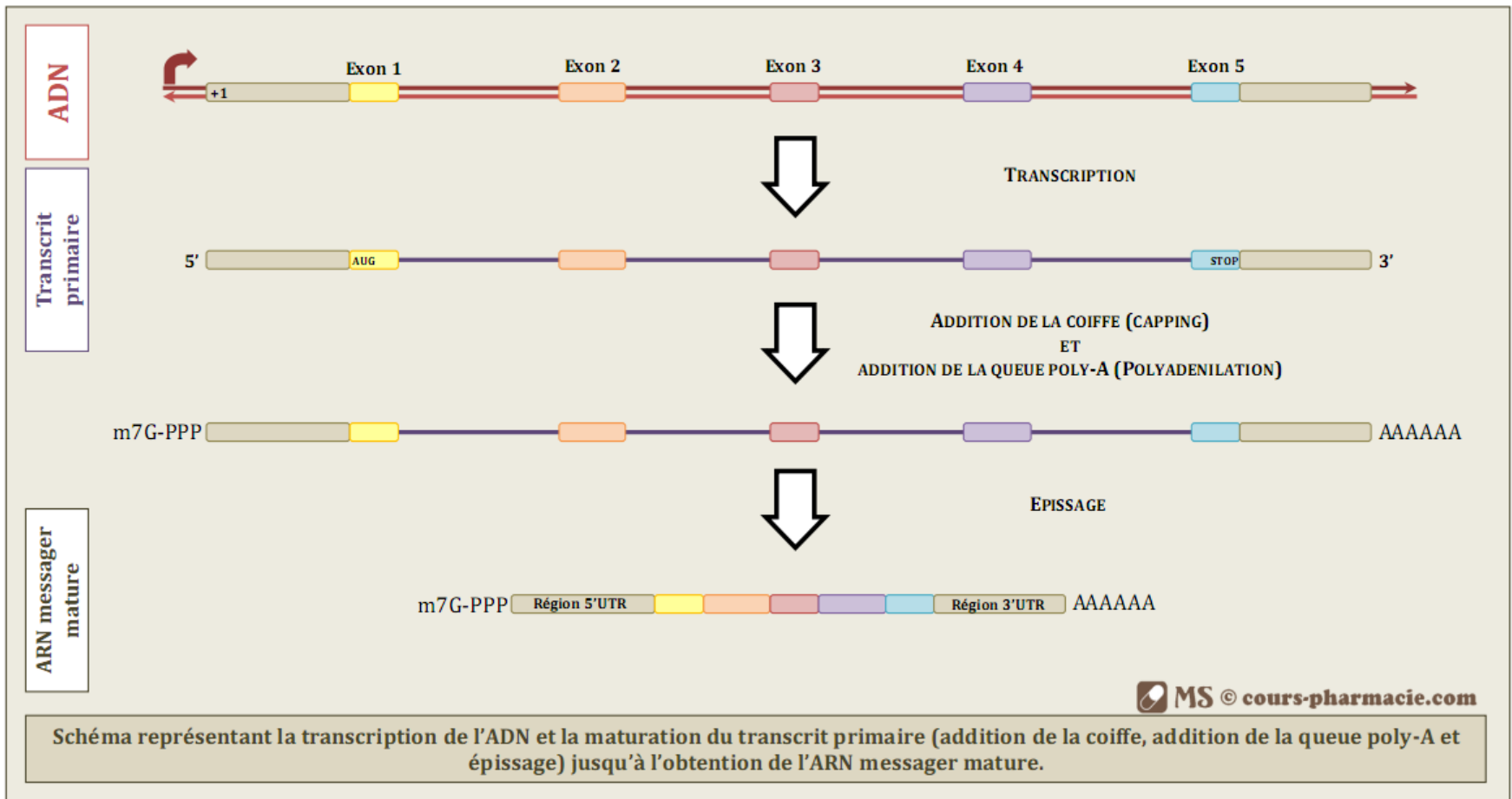
- Le **spliceosome** se forme sur les introns grâce à la reconnaissance de sites spécifique d'épissage par les snRNPs:
  - snRNP U1 se lie au site donneur
  - snRNP U2 se fixe sur le site de branchement et le site accepteur
- L'intron est excisé sous forme de lasso (lariat) qui sera ensuite dégradé
- Les deux exons sont reliés à la fin de la réaction.

# Épissage alternatif

- Permet d'avoir deux ou plusieurs ARNm matures à partir d'un seul transcrit primaire
- Il aboutit à la formation de protéines isoformes



# Transcription eucaryote: Résumé



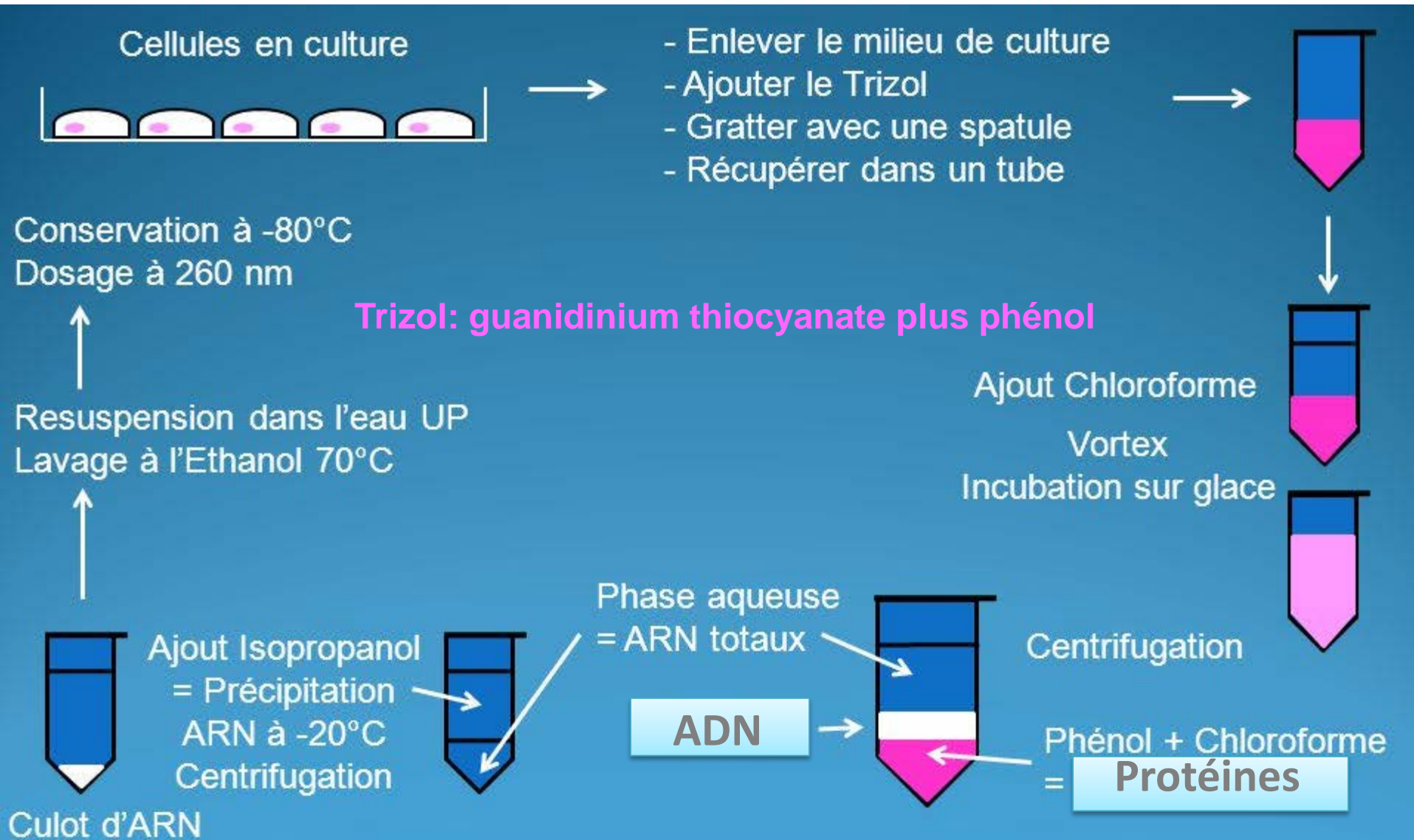
# Résumé

Procaryotes	Eucaryotes
Transcription dans le cytoplasme	Transcription dans le noyau
Transcrit Polycistronique (Plusieurs gènes donnent un ARNm)	Transcrit Monocistronique (un gène donne un ARNm primaire)
ARNm est mature	Pré-ARNm n'est pas mature
Un seul type d'ARN-Polymérase	Plusieurs ARN-polymérases
Transcription et Translation sont couplées	Transcription et Translation ne sont pas couplées

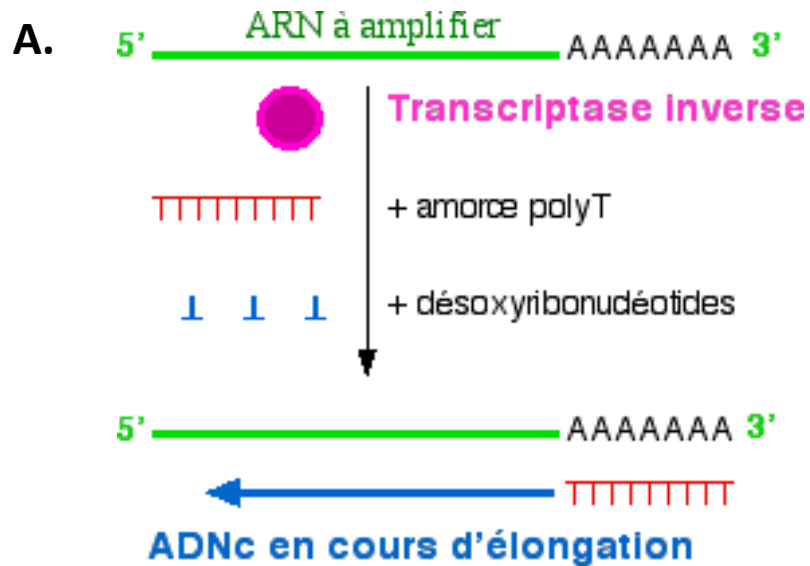
# Comment étudier la Transcription au Laboratoire?



# Extraction de l'ARN par le Trizol



# RT-PCR



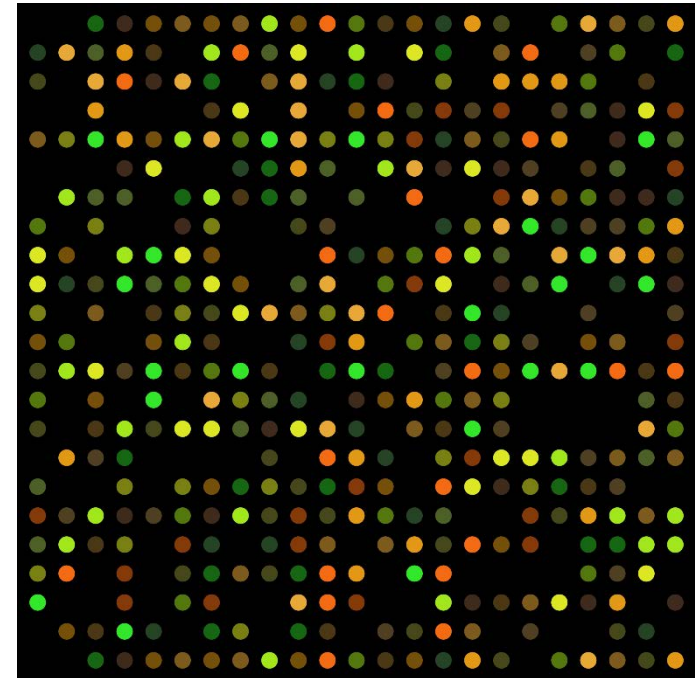
**B.** **ADNc obtenu par transcription Inverse**



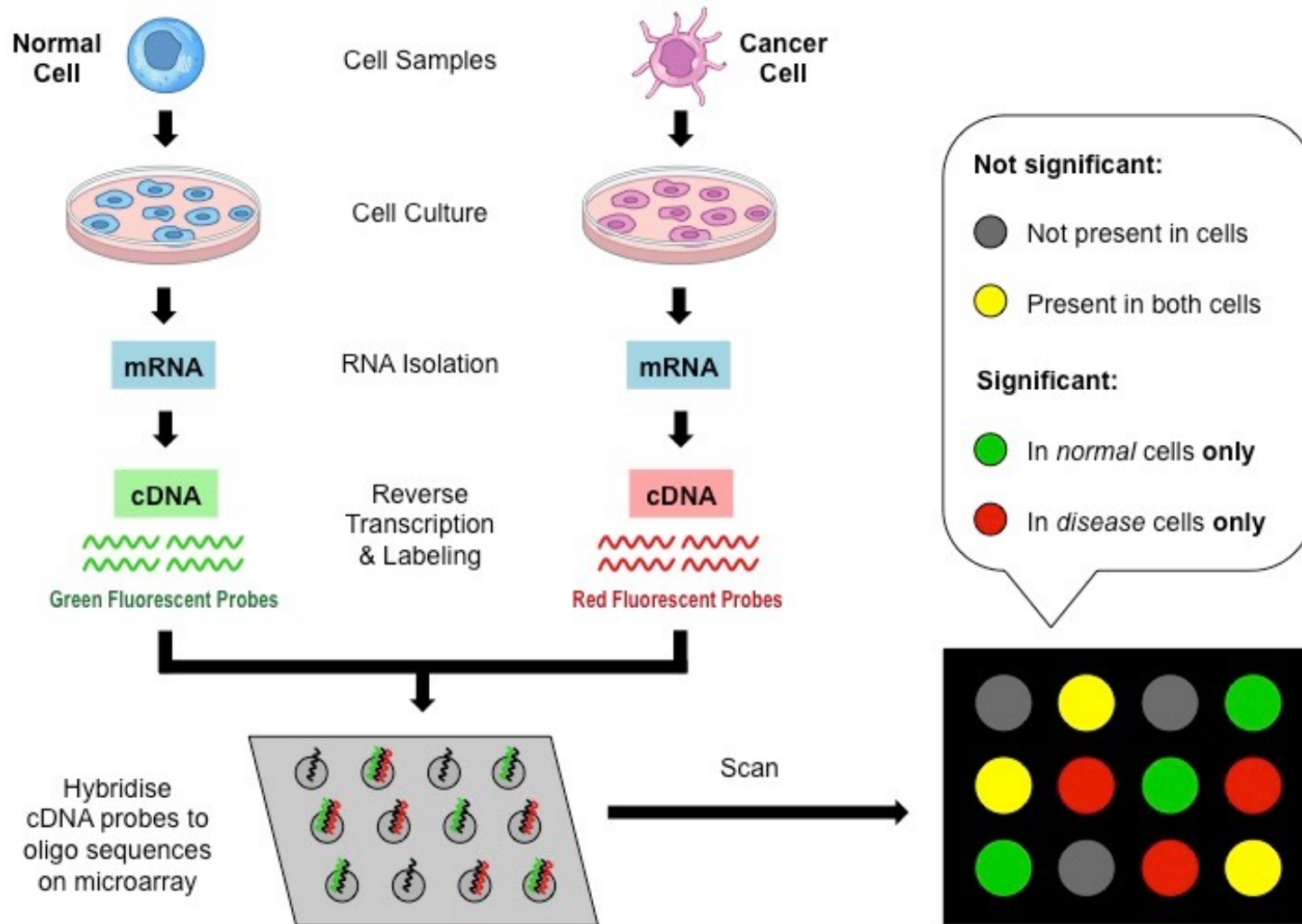
- La RT-PCR est une technique qui permet de faire une PCR (réaction en chaîne par polymérase: amplification) à partir d'un échantillon d'ARN.
- L'ARN est initialement rétrotranscrit grâce à une enzyme appelée transcriptase inverse, qui permet la synthèse de l'ADN complémentaire (ADNc; schema A). Ce dernier est ensuite utilisé pour réaliser une PCR (schema B).
- La RT-PCR peut être quantitative (qRT-PCR). Ceci permet de quantifier un type d'ARN initialement présent dans un échantillon (pour connaître le niveau d'expression d'un gène).
- La RT-PCR se fait dans des appareils de PCR particuliers qui permettent de visionner en temps réel la synthèse des fragments d'ADN.
- Pour cela, des molécules fluorescentes qui se fixent sur l'ADN sont utilisées. La fluorescence de l'échantillon augmente proportionnellement en fonction du nombre de molécules d'ADN formées.

# Micro-array: Biopuce à ADN

- Une biopuce à ADN est une collection de séquences d'ADN microscopiques (oligos) attachées à une surface solide (verre, plastique ou silicium)
- Ces séquences représentent des fractions d'une grande bibliothèque de gènes présents dans une cellule
- Si un gène est actif dans une cellule, alors l'ADNc (produit à partir de la transcription de l'ARNm) se liera à son oligo complémentaire
- Si l'ADNc a été marqué par fluorescence, alors l'oligo complémentaire peut être identifié (avec le gène qu'il représente)
- Si l'ADNc de cellules saines et malades sont marqués avec différents fluorophores, des comparaisons d'expression de gène peuvent être faites
- Les gènes qui ne sont actifs que dans un état malade ou normal intéresseront particulièrement les scientifiques.



# Micro-array en Image



# Questions?