

### III. MECANISMES MOLECULAIRES DE L'EXPRESSION GENETIQUE

L'étude de la transcription et de la traduction est l'occasion de présenter les divers types d'ARN produits.

La fidélité de la copie transcrite est comparée à celle de la réplication. L'existence de signaux d'initiation et de fin de transcription est présentée.

Le rôle des amino-acyl-ARNt-synthases est indiqué.

#### A-Transcription de l'ADN en ARN

##### - Mee d'un flux d'information du noyau vers le cytoplasme

- Localisation expérimentales de la synthèse protéique :

Expérience de Pallade en 1950 avec Leucine tritiée incorporée dans cellule de coupe épaisse de pancréas, lavage et prélèvement à différents temps pour réalisation de préparations microscopiques; autoradiographie et suivi de l'incorporation des acides aminés : dans le cytoplasme.

##### - Nécessité et identification d'un intermédiaire entre ADN et protéines : l'ARNm

- Recherche d'un intermédiaire capable de se déplacer du noyau au cytoplasme et capable de contenir une information.

- Rappel sur les différentes familles d'ARN r(80%, à 23, 16 et 5S chez Procaryotes; 28, 18, 5.8 et 5 S chez Eucaryotes) et t(15%, environ 60 différents) trouvés dans le noyau (coloration vert de méthyle pour l'ADN, pyronine colore en rose l'ARN), avec leur coefficient de sédimentation.

Mais :

- trop peu diversifiés pour représenter la diversité des protéines

- Rapport AT/CG différent entre ARNr et ADN, et trop stables .

- Les ARN sont néosynthétisés dans le noyau (mee par marquage spécifique de l'ARN à l'uridine tritiée)

- Le taux d'ARN est corrélé à l'activité de synthèse protéique

- Mee d'un nouvel ARN messenger spécifique d'une protéine

- Test de Monod et Jacob (1960) sur système inductible : utilisation du lactose par bactéries (utilisent normalement le glucose); marquage à l'uridine tritiée pour repérer ARN néosynthétisés

- Résultats : apparition d'une nouvelle enzyme (bêta galactosidase) et apparition d'un nouveau ARN ni r ni t mais m, inductible, à durée de vie courte (demi vie e 2 min chez E.coli)

- Vérification : on dénature ADN et ARN radioactif (à 85°C) puis renaturation aléatoire à 60°C / on hydrolyse tous les ARN libres par RNase et on recueille sur un filtre toutes les molécules d'ARN associées à ADN : protégées par association avec ADN. Donc ARN est bien en partie complémentaire de ADN

#### 1- Modalités de la transcription chez les Procaryotes

##### a) Observation et localisation de la transcription

ADN étalé montre des structures orthogonales portant des ribosomes

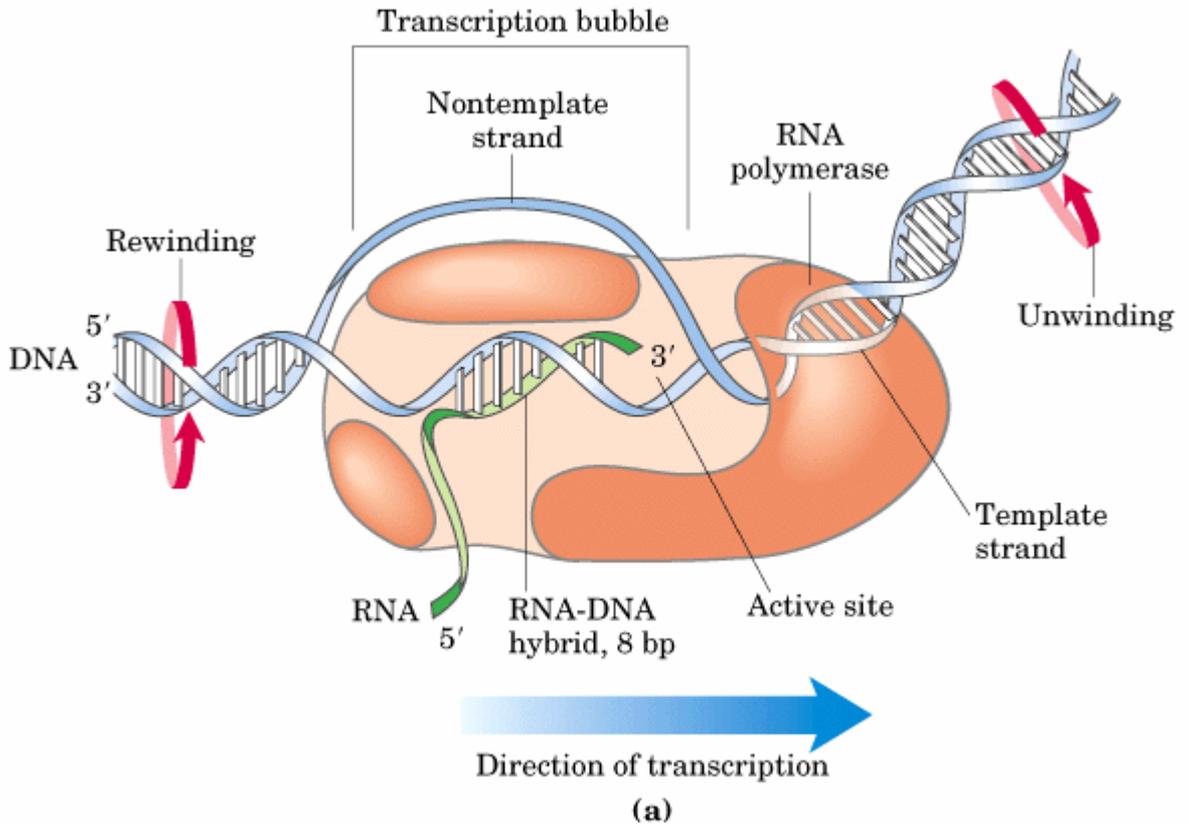
## b) Initiation

Fonction de l'architecture de la région promotrice et des propriétés de l'ARN polymérase ADN dépendante

### b1) Propriétés de l'ARN polymérase

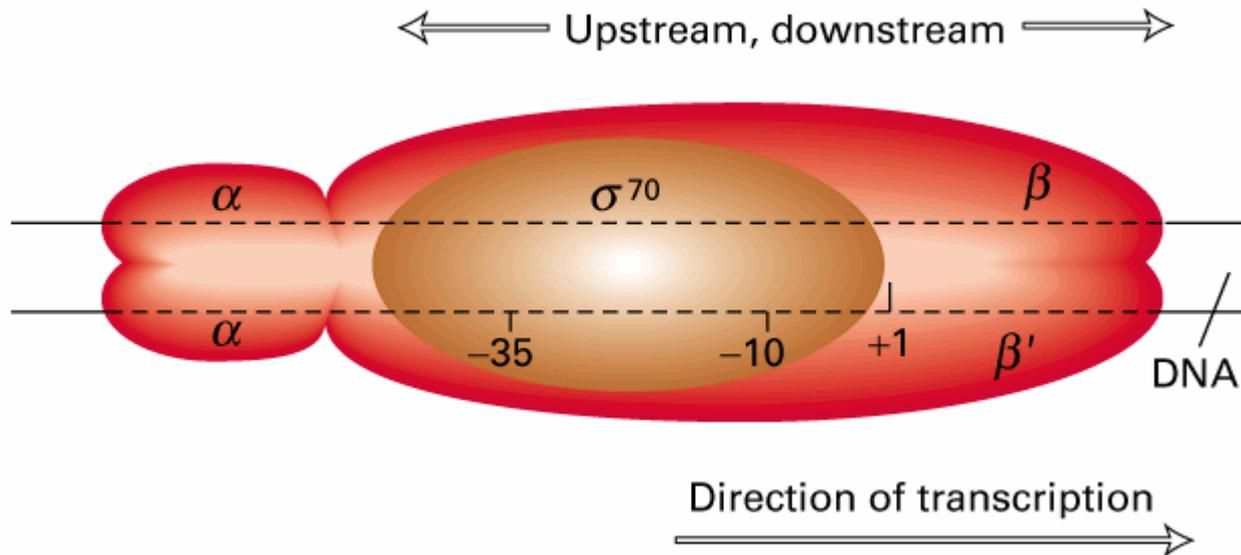
#### - Structure :

- structure quaternaire :  $\alpha_2$ ,  $\beta$ ,  $\beta'$ ,  $\sigma$
- 480 kDa;
- se fixent sur 5 tours d'hélice (15 nm)
- durée de demi-vie : 60 min.



#### - Rôle des sous-unités :

- $\alpha_2$  : permet association à  $\beta/\beta'$
- $\beta$  :
  - autocontrôle de la synthèse des sous-unités  $\beta$  et  $\beta'$
  - liaison aux ribonucléotides
  - liaison à l'ARN en cours de synthèse
  - permet la formation des liens phosphodiester (catalyse initiation et élongation); sensible aux inhibiteurs de la transcription : rifampycine pour initiation, streptolydine pour élongation)
  - reconnaissance des terminateurs
- $\beta'$  : liaison à l'ADN et au facteur  $\sigma$
- $\sigma$  : reconnaissance des promoteurs et donc précision de l'initiation (le core enzyme sans facteur  $\sigma$  se fixe sur l'ADN dans 70% des cas et peut synthétiser ARN mais à partir de n'importe où!).

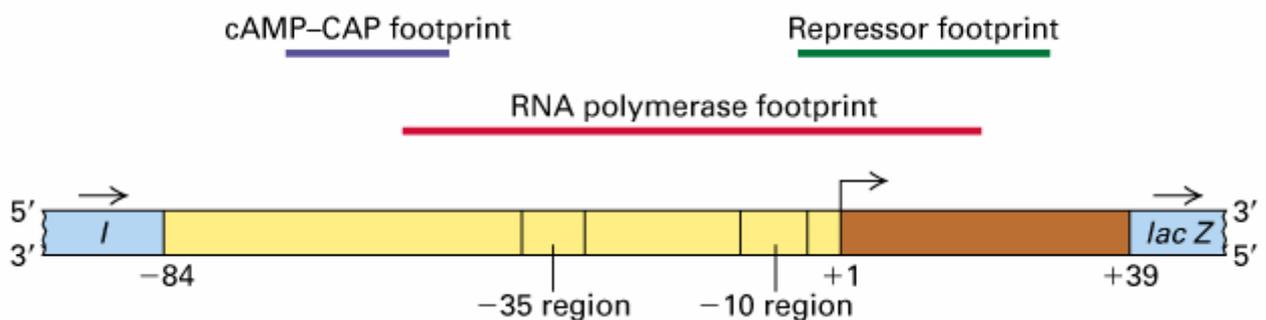


- Rôles du facteur  $\sigma$  :

- **Variabilité** du facteur  $\sigma$  :
  - $\sigma^{43}$  pour phase végétative
  - devient  $\sigma^{37}$  (présent mais non utilisé; on ignore les mécanismes du remplacement) qui permet de
  - coder pour  $\sigma^{29}$  qui code un nouveau jeu de gènes pour la phase de sporulation
- **Reconnaissance de séquences promotrices** par holoenzyme : en 1 sec. (on ne connaît pas les mécanismes de passage : glissement, diffusion?)
- **Complexe ouvert/fermé** : rupture des liaisons H sur 10 à 14 nt
- **Incorporation des 6 à 9 premiers ribonucléotides** diphosphate (coenzymes et substrat à la fois): avec établissement des liaisons phosphodiester
- La suite de la transcription est effectué par ARNpol sans facteur  $\sigma$ .

b2) Architecture de la région promotrice

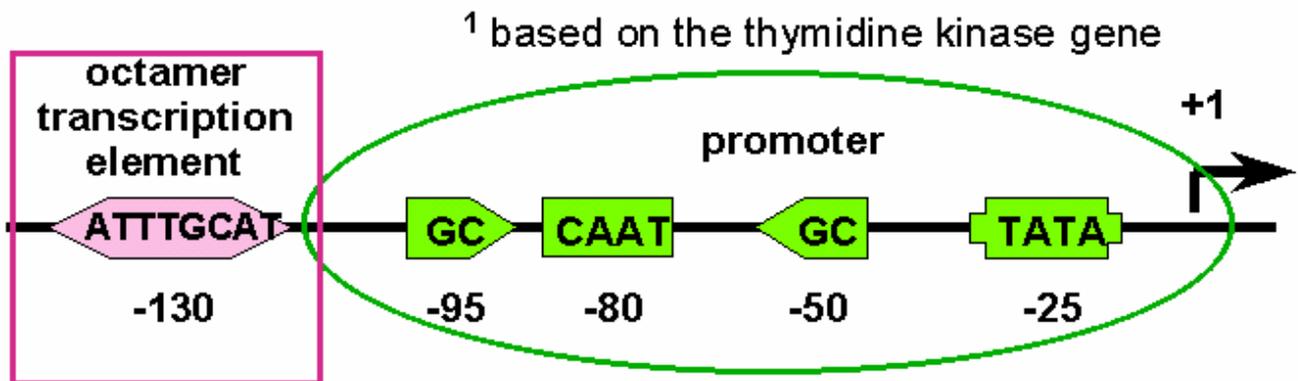
- Région **protégée** par ARNpol : de -60 à +20



- Régions d'**interactions fortes** : de -20 à +20 (contacts mée par méthylations de l'ADN protégé : tout ce qui est en contact ne peut être méthylé)
- **Boîtes** (régions très conservées) **ubiquistes** (se retrouvent sur très nombreux gènes et, associées, suffisantes pour promouvoir un nouveau gènes?) du promoteur **proximal** :
  - séquence CAT (avec A en position +1 du brin non matrice) et
  - **boîte TATA** (TAAAT environ en -10, correspond à zone d'ouverture du complexe) mais peut être plus en amont (-30)
  - séquence TTGACA en -35 qui permet reconnaissance par ARNpol
  - **boîte CAT** en -70

- **boîte CAC** en - 90
- il existe **boîtes spécialisées**, présentes sur un petit nombre de gène, mais à fonctions analogues : ex octamère ATGCAAAT en position -70 spécifique des immunoglobulines (le promoteur des Ig ne fonctionne que dans les cellules lymphoïdes, quelle que soit l'orientation de l'octamère par rapport au sens de transcription)
- il existe **éléments inductibles** : HSE, IRE, GRE, MRE
  - **HSE** (Heat Shock Element):
    - 14 pb.
    - Trouvé dans le promoteur minimal de choc thermique
    - Position variable par rapport à la boîte TATA d'un gène à l'autre
    - Active des gènes qui codent pour des protéines qui protègent la cellule contre les réactions dues à la chaleur (protéines chaperonnes, protéines impliquées dans l'immunité lors de la fièvre...)
    - ces éléments inductibles sont suffisants pour conférer à n'importe quel promoteur des caractères de choc thermique
  - **IRE** (Interferon Regulatory element): 10 nt induits par interféron (réaction immunitaire)
  - **GRE** : induits par glucocorticoïdes
  - **MRE** : répondent aux traces métalliques
- Promoteurs forts/faibles : fonction des séquences entre domaines conservés

## Promoter elements within a typical eukaryotic gene<sup>1</sup>



## Elements that promote transcription at the +1 site:

### TATA box (TATAAAA)

- located approximately 25-30 bp upstream of the +1 start site
- determines the exact start site (not in all promoters)
- binds the TATA binding protein (TBP) which is a subunit of TFIID

### Initiator: 5' -YYA<sup>(+1)</sup>N(T/A)YYY -3'

- Y = pyrimidine, (T/A) = T or A, A<sup>(+1)</sup> = A at transcription initiation site.

## Elements that can promote transcription at other sites (e.g. promoter-proximal elements) may contain:

### GC box (CCGCCC) (CpG islands)

- binds Sp1 (Specificity factor 1)
- (Used for housekeeping genes, no TATA box)

### CAAT box (GGCCAATCT)

- binds CTF (CAAT box transcription factor)

### Octamer (ATTGTCAT)

- binds OTF (Octamer transcription factor)

#### - il existe promoteurs sans boîte TATA :

##### - Boîtes C/G :

- séquences (GGGCGG) un certain nombre de fois
- appartiennent aux gènes de ménage : gènes exprimés dans toutes les cellules (contrairement aux gènes tissu-spécifiques)

##### - promoteur sans boîte C/G :

- impliqués dans le contrôle du développement embryonnaire

#### - il existe gènes possédant plusieurs promoteurs

- ex. gène de l'amylase : 2 Promoteurs P1 et P2 différents (chez la souris, les 2 sont à boîte TATA). Mécanisme du choix inconnu mais P1 utilisé dans les glandes salivaires et P2 dans le foie. Quel intérêt puisque produisent la même

protéine? Les 2 promoteurs peuvent fonctionner en même temps dans une même cellule

- **ex. Gène PBGD** (PorphobilinoGene Déaminase) : deux promoteurs dont un est un gène de ménage et l'autre un gène tissu spécifique qui code pour une enzyme participant à la synthèse de l'hème dans les cellules érythroïdes. On a observé que les deux promoteurs ne pouvaient pas fonctionner en même temps

- Importance de la topologie de l'ADN sur l'efficacité de la transcription :

- in-vitro : l'ADN super enroulé est transcrit plus vite que ADN déroulé
- in-vivo : mutants de topoisomérase II (gyrase : provoque les super enroulements ) ont transcription diminuée alors que mutants de topoisomérase I (catalyse la relaxation des supertours) ont transcription augmentée.

### c) Elongation

- Co-substrat et co-enzyme : nucléotide triphosphate
- Obligation d'un brin patron pour l'ARNpolymérase, mais pas de nécessité d'amorce (contrairement à ADNpol)
- Taux d'erreur :  $10^{-4}$  à  $10^{-5}$  très élevé (idem virus à ARN comme grippe ou SIDA qui sont très changeants). L'ARNpol ne sait pas se corriger car pas d'activité exonucléasique.
- Vitesse : 40 nt/s chez E.Coli; 200 nt/s Phage d'E.Coli

### d) Terminaison

- Architecture du signal de détachement de l'ARNpol : séquence palindromique G/C suivie d'une séquence quelconque riche en A/T (faible liaisons)
- Permet la réalisation d'une épingle à cheveux (tige/boucle) qui permet le détachement de l'enzyme.
- Certains terminateurs sont dépendant d'un facteur protéique  $\rho$  qui agit lorsque l'ARNpol fait une pause au niveau du terminateur
- Il existe des antiterminateurs dans boîtes NUT en amont de terminateur qui fixent prot NUS A et prot N (20 nt) ce qui modifie l'ARNpol qui ne reconnaîtra plus le terminateur quand elle passera dessus.

## 2- Modalités de la transcription chez les Eucaryotes

3 types d'ARNpol

### a) ARNpol I

- transcription des gènes codant pour ARNr<sub>45S</sub> dans le nucléole (où aura également lieu la maturation des ARNr et l'association avec sous-unités ribosomiales)

### b) ARNpol II

- transcription des gènes de structure + gènes ARNhn des RNP

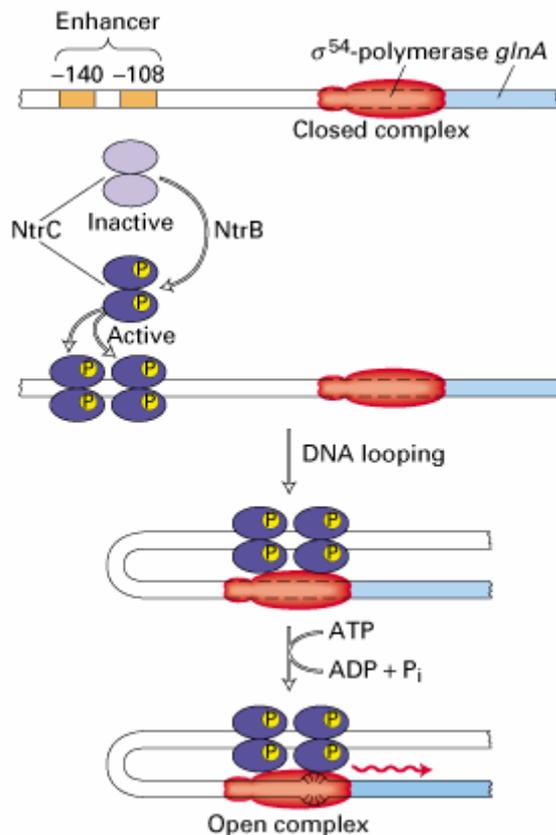
### c) ARNpol III

- transcription des gènes ARNt et ARNr<sub>5S</sub> et autres petits ARN



- système atténuateur

#### **d) Architecture des sites de régulation**



## **2- contrôle chez les eucaryotes**

### **a) Etat de la chromatine**

#### **a1) Mee d'une régulation**

- Expériences de Gurdon (cf. Concours communs 2000, épreuves B)

- Interprétation :

- Il existe une substance cytoplasmique régulée capable de rendre efficaces tous les gènes
- Le noyau contient toute l'info (sauf mitochondries et CP) sauf chez *Ascaris* où 50% de l'info est éliminée (sauf dans lignée germinale), surtout aux extrémités hétérochromatiques.
- Dans les cellules différenciées, une partie de l'info est non lue (mise sous silence, par méthylation notamment et ploïement de l'ADN en hélice Z par exemple)

#### **a2) Etat de la chromatine et activité transcriptionnelle des gènes**

- Mee de l'état de condensation par Action de la DNase I : ne concerne pas toutes les zones de l'ADN. Certaines zones sont protégées par état de condensation du nucléofilament en hétérochromatine dense
- Mee du rôle rétro-inhibiteur des protéines histones dans la transcription :
  - En système acellulaire : plus il y a d'histones, moins la transcription se fait
  - In vivo : modifie les contraintes de l'ADN

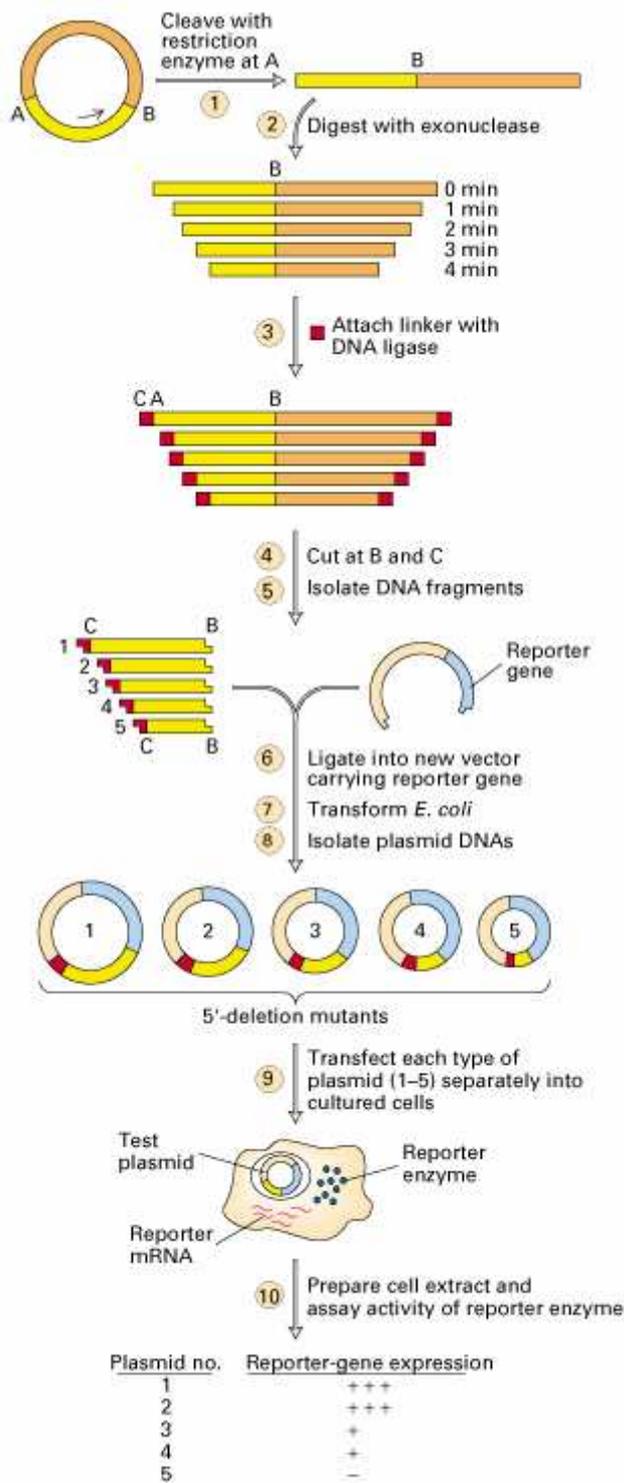
- Méthylation de l'ADN de la queue 3' de l'histone H3 : perturbe la reconnaissance et la transcription par ARNpol

**b) site d'initiation de la transcription et définition fonctionnelle des gènes chez les eucaryotes**

**b1) Position des sites de régulation**

- Upstream Regulatory Sequences (URS) :

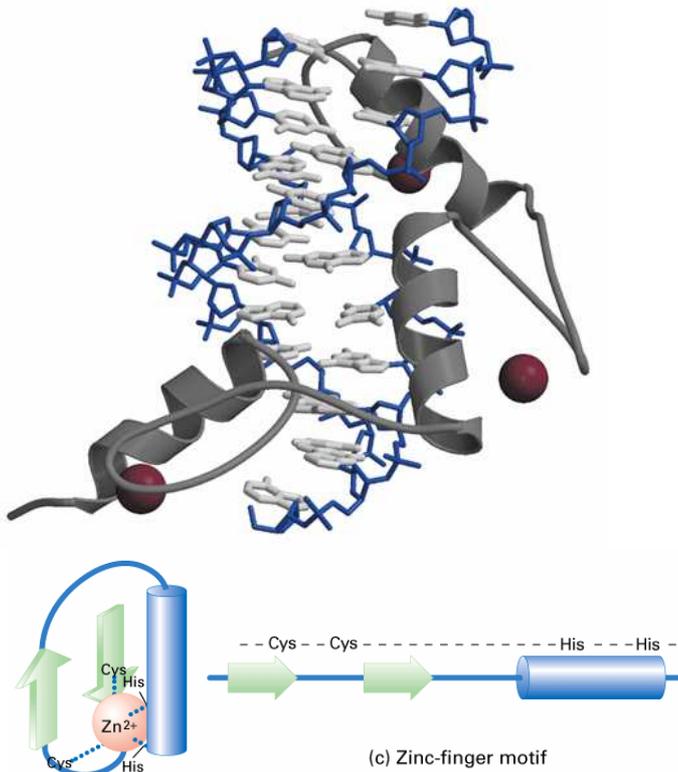
- Mee :

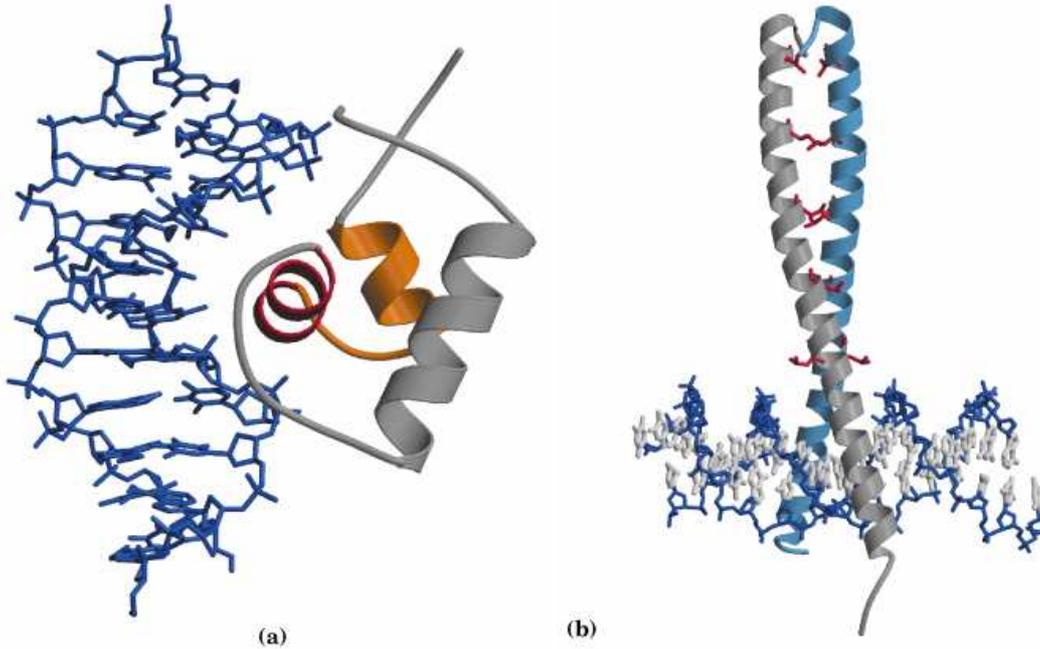


- Matériel et méthodes : Par transfert d'un gène (de globine par ex., associé à son site promoteur) dans une cellule hôte (de souris par exemple) par l'intermédiaire d'un vecteur (ex. virus SV40 modifié : non pathogène mais capable d'autoréplication grâce à son Ori)
- Résultats : On observe que certains gènes transférés s'expriment très peu. D'autre part, certains gènes de l'hôte voient leur niveau de transcription modifié.
- Interprétation : il manque la séquence URS dans le vecteur, en amont du promoteur. D'autre part, l'insertion du virus vecteur a pu se faire dans l'ADN hôte entre une séquence URS et son promoteur d'où la chute du taux de transcription.
- Action en CIS ; par interaction "en ligne" en amont du promoteur (modèle "scanning")
- Régulation positive ou négative
- Lieu de fixation de dimères (Trans)
- Interactions avec complexes de pré initiation.
- Enhancers et Silencers :
- Séquences activatrices ou inhibitrices, respectivement
- Séquences éloignées du promoteur
- Peuvent avoir action en Trans par reploiement de l'ADN et rapprochement du promoteur (modèle "looping")
- site de fixation de protéines régulatrices

## **b2) Facteurs protéiques de transcription (facteurs Trans)**

- Propriétés générales des facteurs de transcription : protéines possédant différents domaines
  - Domaine de liaison à l'ADN





homeodomaine (a) et leucine zipper (b)

- Domaine d'interaction avec les enzymes de transcription
- Eventuellement : domaine de réception d'un signal extérieur

### - 3 Types de facteurs de transcription

- **TATA Binding Protéines** : Facteurs non spécifiques, analogues au facteur  $\sigma$ , qui positionnent l'ARNpol et ouvrent l'hélice : Facteurs de transcription du site d'initiation des Eucaryotes : TFII D sur boîte TAT puis + TFIIA en amont puis ARNpolIII (sur complexe fermé à cheval sur position +1) + TFII B non directement associé à ADN + TFII E en aval de ARNpolIII.
- **Coactivateurs** : Protéines de liaison entre facteurs non spécifiques et activateurs ou inhibiteurs
- **Activateurs** ou inhibiteurs : liaison spécifique (sur séquences particulières) sur séquences d'ADN éloignée en amont ou en aval du promoteur

### - Structure du domaine de liaison à l'ADN des facteurs de transcription spécifiques

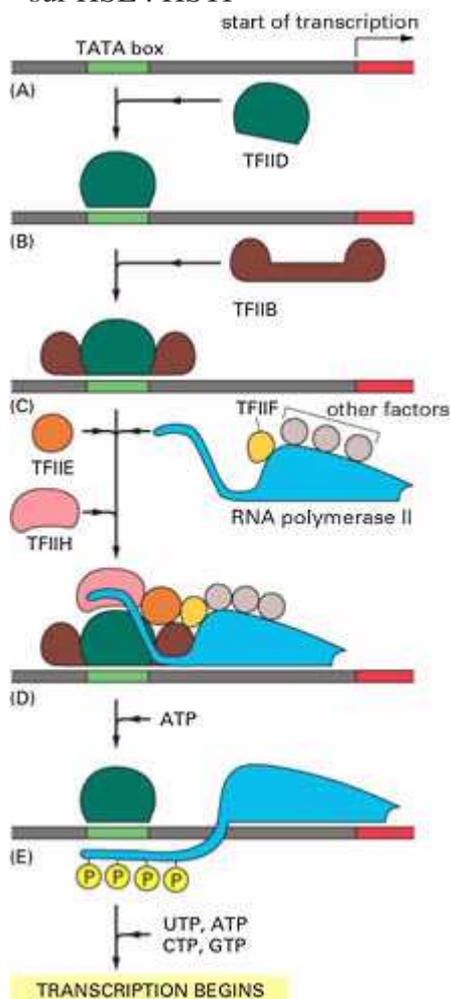
- homéobox de 60 aa soit 180 pb Il existe une grande unité structurale entre les différents homéodomaines : en effet, ils sont pratiquement tous constitués par 3 hélices alpha et parfois par une quatrième au niveau de l'extrémité C-terminale.

La reconnaissance spécifique d'une partie de l'ADN par l'homéodomaine est principalement assurée par 4 résidus en position 47, 50, 51 et 54 qui entrent en contact avec les bases du grand sillon de l'ADN. Quelques données montrent une interaction au niveau du petit sillon de résidus de l'extrémité N-terminale.

- Protéines à doigt de zinc : type cystéine/Histidine ou Cys/cys : ex. des récepteurs aux stéroïdes
- Protéines à hélices amphipathiques (une face hydrophile, une face hydrophobe) comme dans agrafes à leucine (leucine tous les 7 aa, tous les 2 tours alignés; les 2 hélices s'agrafent)

- **Liste non exhaustive des facteurs de transcription :**

- sur TATA : TFIIA, TFIIB, TFIID, TFIIE (Transcription factor de polymérase II, trouvé en poussant la purification de polymérase II, qui fonctionnait de moins en moins bien)
- sur CAT : CP1 (CAT prot n°1), CDP (CAT Displacement Prot n°1)
- sur C/G : SP1
- sur octamère : OTF1=oct1, OTF2=oct2
- sur HSE : HSTF



**b3) Gènes de commandes codant pour des protéines facteurs de transcription**

- 5% des gènes; jouent lors du développement, à divers moments

Defintion d'un gène homéotique

Un gène homéotique est un gène dont la mutation produit une homéose, c'est-à-dire l'apparition d'un organe bien formé, mais à un mauvais emplacement du corps. Un gène homéotique est donc défini par rapport au phénotype qu'il entraîne lorsqu'il est muté, et non par sa séquence nucléotidique. Les premières observations de phénotypes homéotiques remontent à 1894, bien avant la découverte des gènes et la définition du mot phénotype ! C'est William Bateson, qui en étudiant les variations intraspécifiques chez un coléoptère, observa des mutations homéotiques, notamment l'apparition de pattes à la place des antennes. Il fit d'ailleurs des observations similaires chez les végétaux, où les étamines pouvaient par

exemple être remplacées par des pétales. Bateson comprit qu'un segment de l'organisme possède toutes les potentialités, et qu'au cours du développement, un « choix » s'opère ; lorsque ce choix est faux, le segment ne possède pas les appendices ou organes attendus. On apprendra plus tard que ce sont des gènes qui déterminent ces choix.

### Les gènes homéotiques ne possèdent pas nécessairement la séquence « homéoboîte »

Mais la confusion est fréquente : en 1983, deux équipes de recherche distinctes, celle de Walter J. Gehring et celle de Thomas C. Kaufman, découvrent que tous les gènes homéotiques connus chez la drosophile possèdent une séquence commune de 180 paires de bases, qu'on nomme « homéoboîte » (en anglais homeobox). La tentation est grande d'appeler alors « gène homéotique » tout gène à homéoboîte... Cette définition s'applique effectivement chez la drosophile, et chez tous les animaux, mais il faut bien préciser deux points.

D'une part, la réciproque est fautive : les chercheurs se sont très vite rendus compte que tous les gènes à homéoboîte ne sont pas des gènes homéotiques, c'est-à-dire que leurs mutations ne donnent pas nécessairement un phénotype homéotique. C'est le cas des gènes bicoïd (bicoïd code une protéine déterminant la polarité antéro-postérieure de l'embryon de drosophile), engrailed, fushi tarazu (engrailed et fushi tarazu sont des gènes intervenant dans la segmentation de l'embryon de Drosophile), ou encore des gènes Pax (Parmi les gènes Pax des vertébrés, on trouve une famille multigénique qui possède une paired-box, séquence présente dans trois gènes de segmentation de la drosophile. Cette famille de gènes Pax interviendrait dans la formation du tube neural, du rein et de l'œil. Pax 6 est un gène de cette famille, homologue de eyeless et aniridia), qui sont des gènes du développement, mais pas des gènes homéotiques.

D'autre part, cette définition ne s'applique pas aux végétaux : les gènes entraînant des phénotypes homéotiques chez les végétaux ne possèdent pas l'homéoboîte.

### rôles des gènes homéotiques chez les différents organismes

Chez les arthropodes comme la drosophile, animal modèle des laboratoires, les gènes homéotiques déterminent l'identité de chaque segment de l'animal : ils dirigent ainsi l'apparition de pattes, d'antennes, ou de balanciers suivant les segments. Le phénotype « pattes à la place des antennes », décrit par Bateson, est dû chez la drosophile à une mutation dans le complexe de gènes homéotiques Antennapedia (*Drosophila melanogaster* possède deux grands complexes de gènes homéotiques : Antennapedia et Bithorax, qui correspondent à la scission d'un complexe ancestral HOX). Cette spécification des segments se fait suivant un axe antéro-postérieur, et il est fortement corrélé à l'ordre des gènes sur les chromosomes : c'est la règle de colinéarité. En parcourant l'ADN de 3' vers 5', on trouve des gènes dont les lieux d'action s'échelonnent de l'avant vers l'arrière de l'animal.

Chez les vertébrés, la majeure partie des gènes homéotiques intervient dans l'identité des différentes parties du corps, suivant l'axe antéro-postérieur, -mais pas dans celle de l'axe dorso-ventral. Une partie d'entre eux joue également un rôle dans l'édification des membres. Chez la souris par exemple, les gènes homéotiques, regroupés en 4 complexes appelés HOX A à HOX D, interviennent dans la régionalisation de la colonne vertébrale et du système nerveux central. Des mutations de ces gènes entraînent par exemple l'apparition de vertèbres cervicales à l'endroit des vertèbres dorsales ou encore des lombaires à la place de dorsales. Chez les végétaux, les gènes à MADSbox déterminent la mise en place des différentes verticilles de pièces florales : sépales, pétales, carpelle, étamines...

### mode d'action des gènes à homéoboîte

Les gènes à homéoboîte sont des régulateurs de transcription, activateurs ou inhibiteurs suivant les cas : ils sont en quelque sorte des gènes « maîtres », qui dirigent l'expression d'autres gènes. Prenons l'exemple des gènes homéotiques animaux, qui possèdent tous l'homéoboîte : cette séquence code un peptide de 60 acides aminés, appelé homéodomaine, qui possède dans sa structure secondaire trois hélices à lui permettant de se fixer à l'ADN. L'homéodomaine régule ainsi la transcription d'autres gènes. Chez les végétaux, la MADSbox joue un rôle similaire : elle code un peptide possédant un site de fixation à l'ADN.

- Expression séquentielle dans le temps et l'espace :

- ex : Axe antéro-postérieur : information de position due à gradient protéique lié à expression des gènes en cascade (Tête par gradient de protéines héritées de la mère)
- ex : gènes à homéodomains responsables de la différenciation des prosegments (métamères).

Chez la *Drosophile* on connaît 2 familles multigéniques : Antennapedia et Bithorax (épissage alternatif suivant le segment thoracique)

(b)

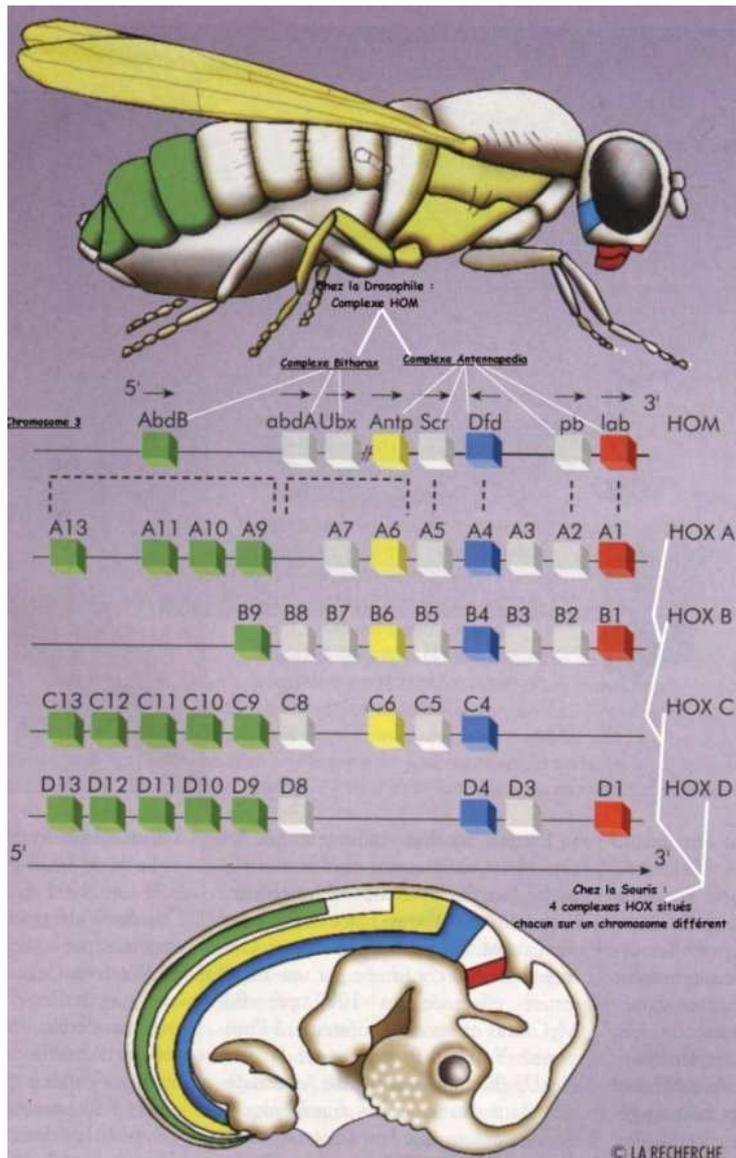


Normal



*Ubx* mutant

- Correspondance entre la place du gène sur le chromosome et l'ordre géographique d'expression : Antennapedia (Mutation qui fait apparaître une paire de pattes au niveau des antennes. Cette mutation fut observée pour la première fois en 1948. Cette mutation provoque l'expression du gène **antennapedia** au niveau de la tête ce qui provoque l'apparition de pattes à la place des antennes)



L'étude de la transmission héréditaire des mutations homéotiques a ensuite permis de **localiser les gènes homéotiques**. C'est ainsi que Lewis a localisé (1978) sur le chromosome 3 de la drosophile deux groupes de gènes homéotiques : le complexe Antennapediale et le complexe Bithorax . Le **complexe Antennapedia** contient les gènes homéotiques labial (lab), Antennapedia (Antp), Sex comb reduced (Scr), Deformed (Dfd) et proboscipedia (pb). Les gènes labial et deformed spécifient les segments céphaliques, tandis que Sex comb reduced et Antennapedia participent à l'établissement de l'identité des segments thoraciques. Le gène proboscipedia semble n'agir que dans les adultes ; toutefois, en son absence, les palpes labiaux de la bouche sont transformés en pattes. Le **complexe Bithorax** comporte trois gènes : le gène Ultrabithorax (ubx) qui est nécessaire à l'identité du troisième segment thoracique ; les gènes abdominal A (abdA) et abdominal B (abdB) qui sont responsables de l'identité segmentaire des segments abdominaux. Lewis a également découvert que l'ordre des gènes homéotiques du complexe Bithorax sur le chromosome 3 correspond à l'ordre dans lequel ils sont activés le long de l'axe antéro-postérieur du corps.

La structure de l'homéodomaine du gène Antennapedia et les 14 paires de bases de l'ADN avec lesquelles il interagit n'a été que récemment observée. Cet homéodomaine interagit avec 14 nucléotides de l'ADN "GAAAGCCATTAGAG". Les 6 nucléotides en gras étant ceux qui

interagissent directement avec les 4 acides aminés de l'homéodomaine (Ile47, Gln50, Asn51, Met54). De plus, Arg5 vient se placer dans le petit sillon. L'Asn51 est strictement conservé dans l'ensemble des homéodomaines.

### Conservation des gènes homéotiques au sein du monde vivant

Lorsqu'en 1983, Gehring et Kaufman découvrent « l'homéoboîte », on crible des banques connues de séquences de gènes pour savoir si cette séquence est présente chez d'autres organismes que la drosophile : on trouve d'abord des gènes à homéoboîte chez la souris, chez l'homme, puis chez tous les vertébrés. La plupart de ces gènes s'avèrent homologues chez les arthropodes et chez les vertébrés. De plus, ils sont regroupés en complexes dans les deux cas. Or les paléontologues nous disent que les arthropodes et les cordés se sont séparés il y a 540 millions d'années. L'ancêtre commun à ces deux embranchements possédait donc déjà des gènes à homéoboîte, qui se sont probablement dupliqués au sein d'un complexe, qui se scindera plus tard en deux sous-complexes chez la drosophile, et se dupliquera pour en donner 4 chez les vertébrés, créant de la diversité.

Les gènes à homéoboîte forment une famille multigénique, les gènes homéotiques en sont une sous-famille, et les gènes Hox des animaux constituent encore un sous-ensemble homogène de la famille des gènes homéotiques.

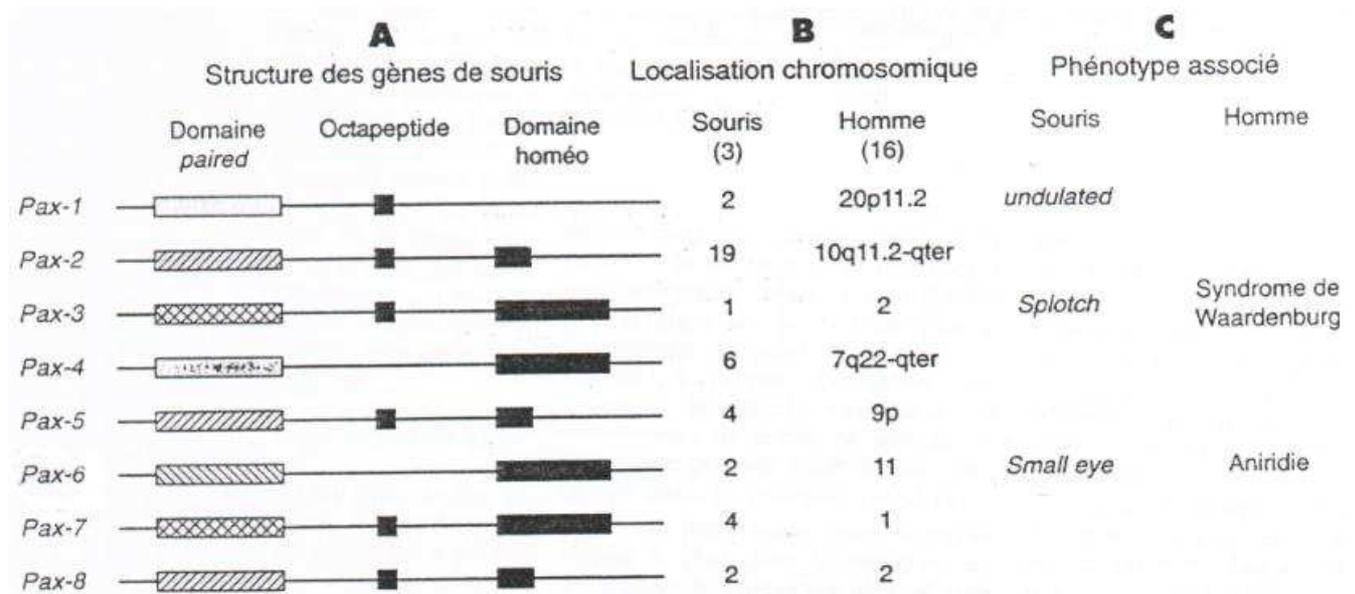
Les gènes à homéoboîte sont très anciens, puisqu'on les trouve chez les plantes, chez les champignons.... Il en va de même pour les gènes à MADSbox découverts chez les végétaux, et qui sont présents chez presque tous les Eucaryotes, des unicellulaires aux vertébrés (le facteur immunitaire SRF, chez les vertébrés, possède une MADSbox) Des gènes très anciens, par duplication et remaniements, ont été « recrutés » au cours de l'évolution pour accomplir de nouvelles fonctions.

### Une catégorie de gènes du développement les gènes Pax

*D'après Medecine/Science, 1993, V 9, N°1.*

Les gènes Pax constituent, parmi les gènes du développement, une catégorie distincte de celle des gènes Hox.

Du point de vue de leur **structure**, les gènes Pax possèdent généralement, en plus de l'homéobox déjà caractéristique des gènes Hox, une séquence dénommée *paired-box* ainsi qu'un motif octapeptide intercalé entre l'homéobox et la *paired-box*. La figure ci-dessous précise la structure des gènes Pax, leur localisation et les phénotypes associés.



Du point de vue **fonctionnel**, la plupart des gènes Hox s'expriment uniquement pendant l'embryogénèse, lors de la formation du système nerveux. Cette idée est corroborée par l'effet des mutations des gènes Pax sur les phénotypes des individus. Ainsi une mutation du gène Pax-3 chez la souris entraîne, à l'état homozygote, une malformation de l'axe neural ; une mutation du gène Pax-6 entraîne, à l'état homozygote, une absence totale d'oeil et de cavité nasale. Le tableau ci-dessous renseigne sur le profil d'expression des différents gènes Pax chez la souris.

Gènes	Période et lieu d'expression chez l'embryon	Expression chez l'adulte
Pax-1	j10-17 : Mésoderme somitique: sclérotome	Non
Pax-2	j10-18 : Rein en développement. Tube nerveux, rhombencéphale, vésicule otique, vésicule optique	Non
Pax-3	j8.5-16 : Certaines régions du cerveau, partie dorsale du tube nerveux, cellules de la crête neurale, dermomyotome	Non
Pax-4	Non déterminé	Non déterminé
Pax-5	j10-14 : Mésencéphale, tube nerveux, foie foetal	Rate, ganglions lymphatiques, sang
Pax-6	j8-18 : Cerveau antérieur et postérieur, hypophyse, épithélium olfactif, oeil, tube nerveux (zone ventrale)	Non
Pax-7	j8-17 : Groupe de cellules dans tout le cerveau, puis limité au mésencéphale, tube neural (zone dorsale)	Non
Pax-8	À partir de j11.5 : Expression transitoire dans tout le tube nerveux et le myélocéphale(j11.5 à j12.5). Thyroïde et rein en développement	Thyroïde, rein

Le **mécanisme d'action** des gènes Pax a été partiellement élucidé. Le motif protéique codé par le domaine paired se lie à l'ADN et agit comme un facteur de transcription. Dans le cas du gène Pax-6 chez la souris, ce mécanisme a pu être précisé : le gène Pax-6 est un gène "maître", situé au sommet d'un édifice génétique d'où il régule un ensemble de gènes "secondaires" influençant eux-mêmes l'activité d'autres gènes-cibles d'un niveau inférieur dans la hiérarchie. De cette activation en cascade, impliquant 2000 à 3000 gènes, résulte la mise en place d'une structure parfaite comme l'oeil.

### **c) Gènes en mosaïque et maturation des ARNm :**

Les mécanismes de l'excision et de l'épissage ne sont pas au programme. Les processus de maturation sont étudiés dans le cadre de la compartimentation des cellules eucaryotes. La maturation post-traductionnelle et l'adressage des protéines sont présentés. On se limite aux mécanismes simplifiés de translocation co-traductionnelle dans le réticulum et aux seules mentions et localisations des maturations et modifications par établissement des ponts disulfures, glycosylations, clivages protéiques et phosphorylations.

#### *a) Mee d'une maturation des ARNm par épissage Chambon, 1987*

- Matériel d'étude : travaille sur gènes de l'ovalbumine de poule qui s'exprime dans paroi de l'oviducte

- Méthodes d'étude :

- Réalisation de profils de restriction : découpe fragments d'ADN ou d'ADNc (obtenue par transcription inverse d'ARNm) par enzymes de restriction puis séparation par chromatographie.

On obtient profils différents selon que l'on part d'ADN ou d'ADNc.

- Hybridation moléculaire entre ADN et ARNm

- Résultats : Pas d'identité entre ADN et ARNm

- Interprétation :

- Notion de gène morcelé ou gène mosaïque

- 25% sous forme d'exons (transcrits) et 75% sous forme d'introns (non exprimés : portion intercalées qui "ne veulent rien dire", de longueur variable entre 60 nt et 30 kb)

- remise en question de la notion de gène : longueur du gène = 4 fois longueur de ARNm

#### *b) Etapes de la maturation des ARN<sub>prémessagers</sub> par excision/épissage*

- Excision-épissage alternatif

- NB: il existe des gènes sans introns en général : histones (mais coupure quand même par SNURP U7) , interféron, petits ARN nucléaires

- Mécanismes de l'épissage pas au programme

- Un même gène peut donner des ARNm différents par épissage alternatif (diversité des anticorps, par exemple). Intérêt : condense stockage de l'info mais cas rares

- après l'excision, il peut rester un intron qui donnera lieu à un épissage alternatif : ex. du gène chaîne légère de myosine qui possède deux sites TATA permettant d'associer les introns 1456789 (présentes dans tous les muscles) ou 12356789 (muscles rapides de l'adulte seulement)

- Il existe transplicing : épissage d'exons appartenant à deux ARN<sub>prém</sub> différents

- Chapeutage en 5' par coiffe de 7méthylguanosine

- Lien diphosphate 5'-5' par estérification de sa fonction libre ribose en 5'
- Guanosine porte -CH<sub>3</sub> en 5'
- Intérêts de la coiffe :
  - Stabilise ARNm
  - Protège ARNm des ribonucléases
  - Reconnaissance par petite sous-unité ribosomiale

#### - Méthylations

- sur Adénosine (0,1%) en position 6
- Pendant ou juste après ajout de coiffe
- Intérêts : reconnaissance par sous-unité ribosomiale

#### - Polyadénylation (queue polyA)

- Rappel : le détachement de l'ARNpol s'est fait sur séquence riche en A et T
- La séquence correspondante riche en A et U est coupée en un point spécifique (facteurs tissus spécifiques et intervention SNURP U11?) et remplacée par Polymérase qui met polyA sur 40 à 40 résidus
- Intérêts : La durée de vie de l'ARNm est fonction de la longueur de la queue polyA : de plusieurs jours à moins d'une seconde (si moins de 40 A), suivant le degré de protection par Poly A Binding Protéines (PABP) de 600 aa avec 4 domaines de liaison Nterminal à l'ARN très conservés de la levure à l'homme.
- Il existe séquence non traduite entre codon STOP et queue Poly A (10 à 1000 nt) : responsable de la grande instabilité des ARNm, surtout chez les oncogènes (messagers de facteurs de croissance)
- Il existe des gènes non polyadénylés comme histones, interféron, petits ARN nucléaires

#### - "editing"

- Connus chez gènes mitochondriaux (cytochrome oxydase avec ajouts de U), chez le blé (modification provoquant échange Tryptophane / Arginine) et chez trypanosomas (Insertions et délétions de U plus échange) d'où clonage très difficile (fabrication d'ADN à partir de l'ARNm)

#### c) Maturation des ARN<sub>pré</sub>r et ARN<sub>pré</sub>t

- ARN<sub>pré</sub>t : perte de la **tête**, modification de la **queue**, excision d'un **intron** dans la boucle anticodon
- ARN<sub>pré</sub>r : épissage alternatif d'un grand transcrit primaire de 41 Svedbergs en 2 transcrits primaires plus courts de 20 S et 32 S qui donneront respectivement ARNr 18S, et ARNr 5.8 S + ARNr 28S