

## Biologie moléculaire, Chapitre 1 :

# Les génomes.

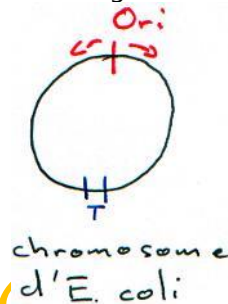
## I\ Généralités.

On peut dire que l'on a du vivant quand une cellule peut donner deux cellules identiques de manière autonome.

L'information est l'ADN ; il doit être dupliqué ainsi que l'enveloppe (membrane + phospholipides). Doivent être synthétisées, des protéines pour la synthèse de l'enveloppe (phospholipides) et pour la synthèse de l'ADN (informations).

Le plus petit génome procaryote connu est composé de  $2.10^6$  pb. 260 gènes devraient être nécessaires pour le métabolisme de base.

Chez E. Coli, on trouve  $4,2.10^6$  pb, soit, 4800 gènes.



Lorsque la distance entre deux gènes est entre 4 et 5 acides nucléiques, on a une structure compacte

On peut aussi avoir des gènes polycistroniques.

## A\ Les levures (eucaryotes).

Les levures ont un noyau et des organites. Les mitochondries vont fournir de l'ATP grâce à la respiration. Le génome mitochondrial est composé de  $2.10^7$  pb, soit, 6200 gènes. La levure a des chromosomes. Les procaryotes n'ont qu'une molécule d'ADN alors que les eucaryotes en ont plusieurs (chromosomes).

## B\ La drosophile.

Le génome de la drosophile est constitué de  $10^8$  pb. Une cellule va donner des millions de cellules mais différenciées.

## C\ Le génome humain.

Le génome humain est composé par  $3.10^9$  pb, soit, 50 000 gènes.

## D\ Les batraciens.

Chez les batraciens, on trouve  $10^{11}$  paires de bases.

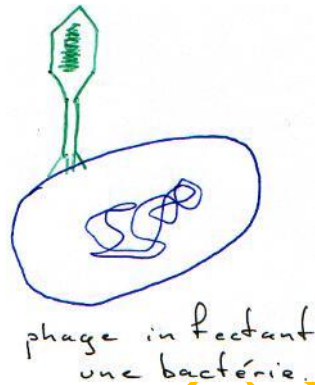
## E\ Les parasites.

Parmi les parasites, on trouve : les mycoplasmes ( $2.10^5$  pb) et les phages (ex : le phage  $\lambda$  de 50 000 pb).

### 1\ Les phages.

Une fois le phage dans la bactérie, selon l'espèce, il a différents devenir :

- L'ADN phagique peut s'insérer dans l'ADN bactérien.
- L'ADN va servir à produire du phage : à un faible niveau, il n'entraîne pas la mort de la bactérie ; à un haut niveau, il entraîne la mort de la bactérie.



### 2\ Les virus.

On trouve des virus à ADN ou à ARN. Les virus à ADN ont une taille variable comprise entre 4000 pb et  $10^5$  pb. Parmi les virus à ARN (rétrovirus), on trouve le HIV ( $14.10^3$  nt) et le virus de la grippe.

### 3\ Les plasmides.

Les plasmides sont de petites molécules d'ADN autonome qui vivent dans les bactéries. En général, on utilise des plasmides de l'ordre de 4000 pb. Il existe toutefois de plus gros plasmides qui peuvent aller jusqu'à  $10^5$  pb. Ils possèdent tous une séquence « Ori » qui leur permet une réplication autonome.

On a deux familles de plasmides en fonction « d'Ori » :

- de 3 à 5 copies par bactérie,
- plus de 100 copies par bactérie.

Les résistances aux antibiotiques sont dues aux plasmides (gènes de résistance).

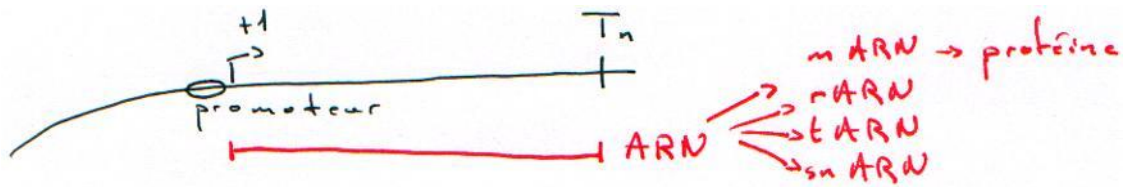
## F\ Temps de réplication et chromosome.

Chez E. Coli, le temps de génération est de 20 minutes (doublement de population). Pour une cellule eucaryote, une mitose a une durée comprise entre 20 et 24 heures.

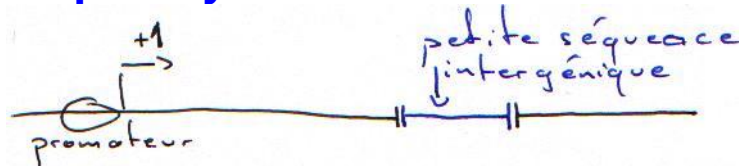
La réplication du chromosome bactérien est bidirectionnelle avec une seule origine de réplication alors que chez les eucaryotes, il y a plusieurs « Ori » le long des chromosomes.

Chez les eucaryotes, centromère et télomères sont des séquences particulières. Les télomères sont reliés au vieillissement : ils forment un système de protection.

## II\ Structure d'un gène.



## A\ Chez les procaryotes.

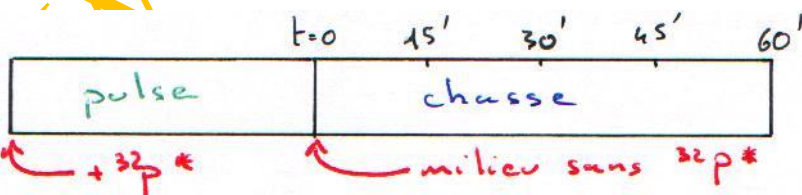
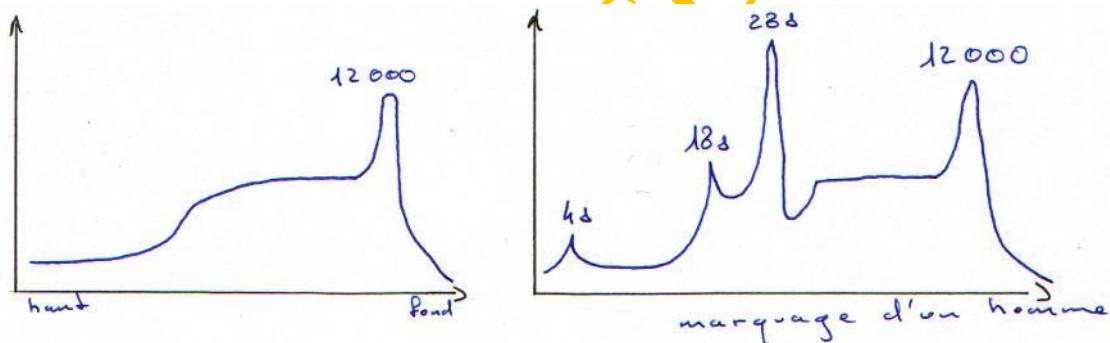


Dans le cas d'un opéron, on a un même promoteur qui code pour plusieurs gènes collés

## B\ Chez les eucaryotes.

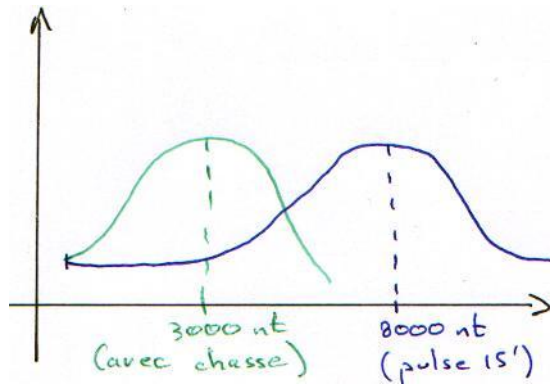
On a essayé de directement travailler sur l'ARN. Pour visualiser cet ARN, on réalise un marquage in vivo par un précurseur radioactif (NTP). Le précurseur peut être du  $^{32}\text{PO}_4^{2-}$  qui est incorporé dans les bases ou du  $^3\text{H}$  ou  $^{14}\text{C}$  qui marque les bases.

Un marquage court (15 minutes) est appelé un « pulse ». Le résultat du marquage est exprimé en fonction des fractions de gradient.

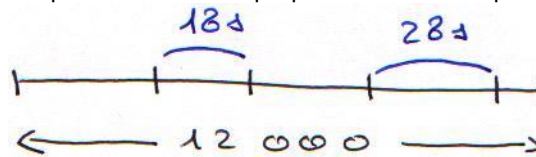


L'ARN synthétisé dans le noyau (grande taille) donne un ARN plus petit dans le cytoplasme (à cause de la maturation ou « processing »).

On isole les messagers sur une colonne d'oligodT cellulose (liaison des ARNm par leur queue polyA).



L'ARNr synthétisé pour le pré-rRNA est coupé par des RNAses spécialisées.

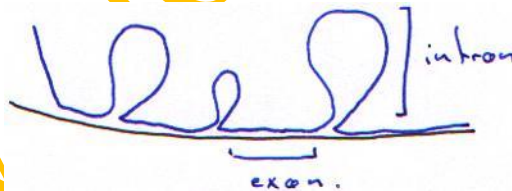


### III\ La maturation des messagers.

Les globines et l'ovalbumine sont des messagers. Les premiers cDNA ont été réalisés sur ces molécules. On est allé chercher les gènes et on en a fait la carte de restriction mais on n'obtient rien de correct sans coupures ni ligations.

ON peut visualiser l'ADN double brin de ces molécules au microscope électronique par des dépôts d'ions métalliques.

Lorsque l'on hybride le gène « ova » et son cDNA, on obtient :



La structure est formée de parties conservées et de parties éliminées.

→ Les gènes eucaryotes sont dits gènes mosaïques.

Chez la levure, 95% des gènes n'ont pas d'intron. Les 5% restant n'en ont qu'un.

Chez les mammifères supérieurs, 95% des gènes ont des introns (les histones n'ont pas d'introns). On distingue un groupe de gènes avec 3 à 5 introns (30 à 40%), un groupe avec 5 à 10 introns (2 à 30%) et un groupe de gènes de très grande taille pouvant avoir jusqu'à 60 introns.

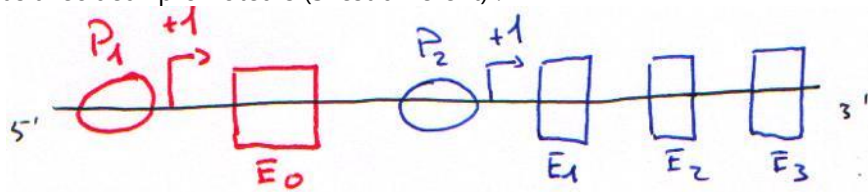
Les maladies génétiques sont souvent dans les gènes de grande taille.

La taille moyenne d'un ARN est de 3000pb alors que celle d'un gène est de 9000pb. Les exons ont une taille moyenne autour de 400pb et les introns voient leur taille varier de 100pb à plusieurs milliers de paires de bases. Les séquences sans gène sur le chromosome représentent  $10^5$  à  $10^6$ pb.

### A\ Différences d'extrémités 5' et 3'.

A partir d'un gène, on peut avoir plusieurs protéines.

- Le cas avec deux promoteurs (5' est différent) :



On a un même gène avec deux extrémités 5' différentes. Suivant de premier AUG, on a deux extrémités NH<sub>2</sub> distinctes et donc, des signaux d'adressage différents.

- Le cas où 3' est différente :

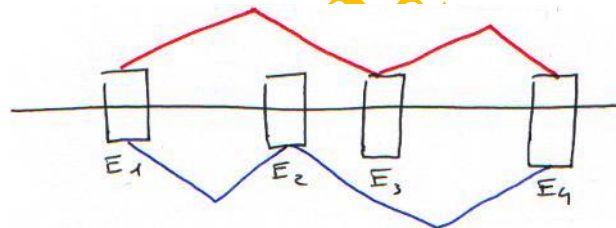


Pour aller chercher les exons en plus, on fait de la « marche sur le chromosome ».

On prend une sonde 5' du premier cosmide. Si elle chevauche avec le cosmide portant E0, c'est gagné. On fait ça plusieurs fois, jusqu'à trouver le cosmide portant E0. On peut « marcher » en 5' ou en 3'.

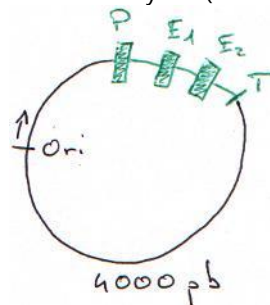
## B\ Epissage alternatif.

Dans ce cas, on compare l'expression du même gène dans différents tissus. On obtient des ARNm distincts.



Dans ce cas, les extrémités NH<sub>2</sub> et COOH sont identiques mais le milieu est différent.

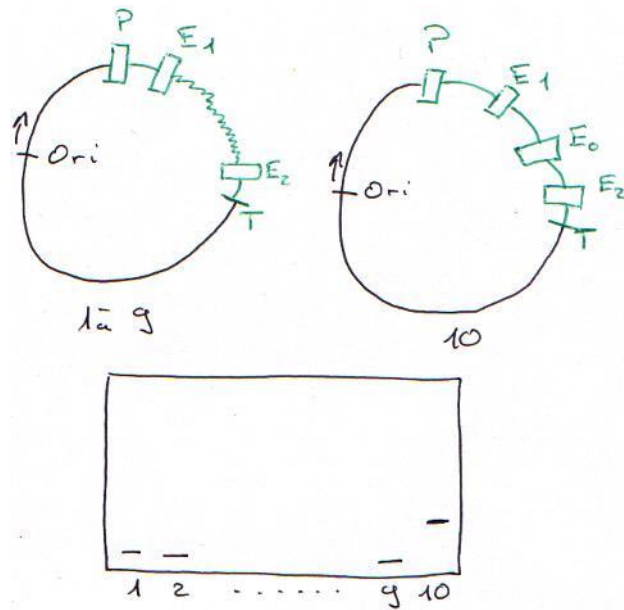
Dans un cosmide, on va essayer d'aller attraper un exon E0. La technique utilisée est « l'exon trapping ». Pour cela, on va utiliser un vecteur eucaryote (souvent le virus SV40).



On rajoute sur le vecteur : un promoteur, un exon connu E1 et un autre E2, un terminateur et des sites de restriction connus.

La transfection donnera l'ensemble constitué de E1 et E2 ainsi que de la partie intermédiaire. L'épissage donnera E1-E2.

On va réaliser un northern avec les sondes (E1, E2 ou E1E2). On découpe ensuite le cosmide portant l'exon inconnu E0 en 10 fragments d'environ 4kb que l'on insère entre les exons E1 et E2.



Dans le complexe 10, on fait un northern avec E1E2.  
La migration permet de récupérer l'exon inconnu. Il ne reste plus qu'à le séquencer.

## IV\ Les familles de gènes.

### A\ Les gènes globiques.

Ces gènes vont permettre la formation des tétramères de globine de  $Fe^{2+}$  (isolés par les biochimistes). Ces tétramères sont composés de 2 chaînes  $\alpha$  et de deux  $\beta$  dans 95% des cas. Dans les 5% restant, on trouve deux chaînes  $\alpha$  et deux  $\delta$  ( $\delta$  ne diffère de  $\beta$  que par quelques acides aminés).

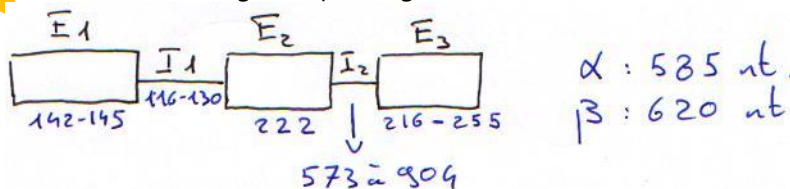
Chez l'embryon, on trouve  $\zeta_2\epsilon_2$ ,  $\zeta_2\gamma_2$ ,  $\alpha_2\epsilon_2$ .

Chez le fœtus, on a :  $\alpha_2\gamma_2$ .

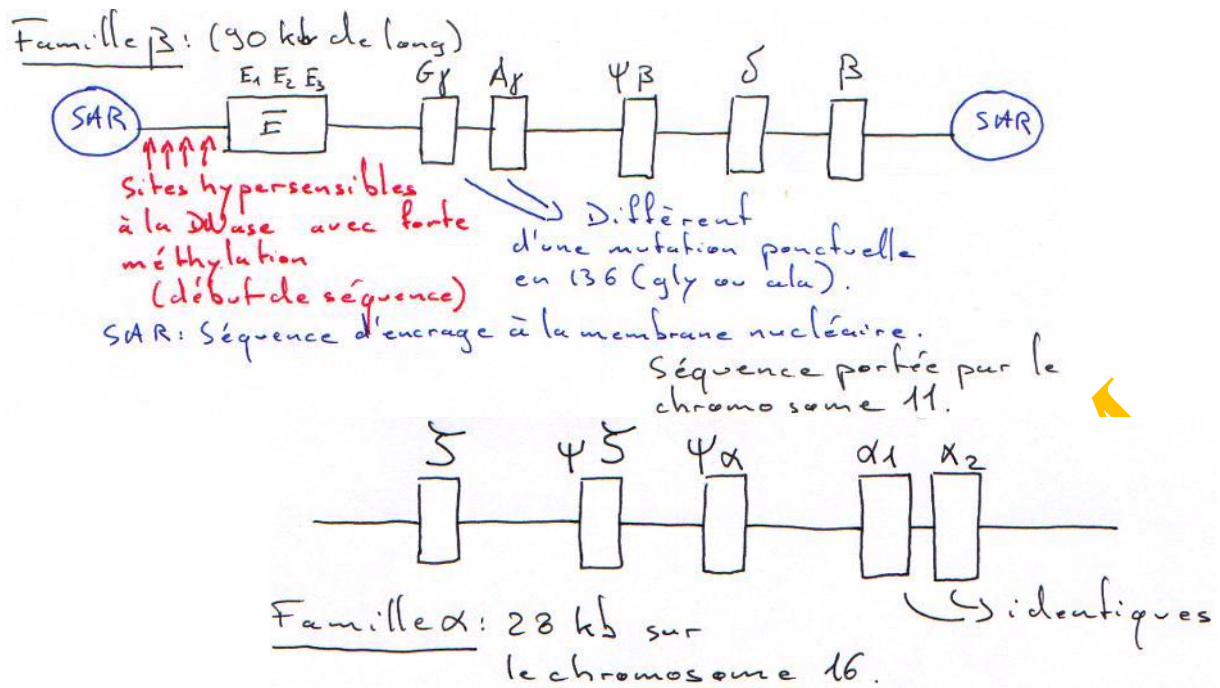
Chez l'adulte, il y a :  $\alpha_2\beta_2$  ou  $\alpha_2\delta_2$ .

### 1\ Constitution des gènes de globines.

La thalassémie est une maladie génétique de globines.







$\psi$ : ce sont des pseudogènes. En réalité, c'est un cDNA qui s'est réinséré: on a E1E2E3; ils ne sont pas fonctionnels car ils n'ont pas de promoteur et portent des mutations.  
Aux bornes des  $\psi$ , on trouve une séquence répétée directe (4 à 6nt).

## 2\ L'hybridation gène/cDNA.

A haute stringence, on obtient des hybrides parfaits (si 5 gènes).  
A basse stringence, on a des hybrides non parfaits (si 50 gènes).

## 3\ Evolution moléculaire.

Les mutations sont des horloges moléculaires qui permettent de reconstituer un arbre phylogénétique. On devrait avoir un ancêtre commun pour  $\alpha$  et  $\beta$  datant de 800 millions d'années.  
Il y a 500 millions d'années, il y a duplication de gène:  $\alpha \rightarrow \alpha + \beta$ .  
Voilà 200 millions d'années,  $\alpha$  est sur le chromosome 16 et  $\beta$  sur le 11.

## 4\ La thalassémie.

On a deux types:  $\alpha$  et  $\beta$ , qui touchent leur gène respectif.

### • $\beta$ :

Mutation ponctuelle  $\rightarrow$  non-sens  $\rightarrow$  protéine incomplète.  
Mutation ponctuelle des points d'épissage  $\rightarrow$  traduction de l'intron  $\rightarrow$  globine inactive.

### • $\alpha$ :

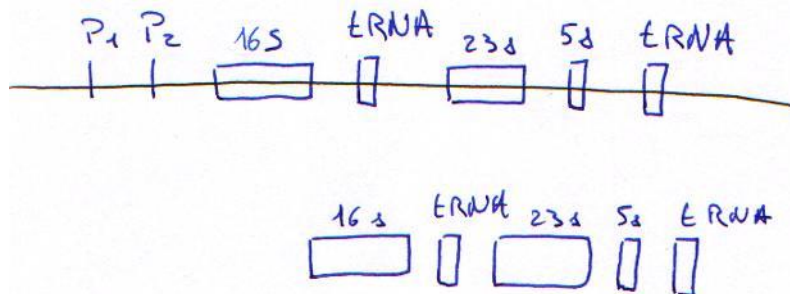
Pertes de gènes: crossing over inégal:  
 $\rightarrow$  Individus avec un gène  $\alpha$  manquant.  
 $\rightarrow$  Individus avec deux gènes  $\alpha$  manquants.  
 $\rightarrow$  Individus sans gène  $\alpha \rightarrow$  mort à la naissance.

## V\ Les rDNA (la famille).

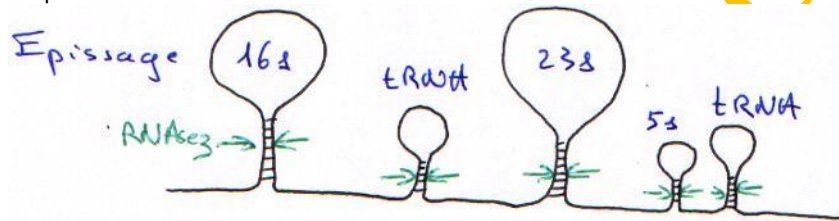
Ces rDNA donnent des ARN de structure.

## A\ Chez les procaryotes.

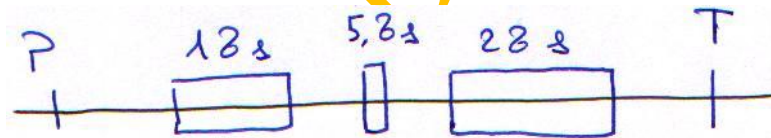
Chez E. coli, le gène codant rDNA est présent sept fois.



On trouve deux promoteurs → il y a une différence de la quantité transcrite.  
Le précurseur est un rRNA qui après maturation donne la forme juste au dessus.  
ARNr + ≈60 protéines → ribosomes.



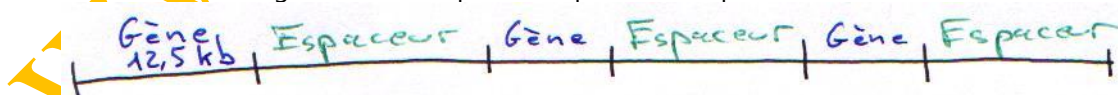
## B\ Chez l'Homme.



La partie de 5,8s correspond à l'extrémité 5' du 23s (procaryote). Les tRNA ne sont plus présents.  
Le 5,8s va s'associer au ribosome sauf chez les eucaryotes supérieurs où il sert à la régulation.  
Ici, on a 280 copies de ce gène.

## C\ Organisation.

Chez l'Homme, le gène fait 12000pb. Les copies sont disposées de la manière suivante :

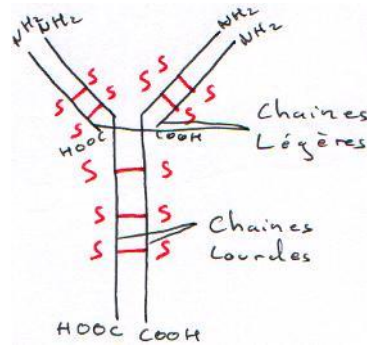


On les trouve sous la forme d'une batterie d'une cinquantaine de gènes. En microscopie photonique, on voit des zones rétrécies sur les chromatides.

Ce sont les gènes les plus répétés.

## VI\ Les Immunoglobulines.





Les immunoglobulines sont composées de deux parties : une partie constante (C-terminal) et une partie variable (NH<sub>2</sub>-terminal). Les Ig peuvent être ancrées à la membrane.

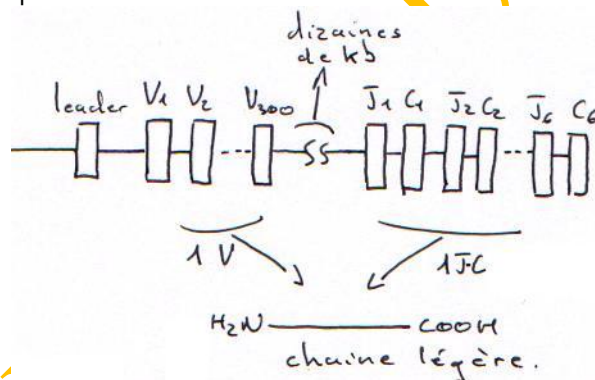
Tonegawa a donné la structure des Ig en travaillant à partir de mélanomes (un lymphocyte tumoral). Le lymphocyte produit dans ce cas une grande quantité d'anticorps du même type et 30% de ses mRNA produits sont utilisés à la synthèse d'IG.

Si l'on fait un southern génomique, les résultats sont in-interprétables.

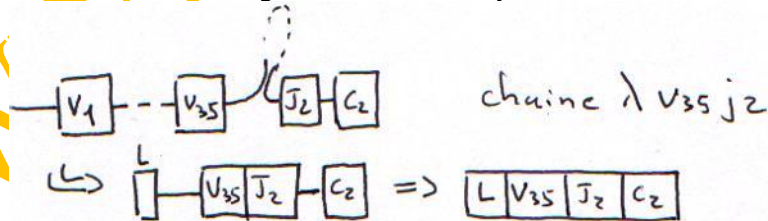
Un southern de fibroblastes donne des résultats différents de ceux du mélanome.

Un southern de lymphocyte tumoral donne une organisation différente de celle d'un lymphocyte activé.

Les lymphocytes L<sub>b/t</sub> non activés ont une organisation classique du génome. Les activés subissent un réarrangement génomique.



→ C'est la chaîne  $\lambda$ . Quand un antigène est reconnu, il y a recombinaison. Par exemple :



Un brin se replie avec 19 appariements qui permettent la mise en contact de V35 et de J2. On obtient d'abord un pré-mARN puis un mARN.

Pour les « V », les signaux sont toujours e, 5'. Quand l'Ig est recombinée, la reconnaissance se fait de 5' vers 3'.