

## Leçon n° 8

### Expression du patrimoine génétique

#### I. Relation séquence des gènes (génom) et protéines (protéome) => activité n° 13

##### 1) Gènes et protéines, des territoires et des séquences différents qui posent le problème de leur relation

+ Les **gènes** qui ont pour support l'ADN ont pour siège le **noyau** : la méthode de **coloration de Feulgen** met en évidence cet ADN par une **coloration rouge** dans ce compartiment nucléaire. Le gène est une séquence bien particulière de nucléotides, séquence constituée à partir d'un **alphabet à 4 lettres** (**A**dénine, **T**hymine, **G**uanine et **C**ytosine).

+ Les **protéines** ont pour siège le **cytoplasme** : on peut mettre en évidence cette localisation par autoradiographie, en apportant à la cellule des **acides aminés marqués** (= radioactifs), on constate une radioactivité ayant pour siège le cytoplasme même après lavage de la préparation : l'incorporation de ces acides aminés dans les protéines cytoplasmiques lors de leur synthèse en est la cause. Les protéines sont constituées d'une succession d'**acides aminés** reliés entre eux par des **liaisons peptidiques**. Dans la réalité un enchaînement d'acides aminés forme un **polypeptide** et une **protéine** est constituée par l'agencement de quelques –voire plusieurs – polypeptides. Par exemple dans l'hémoglobine, quatre polypeptides nommés globines participent à former cette molécule. Il existe **20 acides aminés** différents pouvant constituer ces polypeptides, donc ces protéines.

+ Un des concepts majeurs de la biologie est qu'« **un gène code pour une protéine** ». Compte tenu des **différences de localisation et de structure de ces molécules** se pose donc la question de la relation entre elles.

##### 2) Relation ADN / Protéine formée : colinéarité gène / protéine

+ Des expériences réalisées sur des **virus** – pour lesquels le **caractère infectieux est testé** – montrent que toutes **modification de séquences nucléotidiques** d'un **multiple de trois** (addition ou délétion) ou substitutions ne **change pas la virulence**. Le chiffre 3 joue sans doute dans la correspondance ADN / ARN un rôle important.

+ Une seule substitution localisée au **6<sup>ème</sup> codon** (ensemble de 3 nucléotides) du gène codant pour la  $\beta$  globine se traduit par le changement du **6<sup>ème</sup> acide aminé** de la séquence de ce polypeptide : l'acide aminé « valine » remplace l'acide aminé « acide glutamique ». **L'enchaînement des acides aminés** d'une protéine suit donc **l'enchaînement des nucléotides** de l'ADN constituant le gène : c'est la **colinéarité gène / protéine**.

##### 3) Echanges entre le noyau et le cytoplasme : l'ARN, une molécule qui pourrait être un intermédiaire

+ On connaît un **acide nucléique** nommé **ARN** (Acide Ribo Nucléique) qui pourrait être **l'intermédiaire** entre les deux molécules ADN et protéines.

+ Cette molécule ARN, comme l'ADN, est constituée d'un enchaînement de nucléotides mais présente 3 différences :

- L'ARN n'est constituée que d'un **seul brin** => il n'y a **pas de liaisons « hydrogène »**,
- Le sucre constituant les montants de la molécule est toujours un pentose (= 5 atomes de carbone), mais c'est le **ribose** au lieu du **désoxyribose** (ce qui signifie que le sucre de l'ARN incorpore un oxygène supplémentaire),
- Enfin, le **nucléotide Thymine n'existe pas** et est remplacé par l'**Uracile**.

+ La coloration au vert de **méthyl-pyronine** est un colorant qui permet de **différencier** l'ADN (coloré en bleu-vert) et l'ARN (coloré en rose) ; appliquée à l'épiderme d'oignon, on constate alors que l'ADN a bien pour siège le noyau alors que l'ARN est localisé à la fois dans le noyau et le cytoplasme.

+ Mais la présence de l'**Uracile**, nucléotide spécifique de l'ARN va permettre par autoradiographie après marquage de l'Uracile et lavage de la préparation, de montrer que **l'ARN est d'abord présent** dans le noyau **où il se forme** à partir des nucléotides fournis puis **migre du noyau vers le cytoplasme**. Les électrographies de cellules réalisées après cryodécapage montrent l'existence de **pores nucléaires**, orifices par lesquels **l'ARN s'échappe du noyau**.

+ D'autres expériences montrent une nouvelle propriété de l'ARN confirmant son implication dans la relation gène / protéine : au fur et à mesure que l'ARN est introduit dans un milieu *in vitro* en présence d'acides aminés, la chaîne polypeptidique s'allonge par intégration de nouveaux acides aminés. Par ailleurs l'ARN disparaît vite ce qui signifie que cette **molécule est peu stable** et donc **éphémère**.

##### 4) Correspondance ADN / ARN / Protéines : le code génétique

+ **Correspondance ADN / ARN**. Si on compare les séquences des 2 brins complémentaires de l'ADN à l'ARN correspondant, on constate que la succession de nucléotides dans les deux types de molécules est tout à fait la même, les « U » remplaçant les « T » dans l'ARN comme vu au paragraphe précédent. On nomme « **brin non transcrit** » de l'ADN celui qui a rigoureusement la même séquence que l'ARN (au U / t près) et « **brin transcrit** » celui qui est le complémentaire de l'ARN. On verra en effet dans le chapitre suivant que la **transcription** est le mécanisme par lequel un ARN est fabriqué à partir du brin de l'ADN **qualifié de transcrit**.  
+ **Correspondance ARN / Protéines**. Nous avons vu plus haut que un **triplet de nucléotides** (= codon) semblait **correspondre à un acide aminé**<sup>1</sup>. En effet l'ARN (et ADN) est construit à partir d'un alphabet de 4

<sup>1</sup> Un seul nucléotide ne peut placer que 4 (car 4<sup>1</sup>) acides aminés et non 20 ; de même une association de 2 nucléotides peut placer 16 acides aminés (car 4<sup>2</sup>) ne répondant pas à la nécessité de placement des 20 disponibles. C'est donc un triplet de **3 nucléotides** (= codon) qui code pour un **acide aminé**.

**lettres** alors que les protéines sont construites à partir de **20 acides aminés**. On nomme « **code génétique** », le système de correspondance entre chaque codon de nucléotide et l'acide correspondant. Or un codage par « triplets » offre  $4^3 = 64$  **combinaisons** possibles, donc bien supérieure aux 20 acides aminés à placer. Ce qui signifie donc que **plusieurs combinaisons de codons** codent pour **le même acide aminé** (exemple 6 codons pour l'acide aminé LEU (leucine)) : on nomme

**redondance** cette propriété du code génétique. Cependant un codon donné code toujours pour le même acide aminé : on dit que le code génétique est **univoque**. Enfin, 3 codons particuliers (UAA, UAG et UGA) ne codent pour aucun acide aminé mais ont valeur de **fin de message** : ce sont les « **codons-stop** ».

+ Enfin, les exemples de **transgénèse** vus en 2<sup>nde</sup> montrent que ce code génétique est un système appliqué par l'ensemble du vivant : on dit que le **code génétique est universel**<sup>2</sup> (le gène codant pour l'insuline humaine introduit dans le génome des bactéries fait fabriquer par ces bactéries... de l'insuline humaine). Nous avons vu à ce propos que cette caractéristique était une preuve supplémentaire de **parentés du vivant** et d'une **origine commune**. L'**ancêtre commun** appliquant déjà ce code génétique.

		2 <sup>ème</sup> nucléotide								
		U	C	A	G					
1 <sup>er</sup> nucléotide	U	UUU	phénylalanine [PHE]	UCU	sérine [SER]	UAU	tyrosine [TYR]	UGU	cystéine [CYS]	U
		UUC		UCC		UAC		UGC		C
		UUA	leucine [LEU]	UCA		UAA	codons-stop	UGA	codon-stop	A
	C	CUU	leucine [LEU]	CCU	proline [PRO]	CAU	histidine [HIS]	CGU	arginine [ARG]	U
		CUC		CCC		CAC		CGC		C
		CUA		CCA		CAA	glutamine [GLN]	CGA		A
	A	CUG		CCG		CAG		CGG		G
		AUU	isoleucine [ILE]	ACU	thréonine [THR]	AAU	asparagine [ASN]	AGU	sérine [SER]	U
		AUC		ACC		AAC		AGC		C
	G	AUA		ACA		AAA	lysine [LYS]	AGA	arginine [ARG]	A
		AUG	méthionine [MET]	ACG		AAG		AGG		G
		GUU		GCU	alanine [ALA]	GAU	acide aspartique [ASP]	GGU	glycine [GLY]	U
GUC	valine [VAL]	GCC		GAC		GGC		C		
GUA		GCA		GAA	acide glutamique [GLU]	GGA		A		
GUG		GCG		GAG		GGG		G		

« **codons-stop** ».

+ Enfin, les exemples de **transgénèse** vus en 2<sup>nde</sup> montrent que ce code génétique est un système appliqué par l'ensemble du vivant : on dit que le **code génétique est universel**<sup>2</sup> (le gène codant pour l'insuline humaine introduit dans le génome des bactéries fait fabriquer par ces bactéries... de l'insuline humaine). Nous avons vu à ce propos que cette caractéristique était une preuve supplémentaire de **parentés du vivant** et d'une **origine commune**. L'**ancêtre commun** appliquant déjà ce code génétique.

## II. Du génome au protéome : variété des ARNm et mécanismes de transcription / maturation / traduction (=> activité n° 14)

### 1) Du génomes au protéome : présentation des mécanismes de transcription / (maturation) / traduction

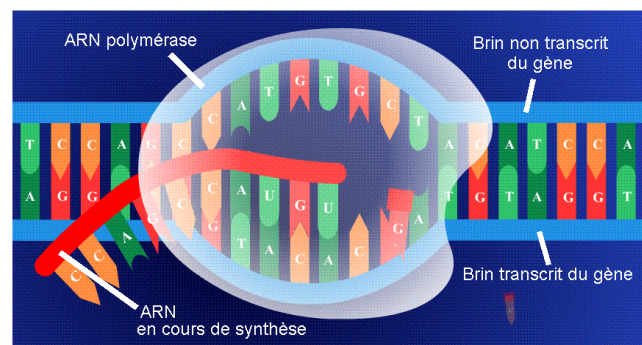
+ Le chapitre précédent a montré que chez les **eucaryotes**<sup>3</sup>, la **synthèse des protéines** se faisait dans deux compartiments différents, successivement (noyau puis cytoplasme), et impliquait des molécules différentes (ADN, ARN et protéine). Le **génome** est l'ensemble des gènes et le **protéome** est l'ensemble des protéines produites.

+ Dans le noyau, l'un des brin de l'ADN (**brin transcrit**) est **copié en ARN** : c'est la **transcription**. Dans le cytoplasme, l'ARN est **traduit** en polypeptide (protéine) selon de code de correspondance vu précédemment (= code génétique) : c'est la **traduction**.

+ Chez les procaryotes, l'ARN produit par transcription est **directement** traduit en protéines. Cet ARN porte le nom d'ARN messager (**ARNm**). Chez les eucaryotes, un mécanisme un peu plus complexe s'**immisce** entre transcription et traduction : c'est le phénomène de **maturation de l'ARN** qui à partir d'un **ARN pré-messager** (résultant de la **transcription** au niveau du noyau) produit un ou parfois plusieurs ARNm différents.

### 2) Transcription au niveau du noyau : fabrication d'un ARN pré-messager

+ Une enzyme intervient lors de la transcription : c'est l'**ARN polymérase** (qui présente des points communs avec l'ADN polymérase déjà dans les leçons précédentes). Cette **enzyme** après avoir **reconnu la séquence du gène à transcrire**, **dégrade** les deux brins de l'ADN (comme le faisait l'ADN polymérase) en supprimant localement les liaisons hydrogène, puis au fur et à mesure de sa progression le long du gène, **associe** à l'un des brins dégradé (**brin transcrit**) des **ribonucléotides libres** par **complémentarité** des nucléotides constituant ce brin transcrit (A => U, T => A, C => G et G => C). C'est grâce à cette **complémentarité des**

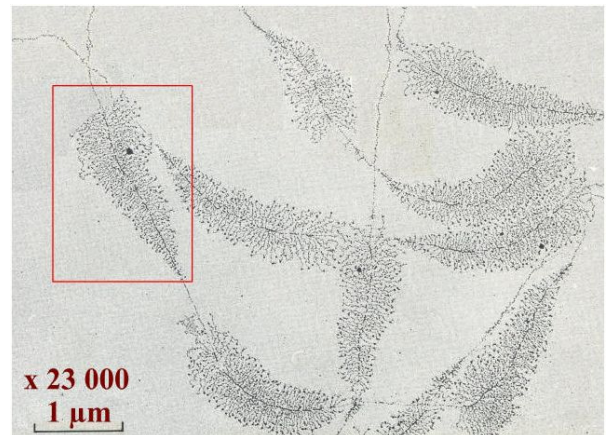


<sup>2</sup> **Remarque** : les élèves confondent souvent **génome** (patrimoine génétique = ensemble des gènes possédés par un individu) et **code génétique** (système de correspondance entre codons des gènes et les acides aminés des protéines que ces gènes permettent –via L'ARN – de faire fabriquer).

<sup>3</sup> Chez les procaryotes, le génome est disséminé dans toute la cellule et la synthèse des protéines (transcription puis traduction) se fait donc dans le même compartiment cytoplasmique.

**nucléotides** lors de la transcription que l'**ARN produit** est identique (à la différence du remplacement de T par U) au **brin non transcrit**. Au fur et à mesure de la progression de l'ARN polymérase sur le gène, l'ARN croit. A la fin du gène, l'ARN polymérase **se détache du gène** ainsi que l'ARN pré-messager ainsi formé, qui va subir une **maturation** avant de migrer (c'est alors un **ARN messager**) vers les pores nucléaires puis vers le noyau.

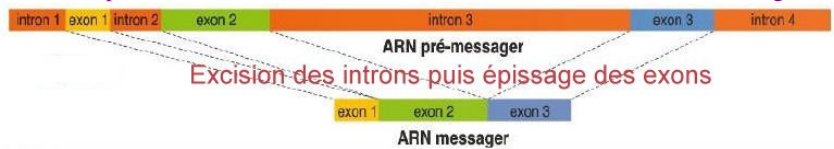
+ Notons que dans le noyau d'une cellule, **plusieurs gènes sont transcrits en même temps** (10 sur le cliché ci-contre) et pour un **gène donné** (voir encadré rouge) plusieurs ARNm sont produits en même temps du fait que les ARN polymérases se suivent le long du gène. On a ainsi un instantané sur lequel pour ce gène, on peut observer des ARN en cours de croissance et à tous les stades de croissance. Ce mécanisme de transcription produit donc des **quantités importantes de copies identiques du gène** (quantités d'ailleurs nécessaires compte tenu de la vie éphémère de chaque molécule d'ARN).



### 3) Maturation de l'ARN pré-messager en ARN messager

#### a) complexité du gène des eucaryotes : le gène morcelé

+ L'**ARN pré-messager** est une copie fidèle du brin (non) transcrit. Il est constitué d'un ensemble de fragments ou tronçons. Certains de ces tronçons se retrouvent dans l'ARNm : ce sont des **tronçons codants** ou **exons** et d'autres,

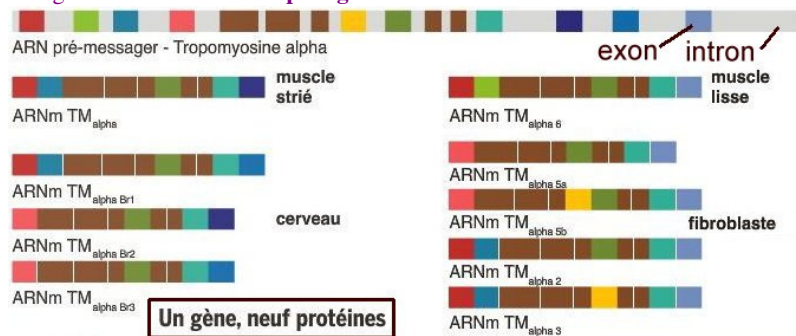


éliminés (tronçons non codants) ne se retrouvent pas dans l'ARN messager. Ainsi lors de la maturation de l'ARN, il y a une **excision** des introns (suppression, élimination) puis un **épissage des exons** (qui sont raccordés entre eux). Le gène est dit **morcelé** car les fragments codants d'ADN sont séparés par des fragments non codants. Il n'y a pas de **morcellement des gènes** chez les procaryotes.

#### b) un gène, des protéines ou l'épissage alternatif

+ Un même gène après **excision des introns** peut aboutir à **plusieurs ARNm** selon la façon dont sont **assemblés les exons**. Cet assemblage variable est nommé **épissage alternatif**.

+ Remarquons que l'épissage alternatif est une sérieuse entorse au dogme « un gène, une protéine » puisque un même gène transcrit un ARN pré-messager qui après excision des introns et selon l'assemblage des exons aboutit à plusieurs ARN messager qui traduits donneront plusieurs protéines différentes.

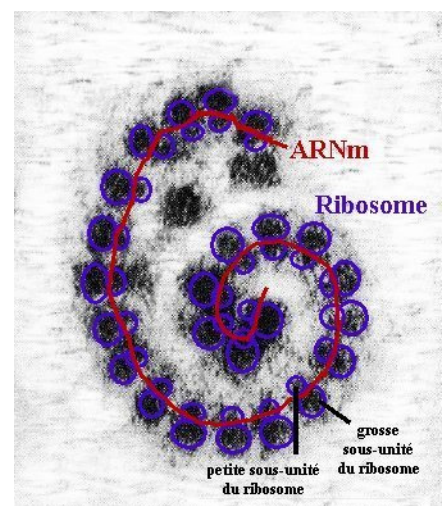


Remarquons que le mécanisme d'épissage alternatif conduit aussi à une **excision de certains exons** (dans la confection de l'ARNm) mais que les exons sont toujours assemblés dans l'ordre qu'ils ont dans l'ARN pré-messager.

+ Chez l'Homme, l'épissage alternatif concernerait plus de **60 % de nos gènes** et expliquerait les **disproportions entre le génome** (évalué actuellement entre 20000 et 25000 gènes) et le **protéome** (évalué entre 500 000 et 5 000 000 protéines). C'est cet épissage alternatif qui semble jouer au **cours de l'évolution un rôle important**.

### 4) Traduction au niveau du cytoplasme de(s) ARNm en protéine(s)

+ Dans le cytoplasme des organites, les **ribosomes** jouent un rôle essentiel dans la traduction. Ces ribosomes sont le plus souvent regroupés en ensembles alignés (= **polysomes**). Ces ribosomes sont constitués de deux sous-unités qui s'associent au moment de la traduction. La **petite sous-unité** (30 s) est le site de reconnaissance de l'ARNm et c'est donc à ce niveau que les



ribosomes se fixent sur le messenger. La **grosse sous-unité** (50s) est la partie qui lie les **acides aminés** pour former entre eux une **liaison peptidique** ; deux sites sur cette grosse sous-unité permettent la **réception de deux acides aminés**, acides aminés qui font face à deux codons localisés au niveau de petite sous-unité.

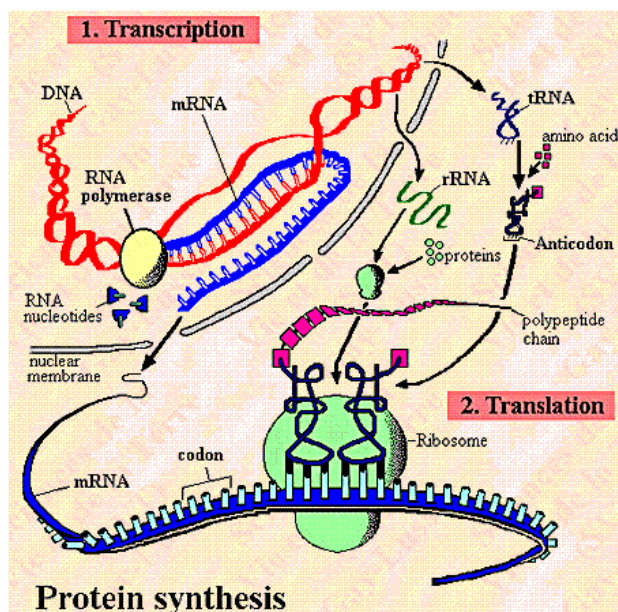
+ La synthèse des protéines se fait en trois temps.

- Au cours de l'**initiation**, c'est le **codon AUG** (codant pour la méthionine) et premier codon de tout ARNm sert d'**initiateur** et c'est à ce niveau que les deux sous-unités s'assemblent pour former le **ribosome**.

- Puis vient l'**élongation** : le ribosome glisse le long de l'ARNm de codons en codons tout en associant les acides aminés correspondants (selon le code génétique) avec création d'une **liaison peptidique** entre acides aminés au fur et à mesure qu'ils se placent sur le ribosome.

- Puis vient le **codon-stop**. Les **deux sous-unités se dissocient** et le **polypeptide se détache** et aussitôt se **spatialise** (en effet les acides aminés qui constituent ce polypeptide ont des groupements chimiques qui interagissent avec leurs voisins en créant des attractions et des répulsions dans la molécule formée lui conférant une forme spatiale).

+ Comme pour la transcription, où plusieurs ARN polymérase glissent le long du gène, dans le cas de la traduction plusieurs ribosomes glissent le long de l'ARNm pour produire en même temps plusieurs chaînes polypeptidiques. Notons que **cet effet d'amplification** cumulé par les **ARN polymérase** et les **ribosomes** contribue à produire d'un point de vue quantitatif à une **masse importante de protéines**.



### III. Influence des gènes et de l'environnement sur les phénotypes aux différentes échelles du vivant (=> activité n° 15)

#### 1) Du génotype au phénotype macroscopique : les phénotypes aux différentes échelles du vivant (exemple de la drépanocytose)

+ On nomme **génotype**, l'**assortiment d'allèles** d'un gène donné. Le phénotype est **ce qui se voit**, donc l'ensemble des caractéristiques aux trois échelles du vivant : molécule, cellule et organisme :

- Le **phénotype moléculaire**. Les effets du phénotype s'observent par les **mutations du génome**. Dans le cas de la drépanocytose, une simple **substitution** du gène codant pour la  $\beta$  globine conduit à modifier légèrement la protéine en substituant un **seul acide aminé** (VAL remplace GLU) parmi les 146 acides aminés du polypeptide. Peu de modification de forme mais une affinité particulière de VAL qui conduit à une polymérisation. Alors que normalement 2 globines  $\beta$  s'associent à 2 globine  $\alpha$  pour former une protéine complexe (l'hémoglobine), la Valine en position 6 du polypeptide conduit à une polymérisation en baguettes des hémoglobines drépanocytaires (HbS), baguettes insolubles : c'est le **phénotype moléculaire pathologique** (mauvaise fixation du dioxygène => **anémie**) alors que les hémoglobines « normales » (HbA) solubles assurent correctement cette fixation du dioxygène.

- Le **phénotype cellulaire**. C'est au niveau des globules rouges (hématies) que ce phénotype s'observe bien dans cette pathologie. Alors que les **hématies normales** sont des **disques biconcaves** presque exclusivement constituées d'hémoglobine HbA [le noyau disparaît progressivement dès lors que l'hémoglobine est produite par synthèse des globines  $\alpha$  et  $\beta$  (et  $\delta$ )], les **hématies drépanocytaires** sont en **forme de faucille** (du grec « drepano », faucille et « cyto », cellule) car **déformées** et **déchirées** par les fibres insolubles d'HbS polymérisée. Alors que le diamètre d'une hématie est habituellement de 7  $\mu\text{m}$ , les hématies drépanocytaires peuvent atteindre plus de 10  $\mu\text{m}$ . Or certains petits vaisseaux ont des diamètres de 8  $\mu\text{m}$  et si les hématies normales y circulent ce n'est pas le cas des hématies drépanocytaires qui forment des amas ou **thromboses**. **Remarque importante**. Notons avec cet exemple que si les cellules ont **toutes le même génome**, elles n'en expriment qu'une partie qui diffère selon la nature de la cellule (exemple : les globules rouges expriment les gènes qui codent pour les globines, alors que les photorécepteurs expriment les gènes qui codent pour les pigments rétiniens...). La spécialisation d'une cellule provient donc de la partie du génome que cette cellule exprime, donc dépend des **protéines codées par les gènes exprimés**.

- Le **phénotype macroscopique** est la conséquence des **deux autres phénotypes**. Le sujet drépanocytare consulte pour des troubles respiratoires raisons d'anémies et des troubles circulatoires qui sans adaptation environnementales sont difficiles à vivre.

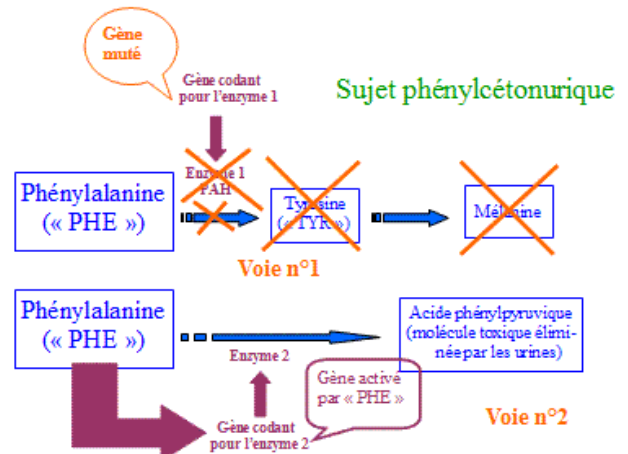
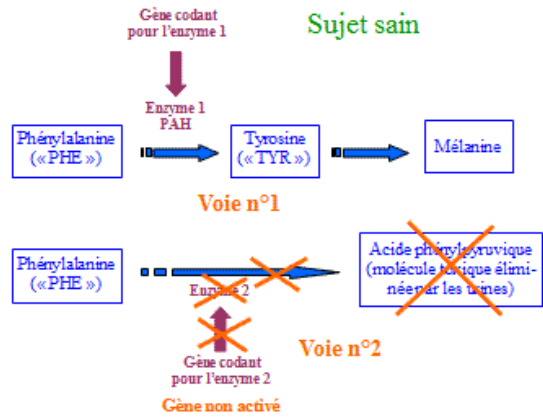
#### 2) Les gènes déterminent le phénotype moléculaire

+ Les mutations qui affectent les gènes ont pour conséquence la **synthèse d'une protéine peu différente** de la protéine normale (cas de la **drépanocytose**, un seul acide aminé modifié qui aboutit malgré tout à une pathologie grave), ou **très différente** comme c'est le cas de la **phénylcétonurie** (452 acides aminés pour

l'enzyme PAH normale, et 59 acides aminés pour l'enzyme PAH exprimée du fait de la mutation). Dans ce cas la protéine **n'est plus fonctionnelle** et comme il s'agit d'une enzyme, la réaction utilisant la phénylalanine apportée par l'alimentation ne se fait plus et cette phénylalanine d'origine alimentaire s'accumule ce qui **déclenche** une réaction n°2 (par activation d'un gène jusque là inactif) ce qui dégrade cette phénylalanine mais aboutit à une molécule hautement toxique pour le système nerveux. (réaction jamais activée chez le sujet normal). On dit que **l'expression de tels gènes** modifiables par des **signaux internes** (voire **externes**) est **inductible** au contraire de gènes qui sont toujours exprimés et ne sont donc pas dépendants de tels facteurs : leur **expression est alors dite constitutive**.

+ Notons aussi que dans le cas de la **phénylcétonurie**, la protéine affectée par la mutation est ici **indirectement** responsable de la pathologie (c'est la conséquence, donc l'accumulation de phénylalanine qui est cause de la pathologie) alors que dans la **drépanocytose**, c'est la  $\beta$  glogine du fait de son comportement polymérisant cause la pathologie.

+ Les **duplication / transpositions et mutations** ne conduisent pas forcément à une détérioration des phénotypes moléculaires. Au contraire, par ces mécanismes, la **famille multigénique** des globines s'enrichit et aboutit à de nouvelles globines telles que les globines  $\gamma$  qui associées aux globine  $\alpha$  forment des **hémoglobines fœtales**. Ces hémoglobines fœtales ( $2\alpha + 2\gamma$ ) ont beaucoup **plus d'affinité** avec le dioxygène dissous que l'hémoglobine adulte ( $2\alpha + 2\beta$ ) qui a la maximum d'efficacité avec le dioxygène gazeux.



### 3) L'environnement influence le phénotype moléculaire

+ En **jouant sur l'environnement**, on peut empêcher la mise en place d'un **phénotype macroscopique pathologique**. Deux exemples. Pour **limiter les effets du phénotype drépanocytaire**, on peut aérer les locaux pour **augmenter le taux de dioxygène** (plus il est élevé, plus le taux d'hématies falciformes chute), **boire de l'eau** pour liquéfier le sang et **porter des vêtements non serrés** pour lutter contre les thromboses... Pour la **phénylcétonurie**, on peut **limiter les apports de phénylalanine** dans le lait de régime donné au nourrisson. On évite ainsi le déclenchement de voie n°2 et donc la formation de la toxine.

+ Au niveau moléculaire, l'excès de phénylalanine, au contraire, provoque la formation de **signaux internes** par exemple des protéines qui en se fixant sur l'ARN polymérase facilitent la fixation de l'ADN sur cette ARN polymérase et permettent ainsi d'initier la transcription. Ce sont donc des **facteurs transcriptionnels**.

