

# FLUX D'INFORMATION GÉNÉTIQUE

## 1-NOTION DE CODE GÉNÉTIQUE

L'hypothèse «un gène une protéine» (évoquée au chapitre consacré à la découverte de la fonction du gène) a été confortée bien avant le développement des technologies de l'ADN recombinant. Ingram, vers 1957 a été l'un des premiers à établir la relation précise entre une mutation bien caractérisée ayant comme manifestation phénotypique une maladie héréditaire grave : l'anémie falciforme (ainsi nommée en raison de la forme en faucille des hématies des sujets homozygotes pour la maladie) et une protéine spécifique : l'hémoglobine. L'hémoglobine normale, de l'adulte, est dite de type A. Elle est composée de quatre motifs polypeptidiques : deux chaînes alpha et deux chaînes beta, l'hémoglobine S, des sujets atteints est également composée des quatre sous unités mais elle se distingue de la A par sa mobilité électrophorétique.



Par des méthodes d'analyse d'oligopeptides et de séquençage protéique que nous ne détaillerons pas, il fut montré que l'hémoglobine S diffère de la normale (A) par un seul acide aminé, en position 6, dans les chaînes  $\beta$  :

	NH <sub>2</sub>							
<b>Hb A :</b>	Val	His	Leu	Thr	Pro	Glu	Glu	....
<b>Hb S :</b>	Val	His	Leu	Thr	Pro	Val	Glu	....

Le remplacement d'un acide aminé diacide (acide glutamique) par un neutre explique la différence de mobilité à l'électrophorèse. Par ailleurs, tous les autres acides aminés sont identiques, autrement dit, l'unité d'effet d'une mutation peut être le changement d'un seul acide aminé dans une protéine.

\* Remarque : La biochimie des protéines a permis de comprendre qu'une très légère altération de la structure primaire pouvait bouleverser les structures secondaires et ter

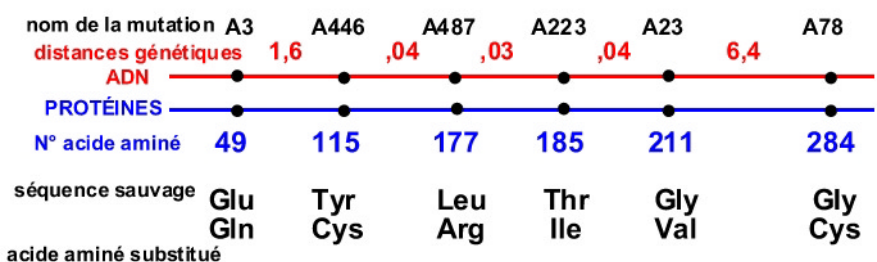
tiaires avec des conséquences physiologiques considérables. Ce n'est pas toujours le cas, tout dépend de la position de la modification dans la protéine, certaines substitutions d'acides aminés ne peuvent avoir que peu ou pas d'effet.

Quelques années plus tard (en 1967), Yanofsky établit une relation entre la structure du gène et la structure primaire de la protéine.

Par les techniques de cartographie à haute résolution adaptées aux bactéries, il positionne plusieurs mutations supposées ponctuelles dans un cistron codant pour la sous unité A de la tryptophane synthétase d'*E. coli*. Il parvient également à séquencer le produit de ce cistron (la protéine) pour le gène sauvage et les variants. L'étude de la séquence en acides aminés de la tryptophane synthétase A des mutants confirme les conclusions précédentes : le plus souvent, un seul acide aminé est modifié, cependant, dans certains cas, les protéines sont tronquées, l'extrémité COOH est prématurée. Yanofsky peut donc construire une carte des effets des mutations dans la séquence protéique et la confronter à la carte génétique des mutations.

La figure ci-dessous montre une excellente colinéarité entre les deux. Ceci suppose que la structure primaire de toute protéine soit spécifiée par un «gène», c'est à dire une séquence d'ADN.

\* Remarque : excellente ne veut pas dire parfaite quant aux distances, il faut bien comprendre la différence technique de réalisation des deux cartes : pour les protéines, il s'agit de carte physique, positionnant les acides aminés les uns après les autres. Pour le cistron, il s'agit de carte génétique reposant sur des phénomènes de recombinaison in vivo. Les méthodes de séquençage de l'ADN ne seront mises au point que plus tard.



Ces découvertes montrent clairement que l'information est codée. Chaque «gène» code une protéine particulière.

La découverte de l'ADN en tant que support de l'information laisse supposer que la séquence des nucléotides n'a pas de rôle prépondérant dans la structure physique de la molécule mais constitue le code de la séquence des acides aminés dans les protéines.

Il existe donc un transfert d'information. Le code génétique représente le système de correspondance entre la séquence des nucléotides dans les acides nucléiques et celle des acides aminés dans les protéines.

La correspondance n'est pas un nucléotide pour un acide aminé puisqu'il n'existe que 4 nucléotides différents pour 20 acides aminés différents. Le code génétique repose sur une combinaison de nucléotides. Une combinaison de 2 parmi 4 possibles ne peut suffire ( $4^2 = 16$ ), une combinaison de 3 (64 possibilités) est vraisemblable. Le système de triplets (3 nucléotides pour spécifier un acide aminé) est effectivement celui qui est universellement utilisé.

Bien avant que l'on ait réussi à déchiffrer l'intégralité du code, l'étude de certaines mutations a apporté des preuves génétiques à la colinéarité entre la séquence des triplets dans un cistron et celle des acides aminés dans la protéine qu'il spécifie. Peu à peu, le terme de cistron fut délaissé et remplacé par celui de gène. Le terme de gène reste associé à l'unité de fonction génétique mais comprend maintenant la notion de séquence de nucléotides dans l'ADN capable d'être traduite en séquence d'acides aminés dans une protéine.

Le mécanisme de traduction implique la lecture des triplets les uns après les autres si l'on suppose le code non chevauchant (le fait que des mutations n'entraînent la modification que d'un seul acide aminé dans une protéine le laisse supposer). Cette traduction implique une notion très importante, celle de cadre de lecture : dans toute séquence d'ADN, si l'on admet le système de triplets, il existe 3 cadres de lecture potentiels.

Par exemple, la séquence :

... A C G A C G A C G A C G A C G A...

peut être décomposée en

ACG ACG ACG ACG ACG etc

soit, comme on le verra plus tard, une série de triplets pour l'acide aminé thréonine.

*\*Remarque : jamais on ne trouve de telles séquences monotones dans l'ADN, il s'agit d'un exemple théorique.*

on peut également la décomposer en

CGA CGA CGA CGA CGA etc

soit arginine

ou encore en

GAC GAC GAC GAC GAC etc

soit acide aspartique

Bien entendu, pour un gène donné, une seule possibilité peut conduire à une protéine normale, la traduction assurera une lecture correcte.

Des mutations par insertion ou délétion d'un seul nucléotide dans une séquence sauvage (correspondant à l'allèle habituel) ont permis de montrer l'importance du cadre de lecture.

Soit une séquence sauvage hypothétique (et peu vraisemblable):

5' GCT GCT GCT GCT GCT GCT GCT 3'

pour le brin codant. Selon les conventions, le produit de transcription du brin complémentaire de celui-ci est :

5' GCU GCU GCU GCU GCU GCU GCU 3'

et la protéine :

NH<sub>2</sub> Ala Ala Ala Ala Ala Ala Ala COOH

Supposons une mutation par insertion d'une adénosine en position 7, les répercussions sont les suivantes:

ADN : \*  
GCT GCT AGC TGC TGC TGC TGC

ARN :  
GCU GCU AGC UGC UGC UGC UGC

Protéine :  
Ala Ala Ser Cys Cys Cys Cys

À partir de l'insertion, toute la protéine est fautive, et certainement non fonctionnelle

De même, une mutation par **délétion** d'une Guanine en 16 a également un effet de décalage du cadre de lecture :

ADN : \*(-G)  
GCT GCT GCT GCT CTG CTG  
ARN :  
GCU GCU GCU GCU GCU CUG CUG  
Protéine :  
Ala Ala Ala Ala Ala Leu Leu

Par contre, si l'on examine le double mutant (insertion suivie d'une délétion), on s'aperçoit que la 2ème mutation rétablit le cadre de lecture, elle supprime l'effet de la première à partir du point de mutation.

ADN : +A -G  
GCT GCT AGC TGC TGC TCT GCT  
ARN :  
GCU GCU AGC UGC UGC UCU GCU  
Protéine :  
Ala Ala Ser Cys Cys Ser Ala

*\* Remarque : Cette **suppression** de l'effet d'une mutation par une autre est une notion importante. Elle est, ici, **intra génique**, la deuxième mutation ayant lieu dans le même cistron que la première. Ce n'est pas toujours le cas, le phénomène de **suppression** peut être **inter génique**, physiologique ...*

Trois mutations auraient des effets différents selon leur signe (par convention : + = insertion, - = délétion)

Trois mutations successives de même signe rétablissent le cadre de lecture et le résultat est l'insertion ou la délétion d'un seul acide aminé dans la protéine.

**Un acide aminé est donc bien spécifié par un système à trois nucléotides.**

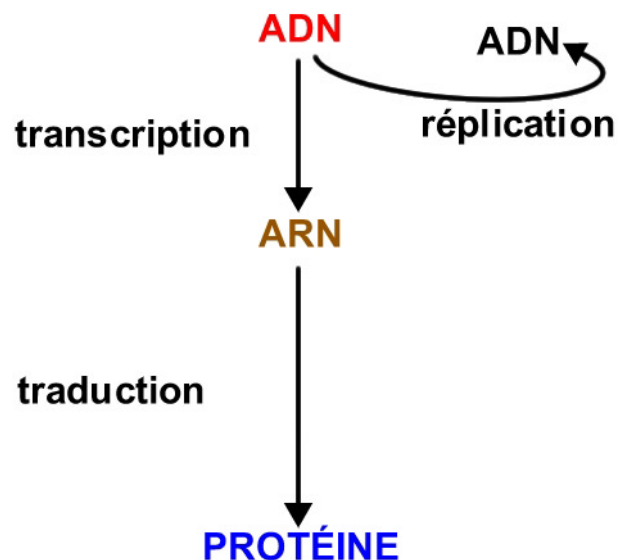
Par la suite, grâce à des expériences reposant sur des systèmes de synthèse protéique *in vitro*, la signification des 64 combinaisons possibles de trois nucléotides (parmi un choix de quatre) a pu être élucidée. Les résultats sont compilés dans le tableau présenté page suivante.

Trois seulement n'ont pas de sens en termes d'acides aminés et des arguments génétiques montreront qu'ils agissent de signaux d'arrêt de la traduction. 61 triplets ont une signification pour l'un des 20 acides aminés et seront appelés **codons**.

Le tableau met en évidence de nombreux **synonymes**. Seuls la méthionine et le tryptophane ne sont codés que par un seul triplet, les autres le sont par des familles de 2, 3, 4 et même 6 (arginine). A part cette exception (l'arginine), il est important de remarquer que les deux premiers nucléotides de triplets codant pour un acide aminé donné sont les mêmes, seul le troisième est variable.

Comment se fait la traduction ?

Etant donné que l'ADN ne sert pas physiquement de support à la synthèse des protéines, il existe un flux d'information génétique dans la cellule.



Ce flux d'information se fait en réalité en deux étapes: il existe un intermédiaire entre la séquence d'ADN (unité d'information) et la protéine spécifiée : l'ARN messager. Par **transcription**, la séquence d'ADN est reproduite dans une séquence d'ARN qui repose, elle aussi sur un système à 4 nucléotides.

La seconde étape est la **traduction** c'est-à-dire la réalisation d'une protéine spécifiée par l'ARN messager avec passage d'un alphabet à 4 lettres à un alphabet à 20 lettres.

*\*Remarque : Le premier flux d'information a déjà été étudié avec la **réplication** de l'ADN, il représente l'état «reproduction» de la cellule, le second flux représente l'aspect «métabolisme».*

	<b>U</b>	<b>C</b>	<b>A</b>	<b>G</b>	
<b>U</b>	UUU <b>Phe</b> UUC UUA UUG	UCU <b>Ser</b> UCC UCA UCG	UAU <b>Tyr</b> UAC UAA UAG	UGU <b>Cys</b> UGC UGA UGG <b>Trp</b>	U C A G
<b>C</b>	CUU <b>Leu</b> CUC CUA CUG	CCU <b>Pro</b> CCC CCA CCG	CAU <b>His</b> CAC CAA <b>Gln</b> CAG	CGU <b>Arg</b> CGC CGA CGG	U C A G
<b>A</b>	AUU <b>Ile</b> AUC AUA AUG <b>Met</b>	ACU <b>Thr</b> ACC ACA ACG	AAU <b>Asn</b> AAC AAA <b>Lys</b> AAG	AGU <b>Ser</b> AGC AGA <b>Arg</b> AGG	U C A G
<b>G</b>	GUU <b>Val</b> GUC GUA GUG	GCU <b>Ala</b> GCC GCA GCG	GAU <b>Asp</b> GAC GAA <b>Glu</b> GAG	GGU <b>Gly</b> GGC GGA GGG	U C A G