

1 ère partie

**Du génome aux phénotypes :
comment les caractères héréditaires sont-ils contrôlés ?**

ou

le gène est une unité de fonction qui peut muter.

Chapitre 3 : De manière semble-t-il contradictoire avec les mécanismes assurant la stabilité au cours de la réplication, on observe qu'un gène peut exister sous plusieurs formes dans les populations naturelles.

- 1. La diversité des caractères.**
- 2. Un exemple de la diversité des polypeptides.**
- 3. Un exemple de la diversité de l' ADN.**
- 4. Conclusions.**

Chapitre 3 : De manière semble-t-il contradictoire avec les mécanismes assurant la stabilité au cours de la réplication, on observe qu'un gène peut exister sous plusieurs formes alléliques dans les populations naturelles.

L'étude de la synthèse d'un polypeptide vient de nous permettre de définir l'unité génétique : **un gène contrôle la synthèse d'un ARN messager, un ARN messager détermine la séquence d'un polypeptide, un polypeptide remplit une fonction**

Nous reviendrons très bientôt sur cette chaîne d'événements qui va des gènes aux caractères (1).

Nous savons par ailleurs que les gènes sont doués de continuité grâce à la réplication semi conservative.

Tout cela paraît **inéluçtable et contradictoire** avec l'immense **diversité** des êtres vivants,

Dans ce chapitre, nous allons illustrer cette contradiction au niveau de l'espèce, groupe naturel dont nous avons donné la définition dans le chapitre 1 (note 2). Les différences entre espèces seront envisagées dans un autre chapitre.

On verra que ce chapitre est assez court : il a été séparé du chapitre suivant, afin de bien **distinguer** la simple **constatation** de la diversité génétique et les **causes** de cette diversité, que nous envisagerons dans le chapitre suivant.

1. La diversité des caractères

Faut-il le préciser : tous les hommes se ressemblent suffisamment pour que nous ayons le sentiment d'appartenir à la même espèce (encart 10). Pourtant, à l'exception des vrais jumeaux, nous présentons des différences évidentes avec notre voisin et même avec les membres de notre famille.

Cette diversité des caractères est la règle générale dans l'ensemble des espèces, quel que soit le caractère que l'on étudie.

Jusqu'au XIX ème siècle, on s'est borné à la constater, en établissant des **collections** sur la base des **apparences**, ce qui avait le mérite d'établir des **classifications** en rapprochant les espèces qui se ressemblent.

Nous allons maintenant examiner la diversité au niveau des macromolécules, ce qui nous donnera des renseignements bien plus solides (2).

Encart 10 : Il n'y a pas de races chez l'Homme.

En ce qui concerne l'espèce humaine, seuls quelques maniaques dangereux continuent à vouloir distinguer des races humaines, dans un but d'une hiérarchisation idéologique dont on connaît les conséquences (3). Personne, évidemment, ne nie que des êtres humains présentent des différences de couleur de peau : il s'agit d'un caractère d'adaptation au climat, fixé au cours de l'évolution. Par contre, il est **impossible de déterminer un ensemble de propriétés biologiques** et de caractères distinguant les noirs, les rouges, les jaunes et les blancs.

Classer par la couleur de la peau est à peu près aussi précis que de regrouper des automobiles et des tondeuses à gazon qui fonctionnent à l'essence, en les opposant à des bateaux et des camions marchant au diesel !!!

En réalité, il n'y a pas d'ensembles naturels comparables aux races des animaux domestiqués ou des plantes cultivées : les lignées pures n'existent pas dans la nature (chapitre 1 note 9). Ces races sont des ensembles artificiels. Pour les obtenir et les conserver, le sélectionneur ne conserve à chaque génération que les géniteurs qui l'intéressent. Il en résulte un **appauvrissement du polymorphisme génétique globalement mauvais pour l'espèce** (voir ce chapitre en 1.4 et plus loin (4)).

Pour l'Homme, « **tous parents, tous différents** » est l'excellent résumé des choses, titre d'une exposition mémorable, organisée par André Langaney au Musée de l'Homme, à Paris:

Si l'on utilise le mot race pour l'Homme, il faut donc savoir qu'il ne représente rien dans la nature, sur le plan biologique.

On peut même constater dans le « Petit Robert » que le mot, né à la fin du XV^{ème} puis utilisé à partir de 1684 pour qualifier des groupes humains est d'une remarquable imprécision. Le meilleur résumé de la situation est de dire avec Gaxotte, cité dans ce dictionnaire : « Il n'existe pas ...de race française, ni de race bretonne, ni de race aryenne, mais...une nation française, un peuple breton, des langues aryennes. »

(1) : attention, ne toujours pas faire l'égalité « un polypeptide / un caractère », précaution déjà réclamée dans le chapitre 1 et dans le chapitre 2 : si nous en parlons sans cesse, cela doit avoir une certaine importance !!

(2) : allez voir la Grande Galerie de l'Évolution, au Muséum National d'Histoire Naturelle, à Paris. Si vous découvrez la liste des membres du Comité de Suivi Scientifique, vous verrez que les généticiens étaient bien représentés dans ce groupe.

(3) : et il ne s'agit pas d'un « détail ».

(4) : un généticien digne de ce nom ne *peut pas* être raciste : indépendamment des aspects moraux, le racisme est une imbécillité biologique. Donc, quelque'un se réclamant de la génétique pour justifier le racisme est, au moins, un imbécile.

2 . Un exemple de la diversité des polypeptides.

Le transport de l'oxygène dans le sang est réalisé par une molécule complexe : l'hémoglobine. Chez l'adulte, la molécule active comporte en particulier quatre polypeptides identiques deux à deux (5) : les chaînes polypeptidiques appelées globines α et β . Les globines ont chacune une séquence particulière (α est faite de 141 acides aminés, β en comporte 146).

Toutes les molécules α d'une personne donnée ont la même séquence d'acides aminés. Il en est de même pour β .

En 1949, Pauling montre que, pour chacune de ces deux molécules, on trouve très généralement la même séquence chez la plupart des personnes en bonne santé.

On appelle cette forme majoritaire : **séquence de référence** (7).

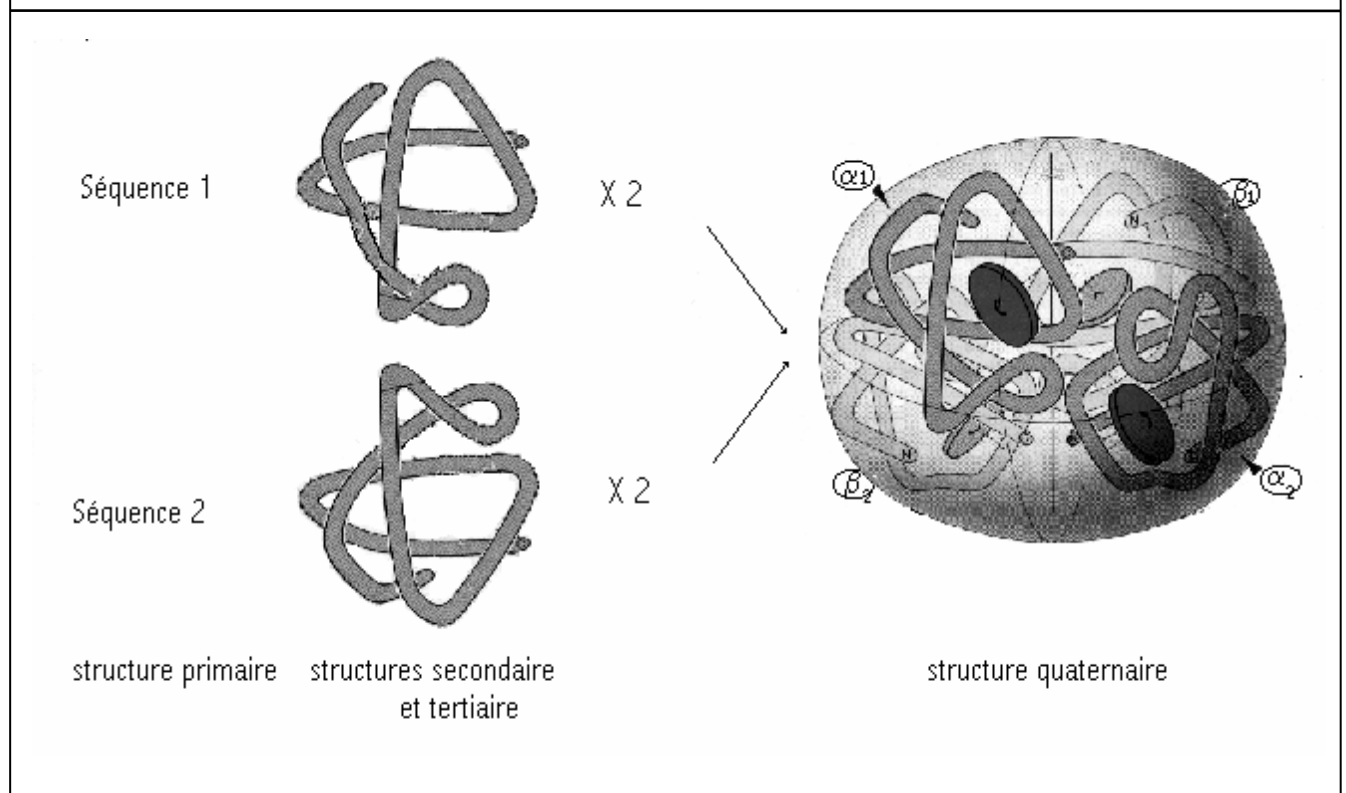
Pauling étudie ensuite l'hémoglobine d'un patient souffrant d'une **anémie héréditaire**. Il constate que les chaînes α sont strictement identiques à la séquence de référence. Par contre, Ingram découvre que la chaîne β du patient anémié se distingue de la séquence de référence par **un acide aminé**. En position 6 de la chaîne polypeptidique, le résidu est l'acide glutamique dans la forme de référence, tandis que l'hémoglobine β de la personne anémiée possède la valine à cette position.

Cette découverte est très importante : une différence héréditaire au niveau du caractère (bonne santé / anémie) est liée à une différence au niveau d'une unité d'action biochimique, d'un polypeptide donc.

D'autres patients anémiés sont étudiés. Certains présentent également des différences de la chaîne β , d'autres de la chaîne α : nous nous intéresserons seulement aux premiers par la suite.

Figure 24 : la molécule d'hémoglobine.

Les quatre disques sont des molécules organiques assurant la fixation transitoire de l'oxygène. La structure primaire est l'enchaînement des acides aminés d'un polypeptide. Les structures secondaires et tertiaires sont des repliements du polypeptide dans l'espace. La structure quaternaire est représentée ici par l'addition des quatre chaînes polypeptidiques. Cet ensemble est le seul édifice fonctionnel (6).



(5) : on parle d 'un **tétramère** : cela indique qu 'il y a quatre chaînes polypeptidiques , identiques ou différentes, comme c 'est le cas ici.

(6) : tiens !! Voilà un cas très net où un caractère (le transport de l 'oxygène) ne dépend pas d 'un mais de deux polypeptides ... Est-ce que cela ne commencerait pas à justifier les notes telles que la note 1 de ce chapitre ? !!

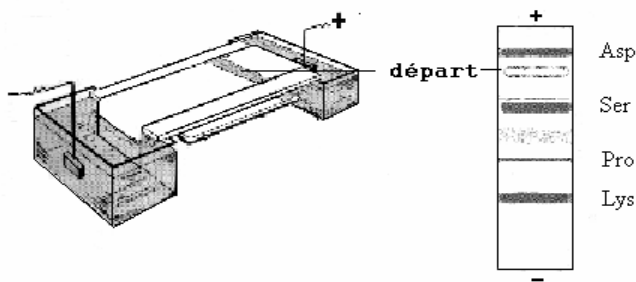
(7) : voir chapitre suivant .

Puis les chaînes β d'un grand nombre de personnes sont étudiées de manière systématique à la suite de simples prises de sang, grâce à une technique facile à mettre en oeuvre, l'**électrophorèse** (encart 11).

On découvre alors de nouvelles **différences** y compris chez des personnes ne présentant **aucun signe clinique**. Dans un bilan établi en 1990, 253 différences, par rapport à la séquence de référence sont enregistrées : on les appelle des **variants**.

La plupart du temps (234 fois), la différence par rapport à la séquence de référence concerne un seul acide aminé : on parle de **substitution**. Quelques unes sont illustrées dans la figure 25. D'autres cas montrent des bouleversements plus importants avec un ou quelques acides aminés en plus (**addition** : 8 fois) ou en moins (**délétion**: 11 fois). Nous n 'en parlerons pas dans la suite de ce chapitre.

Encart 11 : l 'électrophorèse (8). Lorsqu 'elles sont en présence d 'un solvant, les molécules biologiques migrent sur un papier ou sur un gel en fonction de caractéristiques physico-chimiques complexes, chacune ayant une vitesse de migration qui lui est propre. Cela conduit donc à leur séparation. Cette séparation est facilitée dans un champ électrique, car la migration est également fonction de leur charge positive ou négative. C 'est ainsi que l 'on peut facilement séparer les acides aminés. Bien entendu, cette technique est insuffisante pour connaître la séquence : d 'autres dispositifs sont nécessaires. Ils sont maintenant souvent automatisés.



Ici , on montre la séparation de 4 acides aminés. Le principe et le résultat sont les mêmes pour des protéines qui peuvent être séparées en fonction de leurs caractéristiques physico- chimiques.

Figure 25 : exemples de substitutions d 'acides aminés dans des variants.

Les positions 92 et 99 seront étudiées en détail plus loin.

Colonne 1 : Numéro du résidu d 'acide aminé modifié (Exemple 75).

Colonne 2 : Résidu de la forme de référence et résidu du variant (Exemple : *leu* dans la forme de référence, *pro* dans le variant).

Colonne 3 : Lieu (hôpital) ou le variant a été repéré (Exemple : *Atlanta*).

Colonne 4 : Caractéristiques de l 'hémoglobine du variant (Exemple : *instable*)

1	2	3	4	1	2	3	4	1	2	3	4
75	<i>leu-pro</i> <i>leu-arg</i>	<i>Atlanta</i> <i>Pasadena</i>	<i>instable</i> <i>instable</i>	85	<i>phe-ser</i>	<i>Bryn Mayr</i> <i>Buenos Aires</i>	<i>instable</i> <i>instable</i>	94	<i>asp-his</i> <i>asp-asn</i>	<i>Barcelona</i> <i>Bunbury</i>	<i>affinité 02</i> <i>affinité 02</i>
76	<i>ala-asp</i>	<i>Chicago</i>	normale (9)	87	<i>thr-lys</i>	<i>Ibadan</i>	<i>cf falciforme</i>	95	<i>lys-glu</i>	<i>Baltimore</i>	<i>affinité 02N1</i>
77	<i>his-asp</i>	<i>Iran</i>	normale (9)	88	<i>leu-arg</i>	<i>Boras</i>	<i>instable</i>		<i>lys-asn</i>	<i>Detroit</i>	<i>affinité 02N1</i>
78	<i>leu-arg</i>	<i>Quin Hai</i>	<i>pas étudiée</i>		<i>leu-pro</i>	<i>Santa Anna</i>	<i>instable</i>	96	<i>leu-val</i>	<i>Regina</i>	<i>affinité 02N1</i>
79	<i>asp-gly</i> <i>asp-tyr</i>	<i>Hsi-Tsou</i> <i>Tampa</i>	<i>affinité 02</i> <i>affinité 02</i>	89	<i>ser-asn</i>	<i>Créteil</i>	<i>affinité 02</i>	97	<i>his-gln</i> <i>his-leu</i>	<i>Malmö</i> <i>Wood</i>	<i>affinité 02</i> <i>affinité 02</i>
80	<i>asn-lys</i>	<i>Szuhu</i>	<i>affinité 02N1</i>	90	<i>glu-lys</i>	<i>Anenogi</i>	<i>affinité 02</i>		<i>his-pro</i>	<i>Nagoya</i>	<i>instable</i>
81	<i>leu-arg</i>	<i>Baylor</i>	<i>instable</i>	91	<i>leu-pro</i>	<i>Sabine</i>	<i>instable</i>	98	<i>val-met</i>	<i>Köln</i>	<i>instable</i>
82	<i>lys-asn</i> <i>lys-asp</i> <i>lys-thr</i> <i>lys-met</i>	<i>Providence</i> <i>Providence</i> <i>Rahere</i> <i>Helsinski</i>	<i>affinité 02</i> <i>affinité 02</i> <i>affinité 02</i> <i>affinité 02</i>	92	<i>his-tyr</i> <i>his-gln</i>	<i>Hyde Park</i> <i>St Etienne</i>	<i>affinité fer</i> <i>instable</i>		<i>val-gly</i> <i>val-ala</i>	<i>Nottingham</i> <i>Djelfa</i>	<i>instable</i> <i>instable</i>
					<i>his-asp</i>	<i>Istamboul</i> <i>Altgeld Gardens</i>	<i>instable</i> <i>affinité 02N1</i>	99	<i>asp-asn</i> <i>asp-his</i>	<i>Kempsey</i> <i>Yakima</i>	<i>affinité 02</i> <i>affinité 02</i>
83	<i>gly-cys</i> <i>gly-asp</i>	<i>Ta-Li</i> <i>Pyrgos</i>	<i>polymérisation</i> <i>affinité 02</i>		<i>his-arg</i> <i>his-pro</i>	<i>Mozhaisk</i> <i>Newcastle</i>	<i>instable</i> <i>instable</i>		<i>asp-ala</i> <i>asp-tyr</i>	<i>Radcliffe</i> <i>Ypsilanti</i>	<i>affinité 02</i> <i>affinité 02</i>
84	<i>lys-glu</i>	<i>Tubingen</i>	<i>instable</i>	93	<i>cys-arg</i>	<i>Okazaki</i>	<i>affinité 02</i>		<i>asp-gly</i> <i>asp-val</i>	<i>Hôtel Dieu</i> <i>Chemilly</i>	<i>affinité 02</i> <i>affinité 02</i>

(8) : les figures 21 et 22 ont été remaniées de Durand , Hermann 1967. Nous rendons ainsi également hommage à cet auteur , qui a participé à la refonte complète des enseignements de premier cycle , en 1967 , avec Prévost , (chapitre 2 , note 5).

(9) : on remarque qu ' une hémoglobine portant une différence d 'acide aminé sur sa chaîne β peut avoir une activité semblable à celle de la forme de référence. Voir encart 14.

En ce qui concerne les substitutions, la première constatation que l'on peut faire est que les différences affectent presque toutes les positions (encart 12).

La deuxième constatation est que l'on peut trouver **différents** acides aminés à la place de l'acide aminé habituel si l'on examine l'ensemble des variants. La troisième est que le **nombre** d'acides aminés différents est cependant **limité** lorsqu'on observe une position particulière (par exemple 6 en position 99).

Examinons en détail les positions 92 et 99.

Dans le premier cas, on observe que l'histidine habituel est remplacé par 5 acides aminés différents dans 5 hémoglobines isolées dans différents hôpitaux. Dans le deuxième cas, l'acide aspartique habituel est remplacé par 6 acides aminés différents chez les variants.

Si l'on prend en compte le code génétique, on obtient le tableau de la figure 27. La dégénérescence du code complique un peu les analyses : dans le cas de la position 92, on ignore si le codon de référence est CAU ou CAC puisque ces deux codons sont synonymes et correspondent à l'histidine. De même, l'acide aspartique en 99 de la référence peut être codé soit par GAU soit par GAC (10).

Faisons l'**hypothèse** la plus simple : chacune des bases de l'ARN messager pouvant exister sous 4 états correspondants aux 4 bases, **le plus petit changement possible** entre le messager de référence et le messager d'un variant correspond au changement d'**un seul nucléotide**. Ce remplacement peut se produire au niveau de chacune des trois bases d'un codon, unité de changement potentiel d'un seul acide aminé.

Si cette hypothèse est exacte, le codon qui correspond à l'acide aminé 92 peut donner à la suite d'un seul changement un des codons correspondant à 7 acides aminés (et seulement 7, à cause de la dégénérescence, que le codon de référence soit CAC ou CAU).

Le résultat expérimental est cohérent avec ce résultat théorique, puisque 5 acides aminés seulement sont observés. Plus spectaculaire encore que cette proximité quantitative entre résultat attendu dans le cadre de l'hypothèse et résultats expérimentaux, on observe que seuls les acides aminés « possibles » dans le cadre de l'hypothèse existent dans les séquences de variants.

Et le même type de résultats est observé pour la position 99 !

Comme ce jeu « marche » pour toutes les substitutions observées, on en conclut que notre hypothèse est très hautement probable.

Lorsque les polypeptides de deux individus se distinguent par un seul acide aminé, la différence peut être due au changement d'un seul nucléotide dans les ARN messagers correspondants

(10) : Bien sûr, la séquence de l'ARN messager ne comporte qu'un seul des 2 codons possibles, ce qui correspond à un seul triplet de l'ADN. Mais ici, nous ne savons pas lequel ...

Figure 26 : Différences chez des variants, par rapport à la forme de référence β .

Pour un résidu donné, on observe de 0 à 6 résidus différents de la référence. En abscisse : numéros des résidus, de 1 à 146. En ordonnée : nombre de résidus différents du résidu de référence, observés dans l'ensemble des variants.

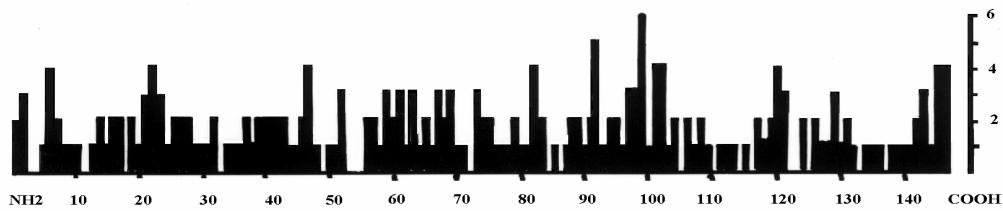


Figure 27 : codons et acides aminés en positions 92 et 99 de la chaîne β .

On fait l'hypothèse qu'un changement affecte une seule base, en petit caractère. Les changements d'acides aminés réellement observés sont soulignés, dans les bilans.

92	base 1	base 2	base 3	bilan
CAC	uAC tyr	CuC leu	CAu his	his ↓
	aAC asn	CcC pro	CAa gln	
	gAC asp	CgC arg	CAG gln	
his				tyr asn asp
CAU	uAU tyr	CuU leu	CAC his	<u>leu</u> <u>pro</u> <u>arg</u> <u>gln</u>
	aAU asn	CcU pro	CAa gln	
	gAU asp	CgU arg	CAG gln	
99	base 1	base 2	base 3	bilan
GAU	uAU tyr	GuU val	GAC asp	asp ↓
	cAU his	GcU ala	GAA glu	
	aAU asn	GgU val	GAC asp	
asp				<u>tyr</u> <u>his</u> <u>asn</u>
GAC	uAC tyr	GuC val	GAu asp	<u>val</u> <u>ala</u> <u>gly</u> glu
	cAC tyr	GuC val	GAu glu	
	aAC asn	GgC gly	GAG glu	

Encart 12 : L'absence de certains acides aminés dans la collection de variants peut correspondre à deux situations : soit le nombre de séquences étudiées est insuffisant (un jour ou l'autre, on isolera les variants qui « manquent »), soit ces variants ne peuvent pas exister. Cette deuxième hypothèse est simple à comprendre, si l'on se réfère à la structure fonctionnelle d'un polypeptide. Cela a déjà été rapidement évoqué dans la note 9 : certains acides aminés sont strictement indispensables à certaines positions, car ils conditionnent très fortement la fonction (c'est le site actif, constellation d'acides aminés proches dans l'espace - et pas forcément contigus dans la séquence). Par contre le reste des positions peut admettre une certaine **souplesse**, à la condition que les différences d'acides aminés permettent les repliements complexes de l'édifice sous sa forme tertiaire (et ici quaternaire, voir figure 24). En gros, on peut dire que certains acides aminés sont des **synonymes biochimiques**. Cette souplesse peut être plus ou moins grande : elle est importante en positions 92 et 99 de l'hémoglobine β . Cependant, elle n'est pas totale : si un changement existe au niveau du messager et qu'il correspond à un acide aminé non synonyme, l'hémoglobine ne sera pas fonctionnelle et l'individu ne pourra pas vivre, d'où l'impossibilité manifeste ... de l'étudier !!

On se souviendra, à l'inverse, qu'un changement d'acide aminé peut ne pas affecter l'activité d'un polypeptide (exemple, note 9).

D'où provient cette différence au niveau de l'ARN messenger ? Puisque ce dernier provient de la transcription, il est facile d'imaginer que cela est dû à une différence dans l'ADN en cause, au niveau du gène responsable de la synthèse de cet ARN messenger. Pour des raisons variées, cette analyse n'a pas été faite de manière systématique sur le gène de l'hémoglobine : nul doute que les résultats seraient semblables à ceux que nous allons obtenir en examinant un exemple qui est rapidement devenu classique.

3. Un exemple de la diversité au niveau de ADN.

La connaissance directe du gène est évidemment la meilleure manière permettant de transformer nos hypothèses en conclusions.

Dans la deuxième partie de cet ouvrage nous verrons comment il est devenu possible d'isoler un gène, depuis le premier succès enregistré en 1976 (11) puis de le séquencer.

Dans ces vingt dernières années, bien des progrès ont été réalisés. Mais l'isolement d'un gène reste souvent laborieux. Cependant, une fois l'isolement réussi, les difficultés ont leur récompense : la connaissance directe de la séquence des gènes.

C'est ainsi que l'on a isolé un gène dont la mutation peut entraîner une maladie grave, la *mucoviscidose*, lorsque le polypeptide CFTR n'est pas fonctionnel (12)

L'étude du **gène** de 16 **malades** (figure 28) a montré 16 **formes différentes**, dont 12 sont des **substitutions** de nucléotides conduisant à de simples substitutions d'acides aminés, comme celles que nous avons analysées dans le cas de la chaîne β de l'hémoglobine ou à la création d'un codon stop, directement ou indirectement. Les 4 autres formes sont des **additions** ou des **délétions**.

Toute cette diversité des modifications du gène n'empêche pas que le résultat soit le même. Malgré leur diversité moléculaire, les 16 séquences différentes de la référence, produisent le même résultat : les personnes étudiées sont atteintes par la maladie.

Figure 28: 16 changements affectant le gène contrôlant le polypeptide CFTR.

Les modifications sont indiquées au niveau de l'ADN pour le brin non codant. Cela permet de lire ces résultats en termes de codons en remplaçant simplement T par U (13).

Type de mutation	malade	modification du gène	modification du polypeptide
substitutions conduisant à un changement d'acide aminé	RP AE SR GD NK	G-->A (nucl. 1172) C-->A (nucl. 1496) T-->G (nucl. 1779) G-->A (nucl. 1784) C-->G (nucl. 4041)	arg-->pro (ac. aminé 347) ala-->glu (ac. aminé 455) ser-->arg (ac. Aminé 549) gly-->asp (ac. Aminé 551) asn-->lys (ac. Aminé 1303)
substitutions produisant un « stop »	YX GX RX	T-->A (nucl. 498) G-->T (nucl. 1376) C-->T (nucl. 1756)	tyr--> stop polypeptide gly--> stop plus ou moins arg--> stop raccourci (voir encart 13)
substitutions qui conduisent à un message anormal	MA MB MC	G-->T (nucl. 712) G-->A (nucl. 2794) A-->G (nucl. 3853)	le produit de transcription est mal épissé. (voir encart 14)
délétion ou addition changeant la phase	1078 3905	nucl. 1078 en moins T entre nucl.3905/3906	stop dans la nouvelle phase polypeptide raccourci
délétions de trois nucléotides	F 508 1507	nucl. 1652/3/4 en moins nucl.1649/50/51 en moins	en absence de phe 508 (14) en absence de ile 507 (14)

Encart 13 : il n'y a pas de stop dans la phase codant pour le polypeptide (chapitre 2). Ceci est dû à la nature même du message, qui doit pouvoir être lu. Par contre, les deux autres phases, non traduites, peuvent en contenir, sans gêner la traduction. En première approximation, la probabilité d'un codon stop dans les deux phases non ouvertes est de 3 / 64, soit le nombre de codons stop -3- divisé par le nombre de codons possibles -64-. Comme un polypeptide est en général fait d'au moins 100 acides aminés, il lui correspond un ARN m d'au moins 100 codons, et deux phases non traduites, contenant en moyenne 5 codons stop chacune. La création par mutation d'un codon stop provoque une terminaison de la traduction avant la synthèse du polypeptide entier. Il y a synthèse d'un fragment de polypeptide de séquence normale jusqu'à l'arrêt non programmé dans la séquence de référence. L'addition ou la délétion d'une base modifiée décale la lecture. Un stop peut exister dans la nouvelle phase ce qui interrompt la traduction.

(11) : les premiers gènes que l'on a pu isoler sont ceux des chaînes α et β , nous verrons pourquoi dans la 3^{ème} partie.

(12) : il s'agit d'un gène du chromosome 7, long de 230 kb, avec 27 exons (encart 13). Il code pour un polypeptide responsable du contrôle de la conductibilité membranaire. Il est désigné ici par les initiales de sa dénomination anglo-saxonne.

(13) : admettons qu'un fragment d'ADN soit ...AATTGTC...
et que le brin du haut soit le brin codant.TTAACAG... Il donnera naissance à un messager
....UUAACAG ... semblable au brin du bas, à U / T près.

(14) : Une délétion ou une addition peut concerner plusieurs nucléotides : un gène peut s'en trouver très allongé ou très raccourci. Dans ce dernier cas, le gène peut même être complètement absent. Les deux délétions présentées ici ont la caractéristique de correspondre à une perte de trois nucléotides, ce qui ne modifie pas la phase. Cependant, le polypeptide est raccourci d'un acide aminé... et cette différence le rend inactif même si tous les autres acides aminés sont identiques à ceux de la forme de référence.

4. Conclusions.

Nous avons vu précédemment que la **diversité des caractères gouvernés par les gènes peut être due au milieu**. C'est par exemple le cas des plantes cultivées en plaine ou en montagne que nous avons rencontrées dans le chapitre 1. Cette variabilité est due à des phénomènes de régulation : **l'expression des gènes** (transcription + traduction) peut être modulée en fonction des conditions de **milieu**. Nous étudierons ces phénomènes à la fin de l'ouvrage.

La cause de la **variabilité des caractères** que nous avons étudiée ici est complètement **différente** car elle est **directement liée à la structure même du gène**.

Le schéma ci contre (figure 29) très simple, presque naïf (17), résume ces deux possibilités de variabilités des caractères, toutes deux liées à des propriétés des gènes.

Nous avons étudié dans ce chapitre, les modifications de la structure des gènes de manière indirecte ou directe.

L'analyse des polypeptides nous a permis d'émettre des hypothèses, indirectes mais très solides sur la structure des gènes grâce à des déductions possibles à l'aide du code génétique.

Nous avons aussi examiné directement des modifications de la séquence d'un gène, ce qui suppose qu'il ait préalablement été isolé (18).

L'idée principale que nous devons déduire de ces observations est que les choses ne sont pas simples, même si l'on se contente d'étudier un caractère assez bien défini.

Par exemple, nous avons vu qu'une anémie peut être due à des causes biochimiques diverses : parmi ces causes nous avons seulement étudié le transport de l'oxygène. Dans ce sous ensemble, nous avons encore distingué des anémies dues à des modifications de la chaîne α de celles liées à des défauts de la chaîne β de l'hémoglobine.

Tout cela signifie qu'une analyse génétique qui se contenterait d'observer les caractères serait largement insuffisante car elle ne percevrait pas toute cette complexité.

Il est, dès à présent, manifeste **qu'un caractère** (« ne pas être anémié ») peut impliquer que de **nombreux gènes** soient fonctionnels et **qu'une différence de caractère** (« être anémié ») peut être due à des **modifications génétiques diverses**.

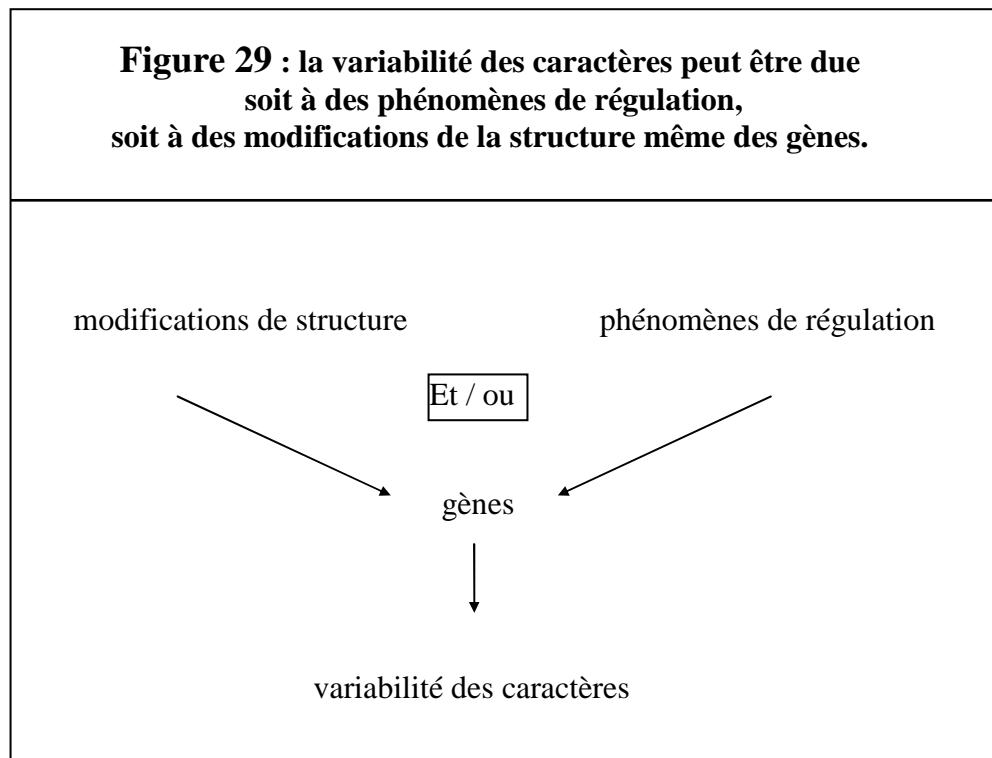
Avant de revenir à l'analyse de cette complexité, nous allons tout d'abord, dans le prochain chapitre, nous préoccuper de l'origine de la diversité génétique que nous venons de constater ici.

Encart 14 : Introns et exons : les gènes mosaïques.

Chez les eucaryotes, le produit de transcription d'un gène subit une **maturation** : il y a élimination de portions internes du polynucléotide. Ces parties correspondant à des zones non traduites du gène sont appelées **introns**. Les enzymes de maturation reconnaissent très précisément certaines séquences, de quelques nucléotides, situées aux frontières exons / introns et coupent l'ARN entre deux nucléotides précis de la séquence reconnue. Puis la continuité du polynucléotide est rétablie. Les parties du gène qui seront finalement traduites sont nommées **exons**. Les gènes ayant une telle structure sont appelés gènes mosaïques ou morcelés.

Lorsque toutes les enzymes de maturation ont pu agir, on obtient le véritable ARN messager qui est ensuite traduit en polypeptide. Par contre si une séquence de reconnaissance et d'action d'un enzyme de maturation est modifiée à la suite d'une mutation, une maturation peut ne plus se faire. L'ARN obtenu est alors très différent de celui de référence. Il peut être parfois traduit mais il ne donne de toutes façons pas de polypeptide fonctionnel : non seulement celui-ci peut posséder quelques acides aminés supplémentaires, mais de plus il peut se produire un décalage du cadre de lecture et il peut exister un codon stop dans la nouvelle phase (14).

Figure 29 : la variabilité des caractères peut être due soit à des phénomènes de régulation, soit à des modifications de la structure même des gènes.



(17) : voir note 25 du chapitre précédent !

(18) : voir la troisième partie

