

1. Origine et développements des biotechnologies

Bref historique

On distingue trois étapes dans l'histoire des « biotechnologies

1° Du Néolithique au début du 20ème siècle

- utilisation des bactéries, levures, moisissures (aliments, boissons, textiles)

- standardisation des procédés de fermentation

1650 : procédé d'Orléans pour la fabrication du vinaigre

1664 : création d'une bière alsacienne réputée.

1890 : Louis Pasteur et Robert Koch développent les premiers vaccins.

2° Des années 1920 aux années 1970

- industrie des antibiotiques, des vitamines, etc ...

1927 : Alexander Flemming découvre la pénicilline

1953 : l'ADN est le support des gènes

3° Depuis le début des années 1970

- progrès en génétique, biologie cellulaire, immunologie

- « maîtrise » du génome : clonage moléculaire

1972 : Stanley Cohen et Georges Köhler : débuts du « génie génétique »

2002 : Annonce du décodage complet du génome humain

L'identification moléculaire de la structure d'un gène (procaryote ou eucaryote) se fait en établissant la séquence nucléotidique de l'unité de transcription, et en caractérisant les séquences régulatrices qui en permettent l'expression. Ceci se fait en utilisant les techniques de biologie moléculaire. L'identification fonctionnelle d'un gène se fait par l'utilisation de la génétique (*in vivo*), et/ou par l'étude de son expression *in vitro* dans des cellules en culture.

Identifier un gène, c'est aussi connaître ses produits d'expression : ARN messager et/ou protéine. Savoir dans quel organe le gène est exprimé, ou si son expression est induite par une hormone, etc... Ces questions peuvent être abordées par les biotechnologies.

1-1. Définition du gène

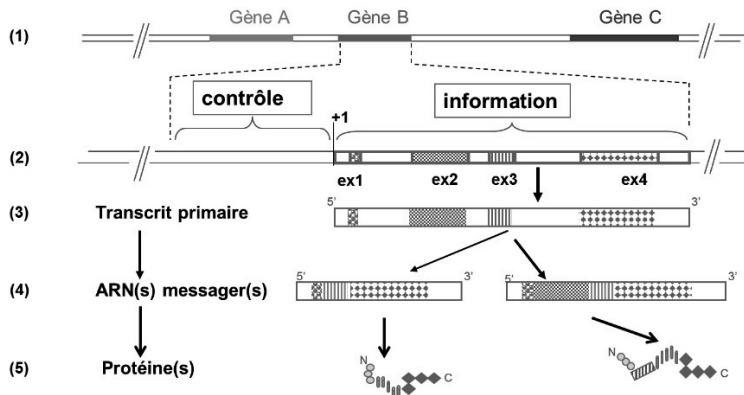
C'est une notion qui provient de la **génétique traditionnelle**

"Le gène est une unité fonctionnelle sur le génome".

Ce terme a été donné par Johannsen en 1909 pour désigner l'unité d'hérédité qui détermine le développement d'un caractère particulier. Puis Benzer (1957) l'a utilisé pour définir l'unité de fonction, pour laquelle deux mutations alléliques différentes présentent une non-complémentation absolue.

Pour la biologie moléculaire : un gène est une unité de transcription (dont toute la séquence est transcrite en ARN) ET les séquences régulatrices qui permettent cette transcription dans les cellules prévues, au moment requis, et en quantité suffisante. (Figure 1-1a). Généralement, le produit fonctionnel est une protéine, traduite à partir de l'ARN messager.

Figure 1-1a



Dans le cas présenté, le gène B est arbitrairement placé plus près du gène A que du gène C sur le chromosome (1). Le gène B est constitué au minimum d'une région « contrôle » en amont d'une région « information », qui est l'unité de transcription (2). Le transcrit primaire (3) subit une maturation soit dans la voie de droite (flèche épaisse, dans laquelle le transcrit de l'exon 3 (ex3) est conservé dans l'ARN messager (4), soit dans la voie de gauche, dans laquelle l'ARN messager ne contient pas le transcrit de l'exon 3. Il existe deux variants protéiques (5) produits d'expression de ce gène.

La plupart de ces séquences "contrôle", régulatrices, sont proximales, c'est à dire qu'elles sont situées entre 10 et 6 000 paires de bases en amont de l'unité de transcription. Mais de très nombreux gènes ont des séquences régulatrices additionnelles situées dans l'unité de transcription, ou en amont, ou en aval. Certaines séquences régulatrices peuvent être très distales, distantes de plusieurs milliers de paires de bases, d'autres peuvent être communes à plusieurs gènes.

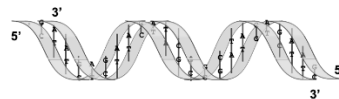
Rappels succins de biologie moléculaire (Figure 1-1b)

L'ADN est une molécule bicaténaire, dont les deux brins d'orientation antiparallèle sont liés par des liaisons hydrogène (1). Ce sont les ADN-polymérase ADN-dépendantes qui assurent la réplication de l'ADN génomique, initiée à partir d'amorces (représentées en traits épais), et utilisant des dNTP (désoxynucléotides-triphosphates) comme substrats; les brins néosynthétisés sont figurés en tirets (2).

Figure 1-1b Rappel de notions fondamentales

(1) L'ADN est une hélice bicaténaire

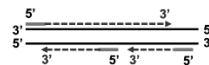
dont les brins de polarité antiparallèle sont liés par des liaisons hydrogène.



(2) Réplication de l'ADN

- bidirectionnelle : sur les 2 brins de l'ADN
- nécessite des amorces
- substrats : les 4 dNTP
- enzyme : ADN-polymérase ADN-dépendante.

La synthèse progresse de 5' vers 3' sur le brin néosynthétisé.



(3) Transcription de l'ADN en ARN

- le brin matriciel est parcouru de 3' vers 5'
- nécessité d'un promoteur
- substrats : les NTP
- enzyme : ARN-polymérase ADN-dépendante.

La synthèse progresse de 5' vers 3' sur le brin d'ARN.



Pour la transcription, les ARN-polymérase ADN-dépendantes reconnaissent le promoteur situé en amont de l'unité de transcription, et parcourent le brin matriciel de l'ADN génomique de 3' en 5' et le transcrivent en ARN pré-messager, en utilisant les NTP comme substrats ; la séquence transcrite est figurée par une flèche en trait pointillé (3). La séquence

transcrite (séquence « informative »), est comprise entre le premier nucléotide transcrit (noté traditionnellement +1) et le dernier nucléotide transcrit (environ 20 à 30 nucléotides en aval du signal de polyadénylation).

La séquence codante d'un gène est comprise entre les signaux d'initiation et de terminaison de traduction, qui sont le(s) codon(s) d'initiation (AUG) et de terminaison (UAA, ou UAG ou UGA) de synthèse protéique. C'est la séquence du brin codant qui est donnée, dans le sens conventionnel de 5' en 3'. Dans de nombreux gènes, il existe des séquences transcrites mais non traduites (exons ou parties d'exons) en amont et/ou en aval des séquences codantes. Ces séquences peuvent être reconnues par des protéines particulières et contribuer à l'expression correcte du gène par différents types de fonctions : exemple : ancrer l'ARN dans une région de la cellule (ex : l'ARN *bicoid* est localisé au pôle apical de l'embryon de drosophile grâce à sa séquence 3' non-traduite, sur laquelle se fixent des protéines.

Remarque : dans de nombreuses banques de données informatiques, c'est la séquence codante seule qui est donnée comme début +1 du gène, par ignorance du nucléotide d'initiation de transcription, et facilité de reconnaissance informatique inter-espèces, du fait de la conservation du code génétique.

Rappel : on sait depuis les travaux de Marshall Nirenberg et de Gobind Khorana (prix Nobel en 1968), que l'enchaînement des bases sur un brin d'ADN renferme à lui seul l'information qui spécifie la séquence des acides aminés dans les chaînes polypeptidiques et que le code génétique fait correspondre au moins un codon de trois bases à un acide aminé donné.

1-2. Les Biotechnologies

- 1) Elles servent à élargir la connaissance générale du fonctionnement des êtres vivants. Nous aurons l'occasion de voir que ces techniques ont permis de montrer que bien des mécanismes moléculaires sont conservés dans des espèces animales aussi différentes que la mouche du fruit (*Drosophila melanogaster*) et l'homme (*Homo sapiens*). L'étude de modèles simples (ver nématode, drosophile, poulet), permet d'aborder l'étude des organismes plus complexes : l'observation et l'expérimentation sur des organismes au développement embryonnaire extérieur au corps de la mère (œuf), a permis d'identifier un certain nombre des acteurs moléculaires du développement embryonnaire chez ces organismes. Les gènes identifiés ont servi d'hameçons moléculaires

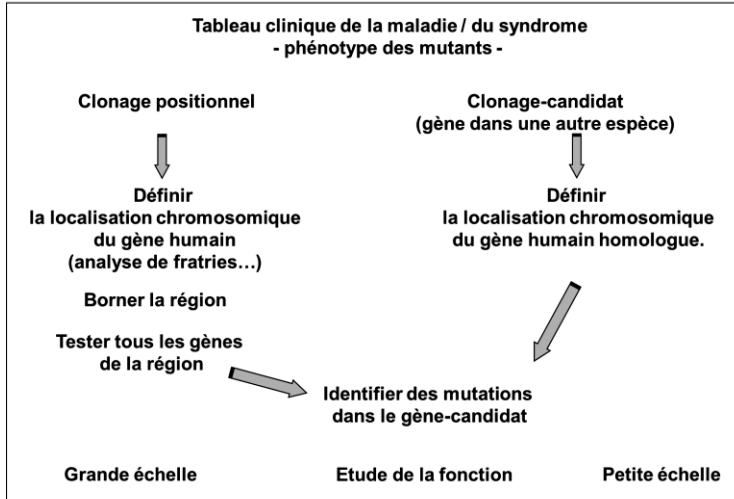
pour identifier les gènes équivalents, puis d'autres, chez les organismes plus complexes.

- 2) Elles permettent d'établir une "carte d'identité moléculaire" des espèces animales. Mais aussi à identifier des individus. Les applications pratiques en sont : recherche de paternité, identification de victimes d'accidents, ou de coupables d'actes criminels.
- 3) Elles servent à définir quelles sont les bases moléculaires de maladies humaines, et d'envisager des stratégies thérapeutiques. Exemple : traitement de certaines pathologies par l'utilisation de virus-recombinants, qui peuvent apporter un gène fonctionnel de façon ciblée, pour pallier à la dysfonction du gène résident du patient.
- 4) Elles ont des applications économiques, commerciales, industrielles. Les biotechnologies fournissent depuis quelques années des molécules d'intérêt pharmaceutique (insuline), ou permettent de créer des animaux-modèles de maladies humaines, afin de trouver des thérapies adaptées. Mentionnons des applications agricoles (obtention de plantes transgéniques). Les principes et techniques d'identification moléculaires sont similaires ou ceux qui sont appliqués pour les études des espèces animales. Dans ce fascicule, nous limiterons toutefois nos exemples aux applications de biologie animale.

Plus précisément orientées vers les disciplines médicales ou les applications cliniques ou thérapeutiques, les biotechnologies vont permettre d'identifier les gènes dont la dysfonction induit une pathologie. A ceci, deux types d'approches sont couramment utilisées (Figure 1-2) :

Chez l'homme, c'est essentiellement le clonage positionnel qui a permis l'identification des gènes dont les mutations sont à l'origine de pathologies. Mais l'approche de gènes-candidats a été également fructueuse, en particulier lorsque des animaux-modèles « construits » pour des études de recherche fondamentale ont développé des anomalies morphologiques ou métaboliques qui miment des pathologies humaines.

Figure 1-2



Avec les immenses avancées de programmes informatiques dédiés, un ensemble d'automates, d'imprimantes 3D, mais aussi divers traitements de milliers de données enregistrées ont été développés. Toutefois, le recueil des informations biologiques repose toujours sur les techniques de base exposées dans ce fascicule.

1-3. Limites des techniques traditionnelles

Pour identifier un gène, il faut isoler physiquement le segment d'ADN correspondant, à l'état pur, en quantité pondérale suffisante pour pouvoir établir sa séquence nucléotidique. Cette entreprise est délicate, et parfois impossible directement, à partir de fragments du génome, en particulier du fait de la taille importante des génomes eucaryotes, donc du très grand nombre de fragments génomiques.

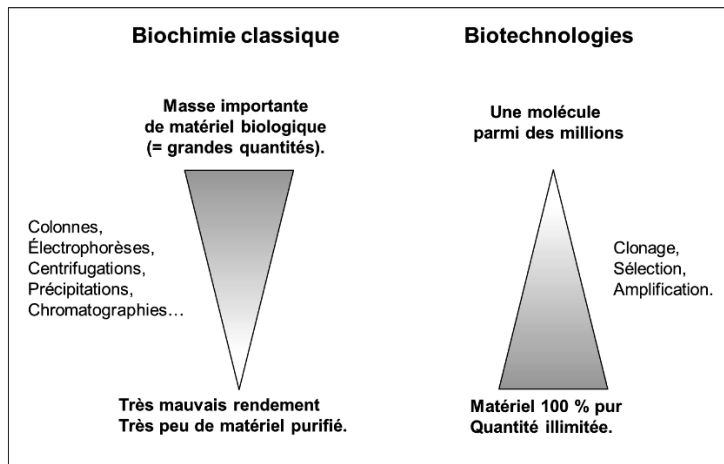
Il est toutefois possible de contourner le problème en isolant d'abord les produits d'expression du gène (ARNm ou protéine), qui serviront d'"hameçons" à partir desquels sera isolé le gène lui-même. A condition que ces produits existent en grande quantité dans un organe donné, ou à un moment précis.

Les méthodes de biochimie classique telles que les centrifugations, chromatographies, électrophorèses, précipitations sélectives, etc..., permettent l'extraction et le tri des macromolécules biologiques (ADN, ARN, protéines), en fonction de leur nature, de certaines de leurs propriétés, ou de leur taille. Ces méthodes sont fines et fiables. Elles avaient permis l'étude et la caractérisation des gènes des procaryotes : en 1969, Jonathan Beckwith et Jack Shapiro réussissent à isoler un gène d'*Escherichia coli*.

Mais lorsqu'elles sont utilisées à des fins préparatives, ces méthodes présentent deux inconvénients majeurs :

1) le rendement final de la fraction enrichie est faible : il est souvent nécessaire d'utiliser différentes techniques en cascade, qui mettent en jeu différentes propriétés physiques ou chimiques des molécules à fractionner. Avec à chaque étape des pertes de matériel, et donc un rendement final faible (Figure 1-3).

Figure 1-3



Par exemple : à partir d'un broyat de cellules, si l'on veut préparer de l'ADN, on utilisera successivement une précipitation sélective (qui permet d'éliminer la majorité des protéines et de garder les acides nucléiques), puis un gradient de densité à l'équilibre, qui permet de séparer les ADN des ARN par leur densité, et enfin une électrophorèse, qui fractionne les ADN en fonction de leur poids moléculaire en utilisant leur charge électrique.

2) ces techniques permettent d'enrichir une fraction en une population de molécules de même nature. *Par exemple* : on obtient 90% de fragments d'ADN de 3 000 paires de bases. Mais par ces techniques, il est impossible d'obtenir une population pure à 100% de molécules de même nature. De plus, elles nécessitent des équipements lourds, du personnel et du temps d'expérimentation.

La puissance des biotechnologies explique leur formidable développement et leurs applications : en partant de très peu de matériel biologique, les outils et techniques développés permettent d'obtenir des produits purs à 100 %, et en quantité illimitée.

1-4. Idée, Arme et Méthode des biotechnologies

Il était techniquement impossible, jusqu'en 1970, d'isoler et d'analyser un gène d'eucaryote. Cette impossibilité tenait à la disproportion qui existe entre la longueur d'un gène moyen, estimée à seulement quelques milliers de paires de bases, et celle du génome, estimée pour l'homme à 3 milliards de paires de bases. Un gène unique devrait donc représenter en moyenne un millionième (10^{-6}) de l'ensemble de l'ADN humain.

Les techniques biochimiques classiques ne pouvaient NI réaliser la PURIFICATION d'un fragment particulier parmi un million d'autres, NI permettre l'enrichissement d'un fragment en QUANTITE PONDERALE SUFFISANTE pour en faire l'analyse.

L'étude des bactéries a fourni les outils qui ouvrirent la voie aux biotechnologies. Ces outils sont les enzymes de restriction et les plasmides bactériens.

Les bactéries ont été asservies et utilisées pour les biotechnologies.

Ce que la biochimie classique ne pouvait pas réaliser est maintenant obtenu par l'utilisation des biotechnologies. En maîtrisant le clonage des fragments d'ADN, par l'utilisation des bactéries et des outils moléculaires qu'elles fournissent, il est possible d'obtenir de façon routinière la quantité désirée de fragment d'ADN, « moléculairement » pure. "Les bactéries arrivent à faire le travail que les biochimistes n'arrivaient pas à réaliser".

Les travaux de Werner Arber, Hamilton Smith et Daniel Nathans (Prix Nobel de Chimie en 1978) ont signé le début des biotechnologies.