

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de  
La Recherche Scientifique  
Université d'Alger  
Faculté de Médecine.

## **INTRODUCTION A LA GENETIQUE MOLECULAIRE**

**DR MAKRI-MOKRANE Samira , PR TAZIR Meriem**  
**Département de Médecine**

**ANNEE 2006**

# INTRODUCTION A LA GENETIQUE MOLECULAIRE

DR MAKRI-MOKRANE Samira , PR TAZIR Meriem

## Introduction

La pratique neurologique a été modifiée par les progrès de la génétique moléculaire ; le nombre des gènes identifiés dans les pathologies héréditaires neurologique est impressionnant ainsi que la liste des mutations.

Le diagnostic génétique moléculaire est tributaire de:

- Un diagnostic clinique correct (examen physique, examens complémentaires)
- La transmission héréditaire : arbre généalogique
- Prélèvement sanguin qui permet l'extraction d'ADN (technique d'extraction au phénol-chloroforme).

## I) Structure et fonction des gènes.

La compréhension de la vraie nature des maladies génétiques nécessite une connaissance détaillée de la structure et de la fonction des gènes humains.

### a) Structure des gènes

Le génome (séquence complète d'ADN) humain est composé de 3,3 milliards de paires de bases, le nombre des gènes est estimé à environ 60.000.

La taille des gènes est très variable, le plus petit gène comprenant quelques centaines de paires de bases, alors que le plus grand peut avoir plus d'un million de paires de bases de longueur ; un kilo base (1 kb) équivaut à 1000 paires de bases (pb).

Les **introns** sont des segments d'ADN localisés au sein du gène, transcrit en ARN, puis **épissés** avant que l'ARN messager ne soit traduit en protéine.

Les **exons** sont les régions du gène qui contiennent l'information codante.

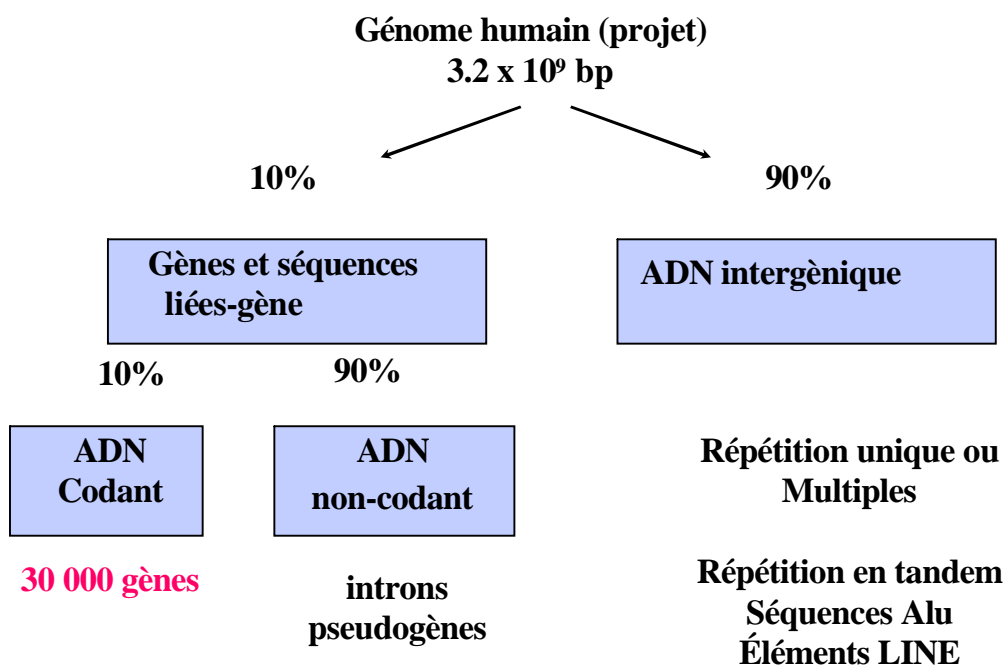
La **région 5'** et la **région 3'**, situées aux deux extrémités du gène sont des régions non codantes (non traduites).

Les gènes sont parfois rassemblés en groupes, les gènes qui partagent une même fonction étant proche, les uns des autres.

Les **pseudogènes** sont des séquences d'ADN qui ont certaines des caractéristiques de structure des gènes exprimés ont probablement été fonctionnels, mais ils ont acquis au cours de l'évolution, un nombre de mutations inactivatrices, les empêchant de produire une protéine.

L'ADN non transcrit est appelée **ADN intergénique ou extragénique**, certaines séquences d'ADN intergénique situées à proximités des gènes exprimés sont indispensables au contrôle de l'expression des gènes, mais une grande quantité ne parait pas indispensable et ne possède pas de fonction connue.

Dans l'ADN intergénique et parfois dans les introns, sont situées des séquences répétées, dispersées au sein du génome, en une ou en plusieurs copies et n'ayant pas de fonction apparente ; des séquences répétées *alu*, les plus répandues, présentes environ 500 000 fois dans le génome humain, des répétition en tandem (CA)<sub>n</sub>, (CT)<sub>n</sub> et des éléments LINE.



Les **allèles mutés** de locus sont considérés **rares** quand la grande majorité de la population possède un unique allèle normal (sauvage).

Le **polymorphisme** est l'existence pour de nombreux locus chez l'homme, de deux allèles ou plus s'exprimant avec une fréquence non négligeable dans la même population.

Plus d'un 1/3 des gènes sont polymorphes, gènes codant des groupes sanguins (système ABO, Rh, protéines sériques...).

Les polymorphismes, en tant que marqueurs de diversité génétique, sont des outils très utiles pour cartographier le génome humain.

## b) Structure de l'ADN

L'**ADN** (acide désoxyribonucléique) a une structure en double hélice. Il est formé de quatre bases, adénine (A) et guanine (G) qui sont des purines, cytosine (C) et thymine (T) qui sont des Pyrimidines. Les bases sont maintenues par une armature sucre-phosphate, les phosphates sont attachés par des liaisons esters aux groupes hydroxyl en 3' et 5'.

Par convention, le groupe phosphorique représente le bout 5', le groupe OH le bout 3' ; c'est également dans ce sens que se lit une molécule d'ADN.

Le nucléotide (sous unité de l'ADN) est composé de l'une des 4 bases, du désoxyribose et du groupe phosphate.

Les couples A-T et G-C sont appelés « paires de bases », de ce fait les deux brins constitutifs sont complémentaires.

Cette complémentarité des bases permet :

- 1) La réplication, chaque brin contient la totalité de l'information de la molécule d'ADN et peut servir de matrice pour la synthèse du brin complémentaire après ouverture de la double hélice.
- 2) La réparation d'erreurs donc la protection d'information liée à une lésion de l'ADN.
- 3) Aux deux brins de se retrouver et de se réassocier dans un mélange moléculaire complexe. Cette « réassociation » ou « hybridation » est utilisée pour assurer la régulation de l'expression génétique.

### c) Structure et fonction de l'ARN

**La transcription** consiste en la génération d'une molécule d'ARN (acide ribonucléique) à simple brin à partir d'une matrice d'ADN à double brin présente dans le noyau cellulaire.

L'ARN de siège intracytoplasmique présente trois différences avec l'ADN, il est monobrin, le sucre est le ribose et l'uracile (U) remplace la base T (ou thymine).

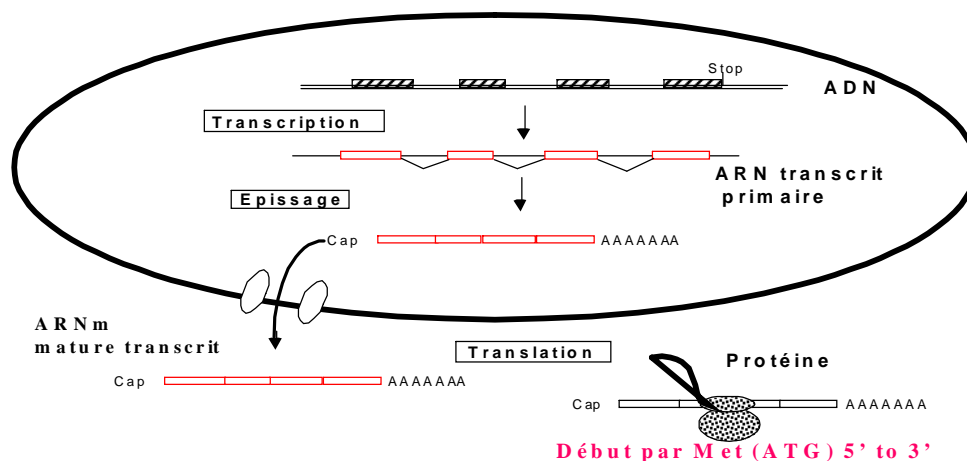
L'ARN a de multiples fonctions cellulaires, les ARN ribosomiaux (ARNr), de transfert (ARNt), messagers (ARNm) et une forme particulière retrouvée principalement dans le noyau, les small nuclear ribosomes (snARN).

Le processus de transcription est réalisé par l'ARN polymérase qui reconnaît une séquence spécifique de l'extrémité 5' du gène (**le promoteur**) et commence la transcription en ARN messenger (ARNm), ce dernier se rend vers les ribosomes ou il est traduit en une séquence protéique.

Le mécanisme de **traduction** repose sur l'utilisation de trois bases (codon) pour décrypter le code génétique ; plusieurs codons différents peuvent conduire au même acide aminé ; trois codons, UUA, UAG et UGA, sont des codons-stop (entraînent un arrêt prématuré de la traduction).

L'ARN de transfert (ARNt) se lie de façon covalente à un acide aminé particulier complémentaire du codon spécifique de l'ARNm.

L'ordre d'addition des acides aminés est tel que l'extrémité 5' de l'ARNm correspond à l'extrémité NH<sub>2</sub>-terminal de la protéine, et l'extrémité 3' de l'ARNm à l'extrémité COOH-terminal.



#### d) Les modifications génétiques

Une **mutation** est une modification du matériel génétique ; le taux de mutation est la fréquence de cette mutation, deux maladies génétiques importantes ont des taux élevés de mutation: la neurofibrose de Recklinghausen et la DMD (cette dernière s'expliquant par la grande taille du gène).

Différents mutations sont observés :

- **Mutation faux-sens (missense mutation)** consiste en un changement du triplet du codon en celui d'un acide aminé différent mais conservation du cadre de lecture, la protéine résultante aura soit une structure différente soit une fonction anormale.

- **Mutation non-sens** transforme un codon normal en UAA, UGA ou UAG, créant un codon stop prématuré conduisant à une protéine tronquée non fonctionnelle.

- **Mutation décalant le cadre de lecture (frameshift mutation)**. Si elle n'est pas un multiple de trois, une insertion ou une délétion d'un petit nombre de nucléotide dans une région codante va altérer le cadre de lecture de la traduction en aval de ce point, conduisant à une extrémité COOH de la protéine anormale.

- **Expansions de trinuécléotides**. Le génome humain contient des répétitions de trinuécléotides en tandem n'excédant pas 5 à 35 répétitions. Quand leur nombre est plus important ou quand ils sont à proximité ou à l'intérieur d'un gène, ils provoquent des maladies.

### **III) Principales techniques de génétique moléculaire.**

La connaissance des techniques d'analyse génétique est imposée actuellement au clinicien afin de comprendre leurs interprétations. Ces techniques reposent sur quelques principes fondamentaux hérités du métabolisme cellulaire.

#### **A) Principes fondamentaux de l'ADN**

##### **1) Dénaturation/renaturation**

**La dénaturation** est la dissociation des deux brins de l'ADN qui deviennent simple brin, lorsqu'ils sont chauffés à une certaine température. La dénaturation est progressive, les régions à faibles forces d'association (riches en A-T, ayant deux liaisons hydrogènes) lâchant d'abord, suivies des régions plus stables (riches en C-G, ayant trois liaisons hydrogène).

**La renaturation ou hybridation** : quand une solution d'ADN dénaturé est refroidie progressivement, les deux brins se rejoignent parfaitement et l'ADN double brin devient fonctionnel. La renaturation n'est pas réservée à l'ADN simple brin mais peut aussi avoir lieu entre des molécules d'ADN et d'ARN. Il est également possible d'hybrider des ADN d'espèces différentes. Dans l'hybridation, seul compte, la séquence des nucléotides.

##### **2) Réplication de l'ADN**

La réplication se fait par l'intermédiaire d'une enzyme, l'ADN polymérase qui, à partir d'une amorce, va synthétiser un second brin en prenant le premier comme exemple. Les ADN polymérase travaillent de façon unidirectionnelle, allant du 5' vers le 3', rendant difficile la synthèse du second brin qui est antiparallèle ; la cellule utilise donc une technique différente qui consiste en la synthèse « a reculons » de petite brins d'ADN, ces petits brins sont ensuite attachés les uns des autres par une ligase.

## **B) Les outils de la cellule.**

- 1) **L'ADN polymérase** a comme action la néosynthèse complémentaire de l'ADN basée sur une matrice simple brin préexistante et d'une amorce. Ils existent plusieurs types d'ADN polymérase, mais celles utilisées en biologie moléculaire est la Taq polymérase, extraite d'une bactérie *thermophilus aquaticus* vivant dans les sources chaudes, supporte des températures élevées sans perte importante de son activité. Une autre ADN polymérase particulière est représentée par le produit de gène *pol* des rétrovirus. Cette polymérase copie une matrice d'ARN en un brin d'ADN complémentaire. Elle est utilisée dans les RT-PCR.
  
- 2) **Les enzymes de restriction**, appelés également **endonucléases** coupent une molécule d'ADN double brin au niveau d'une séquence nucléotidique spécifique. De nombreuses endonucléases, purifiées à partir de bactéries, sont actuellement disponibles.
  
- 3) **Les ligases** effectuent une réaction inverse, elles attachent les petits brins d'ADN l'un à l'autre, elles permettent l'élongation du bras anti-sens et sont également utilisées dans les processus de réparation de l'ADN. En biologie moléculaire, elles sont utilisées comme « colle », lors de constructions : insertion de fragments d'ADN dans des plasmides.



## **C) Principales techniques de génétique moléculaire**

### **1) Analyse par amplification**

#### **1.1. Polymérase chain reaction (PCR)**

La PCR, permet l'amplification de fragments d'ADN (plus d'un million de fois). La réaction en chaîne est enclenchée par l'utilisation de deux amorces complémentaires à des régions du génome proches l'une de l'autre. Une PCR, se compose de 20 à 30 cycles environ dans lesquels alternent dénaturation/hybridation des amorces et synthèse de nouveau brin (par l'ADN polymérase), appelé également « extension ». A chaque cycle, la PCR double l'ADN amplifiée. Les produits d'amplification (appelés « amplimères ») sont séparés sur gel d'agarose ou de polyacrylamide afin de déterminer leur taille et parfois leur abondance.

#### **1.2. Reverse Transcription PCR (RT-PCR)**

La RT-PCR se compose de deux réactions : une première transcrit de l'ARNm en ADN complémentaire simple brin (ADNc). Ce dernier sert alors de support à la seconde réaction qui est une PCR conventionnelle. En génétique moléculaire, cette méthode est utilisée pour détecter l'expression de certains gènes en mesurant la longueur du messager ou en quantifiant le produit d'amplification. Elle permet également de rechercher des mutations dans des gènes dont la structure génomique n'est pas connue, ce qui évite la recherche systématique, exon par exon.

#### **1.3. Ligation Chain Reaction (LCR)**

La LCR est une variante de la PCR basée non sur l'amplification d'une région génomique située entre deux amorces, mais sur l'union de deux amorces adjacentes. Une LCR alterne dénaturation/hybridation et ligase. La présentation correcte des deux amorces et leur ligation ne sont possibles que lorsqu'elles sont parfaitement complémentaires à la matrice d'ADN.

## 2) Analyse par hybridation : analyse de Southern.

Portant le non de son inventeur, l'analyse de *Southern*, consiste à identifier une séquence d'ADN génomique à l'aide d'une séquence d'ADN complémentaire et repose sur le principe de dénaturation-hybridation de deux molécules d'ADN. Des fragments de taille connue sont obtenus après digestion enzymatique de l'ADN génomique puis, ils sont séparés dans un gel d'agarose, dénaturé puis transféré et fixés sur un support solide (membrane de nitrocellulose). Il est alors possible d'hybrider à l'ADN monobrin se trouvant sur cette membrane une molécule simple brin. Cette dernière molécule est appelée « sonde » car elle va « retrouver » sa séquence complémentaire et s'y fixer. La taille du fragment ainsi déterminé est estimée par rapport à des marqueurs de poids moléculaire connu qui se trouvent également sur le gel. De façon analogue, l'analyse par transfert d'ARN et hybridation s'appelle *Northern* et l'analyse de protéines et marquage par anticorps s'appelle *Western*.

## 3) Analyse de séquences.

Le séquençage reste souvent l'étape ultime d'une analyse moléculaire et permet de connaître la suite des bases d'un segment d'ADN.

Plusieurs techniques de séquençage ont actuellement mises au point.

### 3.1. Méthode de Sanger.

La méthode de séquençage de Sanger (1977) est la plus couramment employée. Ce procédé permet de caractériser un gène en termes de séquence linéaire de bases ACGT. Cette séquence étant déterminée, il est facile de prédire la séquence d'acides aminés de la protéine correspondante en se servant du code génétique.

### 3.2. Méthode automatique

Plusieurs séquenceurs automatiques sont disponibles, ils sont basés sur l'utilisation de l'analyse automatique de fluorochromes. Les pics de fluorescence sont détectés en continu par un système de laser et de détecteur. Un logiciel transforme les pics en bases.

## D) Approches diagnostiques.

### Analyse de liaison

L'analyse de liaison est la plus utilisée aujourd'hui, elle permet la localisation de gènes sur le génome. Elle consiste à rechercher une transformation méiotique non aléatoire d'un marqueur et d'une maladie. Durant la méiose, chaque chromosome recombine avec son homologue lors d'un ou de plusieurs crossing-over. De ce fait, si le marqueur analysé n'est pas sur le même chromosome que le locus de la maladie, sa transmission sera en équilibre de liaison. Par contre, la découverte d'un déséquilibre de liaison indique, avec plus ou moins de force, que le marqueur est associé au locus de la maladie. La distance qui sépare deux marqueurs ou un marqueur et le locus d'une maladie peut être estimé par le nombre de recombinaisons survenant entre ces derniers. Une distance génétique d'un centiMorgan (cM) signifie que la probabilité de survenue d'une recombinaison en deux loci lors d'une méiose est de 1%.

L'analyse de liaison se déroule en trois étapes :

- a. Investigation d'une ou de plusieurs familles présentant le même diagnostic, le mode de transmission de la maladie est déterminé et les individus sont classés selon leur phénotype.
- b. Analyse des microsatellites. Une collection de marqueurs polymorphes est analysée par PCR. Les marqueurs actuellement utilisés sont des microsatellites génétiquement bien étudiés et déjà bien localisés avec précision sur les différents chromosomes. Ils recouvrent la totalité du génome et sont à une distance moyenne de 10 CM. Des oligonucleotides déjà bien marqués permettant leur amplification optimale par PCR sont actuellement disponibles
- c. Analyse statistique comme l'existence de liens entre un marqueur et le locus d'une maladie peut être le siège du hasard, des méthodes statistiques d'analyse sont nécessaires. La méthode la plus utilisée est celle du lod score, basée sur le calcul du rapport de deux vraisemblances. En pratique, ce rapport est calculé pour plusieurs valeurs de  $\theta$ . Des lod scores positifs ( $\geq 3$ ) indiquent l'existence d'une liaison, alors que des lod scores négatifs ( $-2$ ) suggèrent une absence de liaison.

## **Références**

- 1) *Biologie moléculaire et médecine*, Kaplan J C et Delpuch M. Flammarion, 2ed édition, Paris : Médecine Sciences, Flammarion 1993.
- 2) *Principes de génétiques moléculaire et médicale*. Gelehrter TD, Collins. FS. Editions Pradel 1992.
- 3) *Neurogénétique*. Collection Traité de Neurologie. Brice.A, Schorderet.DF. Doin 1999.
- 4) *Outils et stratégie de biologie moléculaire dans le diagnostic génotypique prénatal* Pinto. M. Tec & Doc Lavoisier.1997.
- 5) *Handbook of human genetic linkage*. The Johns Hopkins University Press 1994.
- 6) *Molecular Biology of the gene*. J.D.Watson, N.H. Hopkins, J.W.Roberts et al. 4th ed. Mento Park1986.