

La découverte du gène de la mucoviscidose

La mucoviscidose est certainement la maladie héréditaire autosomique récessive la plus fréquente (une naissance sur 2 000, un transmetteur sur 20) dans la population européenne. Les manifestations cliniques de cette affection grave, essentiellement pulmonaires et pancréatiques, résultent d'un épaissement des sécrétions exocrines qui obstruent les voies aériennes supérieures et prédisposent aux infections bronchiques répétées. Ces atteintes détruisent progressivement le poumon et conduisent à une insuffisance respiratoire létale, le plus souvent avant 30 ans. Les études électrophysiologiques les plus récentes indiquent que, chez les malades, le transport des ions à travers les membranes des cellules épithéliales est anormal. L'activité du canal à chlore (Cl⁻), qui semble sous le contrôle d'un processus complexe de régulation, est normalement stimulée par les protéines kinases C et A (c'est-à-dire dépendantes de l'AMPC); cette activation fait défaut chez les patients. Malgré ces progrès, la lésion biochimique primaire n'est toujours pas connue, et une approche alternative, celle de la génétique inverse [1], a été mise en œuvre par plusieurs laboratoires dans le but d'isoler le gène responsable en partant de sa localisation chromosomique.

Il y a un peu moins de 18 mois, en avril 1988, la revue scientifique nord-américaine *Science*, dans un article intitulé « La course pour le gène de la mucoviscidose touche à sa fin » [2], commentait longuement les enjeux de cette intense compétition et les rivalités des quelques laboratoires encore en piste. Le groupe dirigé par R. Williamson (Londres, GB), avait en effet laissé entendre un an plus tôt, qu'il avait isolé le gène CF (de *cystic fibrosis*, pour fibrose kystique du pancréas, ou mucoviscidose), une nouvelle qui avait découragé certains des protagonistes. Il s'agissait en fait d'une séquence codant pour une protéine (*int related protein*, IRP) sans aucun rapport avec la maladie, mais située, on le sait aujourd'hui, à quel-

ques 150 kilopaires de base (kpb) du gène tant convoité. Trois ans auparavant, la découverte de la localisation du gène CF sur le chromosome 7 avait brutalement accéléré la course que viennent de gagner L.C. Tsui, J.R. Riordan (Université de Toronto, Canada), F.S. Collins (Université du Michigan, USA) et leurs collaborateurs, qui ont, en 1987, regroupé efforts et compétences sur le sujet. Dans sa parution du 8 septembre 1989, le même journal, *Science*, publie trois articles impressionnants décrivant l'identification du gène, la nature de son produit, et la caractérisation de la principale mutation responsable de la maladie [3-5]. Il s'agit d'un événement scientifique majeur. Pour les malades et leurs familles, cette découverte donne immédiatement accès à de nouveaux moyens de diagnostic et permet d'espérer des jours meilleurs en matière de prévention et de traitement. Pour ceux qui essaient, à l'aide de méthodes semblables, de caractériser d'autres maladies génétiques, c'est un encouragement qui vient à point tant sont ardues ces approches et peuvent sembler aléatoires leurs succès.

La marche vers le gène. Depuis 1985, à la suite des travaux des groupes de L.C. Tsui, de R. White et R. Williamson, on savait que le gène CF est situé entre deux marqueurs liés, MET et D7S8 (J3.11), séparés par approximativement un à deux millions de paires de bases [6]. Pour pouvoir, avec quelques chances de succès, commencer à marcher sur la région, c'est-à-dire cloner toutes les séquences qu'elle contient, il fallait réduire cette distance en trouvant des marqueurs plus proches. De la bouche même de Lap-Chee Tsui, dont la détermination et la ténacité tout au long de l'entreprise sont à saluer, le succès de la stratégie choisie (*a brute force approach*) dépendait avant tout des moyens massifs mis en jeu. Pour se rapprocher au maximum du gène convoité, son groupe entreprit de saturer la région cible en

marqueurs ADN qui furent localisés et ordonnés à l'aide d'analyses physiques et génétiques. Pas moins de 258 séquences spécifiques du chromosome 7 furent isolées [7], dont 53 localisées en 7q31-32. Deux de ces marqueurs (D7S122 et D7S340), étroitement liés au gène CF, se révélèrent contenus dans la région délimitée par MET et D7S8. L'ordre des marqueurs déterminé (MET-D7S340-DS7122-J3.11), la carte physique de la région établie à l'aide des techniques d'électrophorèse en champ pulsé (*m/s n° 1, vol. 4, p. 46*), il restait à passer au peigne fin cette région du génome (environ 1 500 kpb).

Pour accélérer la recherche et restreindre l'étendue de la zone à cibler, les méthodes classiques de marche sur le chromosome furent doublées de tentatives de saut à partir de chaque extrémité de la région. Cette seconde approche, outre son intérêt pour rejoindre plus vite le but, permet de s'affranchir des embûches dressées par les régions « inclonables » qui, on le sait maintenant, peuvent sérieusement handicaper, sinon arrêter la progression des marcheurs. A cet effet, F. Collins avait construit en 1987 une banque de saut* constituée d'une collection de séquences d'ADN génomique distantes d'une centaine de kpb en moyenne. Des séquences contiguës d'ADN s'étendant sur une région de 280 kpb furent clonées à l'aide de ces deux approches qui permirent d'isoler quelques 49 clones de marche (cosmides ou phages) et neuf clones de saut, chacun permettant de fran-

* Banque de saut : technique qui consiste à couper de grands fragments d'ADN (autour de 100 kpb), à les recirculariser en liant leurs deux extrémités à un marqueur de sélection, puis à cloner, après une nouvelle coupure, l'ADN des deux extrémités maintenant réunies au marqueur. Si un tel clone est détecté par hybridation avec une sonde de l'une des extrémités, l'ADN de l'autre extrémité fournira lui-même une sonde correspondant à une région du chromosome située, dans l'exemple pris ici, 100 kpb plus loin.

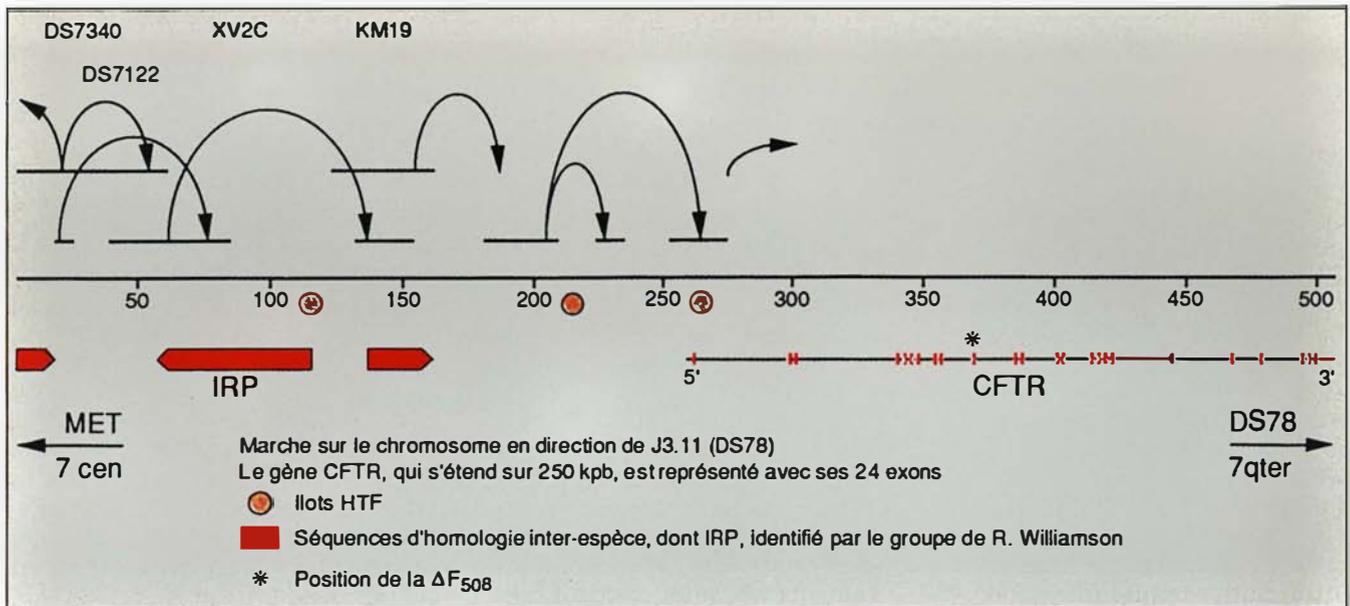


Figure 1. **Marche vers le gène de la mucoviscidose.** Le point de départ est représenté à gauche sur le schéma et correspond à la localisation de la sonde caractérisée par le groupe de L.C. Tsui comme la plus proche du gène CF. Les 280 kpb séparant ce point du début du gène CFTR ont été franchies en marchant et en sautant (flèches courbes) sur le génome en direction de DS 78. Le gène CFTR est représenté avec ses 24 exons et, sur la gauche du schéma, sous la ligne d'échelle, sont positionnées les trois séquences codantes conservées entre les espèces, dont IRP, caractérisée par le groupe de R. Williamson. Les cercles roses représentent les îlots HTF retrouvés dans la région qui, souvent, balise le début des gènes. 7 cen = vers le centromère ; 7 qter = vers le télomère du bras long.

chir entre 30 et 70 kpb (figure 1). Pour épingler les séquences codantes CF, la stratégie des *zoo-blots*** fut choisie en raison de sa relative facilité et de ses succès antérieurs, notamment pour la caractérisation du gène de la myopathie de Duchenne. Quatre régions contenant des séquences nucléotidiques homologues à celles d'autres espèces (en l'occurrence le bœuf, la souris et le poulet) furent retenues comme de possibles candidats, et les séquences conservées utilisées comme sondes pour rechercher leurs copies dans sept banques d'ADNc construites à partir d'ARN messagers de plusieurs tissus : lignée tumorale T84, cultures de cellules de glandes sudoripares normales ou de malades atteints de mucoviscidose, pancréatiques ou pulmonaires. Les trois premières zones d'homologie inter-espèces révélèrent la présence d'une séquence répétée transcrite et de deux gènes, dont IRP, déjà épinglé par le groupe de Londres grâce à la présence d'îlots HTF*** et, un temps candidat, CF. C'est la quatrième région conservée qui autorisa la pêche, difficile mais réussie, d'un seul clone contenant seulement 113 paires de bases (pb) du premier

exon du gène CF. Cette séquence permit de reconstituer, en agencant plus de 20 recombinants isolés grâce à elle, le *puzzle* de l'ADNc complet (6 129 pb) et, partant, de positionner les 24 exons du gène sur 250 kpb de génome. La distribution tissulaire des ARN messagers s'avère compatible avec les données physiopathologiques, de même que le produit du gène, une protéine de transport membranaire, que ses découvreurs proposent d'appeler CFTR (*cystic fibrosis transductor regulator*).

Le CFTR. Le gène ainsi caractérisé code pour une protéine de 1 480 acides aminés dont la séquence indique à l'évidence son appartenance à une super-famille de protéines membranaires. Le CFTR présente une homologie structurale considérable avec un ensemble de transporteurs dépendants de l'ATP décrits chez les procaryotes, et plus récemment chez les eucaryotes [8], telles les protéines MDR (*multidrug resistance*), CQR (*chloroquine resistance*) ou STE6 (de levure). Il associe deux motifs répétés formés d'un domaine hydrophobe transmembranaire, et une importante région hydrophile exposant des séquences consensus N B F

(*nucleotide (ATP)-binding fold*) pouvant lier l'ATP. Ces deux parties de la protéine sont reliés par un domaine cytoplasmique (domaine R), codé par un grand exon unique, présentant de nombreux résidus chargés, contenant la majorité des sites potentiels de phosphorylation par les protéines kinase A ou C, d'où sa fonction régulatrice probable (figure 2).

Le rôle exact joué par cette protéine reste à déterminer. Elle est très probablement impliquée dans le transport transmembranaire actif des ions au niveau du pôle apical des cellules épithéliales. Il reste à établir s'il

** Zoo blot : technique qui se base sur le contraste entre la conservation au cours de l'évolution des séquences codantes et la non-conservation des séquences introniques et intergéniques pour identifier des fragments exoniques ; seuls de tels exons peuvent hybrider en Southern blot avec des fragments d'ADN d'espèces variées.

*** Îlots HTF : Hpa II tiny fragments, îlots contenant de nombreux sites de coupure par l'enzyme de restriction, Hpa II qui reconnaît la séquence 5'CCGG3' lorsque les C sont méthylés. Des régions riches en CG méthylées sont caractéristiques des séquences bordantes 5' des gènes actifs, principalement ceux à expression ubiquitaire.

s'agit d'un canal ionique, ou plutôt d'un régulateur de l'activité d'un tel canal, et si elle a par ailleurs d'autres fonctions cellulaires sans rapport avec le transport actif des ions. Quoiqu'il en soit, comme le suggèrent les données structurales, l'hydrolyse de l'ATP et la phosphorylation par les protéines kinases semblent nécessaires à cette activité de transport actif, et à sa régulation.

La comparaison de la structure des clones d'ADN complémentaire obtenus à partir de cellules de glandes sudoripares de sujets sains ou de malades confirme ce que les études génétiques antérieures avaient permis de postuler, à savoir qu'une majorité des patients sont porteurs de la même mutation. Il s'agit d'une délétion de trois nucléotides de l'exon 10 qui ampute la protéine d'une phénylalanine en position 508 (ΔF_{508}), c'est-à-dire au niveau d'une des séquences NBF. Cette anomalie pourrait empêcher la liaison correcte de l'ATP, ou perturber la dynamique conformationnelle indispensable à l'activité du CFTR, expliquant par là les observations faites sur l'épithélium bronchique des malades grâce aux expériences de *patch-clamp* [9].

Génétique de la mucoviscidose : prévention et traitement. Les données de l'analyse génétique apportent les arguments les plus solides en faveur du fait que le gène codant pour le CFTR est bien le gène CF. Environ 70 % des défauts moléculaires observés chez les patients ou les transmetteurs sont constitués par la délétion ΔF_{508} . Une étude exhaustive des marqueurs ADN de la région a permis d'établir une série d'haplotypes utiles à la caractérisation des divers chromosomes sains ou CF. La délétion ΔF_{508} est presque toujours associée, sur les chromosomes CF, à un haplotype majoritaire ou à ses dérivés ; en revanche, elle n'est jamais retrouvée sur les chromosomes sains portant cet haplotype. Cette observation rend extrêmement probable l'hypothèse qu'il s'agit d'une mutation délétère, et non d'un polymorphisme. En une occasion, elle est retrouvée associée à un haplotype différent, ce qui suggère la survenue d'un événement mutational récurrent, phénomène déjà observé pour le gène β -globine. Les autres lésions moléculaires constituent un groupe hétérogène et

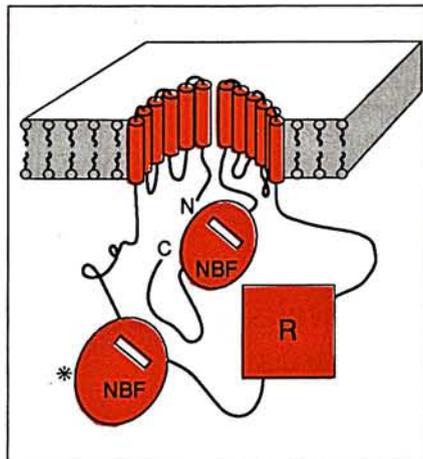


Figure 2. **Schéma de la protéine CFTR, produit du gène CF.** Les deux régions transmembranaires sont séparées par une région de liaison potentielle de l'ATP (NBF, nucléotide binding fold) et par la grande sous-unité R. Il existe un autre domaine NBF à proximité de l'extrémité carboxyterminale de la protéine. L'astérisque indique la position de la ΔF_{508} . (Modifié d'après [4].)

restent à caractériser, mais l'analyse génétique donne déjà quelques clés importantes. Deux formes peuvent être distinguées cliniquement, PI (insuffisance pancréatique), associée à ΔF_{508} , et PS (sans insuffisance pancréatique). L'analyse génétique fait apparaître que les mutations responsables de ces deux formes sont de gravité différente, et que le phénotype PS est dominant sur PI. L'iléus méconial****, qui survient chez 5 à 10 % des nouveaux-nés, surviendrait chez des sujets porteurs de la délétion ΔF_{508} , dans un contexte génétique ou d'environnement qui reste à élucider. Ces découvertes sont d'une très grande importance, à terme, pour la compréhension de la biologie et de la physiologie du système. Elles concernent déjà les patients et leurs familles en bouleversant les données sur lesquelles s'appuient les moyens de prévention. Près de 70 % des transmetteurs de la mucoviscidose peuvent d'ores et déjà être identifiés facilement, à l'aide de la sonde spécifique ΔF_{508} , par des techniques maintenant bien standardisées. En dépit des problèmes posés par la non-caractérisation de 30 % des muta-

**** Iléus méconial : occlusion intestinale néonatale due à l'épaississement du méconium qui ne peut s'évacuer.

tions, rien ne s'oppose à ce que le dépistage des transmetteurs soit mis en place dès maintenant, en commençant avec des échantillons ciblés, judicieusement définis, de la population. Bien sûr, pour les mutations encore inconnues du gène, il n'est pas encore possible de s'affranchir des méthodes d'analyse de liaison génétique, et le diagnostic prénatal, notamment, requiert pour sa mise en œuvre la maîtrise des deux approches. Cependant, la mucoviscidose est désormais accessible au dépistage à grande échelle.

Cette percée laisse-t-elle envisager, à moyen terme, la mise au point d'un traitement curatif ? Certes, il est d'autres maladies pour lesquelles le défaut moléculaire est connu de longue date et que l'on ne sait toujours pas guérir. Mais la connaissance de la structure et du fonctionnement du CFTR sont maintenant une question de temps, et l'on peut raisonnablement espérer que l'on pourra à terme agir avec efficacité contre les effets délétères du gène anormal en modulant l'activité de son produit.

RÉFÉRENCES

1. Jordan B. Grandeur et servitudes de la génétique inverse. *médecine/sciences* 1988 ; 4 : 138-40.
2. Roberts L. The race for the cystic fibrosis gene. *Science* 1988 ; 240 : 141-4 ; 282-5.
3. Rommens JM, Tannuzzi MC, Kerem BS, et al. Identification of the cystic fibrosis gene : chromosome walking and jumping. *Science* 1989 ; 245 : 1059-64.
4. Riordan JR, Rommens JM, Kerem BS, et al. Identification of the cystic fibrosis gene : cloning and characterization of complementary DNA. *Science* 1989 ; 245 : 1066-72.
5. Kerem BS, Rommens JM, Buchanan JA, et al. Identification of the cystic fibrosis gene : genetic analysis. *Science* 1989 ; 245 : 1073-80.
6. Kitzis A, Warren P, Kaplan JC. *médecine/sciences* 1988 ; 4 : 151-6.
7. Rommens JM, Zengerling S, Burns J, et al. Identification and regional localization of DNA markers on chromosome 7 for the cloning of the cystic fibrosis gene. *Am J Hum Genet* 1988 ; 43 : 645-63.
8. Higgins CF. Export-import family expands. *Nature* 1989 ; 340 : 342.
9. Sauvé R. Le *patch clamp*, une nouvelle façon de voir le canaux ioniques. *médecine/sciences* 1987 ; 3 : 538-45.

Michel Goossens

Professeur à la faculté de médecine, Université Paris XII.

Laboratoire de biochimie et Inserm U. 91, hôpital Henri-Mondor, 51, avenue du Maréchal-de-Lattre-de-Tassigny, 94010 Créteil, France.