

Génétique inverse et mucoviscidose

La mucoviscidose est la plus fréquente des maladies héréditaires. L'approche de la «génétique inverse», allant du gène vers ce pour quoi il code et non l'inverse, a mis sur la piste du gène qui est modifié chez les malades. Le diagnostic prénatal de l'affection est d'ores et déjà possible et la prochaine identification de la protéine anormale en cas de mucoviscidose permettra probablement d'en comprendre le mécanisme et peut-être de proposer des traitements nouveaux.

Alain Kitzis

Chef de travaux au CHU Cochin

Paule Warren

*Licenciée ès sciences,
université Laval, Québec*

Jean-Claude Kaplan

Professeur au CHU Cochin

La mucoviscidose ou fibrose kystique du pancréas (CF pour *cystic fibrosis* des anglo-saxons) est la maladie héréditaire la plus fréquente dans les populations d'origine européenne. Un enfant sur 1 600 à 2 000 naît atteint de cette maladie. L'affection est transmise selon un mode autosomique récessif, et environ une personne sur 20 est porteuse du gène morbide*, soit près de trois millions de personnes en France.

Il s'agit d'une exocrinopathie touchant essentiellement les voies respiratoires et digestives: un épaissement des sécrétions entraîne des phénomènes d'obstruction touchant les canaux excréteurs.

La lésion moléculaire de la mucoviscidose demeure totalement inconnue et, aucune thérapeutique efficace n'étant disponible, seuls sont mis en œuvre des traitements symptomatiques à visée palliative: kinésithérapie, antibiothérapie ou encore enzymes pancréatiques de substitution (lipase et amylase).

Le diagnostic de l'affection repose sur un tableau clinique fait de manifestations pulmonaires et digestives, ces dernières pouvant être très précoces au

point de revêtir l'aspect d'un iléus méconial (10 à 15 % des nouveau-nés mucoviscidosiques). Il est étayé par des tests biologiques: dosage de la trypsiniémie, et surtout test à la sueur dont la fiabilité est quasi absolue. En revanche les hétérozygotes sont cliniquement et biologiquement normaux et, à l'heure actuelle, il n'existe aucun moyen de les dépister s'ils n'ont déjà eu un enfant atteint de mucoviscidose. En l'absence de traitement efficace, plus de la moitié des enfants mucoviscidosiques meurent avant l'âge de 15-20 ans. A défaut de moyens de détection des hétérozygotes, la seule attitude en vigueur était, jusqu'en 1983, le conseil génétique classique, consistant en une explication du risque a priori dans les familles qui ont déjà eu un enfant atteint, c'est-à-dire une chance sur quatre à chaque naissance.

ADRESSE

A. Kitzis, P. Warren, J.-C. Kaplan: institut de pathologie moléculaire, CHU Cochin, 24, rue du Faubourg-Saint-Jacques, 75014 Paris, France.

TIRÉS A PART

A. Kitzis: institut de pathologie moléculaire, CHU Cochin, 24, rue du Faubourg-Saint-Jacques, 75014 Paris, France.

m/s n° 3 vol. 4, mars 88

* La mucoviscidose touche un nouveau-né sur 1 600. En appliquant la loi de distribution binomiale d'Hardy-Weinberg ($p^2 + 2pq + q^2 = 1$, où p et q représentent la fréquence respective de deux allèles exclusifs, p^2 et q^2 expriment la fréquence des homozygotes et $2pq$ celle des hétérozygotes), on calcule la fréquence du gène CF = $1/40$, et celle des individus porteurs = $1/20$.

RÉFÉRENCES

1. Brock DJH. Amniotic fluid alkaline phosphatase isoenzyme in early prenatal diagnosis in cystic fibrosis. *Lancet* 1983 ; ii : 941-3.
2. Muller F, Berg S, Frot JC, Boue J, Boue A. Prenatal diagnosis of cystic fibrosis. I. Prospective of 51 pregnancies. *Prenat Diagn* 1985 ; 5 : 97-108.
3. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Human Genet* 1980 ; 32 : 314-31.
4. Eiberg H, Mohr J, Schmiegelow K, Nielsen LS, Williamson R. Linkage relationships of paraoxonase (PON) with other markers: indication of PON-cystic fibrosis synten. *Clin Genet* 1985 ; 28 : 265-71.
5. Tsui LC, Buchwald M, Barker D, et al. Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science* 1985 ; 230 : 1054-7.
6. Knowlton RG, Cohen-Haguenauer O, Cong NV, et al. A polymorphic DNA marker linked to cystic fibrosis is located on chromosome 7. *Nature* 1985 ; 318 : 380-2.
7. Wainwright BJ, Scambler PJ, Schmidtke J, et al. Localization of cystic fibrosis locus to human chromosome 7cenq22. *Nature* 1985 ; 318 : 384-5.
8. White R, Woodward S, Leppert M, et al. A closely linked genetic marker for cystic fibrosis. *Nature* 1985 ; 318 : 382-4.

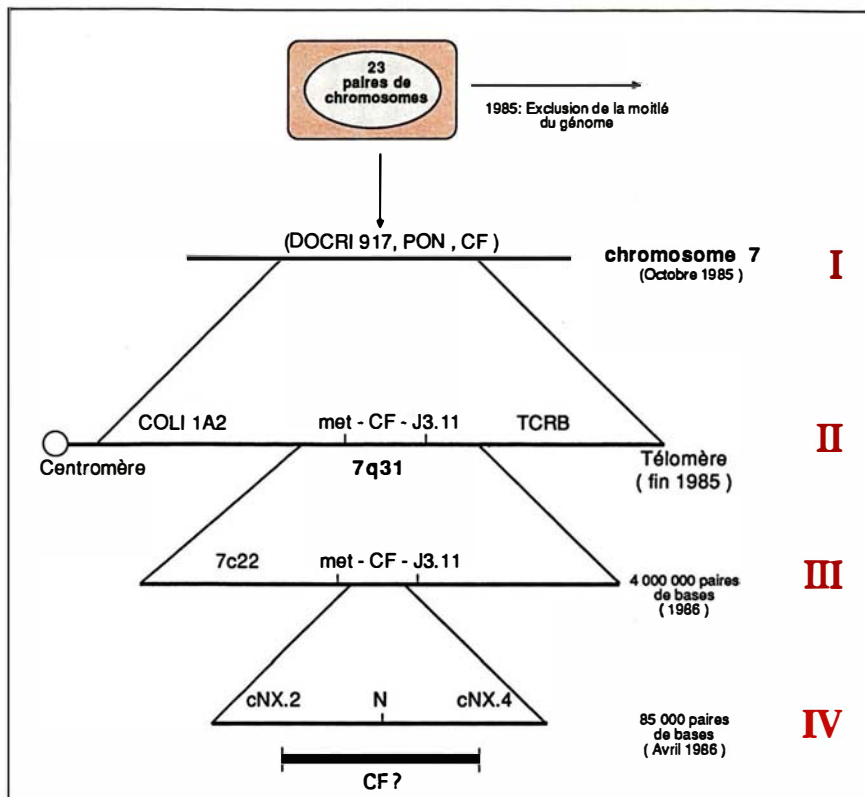


Figure 1. **Les étapes de la recherche du gène de la mucoviscidose.** Les positions des différents loci définis par les sondes sont indiquées au-dessus des fragments d'ADN. N : site de restriction Not 1.

Depuis 1983 un diagnostic prénatal par études enzymologiques du liquide amniotique est possible [1, 2], fondé sur le dosage des phosphatases alcalines d'origine intestinale, fortement diminuées dans le liquide amniotique des fœtus atteints de mucoviscidose. Ce diagnostic a cependant l'inconvénient de ne pouvoir être pratiqué que pendant une fenêtre étroite: la 17^e et la 18^e semaine d'aménorrhée. Des erreurs portant sur l'estimation de l'âge de la grossesse pourraient expliquer le taux élevé de faux négatifs, compris entre 1 et 10 % selon les laboratoires (A. Beudet, communication à la conférence nord-américaine sur la mucoviscidose, Toronto, octobre 1987). En fait, la génétique moléculaire offre désormais la possibilité de rechercher directement sur le génome les gènes responsables de maladies monofactorielles dues à une lésion d'un gène non encore

identifié. Cette démarche a bien entendu été appliquée à la mucoviscidose. Elle recueille ses premiers succès en ce qui concerne le diagnostic prénatal à l'aide de marqueurs génotypiques, et l'arrivée au gène lui-même est en vue.

Localisation chromosomique du locus CF

La mucoviscidose fait partie des maladies monofactorielles à gène inconnu où toutes les recherches antérieures visant à identifier une protéine responsable étaient demeurées vaines. En 1980, il est apparu que les innombrables marqueurs génotypiques que sont les polymorphismes de restriction, ou RFLP (*restriction fragment length polymorphism*) (*m/s suppl. au n° 7, vol. 3, p. 4*) offraient la possibilité de sortir de l'impasse [3]. La stratégie de la génétique inverse consiste: (a) à utiliser ces marqueurs pour des

études de *linkage* permettant dans un premier temps de localiser le locus morbide sur un chromosome, puis sur une région chromosomique ; (b) à le cerner ; (c) à l'isoler à partir d'un ADN normal ; (d) à le cloner et le séquencer ; (e) en fin de compte à déchiffrer la protéine correspondant au message génétique. Cette démarche nouvelle est appelée « génétique inverse » car elle procède d'une stratégie qui est l'inverse de celle de la génétique classique : alors que celle-ci part de la protéine pour arriver au gène (exemples : les hémoglobinopathies, les hémophilies), la génétique inverse part du gène pour descendre vers la protéine. La première étape a été particulièrement longue à franchir. En effet, en l'absence de tout indice

permettant de suspecter un chromosome plutôt qu'un autre, il a fallu adopter la démarche aléatoire de la stratégie d'exclusion. En bref, la méthode consiste à rechercher dans les familles de mucoviscidosiques la co-ségrégation d'un RFLP avec le locus CF. Cette étude permet l'évaluation quantitative de la distance génétique séparant le site polymorphe du locus morbide. Elle est d'autant plus faible qu'il y a moins de recombinaison entre les deux. On l'exprime en centimorgans, unité de distance génétique correspondant à une fréquence de recombinaison de 1 % par méiose et assimilable, en première approximation, à une longueur d'ADN de 1 million de paires de bases (1 Mb). Il faut étudier de nombreuses familles comportant

au moins deux enfants mucoviscidosiques, et effectuer un calcul statistique de la probabilité de *linkage*. Lorsqu'un résultat négatif (non liaison) est obtenu pour une sonde donnée, la localisation proche du locus morbide est exclue. La procédure est lourde et, même si le succès est en principe garanti, il est soumis aux lois du hasard. C'est ainsi qu'il a fallu 4 ans pour exclure environ 50 % du génome (R. Williamson, *St Mary's Hospital*, Londres). Paradoxalement, le premier *linkage* significatif a été obtenu par une équipe danoise à l'aide d'un marqueur protéique polymorphe : la paraxonase plasmatique* [4]. Cependant la localisation chromosomique du locus correspondant (locus PON) demeurait inconnue, ce qui réduisait la portée de cette découverte. L'intérêt de ce *linkage* a été cependant de montrer que dans les 68 familles de mucoviscidosiques étudiées le locus CF paraissait unique. En cette même année, à l'automne 1985, une étude américano-canadienne montrait enfin une liaison génétique entre une sonde anonyme, Docri 917 et le gène CF [5] (*figure 1*). Cette sonde, qui était aussi liée au locus PON, fut ensuite localisée par l'équipe de J. Frézal sur le bras long du chromosome 7 [6]. Simultanément, l'équipe de R. Williamson à Londres et celles de R. White et de G. Vande Woude aux États-Unis décrivaient deux sondes liées au gène CF : pJ 3.11 d'une part et le proto-oncogène *met* d'autre part [7, 8]. L'évaluation des distances génétiques devait ensuite montrer que la sonde Docri 917 était située relativement loin du gène CF (environ 15 centimorgans), tandis que pJ 3.11 et *met* étaient situées beaucoup plus près (respectivement à 0,3 et 0,4 centimorgan du locus CF). De plus ces sondes étaient situées de part et d'autre de ce locus [9, 10]. On pou-

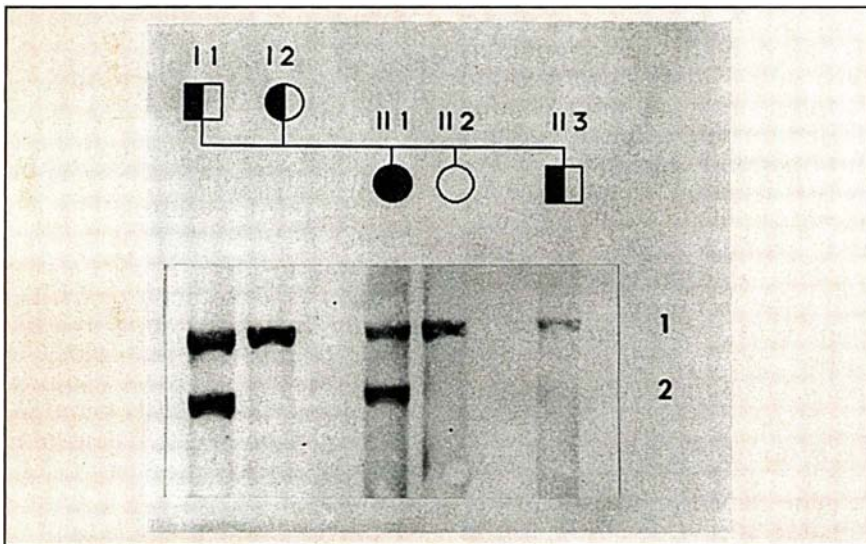


Figure 2. Étude familiale par la sonde *met D* en vue d'un diagnostic prénatal de mucoviscidose. L'ADN préparé à partir des globules blancs ou des villosités chorales prélevées à la dixième semaine de grossesse a été digéré par l'enzyme *Taq I*, soumis à une électrophorèse en gel d'agarose à 0,8 % et transféré sur un filtre de nylon selon la technique de Southern (I1, père ; I2, mère ; II1, enfant mucoviscidosique ; II2, enfant normal ; II3, prélèvement de trophoblaste). L'enfant mucoviscidosique est hétérozygote pour les allèles 1 et 2. Le père est également hétérozygote pour ces allèles. La mère est homozygote pour l'allèle 1. L'analyse du trophoblaste montre une bande au niveau de l'allèle 1 et pas de bande au niveau de l'allèle 2. Cet allèle, que possède l'enfant mucoviscidosique et qui vient du père, signe le chromosome atteint : le fœtus ne le possède pas et il sera soit homozygote normal, s'il a reçu le chromosome normal de la mère, soit hétérozygote sain, s'il a reçu le chromosome atteint de la mère, ce que cette seule étude ne peut dire. L'utilisation de la sonde pJ 3.11 (non représentée ici) a confirmé que le fœtus a bien reçu le gène normal du père, ce qui permet d'éliminer le risque de simple recombinaison et a montré qu'il a reçu l'allèle atteint de la mère, ce qui indique qu'il sera donc hétérozygote sain. Ce diagnostic a été confirmé par la naissance d'un enfant sain (trypsinémie normale).

m/s n° 3 vol. 4, mars 88

* Enzyme clivant le paroxon, substance anticholinestérasique utilisée dans le traitement du glaucome et dont il existe un polymorphisme phénotypique.

RÉFÉRENCES

9. Farrall M, Rodeck CH, Stanier P, *et al.* First trimester prenatal diagnosis of cystic fibrosis with linked DNA probes. *Lancet* 1986; i: 1402-5.
10. Beaudet A, Bowcock A, Buckwald M, *et al.* Linkage of cystic fibrosis to two tightly linked DNA markers: joint report from a collaborative study. *Am J Hum Genet* 1986; 39: 681-93.
11. Porteous DJ. Chromosome mediated gene transfer: a functional assay for complex loci and an aid to human genome mapping. *Trends Genet* 1987; 3: 177-82.
12. Cooper CS, Park M, Blair DG, *et al.* Molecular cloning of a new transforming gene from a chemically transformed human cell line. *Nature* 1984; 311: 29-33.
13. Cooper CS, Blair DG, Oskarsson MK, Tainsky MA, Eader LA, Vande Woude G. Characterization of human transforming genes from chemically transformed teratocarcinoma and pancreatic carcinoma cell lines. *Cancer Res* 1984; 44: 1-10.
14. Scambler PJ, Law HY, Williamson R, Cooper CS. Chromosome mediated gene transfer of six DNA markers linked to the cystic fibrosis locus on human chromosome seven. *Nucleic Acids Res* 1986; 14: 7159-74.
15. Estivill X, Farrall M, Scambler PJ, *et al.* A candidate for the cystic fibrosis locus isolated by selection for methylation-free islands. *Nature* 1987; 326: 840-5.
16. Bird AP. CpG-rich islands and the function of DNA methylation. *Nature* 1986; 321: 209-13.

vait dès lors considérer que ce dernier était cerné dans un domaine d'ADN d'environ 1 Mb. Dès lors les recherches se sont orientées simultanément dans deux directions: d'une part, la mise en œuvre immédiate d'un diagnostic prénatal génotypique, d'autre part la quête du gène CF lui-même.

Diagnostic prénatal et sondes indirectes

Les sondes pJ3.11, *metD* ou *metH* définissent 5 RFLP (figure 2, page 153). Leur application au diagnostic prénatal était possible en raison: (a) de leur proximité, ce qui minimise les risques de recombinaison; (b) de leur situation de part et d'autre du locus CF, ce qui augmente considérablement la fiabilité de la méthode, en permettant de visualiser à coup sûr les recombinaisons simples, dont la probabilité est inférieure à 1 % pour chaque RFLP; (c) de l'informativité qui est importante, puisque 77 % des couples ayant un enfant mucoviscidose vivant sont informatifs**, parmi lesquels 55 % le sont des deux côtés du gène CF. C'est par cette technique que nous avons réalisé notre premier diagnostic prénatal de mucoviscidose en 1986 (figure 2).

La quête du gène

Le gène étant circonscrit dans un domaine d'environ 1 Mb, vingt fois plus grand que la taille maximale d'un cosmide, il fallait trouver un moyen d'isoler ce territoire

* La fiabilité du diagnostic prénatal est donc très grande puisqu'elle serait seulement mise en défaut par une double recombinaison; or la probabilité de celle-ci est égale au produit des probabilités individuelles de recombinaison simple, soit $0,3\% \times 0,4\% \ll 0,01\%$, ce qui est inférieur au risque d'avoir un enfant mucoviscidose dans la population générale!

** L'informativité résulte de l'hétérozygotie d'un RFLP permettant de distinguer les deux chromosomes d'une même paire et d'attribuer un locus morbide à l'un des deux.

par d'autres moyens. Ceux-ci ont été fournis par le procédé dit de *chromosome mediated gene transfer* (CMGT) (voir nouvelle m/s n° 1, vol. 3, p. 47) [11]. Il consiste à transférer in vitro des chromosomes métaphasiques dans une cellule hôte appartenant à une espèce différente. En fait les chromosomes ne demeurent pas entiers lors de la manipulation et ce sont des fragments qui pénètrent dans la cellule réceptrice. Ce système réclame la présence d'un marqueur qui permet la sélection positive de la cellule ayant incorporé le fragment recherché. En ce qui concerne le gène CF, Williamson a eu l'idée que l'oncogène *met* pourrait justement constituer un tel marqueur, en raison de sa localisation providentiellement proche. Il a donc utilisé comme source de chromosome humain une lignée d'ostéosarcome portant une translocation chromosomique t(1; 7) dans laquelle on savait que *met* était activé, donc capable de transformer in vitro les cellules de la lignée de souris NIH 3T3 [12, 13]. C'est ainsi qu'a été obtenue la lignée CII, où les cellules transformées par *met* activé se sont avérées contenir de l'ADN humain hybridant aussi avec la sonde pJ3.11 [14]. Le gène CF étant situé entre ces deux séquences (figure 1), l'opération de transfert du gène CF dans une cellule de souris était réussie. Ceci équivaut à un clonage moléculaire, avec une nuance importante: le fragment d'ADN humain cloné ne peut être séparé en bloc de l'ADN de souris, à la différence de ce qui se passe dans un clonage classique. Or le fragment intéressant représente environ 4 Mb (figure 1, étape III), ultérieurement réduits à 1 Mb dans le sous-clone cellulaire CIId renfermant toujours *met*. C'est dans ce dernier que Williamson a fait le pari que se trouvait le gène CF.

Le saut vers le gène CF

L'ADN de la lignée CIId a été cloné dans une banque de cosmides où les rares clones d'ADN

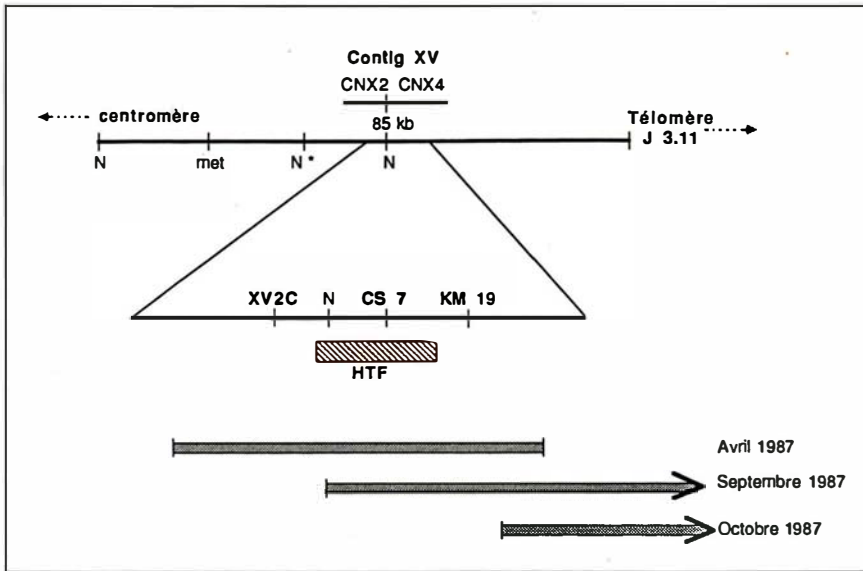


Figure 3. **Caractérisation du segment d'ADN susceptible de contenir le gène CF.** Représentation du fragment d'ADN de 85 kilobases (kb) (contig), centré sur le site Not I (N), et situé à environ 650 kb du locus *met*. Le site Not I facultatif, situé entre *met* et XV 2c, est représenté par N*. La localisation proposée du locus CF à différentes étapes de sa recherche est représentée par les rectangles gris.

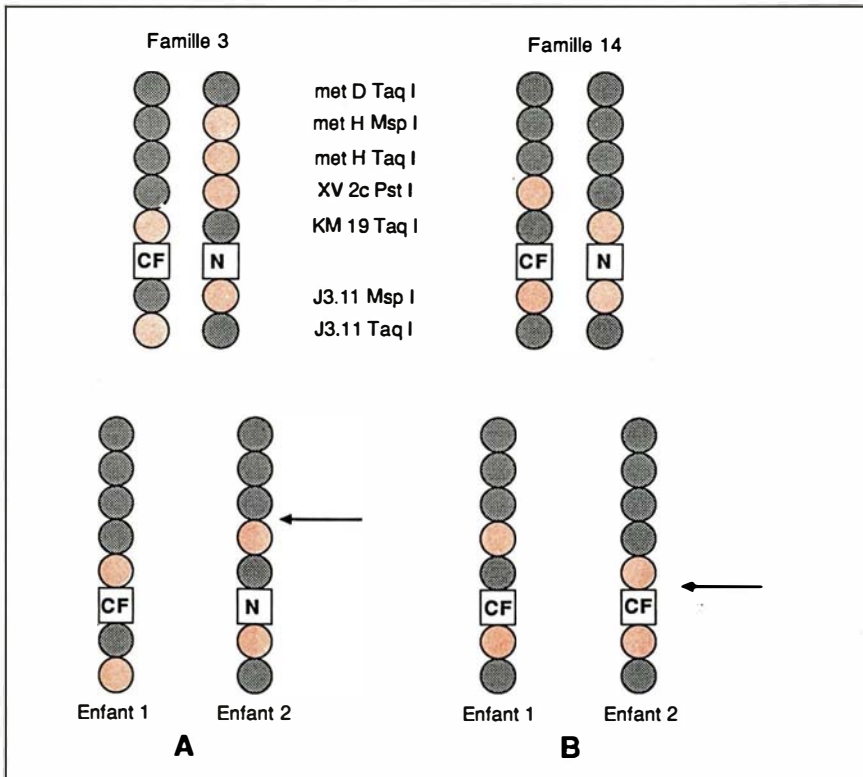


Figure 4. **Représentation schématique des crossing-over proches du locus CF.** (A), entre les sondes *met* et le locus CF; (B), entre les sondes XV2c-KM 19 et le locus. La flèche indique le lieu du crossing-over chez le deuxième enfant des familles 3 et 14 (résultats personnels non publiés). Grands allèles = cercles gris; petits allèles = cercles rouges.

m/s n° 3 vol. 4, mars 88

humain ont été repérés, puis utilisés pour cribler une banque spécifique du chromosome 7 humain. Ceci a permis de repérer deux cosmides contigus, car hybridant avec une même sonde. Cette stratégie dite des *contigs* consiste à reconstituer une longue séquence d'ADN génomique à partir de clones chevauchants et ordonnés (figure 1, étape IV et figure 3). Grâce à la méthode d'électrophorèse en champ pulsé, le *contig* XV, constitué par les deux cosmides CNX2 et CNX4 partageant le même site NotI (figure 3), a été localisé à environ 650-700 kilobases (kb) du gène *met* [15]. L'étape suivante a consisté à rechercher systématiquement les régions riches en doublet CpG dont la cytosine n'est pas méthylée, ce qui définit des îlots dits HTF (*HpaII* tiny fragments) qui caractérisent les régions de l'ADN renfermant des gènes activement transcrits dans tous les types cellulaires [16]. Le clone génomique CS7 répondait à ces critères (figure 3).

Le gène de la mucoviscidose ?

Comment démontrer que la région CS7 est le bon candidat ? Les critères suivants ont été rassemblés [15] : (a) séquence conservée dans le règne animal (technique dite du *zoo-blotting*); (b) séquence exprimée dans un grand nombre de tissus dont le poumon; (c) présence d'un RFLP ne recombinant plus avec le locus CF dans les familles où une recombinaison avait été observée avec les sondes *met* et pJ 3.11 (figure 4). Deux nouveaux RFLP du contig XV, définis par les sondes XV2c et KM19 situées de part et d'autre de CS7 (figure 3), ont également montré leur plus grande proximité génétique du locus CF, notamment dans une famille que nous avons étudiée (figure 4A).

Ces critères, tous indirects, ne sont cependant pas suffisants pour affirmer que le gène est atteint. La preuve définitive, et non encore apportée, serait de démontrer que

Tableau I						
DISTRIBUTION DES HAPLOTYPES CONSTITUÉS A PARTIR DES SONDAS pXV2c et pCS7 [15]						
Les allèles 1 et 2 codent respectivement pour les grands et les petits fragments de restriction définis pour chaque couple sonde/enzyme.						
Haplotypes	pXV-2c Taq 1	pCS.7 Hha 1	Normaux	(%)	CF	(%)
A	1	1	9	14	0	0
B	1	2	22	34	66	94
C	2	1	28	43	0	0
D	2	2	6	9	4	6
Total			65	100	70	100

le gène candidat est muté chez les mucoviscidosiques. Nos résultats très récents et non encore publiés ont cependant montré l'existence d'une recombinaison entre KM19-XV2c et le locus CF (figure 4B), ce qui exclut le gène candidat CS7. Ceci illustre bien la difficulté de l'entreprise lorsque l'on ne dispose pas d'un fil conducteur*. Le gène CF ne doit pas être loin de la région étudiée, et même si l'absence d'indice complique singulièrement la tâche en rendant nécessaire un fastidieux travail de séquençage dans un génome mucoviscidosique, le succès est garanti à terme.

Un résultat inattendu

La mutation CF résulterait d'un effet fondateur. Le travail publié par l'équipe de R. Williamson en avril 87 [15] contient en outre des résultats tendant à prouver une origine monocentrique de la mutation CF. En effet, dans la population d'Europe du Nord, il

existe un déséquilibre de liaison significatif entre un certain haplotype et le gène CF muté (Tableau I): l'haplotype B est retrouvé dans 94 % des chromosomes portant l'anomalie CF et 34 % des chromosomes normaux, alors que l'haplotype C est retrouvé dans 0 % des chromosomes portant CF et 43 % des normaux. Ce très fort déséquilibre de liaison est en faveur d'un événement fondateur récent. Il reste à déterminer si cet événement est unique ou s'il est survenu indépendamment dans des populations différentes**. L'existence d'un même déséquilibre de liaison dans différentes populations d'Europe serait très en faveur d'un fondateur originel unique, et donc d'une mutation univoque. Cette découverte est d'une portée considérable, d'une part parce que l'existence d'un allèle pathologique unique permet d'envisager un diagnostic prénatal génotypique, et à terme phénotypique (lorsque le produit du gène sera identifié) considérablement simplifié, d'autre part parce que, pour la première fois, une détection pourrait être envisagée chez les porteurs sains, sans réclamer le « préavis » d'un cas index. Il reste à expliquer la remarquable dissémination de ce gène pathologique dans la population européenne, laquelle suggère un avantage sélectif des hétérozygotes ■

* Rappelons que dans le cas de la myopathie de Duchenne l'existence de délétions a représenté un atout majeur.

** Nos résultats préliminaires tendent à montrer que ce déséquilibre de liaison existe aussi en France.

Summary

Cystic fibrosis (CF) is one of the most frequent autosomal recessive disorder in man. One child is affected among 1 600 newborns; one individual among 20-25 is a silent heterozygous carrier. Classical biochemical methods have failed to uncover the abnormal defective gene product, and the responsible gene is not known. Molecular genetics gives us now the possibility to reach this unknown gene, to read its sequence and to deduce the nature and the function of the corresponding protein. The molecular pathology of the abnormal gene will thus be deciphered, and ultimately the malfunction of the affected cells understood. This strategy, going from the disease to the gene and then to its function is called « reverse genetics ». We describe the different steps of this search: (1) localization of the morbid locus, called CF, on chromosome 7; (2) regional mapping of the CF locus; (3) approaching the CF gene itself. The final goal, ie reaching and deciphering the CF gene, is in view but not reached yet. Nonetheless several useful restriction fragment length polymorphisms (RFLP) closely linked to the CF locus, allow us already to perform a safe prenatal detection of foetuses at risk. A strong linkage disequilibrium between a particular haplotype and the abnormal CF gene was unexpectedly found in Northern — and Western — Europe, suggesting that a same abnormal founding allele has widespread in these populations.

Remerciements

Paule Warren est boursière de l'association française de lutte contre la mucoviscidose.