

m/s

médecine/sciences 1988 ; 4 : 138-140

## GRANDEUR ET SERVITUDES DE LA GÉNÉTIQUE INVERSE

**Bertrand Jordan**  
Directeur de recherche  
au Cnrs

**O**n parle beaucoup de «génétique inverse» (ou génétique reverse en français), et c'est à juste titre. Comme le montrent plusieurs articles de ce numéro, cette voie d'approche aboutit à des succès spectaculaires pour des maladies génétiques comme la mucoviscidose, la myopathie de Duchenne ou bientôt la chorée de Huntington.

De quoi s'agit-il en fait ? Tout simplement d'inverser la démarche utilisée jusqu'à récemment en pathologie génétique humaine. Dans le passé, par exemple pour les hémoglobinopathies ou les hémophilies, l'on est parti de la connaissance de la protéine (globine, facteur VIII ou IX de coagulation) dont l'absence ou la défectuosité sont responsables de l'affection en cause. La connaissance de la protéine a permis, par diverses voies et après un délai plus ou moins long, l'isolement du gène correspondant. Une fois ceci réalisé, l'analyse moléculaire du gène porté par les individus atteints peut être entreprise et un diagnostic prénatal peut, le cas échéant, être mis en place ; les possibilités de traitement de la maladie peuvent être améliorées si la production du facteur manquant chez l'individu atteint est possible par génie génétique — en dehors de perspectives de thérapie génétique dont les possibilités et l'impact réel me paraissent largement surestimés dans les médias et même dans le milieu scientifique.

**Au départ, la liaison génétique.** La «génétique inverse», elle, part du caractère héréditaire de la maladie, sans qu'il soit nécessaire au départ d'avoir une idée précise sur la nature du défaut moléculaire ni même du tissu dans lequel ce défaut se manifeste. Une analyse purement génétique menée sur des familles aussi larges et complètes que possible étudie les liaisons génétiques éventuelles entre la maladie et un ensemble de points de repères répartis sur l'ensemble des chromosomes. Ces points de repères sont constitués dans la plupart des cas par des polymorphismes au niveau de l'ADN (les RFLPs, *restriction fragment length polymorphisms*) détectés par des sondes clonées ; leur organisation en une carte génétique complète couvrant l'ensemble des chromosomes humains, proposée en 1980 [1], est maintenant bien avancée [2] et permet de disposer d'une carte de référence et d'un jeu étendu de sondes.

### RÉFÉRENCES

1. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 1980 ; 32 : 314-31.
2. Donis-Keller H, Green P, Helms C, et al. A genetic linkage map of the human genome. *Cell* 1987, 51 : 319-37.
3. Jordan BR. Megabase methods: a quantum jump in recombinant DNA techniques. *Bioessays* 1988 (sous presse).
4. Schwartz DC, Cantor CR. Separation of yeast chromosome sized DNAs by pulsed field gradient gel electrophoresis. *Cell* 1984 ; 37 : 67-75.
5. Burke DT, Carle GF, Olson MV. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 1987, 236 : 806-12.
6. Collins FS, Weissman SM. Directional cloning of DNA fragments at a large distance from an initial probe: a circularization method. *Proc Natl Acad Sci USA* 1984 ; 81 : 6812-6.
7. Hoffman EP, Brown RH, Kunkel LM. Dystrophin: the protein product of the Duchenne muscular dystrophy locus. *Cell* 1987 ; 51 : 919-28.
8. Bird AP. CpG rich islands and the function of DNA methylation. *Nature* 1986 ; 321 : 209-13.

### ADRESSE

B. Jordan : CIML, Inserm, Cnrs, case 906, 13288 Marseille Cedex 9, France.

---

L'analyse menée dans les familles détermine dans un premier temps le chromosome sur lequel doit se trouver le gène impliqué dans la maladie en question ; la région chromosomique peut ensuite être précisée en continuant l'étude de liaison. Parfois, des données issues de la cytogénétique permettent d'accélérer ce processus en associant, par exemple, le phénotype atteint et une translocation spécifique. On en arrive finalement à situer le gène recherché à quelques centimorgans d'un repère connu (un gène ou un segment d'ADN cloné).

**Problèmes d'échelle.** Le problème est alors que cette distance de quelques centimorgans (un centimorgan correspond à une fréquence de recombinaison de 1 %), petite aux yeux du généticien, est énorme à ceux du biologiste moléculaire puisqu'en moyenne une distance génétique d'un centimorgan correspond chez l'homme à une distance physique d'un million de bases le long de l'ADN. Or les techniques classiques de génétique moléculaire ne permettent d'étudier l'ADN (par cartographie ou clonage) que sur des distances de quelques dizaines de milliers de bases. Le passage de la proximité génétique à l'identification effective et au clonage du gène n'est donc pas aisé. Dans le passé récent, il a souvent pu être réalisé grâce à l'apport de la cytogénétique : l'existence de petites délétions ou de translocations entraînant le phénotype atteint indiquait que la délétion ou la translocation avaient interrompu le gène impliqué et donc localisaient précisément ce gène, en facilitant souvent l'isolement. Un exemple maintenant classique est l'isolement de séquences du gène DMD (dystrophie musculaire de Duchenne) par le groupe de Worton à partir de telles translocations et par celui de Kunkel à partir de délétions (*voir l'article de synthèse de J. Chelly et J.-C. Kaplan dans ce numéro, p. 141*).

*m/s n° 3 vol. 4, mars 88*

Mais en l'absence de tels accidents chromosomiques, le biologiste moléculaire en était réduit à l'utilisation de techniques comme la marche le long du chromosome à l'aide de bibliothèques d'ADN dans des cosmides, méthode lourde permettant d'avancer par pas de 20 ou 30 kilobases (kb) (1 kb = 1 000 bases) au prix d'un important travail : parcourir ainsi 1 000 ou 2 000 kilobases semblait hors de question.

**Les méthodes mégabase.** Cependant depuis peu de temps un ensemble de nouvelles méthodes que je regroupe sous le nom de « méthodes mégabase » [3] permet d'envisager le passage de la liaison génétique au clonage effectif avec plus de sérénité. Il s'agit, au niveau analytique, de l'électrophorèse en champ pulsé [4]; au niveau du clonage, de la construction de chromosomes artificiels dans la levure [5], et enfin, au niveau de la marche le long du chromosome, des techniques de « bibliothèques de saut » [6]. Le point commun entre ces trois techniques d'apparition récente est qu'elles travaillent à une échelle de l'ordre de la mégabase ( $10^6$  paires de bases, 1 000 kb) donc à l'échelle requise pour la génétique inverse : l'électrophorèse en champ pulsé, maintenant répandue dans de nombreux laboratoires, permet de séparer des fragments d'ADN allant de quelques centaines à quelques milliers de kb, obtenus par coupure avec les enzymes de restriction dont le site est très peu fréquent dans l'ADN ; les YAC (*yeast artificial chromosomes*) permettent de construire des bibliothèques d'ADN humain dans la levure avec des clones contenant chacun un fragment d'ADN long de plusieurs centaines de kb, enfin les bibliothèques de saut permettent de se déplacer le long du chromosome avec un « pas » du même ordre.

Il est donc vraisemblable que, grâce à ces méthodes, l'isolement du gène impliqué dans une mala-

die génétique donnée ne sera plus aussi dépendante du repérage (toujours aléatoire) de délétions ou de translocations induisant le phénotype atteint et dont on peut donc penser qu'elles interrompent le gène.

Un point à ne pas négliger, enfin, est que l'approche par la génétique inverse peut révéler la nature du défaut moléculaire responsable d'une affection donnée : dans une maladie dont on ne connaissait que les symptômes, elle peut indiquer la protéine dont la structure est altérée ou l'expression modifiée. Il y a donc réel progrès dans la compréhension de la maladie et par là même espoir d'une thérapeutique dont l'amorce était précédemment impossible puisqu'on n'en connaissait pas la cible. Il semble donc que les nouveaux outils du génie génétique, depuis l'analyse de liaison à l'aide de sondes révélant des polymorphismes jusqu'aux méthodes mégabase, permettent d'étudier systématiquement et avec de bonnes chances d'aboutir tous les cas de maladies héréditaires humaines.

**Tout n'est pas si simple...** On constate pourtant, à côté de nombreux succès, des piétinements persistants, des maladies clairement héréditaires dont l'élucidation n'avance pas contrairement aux espoirs. Un exemple : le retard mental lié à la fragilité du chromosome X, retard mental profond et fréquent (un cas sur 2 000 garçons) dont on sait maintenant depuis plusieurs années qu'il doit impliquer un ou des gènes situés dans une région bien précise (interface q27-q28) du chromosome humain le mieux connu, le chromosome X. Malgré les efforts déployés par de nombreux groupes et l'utilisation des techniques les plus performantes l'étude moléculaire de ce syndrome n'a guère progressé depuis deux ou trois ans. Je pense que cet exemple comme beaucoup d'autres manifeste une des

grandes faiblesses de la technologie actuellement utilisée : celle-ci est très performante sur le plan structural et capable de définir assez rapidement — moyennant des efforts importants et la mise en œuvre simultanée de toute une série de techniques — la région chromosomique dans laquelle doit se trouver le gène dont le dysfonctionnement est responsable d'une maladie donnée. Cette technologie peut aussi, et de plus en plus avec les méthodes mégabase, aboutir au clonage de la région (quelques centaines de kilobases) dans laquelle doit se trouver le gène impliqué. Mais l'identification du gène reste problématique si l'on n'a pas quelques informations sur la nature probable du produit affecté et sur le tissu dans lequel le gène qui le code s'exprime. Si, par exemple, l'on a pu identifier le produit du gène de la myopathie de Duchenne [7] (en fait, du gène qui, lorsqu'il est défectueux, est responsable de cette myopathie), c'est en grande partie parce qu'on pouvait s'attendre presque à coup sûr à ce qu'il s'agisse d'une protéine exprimée dans le tissu musculaire. De plus la taille du gène et sa coïncidence avec le locus défini par la génétique et la cytogénétique ne laissent guère place au doute sur le fait que c'est bien le produit codé par ce gène qui est impliqué. Dans le cas du retard mental lié à la fragilité de l'X, comme dans celui de la chorée de Huntington, la définition du gène est rendue beaucoup plus difficile du fait que le type du produit et le tissu dans lequel le gène s'exprime ne sont a priori pas définis.

**Mais l'espoir est permis !** Des amorces de réponse à ces questions existent. Sur un plan purement structural, il se confirme que des séquences d'ADN particulières et facilement reconnaissables, les îlots HTF\* [8], signalent la présence de gènes, souvent d'expression ubiquitaire. Ces séquences permettent donc d'accéder directement à au moins une

partie des gènes contenus dans la région étudiée. D'autre part, le séquençage systématique de l'ADN permet de détecter les cadres de lecture ouverts, les ORFs (*open reading frames*), qui ont toutes les chances de correspondre à des gènes ; ce séquençage donne aussi la structure des protéines correspondantes et peut, après recherche d'homologie dans les banques de données, donner une idée sur leur fonction possible : cette approche par la séquence prendra évidemment beaucoup d'importance si le projet de séquençage total du génome humain se réalise. L'apport de ces techniques n'est pas négligeable puisque les 50 000 gènes (environ) du génome humain n'occupent vraisemblablement qu'environ 10 % de ce dernier, le reste étant constitué de séquences dont la fonction (si elle existe) reste inconnue. Toute approche menant directement aux gènes est donc bonne à prendre ! Mais cela ne suffit pas, et je crois beaucoup au développement des techniques permettant de détecter une expression différentielle (d'un tissu à l'autre, ou du sain au malade) de gènes codés dans un grand segment d'ADN. C'est à ce niveau-là, que le clonage dans la levure — s'il se révèle réellement opérationnel — pourra jouer un rôle essentiel : ce système permet en effet de disposer sous une forme clonée (donc à la fois pure et abondante) de l'ensemble des séquences d'ADN contenues dans une région de quelques centaines de kilobases. Il devient alors possible avec des sondes très hétérogènes correspondant à l'ARN messager exprimé dans tel ou tel tissu d'obtenir des signaux interprétables et

par exemple de déterminer si le grand segment cloné contient une séquence exprimée chez l'individu sain mais non chez le malade.

Enfin, il faut bien voir que la pratique de la génétique inverse impose des contraintes nouvelles. Elle demande, pour être efficace, l'application à un problème donné d'une très vaste panoplie de techniques dont chacune est relativement lourde à établir et à maintenir en bon état de fonctionnement : l'analyse de liaison suppose à la fois un gros travail expérimental et une informatique performante ; la cytogénétique, l'établissement d'hybrides somatiques classiques ou par transfert de chromosomes et leur conservation demandent beaucoup de technicité et de rigueur ; enfin les « méthodes mégabase » pour puissantes et séduisantes qu'elles soient restent d'un maniement délicat. Et bien sûr toute la panoplie du génie génétique « classique » reste nécessaire. Il semble que l'ère du « bricolage » est passée, et que dans ce domaine en évolution rapide il est nécessaire soit de constituer des laboratoires importants et actifs réunissant un ensemble de connaissances complémentaires (un peu à l'image du pôle que tend à devenir, pour prendre un exemple en Europe, le laboratoire de l'*Imperial Cancer Research Fund* à Londres), soit de développer des collaborations beaucoup plus intenses et étroites que nous n'en avons l'habitude en France, entre chercheurs des différents domaines mais aussi entre chercheurs et cliniciens. Un des freins au développement de la génétique inverse en France a certainement été non seulement l'absence des « registres des malades » aussi précis et performants que dans d'autres pays, mais aussi les difficultés de collaboration et d'accès aux malades. Il est urgent que cela change si nous voulons contribuer de façon significative aux succès de la génétique inverse ■

\* Les îlots HTF sont des zones de l'ADN riches en séquences CpG non méthylées et qui sont découpées en petits fragments par l'enzyme *Hpa II*, d'où leur nom (*Hpa II* Tiny Fragments).