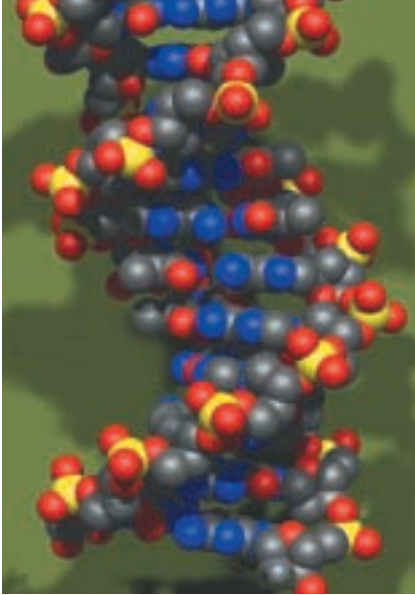




# La course aux gènes : et après?

**BRIGITTE CHAMAK**

*Bien que le génome humain soit sur le point d'être séquencé, l'épopée de la génétique, débutée il y a 100 ans, est loin d'être achevée. Se mêlent toujours espoirs et inquiétudes.*

**S**équençage, génomique, postgénomique, bio-informatique... Le sens des termes n'est pas toujours explicite, mais, à l'évidence, les enjeux économiques, juridiques et sociaux sont considérables. Le séquençage de l'ADN, l'acide désoxyribonucléique, ne représente qu'une première étape : si l'on dispose désormais du répertoire des fragments d'ADN, il reste à déterminer – tâche considérable – les fonctions qui leur sont associées. L'industrie pharmaceutique, à l'affût de nouvelles stratégies thérapeutiques, finance le développement de ce secteur. Au début des années 1980, les laboratoires pharmaceutiques comprennent que l'identification des gènes est un moyen efficace pour concevoir de nouvelles cibles thérapeutiques. Depuis quand, toutefois, nous intéressons-nous aux gènes?

Les «lois de l'hérédité» énoncées par Gregor Mendel, en 1866, formalisent la transmission, par le père et par la mère, des éléments qui déterminent les caractéristiques observables des êtres vivants, ou, plus simplement, la transmission de ce que l'on nommera plus tard les gènes. À cette époque, on ignore encore que le support matériel de ces éléments sont les chromosomes. La distinction entre le génotype – l'ensemble des gènes – et le phénotype – l'ensemble des caractères – est introduite en 1909 par le généticien danois Wilhelm Johannsen, mais la génétique ne se développe vraiment qu'avec les travaux du biologiste américain Thomas Morgan. Ce dernier a choisi d'étudier la drosophile (la mouche du vinaigre), dont la rapidité de reproduction facilite l'étude génétique de ses caractères. Les résultats sont nets : les caractères héréditaires sont portés et transmis par les chromosomes. En étudiant les phénomènes de recombinaison génétique (au cours de la division des cellules, certains fragments de chromosomes s'échangent), les «généticiens» du groupe de

Morgan établissent des cartes chromosomiques et localisent certains gènes.

En 1944, l'Américain Oswald Avery et ses collègues démontrent que les gènes sont formés d'ADN, mais la structure en reste inconnue. Or, juste avant la Seconde Guerre mondiale, le groupe de Lawrence Bragg, en Angleterre, et celui de Linus Pauling, en Californie, avaient montré que, par diffraction des rayons X, on peut décrire la structure tridimensionnelle des molécules cristallisées. À partir des données cristallographiques de molécules d'ADN, obtenues par diffraction des rayons X par Maurice Wilkins et Rosalind Franklin du King's College de Londres, James Watson et Francis Crick, qui travaillent dans le laboratoire Cavendish, à Cambridge, commencent à élaborer des modèles de la molécule d'ADN. S'inspirant des travaux de Linus Pauling, qui avait montré, en 1950, que l'hélice est une structure importante, ils envisagent plusieurs modèles d'hélices, et, en avril 1953, proposent le modèle de la double hélice avec appariement de bases complémentaires. En mai 1953, ils publient un article dans la revue *Nature*, où ils proposent, d'après la structure, un modèle simple de réplication de l'ADN.

## LES DÉBUTS DU GÉNIE GÉNÉTIQUE

Dès 1962, les biologistes tentent de transformer directement des cellules d'organismes supérieurs par un ADN étranger, mais la manipulation des gènes reste difficile. Ce n'est qu'en 1972 qu'est produite, *in vitro*, la première recombinaison génétique entre molécules d'ADN d'origines différentes. Cette expérience, réalisée par le biochimiste américain Paul Berg et ses collègues de l'Université Stanford, préfigure l'avènement du génie génétique. Dès l'annonce de ces résultats, l'inquiétude s'empare des chercheurs, qui craignent la fabrication et la dissémination de bactéries ayant intégré des gènes res-

ponsables, par exemple, de cancers. En 1974, Berg propose un moratoire sur ces recherches, et l'organisation d'une conférence chargée de définir les conditions de réalisation des expériences futures. Au cours de cette conférence, qui se tient à Asilomar, en Californie, en février 1975, les biologistes précisent la nature des risques liés à ces manipulations et les précautions nécessaires : les expériences devront être faites dans des conditions de confinement physique et biologique strictes.

En 1977, l'Anglais Frederick Sanger, à Cambridge, met au point une méthode enzymatique de séquençage de l'ADN : grâce à un marquage radioactif, il identifie le nombre et la position des nucléotides qui constituent l'ADN. La même année, Allan Maxam et Walter Gilbert proposent une autre méthode de séquençage qui utilise aussi un marquage radioactif, et des réactifs chimiques qui découpent l'ADN en des positions précises, assurant le repérage des nucléotides. La mise au point de ces méthodes de séquençage, ainsi que d'autres techniques de génie moléculaire autorisent la première production industrielle par génie génétique, d'insuline par la bactérie *Escherichia coli*. La bataille fait rage entre les laboratoires de la côte Ouest des États-Unis et ceux de la côte Est, notamment entre l'Université de Californie et la Société *Genentech*, à San Francisco, et l'Université Harvard, près de Boston.

En 1981, l'Américain Leroy Hood fonde la Société *Applied BioSystems*, spécialisée dans la production d'appareils de laboratoire pour la biologie moléculaire ; en 1986, il fabrique un appareil de séquençage automatique. Dès lors, il multiplie les démarches auprès du Congrès américain pour le convaincre de l'importance du séquençage du génome humain.

En France, Jean Dausset, qui a découvert les antigènes majeurs d'histocompatibilité, les molécules du «soi», crée,

en 1982, grâce à un legs important, le Centre d'étude du polymorphisme humain, le CEPH. Daniel Cohen, un de ses élèves, Jean Weissenbach et Marc Lathrop travaillent dans ce centre et y étudient les gènes qui codent les antigènes majeurs d'histocompatibilité. En 1987, Bernard Barataud, président de l'Association française contre les myopathies, l'AFM, organise le premier téléthon et décide de faire construire, avec les fonds obtenus, le Généthon, à Évry. Daniel Cohen l'avait persuadé que le séquençage du génome humain accélérerait la connaissance des maladies génétiques. Au cours de la première étape, nommée Généthon I, les généticiens du Centre d'étude du polymorphisme humain cartographient le génome humain : à la façon des cartographes classiques ils repèrent des «marqueurs» qui jalonnent les chromosomes. Cette cartographie est établie en 1991 et en 1992. La deuxième étape, le Généthon II, consiste alors à détecter des gènes dits morbides associés aux maladies d'origine génétique : c'est le clonage positionnel. La troisième étape est le séquençage du génome humain.

Au début des années 1990, aux États-Unis, le Congrès américain lance le *Programme génome humain*. Craig Venter travaille alors à l'Institut américain de recherche médicale, le NIH – organisme public de recherche. En juin 1991, il dépose une demande de brevets pour des séquences d'ADN dont on ignore la fonction. La communauté scientifique internationale réproue de façon quasi unanime cette demande : elle s'inquiète de voir récompenser un travail purement technique au détriment de travaux de recherche plus fondamentaux. Le bureau américain des brevets rejette la demande, au motif que G. Venter n'a pas été capable de démontrer l'utilité commerciale de ces séquences. En 1992, ce dernier quitte le NIH pour fonder un centre privé de

séquençage ; avec l'aide de la société *Perkin-Elmer*, qui fabrique des automates de séquençage, il crée, en 1998, la société *Celera Genomics*. *Perkin-Elmer* qui, en 1997, avait racheté la société créée par L. Hood, met à la disposition de G. Venter 300 machines de séquençage, un ordinateur très puissant et des moyens considérables. Leur objectif : séquencer le génome humain avant les organismes publics. En septembre 1999, le laboratoire *Celera Genomics* achève le séquençage du génome de la drosophile et, le 6 avril 2000, C. Venter annonce que sa société a établi la séquence complète du génome d'un être humain.

En 1996, le *Human Genome Project* (Projet génome humain), organisme public qui rassemble divers instituts de recherche américains et un institut britannique, *Wellcome Trust*, s'était fixé cet objectif. Il semble que *Celera Genomics* l'ait devancé. Pourtant, le séquençage n'est pas achevé : si *Celera Genomics* a bien fragmenté, de façon aléatoire, un génome, dont chaque segment a été séquencé, il reste à ordonner toutes ces séquences. *Celera Genomics* profite des travaux réalisés dans les organismes publics du monde entier qui, eux, partagent leurs données. Grâce à un ordinateur puissant, C. Venter pourra affirmer, dans quelques mois, qu'il a établi la séquence complète du génome humain.

## DU GÉNOME AU PROTÉOME

Ce séquençage n'est qu'une toute première étape. Parmi les trois milliards de paires de bases séquencées, moins de cinq pour cent correspondent à des gènes codants. Comme il n'est pas très intéressant de disposer de séquences d'ADN dont on ne connaît pas les fonctions dans l'organisme, la prochaine étape sera de repérer les séquences significatives : nous entrerons dans l'ère de la postgénomique.

Les généticiens détermineront quelles sont les séquences géniques actives selon les tissus et leur stade de développement, et préciseront les fonctions de milliers de gènes et celles des protéines correspondantes. Les laboratoires pharmaceutiques y puiseront de nouvelles cibles thérapeutiques.

Ainsi la postgénomique ne concerne plus seulement l'ADN et les gènes, mais tout ce qui lie les gènes aux protéines. L'ADN est transcrit par une enzyme, l'ARN polymérase, en molécules d'ARN messager, elles-mêmes traduites en protéines. Cette traduction est fondée sur une correspondance entre les nucléotides qui constituent les ARN et les acides aminés des protéines. Si l'ADN du génome humain comprend entre 35 000 et 40 000 gènes, l'ensemble des ARN messagers, qui constitue le transcriptome, compte 100 000 à 150 000 entités ; quant au nombre de protéines du protéome, il est de l'ordre du million.

La première étape est d'identifier les séquences des ARN messagers activés dans un tissu donné. Pour ce faire, on utilise l'analyse en série de l'expression des gènes (SAGE, d'après les initiales de *Serial Analysis of Gene Expression*) mise au point en 1995 et les puces à ADN. La méthode SAGE permet une analyse de la fréquence d'un ARN messager parmi les milliers produits dans une cellule à un moment donné : une séquence d'une dizaine de nucléotides suffit à constituer une «étiquette» caractéristique de chacun des ARN messagers d'une cellule. Les puces à ADN sont de minuscules surfaces de verre ou de silicium où sont ancrées des milliers de séquences d'ADN connues ; ces dernières fixent leurs séquences complémentaires, de sorte que l'on peut ainsi identifier les séquences contenues dans l'ADN analysé.

Pour être exploitables, toutes ces connaissances doivent être classées sous forme de banques de données.



AFM



AFM

Pour étudier les gènes responsables de maladies génétiques, on prélève des cellules aux malades et on les place dans des tiroirs d'échantillons (a), lesquels sont conservés dans des cuves emplies d'azote



AFM / L. Morvan

liquide (b). Pour préparer les «méduses d'ADN» (c), on multiplie, *in vitro*, des globules blancs, puis on les fait éclater, on sépare par extraction l'ADN, et on le fait précipiter grâce à une solution d'alcool et de sel.