

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 réproouve 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.



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



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.