

loppement d'un être humain, de l'œuf fécondé à l'adulte, résulte d'une modification bien orchestrée de l'expression des gènes dans les tissus.

Notre raisonnement était donc le suivant : si nous connaissions les gènes exprimés dans les tissus sains et dans les tissus pathologiques, nous pourrions identifier les protéines nécessaires au bon fonctionnement de ces tissus et les anomalies qui déclenchent les maladies. Nous pourrions alors mettre au point de nouveaux tests diagnostiques et des traitements qui compenseraient l'activité des protéines ou des gènes défectueux.

Peut-on ainsi établir l'anatomie humaine, non plus à l'échelle de l'organe, mais à celle des molécules ? L'identification exhaustive de tous les gènes exprimés dans tous les tissus de l'organisme serait une tâche titanesque : chaque cellule contient quelque 100 000 gènes, et les gènes (environ 15 000) qui s'expriment dans chaque cellule changent d'un type de cellules à l'autre. L'étude d'un ou de deux types cellulaires seulement ne fournirait aucune information sur les gènes exprimés ailleurs. En outre, l'anatomie moléculaire ne serait véritablement connue que

si l'on explorait les tissus aux stades successifs du développement, et le tout aussi bien dans les tissus malades que dans les tissus sains.

Les progrès techniques récents ont rendu cette tâche faisable : les biologistes peuvent trouver rapidement quels sont les gènes exprimés dans n'importe quel tissu. Notre programme s'est révélé le plus rapide pour l'identification de gènes thérapeutiques potentiels.

Prenons l'exemple de l'athérosclérose : au cours de cette maladie, une substance grasseuse s'accumule dans les vaisseaux qui alimentent le cœur. Les techniques que nous avons mises au point procurent une liste des gènes exprimés dans les artères normales, avec une mesure de cette expression pour chaque gène. On peut ensuite comparer cette liste à celle établie chez des personnes atteintes d'athérosclérose : les gènes (et donc les protéines) qui diffèrent entre la personne saine et la personne malade sont responsables de la maladie. On détermine aussi dans quelle mesure la maladie agit sur l'expression de ces gènes.

Connaissant les gènes, on fabrique ensuite les protéines qu'ils codent et, une fois les protéines purifiées, les bio-

logistes préparent des tests pour les doser chez les patients. Un test qui révélerait une production excessive d'une protéine des plaques d'athérome indiquerait une prédisposition à l'athérosclérose : une détection précoce autoriserait un traitement plus efficace. De plus, les pharmacologues pourraient utiliser des protéines purifiées pour trouver de nouveaux médicaments : tout composé inhibant la production d'une protéine de la plaque d'athérome est un médicament potentiel.

Notre méthode est un peu à l'écart des grandes recherches actuelles en génétique humaine. De nombreux scientifiques participent au séquençage du génome humain, une collaboration internationale destinée à établir la séquence complète des bases de l'ADN humain. Ces informations seront essentielles pour l'étude du rôle des gènes et de leur évolution, et pour celle des maladies héréditaires. Toutefois, ce séquençage n'est pas le moyen le plus rapide pour découvrir des gènes, car les gènes ne représentent qu'une petite partie de l'ADN. De plus, le programme ne précise pas quels sont les gènes impliqués dans des maladies.

La fabrication des molécules d'ADN complémentaires

Pour fabriquer les protéines, les cellules utilisent des ARN messagers. Pour trouver des gènes, nous n'avons pas utilisé les molécules d'ARN messager, trop fragiles, mais des copies de cet ARN messager, l'ADN complémentaire. Nous les obtenons en mélangeant une molécule d'ARN messager avec des nucléotides (les sous-unités de ces molécules) et une enzyme, la transcriptase inverse, qui recopie l'ARN sous forme d'ADN. Nous obtenons un ADN complémentaire que nous cherchons ensuite à recopier en plusieurs exemplaires, afin de déterminer les bases qui les composent. Les biologistes moléculaires ont mis au point des techniques pour insérer les molécules d'ADN complémentaires dans le génome de bactéries ; étalées sur un gel, elles se multiplient, fabriquant des copies de l'ADN complémentaire intégré.

Puis un robot repère les colonies qui ont effectivement intégré les ADN complémentaires et il les prélève. Le robot reconnaît les colonies par leur couleur : les bactéries qui n'ont pas intégré l'ADN complémentaire produisent un pigment bleu. Le robot prélève chaque jour environ 10 000 colonies bactériennes. L'ADN complémentaire des colonies sélectionnées est assez abondant pour être analysé, puis purifié automatiquement. L'ensemble des molécules d'ADN complémentaire provenant d'un même tissu forme ce que l'on nomme une banque.

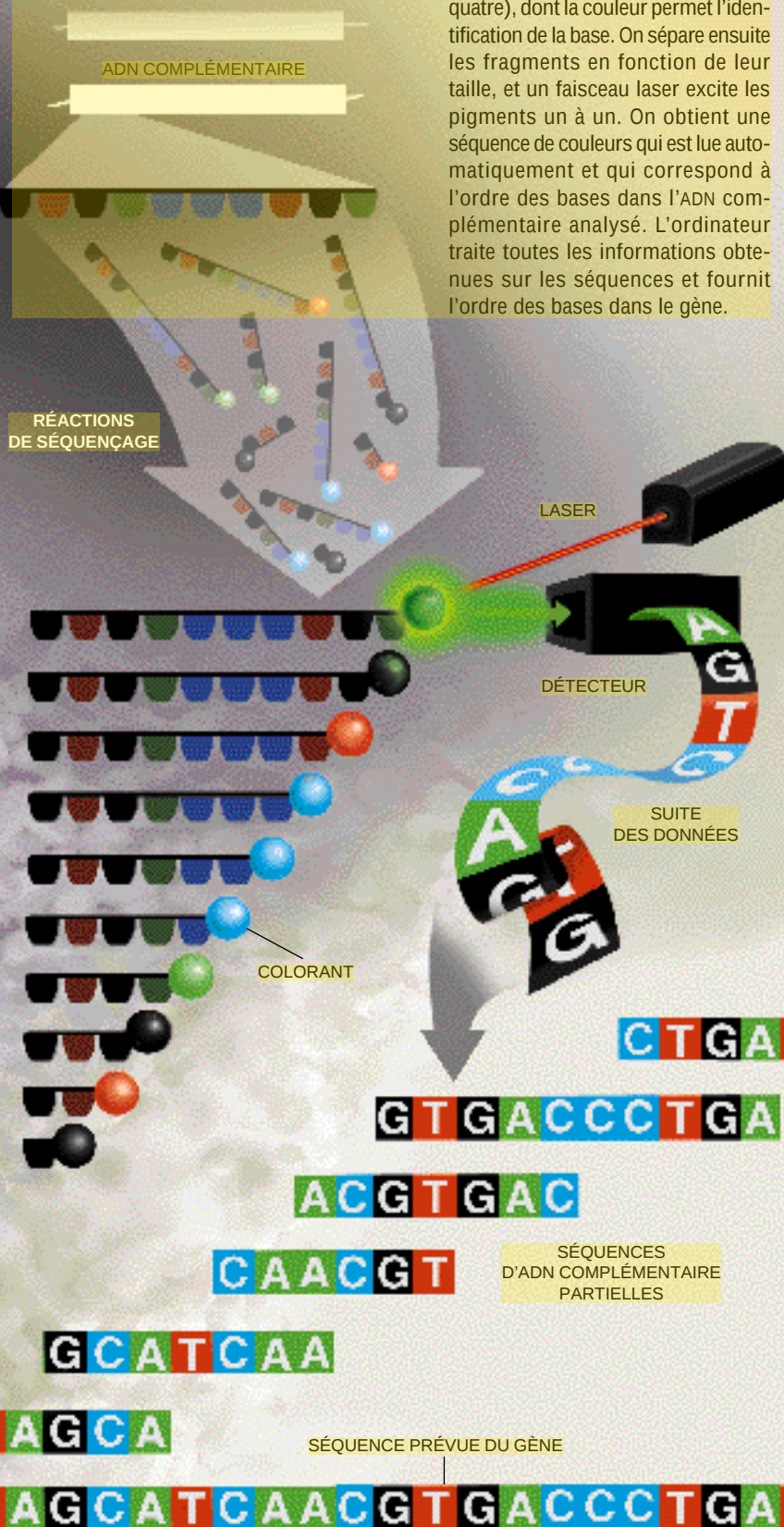
Les chercheurs de la Société HGS ont préparé des banques d'ADN complémentaires humains pour presque tous les organes et tous les tissus sains, ainsi que pour de nombreux organes et tissus pathologiques.



Comment trouver une séquence partielle d'ADN complémentaire?

Pour obtenir des séquences d'ADN complémentaires, les biologistes «découpent» des copies d'une molécule d'ADN complémentaire, fabriquant ainsi toute une gamme de fragments dont les longueurs diffèrent seulement d'une base. La base située à une extrémité de chaque fragment est associée à un pigment fluorescent (il y en a

quatre), dont la couleur permet l'identification de la base. On sépare ensuite les fragments en fonction de leur taille, et un faisceau laser excite les pigments un à un. On obtient une séquence de couleurs qui est lue automatiquement et qui correspond à l'ordre des bases dans l'ADN complémentaire analysé. L'ordinateur traite toutes les informations obtenues sur les séquences et fournit l'ordre des bases dans le gène.



En 1992, nous avons créé la société HGS et commencé à collaborer avec l'Institut pour la recherche génomique, dirigé par Craig Venter ; six mois plus tard, le Laboratoire pharmaceutique SmithKline Beecham nous rejoignait. Puis, d'autres laboratoires se joignent à nous : Schering-Plough, Merck, Synthelabo et Takeda.

La fabrication de gènes

Pourquoi s'intéresse-t-on aux gènes, puisque les médicaments sont constitués de protéines et non de gènes? En théorie, on pourrait analyser directement les protéines cellulaires. Toutefois, on ne sait pas les fabriquer, même quand on en connaît la composition ; or, pour produire des médicaments, il faudrait synthétiser des quantités notables de protéines identiques. On atteint cet objectif en isolant les gènes intéressants et en les transplantant dans des cellules qui les expriment.

Notre méthode d'identification des gènes est fondée sur l'analyse d'un produit intermédiaire, formé lorsque les gènes s'expriment : l'ARN messager. Les enzymes qui synthétisent les ARN messagers copient les gènes : le message codé est identique sur le brin d'ADN copié et sur la molécule d'ARN synthétisée. L'ARN messager sert ensuite de modèle pour la synthèse de la protéine spécifique codée par le gène. Les ARN messagers nous sont utiles, parce qu'ils sont synthétisés uniquement lorsque les gènes correspondants sont actifs. La séquence de l'ARN, copie de la séquence du gène, donne suffisamment d'informations pour que l'on isole le gène de l'ADN cellulaire et pour que l'on fabrique la protéine correspondante.

Comme l'ARN messager est très fragile, nous avons plutôt manipulé des ADN complémentaires, c'est-à-dire des molécules d'ADN qui sont des copies stables des ARN messagers. Pour synthétiser les molécules d'ADN complémentaires, nous mélangeons les molécules d'ARN messager avec des nucléotides (les sous-unités de l'ADN) et une enzyme, la transcriptase inverse, qui lit les sous-unités de l'ARN messager et les convertit en sous-unités d'ADN : l'enzyme construit le «collier» de l'ADN complémentaire en transcrivant les «perles» ARN en «perles» ADN.

Les ADN complémentaires ainsi obtenus sont généralement des répliques de segments d'ARN messa-