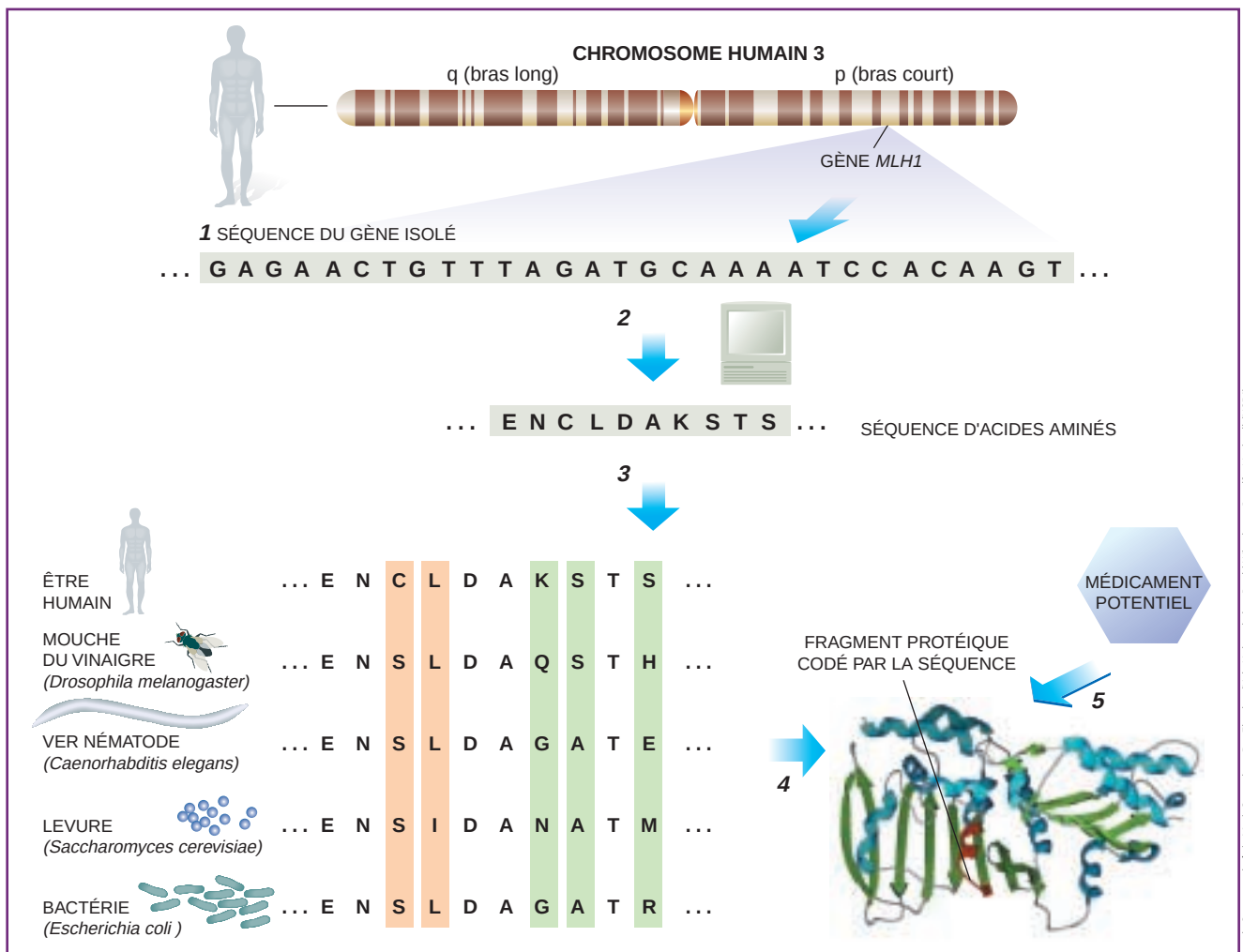




Laurie Grace, assisté de Mark Gerstein et Pat Fleming (Yale University) et David Wheeler et Jennifer Viskochil (NCBI)

3. LA COMPARAISON DE GÈNES HUMAINS avec ceux d'animaux modèles offre parfois des pistes pour la recherche de nouveaux médicaments. Par exemple, quand on connaît la séquence du gène humain *MLH1* (1), associé au cancer du côlon, un programme informatique peut traduire cette séquence d'ADN en une séquence d'acides aminés, qui compose la protéine codée (2). On cherche ensuite des séquences similaires dans des banques de données

de protéines synthétisées par des organismes modèles (3). Les bandes vertes indiquent des zones de grande variabilité (les acides aminés sont peu conservés d'un organisme à un autre), et les bandes orange, des zones de faible variabilité. On modélise la protéine humaine à partir des structures des protéines similaires des organismes modèles (4) et, enfin, on crée une molécule qui se fixe à la protéine modélisée (5).

produits, de la formulation, de la toxicologie et des essais cliniques.

Naguère, ces services étaient excessivement cloisonnés, et les données d'un service restaient inaccessibles aux autres équipes. La bio-informatique donnera la même information à tous, et chacun fera un usage individuel des données.

En outre, les services internes communs de bio-informatique réduisent les investissements informatiques. Par exemple, les moyens logiciels individuels, utilisés par les différents chercheurs d'une même entreprise pour l'accès aux bases de données et le traitement de ces données, sont remplacés par une plate-forme logicielle unique : les économies se comptent en millions de francs.

Afin d'intégrer pleinement la bio-informatique à leurs sociétés, des

groupes pharmaceutiques concluent parfois des alliances stratégiques ou même acquièrent des sociétés de génie génétique plus petites ; ainsi munis de partenaires et de réseaux commerciaux, ils augmentent leurs compétences bio-informatiques et s'adaptent aux nouvelles techniques sans réorganiser leurs propres structures.

Toutefois, le regroupement des banques de données est notoirement périlleux. On cherche à éviter les mauvaises communications à l'aide de systèmes de nomenclature des données comprise aussi bien des diverses banques de données que des systèmes d'appellation. Imaginons qu'une banque A communique avec une banque B, celle-ci avec une banque C, C avec D, et ainsi de suite. Si A change ses données, les références de la banque D

risquent d'être périmées, surtout quand le renouvellement des données est permanent. Ce problème s'aggrave, car les connaissances en biologie et les analyses par ordinateur sont de plus en plus complexes.

Des améliorations systématiques seront d'une grande aide, mais le vrai progrès semble reposer sur l'ingéniosité des utilisateurs : le succès de la bio-informatique n'est pas une affaire de matériel, ni de logiciel, mais bien d'intelligence.

Ken HOWARD est journaliste scientifique.

D. B. SEARLS, *Using Bioinformatics in Gene and Drug Discovery*, in *Drug Discovery Today*, vol. 5, n° 4, pp. 135-143, avril 2000.

Au-delà du génome

CAROL EZZELL

Le séquençage de l'ADN d'un être humain n'est qu'une première étape dans l'étude du fonctionnement d'une cellule : on doit aujourd'hui explorer les protéines codées par ces gènes.

Au moment où les gènes sont d'actualité, ils sont supplantés par les protéines. Un savoir est acquis : les séquences des milliers de gènes humains. Les biologistes doivent en conquérir un autre : le fonctionnement de ces gènes ; on cherche où et quand ils s'expriment, et quelles sont les propriétés des protéines qu'ils codent. Ce nouveau champ du savoir, nommé protéomique concentre la majorité des efforts financiers et humains des organismes publics et des sociétés privées.

Pour comprendre la protéomique, examinons comment un gène code une protéine. Dans le noyau d'une cellule, un segment d'ADN qui correspond au gène est d'abord transcrit en une molécule d'ARN, également formée de nucléotides. Cet ARN est composé de régions codantes, nommées exons, et de régions non codantes, les introns. L'ARN est débarrassé des introns par des enzymes d'«épissage» : ainsi se forme l'ARN messenger. Si l'ADN est la série de plans qu'une cellule utilise pour synthétiser les protéines, l'ARN messenger est la copie d'une partie du plan que l'entrepreneur emmène sur le chantier chaque jour. Le chantier est à l'extérieur du noyau, dans le cytoplasme, où la séquence en nucléotides de l'ARN messenger est traduite en une séquence en acides aminés, qui composent la protéine.

Bien que chaque cellule de l'organisme contienne l'ensemble de l'ADN utile à la fabrication et au fonctionnement d'un être humain, beaucoup de ces gènes ne s'expriment jamais, c'est-à-dire ne sont jamais transcrits en ARN messagers, lorsque le développement embryonnaire est achevé. En outre, les gènes ne s'expriment qu'en fonction de leur situation dans les tissus et de leur rôle dans l'organisme. Par

exemple, dans une cellule bêta du pancréas, le gène de l'insuline s'exprime intensément (la cellule est pleine d'ARN messagers qui codent l'insuline), alors qu'il reste muet dans un neurone.

Jusqu'au début des années 1980, les biologistes pensaient qu'à un gène correspondait un seul ARN messenger et une seule protéine. Il n'en est rien : un gène peut être transcrit par segments, qui sont ensuite coupés et regroupés en une variété d'ARN messagers ; en outre, les protéines ainsi synthétisées sont parfois modifiées : ajout de groupes méthyle ($-CH_3$), liaison à des sucres, etc. Ainsi la connaissance du génome humain n'est pas celle du fonctionnement cellulaire : les biologistes doivent étudier le «transcriptome», c'est-à-dire l'ensemble des ARN messagers produits par les cellules, et le «protéome», c'est-à-dire l'ensemble des protéines synthétisées à partir des ARN messagers.

Des puces biologiques

L'étude du transcriptome bénéficie d'un progrès technique récent : la mise au point de biopuces. Sur des plaques de verre d'environ un centimètre carré, on ancre les molécules d'ADN complémentaires des ARN messagers que l'on recherche. Ces molécules d'ADN complémentaires sont obtenues à partir des ARN messagers. Ces derniers s'associent spécifiquement à leurs ADN complémentaires selon le même principe que les deux brins de la double hélice d'ADN. L'ARN messenger d'un gène ne s'associera qu'au seul ADN qui lui est complémentaire. Celui-ci est une sorte de copie du gène sans intron.

Pour utiliser ces biopuces, les biologistes extraient les ARN messagers d'un échantillon, ils les marquent par une molécule fluorescente, puis les ver-

sent sur la puce. La lame de verre de la biopuce est divisée en microsurfaces (jusqu'à plusieurs dizaines de milliers). Sur chacune d'elles, on n'ancre qu'un seul ADN complémentaire bien identifié. Chaque ARN fluorescent de l'échantillon s'associe à l'ADN qui lui est complémentaire. À l'aide d'un microscope, on repère les microsurfaces fluorescentes : selon les parties des biopuces auxquelles les ARN messagers fluorescents se lient, les biologistes déduisent, de la séquence des ADN complémentaires, les séquences des ARN messagers de l'échantillon. De telles puces détectent simultanément plus de 60 000 ARN messagers ; par exemple, des biopuces détectent spécifiquement les ARN messagers humains associés au cancer.

Depuis 1998, des centres de recherche publics et privés utilisent de tels systèmes pour identifier les ARN messagers produits par divers types de cellules cancéreuses. Grâce aux biopuces, les généticiens savent, par exemple, que 5 692 gènes sont actifs dans les cellules de cancer du sein et que, parmi eux, 277 sont inactifs dans les cellules des autres tissus. Des composés qui bloqueraient les protéines codées par ces 277 gènes pourraient être des médicaments anticancéreux qui auraient moins d'effets secondaires que les chimiothérapies actuelles. Aussi des millions de francs ont-ils été récemment investis dans l'analyse des protéines impliquées dans le cancer.

L'étude des ARN messagers n'est qu'un moyen : on veut surtout mieux connaître les protéines qui, replacées dans le fonctionnement d'une

1. APRÈS AVOIR ANALYSÉ ses chromosomes, l'espèce humaine doit aujourd'hui déterminer où et quand les gènes sont actifs, et identifier les protéines que ces gènes codent.