

cellule, constituent les cibles des médicaments. Seules, mais nombreuses! On suppose que les milliers de gènes humains produisent plus d'un million de protéines différentes. Certains experts estiment qu'au cours de la prochaine décennie l'industrie pharmaceutique disposera de plus de 10 000 protéines humaines contre lesquelles on pourra chercher des principes actifs : c'est 25 fois plus de cibles que l'industrie pharmaceutique n'en a évaluées depuis qu'elle existe. Pour satisfaire leurs actionnaires, les

grandes sociétés pharmaceutiques doivent identifier trois à cinq nouveaux médicaments potentiels chaque année, soit deux à cinq fois plus qu'aujourd'hui. Aussi, elles s'associent avec des petites sociétés spécialisées qui leur fournissent des cibles potentielles de médicaments déjà testées.

Des machines à protéines

Des sociétés privées qui ont participé au séquençage du génome s'intéressent aujourd'hui à l'étude de

l'expression protéique. La Société *Celera genomics*, en association avec l'Institut suisse de bio-informatique, a créé une entreprise qui veut identifier tout le protéome humain. Simultanément, les techniques d'analyse se perfectionnent : en 1999, une équipe a ainsi publié les plans d'un scanner moléculaire qui automatiserait le processus fastidieux de séparation et d'identification des milliers de protéines d'une cellule.

Aujourd'hui, pour étudier les protéines, on utilise souvent la technique

La protéomique fonctionnelle

L'étape suivant le séquençage du génome et l'identification des gènes consiste à étudier les protéines que ces derniers codent. Cependant, les protéines n'interviennent pas de façon isolée dans les processus biologiques cellulaires : elles constituent avec d'autres protéines une ou plusieurs cascades d'interactions, qui sont l'expression de fonction biologiques. La caractérisation de ces interactions entre protéines est le champ d'application de la protéomique fonctionnelle qui met en correspondance les données du génome et les observations biologiques.

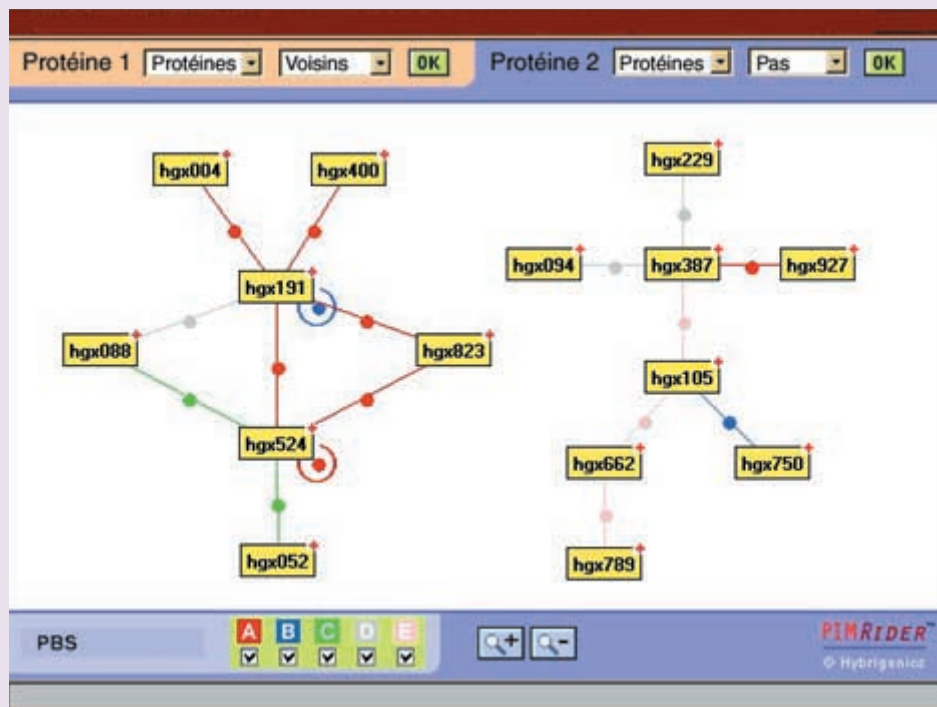
Les techniques de la protéomique fonctionnelle s'appliquent à trois niveaux de complexité croissante : l'identification des protéines qui agissent directement avec quelques protéines choisies ; l'identification des cas-

cades d'interactions complètes ; enfin, l'identification des connexions entre les différentes cascades d'interactions.

En France, la Société *Hybrigenics* met au point des méthodes d'étude de la protéomique fonctionnelle. Par exemple, elle a amélioré la technique de double hybride de la levure décrite dans l'article. Ainsi, au lieu de mettre en présence dans une levure deux protéines entières qui déclencheront une réaction observable lorsqu'elles interagissent, on segmente les protéines «proies» en de nombreux fragments, puis on cherche les éventuelles interactions de la protéine «appât» et ces nombreux «fragments proies». Les résultats obtenus sont plus fiables et plus précis. Grâce à cette méthode, on détermine quel fragment d'une «protéine proie» agit avec une protéine «appât» donnée.

En contrepartie, le nombre d'expérimentations et de

la complexité du traitement des données issues de ces expérimentations, augmente, en particulier lors de l'identification des réseaux d'interactions. Le nombre d'interactions étudiées atteint souvent plusieurs centaines de millions à quelques milliards : l'exploitation des données requiert alors des systèmes informatiques adaptés. Les systèmes bio-informatiques mis en œuvre effectuent un «criblage à haut débit», c'est-à-dire le tri de plusieurs centaines de millions d'interactions testées en quelques semaines. On obtient ainsi une cartographie des réseaux d'interactions protéiques (voir la figure) que les utilisateurs visualisent directement à partir de leur ordinateur via le réseau Internet.



Sur son ordinateur, un utilisateur obtient une cartographie d'interactions protéiques. Les rectangles jaunes représentent les protéines et les lignes de couleur, les interactions.

Donny STROSBURG, PDG de la Société *Hybrigenics*

d'électrophorèse bidimensionnelle, qui sépare les protéines selon leur charge électrique et leur taille : on dépose une solution qui contient de nombreuses protéines sur une étroite bande de polymère dont l'acidité varie d'un bord à l'autre. Quand on applique au gel une différence de potentiel électrique, chaque protéine du mélange se déplace à une vitesse qui dépend de sa charge et de l'acidité du milieu. On place ensuite la bande au bord d'un gel plat et on applique une nouvelle différence de potentiel électrique, perpendiculaire au premier ; les protéines migrent sur le gel à une vitesse qui dépend de leur masse moléculaire. Finalement, chacune des taches obtenues (après une réaction chimique de «révélation») contient une protéine différente.

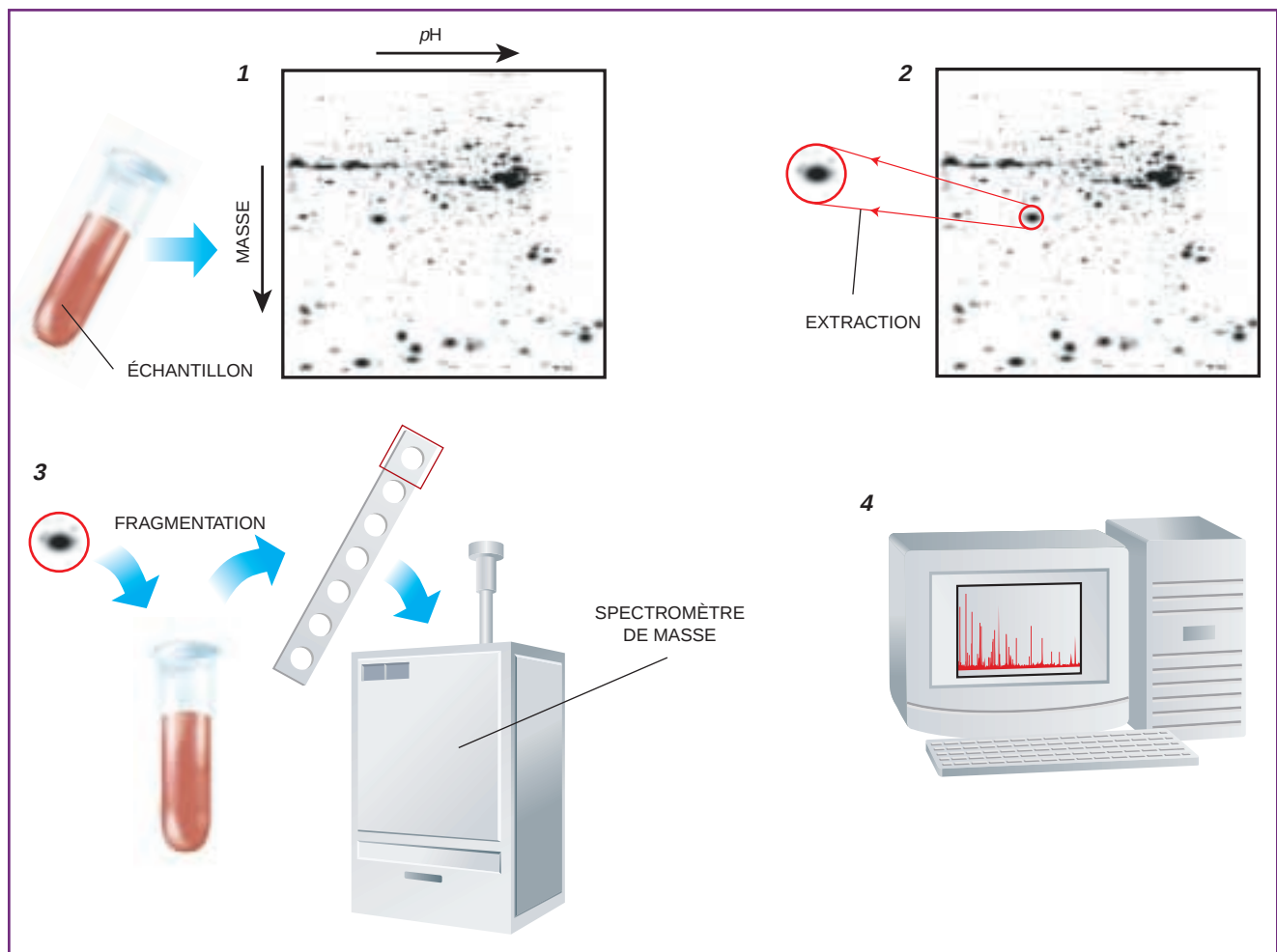
Dans les laboratoires universitaires, les taches sont généralement extraites du gel grâce à un perforateur, et les protéines de chaque tache sont analysées par spectrométrie de masse : à

l'aide d'enzymes, on découpe les protéines en peptides (des courtes chaînes d'acides aminés) dont on identifie la masse et la charge électrique ; à partir de ces informations, des programmes d'ordinateur retrouvent dans les banques de données de séquences le gène dont la séquence peut coder pour de tels peptides. Un gène est ainsi relié à une tache : on déduit enfin la séquence de la protéine de la tache analysée (voir la figure 2). Aujourd'hui, les robots remplacent progressivement les biologistes dans cette tâche : des machines entièrement automatiques extraient les taches des gels, utilisent les enzymes pour couper les protéines en morceaux, transmettent les fragments à un spectromètre de masse laser et transfèrent les informations à un ordinateur en vue de leur analyse.

Toutefois, l'automatisation ne pallie pas les inconvénients de l'électrophorèse bidimensionnelle. Notamment, leur préparation est délicate, et les

protéines fortement chargées, ou de masse faible, sont mal séparées. Les protéines qui ont des régions hydrophobes, telles les protéines membranaires, sont également difficiles à étudier par cette technique ; or, les récepteurs transmembranaires sont des cibles thérapeutiques importantes.

Dans les cas complexes, les biologistes utilisent alors la méthode de «culpabilité par association» : on a des indications sur la fonction d'une protéine en connaissant ses interactions avec une autre protéine déjà connue. Ainsi, en février 2000, des généticiens ont identifié 957 interactions parmi les 1 004 protéines de la levure de boulangerie *Saccharomyces cerevisiae*. Ils ont employé un système de double hybride : une protéine qui active un gène est coupée en deux ; une moitié est fixée à une protéine «appât» dont on connaît la fonction, l'autre est fixée à la protéine «proie» que l'on étudie. Quand les protéines



2. L'IDENTIFICATION DE PROTÉINES INCONNUES s'effectue sur un gel bidimensionnel. On sépare les protéines selon leur charge (pH) dans une direction et leur masse dans la direction perpendiculaire (1). Chaque tache obtenue contient une protéine (2) que des enzymes

découpent en plusieurs fragments (3). L'ordinateur d'un spectromètre de masse établit un diagramme des fragments de protéines en fonction de leur masse (4). Ce diagramme est caractéristique de la protéine initiale.