

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.

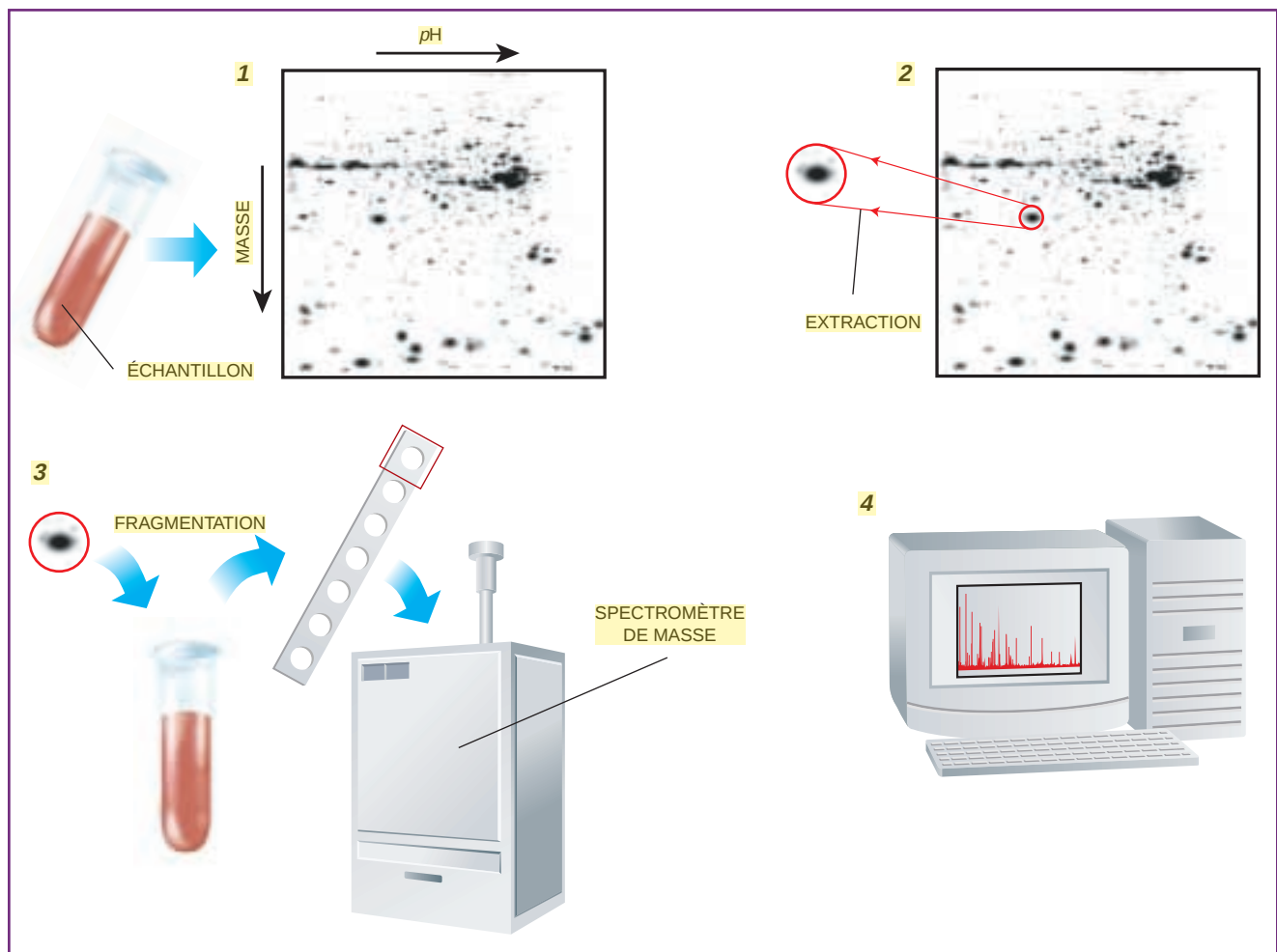
Dans les laboratoires universitaires, les taches sont généralement extraites du gel grâce à un perforateur, et les protéines de chaque tache sont analysées par spectrométrie de masse : à

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

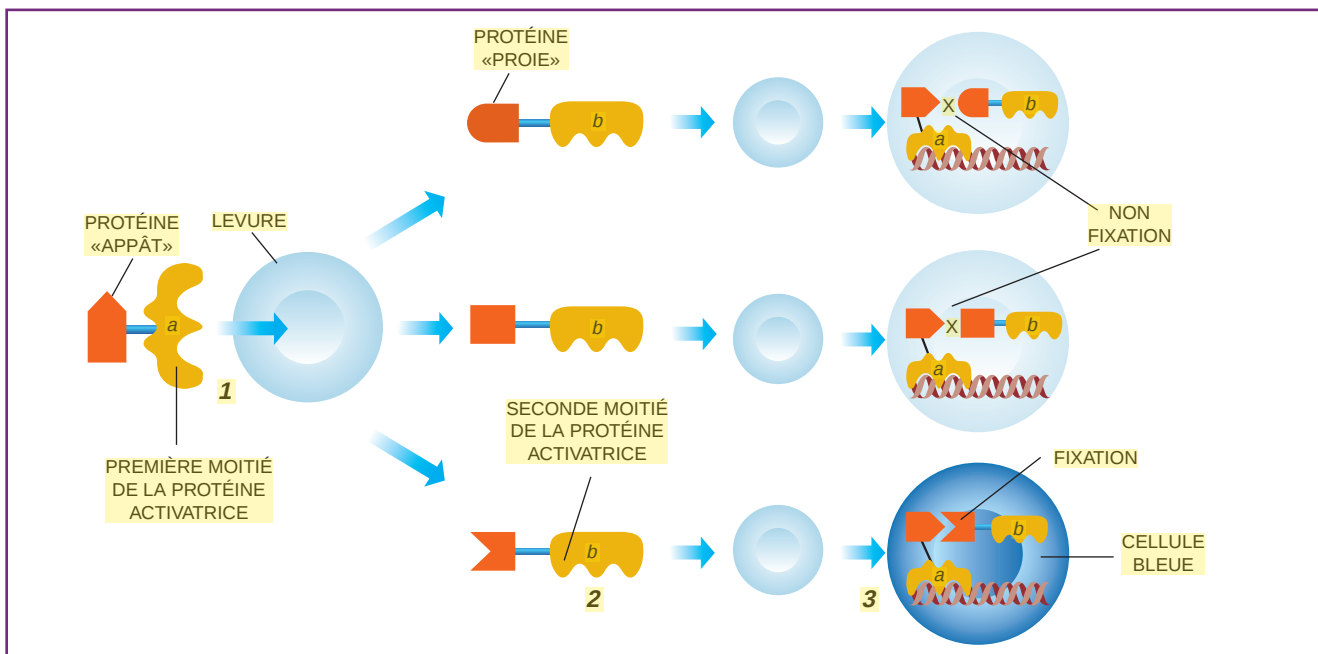
protéines fortement chargées, ou de masse faible, sont mal séparées. Les protéines qui ont des régions hydrophobes, telles les protéines membranaires, sont également difficiles à étudier par cette technique ; or, les récepteurs transmembranaires sont des cibles thérapeutiques importantes.

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.



Laure Grace, assistée de Curagen

3. DANS LA TECHNIQUE DE «DOUBLE HYBRIDE» de la levure, on utilise deux moitiés (a et b) d'une protéine qui active un gène, tel celui qui active la synthèse d'un pigment bleu. Grâce à cette technique, on détermine quelle protéine d'un ensemble de protéines «proies» inconnues se lie à une protéine «appât» connue : l'ADN codant la protéine «appât»,

lié à l'ADN codant la moitié (a) de la protéine activatrice, est introduit dans le noyau de la levure (1). Puis, l'ADN codant l'autre moitié (b) de la protéine activatrice liée à l'ADN codant une des protéines «proies» est également introduit dans le noyau de la levure (2). Lorsque la couleur de la cellule change, la protéine «proie» s'est liée à la protéine «appât» (3).

«proie» et «appât» interagissent, les deux moitiés de la protéine activatrice sont réunies : le gène s'exprime, et la présence de la protéine codée par le gène révèle l'association (voir la figure 3).

Depuis le séquençage du génome de la levure en 1996, les fonctions d'un tiers de ses 6 000 gènes restent mystérieuses. Cependant, grâce au système de double hybride, une partie des protéines codées par ces gènes ont été classées en catégories fonctionnelles, telles la production d'énergie, la réparation de l'ADN, le vieillissement... Le plan des interactions protéiques de la drosophile est également en cours de déchiffrement.

La levure était un prototype ; avec la drosophile, on obtient des informations sur un organisme multicellulaire, et on utilise la protéomique pour trouver de nouveaux médicaments.

Certaines sociétés appliquent une variante du système de double hybride de la levure à l'étude des voies métaboliques spécifiques de maladies. Par exemple, l'une d'elles teste les interactions biochimiques des protéines codées par le gène *MMAC1*, qui, quand il est muté, conduit parfois à un cancer du cerveau et de la prostate.

Enfin, des biopuces d'un type nouveau apparaissent. Certaines, sous la forme de bandelettes, adsorbent, c'est-

à-dire retiennent à leur surface, les protéines selon des propriétés telles que leur solubilité dans l'eau, leur affinité pour des ions métalliques, etc. Ensuite placées dans un lecteur de puces, qui inclut un spectromètre de masse, les protéines des bandes sont identifiées.

En mars 2000, une telle biopuce à protéines a été utilisée pour la découverte de 12 marqueurs potentiels d'une maladie bénigne de la prostate et de six marqueurs du cancer de la prostate. On espère que les tests fondés sur les marqueurs ainsi identifiés seront plus fiables et précis que le test utilisé aujourd'hui.

De la structure à la fonction

L'identification des protéines humaines sera un premier travail, qui devra se poursuivre par l'étude de la fonction des protéines, l'élucidation de leur forme et de leur structure. Un groupe de biologistes a récemment proposé un programme d'exploration de la structure des protéines normales et anormales par des techniques de radio-cristallographie quasi automatisées.

Classiquement, pour déterminer la structure d'une protéine, on doit utiliser un cristal pur de cette protéine, que l'on irradie avec des rayons X. Ces derniers rebondissent sur les atomes

de la molécule et forment une figure de diffraction dont on déduit la forme tridimensionnelle de la molécule entière. Le programme proposé veut perfectionner et automatiser les techniques actuelles.

Connaissant la structure exacte de chacune des protéines du protéome humain, les concepteurs de médicaments élaboreront des molécules qui s'adapteront aux sites protéiques et les activeront ou les empêcheront d'agir. À ce jour, de telles tentatives de conception rationnelle ont eu peu de succès. Cependant, l'espoir subsiste pour les 99 pour cent de protéines humaines dont on ne connaît pas encore la structure.

Carol EZZELL est journaliste à *Scientific American*.

The Chipping Forecast, supplément spécial de *Nature Genetics*, vol. 21, n° 1, janvier 1999.

A. ABBOTT, *A Post-Genomic Challenge : Learning to Read Patterns of Protein Synthesis*, in *Nature*, vol. 402, pp. 715-720, 16 décembre 1999.

R. JAMES, *Differentiating Genomics Companies*, in *Nature Biotechnology*, vol. 18, pp. 153-155, février 2000.

R.F. SERVICE, *Structural Genomics Offers High-Speed Look at Proteins*, in *Science*, vol. 287, pp. 1954-1956, 17 mars 2000.