

Protéome et analyse protéomique : de nouveaux concepts pour de nouveaux champs d'applications biomédicales

Le terme de protéome, proposé en 1995 pour désigner l'ensemble des produits fonctionnels exprimés par un génome, atteste de la nécessité d'individualiser un domaine nouveau promettant une meilleure compréhension de la complexité du fonctionnement cellulaire à partir de l'expression protéique globale [1, 2]. Ce concept a pu se développer grâce à l'évolution de technologies nées dans les années 1980 telles que l'électrophorèse bidimensionnelle qui permet de séparer simultanément des centaines, voire des milliers, de polypeptides [3] et les méthodes de microcaractérisation des protéines. Cependant, il fallu attendre la fin des années 1990 pour que ces méthodes soient améliorées et associées de façon systématique, rendant possible l'analyse de l'expression protéique dans sa globalité pour une cellule, un tissu ou un organisme. La chimie des protéines a ainsi connu une évolution – si ce n'est une révolution – passant de l'analyse individuelle des protéines à la caractérisation de masse à partir d'un échantillon complexe. Il faut souligner que les avancées dans le décryptage du génome de différents organismes – en particulier le génome humain – au cours de ces mêmes années, ont grandement facilité l'émergence du protéome. Un autre terme – *proteomics* en anglais – dont l'équivalent français peut être analyse protéomique, est né *a posteriori* pour désigner un concept dynamique et ambitieux. Le protéome représente une somme statique de données – ce qui est également le cas du génome – alors que l'analyse protéomique est un

domaine d'étude à part entière, comme l'est l'analyse génomique. NL Anderson et NG Anderson définissent l'analyse protéomique comme « l'utilisation de la quantification au niveau de la protéine comme mesure objective de l'expression génique caractérisant un processus biologique donné (processus physiologique ou pathologique, effets des médicaments, de l'environnement...) et comme moyen de décodage des mécanismes contrôlant cette expression » [4]. L'analyse protéomique dans l'acceptation de cette définition, met l'accent sur une description dynamique de la régulation des gènes. Nous nous proposons, après avoir précisé le pourquoi et le comment du protéome et de l'analyse protéomique, de montrer à l'aide de quelques exemples les potentialités de ces nouveaux concepts dans leurs applications biomédicales.

Pourquoi étudier le protéome

L'importance prise par l'analyse génomique engendre une somme croissante d'informations sur les séquences codantes à laquelle s'ajoutent les données quantitatives portant sur l'expression des ARN messagers (ARNm). Prenant en compte cette somme d'informations, on peut se demander pourquoi des projets protéome technique complexes doivent s'y surajouter, et quelles informations spécifiques en attendre qui ne soient obtenues directement par l'étude génomique. Au moins trois raisons peuvent être avancées pour faire de l'analyse protéomique une composante essentielle à la com-

préhension des systèmes biologiques :

- les niveaux d'expression protéique ne sont pas un simple reflet des niveaux d'expression des ARNm. Ainsi, des études menées sur des organismes eucaryotes montrent que, même pour une population de gènes considérés comme relativement homogènes en termes d'expression des ARNm et de demi-vie des protéines codées, des variations significatives du niveau d'expression de celles-ci peuvent être observées [4, 5] ;
- les protéines peuvent subir une série de modifications qui ne sont pas nécessairement identifiables à partir de la seule séquence de leurs gènes. Beaucoup de ces molécules ne parviennent à leur forme biologiquement active qu'à la suite d'étapes de maturation co- et post-traductionnelles telles que glycosylation, phosphorylation, isoprénylation, désamination... Globalement, l'état de modification de l'ensemble des polypeptides qui constituent un système biologique donné représente une information importante sur l'état de ce système. En outre, l'ajout ou l'élimination de groupements chimiques sur la chaîne polypeptidique constitue parfois un signal de ciblage que l'étude de l'expression protéomique dans différents compartiments cellulaires permettra de déchiffrer ;
- l'analyse protéomique est dynamique. Le même génome peut ainsi conduire à différents protéomes en fonction des étapes du cycle cellulaire ou de la différenciation, de la réponse à des signaux biologiques ou physiques, de l'état physiopathologique... Le protéome reflète les

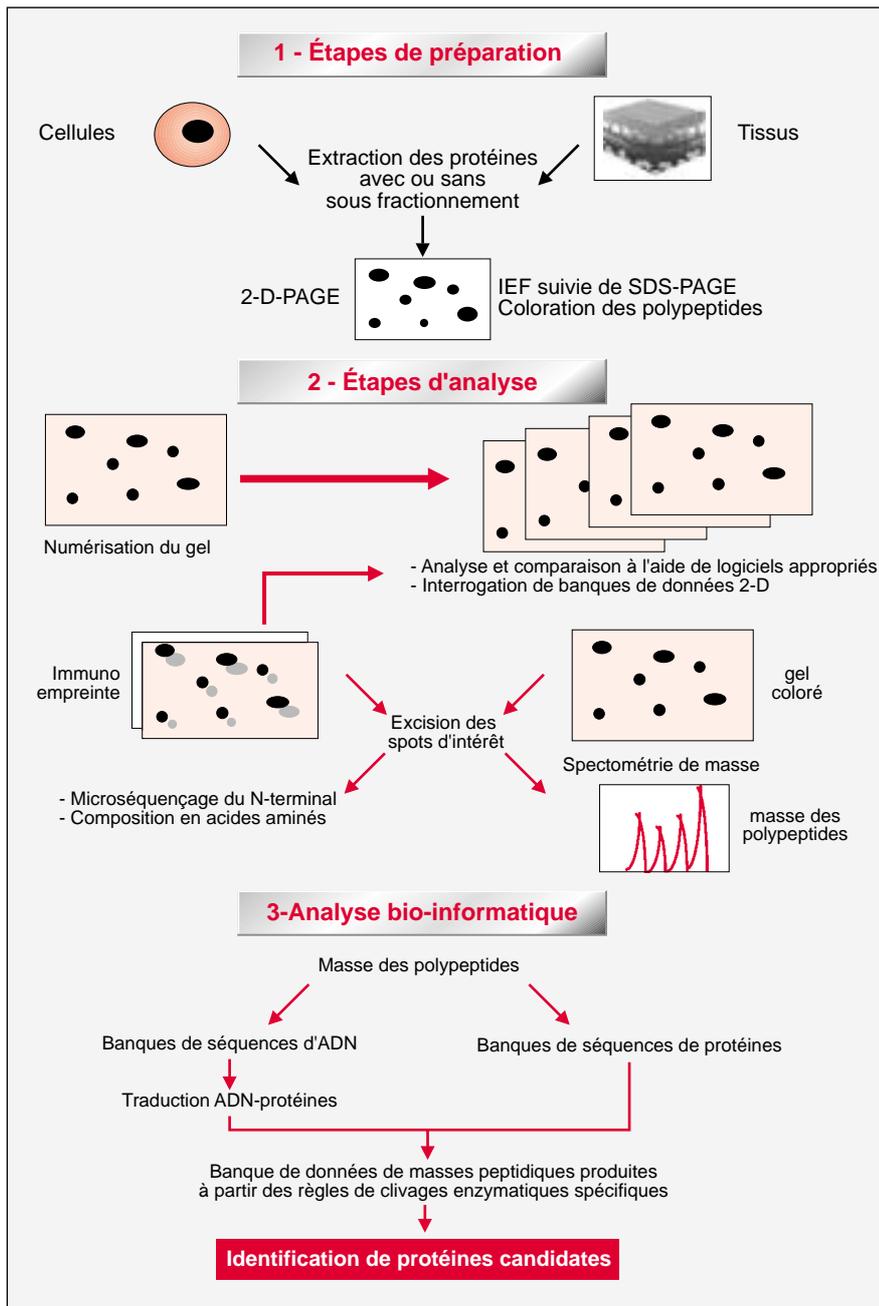


Figure 1. **Stratégie de l'analyse protéomique.** Schématiquement, la stratégie de l'analyse protéomique peut se décomposer en trois grandes phases. **1.** Au cours des étapes de préparation, les protéines obtenues à partir d'extraits tissulaires ou cellulaires sont séparées par électrophorèse bidimensionnelle. Le gel est généralement coloré par une méthode au nitrate d'argent. **2.** Les étapes analytiques peuvent être variées. Une première approche consiste, après numérisation du gel, à analyser d'un point de vue qualitatif et quantitatif à l'aide de logiciels d'analyse d'images spécialisés les spots résolus. Cela conduit à l'obtention des coordonnées (point isoélectrique et masse) des polypeptides détectés, ainsi qu'à la détermination de leur degré d'expression utilisant des paramètres densitométriques de volume, surface... L'interrogation de banques de données telles que SWISS 2D-PAGE peut aider à l'analyse. La caractérisation de certains polypeptides au moyen de méthodes immunochimiques (nécessitant au préalable un transfert des protéines du gel vers une matrice semi-rigide ainsi que la détection par un anticorps spécifique) ne peut s'appliquer que dans le cas d'un nombre limité de protéines à analyser. Les spots d'intérêt peuvent également être excisés du gel ou de la matrice de transfert et soumis à une dégradation d'Edman visant à obtenir une information sur la séquence amino-terminale. Une détermination de la composition globale en acides aminés est également envisageable à cette étape. La spectrométrie de masse des peptides obtenus après digestion par une endoprotéase tend à s'imposer. **3.** Une analyse bio-informatique permet, en étape finale, de comparer la masse des peptides obtenus en spectrométrie de masse,

aux cartes de masses théoriques des protéines répertoriées dans les banques de données (protéomiques ou nucléotidiques). Un certain nombre de protéines candidates sont alors proposées. La corrélation avec d'autres données (point isoélectrique, masse, séquence amino-terminale) permet de trancher entre les différentes possibilités.

répercussions de ces événements cellulaires au niveau tant traductionnel que post-traductionnel. De ce point de vue, seule une analyse protéique directe peut donner une image globale des systèmes biomoléculaires dans leur complexité.

Les outils de l'analyse protéomique

La stratégie de l'analyse protéomique est résumée dans la figure 1. La méthodologie mise en œuvre au cours de cette analyse peut schématiquement être séparée en trois

grandes phases : des étapes de préparation conduisant à la production de gels d'électrophorèse bidimensionnelle, puis d'analyse dans lesquelles on peut inclure la caractérisation de certains polypeptides au moyen de méthodes classiques de la chimie de

protéines (immunodétection après transfert, microséquençage de spots excisés du gel...) ou spectrométries (spectrométrie de masse). Enfin, une analyse bio-informatique à l'aide de logiciels spécialisés associée à la consultation des banques de données de séquences protéiques ou nucléotidiques permettra d'identifier les protéines candidates.

Lors des étapes de préparation, les protéines solubilisées à partir de cellules isolées ou de tissus sont soumises à une séparation électrophorétique selon deux dimensions. La préparation/solubilisation de l'échantillon est jusqu'à aujourd'hui un facteur crucial et parfois limitant de cette étape [6]. L'électrophorèse bidimensionnelle utilisée lors de la séparation est une méthode déjà ancienne puisque sa description date de 1975 [7], remise au goût du jour grâce à la commercialisation de gels d'immobilines coulés sur un support rigide (*gelbond*). Les immobilines par leur co-polymérisation avec l'acrylamide permettent d'établir un gradient de pH immobilisé pour la première dimension rendant l'isoélectrofocalisation des protéines extrêmement résolutive et reproductible [3].

Les gels obtenus sont ensuite colorés, généralement par le nitrate d'argent, puis numérisés, ce qui permet l'analyse des variations d'expression des polypeptides à l'aide de logiciels spécialisés. Dans la majorité des cas, une identification directe des polypeptides observés et de leurs variations n'est pas envisageable. La détermination, même précise, de leurs coordonnées (en point isoélectrique et masse) sur le gel ne peut être assimilée à une méthode d'identification. L'identification des protéines séparées par électrophorèse 2D est donc un problème central pour le projet protéome [8]. Dans le cas d'études ponctuelles portant sur un petit nombre de protéines connues, des approches telles que l'immunodétection après transfert, ou la co-migration avec un excès d'une protéine purifiée à partir du tissu peuvent s'appliquer. Mais dans le cadre des programmes plus ambitieux, les différentes stratégies d'identification reposent sur le couplage de tech-

niques analytiques et de la recherche dans des banques de données. Les principales techniques analytiques utilisées sont la détermination de la composition globale en acides aminés, et surtout :

– l'établissement de cartes de masses peptidiques par spectrométrie de masse MALDI-TOF (*matrix-assisted laser desorption ionization time of flight*). La masse des peptides obtenus après digestion de la protéine par une endoprotéase est comparée aux cartes de masses théoriques des protéines répertoriées dans les banques de données. Cette stratégie est actuellement une des plus largement explorées ;

– l'établissement de courtes séquences peptidiques terminales ou internes (*tags*) par séquençage suivant une méthode conventionnelle (dégradation d'Edman...) ou spectrométrie de masse (MS-MS). La présence d'un nombre croissant de séquences génomiques dans les banques de données rend possible une analyse systématique des protéines séparées lors de l'électrophorèse bidimensionnelle par détermination des séquences amino-ou carboxy-terminales [9]. Ainsi, la séquence des quatre résidus amino-ou carboxy-terminaux conduit à l'identification d'un unique polypeptide pour environ 80 % des protéines humaines retrouvées dans les banques. Les enchaînements plus fréquents correspondent à des protéines codées par des familles de gènes, principalement les antigènes d'histocompatibilité et les chaînes d'immunoglobulines. Dans le cas où ces séquences terminales sont communes à un groupe de protéines, le point isoélectrique et la masse apparente déterminés lors de l'électrophorèse (éventuellement la masse réelle déterminée par spectrométrie) permettent le plus souvent de trancher entre les différentes possibilités. D'une façon générale, le recoupement de plusieurs informations (séquence de 3 à 6 acides aminés, localisation d'une fenêtre comprenant la tache en électrophorèse 2D, espèce animale d'où provient l'échantillon) offre un maximum de sécurité pour l'identification d'un polypeptide.

Quelle que soit la stratégie choisie, l'analyse bio-informatique constitue

une étape capitale. Pour se donner les moyens de son ambition, l'analyse protéomique doit se doter d'un nombre important de résultats expérimentaux regroupés dans des banques de données ou serveurs, actuellement en cours de constitution et accessibles aux expérimentateurs *via* le multimédia. Les recherches de séquences connues utilisent généralement la banque SWISSPROT et la traduction de la banque de séquences d'ADN de l'EMBL en séquences protéiques (TrEMBL). Ces banques sont accessibles par <http://expasy.proteome.org.au>

Utilisation de l'analyse protéomique en biomédecine

Les avancées de l'analyse protéomique ont porté dans un premier temps sur des organismes simples, principalement procaryotes, comportant un nombre restreint de protéines et de modifications post-translationnelles. Mais aujourd'hui, le pari majeur de cette analyse est le développement de nouveaux champs d'application dans le domaine biomédical. Le nombre de banques de gels 2D consacrées aux cellules ou tissus humains (*Tableau 1*) témoigne de cet intérêt. Plusieurs de celles-ci ne se limitent pas à un inventaire des protéines identifiées, mais fournissent des données comparatives sur les variations induites par la différenciation cellulaire, l'action de cytokines, de molécules utilisées en chimiothérapie... Parmi les domaines dans lesquels des analyses protéomiques systématiques sont en cours, on peut citer : l'établissement de cartes des fluides biologiques à visée diagnostique, la recherche d'une vision globale des modifications induites par le vieillissement, par la tumorigénèse, ou par l'action d'agents chimiques ou biologiques.

Les données concernant les fluides biologiques sont principalement accessibles au travers de deux bases de données : NIMH-NCI *protein-disease database* (PDD) et ExPASy. La base PDD vise à corréler les variations d'expression des protéines de l'inflammation aiguë dans le plasma et le liquide céphalorachidien avec

Tableau I		
BANQUES CONTENANT DES DONNÉES SUR LE PROTÉOME DE CELLULES OU TISSUS D'ORIGINE HUMAINE		
Accès http	Localisation	Contenu de la banque
http://expasy.hcuge.ch/ch2d	Université de Genève (Suisse)	Plasma, LCR, foie, rein, globules rouges, plaquettes, hépatome, lignées hématopoïétiques
http://www.univ-paris13.fr/biochim.htm	Laboratoire de biochimie Bobigny (France)	Lignées hématopoïétiques et lymphoïdes : TF1, K562, KG1a, HL60, PRI, DG75
http://www.harefield.nthames.nhs.uk/nhli/protein	Heart science centre Harefield (Grande-Bretagne)	Cœur, cellules endothéliales
http://userpage.chemie.fu-berlin.de/~pleiss/dhzb.html	German Heart Institute Berlin (Allemagne)	Cœur
http://www.mdc-berlin.de/~emu/heart	MDC, Berlin (Allemagne)	Cœur
http://www-pdd.ncifcrf.gov	NIMH-NCI, PDD, Washington (États-Unis)	Plasma, LCR, urine
http://www.an1.gov/CMB/PMG	Argonne University, Chicago (États-Unis)	Lignée cancer du sein
http://iupucbio1.iupui.edu/frankw/molan.htm	Molecular Anatomy Laboratory Columbus (États-Unis)	Plasma, foie
http://biobase.dk/cgi-bin/celis	Danish Center for Human Genome, Aarhus (Danemark)	Kératinocytes, carcinomes prostatiques, urine, fibroblastes
http://www.tmig.or.jp/2D/2D-HOME.html	Metropolitan Institute of gerontology, Tokyo (Japon)	Fibroblastes (syndrome de Werner)
http://www.ed.ac.uk/~nh/2D/PAGE.html	University of Edinburgh (Grande-Bretagne)	Cellules souches hématopoïétiques
http://www.ludwig.edu.au/www/jpsl/jpslhome.html	Ludwig Institute for Cancer Research Melbourne (Australie)	Placenta, lignées cellulaires colorectales
http://www-lecb.ncifcrf.gov/ips-databases.html	LECB, NCI-FCRDC, Frederick (États-Unis)	Méta-base regroupant des données issues de diverses banques

différentes affections neurologiques. De nombreuses autres protéines des fluides biologiques ont été répertoriées par l'équipe de Denis F. Hochstrasser à l'hôpital universitaire de Genève et sont accessibles *via* le serveur ExPASy/SWISS-2DPAGE [10]. Des travaux portant sur la

caractérisation des *spots* d'immunoglobulines, en particulier dans les gammopathies monoclonales sont réalisées par J.D. Tissot, au centre de transfusion sanguine de Lausanne [11]. Parmi les données concernant les protéines urinaires, la base du *Danish Centre for Human*

Genome Research (université d'Aarhus) recense des marqueurs tumoraux pouvant servir de facteurs pronostiques dans le cancer de la vessie [12]. La banque TMIG-2DPAGE, développée par l'institut de gérontologie de Tokyo est spécifiquement consacrée aux recherches

sur le vieillissement, au travers de l'étude du protéome des fibroblastes [13]. Par ailleurs, l'étude protéomique a permis récemment de mieux caractériser les effets induits par différents agents de l'activation cellulaire ou de la différenciation, en particulier sur des lignées hématopoïétiques [14, 15].

Perspectives

Les avancées futures de l'analyse protéomique seront étroitement liées à l'approfondissement de la connaissance des séquences génomiques. Un exemple des progrès réalisés dans l'établissement du protéome d'un organisme simple est celui du protéome de *Saccharomyces cerevisiae*, découlant directement de la connaissance de son génome [16, 17]. La perspective d'un décryptage complet du génome humain, et par conséquent des séquences polypeptidiques correspondantes, dans la prochaine décennie, constitue un point de départ pour les programmes protéome appliqués au domaine médical. Mais l'impact de ces programmes en biochimie et biologie fondamentales est aussi important. De nouvelles technologies de protéome différentiel devraient, par exemple, fournir une voie d'étude de la biochimie des cellules cancéreuses avec un niveau de raffinement inenvisageable il y a quelques années. Un immense travail de caractérisation fonctionnelle reste aussi à mener sur les protéines et particulièrement sur leurs variants post-traductionnels mis en évidence par l'approche protéomique ■

RÉFÉRENCES

1. Kahn P. From genome to proteome: looking at a cell's proteins. *Science* 1995; 270: 369-70.
2. Wilkins MR, Pasquali C, Appel RD, *et al*. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Bio/Technology* 1996; 14: 61-5.
3. Görg A. Two-dimensional electrophoresis. *Nature* 1991; 349: 545-6.
4. Anderson NL, Anderson NG. Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* 1998; 19: 1853-61.
5. Haynes PA, Gygi SP, Figeys D, Aebersold R. Proteome analysis: biological assay or data archive? *Electrophoresis* 1998; 19: 1862-71.
6. Rabilloud T. Solubilization of proteins for electrophoretic analyses. *Electrophoresis* 1996; 17: 813-29.
7. O'Farrell P. High resolution of 2-dimension electrophoresis of proteins. *J Biol Chem* 1975; 250: 4007-21.
8. Wilkins MR, Gooley AA. Protein identification in proteome project. In: Wilkins MR, Williams KL, Appel, RD Hochstrasser DF, eds. *Proteome Research: New Frontiers in Functional Genomics*. Berlin: Springer-Verlag, 1997: 35-64.
9. Wilkins MR, Gasteiger E, Tonella L, *et al*. Protein identification with N and C-terminal sequence tags in proteome projects. *J Mol Biol* 1998; 278: 599-608.
10. Appel RD, Sanchez JC, Bairoch A, *et al*. The SWISS-2DPAGE database of two-dimensional polyacrylamide gel electrophoresis, its status in 1995. *Nucleic Acids Res* 1996; 24: 180-1.
11. Spertini F, Tissot JD, Dufour N, Francillon C, Frei PC. Role of two-dimensional electrophoretic analysis in the diagnosis and characterization of IgD monoclonal gammopathy. *Allergy* 1995; 50: 664-70.
12. Rasmussen HH, Orntoft TF, Celis JE. Towards a comprehensive database of proteins from the urine of patients with bladder cancer. *J Urol* 1996; 155: 2113-9.
13. Toda T, Kaji K, Kimura N. TMIG-2DPAGE: a new concept of two-dimensional gel protein database. *Electrophoresis* 1998; 19: 344-8.
14. Hanash SM, Teichroew D. Mining the human proteome: experience with the human lymphoid protein database. *Electrophoresis* 1998; 19: 2004-9.
15. Lutomski D, Fouillit M, Bourin P, *et al*. Externalization and binding of galectin-1 on cell surface of K562 cells upon erythroid differentiation. *Glycobiology* 1997; 7: 1193-9.
16. Boucherie H, Sagliocco F, Joubert R, Maillet I, Labarre J, Perrot M. Two-dimensional gel protein database of *Saccharomyces cerevisiae*. *Electrophoresis* 1996; 17: 1683-99.
17. Payne WE, Garrels JI. Yeast protein database (YPD): a database for the complete proteome of *Saccharomyces cerevisiae*. *Nucleic Acids Res* 1997; 25: 57-62.

Raymonde Joubert-Caron

Docteur ès sciences, ingénieur de recherche, Université Paris 13.

Michel Caron

Docteur ès sciences, maître de conférence, praticien hospitalier, Université Paris 13. Biochimie cellulaire des hémopathies lymphoïdes (EA 1625), UFR SMBH-Léonard-de-Vinci, 74, rue Marcel-Cachin, 93017 Bobigny Cedex, France. E-mail: caron@smbh.univ-paris13.fr

TIRÉS À PART

M. Caron.