

Dominique Porquet*

Analyse protéomique : stratégies méthodologiques et applications en biologie clinique

RÉSUMÉ

Le protéome correspond à l'ensemble des protéines produites par un génome donné. Son analyse s'est considérablement développée au cours des 10 dernières années et ouvre de larges perspectives sur la compréhension du fonctionnement d'une cellule ou d'un organe donné. Parmi les différentes approches protéomiques envisageables, n'est abordé dans cet article que l'analyse protéomique qualitative qui seule actuellement présente des applications à court terme en biologie clinique. Pour l'analyse simultanée de nombreuses protéines, la méthode de référence actuelle reste l'électrophorèse 2D suivie d'une analyse des protéines d'intérêt par spectrométrie de masse après digestion partielle par la trypsine. Cette approche permet l'identification de différentes protéines en établissant une véritable carte « d'identité » peptidique de chacune d'entre elles. En cas d'échec d'identification d'une protéine, il est alors fait appel à la spectrométrie de masse en tandem qui permet d'établir la séquence en acides aminés d'un fragment peptidique. Actuellement, le développement très rapide des méthodes de l'analyse protéomique s'oriente vers la mise au point de méthodes de séparation en HPLC ou en électrophorèse capillaire (en alternative à l'électrophorèse en 2D) couplées à la spectrométrie de masse, ou vers l'utilisation de biopuces à protéines. Certaines approches proposent d'ailleurs une stratégie en deux temps avec le fractionnement d'un mélange protéique à l'aide de biopuces suivi d'une analyse de chaque fraction en spectrométrie de masse. De nombreuses applications de la protéomique en biologie clinique sont en cours de développement. Elles correspondent à des études à visée diagnostique ou à visée thérapeutique sur des agents pathogènes ainsi qu'à l'établissement d'outils diagnostiques dans différents domaines : maladies auto-immunes, pathologies cancéreuses, pathologies rénales, pathologies du système nerveux central.

MOTS CLÉS

Protéome, électrophorèse 2D, spectrométrie de masse, biopuces à protéines, biologie clinique

Proteomic analysis : methods and clinical applications

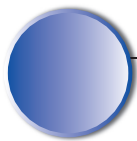
SUMMARY :

Proteome corresponds to all the proteins produced by a given genome. Its analysis considerably developed during the last 10 years and opens a lot of perspectives on the understanding of the functioning of a cell or a given organ. Among various approaches, we consider qualitative aspects of proteomic analysis, which only presents short-term applications in clinical biology. For the simultaneous analysis of numerous proteins, the current reference method is the 2D electrophoresis followed by a mass spectrometry analysis. This approach allows the identification of various proteins by establishing a mass peptide finger print of each. When identification is not possible, a tandem mass spectrometry is required to establish the aminoacid sequence of a peptidic fragment. At present, the very fast development of methods of proteomic analysis, turns to the development of separation methods based on HPLC or capillary electrophoresis (as a replacement of 2D electrophoresis) coupled with the mass spectrometry or towards the use of protein arrays. Moreover, a lot of approaches propose a two step strategy with a separation step of a protein mixture by means of biochips, followed by a treatment of every fraction by mass spectrometry. Clinical application concern mainly studies with diagnostic or therapeutic aim on pathogenic agents and also the establishment of diagnostic tools in various fields: autoimmune diseases, cancer diseases, renal diseases, and pathologies of central nervous system.

KEYWORDS

Proteome, 2D electrophoresis, mass spectrometry, protein biochips, clinical biology

* Service de Biochimie-Hormonologie - Hôpital Robert Debré - 48, Bd. Sérurier - 75935 Paris Cedex 19



I - Introduction

Voilà maintenant 10 ans que Marc Wilkins lors du premier congrès de Sienna sur la protéomique utilisait pour la première fois le terme de protéome. Depuis lors, c'est tout un champ de recherche qui s'est ouvert et cette évolution a permis à la fois des progrès méthodologiques très importants et des avancées très prometteuses dans de nombreux domaines physiopathologiques, ce qui est également de nature à susciter de grands espoirs pour la mise en évidence de nouvelles cibles thérapeutiques.

Quelques définitions pour commencer. Dans son acception la plus large, le terme de protéome caractérise l'ensemble des PROTEINES exprimées par le génOME. De façon plus restrictive mais également plus accessible, le protéome est l'ensemble des protéines « exprimées » par une cellule (ou un tissu, ou un organe) à un instant donné. Si l'analyse d'un protéome est toujours, et cela va être abordé un peu plus loin, un challenge complexe, l'intérêt de cette approche n'est plus à démontrer :

- seule la protéine (par rapport à une séquence d'ADN ou un ARNm) présente un effet biologique ;
- un même génome peut donner naissance à de nombreux protéomes différents ;
- le niveau d'expression d'une protéine ne peut que rarement être prédit à partir du niveau d'expression du message correspondant ;
- la (ou plutôt) les protéines « issues » d'un même gène sont variables du fait, entre autre, des très nombreuses modifications post-traductionnelles dont une protéine peut faire l'objet.

Challenge complexe est-il dit un peu plus haut. En effet, une approche « protéomique » suppose dans tous les cas l'analyse simultanée d'un très grand nombre de protéines, en vue de caractériser leurs structures, leurs fonctions, leurs interactions avec d'autres composants cellulaires (et en particulier d'autres protéines) ainsi que les variations de ces caractéristiques en fonction du contexte cellulaire. Cette analyse simultanée d'un grand nombre de protéines se heurte à un certain nombre de difficultés : il n'est pas possible à l'instar de ce qu'il est habituel de faire pour des séquences nucléotidiques d'« amplifier » des protéines pour mieux les étudier ; les protéines sont parfois difficiles à extraire ; au sein d'une même solution, des protéines peuvent être présentes à des concentrations extrêmement différentes (l'écart de concentration pouvant atteindre 6 logs) ; enfin, le nombre total de protéines à analyser dans un échantillon, et ce quelle que soit son origine, est très élevé (une cellule à un instant donné peut exprimer environ 10^4 protéines différentes et un individu à un instant donné jusqu'à 10^6 protéines différentes).

II – Les différents types d'analyses protéomique

On peut distinguer de façon conceptuelle deux

grands types d'analyses protéomiques : la protéomique descriptive et la protéomique fonctionnelle. Dans le premier cas, celui de la **protéomique descriptive**, il s'agit de répondre aux quatre questions suivantes : quelles protéines sont exprimées ; quel est le niveau de cette expression ; quelle est la localisation subcellulaire de ces différentes protéines et quelles sont les modifications post-traductionnelles dont elles font l'objet. Dans le second cas, celui de la **protéomique fonctionnelle**, et afin d'approcher la fonction des protéines, on cherchera à répondre à des questions du type : quelles sont les molécules qui au sein de la cellule s'associent à nos protéines d'intérêt ; quel type de modifications structurales est responsable de quel type de modifications d'activité.

En ce qui concerne la protéomique descriptive, la réponse à la question : quelles protéines ? relève plus précisément de la **protéomique qualitative** et ses techniques font appel à l'électrophorèse bidimensionnelle couplée à la spectrométrie de masse, à des techniques séparatives diverses couplées à une analyse en spectrométrie de masse en tandem ou à des biopuces à protéines. L'aspect quantitatif de l'expression des protéines (**protéomique quantitative**) repose au plan méthodologique sur le principe du marquage différentiel effectué soit avant la phase d'extraction des protéines (approche *in vivo* par la méthode SILAC (Stable Isotope Labelling with Aminoacids in Cell culture) soit après l'extraction de ces mêmes protéines (approche *in vitro* par marquage par des fluorochromes, par la méthode ICAT (Isotope Coded Affinity Tag) ou par la méthode ALICE (Acid-Labile Isotope-Coded Extractant). En ce qui concerne la localisation subcellulaire de la production des protéines, elle relève de la **protéomique « topologique »** et s'appuie essentiellement sur la technologie MELK (Multiple-Epitope-Ligand « Kartographie »). Enfin, pour ce qui est des modifications post-traductionnelles, il est fait appel à une approche dite de Shotgun MS consistant en une chromatographie en phase liquide (LC) suivie d'une spectrométrie de masse en tandem (MS/MS) sur un échantillon biologique correspondant à des peptides obtenus après clivage par différentes protéases.

Dans le cas de la protéomique fonctionnelle, la protéomique d'interaction, qui vise à mettre en évidence les différents partenaires d'une protéine au sein d'un milieu complexe ou d'une cellule, repose sur plusieurs approches méthodologiques : le système double hybride, le FRET (Fluorescence Resonance Energy Transfer), l'analyse de complexes après purification (par des méthodes biochimiques conventionnelles ou par Tandem Affinity Purification « TAP ») ou l'utilisation de sondes chimiques multifonctionnelles. La **protéomique structurale** quant à elle fait largement appel à des techniques telles que la cristallographie aux rayons X, la spectrométrie par résonance magnétique nucléaire et la tomographie d'électrons.

Analyse protéomique : stratégies méthodologiques et applications en biologie clinique

Dans la suite de cet article, seules les techniques de la protéomique qualitative (descriptive) vont être détaillées. En effet, dans le cadre de l'application des approches protéomiques à la biologie clinique, seule la protéomique descriptive présente actuellement des applications à court terme.

III – Les techniques de l'analyse protéomique qualitative

Très classiquement, et comme cela a été brièvement cité ci-dessus, l'approche utilisée est celle de l'électrophorèse bidimensionnelle (2D) suivie par une identification des protéines d'intérêt par spectrométrie de masse.

1. Extraction et séparation des protéines

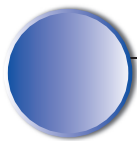
La première étape de cette approche, en fait la plus délicate et la plus limitante, est la phase d'extraction des protéines. Cette extraction reste assez simple dans le cas d'un liquide biologique ; cela devient d'emblée beaucoup plus délicat dans le cas d'un tissu. Beaucoup de protéines sont en effet hydrophobes et donc difficiles à extraire, ce qui est également le cas de protéines très fortement ancrées à des structures cellulaires. Il est alors fait appel à des solutions d'extraction associant des agents chaotropes, des détergents neutres et des réducteurs. De plus, l'addition d'anti-protéases permet de réduire la dégradation des protéines lors de cette phase d'extraction. La deuxième étape correspond à la séparation en 2D proprement dite. Dans la première dimension, la séparation s'effectue selon le principe de la focalisation isoélectrique et de plus en plus de gradients de PH préformés, étroits ou larges, sont disponibles sur le marché. Pour la seconde dimension, il s'agit d'un classique gel de polyacrylamide en présence d'agents dénaturants (SDS). En définitif, les protéines sont donc séparées en fonction de leurs masses et charges respectives. En ce qui concerne la masse, l'intervalle de 7000 Da à 200 000 Da peut être retenu. La troisième étape de l'analyse consiste à révéler et analyser les spots protéiques obtenus. La révélation peut faire appel à l'utilisation de colorants (Bleu de Coomassie), de marquages fluorescents, à la chimioluminescence ou au marquage isotopique. La détection se fait ensuite par densitométrie, ou toute autre méthode appropriée utilisant un scanner, une caméra, un imager ...

2. Analyse des gels d'électrophorèse

Après révélation, on obtient ainsi une véritable **carte protéique** de l'échantillon étudié. Cette carte peut alors être analysée aux moyens d'outils bioinformatiques permettant la comparaison de cartes protéiques entre elles et également la

comparaison avec des cartes « témoins », ce qui peut permettre dans des situations favorables l'identification d'un certain nombre de protéines. Dans les cas les moins favorables, la comparaison de deux cartes protéiques obtenues sur un même type d'échantillon, mais correspondant à des « situations » différentes, va cependant permettre de visualiser des protéines dont l'expression est fortement modifiée entre ces deux situations. La façon la plus simple d'identifier ces protéines est de recourir à la spectrométrie de masse. Classiquement, chaque protéine à identifier est éluee à partir du gel en 2D, puis soumise à une digestion trypsique et les peptides ainsi obtenus sont analysés en MALDI-TOF. Le MALDI-TOF est un spectromètre de masse dont le système d'ionisation repose sur la cristallisation de l'échantillon à analyser dans une matrice (MALDI = Matrix-Associated Laser Desorption-Ionization) et la mesure du rapport m/z des peptides est effectuée en temps de vol (TOF = Time Of Flight). Cette analyse permet l'obtention d'une **carte peptidique** où chaque peptide obtenu après l'hydrolyse trypsique est représenté par son rapport m/z. Cette carte peptidique peut à son tour être comparée à de très nombreuses cartes peptidiques contenues dans des bases de données bioinformatiques, bases facilement accessibles sur Internet. Assez fréquemment, cela va permettre l'identification de la protéine en question, pour peu que cette protéine ait déjà été « cartographiée » selon cette approche. Dans les cas où aucune banque de données n'est susceptible de permettre l'identification de la protéine, on a alors recours à l'utilisation d'une spectrométrie de masse en tandem. Le principe de cette approche méthodologique est la suivante : dans le premier analyseur de masse, il y a séparation des peptides issus de la digestion trypsique ; ces peptides sont alors introduits dans une chambre de fragmentation permettant de générer à partir d'un peptide donné (sélectionné spécifiquement par son rapport m/z) des peptides fils ayant perdu un certain nombre d'acides aminés (la fragmentation doit permettre de cibler spécifiquement les liaisons peptidiques). Les peptides fils ainsi obtenus sont alors analysés dans un second spectromètre de masse qui les sépare à nouveau en fonction de leur rapport m/z ; enfin, les masses des peptides fils sont comparées à la masse du peptide parental, permettant ainsi par différence de masse d'établir la séquence d'acides aminés. Cette séquence peut alors être rentrée dans des banques de données où sont stockées des séquences primaires de protéines obtenues par séquençage ou par traduction à partir des séquences nucléotidiques d'ADNc.

Plusieurs types de spectromètres de masse en tandem peuvent être utilisés pour l'analyse protéomique combinant des ionisations de type MALDI



ou de type ESI (Ionisation par ElectroSpray) et quatre types d'analyse de masse (combinés 2 à 2) : la capture d'ions, le temps de vol, le quadropole, les transformées de Fourier/cyclotron.

IV – Evolutions récentes en analyse protéomique qualitative

L'approche classique décrite précédemment (électrophorèse 2D suivie d'une analyse par spectrométrie de masse) reste une méthode de référence mais fait actuellement l'objet de nombreuses évolutions.

1. L'automatisation de l'étape séparative

Une première série d'évolutions consiste à automatiser toutes les étapes de l'analyse en 2D et de nombreuses plateformes de ce type sont maintenant proposées par différents fabricants. Afin toujours de simplifier l'étape séparative préalable à l'analyse en spectrométrie de masse, certains auteurs proposent également de recourir à une séparation monodimensionnelle reposant sur une seule électrophorèse en gel de polyacrylamide (PAGE-SD). Enfin et il s'agit là d'une évolution majeure, de plus en plus d'expérimentations sont conduites pour coupler de façon automatique et directe l'étape séparative et l'analyse en spectrométrie de masse. Cette étape séparative repose alors sur des chromatographies en HPLC ou en nanoHPLC dites « multidimensionnelles » car associant plusieurs types de séparation chromatographique (exclusion/phase reverse ; affinité/phase reverse ; échange d'ions/affinité ...).

De façon alternative aux séparations par chromatographie, il est également possible d'utiliser des techniques électrophorétiques et en particulier l'électrophorèse capillaire qui peut être réalisée sur des supports miniaturisés du type « laboratoire sur une puce » (Lab on a Chip) comme le propose plusieurs fabricants (Agilent Technologies par exemple).

Les techniques séparatives ont parfois atteint un tel niveau d'efficacité qu'elles peuvent être, dans certaines conditions particulièrement favorables, conduites directement sur des protéines qui n'ont pas été soumises à une digestion trypsique préalable.

2. Les biopuces dédiées à la protéomique

A côté des méthodologies précédemment citées se développe également la technologie des biopuces à protéines ou protein array. Le principe se veut similaire à ce qui est maintenant largement développé pour les études du transcriptome et qui consiste à immobiliser sur des puces des ADNc ou des oligonucléotides de synthèse ; les ARNm

à identifier sont alors déposés sur ces puces, le principe général de l'hybridation par complémentarité de bases permettant la fixation des ARNm sur les séquences nucléotidiques qui leur sont complémentaires ; le marquage préalable de ces ARNm, lors de leur amplification par RT-PCR, permet par la suite la localisation sur la puce de ces hybridations et ainsi l'identification des ARNm correspondants. Dans le cas des biopuces à protéines, le réactif déposé sur ces puces, réactif qui va permettre la fixation sélective de ces protéines, peut être variable et dépend initialement du type de biopuces développées, biopuces analytiques ou biopuces fonctionnelles. Le réactif déposé peut ainsi être un anticorps, un antigène, un allergène, un aptamère, un substrat ou plus généralement tout « ligand » susceptible de reconnaître et de lier spécifiquement une protéine donnée.

S'il n'y a encore actuellement que peu d'exemples en pratique de l'utilisation de tels systèmes, en revanche certains fabricants ont développé des systèmes associant l'utilisation de puces à protéines, en tant que méthode initiale de séparation et d'isolement de protéines à partir d'un mélange, à une identification plus ou moins poussée de ces protéines par spectrométrie de masse selon la technologie SELDI-TOF-MS (Surface Enhanced Laser Desorption/Ionization-Time Of Flight-Mass Spectrometry). C'est le cas par exemple du système développé par la société CIPHERGEN. Le principe en est le suivant : dans une première étape, les échantillons biologiques à analyser sont déposés sur des surfaces réactives (puces) sur lesquelles ont été fixés des groupements chimiques permettant une fixation selon des principes classiques de chromatographie par rétention ou de chromatographie d'affinité. Après une étape de lavage qui permet l'élimination des protéines qui n'ont pas été fixées sur le support, un réactif spécifique est ajouté qui va permettre de stabiliser les protéines fixées sur la puce dans une matrice cristalline. Dans une seconde étape, la puce est soumise à un rayonnement LASER qui va entraîner la désorption et l'ionisation des protéines qui vont être analysées dans un spectromètre de masse qui mesure leur rapport m/z en temps de vol. Ceci va conduire à établir un profil protéique particulier de l'échantillon biologique. Ce système est plus précisément destiné à la biologie clinique où il peut servir d'outil diagnostique ; en effet, il est possible avec cette approche de comparer deux profils protéiques correspondant à un même type d'échantillons biologiques mais préparés à partir de deux individus différents ou chez un même individu mais à des moments différents. Par rapport au profil de l'un de ces deux échantillons qui va être considéré comme « témoin normal », toute différence qualitative ou quantitative du profil du deuxième échantillon va « signer » une situation physio-pathologique particulière et a priori différente de la normalité. Il est même possible, pour un

Analyse protéomique : stratégies méthodologiques et applications en biologie clinique

type d'échantillon biologique parfaitement défini (plasma, urines, LCR) et traité dans des conditions standardisées, d'établir un profil type de normalité auquel vont être comparés tous les profils établis dans les mêmes conditions chez une population de sujets à explorer. La société CIPHERGEN propose d'ores et déjà des protocoles et des biopuces dédiés à des activités diagnostiques dans différents domaines comme par exemple le cancer, les pathologies neuropsychiatriques (dépression, maladie de Parkinson, maladie d'Alzheimer ...), les pathologies infectieuses, les pathologies rénales

V - Les applications de la protéomique en biologie clinique

De façon très générale, les principaux domaines d'application de la protéomique en biologie clinique peuvent être regroupés de la façon suivante :

1) Etudes à visée diagnostique ou thérapeutique (cibles vaccinales) sur des agents pathogènes

Ce type d'études permet de caractériser différentes souches d'agents pathogènes, d'identifier plus précisément des protéines impliquées, soit dans la pathogénicité de ces souches, soit dans les interactions hôtes-pathogènes, et d'évaluer différents types de mécanisme de résistance.

2) Evaluation et développement de nouveaux médicaments

Cela correspond à des possibilités d'évaluer de nouvelles cibles pharmacologiques, d'étudier l'effi-

cacité d'un médicament et également d'explorer sa toxicité potentielle.

3) Constitution d'outils diagnostiques

Il s'agit là d'un champ d'investigation extrêmement étendu. En effet, la capacité à établir et à comparer des profils protéomiques doit contribuer à la mise en évidence et au développement de marqueurs spécifiques de pathologies. C'est le cas par exemple, pour les maladies auto-immunes, les pathologies cancéreuses, les pathologies rénales (étude en particulier de protéomique « urinaire »), et les pathologies du système nerveux central, tout particulièrement les pathologies neurodégénératives (étude de protéomique « rachidienne » à partir de LCR).

VI - Conclusion

Quoique encore délicates et relativement lourdes à mettre en œuvre, les techniques de l'analyse protéomique évoluent cependant très rapidement et leurs simplifications et leurs automatisations laissent entrevoir qu'une approche protéomique globale sera de plus en plus souvent à l'avenir évoquée pour toute démarche diagnostique ou de suivi thérapeutique en biologie clinique. D'ores et déjà, la mise en évidence et la caractérisation de marqueurs tumoraux totalement spécifiques d'une tumeur donnée constitue une des applications privilégiées de l'analyse protéomique. De nombreux autres domaines de la physiopathologie peuvent également être abordés via l'analyse protéomique comme la recherche de marqueurs dans les pathologies rénales, les pathologies neuro-dégénératives, les intoxications et les pathologies infectieuses.

POUR EN SAVOIR PLUS

CAHILL DJ, NORDHOFF E. Protein arrays and their role in proteomics. *Adv. Biochem. Eng. Biotechnol.*, 2003, **83**, 177-187.

ZHU H, BILGIN M, SNYDER M. *Proteomics. Annu. Rev. Biochem.*, 2003, **72**, 783-812.

GERLING IC, SOLOMON SS, BRYER-ASH M. Genomes, transcriptomes and proteomes: molecular medicine and its impact on medical practice. *Arch. Intern. Med.*, 2003, **163**(2), 190-198.

PATTERSON SD, AEBERSOLD RH. Proteomics: the first decade and beyond. *Nat. Genet.* 2003, **33**, 311-323.

THÉBAULT S, MACHOUR N, PERRON F, JOUENNE T, LANGE C, HUBERT M, FONTAINE M, TRON F, CHARLIONET R. Objet et évolution méthodologique de l'analyse protéomique. *Médecine/Sciences* 2001, **17**, 609-618.

RABILLOUD T., Progrès récents et évolutions en analyse protéomique. *Spectra Analyse* 2002, **31**, 225, 26-31.

A. BERNOT. Analyse des génomes – Transcriptomes et Protéomes. Editeur Dunod. Collection : Masson sciences. 2001 (3ème édition).

RIGHETTI PG, CASTAGNA A, ANTONUCCI F, PIUBELLI C, CECCONI D, CAMPOSTRINI N, ZANUSSO G, MONACO S. The proteome: anno Domini 2002. *Clin. Chem. Lab. Med.*, 2003, **41**, 425-438.

LILL J., Proteomic tools for quantitation by mass spectrometry. *Mass Spectrom. Rev.*, 2003, **22**, 182-194.

LOPEZ MF, PLUSKAL MG., Protein micro- and macroarrays: digitizing the proteome. *J. Chromatogr. B Analyt Technol Biomed Life Sci*, 2003, **787**, 19-27.

LEFKOVITS I., Functional and structural proteomics: a critical appraisal. *J. Chromatogr. B Analyt Technol Biomed Life Sci*, 2003, **787**, 1-10.

PETRICCOIN EE, PAWELETEZ CP, LIOTTA LA., Clinical applications of proteomics: proteomic pattern diagnostics. *J. Mammary Gland Biol. Neoplasia*, 2002, **7**, 433-440.

MORTUAIRE G, MARCHETTI P, FORMSTECHEP P, DANZÉ PM. Nouvelle approche globale en protéomique : les biopuces à protéines. Techniques actuelles et applications. *Ann. Biol. Clin.*, 2004, **62**, 139-148.