

Direction de l'Évaluation des Médicaments Et des  
produits Biologiques  
Département de toxicologie  
Chef de département : D. Masset  
Secrétaire scientifique : A. Sanh

## POSITION PAPER SUR LA PROTEOMIQUE

Version n°1 du 29 avril 2009

Document préparé par le Groupe de Travail « Innovation non Clinique »

- **Coordinateur de la rédaction**
  - T. RABILLOUD
- **Président**
  - J.-R. CLAUDE
- **Membres**
  - L. DOMENJOUR
  - E. FATTAL
  - J. GUILLEMAIN
  - A. GUILLOUZO
  - S. LE CROM
  - A. LE PAPE
  - S. LERONDEL
  - P. LESCUYER
  - F. MOREL
  - M. PALLARDY
  - G. PELTRE
  - C. PINEAU
  - T. RABILLOUD
  - R. RAHMANI
  - D. TREMBLAY<sup>†</sup>
- **Représentants de l' Afssaps invités**
  - D. ABDON
  - D. SAUVAIRE

# **Le point sur l'analyse protéomique dans l'industrie pharmaceutique: le défi analytique n'est-il pas tout simplement trop grand ?**

## **Version n°1 du 29 avril 2009**

### **1. Objet**

Les techniques à grande échelle comme les "omiques", la chimie combinatoire et le criblage à haut débit, sont très utilisées dans l'industrie pharmaceutique aux cotés d'approches plus ciblées. Parmi ces approches à grande échelle, l'analyse protéomique représente un cas particulier. Il semble en effet y avoir un fossé grandissant entre les espoirs placés dans la technique et les réalisations, ce qui conduit à un fort questionnement sur l'importance et le positionnement de l'analyse protéomique dans l'industrie pharmaceutique.

Le but de ce document est de rappeler les forces et les faiblesses de l'analyse protéomique. Leur mise en perspective permettra de recadrer les espoirs que l'on peut raisonnablement placer dans la technique et de mieux définir des stratégies dans lesquelles l'analyse protéomique pourra apporter sa pleine contribution.

### **2. Analyse de la situation**

L'industrie pharmaceutique et du diagnostic est engagée dans une rude compétition pour trouver de nouvelles molécules, de nouvelles cibles et de nouveaux marqueurs. Dans cette course permanente, les criblages intenses et parallèles sont de plus en plus utilisés (e.g., la chimie combinatoire ou le criblage à haut débit). Dans cet ordre d'idée, il est logique que l'analyse protéomique, que l'on peut définir comme un criblage à haut débit de protéines, ait suscité un fort intérêt. Cependant, de plus en plus de voix se font entendre pour exprimer leur déception quant aux résultats réels de l'analyse protéomique dans l'industrie pharmaceutique.

Comme souvent, ce décalage entre attentes et réalité traduit plus des espoirs inconsidérément élevés qu'une réelle faillite de la technique, et en ce sens, il est fondamental d'avoir une perception saine de ce que l'analyse protéomique peut réellement faire ou ne pas faire. Il sera alors possible d'intégrer l'analyse protéomique à sa juste place dans le corpus méthodologique utilisé par l'industrie pharmaceutique, et d'adapter les investissements en conséquence.

#### *2.1 Contraintes d'ordre analytique*

L'analyse protéomique n'est rien d'autre que de la chimie analytique des protéines à grande échelle, mais chaque mot de cette définition a son importance. "A grande échelle" implique une notion de robustesse, en ce sens qu'un nombre limité de conditions analytiques (par exemple de solvants) doit pouvoir couvrir une proportion maximale d'analytes (ici les protéines). Cette notion s'applique pleinement aux "omiques" mettant en jeu les acides nucléiques, car ceux ci sont toujours très solubles dans les solutions aqueuses, quelle que soit leur séquence. Mais ce n'est pas vrai pour les protéines. Certaines d'entre elles sont très solubles dans l'eau, certaines préfèrent les solutions salines alors que d'autres nécessitent à la fois un environnement aqueux et lipidique pour être pleinement structurées et actives. De ce simple fait, il n'existe pas de solvant "universel" pour les protéines, ce qui implique de facto qu'il n'y a pas d'exhaustivité possible en analyse protéomique, au rebours de la situation dans les autres "omiques".

On peut penser que ce problème de conditions de solubilité est vrai pour les protéines, mais devrait être nettement moins aigu pour les peptides, qui sont des analytes plus simples. En conséquence, les techniques de protéomique avec la plus grande homogénéité se trouvent dans les approches où ce ne sont pas les protéines entières qui sont analysées, mais uniquement des peptides issus directement de la digestion protéolytique des protéines d'intérêt [1,2]. Bien que cette assertion soit vraie dans l'ensemble, les peptides présentent une diversité chimique suffisante pour ne pas être tous solubles dans les quelques conditions analytiques compatibles avec une nécessaire analyse à très haute sensibilité.

#### *2.2. Contraintes d'origine biologique*

Cette très grande sensibilité est absolument indispensable du fait que de nombreuses protéines sont présentes dans les échantillons biologiques à de très faibles concentrations, alors même que certaines protéines sont présentes à des concentrations bien plus grandes. De façon évidente, cette

dynamique d'expression (*i.e.*, le rapport quantitatif entre les protéines les plus rares et les plus abondantes) varie considérablement selon le type d'échantillon biologique considéré. Celle-ci s'étend sur 4 ordres de grandeur chez *E. coli*, 5 sur la levure [3], 6 dans une cellule de mammifère et jusqu'à 12 dans les fluides biologiques complexes tels que le plasma [4].

En plus du défi lié à la dynamique d'expression, se pose le problème de la complexité et plus particulièrement du nombre d'analytes différents à prendre en compte. Alors que nous avons maintenant une notion claire de la complexité des génomes, celle des protéomes est nettement moins bien appréhendée. On est sûr cependant que cette complexité excède très largement celle des génomes. Par exemple, les modifications posttraductionnelles sont connues pour moduler, parfois fortement, la fonctionnalité et la localisation des protéines, de sorte que des variants posttraductionnels du même produit génique doivent être considérés comme différents et analysés en tant que tels.

Là encore, la situation est très différente d'un organisme à l'autre. Les procaryotes les plus simples présentent probablement au plus 3000 formes protéiques différentes [5,6], mais on attend de l'ordre de 20.000 formes protéiques dans une cellule de mammifère [7] alors que les fluides biologiques complexes doivent contenir plusieurs centaines de milliers de formes protéiques différentes.

### 2.3. Conséquences pratiques

La mise en correspondance de toutes ces contraintes d'origine biologique d'une part, et des capacités analytiques de la protéomique d'autre part montre immédiatement que la puissance de l'analyse protéomique va varier considérablement selon le système biologique étudié. La protéomique bactérienne (par exemple d'organismes pathogènes) n'est certes pas exhaustive, mais couvre sans problème 70-80% du protéome. En revanche il n'est possible d'analyser que moins de la moitié du protéome des cellules de mammifère, et une toute petite fraction de celui des fluides biologiques [8]. De fait, quand on part de 12 ordres de grandeurs, éliminer parfaitement les quelques protéines qui représentent 99% de la masse protéique réduit la complexité à 10 ordres de grandeur, ce qu'aucune technique analytique ne peut aborder à ce jour. De plus, dans la plupart des applications, une simple liste de protéines n'est pas suffisante, et il est nécessaire d'effectuer des mesures quantitatives comparatives entre différents échantillons. Ces mesures quantitatives compliquent encore l'analyse et en diminuent les performances en termes de nombre d'analytes pris correctement en compte.

Ce manque criant d'exhaustivité empêche d'utiliser l'analyse protéomique dans l'industrie pharmaceutique de la même façon que les autres approches de criblage massivement parallèles. De même, il est évident que les usages possibles de la protéomique dépendront fortement de l'échantillon biologique considéré. Il est ainsi envisageable de catégoriser les usages possibles de la protéomique en fonction des champs thématiques d'intérêt.

Dans le processus de découverte, la non-exhaustivité de la protéomique pose des problèmes d'acuité variés selon le système considéré. Si des systèmes simples comme les bactéries peuvent être étudiés très en détail, l'incapacité à analyser les protéines minoritaires constitue un handicap majeur pour trouver des cibles pertinentes dans les systèmes plus complexes comme les cellules de mammifères. Ainsi pour pouvoir analyser les protéines minoritaires, il sera nécessaire de les enrichir considérablement, ce qui implique de devoir poser des hypothèses de recherche restrictives. En conséquence, la protéomique perdra de son caractère général et donc une partie importante de son intérêt.

Dans la recherche de biomarqueurs, il apparaît essentiel d'affirmer que les stratégies évidentes, comme de comparer des échantillons sanguins de personnes saines vs. atteintes d'une pathologie d'intérêt, sont tout simplement vouées à l'échec. Ces stratégies ont déjà été abondamment appliquées, et n'ont donné pour tout résultat que des protéines de l'inflammation, certes non spécifiques d'une pathologie, mais facilement analysables.

## 3. Propositions pour le futur

A partir de ce constat, une analyse rétrospective devrait permettre de mettre en place de meilleures stratégies. La plupart des marqueurs protéiques sanguins utilisés aujourd'hui sont des protéines relarguées par les tissus dans le flux sanguin, souvent à basse concentration [4]. Même s'il est probable que ces protéines seront détectables par spectrométrie de masse dans un futur assez

proche [10] ces protéines sont actuellement détectées par immunoessai ou via leur activité enzymatique, quand elles en ont une aisément mesurable. Là encore, l'exemple est instructif. Il est probable que toutes les activités enzymatiques aisément dosables ont été testées en tant que biomarqueurs, et seules quelques unes sont effectivement utilisées. Les causes d'échec sont multiples, comme par exemple une variabilité individuelle excessive, une stabilité insuffisante, et bien sûr l'absence de corrélation avec un état pathologique.

Il n'y a aucune raison pour que le taux d'attrition observé sur les marqueurs enzymatiques ne se retrouve pas sur d'autres marqueurs protéiques trouvés par d'autres moyens, par exemple en analyse protéomique. Mais ce taux d'attrition fait partie de la règle du jeu dans la recherche de biomarqueurs, quelle que soit l'approche utilisée.

En conséquence, il est absolument nécessaire de mettre en place des stratégies moins générales, mais plus adaptées et plus efficaces.

Une première stratégie est de partir d'un fluide plus simple que le plasma, comme le LCR. La recherche de marqueurs de la maladie de Creutzfeldt-Jakob dans le LCR est un des rares exemples de succès de l'analyse protéomique [11] bien que le marqueur mis en évidence soit loin d'être parfait [12]. Cependant, un tel exemple valide l'approche stratégique choisie, qui reste toutefois d'application limitée.

Si les biomarqueurs doivent être recherchés dans le sang, ce qui est le cas le plus répandu, deux stratégies peuvent être mises en place. La première place comme présupposé que les marqueurs seront issus de "fuites" tissulaires. Il s'agit alors de rechercher des protéines relarguées différenciellement par le tissu malade par rapport au tissu sain et aux autres tissus de l'organisme. Une telle approche peut être menée en première intention sur des modèles cellulaires, puis validée sur le fluide biologique d'individus malades (voir par exemple [13]). Cependant, une telle stratégie gagnerait probablement à être conduite directement sur des "sécrétomes" (*i.e.*, sur des protéines relarguées par les cellules dans le milieu de culture) plutôt que sur des extraits cellulaires totaux [14].

La seconde stratégie consiste au contraire en l'analyse du fluide biologique d'intérêt *in toto*. Cependant, pour accéder aux composants minoritaires, il est illusoire d'utiliser une analyse totale ou des stratégies de déplétion. Il faudra dans ce cas utiliser des stratégies de sélection positive. Un bon exemple récemment publié porte sur l'enrichissement et l'analyse ciblée de glycoprotéines plasmatiques [15], mais d'autres moyens de sélection positive peuvent être envisagés, comme la liaison à des cofacteurs (*e.g.*, ATP). A condition de ne pas sélectionner une ou des protéines majoritaires du plasma, ces stratégies de sélection positive devraient permettre d'étudier les composants minoritaires des fluides biologiques, au prix cependant d'une hypothèse *a priori* sur certaines caractéristiques structurales des protéines recherchées.

L'analyse protéomique peut aussi être utilisée en toxicologie, domaine où elle a montré son potentiel très tôt [16]. Même si les études toxicologiques sont très lourdes, elles peuvent utiliser des marqueurs indirects, ce qui diminue l'impact du manque d'exhaustivité de l'analyse protéomique. Ici aussi se posent les problèmes de gamme dynamique observés sur les fluides biologiques. La gamme dynamique plus faible observée sur les cellules et les tissus rend cependant la protéomique plus attractive [17,18] en analyse toxicologique. Toutefois, la lourdeur combinée des études toxicologiques et de la protéomique a considérablement limité l'utilisation de cette approche, alors même que ce champ est probablement un de ceux où le potentiel de l'analyse protéomique est le plus grand.

L'analyse protéomique peut aussi être mise à profit dans des domaines plus marginaux, comme le contrôle des bioproductions [19] ou le contrôle de qualité des produits biotechnologiques. Dans ce dernier domaine, des stratégies d'"égalisation" permettant une meilleure analyse des composants traces devraient se révéler très puissantes et précieuses [20].

Enfin le diagnostic fin de l'allergie pourrait être un cas particulier d'utilisation des techniques protéomiques. En effet, le répertoire des allergènes ou allergome tel que nous le connaissons est très limité. Chaque grande source d'allergènes telles que les acariens, les pollens d'arbres ou de graminées, le lait, l'oeuf, etc. ne contiennent que moins d'une vingtaine d'allergènes ; un sous-ensemble bien moins complexe que le protéome de ces mêmes sources, riche de plusieurs milliers de protéines. La majorité de ces allergènes se trouve de plus parmi les protéines les plus abondantes. Ainsi une analyse par électrophorèse bidimensionnelle, même de faible résolution, suivie d'empreinte et de détection des allergènes reconnus par les IgE du sérum d'un seul patient pourrait constituer une alternative à un test d'exposition ou de réintroduction alimentaire, souvent praticable uniquement en

milieu hospitalier et potentiellement dangereux. Cette approche est susceptible d'intéresser, dans l'industrie pharmaceutique, les firmes préparant des allergènes ou fabriquant des dispositifs ou des bioréactifs.

#### **4. Conclusion**

En conclusion, l'analyse protéomique peut être un outil utile dans l'arsenal méthodologique de l'industrie pharmaceutique et diagnostique. Cependant, les contraintes rencontrées, et esquissées ci-dessus, qu'elles soient d'ordre analytique ou d'origine biologique, doivent absolument être prises en compte pour éviter de cruelles désillusions. C'est à cette condition que l'analyse protéomique pourra être utilisée de façon efficace dans l'industrie pharmaceutique et diagnostique, comme l'évoquent les quelques pistes mentionnées dans ce document.

## Références

- [1] Schirle M, Heurtier MA, Kuster B.  
Profiling core proteomes of human cell lines by one-dimensional PAGE and liquid chromatography-tandem mass spectrometry.  
Mol Cell Proteomics. 2003; 2: 1297-1305
- [2] Wolters DA, Washburn MP, Yates JR 3rd.  
An automated multidimensional protein identification technology for shotgun proteomics.  
Anal Chem. 2001; 73: 5683-5690
- [3] Lu P, Vogel C, Wang R, Yao X, Marcotte EM.  
Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation.  
Nat Biotechnol. 2007; 25: 117-124
- [4] Anderson NL, Anderson NG.  
The human plasma proteome: history, character, and diagnostic prospects.  
Mol Cell Proteomics. 2002; 1: 845-867
- [5] Wilkins MR, Sanchez JC, Williams KL, Hochstrasser DF.  
Current challenges and future applications for protein maps and post-translational vector maps in proteome projects.  
Electrophoresis. 1996; 17: 830-838
- [6] Tonella L, Hoogland C, Binz PA, Appel RD, Hochstrasser DF, Sanchez JC.  
New perspectives in the Escherichia coli proteome investigation.  
Proteomics. 2001 ; 1: 409-423
- [8] Lescuyer P, Hochstrasser D, Rabilloud T.  
How shall we use the proteomics toolbox for biomarker discovery?  
J Proteome Res. 2007; 6: 3371-3776
- [7] Duncan R, McConkey EH.  
How many proteins are there in a typical mammalian cell?  
Clin Chem. 1982; 28: 749-755
- [9] Chao CC, Chelius D, Zhang T, Mutumanje E, Ching WM.  
Insight into the virulence of Rickettsia prowazekii by proteomic analysis and comparison with an avirulent strain.  
Biochim Biophys Acta. 2007; 1774: 373-381
- [10] Anderson, L.; Hunter, C.L.  
Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins.  
Mol Cell Proteomics. 2006; 5: 573-588
- [11] Harrington, M.G.; Merril, C.R.; Asher, D.M.; Gajdusek, D.C.  
Abnormal proteins in the cerebrospinal fluid of patients with Creutzfeldt-Jakob disease.  
N Engl J Med. 1986; 315: 279-283
- [12] Burkhard, P.R.; Sanchez, J.C.; Landis, T.; Hochstrasser, D.F.  
CSF detection of the 14-3-3 protein in unselected patients with dementia.  
Neurology. 2001; 56: 1528-1533
- [13] Roessler M, Rollinger W, Mantovani-Endl L, Hagmann ML, Palme S, Berndt P, Engel AM, Pfeffer M, Karl J, Bodenmüller H, Rüschoff J, Henkel T, Rohr G, Rossol S, Rösch W, Langen H, Zolg W, Tacke M.  
Identification of PSME3 as a novel serum tumor marker for colorectal cancer by combining two-dimensional polyacrylamide gel electrophoresis with a strictly mass spectrometry-based approach for data analysis.

Mol Cell Proteomics. 2006 ; 5: 2092-2101

[14] Chevallet M, Diemer H, Van Dorssealer A, Villiers C, Rabilloud T.  
Toward a better analysis of secreted proteins: the example of the myeloid cells secretome.  
Proteomics. 2007; 7: 1757-1770

[15] Zhang H, Liu AY, Loriaux P, Wollscheid B, Zhou Y, Watts JD, Aebersold R.  
Mass spectrometric detection of tissue proteins in plasma.  
Mol Cell Proteomics. 2007; 6: 64-71

[16] Aicher L, Meier G, Norcross AJ, Jakubowski J, Varela MC, Cordier A, Steiner S.  
Decrease in kidney calbindin-D 28kDa as a possible mechanism mediating cyclosporine A- and FK-506-induced calciuria and tubular mineralization.  
Biochem Pharmacol. 1997; 53: 723-731

[17] Witzmann FA, Richardson MR.  
Two-dimensional gels for toxicological drug discovery applications.  
Expert Opin Drug Metab Toxicol. 2006; 2: 103-111

[18] Merrick BA, Bruno ME.  
Genomic and proteomic profiling for biomarkers and signature profiles of toxicity.  
Curr Opin Mol Ther. 2004; 6: 600-607

[19] Wang YX, Yuan YJ.  
Direct proteomic mapping of Streptomyces Luteogriseus Strain 103 and cnn1 and insights into antibiotic biosynthesis.  
J Proteome Res. 2005; 4: 1999-2006

[20] Fortis F, Guerrier L, Areces LB, Antonioli P, Hayes T, Carrick K, Hammond D, Boschetti E, Righetti PG.  
A new approach for the detection and identification of protein impurities using combinatorial solid phase ligand libraries.  
J Proteome Res. 2006 ; 5: 2577-2585