



HAL
open science

Recherche de biomarqueurs d'exposition et d'effet à des cancérogènes de l'environnement par spectrométrie de masse

Marianne Ibrahim

► **To cite this version:**

Marianne Ibrahim. Recherche de biomarqueurs d'exposition et d'effet à des cancérogènes de l'environnement par spectrométrie de masse. Sciences agricoles. Université de Strasbourg; Université Libanaise, 2013. Français. NNT : 2013STRAF038 . tel-01018695

HAL Id: tel-01018695

<https://theses.hal.science/tel-01018695>

Submitted on 4 Jul 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE STRASBOURG
ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES
(UMR 7140)
ET
UNIVERSITÉ LIBANAISE
ÉCOLE DOCTORALE DES SCIENCES ET DE TECHNOLOGIE

COTUTELLE DE THÈSE présentée par :

Marianne IBRAHIM

Soutenue le : 05 décembre 2013

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**
et de l'Université Libanaise

Discipline/ Spécialité : **Chimie biologique et thérapeutique**

<p>RECHERCHE DE BIOMARQUEURS D'EXPOSITION ET D'EFFET À DES CANCÉRIGÈNES DE L'ENVIRONNEMENT PAR SPECTROMÉTRIE DE MASSE</p>
--

THÈSE dirigée par :

Madame LEIZE-WAGNER Emmanuelle

Docteur, Université de Strasbourg

CO-DIRECTEURS :

Monsieur CHAHINE Ramez
Madame DAGHER Zeina

Professeur, Université Libanaise
Docteur, Université Libanaise

RAPPORTEURS :

Madame CHAMOT-ROOKE Julia
Madame FERRO Myriam

Docteur, Institut Pasteur
Docteur, CEA Grenoble

AUTRES MEMBRES DU JURY :

Monsieur BECHINGER Burkhard

Président du jury
Professeur, Université de Strasbourg

Madame GREIGE-GERGES Hélène
Madame HAMADÉ Eva

Professeur, Université Libanaise
Docteur, Université Libanaise

A ceux sans qui je ne suis rien :
*mon père Camille, ma mère Maguy
et mes deux frères Hicham et Elie*

« Qui vit content de rien possède toute chose »
N. Boileau

Remerciements

Ca y est, c'est la fin d'une belle aventure et le moment tant attendu de vous remercier. Je tiens à exprimer ici toute ma gratitude aux personnes qui ont contribué à l'aboutissement de ce travail.

J'adresse tout d'abord mes vifs remerciements au Pr. Burkhard Bechinger pour avoir accepté de présider ce jury de thèse. J'exprime également toute ma reconnaissance au Dr. Julia Chamot-Rooke et Dr. Myriam Ferro d'avoir accepté d'être les rapporteurs de ce travail, ainsi que Dr. Hélène Greige et Dr. Eva Hamadé pour leur présence dans ce jury de thèse.

Les travaux présentés dans ce manuscrit ont été réalisés au sein du Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (LSMIS) de l'UMR 7140 à Strasbourg et sous la direction du Dr. Emmanuelle Leize-Wagner.

Emmanuelle, je tiens à te remercier chaleureusement pour la confiance que tu m'as témoignée en m'accueillant en thèse dans ton équipe, pour tout ton soutien, ta disponibilité et ta gentillesse tout au long de mes trois années de thèse. Un manuscrit entier ne suffirait pas à t'exprimer toute ma gratitude. Je te remercie donc simplement d'être aussi formidable humainement et professionnellement et d'avoir fait en sorte que cette thèse se déroule dans les meilleures conditions possibles. Merci pour m'avoir fait découvrir la spectrométrie de masse et apprécier plusieurs congrès scientifiques. Je n'oublie pas également les soirées que tu m'as consacrées pour m'aider à analyser mes résultats et surtout la « nuit blanche » pour la soumission de notre article.

Cette thèse n'aurait pas pu être réalisée sans le soutien financier de l'Agence Universitaire de la Francophonie, le CNRS libanais et l'ANR.

Tous mes respects et ma profonde reconnaissance à mon co-directeur de thèse Pr. Ramez Chahine.

Je tiens à remercier ma co-directrice, Dr. Zeina Dagher, pour avoir proposé ce sujet et pour son encadrement dans les parties culture cellulaire et collection du sang des patients.

Les travaux de culture cellulaire ont été effectués au Laboratoire d'Immunologie à la Faculté de Santé Publique II à l'Université Libanaise. Merci à l'ensemble de l'équipe, merci au Dr. Souleima Chamat et Dr. Hasnaa Bou-Haroun Tayoun pour l'accueil que vous m'avez réservé au sein de votre Laboratoire, merci pour les conseils et les encouragements que vous m'avez octroyés chaque fois que cela fût nécessaire.

Les recherches et réflexions sont aussi le fruit d'une collaboration avec la « Plateforme Protéomique Strasbourg Esplanade » (PPSE) dirigée par Philippe Hammann. Merci Philippe pour ta précieuse aide technique surtout dans la partie 'préparation d'échantillons', merci pour tes idées géniales et pour m'avoir supportée durant tout ce temps! Je t'ai laissé un petit souvenir au labo (Mot-clé : perceuse), nous avons tellement rigolé ce jour-là ! De même l'histoire du fameux gel (cf : photos souvenirs)...

Mes remerciements s'adressent aussi à Lauriane Kuhn, deuxième membre de la PPSE. Merci Lauriane pour ton soutien de tous les instants, pour tous les moments de bonne humeur

que l'on a partagés, pour tous tes précieux conseils en protéomique et bioinformatique, pour ta « précision en lkp » et pour ta présence à côté de moi surtout pendant les derniers jours les plus durs de la rédaction. Merci pour toutes tes discussions qui m'ont augmenté la morale et « rechargé les batteries ». En travaillant avec toi, t'es devenue une amie. Et tu sais qu'un ami, rien qu'un ami, c'est aussi précieux qu'une vie ?!

A Hélène que je remercie du fond du cœur surtout pour mon premier jour à Strasbourg, je n'oublierai jamais le petit tour que nous avons fait ensemble à l'Esplanade. Merci pour ta gentillesse et ton encouragement au cours de ma rédaction.

Une pensée va aussi à Reynald avec qui j'ai découvert le MALDI-TOF. Tu m'as appris à faire les premiers pas dans la spectrométrie de masse, Merci !

Mes remerciements vont aux permanents du LSMIS : Yannis (Monsieur « j'aime les ... »), Armelle (pour ta gentillesse et tes mails encourageants), Noelle (pour ton idée géniale au début de ma thèse) et Nathalie (pour ton occupation des papiers administratifs).

Un merci spécial va à Mélanie (et sa bonne humeur), Mickael (plutôt Mickayyyel pour tes encouragements pendant la rédaction, pour les pauses 'chocolat' à minuit, ... Vive le flan au « p..... », Vive le Liban !), Rabah (le Monsieur du WE, « Au secours Rabah !! »), Cécile (Madame the best, Madame la Reine des ronds, je n'oublie jamais les tiktiktik), Nha-Thi (xin chèo et bye bye avec ta spéciale tonalité) et Margaux (Madame la spécialiste des tartes au chocolat malgré que je n'aime pas du tout le chocolat !). Merci à vous tous pour votre bonne humeur du bureau.

Merci à tous les stagiaires qui ont égayé la vie du labo durant leur passage. Une pensée va à Marion et Marc. Merci Maxime de m'avoir aidée à débloquer des trucs informatiques.

De même, je n'oublie pas Johana et Philippe, membres de la PPSE. Je vous remercie pour toutes vos discussions.

Je tiens à remercier les membres du laboratoire d'Immunologie : Dr. Nabil, Najibé, Catherine, Stéphanie, Leila, Hanadi et Nathalie pour les bénéfiques moments de détente. Merci Najibé pour la formation complète prodiguée en culture cellulaire. Stephy, je n'oublierai jamais nos manipulations dans la salle de culture cellulaire, chacune sous une hotte et ...

Je ne peux pas oublier mes amis les plus proches qui ont subi ces années par ma disponibilité réduite. Je pense à Joëlle, ma copine d'enfance que je connaisse depuis l'âge de 5 ans, la deuxième Joëlle que j'ai rencontrée pendant ma licence en biochimie, maintenant au Canada et mère de 3 enfants et qui sait, peut-être plus ... Merci à Myriam, t'es une amie précieuse. Merci à mes amies parisiennes : Stephy, Fabienne, Aline et Rasha pour les bons moments vécus ensemble pendant nos voyages. Merci à Malak, Mireille, Céline, Rémi et Omar.

Un grand merci à la troisième Joëlle: je t'ai rencontrée depuis deux ans, t'es devenue ma soeurette qui signifie tant pour moi. Je n'oublierai jamais tous les midis et les bons moments passés ensemble pendant notre séjour à Strasbourg. Je garde de toi un très beau souvenir.

Merci à Tartar, Mirna, Ghada et leurs familles.

Et à tous ceux et celles qui auraient échappé malencontreusement à mon attention dans ces lignes, Merci !

Et pour finir, une part conséquente de remerciements va à mes parents et mes deux frères pour leur soutien inébranlable tout au long de mon parcours, leurs encouragements inconditionnels, leur gentillesse intarissable et leur présence malgré la distance! Merci papa pour ta générosité. Merci maman pour m'avoir rendu visite à Strasbourg pendant ma troisième année de thèse, merci pour m'avoir supportée à tous niveaux et jusqu'au bout malgré les événements. Merci cousine.

Ce n'est qu'une page qui se tourne mais il en reste beaucoup à écrire...

Valorisation scientifique

Communications orales

1. **M. IBRAHIM**, L. KUHN, P. HAMMANN, Z. DAGHER, R. CHAHINE, E. LEIZE-WAGNER. « Characterization of exposure biomarkers to benzo(a)pyrene by quantitative mass spectrometry ». *19th LAAS International Science Conference*, Beyrouth (Liban), 5-6 Avril 2013.
2. **M. IBRAHIM**, L. KUHN, P. HAMMANN, Z. DAGHER, R. CHAHINE, E. LEIZE-WAGNER. « A la recherche de biomarqueurs d'exposition au benzo(a)pyrène par approche Label-Free : Evaluation de 3 suites logicielles pour la quantification MS1 ». *9^{èmes} Journées Club-Jeunes de la SFEAP*, Strasbourg, 29-31 Mai 2013.
3. **M. IBRAHIM**, L. KUHN, P. HAMMANN, Z. DAGHER, R. CHAHINE, E. LEIZE-WAGNER. « Evaluation of 3 softwares for MS1 quantification ». *Workshop "Systèmes complexes : parallélisme, génomique et protéomique clinique"*, Strasbourg, 2 Mai 2013.
4. **M. IBRAHIM**, L. KUHN, P. HAMMANN, Z. DAGHER, R. CHAHINE, E. LEIZE-WAGNER. « Caractérisation de biomarqueurs d'exposition au benzo(a)pyrène par spectrométrie de masse quantitative ». *Journées UMR 7140 "Chimie de la matière complexe"*, Strasbourg, 20 Février 2013.
5. **M. IBRAHIM**, L. KUHN, P. HAMMANN, Z. DAGHER, R. CHAHINE, E. LEIZE-WAGNER. « Développement de la spectrométrie de masse pour la recherche de biomarqueurs d'exposition et d'effet à des cancérigènes ». *Journée des doctorants en Chimie (JDC 2011)*, Collège Doctoral Européen et Faculté de Chimie, Campus de l'Esplanade, Strasbourg, 10 Novembre 2011.

Communications écrites

1. **M. IBRAHIM**, L. KUHN, P. HAMMANN, Z. DAGHER, R. CHAHINE, E. LEIZE-WAGNER. « Time-course secretome analysis using label-free MS1 quantitative proteomic method to reveal differentially expressed proteins in hepatic human cells exposed to benzo(a)pyrene ». *Congrès International EuPA*, Saint Malo, 14-17 Octobre 2013.
2. **M. IBRAHIM**, L. KUHN, P. HAMMANN, Z. DAGHER, R. CHAHINE, E. LEIZE-WAGNER. « Caractérisation des biomarqueurs d'exposition à des cancérigènes par analyse protéomique ». *29^{èmes} Journées Françaises de Spectrométrie de Masse (JFSM)*, Orléans, 17-20 Septembre 2012.
3. **M. IBRAHIM**, L. KUHN, P. HAMMANN, Z. DAGHER, R. CHAHINE, E. LEIZE-WAGNER. « Analyse quantitative 'label-free' des biomarqueurs d'effet induits par le benzo(a)pyrène ». *Troisième Forum Doctoral, EDST*, Beyrouth (Liban), 25 Juin 2013.

4. **M. IBRAHIM**, L. KUHN, P. HAMMANN, Z. DAGHER, R. CHAHINE, E. LEIZE-WAGNER. « Caractérisation des biomarqueurs d'exposition à des cancérigènes par analyse protéomique ». Poster vulgarisé - *Doctoriales d'Alsace*, Mittelwihr, 24-29 Juin 2012 (Poster 1^{er} prix).

5. **M. IBRAHIM**, L. KUHN, P. HAMMANN, Z. DAGHER, R. CHAHINE, E. LEIZE-WAGNER. « Caractérisation des biomarqueurs d'exposition à des cancérigènes par analyse protéomique ». *Deuxième Forum Doctoral : Ecole Doctorale des Sciences et de Technologie – EDST* Liban, 19 Juin 2012.

6. **M. IBRAHIM**, L. KUHN, P. HAMMANN, Z. DAGHER, R. CHAHINE, E. LEIZE-WAGNER. « Étude par analyse protéomique de biomarqueurs d'exposition à des cancérigènes ». *19^{èmes} Journées Jeunes Chercheurs de la Société de chimie Thérapeutique(SCT)*, Romainville, 02-03 Février 2012.

7. **M. IBRAHIM**, L. KUHN, P. HAMMANN, Z. DAGHER, R. CHAHINE, E. LEIZE-WAGNER. « Étude par analyse protéomique de biomarqueurs d'exposition à des cancérigènes ». *Congrès de spectrométrie de Masse Appliquée à la Protéomique (SMAP)*, Avignon, 19-22 Septembre 2011.

Article soumis dans Molecular and Cellular Proteomics (MCP), 30 Octobre 2013 :

« Secretome analysis using Label-Free quantitative proteomics to discover potential cancer biomarkers of Benzo(a)Pyrene exposure ». **M. IBRAHIM, L. KUHN, Z. DAGHER, J. CHICHER, R. CHAHINE, P. HAMMANN, E. LEIZE-WAGNER.**

Résumé

Le Benzo(a)pyrène (BaP), micropolluant ubiquitaire de l'environnement, appartient à la famille des hydrocarbures aromatiques polycycliques (HAP) : il est cancérigène pour l'homme. Dans cette étude, nous avons développé une approche protéomique quantitative nanoLC-MS/MS label-free pour identifier des biomarqueurs liés à l'exposition au BaP dans le sécrétome des cellules hépatiques humaines exposées au BaP *vs.* des cellules non exposées et exposées au Benzo(e)pyrène (BeP). Les cellules traitées au BeP ont été choisies comme contrôle négatif afin de distinguer les protéines spécifiques du BaP de celles spécifiques des HAP : le BeP est un isomère du BaP mais il n'est pas classifié comme étant cancérigène pour l'homme. Les cellules ont été exposées à 10µM de BaP ou BeP et incubées pendant 2h et 72h afin de différencier les protéines précoces de celles tardives, respectivement. 847 protéines, dont 51,95% sont prédites sécrétées, ont été identifiées et quantifiées en utilisant le logiciel Progenesis LC-MS. 55 protéines ont été fortement surexprimées avec un ratio supérieur à 5. D'après la littérature et la fonction moléculaire de GO, la plupart de ces protéines surexprimées sont précoces et liées au cancer. Une validation ultérieure de l'expression de ces protéines dans le plasma de la population exposée au BaP aidera dans le développement de biomarqueurs qui permettront d'améliorer considérablement la détection précoce, le pronostic et prévention.

Mots-clés : Benzo(a)pyrène, Benzo(e)pyrène, biomarqueurs, cancer, sécrétome, protéomique, quantification label-free, spectrométrie de masse

Summary

Benzo(a)pyrene (BaP) belongs to a class of polycyclic aromatic hydrocarbon and is a well-recognized environmental micropollutant, previously reported as a potent human carcinogen. In the present study, we performed a nanoLC-MS/MS-based label-free quantitative proteomics approach to identify potential biomarkers of exposure in the secretome of BaP-treated *vs.* non-treated and Benzo(e)pyrene (BeP)-treated human hepatoma cell line HepG2. BeP-treated cells were chosen as a negative control to distinguish the BaP-specific from the HAP-specific regulated proteins: BeP is a BaP isomer but unlike BaP, it is not classifiable as to its carcinogenicity to humans. Cells were exposed to 10 μ M BaP or BeP and incubated for 2h and 72h to better differentiate between proteins having an early onset and a delayed onset, respectively. 847 proteins, with 51.95% of them predicted as secreted, have been identified and quantified using Progenesis LC-MS software. 55 proteins were seen as being highly up-regulated with a fold change of at least 5. Most of these up-regulated proteins were focused in cancer-related activities as revealed by literature and by a Gene Ontology (GO) molecular function classification. The potential use of the regulated proteins found in this proteomic study as early indicators of BaP exposure is discussed: further validation of expression of these proteins in the plasma of BaP-exposed population will assist in the development of biomarkers that will greatly improve early detection, prognosis, prediction of treatment response and prevention.

Keywords: Benzo(a)pyrene, Benzo(e)pyrene, biomarkers, cancer, secretome, proteomics, label-free quantification, mass spectrometry

Liste des abréviations

ACN : Acétonitrile
ADN : Acide désoxyribonucléique
AhR : Aryl hydrocarbon Receptor
AKR : Aldo-Kéto Réductases
BaP : Benzo(a)pyrène
BeP : Benzo(e)pyrène
BPDE : Benzo(a)Pyrène-7,8-dihydrodiol-9,10-époxyde
BSA : Bovine Serum Albumin
CIRC : Centre International de Recherche sur le Cancer
CYP450 : Cytochromes P450
DTT : Dithiothreitol
EFSA : European Food Safety Authority
EH : Epoxyde Hydrolase
ESI : Electrospray ionization
eV : Electron-volt
GO : Gene Ontology
HAP : Hydrocarbures Aromatiques Polycycliques
IA : Iodoacétamide
IARC : International Agency for Research on Cancer
nanoLC : nanoLiquid Chromatography
MALDI : Matrix-Assisted Laser Desorption/Ionisation
MPTs : modifications post-traductionnelles
m/z : masse sur charge
ng : nanogramme
nm : nanomètre
PM : Poids Moléculaire
rpm : rotation per minute
SVF : Sérum du veau fœtal
TFA : Trifluoroacetic acid
TIC : Total Ion Chromatogram
TOF : Time Of Flight
UGT : UDP-Glucuronosyltransférases
USEPA : United States Environmental Protection Agency
UTC : Urée, Thiourée, Chaps
UV : Ultraviolet
XIC : eXtracted ion chromatogram
µg : microgramme

Liste des tableaux

- Tableau 1 : Les principaux HAP classés selon différents critères
- Tableau 2 : Matrices utilisées en MALDI
- Tableau 3 : Principales performances et caractéristiques des analyseurs
- Tableau 4 : Résultats du dosage des extraits protéiques (condition BaP 24h et 72h)
- Tableau 5 : Résultats d'analyses par 2-DE et MALDI-TOF(/TOF) des spots de protéines totales détectées, sous ou surexprimées après traitement au BaP pendant 24 et 72h
- Tableau 6 : Résultats d'analyses par 2-DE et NanoLC-MS/MS des spots de protéines totales détectées, sous ou surexprimées après traitement au BaP pendant 24 et 72h
- Tableau 7 : Résultats du dosage Bradford des extraits protéiques (conditions Control vs. BeP vs. BaP pendant 6h, 24h et 72h)
- Tableau 8 : Protéines différentiellement exprimées
- Tableau 9 : Résultats d'analyses des spots de protéines totales détectées, sous ou surexprimées après traitement au BaP/BeP pendant 24 et 72h par nanoLC-MS/MS
- Tableau 10 : Résultats du dosage Bradford des extraits protéiques des sécrétomes des cellules exposées à 10 μ M de BaP pendant 6h et 72h.
- Tableau 11 : Résultats du dosage Bradford des extraits protéiques des sécrétomes
- Tableau 12 : Valeurs des XICs du peptide ELDESLQVAER de la protéine CLUS_Human pour chaque condition d'exposition
- Tableau 13 : Caractéristiques des 3 suites logicielles testées sur les mêmes données
- Tableau 14 : Densité des éléments du sang
- Tableau 15 : Programmation de l'appareil IEF
- Tableau 16 : Paramètres d'acquisition des spectres de masse avec le logiciel FlexControl v3.0

Table des figures

- Figure 1 : Quantité moyenne de Benzo(a)pyrène absorbé chaque jour à partir de différentes sources
- Figure 2 : Structure chimique du benzène
- Figure 3 : Présentation des 16 HAP classés prioritaires par l'USEPA
- Figure 4 : Caractéristique structurale de deux HAP ayant un potentiel cancérigène
- Figure 5 : Métabolisme simplifié du BaP
- Figure 6 : Les 3 voies principales de métabolisation du BaP
- Figure 7 : Schéma simplifié du mécanisme de l'induction des gènes cyp par le BaP
- Figure 8 : Conjugaison du glutathion au BPDE
- Figure 9 : Principe de l'immunodéplétion
- Figure 10 : Principe de la technologie « Equalizer »
- Figure 11 : Les 3 grandes voies de sécrétion des protéines
- Figure 12 : Nombre d'articles publiés en protéomique de 1995 jusqu'au juillet 2013
- Figure 13 : Les « Omiques »
- Figure 14 : Différentes étapes de l'approche protéomique
- Figure 15 : Schéma du principe de fonctionnement d'un spectromètre de masse
- Figure 16 : Effet de la résolution sur le spectre de masse d'un peptide de 12 acides aminés
- Figure 17 : Photographie d'une cible MALDI (Bruker)
- Figure 18 : Principe de la désorption/ionisation par MALDI
- Figure 19 : Illustration schématique de la production de gouttelettes chargées
- Figure 20 : Modèle de la charge résiduelle
- Figure 21 : Modèle d'évaporation des ions
- Figure 22 : Résolution d'un pic en spectrométrie de masse
- Figure 23 : Trajet d'un ion en mode linéaire dans l'analyseur TOF
- Figure 24 : Représentation schématique d'un couplage MALDI-TOF en mode réflecteur comportant un système d'extraction retardée
- Figure 25 : Principe d'un quadripôle à section cylindrique
- Figure 26 : Trajectoire oscillante stable des ions détectés
- Figure 27 : Appareils hybrides
- Figure 28 : Représentation d'un triple quadripôle
- Figure 29 : Spectrométrie de masse en mode tandem
- Figure 30 : Structure chimique d'un dipeptide
- Figure 31 : Fragments des peptides d'après Biemann
- Figure 32 : Structure chimique d'un ion immonium
- Figure 33 : Nomenclature des ions fragments des peptides
- Figure 34 : Vue d'ensemble du couplage nanoLC-MS/MS
- Figure 35 : Principe d'identification des protéines par interrogation de banques protéiques
- Figure 36 : Stratégies utilisées pour l'analyse protéomique quantitative
- Figure 37 : Approche protéomique des extraits totaux des cellules
- Figure 38 : Cytotoxicité du BaP à des concentrations allant jusqu'à 50µM dans les cellules HepG2 exposées pendant 24h et 72h
- Figure 39 : Duplicats de gels 2D des extraits protéiques totaux des cellules HepG2 exposées à 10µM de BaP pendant 24h, lavées et remises en culture avec sérum pendant 24h (A) ou 72h (B).
- Figure 40 : Focalisation avec grandissement sur une partie des gels permettant de bien visualiser les spots d'intérêts marqués en vert

- Figure 41 : Triplicats de gels 2D des extraits protéiques totaux des cellules HepG2 exposées à 10 μ M de BaP (A-B) ou de BeP (C-D) pendant 24h, lavées et remises en culture dans un milieu sans sérum pendant 24h (A-C) ou 72h (B-D).
- Figure 42 : Quantification des protéines par PD-Quest
- Figure 43 : Classification des 20 protéines différemment exprimées
- Figure 44 : Fonctions moléculaires des 20 protéines différemment exprimées
- Figure 45 : Résultats des gels 2D montrant une diminution de la quantité d'albumine après utilisation d'un supplément B27 au lieu du sérum SVF : A- avec SVF. B- avec B27
- Figure 46 : Gels 2D des extraits protéiques du sécrétome des cellules exposées à 10 μ M de BaP pendant 6h (A) et 72h(B)
- Figure 47 : Gels 2D des extraits protéiques après concentration (50 kDa) du sécrétome des cellules exposées à 10 μ M de BaP pendant 6h (A) et 72h (B)
- Figure 48 : Gels 1D des extraits protéiques des sécrétomes des cellules HepG2 exposées à 10 μ M de BaP pendant 6h et 72h, (A) sans filtre de concentration, (B) avec filtre de concentration 50 kDa
- Figure 49 : Gels 1D des extraits protéiques des sécrétomes des cellules HepG2 exposées à 10 μ M de BaP pendant 6h (A) et 72h (B) après enrichissement des protéines faiblement abondantes
- Figure 50 : Zones de découpage de la piste de l'éluat protéique du sécrétome des cellules HepG2 exposées à 10 μ M de BaP pendant 72h
- Figure 51 : Gels 1D de l'extrait protéique du sécrétome des cellules HepG2 exposées à 10 μ M de BaP 72h
- Figure 52 : Méthode Paragon de ProteinPilot
- Figure 53 : Ajout de critères spécifiques et fichiers « .wiff » chargés dans PeakView
- Figure 54 : XIC d'un peptide
- Figure 55 : Fichiers générés par MarkerView (1.2.1)
- Figure 56 : Représentation d'une PCA sur MarkerView
- Figure 57 : PCA individuelles des 4 temps d'incubation (2h, 6h, 24h et 72h) réalisées sur MarkerView. (A) 2h, (B) 6h, (C) 24h et (D) 72h
- Figure 58 : PCA globale des 4 temps d'incubation (2h, 6h, 24h et 72h)
- Figure 59 : Diagramme de Venn de toutes les protéines identifiées pendant 2h, 6h, 24h et 72h.
- Figure 60 : Exemple de profil de régulation (appelé plot profile)
- Figure 61 : Plot profile du peptide ELDESLQVAER de la protéine CLUS_Human
- Figure 62 : XICs manuels et automatiques du peptide ELDESLQVAER de la protéine CLUS_Human générés par PeakView
- Figure 63 : Création d'une librairie de spectres MS/MS à partir d'une recherche en banque de données par l'algorithme Paragon
- Figure 64 : Paramétrage des peptides avec le détail relatif aux 5 onglets composant cette table (Digestion, Prediction, Filter, Library, Modifications)
- Figure 65 : Paramétrage des transitions avec le détail relatif aux 5 onglets composant cette table (Prediction, Filter, Library, Instrument, Full-Scan)
- Figure 66 : Arborescence générée par l'import du fichier .fasta dans Skyline
- Figure 67 : Reconstitution d'un pic d'éluat peptidique (XIC) à partir de la masse d'un précurseur donné ainsi que la masse des 2 isotopes suivants
- Figure 68 : Reconstitution des 9 XICs d'un peptide de la protéine AGRIN_HUMAN
- Figure 69 : Reconstitution des 9 XICs d'un peptide de la protéine CH60_HUMAN
- Figure 70 : Outil « Isotopic dot plot »
- Figure 71 : Outil « Retention time alignment »

- Figure 72 : Recherche de protéines et de peptides montrant un profil de régulation (sur- ou sous-) à l'aide de l'outil « Peak Areas Comparison »
- Figure 73 : Durée d'alignement des 36 « runs »
- Figure 74 : Réalignement des Cartes 2D « massiques »
- Figure 75 : Représentation schématique du XIC : (A) 1D, (B) 2D, (C) 3D
- Figure 76 : Illustration des protéines regroupées sous forme d'un dendrogramme
- Figure 77 : Illustration des protéines regroupées sous forme d'une PCA
- Figure 78 : PCA individuelles des 4 temps d'incubation (2h, 6h, 24h et 72h) réalisées sur Progenesis. (A) 2h, (B) 6h, (C) 24h et (D) 72h
- Figure 79 : PCA globale des 4 temps d'incubation (2h, 6h, 24h et 72h) en prenant la moyenne des abondances protéiques de chaque triplicat
- Figure 80 : Comparaison de l'intensité maximale des XICs générés de manière manuelle ou automatique par PeakView sur le précurseur avec $m/z=644,820$
- Figure 81 : Exemple d'une mauvaise intégration d'un XIC par Skyline.
- Figure 82 : Correction manuelle de l'intégration automatique des XIC pour un peptide donné de la protéine CH60_HUMAN
- Figure 83 : Modification manuelle du choix du « peak group »
- Figure 84 : Progenesis vs. MarkerView et Skyline
- Figure 85 : Spectres des matrices seules HCCA et DHB
- Figure 86 : Comparaison des spectres du plasma d'un fumeur mélangé avec l'une des deux matrices suivantes : HCCA ou SA
- Figure 87 : Comparaison des spectres du plasma de deux fumeurs à ceux de la matrice seule et du plasma d'un non-fumeur
- Figure 88 : Pics apparus seulement chez les fumeurs
- Figure 89 : Schéma du montage de la pré-colonne C18 sur les vannes de la nanoLC
- Figure 90 : Montage pour chromatographie en phase C18 : Gradients en solvants au cours du temps
- Figure 91 : Système nano cHiP HPLC Eksigent
- Figure 92 : Configuration Trap elute du système cHiPLC
- Figure 93 : Schéma du gradient chromatographique utilisé
- Figure 94 : (A) Illustration détaillée de l'instrument Triple TOF 5600 (AB SCIEX)
(B) Image de la machine
- Figure 95 : Représentation schématique d'un cycle MS et MS/MS (Top 20) sur le triple TOF 5600
- Figure 96 : Approches protéomiques développées

Sommaire

Introduction générale et présentation de la thèse	7
Partie I - Introduction bibliographique	12
I) Les hydrocarbures aromatiques polycycliques	12
1) Définition.....	12
1.1) Sources d'exposition aux HAP.....	12
1.2) Des molécules à plusieurs cycles : Classification et Règlementation	13
1.3) Effets des HAP sur la santé	16
2) Devenir des HAP dans l'organisme	16
2.1) Absorption	16
2.1.1) Absorption respiratoire.....	16
2.1.2) Absorption digestive.....	17
2.1.3) Absorption cutanée.....	17
2.2) Distribution.....	17
2.3) Métabolisation	17
2.3.1) Voie de mono-oxygénation par les CYP450 : voie des diol-époxydes	19
2.3.2) Voie d'oxydation mono-électrique par les peroxydases : voie du radical cation ..	21
2.3.3) Voie des o-quinones	22
2.3.4) Autre voie : voie mineure	22
2.4) Elimination : réactions de conjugaison.....	22
2.4.1) Conjugaison par les UDP-Glucuronyltransférases	22
2.4.2) Conjugaison par les glutathion-S-transférases	23
2.4.3) Conjugaison par les Sulfotransférases.....	23
3) Toxicoprotéomique : Analyse protéomique de la toxicité induite par le BaP.....	23
3.1) Effets carcinogènes.....	24
3.1.1) Carcinogénèse pulmonaire	24
3.1.2) Carcinogénèse cervicale	25
3.1.3) Cancérogenèse de la vessie	25
3.1.4) Cancérogenèse du sein	26
3.1.5) Cancérogenèse de la prostate.....	26
3.2) Autres effets toxiques	27
3.2.1) Hépatotoxicité	27
3.2.2) Immunotoxicité	27
3.2.3) Neurotoxicité.....	28
II) Les biomarqueurs : vers la médecine personnalisée	28

1)	Notion de biomarqueurs	28
2)	Les défis liés à la découverte de biomarqueurs protéiques dans le plasma	31
3)	Découverte de biomarqueurs à partir des sécrétomes des modèles cellulaires	32
4)	De la découverte à la validation d'un biomarqueur en passant par sa vérification	35
III)	Analyse protéomique, réservoir de biomarqueurs	35
3.1)	Introduction et stratégie générale	35
3.2)	Outils pour l'étude protéomique.....	38
3.2.1)	Spectrométrie de masse	38
3.2.1.1)	Mesure de masse	40
3.2.1.2)	Sources d'ionisation.....	40
3.2.1.2.1)	Désorption-ionisation laser assistée par matrice (MALDI)	40
3.2.1.2.2)	Source par ionisation Electrospray.....	43
3.2.1.3)	Analyseurs de masse	47
3.2.1.3.1)	Analyseur à temps de vol (TOF).....	49
3.2.1.3.2)	Le quadripôle (Q).....	50
3.2.1.3.3)	Les analyseurs hybrides	51
3.2.1.3.3.1)	TOF-TOF	51
3.2.1.3.3.2)	Q-TOF	52
3.2.1.3.3.3)	QqQ.....	52
3.2.2)	Spectrométrie de masse en mode tandem.....	53
3.2.2.1)	Principe	53
3.2.2.2)	Fragmentation peptidique.....	54
3.2.2.2.1)	Fragmentation de basse énergie	54
3.2.2.2.2)	Fragmentation de haute énergie	56
3.3)	Fractionnement du protéome.....	56
3.3.1)	Séparation des protéines : Electrophorèse.....	57
3.3.2)	Séparation des peptides : techniques chromatographiques.....	58
3.4)	Analyse bioinformatique des données MS/MS	59
3.5)	Protéomique quantitative.....	62
3.5.1)	Quantification par électrophorèse bidimensionnelle	63
3.5.2)	Quantification par spectrométrie de masse.....	63
3.5.2.1)	Techniques avec marquage	64
3.5.2.2)	Techniques sans marquage.....	64
IV)	Objectifs de ce travail de thèse	66
	Références bibliographiques	67
	Partie II - Résultats : De la préparation d'échantillons à l'analyse protéomique quantitative....	76

Chapitre 1 : Analyse différentielle du protéome total et du sécrétome des cellules HepG2 par électrophorèse bidimensionnelle couplée à la spectrométrie de masse	77
A) Analyse différentielle du protéome total	77
1) Cellules cultivées dans le milieu RPMI complété par du sérum	77
1.1) Culture cellulaire	78
1.2) Extraction des protéines à partir des culots cellulaires et dosage	79
1.3) Electrophorèse bidimensionnelle.....	79
1.4) Identification par spectrométrie de masse	81
1.5) Discussion biologique	86
2) Cellules cultivées dans le milieu RPMI complété par le supplément B27	88
2.1) Culture cellulaire	88
2.2) Dosage des protéines extraites à partir des culots cellulaires	88
2.3) Electrophorèse bidimensionnelle.....	89
2.4) Identification par spectrométrie de masse	92
2.5) Discussion biologique	95
B) Analyse différentielle du sécrétome	96
1) Séparation sur gels 2D.....	96
2) Séparation sur gels 1D.....	98
3) Optimisation de l'étape de culture cellulaire.....	100
Références bibliographiques	104
Chapitre 2 : “Secretome analysis using Label-Free quantitative proteomics to discover potential cancer biomarkers of Benzo(a)Pyrene exposure” Article soumis dans MCP	107
Chapitre 3 : Analyse différentielle par spectrométrie de masse quantitative (label-free)	132
1) Choix des logiciels	133
2) Quantification relative à l'aide de la suite logicielle PeakView/ MarkerView (AB Sciex)	133
2.1) Pre-processing	133
2.2) Calcul des valeurs d'aires des pics peptidiques.....	134
2.3) Analyse statistique.....	136
2.4) Résultats et discussion.....	137
3) Quantification relative à l'aide du logiciel Skyline (Université de Washington, Laboratoire MacCoss).....	142
3.1) Prétraitement des données	142
3.1.1) Création d'une librairie spectrale MS/MS.....	143
3.1.2) Paramétrage des données relatives aux peptides et aux transitions.....	144
3.1.3) Import des séquences protéiques	145
3.1.4) Import des fichiers bruts issus de l'acquisition sur le spectromètre de masse.....	146

3.2)	Calcul des valeurs d'aires peptidiques	146
3.2.1)	Outils complémentaires à l'analyse des données	147
3.2.1.1)	Outil « Peak Areas »	147
3.2.1.2)	Outil « isotopic dot plot ».....	149
3.1.2.3)	Outil « Retention time alignment »	149
3.3)	Analyse statistique.....	150
4)	Quantification relative à l'aide de Progenesis LC-MS (Nonlinear Dynamics).....	151
4.1)	Pre-processing des données.....	151
4.2)	Calcul des valeurs d'aires des pics peptidiques.....	153
4.3)	Analyse statistique.....	153
4.4)	Résultats et discussion.....	154
5)	Retour d'expérience sur les logiciels pour la quantification relative des protéines par approche MS1 label-free	156
5.1)	ProteinPilot/PeakView/MarkerView (AB Sciex).....	156
5.2)	Skyline (Laboratoire MacCoss, Université de Washington).....	159
5.3)	Progenesis (Nonlinear Dynamics).....	161
	Références bibliographiques	164
	Chapitre 4 : La spectrométrie de masse, un outil pour la recherche d'adduits BPDE à l'ADN et aux protéines	166
1)	Collection du plasma des patients	166
2)	Recherche de BPDE, sous sa forme libre ou liée, dans le plasma des fumeurs	168
3)	Perspectives	171
	Références bibliographiques	172
	Conclusion générale et perspectives.....	174
	Références bibliographiques	176
	Annexe A - Mise en œuvre expérimentale	178
1)	Culture cellulaire	178
2)	Extraction au trizol	180
3)	Electrophorèse	180
3.1)	Electrophorèse monodimensionnelle sur gel de polyacrylamide	181
3.2)	Electrophorèse bidimensionnelle.....	181
4)	Préparation des échantillons pour la spectrométrie de masse.....	183
4.1)	Excision des spots ou bandes	183
4.2)	Lavages et réduction-alkylation des cystéines.....	183
4.3)	Digestion trypsique.....	183
4.4)	Extraction des peptides.....	184

4.5) Conditionnement MALDI ou nanoLC	184
5) Acquisition des spectres de masse MALDI.....	184
6) Analyses nanoLC-ESI-MS/MS	186
6.1) MicroToF-QII.....	186
6.2) Triple TOF 5600.....	187
6.2.1) Description technique du couplage nanoLC-ESI-MS/MS utilisé: système nano cHiP HPLC (Eksigent)	187
6.2.2) Description technique du couplage nanoLC-ESI-MS/MS utilisé: système Triple TOF 5600 (AB Sciex)	189
6.2.3) Séquence d'acquisition MS et MSMS combinant identification et quantification : méthode dite « TOP 20 IDA ».....	191
Annexe B - Supplemental data 3 - Article	193

Introduction générale

et

présentation de la thèse

Introduction générale et présentation de la thèse

Depuis plusieurs années, les Hydrocarbures Aromatiques Polycycliques (HAP) sont très étudiés à cause de leur forte toxicité et de leur présence dans l'environnement (air, eau et alimentation). Ils proviennent, en majorité, de la combustion incomplète du charbon et des produits pétroliers. Dès 1976, ils sont ajoutés à la liste des polluants prioritaires par l'US EPA (United States Environmental Protection Agency). Ces composés appartiennent à une famille contenant de molécules d'atomes de carbone et d'hydrogène et dont la structure comprend, au moins, deux cycles aromatiques. Quelle que soit la voie d'exposition, la population est exposée à un mélange d'HAP. Une fois associés aux poussières, ils sont capables de pénétrer dans les alvéoles pulmonaires et d'exercer des effets nocifs sur la santé. Parmi ces HAP, le Benzo(a)pyrène (BaP), à cinq cycles aromatiques, est le composé le plus étudié. Il est classé groupe 1 selon l'« International Agency for Research on Cancer » (IARC) c'est-à-dire cancérigène avéré pour l'homme à cause de sa métabolisation en diol époxyde (BPDE : Benzo(a)Pyrène-7,8-dihydrodiol-9,10-époxyde). Le BPDE est capable de se lier à l'acide désoxyribonucléique (ADN) des cellules pour former des adduits stables BPDE-ADN et entraîner des mutations pouvant aboutir à un cancer : c'est la voie d'activation du BaP la plus importante.

Dans ce contexte, la recherche de biomarqueurs liés à une exposition à un composé particulier est devenue un des points centraux de l'analyse protéomique. Elle permet de comprendre les causes d'apparition d'une maladie, voire du cancer et d'améliorer sa prévention. En effet, les tumeurs libèrent plusieurs facteurs impliqués dans leur développement. Ceux-ci sont généralement retrouvés dans les fluides biologiques (sang, urine). Leur découverte ou leur quantification dans ces milieux constituent de nouveaux enjeux pour les développements pharmaceutiques et le suivi des pathologies. Le sang possède un contenu très riche et son prélèvement est un acte peu invasif pour le patient. Cela en fait une source de biomarqueurs idéale. Leur détection directe (protéines et métabolites) se heurte actuellement à la gamme dynamique du fluide sanguin. Par conséquent, le passage préalable par un modèle cellulaire constitue une approche pertinente afin d'établir une liste de marqueurs potentiels à rechercher et valider dans un deuxième temps dans des cohortes de plasma de patients.

Ces dernières années, un nouveau champ d'investigation a explosé : c'est l'analyse protéomique du sécrétome. Le terme « sécrétome » correspondant au surnageant de culture cellulaire, a été proposé par Tjalsma *et al.* (2000). C'est un modèle séduisant car les protéines libérées dans le milieu extracellulaire sont souvent retrouvées dans le plasma. Il qualifie la composition des protéines sécrétées dans un environnement particulier à un instant donné. L'approche protéomique du surnageant cellulaire intéresse surtout l'oncologie. Le sécrétome des cellules cancéreuses est une source prometteuse pour la découverte de biomarqueurs : il a été défini comme le « réservoir de biomarqueurs » du cancer (Xue, Lu et al. 2008). Pour la recherche d'un biomarqueur, plusieurs critères doivent être respectés :

- ❖ Présenter une sensibilité, une spécificité et une reproductibilité satisfaisante
- ❖ Permettre la détection des premiers stades de la pathologie
- ❖ Concorder avec la progression de la maladie
- ❖ Etre capable de suivre les effets de traitements thérapeutiques

Dans le cadre de ce travail, nous nous sommes intéressés à l'étude du sécrétome des cellules hépatique humaines HepG2 exposées au BaP et à son isomère toxique (Benzo(e)pyrène (BeP) utilisé comme contrôle afin de rechercher des biomarqueurs spécifiques au BaP et non aux HAP. Le choix du modèle cellulaire s'est porté sur les cellules hépatiques humaines HepG2, modèle représentatif pour l'évaluation de la génotoxicité et de la mutagénicité des génotoxiques. Nous avons donc privilégié l'utilisation de cette lignée puisqu'elle possède de nombreuses enzymes impliquées dans le métabolisme des xénobiotiques dans le foie humain.

Deux axes d'étude ont donc été définis au début de ma thèse :

- 1) La recherche, *in vitro*, de nouveaux biomarqueurs protéiques suite à l'exposition au benzo(a)pyrène et, dans un deuxième temps, leur validation dans le plasma des patients exposés aux HAP, notamment les fumeurs, les agents de polices exposés professionnellement aux HAP et les patients ayant des troubles respiratoires,
- 2) La recherche, *in vivo*, des adduits du BPDE à l'ADN et aux protéines dans les plasmas de fumeurs.

Ce travail a été réalisé en cotutelle entre le Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (LSMIS, Université de Strasbourg) sous la direction du Dr. Emmanuelle LEIZE-WAGNER et le Laboratoire d'Antioxydants et Stress Oxydatif (Université Libanaise) sous la direction du Dr. Ramez CHAHINE et l'encadrement du Dr. Zeina DAGHER. Les manipulations de culture cellulaire ont été réalisées au Liban au Laboratoire d'Immunologie – Faculté de Santé Publique II, Université Libanaise – dirigé par Dr. Souleima CHAMAT et Dr. Hasnaa BOU HAROUN-TAYOUN alors que les développements protéomiques ont été effectués à Strasbourg au LSMIS et en collaboration avec Philippe HAMMANN, responsable de la Plateforme Protéomique Strasbourg Esplanade (PPSE) à l'Institut de Biologie Moléculaire et Cellulaire.

Ce manuscrit est divisé en deux parties :

La première partie concerne la partie bibliographique de ce manuscrit; elle a pour but de:

- décrire les HAP et leurs effets toxiques sur l'individu,
- présenter les différentes techniques de spectrométrie de masse utilisées,
- résumer les différentes approches de l'analyse protéomique et leur utilisation pour la recherche de biomarqueurs,
- et présenter les objectifs de ma thèse.

La deuxième partie du manuscrit concerne toutes les optimisations et les résultats obtenus au cours de ce travail de thèse, depuis la préparation des échantillons jusqu'à l'analyse protéomique quantitative et l'identification des biomarqueurs liés à l'exposition au BaP. Cette partie est divisée en plusieurs chapitres :

- l'analyse différentielle du protéome total et du sécrétome des cellules hépatiques humaines HepG2 par électrophorèse bidimensionnelle couplée à la spectrométrie de masse,
- l'analyse différentielle du sécrétome par spectrométrie de masse quantitative label-free, ayant fait l'objet d'une publication soumise dans 'Molecular and Cellular Proteomics' : « Secretome analysis using Label-Free quantitative proteomics to discover potential cancer biomarkers of Benzo(a)Pyrene exposure »,
- le développement d'une approche de détection du BPDE dans le plasma, sous sa forme libre ou liée, par spectrométrie de masse MALDI-TOF.

Ces chapitres résumant :

- l'étude cytotoxique avec l'optimisation de la culture des cellules HepG2 et le choix des temps d'incubation et de la concentration des agents toxiques,
- l'évaluation du mode d'obtention des protéines d'intérêts, soit par extraction des protéines des culots cellulaires, soit par leur précipitation dans le sécrétome,
- le développement d'une stratégie d'analyse protéomique adaptée aux extraits protéiques, soit par une séparation des protéines sur gels 2D et identification par nano-LC/MSMS, soit par une approche quantitative label-free sur des digestions liquides,
- l'évaluation de différents algorithmes de quantification : ProteinPilot/PeakView/MarkerView (AB Sciex), Skyline (Laboratoire MacCoss, Université de Washington), Progenesis (Nonlinear dynamics),
- l'ensemble des protéines obtenues pouvant constituer des biomarqueurs potentiels de pathologies, en l'occurrence de type cancer, liées à l'exposition au BaP.

Partie I

Introduction bibliographique

I) Les hydrocarbures aromatiques polycycliques

1) Définition

Les Hydrocarbures aromatiques polycycliques (HAP) sont ubiquitaires dans l'environnement : ils sont parmi les premiers polluants atmosphériques à avoir été identifiés comme cancérogènes (Grimmer, Brune et al. 1991). Ils sont des composés organiques constitués de carbone et d'hydrogène et caractérisés par la présence de deux à six cycles aromatiques juxtaposés. Ils sont générés essentiellement lors de la combustion incomplète de la matière organique (charbon, carburants fossiles, bois, ...).

1.1) Sources d'exposition aux HAP

Les principales sources d'exposition des populations aux HAP sont les sources anthropiques (imputables aux activités humaines) représentant 90% des HAP émis dans l'atmosphère et les sources naturelles (feux de forêt, volcans). Les sources anthropiques se divisent en sources domestiques, sources industrielles, et transport.

Les sources domestiques, malgré leurs petites sources individuelles, sont très nombreuses et répandues. Elles conduisent à l'émission d'une quantité très importante d'HAP. L'alimentation représente la principale source d'exposition tout en considérant les diverses sources d'exposition humaine, suivie par les tabagismes actif et passif (Figure 1). Une personne à alimentation équilibrée ingère 235 ng/jour de BaP en moyenne. Les sources alimentaires proviennent aussi bien des aliments eux-mêmes (céréales, huiles, graisses) que des modes de cuisson (grillage des graisses). Les céréales et les produits de la pêche sont consommés en grande quantité : les céréales tiennent le haut du pavé, l'ingestion quotidienne moyenne de BaP a été estimée entre 50 à 290 ng/adulte (EFSA 2008). Les aliments grillés et fumés contiennent des taux d'HAP de l'ordre de 165 ng/g. Les végétaux contiennent aussi des quantités non négligeables d'HAP de l'ordre de 25 ng/g : les HAP absorbés de l'atmosphère se concentrent dans les parties feuillues (Phillips 1999). En revanche, l'absorption du BaP par le biais de l'eau est insignifiante, elle se chiffre à 2 ng/jour (EFSA 2008). Quant au tabagisme, il constitue aussi une source importante d'exposition aux HAP. La plupart des HAP qu'on respire proviennent de la fumée de tabac. Plus de 150 HAP ont été retrouvés dans la fumée de tabac (Lee, Novotny

et al. 1976). Une personne fumant 20 cigarettes par jour absorbe en moyenne 105 ng/jour de BaP et un fumeur passif 40 ng/jour (EFSA 2008).

Parmi les plus importantes sources industrielles figurent les activités industrielles pétrochimiques, les activités basées sur la fabrication de pneus, l'emploi de l'asphalte, la production du fer et d'aluminium et l'incinération des déchets.

Le transport est une source importante d'introduction des HAP dans l'espace urbain. Ces HAP proviennent essentiellement des gaz d'échappement et, dans une moindre mesure, de l'abrasion des pneus des voitures. Les doses de HAP absorbés à l'extérieur se montent à 20 ng/jour de BaP, ce qui reste relativement peu comparé aux autres sources de HAP.

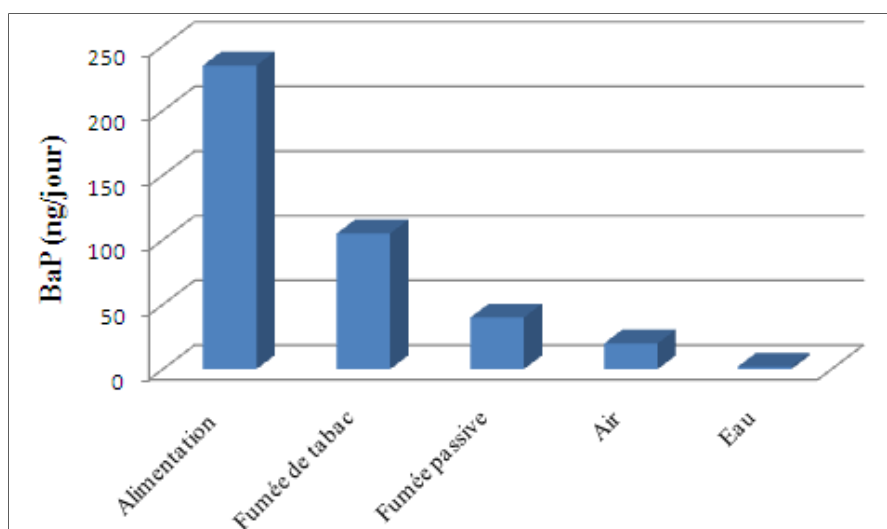


Figure 1 : Quantité moyenne de Benzo(a)pyrène absorbé chaque jour à partir de différentes sources (EFSA, 2008)

1.2) Des molécules à plusieurs cycles : Classification et Règlementation

Comme leur nom l'indique, les HAP sont composés d'au moins deux cycles de benzène accolés, le benzène étant l'élément de base des HAP (Figure 2). Selon leur nombre de cycles et donc leur poids moléculaires, ils sont classés en HAP gazeux ou légers (jusqu'à trois cycles) et particulaires ou lourds (quatre cycles et plus) qui ont des caractéristiques physico-chimiques et toxicologiques différentes. Les HAP particulaires représentent entre 70 et 80% des HAP totaux de l'atmosphère.

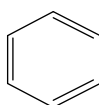


Figure 2 : Structure chimique du benzène

Du fait de leur présence dans l'environnement sous forme de mélanges complexes, seuls certains HAP peuvent être identifiés par les analyses. Dans les années 80, l'USEPA a classé 16 HAP dans sa liste des polluants prioritaires (Tableau 1). Ces 16 composés (Figure 3) sont considérés comme représentatifs de toute la famille des HAP. La concentration en HAP dans un mélange peut être déduite facilement puisque le BaP se trouve en proportion relativement constante, de l'ordre de 10%.

Le CIRC (Centre International de Recherche sur le Cancer) propose des classifications des produits cancérigènes. Ces classifications sont présentées dans le tableau 1.

Tableau 1 : Les principaux HAP classés selon différents critères

HAP	Polluants Prioritaires (USEPA)	Cancérogène pour l'homme (CIRC, 2010)	Cancérogène dans l'alimentation (EFSA, 2008)
Benzo(a)pyrène	X	2A	X
Cyclopenta(c,d)pyrène		3	
Dibenzo(a,h)anthracène	X	2A	X
Dibenzo(a,l)pyrène		2B	
Benz(a)anthracène	X	2A	X
Benzo(b)fluoranthène	X	2B	X
Benzo(j)fluoranthène		2B	
Benzo(k)fluoranthène	X	2B	X
Benzo(c)phenanthrène		3	
Benzo(e)pyrène		3	
Chrysène	X	3	X
Dibenzo(a,h)pyrène		2B	
Dibenzo(a,i)pyrène		2B	
Indéno(1,2,3,c,d)pyrène	X	2B	X
5-méthylchrysène		2B	
Benzo(g,h,i)pérylène	X	3	X
Anthracène	X	3	
Acénaphène	X	-	
Acénaphylène	X	-	
Fluoranthène	X	3	
Fluorène	X	3	
Naphtaline	X	-	
Phénanthrène	X	3	
Pyrène	X	3	

Classification du CIRC :

Catégorie 1 : cancérigène pour les humains.

Catégorie 2A : cancérigène probable pour les humains.

Catégorie 2B : cancérogène possible pour les humains.

Catégorie 3 : classification impossible quant au pouvoir cancérogène pour les humains.

En 2005, les données épidémiologiques obtenues chez l'homme pour le BaP ont permis au CIRC de classer ce composé dans le groupe 1 (Straif, Baan et al. 2005).

L'EFSA (Autorité Européenne de Sécurité des Aliments) répertorie 8 HAP cancérogènes pour l'homme lorsqu'ils sont présents dans les denrées alimentaires (Tableau 1).

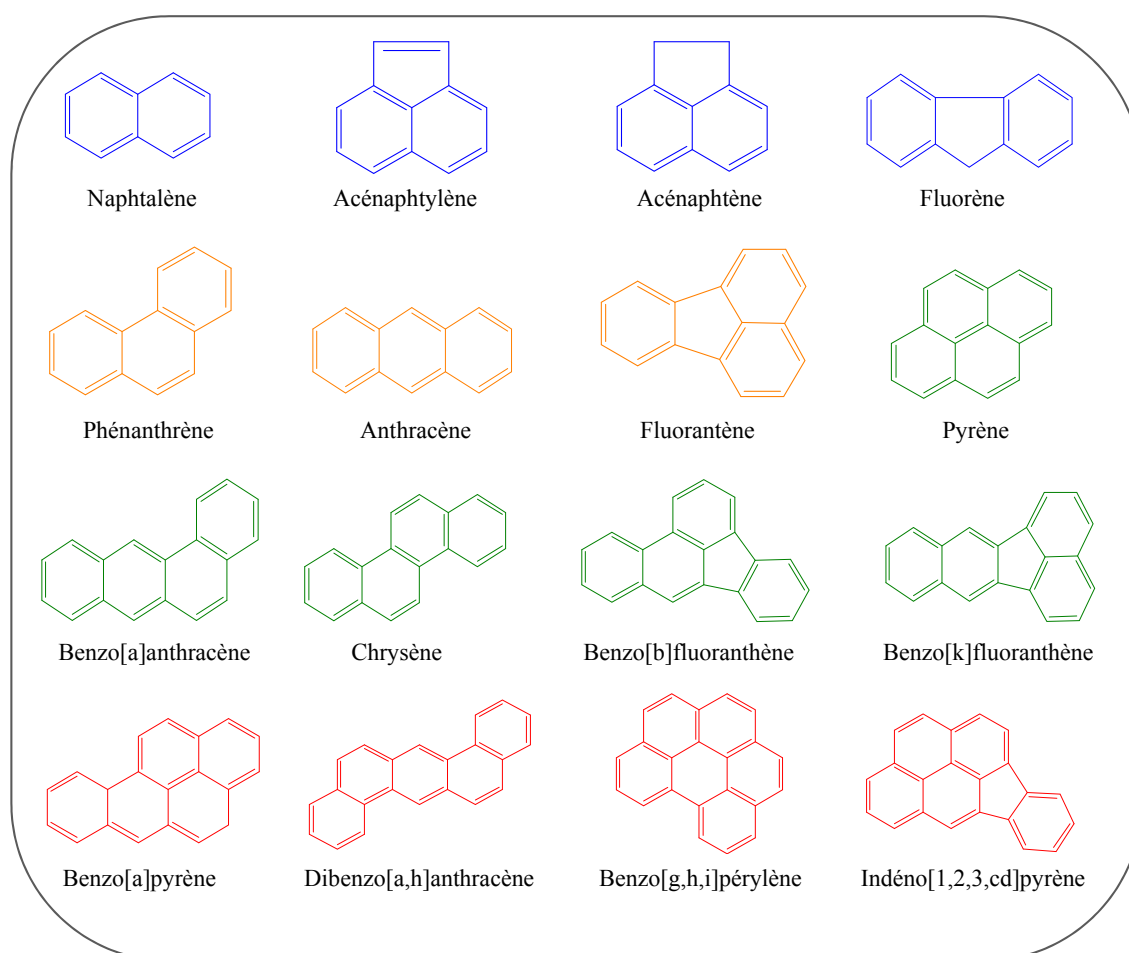


Figure 3 : Présentation des 16 HAP classés prioritaires par l'USEPA
(2cycles, 3cycles, 4cycles, 5cycles)

1.3) Effets des HAP sur la santé

Les HAP peuvent être à l'origine d'une toxicité aiguë. Si l'homme est exposé simultanément aux HAP et aux UV, une dermatite phototoxique accompagnée de conjonctivite peuvent apparaître (Wang, Sheng et al. 2007). Ils peuvent aussi entraîner une diminution de la réponse du système immunitaire et donc le risque d'infection : cette étude a été rapportée chez des salariés travaillant dans une fonderie (Szczeklik, Szczeklik et al. 1994). De plus, les HAP peuvent être à l'origine de la réduction du développement de tous les organes.

Les effets des HAP sur la santé sont principalement chroniques. Les HAP sont des molécules biologiquement actives, la structure de certains leur permet de se transformer dans les tissus organiques en époxydes, composés extrêmement toxiques. Ces époxydes réagissent avec l'ADN pour former des adduits à l'ADN et provoquer des dysfonctionnements cellulaires, ce qui entraîne des mutations génétiques, pouvant à terme, aboutir au développement d'un cancer. Parmi ces HAP, le BaP, le composé auquel nous nous intéressons dans le cadre de la présente thèse, est particulièrement toxique : ses effets cancérigènes sont déjà prouvés pour l'homme et il est considéré comme mutagène et tératogène. Des études épidémiologiques ont démontré, dans plusieurs secteurs professionnels, que l'exposition aux HAP entraînent une augmentation du risque d'apparition des cancers du poumon, de la peau et de la vessie (Mastrangelo, Fadda et al. 1996, Boffetta, Jourenkova et al. 1997, Bosetti, Boffetta et al. 2007, Chen, Tsai et al. 2008). En environnement, une corrélation significative entre les HAP contenus dans les cigarettes et le cancer du poumon a également été établie (Denissenko, Pao et al. 1996). De plus, certains HAP peuvent affecter la reproduction ou le développement fœtal (effets tératogènes démontrés en expérimentation animale).

2) Devenir des HAP dans l'organisme

2.1) Absorption

Le passage des HAP dans l'organisme humain s'effectue par inhalation, par ingestion, et par transfert à travers la peau. Leur pénétration en passant la barrière cutanée, digestive et pulmonaire et leur diffusion passive dans l'organisme, à travers les membranes plasmiques, est très facile du fait de leur structure aromatique qui leur assure une très grande lipophilie.

2.1.1) Absorption respiratoire

50% des HAP totaux sont adsorbés sur les particules fines de moins de $0,5\mu\text{M}$ (Rehwagen, Müller et al. 2005) qui sont capables de pénétrer en profondeur dans le système

broncho-pulmonaire, tandis que les plus grosses particules sont déposées au niveau des voies aériennes supérieures où les HAP seront lentement absorbés.

2.1.2) Absorption digestive

Lors des expositions par voie orale, les HAP sont rapidement absorbés au niveau intestinal (facilité par les aliments riches en graisses).

2.1.3) Absorption cutanée

C'est surtout dans le cas des expositions professionnelles que cette voie d'entrée est très importante pour les HAP. L'absorption cutanée est observée lors du contact direct avec les produits ou lors du contact indirect avec des éléments souillés par les HAP, de même par retombées atmosphériques de leurs poussières en suspension.

2.2) Distribution

Quelle que soit la voie d'entrée des HAP, des études sur l'animal ont montré qu'ils sont rapidement transportés, via le sang et les vaisseaux lymphatiques, vers d'autres organes : leur stockage se fait dans le foie, les reins et le tissu adipeux.

Dans le cas d'une exposition respiratoire et après injection intraveineuse, le poumon sera aussi un site de stockage, cet organe recevant la totalité du débit cardiaque est le premier à être en contact avec la dose intraveineuse après le cœur (Marie, Bouchard et al. 2010).

De même, le compartiment gastro-intestinal contient des niveaux assez élevés d'HAP.

Vu cette large diffusion des HAP dans l'organisme, leur cinétique d'élimination sera limitante.

2.3) Métabolisation

Parmi les HAP classés pour leur carcinogénicité, ne figurent aucun HAP gazeux. La carcinogénicité est fréquemment associée à des HAP de quatre à six cycles et possédant en effet des configurations structurales particulières appelées régions baie et fjord (Figure 4) qui sont des structures angulaires formées par les anneaux et constituant une gêne stérique favorisant les réactions d'oxydation et la formation de radicaux. Le processus d'oxydation qui conduit à des composés réactifs nécessite la présence de l'une de ces deux régions (Baird, Hooven et al. 2005). Ainsi, les HAP symétriques, ne possédant ni une région baie ni fjord, sont éliminés en un nombre restreint de métabolites mono-hydroxylés, comme le pyrène ou le naphthalène.

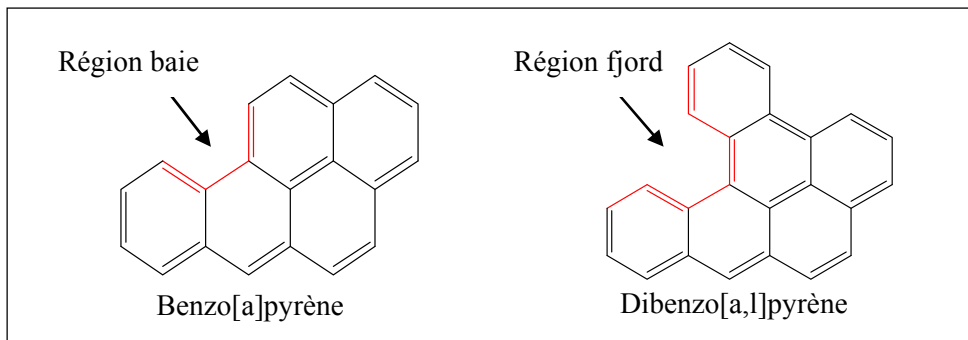


Figure 4 : Caractéristique structurale de deux HAP ayant un potentiel cancérigène

Par contre, le métabolisme des molécules asymétriques conduit à des composés secondaires de différents types : époxydes, dihydrodiol-époxydes, phénols, dihydrodiols, phenol-dihydrodiols, quinones, hydroquinones, tétrahydrodiols qui peuvent soit réagir avec des composants cellulaires soit être conjugués et éliminés sans exercer leur toxicité.

Le BaP est un procarcinogène nécessitant une activation en intermédiaires réactifs pouvant réagir sur l'ADN et exercer des effets toxiques en absence de leur élimination par conjugaison (Miller and Ramos 2001) (Figure 5).

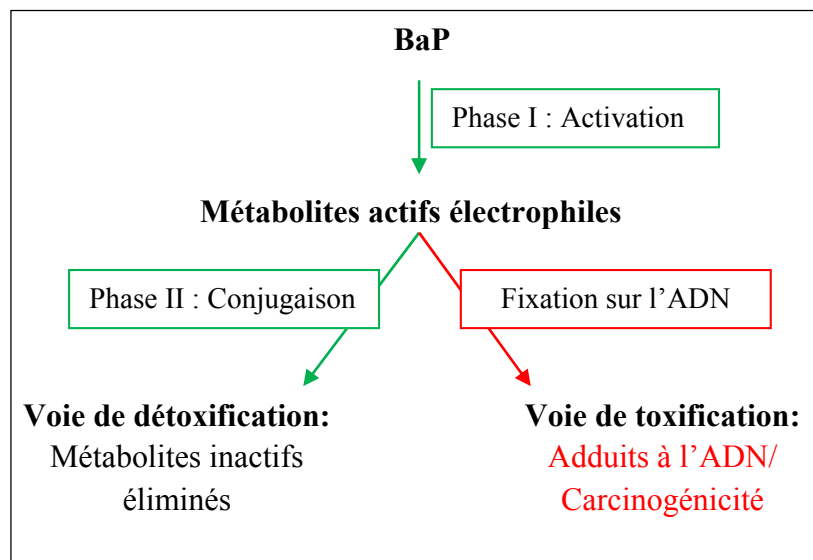


Figure 5 : Métabolisme simplifié du BaP

Les principales voies de métabolisation des HAP sont au nombre de trois (Xue and Warshawsky 2005), toutes impliquant l'action de plusieurs enzymes. La première voie est capable de métaboliser les HAP en époxydes par une famille de cytochromes P450 (CYP450). La deuxième est catalysée par des peroxydases tout en générant un radical cation. Et la dernière, catalysée par des dihydrodiols déshydrogénases, repose sur la formation d'o-quinones à partir de catéchols.

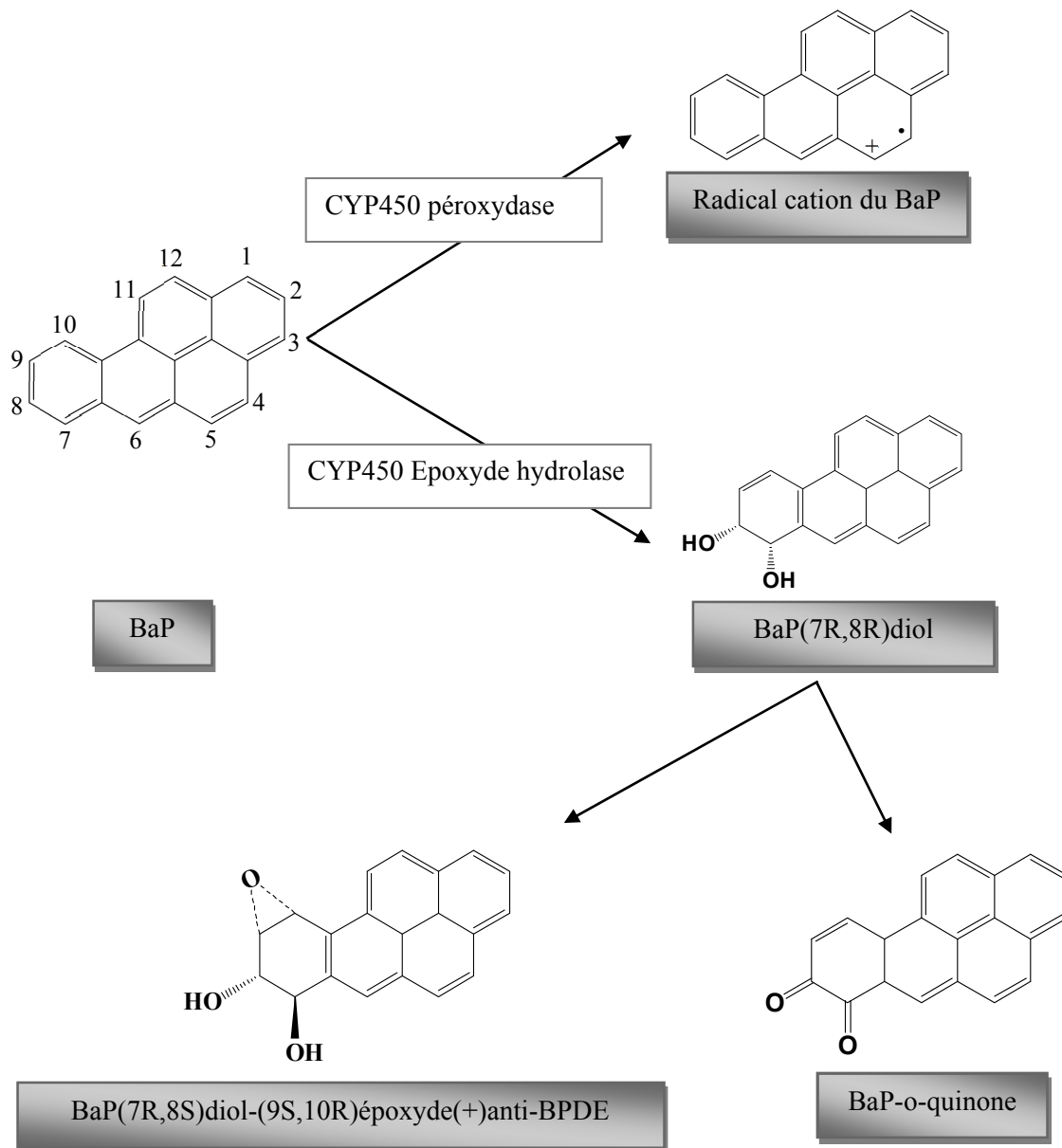


Figure 6 : Les 3 voies principales de métabolisation du BaP

2.3.1) Voie de mono-oxygénation par les CYP450 : voie des diol-époxydes

Les CYP450 constituent 17 familles multigéniques d'enzymes chez l'homme : elles catalysent des réactions d'oxydation et de réduction du métabolisme des xénobiotiques, surtout les familles CYP1, 2 et 3. Lors de cette étape, les HAP sont pris en charge par les mono-oxygénases à cytochrome P 450, surtout par les isoformes : 1A1, 1A2 et 1B1 (Shimada, Oda et al. 2001). Le CYP1A2 est exprimé particulièrement dans le foie, tandis que les CYP 1A1 et 1B1 le sont préférentiellement dans les organes extra-hépatiques. L'activation du BaP est plutôt catalysée par le CYP1B1, alors que le CYP1A1 est plutôt impliqué dans la détoxification du BaP (Uno, Dalton et al. 2006). Bien que le CYP1A2 soit exprimé dans le foie, il catalyse une production de dihydrodiols beaucoup plus faible que celle catalysée par les CYP 1A1 et 1B1.

La voie de mono-oxygénation par les CYP450, très étudiée dans le cas du BaP, est considérée comme majoritaire (Xue and Warshawsky 2005). La métabolisation du BaP, notamment l'oxydation d'une double liaison catalysée par les CYP450, entraîne la formation des époxydes : 4,5-, 7,8- et 9,10-époxyde BaP (Jacob and Grimmer 1995). Leur hydrolyse par l'Époxyde Hydrolase (EH) produit des *trans*-dihydrodiols, le 7,8-*trans*-dihydrodiol-BaP étant le majoritaire. Ce composé subit une nouvelle époxydation sur C9 et C10 par les CYP450 et donne un diol époxyde très réactif vis-à-vis de l'ADN, le BPDE (7,8-Dihydro-7,8-dihydroxybenzo(a)pyrène 9,10-oxyle), considéré comme le carcinogène ultime du BaP. Cette époxydation mène à quatre configurations isomériques : 2 paires d'énantiomères (\pm)-anti-isomères et (\pm)-syn-isomères. Ces métabolites sont produits suivant une grande stéréosélectivité, notamment la formation presque exclusive de l'isomère 7R,8S-oxyle qui sera préférentiellement converti en 7R,8R-diol : c'est l'isomère (+)-(7R,8S,9S,10R)-BPDE [(+)-anti-BPDE].

Le BPDE peut être hydrolysé en tétraols inactifs (Xue and Warshawsky 2005) ou former des adduits à l'ADN (Chen, Devanesan et al. 1996) : c'est l'anneau époxyde qui est le site à plus grande réactivité électrophile et qui sera préférentiel pour l'interaction covalente BPDE-bases puriques de l'ADN. Le terme « adduit à l'ADN » désigne le produit généré de l'addition d'un xénobiotique sur un nucléotide.

Outre leur particularité d'être des substrats, les HAP sont aussi des inducteurs des gènes des CYP450. Cette induction est régulée par l'Ahr (ArylhydrocarbonReceptor), un récepteur faisant partie de la famille bHLH/PAS (basic Helix-Loop-Helix/Per-Arnt-Sim) et spécifique des composés aromatiques plans.

Mécanismes de l'induction du gène *cyp1a1* par le BaP :

Les récepteurs Ah (Aryl hydrocarbon), situés dans le cytoplasme, sont associés à des protéines chaperonnes du type HSP90 (Heat Shock Protein). Une fois que le BaP traverse la membrane plasmique, il se lie directement à l'Ahr : ce dernier se dissocie de l'HSP90 et le complexe BaP-Ahr sera transloqué dans le noyau où il forme des dimères avec des protéines partenaires, les Arnt (Ah receptor nuclear translocator). Le complexe BaP-Ahr-Arnt se fixe au niveau d'un élément de réponse aux xénobiotiques (XRE) pour activer la transcription des gènes *cyp*. L'ARNm sera transloqué dans le cytoplasme et traduit par les ribosomes pour donner une protéine néo-synthétisée CYP1A1 capable de métaboliser une molécule de BaP en produits diols et diol-époxyde (Figure 7).

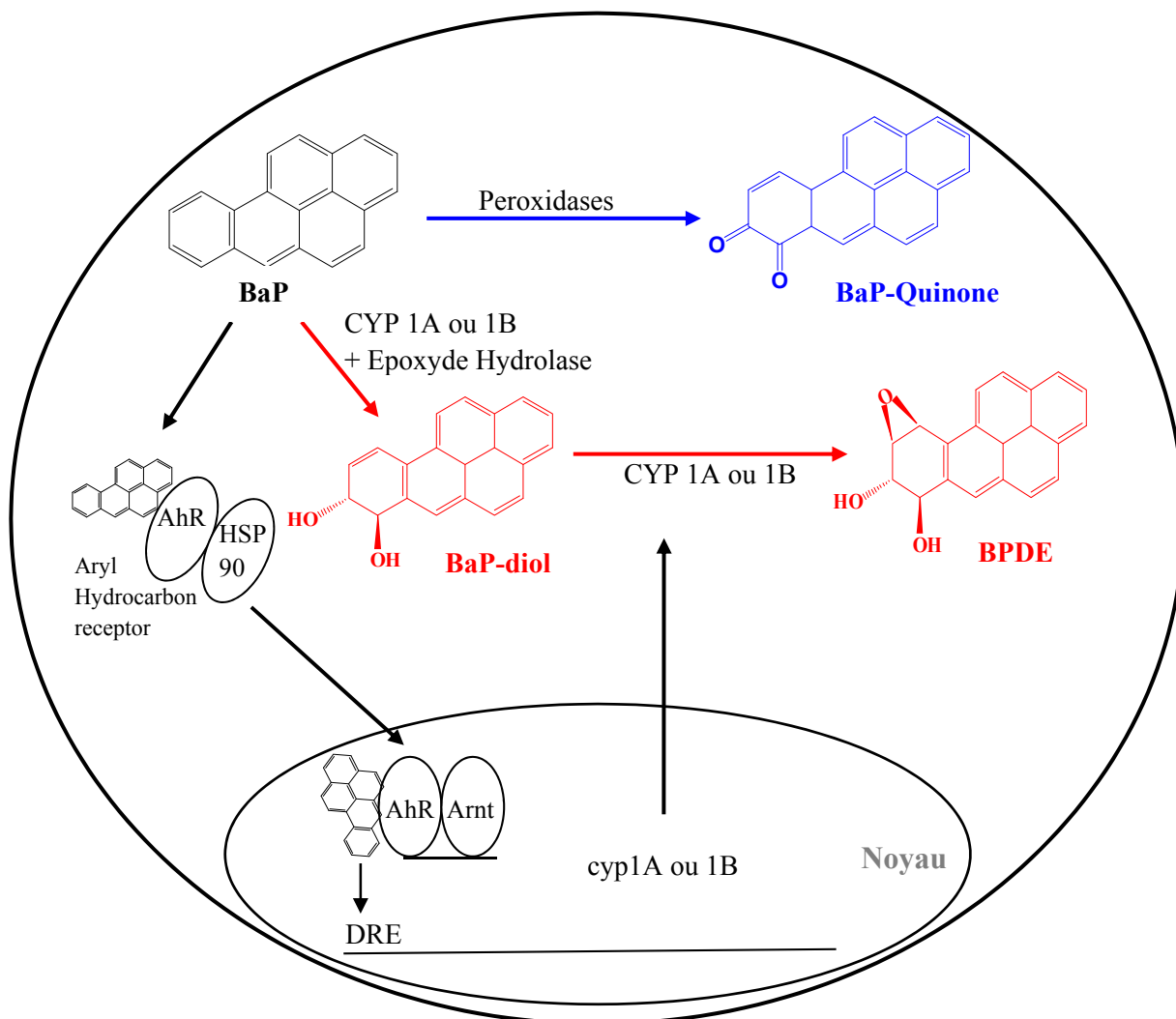


Figure 7 : Schéma simplifié du mécanisme de l'induction des gènes cyp par le BaP

2.3.2) Voie d'oxydation mono-électronique par les peroxydases : voie du radical cation

Cette voie, également appelée voie du carbocation, implique la biotransformation des HAP en radical cation. Dans le cas du BaP, le C6, le carbone le plus électrophile, subit une oxydation mono-électronique par des peroxydases dépendantes des CYP450 (Penning, Burczynski et al. 1999), ce qui conduit à avoir des radicaux cations du BaP électrophiles très réactifs vis-à-vis de l'ADN, cette réactivité étant due aux caractéristiques suivantes des HAP :

- ❖ Un potentiel d'ionisation bas permettant l'arrachement d'un électron
- ❖ La localisation de la charge sur le radical cation rendant l'intermédiaire plus réactif vis-à-vis d'espèces nucléophiles
- ❖ Une configuration géométrique optimale leur permettant de s'insérer dans l'ADN en formant des liaisons covalentes avec les sites nucléophiles de l'ADN

2.3.3) Voie des o-quinones

CYP450 et EH sont les enzymes impliquées dans cette voie de biotransformation puisque les premières étapes de formation de ces o-quinones sont celles de la voie des diol-époxydes jusqu'à ce que le 7,8-dihydrodiol-BaP se forme. Puis interviennent les « dihydrodiol déhydrogénase » qui appartiennent à la famille des « Aldo-Keto Réductases » (AKR). Le BaP-diol sera donc métabolisé en catéchol qui, suite à deux réactions d'oxydation, se transforme en BaP-7.8-dione ou BaP-o-quinone (Xue and Warshawsky 2005). Ce produit électrophile est très réactif vis-à-vis de l'ADN : il conduit à la génération de bases oxydées notamment au niveau de la guanine : 7,8-dihydro-8-oxo-guanine. Cette voie métabolique peut former des adduits à l'ADN comme elle peut produire d'espèces réactives à l'Oxygène.

2.3.4) Autre voie : voie mineure

Cette voie de biotransformation, catalyse l'estérification d'acide sulfurique par l'action de sulfotransférases. Elle est considérée comme voie mineure (Xue and Warshawsky 2005).

2.4) Elimination : réactions de conjugaison

L'élimination des métabolites des HAP se fait majoritairement par les urines et les fèces. L'excrétion urinaire des métabolites du BaP est une voie mineure : l'élimination se fait majoritairement par le biais du système hépatobiliaire et par le tractus intestinal.

Les enzymes de phase I (activation) introduisent ou modifient, par des réactions d'hydrolyse ou d'oxydation, des groupements fonctionnels au niveau des xénobiotiques afin qu'ils deviennent plus hydrosolubles et puissent s'éliminer dans les urines, les fèces et la bile par des réactions de conjugaison catalysées par l'action de plusieurs enzymes.

Les principales voies d'élimination sont les suivantes :

2.4.1) Conjugaison par les UDP-Glucuronyltransférases

Les UGT (UDP-Glucuronyltransférases), situées dans le réticulum endoplasmique sont dépendantes du récepteur Ah et catalysent la conjugaison des dihydrodiols avec le glucuronide. De plus, les diphénols formés à partir de la réduction des quinones peuvent se conjuguer au glucuronide afin d'être excrétés.

L'activité des UGT semble être liée à l'activité des CYP450 : le CYP1A peut interagir avec les UGT responsables de la conjugaison des phénols et peut donc intervenir dans le mécanisme d'action de ce type de conjugaison (Taura, Ishii et al. 2004).

2.4.2) Conjugaison par les glutathion-S-transférases

Fjellstedt *et al.* ont démontré que cette famille d'enzymes catalyse la conjugaison du glutathion avec les époxydes, notamment le BPDE, leur substrat préférentiel (Figure 8) (Fjellstedt, Allen *et al.* 1973, Jernström, Funk *et al.* 1996).

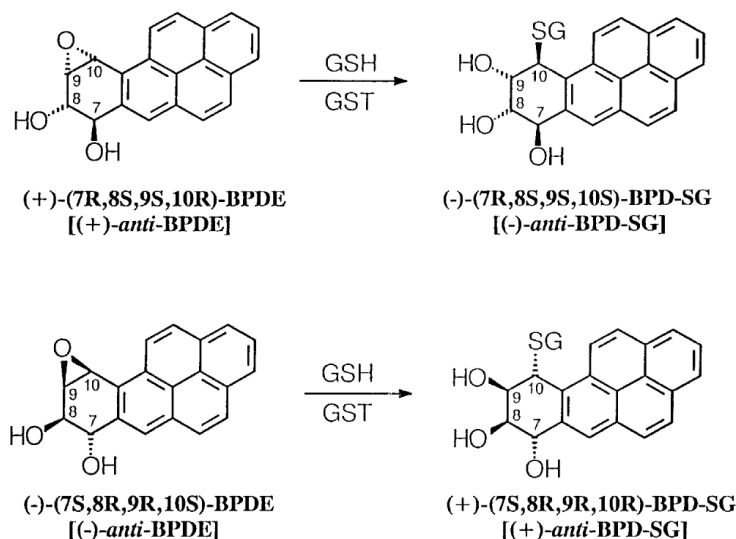


Figure 8 : Conjugaison du glutathion au BPDE
(-)-anti-BPD-SG : (+)-anti-BPDE conjugué au glutathion;
(+)-anti-BPD-SG : (-)-anti-BPDE conjugué au glutathion

2.4.3) Conjugaison par les Sulfotransférases

Les substrats des sulfotransférases sont, dans le cas des HAP, les hydroquinones, les phénols et les dihydrodiols (Glatt 2000). Le transfert du groupe sulfonate, qui provient du 3'-phosphoadénosine-5'-phosphosulfate, à un groupe hydroxyle ou amine est catalysé par ces enzymes. Une charge négative est introduite sur le composé au cours de sa conjugaison, ce qui permet d'augmenter son hydrosolubilité et son excrétion par les voies biliaires et urinaires (Glatt 2000).

3) Toxicoprotéomique : Analyse protéomique de la toxicité induite par le BaP

La capacité du BaP à former des adduits à l'ADN le rend un composé génotoxique très puissant. Cet effet génotoxique a été démontré dans des études à la fois *in vivo* et *in vitro* (Talaska, Jaeger *et al.* 1996, van Leeuwen, Gottschalk *et al.* 2005). Les dommages à l'ADN pourront se produire sur les 4 bases de l'ADN, les plus étudiés sont ceux liés à la guanine, la base la plus sensible aux oxydations (Cadet, Delatour *et al.* 1999).

Hockley *et al.* ont démontré que l'effet génotoxique est dose-dépendant et que le nombre d'adduits à l'ADN augmente avec la concentration du BaP (Hockley, Arlt *et al.* 2006).

Les dommages d'ADN et les altérations des voies de transduction du signal peuvent affecter l'expression de plusieurs gènes, ce qui permet de les analyser au niveau protéique par des techniques de protéomique. Ce qui suit met en évidence des exemples dans lesquels la protéomique a été utilisée pour l'analyse de la toxicité du BaP et la découverte de biomarqueurs.

3.1) Effets carcinogènes

3.1.1) Carcinogénèse pulmonaire

Les cellules épithéliales respiratoires sont parmi les premières à entrer en contact avec les agents exogènes de l'atmosphère tels que le BaP. Il a été établi, depuis les années 90, que 60% des cancers du poumon sont dus aux mutations causées par le BaP et autres HAP (Hollstein, Shomer et al. 1996, Hecht 1999). Puisqu'il n'y a pas de tests de diagnostic fiables pour la détection précoce de ce type de cancer, les biomarqueurs pourront être utiles afin de pallier cette maladie. Minet *al.* ont analysé les modifications d'expression des protéines des cellules épithéliales des voies respiratoires humaines (A549) pour décrire la réponse cellulaire précoce de ces cellules au stress oxydatif induit par le BaP (Min, He et al. 2011). Les protéines du protéome total de ces cellules exposées ou non à 1µM de BaP ont été séparées par électrophorèse bidimensionnelle. 23 protéines parmi 39 différentiellement exprimées ($\geq 1,5$ -fold) entre les cellules exposées et les contrôles ont été caractérisées par spectrométrie de masse (MALDI-TOF). Un différentiel d'expression de la superoxyde dismutase (SOD) a été confirmé par Western Blot: un changement d'expression dose-dépendante de cette enzyme a été observé. Ces résultats confirment l'hypothèse de cette étude : le BaP provoque un stress oxydatif en induisant un dysfonctionnement au niveau de la mitochondrie et une perturbation de l'effet antioxydant grâce à une altération d'expression des protéines comme la SOD.

Depuis une décennie, il a été démontré qu'une exposition simultanée à l'arsenic potentialise la génotoxicité du BaP (Maier, Schumann et al. 2002). Dans une tentative d'obtenir d'autres renseignements sur les effets globaux déployés dans les cellules, au niveau protéomique, après exposition simultanée au BaP et à l'arsenic, Lau *et al.* analysent l'expression protéique d'une lignée cellulaire épithéliale pulmonaire de rat (LEC pour Lung Epithelial Cell) exposée ou non au BaP seul (100nM), à l'arsenic seul (1,5M) et aux deux produits simultanément pendant 12 semaines (Lau and Chiu 2006). Pour démontrer les effets toxiques synergiques du BaP et de l'arsenic, des protéines ont été séparées par électrophorèse bidimensionnelle et celles différentiellement exprimées entre les 3 groupes ($\geq 1,5$ -fold) ont été caractérisées par MALDI-TOF : l'expression de la périphérine a été observée pour la première fois dans des cellules non neuronales. Cette expression a diminué seulement dans les LEC

exposées au BaP seul ou au BaP et à l'arsenic (révélation par Western Blot). La périphérine est nécessaire à l'organisation du cytosquelette et à l'architecture cellulaire : une perte sévère aboutit à une forme anormale des LEC. Les auteurs ont aussi démontré que les LEC co-exposées au BaP et à l'arsenic avaient un niveau élevé d' α -actinine qui pourrait être associée à une perte d'expression de CD14 (cluster of differentiation14) indiquant que les LEC subissent une dédifférenciation et donc une transformation. Cette étude montre comment les techniques protéomiques peuvent aider à identifier les voies de signalisation impliquées dans la transformation cellulaire et à comprendre les mécanismes de l'action synergique de BaP et de l'arsenic au cours de la cancérogenèse.

3.1.2) Carcinogenèse cervicale

Les métabolites du BaP ont été trouvés dans le mucus cervical et les adduits à l'ADN dans le tissu cervical chez les humains (Melikian, Sun et al. 1999). Yan *et al.* ont utilisé les cellules HeLa du carcinome cervical humain pour évaluer les changements de l'expression des protéines nucléaires induits par le BaP (Yan, Wu et al. 2010). Ils ont enrichi les protéines nucléaires de ces cellules, exposées ou non à 10 μ M de BaP, par fractionnement subcellulaire et ont séparé les protéines extraites sur gels bidimensionnels. L'analyse statistique ($p < 0,05$) indique une expression différentielle de 128 protéines, parmi lesquelles 45 spots ont été identifiés par spectrométrie de masse (LC-ESI). Plusieurs protéines identifiées ont été impliquées dans l'épissage alternatif. Pour valider leurs résultats, ils ont réalisé une analyse par RT-PCR (Reverse Transcriptase-Polymerase Chain Reaction) des gènes Fas et CD44 connus pour être épissés alternativement en réponse aux dommages de l'ADN (Filippov, Filippova et al. 2007, Filippov, Schmidt et al. 2008) et n'ont constaté que des formes variantes de CD44. Cela indique que l'activation de l'épissage alternatif est impliquée dans l'apoptose induite par le BaP.

3.1.3) Cancérogenèse de la vessie

Le risque que l'exposition au BaP ou à un autre HAP provoque un cancer de la vessie est encore incertain (Kogevinas, Mannelje et al. 2003). Cependant, dans la plupart des populations, le tabagisme est la principale cause connue de ce type de cancer ; il est associé à plus de la moitié des cas de cancer de la vessie chez les hommes (Castelao, Yuan et al. 2001). Pour analyser le rôle possible du BaP dans le développement du cancer de la vessie, Verma *et al.* ont examiné les changements d'expression des protéines dans les cellules urothéliales exposées au BaP (Verma, Rettenmeier et al. 2010). Pour obtenir une vue d'ensemble des protéines exprimées dans les cellules épithéliales de la vessie, Verma *et al.* ont mis en place une carte du protéome de ces cellules isolées du porc (Verma, Bäuerlein et al. 2011). 42 protéines différemment

exprimées sur gels bidimensionnels ont été obtenues après exposition de ces cellules à de faibles doses de BaP (0,5µM). L'analyse de ces protéines par MALDI-TOF a révélé des protéines apoptotiques, des protéines impliquées dans la réparation de l'ADN et quelques protéines nucléaires. Jusqu'à présent, ce sont les seuls résultats qui montrent que la formation d'une tumeur dans la vessie pourrait être induite par le BaP.

3.1.4) Cancérogenèse du sein

Le tabagisme semble aussi augmenter le risque du cancer du sein, et des adduits à l'ADN liés au tabac ont été trouvés dans les tumeurs du sein et des tissus sains environnants (Perera, Estabrook et al. 1995, Li, Wang et al. 1996).

Hooven *et al.* ont utilisé une approche protéomique pour distinguer entre l'expression protéique des cellules des tissus mammaires du carcinome humain (MCF-7) exposées à un HAP individuel (BaP, dibenzo [a,l] pyrène) et à un mélange de HAP (Hooven and Baird 2008). Après 24h d'exposition à ces divers composés, les protéines extraites de ces cellules ont été séparées par électrophorèse bidimensionnelle. Les résultats de l'analyse MALDI-TOF/TOF révèlent que chaque exposition à un HAP montre une empreinte protéomique unique et peuvent donner une idée plus précise des mécanismes possibles impliqués dans la cancérogenèse induite par les HAP. Cependant, ces résultats sont préliminaires puisqu' aucune de ces protéines identifiées n'a encore été validée.

3.1.5) Cancérogenèse de la prostate

Plusieurs études épidémiologiques ont prouvé que la consommation de viande grillée est liée à un risque élevé du cancer de la prostate (Koutros, Cross et al. 2008, Venkateswaran and Klotz 2010). Le BaP, formé en grandes quantités dans la viande grillée, est considéré comme agent causal de cet effet cancérigène (Fatma Aygün and Kabadayi 2005) : ses métabolites potentiellement cancérigènes ont été observés dans une culture primaire de cellules de la prostate (Williams, Martin et al. 2000). Chaudhary *et al.* ont utilisé une approche protéomique pour élucider les mécanismes de la carcinogenèse médiée par le BaP et le potentiel anti-cancéreux de la quercétine dans le cas de ce cancer (Chaudhary, Pechan et al. 2007). Des cellules épithéliales humaines de la prostate, normales et cancéreuses, ont été exposées à 1µM BaP ou 5µM de la quercétine. Compte tenu du rôle du stress oxydatif dans le développement du cancer, et après avoir analysé le lysat cellulaire par électrophorèse bidimensionnelle et MALDI-TOF, les auteurs se sont focalisés sur l'expression différentielle des enzymes anti-oxydantes : les peroxyrédoxines (Prx I et II). L'exposition au BaP a provoqué une augmentation de l'expression de la Prx I et une diminution de celle de la Prx II dans les cellules saines et

cancéreuses. La quercétine a seulement augmenté l'expression de la Prx II dans les cellules cancéreuses, favorisant ainsi l'apoptose.

3.2) *Autres effets toxiques*

3.2.1) Hépatotoxicité

Le foie est l'usine centrale et principale du métabolisme de toute substance étrangère à l'organisme. La biotransformation de ces composés dans le foie implique des réactions d'oxydation principalement catalysées par les CYP450 : le BaP est un inducteur de ces enzymes comme il a été décrit dans la partie 2 et il est capable de déclencher des carcinomes hépatocellulaires chez la souris et le rat (Tomoyuki, Tadashi et al. 1980).

Dautel *et al.* ont utilisé une technique d'analyse différentielle de protéines sur gels bidimensionnels afin d'étudier les effets du BaP sur l'expression protéique des cellules d'hépatome murin (Hepa1c1c7) (Dautel, Kalkhof et al. 2010). Les cellules ont été exposées à 5µM et 50nM de BaP pendant 2 et 24 h. 112 protéines différemment exprimées ont été identifiées par MALDI-TOF et nanoLC-MS/MS. L'exposition à ces deux concentrations a induit l'expression de plusieurs enzymes antioxydantes et une augmentation de l'expression de protéines anti-apoptotiques comme la nucléophosmine et l'Annexine A1. Cette étude a montré que l'exposition au BaP cause un stress oxydatif et une altération dans la progression du cycle cellulaire.

3.2.2) Immunotoxicité

Le système immunitaire est indispensable pour lutter contre les infections : tout dysfonctionnement de ce système peut entraîner de graves conséquences pour la santé humaine. Le BaP, à forte concentration, active les gènes impliqués dans l'apoptose des cellules lymphoïdes (Burchiel and Luster 2001). Pour étudier les effets du BaP sur le système immunitaire, Oh *et al.* ont exposé une lignée cellulaire de lymphome T à 2,5 et 10µM de BaP pour 24 et 48h (Oh, Im et al. 2004). Une séparation sur gels bidimensionnels et une analyse par MALDI-TOF ont montré une expression différentielle, dose-dépendante, de 46 protéines : les protéines impliquées dans l'apoptose et la suppression de la tumeur sont surexprimées alors que celles impliquées dans la synthèse de l'ADN et la structure cellulaire sont sous-exprimées. Cette étude nécessite une poursuite d'expériences sur les protéines identifiées afin de déterminer leur rôle dans l'immunotoxicité médiée par le BaP.

3.2.3) Neurotoxicité

Comme la plupart des composés lipophiles, le BaP peut traverser facilement la barrière hémato-encéphalique et avoir un accès direct au système nerveux central. Des symptômes neurologiques, comme une perte de mémoire à court terme, ont été signalés dans une étude sur des travailleurs exposés professionnellement au BaP (Zhang, Nie et al. 2008). Afin de comprendre le mécanisme lié à cette perte de mémoire et étudier la neurotoxicité du BaP, Chengzhi *et al.* ont étudié l'expression des protéines liées au déficit de la mémoire spatiale dans l'hippocampe du cerveau de rats nourris de l'huile d'arachide et l'ont comparée à celui des rats nourris de BaP (2 mg/kg) (Chengzhi, Yan et al. 2011). La mémoire spatiale est la partie de la mémoire responsable de l'enregistrement des informations concernant l'espace environnant, et l'hippocampe est une zone du cerveau spécifiquement dédiée à l'apprentissage et à la mémoire spatiale. Le changement de l'expression de deux protéines a attiré l'attention des auteurs : le récepteur de l'acide rétinoïque et le facteur neurotrophique dérivé du cerveau, le premier étant impliqué dans l'inhibition de la mémoire à long terme et le second ayant une action sur certains neurones du système nerveux central. L'hypothèse de ces auteurs est : la sous-expression de ces deux protéines peut aboutir au déficit de la mémoire spatiale induite par le BaP. Ces protéines identifiées constituent de nouvelles cibles thérapeutiques pour la prévention et le traitement des maladies neuro-dégénératives.

Dans cette partie de toxicoprotéomique, des termes en rapport avec la spectrométrie de masse ont été mentionnés plusieurs fois : MALDI-TOF-MS, nanoLC-MS/MS et protéomique : ces techniques sont décrites dans la partie III.

II) Les biomarqueurs : vers la médecine personnalisée

La recherche et la découverte de biomarqueurs par approche protéomique pour le diagnostic d'une maladie et son suivi thérapeutique est un enjeu vers la médecine moderne, dite personnalisée.

1) Notion de biomarqueurs

Le terme « Biomarqueur » a pris son émergence avec celle des disciplines en « omique ». Par définition, un biomarqueur est une caractéristique, mesurée avec précision et reproductibilité, et évaluée comme indicateur physiologique ou pathologique. Il peut-être aussi un indicateur de l'action des médicaments. Autrement dit, un biomarqueur est un changement observable et/ou mesurable aux niveaux suivants :

- moléculaire
- biochimique
- cellulaire
- physiologique ou comportemental

Il révèle l'exposition d'un individu (actuelle ou passée) à une ou plusieurs substances chimiques considérées comme polluantes de l'environnement.

Il ne s'agit pas d'un concept nouveau. Par exemple, le cholestérol, un terme très courant en pratique médical, est un indicateur du risque cardiovasculaire.

En médecine, la détection de biomarqueurs permet de réaliser un dépistage, un diagnostic d'une pathologie (biomarqueurs diagnostiques), suivre l'évolution de la maladie (biomarqueurs pronostiques) ou encore l'efficacité ou la toxicité d'un traitement (biomarqueurs prédictifs).

On distingue trois types de biomarqueurs :

- Les biomarqueurs d'exposition (polluant dans l'organisme) : correspondent à une mesure quantitative du produit cancérigène ou d'un de ses métabolites dans un milieu biologique (exemple : les adduits à l'ADN sont utilisés comme biomarqueurs d'exposition à des cellules cancérigènes).
- Les biomarqueurs d'effet (réponse de l'organisme) : traduisent les conséquences physio-pathologiques de l'exposition à un polluant, ils permettent de détecter des effets biologiques plus ou moins proches de la survenue de la maladie (exemple : protéines du choc thermique, enzymes du stress oxydant...).
- Les biomarqueurs de sensibilité aux effets : traduisent les variations de la sensibilité et permettent de mettre en évidence les variations d'origine génétique de la réponse (liées au polymorphisme génétique des individus d'une même population ou communauté) (exemple : résistance aux contaminants).

Un tissu en lésion ou une tumeur peuvent libérer de façon aberrante des protéines spécifiques : on les retrouve dans les fluides biologiques cliniquement accessibles (le sang, l'urine ou le liquide céphalo-rachidien). Ces protéines sont des marqueurs biologiques ou biomarqueurs potentiels. Leur dosage sanguin utilisé en clinique depuis plusieurs années permet de refléter l'état physio-pathologique de l'individu. Ainsi plusieurs marqueurs protéiques ont été découverts et approuvés par la FDA (Food and Drug Administration), mais sans utiliser des méthodes protéomiques :

- cancer du colon : CEA (CarcinoEmbryonicAntigen)
- cancer de la prostate : PSA (ProstaticSpecificAntigen)
- cancer de l'ovaire : CA 125 (Cancer Antigen 125)
- cancer du sein : CA 15.3
- cancer du pancréas : CA 19.9
- Infarctus du myocarde : cTnI (cardiac Troponine I)

Pour la recherche d'un biomarqueur, plusieurs critères doivent être respectés :

- ❖ Présenter une sensibilité, une spécificité et une reproductibilité satisfaisante
- ❖ Permettre la détection des premiers stades de la pathologie
- ❖ Concorder avec la progression de la maladie
- ❖ Etre capable de suivre les effets de traitements thérapeutiques

La plupart des biomarqueurs protéiques déjà approuvés ne satisferont pas ces critères (Kulasingam, Pavlou et al. 2010) : il apparaît donc que l'usage combiné de plusieurs biomarqueurs est nécessaire, les pathologies étant multifactorielles (Visintin, Feng et al. 2008).

Le développement de tests cliniques de dépistage, simples et non-invasifs, est le but ultime de la recherche de biomarqueurs : ils permettent de détecter ou différencier des maladies avant qu'elles ne soient symptomatiques. En raison de la gamme dynamique étendue des protéines plasmatiques et de la très faible concentration des biomarqueurs dans les fluides biologiques très complexes, les techniques analytiques doivent être très sensibles et spécifiques.

L'approche protéomique offre la perspective de développer de nouveaux tests cliniques plus sensibles et plus spécifiques, la sensibilité étant la proportion de vrais positifs qui est détectée par le test comme ayant la « maladie » (Les faux négatifs sont les patients ayant la « maladie » mais non détectés par le test) et la spécificité étant la proportion de vrais négatifs qui est détectée par le test comme n'ayant pas la « maladie » (Les faux positifs sont les patients n'ayant pas la « maladie », mais détectés par le test comme étant positifs). Un test parfait serait celui qui posséderait 100% de sensibilité (Faux-négatifs = 0) ainsi que 100% de spécificité (Faux-positifs = 0). L'OVA1, le premier et unique test diagnostique autorisé par la FDA en 2010, permet d'identifier les tumeurs de l'ovaire avant de faire une chirurgie, et ceci en s'appuyant sur le dosage sérique combiné de cinq biomarqueurs protéiques (Fung 2010).

2) Les défis liés à la découverte de biomarqueurs protéiques dans le plasma

Le choix de la nature des prélèvements à effectuer est très important et dépend de la maladie à étudier. En effet, c'est plus préférable de choisir un fluide biologique facilement accessible, abondamment disponible, et peu invasif comme le sang ou l'urine que de prendre, par exemple, le liquide cébrospinal ou une biopsie (Rifai, Gillette et al. 2006).

Le sang était la première source de découverte des biomarqueurs par approche protéomique (Liotta, Ferrari et al. 2003). Il est recommandé d'utiliser le plasma, plutôt que du sérum dégradable *ex vivo*, avec l'EDTA (Ethylene diamine tetraacetic Acid) comme anticoagulant. Le plasma étant la matrice biologique de choix pour la mesure de biomarqueurs, il constitue un milieu physiologique circulant représentatif de la composition des tissus et de processus pathologiques. C'est un protéome très complexe, son problème majeur est sa teneur en protéines qui atteint 80 mg/mL : seule l'albumine est dosée à 40-50 mg/mL, elle représente plus de la moitié du taux protéique plasmatique (Anderson and Anderson 2002). Or les biomarqueurs d'intérêt ne constituent qu'1% de ce taux (1-100 ng/mL) puisque l'albumine avec les 22 protéines les plus abondantes en constituent 99% (Tirumalai, Chan et al. 2003). C'est le grand défi de la recherche de biomarqueurs protéiques dans le plasma sanguin.

De ce fait, des techniques de pré-fractionnement du protéome plasmatique ont été mises en place comme la déplétion par immuno-affinité des protéines les plus abondantes (Yocum, Yu et al. 2005, Pernemalm, Lewensohn et al. 2009) qui est réalisée à l'aide d'un mélange d'anticorps spécifiques immobilisés sur une matrice chromatographique (Figure 9), retenant, simultanément, jusqu'à 20 protéines majoritaires. Néanmoins, l'efficacité de cette technologie reste relative puisque l'élimination des protéines ciblées n'est jamais totale et des protéines de faible abondance et donc potentiellement les plus intéressantes peuvent-être aussi co-déplétées, notamment les protéines liées à des protéines majoritaires de transport comme l'albumine.

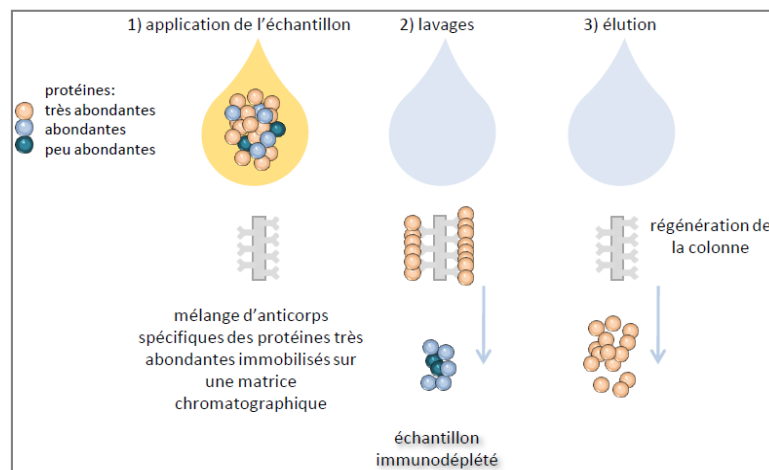


Figure 9 : Principe de l'immunodéplétion

Une autre stratégie utilisant différentes affinités de liaison aux protéines a été proposée : c'est la technologie d'«égalisation» (Thulasiraman, Lin et al. 2005) comme Proteominer (Bio-Rad) permettant de réduire l'abondance des protéines majoritaires et de concentrer les protéines minoritaires. Dans ce cas, l'échantillon est incubé avec des hexapeptides greffés sur des billes : les protéines majoritaires saturent leur peptide-ligand rapidement alors que les protéines minoritaires s'adsorbent, sans limite, sur les billes. L'excès des protéines majoritaires sera donc éliminé par lavage (Figure 10).

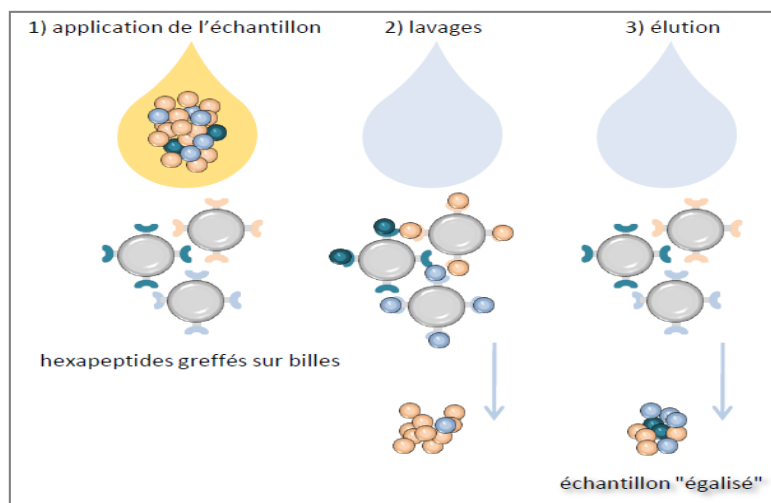


Figure 10 : Principe de la technologie « Equalizer »

Etant donné l'énorme complexité du protéome plasmatique, d'autres échantillons biologiques ont été envisagés pour la découverte de biomarqueurs, notamment les cellules en culture et leur sécrétome.

3) Découverte de biomarqueurs à partir des sécrétomes des modèles cellulaires

Le passage préalable par un modèle cellulaire constitue une approche pertinente afin de rechercher de marqueurs potentiels et de les valider dans des cohortes de plasmas de patients. Ces dernières années, un nouveau champ d'investigation a explosé : c'est l'analyse protéomique du sécrétome. Le terme « sécrétome » correspondant au surnageant de culture cellulaire, a été proposé par Tjalsma *et al.* (Tjalsma, Bolhuis et al. 2000). C'est un modèle séduisant car les protéines libérées dans le milieu extracellulaire sont souvent retrouvées dans le plasma. Il qualifie la composition des protéines sécrétées dans un environnement particulier à un instant donné.

Trois voies de sécrétion des protéines vers l'espace extracellulaire existent (Nickel and Rabouille 2009) (Figure 11):

- La voie classique dépendante du réticulum endoplasmique (RE)/Golgi : la plupart des protéines possédant un peptide signal en N-ter sont adressées au RE puis sont transportées vers l'appareil de Golgi dans des vésicules recouvertes de « Coat Protein Complex II » (COPII)
- La voie non classique des protéines possédant un peptide signal : certaines protéines possédant aussi un peptide signal traversent le RE et atteignent la membrane cellulaire mais sans passer par les vésicules recouvertes de COPII ou le Golgi
- La voie non classique des protéines ne possédant pas un peptide signal et indépendante de RE/Golgi : ces protéines, cytoplasmiques ou nucléaires, peuvent être relarguées sous l'effet du stress pour médier différentes fonctions extracellulaires.

L'approche protéomique du surnageant cellulaire intéresse surtout l'oncologie. Le sécrétome des cellules cancéreuses est une source prometteuse pour la découverte de biomarqueurs. Il a été défini comme le « réservoir de biomarqueurs » du cancer (Xue, Lu et al. 2008). Il a été démontré, par LC-MS/MS, que des protéines sécrétées par des cellules cancéreuses pulmonaires (*in vitro*) peuvent être détectées dans le plasma sanguin d'un modèle animal (*in vivo*) (Chenau, Michelland et al. 2009).

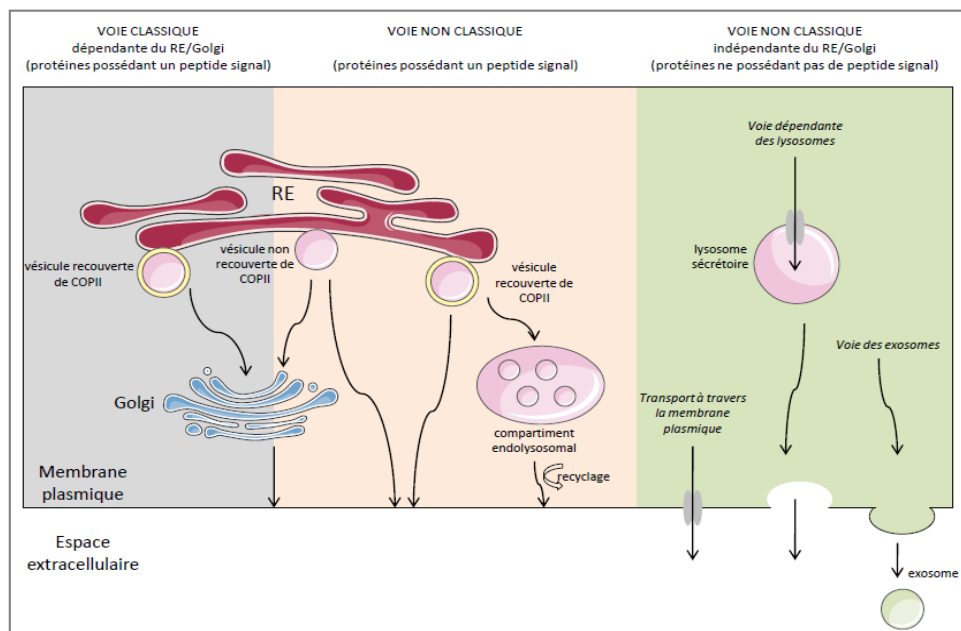


Figure 11 : Les 3 grandes voies de sécrétion des protéines

Plusieurs études ont utilisé l'approche protéomique du sécrétome pour découvrir de nouveaux biomarqueurs du cancer. La spectrométrie de masse a donc nettement amélioré les

capacités de détection, de quantification, et d'identification des biomarqueurs avec des analyses plus rapides et plus sensibles.

Xiao *et al.* ont analysé les protéines du sécrétome des cellules du cancer pulmonaire par SDS-PAGE et nanoLC-MS/MS. 13 protéines intéressantes ont été sélectionnées pour être validées dans le plasma des patients atteints du cancer pulmonaire par ELISA (Enzyme-Linked Immuno Sorbent Assay), une méthode de quantification ciblée très sensible utilisant des anticorps spécifiques et ayant une limite de détection des protéines inférieure à quelques ng/mL de plasma. Les protéines « Fascin, ezrin, Cd98, annexin A4, 14-3-3 ϕ , 14-3-3 β et 14-3-3 η » ont été détectées dans le plasma humain pour la première fois par ELISA (Xiao, Ying et al. 2005). Chen *et al.* ont analysé le sécrétome d'une lignée pulmonaire par SDS-PAGE. 12 protéines différenciellement exprimées ont été identifiées par la spectrométrie de masse en tandem. L-lactate dehydrogenase B chain (LDHB) a été trouvée associée au phénotype métastatique et a été ensuite détectée, par ELISA, à des taux élevés dans le sérum des patients atteints de ce cancer (Chen, Zhang et al. 2006). Planque *et al.* ont identifié, par 2D et LC-MS/MS, 5 protéines surexprimées dans le sécrétome de 4 lignées cancéreuses pulmonaires et les ont validées, par ELISA, dans le sérum des patients atteints d'un cancer pulmonaire (Planque, Kulasingam et al. 2009). Wu *et al.* ont étudié les sécrétomes de 21 lignées cellulaires dérivées de 12 types de cancer. Les analyses par SDS-PAGE et MALDI-MS ont révélé 320 protéines. Deux protéines, dont la CRMP-2 (collapsin response mediator protein 2), ont été validées par ELISA dans le plasma des patients en tant que biomarqueurs du cancer colorectal (Wu, Chen et al. 2008). Yamashita *et al.* ont identifié 86 protéines dans le sécrétome des cellules hépatiques HepG2 par 2D-LC-MS/MS. 10 protéines ont été trouvées comme de nouvelles protéines secrétées de ces cellules qui n'ont même pas été identifiées précédemment par des stratégies immunologiques (cathepsin H, cathepsin L, cyclophilin B, cystatin C, growth differentiation factor 15, nucleobindin 1, osteopontin, prosaposin, RNase 4, and stanniocalcin2). Ces protéines secrétées peuvent être des biomarqueurs pour le diagnostic et le traitement des maladies spécifiques du foie (Yamashita, Fujiwara et al. 2007).

La protéomique a permis la découverte de potentiels biomarqueurs ces dernières années. Cependant, les biomarqueurs découverts doivent passer par une phase de validation pré-clinique. Classiquement, un projet de recherche de biomarqueurs est constitué en trois étapes : la découverte de candidats potentiels, la validation des candidats choisis à l'étape de découverte et la confirmation des biomarqueurs à plus grande échelle pour des applications cliniques.

4) De la découverte à la validation d'un biomarqueur en passant par sa vérification

Un biomarqueur, à l'image d'un médicament, suit un certain nombre d'étapes pour être validé. Durant la phase de découverte de « candidats biomarqueurs », le taux de faux positifs est élevé surtout pour les protéines d'intérêt de faible abondance, et ceci est à cause des faibles intensités des signaux des peptides pendant l'analyse MS/MS. Une étape de vérification est donc nécessaire, elle met à profit la spectrométrie de masse pour faire le pont entre la découverte et la validation (Boja, Rivers et al. 2010).

Dans l'étape de « vérification », l'analyse est étendue à une centaine d'échantillons plasmatiques, et prend en compte des critères environnementaux, génétiques et biologiques pour grouper les échantillons tout en les classant en deux groupes : « cas » et « contrôles ». A travers cette étape de vérification, la sensibilité du candidat biomarqueur est confirmée (la capacité du test à être positif pour un échantillon malade), et sa spécificité commence à être évaluée (la capacité du test à être négatif pour un échantillon sain). Les candidats sont dits « vérifiés » quand l'expression différentielle est confirmée (Surinova, Schiess et al. 2010). C'est l'essor des approches protéomiques permettant la quantification des protéines par spectrométrie de masse qui a permis de vérifier les différences d'expression.

Les quelques « candidats biomarqueurs » vérifiés seront soumis à une phase de validation pré-clinique qui se fait sur plusieurs milliers d'échantillons plasmatiques (Rifai, Gillette et al. 2006). Cette phase repose sur l'ELISA comme le montrent les études décrites précédemment. Actuellement, la quasi-totalité des dosages de biomarqueurs protéiques se font par immunodosage pour des raisons de coût et de facilité de mise en œuvre. Une fois les biomarqueurs validés, généralement protégés par un brevet, entrent dans la dernière phase : phase d'évaluation clinique.

III) Analyse protéomique, réservoir de biomarqueurs

3.1) Introduction et stratégie générale

Le terme « protéome » a été introduit par Wilkins *et al.* en 1996 (Wilkins, Sanchez et al. 1996). Il l'a défini comme l'ensemble complet des PROTEINES codées par un gÉNOME. Aujourd'hui, le terme « protéomique » est utilisé pour l'étude globale de l'expression des protéines d'une cellule ou d'un organisme à un temps donné et selon un ensemble de conditions connues (Kenyon, DeMarini et al. 2002). Le but de la protéomique n'est pas limité à l'identification des protéines, elle permet aussi de créer un réseau multidimensionnel des protéines d'une cellule avec leurs localisations et interactions. L'analyse protéomique est donc

dynamique : un organisme possède une très grande diversité de protéomes qui sont directement liés à leur environnement dans lequel ils sont étudiés. Les protéines peuvent être sécrétées, synthétisées, dégradées, elles peuvent subir des modifications post-traductionnelles (MPTs) ou changer de compartiment cellulaire en réponse à des signaux internes ou externes.

Le concept de la protéomique a connu son véritable point de départ en 1975 quand O'Farrell a pu séparer 1100 composés protéiques par électrophorèse bidimensionnelle chez *Escherichia coli* (O'Farrell 1975). De même et à la même période, Klose et Scheele ont séparé les protéines de la souris et du cochon de Guinée par la même technique (Klose 1975, Scheele 1975). Néanmoins, à cette époque-là, les limites de la technologie ne permettaient pas d'identifier les protéines et donc, et l'électrophorèse bidimensionnelle n'était qu'une simple technique purement descriptive. C'est le développement simultané des techniques de séparation des protéines et des peptides et leur analyse par spectrométrie de masse qui ont permis l'essor de l'analyse protéomique. Il faut néanmoins souligner que les grandes performances de cette approche reposent sur l'ensemble des données acquises en génomique couplée au développement simultané de différentes techniques allant de la préparation de l'échantillon (fractionnement, gel d'électrophorèse, chromatographie liquide...) aux analyses bioinformatiques (banque de données génomique et protéique...) tout en passant par la spectrométrie de masse qui a été décrite dans la partie précédente.

Depuis 1995, le nombre de publications croît d'une manière exponentielle en protéomique : on peut dénombrer plus de 62 900 articles publiés en juillet 2013 (Figure 12).

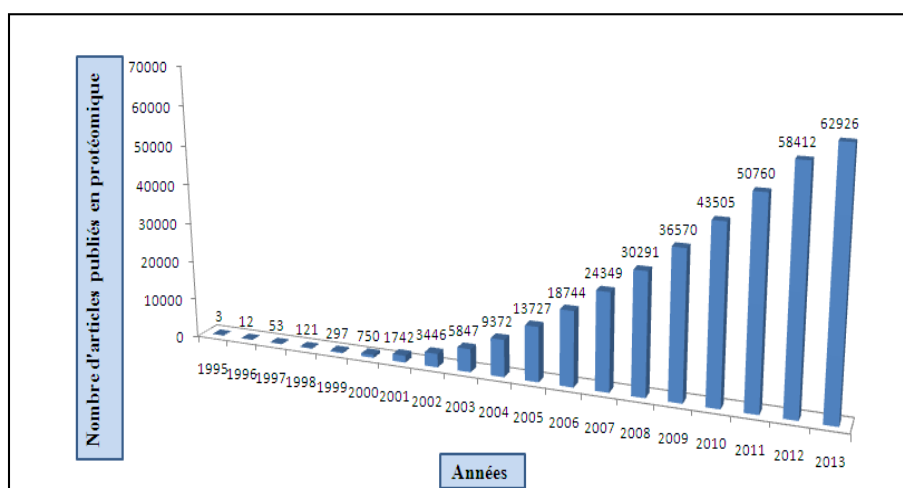


Figure 12 : Nombre d'articles publiés en protéomique de 1995 jusqu'au juillet 2013. Recherche réalisée dans PubMed en utilisant les mots clefs suivants : Proteomics/Proteomic/Proteome

Contrairement au génome restant constant dans les cellules, le protéome subit des modifications en réponse aux conditions environnementales. La protéine, produit final du gène, est plus représentative de la fonction d'une cellule que le gène lui-même. L'exemple le plus connu est celui exposé par Lottspeich : il prend la chenille comme exemple : avec les mêmes gènes, elle présente une morphologie différente selon son stade de développement en papillon (Lottspeich 1999). Ainsi, la complexité du protéome dépasse de très loin la séquence du génome.

A l'heure actuelle, le décodage du génome humain est constitué de 22 000 gènes.

La transcriptomique étudie l'expression globale des gènes. Mais l'analyse des transcrits n'est pas suffisante : les niveaux d'ARNm exprimés (acides ribonucléiques messagers) ne corrélient pas toujours avec les niveaux protéiques (Gygi, Rochon et al. 1999). La protéomique est donc différente, mais complémentaire de la génomique et de la transcriptomique (Figure 13). Les protéines ne sont pas isolées dans les cellules, elles interagissent avec plusieurs facteurs : on parle donc d'interactomique. Quant à la métabolomique, elle étudie l'ensemble des métabolites d'une cellule (sucres, lipides...).

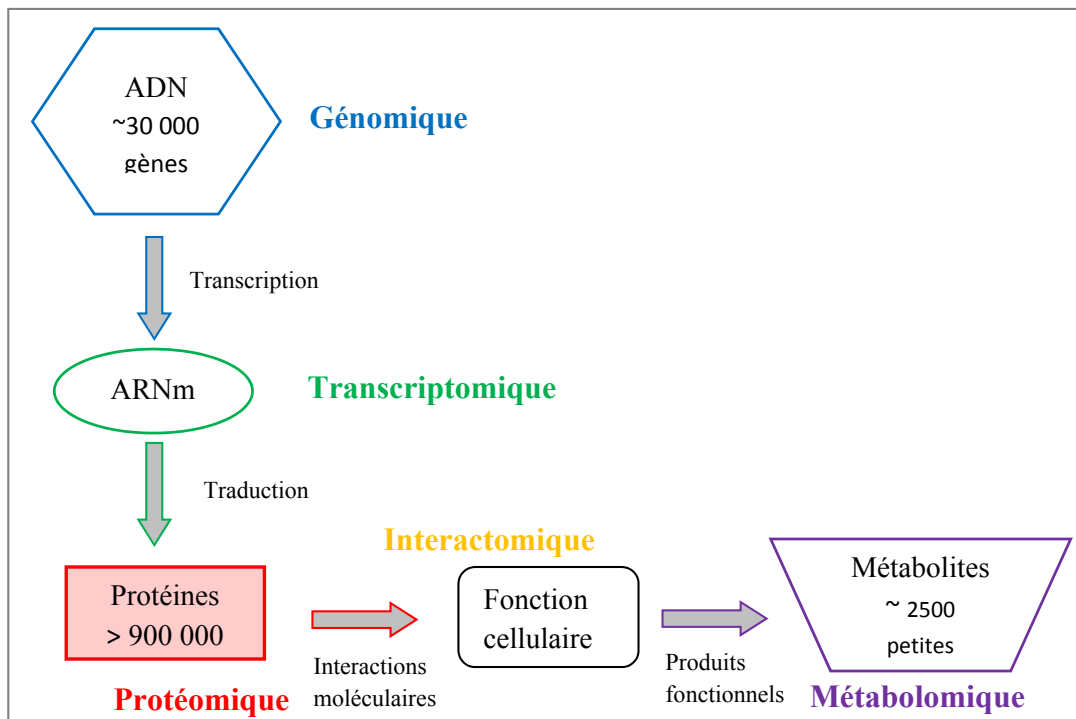


Figure 13 : Les « Omiques »

Dans ce travail, juste la partie protéomique sera décrite. Elle est schématisée dans la figure 14. L'étape de la préparation de l'échantillon protéique est l'étape la plus importante. Après cette étape, cet échantillon est fractionné grâce à des techniques séparatives de protéines, comme les gels d'électrophorèse à une ou deux dimensions, et de peptides, comme les

chromatographies liquides. Les peptides digérés par des protéases sont ensuite analysés par spectrométrie de masse qui permet l'identification des protéines. L'exploitation des spectres se fait ensuite grâce à des outils de bioinformatique très spécialisés.

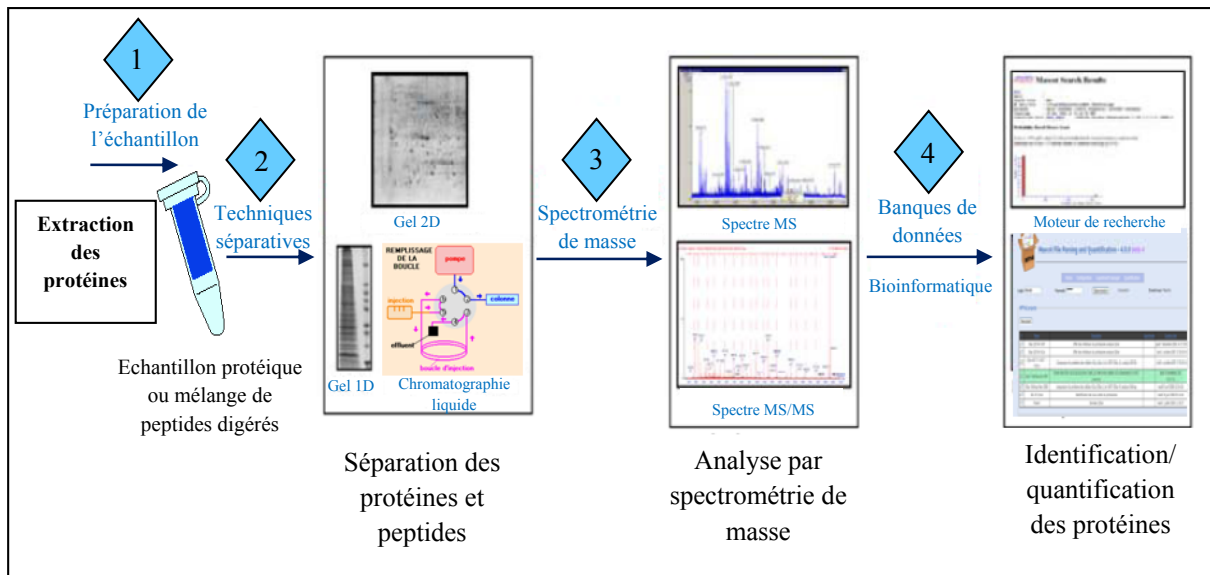


Figure 14 : Différentes étapes de l'approche protéomique

Deux stratégies d'analyse des protéines existent: le « top-down » et le « bottom-up » (Kelleher, Lin et al. 1999). Dans l'approche « top-down », des spectromètres de masse hautement résolutifs sont nécessaires pour l'identification des protéines par leur masse moléculaire (Chait 2006).

Quant à l'approche « bottom-up », elle consiste à passer par une étape de digestion des protéines soit après une étape de séparation des protéines, soit sans séparation préliminaire (« Shotgun »). Seule la stratégie « bottom-up » sera décrite dans la suite.

3.2) Outils pour l'étude protéomique

3.2.1) Spectrométrie de masse

Depuis une décennie, la spectrométrie de masse (mass spectrometry, MS) a effectué un bond prodigieux (prix Nobel de chimie en 2002) : elle occupe une place privilégiée parmi les techniques d'analyse. Elle est appliquée dans tous les domaines scientifiques : physique, chimie en phase gazeuse, chimie organique, biologie, médecine ... Que ce soit au niveau quantitatif ou qualitatif, ces diverses applications sont liées à ses deux caractéristiques : spécificité et sensibilité.

Les avancées qu'a connues la spectrométrie de masse depuis une vingtaine d'année permettent aujourd'hui l'étude de macromolécules complexes et la compréhension du vivant. Elle offre

actuellement la possibilité d'analyser des protéines, des polysaccharides ou des oligonucléotides, notamment l'enchaînement en acides aminés, en monosaccharides ou en nucléotides.

La MS est une méthode d'analyse permettant de détecter et d'identifier les molécules par mesure de leur masse, et de caractériser leur structure chimique. Le principe général de cette technique réside dans la séparation en phase gazeuse de molécules chargées, les ions, en fonction leur rapport masse sur charge (m/z).

Un spectromètre de masse comprend les éléments suivants (Figure 15) :

- ❖ Un système d'introduction de la substance à analyser
- ❖ Une source (sous vide ou à pression atmosphérique, par ionisation ou par désorption)
- ❖ Un ou plusieurs analyseurs (qui séparent les ions en fonction de leur rapport m/z)
- ❖ Un détecteur qui permet la détection des ions et l'envoi d'un signal électrique vers le système d'enregistrement de données
- ❖ Un système de pilotage de l'instrument et de traitements des données acquises fournissant le spectre de masse

Dans le cadre de ce travail relevant de la recherche de biomarqueurs d'exposition et d'effet à des cancérigènes de l'environnement, nous nous sommes dirigés vers deux techniques de désorption-ionisation dites douces : les sources d'Electrospray (ESI) (Fenn, Mann et al. 1989) et désorption/ionisation assisté par matrice (MALDI) (Karas and Hillenkamp 1988) qui sont adaptées à l'analyse de biomolécules.

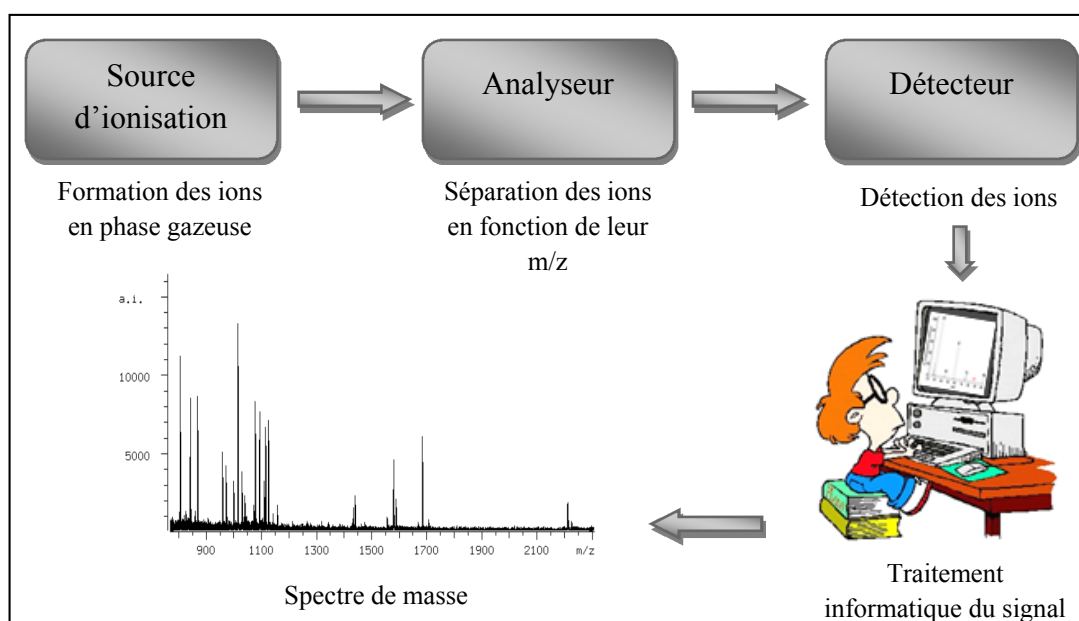


Figure 15 : Schéma du principe de fonctionnement d'un spectromètre de masse

3.2.1.1) Mesure de masse

Un spectre de masse donne l'abondance des ions en fonction de leur rapport m/z . Le type de masse mesuré par MS peut correspondre à la masse moyenne ou aux masses isotopiques : il dépend de la résolution de l'analyseur. Une résolution faible (de 1000 par exemple) ne peut pas résoudre les isotopes, un seul pic sera formé à partir de la combinaison des différents pics du massif isotopique : la masse mesurée correspond donc à la masse moyenne. Par contre, une résolution importante (de 10 000) permettra de mesurer la masse monoisotopique (Figure 16).

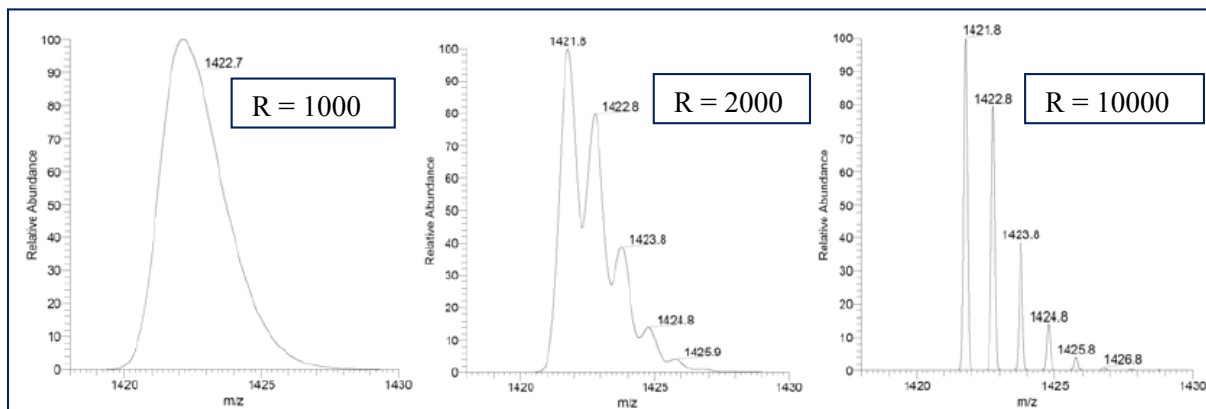


Figure 16 : Effet de la résolution ($R=1000, 2000$ et 10000) sur le spectre de masse pour un peptide de 12 acides aminés : RPRLSHKGPMMPF

Les isotopes sont des atomes ayant le même nombre de protons mais qui diffèrent par leur nombre de neutrons, ce qui entraîne une différence de masse atomique. L'écart entre chaque pic du massif isotopique est dû à la différence de masse des isotopes et à l'état de charge de la molécule, ce massif étant dépendant de la composition élémentaire de la molécule donnée. L'écart m/z entre chaque isotope pour un peptide monochargé sera donc de 1. S'il est dichargé, l'écart m/z sera de 0,5 et trichargé : 0,333.

3.2.1.2) Sources d'ionisation

Seules les formes ionisées des molécules en phase gazeuse peuvent être séparées et analysées selon leur rapport m/z . Pour passer les biomolécules présentes en solution à des ions en phase gazeuse, deux types de désorption-ionisation sont utilisées : MALDI et ESI.

3.2.1.2.1) Désorption-ionisation laser assistée par matrice (MALDI)

La méthode MALDI (Matrix Assisted Laser Desorption Ionisation), introduite en 1988 par Karas, Hillenkamp et Tanaka, permet l'analyse de biomolécules (biopolymères comme les protéines, les peptides, les sucres) et des grosses molécules organiques (polymères,

dendrimères) (Karas and Hillenkamp 1988, Tanaka, Waki et al. 1988). Elle consiste à irradier un dépôt cristallin d'une matrice organique co-cristallisant avec l'analyte.

L'ionisation par MALDI est provoquée par un faisceau laser. Ce phénomène comporte deux phases : tout d'abord, la substance à analyser, mélangée avec la matrice, est déposée sur un support métallique, la cible MALDI (Figure 17). Le solvant s'évapore, ce qui va rendre le dépôt matrice/analyte co-cristallisé. Puis la plaque MALDI sera transférée dans la chambre d'ionisation du spectromètre où le processus d'ionisation s'effectue sous vide. La surface cristalline de la matrice qui a piégé l'analyte sera irradiée par des impulsions laser intenses d'une durée de quelques nanosecondes : le laser le plus répandu à l'heure actuelle est le laser à azote irradiant dans l'ultraviolet à 337nm. Les ions sont donc produits en phase gazeuse sous l'effet de ces impulsions. Le faisceau laser excite et ionise la matrice par l'absorption d'un photon. C'est la première phase d'ionisation (Glückmann and Karas 1999). La matrice va alors se dissocier et se sublimer : l'analyte va être entraîné dans ce panache d'expansion où il va y avoir transfert de charge entre la matrice et l'analyte (Figure 18), c'est la deuxième phase d'ionisation (Knochenmuss and Zenobi 2003).

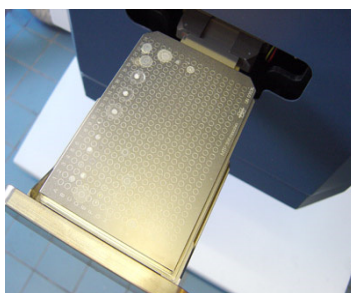


Figure 17 : Photographie d'une cible MALDI (Bruker)

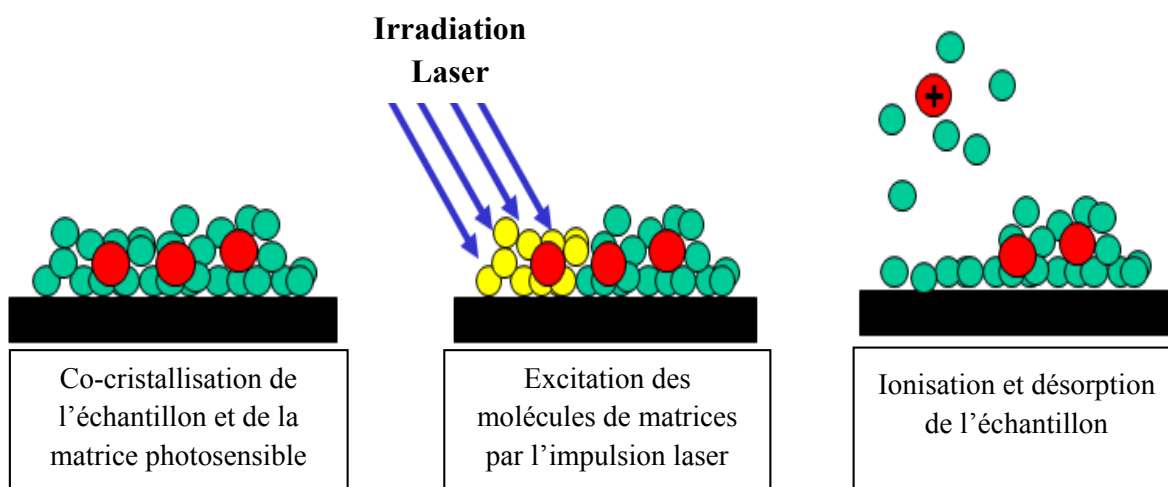


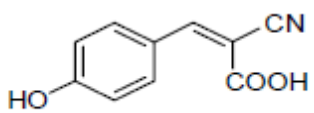
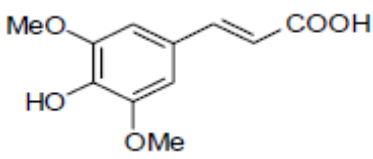
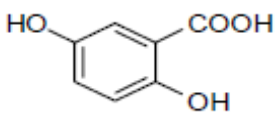
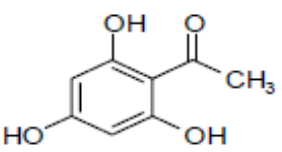
Figure 18 : Principe de la désorption/ionisation par MALDI

a- Choix et rôles de la matrice

Il est nécessaire de tester plusieurs matrices pour analyser chaque molécule parce qu'il n'existe pas une matrice universelle : celle donnant le meilleur résultat sera choisie. L'alpha-cyano-4-hydroxynamique est la matrice la plus utilisée pour les analyses protéomiques (Kusmann, Nordhoff et al. 1997). Les matrices les plus courantes sont répertoriées dans le tableau 2.

La matrice est une petite molécule capable d'absorber les radiations UV afin de protéger l'analyte de l'effet destructeur du faisceau laser. Elle empêche, avec son effet dénaturant, la formation d'agrégats entre les molécules à analyser, ce qui va permettre l'analyse des ions moléculaires.

Tableau 2 : Matrices utilisées en MALDI-MS

Nom de la matrice	Structure	Principale Application
Acide alpha-cyano-4-hydroxynamique (HCCA)		Protéomique (peptides et petites protéines)
Acide 3,5-dimethoxy-4-hydroxycinnamique (acide sinapinique)		Protéines entières
Acide 2,5-dihydrobenzoïque (DHB)		Phosphopeptides
2,4,6-Trihydroxyacetophenone (THAP)		Acides nucléiques

b- Mode de dépôt

Les différents types de dépôt doivent être testés afin de choisir le meilleur pour l'analyse d'une molécule. On distingue 3 modes de dépôt : goutte séchée, couche mince, sandwich.

b-1) Goutte séchée

Cette technique consiste à mélanger une solution de matrice à la solution contenant la substance à analyser. Une goutte de ce mélange sera déposée sur la cible MALDI et évaporée

à l'air ambiant ou sous vide pour cristallisation. Ce mode de dépôt est le plus rapide et le plus facile, par contre les cristaux générés sont grossiers et très hétérogènes, ce qui est un point négatif quant à la résolution et la reproductibilité des spectres.

Cette méthode est adaptée à l'analyse des peptides et petites molécules.

b-2) Couche mince

Le dépôt couche mince est réalisé en deux étapes. Dans un premier temps, une goutte de matrice est déposée sur la cible. L'évaporation du solvant de la matrice étant très rapide, une couche mince de cristaux homogènes sera donc obtenue, améliorant ainsi la résolution et la reproductibilité des spectres. Dans un deuxième temps, une goutte de la solution de l'analyte sera déposée sur la couche mince et évaporée à l'air.

Ce type de dépôt est très fragile, le nombre de tirs laser sera donc limité.

b-3) Sandwich

Cette méthode présente une étape supplémentaire à celle de la couche mince : elle consiste à recouvrir la solution d'analyte par une solution saturée de matrice tout en utilisant des solvants qui ne redissolvent pas la première couche de matrice déposée et séchée. Des cristaux très homogènes de petite taille seront obtenus.

Ce type de dépôt est adapté à l'étude des protéines et des macromolécules.

Le dépôt MALDI est facilement lavé puisque la matrice est peu soluble dans l'eau. Une goutte d'eau peut être ajoutée sur le dépôt, elle sera ré-aspirée ou chassée par un flux d'air : la technique MALDI-MS est donc très tolérante aux sels.

3.2.1.2.2) *Source par ionisation Electrospray*

En 1968, Dole avait décrit le phénomène d'ionisation par Electrospray (Dole, Mack et al. 1968). Le couplage de ce mode d'ionisation à un analyseur quadripolaire a eu lieu dans les années 80 par le groupe de Fenn (Yamashita and Fenn 1984) : il démontre que des composés, non volatils, pourront être ionisés et transmis en phase gazeuse, et ceci sans fragmentation ni destruction du composé ; et c'est cette découverte qui va entraîner l'application à l'ionisation de biomolécules par Electrospray. Cette technique a été aussi couplée à la chromatographie en phase liquide (LC-MS) (Whitehouse, Dreyer et al. 1985).

a- **Mécanisme d'ionisation Electrospray**

Cette technique d'ionisation, permettant d'obtenir des ions multichargés, génère, sous l'effet d'un champ électrique intense, des ions en phase gazeuse à partir d'analytes en solution. Ceci est décomposé en trois étapes :

- ❖ Production de gouttelettes chargées à partir de l'électrolyte en solution par l'application d'une haute tension sur un capillaire métallique.
- ❖ Fission des gouttelettes chargées en gouttelettes filles par explosions coulombiennes successives
- ❖ Emission des ions désolvatés en phase gazeuse

a-1) Production de gouttelettes chargées à partir d'une solution

Un champ électrique intense, de l'ordre de 10^6 V/m, est appliqué entre la pointe du capillaire métallique et une contre électrode, ce qui va entraîner la polarisation du liquide et la séparation des charges positives et négatives.

En mode d'ionisation positif, mode le plus utilisé pour l'étude des biomolécules, le capillaire qui est chargé positivement joue le rôle d'anode : il va donc attirer les charges négatives pour maintenir l'équilibre des charges et va les neutraliser par oxydation du solvant ou corrosion du capillaire (Mora, Van Berkel et al. 2000). Les charges vont donc s'accumuler à la pointe du capillaire. La surface du liquide se déstabilise en s'allongeant pour former un cône, « le cône de Taylor » qui va s'étirer afin d'émettre des gouttelettes minces contenant des ions positifs (Hayati, Bailey et al. 1987, Blades, Ikonomou et al. 1991) (Figure 19). Le phénomène de formation de gouttelettes chargées est assisté par un gaz coaxial (gaz de nébulisation, en général l'azote) afin de favoriser la stabilisation de la production de ces gouttelettes.

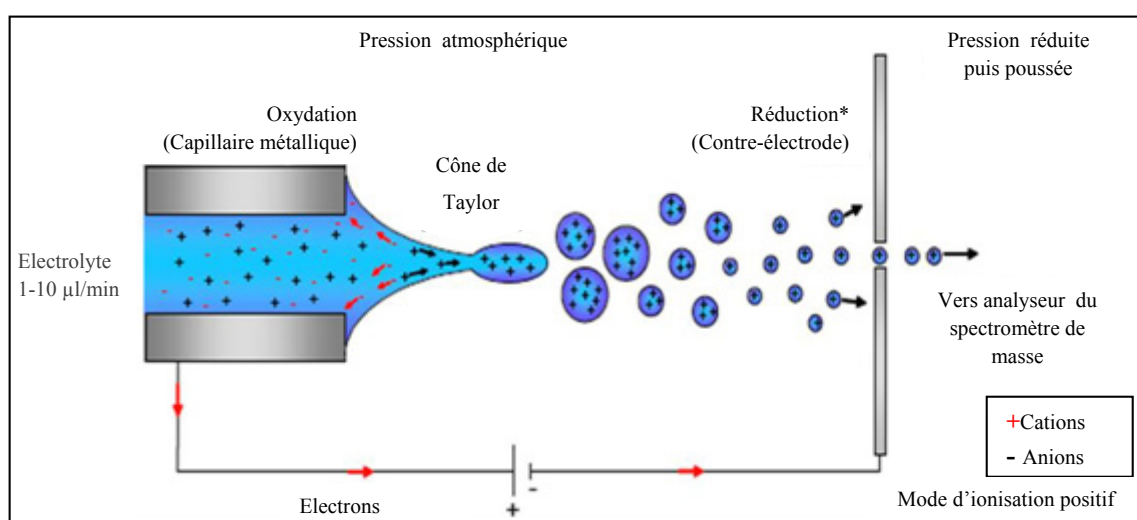


Figure 19 : Illustration schématique de la production de gouttelettes chargées

a-2) Fission des gouttelettes chargées : Explosion Coulombienne

Le solvant contenu dans les gouttelettes chargées, qui sont mises en contact avec un gaz chaud (l'azote), va s'évaporer progressivement, ce qui va entraîner une diminution de la taille des gouttelettes. Leur charge reste constante et leur densité de charge augmente jusqu'à atteindre « le rayon critique de Rayleigh » (les forces de répulsion électrostatiques sont égales aux forces de tension de surface). Lorsque les forces de répulsion coulombiennes l'emportent sur les forces de cohésion du liquide, la gouttelette instable va s'exploder en gouttelettes filles de densité de charge différente. Ces gouttelettes filles subiront le même processus qui se reproduit sur plusieurs générations jusqu'à l'émission des ions complètement désolvatés.

a-3) Emission des ions en phase gazeuse

Le principe d'émission des ions en phase gazeuse est controversé ; deux théories sont proposées : le modèle de Dole (modèle de charge résiduelle) (Dole, Mack et al. 1968) et le modèle d'Iribarne et Thomson (principe de l'évaporation ionique) (Iribarne and Thomson 1976).

- Modèle de Dole :

Après une succession d'explosions coulombiennes, une gouttelette fille ne contenant qu'un seul ion se forme. L'ion complètement désolvaté est généré par l'évaporation du solvant résiduel de cette gouttelette (Figure 20).

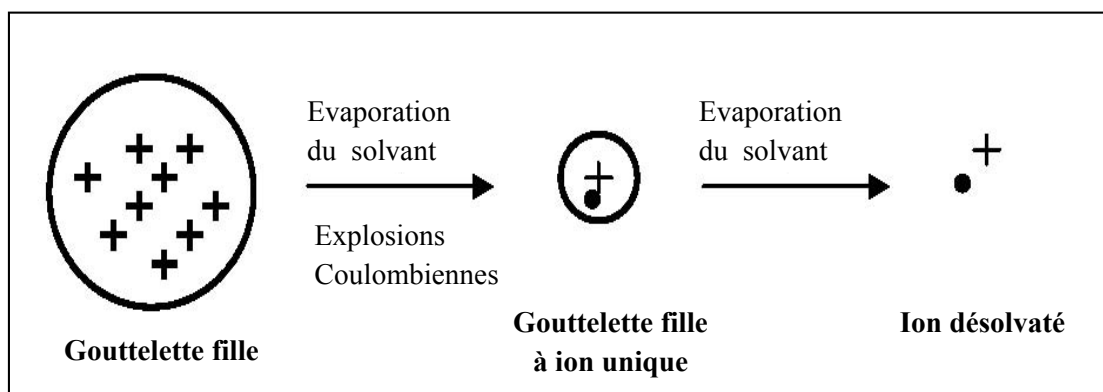


Figure 20 : Modèle de la charge résiduelle

- Modèle d'Iribarne et de Thomson :

Ce modèle propose qu'avant même que le rayon de la gouttelette n'atteigne la valeur du rayon de Rayleigh, sa densité de charge est suffisante pour émettre directement les ions en phase gazeuse (Figure 21).

Il est admis que les gros ions (macromolécules, protéines...) sont produits majoritairement d'après le modèle de Dole (Cole 2000, Mora, Van Berkel et al. 2000), et les molécules de petites tailles d'après le modèle d'Iribarne et Thomson (Kearle and Tang 1993, Gamero-Castano and Fernández de la Mora 2000). A l'heure actuelle, aucune preuve claire validant le premier ou le second modèle n'existe.

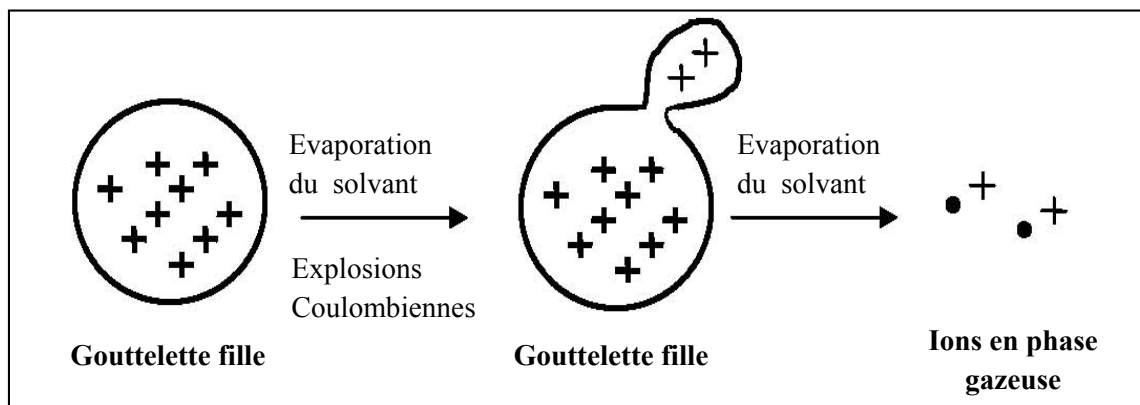


Figure 21 : Modèle d'évaporation des ions

b- Miniaturisation des sources électrospray en nanospray

La miniaturisation des sources électrospray a permis de diminuer le débit d'analyte délivré au niveau de la source. L'ionisation par électrospray se réalisait à des débits de 4 à 50 $\mu\text{L}/\text{min}$. En 1994, un dispositif micro-électrospray a été développée par l'équipe de Caprioli, avec un débit inférieur à 1 $\mu\text{L}/\text{min}$ (Emmett and Caprioli 1994). Puis, en 1996, l'équipe de Mann introduit la source nano-électrospray avec un débit de quelques centaines de nL/min : des gouttelettes de taille plus petite vont alors se former en introduisant une quantité d'analyte beaucoup plus faible (Wilm and Mann 1996). En diminuant le débit, la densité de charge de ces petites gouttelettes (diamètre environ de 200nm) va augmenter favorisant ainsi le rendement d'ionisation/désorption de l'analyte : la sensibilité de la mesure va donc s'accroître. L'assistance pneumatique contribuant à la désolvatation de l'échantillon n'est plus obligatoire dans ce cas.

L'analyse protéomique est une des applications de la source nanospray grâce à sa plus grande sensibilité et grande précision en mesure de masse. On distingue deux types d'ionisation nanospray :

- Nanospray offline : technique lente non automatisable nécessitant une analyse d'échantillon par échantillon
- Nanospray online : couplé à un système de chromatographie liquide (automatisable).

c- L'Interface ESI

L'interface assure la transmission des ions de la pression atmosphérique au niveau de la source à un vide poussé de l'ordre de 10^{-6} mbar au niveau de l'analyseur. L'accélération des ions est due aux différences de potentiel appliquées entre la source et les différentes lentilles successives de l'interface. Cette accélération va augmenter leur énergie cinétique. La pression au niveau de l'interface est suffisamment faible pour que les ions acquièrent une énergie cinétique. De même, elle est suffisamment élevée pour provoquer des collisions entre les ions et des molécules de gaz résiduel (azote ou solvant). Lors de ces collisions, l'énergie cinétique des ions se transforme en énergie interne, ce qui va contribuer à l'amélioration de la désolvatation des ions. Mais si l'énergie interne devient trop élevée, elle pourra provoquer la rupture des liaisons les plus fragiles de la molécule : c'est la décomposition induite par collision (CID : collision-induced dissociation). Donc il faudra bien contrôler, au niveau de l'interface, les paramètres de pression et de tension pour avoir la meilleure désolvatation.

3.2.1.3) Analyseurs de masse

Différents types d'analyseurs de masse et d'instruments hybrides peuvent être couplés aux sources d'ionisation ESI et MALDI.

On distingue deux catégories :

- les analyseurs à faisceaux d'ions : les quadripôles (Q) et les analyseurs à temps de vol (TOF : Time Of Flight)
- les analyseurs à piégeage d'ions : les analyseurs de type résonance cyclotronique ionique à transformée de Fourier (FT-ICR : Fourier Transform Ion Cyclotron Resonance), l'Orbitrap et les pièges à ions (IT : Ion Trap)

Ces analyseurs sont utilisés seuls ou assemblés (Q-TOF ou TOF-TOF) afin de former des hybrides possédant des avantages de chaque analyseur. Chaque analyseur possède des caractéristiques et performances analytiques propres à lui. Les principales caractéristiques d'un analyseur sont la résolution, la précision sur la mesure de masse, la sensibilité et la gamme de masse détectable, c'est-à-dire la gamme de rapports m/z qu'il peut identifier.

- La résolution représente la capacité de l'analyseur à séparer deux pics proches, c'est-à-dire deux ions de m/z voisins.

R (résolution) = $M/\Delta M$ correspond au rapport m/z mesuré (en Da)
et ΔM correspond à la largeur du pic à mi-hauteur

Si $M/\Delta M < 5000$: résolution faible

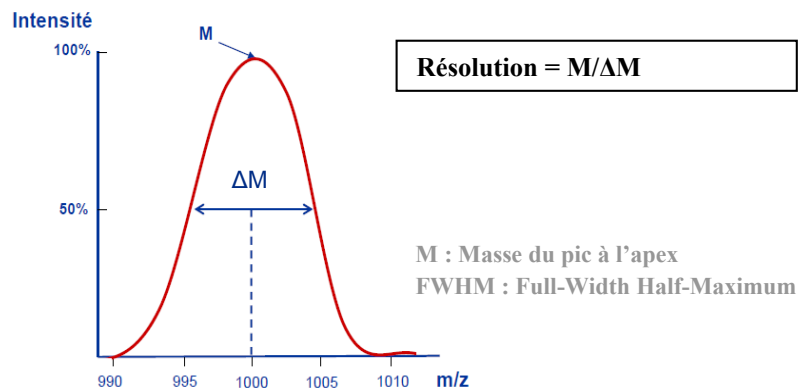


Figure 22 : Résolution d'un pic en spectrométrie de masse

Plus la résolution est élevée, plus le pic sera fin et plus la mesure en masse sera précise.

- La précision sur la masse de m/z représente la capacité du spectromètre à mesurer la masse la plus proche possible de la masse théorique. Elle est exprimée en ppm (partie par million). Elle est définie par le rapport : $(\Delta M/M) \times 10^6$, ΔM étant l'erreur commise sur la mesure de la masse théorique (M)
- La sensibilité correspond à la quantité minimale d'échantillon détectable

Les analyseurs sont également caractérisés par leur vitesse de balayage qui correspond au temps consacré pour l'établissement d'un spectre dans une gamme de m/z donnée.

Le tableau suivant compare les performances des analyseurs les plus courants :

Tableau 3: Principales performances et caractéristiques des analyseurs

Analyseur	Gamme de masse	Résolution	Précision sur la mesure de masse (ppm)
Quadripôle	10^3 à 10^4	10^2 à 10^3	100
Piège à ions	10^3 à 10^5	10^2 à 10^3	50 à 100
Temps de vol	$>10^5$	10^3 à 10^4	5 à 20
Résonance cyclotronique	10^4	10^4 à 10^6	0,1 à 5
Orbitrap	10^5	10^6	1 à 2

Dans ce travail, les analyseurs Q et TOF ont été principalement utilisés. Ils sont décrits dans la partie suivante.

3.2.1.3.1) Analyseur à temps de vol (TOF)

L'analyseur TOF est, par son concept de base, le plus simple des analyseurs. Il consiste à mesurer le temps que met un ion, accéléré par une tension électrique, à parcourir une distance donnée dans un tube de vol (région libre de champ) pour atteindre le détecteur. La mesure du temps de vol des ions est corrélée à leur rapport m/z . L'analyseur TOF est adapté aux sources pulsées (comme MALDI) puisqu'il nécessite que les ions soient produits par paquets successifs. Sous l'action du champ électrique, tous les ions de même charge ont la même énergie cinétique, ils vont se séparer en fonction de leur vitesse acquise dans la zone d'accélération. Plus l'ion sera léger, plus il sera rapide et arrive au détecteur plus rapidement qu'un ion lourd. Les analyseurs TOF ont une très haute sensibilité puisque tous les ions sont détectés. Mais la limitation de cette technique à ses débuts était sa faible résolution à cause de la dispersion énergétique des ions à l'entrée de l'analyseur.

En mode de détection linéaire, la résolution est faible, de l'ordre de 5000. Ce mode de détection n'est pas limité en m/z . Il est adapté particulièrement à l'étude des molécules de haute masse moléculaire comme les protéines entières.

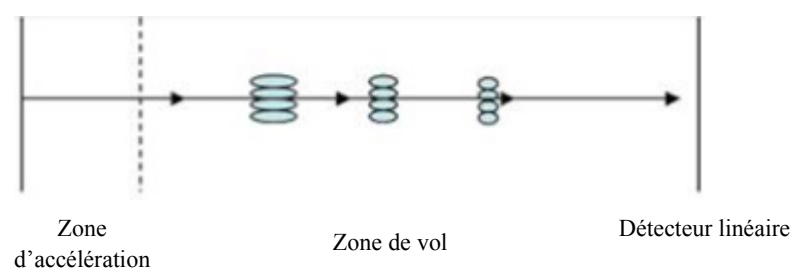


Figure 23 : Trajet d'un ion en mode linéaire dans l'analyseur TOF

Le mode linéaire n'est pas assez résolu pour identifier une protéine à partir de ses peptides tryptiques. Pour corriger les dispersions énergétiques, deux développements ont permis d'améliorer les performances instrumentales : l'extraction retardée et l'utilisation d'un réflectron (Cotter 1992). L'extraction retardée des ions consiste à homogénéiser les énergies cinétiques des ions dans la source avant leur injection dans l'analyseur. Le mode réflectron correspond à l'interposition de miroirs électrostatiques sur le trajet des ions. Le miroir électrostatique impose un champ électrique de direction opposée à celle du champ accélérateur initial. La taille de l'analyseur restant la même, c'est la distance de vol qui va s'allonger (Figure 24). Les ions vont donc mettre plus de temps pour arriver au détecteur. La dispersion a pu être réduite par une refocalisation des ions sur les plans temporel et énergétique, ce qui va améliorer

la résolution ($R=20\ 000$). Ce mode est plus adapté à l'analyse des peptides. Mais il est limité à des m/z allant jusqu'à 10 000.

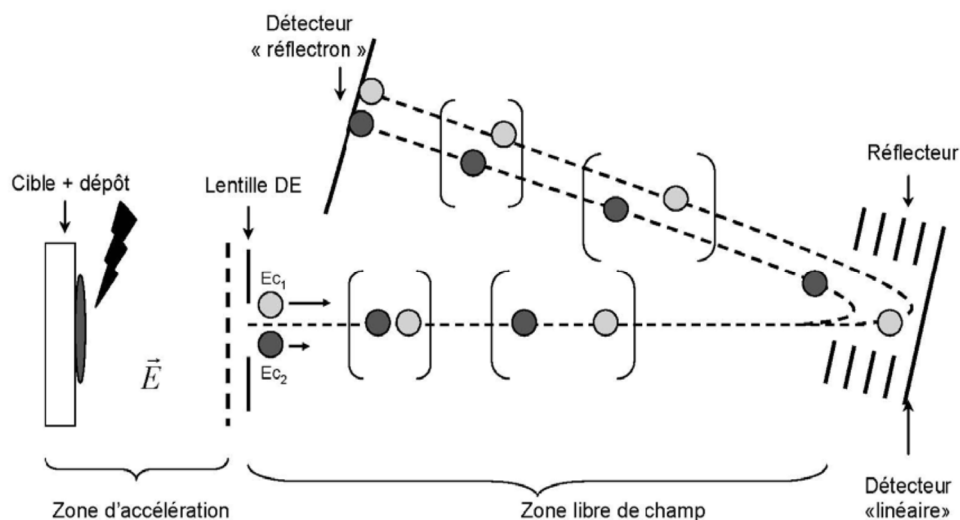


Figure 24 : Représentation schématique d'un couplage MALDI-TOF en mode réflecteur comportant un système d'extraction retardée (DE)

3.2.1.3.2) Le quadripôle (Q)

Un quadripôle est formé de deux paires de barres conductrices parallèles de section hyperbolique ou circulaire jouant le rôle d'électrodes. Il fonctionne comme un filtre en ne laissant passer à un temps donné que des ions d'un rapport m/z déterminé.

Les barres sont reliées électriquement deux à deux de manière à créer un champ quadripolaire : elles sont portées à des potentiels opposés deux à deux (Φ_0 , et $-\Phi_0$) sous la forme d'une tension électrique constituée d'une composante alternative haute fréquence V et d'une tension continue U .

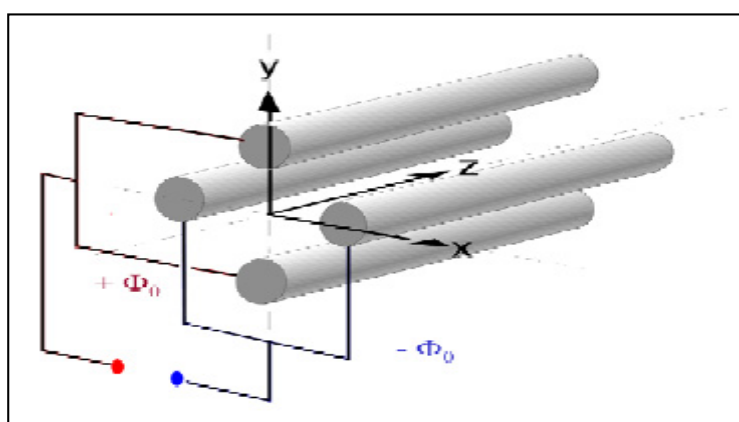


Figure 25 : Principe d'un quadripôle à section cylindrique

Les ions accélérés suivant l'axe z du quadripôle pénètrent entre les quatre barres et conservent leur vitesse selon cet axe. Ils sont soumis à des accélérations résultantes des forces

dues aux champs électriques suivant les axes x et y. Selon des valeurs précises de U et V, seuls les ions avec un rapport m/z donné auront une trajectoire stable et oscillante et pourront donc franchir le filtre et atteindre l'extrémité des quadripôles pour se diriger vers le détecteur. Les autres ions se déchargent par contact sur les barres du quadripôle et ne sont pas détectés (Figure 26). Le mouvement d'un ion de rapport m/z donné, autrement dit, la stabilité de sa trajectoire est décrite par les équations de Mathieu qui les a établies en 1866 pour décrire la propagation d'ondes dans des membranes.

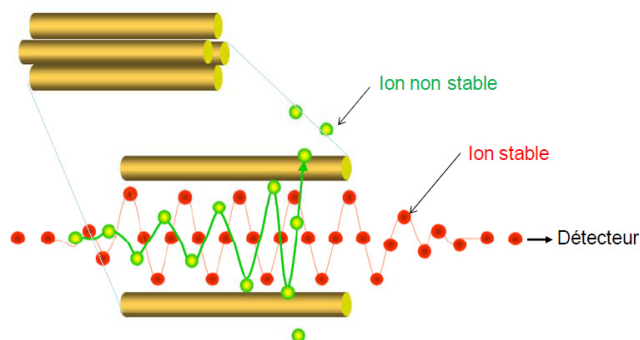


Figure 26 : Trajectoire oscillante stable des ions détectés

Le spectre de masse sera alors obtenu en variant U et V selon un mode de balayage adéquat pour faire défiler les ions les uns après les autres dans le détecteur. Le mode Full Scan permet de détecter tous les ions produits dans la gamme m/z de l'analyseur. Le mode SIM (Single Ion Monitoring ou SIR pour Single Ion Recording) existe aussi, il assure une meilleure sensibilité puisqu'il ne détecte que les rapports m/z sélectionnés, caractéristiques de l'analyte.

3.2.1.3.3) Les analyseurs hybrides

Les appareils hybrides présentent des propriétés bien supérieures aux analyseurs simples puisqu'ils permettent de combiner les avantages de chaque analyseur. Plusieurs combinaisons sont possibles, mais seuls TOF-TOF, Q-TOF et QqQ (triple quadripôle) seront spécifiquement développés dans cette partie.

3.2.1.3.3.1) TOF-TOF

Décrit pour la première fois en 2000 (Medzihradsky, Campbell et al. 2000), le TOF-TOF a été principalement utilisé pour le séquençage des peptides et des protéines en analyse protéomique et pour la caractérisation des modifications post-traductionnelles (Yergey, Coorsen et al. 2002). Les avantages du TOF-TOF, en tant que résolution et sensibilité, sont les mêmes que ceux du Q-TOF. C'est la nature des ions fragments formés qui diffère entre les deux appareils hybrides. Les collisions à haute énergie génèrent des clivages internes dans le cas du TOF-TOF, qui ne sont pas observés en Q-TOF : les spectres MS/MS sont plus complexes, ce

qui va les rendre plus informatifs (permettant de distinguer par exemple une leucine d'une isoleucine (Medzihradzsky, Campbell et al. 2000)).

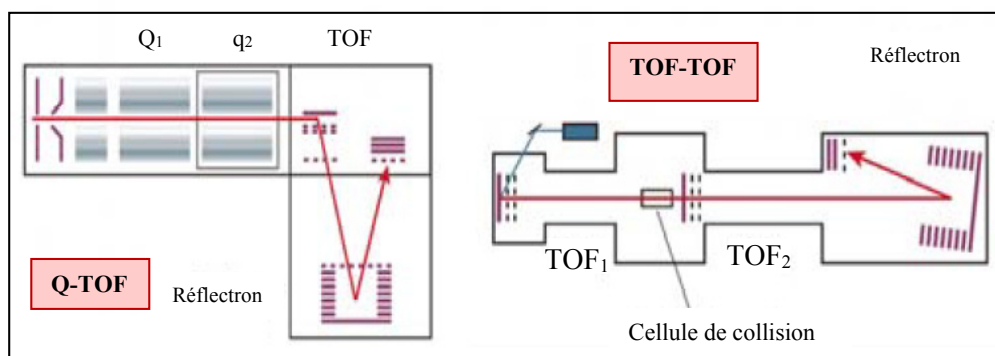


Figure 27 : Appareils hybrides

3.2.1.3.3.2) *Q-TOF*

Le principe du Q-TOF a été revu en 2001 (Chernushevich, Loboda et al. 2001). Le quadripôle et le TOF, séparés par la cellule de collision, sont perpendiculaires. Les ions fragments sont focalisés et envoyés par paquets dans le TOF par un système de lentilles qui permet le ralentissement ainsi que la refocalisation des ions sortant de la cellule de collision. Une impulsion est nécessaire aux ions pour les réaccélérer dans une direction perpendiculaire, elle sera donc stimulée par le "pusher" situé à l'entrée du TOF. Les ions générés en fragments sont ensuite injectés dans le TOF.

Des sensibilités de l'ordre de l'attomole peuvent être atteintes puisque ce Q-TOF combine la simplicité d'un quadripôle avec la sensibilité, la résolution et la précision des mesures de masse d'un réflectron-TOF (Kristensen, Imamura et al. 2000).

Cet appareil est très adapté au séquençage de peptides (analyses en mode MS/MS) (Shevchenko, Wilm et al. 1997), à l'identification des modifications post-traductionnelles et aux analyses quantitatives. Son inconvénient est qu'il ne peut pas faire de balayage sur des ions spécifiques puisque le TOF mesure le temps de vol de tous les ions donc sur toute la gamme de masse et ceci sans moyen de contrôle.

3.2.1.3.3.3) *QqQ*

Le triple quadripôle (TQ ou QqQ) permet de faire de la spectrométrie de masse en tandem (MS/MS) (Figure 28).

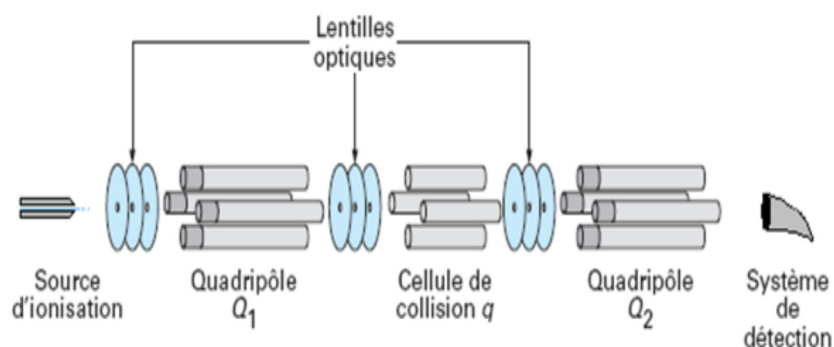


Figure 28: Représentation d'un triple quadripôle

Il est formé de deux analyseurs (quadripôles) en série séparés par une cellule de collision contenant un gaz de collision inerte. Les quadripôles Q_1 et Q_2 fonctionnent comme des filtres de masse qui, selon le mode de fonctionnement, laissent passer tous les ions ou certains ions présélectionnés. Q_1 sélectionne l'ion parent. Après dissociation de cet ion par collision, les ions fils ou fragments sont analysés par balayage sur le second analyseur Q_2 . L'avantage de ce couplage est l'augmentation de la sélectivité de deux filtres en série. C'est une technique très sensible utilisée pour la quantification.

Parmi les différents modes d'acquisition disponibles lors du couplage de la chromatographie liquide avec un spectromètre du type triple quadripôle, on cite le mode SRM (*Selected Reaction Monitoring*) qui permet de détecter un ion fragment provenant d'un ion précurseur donné. Il permet d'obtenir un gain en sensibilité en diminuant le bruit de fond et d'avoir un mode de détection très sélectif pour l'identification des protéines.

3.2.2) Spectrométrie de masse en mode tandem

3.2.2.1) Principe

La découverte de la spectrométrie de masse en tandem (MS/MS ou MS^2) a représenté une révolution dans les milieux analytiques. Elle combine deux analyseurs et permet de réaliser le séquençage des peptides d'intérêt.

Les appareillages construits pour réaliser ces expériences ont été appelés "tandem" car ils sont conçus de la manière suivante: une première étape (MS1) consistant à sélectionner l'ion « parent » précurseur (un ion stable particulier issu de la source d'ionisation) dans le premier analyseur et une seconde étape (MS2) représentant l'analyse des ions « fils » ou « produits » issus de sa décomposition dans un second analyseur. Dans l'intermédiaire se situe la cellule de collision contenant des molécules sous forme gazeuse et permettant l'activation de l'ion précurseur qui conduit à son excitation (augmentation de son énergie interne) et à sa fragmentation (Figure 29).

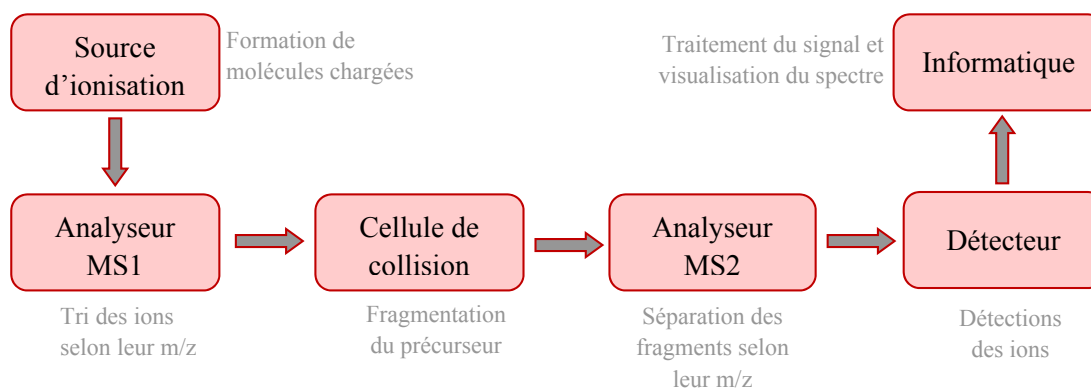


Figure 29 : Spectrométrie de masse en mode tandem

3.2.2.2) Fragmentation peptidique

La molécule à étudier doit subir une fragmentation d'une ou plusieurs liaisons afin d'obtenir des informations de séquences peptidiques : la masse des fragments obtenus sera reliée à la structure chimique des peptides.

La méthode la plus utilisée permettant de transférer aux ions stables, produits lors de l'ionisation, de l'énergie nécessaire à la fragmentation est la dissociation induite par collision (CID : collision induced dissociation ou CAD : collision activated dissociation).

Deux types de collision CID existent :

- Les fragmentations de basse énergie (quelques électron-Volts (eV)) : ils ont lieu, par exemple, dans les analyseurs de type quadripôle
- Les fragmentations de haute énergie (quelques kilo-eV) : analyseurs type TOF-TOF.

3.2.2.2.1) Fragmentation de basse énergie

Un peptide est un copolymère d'acides aminés reliés par des liaisons peptidiques $-\text{CO}-\text{NH}-$ (liaisons amides). Il possède deux extrémités : N-terminale ($-\text{NH}_2$) et C-terminale ($-\text{COOH}$) (Figure 30).

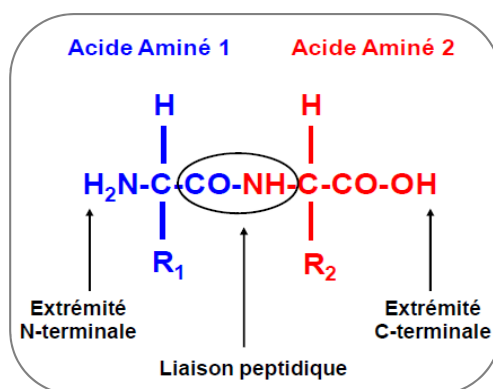


Figure 30 : Structure chimique d'un dipeptide

Lors de l'analyse MS/MS, les peptides se fragmentent majoritairement au niveau des liaisons peptidiques. Une nomenclature des ions fragments a été proposée par Roepstorff et Fohlman (Roepstorff and Fohlman 1984) puis elle a été modifiée par Biemann (Biemann 1992) (Figure 31). La fragmentation de la liaison $-\text{CO}-\text{NH}-$ aboutit à la formation des ions **b** et **y** :

- **b** si la charge de l'ion est conservée par le fragment N-terminal du peptide
- **y** si la charge est conservée par le fragment C-terminal

D'autres fragments peuvent aussi se former : ce sont les ions **a** et **c** (N-terminal), **x** et **y** (C-terminal) qui sont obtenus selon le lieu de fragmentation par rapport à la liaison peptidique. Les fragments N et C-terminaux sont notés avec un indice qui précise le nombre de résidus portés par l'ion fragment.

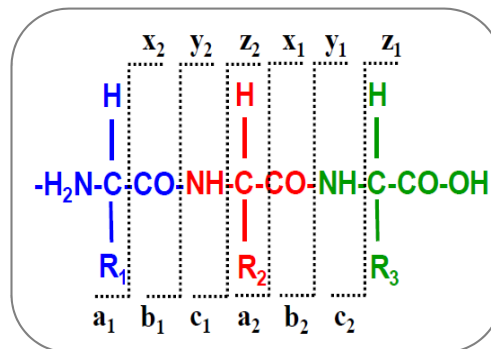


Figure 31 : Fragments des peptides d'après Biemann

Les peptides les plus couramment analysés en protéomique sont les peptides tryptiques : dans ce cas, la majorité des ions formés par ESI-MS sont dichargés, ils vont générer essentiellement des ions **y** et/ou **b** monochargés (Steen and Mann 2004).

La double coupure (fragmentation à la fois en C- et en N-terminal des acides aminés) produit des ions immonium (Figure 32). Ce sont des fragments internes retrouvés dans la région de faible m/z du spectre MS/MS.

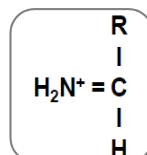


Figure 32 : Structure chimique d'un ion immonium

La séquence du peptide est déduite par la différence de masse entre deux ions consécutifs d'une même série (sauf les ions isobares leucine/isoleucine et les non-isobares

Glutamine/Lysine ayant le plus faible intervalle de masse existant entre deux acides aminés, de l'ordre de 0,0363 Da).

3.2.2.2.2) Fragmentation de haute énergie

Deux types d'ions peuvent être générés à haute énergie:

- Les ions qui portent la charge positive au niveau de l'acide aminé N-terminal : ce sont les ions **d**. La coupure de la chaîne latérale se fait après le carbone β .
- Les ions qui portent la charge positive au niveau de l'acide aminé C-terminal : ce sont les ions **v** (coupure de la chaîne latérale après le carbone α) et **w** (coupure après le carbone β).

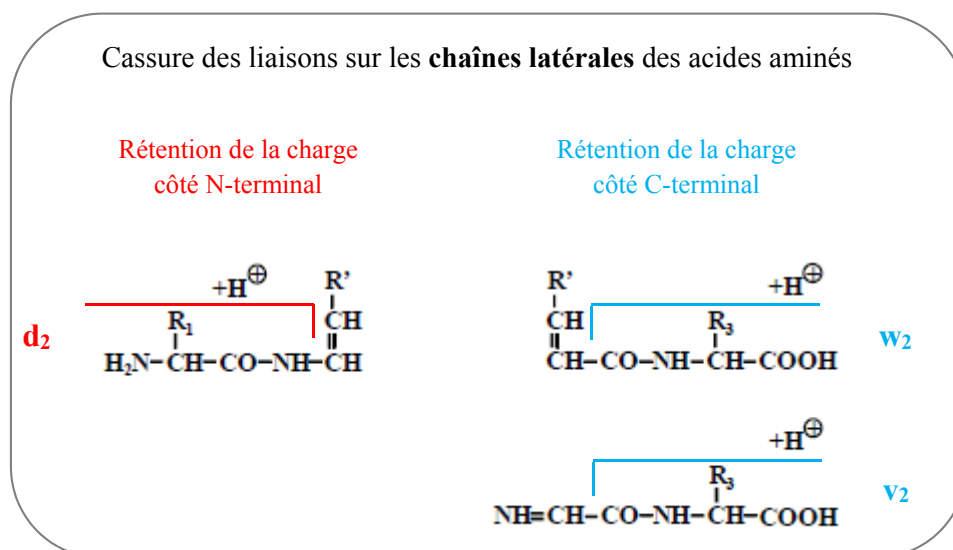


Figure 33 : Nomenclature des ions fragments des peptides (haute énergie)

3.3) Fractionnement du protéome

L'échantillon biologique à analyser peut être soit un protéome entier (cellules), soit un sous-protéome (phosphoprotéome, compartiments cellulaires, sécrétome, ...) soit des protéines isolées. La protéomique doit caractériser l'ensemble de protéines de cet échantillon. Cet objectif s'oppose à des limites de biologie qui constituent des défis à relever : la dynamique de concentration, la complexité qualitative et l'impureté de l'échantillon biologique qui est contaminé par des macromolécules biologiques comme les acides nucléiques, les lipides, ou des interférents chimiques comme les sels et les détergents. De cette façon, l'ajout d'un sérum du veau fœtal au milieu de culture cellulaire, *in vitro*, compromet la pureté du sécrétome.

Il est indispensable que la préparation de l'échantillon soit compatible avec les techniques de séparation des protéines utilisées ultérieurement.

A un échantillon peuvent être appliquées plusieurs stratégies de fractionnement : il faut prendre en compte différents paramètres physico-chimiques (point isoélectrique, poids

moléculaire...) des protéines intactes ou des peptides tryptiques. Ce qui va permettre de concentrer les analytes dans une fraction particulière facilitant ainsi la détection des protéines ou peptides les moins représentés dans l'échantillon.

La combinaison de plusieurs séparations améliore la couverture protéomique : aucune stratégie de fractionnement n'est privilégiée. La séparation la plus traditionnelle des protéines est l'électrophorèse bidimensionnelle (2-DE : Two-dimensional electrophoresis). Alors que le fractionnement des peptides tryptiques a été favorisé par la miniaturisation et le couplage des systèmes de chromatographie liquide à haute performance aux spectromètres de masse.

3.3.1) Séparation des protéines : Electrophorèse

L'électrophorèse à une seule dimension (1D) ou SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) sépare, sous l'influence d'un champ électrique, les protéines en fonction de leur poids moléculaire (PM), tandis que l'électrophorèse bidimensionnelle permet la séparation des protéines selon leur point isoélectrique (pI) par isoélectrofocalisation (IEF) en première dimension puis selon leur PM par SDS-PAGE.

La 2-DE est une approche puissante à haute résolution permettant de séparer des milliers de protéines sur un même gel : elle a longtemps été la méthode unique pour l'analyse du protéome (O'Farrell 1975) et reste l'approche séparative la plus adaptée à l'identification des protéines par spectrométrie de masse (MALDI-TOF). Elle permet de fournir une carte protéique où chaque protéine apparaît sous la forme d'un « spot » (tâche) après sa migration, et de différencier, qualitativement et quantitativement, deux ou plusieurs échantillons protéiques (Rabilloud 2002).

Malgré sa robustesse et sa performance, cette technique possède quelques limites :

- séparation de certaines catégories de protéines « extrêmes » ayant un pI inférieur à 3 ou supérieur à 9, et un pM < 10 kDa ou > 150 kDa. Ces protéines sont mal résolues par cette technique classique
- Technique difficile à automatiser et à reproduire malgré les gels pré-coulés facilitant la tâche)
- Problèmes de solubilité des protéines qui découlent de leur hétérogénéité chimique : la détection des protéines les plus hydrophobes, ayant tendance à précipiter dans la première dimension, est défavorisée malgré l'optimisation des protocoles de solubilisation des protéines à l'aide de détergents compatibles (Chevalier 2010)

- Dynamique de l'expression protéique : il est difficile de visualiser les protéines faiblement représentées dans un extrait total, leur détection dépend aussi de la méthode de révélation utilisée. Les colorations fluorescentes et à l'argent opèrent jusqu'à 0,5ng de protéines/spot (Rabilloud, Chevillet et al. 2010). L'apparition des gels avec un gradient de pH restreint (« les zoom gels ») offre une plus grande résolution permettant ainsi la détection des protéines faiblement abondantes dans cette zone de pH réduite.

En revanche, l'électrophorèse bidimensionnelle demeure une méthode très attractive pour différencier les protéines modifiées post-traductionnellement.

3.3.2) Séparation des peptides : techniques chromatographiques

Grâce à l'avènement de l'électrospray, les analyses par spectrométrie de masse ont pu être appliquées aux protéines. Mais en raison de la complexité de l'échantillon, d'autres molécules présentes en solution peuvent aussi s'ioniser ce qui va diminuer le rendement d'ionisation de l'analyte : c'est le cas de la suppression d'ionisation.

Pour pallier ces problèmes, la chromatographie en phase liquide à haute performance (HPLC : High Performance Liquid Chromatography) sera l'outil de choix grâce à sa résolution, son fort pouvoir séparatif et ses possibilités de couplage direct, via la source ESI, aux spectromètres de masse. Ce couplage est en fort développement pour la séparation et l'analyse des peptides.

La séparation de petites molécules par chromatographie repose sur le partage de celles-ci entre une phase stationnaire et une phase mobile. Différentes techniques de LC existent ; elles sont dépendantes de la nature de la phase stationnaire utilisée. Chacune de ces phases stationnaires permet la séparation des analytes en fonction des paramètres physico-chimiques particuliers. En analyse protéomique, la chromatographie d'adsorption en phase inverse (RPLC, Reverse Phase Liquid Chromatography) est la plus utilisée pour le couplage avec la spectrométrie de masse : elle permet de travailler avec des pourcentages importants de solvants organiques favorisant ainsi la bonne ionisation des molécules dans la source électrospray. Elle sépare les peptides, issus de la digestion trypsique des protéines préalablement séparées, selon leur taille et leur hydrophobicité (Shi, Xiang et al. 2004). Les peptides retenus dans la colonne en fonction de leur hydrophobicité au cours du gradient chromatographique sont élués à un temps de rétention donné (RT pour Retention Time) sous forme d'un pic chromatographique détecté par le spectromètre de masse.

Les avantages de l'utilisation de la LC en amont de l'analyse MS sont :

- fractionnement de l'échantillon limitant ainsi la compétition entre les peptides au moment de l'ionisation ainsi qu'au moment de la sélection des ions pour la fragmentation MS/MS.
- diminution de l'effet de suppression d'ion
- réduction du nombre d'ions analysés simultanément par le spectromètre de masse.

3.4) Analyse bioinformatique des données MS/MS

Au fur et à mesure de leur élution au cours du gradient chromatographique, les peptides sont ionisés et injectés dans le spectromètre de masse pour être analysés en MS générant ainsi des spectres MS qui représentent l'empreinte massique de l'ensemble des peptides élués à un instant donné (Figure 34). Lors d'un scan MS, les ions détectés sont les plus intenses, chacun d'entre eux va être sélectionné et fragmenté pour être séquencé. De ce fait, on obtient des spectres MS/MS contenant les différents fragments de l'ion parent.

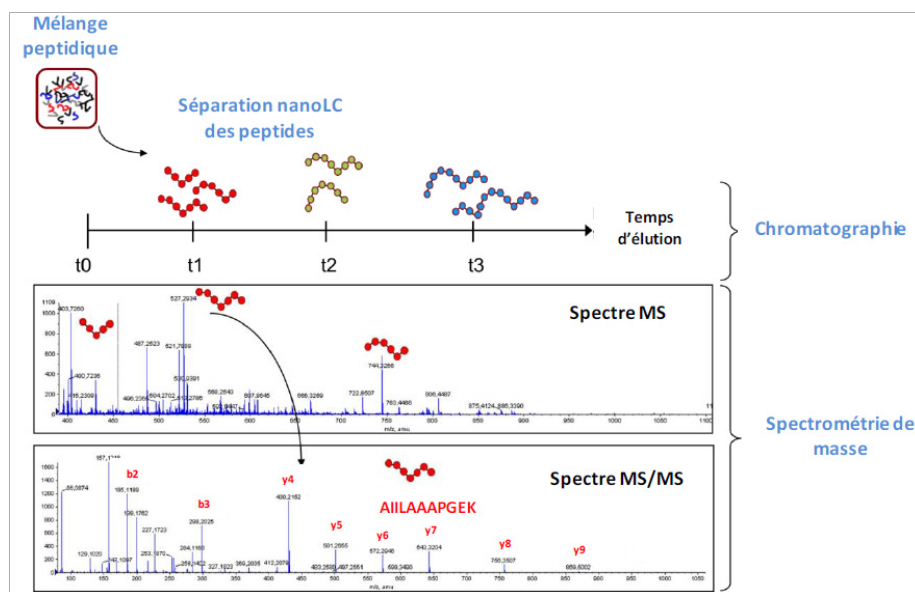


Figure 34: Vue d'ensemble du couplage nanoLC-MS/MS

Un logiciel de pilotage du spectromètre de masse permet le passage automatique du mode MS au mode MS/MS : il faut qu'on définisse des critères dans la méthode d'acquisition. Avant toute programmation, il faut tenir compte du réglage des paramètres de spray et de la vitesse de séparation chromatographique (temps moyen d'élution d'un peptide = 30 secondes).

Parmi les nombreux moteurs de recherche utilisant cette approche, l'on cite MASCOT (Perkins, Pappin et al. 1999), OMSSA (Geer, Markey et al. 2004) ou SEQUEST (Eng, McCormack et al. 1994). Ces moteurs de recherche permettent d'identifier les séquences

peptidiques par comparaison entre les spectres MS/MS expérimentaux bruts (peak list) et des spectres MS/MS théoriques issus de la digestion *in silico* de l'ensemble des protéines présentes dans une banque (Figure 35). C'est l'approche PFF (Peptide Fragment Finger printing). L'ensemble des peptides identifiés permet de remonter à l'identité des protéines initialement présentes dans le mélange complexe.

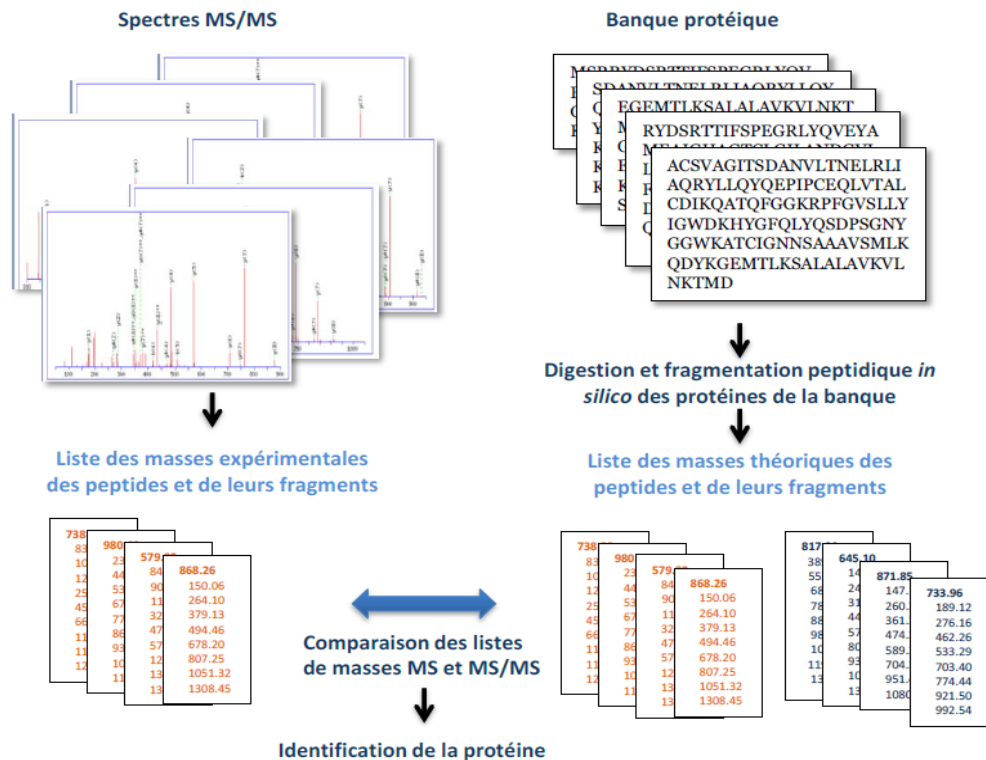


Figure 35: Principe d'identification des protéines par interrogation de banques protéiques

Les moteurs de recherche effectuent cette corrélation entre les spectres expérimentaux et théoriques en tenant compte de plusieurs paramètres qui sont réglés par l'utilisateur :

- Base de données des protéines
- L'enzyme de digestion utilisée
- Tolérance sur la masse du précurseur
- Modifications des acides aminés
- Tolérance sur la masse des ions fragments
- Types de fragments recherchés

Ces programmes permettent de rechercher dans toutes les séquences peptidiques, obtenues par digestion trypsique théorique de toutes les protéines présentes dans la banque interrogée, celles qui correspondent au poids moléculaire des ions précurseurs fragmentés, et ceci avec une certaine tolérance d'erreur de masse (Steen and Mann 2004). Plus la série d'ions

fragments théoriques se rapproche de la série expérimentale, plus l'identification est probable et plus le score du peptide est élevé. La séquence du peptide correspondant au meilleur score (meilleure superposition) est généralement attribuée à l'ion précurseur analysé.

La principale différence entre ces logiciels est la vraisemblance de l'identification. Autrement dit, la fonction de corrélation utilisée pour déterminer le degré de similarité entre un spectre MS/MS expérimental et un spectre MS/MS théorique diffère d'un logiciel à un autre. Par exemple, le moteur de recherche MASCOT fournit un score de corrélation (de similarité) peptide-spectre MS/MS associé à une valeur de probabilité.

Certaines correspondances peptidiques pour des spectres MS/MS enregistrés ne sont pas justes, soit parce que :

- le spectre est de mauvaise qualité et peu informatif
- la séquence protéique n'est pas présente dans la base de données
- les paramètres de recherche spécifiés (nombre de modifications post-traductionnelles recherchées, valeur de la tolérance d'erreur de masse de l'ion précurseur, spécificité de l'enzyme) ne sont pas suffisamment larges pour inclure le peptide réellement mesuré comme candidat potentiel.

Il faut donc fixer une valeur seuil de score (seuil d'identité ou « identity threshold ») pour les peptides identifiés afin d'exclure les faux-positifs (FP).

Une fois les peptides identifiés, on passe à l'objectif principal de l'étude qui est l'identification des protéines présentes dans l'échantillon étudié. Cette étape est la plus complexe : elle consiste à assembler les séquences peptidiques correspondant à une même protéine. Mais plusieurs protéines peuvent être associées au même ensemble de séquences peptidiques comme les protéines isoformes et homologues, de même un peptide peut être commun à plusieurs séquences protéiques. Les moteurs de recherche, comme MASCOT, consistent à trouver la plus petite liste de protéines correspondant à l'ensemble des séquences peptidiques identifiées. De cette façon et après une recherche dans les bases de données, l'ensemble des numéros d'accession protéiques associés exactement au même ensemble de peptides (peptides same-set) sont regroupés et restitués dans la liste comme une identification. Les numéros d'accession qui auraient pu être associés uniquement à un sous-ensemble des mêmes peptides sont généralement éliminés (peptide sub-set).

Même après l'assemblage des peptides en protéines, il faut procéder ensuite à la validation de groupes de protéines : il faut calculer le pourcentage de faux-positifs au niveau protéique. Par exemple, si l'on décide de conserver toutes les protéines à un seul peptide validé, alors le pourcentage de FP protéique sera supérieur à celui calculé au niveau peptidique. Ce pourcentage de FP protéique doit être inférieur à 1%.

3.5) Protéomique quantitative

La protéomique quantitative est en plein essor : elle est très utile dans la recherche des biomarqueurs. Elle permet de déterminer la quantité absolue de protéines dans un échantillon, et plus spécifiquement, comparer les quantités relatives de protéines présentes dans multiples conditions, et ceci au niveau d'une cellule, d'un organe ou organisme et à un moment donné. L'apparition, la disparition et la variation des quantités en protéines permettent la comparaison des protéomes entre deux conditions distinctes (patients malades et des témoins normaux) (Ong and Mann 2005).

Deux stratégies ont été développées en protéomique quantitative : une est basée sur l'analyse 2-DE et l'autre sur l'analyse par spectrométrie de masse (Figure 36).

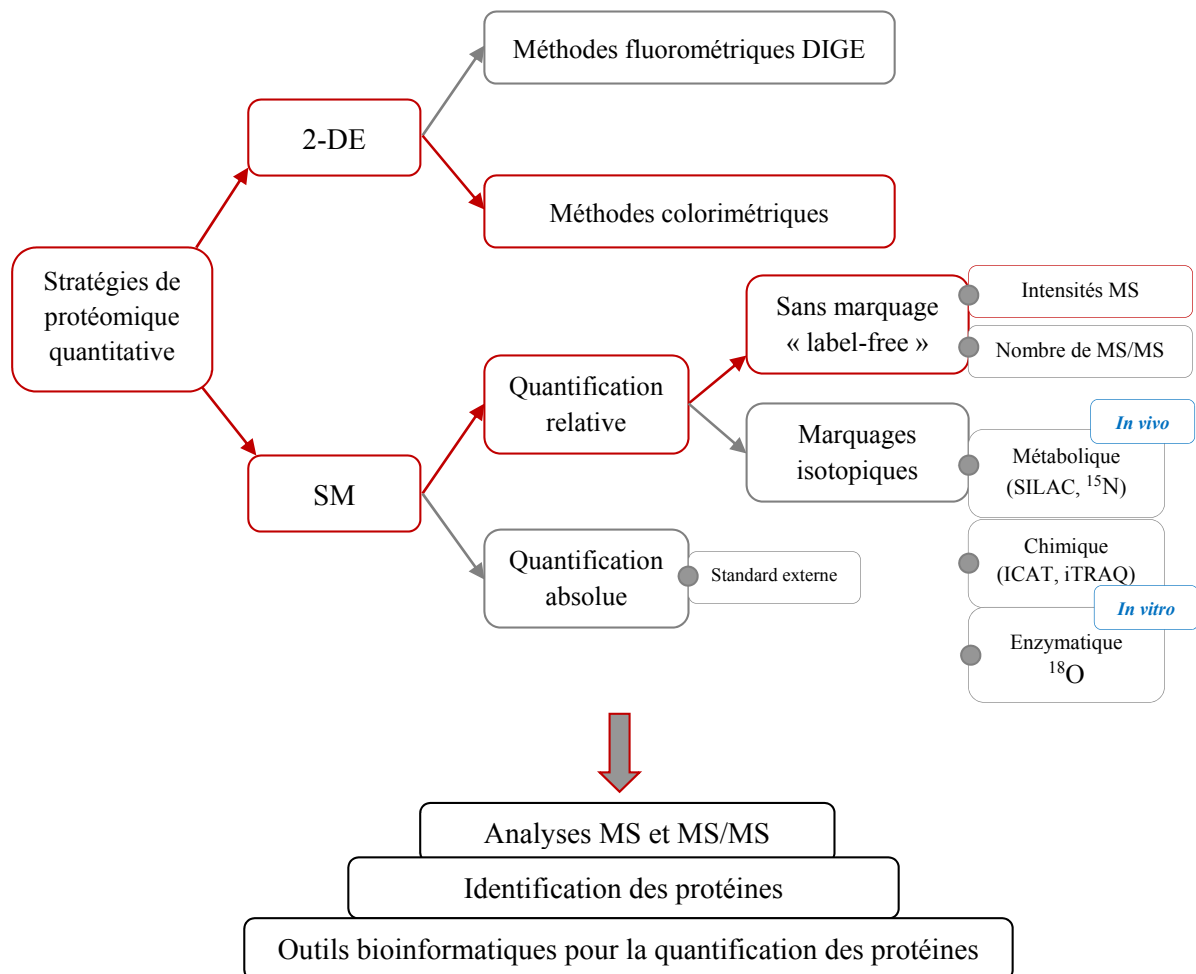


Figure 36: Stratégies utilisées pour l'analyse protéomique quantitative

3.5.1) Quantification par électrophorèse bidimensionnelle

Cette méthode utilisant les fluorophores ou les colorants a été la seule méthode disponible pendant plusieurs années permettant d'étudier les différences d'expression protéique dans des échantillons biologiques différents.

Les protéines séparées sont classiquement révélées par un colorant. Plusieurs techniques de visualisation des protéines possédant différentes sensibilités et caractéristiques dans les gels d'acrylamide ont été décrites : les plus utilisées en protéomique sont le bleu de coomassie colloïdal et la coloration à l'argent qui sont compatibles avec une analyse ultérieure par spectrométrie de masse. Une fois les gels colorés, ils seront digitalisés puis analysés par traitement informatique à l'aide de logiciels. Les spots protéiques d'intérêt sont ensuite découpés et analysés par spectrométrie de masse. Cette technique a une bonne résolution (5000 protéines séparées sur gel), mais son manque de reproductibilité entre les gels, limite l'analyse en protéomique.

Cette approche a connu une avancée significative avec le développement de la 2D-DIGE (2 Dimensional Differential In-Gel Electrophoresis) qui permet de séparer et comparer différents échantillons protéiques sur un même gel (résolvant ainsi le problème dû à la différence de migration des protéines d'un gel à un autre). Elle se base sur le marquage du groupe amine en epsilon des chaînes latérales des lysines à l'aide de fluorophores de type cyanine 2 (utilisé comme standard interne), cyanine 3 et 5, de taille et de charge similaires mais de propriétés spectrales distinctes permettant de visualiser l'émission de leur fluorescence indépendamment l'une de l'autre. 3 images superposables sont obtenues, par gel, grâce à un scanner de fluorescence. L'abondance de chaque spot est mesurée relativement à celle du spot correspondant dans le standard interne présent sur le même gel (Marouga, David et al. 2005).

Malgré toutes ces innovations, ces méthodes souffrent d'une gamme dynamique limitée, rendant ainsi difficile l'analyse des protéines de faible abondance, mais aussi d'un manque de résolution pour les protéines acides ou basiques, sans parler des protéines membranaires rarement observées. De plus, elles ne permettent pas de déterminer, directement et avec exactitude, plusieurs protéines dans un même spot.

3.5.2) Quantification par spectrométrie de masse

L'analyse protéomique quantitative par spectrométrie de masse est l'approche la plus répandue actuellement. Deux stratégies d'analyse existent : quantification absolue des protéines ou peptides et quantification relative.

La quantification relative consiste en une quantification sans marquage (label-free) ou avec marquage isotopique.

3.5.2.1) Techniques avec marquage

Le marquage isotopique stable des peptides a été introduit en protéomique en 1999 (Gygi, Rochon et al. 1999). Puis il a pris une grande ampleur ces dernières années (Patterson and Aebersold 2003, Ong and Mann 2005). Ce marquage repose sur l'introduction d'éléments isotopiques de composition chimique équivalente mais de masses différentielles, garantissant une co-élution en nano-LC et une même efficacité d'ionisation pour les peptides comparés. L'information quantitative sera obtenue par le rapport des intensités des peptides marqués avec des isotopes différents selon les échantillons. Plusieurs protocoles de marquage isotopique existent : les isotopes peuvent être introduits, au niveau protéique ou peptidique et selon la méthode, directement dans les séquences en acides aminés (marquage métabolique comme SILAC pour Stable Isotope Labelling by Aminoacids in Cell culture (Ong, Blagoev et al. 2002) et marquage enzymatique ^{18}O (Yao, Freas et al. 2001) ou par l'intermédiaire d'étiquettes isotopiques liées par réaction chimique aux protéines/peptides (marquage chimique comme iTRAQ pour Isobaric Tags for Relative and Absolute Quantification (Ross, Huang et al. 2004). Bien que très performantes, l'inconvénient non négligeable de ces approches est leur coût relativement élevé.

3.5.2.2) Techniques sans marquage

Quant aux approches sans marquage (label-free, Spectral counting et MRM pour Multiple Reaction Monitoring (Ang and Nice 2010)), elles sont de plus en plus répandues. Moins onéreuses que les précédentes, elles ont été privilégiées pour l'analyse de nos échantillons biologiques en triplicats.

Ces approches se basent sur la comparaison des intensités des signaux MS afin d'obtenir l'abondance relative des protéines de chaque échantillon biologique. Etant prometteuses, elles suscitent un grand intérêt puisqu'elles permettent, tout en réduisant les coûts d'analyses, d'obtenir une quantification fiable des protéines présentes dans un mélange complexe (Olsen, Nielsen et al. 2007). On ne peut pas dire que l'identification des protéines par spectrométrie de masse renseigne sur leur abondance dans l'échantillon : les scores d'identification obtenus, après une recherche dans les banques de données par des logiciels, ne sont pas directement corrélés à la quantité de protéines analysées (Ong and Mann 2007). Les algorithmes des banques de données fournissent des scores de peptides basés sur le nombre de séquences retrouvées dans les banques plutôt que sur leur intensité absolue. Ainsi, un peptide ayant une

faible abondance et présentant un signal faible peut être identifié avec un score très important et sans ambiguïté de sa séquence (Steen and Mann 2004).

Il faut faire attention à l'analyse des peptides par spectrométrie de masse : on n'arrive pas à trouver les peptides très petits et trop grands dans la gamme de masse analysée, de même, il y a des peptides qui ne sont pas retenus sur la colonne de la chromatographie, c'est pour cela que tous les peptides ne sont pas analysables en MS.

Dans cette partie (techniques sans marquage), seule l'approche « label-free » sera détaillée dans ce qui suit.

La quantification « label-free » utilise l'information contenue dans le courant ionique MS (TIC pour Total Ion Chromatogram) à partir duquel on extrait les peptides de masse donnée en fonction d'une fenêtre de tolérance massique et en fonction de son temps de rétention. L'information quantitative est donc obtenue par mesure de l'aire des pics peptidiques MS, spécifiques d'une protéine. Notons que l'aire sous la courbe du pic chromatographique est le courant ionique extrait (XIC pour eXtracted Ion Chromatogram). Ainsi, le XIC est dépendant de la quantité du même peptide et dans les mêmes conditions expérimentales (Ong and Mann 2005). Parallèlement au développement des techniques de quantification, des outils bioinformatiques ont été mis en place pour le traitement des données obtenues qui sont souvent très complexes. Ces outils seront décrits ultérieurement.

Cette approche, ne nécessitant aucune stratégie de marquage, est donc moins fastidieuse à mettre en œuvre et peut s'appliquer à tout type d'échantillons. Cependant, il faut être rigoureux dans la préparation des réplicats biologiques, étape nécessaire pour assurer une bonne reproductibilité entre les échantillons et il faut utiliser des spectromètres de masse hautement résolutifs pour une comparaison précise des signaux.

Les limites de la technique sont la co-élution de peptides isobares (fréquence événementielle augmentant avec la complexité de l'échantillon à analyser) qui perturberont la mesure de l'aire et la reproductibilité des séparations chromatographiques des triplicats (alignement des temps de rétention). Un retraitement des données après acquisition des spectres est donc requis et n'est possible qu'à l'aide d'outils bioinformatiques très performants.

IV) Objectifs de ce travail de thèse

Les données actuelles montrent que l'analyse protéomique permet l'identification, la quantification et la caractérisation d'un grand nombre de protéines, et ceci que ce soit par des analyses MALDI-MS ou nanoLC-MS/MS. De plus, la recherche de biomarqueurs liés à une exposition à un composé particulier est devenue un des points centraux de l'analyse protéomique. Mais leur détection directe se heurte actuellement à la gamme dynamique du fluide sanguin. Par conséquent, le passage préalable par le sécrétome d'un modèle cellulaire constitue une approche pertinente afin d'établir une liste de marqueurs potentiels à rechercher et valider dans un deuxième temps dans des cohortes de plasma de patients.

Durant ce travail de thèse, nous avons donc :

- Optimisé les conditions de culture cellulaire (choix de la lignée cellulaire, choix des contrôles, choix de la concentration des composés toxiques, choix du temps)
- Analysé le protéome total et le sécrétome des cellules hépatiques humaines HepG2 par électrophorèse bidimensionnelle couplée à la spectrométrie de masse
- Analysé le sécrétome par spectrométrie de masse quantitative label-free pour augmenter les chances de détecter les protéines mineures, difficiles à observer par une approche basée sur l'utilisation des gels-2D. Cette partie a fait l'objet d'une publication soumise dans 'Molecular and Cellular Proteomics': « Secretome analysis using label-free quantitative proteomics to discover potential cancer biomarkers of Benzo(a)Pyrene exposure »
- initié une approche de détection du BPDE dans le plasma, sous sa forme libre ou liée, par spectrométrie de masse MALDI-TOF.

Références bibliographiques

A

- Anderson, N. L. and N. G. Anderson (2002). "The human plasma proteome history, character, and diagnostic prospects." Molecular & Cellular Proteomics **1**(11): 845-867.
- Ang, C.-S. and E. C. Nice (2010). "Targeted in-gel MRM: a hypothesis driven approach for colorectal cancer biomarker discovery in human feces." Journal of proteome research **9**(9): 4346-4355.

B

- Baird, W. M., L. A. Hooven and B. Mahadevan (2005). "Carcinogenic polycyclic aromatic hydrocarbon-DNA adducts and mechanism of action." Environmental and Molecular Mutagenesis **45**(2-3): 106-114.
- Biemann, K. (1992). "Mass spectrometry of peptides and proteins." Annual review of biochemistry **61**(1): 977-1010.
- Blades, A. T., M. G. Ikonomou and P. Kebarle (1991). "Mechanism of electrospray mass spectrometry. Electrospray as an electrolysis cell." Analytical chemistry **63**(19): 2109-2114.
- Boffetta, P., N. Jourenkova and P. Gustavsson (1997). "Cancer risk from occupational and environmental exposure to polycyclic aromatic hydrocarbons." Cancer Causes & Control **8**(3): 444-472.
- Boja, E., R. Rivers, C. Kinsinger, M. Mesri, T. Hiltke, A. Rahbar and H. Rodriguez (2010). "Restructuring proteomics through verification." Biomarkers in medicine **4**(6): 799-803.
- Bosetti, C., P. Boffetta and C. La Vecchia (2007). "Occupational exposures to polycyclic aromatic hydrocarbons, and respiratory and urinary tract cancers: a quantitative review to 2005." Annals of Oncology **18**(3): 431-446.
- Burchiel, S. W. and M. I. Luster (2001). "Signaling by environmental polycyclic aromatic hydrocarbons in human lymphocytes." Clinical Immunology **98**(1): 2-10.

C

- Cadet, J., T. Delatour, T. Douki, D. Gasparutto, J.-P. Pouget, J.-L. Ravanat and S. Sauvaigo (1999). "Hydroxyl radicals and DNA base damage." Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis **424**(1): 9-21.
- Castelao, J. E., J.-M. Yuan, P. L. Skipper, S. R. Tannenbaum, M. Gago-Dominguez, J. S. Crowder, R. K. Ross and C. Y. Mimi (2001). "Gender-and smoking-related bladder cancer risk." Journal of the National Cancer Institute **93**(7): 538-545.
- Chait, B. T. (2006). "Chemistry. Mass spectrometry: bottom-up or top-down?" Science **314**(5796): 65-66.
- Chaudhary, A., T. Pechan and K. L. Willett (2007). "Differential protein expression of peroxiredoxin I and II by benzo (a) pyrene and quercetin treatment in 22Rv1 and PrEC prostate cell lines." Toxicology and applied pharmacology **220**(2): 197-210.
- Chen, L., P. D. Devanesan, S. Higginbotham, F. Ariese, R. Jankowiak, G. J. Small, E. G. Rogan and E. L. Cavalieri (1996). "Expanded analysis of benzo [a] pyrene-DNA adducts formed in vitro and in mouse skin: their significance in tumor initiation." Chemical research in toxicology **9**(5): 897-903.
- Chen, M.-R., P.-J. Tsai and Y.-F. Wang (2008). "Assessing inhalatory and dermal exposures and their resultant health-risks for workers exposed to polycyclic aromatic hydrocarbons (PAHs) contained in oil mists in a fastener manufacturing industry." Environment international **34**(7): 971-975.

Chen, Y., H. Zhang, A. Xu, N. Li, J. Liu, C. Liu, D. Lv, S. Wu, L. Huang and S. Yang (2006). "Elevation of serum l-lactate dehydrogenase B correlated with the clinical stage of lung cancer." Lung Cancer **54**(1): 95-102.

Chenau, J. r. m., S. Michelland, F. de Fraipont, V. r. Josserand, J.-L. Coll, M.-C. Favrot and M. Seve (2009). "The cell line secretome, a suitable tool for investigating proteins released in vivo by tumors: application to the study of p53-modulated proteins secreted in lung cancer cells." Journal of proteome research **8**(10): 4579-4591.

Chengzhi, C., T. Yan, C. Shuqun, J. Xuejun, Q. Youbin, X. Yinyin, T. Qian and T. Baijie (2011). "New candidate proteins for Benzo (a) pyrene-induced spatial learning and memory deficits." The Journal of toxicological sciences **36**(2): 163-171.

Chernushevich, I. V., A. V. Loboda and B. A. Thomson (2001). "An introduction to quadrupole–time-of-flight mass spectrometry." Journal of Mass Spectrometry **36**(8): 849-865.

Chevalier, F. (2010). "Review Highlights on the capacities of " Gel-based" proteomics." Proteome science **8**(23).

Cole, R. B. (2000). "Some tenets pertaining to electrospray ionization mass spectrometry." Journal of Mass Spectrometry **35**(7): 763-772.

Cotter, R. J. (1992). "Time-of-flight mass spectrometry for the structural analysis of biological molecules." Analytical chemistry **64**(21): 1027A-1039A.

D

Dautel, F., S. Kalkhof, S. Trump, J. Michaelson, A. Beyer, I. Lehmann and M. von Bergen (2010). "DIGE-based protein expression analysis of B [a] P-exposed hepatoma cells reveals a complex stress response including alterations in oxidative stress, cell cycle control, and cytoskeleton motility at toxic and subacute concentrations." Journal of proteome research **10**(2): 379-393.

Denissenko, M. F., A. Pao, M.-s. Tang and G. P. Pfeifer (1996). "Preferential formation of benzo [a] pyrene adducts at lung cancer mutational hotspots in P53." Science **274**(5286): 430-432.

Dole, M., L. Mack, R. Hines, R. Mobley, L. Ferguson and M. d. Alice (1968). "Molecular beams of macroions." The Journal of Chemical Physics **49**: 2240.

E

EFSA (2008). "Polycyclic Aromatic Hydrocarbons in Food [1]. Scientific Opinion of the Panel on Contaminants in the Food Chain " The EFSA Journal **724**: 1-114.

Emmett, M. R. and R. M. Caprioli (1994). "Micro-electrospray mass spectrometry: ultra-high-sensitivity analysis of peptides and proteins." Journal of the American Society for Mass Spectrometry **5**(7): 605-613.

Eng, J. K., A. L. McCormack and J. R. Yates Iii (1994). "An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database." Journal of the American Society for Mass Spectrometry **5**(11): 976-989.

F

Fatma Aygün, S. and F. Kabadayi (2005). "Determination of benzo [a] pyrene in charcoal grilled meat samples by HPLC with fluorescence detection." International journal of food sciences and nutrition **56**(8): 581-585.

- Fenn, J. B., M. Mann, C. K. Meng, S. F. Wong and C. M. Whitehouse (1989). "Electrospray ionization for mass spectrometry of large biomolecules." Science **246**(4926): 64-71.
- Filippov, V., M. Filippova and P. J. Duerksen-Hughes (2007). "The early response to DNA damage can lead to activation of alternative splicing activity resulting in CD44 splice pattern changes." Cancer research **67**(16): 7621-7630.
- Filippov, V., E. L. Schmidt, M. Filippova and P. J. Duerksen-Hughes (2008). "Splicing and splice factor SRp55 participate in the response to DNA damage by changing isoform ratios of target genes." Gene **420**(1): 34-41.
- Fjellstedt, T. A., R. H. Allen, B. K. Duncan and W. B. Jakoby (1973). "Enzymatic conjugation of epoxides with glutathione." Journal of Biological Chemistry **248**(10): 3702-3707.
- Fung, E. T. (2010). "A recipe for proteomics diagnostic test development: the OVA1 test, from biomarker discovery to FDA clearance." Clinical chemistry **56**(2): 327-329.

G

- Gamero-Castano, M. and J. Fernández de la Mora (2000). "Mechanisms of electrospray ionization of singly and multiply charged salt clusters." Analytica Chimica Acta **406**(1): 67-91.
- Geer, L. Y., S. P. Markey, J. A. Kowalak, L. Wagner, M. Xu, D. M. Maynard, X. Yang, W. Shi and S. H. Bryant (2004). "Open mass spectrometry search algorithm." Journal of proteome research **3**(5): 958-964.
- Glatt, H. (2000). "Sulfotransferases in the bioactivation of xenobiotics." Chemico-biological interactions **129**(1): 141-170.
- Glückmann, M. and M. Karas (1999). "The initial ion velocity and its dependence on matrix, analyte and preparation method in ultraviolet matrix-assisted laser desorption/ionization." Journal of mass spectrometry **34**(5): 467-477.
- Grimmer, G., H. Brune, G. Dettbarn, J. Jacob, J. Misfeld, U. Mohr, K.-W. Naujack, J. Timm and R. Wenzel-Hartung (1991). "Relevance of polycyclic aromatic hydrocarbons as environmental carcinogens." Fresenius' journal of analytical chemistry **339**(10): 792-795.
- Gygi, S. P., Y. Rochon, B. R. Franza and R. Aebersold (1999). "Correlation between protein and mRNA abundance in yeast." Molecular and cellular biology **19**(3): 1720-1730.

H

- Hayati, I., A. Bailey and T. F. Tadros (1987). "Investigations into the mechanisms of electrohydrodynamic spraying of liquids: I. Effect of electric field and the environment on pendant drops and factors affecting the formation of stable jets and atomization." Journal of Colloid and Interface Science **117**(1): 205-221.
- Hecht, S. S. (1999). "Tobacco smoke carcinogens and lung cancer." Journal of the National Cancer Institute **91**(14): 1194-1210.
- Hockley, S. L., V. M. Arlt, D. Brewer, I. Giddings and D. H. Phillips (2006). "Time- and concentration-dependent changes in gene expression induced by benzo (a) pyrene in two human cell lines, MCF-7 and HepG2." BMC genomics **7**(1): 260.
- Hollstein, M., B. Shomer, M. Greenblatt, T. Soussi, E. Hovig, R. Montesano and C. Harris (1996). "Somatic point mutations in the p53 gene of human tumors and cell lines: updated compilation." Nucleic acids research **24**(1): 141-146.
- Hooven, L. A. and W. M. Baird (2008). "Proteomic analysis of MCF-7 cells treated with benzo [a] pyrene, dibenzo [a,h] pyrene, coal tar extract, and diesel exhaust extract." Toxicology **249**(1): 1-10.

I

Iribarne, J. and B. Thomson (1976). "On the evaporation of small ions from charged droplets." The Journal of Chemical Physics **64**: 2287.

J

Jacob, J. and G. Grimmer (1995). "Metabolism and excretion of polycyclic aromatic hydrocarbons in rat and in human." Central European journal of public health **4**: 33-39.

Jernström, B., M. Funk, H. Frank, B. Mannervik and A. Seidel (1996). "CARCINOGENESIS: Glutathione S-transferase A1-1-catalysed conjugation of bay and fjord region diol epoxides of polycyclic aromatic hydrocarbons with glutathione." Carcinogenesis **17**(7): 1491-1498.

K

Karas, M. and F. Hillenkamp (1988). "Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons." Analytical chemistry **60**(20): 2299-2301.

Kebarle, P. and L. Tang (1993). "From ions in solution to ions in the gas phase-the mechanism of electrospray mass spectrometry." Analytical chemistry **65**(22): 972A-986A.

Kelleher, N. L., H. Y. Lin, G. A. Valaskovic, D. J. Aaserud, E. K. Fridriksson and F. W. McLafferty (1999). "Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry." Journal of the American Chemical Society **121**(4): 806-812.

Kenyon, G. L., D. M. DeMarini, E. Fuchs, D. J. Galas, J. F. Kirsch, T. S. Leyh, W. H. Moos, G. A. Petsko, D. Ringe and G. M. Rubin (2002). "Defining the Mandate of Proteomics in the Post-Genomics Era: Workshop Report." Molecular & Cellular Proteomics **1**(10): 763-780.

Klose, J. (1975). "Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues." Humangenetik **26**(3): 231-243.

Knochenmuss, R. and R. Zenobi (2003). "MALDI ionization: the role of in-plume processes." Chemical reviews **103**(2): 441-452.

Kogevinas, M., A. t. Mannetje, S. Cordier, U. Ranft, C. A. González, P. Vineis, J. Chang-Claude, E. Lynge, J. Wahrendorf and A. Tzonou (2003). "Occupation and bladder cancer among men in Western Europe." Cancer Causes & Control **14**(10): 907-914.

Koutros, S., A. J. Cross, D. P. Sandler, J. A. Hoppin, X. Ma, T. Zheng, M. C. Alavanja and R. Sinha (2008). "Meat and meat mutagens and risk of prostate cancer in the Agricultural Health Study." Cancer Epidemiology Biomarkers & Prevention **17**(1): 80-87.

Kristensen, D. B., K. Imamura, Y. Miyamoto and K. Yoshizato (2000). "Mass spectrometric approaches for the characterization of proteins on a hybrid quadrupole time-of-flight (Q-TOF) mass spectrometer." Electrophoresis **21**(2): 430-439.

Kulasingam, V., M. P. Pavlou and E. P. Diamandis (2010). "Integrating high-throughput technologies in the quest for effective biomarkers for ovarian cancer." Nature Reviews Cancer **10**(5): 371-378.

Kussmann, M., E. Nordhoff, H. Rahbek-Nielsen, S. Haebel, M. Rossel-Larsen, L. Jakobsen, J. Gobom, E. Mirgorodskaya, A. Kroll-Kristensen and P. Roepstorff (1997). "Matrix-assisted laser desorption/ionization mass spectrometry sample preparation techniques designed for various peptide and protein analytes." Journal of Mass Spectrometry **32**(6): 593-601.

L

- Lau, A. T. and J. F. Chiu (2006). "Proteomic and biochemical analyses of in vitro carcinogen-induced lung cell transformation: Synergism between arsenic and benzo [a] pyrene." Proteomics **6**(5): 1619-1630.
- Lee, M., M. Novotny and K. Bartle (1976). "Gas chromatography/mass spectrometric and nuclear magnetic resonance determination of polynuclear aromatic hydrocarbons in airborne particulates." Analytical Chemistry **48**(11): 1566-1572.
- Li, D., M. Wang, K. Dhingra and W. N. Hittelman (1996). "Aromatic DNA adducts in adjacent tissues of breast cancer patients: clues to breast cancer etiology." Cancer Research **56**(2): 287-293.
- Liotta, L. A., M. Ferrari and E. Petricoin (2003). "Clinical proteomics: written in blood." Nature **425**(6961): 905-905.
- Lottspeich, F. (1999). "Proteome analysis: a pathway to the functional analysis of proteins." Angewandte Chemie International Edition **38**(17): 2476-2492.

M

- Maier, A., B. L. Schumann, X. Chang, G. Talaska and A. Puga (2002). "Arsenic co-exposure potentiates benzo [a] pyrene genotoxicity." Mutation Research/Genetic Toxicology and Environmental Mutagenesis **517**(1): 101-111.
- Marie, C., M. Bouchard, R. Heredia-Ortiz, C. Viau and A. Maitre (2010). "A toxicokinetic study to elucidate 3-hydroxybenzo(a)pyrene atypical urinary excretion profile following intravenous injection of benzo(a)pyrene in rats." Journal of Applied Toxicology **30**(5): 402-410.
- Marouga, R., S. David and E. Hawkins (2005). "The development of the DIGE system: 2D fluorescence difference gel analysis technology." Analytical and bioanalytical chemistry **382**(3): 669-678.
- Mastrangelo, G., E. Fadda and V. Marzia (1996). "Polycyclic aromatic hydrocarbons and cancer in man." Environmental health perspectives **104**(11): 1166.
- Medzihradzsky, K. F., J. M. Campbell, M. A. Baldwin, A. M. Falick, P. Juhasz, M. L. Vestal and A. L. Burlingame (2000). "The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer." Analytical Chemistry **72**(3): 552-558.
- Melikian, A. A., P. Sun, B. Prokopczyk, K. El-Bayoumy, D. Hoffmann, X. Wang and S. Waggoner (1999). "Identification of benzo [a] pyrene metabolites in cervical mucus and DNA adducts in cervical tissues in humans by gas chromatography-mass spectrometry." Cancer letters **146**(2): 127-134.
- Miller, K. P. and K. S. Ramos (2001). "Impact of cellular metabolism on the biological effects of benzo a pyrene and related hydrocarbons." Drug Metabolism Reviews **33**(1): 1-35.
- Min, L., S. He, Q. Chen, F. Peng, H. Peng and M. Xie (2011). "Comparative proteomic analysis of cellular response of human airway epithelial cells (A549) to benzo (a) pyrene." Toxicology Mechanisms and Methods **21**(5): 374-382.
- Mora, J. F. d. I., G. J. Van Berkel, C. G. Enke, R. B. Cole, M. Martinez-Sanchez and J. B. Fenn (2000). "Electrochemical processes in electrospray ionization mass spectrometry." Journal of mass spectrometry **35**(8): 939-952.

N

- Nickel, W. and C. Rabouille (2009). "Mechanisms of regulated unconventional protein secretion." Nature Reviews Molecular Cell Biology **10**(3): 234-234.

O

- O'Farrell, P. H. (1975). "High resolution two-dimensional electrophoresis of proteins." Journal of biological chemistry **250**(10): 4007-4021.
- Oh, S., H. Im, E. Oh, J. Lee, J. y. Khim, J. Mun, Y. Kim, E. Lee, J. Kim and D. Sul (2004). "Effects of benzo (a) pyrene on protein expression in Jurkat T-cells." Proteomics **4**(11): 3514-3526.
- Olsen, J. V., P. A. Nielsen, J. R. Andersen, M. Mann and J. R. Wiśniewski (2007). "Quantitative proteomic profiling of membrane proteins from the mouse brain cortex, hippocampus, and cerebellum using the HysTag reagent: mapping of neurotransmitter receptors and ion channels." Brain research **1134**: 95-106.
- Ong, S.-E., B. Blagoev, I. Kratchmarova, D. B. Kristensen, H. Steen, A. Pandey and M. Mann (2002). "Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics." Molecular & cellular proteomics **1**(5): 376-386.
- Ong, S.-E. and M. Mann (2005). "Mass spectrometry-based proteomics turns quantitative." Nature chemical biology **1**(5): 252-262.
- Ong, S.-E. and M. Mann (2007). "A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC)." Nature protocols **1**(6): 2650-2660.

P

- Patterson, S. D. and R. H. Aebersold (2003). "Proteomics: the first decade and beyond." Nature genetics **33**: 311-323.
- Penning, T. M., M. E. Burczynski, C.-F. Hung, K. D. McCoull, N. T. Palackal and L. S. Tsuruda (1999). "Dihydrodiol dehydrogenases and polycyclic aromatic hydrocarbon activation: generation of reactive and redox active o-quinones." Chemical research in toxicology **12**(1): 1-18.
- Perera, F. P., A. Estabrook, A. Hewer, K. Channing, A. Rundle, L. A. Mooney, R. Whyatt and D. H. Phillips (1995). "Carcinogen-DNA adducts in human breast tissue." Cancer Epidemiology Biomarkers & Prevention **4**(3): 233-238.
- Perkins, D. N., D. J. C. Pappin, D. M. Creasy and J. S. Cottrell (1999). "Probability-based protein identification by searching sequence databases using mass spectrometry data." ELECTROPHORESIS **20**(18): 3551-3567.
- Pernemalm, M., R. Lewensohn and J. Lehtiö (2009). "Affinity prefractionation for MS-based plasma proteomics." Proteomics **9**(6): 1420-1427.
- Phillips, D. H. (1999). "Polycyclic aromatic hydrocarbons in the diet." Mutation Research/Genetic Toxicology and Environmental Mutagenesis **443**(1): 139-147.
- Planque, C., V. Kulasingam, C. R. Smith, K. Reckamp, L. Goodglick and E. P. Diamandis (2009). "Identification of five candidate lung cancer biomarkers by proteomics analysis of conditioned media of four lung cancer cell lines." Molecular & Cellular Proteomics **8**(12): 2746-2758.

R

- Rabilloud, T. (2002). "Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains." Proteomics **2**: 3-10.
- Rabilloud, T., M. Chevallet, S. Luche and C. Lelong (2010). "Two-dimensional gel electrophoresis in proteomics: past, present and future." Journal of proteomics **73**(11): 2064-2077.

Rehwagen, M., A. Müller, L. Massolo, O. Herbarth and A. Ronco (2005). "Polycyclic aromatic hydrocarbons associated with particles in ambient air from urban and industrial areas." Science of the total Environment **348**(1): 199-210.

Rifai, N., M. A. Gillette and S. A. Carr (2006). "Protein biomarker discovery and validation: the long and uncertain path to clinical utility." Nature biotechnology **24**(8): 971-983.

Roepstorff, P. and J. Fohlman (1984). "Proposal for a common nomenclature for sequence ions in mass spectra of peptides." Biomed Mass Spectrom **11**(11): 601.

Ross, P. L., Y. N. Huang, J. N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey and S. Daniels (2004). "Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents." Molecular & cellular proteomics **3**(12): 1154-1169.

S

Scheele, G. (1975). "Two-dimensional gel analysis of soluble proteins. Characterization of guinea pig exocrine pancreatic proteins." Journal of Biological Chemistry **250**(14): 5375-5385.

Shevchenko, A., M. Wilm and M. Mann (1997). "Peptide sequencing by mass spectrometry for homology searches and cloning of genes." Journal of protein chemistry **16**(5): 481-490.

Shi, Y., R. Xiang, C. Horváth and J. A. Wilkins (2004). "The role of liquid chromatography in proteomics." Journal of Chromatography A **1053**(1): 27-36.

Shimada, T., Y. Oda, E. M. Gillam, F. P. Guengerich and K. Inoue (2001). "Metabolic activation of polycyclic aromatic hydrocarbons and other procarcinogens by cytochromes P450 1A1 and P450 1B1 allelic variants and other human cytochromes P450 in *Salmonella typhimurium* NM2009." Drug metabolism and disposition **29**(9): 1176-1182.

Steen, H. and M. Mann (2004). "The ABC's (and XYZ's) of peptide sequencing." Nature Reviews Molecular Cell Biology **5**(9): 699-711.

Straif, K., R. Baan, Y. Grosse, B. Secretan, F. El Ghissassi and V. Coglianò (2005). "Carcinogenicity of polycyclic aromatic hydrocarbons." The lancet oncology **6**(12): 931-932.

Surinova, S., R. Schiess, R. Hüttenhain, F. Cerciello, B. Wollscheid and R. Aebersold (2010). "On the development of plasma protein biomarkers." Journal of proteome research **10**(1): 5-16.

Szczeklik, A., J. Szczeklik, Z. Galuszka, J. Musiał, E. Kolarzyk and D. Targosz (1994). "Humoral immunosuppression in men exposed to polycyclic aromatic hydrocarbons and related carcinogens in polluted environments." Environmental health perspectives **102**(3): 302.

T

Talaska, G., M. Jaeger, R. Reilman, T. Collins and D. Warshawsky (1996). "Chronic, topical exposure to benzo [a] pyrene induces relatively high steady-state levels of DNA adducts in target tissues and alters kinetics of adduct loss." Proceedings of the National Academy of Sciences **93**(15): 7789-7793.

Tanaka, K., H. Waki, Y. Ido, S. Akita, Y. Yoshida, T. Yoshida and T. Matsuo (1988). "Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry." Rapid Communications in Mass Spectrometry **2**(8): 151-153.

Taura, K.-i. N., Eri, Y. I. Ishii, M.-a. O. Mori, Kazuta and H. Yamada (2004). "Cytochrome P450 1A1 (CYP1A1) Inhibitor α -Naphthoflavone Interferes with UDP-Glucuronosyltransferase (UGT) Activity in Intact but Not in Permeabilized Hepatic Microsomes from 3-Methylcholanthrene-Treated Rats: Possible Involvement of UGT-P450 Interactions." Biol. Pharm. Bull **27**(1): 56-60.

Thulasiraman, V., S. Lin, L. Gheorghiu, J. Lathrop, L. Lomas, D. Hammond and E. Boschetti (2005). "Reduction of the concentration difference of proteins in biological liquids using a library of combinatorial ligands." Electrophoresis **26**(18): 3561-3571.

Tirumalai, R. S., K. C. Chan, D. A. Prieto, H. J. Issaq, T. P. Conrads and T. D. Veenstra (2003). "Characterization of the low molecular weight human serum proteome." Molecular & Cellular Proteomics **2**(10): 1096-1103.

Tjalsma, H., A. Bolhuis, J. D. Jongbloed, S. Bron and J. M. van Dijk (2000). "Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome." Microbiology and Molecular Biology Reviews **64**(3): 515-547.

Tomoyuki, K., H. Tadashi, I. Takatoshi, N. Nobuo and T. Shozo (1980). "Induction of hepatocellular carcinoma in rat liver by initial treatment with benzo (a) pyrene after partial hepatectomy and promotion by phenobarbital." Toxicology Letters **6**(3): 167-171.

U

Uno, S., T. P. Dalton, N. Dragin, C. P. Curran, S. Derkenne, M. L. Miller, H. G. Shertzer, F. J. Gonzalez and D. W. Nebert (2006). "Oral benzo [a] pyrene in Cyp1 knockout mouse lines: CYP1A1 important in detoxication, CYP1B1 metabolism required for immune damage independent of total-body burden and clearance rate." Molecular pharmacology **69**(4): 1103-1114.

V

van Leeuwen, D. M., R. W. Gottschalk, M. H. van Herwijnen, E. J. Moonen, J. C. Kleinjans and J. H. van Delft (2005). "Differential gene expression in human peripheral blood mononuclear cells induced by cigarette smoke and its constituents." Toxicological Sciences **86**(1): 200-210.

Venkateswaran, V. and L. H. Klotz (2010). "Diet and prostate cancer: mechanisms of action and implications for chemoprevention." Nature Reviews Urology **7**(8): 442-453.

Verma, N., C. B auerlein, M. Pink, A. W. Rettenmeier and S. Schmitz-Spanke (2011). "Proteome and phosphoproteome of primary cultured pig urothelial cells." Electrophoresis **32**(24): 3600-3611.

Verma, N., A. W. Rettenmeier and S. Schmitz-Spanke (2010). "Proteomic analysis of primary porcine bladder epithelial cells after BaP exposure." Toxicology Letters **196**: S71.

Visintin, I., Z. Feng, G. Longton, D. C. Ward, A. B. Alvero, Y. Lai, J. Tenthorey, A. Leiser, R. Flores-Saaib and H. Yu (2008). "Diagnostic markers for early detection of ovarian cancer." Clinical Cancer Research **14**(4): 1065-1072.

W

Wang, S., Y. Sheng, M. Feng, J. Leszczynski, L. Wang, H. Tachikawa and H. Yu (2007). "Light-induced cytotoxicity of 16 polycyclic aromatic hydrocarbons on the US EPA priority pollutant list in human skin HaCaT keratinocytes: Relationship between phototoxicity and excited state properties." Environmental toxicology **22**(3): 318-327.

Whitehouse, C. M., R. N. Dreyer, M. Yamashita and J. B. Fenn (1985). "Electrospray interface for liquid chromatographs and mass spectrometers." Analytical Chemistry **57**(3): 675-679.

Wilkins, M. R., J.-C. Sanchez, A. A. Gooley, R. D. Appel, I. Humphery-Smith, D. F. Hochstrasser and K. L. Williams (1996). "Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it." Biotechnology and genetic engineering reviews **13**(1): 19-50.

Williams, J., F. Martin, G. Muir, A. Hewer, P. Grover and D. Phillips (2000). "Metabolic activation of carcinogens and expression of various cytochromes P450 in human prostate tissue." Carcinogenesis **21**(9): 1683-1689.

Wilm, M. and M. Mann (1996). "Analytical properties of the nanoelectrospray ion source." Analytical chemistry **68**(1): 1-8.

Wu, C. C., H. C. Chen, S. J. Chen, H. P. Liu, Y. Y. Hsieh, C. J. Yu, R. Tang, L. L. Hsieh, J. S. Yu and Y. S. Chang (2008). "Identification of collapsin response mediator protein-2 as a potential marker of colorectal carcinoma by comparative analysis of cancer cell secretomes." Proteomics **8**(2): 316-332.

X

Xiao, T., W. Ying, L. Li, Z. Hu, Y. Ma, L. Jiao, J. Ma, Y. Cai, D. Lin and S. Guo (2005). "An approach to studying lung cancer-related proteins in human blood." Molecular & Cellular Proteomics **4**(10): 1480-1486.

Xue, H., B. Lu and M. Lai (2008). "The cancer secretome: a reservoir of biomarkers." J Transl Med **6**(1): 6-52.

Xue, W. and D. Warshawsky (2005). "Metabolic activation of polycyclic and heterocyclic aromatic hydrocarbons and DNA damage: a review." Toxicology and applied pharmacology **206**(1): 73-93.

Y

Yamashita, M. and J. B. Fenn (1984). "Electrospray ion source. Another variation on the free-jet theme." The Journal of Physical Chemistry **88**(20): 4451-4459.

Yamashita, R., Y. Fujiwara, K. Ikari, K. Hamada, A. Otomo, K. Yasuda, M. Noda and Y. Kaburagi (2007). "Extracellular proteome of human hepatoma cell, HepG2 analyzed using two-dimensional liquid chromatography coupled with tandem mass spectrometry." Molecular and cellular biochemistry **298**(1-2): 83-92.

Yan, C., W. Wu, H. Li, G. Zhang, P. J. Duerksen-Hughes, X. Zhu and J. Yang (2010). "Benzo [a] pyrene treatment leads to changes in nuclear protein expression and alternative splicing." Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis **686**(1): 47-56.

Yao, X., A. Freas, J. Ramirez, P. A. Demirev and C. Fenselau (2001). "Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus." Analytical chemistry **73**(13): 2836-2842.

Yergey, A. L., J. R. Coorsen, P. S. Backlund Jr, P. S. Blank, G. A. Humphrey, J. Zimmerberg, J. M. Campbell and M. L. Vestal (2002). "De novo sequencing of peptides using MALDI/TOF-TOF." Journal of the American Society for Mass Spectrometry **13**(7): 784-791.

Yocum, A. K., K. Yu, T. Oe and I. A. Blair (2005). "Effect of immunoaffinity depletion of human serum during proteomic investigations." Journal of proteome research **4**(5): 1722-1731.

Z

Zhang, H.-M., J.-S. Nie, F. Wang, Y.-T. Shi, L. Zhang, A. Antonucci, H.-J. Liu, J. Wang, J. Zhao and Q.-L. Zhang (2008). "Effects of benzo [a] pyrene on autonomic nervous system of coke oven workers." Journal of occupational health(0): 0805120012.

Partie II

Résultats

De la préparation d'échantillons à l'analyse protéomique quantitative

Chapitre 1: Analyse différentielle du protéome total et du sécrétome des cellules HepG2 par électrophorèse bidimensionnelle couplée à la spectrométrie de masse

A) Analyse différentielle du protéome total

1) Cellules cultivées dans le milieu RPMI complété par du sérum

Contexte : Avant de trouver des biomarqueurs protéiques liés à l'exposition au benzo(a)pyrène, il convient d'évaluer la cytotoxicité de ce composé afin de déterminer la concentration qui altère peu la viabilité. Le choix de la concentration nous a permis d'exposer les cellules HepG2, mises en culture dans le milieu RPMI1640 complété avec du sérum, au BaP pendant 24h et 72h. La totalité des protéines a été extraite à partir du culot cellulaire. Elles ont été séparées et quantifiées par gels 2D puis identifiées par MALDI-TOF/TOF et nanoLC-MS/MS (Figure 37). Dans cette partie, des contrôles négatifs (cellules non exposées) n'ont pas été utilisés dans la partie analyse protéomique puisque ce sujet venait d'être démarré et qu'il était nécessaire d'optimiser des protocoles sur quelques échantillons en réplicats afin de ne pas gaspiller la totalité d'échantillons.

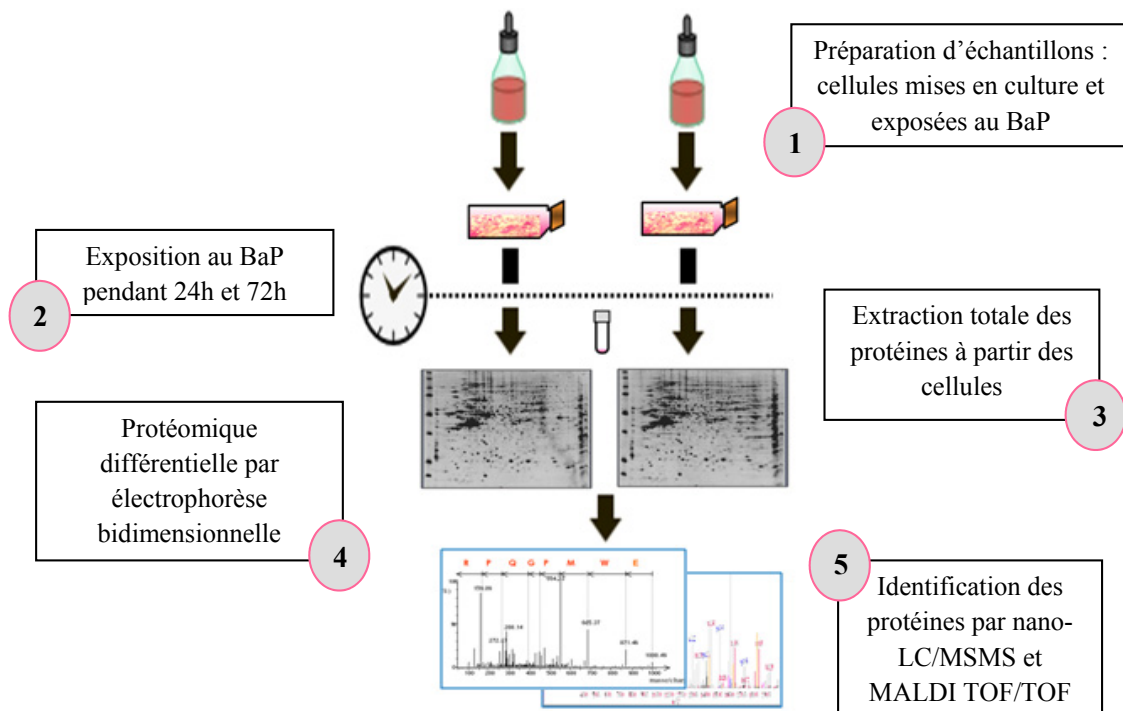


Figure 37 : Approche protéomique des extraits totaux des cellules

1.1) Culture cellulaire

Le choix du modèle cellulaire s'est porté sur les cellules hépatiques humaines HepG2, modèle représentatif pour l'évaluation de la génotoxicité et de la mutagénicité des génotoxiques. Nous avons donc privilégié l'utilisation de cette lignée puisqu'elle possède de nombreuses enzymes de phase I et de phase II impliquées dans le métabolisme des xénobiotiques dans le foie humain (Xu, Ma et al. 2003, Goldring, Kitteringham et al. 2006).

La première culture de cellules HepG2 s'est basée sur l'utilisation du milieu RPMI1640 supplémenté de 10% du sérum du veau fœtal (SVF). Comme la majorité des milieux de culture, le supplément de nutriments nécessaires est apporté par le SVF. Ces cellules ont été exposées, pendant 24h et 72h, à différentes concentrations de Benzo(a)pyrène en solution dans du diméthylsulfoxyde (DMSO) dilué dans le milieu de culture : 0 ; 0,125 ; 0,25 ; 0,5 ; 1 ; 10 et 50 μ M. Pour déterminer le dosage d'exposition au BaP, un test de cytotoxicité MTT a été réalisé (Figure 38), déterminant au final une exposition à 10 μ M de BaP, une concentration correspondant à 60-80% de viabilité cellulaire quelque soit le temps d'exposition.

Le pourcentage de viabilité cellulaire est calculé selon l'équation suivante :

$$\% \text{ de viabilité cellulaire} = (\text{moyenne DO cellules exposées} / \text{moyenne DO cellules contrôles}) \times 100$$

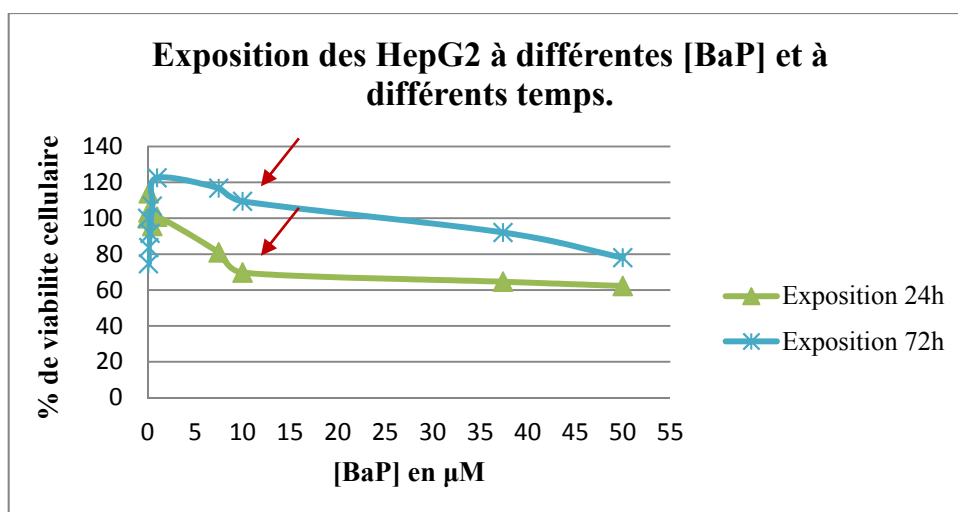


Figure 38 : Cytotoxicité du BaP à des concentrations allant jusqu'à 50 μ M dans les cellules HepG2 exposées pendant 24h et 72h

Suite au test MTT, les cellules HepG2 ont été exposées à 10 μ M de BaP puis lavées deux fois au PBS et incubées pendant 24h et 72h dans un milieu RPMI1640 complété avec du SVF à 10%. Après incubation, les cellules, au nombre de $10 \cdot 10^6$, sont décollées et le culot cellulaire est lavé au PBS puis figé dans l'azote liquide.

Le mot clé pour la préparation d'échantillons est reproductibilité. Cette condition indispensable pour la réalisation de différentiel d'expression protéique nous a guidés tout au long des étapes préparatoires, accompagnées de la réalisation de triplicats biologiques pour chaque condition. Les techniques de culture cellulaire ont été la base de la préparation du matériel protéomique pour la suite du projet.

1.2) Extraction des protéines à partir des culots cellulaires et dosage

Deux réplicats de chaque condition de temps (24h et 72h) ont été utilisés pour l'extraction protéique du lysat cellulaire au trizol. La quantité de protéine de l'extrait est dosée par la méthode de Bradford afin de pouvoir mener l'électrophorèse sur gel avec 150µg de protéines totales par gel.

Les résultats du dosage des 4 extraits protéiques sont représentés dans le tableau suivant :

Tableau 4 : Résultats du dosage des extraits protéiques (condition BaP 24h et 72h)

Echantillon	Concentration µg/µL	Rendement µg (dans 200µL)
1 ^{er} réplicat 24h	7,57	1514
2 ^{ème} réplicat 24h	8,39	1678
1 ^{er} réplicat 72h	5,75	1150
2 ^{ème} réplicat 72h	6,40	1280

1.3) Electrophorèse bidimensionnelle

150µg des extraits protéiques totaux de chaque échantillon ont été déposés sur gels à gradient de pH immobilisé (strip IPG 17cm pH5-8). Cette isoélectrofocalisation (première dimension) a été suivie d'un gel SDS-PAGE 5-13% avant d'être coloré au bleu colloïdal. Les résultats des 4 gels 2D digitalisés correspondant à deux réplicats par condition sont figurés ci-dessous.

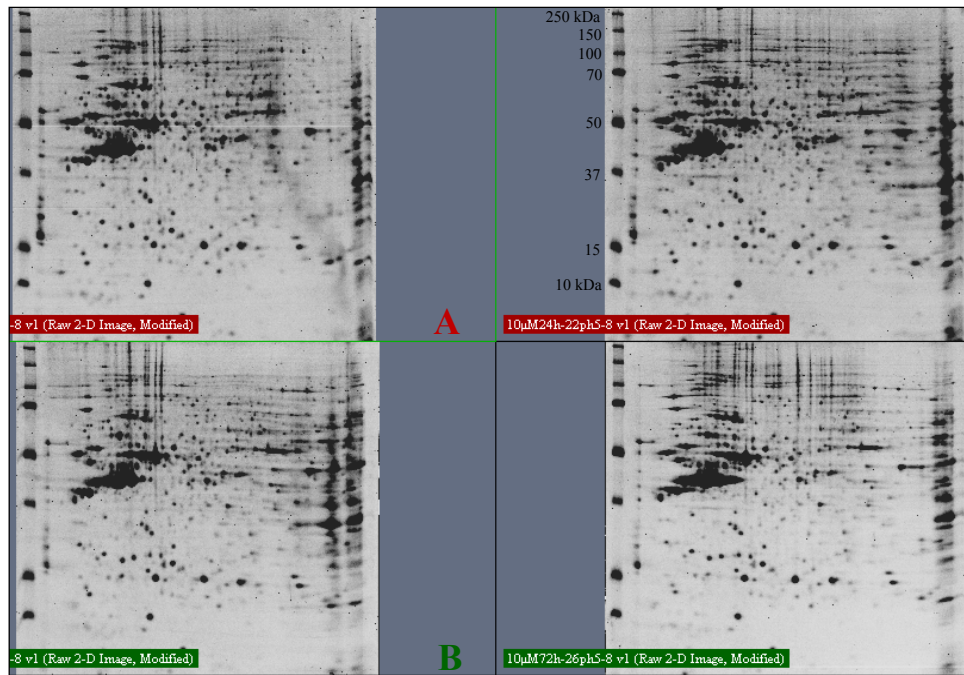


Figure 39 : Duplicats de gels 2D des extraits protéiques totaux des cellules HepG2 exposées à 10µM de BaP pendant 24h, lavées et remises en culture avec sérum pendant 24h (A) ou 72h (B).

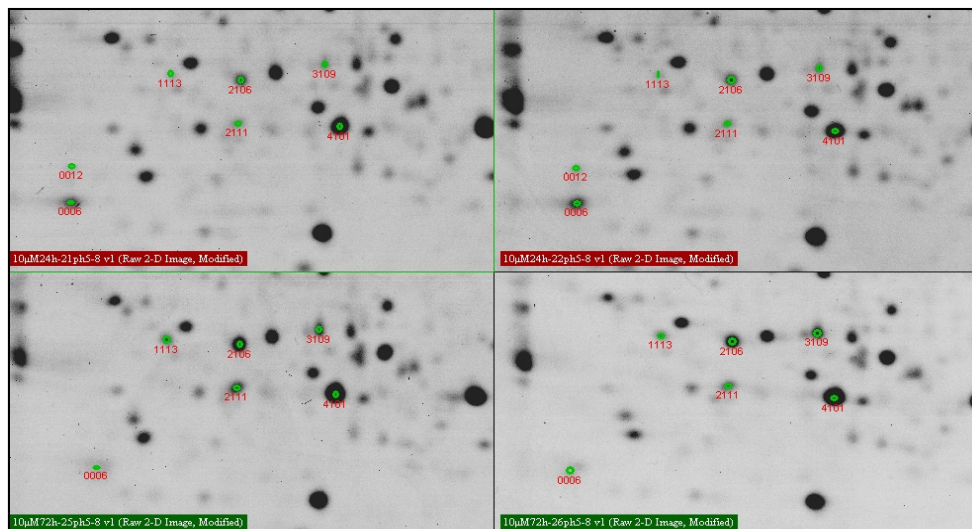


Figure 40 : Focalisation avec grandissement sur une partie des gels permettant de bien visualiser les spots d'intérêts marqués en vert

L'information quantitative est obtenue par analyse densitométrique comparative du volume des spots protéiques à l'aide du logiciel PD-Quest (8.0.1). Au total, 31 protéines sont différentiellement exprimées : 3 sont seulement présentes à la condition 24h, 13 sont surexprimées (>2-fold) et 15 sous-exprimées à la condition 72h (<0,5-fold). Les ratios de ces protéines sont présentés dans les tableaux 5 et 6.

1.4) Identification par spectrométrie de masse

Les protéines d'intérêt, exprimées différentiellement, sont excisées du gel, digérées par la trypsine et identifiées par spectrométrie de masse MALDI-TOF/TOF.

Elles sont tout d'abord identifiées en MS (approche PMF pour Peptide Mass Fingerprinting) en mode réflectron positif, et ceci à l'aide du moteur de recherche Mascot (Matrix Science). Les spectres de masse fournis par l'appareil ont été comparés à des spectres théoriques obtenus en simulant *in silico* le comportement de la trypsine, utilisée expérimentalement, sur la banque de protéines NCBI (non-redundant). Un spectre théorique sera donc calculé pour chacune des protéines de cette banque. La comparaison consiste à évaluer la similarité entre un spectre expérimental et un spectre théorique.

Si la protéine n'est pas identifiée en MS, son mélange peptidique sera soumis à la fragmentation MS/MS (approche PFF pour Peptide Fragment Fingerprinting) en mode LIFT. La fragmentation en MS/MS va couper le peptide à différents endroits, et ainsi créer une série d'ions. Comme dans les spectres MS, Mascot est capable de comparer les spectres MS/MS expérimentaux bruts aux spectres MS/MS théoriques issus de la digestion *in silico* des protéines de la banque. Les pics des spectres MS/MS sont caractérisés par leur position (m/z d'un fragment) et leur intensité. La comparaison s'arrête cependant là : un spectre MS/MS est plus complexe, et comporte potentiellement de très nombreux pics significatifs ayant des origines variées, par opposition au spectre MS dans lequel un pic significatif représente nécessairement le m/z d'un peptide. L'ensemble des peptides identifiés permet de remonter à l'identité des protéines initialement présentes dans le mélange complexe.

Les résultats des 10 protéines identifiées par approche PMF ou PFF sont représentés dans le tableau 8.

Les protéines, non identifiées par MALDI, ont été passées en nanoLC-MSMS (nanoLC-U3000, Dionex, US / MicroTOFQ-II, Bruker, Germany). Dans ce cas là, le logiciel Proteinscape (version 3.1) de la société Bruker a été utilisé : il est capable d'exploiter l'ensemble des données de masse, quelque soit la provenance instrumentale.

Proteinscape utilise le moteur de recherche Mascot pour l'identification des protéines. Les paramètres utilisés sont les suivants :

- tolérance sur la masse du peptide : 25 ppm
- tolérance MSMS : 0,5 dalton
- nombres de sites de coupure tryptique manqués: 2

- modifications variables : carbométhylation des cystéines, acétylation de l'extrémité N-terminale protéique, oxydation des méthionines.

L'identification des protéines contenues dans les échantillons s'est faite par comparaison avec la banque de séquences protéiques NCBIInr, taxonomie *Homo sapiens* (human).

Proteinscape valide automatiquement les résultats en appliquant les paramètres suivants : un peptide de score supérieur à 10 est importé dans la base Proteinscape mais seuls les peptides ayant un score supérieur à l'Identity Threshold (calculé par Mascot) sont validés de manière à avoir un taux de faux-positifs inférieur à 1%. Une validation manuelle est parfois nécessaire et les critères suivants ont été appliqués : les pics majeurs du spectre doivent correspondre à des fragments du peptide théorique, une série successive d'ions Y doit être lisible correctement (minimum de 5 acides aminés consécutifs), le signal des pics fragments doivent ressortir du bruit de fond ($S/N > 3$), si le peptide contient une proline alors le fragment y se terminant par la proline doit être un des pics les plus intenses sur le spectre MS/MS.

Les résultats des 31 protéines identifiées par Proteinscape sont représentés dans le tableau 6.

Tableau 5 : Résultats d'analyses par 2-DE et MALDI-TOF(/TOF) des spots de protéines totales détectées, sous ou surexprimées après traitement au BaP pendant 24 et 72h

Spot	72H/24H	PMF	PFF	MSMS	Identification	Mw (kDa)	pI	N°Access ncbi	Score Mascot	% recouvrement	Erreur (ppm)
2303	Présente 24h	x			karyopherin alpha 2 [Homo sapiens]	58,17	5,25	gi 4504897	81	18	11
2111	2,71	x			heat shock protein 27 [Homo sapiens]	22,48	7,83	gi 662841	142	39	7
4101	2,16	x			heat shock protein 27 [Homo sapiens]	22,43	7,83	gi 662841	92	37	26
1403	0,5	x			tubulin alpha 6 [Homo sapiens]	50,55	4,96	gi 14389309	107	28	21
3501	0,46	x			Chain A, Heat-Shock 70kd Protein 42kd Atpase N-Terminal Domain	41,8	6,8	gi 6729803	213	45	10
1307	0,46	x			tubulin alpha 6 [Homo sapiens]	50,55	4,96	gi 14389309	108	28	13
2216	0,43	x	x	1	heterogeneous nuclear ribonucleoprotein F [Homo sapiens]	45,98	5,38	gi 4826760	77	16	24
2708	0,27	x			heat shock 105kDa/110kDa protein 1 [Homo sapiens]	93,05	5,29	gi 55957726	103	15	49
1301	0,20	x			heat shock 70kDa protein 8 isoform 2 variant [Homo sapiens]	53,58	5,62	gi 62896815	126	30	30
6404	0,19	x			Chaperonin containing TCP1, subunit 3 (gamma) [Homo sapiens]	60,93	6,1	gi 14124984	129	28	28

Tableau 6 : Résultats d'analyses par 2-DE et NanoLC-MS/MS des spots de protéines totales détectées, sous ou surexprimées après traitement au BaP pendant 24 et 72h

Spot	72H/24H	Mixte	nanoLC-MSMS	PEPTIDES MSMS	Identification	Mw (kDa)	pI	N°Access	Score Mascot	% recouvrement	Erreur (ppm)
12	Présente 24h		2		Aucune identification par nanoLC-MS/MS						
2303	Présente 24h				karyopherin alpha 2 [Homo sapiens]	58,17	5,25	gi 4504897	255	33,6	9
2405	Présente 24h		1	5	RPD3 protein [Homo sapiens]	55,0	5,3	gi 1665723	274,3	14,7	1,44
1113	9,66		1	5	cathepsin D preproprotein [Homo sapiens]	44,5	6,1	gi 4503143	416,7	14,3	9,50
6405	6,32		1	5	chaperonin containing TCP1, subunit 3 isoform c [Homo sapiens]	56,4	6,1	gi 58761484	651,4	24,1	23,68
3308	3,15		2	2	RuvB-like 2 [Homo sapiens]	51,23	5,49	gi 5730023	105	7	30
3602	3,05		1	5	heat shock protein gp96 precursor [Homo sapiens]	90,1	4,6	gi 15010550	322,4	9,5	18,92
2106	2,92		1	15	cathepsin D preproprotein [Homo sapiens]	44,5	6,1	gi 4503143	880,9	35,7	3,76
4606	2,85		1	9	gelsolin isoform b [Homo sapiens]	80,6	5,5	gi 38044288	631,2	19,2	29,53
2111	2,71		1	2	heat shock protein 27 [Homo sapiens]	22,3	9,1	gi 662841	312,5	36,7	53,36
3109	2,65		1	6	Chain B, Human Carboxypeptidase A4 In Complex With Human Latexin.	25,8	5,5	gi 66361584	326,40	27,70	6,76
4304	2,4		2	1	histidyl-tRNA synthetase [Homo sapiens]	57,4	5,6	gi 62766469	85,4	4,3	6,65
5210	2,34		2		Aucune identification par nanoLC-MS/MS						
5108	2,22		2		Aucune identification par nanoLC-MS/MS						
4101	2,16		1	6	heat shock protein 27 [Homo sapiens]	22,3	9,1	gi 662841	414,7	35,2	27,33
6210	0,49		1	3	Chain A, The Crystal Structure Of The Exon Junction Complex At 3.2 A Resolution	46,7	6,3	gi 114794876	393,3	16,3	52,87
1403	0,49		1	8	alpha-tubulin [homo sapiens]	50,8	5,02	gi 37492	291	20	35,15
3501	0,46	oui	1	3	Chain A, Heat-Shock 70kd Protein 42kd Atpase N-Terminal Domain	41,8	6,8	gi 6729803	180,6	10,3	32,26
			1	23	heat shock 70kDa protein 8 isoform 1 [Homo sapiens]	70,9	5,2	gi 5729877	1679,6	35,9	30,19

Spot	72H/24H	Mixte	nanoLC-MSMS	PEPTIDES MSMS	Identification	Mw (kDa)	pI	N°Access	Score Mascot	% recouvrement	Erreur (ppm)
1307	0,46		1	10	alpha-tubulin [homo sapiens]	50,8	5,02	gi 37492	1021	36	31
5406	0,45	oui	1	14	chaperonin containing TCP1, subunit 3 (gamma) [Homo sapiens]	57,9	6,5	gi 55960506	833,60	35,6	3,63
				5	asparaginyl-tRNA synthetase variant [Homo sapiens]	62,9	5,9	gi 62897229	247,1	11,5	1,15
2709	0,43		1	11	HSP105 beta [Homo sapiens]	92,1	5,3	gi 3970829	777,40	22,5	7,44
2216	0,43		1	7	heterogeneous nuclear ribonucleoprotein F [Homo sapiens]	45,6	5,3	gi 4826760	483,9	21,4	34,13
5708	0,41		1	12	vinculin, isoform CRA_a [Homo sapiens]	108,7	6,0	gi 119574932	577,70	16,7	2,30
2007	0,32		1	1	HP1Hs-gamma	19,7	4,9	gi 1773227	86,90	15,00	9,98
2708	0,27		1	5	HSP105 beta [Homo sapiens]	92,1	5,3	gi 3970829	214,0	7,4	42,22
7515	0,27		1	14	lamin A/C, isoform CRA_c [Homo sapiens]	87,3	9,7	gi 119573383	802,3	21,3	37,37
4510	0,25		1		Aucune identification par nanoLC-MS/MS						
6	0,24		1	3	TCTP [Homo sapiens]	19,6	4,8	gi 33285832	156,20	30,2	7,19
1301	0,2		1	12	heat shock 70kDa protein 8 isoform 2 variant [Homo sapiens]	53,58	5,62	gi 62896815	983,5	28,1	12
6404	0,19		1	7	chaperonin containing TCP1, subunit 3 isoform c [Homo sapiens]	56,4	6,1	gi 58761484	731	26,4	29,78

1.5) Discussion biologique

La principale approche de lyse cellulaire utilise des tampons contenant des détergents comme les agents chaotropes (Urée). Ces solutions créent un choc osmotique, faisant entrer l'eau dans les cellules et provoque la rupture de la membrane plasmique. Une ultracentrifugation permet de culotter les débris cellulaires au fond du tube Eppendorf et le surnageant est récupéré pour une analyse protéomique. Après séparation des protéines sur gels 2D, on a remarqué que 85 à 90 % de l'extrait protéique correspond à la BSA. Cette protéine provient du sérum puisque les cellules ont été incubées après deux lavages au PBS dans un milieu avec sérum. Comme illustré dans la figure 49, malgré la contamination de la BSA qui pourra masquer des protéines intéressantes, les gels 2D ont été analysés sur PD-Quest : 31 protéines sont différentiellement exprimées. Trois protéines sont exprimées seulement dans les cellules traitées pendant 24h à 10 μ M de BaP, 15 protéines sont sous-exprimées (ratio BaP72h/BaP24h < 0,5) et 13 surexprimées (ratio BaP72h/BaP24h > 2) dans celles traitées pendant 72h par rapport à celles traitées pendant 24h. Parmi ces 31 protéines, 2 protéines surexprimées impliquées dans le cancer nous ont paru intéressantes : la cathepsin D et la RuvBl2. Tandis que les autres protéines sont impliquées dans d'autres fonctions comme le stress cellulaire (Heat Shock Protein 27, Chaperonin containing TCP1) et le transport (Karyopherin α 2).

La production excessive de la cathepsin D est associée à l'apparition des métastases cancéreuses. C'est un marqueur de mauvais pronostic du cancer du sein qui stimule la prolifération des cellules cancéreuses (Masson, Prébois et al. 2011). Cette protéine a été identifiée comme surexprimée dans différentes études : l'une, *in vivo*, sur des souris exposées à 2 μ M de BaP, l'autre, *in vitro*, sur des cellules urothéliales traitées avec 0,5 μ M de BaP, et la dernière, aussi *in vitro*, sur des cellules épithéliales des voies respiratoires humaines exposées à 1 μ M de BaP (Kassie, Anderson et al. 2008, Verma, Rettenmeier et al. 2010, Min, He et al. 2011).

Quant à la RuvBl2 (RuvB-like2), connue sous le nom de Reptin52, elle appartient à la superfamille AAA+ (ATPases associated with various cellular activities) et conserve 2 domaines : Walker A et Walker B (Gorbalenya *et al.*, 1989), le premier permettant sa liaison à l'ATP et le second l'hydrolyse de l'ATP. Donc la RuvBl2 humaine a une activité ATPase. Blanc *et al.* ont démontré qu'elle est surexprimée dans les tumeurs chez des patients atteints du carcinome hépatocellulaire. Cette expression augmente au niveau de l'ARNm et au niveau protéique (Haurie, Ménard et al. 2009). Mais le mécanisme de cette augmentation n'est pas

encore connu. Par contre, elle interagit avec des médiateurs de cancer, comme des facteurs de transcription qui sont impliqués dans la carcinogenèse (c-myc), ce qui va favoriser leur rôle oncogénique (Menssen and Hermeking 2002). Elle est aussi capable d'interagir avec la β -caténine (clé de voûte de la voie de carcinogenèse : voie Wnt) (Bauer, Chauvet et al. 2000). Comme elle peut coopérer avec la β -caténine pour réprimer l'expression du gène anti-métastatique KAI-1 (un gène sous-exprimé dans le carcinome hépatocellulaire) (Kim, Kim et al. 2005). De même, elle peut interagir avec Hint-1 en inhibant l'apoptose (Hint-1 : histidine triad protein qui déclenche l'apoptose) (Weiske and Huber 2005). Des études ont aussi démontré que la sous-expression de la RuvB12 est associée avec une augmentation de l'activation des protéines proapoptotiques Bak-1 et Bax, ce qui a permis aux auteurs de proposer que l'apoptose induite par la déplétion de RuvB12 est liée à l'activation des membres proapoptotiques de la famille de Bcl2 (Rousseau, Ménard et al. 2007). Donc la RuvB12 favorise la prolifération cellulaire et inhibe l'apoptose. Elle peut être qualifiée comme cible thérapeutique intéressante contre le cancer.

Mais cette protéine n'a pas été encore trouvée comme biomarqueur lié à l'exposition au BaP. C'est pour cela que nous nous sommes intéressés à la découper des autres gels déjà préparés puisqu'elle a été identifiée avec un seul peptide dans le premier gel. Donc deux spots ont été découpés, digérés puis analysés en MALDI-TOF/TOF en mode LIFT. La protéine RuvB12, surexprimée à 72h, a été validée avec 17 peptides et un score Mascot de 136 dans le premier résultat puis avec 14 peptides et un score Mascot de 138 dans le second.

Les résultats de cette analyse protéomique étant encourageants, d'autres expériences ont été envisagées:

- Puisque la BSA occupait la majorité de la quantité des protéines de l'extrait total, nous avons pensé à un supplément « serum free » capable de remplacer le SVF que nous avons utilisé dans notre première culture cellulaire.
- Comparer les résultats de la RuvB12 apparus dans ces cellules hépatiques humaines exposées au BaP à des cellules hépatiques contrôles
- Travailler sur d'autres temps d'exposition (2h et 6h) afin d'identifier des biomarqueurs précoces et de les comparer aux temps 24h et 72h (biomarqueurs tardifs)

- Envisager un fractionnement : prendre le protéome total et le sécrétome, ce dernier étant une source potentielle appropriée pour la recherche de nouvelles cibles thérapeutiques ou de candidats biomarqueurs de nombreuses maladies.

Les optimisations de la culture cellulaire seront détaillées dans les parties qui suivent.

2) *Cellules cultivées dans le milieu RPMI complété par le supplément B27*

Contexte : Le but de cette partie était d'optimiser les conditions de culture cellulaire afin de minimiser la quantité de protéines provenant du sérum et d'exposer les cellules HepG2 à un agent toxique non cancérigène jouant le rôle d'un contrôle, le Benzo(e)pyrène (BeP). Les protéines ont été extraites du culot cellulaire des cellules HepG2 exposées à 10µM de l'agent cancérigène BaP et de l'agent toxique BeP, séparées par gels 2D et identifiées par nanoLC-MS/MS.

2.1) Culture cellulaire

La culture cellulaire s'est basée, cette fois-ci, sur l'utilisation du supplément B27 (milieu sans sérum) au lieu du SVF. La cytotoxicité du BaP étant évaluée par le test MTT, nous avons exposé les cellules HepG2 à 10µM de BaP et BeP pendant 24h. Après 24h d'exposition au BaP et au BeP, les cellules ont été lavées deux fois par du PBS afin de réduire la quantité de protéines ayant pour origine le SVF, et remises en culture dans un milieu supplémenté de B27 pendant 6h, 24h et 72h (3 réplicats/condition). Après chaque temps d'incubation, les sécrétomes ont été stockés à -80°C et les culots cellulaires (nombre de cellules $10 \cdot 10^6$) figés dans l'azote liquide.

2.2) Dosage des protéines extraites à partir des culots cellulaires

Les résultats du dosage Bradford des extraits protéiques (trois réplicats par condition d'exposition et temps) sont représentés dans le tableau suivant:

Tableau 7 : Résultats du dosage Bradford des extraits protéiques (conditions Control vs. BeP vs. BaP pendant 6h, 24h et 72h)

Echantillon	Concentration µg/µL	Rendement µg (dans 200µL)
Control 6h (1)	16,49	3298
Control 6h (2)	17,91	3582
Control 6h (3)	19,10	3820

Echantillon	Concentration $\mu\text{g}/\mu\text{L}$	Rendement μg (dans 200 μL)
BeP 6h (1)	2,94	588
BeP 6h (2)	3,68	736
BeP 6h (3)	2,29	458
BaP 6h (1)	7,78	1556
BaP 6h (2)	8,09	1618
BaP 6h (3)	8,67	1734
Control 24h (1)	14,91	2982
Control 24h (2)	15,69	3138
Control 24h (3)	13,98	2796
BeP 24h (1)	5,74	1148
BeP 24h (2)	6,30	1260
BeP 24h (3)	8,53	1706
BaP 24h (1)	6,01	1202
BaP 24h (2)	6,19	1238
BaP 24h (3)	7,25	1450
Control 72h (1)	9,84	1968
Control 72h (2)	15,78	3156
Control 72h (3)	14,12	2824
BeP 72h (1)	6,88	1376
BeP 72h (2)	7,88	1576
BeP 72h (3)	9,36	1872
BaP 72h (1)	2,46	492
BaP 72h (3)	2,94	588
BaP 72h (3)	3,05	610

2.3) Electrophorèse bidimensionnelle

150 μg des extraits protéiques ont été déposés sur gels à gradient de pH immobilisé (strip IPG 17cm pH5-8), puis migrés sur gel SDS-PAGE 5-13% et colorés au bleu colloïdal.

Dans cette partie, nous montrons juste les résultats des gels 2D des conditions suivantes : BaP/BeP, 24h/72h (Figure 41). Les manipulations étant séparées dans le temps, nous n'avons pas pu comparer les autres conditions d'exposition et de temps avec celles des BaP et BeP (24 et 72h).

A l'aide du logiciel PD-Quest, nous avons sélectionné 20 spots différentiellement exprimés (tableau 8) : 5 protéines sont seulement présentes à BaP 72h, 12 sont up-régulées (ratio BaP72h/ BaP24h > 2) et 3 sous-exprimées (ratio BaP72h/ BaP24h < 0,5) (Figure 42).

Tableau 8 : Protéines différentiellement exprimées

Spot number	Ratio BaP 72h/24h	Ratio BeP 72h/24h	Spot number	Ratio BaP 72h/24h	Ratio BeP 72h/24h
Protéines présentes à BaP 72h			Protéines surexprimées		
0223	pres 72h	-	8008	12,33	3,835
1010	pres 72h	-	1604	4,58	3,156
1810	pres 72h	-	1603	4,21	1,313
2712	pres 72h	-	5706	3,95	1,563
4306	pres 72h	-	0214	3,26	1,709
Protéines sous-exprimées			4609	3,09	1,342
4507	0,25	0,677	4101	2,59	1,244
8701	0,41	0,929	9503	2,54	1,080
6312	0,41	0,568	8407	2,36	1,388
			8207	2,35	2,173
			8408	2,27	1,512
			5107	2,19	1,138
			2605	2,11	0,539

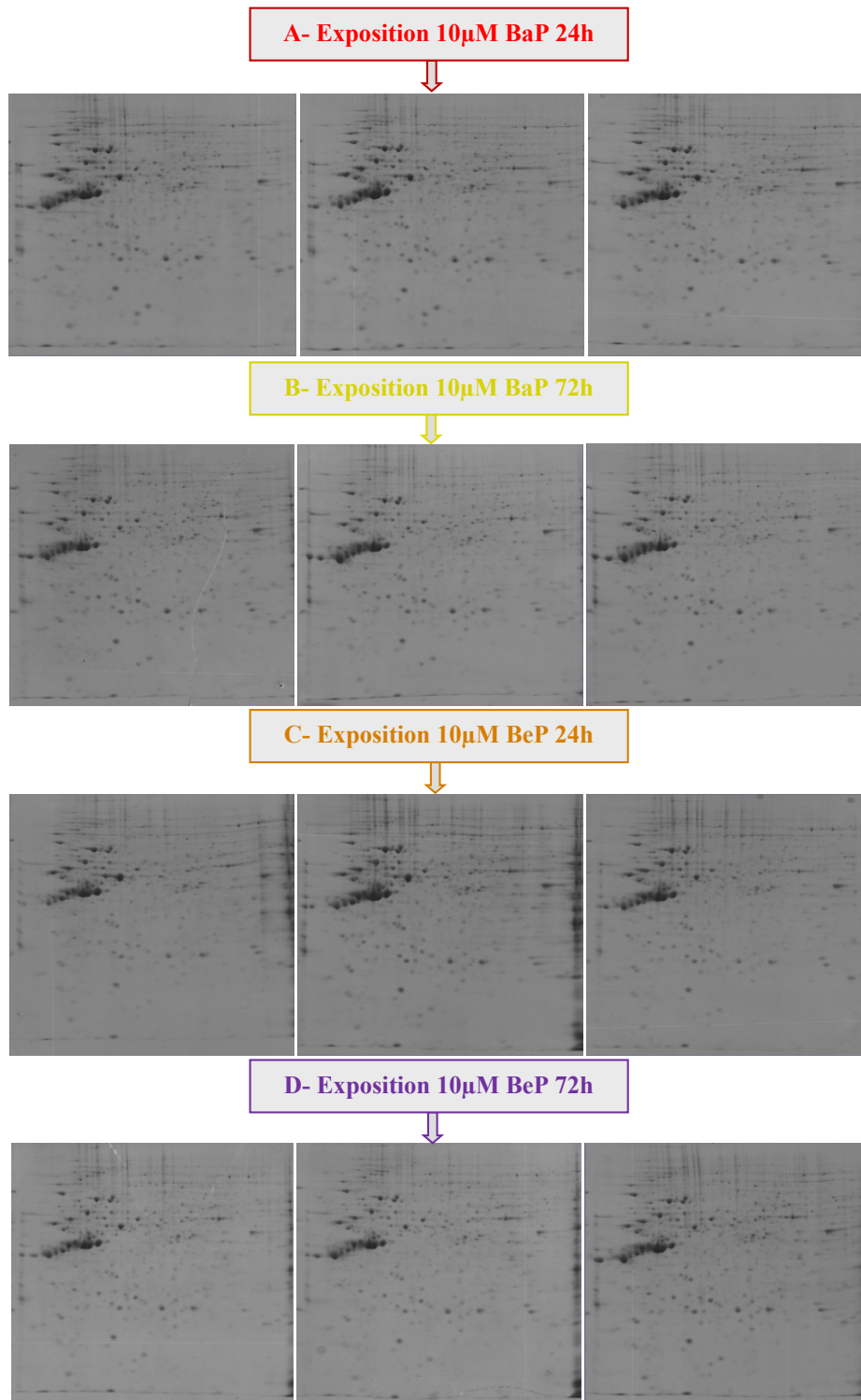
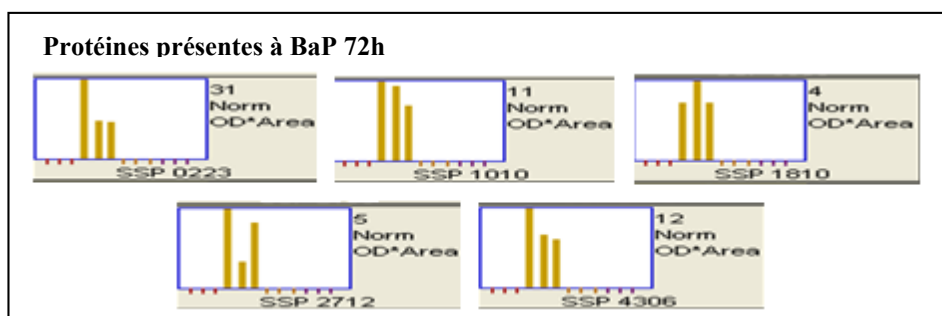
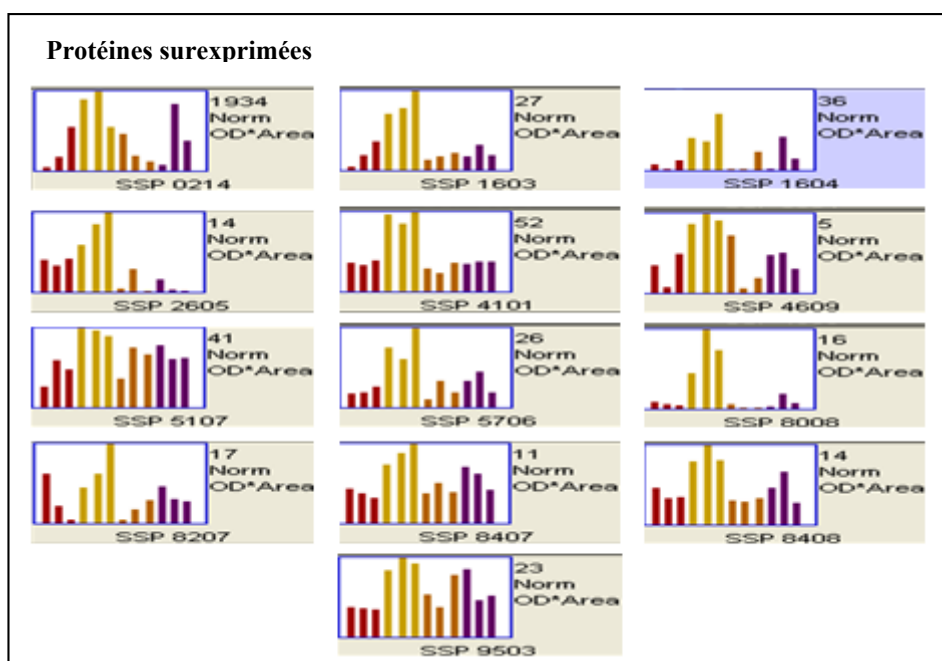


Figure 41 : Triplicats de gels 2D des extraits protéiques totaux des cellules HepG2 exposées à 10 μ M de BaP (A-B) ou de BeP (C-D) pendant 24h, lavées et remises en culture dans un milieu sans sérum pendant 24h (A-C) ou 72h (B-D).



◆ BaP 24h

◆ BaP 72h



◆ BeP 24h

◆ BeP 72h

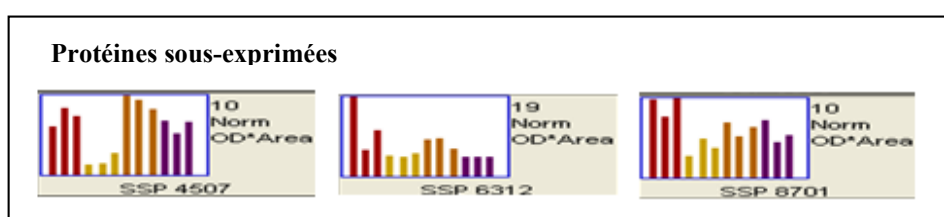


Figure 42 : Quantification des protéines par PD-Quest

2.4) Identification par spectrométrie de masse

Les résultats des 20 spots digérés, analysés par nanoLC-MS/MS et identifiés à l'aide de Proteinscape sont présentés dans le tableau 9. Leurs classifications et leurs fonctions moléculaires sont représentées dans les figures 43 et 44, respectivement.

Tableau 9 : Résultats d'analyses des spots de protéines totales détectées, sous ou surexprimées après traitement au BaP/BeP pendant 24 et 72h par nanoLC-MS/MS

Spot	72H/24H	Mixt	PEPTIDES MSMS	Identification	Mw (kDa)	pI	N°Access ncbi	Score Mascot	Erreur (ppm)	% recouvrement
Protéines présentes à BaP 72h										
0223	Présente 72h		6	PIH1 domain-containing protein 1	32,3	4,9	PIHD1_HUMAN	357,9	2,95	23,8
1010	Présente 72h		2	Heat shock protein beta-1	22,8	6,0	HSPB1_HUMAN	123,1	2,06	8,3
1810	Présente 72h		29	Plectin	531,5	5,7	PLEC_HUMAN	2669,0	1,11	6,2
2712	Présente 72h		19	Transitional endoplasmic reticulum ATPase	89,3	5,0	TERA_HUMAN	1896,5	2,39	26,9
4306	Présente 72h		27	Prelamin-A/C	74,1	6,6	LMNA_HUMAN	1860,4	3,01	32,7
Protéines surexprimées										
8008	12,33		12	Peroxiredoxin-1	22,1	9,2	PRDX1_HUMAN	755,5	2,36	46,7
1604	4,58	oui	10	78 kDa glucose-regulated protein	72,3	4,9	GRP78_HUMAN	860,9	6,08	17,3
			11	Protein disulfide-isomerase A4	72,9	4,8	PDIA4_HUMAN	855,4	4,41	14,7
1603	4,21	oui	4	Cytoplasmic dynein 1 intermediate chain 2	71,4	4,9	DC1I2_HUMAN	449,3	6,02	7,1
			5	78 kDa glucose-regulated protein	72,3	4,9	GRP78_HUMAN	353,7	5,95	8,7
5706	3,95		19	Gelsolin	85,6	5,9	GELS_HUMAN	1586,7	6,41	26,9
4609	3,09	oui	5	Kinesin light chain 4	68,6	5,8	KLC4_HUMAN	543,9	2,17	9,2
			3	Serum albumin	69,3	5,9	ALBU_HUMAN	300,8	2,01	3,9
			2	Fragile X mental retardation syndrome-related protein 1	69,7	5,8	FXR1_HUMAN	253,0	1,88	2,6
4101	2,59		4	Latexin	25,7	5,5	LXN_HUMAN	369,3	2,10	15,3
9503	2,54		14	Delta-1-pyrroline-5-carboxylate dehydrogenase, mitochondrial	61,7	9,1	AL4A1_HUMAN	993,9	8,08	26,3
8407	2,36		22	Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial	61,3	8,5	MCCB_HUMAN	1859,4	2,17	35,9
8207	2,35		17	Alcohol dehydrogenase [NADP+]	36,5	6,4	AK1A1_HUMAN	1379,4	3,15	39,7
8408	2,27	oui	28	Alpha-aminoadipic semialdehyde dehydrogenase	58,5	9,1	AL7A1_HUMAN	2075,2	4,67	40,4
			9	Cytosol aminopeptidase	56,1	9,0	AMPL_HUMAN	680,4	5,57	23,7
5107	2,19		3	Isopentenyl-diphosphate Delta-isomerase 1	26,3	5,9	IDII_HUMAN	253,1	3,12	14,5
2605	2,11		23	Protein disulfide-isomerase A4	72,9	4,8	PDIA4_HUMAN	1784,9	2,19	23,4
Protéines sous-exprimées										
6312	0,41		23	Elongation factor 2	95,3	6,4	EF2_HUMAN	1962,7	1,97	22,5
8701	0,41		9	Proliferation-associated protein 2G4	43,8	6,1	PA2G4_HUMAN	678,7	1,35	19,3
4507	0,25		8	Methionine aminopeptidase 2	52,9	5,5	AMPM2_HUMAN	647,6	2,20	21,8

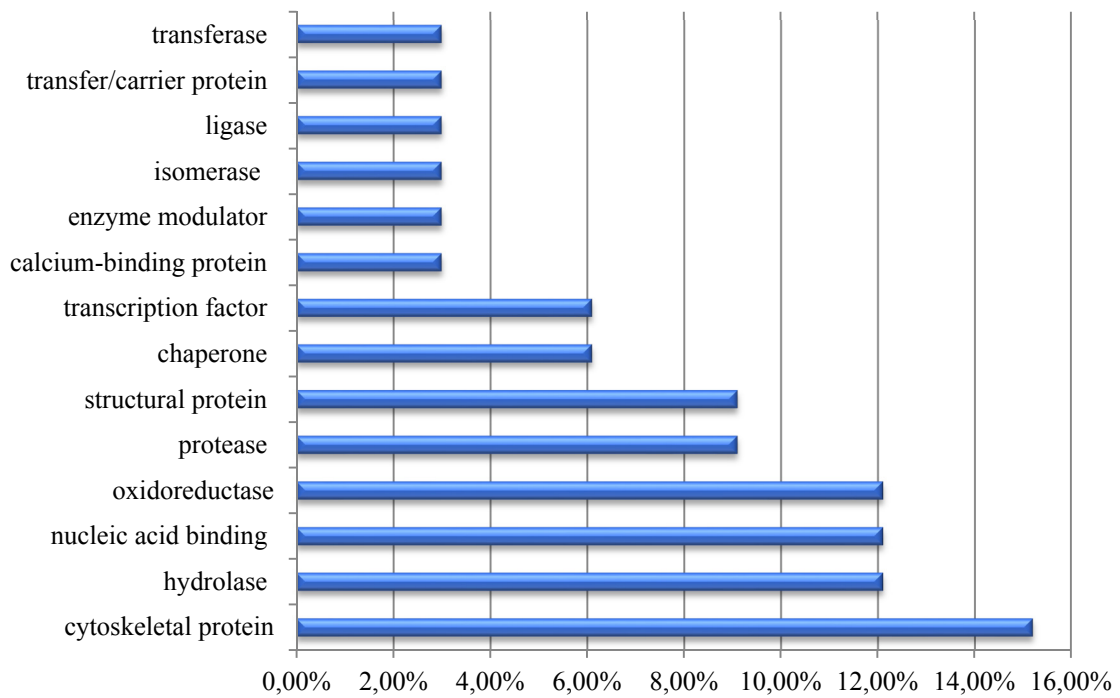


Figure 43 : Classification des 20 protéines différemment exprimées

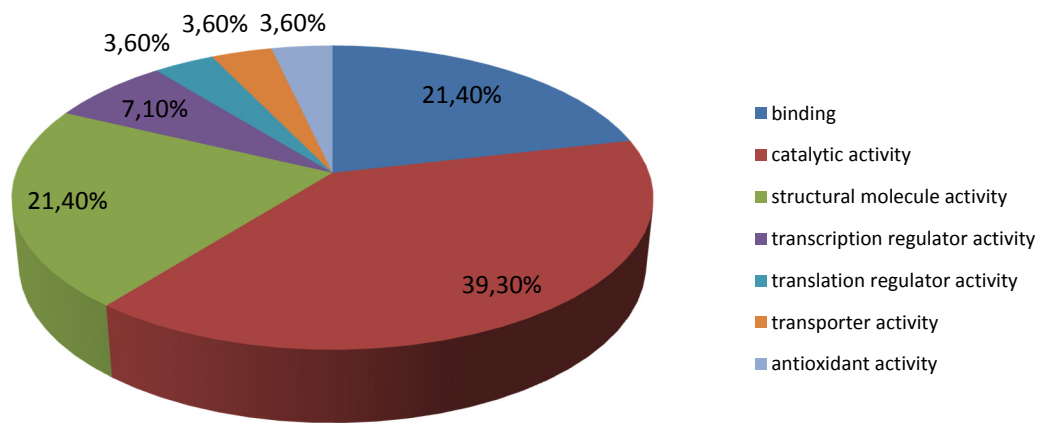


Figure 44 : Fonctions moléculaires des 20 protéines différemment exprimées

2.5) Discussion biologique

En comparant les résultats des gels 2D (Figure 45), on remarque qu'avec l'ajout du supplément B27 au milieu de culture cellulaire à la place du SVF, la quantité de BSA provenant du sérum et contaminant le gel diminue, mais n'est pas éliminée totalement.

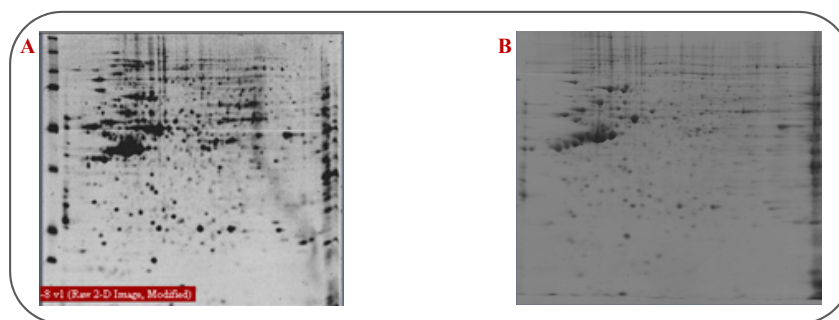


Figure 45 : Résultats des gels 2D montrant une diminution de la quantité d'albumine après utilisation d'un supplément B27 au lieu du sérum SVF : A- avec SVF. B- avec B27

Malgré les avantages qu'offrent les gels 2D, il est impossible de séparer les protéines ayant des pI et des masses très proches dans un même spot, comme le cas du spot 4609. Les analyses sur Proteinscape ont permis d'identifier 3 protéines dans ce spot ayant des pI de 5,8-5,9 et des masses entre 68,6 et 69,7 kDa.

20 protéines différemment exprimées ont été détectées sur gels 2D à l'aide du logiciel PD-Quest: 5 sont apparues, 12 sont surexprimées ($BaP72h/BaP24h > 2$) et 3 sous-exprimées ($BaP72h/BaP24h < 0,5$) dans les cellules traitées au BaP pendant 72h. Ces protéines sont impliquées dans la prolifération cellulaire (Peroxyredoxin-1), l'angiogenèse (Heat shock protein beta-1), l'apoptose (78 kDa glucose-regulated protein).

Parmi ces 20 protéines, la gelsoline ($BaP72h/BaP24h = 3,95$) et la « proliferation-associated protein 2G4 » ($BaP72h/BaP24h = 0,41$) sont des protéines d'intérêt impliquées dans le cancer.

La gelsoline semble avoir des rôles complexes dans la tumorigenèse, avec preuves à l'appui de son implication contradictoire dans la suppression des tumeurs et la progression maligne à la fois. Elle est sous-exprimée dans 60 à 90% des tumeurs lors de la carcinogenèse du sein (Asch, Head et al. 1996), du colon, de l'estomac et de la vessie (Kuzumaki, Fujita et al. 1998), de la prostate (Prasad, Thraves et al. 1997) et des poumons (Dosaka-Akita, Hommura et al. 1998), donc sa perte favorise l'oncogenèse.

Cependant, il a été prouvé que les stades progressivement malignes du cancer sont associés à une expression élevée de gelsoline dans le cancer du poumon à petites cellules (Shieh, Godleski et al. 1999) et le carcinome rénal (Visapää, Bui et al. 2003). Il est probable que le rôle de la gelsoline diffère au cours de la progression tumorale et qu'elle coopère avec d'autres facteurs oncogéniques pour accélérer la progression.

La « prolifération-associated protein 2G4 », connue sous « ErbB3-binding protein1 » (EBP1) est sous-exprimée dans le cancer de la prostate (Zhang, Wang et al. 2005, Zhang, Linn et al. 2008) et de la vessie (He, Ling et al. 2013).

Les résultats de ces analyses comparatives entre le protéome total des cellules HepG2 exposées au BaP et au BeP étant encourageants, nous nous sommes orientés vers l'étude des protéines du sécrétome des cellules HepG2 exposées ou non au BaP et BeP.

B) Analyse différentielle du sécrétome

Contexte : Le sécrétome étant un modèle séduisant puisque les protéines libérées dans le milieu extracellulaire sont très souvent retrouvées dans le plasma, nous avons décidé de précipiter des protéines du sécrétome provenant des cellules cultivées dans le milieu RPMI complété par le supplément B27, de séparer les protéines sur gels 2D et de les identifier par MALDI-TOF(/TOF) et/ou par nanoLC-MS/MS.

1) Séparation sur gels 2D

Les protéines ont été précipitées au TCA (acide trichloroacétique) (Chevallet, Diemer et al. 2007). Ce protocole a été appliqué sur deux échantillons afin d'évaluer sa pertinence à l'aide de ces premiers résultats et afin de ne pas gaspiller tous les échantillons du sécrétome. Le culot protéique obtenu a été resuspendu dans du tampon UTC. Les résultats du dosage Bradford et des gels 2D des deux extraits protéiques du sécrétome des cellules exposées à 10 μ M pendant 6h et 72h sont représentés, respectivement, dans le tableau 10 et la figure 46.

Tableau 10 : Résultats du dosage Bradford des extraits protéiques des sécrétomes des cellules exposées à 10 μ M de BaP pendant 6h et 72h.

Echantillon	Concentration μg/μL	Rendement μg (dans 600μL)
BaP 6h	30,53	18 318
BaP 72h	38	22 800

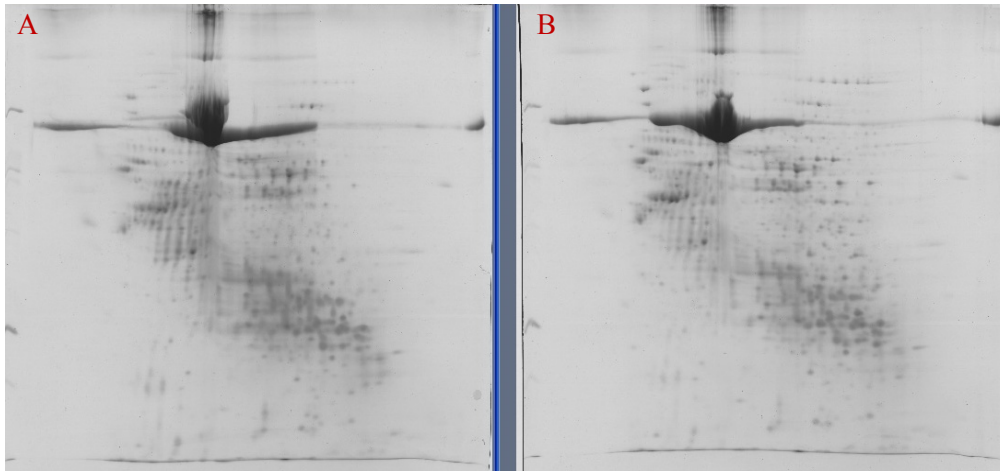


Figure 46 : Gels 2D des extraits protéiques du sécrétome des cellules exposées à 10 μ M de BaP pendant 6h (A) et 72h(B)

Comme le montre la figure 46, malgré l'utilisation du supplément de culture B27 (ne contenant pas de sérum), la BSA est encore majoritairement présente. Sa quantité est estimée à 95% de l'extrait protéique de par l'intensité de coloration des spots correspondant à ce contaminant sur le gel 2D. Néanmoins, l'apparition de nombreux spots révélés à l'argent sous la BSA, sont source d'encouragement.

Pour essayer d'améliorer la visualisation de ces protéines minoritaires sous la BSA, nous avons ensuite employé une technique basée sur l'utilisation de filtres de concentration (Amicon Ultra-15, Millipore), le but étant d'éliminer les protéines ayant un poids moléculaire supérieur à 50 kDa (PM (BSA) = 66,463 kDa). Après passage des sécrétomes sur des filtres de concentration de 50kDa, précipitation et dosage des protéines, les gels d'électrophorèse bidimensionnels obtenus sont représentés ci-dessous (Figure 47). Les concentrations des deux extraits protéiques du sécrétome des cellules exposées à 10 μ M pendant 6h et 72h étaient de 3,04 et 9,65 μ g/ μ L, respectivement.

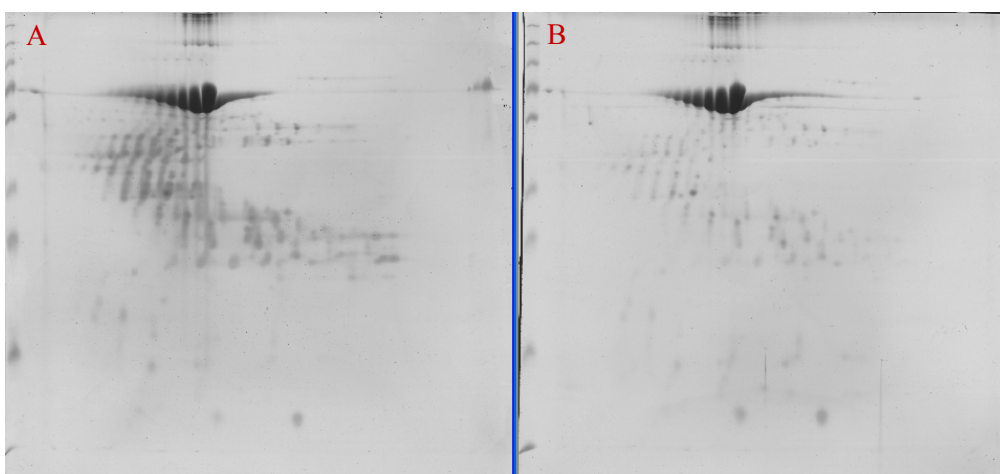


Figure 47 : Gels 2D des extraits protéiques après concentration (50 kDa) du sécrétome des cellules exposées à 10 μ M de BaP pendant 6h (A) et 72h (B)

Le différentiel sur gels 2D était très difficile à réaliser car une contamination importante par l'albumine était toujours présente et ceci malgré l'utilisation des filtres de concentration. Cette contamination peut s'expliquer par le fait que l'albumine a pu tout de même traverser la membrane des filtres. Elle a été identifiée non seulement à ses coordonnées théoriques de poids moléculaire et de pI, mais également dans des spots 2D pouvant correspondre à des fragments de BSA de plus petit poids moléculaire.

2) Séparation sur gels 1D

Suite à ces problèmes rencontrés, nous avons ensuite envisagé de séparer les mêmes extraits protéiques sur gels 1D (gradient d'acrylamide 5-13%) et de découper la piste des protéines ayant un poids moléculaire inférieur à 50 kDa afin de les identifier par nanoLC-MS/MS. Le résultat des gels 1D est représenté dans la figure 48.

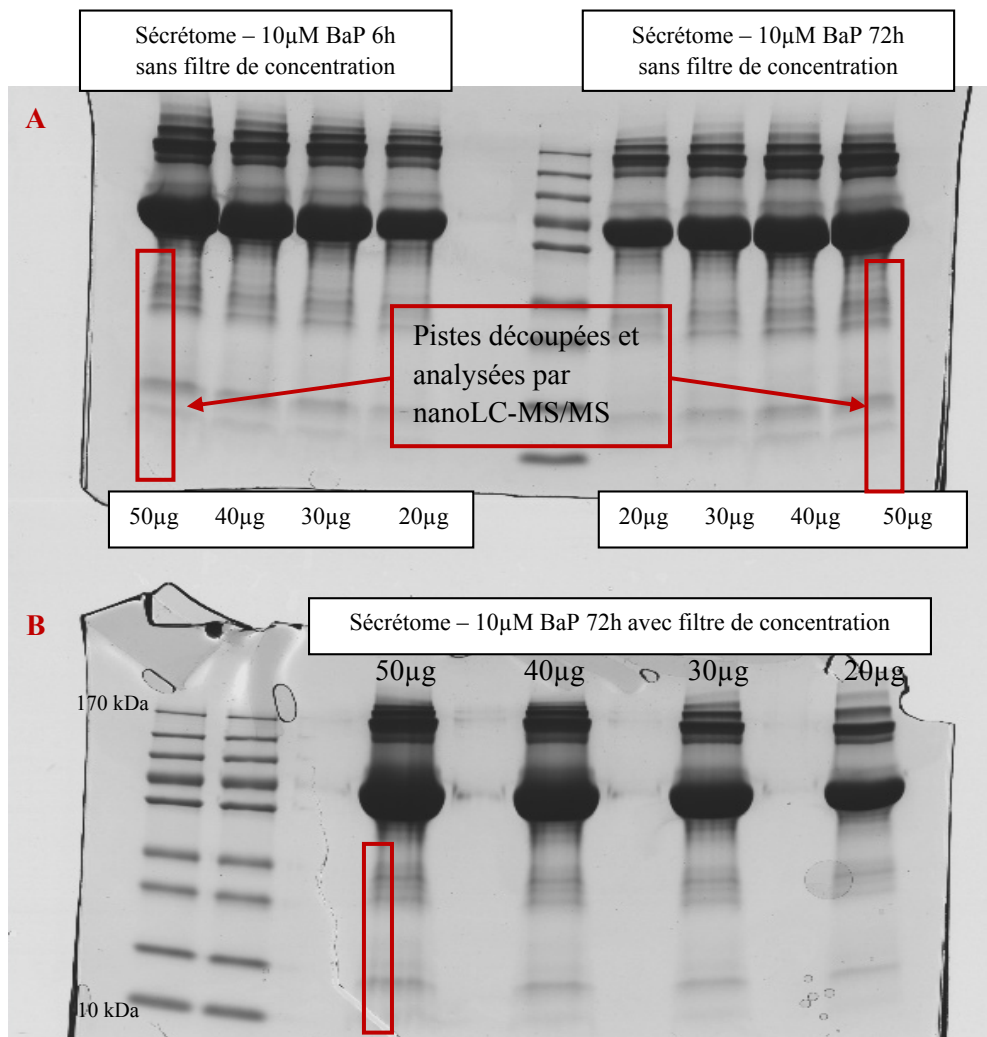


Figure 48 : Gels 1D des extraits protéiques des sécrétomes des cellules HepG2 exposées à 10µM de BaP pendant 6h et 72h, (A) sans filtre de concentration, (B) avec filtre de concentration 50 kDa

Après analyse nanoLC-MS/MS, l'identification des protéines contenues dans les échantillons s'est faite par comparaison avec la banque de séquences protéiques SwissProt (version 20120111, 21918 séquences), en spécifiant la taxonomie *Homo sapiens* (human). Parmi 392 protéines identifiées dans l'extrait protéique du sécrétome passé sur un filtre de concentration 50 kDa (BaP72h), seulement 21 protéines ont été validées avec un minimum de 2 peptides.

L'albumine étant majoritaire (plus haut score Mascot (685,7)), nous nous sommes orientés vers l'utilisation du kit Proteominer (Bio-Rad) afin d'enrichir des protéines moyennement et faiblement abondantes. C'est un outil de préparation des échantillons qui diminue le pourcentage de protéines abondantes, et ceci par un principe d'enrichissement et non de déplétion. Cette technologie utilise une banque large et très diverse d'hexapeptides fixés sur billes de ligand. Quand un mélange biologique complexe est incubé avec les billes, les protéines abondantes saturent leur ligand spécifique et l'excès de protéines est ainsi éliminé par lavage. En revanche, les protéines faiblement représentées sont concentrées sur leur ligand spécifique.

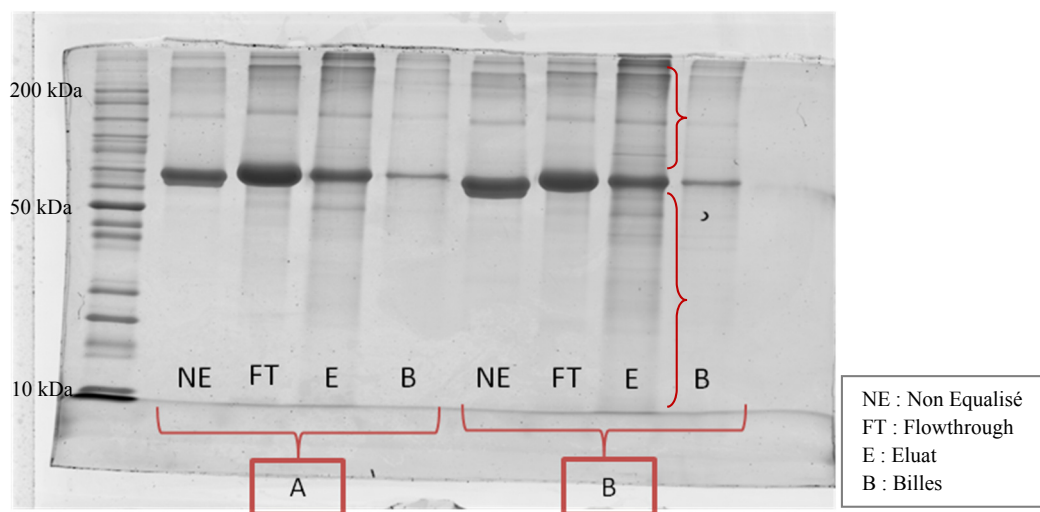


Figure 49 : Gels 1D des extraits protéiques des sécrétomes des cellules HepG2 exposées à 10µM de BaP pendant 6h (A) et 72h (B) après enrichissement des protéines faiblement abondantes

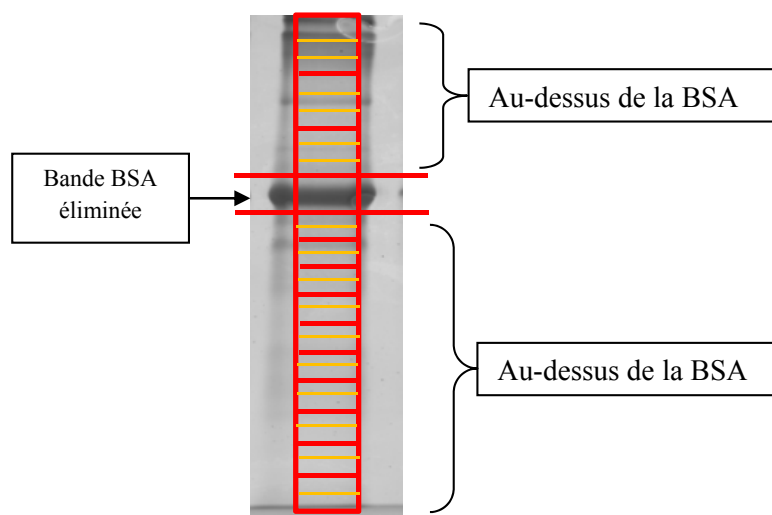


Figure 50 : Zones de découpage de la piste de l'éluat protéique du sécrétome des cellules HepG2 exposées à 10µM de BaP pendant 72h

L'extrait protéique traité par le kit Proteominer a ensuite été déposé sur un gel dénaturant SDS-PAGE. Toutes les bandes du gel (exception faite de celle contenant la BSA donc 56 bandes au total) (Figure 50) ont été lavées, réduites, alkylées et digérées. Les peptides extraits ont été injectés sur la chaîne d'analyse nanoLC-MS/MS. Les fichiers bruts obtenus par l'instrument ont été convertis en fichiers .mgf : l'ensemble de ces fichiers .mgf a été fusionné pour ne donner qu'un seul et même fichier .mgf regroupant l'ensemble des spectres MS/MS acquis sur la totalité de la piste SDS-PAGE. L'identification des protéines reliées à ces spectres de fragmentation a ensuite été réalisée à l'aide du logiciel Mascot pour la recherche en banque de données, par l'intermédiaire de Proteinscape 3.1 : 259 protéines ont été identifiées, mais seulement 18 protéines ont été validées par un minimum de 2 peptides.

3) *Optimisation de l'étape de culture cellulaire*

Des contaminations étant encore rencontrées malgré l'utilisation du supplément de culture B27 et le kit Proteominer, nous avons continué l'optimisation de la préparation du sécrétome en concentrant nos efforts sur l'étape de culture cellulaire. L'objectif de cette étape d'optimisation était de pouvoir éliminer la BSA, la protéine la plus abondante du sérum de veau fœtal ajouté au milieu de culture.

Les cellules ont été exposées ou non à 10µM du BaP ou du BeP, et cette fois-ci, lavées 6 fois au PBS et incubées dans un milieu DMEM/F12 (milieu sans sérum) pendant 2h, 6h, 24h et 72h. Le protocole depuis la culture cellulaire jusqu'à la précipitation des protéines du sécrétome de ces cellules est détaillé dans la partie « méthodes » du chapitre 2.

Les résultats du dosage Bradford des extraits protéiques des sécrétomes des cellules en triplicats exposées ou non à 10 μ M du BaP ou du BeP sont représentés dans le tableau 11.

Tableau 11 : Résultats du dosage Bradford des extraits protéiques des sécrétomes

Echantillon	Concentration μg/μL	Rendement μg (dans 75μL)
Control 2h (1)	0,57	42,72
Control 2h (2)	0,49	36,97
Control 2h (3)	0,46	34,72
BeP 2h (1)	0,66	49,63
BeP 2h (2)	0,43	32,91
BeP 2h (3)	0,79	59,45
BaP 2h (1)	5,16	387,5
BaP 2h (2)	7,4	555,37
BaP 2h (3)	5,31	398,81
Control 6h (1)	1,51	113,64
Control 6h (2)	0,48	35,82
Control 6h (3)	0,59	44,54
BeP 6h (1)	0,83	63,36
BeP 6h (2)	0,71	53,27
BeP 6h (3)	0,63	47,45
BaP 6h (1)	3,27	245,71
BaP 6h (2)	2,08	156,43
BaP 6h (3)	2,6	195
Control 24h (1)	0,77	57,64
Control 24h (2)	0,42	31,45
Control 24h (3)	0,48	36,18
BeP 24h (1)	0,86	64,62
BeP 24h (2)	0,85	64,18
BeP 24h (3)	1,25	94,05
BaP 24h (1)	1,98	148,18
BaP 24h (2)	2,29	171,81
BaP 24h (3)	2,59	194,24
Control 72h (1)	1,70	127,57
Control 72h (2)	1,34	100,91
Control 72h (3)	0,90	67,45
BeP 72h (1)	1,58	118,72
BeP 72h (2)	1,22	91,45
BeP 72h (3)	1,26	94,36
BaP 72h (1)	2,44	182,86
BaP 72h (2)	1,77	132,86
BaP 72h (3)	2,5	187,85

Dix μ g de l'extrait protéique BaP 72h a été migré sur un gel monodimensionnel dénaturant SDS-PAGE (Figure 51). Le gel était super-impressionnant : on peut bien remarquer

la diminution de la quantité de la BSA par rapport à nos résultats précédents, et ceci même sans utiliser le kit Proteominer.

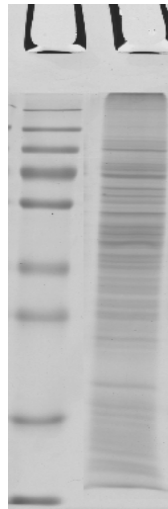


Figure 51 : Gels 1D de l'extrait protéique du sécrétome des cellules HepG2 exposées à 10 μ M de BaP 72h

Toute la piste des protéines a été découpée, digérée et analysée par nanoLC-MS/MS. En utilisant les mêmes paramètres cités dans le paragraphe « séparation sur gels 1D », nous avons identifié 615 protéines dans l'extrait protéique du sécrétome des cellules HepG2 exposées à 10 μ M de BaP72h.

Les résultats de la digestion *in-gel* ont été comparés à ceux de la digestion liquide du même extrait protéique (protocole détaillé dans la partie expérimentale du chapitre 2). 1258 protéines ont été identifiées dans le sécrétome des cellules HepG2 exposées à 10 μ M de BaP72h.

En comparant les résultats des deux digestions, nous avons remarqué que le nombre des protéines identifiées à partir d'une digestion liquide est le double de celui des protéines identifiées à partir de la digestion *in-gel*. De plus, la digestion des protéines en solution est plus reproductible que celle *in-gel*. Elle a donc été envisagée pour la suite de ce projet (chapitre 2).

L'étape de préparation des échantillons est cruciale pour obtenir des échantillons de bonne qualité (séparation électrophorétique, diminution d'éventuelles contaminations par des protéines majoritaires du sérum, choix du type d'électrophorèse, digestion liquide ou *in-gel*, conditions de digestion trypsique, etc).

Une fois ces étapes optimisées, l'analyse des échantillons par le couplage entre la chromatographie liquide nano-débit et le spectromètre de masse en tandem s'avère relativement aisée : en effet, les injections sont réalisées de manière automatisée grâce à un passeur

d'échantillons programmable et à un bon entretien des composantes de la chaîne instrumentale (respect de temps de régénération des colonnes nanoLC ainsi que de leur limite de charge, contrôle de la sensibilité de détection à l'aide d'un échantillon connu injecté régulièrement, outils de diagnostic performants en cas de bouchage des nano-capillaires, etc). L'obtention des fichiers nécessaires à l'interrogation en banque de donnée permet ainsi d'accéder à une autre étape cruciale de l'analyse protéomique : le traitement des données d'identification et/ou de quantification. Cette partie, appelée analyse bio-informatique, doit être menée avec rigueur de manière à faire ressortir l'information biologique pertinente et de manière à générer le moins de fausses interprétations possible. Cette analyse est ensuite complétée par l'intégration de données d'annotations biologiques et de données bibliographiques relatives aux protéines identifiées et quantifiées. Les résultats de ces analyses ont abouti à une soumission d'un article dans MCP (chapitre 2) et ceci par l'intermédiaire d'une ultime étape d'analyse bio-informatique des données de spectrométrie de masse (chapitre 3) qui a été abordée dans mon sujet d'étude.

Références bibliographiques

A

Asch, H. L., K. Head, Y. Dong, F. Natoli, J. S. Winston, J. L. Connolly and B. B. Asch (1996). "Widespread loss of gelsolin in breast cancers of humans, mice, and rats." Cancer research **56**(21): 4841-4845.

B

Bauer, A., S. Chauvet, O. Huber, F. Usseglio, U. Rothbacher, D. Aragnol, R. Kemler and J. Pradel (2000). "Pontin52 and Reptin52 function as antagonistic regulators of β -catenin signalling activity." The EMBO journal **19**(22): 6121-6130.

C

Chevallet, M., H. Diemer, A. Van Dorssealer, C. Villiers and T. Rabilloud (2007). "Toward a better analysis of secreted proteins: the example of the myeloid cells secretome." Proteomics **7**(11): 1757-1770.

D

Dosaka-Akita, H., F. Hommura, H. Fujita, I. Kinoshita, M. Nishi, T. Morikawa, H. Katoh, Y. Kawakami and N. Kuzumaki (1998). "Frequent loss of gelsolin expression in non-small cell lung cancers of heavy smokers." Cancer research **58**(2): 322-327.

G

Goldring, C. E., N. R. Kitteringham, R. Jenkins, C. A. Lovatt, L. E. Randle, A. Abdullah, A. Owen, X. Liu, P. J. Butler and D. P. Williams (2006). "Development of a transactivator in hepatoma cells that allows expression of phase I, phase II, and chemical defense genes." American Journal of Physiology-Cell Physiology **290**(1): C104-C115.

H

Haurie, V., L. Ménard, A. Nicou, C. Touriol, P. Metzler, J. Fernandez, D. Taras, P. Lestienne, C. Balabaud and P. Bioulac-Sage (2009). "Adenosine triphosphatase pontin is overexpressed in hepatocellular carcinoma and coregulated with reptin through a new posttranslational mechanism." Hepatology **50**(6): 1871-1883.

He, H.-c., X.-h. Ling, J.-g. Zhu, X. Fu, Z.-d. Han, Y.-x. Liang, Y.-h. Deng, Z.-y. Lin, G. Chen and Y.-f. Chen (2013). "Down-regulation of the ErbB3 binding protein 1 in human bladder cancer promotes tumor progression and cell proliferation." Molecular biology reports: 1-7.

K

Kassie, F., L. B. Anderson, L. Higgins, Y. Pan, I. Matisse, M. Negia, P. Upadhyaya, M. Wang and S. S. Hecht (2008). "Chemopreventive agents modulate the protein expression profile of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone plus benzo [a] pyrene-induced lung tumors in A/J mice." Carcinogenesis **29**(3): 610-619.

Kim, J. H., B. Kim, L. Cai, H. J. Choi, K. A. Ohgi, C. Tran, C. Chen, C. H. Chung, O. Huber and D. W. Rose (2005). "Transcriptional regulation of a metastasis suppressor gene by Tip60 and β -catenin complexes." Nature **434**(7035): 921-926.

Kuzumaki, N., H. Fujita, M. Tanaka, N. Sakai, M. Ohtsu, H. Maruta and K. Kohama (1998). "Tumor suppressive function of gelsolin." G Proteins, Cytoskeleton and Cancer: 121.

M

Masson, O., C. Pr bois, D. Derocq, A. Meulle, C. Dray, D. Daviaud, D. Quilliot, P. Valet, C. Muller and E. Liaudet-Coopman (2011). "Cathepsin-D, a key protease in breast cancer, is up-regulated in obese mouse and human adipose tissue, and controls adipogenesis." PloS one **6**(2): e16452.

Menssen, A. and H. Hermeking (2002). "Characterization of the c-MYC-regulated transcriptome by SAGE: identification and analysis of c-MYC target genes." Proceedings of the National Academy of Sciences **99**(9): 6274-6279.

Min, L., S. He, Q. Chen, F. Peng, H. Peng and M. Xie (2011). "Comparative proteomic analysis of cellular response of human airway epithelial cells (A549) to benzo (a) pyrene." Toxicology Mechanisms and Methods **21**(5): 374-382.

P

Prasad, S. C., P. J. Thraves, A. Dritschilo and M. R. Kuettel (1997). "Protein expression changes associated with radiation-induced neoplastic progression of human prostate epithelial cells." Electrophoresis **18**(3-4): 629-637.

R

Rousseau, B., L. M nard, V. Haurie, D. Taras, J. F. Blanc, F. Moreau-Gaudry, P. Metzler, M. Hugues, S. Boyault and S. Lemi re (2007). "Overexpression and role of the ATPase and putative DNA helicase RuvB-like 2 in human hepatocellular carcinoma." Hepatology **46**(4): 1108-1118.

S

Shieh, D. B., J. Godleski, J. E. Herndon, T. Azuma, H. Mercer, D. J. Sugarbaker and D. J. Kwiatkowski (1999). "Cell motility as a prognostic factor in stage I nonsmall cell lung carcinoma." Cancer **85**(1): 47-57.

V

Verma, N., A. W. Rettenmeier and S. Schmitz-Spanke (2010). "Proteomic analysis of primary porcine bladder epithelial cells after BaP exposure." Toxicology Letters **196**: S71.

Visapää, H., M. Bui, Y. Huang, D. Seligson, H. Tsai, A. Pantuck, R. Figlin, J. Y. Rao, A. Beldegrun and S. Horvath (2003). "Correlation of Ki-67 and gelsolin expression to clinical outcome in renal clear cell carcinoma." Urology **61**(4): 845-850.

W

Weiske, J. and O. Huber (2005). "The histidine triad protein Hint1 interacts with Pontin and Reptin and inhibits TCF- β -catenin-mediated transcription." Journal of cell science **118**(14): 3117-3129.

Xu, J., M. Ma and W. M. Purcell (2003). "Characterisation of some cytotoxic endpoints using rat liver and HepG2 spheroids as in vitro models and their application in hepatotoxicity studies. I. Glucose metabolism and enzyme release as cytotoxic markers." Toxicology and applied pharmacology **189**(2): 100-111.

Z

Zhang, Y., D. Linn, Z. Liu, J. Melamed, F. Tavora, C. Y. Young, A. M. Burger and A. W. Hamburger (2008). "EBP1, an ErbB3-binding protein, is decreased in prostate cancer and implicated in hormone resistance." Molecular cancer therapeutics **7**(10): 3176-3186.

Zhang, Y., X.-W. Wang, D. Jelovac, T. Nakanishi, M.-h. Yu, D. Akinmade, O. Goloubeva, D. D. Ross, A. Brodie and A. W. Hamburger (2005). "The ErbB3-binding protein Ebp1 suppresses androgen receptor-mediated gene transcription and tumorigenesis of prostate cancer cells." Proceedings of the National Academy of Sciences of the United States of America **102**(28): 9890-9895.

Chapitre 2 : “Secretome analysis using Label-Free quantitative proteomics to discover potential cancer biomarkers of Benzo(a)Pyrene exposure”

Article soumis dans MCP

Contexte : Les dernières optimisations de la préparation d'échantillons et les résultats de l'identification des protéines extraites du sécrétome des cellules hépatiques humaines HepG2 ont abouti à un article soumis dans le journal MCP (Molecular & Cellular Proteomics ; Impact factor 7.398).

Dans cet article, nous avons développé une approche protéomique quantitative nanoLC-MS/MS label-free pour identifier des biomarqueurs liés à l'exposition au BaP dans le sécrétome des cellules HepG2 exposées au BaP vs. des cellules non exposées et exposées au BeP. Les cellules traitées au BeP ont été choisies comme contrôle négatif afin de distinguer les protéines spécifiques du BaP de celles spécifiques des HAP : le BeP est un isomère du BaP mais il n'est pas classifié comme étant cancérigène pour l'homme. De plus, dans cette partie, nous avons utilisé des cellules contrôles non exposées à aucun agent toxique.

Deux temps d'exposition à 10 μ M de BeP ou BaP ont été choisis : 2h et 72h afin de différencier les protéines précoces de celles tardives, respectivement. L'identification et la quantification des protéines s'est faite grâce au logiciel Progenesis LC-MS (Nonlinear Dynamics, Newcastle-upon-Tyne, UK). Le choix de ce logiciel, en le comparant à deux autres : ProteinPilot/PeakView/MarkerView (AB Sciex, CA, USA) et Skyline (MacCoss Laboratory, University of Washington), est développé dans le chapitre 3.

Les résultats des protéines sont détaillés dans l'article qui suit.

Secretome analysis using Label-Free quantitative proteomics to discover potential cancer biomarkers of Benzo(a)Pyrene exposure

Marianne Ibrahim^{1,4}, Lauriane Kuhn², Zeina Dagher³, Johana Chicher², Ramez Chahine⁴,
Philippe Hammann², Emmanuelle Leize-Wagner^{1*}

¹ Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes, CNRS-UMR7140, Université de Strasbourg, 67008, Strasbourg, France

² Plateforme Protéomique Strasbourg Esplanade (IBMC), CNRS-FRC1589, 67084 Strasbourg, France

³ Equipe Molécules Bioactives, EDST, Université Libanaise, Hadath, Liban

⁴ Laboratoire Stress Oxydatif et Antioxydants, EDST et Faculté des Sciences Médicales, Université Libanaise, Hadath, Liban

* Corresponding author :

Emmanuelle LEIZE-WAGNER

Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (LSMIS)

UMR 7140 UDS-CNRS "Chimie de la Matière Complexe"

1 rue Blaise Pascal 67008 Strasbourg

+ 33 (0)3 68 85 16 26

leize@unistra.fr

RUNNING TITLE

Secretome analysis on human cells exposed to BaP

ABBREVIATIONS

BaP, Benzo(a)pyrene; BeP, Benzo(e)pyrene; GO, Gene Ontology; FDR, False Discovery Rate; PCA, Principal Component Analysis; XIC, eXtracted Ion Chromatogram

SUMMARY

Benzo(a)pyrene (BaP) belongs to a class of polycyclic aromatic hydrocarbon and is a well-recognized environmental micropollutant, previously reported as a potent human carcinogen. In the present study, we performed a nanoLC-MS/MS-based label-free quantitative proteomics approach to identify potential biomarkers of exposure in the secretome of BaP-treated *vs.* non-treated and Benzo(e)pyrene (BeP)-treated human hepatoma cell line HepG2. BeP-treated cells were chosen as a negative control to distinguish the BaP-specific from the HAP-specific regulated proteins: BeP is a BaP isomer but unlike BaP, it is not classifiable as to its carcinogenicity to humans. Cells were exposed to 10 μ M BaP or BeP and incubated for 2h and 72h to better differentiate between proteins having an early onset and a delayed onset, respectively. 847 proteins, with 51.95% of them predicted as secreted, have been identified and quantified using Progenesis LC-MS software. 55 proteins were seen as being highly up-regulated with a fold change of at least 5. Most of these up-regulated proteins were focused in cancer-related activities as revealed by literature and by a Gene Ontology (GO) molecular function classification. The potential use of the regulated proteins found in this proteomic study as early indicators of BaP exposure is discussed: further validation of expression of these proteins in the plasma of BaP-exposed population will assist in the development of biomarkers that will greatly improve early detection, prognosis, prediction of treatment response and prevention.

KEYWORDS

Benzo(a)pyrene, Benzo(e)pyrene, cancer biomarkers, secretome, proteomics, label-free quantification, mass spectrometry

INTRODUCTION

Polycyclic aromatic hydrocarbons (PAHs) are widespread environmental contaminants, notably found in ambient air, tobacco smoke, broiled and smoked food (1). Food Processing (such as drying and smoking) and cooking at high temperatures (grilling, roasting, frying) are major sources of PAHs emission (2, 3). People are therefore exposed to PAHs through

smoking, food and medicine consumption and via polluted air and working environments. Depending on the source of exposure, uptake occurs by inhalation, dermal absorption, or ingestion. Ingestion and inhalation are the two major routes of exposure for the general population (4, 5). Many studies have been carried on the carcinogenic quality (6) and mutagenic potential (7, 8) of PAHs especially since they are widespread in the environment.

Among these PAHs, benzo(a)pyrene (BaP), classified as a potent human carcinogen in International Agency for Research on Cancer (IARC) group 1 (6, 9), is one of the most abundant and thus thoroughly studied carcinogenic PAHs in tobacco smoke (10). Due to its clear genotoxic effects, BaP has been used since a long time as a prime example in toxicological and pharmaceutical studies in various systems (11) and has been associated with lung cancer (12, 13).

BaP is often used as an example while illustrating the metabolism of PAHs. After its uptake from food or contaminated aerosols, BaP is diffused into the liver, kidney, lung and bladder (14, 15). Liver is the organ that contains the highest amount of enzymes required for the bioactivation of BaP (16, 17). Several studies have described two major pathways for the metabolic activation of BaP. The first pathway is the activation of BaP into anti-benzo(a)pyrene-trans-7,8-dihydrodiol-9,10-epoxide (BPDE), its ultimate carcinogenic metabolite: this process is catalyzed by liver microsomal enzymes (P-450 family, epoxide hydrolase, prostaglandin H synthase) and successively implies the BP-7,8-oxide and BP-7,8-dihydrodiol intermediates. BPDE can interact with DNA to result in DNA adduct formation and therefore leads to mutations (18, 19). PAH-induced cancer begins with the formation of a covalent bond between DNA bases and the carcinogen. The (+)-anti-isomer form of the dihydrodiol epoxides has the greatest carcinogenic effect. The second pathway involves the formation of radical cation intermediates *via* enzymatically catalyzed one-electron oxidation (20).

In order to effectively and promptly treat diseases related to BaP exposure, especially cancer, it is important to develop early-diagnosis strategies such as looking for specific BaP exposure biomarkers. Biomarkers are endogenous characteristics measured and evaluated as markers of

normal biological processes, pathogenic processes, or responses to an external event. The use of biomarkers is not new in concept but has broadened by the wide-ranging technological advances made in the fields of molecular and cellular biology and biomedical engineering in recent decades. They are of high interest in drug discovery, and have the potential to not only help identify who will develop cancer but also to determine when the event is most likely to occur. These biomarkers could eventually be converted into novel diagnostic marker assays. Proteomic studies have previously identified novel protein biomarkers in various biological matrices from cell cultures (lysates, supernatants) to human clinical samples (serum, plasma, and urine) (21, 22). The secreted protein profiles (secretome or supernatant) of various cells such as hepatocytes have been investigated using proteomic approaches (23). The advantage of studying these soluble proteins, secreted into the extracellular environment of the cells, comes from their use as potential targets for diagnosis and treatment of several diseases. Most of the proteomics studies analyzed cell culture media and detected the secreted proteins by one (1D) or two-dimensional (2D) electrophoresis or nano-Liquid Chromatography (nanoLC), both combined with mass spectrometry (MS and MS/MS). However, the 2D gel-based method has drawbacks with respect to the identification of proteins in the basic pI range, extremely low- or high-molecular weight range, hydrophobicity, or low abundance (24, 25). The method to identify novel biomarkers should be specific and sensitive enough to facilitate early detection of these proteins. Recently, nano-LC coupled with tandem mass spectrometry (MS/MS) has been widely used for shotgun proteomics. Comparative quantitative proteomic analyses reported in the literature still preferably rely on approaches requiring labeling. Most of the labeling-based quantification approaches have numerous limitations thus explaining the decreasing

number of studies using this strategy: reagents are very costly, preparation requires a lengthy and tedious work, labeling is often incomplete and sophisticated bioinformatics tools often lack to correctly quantify paired peptides when one of the 2 entities is close to the background (infinite fold-change). Recent innovations in nanoLC-MS-based methods have facilitated quantitative and functional proteomic analyses of large numbers of proteins derived from complex samples without any need for protein or peptide labeling (26-28). Mass spectrometry-based label-free quantitative proteomics takes advantage of the correlation between high-resolution extracted ion currents (XICs) and peptide abundances (29, 30). These approaches are more sensitive to technical deviations between nanoLC-MS runs as successive measurements must be done instead of a multiplexed analysis, therefore enhancing technical potential deviations. The reproducibility of the analytical platform and especially the nano-LC system is crucial to lead to a successful label-free quantification. The success of these quantification can be accomplished by bioinformatics software tools which allow retention times alignment and an accurate determination of the m/z values of the XICs across numerous samples (31). Thus, the aligned signals of XICs can be compared to measure significant differences between sample groups.

In the present study, we performed a quantitative label-free nanoLC-MS/MS

EXPERIMENTAL PROCEDURES

Materials - Benzo(a)pyrene and Benzo(e)pyrene were purchased from Sigma (St. Louis, MO, USA). Roswell Park Memorial Institute medium (RPMI 1640) and fetal bovine serum (FBS) were purchased from Sigma and Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (DMEM/F12) from Lonza (Basel,

study to identify potential biomarkers in the secretome of BaP-treated vs. non-treated and BeP-treated human hepatoma cell line HepG2. BeP-treated cells were chosen as negative control: BeP is a BaP isomer but unlike BaP, it is not carcinogenic (IARC group 3) (32) (33). To date, proteins in BaP-exposed cell pellet HepG2 or non-exposed cultured media HepG2 (23, 34, 35) have been exclusively analyzed by 2-DE technology and identify by mass spectrometry. In this study, we rather digested the samples in solution with trypsin and submitted the resulting peptide mixtures to a nanoLC-MS/MS analysis with a high-speed mass spectrometer. This strategy, combined with the use of a robust data processing pipeline including retention times alignment, allows us to quantify 847 proteins with False Discovery Rate (FDR) <1%. Proteins were further quantified by MS1 label-free quantification method using features intensities (extracted ion chromatograms) and individual peptides values were averaged at the protein level to allow fold change calculations between all the conditions. This label-free approach allowed us to decipher the BaP-treated HepG2-derived secreted proteins. Moreover, as cells were incubated with both compounds for two different periods, we were able to cluster data into specific protein profiles in order to diagnose biomarkers of early and late stage cancer patients.

Switzerland). DMSO was used to dissolve BaP and its final concentration in cultures was less than 0.2%. All other reagents and plastic ware were obtained from commercial sources.

Cell culture and BaP/BeP exposure - HepG2 cells (human hepatocellular liver carcinoma cell line) were cultured in Roswell Park Memorial Institute medium (RPMI 1640), supplemented with 10% FBS, 1% L-glutamine (200mM) and 1% of

the mixture P/S (10000 U/ml penicillin, 10000µg/ml streptomycin) in a CO₂ humid incubation chamber at 37 °C. Cells were seeded into 75cm² flasks. They were grown for 24h before being exposed to either BaP or BeP for an additional 24h in serum medium.

The concentration of BaP/BeP treatment used in this study was chosen based on the results of cytotoxicity tests (MTT assay). After the chemical exposure, all cultures were rinsed six times with phosphate-buffered saline (PBS 1x), and incubated in 15mL/flask of serum-free DMEM/F12 another 2h and 72h. After incubation, the secretome (conditioned medium) was collected, immediately frozen at -80° and stored for subsequent protein extraction. For the differential secretome study, three biological replicates of HepG2 cells treated or not with BaP or BeP for 2h and 72h were collected.

Cytotoxicity assays - The cytotoxicity of BaP or BeP on HepG2 cells was evaluated with the MTT assay in 96-well plates. Exponentially growing cells were seeded at a density of 10,000 cells per well and allowed to grow overnight. After removing the cell culture medium and washing the cells in phosphate-buffered saline (PBS, pH 7.4), HepG2 cells were incubated in a new medium containing serial dilutions of BaP or BeP. The working solutions were freshly prepared in RPMI. The maximum DMSO concentration in the medium was less than 0.2% v/v including the control groups. The MTT assay involves reduction of tetrazolium salt (3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide) by viable cells to purple formazan. Briefly, after BaP or BeP treatment, the cells were washed twice with PBS and then incubated at 37 °C in the culture culture medium containing MTT at a final concentration of 0.5 mg/mL. The cells were finally lysed with DMSO to solubilize the formazan. Absorbance was measured at 492 nm. Results were expressed as mean values ± SEM. Statistical analysis was performed

using GraphPad Prism 5 software and Microsoft Excel 2007. The results were analyzed by a one-way analysis of variance (ANOVA) with Dunnett's test to compare data from cell cultures treated with increasing concentrations of BaP or BeP with those from non-treated cell cultures. Differences were considered to be statistically significant at $p < 0.05$.

Sample preparation - Secretome samples harvested from the HepG2 cells were concentrated using Amicon Ultra-15 tubes (molecular weight cutoff 10 kDa; Millipore, Billerica, MA). Protein concentration was measured using the Bradford assay (Bio-Rad, Hercules, CA). Bovine serum albumin (Sigma) was used as a standard. 10µg of protein were diluted in 75µL ammonium bicarbonate buffer (50mM) and incubated with 5µL dithiothreitol 0.1M for protein denaturation (10min, 95°C). After cooling down, 10µL iodoacetamide 0.1M were added and the samples were incubated for 20min at room temperature in the dark to allow alkylation of reduced cysteine residues. For protein digestion, 100ng trypsin (Promega, V5111) was added and samples were incubated overnight at room temperature. 3µg of the clear supernatant were transferred into glass inserts and evaporated to dryness using a miVac DNA concentrator (Genevac, NY, USA). The samples were afterwards reconstituted in 15µL Formic Acid (0.1% v/v).

NanoLC-MS/MS analysis - Samples were transferred into glass inserts, compatible with the LC autosampler system (ultra nanoLC-2D plus, Eksigent, UK), and 5µL of the sample were injected. The method consisted in a 90 min run at a flow rate of 300 nL/min using a gradient of two solvents: A (99.9% water: 0.1% formic acid) and B (99.9% acetonitrile: 0.1% formic acid). The system included a nanoFlex cHiP module with a Trap-and-elute jumper, a ChromXP-C18 trap cHiP column (5mm x 300µm i.d., 3µm, 100Å, Eksigent) and a ChromXP-C18 analytical

cHiP column (15cm x 75 μ m i.d., 3 μ m, 100Å, Eksigent).

The LC system was coupled to a Triple TOF™ 5600 system with nanospray source operating in the positive ESI mode for detection (AB Sciex, CA, USA). The following parameters were used: ion spray voltage, 2.3 kV; ion source heater, 150 °C; curtain gas, 25 psi and ion source gas 1, 5 psi. The nanoLC-MS/MS system was auto-calibrated each 10 samples with a Beta-Galactosidase tryptic digest (25fmol loaded on column). MS experiments were performed using an IDA (Information Dependent Acquisition) method. The full MS scan was run with an m/z 400-1250 mass range with a 250ms accumulation time; whereas the MS/MS experiments mass range was m/z 100-1500 with a 60ms accumulation time. The 20 most intense peaks detected on the full MS scan were selected for CID (Collision Induced Dissociation) fragmentation in product ion mode giving raise to 1.5sec total cycle time. Data were acquired with the Analyst 1.6 software (AB Sciex) and the resulting .wiff files were primarily evaluated with PeakView 1.2 software (AB Sciex).

Differential data analysis - The differential secretome analysis was performed with Progenesis LC-MS software (version 4.1; Nonlinear Dynamics Ltd., Newcastle-upon-Tyne, UK). For label-free quantification, the acquired .wiff data files corresponding to the 18 samples across time points were imported into Progenesis for m/z and retention time alignment of the peptide ion currents across the samples. Automated data alignment between replicate runs and different samples was performed on the retention time axis, whereas m/z values stay unaltered. After alignment, samples were divided into the appropriate groups (Control 2h and 72h; BeP 2h and 72h; BaP 2h and 72h). After peptide detection, we only considered features charged with two to seven protons and a matrix was generated for the whole experiment, including all features with their corresponding intensities

and normalized abundances. For protein identification, the Mascot search engine (version 2.2) and the following parameters were used: (a) database: Swiss-Prot with *Homo sapiens* taxonomy (version 20120111, 21918 sequences), (b) enzyme: Trypsin/P, (c) missed cleavages: allow up to 1, (d) variable modifications: acetylation (Protein N-term), oxidation (Met), carbamidomethylation (Cys), phosphorylation (Ser, Thr and Tyr), (e) peptide mass tolerance: 30 ppm, (f) MS/MS tolerance: 0.5 Da. The Mascot results file was converted into xml format and imported into Progenesis. For identification, only proteins with at least one peptide and individual ions scores ≥ 25 were considered, giving rise to a FDR below 1%. For quantification, only unique peptides were tolerated, and the total cumulative abundance was calculated by summing the individual abundances of all peptides assigned to each protein. Proteins showing statistically significant differences in abundance between the groups were selected using ANOVA ($p < 0.05$) and submitted to a Principal Component Analysis (PCA). A report of the analysis results was created for further data processing and data mining.

RESULTS

The major goals of this study were in a first instance to identify, *in vitro*, the threshold concentration of BaP causing cytotoxicity and in a second instance, to quantify, in the secretome of HepG2 cells untreated or treated with either BaP or BeP, the proteins which could be further considered as potential biomarkers of BaP exposure.

Cytotoxic effects - In this study, we investigated the effects of BaP, the most carcinogenic member of the PAHs, on HepG2 cells. These cells were treated with increasing concentrations of BaP and BeP (0, 0.125, 0.25, 0.5, 1, 10 or 50 μ M) for 2h and 72h. MTT assay was carried out to determine relevant BaP treatment concentration (Figure 1). The dose

response study using these varying doses of BaP on HepG2 cells, seeded in 96-well culture plates, revealed that changes in cell viability of HepG2 exposed to BaP for 2h were not found to be significantly different from the control cells. These results indicated a threshold concentration of 10 μ M BaP causing cytotoxicity. Moreover, 1 μ M BaP was shown to induce no toxic effects in MTT assay. Thus, all BaP levels below 1 μ M were assumed to be subacute for HepG2 cells. Moreover, after exposure to 50 μ M BaP or BeP for 72h, the percentage of cell viability decreased to 50%: induction of such a lethal state could lead to a limited available amount of biological material, as well as interferences with proteins regulated by the PAH exposure.

Protein identification - Based on the results shown in Figure 1, we chose one toxic BaP-concentration (10 μ M) and two incubation periods (2 and 72 h) to analyze BaP-induced changes of the subproteome using quantitative proteomic approach. These experiments were performed in biological triplicates and focused on the HepG2 secretome, seeking to identify biomarkers candidates after BaP exposure. Figure 2 summarizes the HepG2 secretome analysis workflow. A total of 897 proteins with at least one unique peptide (supplemental data 1 & 2) with ion scores of ≥ 25 were identified in the secretome of treated and untreated cells. Of these, 847 proteins (supplemental data 3) were validated with a FDR below 1%. 51% of these proteins exhibited a p-value ≤ 0.05 at least in one condition of incubation period (2h or 72h or both of them). All values were extracted using Progenesis LC-MS software. These results showed that BaP treatment leads to significant changes in protein expression. We analyzed the time-dependent BaP-induced alterations in protein expression by a label-free approach. A principal component analysis (PCA) was used to extract the main sources of variations of protein expression in HepG2 secretome (897 proteins identified in all samples) (Figure 3). This examined the

reproducibility among the triplicates, and allowed to distinguish 2 separate clusters, namely the BaP cluster which was segregated from a second cluster including the two controls (non-exposed cell and BeP-exposed cell).

Moreover, the use of computational methods that predict whether a protein is targeted to a specific pathway or not ensured the confident and comprehensive identifications of secreted proteins. Based on the SignalP 4.1 (36), SecretomeP 2.0 (37) and Gene Ontology (GO) analysis, 51.95% proteins were predicted as secreted. 41.2% were highlighted by using the SecretomeP algorithm (score 0.5 or higher), 17.35% were found by SignalP and 11.22% of the proteins were localized to the extracellular region based on the GO analysis.

Protein quantification by MS1 label-free proteomic approach - Protein quantification was carried out using XICs and the area under these reconstructed elution peaks were determined. Fold changes were calculated based on the average normalized abundance of triplicates between 2 of these groups (Control, BeP or BaP). 122 proteins among all the identified protein showed a significant change for a BaP-induced time effect. A protein was considered as positively regulated when the BaP/Control ratio was greater than 2, which is in agreement with values classically used in the literature (38, 39). However we chose to increase this minimal fold-change to 5 to strengthen the statistical relevance of our results. All the proteins selected with a BaP/Control ratio > 5 were also specific to BaP because they were chosen with a similar protein expression between BeP and control conditions ($0.2 < \text{BeP/Control} < 5$). Whereas a protein is specific to HAP when the BaP and the BeP analogously and significantly changed ($\text{BaP/Control} \approx \text{BeP/Control} > 5$), while the control remains unchanged. The BeP was chosen as a negative control in order to better

differentiate between the BaP-specific and the HAP-specific regulated proteins.

Using these parameters to explore the data, the number of proteins being highly up-regulated (fold ≥ 5) was 55, while 67 cells were down-regulated (fold ≤ 0.2). We decided to only select the highly up-expressed proteins in BaP-treated cells, because positive regulation is less error-prone than a decreasing expression level of a protein which can even be lower than the assay's sensitivity. Among the 55 highly up-regulated proteins, 27 have a BaP/Control ratio = ∞ (the protein was not present in the control). Panther Database was then used to cluster these 55 differentially expressed proteins, based on their corresponding Gene Ontology (GO) terms.—The majority of these proteins are nucleic acid binding proteins (28%), cytoskeletal proteins (8 %) and hydrolases (8%). Proteins that were not included in Panther database were listed as unclassified. Regarding molecular function, 36.7% of these highly expressed proteins are implicated in binding, 34.7% have catalytic activity and the remaining proteins are known to have various functions within the cells (structural molecule activity, transcription regulator activity, transporter activity). Table 1 displays the classification and the molecular function of the up-regulated proteins. Molecular function of the unclassified proteins in Panther Database was obtained using a manual curation of the protein entry in Uniprot database (Table 1).

Based on our proteomic studies, a protein was considered as having an “early onset” when it was overexpressed 2 hours after being exposed to BaP, while its level of expression can increase (Figure 4-A), remain constant (Figure 4-B), or decrease (Figure 4-C) after 72 hours of exposure. Whereas, a protein had a “delayed onset” when it was only overexpressed after 72 hours of BaP exposure (Figure 4-D).

Our final criterion for selecting potential exposure biomarkers to BaP was to mine the final dataset to find cancer-related proteins: 14 of the 55 highly up-regulated proteins were cancer-related proteins and thus are potential BaP-exposure biomarkers. First, a protein is related to cancer if it has been already described in the literature as a cancer biomarker. A second hypothesis to link a protein to cancer is based on the fact that its expression is modified in cancer disease. Finally, the last criterion which is used to associate a protein with cancer development comes from the fact that a protein has been previously reported to interact with cancer-related proteins.

DISCUSSION

Why and how to study secretome proteomics - The protein pool secreted by the cells into conditioned media at any given time or under specific physiological conditions composes the secretome. Secreted proteins are studied *in vitro* to better understand the *in vivo* mechanisms. The term secretome refers to a set of proteins that includes extracellular matrix proteins, proteins shed from the cell membrane, and vesicle proteins (e.g., from exosomes and microsomal vesicles) (40, 41). Secretome studies offer a good alternative to body fluid proteomics; they increase specificity of analysis while minimizing biological variability and issues of enormous complexity and dynamic range associated body fluid proteomics become of less importance. In order to optimize protein recovery and minimize protein contamination from serum and dead cells, we rinsed the cells many times with PBS and incubated them in serum free medium. The number of times the cells were rinsed was optimized. Six rinsings were found to remove the majority of serum proteins, as determined by the reduction of their abundance in the gel 1D (data not shown). Six rinsings were found to remove the majority of serum proteins, as determined

by the reduction of their abundance using a SDS-PAGE migration and mass spectrometry identification (data not shown).

Proteins are usually secreted at low concentration in the culture media, which makes their recovery difficult (42). In our study, half of identified proteins (51.95%) were predicted as secreted indicating that 48.05% were liberated in the culture medium upon lysis of a few dead cells and could be cytoplasmic contaminants. Moreover, the computational methods (SignalP and SecretomeP) rely on already defined pathways as an input and cannot be used to infer new pathways. Our quantitative label-free bottom-up approach was used as a strategy to study the secretome of HepG2 cells either untreated or treated, at two incubation periods (2h and 72h), with either 10 μ M BaP or 10 μ M BeP (negative control). Despite being a BaP isomer, BeP is not carcinogenic. Several studies strived to explain this lack of BeP carcinogenicity. Thakker *et al.* showed that the carcinogenic activity of the BeP diol epoxides is less than the tenth of that of the BaP diol epoxides (33). BeP was also reported not to induce enzymes involved in its own metabolism unlike that of BaP in Caco-2 cells (43). Moreover, biological triplicates of sample secretome were used in our study to assure reproducibility and enough sensitivity in biomarker screening. The need for developing more effective cancer biomarkers and therapeutic modalities has led to the study of “cancer cell secretome” (proteins secreted by cancer cells), a promising source for novel biomarkers discovery, as a means to identify and characterize diagnostic and prognostic markers. This type of study has only recently been shown to have significant potential for diverse applications in oncoproteomics (44-46).

Data processing to quantify proteins of interest under BaP exposure - In order to study the differential protein expression in the secretome samples, strategies for rapid,

highly reproducible and accurate quantification are necessary. Non-gel-based quantitative proteomics methods have, therefore, also been developed significantly in recent years (47-50). Thus, we have applied label-free shotgun proteomics technique for faster, cleaner, and simpler results. Quantitative label-free nanoLC-MS/MS is more reproducible and less labor-intensive. It allows automation and peptide/protein identification and can accommodate larger sample sets than gel-based techniques. It is based on the measurements of changes in peptide peak areas. The areas of chromatographic peaks have been shown to correlate linearly in a wide range with the protein abundance which makes their measurement feasible for quantitative studies (51, 52). However, many commercial and open-source software packages are available to achieve label-free quantification like Corra (53), DANTE (54), MassChroQ (55), MaxQuant (56), MFPaQ (57), OpenMS (58), PEAKS Q (59), IDEAL-Q (60), Skyline (61), XPRESS (62), Progenesis and others. We have chosen Progenesis LC-MS which allows chromatographic alignment as an initial step to find differences that are statistically meaningful. This software can determine a *p*-value for the expression ratio of each differential peak, providing an extra measure of confidence. Moreover, the relative nearness of samples in the PCA plot indicates similarity, otherwise reproducibility among the triplicates, whereas large distances between samples indicate dissimilarity in protein abundance and distinguish 2 separate clusters: BaP cluster and controls (non-exposed cell and BeP-exposed cell). On the basis of our results, we identified 847 proteins having a FDR < 1% and including 207 up-regulated proteins with a fold-change above 2 as well as 55 highly up-regulated proteins with an increase of the expression level greater than 5. In the literature, there is no one common criteria defined for changes of the expression level for declaring a protein as differentially expressed.

Biological relevance of the up-regulated proteins highlighted by our study - Within the 55 highly up-regulated proteins, 30 proteins have not been already predicted as secreted neither by SecretomeP nor by SignalP algorithms, while 25 proteins have found to be secreted: 22 proteins are having a SecretomeP score greater than 0.5 thus indicating a high probability of non-classical secretion and 3 proteins are extracellular as determined by GO analysis. Many of our highly up-regulated secreted proteins weren't linked to cancer. Among them, 9 proteins have a BaP/Control ratio = ∞ , therefore the proteins appeared only in the BaP condition. Further work on these proteins should be undertaken to determine their role in BaP-mediated toxicity. Moreover, we divided the list of 55 proteins into two categories: 43 proteins have an "early onset" and only 7 proteins have a "delayed onset". Among the 55 proteins, 14 proteins were linked to cancer: 10 have an "early onset" (EIF3M, VPS35, RS9, DIAP1, GFAP, NUCKS, PSB2, BID, H4, RuvB2) and only 4 proteins have a "delayed onset" (QORX, RL6, CPSF5, NDKA). Several proteins revealed by our study have already been found in the literature to be related to various types of cancer, like EIF3M (early onset, figure 4-C), DIAP1 (early onset, fig 4-B) and CPSF5 (delayed onset, fig 4-D): these proteins were found to be expressed in various cancer tissue in G2SBC database (63). As for VPS35 (early onset, figure 4-C), the results of An *et al.* indicate that frame shift mutations of VPS genes and losses of expression of VPS13A and VPS35 proteins are common in gastric cancers and colorectal cancers with high microsatellite instability and suggest that these alterations might contribute to development of cancers with high microsatellite instability by deregulating vacuolar protein sorting proteins (64). Concerning H4 (early onset, figure 4-A), Fraga *et al.* suggested that the global loss of monoacetylation and trimethylation of histone H4 is a common hallmark of human

tumor cells (65). The last protein highlighted by our study and described as related to several types of cancer is NDKA (early onset, figure 4-A): the protein product of NDKA (NM23-H1) gene has activity of nucleoside diphosphate kinase and acts as a metastasis suppressor for some tumor types (66). The level of NM23-H1 expression is inversely correlated with the tumor's metastatic potential in experimental rodent cells and in human tumors, such as hepatocellular carcinoma, melanomas and breast, ovarian, cervical and gastric cancer (67). NM23-H1 overexpression predicts a favorable patient prognosis because this protein is implicated in the regulation of metastasis in a variety of human cancers. A correlation between the mutation in the nm23 gene and metastasis in colorectal oncogenesis suggests that the nm23 gene plays a role in the causation of metastasis (68). Among the other proteins highlighted to undergo a high positive expression regulation, several are also related to apoptosis, like BID (early onset, figure 4-A): BID is a unique BH3-only pro-apoptotic protein (69). BID knock down protects cancer cells against apoptosis. The expression of BID is up-regulated by the tumor suppressor p53, and BID has been shown to be involved in p53-mediated apoptosis (70). As for QORX (TP53I3) (delayed onset, figure 4-D), which expression is regulated by Tumor suppressor p53 that trigger apoptosis (71). TP53I3 expression is largely regulated through its promoter (72), and polymorphisms within this site have been associated with invasive bladder cancer (73) and leukemia (74). Among the other interesting proteins linked to cancer, several are specific from a type of tissue, like GFAP (early onset, figure 4-A): Glial fibrillary acidic protein (GFAP) was described as a major constituent of certain areas in human brain with pathological conditions. Human brain tumors contain relatively high amounts of GFAP, up to 85 times the concentration in parietal grey substance of normal human brain (75). Whereas other

proteins are described in the literature to be highly overexpressed in breast cancer, like NUCKS (early onset, fig 4-A) (76). Furthermore, the abundance of NUCKS in rapidly growing cells as well as the overexpression of nucks mRNA in ovarian cancer (77) suggests that it may be involved in facilitating and maintaining activity of transcription of some genes during rapid proliferation and in cancer. As for PSB2 (Proteasome subunit beta type-2) (early onset, figure 4-A) and RuvB2 (early onset, figure 4-C): the first has been identified by *Wada et al.* to be involved in ovarian carcinogenesis (78), while Rousseau *et al.* identified the overexpression of the second (RuvB2) in 4 tumors of human hepatocellular carcinoma. Moreover, RuvB2 is known to interact with β -catenin and an oncogene (c-myc) (79) and its overexpression was shown to enhance tumorigenicity, showing that this protein is required for tumor cell viability.

Our study also highlighted ribosomal proteins as being highly up-regulated under BaP exposure, like RS9 (early onset, figure 4-A). Lindström *et al.* indicated in their study, that silencing of RS9 induced suppression of cell proliferation in a p53-independent manner. Inhibiting the expression of RS9 could be one efficient way to reinitiate differentiation processes or to induce senescence or apoptosis in rapidly proliferating tumor cells (80). As for RL6 (delayed onset, figure 4-D), *Gou et al.* demonstrated overexpression of RL6 in gastric cancer and suggested that up-regulation of RL6 could promote growth and cell proliferation through up-regulating cyclin E expression. RL6 is likely to play an important role in tumorigenesis of human gastric cancer and it could be a potential biomarker and therapeutic target for gastric cancer (81).

In conclusion, we used a MS1 quantitative proteomic approach in the present study to decipher proteins which expression is regulated under BaP exposure: in fact, proteomics has become a popular approach

in the field of toxicology, even if the application of label-free quantitative proteomics in this field is relatively new when compared to other fields. The use of a comprehensive experimental design is a crucial element in such studies, especially concerning the choice of the control conditions (which can be multiple): that's why, we carefully chose the non-carcinogenic analog of BaP, named BeP, to better discriminate proteins which were regulated in a HAP-specific manner (including both compounds) from proteins we were interested in as BaP-specific regulated. Although researchers have discovered many BaP exposure-associated biomarkers, none of the biomarkers have been translated into clinical diagnostics till now. Our results can be considered as preliminary: further validation of expression of these proteins in the plasma of BaP-exposed population, by SRM (Selected Reaction Monitoring) or SRM-like methods (MS/MS^{ALL} with SWATH for example) will assist in the development of biomarkers for environment prevention. Compared to the Western Blot approach which depends only on the specificity of the antibody used (if it is commercialized), SRM depends on multiple parameters (retention time, mass-to-charge ratio of the precursor ion and selected fragment ions of the targeted peptide, relative signal intensities of the detected fragment signals). As Aebersold said (82), it is time to "turn the tables" and use SRM, developed for essentially any protein, to validate a subset of interesting proteins highlighted by quantitative proteomics studies. Furthermore, and as rightly quoted by Baker (83): "The ultimate success of biomarkers may only be realized when the focus shifts from finding them to understanding their physiological relevance".

ACKNOWLEDGMENTS

The authors thank Julie Thompson and Olivier Poch (ICUBE Laboratory, UMR CNRS 7357, Strasbourg, France) for their

help in bioinformatic analysis, the AUF (Agence Universitaire de la Francophonie), the National Council for Scientific

Research (Lebanon) and the ANR (Agence Nationale de la Recherche) for the PhD Scholarship.

REFERENCES

1. Zedeck, M. S. (1980) Polycyclic aromatic hydrocarbons: a review. *Journal of environmental pathology and toxicology* 3, 537
2. Phillips, D. H. (1999) Polycyclic aromatic hydrocarbons in the diet. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* 443, 139-147
3. Guillen, M., Sopelana, P., and Partearroyo, M. A. (1997) Food as a source of polycyclic aromatic carcinogens. *Reviews on environmental health* 12, 133-146
4. Boström, C.-E., Gerde, P., Hanberg, A., Jernström, B., Johansson, C., Kyrklund, T., Rannug, A., Törnqvist, M., Victorin, K., and Westerholm, R. (2002) Cancer risk assessment, indicators, and guidelines for polycyclic aromatic hydrocarbons in the ambient air. *Environmental health perspectives* 110, 451
5. Ramesh, A., Walker, S. A., Hood, D. B., Guillén, M. D., Schneider, K., and Weyand, E. H. (2004) Bioavailability and risk assessment of orally ingested polycyclic aromatic hydrocarbons. *International journal of toxicology* 23, 301-333
6. IARC (1983) Polynuclear aromatic compounds, part 1, chemical, environmental and experimental data. Summary of data reported and evaluation. *In: IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. World Health Organization, International Agency for Research on Cancer*
7. Kawanaka, Y., Matsumoto, E., Sakamoto, K., Wang, N., and Yun, S.-J. (2004) Size distributions of mutagenic compounds and mutagenicity in atmospheric particulate matter collected with a low-pressure cascade impactor. *Atmospheric Environment* 38, 2125-2132
8. Villalobos-Pietrini, R., Hernández-Mena, L., Amador-Muñoz, O., Munive-Colín, Z., Bravo-Cabrera, J. L., Gómez-Arroyo, S., Frías-Villegas, A., Waliszewski, S., Ramírez-Pulido, J., and Ortiz-Muñiz, R. (2007) Biodirected mutagenic chemical assay of PM₁₀ extractable organic matter in Southwest Mexico City. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* 634, 192-204
9. IARC (2010) Some non-heterocyclic polycyclic aromatic hydrocarbons and some related exposures. *In: IARC Monographs on the Evaluation of Carcinogenic Risks to Human*
10. Hoffmann, D. H., Ilse (1997) The changing cigarette, 1950-1995. *Journal of Toxicology and Environmental Health Part A* 50, 307-364
11. Ferguson, L. R., and Philpott, M. (2008) Nutrition and mutagenesis. *Annu. Rev. Nutr.* 28, 313-329
12. Armstrong, B., Hutchinson, E., Unwin, J., and Fletcher, T. (2004) Lung cancer risk after exposure to polycyclic aromatic hydrocarbons: a review and meta-analysis. *Environmental Health Perspectives* 112, 970
13. Denissenko, M. F., Pao, A., Tang, M.-s., and Pfeifer, G. P. (1996) Preferential formation of benzo [a] pyrene adducts at lung cancer mutational hotspots in P53. *Science* 274, 430-432
14. Yamazaki, H., Terada, M., Tsuboi, A., Matsubara, C., Hata, T., and Kakiuchi, Y. (1987) Distribution and binding pattern of benzo (a) pyrene in rat liver, lung and kidney constituents after oral administration. *Toxicological & Environmental Chemistry* 15, 71-81
15. Verma, N., Pink, M., Petrat, F., Rettenmeier, A. W., and Schmitz-Spanke, S. (2012) Exposure of primary porcine urothelial cells to benzo (a) pyrene: in vitro uptake, intracellular concentration, and biological response. *Archives of toxicology* 86, 1861-1871

16. Ekström, G., Bahr, C., Glaumann, H., and Ingelman-Sundberg, M. (1982) Interindividual Variation in Benzo (a) pyrene Metabolism and Composition of Isoenzymes of Cytochrome P-450 as Revealed by SDS-Gel Electrophoresis of Human Liver Microsomal Fractions. *Acta Pharmacologica et Toxicologica* 50, 251-260
17. Mitchell, C., Fischer, J., and Dahl, A. (1987) Differential induction of cytochrome P-450 catalyzed activities by polychlorinated biphenyls and benzo [a] pyrene in B6C3F₁ mouse liver and lung. *Toxicology* 43, 315-323
18. Dreij, K., Seidel, A., and Jernström, B. (2005) Differential removal of DNA adducts derived from anti-diol epoxides of dibenzo [a, l] pyrene and benzo [a] pyrene in human cells. *Chemical research in toxicology* 18, 655-664
19. Lagerqvist, A., Håkansson, D., Prochazka, G., Lundin, C., Dreij, K., Segerbäck, D., Jernström, B., Törnqvist, M., Seidel, A., and Erixon, K. (2008) Both replication bypass fidelity and repair efficiency influence the yield of mutations per target dose in intact mammalian cells induced by benzo [a] pyrene-diol-epoxide and dibenzo [a, l] pyrene-diol-epoxide. *DNA repair* 7, 1202-1212
20. Melendez-Colon, V. J., Luch, A., Seidel, A., and Baird, W. M. (2000) Formation of stable DNA adducts and apurinic sites upon metabolic activation of bay and fjord region polycyclic aromatic hydrocarbons in human cell cultures. *Chemical Research in Toxicology* 13, 10-17
21. Frank, R., and Hargreaves, R. (2003) Clinical biomarkers in drug discovery and development. *Nat Rev Drug Discov* 2, 566-580
22. Ptolemy, A. S., and Rifai, N. (2010) What is a biomarker? Research investments and lack of clinical integration necessitate a review of biomarker terminology and validation schema. *Scandinavian journal of clinical & laboratory investigation* 70, 6-14
23. van Greevenbroek, M. M., Vermeulen, V. M. M., and de Bruin, T. W. (2002) Familial combined hyperlipidemia plasma stimulates protein secretion by HepG2 cells identification of fibronectin in the differential secretion proteome. *Journal of lipid research* 43, 1846-1854
24. Alaoui-Jamali, M. A., and Xu, Y.-j. (2006) Proteomic technology for biomarker profiling in cancer: an update. *Journal of Zhejiang University SCIENCE B* 7, 411-420
25. Godovac-Zimmermann, J., and Brown, L. R. (2001) Perspectives for mass spectrometry and functional proteomics. *Mass Spectrometry Reviews* 20, 1-57
26. Zhu, W., Smith, J. W., and Huang, C.-M. (2009) Mass spectrometry-based label-free quantitative proteomics. *BioMed Research International* 2010
27. Moura, H., Terilli, R. R., Woolfitt, A. R., Williamson, Y. M., Wagner, G., Blake, T. A., Solano, M. I., and Barr, J. R. (2013) Proteomic Analysis and Label-Free Quantification of the Large Clostridium difficile Toxins. *International Journal of Proteomics* 2013
28. Fye, H. K., Wright-Drakesmith, C., Kramer, H. B., Camey, S., da Costa, A. N., Jeng, A., Bah, A., Kirk, G. D., Sharif, M. I., and Ladep, N. G. (2013) Protein Profiling in Hepatocellular Carcinoma by Label-Free Quantitative Proteomics in Two West African Populations. *PloS one* 8, e68381
29. Tang, K., Page, J. S., and Smith, R. D. (2004) Charge competition and the linear dynamic range of detection in electrospray ionization mass spectrometry. *Journal of the American Society for Mass Spectrometry* 15, 1416-1423
30. Wang, G., Wu, W. W., Zeng, W., Chou, C.-L., and Shen, R.-F. (2006) Label-free protein quantification using LC-coupled ion trap or FT mass spectrometry: Reproducibility, linearity, and application with complex proteomes. *Journal of proteome research* 5, 1214-1223
31. Deutsch, E. W., Lam, H., and Aebersold, R. (2008) Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics. *Physiological genomics* 33, 18-25

32. Slaga, T., Gleason, G., Mills, G., Ewald, L., Fu, P., Lee, H., and Harvey, R. (1980) Comparison of the skin tumor-initiating activities of dihydrodiols and diol-epoxides of various polycyclic aromatic hydrocarbons. *Cancer Research* 40, 1981-1984
33. Thakker, D., Levin, W., Buening, M., Yagi, H., Lehr, R., Wood, A., Conney, A., and Jerina, D. (1981) Species-specific enhancement by 7, 8-benzoflavone of hepatic microsomal metabolism of benzo [e] pyrene 9, 10-dihydrodiol to bay-region diol epoxides. *Cancer Research* 41, 1389-1396
34. Zwickl, H., Traxler, E., Staettner, S., Parzefall, W., Grasl-Kraupp, B., Karner, J., Schulte-Hermann, R., and Gerner, C. (2005) A novel technique to specifically analyze the secretome of cells and tissues. *Electrophoresis* 26, 2779-2785
35. Yamashita, R., Fujiwara, Y., Ikari, K., Hamada, K., Otomo, A., Yasuda, K., Noda, M., and Kaburagi, Y. (2007) Extracellular proteome of human hepatoma cell, HepG2 analyzed using two-dimensional liquid chromatography coupled with tandem mass spectrometry. *Molecular and cellular biochemistry* 298, 83-92
36. Dyrlov Bendtsen, J., Nielsen, H., von Heijne, G., and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *Journal of molecular biology* 340, 783-795
37. Bendtsen, J. D., Jensen, L. J., Blom, N., Von Heijne, G., and Brunak, S. (2004) Feature-based prediction of non-classical and leaderless protein secretion. *Protein Engineering Design and Selection* 17, 349-356
38. Zeng, G.-Q., Zhang, P.-F., Deng, X., Yu, F.-L., Li, C., Xu, Y., Yi, H., Li, M.-Y., Hu, R., and Zuo, J.-H. (2012) Identification of candidate biomarkers for early detection of human lung squamous cell cancer by quantitative proteomics. *Molecular & Cellular Proteomics* 11
39. Yan, C., Chen, Z., Li, H., Zhang, G., Li, F., Duerksen-Hughes, P. J., Zhu, X., and Yang, J. (2012) Nuclear proteome analysis of benzo (a) pyrene-treated HeLa cells. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 731, 75-84
40. Higashitsuji, H., Higashitsuji, H., Nagao, T., Nonoguchi, K., Fujii, S., Itoh, K., and Fujita, J. (2002) A novel protein overexpressed in hepatoma accelerates export of NF- κ B from the nucleus and inhibits p53-dependent apoptosis. *Cancer Cell* 2, 335-346
41. Makridakis, M., and Vlahou, A. (2010) Secretome proteomics for discovery of cancer biomarkers. *Journal of proteomics* 73, 2291-2305
42. Antelmann, H., Tjalsma, H., Voigt, B., Ohlmeier, S., Bron, S., van Dijl, J. M., and Hecker, M. (2001) A proteomic view on genome-based signal peptide predictions. *Genome Research* 11, 1484-1502
43. Chevallet, M., Diemer, H., Van Dorssealer, A., Villiers, C., and Rabilloud, T. (2007) Toward a better analysis of secreted proteins: the example of the myeloid cells secretome. *Proteomics* 7, 1757-1770
44. Rosenberg, D., and Leff, T. (1993) Regulation of cytochrome P450 in cultured human colonic cells. *Archives of biochemistry and biophysics* 300, 186-192
45. Schaaij-Visser, T., Wit, M. d., Lam, S. W., and Jiménez, C. R. (2013) The cancer secretome, current status and opportunities in the lung, breast and colorectal cancer context. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*
46. Jiménez, C. R., and Slijper, M. (2010) Current status of cancer proteomics, how far are we from clinical applications? *Journal of proteomics* 73, 1787-1789
47. Karagiannis, G. S., Pavlou, M. P., and Diamandis, E. P. (2010) Cancer secretomics reveal pathophysiological pathways in cancer molecular oncology. *Molecular oncology* 4, 496-510
48. Yates III, J. R., and Washburn, M. P. (2013) Quantitative Proteomics. *Analytical chemistry* 85, 8881-8881

49. Yates, J. R., Ruse, C. I., and Nakorchevsky, A. (2009) Proteomics by mass spectrometry: approaches, advances, and applications. *Annual review of biomedical engineering* 11, 49-79
50. Lescuyer, P., Bianchi, L., Hochstrasser, D., Bini, L., and Sanchez, J.-C. (2012) Translational proteomics. *Journal of Proteomics* 75, 4571-4572
51. Bensimon, A., Heck, A. J., and Aebersold, R. (2012) Mass spectrometry-based proteomics and network biology. *Annual review of biochemistry* 81, 379-405
52. Bondarenko, P. V., Chelius, D., and Shaler, T. A. (2002) Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. *Analytical Chemistry* 74, 4741-4749
53. Chelius, D., and Bondarenko, P. V. (2002) Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *Journal of proteome research* 1, 317-323
54. Brusniak, M.-Y., Bodenmiller, B., Campbell, D., Cooke, K., Eddes, J., Garbutt, A., Lau, H., Letarte, S., Mueller, L. N., and Sharma, V. (2008) Corra: Computational framework and tools for LC-MS discovery and targeted mass spectrometry-based proteomics. *BMC bioinformatics* 9, 542
55. Polpitiya, A. D., Qian, W.-J., Jaitly, N., Petyuk, V. A., Adkins, J. N., Camp II, D. G., Anderson, G. A., and Smith, R. D. (2008) DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* 24, 1556-1558
56. Valot, B., Langella, O., Nano, E., and Zivy, M. (2011) MassChroQ: A versatile tool for mass spectrometry quantification. *Proteomics* 11, 3572-3577
57. Cox, J., Matic, I., Hilger, M., Nagaraj, N., Selbach, M., Olsen, J. V., and Mann, M. (2009) A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nature protocols* 4, 698-705
58. Bouyssié, D., de Peredo, A. G., Mouton, E., Albigot, R., Roussel, L., Ortega, N., Cayrol, C., Burlet-Schiltz, O., Girard, J.-P., and Monsarrat, B. (2007) Mascot File Parsing and Quantification (MFPaQ), a New Software to Parse, Validate, and Quantify Proteomics Data Generated by ICAT and SILAC Mass Spectrometric Analyses Application To the Proteomics Study of Membrane Proteins from Primary Human Endothelial Cells. *Molecular & Cellular Proteomics* 6, 1621-1637
59. Sturm, M., Bertsch, A., Gröpl, C., Hildebrandt, A., Hussong, R., Lange, E., Pfeifer, N., Schulz-Trieglaff, O., Zerck, A., and Reinert, K. (2008) OpenMS—an open-source software framework for mass spectrometry. *BMC bioinformatics* 9, 163
60. Xin, L., Lin, H., Rahman, M. Z., Chen, W., Shan, B., and Ma, B. (2013) PEAKS- A Software Tool for Shotgun Label Free Proteomics with High Sensitivity and High Accuracy. *HUPO 12th Annual World Congress*
61. Tsou, C.-C., Tsai, C.-F., Tsui, Y.-H., Sudhir, P.-R., Wang, Y.-T., Chen, Y.-J., Chen, J.-Y., Sung, T.-Y., and Hsu, W.-L. (2010) IDEAL-Q, an automated tool for label-free quantitation analysis using an efficient peptide alignment approach and spectral data validation. *Molecular & Cellular Proteomics* 9, 131-144
62. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26, 966-968
63. Han, D. K., Eng, J., Zhou, H., and Aebersold, R. (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nature biotechnology* 19, 946-951
64. Mosca, E., Alfieri, R., Merelli, I., Viti, F., Calabria, A., and Milanesi, L. (2010) A multilevel data integration resource for breast cancer study. *BMC systems biology* 4, 76

65. An, C. H., Kim, Y. R., Kim, H. S., Kim, S. S., Yoo, N. J., and Lee, S. H. (2012) Frameshift mutations of vacuolar protein sorting genes in gastric and colorectal cancers with microsatellite instability. *Human pathology* 43, 40-47
66. Fraga, M. F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M., Espada, J., Schotta, G., Bonaldi, T., Haydon, C., Ropero, S., and Petrie, K. (2005) Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nature genetics* 37, 391-400
67. De La Rosa, A., Steeg, P. S., and Williams, R. L. (1995) Nm23/nucleoside diphosphate kinase: toward a structural and biochemical understanding of its biological functions. *Bioessays* 17, 53-62
68. MacDonald, N., La Rosa, D., and Steeg, P. (1995) The potential roles of nm23 in cancer metastasis and cellular differentiation. *European Journal of Cancer* 31, 1096-1100
69. Wang, L., Patel, U., Ghosh, L., Chen, H.-C., and Banerjee, S. (1993) Mutation in the nm23 gene is associated with metastasis in colorectal cancer. *Cancer research* 53, 717-720
70. Wang, K., Yin, X.-M., Chao, D. T., Milliman, C. L., and Korsmeyer, S. J. (1996) BID: a novel BH3 domain-only death agonist. *Genes & Development* 10, 2859-2869
71. Sax, J. K., Fei, P., Murphy, M. E., Bernhard, E., Korsmeyer, S. J., and El-Deiry, W. S. (2002) BID regulation by p53 contributes to chemosensitivity. *Nature cell biology* 4, 842-849
72. Levine, A. J., Momand, J., and Finlay, C. A. (1991) The p53 tumour suppressor gene. *Nature* 351, 453-456
73. Contente, A., Dittmer, A., Koch, M. C., Roth, J., and Dobbelsstein, M. (2002) A polymorphic microsatellite that mediates induction of PIG3 by p53. *Nature genetics* 30, 315-320
74. Ito, M., Nishiyama, H., Watanabe, J., Kawanishi, H., Takahashi, T., Kamoto, T., Habuchi, T., and Ogawa, O. (2006) Association of the PIG3 promoter polymorphism with invasive bladder cancer in a Japanese population. *Japanese journal of clinical oncology* 36, 116-120
75. Nomdedéu, J. F., Perea, G., Estivill, C., Badell, I., Lasa, A., and Aventín, A. (2008) Microsatellite instability may involve the pentanucleotide repeat of the PIG3 promoter in bcr/abl acute lymphoblastic leukemia. *Leukemia research* 32, 186-188
76. Rasmussen, S., Bock, E., Warecka, K., and Althage, G. (1980) Quantitation of glial fibrillary acidic protein in human brain tumours. *British journal of cancer* 41, 113
77. Ziolkowski, P., Gamian, E., Osiecka, B., Zougman, A., and Wiśniewski, J. R. (2009) Immunohistochemical and proteomic evaluation of nuclear ubiquitous casein and cyclin-dependent kinases substrate in invasive ductal carcinoma of the breast. *BioMed Research International* 2009
78. Schaner, M. E., Ross, D. T., Ciaravino, G., Sørlie, T., Troyanskaya, O., Diehn, M., Wang, Y. C., Duran, G. E., Sikic, T. L., and Caldeira, S. (2003) Gene expression patterns in ovarian carcinomas. *Molecular biology of the cell* 14, 4376-4386
79. Wada, T., Yamashita, Y., Saga, Y., Takahashi, K., Koinuma, K., Choi, Y. L., Kaneda, R., Fujiwara, S.-I., Soda, M., and Watanabe, H. (2009) Screening for genetic abnormalities involved in ovarian carcinogenesis using retroviral expression libraries. *International journal of oncology* 35, 973
80. Rousseau, B., Ménard, L., Haurie, V., Taras, D., Blanc, J. F., Moreau-Gaudry, F., Metzler, P., Hugues, M., Boyault, S., and Lemièrè, S. (2007) Overexpression and role of the ATPase and putative DNA helicase RuvB-like 2 in human hepatocellular carcinoma. *Hepatology* 46, 1108-1118
81. Lindström, M. S., and Nistér, M. (2010) Silencing of ribosomal protein S9 elicits a multitude of cellular responses inhibiting the growth of cancer cells subsequent to p53 activation. *PLoS One* 5, e9578

82. Gou, Y., Shi, Y., Zhang, Y., Nie, Y., Wang, J., Song, J., Jin, H., He, L., Gao, L., and Qiao, L. (2010) Ribosomal protein L6 promotes growth and cell cycle progression through upregulating cyclin E in gastric cancer cells. *Biochemical and biophysical research communications* 393, 788-793
83. Aebersold, R., Burlingame, A. L., and Bradshaw, R. A. (2013) Western Blots versus Selected Reaction Monitoring Assays: Time to Turn the Tables? *Molecular & Cellular Proteomics* 12, 2381-2382
84. Baker, M. (2005) In biomarkers we trust? *Nature biotechnology* 23, 297-304

FIGURE & TABLE LEGENDS

Figure 1 - *Cell viability assay* - Cell viability as percent viable cells compared to the untreated control in HepG2 cells, monitored by the MTT assay following 2h and 72h exposure to suspensions containing concentrations of BaP or BeP. Each bar represents the mean \pm SEM of at least six replicate measurements for each BaP or BeP concentration and twelve replicate measurements for control group. $P < 0.05$ versus the control cells: ANOVA with Dunnett's test (a: $p < 0.05$; b: $p < 0.01$; c: $p < 0.001$)

Figure 2 - *Experimental workflow*

Figure 3 - *Principal Component Analysis generated with MarkerView (AB Sciex)* - An exposure to benzo(a)pyrene clearly separates the proteomes. 122 proteins were significantly regulated ($p \leq 0.05$) and responsible for the separation of the proteome profiles.

Figure 4 - *Early and delayed onset of highly up-regulated protein* - A protein was considered as having an "early onset" when it was overexpressed 2 hours after being exposed to BaP, while its level of expression can increase (A), remain constant (B), or decrease (C) after 72 hours of exposure. Whereas a protein has a "delayed onset" when it is only overexpressed after 72 hours of BaP exposure (D).

Figure 5 - *Overview of proteins identified and quantified in this study: from the initial dataset to biomarkers selection*

Table 1 - *Up-regulated proteins in 2h and 72h exposure conditions after analysis by label-free quantitative nanoLC-MS/MS*

Note : *Acc#* is the Uni-Prot accession number for the identified proteins. *Protein Name* is the Uni-Prot name for the identified. The *SecP* column indicates the results from the SignalP, SecretomeP and GO analysis: a number in the *SecP* column indicates the numerical score returned by SecretomeP (greater than 0.5 indicates high probability of non-classical secretion), "E" indicates an extracellular protein as determined by GO analysis and "-" indicates that the protein wasn't found neither in SecretomeP/SignalP nor in GO analysis. None of the 55 highly up-regulated proteins has a known or predicted signal peptide.

Table 1

ACC#	Abbreviation	Protein Name	SecP	Score	No. of quantified peptides	p-value	Ratio BaP/Ctrl	Panther Protein Class	Molecular function (from Panther and/or Uniprot)
Secreted proteins									
P62081	RS7_HUMAN	40S ribosomal protein S7	0.905	72,98	2	0,05	63,815 ^{2h}	Nucleic acid binding	Structural constituent of ribosome/ nucleic acid binding
O95336	6PGL_HUMAN	6-phosphogluconolactonase	0.870	123,15	3	0,03	6,885 ^{2h}	Hydrolase	Hydrolase activity
P55957	BID_HUMAN	BH3-interacting domain death agonist	0.824	108,67	2	0,04	119,353 ^{2h}	Unclassified	Death receptor binding
E9PN11	E9PN11_HUMAN	Rho-related GTP-binding protein RhoC (Fragment)	0.824	82,12	1	5,90e-004	∞ ^{2h}	Unclassified	GTP binding
H0YJE9	H0YJE9_HUMAN	26S protease regulatory subunit 10B (Fragment)	0.785	61,57	1	7,42e-004	∞ ^{2h}	Unclassified	Unknown
O00487	PSDE_HUMAN	26S proteasome non-ATPase regulatory subunit 14	0.777	82,53	2	7,89e-003	171,396 ^{2h}	Transcription factor	Transcription factor activity
P62318	SMD3_HUMAN	Small nuclear ribonucleoprotein Sm D3	0.763	37,67	1	0,04	∞ ^{2h}	Nucleic acid binding	RNA splicing factor activity/transesterification mechanism mRNA binding
H7BZ11	H7BZ11_HUMAN	Uncharacterized protein (Fragment)	0.756	38,45	1	0,02	5,421 ^{2h}	Nucleic acid binding	Structural constituent of ribosome/ nucleic acid binding
P62249	RS16_HUMAN	40S ribosomal protein S16	0.735	53,07	1	2,93e-003	∞ ^{2h}	Nucleic acid binding	Structural constituent of ribosome/ RNA binding
I3L3E9	I3L3E9_HUMAN	Flap endonuclease 1 (Fragment)	0.734	52,07	1	2,07e-005	∞ ^{2h}	Unclassified	DNA binding / nuclease activity
Q9HC38	GL0D4_HUMAN	Glyoxalase domain-containing protein 4	0.721	90,12	1	0,04	∞ ^{2h}	Lyase	Lyase activity
A8MWD9	RUXGL_HUMAN	Small nuclear ribonucleoprotein G-like protein	0.690	58,65	1	2,47e-004	∞ ^{2h}	Unclassified	RNA binding
H0YCY6	H0YCY6_HUMAN	FAD-AMP lyase (cyclizing) (Fragment)	0.644	72,60	1	0,05	8,851 ^{2h}	Unclassified	ATP binding/glycerone kinase activity
P36957	ODO2_HUMAN	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	0.634	46,37	1	0,05	5,7588 ^{72h}	Transferase	Acetyltransferase activity
Q53FA7	QORX_HUMAN	Quinone oxidoreductase PIG3	0.622	59,58	1	3,80e-003	∞ ^{72h}	Oxidoreductase	Oxidoreductase activity
P62701	RS4X_HUMAN	40S ribosomal protein S4, X isoform	0.606	94,31	2	0,04	165,228 ^{2h}	Nucleic acid binding	Structural constituent of ribosome/ nucleic acid binding
P62140	PP1B_HUMAN	Serine/threonine-protein phosphatase PP1-beta catalytic subunit	0.604	103,52	2	0,04	∞ ^{2h}	Hydrolase / Calcium-binding protein / Phosphatase	Phosphoprotein phosphatase activity/ calcium ion binding
E9PLG2	E9PLG2_HUMAN	26S protease regulatory subunit 6A (Fragment)	0.592	78,55	2	0,02	6,778 ^{2h}	Unclassified	Unknown
B0QZ18	B0QZ18_HUMAN	Copine I	0.589	39,41	1	0,05	∞ ^{2h}	Unclassified	Unknown
Q01581	HMCS1_HUMAN	Hydroxymethylglutaryl-CoA synthase, cytoplasmic	0.560	63,3	1	0,02	5,561 ^{2h}	Lyase	Transferase activity/ lyase activity
P50991	TCPD_HUMAN	T-complex protein 1 subunit delta	0.535	446,27	6	0,02	5,896 ^{2h}	Unclassified	ATP binding/ unfolded protein binding

Q9UBE0	SAE1_HUMAN	SUMO-activating enzyme subunit 1	0.534	171,89	2	0,02	6,9798 ^{72h}	Ligase / Transfer- carrier protein	Ligase activity
H0Y8L3	H0Y8L3_HUMAN	Transforming growth factor-beta-induced protein ig-h3 (Fragment)	0.513	55,28	1	0,02	∞ ^{2h}	Unclassified	Unknown
P62805	H4_HUMAN	Histone H4	E	202,48	4	0,04	7,073 ^{2h}	Nucleic acid binding	DNA binding
Q71DI3	H32_HUMAN	Histone H3.2	E	227,87	1	0,04	113,162 ^{2h}	Nucleic acid binding	DNA binding
P14136	GFAP_HUMAN	Glial fibrillary acidic protein	E	108,99	2	0,05	34,001 ^{2h}	Cytoskeletal protein / Structural protein	Structural constituent of cytoskeleton

Not predicted as secreted proteins

P46781	RS9_HUMAN	40S ribosomal protein S9	0.498	49,19	1	2,30e-003	∞ ^{2h}	Nucleic acid binding	Structural constituent of ribosome/ nucleic acid binding
Q16836	HCDH_HUMAN	Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial	0.492	134,77	2	6,91e-004	25,344 ^{2h}	Isomerase / Lyase / Oxidoreductase	Oxidoreductase activity/ hydro-lyase activity/ racemase and epimerase activity
E5RJG7	E5RJG7_HUMAN	2,4-dienoyl-CoA reductase, mitochondrial (Fragment)	0.487	67,31	1	0,02	126,733 ^{2h}	Unclassified	Oxidoreductase activity
P62269	RS18_HUMAN	40S ribosomal protein S18	0.459	74,49	2	0,05	192,281 ^{2h}	Nucleic acid binding	Structural constituent of ribosome/ nucleic acid binding
E7EP00	E7EP00_HUMAN	Protein transport protein Sec24C	0.454	64,22	2	0,04	∞ ^{2h}	Unclassified	Zinc ion binding
Q5T446	Q5T446_HUMAN	Uroporphyrinogen decarboxylase (Fragment)	0.448	69,20	1	0,03	6,0769 ^{72h}	Unclassified	Uroporphyrinogen decarboxylase activity
P33908	MA1A1_HUMAN	Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	0.437	67,18	2	0,04	27,137 ^{2h}	Chaperone / Hydrolase	Hydrolase activity
P08243	ASNS_HUMAN	Asparagine synthetase [glutamine-hydrolyzing]	0.427	39,15	1	0,03	∞ ^{2h}	Ligase	Ligase activity
P49721	PSB2_HUMAN	Proteasome subunit beta type-2	0.398	38,89	1	0,03	∞ ^{2h}	Hydrolase / Protease	Peptidase activity
Q14978	NOLC1_HUMAN	Nucleolar and coiled-body phosphoprotein 1	0.385	229,33	4	2,36e-003	873,667 ^{2h}	Unclassified	ATP binding/ GTP binding
Q7L2H7	EIF3M_HUMAN	Eukaryotic translation initiation factor 3 subunit M	0.383	92,97	1	4,05e-004	∞ ^{2h}	Unclassified	Translation initiation factor activity
Q02878	RL6_HUMAN	60S ribosomal protein L6	0.375	65,07	1	0,02	43,219 ^{72h}	Nucleic acid binding	Structural constituent of ribosome/ nucleic acid binding
P15531	NDKA_HUMAN	Nucleoside diphosphate kinase A	0.361	312,14	2	0,01	6,1861 ^{72h}	Kinase / Transferase	Nucleotide kinase activity
O43809	CPSF5_HUMAN	Cleavage and polyadenylation specificity factor subunit 5	0.348	94,15	2	0,05	13,735 ^{72h}	Nucleic acid binding	RNA splicing factor activity, transesterification mechanism mRNA binding
Q9Y230	RUVB2_HUMAN	RuvB-like 2	0.344	70,76	1	0,01	7,338 ^{2h}	Nucleic acid binding/transcription factor	DNA helicase activity/ transcription factor activity/ transcription cofactor activity
Q96QK1	VPS35_HUMAN	Vacuolar protein sorting-associated protein 35	0.331	70,39	1	2,61e-004	∞ ^{2h}	Membrane traffic protein	Unknown
Q92499	DDX1_HUMAN	ATP-dependent RNA helicase DDX1	0.324	54,11	1	0,02	∞ ^{2h}	Nucleic acid binding	RNA helicase activity/ translation factor activity/ nucleic acid binding
Q9H1E3	NUCKS_HUMAN	Nuclear ubiquitous casein and cyclin-dependent kinase substrate 1	0.308	236,44	4	2,49e-003	17,105 ^{2h}	Unclassified	Unknown
Q8TEX9	IPO4_HUMAN	Importin-4	0.253	53,21	2	0,01	∞ ^{2h}	Transporter/ Transfer-carrier protein	Transmembrane transporter activity
O15143	ARC1B_HUMAN	Actin-related protein 2/3 complex subunit 1B	0.222	43,65	1	0,02	∞ ^{2h}	Cytoskeletal protein	Structural constituent of cytoskeleton
Q12996	CSTF3_HUMAN	Cleavage stimulation factor subunit 3	0.211	46,66	1	4,16e-004	∞ ^{2h}	Nucleic acid binding	Poly(A) RNA binding

O94826	TOM70_HUMAN	Mitochondrial import receptor subunit TOM70	0.187	49,41	1	0,04	∞^{2h}	Chaperone	Protein transmembrane transporter activity
O60610	DIAP1_HUMAN	Protein diaphanous homolog 1	0.094	37,26	1	1,61e-003	∞^{2h}	Cytoskeletal protein	Structural constituent of cytoskeleton/ actin binding
F5H4R6	F5H4R6_HUMAN	Nucleosome assembly protein 1-like 1	0.057	58,02	2	5,54e-003	∞^{2h}	Unclassified	Unknown
P68036	UB2L3_HUMAN	Ubiquitin-conjugating enzyme E2 L3	-	83,52	2	0,04	8,668 ^{2h}	Ligase	Ligase activity
Q16401	PSMD5_HUMAN	26S proteasome non-ATPase regulatory subunit 5	-	66,16	1	0,02	∞^{2h}	Unclassified	Unknown
P61326	MGN_HUMAN	Protein mago nashi homolog	-	56,80	1	0,03	123,599 ^{2h}	Unclassified	RNA binding
P60520	GBRL2_HUMAN	Gamma-aminobutyric acid receptor-associated protein-like 2	-	38,52	1	3,77e-003	∞^{2h}	Cytoskeletal protein	Structural constituent of cytoskeleton/ microtubule binding
Q7Z7H5	TMED4_HUMAN	Transmembrane emp24 domain-containing protein 4	-	37,59	1	9,98e-003	∞^{2h}	Membrane traffic protein / Transfer-carrier protein	Signal transducer activity

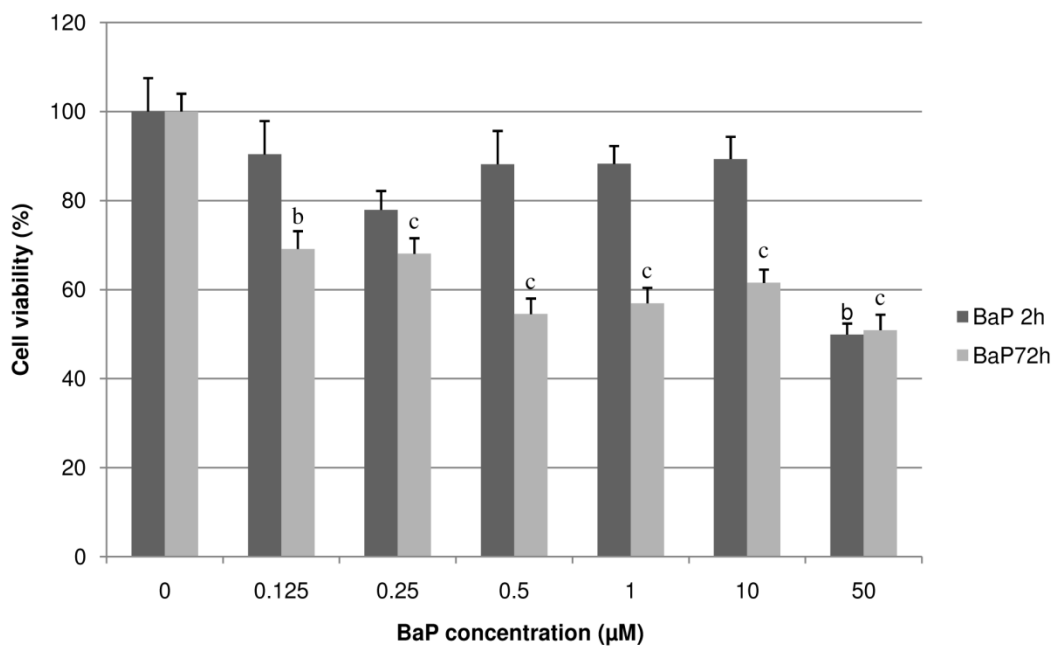
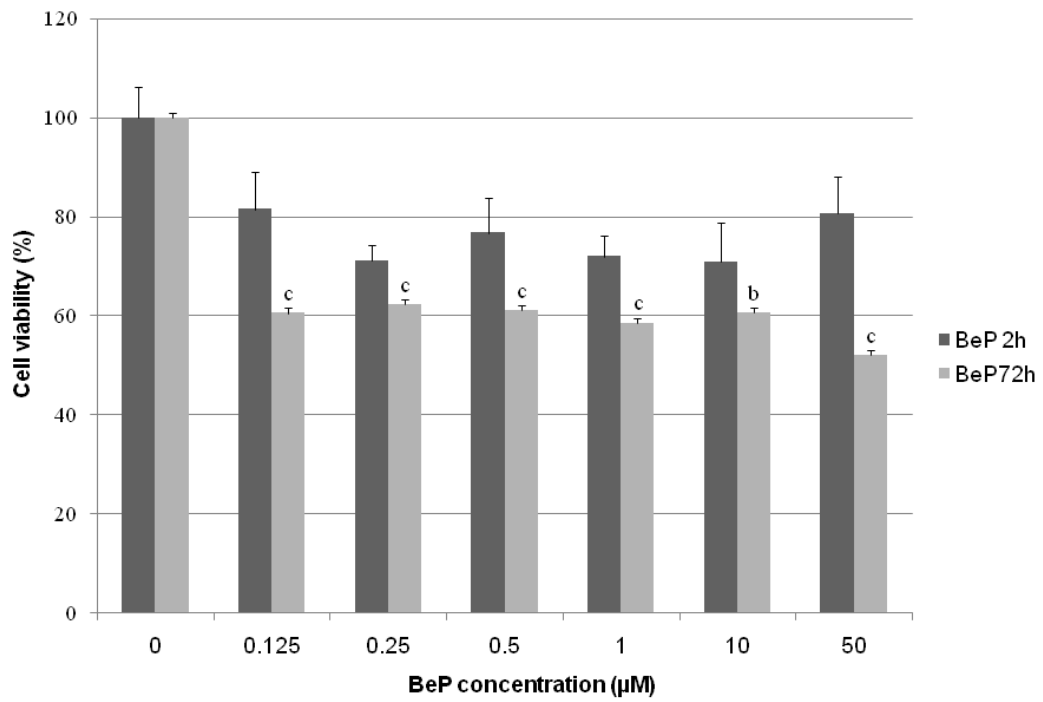


Figure 1

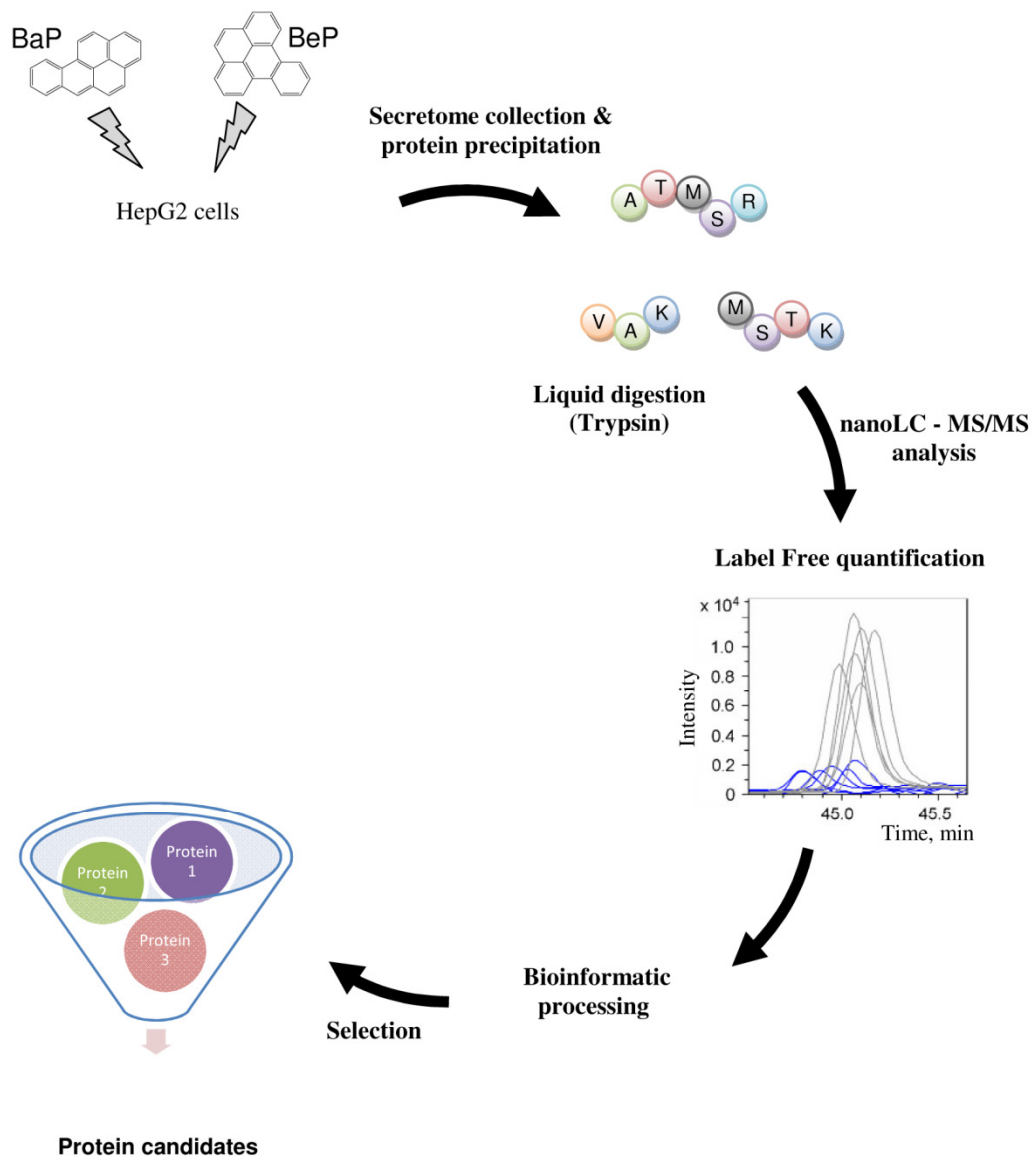


Figure 2

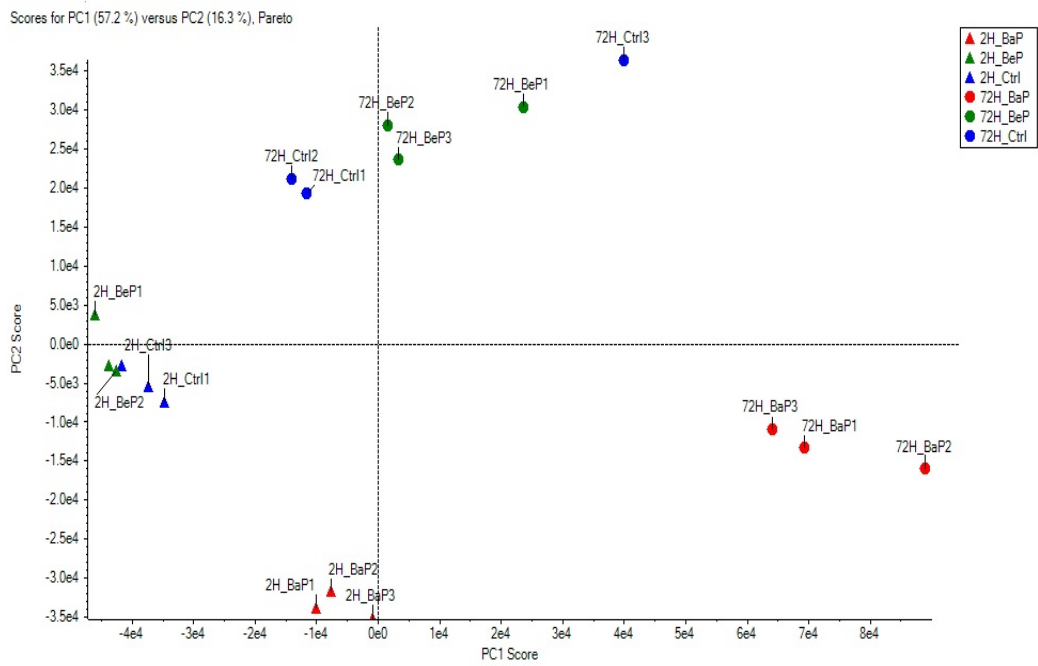


Figure 3

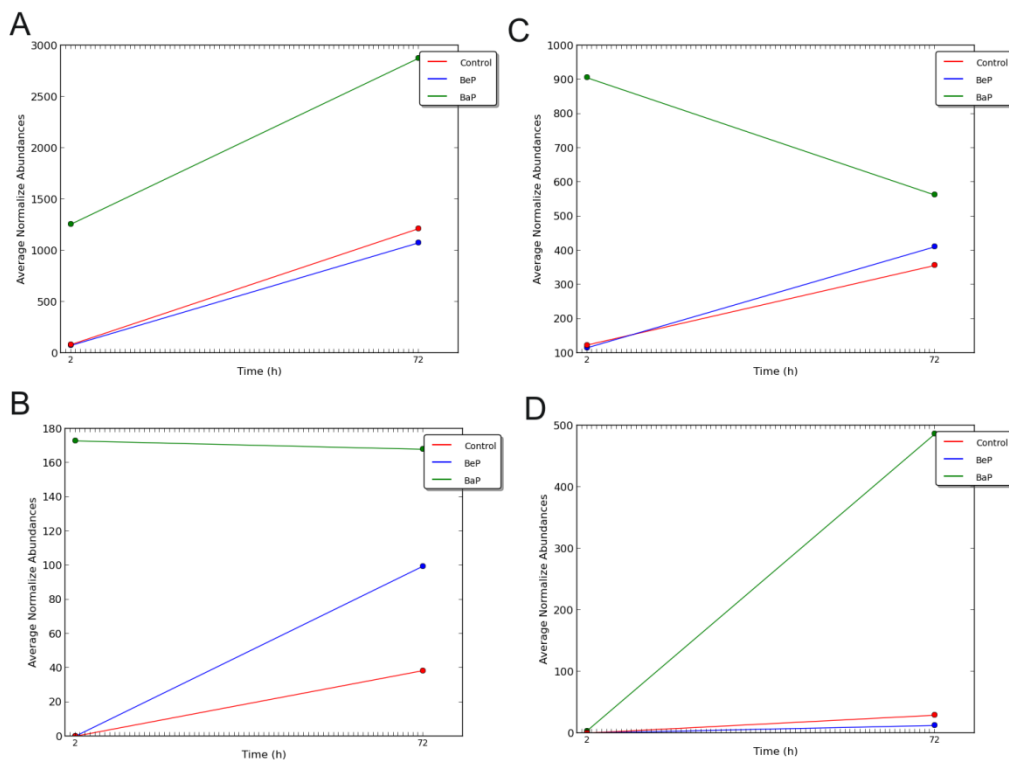


Figure 4

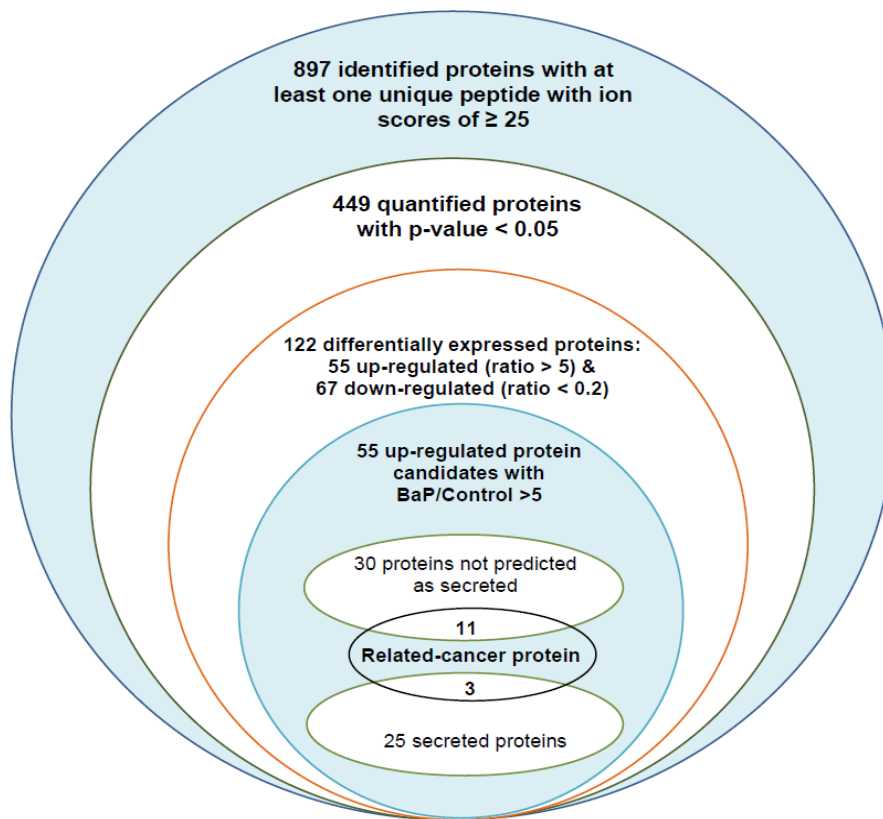


Figure 5

Chapitre 3 : Analyse différentielle par spectrométrie de masse quantitative (label-free)

Les études de protéomique différentielle utilisant la quantification sans marquage (label-free) sont basées sur la quantification MS1 (aires des pics d'élution chromatographique des peptides) ou sur les nombres de spectres MS/MS acquis pour chaque protéine (Griffin, Yu et al. 2009). Les avantages de cette technique sont : son coût puisqu'aucun marquage n'est nécessaire, et sa gamme dynamique importante permettant d'identifier et de quantifier les protéines d'un échantillon complexe. Mais la technique de quantification sans marquage a un inconvénient majeur: en effet, les analyses des divers échantillons à comparer se faisant successivement, il est possible d'observer une variation en temps de rétention pour un même peptide. Cela impliquera alors, si la disparité est trop importante, l'utilisation d'un outil de réaligement des analyses avant toute reconstitution de profils peptidiques. Cette lacune en termes de traitement bioinformatique pour la technique de quantification sans marquage est également retrouvée dans la méthode de quantification avec marquage: en effet, même si les échantillons peuvent être mélangés plus ou moins tôt dans la préparation biologique (multiplexage), il est tout de même nécessaire de repérer au final les peptides appariés (forme isotopiquement légère et alourdie par exemple). Cela peut se révéler très difficile et source d'erreurs, en particulier dans le cadre de fortes régulations (phénomène on/off ou l'une des 2 formes est très proche du bruit de fond). Chacune des 2 techniques (avec ou sans marquage) ayant ses limites, c'est pour cela qu'il n'y a pas de consensus qui se dégage dans la littérature quand au choix de technique à faire pour aborder un sujet d'étude

Pour chacune des conditions de cinétique de temps (2h, 6h, 24h et 72h), et pour limiter les variations d'intensité ionique dues à des différences de configuration du système chromatographique, nous avons passé les échantillons successivement afin d'obtenir des temps de rétention les plus reproductibles possibles. Le processus de détermination des volumes de pics peptidiques est appelé extraction ionique et le pic d'élution est appelé EIC ou XIC (EXtracted Ion Chromatogram). Ce XIC peut être produit pour chaque m/z de chaque expérience nanoLC-MS/MS et les volumes de pics résultants peuvent être comparés de façon quantitative entre les différents échantillons. Trois suites logicielles pour la quantification par label-free MS1 ont été testées : ProteinPilot/PeakView/MarkerView (AB Sciex, CA, USA),

Skyline (MacCoss Laboratory, University of Washington) et Progenesis LC-MS (Nonlinear Dynamics, Newcastle-upon-Tyne, UK).

1) Choix des logiciels

Notre premier test est basé sur la suite ProteinPilot/PeakView/MarkerView. Cette suite est normalement payante, mais elle a été déjà fournie avec l'installation du Triple TOF 5600. Plusieurs articles ont été publiés démontrant ainsi la qualité des résultats obtenus avec (Chen, Zhou et al. 2011, Herrera López, Hernando et al. 2013, Spickett, Reis et al. 2013, Zhu, Schultz et al. 2013).

Skyline, la solution open-source, a été testée dans un second lieu. Son développement ayant commencé plus récemment (MacLean, Tomazela et al. 2010), nous disposons de moins de recul dessus, mais il s'agit d'un logiciel de plus en plus utilisé et cité en référence par d'autres laboratoires (Cilia, Bereman et al. 2012, Roudier, True et al. 2013).

Quant au logiciel Progenesis (vendu par la société Nonlinear Dynamics), il a été utilisé principalement pour de l'analyse 2D-E différentielle. Récemment, il est devenu une méthode en vogue pour la quantification de protéines à partir d'analyses nanoLC-MS (Babaei, Ramalingam et al. 2013, Bostanci, Ramberg et al. 2013, Dagley, Emili et al. 2013).

Ces 3 logiciels permettent d'analyser les données en 3 étapes : pre-processing (traitement des données brutes avant la génération des valeurs d'aires des pics peptidiques), intégration (génération des XICs) et analyse statistique.

2) Quantification relative à l'aide de la suite logicielle PeakView/MarkerView (AB Sciex)

2.1) Pre-processing

ProteinPilot est la première brique de la suite logicielle d'AB SCIEX. C'est un moteur de recherche, permettant d'utiliser soit l'algorithme Mascot (Matrix Science), soit l'algorithme Paragon (ABSciex), soit les 2 à la fois. Durant ma thèse, j'ai utilisé l'algorithme Paragon uniquement. Tout comme Mascot, l'utilisateur doit préciser quelle est la banque de données qui sera utilisée pour la recherche, ainsi que la méthode de validation des protéines (taux de faux positif inférieur à 1%) (Figure 52).

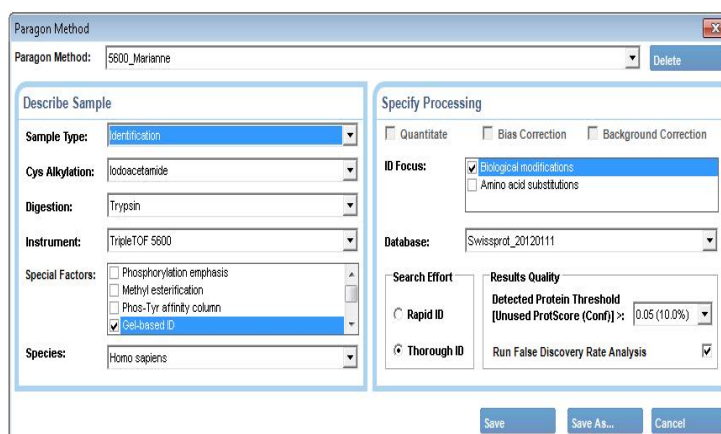


Figure 52 : Méthode Paragon de ProteinPilot

Le logiciel ProteinPilot permet d'effectuer des identifications par injection (1 seul échantillon) mais également de combiner plusieurs injections entre elles en formant un unique « .group ». C'est cette propriété qui est utilisée lors de nos expériences de protéomique quantitative sur les cellules exposées ou non à 10 μ M BaP ou BeP pendant 2h, 6h, 24h et 72h. La construction d'une telle banque de peptides observés dans un ensemble d'injections se fait à partir des données brutes issues du TripleTOF 5600 (fichiers « .wiff » et « .wiff.scan ») et le temps ainsi associé à la recherche est relativement long comparé à Mascot (plusieurs heures pour Paragon à partir des fichiers bruts .wiff dans ProteinPilot contre une dizaine de minutes pour Mascot à partir des fichiers .mgf dans Proteinscape). En confrontant les résultats expérimentaux à la banque de données protéiques UniProt (en spécifiant la taxonomie *Homo sapiens*), nous pouvons ainsi obtenir le répertoire des protéines contenues dans les échantillons avec un seuil de confiance fixé (estimation du taux de faux-positifs).

Un export de l'ensemble des protéines (Protein Summary) ainsi que le détail des peptides (Peptide Summary) est possible: sont notés des paramètres importants pour la suite tels que l'intensité des précurseurs et leur temps de rétention et le pourcentage de confiance attribuée par Paragon à l'identification.

2.2) Calcul des valeurs d'aires des pics peptidiques

PeakView (version 1.1.1) utilise comme données de départ le fichier généré par l'analyse ProteinPilot (fichier avec l'extension « .group », qui est l'équivalent du fichier « .dat » généré par Mascot). Après avoir indiqué à PeakView le nombre de protéines que l'on souhaite importer (correspondant aux nombre de protéines validées tel que FDR < 1%) et le nombre de peptides à considérer par protéine (20 peptides), les données brutes « .wiff » issues du Triple TOF seront importées dans PeakView (Figure 53). Ce dernier va extraire le pic d'élution (XIC) de chaque

peptide (Figure 54) importé à partir des fichiers bruts: une tolérance en masse de $\pm 0,05\text{Da}$ et en temps de rétention $\pm 2,5\text{min}$ ont permis la reconstitution de ces pics d'élution.

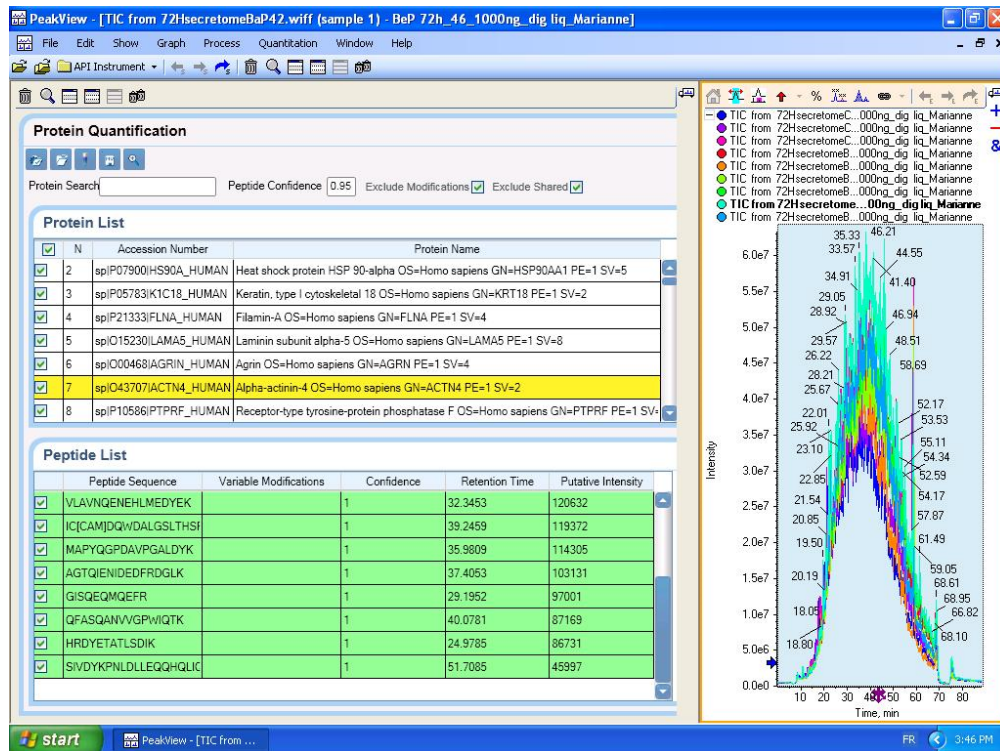


Figure 53 : Ajout de critères spécifiques et fichiers « .wiff » chargés dans PeakView

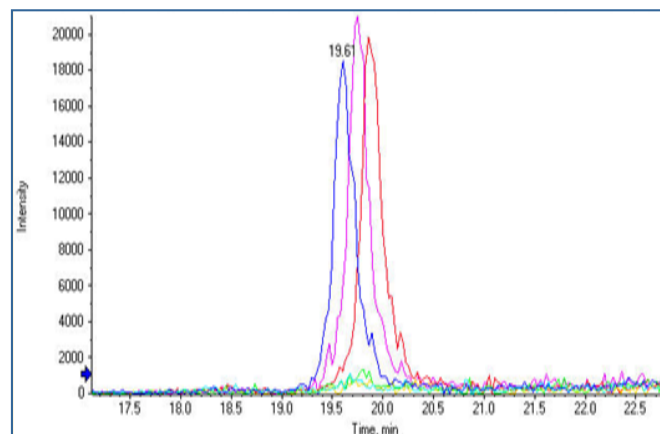


Figure 54 : XIC d'un peptide

Le logiciel PeakView permet d'automatiser la récupération des aires des pics d'élution reconstitués pour chaque peptide d'intérêt. De plus, seuls les peptides présentant un pourcentage de confiance supérieur à 95%, sans modification chimique ou post-traductionnelle et protéotypiques (non partagé par plusieurs protéines) ont été choisis pour être quantifiés (Figure 53). Les aires comprises sous les pics d'élution des peptides d'intérêt sont exportées dans un fichier « .mrkw ».

2.3) Analyse statistique

MarkerView (version 1.2.1) utilise les fichiers « .mrkw » générés par PeakView comme point de départ. On peut choisir de travailler sur le fichier « Peptides » (détail des aires de chaque pic peptidique, dans chacune des conditions d'étude) ou sur le fichier « Protéines » (pour chaque condition étudiée, le logiciel a réalisée la moyenne des aires de l'ensemble des pics peptidiques utilisés pour chaque protéine) (Figure 55).

MarkerView - [Peaks: Ctrl-BaP-BaP-24h-janvier2013_peptides]

Row	Index	Peak Name	m/z	Ret. Time	Group	Use	24HsecretomeCht25	24HsecretomeCht26	24HsecretomeCht27	24HsecretomeBep3
1	1	SRTGSESSQTGT	606.2801	7.92	splP23588IF4B_H	<input checked="" type="checkbox"/>	2.054e4	3.247e4	1.333e4	3.748e4
2	2	AREEEEEQKEGGD	590.2848	7.92	splQ13442HAP28	<input checked="" type="checkbox"/>	3.076e4	2.836e4	3.021e4	1.044e4
3	3	AREEEEEQKEGGD	526.0001	7.92	splQ13442HAP28	<input checked="" type="checkbox"/>	3.926e4	2.394e4	2.801e4	1.608e4
4	4	KAE[CAM]GAQNG	532.7659	7.93	splQ9P2E9FRBP1	<input checked="" type="checkbox"/>	2.481e4	6.214e4	2.942e4	5.791e4
5	5	KDDSHSAEDSED	618.2729	7.95	splQ9H1E3NUCK	<input checked="" type="checkbox"/>	1.029e4	9.217e3	1.409e4	5.984e3
6	6	AREETDKKEEPASK	497.2426	7.96	splP05455ILA_HU	<input checked="" type="checkbox"/>	1.718e4	1.865e4	1.224e4	3.242e4
7	7	PEEVKFR	443.2614	7.96	splP23528ICDF1_	<input checked="" type="checkbox"/>	1.109e5	1.348e5	4.993e4	8.775e4
8	8	RITAEKNER	415.8898	8.05	splP15311IEZRI_H	<input checked="" type="checkbox"/>	9.758e4	8.219e4	5.204e4	3.395e4
9	9	ITHSPQAHDPQE	440.2227	8.14	splP20827IEFNA1	<input checked="" type="checkbox"/>	4.799e4	6.822e4	3.208e4	1.514e3
10	10	CAM_AKE[Na]KIE	526.8140	8.22	splQ12904AIMP1_	<input checked="" type="checkbox"/>	8.150e4	7.702e4	8.381e4	9.898e3
11	11	LAHYNKR	451.2537	8.22	splQ99877IH2B1N	<input checked="" type="checkbox"/>	5.590e4	5.442e4	2.973e4	1.351e5

MarkerView - [t-Test: Ctrl-BaP-BaP-24h-janvier2013_proteines]

Compare: BaP to (All Others) n1 = 3, n2 = 6

Index	Peak Name	m/z	Ret. Time	Group	Use	t-value	p-value	Mean 1	Mean 2	Median 1	Median 2	Sigma 1
1	splP23588IF4B_H	N/A	N/A		<input checked="" type="checkbox"/>	6.28	0.00041	5.100e6	3.145e6	5.170e6	3.090e6	1.643e5
2	splQ13442HAP28	N/A	N/A		<input checked="" type="checkbox"/>	-0.01	0.99367	1.179e6	1.184e6	1.239e6	1.480e6	5.435e5
3	splQ9P2E9FRBP1	N/A	N/A		<input checked="" type="checkbox"/>	-1.26	0.24839	3.216e6	5.306e6	3.377e6	5.530e6	6.440e5
4	splQ9H1E3NUCK	N/A	N/A		<input checked="" type="checkbox"/>	1.37	0.21411	5.730e6	3.213e6	3.573e6	3.154e6	4.477e6
5	splP05455ILA_HU	N/A	N/A		<input checked="" type="checkbox"/>	2.11	0.07252	4.862e5	2.306e5	5.817e5	2.165e5	2.405e5
6	splP23528ICDF1_	N/A	N/A		<input checked="" type="checkbox"/>	0.40	0.70404	8.737e6	7.409e6	5.868e6	7.005e6	7.322e6
7	splP15311IEZRI_H	N/A	N/A		<input checked="" type="checkbox"/>	1.31	0.23150	6.246e6	4.614e6	5.980e6	5.236e6	6.612e5
8	splP20827IEFNA1	N/A	N/A		<input checked="" type="checkbox"/>	-2.16	0.06810	9.787e5	5.232e6	8.960e5	6.412e6	5.957e5
9	splQ12904AIMP1_	N/A	N/A		<input checked="" type="checkbox"/>	-0.36	0.72813	1.612e6	2.866e6	1.483e6	4.582e5	3.472e5
10	splQ99877IH2B1N	N/A	N/A		<input checked="" type="checkbox"/>	0.79	0.45347	3.676e6	2.499e6	3.097e6	1.575e6	1.461e6
11	splP06748NPM_H	N/A	N/A		<input checked="" type="checkbox"/>	-0.36	0.73230	7.913e6	9.288e6	7.556e6	9.437e6	1.295e6

Figure 55 : fichiers générés par MarkerView (1.2.1) : (a) peptides et (b) protéines

La première étape consiste à normaliser les données, étape d'autant plus importante si l'on constate une faible reproductibilité des répliquats biologiques (par exemple en superposant les courants ioniques totaux (Total Ion Chromatogram TIC)). Sur les données des cellules exposées ou non au BaP ou BeP pendant différents temps, une normalisation, basée sur la somme des aires totales des pics d'élution des protéines ou peptides, a été réalisée.

La deuxième étape réside en une analyse PCA (« Principal Component Analysis ») aux niveaux peptidique et protéique: il s'agit d'une méthode de statistique dite « multi-variée » permettant de transformer des variables liées entre elles en de nouvelles variables décorréliées les unes des autres. On réduit ainsi l'information à étudier en un nombre de composantes plus limité que le nombre initial de variables (Figure 56). Dans notre cas, cette analyse va permettre d'évaluer la reproductibilité des triplicats biologiques et la corrélation entre les différentes

conditions étudiées (3 conditions d'exposition (Control, BeP et BaP) et 4 temps d'incubation différents (2h, 6h, 24h et 72h)).

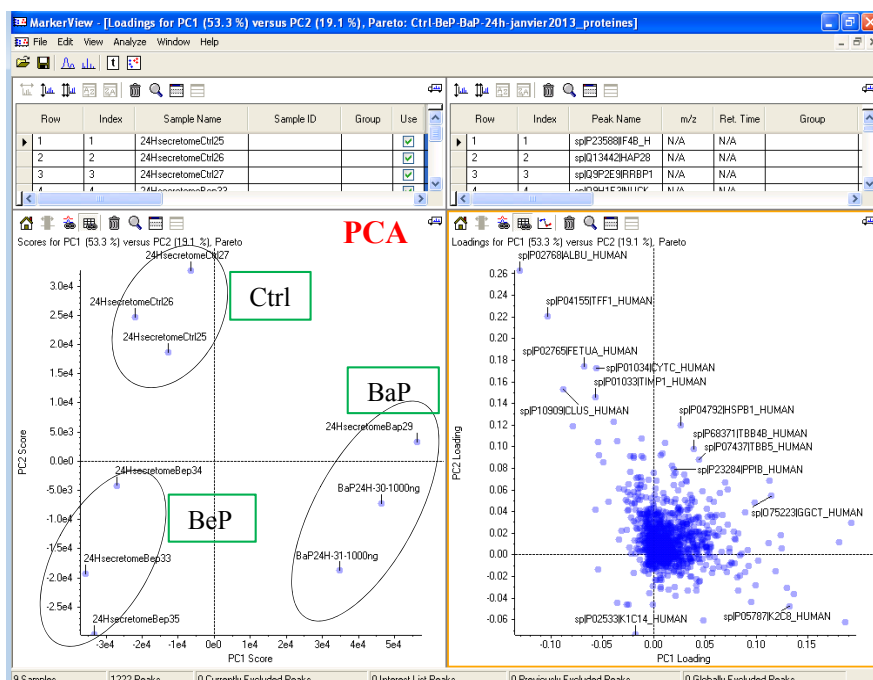


Figure 56 : Représentation d'une PCA sur MarkerView

Finalement, la troisième étape consiste à définir des groupes d'échantillons qui vont être comparés entre eux à l'aide d'un test de Student. Ces groupes d'échantillons correspondent aux 3 conditions d'exposition pour chaque temps d'incubation.

2.4) Résultats et discussion

Les résultats des PCA individuelles (3 conditions d'exposition pour chaque temps d'incubation) et de la PCA globale (3 conditions d'exposition et 4 temps d'incubation) sont représentés dans les figures 57 et 58 respectivement. Nos résultats montrent que la plupart des triplicats sont regroupés entre eux, ce qui permet de confirmer la relative similitude de ces injections. Les distances qui séparent les différents groupes, chacun représentant une condition d'exposition et un temps d'incubation, montrent une différence d'abondance protéique entre les différentes conditions. La figure 58, représentant la PCA globale, montre bien la différence entre les 2 groupes BaP et contrôles (contrôle sans exposition et BeP).

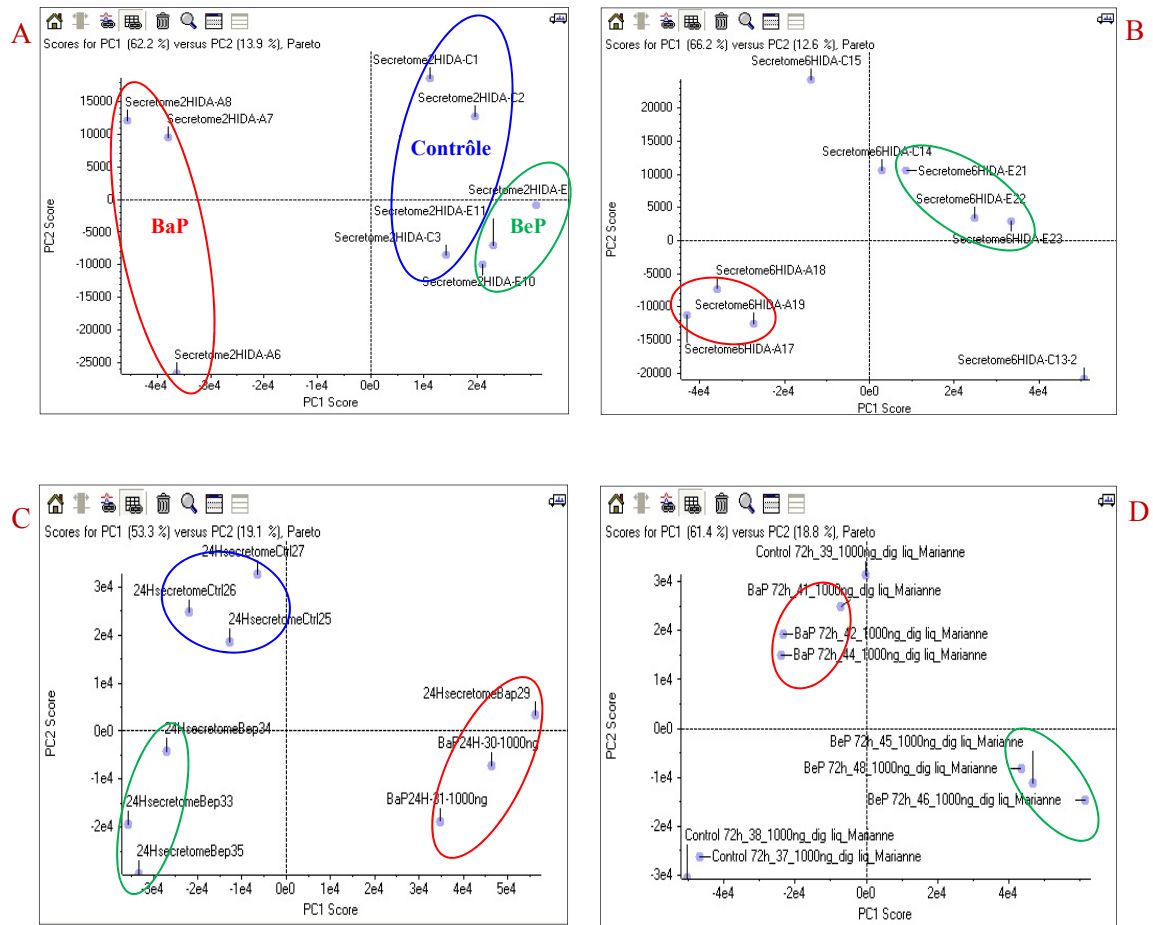


Figure 57 : PCA individuelles des 4 temps d'incubation (2h, 6h, 24h et 72h) réalisées sur MarkerView. (A) 2h, (B) 6h, (C) 24h et (D) 72h

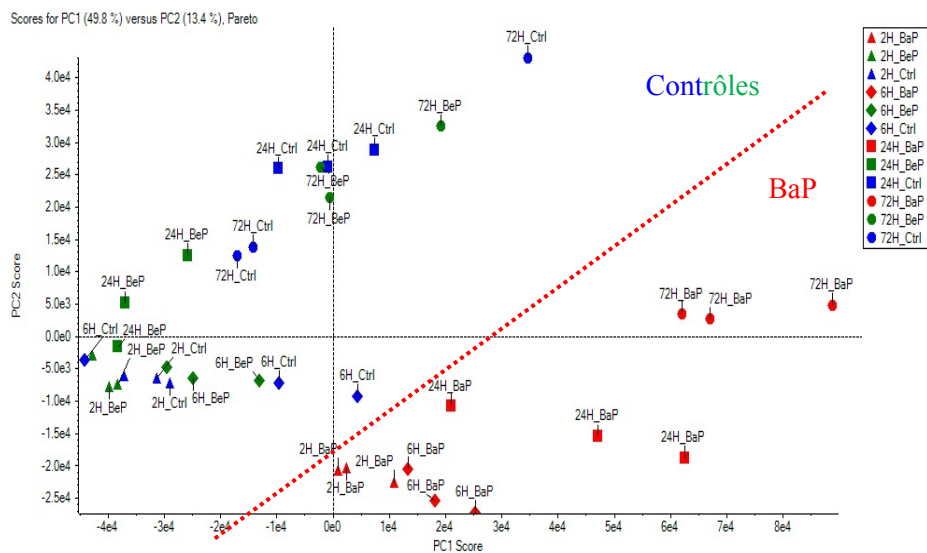


Figure 58 : PCA globale des 4 temps d'incubation (2h, 6h, 24h et 72h)

Chaque « .group » représentait les triplicats d'une seule condition de temps avec 3 conditions d'exposition, donc 9 « .wiff » (exemple: 3 Contrôles-2h, 3 BeP-2h, 3 BaP-2h). Le

nombre de protéines identifiées avec un FDR < 1% et obtenues à partir de chaque « .group » est : 1000 protéines (2h), 885 protéines (6h), 1214 protéines (24h) et 1258 protéines (72h). Le diagramme de Venn suivant montre que 665 protéines sont communes aux 4 conditions de cinétique de temps.

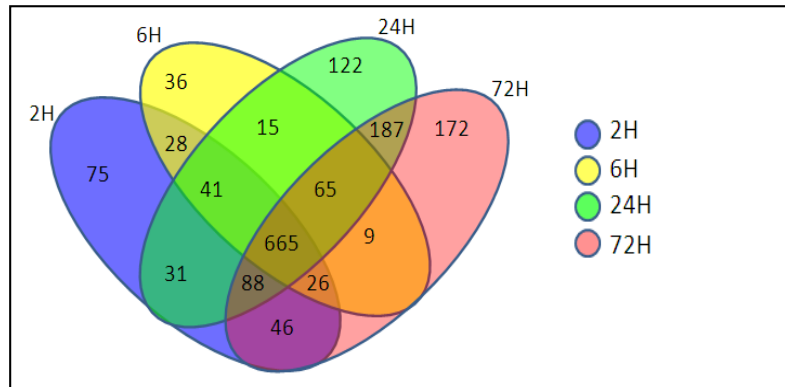


Figure 59: Diagramme de Venn de toutes les protéines identifiées pendant 2h, 6h, 24h et 72h.

Les peptides sont triés selon leur p-value (< 0,05) et leur abondance avec un ratio BaP/Contrôle > 2, tout en prenant en considération les valeurs de l'abondance du BeP (BeP ≈ Contrôle). La quantification sur MarkerView est représentée graphiquement (plot profile) au niveau peptidique et au niveau protéique. Le plot profile au niveau protéique représente la moyenne des aires des pics peptidiques de la même protéine. Notre travail a alors consisté à chercher les protéines montrant un « plot profile » pouvant être qualifié d' « idéal », que ça soit au niveau peptidique ou protéique (Figure 60).

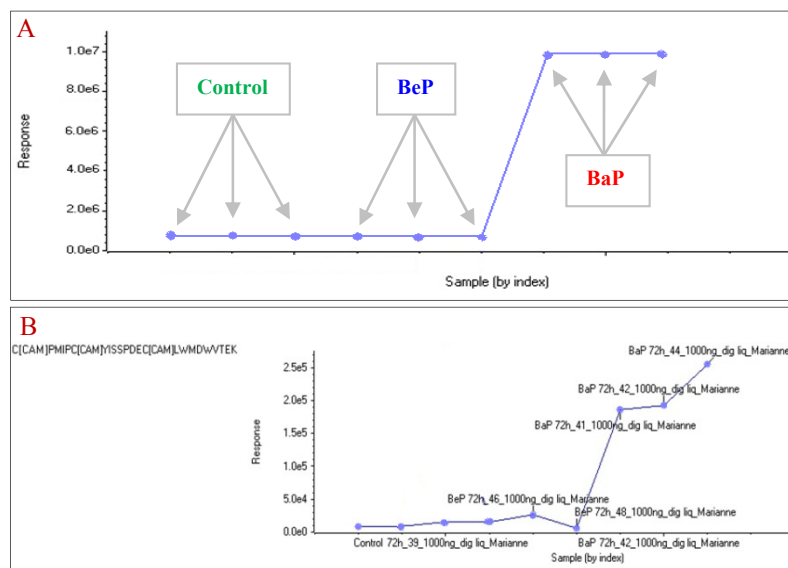


Figure 60 : Exemple de profil de régulation (appelé plot profile) tel que ceux générés dans MarkerView : A = profil caractéristique d'un peptide ou protéine qui ne serait exprimé(e) que dans la condition BaP; B = profil obtenu pour le peptide C*PMIPC*YISSPDEC*LWMDWVTEK de la protéine TIMP2_HUMAN et très proche du profil attendu décrit dans la partie A

Plusieurs peptides ayant le profil (-.-+.-.-+.+.+) nous ont attiré l'attention, parmi eux, le peptide ELDESLQVAER de la protéine CLUS_Human (Figure 61). La valeur de l'aire du pic d'éluion de ce peptide, pour chaque condition d'exposition pendant un temps d'incubation de 72h, est indiquée dans le tableau ci-dessous.

Tableau 12 : Valeurs des XICs du peptide ELDESLQVAER de la protéine CLUS_Human pour chaque condition d'exposition

Row	Index	Peak Name	m/z	Ret. Time	Group	Use
2163	1982	ELDESLQVAER	644.8239	29.42	spP10909[CLUS_	<input checked="" type="checkbox"/>

Control 72h_37_10	Control 72h_38_10	Control 72h_39_10	BeP 72h_45_1000n	BeP 72h_46_1000n	BeP 72h_48_1000n	BaP 72h_41_1000n	BaP 72h_42_1000n	BaP 72h_44_1000n
2.549e5	2.074e5	1.397e7	9.280e5	3.578e4	8.639e5	8.942e6	1.224e7	1.020e7

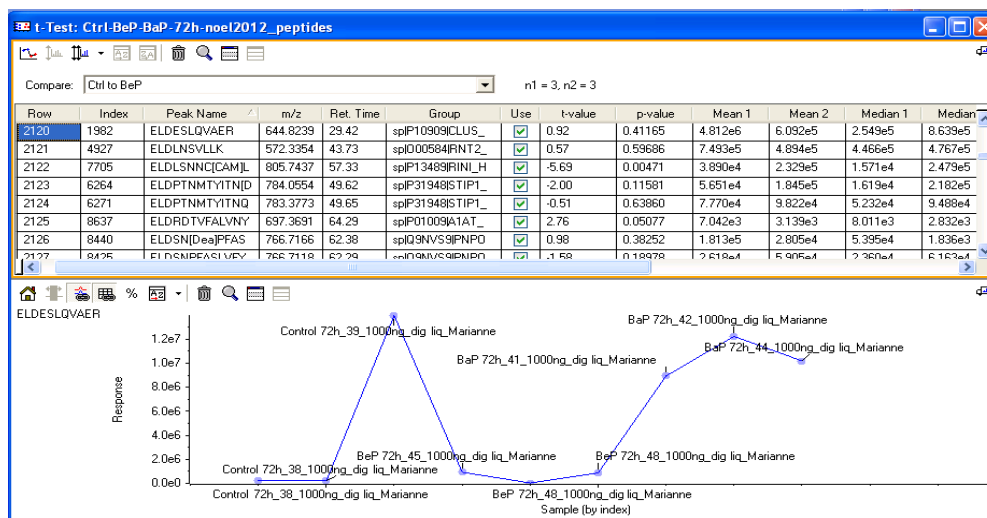
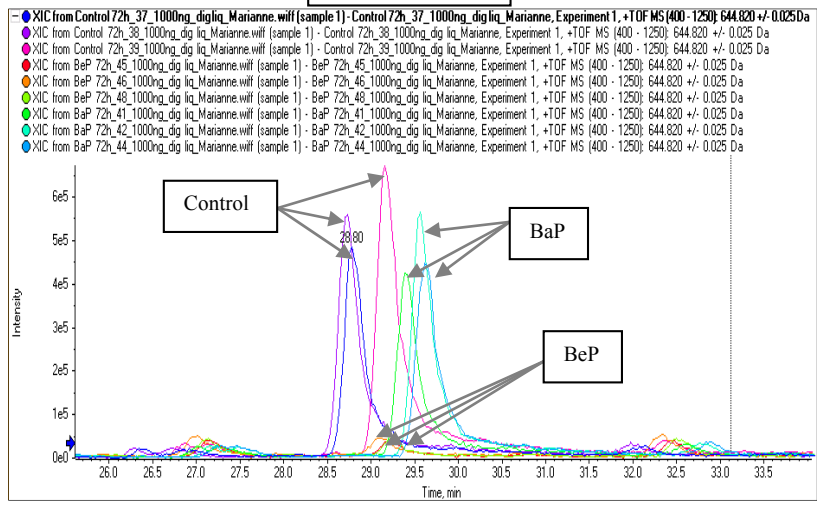


Figure 61 : Plot profile du peptide ELDESLQVAER de la protéine CLUS_Human

Afin de nous assurer de l'allure de ce peptide entre les différents triplicats, nous avons extrait, manuellement et automatiquement de PeakView, les vraies valeurs des XICs (Figure 62). En regardant les intensités, on voit bien que le profil de ce peptide est (+.+.-.-+.+.+) que ça soit au niveau manuel ou automatique. Pour expliquer les mauvaises intégrations constatées sur les deux premiers contrôles, nous pouvons émettre l'hypothèse d'un mauvais alignement des analyses (dérive chromatographique vers les temps de rétention antérieurs)

Protéine = CLUS / Peptide = ELDESLQVAER

XICs manuels



XICs automatiques

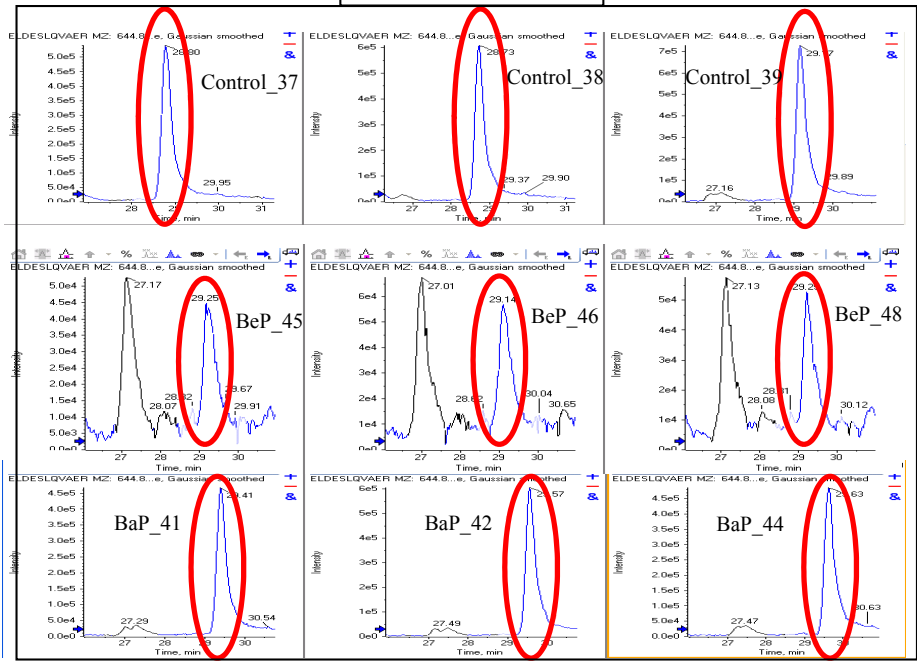


Figure 62 : XICs manuels et automatiques du peptide ELDESLQVAER de la protéine CLUS_Human générés par PeakView

3) Quantification relative à l'aide du logiciel Skyline (Université de Washington, Laboratoire MacCoss)

Le logiciel Skyline est une application client Windows développée en « open-source » (logiciel libre) par le laboratoire du Professeur MacCoss à l'Université de Washington (équipe de Brendan MacLean). Il a été développé dans un premier temps pour faciliter le traitement des données de type SRM/MRM (Selected Reaction Monitoring/Multiple Reaction Monitoring).

A la suite d'une forte demande de la communauté scientifique, de nouvelles fonctionnalités ont été rajoutées au fur et à mesure de la mise en ligne de nouvelles versions (la dernière version disponible au 1er octobre 2013 est la 2.1). Ces nouveautés incluent le traitement des données de type : (1) PRM (Parallel Reaction Monitoring dont la méthodologie MS/MS^{ALL} de chez AB Sciex, également appelée SWATH) et (2) quantification MS1 label-free (reconstitution des XICs et intégration des pics d'élution, méthode qui a été utilisée dans notre cas d'étude).

Skyline est disponible à l'adresse url suivante : <https://brendanx-uw1.gs.washington.edu/labkey/project/home/software/Skyline/begin.view>. La publication de référence du logiciel est parue en 2010 dans le journal Bioinformatics (MacLean, Tomazela et al. 2010). Il est à noter également que ce logiciel est de plus en plus utilisé par la communauté internationale (engouement fort dans les congrès nationaux et internationaux) et de nombreuses publications y font référence (Held, Schilling et al. , Bereman, MacLean et al. 2012, Schilling, Rardin et al. 2012). De plus, de nombreux efforts sont faits par l'équipe de développement de Brendan MacLean pour mettre en ligne des tutoriaux détaillés et très bien illustrés par des exemples. Un forum en ligne permet de poser des questions sur des problèmes rencontrés au cours de l'utilisation de Skyline, et une réponse est rapidement envoyée pour en faire également profiter toute la communauté des utilisateurs du logiciel.

Le traitement des données générées par le couplage nanoLC (Eksigent) et ESI-TripleTOF (AB Sciex) est possible en suivant les étapes décrites dans les paragraphes suivants.

3.1) Prétraitement des données

Le prétraitement des données consiste en l'enchaînement des 4 étapes suivantes : (1) création d'une librairie spectrale MS/MS, (2) paramétrage des données relatives aux peptides et aux transitions, (3) import des séquences protéiques et (4) import des fichiers bruts issus de l'acquisition sur le spectromètre de masse.

3.1.1) Création d'une librairie spectrale MS/MS

La première étape du prétraitement des données est la génération d'une librairie de spectres MS/MS : cette étape est cruciale et est possible grâce aux fichiers de résultats des recherches en banques de données. Dans notre cas, nous avons utilisé les fichiers de résultats générés par les moteurs de recherche Mascot (MatrixScience) et Paragon (ProteinPilot, AB Sciex), donnant respectivement des fichiers avec les extensions .dat et .group. A noter toutefois la conversion au préalable des fichiers .group en extension .group.xml avant toute utilisation. La redondance de la librairie peut être également gardée ou non, selon le choix de l'utilisateur (la redondance reflète dans ce cas le fait que plusieurs spectres MS/MS soient reliés à un même peptide). De plus, un seuil de score peut être spécifié : nous l'avons établi à 0,95 pour les identifications par Paragon et Mascot. La librairie de spectres MS/MS ainsi créée est sauvegardée sous forme d'un fichier avec l'extension .blib et elle peut être explorée à partir du menu View> Spectral Libraries : en choisissant la librairie souhaitée, il est possible de faire défiler tous les spectres MS/MS peuplant cette librairie. Dans l'exemple donné dans la figure 74, on peut ainsi voir le spectre MS/MS attribué au peptide de séquence AAADGDDSLYPIAVLIDELR, issu du fichier brut BaP72h_44_1000ng_digliq_Marianne.wiff (cf « File »). De plus, on peut y voir que le spectre a été acquis à un temps de rétention de 68,653 min et les ions fragments y/b de charge 1+ et 2+ sont annotés sur le spectre.

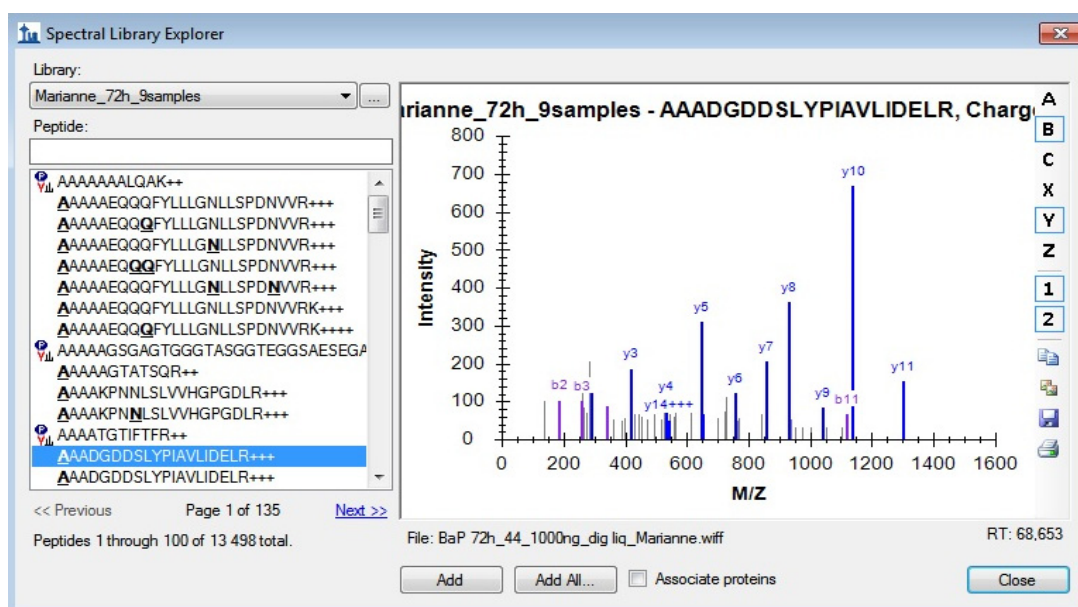


Figure 63 : Création d'une librairie de spectres MS/MS à partir d'une recherche en banque de données par l'algorithme Paragon, en utilisant les 9 fichiers relatifs à l'acquisition des 9 échantillons de la condition 72h de la cinétique d'exposition. Un total de 13498 spectres uniques (un seul spectre par peptide, celui de meilleur score Paragon) peuple cette librairie

3.1.2) Paramétrage des données relatives aux peptides et aux transitions

La seconde étape consiste à paramétrer correctement le logiciel avant de procéder à toute extraction de données. Il faut, pour ce faire, renseigner les paramètres suivants :

(1) View> Peptide settings : la table « Digestion » permet de spécifier que l'enzyme utilisée a été la trypsine, avec une tolérance de 0 mauvaises coupures : les peptides présentant en effet des mauvaises coupures à la trypsine sont susceptibles de modifier la justesse de la quantification. De plus, la table « Modifications » renseigne Skyline sur les modifications structurales tolérées (carbamidométhylation des cystéines, oxydation des méthionines et acétylation N-ter de la protéine dans notre cas) mais celles-ci ne sont pas incluses dans le choix des peptides à utiliser pour la quantification. Quant à la librairie créée à l'étape précédente, elle est ensuite sélectionnée dans la table « Library ». Le détail des paramètres utilisés est disponible dans la figure 64.

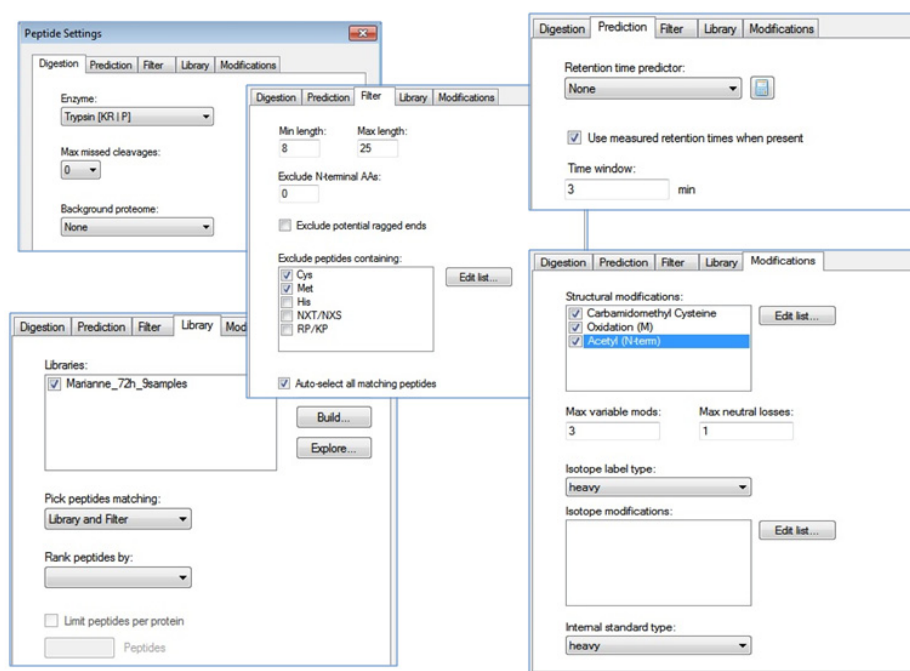


Figure 64: Paramétrage des peptides avec le détail relatif aux 5 onglets composant cette table (Digestion, Prediction, Filter, Library, Modifications)

(2) View> Transition settings : la table « Filter » permet de spécifier que l'on cherche à extraire des données des masses d'ions précurseurs en tapant la lettre « p » dans le champ « Ion types ». Quant au paramétrage pour générer les XICs, il est possible dans la table « Full-scan » dans la partie intitulée « MS1 filtering ». Le détail des paramètres utilisés est disponible dans la figure 65.

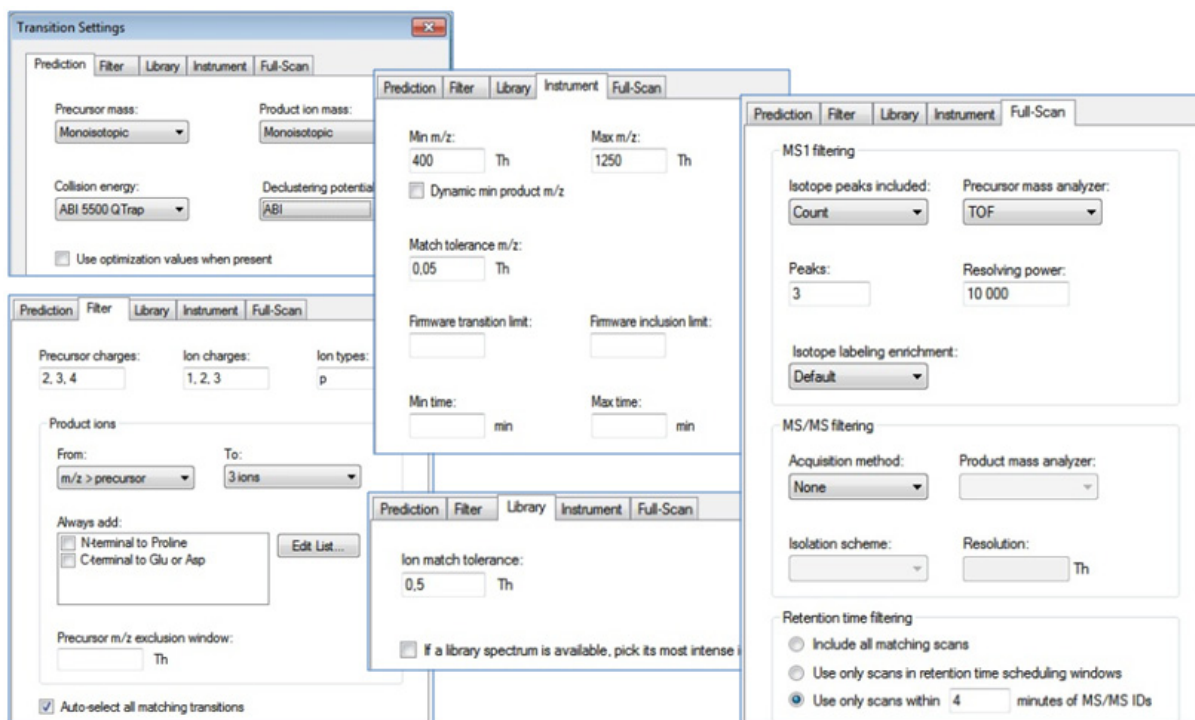


Figure 65 : Paramétrage des transitions avec le détail relatif aux 5 onglets composant cette table (Prediction, Filter, Library, Instrument, Full-Scan)

3.1.3) Import des séquences protéiques

Pour réaliser la troisième étape, nous avons importé dans Skyline le fichier .fasta contenant les séquences protéiques en acides aminés ayant servi à l'interrogation en banque de données. Ceci est indispensable pour relier les spectres MS/MS de la librairie aux protéines dont les peptides lus sur les spectres de fragmentation sont issus. Lorsque l'étape d'import des séquences protéiques est terminée, une arborescence est alors automatiquement créée dans le tableau d'affichage de Skyline figurant sur le côté gauche de la fenêtre (« Targets »). De plus, les critères à appliquer sur les peptides et transitions (rentrés dans l'étape précédente) permettent de digérer *in silico* ces séquences protéiques et de générer ainsi les peptides respectant ces critères. La figure 66 montre les peptides ainsi générés pour la protéine 1433B_HUMAN à l'aide des paramètres décrits dans les figures 64 et 65.

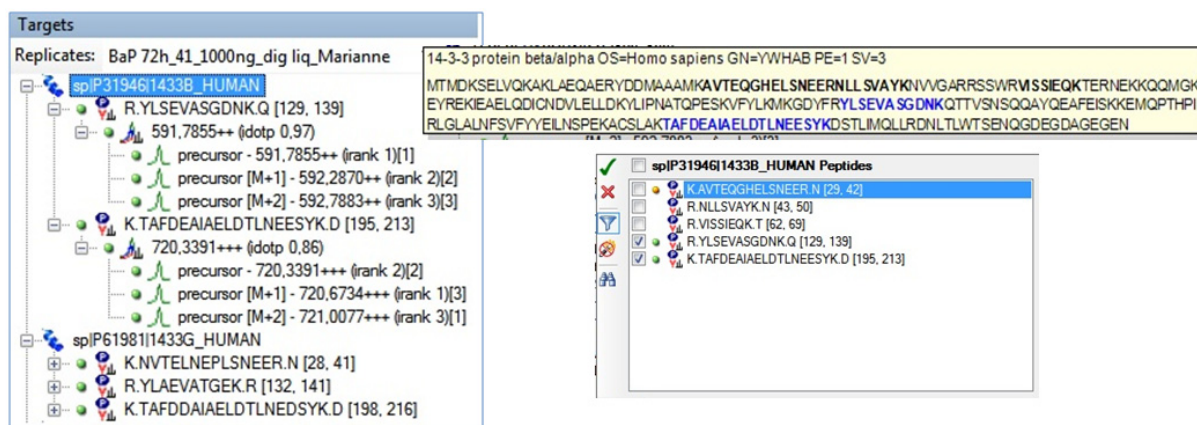


Figure 66 : Arborescence générée par l'import du fichier .fasta dans Skyline. Exemple de la protéine 1433B_HUMAN : 2 peptides présents de la librairie (m/ 591.7855 et m/z 720.3391), 2 et 3 fois chargés respectivement/ La description de la protéine et le positionnement des résidus identifiés est visualisable dans une première fenêtre pop-up. Les autres peptides répondant aux critères de filtre sont mentionnés dans une seconde fenêtre pop-up (ils ne sont cochés que s'ils sont détectés dans la librairie).

3.1.4) Import des fichiers bruts issus de l'acquisition sur le spectromètre de masse

Finalement, la quatrième étape avant la génération des XICs consiste à importer les fichiers bruts issus de l'acquisition des données sur le spectromètre de masse (fichiers .wiff du TripleTOF). Aucune limitation n'est donnée à cette étape et nous avons choisi de commencer l'évaluation du logiciel Skyline en regardant spécifiquement les 9 analyses relatives à l'effet dose 72h.

3.2) Calcul des valeurs d'aires peptidiques

Le temps de calcul du logiciel Skyline est lié à la génération des chromatogrammes extraits pour chaque précurseur listé dans l'arborescence. Le nombre de pics d'élution (XIC) à reconstituer est d'autant plus grand qu'il y a un nombre important de protéines identifiées et validées à $FDR < 1\%$, et qu'il y a de peptides rattachés à ces protéines. Cette étape est décisive pour pouvoir ensuite travailler sur les valeurs d'aires générées et en faire ressortir l'information du sous-ensemble des protéines régulées positivement ou négativement : une mauvaise reconstitution des XIC entraînerait le calcul d'un ratio erroné entre 2 conditions d'étude et fausserait ainsi le processus du choix d'un biomarqueur.

Contrairement aux autres logiciels (PeakView, Progenesis, ou encore d'autres logiciels de quantification MS1), Skyline ne reconstitue pas un pic d'élution unique d'un précurseur donné, mais un XIC est généré pour chacun des 3 premiers isotopes du massif peptidique (Figure 67). De ce fait, l'utilisateur pour vérifier s'il existe une superposition de ces 3 XIC (du fait de la co-élution évidente des isotopes liés à un même peptide). Cela revient au final à ne

valider la quantification d'un peptide que dans le cas d'une superposition de 3 XICs (appelée « peak group »)

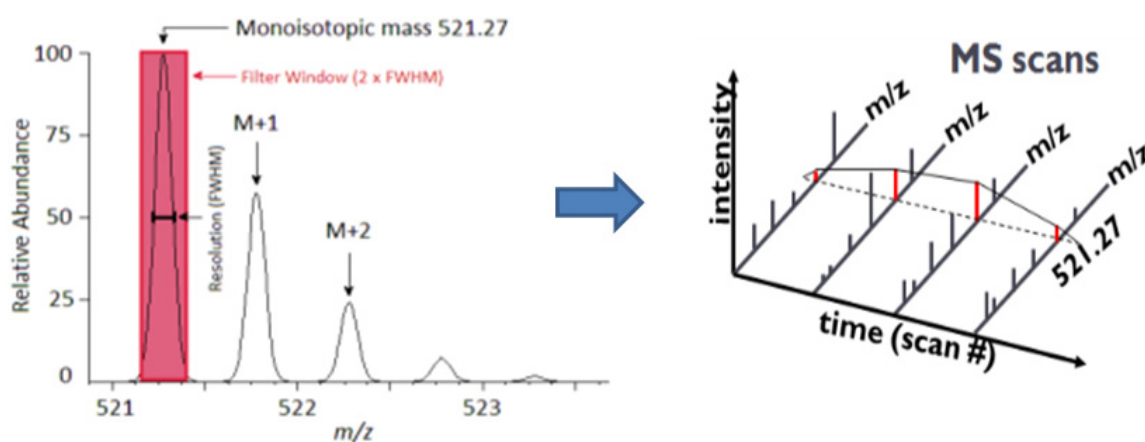


Figure 67 : Reconstitution d'un pic d'élution peptidique (XIC) à partir de la masse d'un précurseur donné ainsi que la masse des 2 isotopes suivants. Figures issues du tutorial rédigé par Brendan LacLean (https://brendanx-uw1.gs.washington.edu/labkey/_webdav/home/software/Skyline/%40files/tutorials/MS1Filtering-1_4.pdf)

3.2.1) Outils complémentaires à l'analyse des données

3.2.1.1) Outil « Peak Areas »

Cette fenêtre supplémentaire permet la visualisation sous forme d'un histogramme de toutes les aires des XICs obtenues pour chacune des conditions d'étude, et ceci pour un peptide donné. Chaque barre individuelle représente une condition et est elle-même la combinaison des aires des 3 isotopes du massif peptidique (Figures 68 et 69, représentant respectivement un peptide peu et très abondant). Une réintégration manuelle est également possible pour chaque XIC en indiquant le début et la fin de l'intégration à l'aide du curseur de la souris. A noter qu'un spectre MSMS acquis pour ce peptide est indiqué par un trait vertical avec l'annotation ID (le spectre utilisé pour la librairie est signalé en rouge).

Grâce à cet outil, nous avons pu rapidement évaluer la reproductibilité des triplicats biologiques (barres de même hauteur pour 3 expériences provenant de la même condition d'étude) ainsi que les différences existant entre plusieurs conditions. Pour plus de clarté, nous avons choisi de travailler sur Skyline avec un sous-ensemble de 9 analyses (3 Contrôles, 3 BeP, 3 BaP), c'est-à-dire une session Skyline par chaque condition de temps de cinétique : cela nous a permis de juger les performances et la convivialité du logiciel Skyline sans surcharger les visualisations graphiques avec le total des 36 conditions.

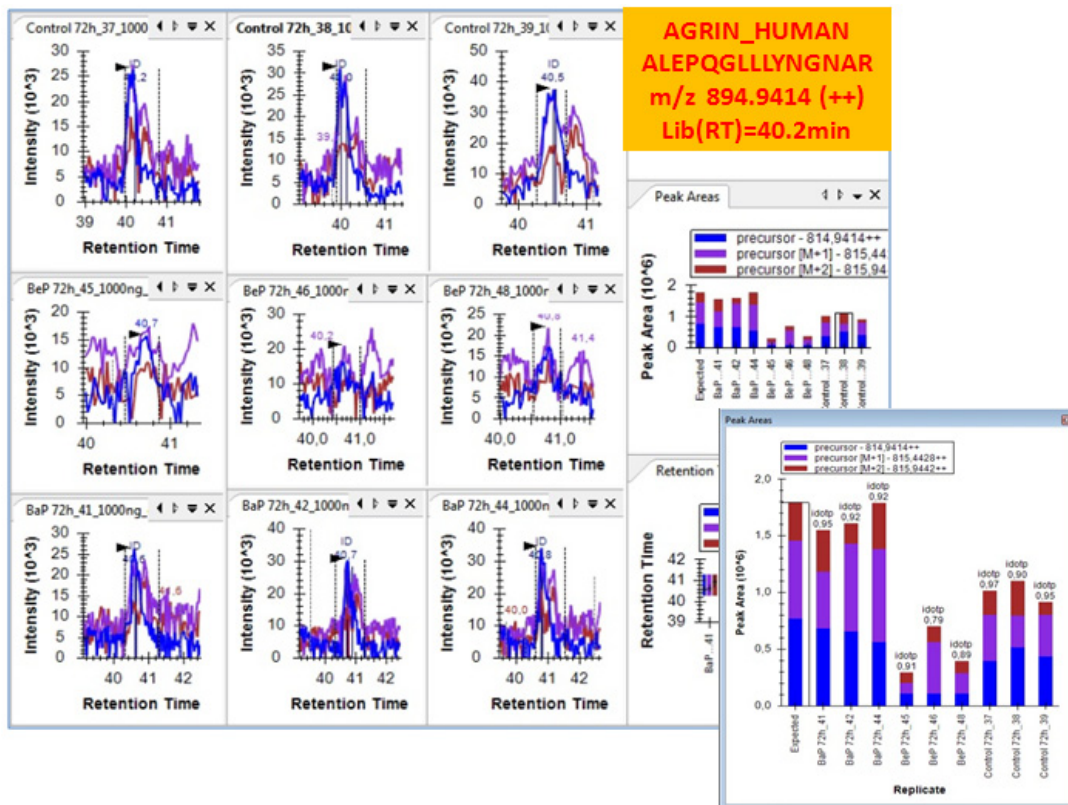


Figure 68 : Reconstitution des 9 XICs d'un peptide de la protéine AGRIN_HUMAN (9 fenêtres de gauche).
 Le diagramme "Peak Areas" (fenêtre du milieu à droite et zoom en bas à droite) montre des pics d'éluion avec une aire plus faible dans les triplicats de la condition BeP (au centre) par rapport aux triplicats des conditions BaP et Contrôle (à gauche et à droite).
 Les plus fortes aires sont celles attribuées à la condition BaP.

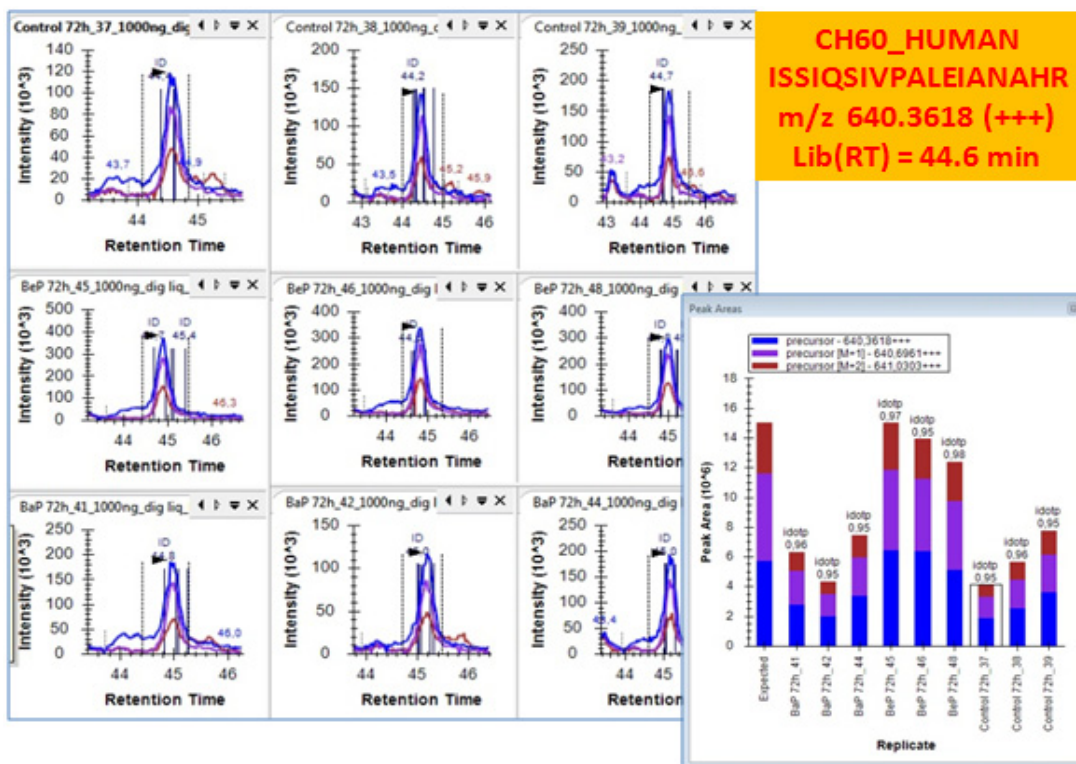


Figure 69 : Reconstitution des 9 XICs d'un peptide de la protéine CH60_HUMAN (9 fenêtres de gauche).
 Le diagramme "Peak Areas" montre une régulation positive dans la condition BeP

3.2.1.2) Outil « isotopic dot plot »

Cet outil a pour but d'évaluer la contribution de chacun des 3 isotopes du massif à la valeur d'aire globale attribuée à chaque peptide. En effet, la distribution isotopique est généralement en faveur d'un pic monoisotopique plus grand pour les précurseurs de faible m/z, tandis que cela s'inverse en faveur du deuxième isotope lorsque la masse du précurseur augmente.

Pour faciliter l'évaluation de la contribution isotopique, le logiciel Skyline reconstitue à gauche de la fenêtre « Peak Areas » la contribution isotopique théorique de chaque peptide en s'aidant de sa séquence en acides aminés (Figure 70). Ainsi, une valeur comprise entre 0 et 1 peut être attribuée à chaque « isotopic dot plot » d'une condition donnée, reflétant la similitude plus ou moins poussée entre la contribution théorique et celle observée expérimentalement (1 si la similitude est totale).

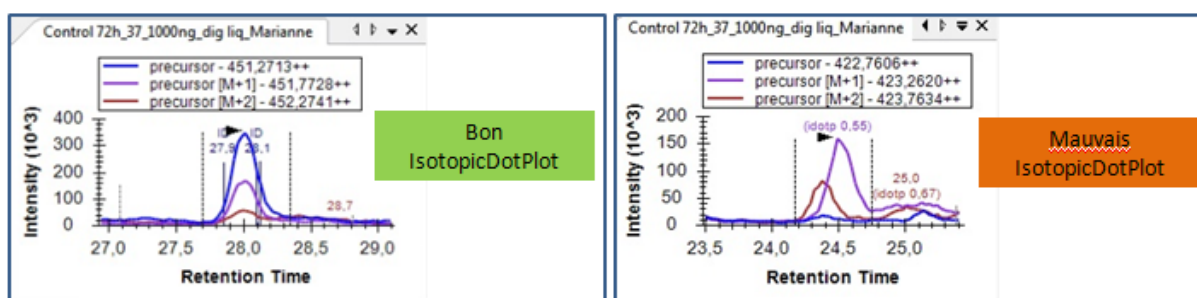


Figure 70 : Outil « Isotopic dot plot ». Le XIC de la fenêtre de gauche montre une répartition isotopique proche de la contribution théorique (idotp=0.96) tandis que la fenêtre de droite montre un XIC pour lequel les 2^{ème} et 3^{ème} isotopes perturbent la contribution isotopique générale du peptide (idotp=0.55).

3.1.2.3) Outil « Retention time alignment »

Ce troisième outil est particulièrement important car il permet de visualiser d'un rapide coup d'œil si tous les XICs ont été générés à un temps de rétention proche ou non. La figure 71 montre l'alignement des temps de rétention pour un peptide donné. Une correction manuelle peut être effectuée au besoin.

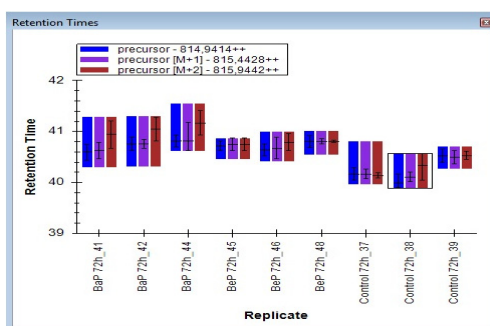


Figure 71 : Outil « Retention time alignment » : intégration automatique des XICs pour un peptide de la protéine AGRIN_HUMAN dont le détail des intégrations est donné dans la figure 6.

3.3) Analyse statistique

Les valeurs d'aires obtenues pour chaque peptide et pour chaque condition peuvent alors servir à calculer les ratios entre 2 conditions données : la visualisation de la régulation d'un peptide à l'aide de l'outil « Peak Areas Comparison » est très simple. De plus, nous avons pu facilement vérifier que la même régulation était observable sur les autres peptides (s'il y en avait bien sûr) de la même protéine (Figure 72).

Un export dans Excel peut être réalisé pour un retraitement plus poussé des données d'aires (calcul de la moyenne, déviation standard, coefficient de variation). Lorsque nous avons testé le logiciel Skyline, celui-ci ne disposait pas encore d'un outil statistique intégré (QuaSAR Quantitative Statistics) comme cela est le cas maintenant dans la version disponible sur le site du laboratoire MacCoss.

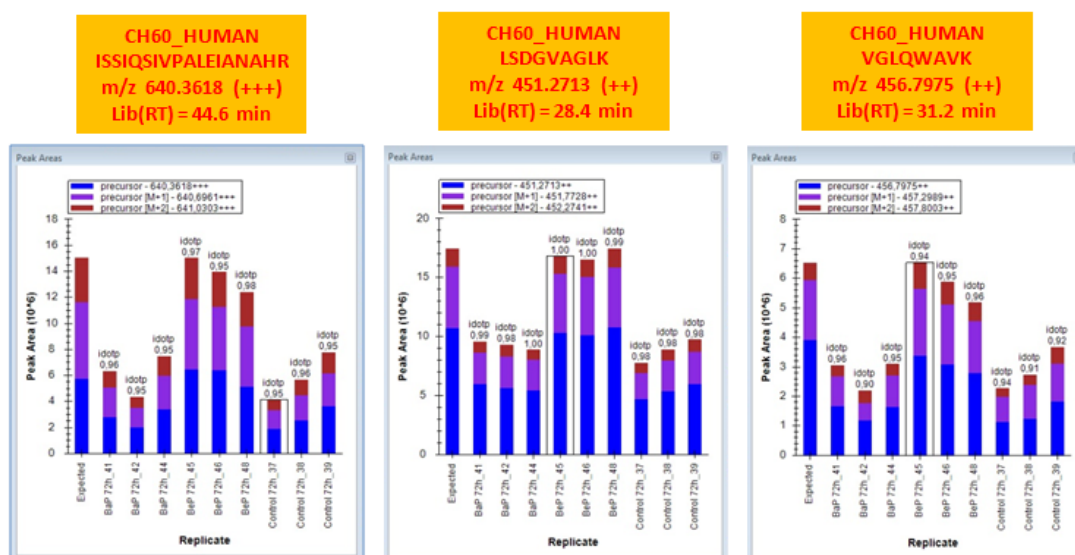


Figure 72 : Recherche de protéines et de peptides montrant un profil de régulation (sur- ou sous-) à l'aide de l'outil « Peak Areas Comparison ». Evaluation du comportement de la protéine CH60_HUMAN à l'aide de l'ensemble de ses peptides. L'exemple montre les 3 peptides les plus intenses utilisés pour la quantification de la protéine.

4) Quantification relative à l'aide de Progenesis LC-MS (Nonlinear Dynamics)

4.1) Pre-processing des données

Les 36 « .wiff » (3 conditions d'exposition * 4 temps d'incubation * 3 réplicats) sont importés directement dans Progenesis. Chaque « data file » est représenté par un diagramme à deux dimensions (RT en abscisses et m/z en ordonnées). Le « run » de référence a été sélectionné automatiquement et tous les autres « runs » ont été alignés à celui-ci. L'étape d'alignement est très rapide, elle a duré 1h38 minutes (Figure 73). Progenesis indique, en %, la valeur de score d'alignement de chaque « run » par rapport à celui de référence (Figure 74). Nous considérons que l'alignement est bon à partir d'une valeur de 60%. Une zone verte correspond à un bon alignement, auquel cas il n'est pas nécessaire de l'examiner manuellement.

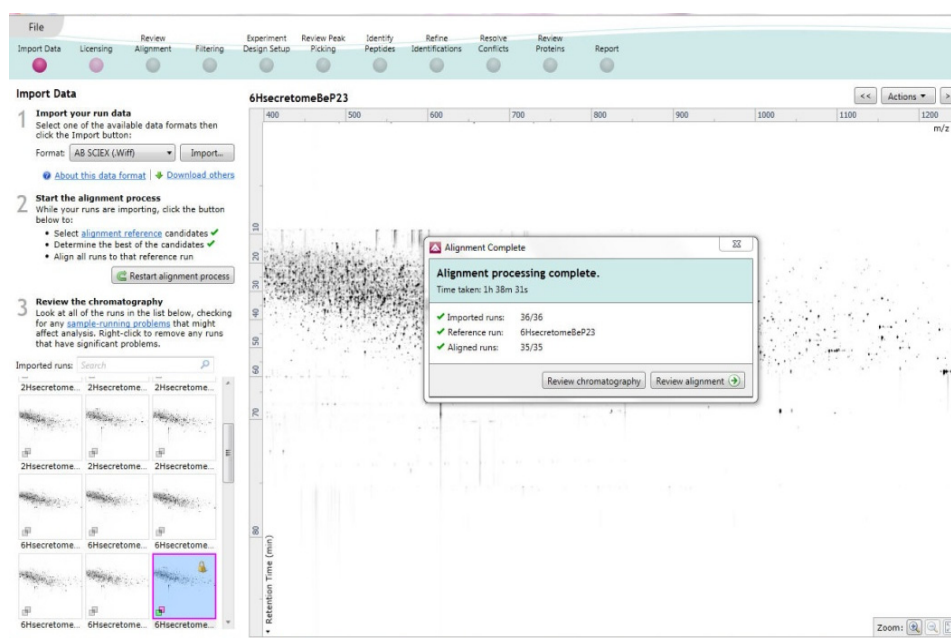


Figure 73 : Durée d'alignement des 36 « runs »

Run	Include?	Vectors	Score
24HsecretomBep35	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	524	86,9%
24HsecretomeBap29	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	422	85,4%
24HsecretomeBaP30	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	275	72,5%
24HsecretomeBaP31	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	310	75,0%
24HsecretomeBep33	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	471	84,9%
24HsecretomeBep34	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	505	85,8%
24HsecretomeCtrl25	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Ref	
24HsecretomeCtrl26	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	565	93,4%
24HsecretomeCtrl27	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	658	86,2%

Table d'alignements des vecteurs et scores de quelques « runs ».

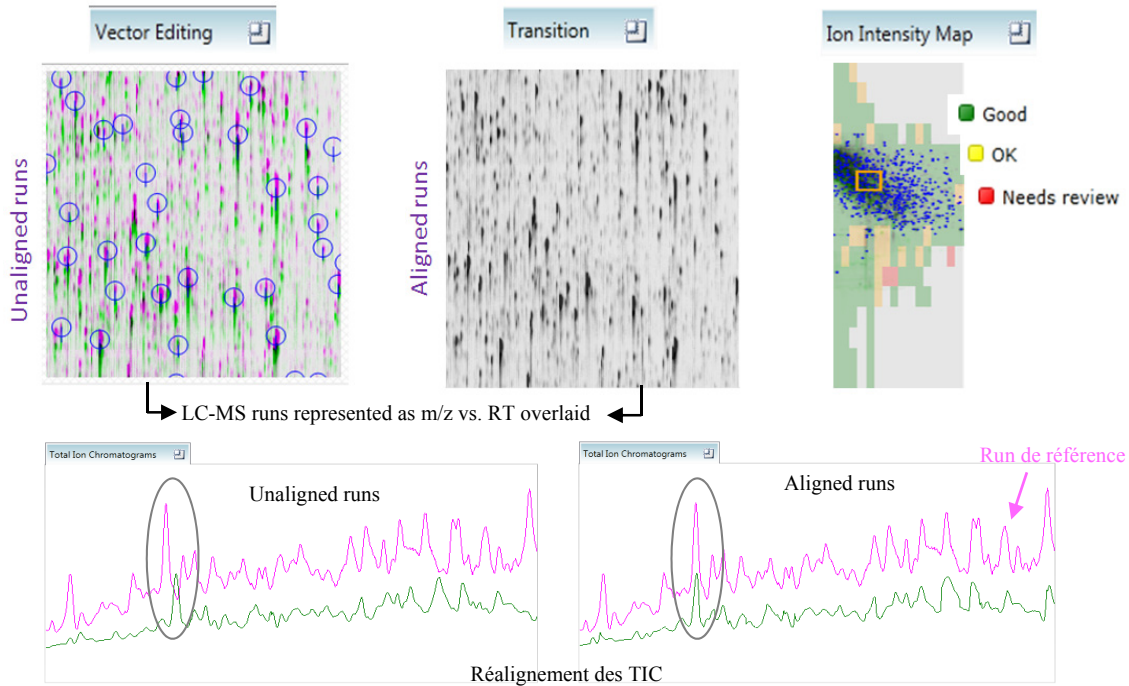


Figure 74 : Réalignement des Cartes 2D « massiques »

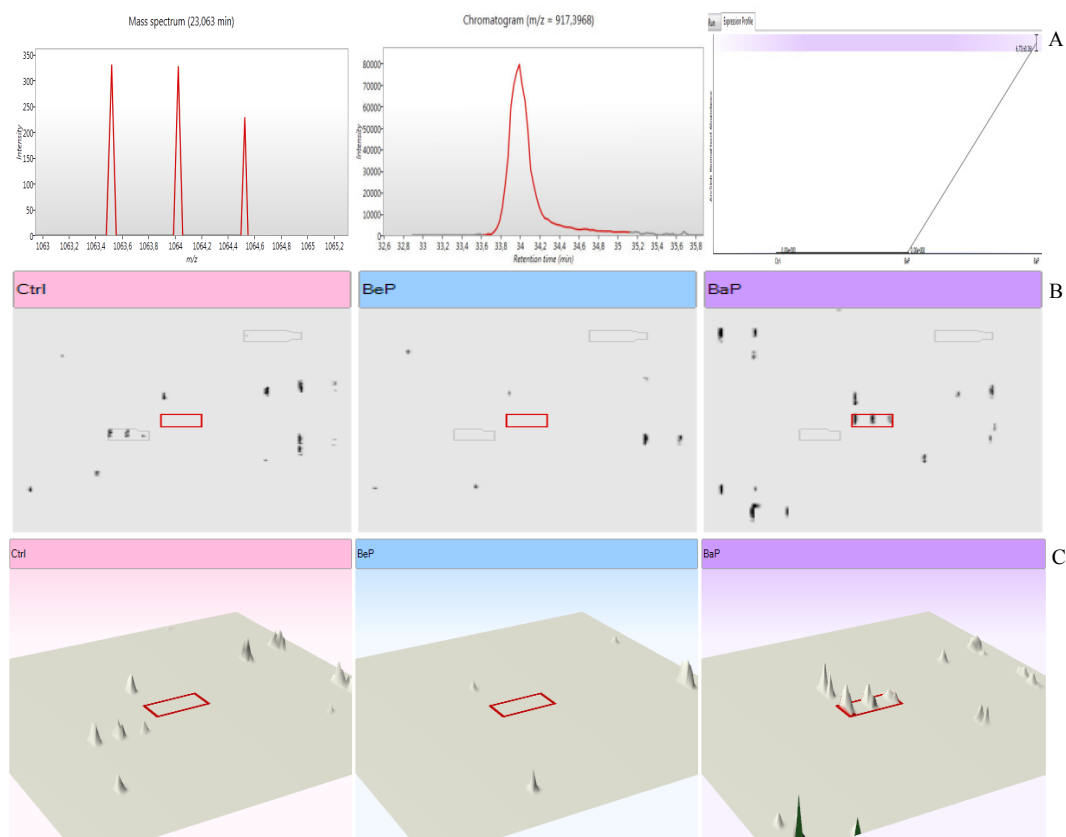
Pour la détection des peptides, des paramètres de « Peak Picking » peuvent être fixés (charge maximale et limites de temps de rétention). Dans notre cas, nous avons laissé les paramètres par défaut (charge maximale = 20, temps de rétention entre 10 et 75min). Après leur détection, les peptides sont filtrés en laissant ceux qui sont 2 à 7 fois chargés. Une normalisation automatique est ensuite appliquée sur tous les peptides de chaque « run ».

Pour l'identification, les spectres MS/MS de notre sélection sont exportés sous forme d'un sous-fichier « .mgf » et identifiés à l'aide du moteur de recherche Mascot. La taille du fichier a été réduite en ne soumettant que les 4 meilleurs spectres MS/MS du même peptide plutôt que tous ses spectres. Les paramètres fixés sur Mascot sont inclus dans la partie « méthodes » du chapitre 2. Une fois la recherche finalisée, les résultats Mascot sont exportés

sous forme d'un fichier « .xml » et importés au sein de Progenesis. Toute identification ayant un score inférieur à 25 a été éliminée.

4.2) Calcul des valeurs d'aires des pics peptidiques

La quantification par Progenesis a été faite à partir des peptides protéotypiques de chaque protéine identifiée. Progenesis extrait le XIC de chaque peptide et utilise la moyenne normalisée des des abondances de chaque peptide par condition. Les volumes de pics résultants peuvent être donc comparés de façon quantitative entre les différentes conditions. Les résultats peuvent être visualisés sous la forme 1D, 2D ou 3D (Figure 75).



4.3) Analyse statistique

A la différence de la suite logicielle de Sciex qui traite statistiquement les données par le *T*-test permettant de comparer les moyennes de deux groupes indépendants, Progenesis utilise l'ANOVA test (ANalysis Of Variance) qui permet de comparer les moyennes de plusieurs groupes indépendants et de représenter les résultats sous forme de profil d'expression (Figure 76). Toutes les protéines que nous avons sélectionnées ont une *p*-value $\leq 0,05$. Progenesis est capable de créer des dendogrammes pour illustrer des regroupements de protéines différemment exprimées entre les groupes (Figure 77). Il est aussi capable de réaliser une

PCA sur les données pour la mise en évidence de similitudes ou de différences entre les différents groupes étudiés (condition d'exposition et temps) (Figure 77).

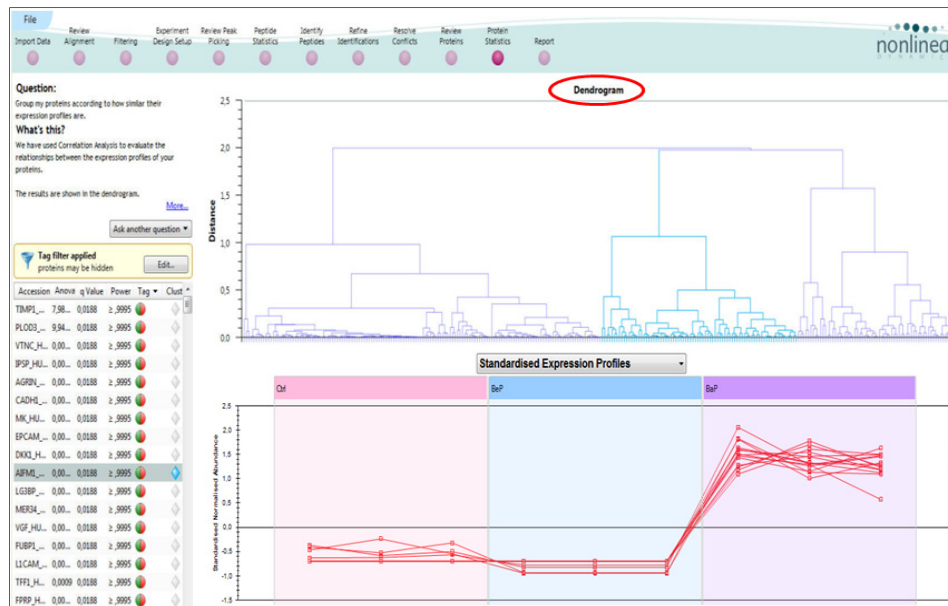


Figure 76 : Illustration des protéines regroupées sous forme d'un dendrogramme

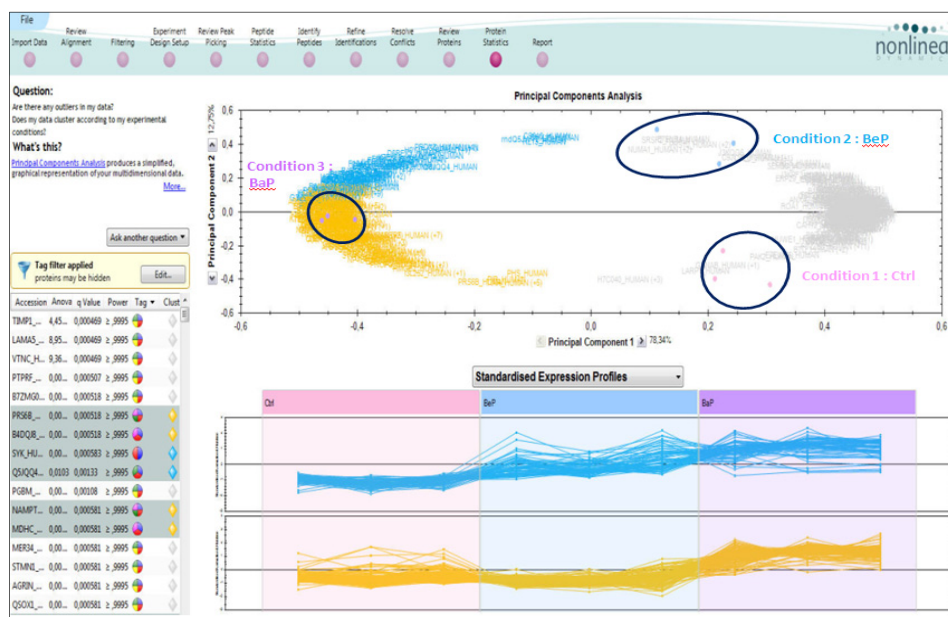


Figure 77 : Illustration des protéines regroupées sous forme d'une PCA

4.4) Résultats et discussion

Les résultats des PCA individuelles (3 conditions d'exposition pour chaque temps d'incubation) et de la PCA globale (3 conditions d'exposition et 4 temps d'incubation) sont représentés dans les figures 78 et 79, respectivement. Ces résultats ressemblent à ceux obtenus par MarkerView (AB Sciex) et montrent une similitude de l'abondance de chaque protéine entre les réplicats. La PCA globale n'est pas exécutable sur Progenesis, chaque condition de temps a été faite

séparément. La PCA globale a été réalisée sur MarkerView mais en prenant les valeurs des moyennes des abondances exportées de Progenesis. La figure 79, représentant la PCA globale, montre bien la différence entre les 2 groupes BaP et contrôles.

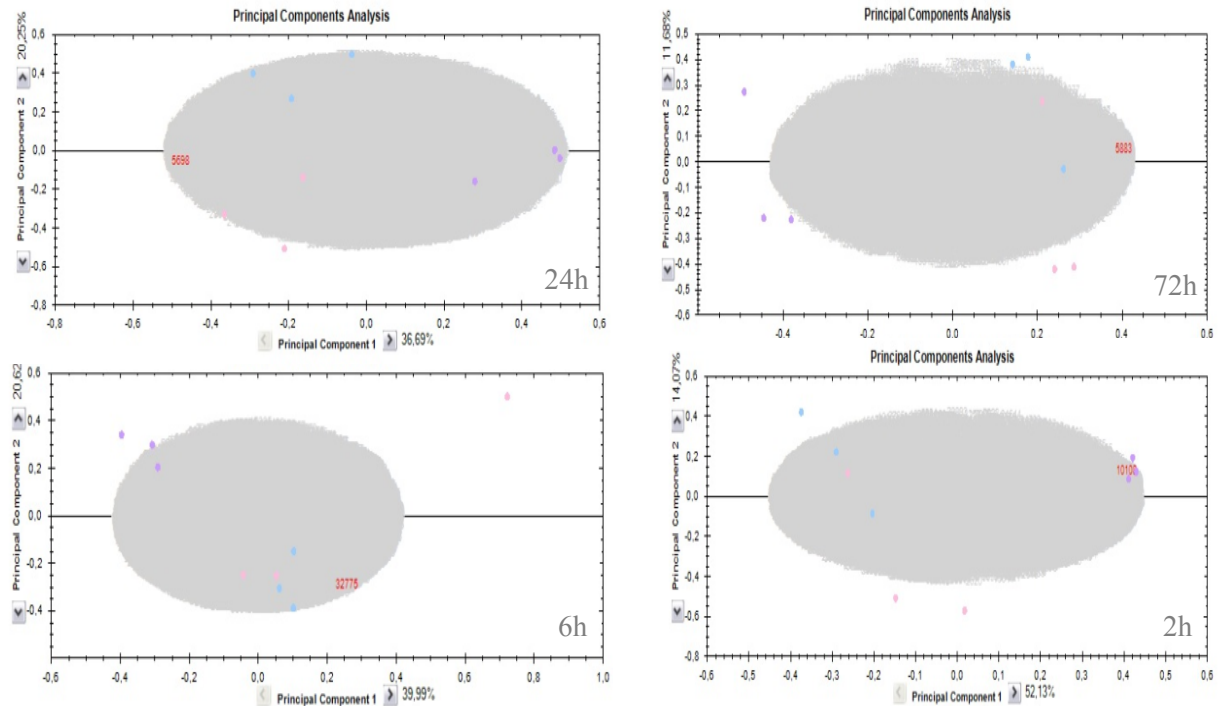


Figure 78 : PCA individuelles des 4 temps d'incubation (2h, 6h, 24h et 72h) réalisées sur Progenesis (A) 2h, (B) 6h, (C) 24h et (D) 72h

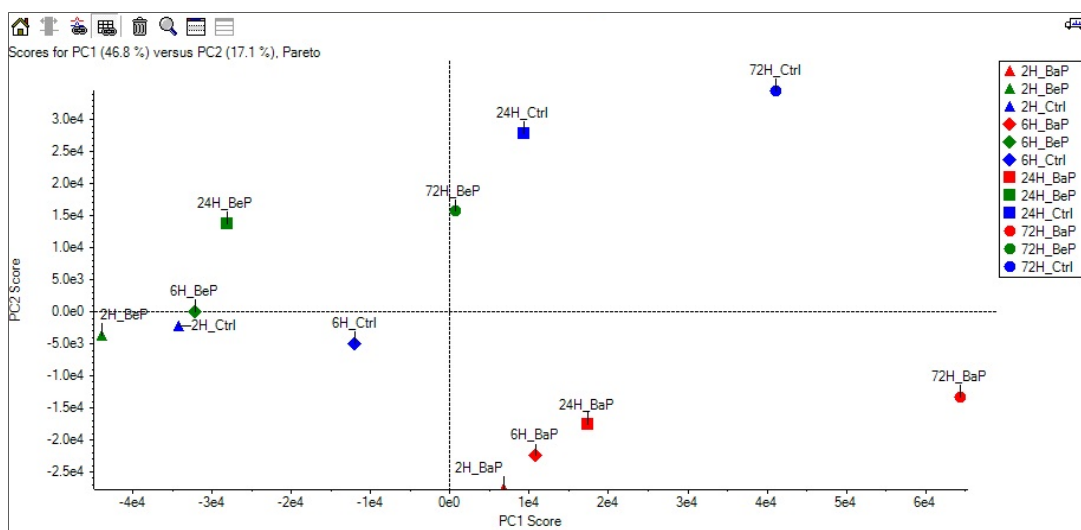


Figure 79 : PCA globale des 4 temps d'incubation (2h, 6h, 24h et 72h) en prenant la moyenne des abondances protéiques de chaque triplicat

Les protéines identifiées sont au nombre de 897 dont 847 répondent au critère FDR < 1%. Les résultats de la quantification de cette partie nous ont permis de rédiger une publication

regroupant les résultats méthodologiques et biologiques obtenus sur deux temps de la cinétique d'exposition et de la soumettre dans le journal MCP. A ce jour, et à nos connaissances, aucune étude sur le benzo(a)pyrène n'a été publiée avec ces méthodes quantitatives et ce design expérimental.

5) Retour d'expérience sur les logiciels pour la quantification relative des protéines par approche MS1 label-free

5.1) *ProteinPilot/PeakView/MarkerView (AB Sciex)*

La suite logicielle AB Sciex, comprenant ProteinPilot (recherche en banque de données), PeakView (génération et intégration des XIC) et MarkerView (validation statistique) a été le premier outil que j'ai évalué. En effet, c'est une solution développée spécialement pour répondre aux utilisateurs des systèmes instrumentaux AB Sciex, dont le TripleTOF fait partie. De plus, de nombreuses publications utilisent cette stratégie, qui nous est dès lors apparue comme la plus adéquate dans un premier temps. Un seul jeu de données a été choisi, celui relatif au temps de cinétique 72h (comparaison de 3 sous-ensembles de 3 conditions), et nous nous sommes rendu compte que la suite logicielle AB Sciex présente de nombreux avantages :

- Une interface très conviviale permet de facilement prendre en main les différents logiciels. En effet, ProteinPilot est très simple à utiliser et son paramétrage ressemble fortement à celui d'une recherche Mascot. L'intégration d'outils tels que le calcul automatique du taux de faux-positifs et l'export de comptes-rendus d'analyse sont d'une grande utilité. Quant au second logiciel, PeakView, les paramètres accessibles sont limités mais représentent les points-clés nécessaires à toute bonne méthode quantitative : cela évite ainsi de se perdre dans une multitude de paramètres (logiciel de type « usine à gaz ») et risquer de générer des erreurs involontaires. Il est important que ces paramètres clés intègrent notamment le choix de certains peptides, de préférence de caractère protéotypique et non modifiés.
- Quant à MarkerView, il est très appréciable d'utiliser un logiciel de validation statistique accessible à un non-spécialiste en statistiques dans une interface simple à utiliser et conviviale. Le test d'Analyse en Composantes Principales (PCA) est très « visuel » et permet d'immédiatement se rendre compte si une expérience a correctement fonctionné ou non : bonne reproductibilité des réplicats et regroupement possible des échantillons selon la condition expérimentale testée. Si certains échantillons paraissent accroître la variabilité des observations au sein d'un groupe de réplicats (triplicats dans notre cas), celui qui génère la disparité peut être mis de côté avant de réaliser le test statistique proprement dit.

- Un autre avantage de la suite logicielle AB Sciex réside dans la possibilité d'effectuer un test de Student : ce test statistique peut être aisément programmé, rendant ainsi les résultats protéomiques générés statistiquement valides. Cette étape est indispensable pour augmenter la confiance dans les protéines qualifiées de « régulées » par l'une des conditions d'exposition.

Outre les avantages précédemment décrits pour cette suite logicielle, les inconvénients suivants ont été observés :

- Même si une recherche ProteinPilot est facile à programmer en utilisant l'algorithme Paragon, la durée de la recherche est très longue : cela est imputable au fait qu'un important nombre de modifications est recherché lors de l'interrogation en banque de données (environ 200 modifications post-traductionnelles, décrites notamment dans Unimod <http://www.unimod.org>). Pour donner un ordre de comparaison, seules 9 modifications sont admises au maximum pour une recherche Mascot. De plus, plus un échantillon est complexe (plusieurs centaines voire un millier de protéines identifiées d'où un nombre important de spectres MS/MS) et plus la banque interrogée est grande, plus le temps de la recherche sera important. Dans notre cas, pour le jeu de données relatif au temps de cinétique 72h, j'ai utilisé une recherche intégrant toutes les modifications potentielles et une banque UniProtKB restreinte à la taxonomie *Homo sapiens* : le temps de recherche Paragon s'est élevé à 1 jour et demi avec les 9 fichiers bruts .wiff (ce temps est monté à 4 jours en intégrant l'ensemble des 36 acquisitions). Bien entendu, cela est dépendant de la performance de la lame du serveur sur laquelle la suite logicielle est installée (depuis, le nombre de processeurs a été doublé et la mémoire vive a été quadruplée).
- Seuls des données générées par des instruments vendus par AB Sciex (avec une extension de type .wiff et .wiff.scan) peuvent être traitées par le logiciel PeakView. En effet, PeakView a besoin des fichiers d'entrée suivants : les données brutes issues du spectromètre de masse (.wiff) et du fichier de résultat généré par ProteinPilot (.group, qui peut être quant à lui généré à partir d'un fichier traité tel qu'une extension .mgf par exemple).
- Toujours par rapport à PeakView, tous les XICs du même peptide de toutes les conditions ne peuvent pas être visualisés en un seul coup d'œil. Cela rend donc impossible le fait d'avoir une vue rapide de l'ensemble des conditions à comparer, pour un même peptide puis à l'échelle de la protéine.
- Les XICs sont longs à générer (environ 1,5min pour le XIC d'un seul précurseur en interrogeant 9 conditions). De plus et d'après nos résultats, l'intégration automatique des

pics et l'export de ces valeurs d'aires dans MarkerView ne correspond pas toujours à un travail manuel de génération et d'intégration des XICs pour le même précurseur. Un exemple de mauvaise intégration est décrit dans la figure 80 : la partie supérieure montre les XICs générés manuellement en spécifiant le rapport m/z d'un précurseur donné à PeakView pour les 3 conditions et 3 réplicats du temps de cinétique 72h. On constate que ce peptide est disparu dans la condition BeP alors que l'intensité des XICs reste quasi constante dans les conditions Contrôle et BaP. Si l'on devait tracer le profil d'expression théorique de ce peptide à partir de ces données, nous obtiendrions le profil représenté dans la partie intermédiaire de la figure. Or, nous constatons nettement qu'une erreur est intervenue dans l'intégration des 2 premiers réplicats de la condition Contrôle alors que le troisième réplicat est correct.

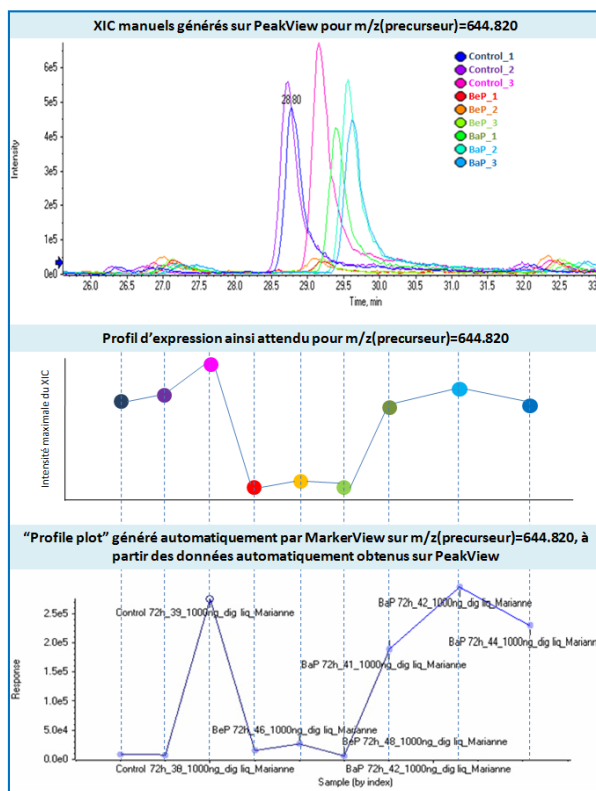


Figure 80 : Comparaison de l'intensité maximale des XICs générés de manière manuelle ou automatique par PeakView sur le précurseur avec m/z=644,820

- Le test statistique de Student n'est pas forcément adapté à notre design expérimental comprenant 3 conditions à comparer avec 3 réplicats par condition, et encore moins sur le jeu de données final composé de 36 analyses. En effet, ce test peut s'avérer trop simpliste dans notre cas car nous devons à la fois comparer 3 groupes d'échantillons (Contrôle, BeP et BaP) mais aussi intégrer une dimension de cinétique.

5.2) Skyline (Laboratoire MacCoss, Université de Washington)

En utilisant le jeu de données relatif au temps de cinétique 72h (comparaison de 3 sous-ensembles de 3 conditions), nous nous sommes aperçus que le logiciel Skyline avait de nombreux avantages qui faisaient défaut dans le logiciel précédemment testé (PeakView/MarkerView) sans toutefois avoir un processus de traitement des données fluide et sans erreurs.

Les avantages que nous avons trouvés à l'utilisation de Skyline dans le cadre de la quantification relative des protéines par l'approche MS1 label-free sont les suivantes :

- Toutes les conditions peuvent être visualisées et comparées en un rapide coup d'œil en alignant les XICs dans des fenêtres les unes à côté des autres (donnant ainsi dans notre cas une matrice à 3 lignes et 3 colonnes, une ligne correspondant à 1 condition, une colonne correspondant à un réplicat). L'environnement des pics d'élution générés peut donc être comparé et servir à corriger manuellement une mauvaise intégration, tout en s'aidant des outils tels que « Peaks Areas Comparison » et « RT alignment ».
- Le développement du logiciel Skyline est assuré par une équipe dynamique (mise à jour régulière de la version du logiciel avec implémentation de nouvelles fonctionnalités) dans un laboratoire reconnu pour ses compétences en protéomique, spectrométrie de masse et bio-informatique (laboratoire MacCoss, Université de Washington). De plus, Skyline est développé sous forme d'un logiciel « open-source » non-payant, permettant ainsi un téléchargement rapide et facile par n'importe quel laboratoire ou plateforme souhaitant utilisée une solution de retraitement des données MS pour la quantification relative des protéines.
- Skyline permet de traiter des données issus d'un nombre important de marques de spectromètres de masse (AB Sciex, Bruker, Thermo, Waters, Agilent) et la quantification est réalisable par plusieurs types d'approche (MS1, SRM/MRM, Independant Data Acquisition tel que la méthodologie SWATH ou MSE).

A l'inverse, les semaines d'évaluation de Skyline passées sur notre jeu de données ont mis en évidence les limitations suivantes :

- Mauvaises intégrations des XICs : la validation manuelle des données est une étape limitante en termes de temps consacré à chaque jeu de données. Une mauvaise intégration des XICs se traduit par une erreur aux limites ou bornes de la fenêtre d'intégration (Figure 81), ce qui rend encore plus compliquée la mise en place d'un processus fiable et rapide.

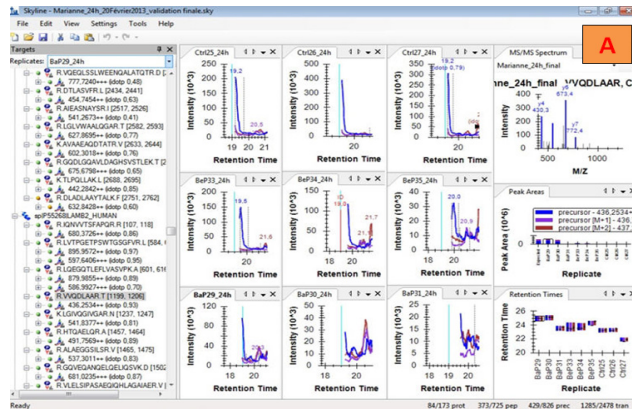


Figure 81 : Exemple d'une mauvaise intégration d'un XIC par Skyline. Le XIC est tronqué vers la gauche de la fenêtre de temps de rétention.

- XICs intégrés à des temps de rétention différents pour un même peptide : dans la figure 82, nous voyons que 7 XICs ont été obtenus à un temps de rétention de 44,6min, correspondant au temps référencé dans la librairie spectrale. Par contre, les conditions control72h_38 et control72h_39 montrent des XICs générés à des temps de rétention antérieurs, 42,1min et 42,5min, respectivement. Cette mauvaise attribution provient du fait qu'un autre « peak group » était mieux visualisé à ce temps de rétention et que le logiciel l'a retenu car il était plus intense que celui détecté au temps de rétention de la librairie. Nous avons alors intégré manuellement le « peak group » au bon RT pour corriger cette mauvaise attribution et l'alignement alors obtenu pour les 9 analyses est tout à fait correct.

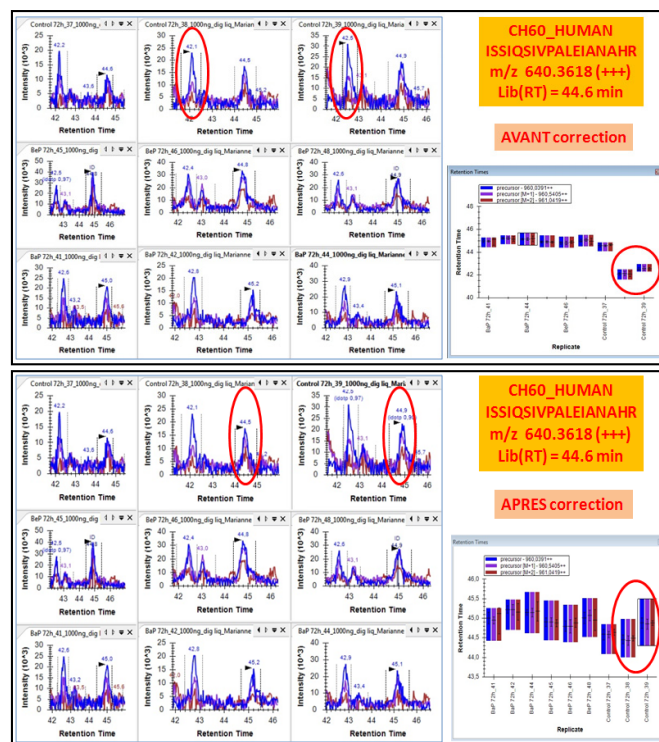


Figure 82 : Correction manuelle de l'intégration automatique des XIC pour un peptide donné de la protéine CH60_HUMAN (le RT ne correspondant pas à celui indiqué dans la librairie spectrale, à savoir 44,6min)

- Mauvais choix du peak group : même en spécifiant une fenêtre de temps relativement réduite, il est possible de trouver plusieurs superpositions de XICs correspondant à 3 masses données (1^{er}, 2^{ème} et 3^{ème} isotope). Ce problème est d'autant plus fréquent que l'échantillon protéique contient un nombre important de protéines et donc de peptides (Figure 83).

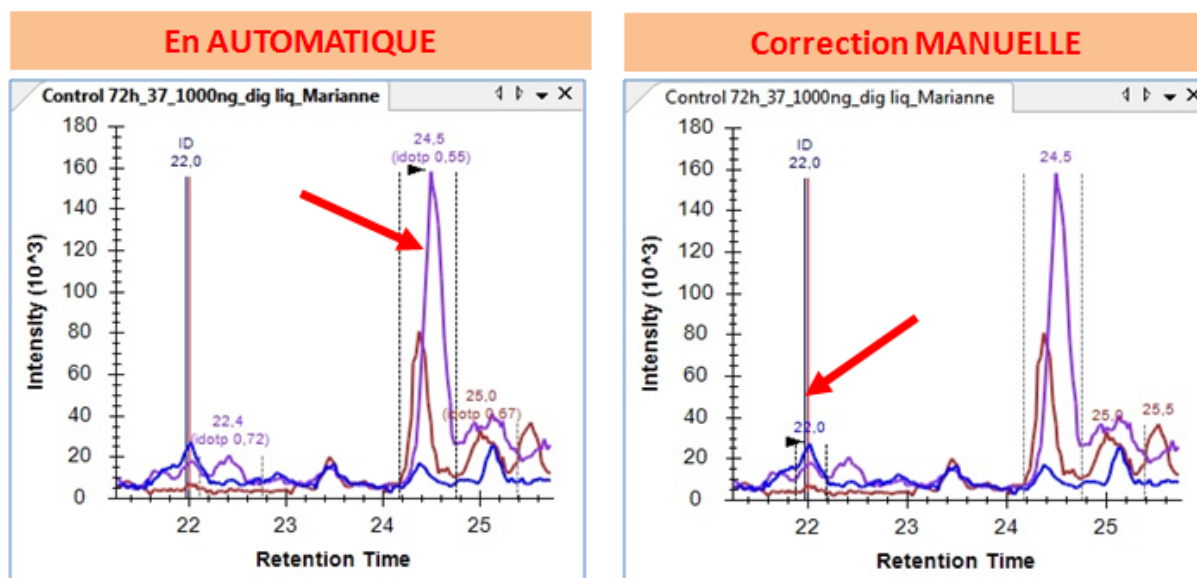


Figure 83 : Modification manuelle du choix du « peak group ». Le processus de « peak picking » automatique a choisi d'intégrer un peak group à 24,5min montrant une distribution isotopique très éloignée de la contribution théorique mais très intense. Manuellement, la correction apportée tient compte de l'observation d'un spectre MS/MS de ce peptide à 22,0min et d'une meilleure valeur attribuée au paramètre idotp.

- Aucun retraitement statistique n'était « à l'époque » intégré dans la suite logicielle. Un tel retraitement était toutefois possible à l'aide d'un export des valeurs d'aires dans un fichier de type Excel, suivi de la conversion de ce fichier tableur en fichier texte (MS-DOS) et import de celui-ci dans le logiciel MarkerView décrit précédemment (suite logicielle AB Sciex).

5.3) Progenesis (Nonlinear Dynamics)

Toutes nos données de cinétique de temps ont été analysées dans un troisième temps sur Progenesis. Les avantages suivants se sont révélés cruciaux pour nous orienter vers le choix de ce logiciel pour la suite de nos analyses :

- L'avantage majeur de Progenesis est qu'il est le seul logiciel parmi ces 3 permettant de réaliser un alignement des TIC (Total Ion Chromatogram). Avec le développement des spectromètres de masse haute résolution, la quantification sans marquage est devenue une méthode de plus en plus répandue et reconnue puisque la précision de masse permet à la fois d'améliorer l'alignement des peptides entre différents échantillons mais aussi d'augmenter

la confiance dans leur identification. Toutefois, la reproductibilité de la séparation chromatographique est critique pour assurer une quantification précise à partir de leur courant d'ions respectif. Une étape d'alignement permet donc grandement de réduire le risque d'erreurs de sélection et d'intégration du « bon » XIC, surtout si d'autres signaux interfèrent avec le XIC du précurseur qui nous intéresse. Un alignement des TIC permet également de réduire la fenêtre de temps de rétention qui sera utilisée pour générer les XICs : plus la fenêtre est restreinte, plus la spécificité augmente car les coordonnées (m/z et RT) du précurseur seront précises. Cet avantage n'est cependant pas spécifique à Progenesis puisque d'autres algorithmes intègrent cette étape (Jeffries 2005). Par contre, il est le seul parmi les 3 logiciels qui étaient en notre possession (PeakView, Skyline, Progenesis) à pouvoir réaliser une étape d'alignement.

- Un test statistique ANOVA est possible à réaliser et ceci de manière simple et conviviale : ce test apparaît plus adapté à notre design expérimental contenant 3 conditions d'exposition ainsi que des réplicats biologiques et une cinétique de 4 temps. En effet, une analyse ANOVA permet de définir si les variances de plusieurs conditions sont égales ou non : ce test statistique peut donc être perçu comme la généralisation d'un t-test de Student à plus de 2 groupes. Si plusieurs t-tests étaient effectués à chaque fois entre 2 conditions, nous pourrions arriver à un résultat similaire mais le risque de générer des erreurs de type I augmenterait (erreur dite également « de première espèce » commise quand l'hypothèse nulle est rejetée, alors qu'elle est vraie).






Malgré ces 2 avantages très importants, nous avons pu relever les inconvénients suivants quant à l'utilisation du logiciel Progenesis :

- Tout comme PeakView, un seul XIC par peptide peut-être visualisé. Cela rend donc fastidieux le fait d'avoir une vue rapide de l'ensemble des conditions à comparer.
- Les fichiers de résultats sont exportés sous forme d'un « .html » : ce sont de gros fichiers, fastidieux à manipuler et qui nécessitent l'utilisation d'un ordinateur disposant d'une forte capacité de mémoire.
- L'inconvénient majeur de Progenesis réside dans le fait qu'il ne peut pas traiter plusieurs conditions : dans notre cas, nous avons pu comparer les conditions Contrôle / BaP / BeP mais chaque condition de temps a dû ensuite être traitée à part (4 fichiers .html au final). Cet inconvénient a été tout de même contourné en concaténant toutes les valeurs d'aires entre elles dans un seul et même tableur, pour une protéine donnée (12 valeurs d'aires, chacune étant la moyenne des 3 aires obtenues pour chaque réplicat). De plus, nous avons utilisé ces

valeurs générées par Progenesis pour réaliser une analyse PCA globale (3 conditions et 4 temps de cinétique) à l'aide de MarkerView : en effet, celui-ci est utilisable à partir de fichier de type .txt sans spécificité aucune de fournisseur.

Le tableau 13 et la figure 84 résument les caractéristiques de chaque logiciel.

Tableau 13 : Caractéristiques des 3 suites logicielles testées sur les mêmes données

MarkerView 1.2.1 (AB Sciex)	Skyline 1.4.0 (MacCoss Lab)	Progenesis LC-MS 4.1 (Nonlinear Dynamics)
Pas d'Alignement	Pas d'Alignement	Alignement 
Pas de réintégration manuelle des XICs	Réintégration manuelle des XICs 	Pas de réintégration manuelle des XICs
Un seul profil d'expression	Histogramme 	Profils d'expression compilés
PCA / Student's T-test 	Pas de tests statistiques	PCA / Anova test 

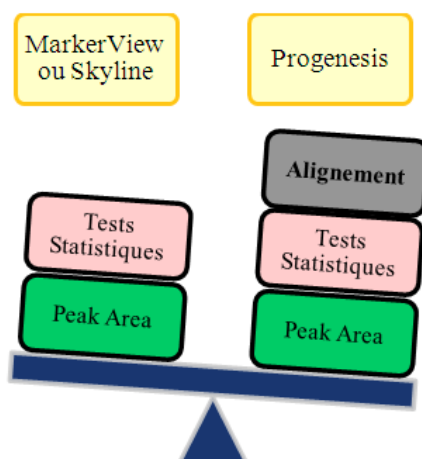


Figure 84: Progenesis vs. MarkerView et Skyline

Références bibliographiques

B

Babaei, F., R. Ramalingam, A. Tavendale, Y. Liang, L. S. K. Yan, P. Ajuh, S. H. Cheng and Y. W. Lam (2013). "Novel Blood Collection Method Allows Plasma Proteome Analysis from Single Zebrafish." Journal of proteome research **12**(4): 1580-1590.

Bereman, M. S., B. MacLean, D. M. Tomazela, D. C. Liebler and M. J. MacCoss (2012). "The development of selected reaction monitoring methods for targeted proteomics via empirical refinement." Proteomics **12**(8): 1134-1141.

Bostanci, N., P. Ramberg, Å. Wahlander, J. Grossman, D. Jönsson, V. M. Barnes and P. N. Papapanou (2013). "Label-Free Quantitative Proteomics Reveals Differentially Regulated Proteins in Experimental Gingivitis." Journal of proteome research **12**(2): 657-678.

C

Chen, L., L. Zhou, E. C. Chan, J. Neo and R. W. Beuerman (2011). "Characterization of the human tear metabolome by LC-MS/MS." Journal of proteome research **10**(10): 4876-4882.

Cilia, M., M. Bereman, T. Fish, M. J. MacCoss and S. Gray (2012). "Homopteran Vector Biomarkers for Efficient Circulative Plant Virus Transmission are Conserved in Multiple Aphid Species and the Whitefly *Bemisia tabaci*." Journal of Integrative Agriculture **11**(2): 249-262.

D

Dagley, L. F., A. Emili and A. W. Purcell (2013). "Application of quantitative proteomics technologies to the biomarker discovery pipeline for multiple sclerosis." PROTEOMICS-Clinical Applications **7**(1-2): 91-108.

G

Griffin, N. M., J. Yu, F. Long, P. Oh, S. Shore, Y. Li, J. A. Koziol and J. E. Schnitzer (2009). "Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis." Nature biotechnology **28**(1): 83-89.

H

Held, J. M., B. Schilling, A. K. D'Souza, T. Srinivasan, D. J. Behring, C. C. Benz, B. W. Gibson and B. W. Gibson "Label-free quantitation and mapping of the ErbB2 tumor receptor by multi-protease digestion with data-dependent (MS1) and data-independent (MS2) acquisitions."

Herrera López, S., M. Hernando, M. Gómez, J. Santiago-Morales, R. Rosal and A. Fernández-Alba (2013). "Investigation of Galaxolide degradation products generated under oxidative and irradiation processes by liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry and comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry." Rapid Communications in Mass Spectrometry **27**(11): 1237-1250.

J

Jeffries, N. (2005). "Algorithms for alignment of mass spectrometry proteomic data." Bioinformatics **21**(14): 3066-3073.

M

MacLean, B., D. M. Tomazela, N. Shulman, M. Chambers, G. L. Finney, B. Frewen, R. Kern, D. L. Tabb, D. C. Liebler and M. J. MacCoss (2010). "Skyline: an open source document editor for creating and analyzing targeted proteomics experiments." Bioinformatics **26**(7): 966-968.

R

Roudier, M., L. True, M. MacCoss and M. Bereman (2013). "Determination of the Intra-individual Variability of Protein Expression in Benign/Cancerous Prostate Tissue via LC-MS—Efforts Towards Molecular Classification." Journal of Biomolecular Techniques: JBT **24**(Suppl): S61.

S

Schilling, B., M. J. Rardin, B. X. MacLean, A. M. Zawadzka, B. E. Frewen, M. P. Cusack, D. J. Sorensen, M. S. Bereman, E. Jing and C. C. Wu (2012). "Platform-independent and Label-free Quantitation of Proteomic Data Using MS1 Extracted Ion Chromatograms in Skyline APPLICATION TO PROTEIN ACETYLATION AND PHOSPHORYLATION." Molecular & Cellular Proteomics **11**(5): 202-214.

Spickett, C. M., A. Reis and A. R. Pitt (2013). "Use of Narrow Mass-Window, High-Resolution Extracted Product Ion Chromatograms for the Sensitive and Selective Identification of Protein Modifications." Analytical chemistry **85**(9): 4621-4627.

Z

Zhu, Z.-J., A. W. Schultz, J. Wang, C. H. Johnson, S. M. Yannone, G. J. Patti and G. Siuzdak (2013). "Liquid chromatography quadrupole time-of-flight mass spectrometry characterization of metabolites guided by the METLIN database." Nature protocols **8**(3): 451-460.

Chapitre 4 : La spectrométrie de masse, un outil pour la recherche d'adduits BPDE à l'ADN et aux protéines

Contexte biologique: en raison de la cancérogénicité avérée de certains HAP, il apparaît primordial de mesurer l'exposition des individus à ces composés afin d'identifier les groupes à risque bien avant la survenue d'une pathologie. L'articulation de différents domaines de compétences (toxicologie, chimie, etc.) permet de mieux caractériser les effets de l'exposition aux polluants depuis leurs sources d'émissions jusqu'à leurs points d'impacts. La mesure de dose externe n'est pas représentative de la dose de polluant qui a réellement pénétré dans l'individu ni de la dose potentiellement à risque quant à la survenue d'un cancer. Il est donc important de développer des outils de surveillance biologique de l'exposition afin de pallier ces manques.

Les HAP sont des molécules biologiquement actives qui, une fois accumulées dans les tissus organiques vont donner lieu à des réactions de transformation. Les métabolites ainsi formés peuvent avoir un effet toxique plus ou moins marqué en se liant à des molécules biologiques fondamentales du corps humain et en provoquant des dysfonctionnements cellulaires. Dans le cadre de la présente thèse, nous nous sommes intéressés à l'étude du métabolite réactif du BaP, le BPDE (BaP-7,8-diol-9,10-époxyde) qui exerce des effets mutagènes et cancérigènes en interagissant avec l'ADN pour former des mutations, conduisant à l'initiation de la cancérogenèse. Par ailleurs, la mesure des adduits à l'ADN est d'un intérêt majeur puisqu'elle permet de quantifier la dose génotoxique. Pour cela, nous nous intéressons au développement d'une méthode par spectrométrie de masse afin de valider ces biomarqueurs d'exposition de dose efficace chez l'homme.

1) Collection du plasma des patients

Avant d'analyser les adduits BPDE à l'ADN ou aux protéines, il était primordial de prélever du sang de personnes ayant été confrontées à une exposition potentielle au BaP.

Des échantillons de sang ont été fournis par des volontaires sains, fumeurs et non fumeurs, dans le but de rechercher les biomarqueurs (adduits BPDE à l'ADN et aux protéines) que nous allons mettre au point au sein du plasma.

La séparation du sang total en ces différents dérivés permet une gestion plus raisonnée pour la recherche de ces biomarqueurs. C'est pourquoi, il apparaît plus intéressant de disposer de dérivés sanguins que du sang total uniquement. En effet, les produits sanguins permettent un traitement plus ciblé et plus avantageux. De plus, la durée de conservation des dérivés sanguins est supérieure à celle du sang total, certains comme le plasma peuvent être congelés et conservés plusieurs mois.

La centrifugation du sang est la technique principale permettant de séparer les différents composants du sang. Elle permet la séparation différentielle des éléments en fonction de leurs densités respectives présentées dans le tableau ci-dessous (Hohenhaus and Rentko 2006).

Tableau 14 : Densité des éléments du sang

Composant	Densité
Sang total	1,053
Plasma	1,023
Plaquettes	1,058
Erythrocytes	1,100

Ainsi, il apparaît logiquement qu'après centrifugation d'un échantillon de sang total, le fond de l'échantillon comprend le composant le plus dense, donc les érythrocytes, recouvert par la couche leuco-plaquettaire. Le surnageant constitue le plasma.

Le plasma est une solution comprenant tous les facteurs de la coagulation ainsi que les différentes protéines sanguines circulantes. Il correspond à la fraction surnageante issue de la centrifugation à froid (entre 1 et 6°C) du sang total (Lucas, Lentz et al. 2004, Hohenhaus and Rentko 2006). La procédure de séparation puis de mise à température de stockage doit être effectuée dans les 6 heures suivant le prélèvement pour que le plasma obtenu ait la qualification de plasma frais. Si le procédé de centrifugation est effectué dans un délai supérieur à 6h après le prélèvement, le plasma est déplété en facteurs de coagulation labiles (facteurs V et VIII) et perd donc sa qualification de plasma frais pour devenir plasma (Stokol and Parry 1998, Murphy and Pamphilon 2009). Celui-ci ne doit pas présenter de coloration rosée, signe d'hémolyse, et être de couleur jaune, relativement transparente (Lucas, Lentz et al. 2004, Murphy and Pamphilon 2009).

Nous avons collecté le sang total de 70 individus (fumeurs et non fumeurs) dans des tubes traités à l'EDTA (acide éthylène diamine tétra-acétique). Les échantillons transportés à

température ambiante ont été traités pendant 2h suite au prélèvement. Après centrifugation (10min à 3000 rpm), le plasma de chaque patient a été collecté et stocké à -80°C.

2) Recherche de BPDE, sous sa forme libre ou liée, dans le plasma des fumeurs

Dans cette partie, nous avons envisagé de détecter le BPDE, sous sa forme libre ($m/z=302,09$) ou liée (adduit à l'ADN ou protéines), dans le plasma de sujets fumeurs par spectrométrie de masse MALDI-TOF.

Le terme adduit à l'ADN désigne le produit résultant de l'addition d'un xénobiotique sur un nucléotide. Dans le cas des HAP, les métabolites électrophiles très réactifs formés au cours de la biotransformation vont interagir par liaison covalente avec les centres nucléophiles des macromolécules cellulaires, telles que l'ADN ou les protéines. Les bases azotées de l'ADN sont considérées comme des cibles privilégiées.

Nos optimisations se sont focalisées sur différents points : (i) la dilution du plasma (2, 5, 10 et 20 fois) dans du tampon Hepes, (ii) la réalisation du dépôt MALDI à l'aide de plusieurs matrices (HCCA, DHB, 3-HPA, dithranol, THAP), (iii) le type de dépôt (goutte séchée, couche mince, sandwich), avec ou sans lavage de dépôt et (iv) le fonctionnement du spectromètre de masse MALDI-TOF en différents modes (réflectron ou linéaire, positif ou négatif).

Les matrices ont tout d'abord été déposées sous forme pure (sans mélange à un échantillon biologique) pour une analyse par spectrométrie de masse MALDI-TOF. Ces matrices génèrent des pics intenses dans la région des faibles masses, englobant de ce fait la zone m/z [300-305] des spectres (Figure 85) : cette zone correspond à la détection théorique de la forme libre du BPDE. De ce fait, les pics de matrice pourront engendrer une diminution ou suppression du signal du BPDE.

En passant les plasmas des fumeurs avec la DHB ou HCCA sur le spectromètre MALDI-TOF, nous avons remarqué que le BPDE, sous sa forme libre, n'est pas détecté, et ceci est dû à la génération de pics intenses des 2 matrices autour de la masse du BPDE.

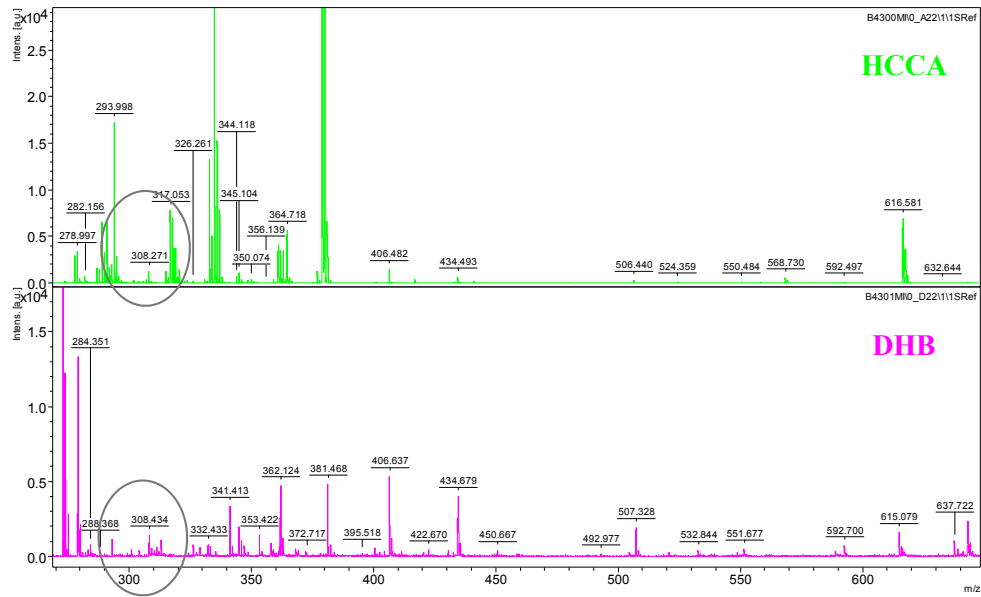


Figure 85 : spectres des matrices seules HCCA et DHB

Nous nous sommes ensuite intéressés à détecter les adduits BPDE liés à l'albumine, protéine majoritaire du plasma. Les résultats des spectres du plasma d'un fumeur sont représentés dans la figure 86 (matrices HCCA et SA, dépôt : goutte séchée (GS), mode linéaire positif).

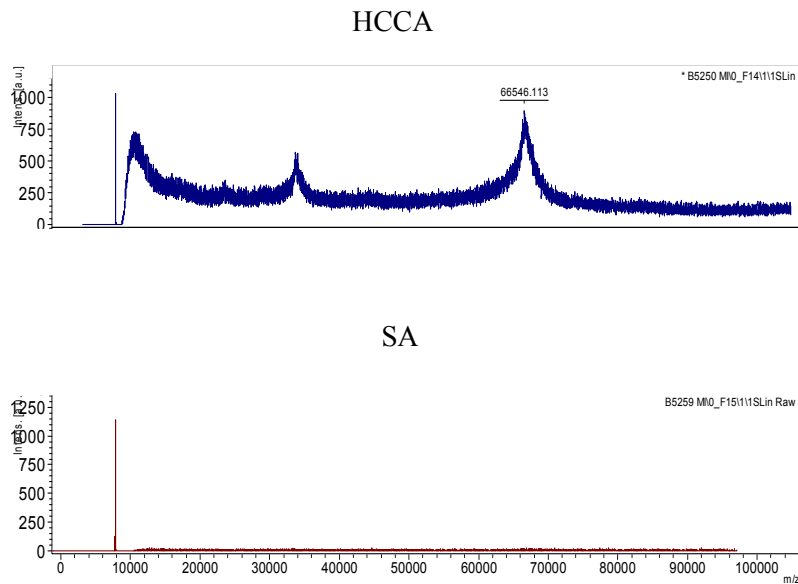


Figure 86 : Comparaison des spectres du plasma d'un fumeur mélangé avec l'une des deux matrices suivantes : HCCA ou SA

En se basant sur les spectres obtenus, l'utilisation de la matrice HCCA est préférable à celle SA. Le shift de masse ne correspondant pas à celui du BPDE, nous avons conclu que le BPDE n'est pas lié à l'albumine.

Dans un deuxième temps, nous avons envisagé la détection du BPDE-glutathion ($m/z = 609.41$). En effet, le glutathion est un tripeptide, γ -L-Glutamyl-L-cystéinyglycine, qui intervient dans la réaction de détoxification et d'élimination du BPDE. Les spectres du plasma de deux patients fumeurs sont comparés à ceux de la matrice déposée pure et du plasma d'un patient non-fumeur : le pic du BPDE lié au glutathion n'apparaît pas (Figure 87), mais un autre pic est détecté à $m/z=616.25$ seulement chez les fumeurs (Figure 88). Ce rapport m/z n'étant pas une masse théorique recherchée au début de notre travail, il nous fallait essayer de chercher si elle était éventuellement tout de même reliée à un adduit BPDE.

Nous avons ainsi recherché les adduits BPDE-ADN (BPDE-désoxyadénosine, BPDE-désoxyguanosine). Les désoxyribonucléosides puriques (désoxyadénosine et désoxyguanosine) étant plus réactifs que les désoxyribonucléosides pyrimidiques (désoxycytidine et désoxythymidine), ils ont un potentiel de formation d'adduits plus important). Malheureusement, aucune détection expérimentale de ces adduits n'a pu être réalisée malgré de nombreux efforts d'optimisation du protocole de dépôt et analyse par spectrométrie de masse MALDI-TOF.

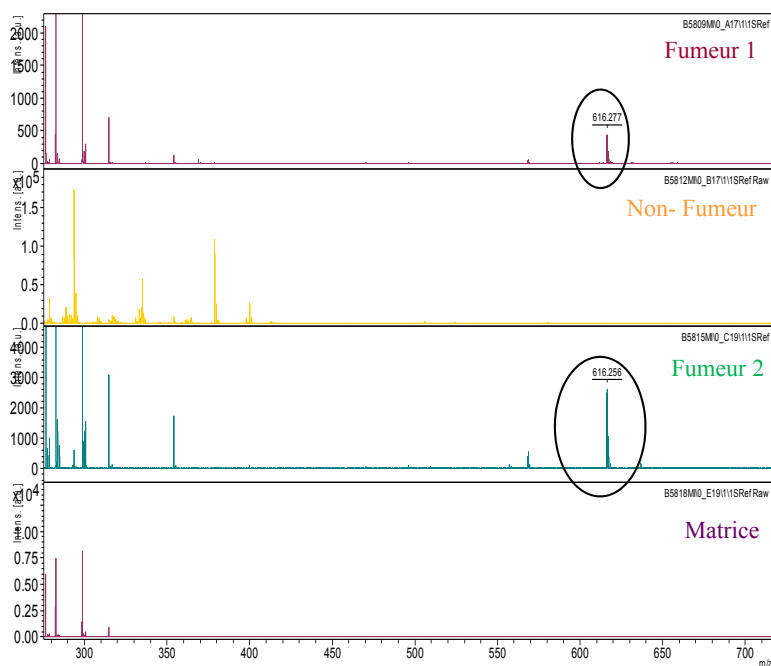


Figure 87 : Comparaison des spectres du plasma de deux fumeurs à ceux de la matrice seule et du plasma d'un non-fumeur

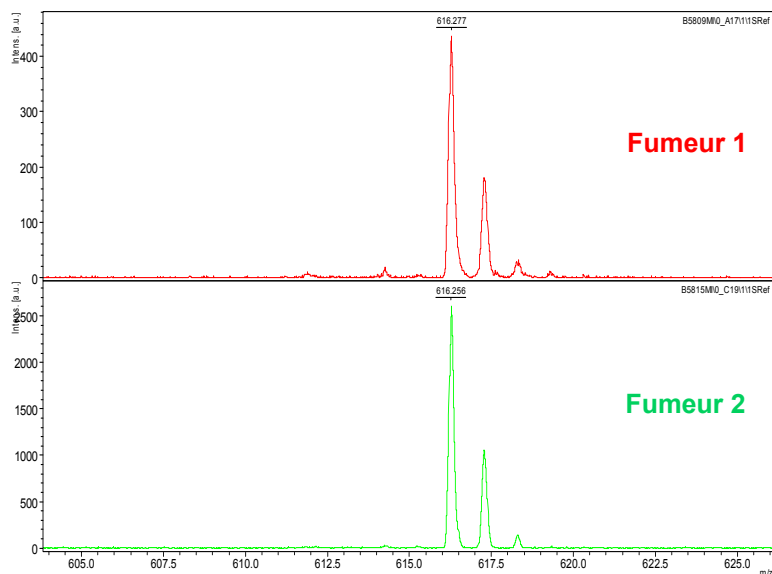


Figure 88 : Pics apparus seulement chez les fumeurs

3) Perspectives

Dans cette étude, nous nous sommes intéressés à rechercher les adduits BPDE-ADN et BPDE-protéines par spectrométrie de masse MALDI-TOF. Nous envisageons de poursuivre cette partie en solubilisant le produit chimique BPDE pour l'analyser sous forme pure par spectrométrie de masse MALDI-TOF : le but de cette expérience serait de définir la masse exacte du composé avec un type de dépôt particulier et ainsi de vérifier qu'elle correspond bien à la masse théorique du BPDE. Dans un deuxième temps, nous souhaiterions mettre au point une méthode en spectrométrie de masse pour détecter le BPDE, sous sa forme libre hydrolysée (Benzo(a)pyrene-tetrahydrotetrol) et sous sa forme liée à des protéines ou à l'ADN dans les cellules HepG2 exposée au BaP, puis le valider dans le plasma de fumeurs. D'autre part, nous allons chercher à identifier la masse à $m/z = 616$ qui n'apparaît que chez les fumeurs en tentant des approches par MS/MS.

Références bibliographiques

H

Hohenhaus, A. and V. Rentko (2006). "Blood transfusions and blood substitutes." Fluid, Electrolyte, and Acid-Base Disorders in Small Animal Practice. 3rd edn. St. Louis: Saunders Elsevier: 567-583.

L

Lucas, R., K. Lentz and A. Hale (2004). "Collection and preparation of blood products." Clinical techniques in small animal practice **19**(2): 55-62.

M

Murphy, M. F. and D. H. Pamphilon (2009). Practical transfusion medicine, John Wiley & Sons.

S

Stokol, T. and B. W. Parry (1998). "Efficacy of Fresh-Frozen Plasma and Cryoprecipitate in Dogs with von Willebrand's Disease or Hemophilia A." Journal of Veterinary Internal Medicine **12**(2): 84-92.

**Conclusion générale
et
perspectives**

Conclusion générale et perspectives

Nous avons développé au cours de ce travail une stratégie protéomique optimisée pour identifier les protéines qui peuvent constituer des candidats biomarqueurs potentiels des pathologies liées à l'exposition au BaP.

Nos optimisations en culture cellulaire nous ont permis de déterminer la dose d'exposition des cellules hépatiques humaines HepG2, à savoir, 10 μ M de BaP pendant différents temps afin de trouver des candidats précoces et tardifs liés à cette exposition. Le BeP a été choisi comme contrôle négatif : c'est un isomère du BaP qui n'est pas classifié comme étant cancérigène pour l'homme. A nos connaissances, aucune étude d'analyse protéomique décrite dans la littérature n'utilise l'exposition au BeP comme expérience de contrôle, ce qui est pourtant la seule manière d'obtenir des biomarqueurs spécifiques à l'exposition au BaP cancérigène. Les biomarqueurs répertoriés dans la littérature sont donc surtout des biomarqueurs d'exposition aux HAP avant d'être des biomarqueurs d'exposition au BaP.

De plus, notre étude est la seule étude consacrée à étudier le sécrétome des cellules hépatiques humaines, alors que l'essentiel des études réalisées sur l'exposition au BaP, n'ont été faites que sur le protéome global. Travailler sur le sécrétome, le réservoir des biomarqueurs recherchés dans les fluides biologiques, et non sur le protéome global, permet de réduire les problèmes liés à la détection des protéines candidates qui sont souvent peu abondantes, grâce notamment à leur enrichissement dans l'échantillon.

Le choix d'une technique de quantification par spectrométrie de masse telle que la méthode label-free, beaucoup plus sensible que celle par gel 2D classiquement utilisée dans les études protéomiques de recherche de biomarqueurs suite à l'exposition au BaP, permet de détecter les protéines très peu abondantes, impossibles à observer par gel-2D.

Trois suites logicielles pour la quantification par label-free MS1 ont été testées : ProteinPilot/PeakView/MarkerView (Sciex), Skyline (MacCoss Laboratory) et Progenesis LC-MS (Nonlinear dynamics). Progenesis permettant d'aligner les TIC a été sélectionné pour l'analyse de nos résultats.

Au total, 897 protéines sont identifiées avec, au minimum, un peptide unique (ion scores ≥ 25) dans le sécrétome des cellules exposées, ou non, au BaP et au BeP pendant 2h et 72h. 847 protéines ont été validées avec un FDR $< 1\%$ et dont 449 ont une p-value $\leq 0,05$ dans 2h et/ou 72h. La PCA a montré la reproductibilité entre les 3 réplicats de chaque condition et la séparation entre 2 groupes: BaP et contrôles. Ces résultats montrent que le traitement par le BaP mène à un changement significatif de l'expression protéique. 207 protéines sont surexprimées avec un ratio BaP/control > 2 dont 55 protéines avec un ratio > 5 ont été sélectionnées pour la suite de l'étude. Nous avons décidé de sélectionner les protéines surexprimées dans les cellules exposées au BaP parce que la surexpression d'une protéine est moins sujette aux erreurs que sa sous-expression puisque sa quantité pourra être en-dessous du seuil de sensibilité de l'instrument.

Le pourcentage de protéines sécrétées parmi les 847 protéines est de 51,95%. Il a été calculé grâce aux algorithmes de prédiction des protéines sécrétées (SecretomeP/SignalP) et l'analyse Extracellular GO. Parmi les 55 protéines surexprimées sélectionnées, 30 n'ont pas été prédites comme sécrétées par SecretomeP/SignalP. 14 ont été trouvées comme liées au cancer d'après la littérature et la classification de la fonction moléculaire de GO.

Nos résultats sont préliminaires, une validation ultérieure de ces protéines dans le plasma de la population exposée au BaP permettra d'améliorer considérablement la détection précoce, le pronostic, la prédiction de la réponse au traitement et la prévention.

Les échantillons de plasma des fumeurs qui étaient destinés à l'étude du BPDE sous sa forme libre et liée à l'ADN et aux protéines ont été difficilement exploitables. Ces mêmes échantillons seront utilisés pour la validation des 14 protéines candidates identifiées dans le sécrétome des cellules exposées au BaP. La validation de leur expression est privilégiée par les méthodes SRM (Selected Reaction Monitoring) ou SRM-like (MS/MS^{all} : SWATH) qui dépendent de plusieurs paramètres (temps de rétention, rapport m/z de l'ion précurseur et des ions fragments du peptide ciblé, intensités des signaux de fragments détectés) contrairement aux approches par Western Blot qui ne dépendent que de la spécificité de l'anticorps utilisé, s'il est commercialisé. Comme Aebersold l'a mentionné : « il est temps d'utiliser la SRM, développée pour pratiquement n'importe quelle protéine, afin de valider un sous-ensemble de protéines intéressantes mises en évidence par des études quantitatives en protéomique » (Aebersold, Burlingame et al. 2013). Par ailleurs, et comme Baker l'a justement cité: «Le succès

ultime des biomarqueurs ne peut être réalisé que lorsque leur découverte permet la compréhension de leur pertinence physiologique » (Baker 2005).

Références bibliographiques

A

Aebersold, R., A. L. Burlingame and R. A. Bradshaw (2013). "Western Blots versus Selected Reaction Monitoring Assays: Time to Turn the Tables?" Molecular & Cellular Proteomics **12**(9): 2381-2382.

B

Baker, M. (2005). "In biomarkers we trust?" Nature biotechnology **23**(3): 297-304.

Annexes

Annexe A

Mise en œuvre expérimentale

1) Culture cellulaire

Cette partie décrit les techniques de base utilisées au sein du Laboratoire d'Immunologie de la Faculté de Santé Publique II (Université Libanaise) pour le traitement des cultures cellulaires. Les manipulations effectuées sur la lignée cellulaire hépatique HepG2 que nous utilisons pour nos cultures sont réparties en étapes distinctes, mais réalisées successivement et de manière standardisée.

La lignée cellulaire HepG2 est dérivée d'un carcinome hépatocellulaire humain. Les cellules sont cultivées par plusieurs changements de milieu de culture et passages successifs, et on multiplie le nombre de flacons de culture jusqu'à avoir un nombre de cellules adéquat (à confluence) pour en extraire la quantité de protéines suffisante pour les analyses en protéomique. Les cellules sont incubées à 37°C (incubateur ThermoForma), à un taux de dioxyde de carbone (CO₂) de 5% et en milieu spécialisé afin d'assurer leur croissance. Les cellules sont manipulées sous poste de sécurité microbiologique (TechGen International) dont la surface est régulièrement nettoyée et décontaminée à l'aide de rayons Ultra-violets.

Trois milieux de culture ont pu être testés : i) RPMI 1640 + SVF, ii) RPMI1640 + B27 et iii) DMEM/F12.

Le sérum de veau fœtal (SVF) (Sigma Aldrich) est le sérum le plus largement utilisé en culture cellulaire. La Bovine Serum Albumin (BSA) est le composant majeur du SVF. L'importante variété de protéines dans le SVF maintient les cellules en culture dans le milieu dans lequel elles peuvent survivre, grandir et se multiplier.

Le supplément B27 (Gibco) est optimisé pour la survie et la croissance des cellules HepG2 dans un environnement exempt de sérum afin de réduire la présence de la BSA dans les extraits protéiques et les sécrétomes.

Le milieu de culture utilisé par défaut est le RPMI 1640 supplémenté de 10% de SVF (Gibco), 1% de L-glutamine (200 mM, Gibco) et 1% d'un mélange P/S : Pénicilline (10000 U/mL) et Streptomycine (10000 µg/mL) (Gibco). L'addition d'antibiotiques permet de prévenir les contaminations d'origine fongique et bactériologique. L'action du mélange P/S utilisé dans notre culture couvre un large spectre d'agents pathogènes.

Les suspensions cellulaires ont été établies dans des flacons stériles de culture en polystyrène (T75) dont la surface de croissance cellulaire est de 75cm².

Exposition au Benzo(a)Pyrène et Benzo(e)Pyrène :

Nous avons étudié l'effet de l'exposition des cellules HepG2 aux BaP et BeP (Sigma Aldrich) à différentes concentrations (0 ; 0,125 ; 0,25 ; 0,5 ; 1 ; 10 et 50µM). Pour ceci, une solution-mère de chaque composé a été préalablement préparée dans du DMSO (Dimethyl sulfoxide). Les solutions-filles sont ensuite obtenues par dilution de la solution-mère dans le milieu de culture RPMI1640 jusqu'à l'obtention de la concentration souhaitée. La concentration du DMSO dans ce milieu ne doit pas excéder 0,5% (v/v).

Afin de mesurer la toxicité de ces deux composés à différentes concentrations, nous avons utilisé une méthode de numération des cellules vivantes par la mesure de l'activité de la déshydrogénase mitochondriale. Le principe de ce test est basé sur la réduction du sel de tétrazolium MTT (bromure de 3-(4,5-diméthylthiazol-2-yl)-2,5-diphényl tétrazolium) (Sigma Aldrich) en formazan, par la succinate déshydrogénase mitochondriale des cellules vivantes. Ceci forme un précipité de couleur violette dans la mitochondrie. La quantité de précipité formée est proportionnelle au nombre de cellules vivantes. Il suffit donc, après incubation des cellules avec le MTT pendant 3h à 37°C, de dissoudre les précipités violets (de formazan) dans du DMSO. La mesure de la densité optique par spectroscopie permet de connaître la quantité relative de cellules vivantes et actives métaboliquement.

Des cellules HepG2 ont étéensemencées dans des microplaques à 96 puits de manière à avoir le jour de la lecture de chaque microplaque 4.10⁴ cellules/puits, le temps de dédoublement des cellules étant de 24h. Elles sont incubées 24h avant leur exposition aux différentes concentrations du BaP ou BeP. Suite à l'exposition au BaP ou au BeP et à une incubation pendant 2, 6, 24 et 72h, le MTT a été ajouté à la microplaque et incubé pour 3h, et la lecture s'est faite à 492nm par un lecteur d'ELISA (MultiskanFC, Thermo scientific).

Après avoir choisi la concentration d'exposition des cellules au BaP et après incubation de ce composé pendant différents temps, les sécrétomes sont prélevés et stockés à -80°C, les cellules sont décollées et le culot cellulaire est lavé 3 fois avec le tampon PBS puis figé dans l'azote liquide. A cette étape, le nombre cellulaire est aux alentours de 10.10⁶ cellules. Tous ces échantillons préparés au Liban ont été transportés dans de la carboglace à Strasbourg afin de procéder à l'analyse protéomique.

2) Extraction au trizol

L'aliquot de $10 \cdot 10^6$ de cellules récupérées, exposées ou non aux agents toxiques, a été préparé par trois lavages au tampon PBS puis congelé sous forme de culot sec dans l'azote liquide.

Les cellules ont été décongelées sur la glace et solubilisées dans un volume de $400 \mu\text{L}$ de tampon de lyse contenant de l'Urée (7M) et Thiourée (2M) (agents chaotropes permettant la dénaturation des protéines afin d'empêcher leur agrégation), du CHAPS 4% (détergent zwitterionique) et du Tris 20mM. Le lysat cellulaire a été purifié par le kit trizol "tri reagent" (Sigma, ref T3934). Celui-ci contient 2 étapes préliminaires d'extraction séquentielle d'ARN et d'ADN, suivies par une étape finale de précipitation de protéines (référence protocole T3934). Le culot obtenu est solubilisé dans du tampon 2D (Urée 7M, Thiourée 2M, Chaps 4%, Tris 20 mM pH8,5).

Les protéines extraites sont ensuite dosées par une méthode d'analyse spectroscopique (méthode de Bradford, Bio-Rad). C'est un dosage colorimétrique, basé sur le changement d'absorbance et se manifestant par le changement de la couleur du bleu de Coomassie après liaison aux acides aminés basiques (arginine, histidine et lysine) et aux résidus hydrophobes des acides aminés présents dans les protéines. Sous sa forme cationique libre (couleur rouge et marron), ce colorant réactif absorbe la lumière à une longueur d'onde de 465-470nm. Lorsqu'il est mis en contact avec une solution protéique à doser, il se lie aux protéines présentes et aux groupements aromatiques (forme anionique liée de couleur bleue) ce qui a pour effet de déplacer sa principale raie d'absorption à 595nm. Le changement d'absorbance est proportionnel à la quantité de colorant lié, indiquant donc la concentration en protéines dans l'échantillon. L'absorbance du rayonnement visible à 595nm est mesurée par un détecteur, et permet de doser indirectement la teneur en protéines de la solution à l'aide d'une courbe d'étalonnage en BSA.

Le dosage des échantillons a été réalisé en microplaque à 96 puits avec utilisation de BSA comme calibrant (P5619, Sigma).

3) Electrophorèse

Tous les réactifs utilisés pour l'électrophorèse et le traitement des gels proviennent de « Euromedex », sauf indication contraire.

3.1) Electrophorèse monodimensionnelle sur gel de polyacrylamide

L'électrophorèse est réalisée sur gel de polyacrylamide en conditions dénaturantes sous l'action du SDS (dodécylsulfate de sodium).

- Gel de séparation (13% acrylamide) : Acrylamide 37,5/Bisacrylamide 1, 10% SDS, Tris-HCl 1,5M pH8,8 et eau milliQ. La polymérisation de l'acrylamide est catalysée par l'ammonium persulfate 20% (w/v) et le Temed (N,N,N,N'-tétraméthyléthylènediamide) ajoutés extemporanément.
- Gel de concentration (5% acrylamide) : préparé comme précédemment sauf avec Tris-HCl 1M pH6,8

Des gels préfabriqués ont été aussi utilisés (Thermo scientific, Tris-Glycine Gels, product number 0025269). Le gradient des ces gels est de 4 à 12% permettant ainsi une séparation optimale des protéines de basse et haute masses.

Les échantillons sont dilués au ½ dans du tampon dénaturant Laemmli 4X (50mM Tris-HCl pH6,8 , 2% SDS, 10% glycérol, 1% β-mercaptoéthanol, 12,5mM EDTA, 0,02 % bleu de bromophénol), et portés à 95°C pendant 5min. Après polymérisation du gel à température ambiante, les mélanges échantillon-tampon Laemmli sont déposés dans les puits en laissant un puits libre entre chaque dépôt. L'équivalent de 10 ou 20µg en protéines totales est déposé dans chaque puits. De même, 1 marqueur de poids moléculaire (PageRuler Plus, Prestained Protein Ladder, Thermo scientific) est déposé dans un des puits de chaque gel.

La migration des protéines en SDS-PAGE est réalisée dans un tampon de migration (Tris 25mM /Glycine 125mM /SDS 0,1%w/v) à température ambiante. La tension appliquée par le générateur pendant les premières 15min est de 70V avant d'être augmentée à 150 voire 200V pendant 1h30 (la migration sera surveillée et le gel est arrêté lorsque le front de migration atteint le bas du gel).

Après l'électrophorèse, le gel est coloré en 2 étapes : fixation des protéines dans une solution 45% méthanol, 1% acide acétique (20 minutes au minimum) puis coloration au bleu colloïdal (34% méthanol, 17% sulfate d'ammonium, 0,5% acide acétique, 0,1% bleu G250) pendant 2h ou par l'Instant Blue (Expedeon) pendant 15-20min. La décoloration est réalisée par plusieurs bains successifs d'eau milliQ chauffée, afin de faire apparaître les bandes protéiques.

3.2) Electrophorèse bidimensionnelle

La première dimension de l'électrophorèse bidimensionnelle consiste en une focalisation isoélectrique (IEF, isoelectric focusing) sur des bandes de gels à gradient de pH immobilisé ou

«IPG strips» (immobilized pH gradient strips) permettant de séparer les protéines selon leur point isoélectrique (pI). Le pI correspond au pH pour lequel la charge totale de la protéine apparaît nulle. Un champ électrique est appliqué au gel d'électrophorèse hydraté par la solution protéique d'intérêt. Les protéines chargées vont migrer dans le gradient de pH du gel jusqu'au pH égal à leur pI. Pour obtenir une meilleure reproductibilité entre focalisations isoélectriques, nous avons utilisé des bandelettes d'IPG commerciales. Dans notre cas, nous avons choisi des strips d'électrophorèse commercialisés par Bio-Rad, de longueur 17cm pour pouvoir charger une plus grande quantité de protéines avec une gamme de pH entre 5 et 8. 200µg de protéines sont séparés ainsi par électrophorèse bidimensionnelle.

Le gradient fixe de pH est réalisé grâce au greffage d'ampholytes directement sur l'acrylamide par liaison covalente. Les gels sont réhydratés avec 300µL d'échantillon, solubilisés dans le tampon, supplémenté par un réducteur, du DTT (20mM), un colorant (bleu de bromophénol 0.5%) et un mélange d'ampholytes (3-10) entraînant les protéines dans leur migration. La première dimension d'électrophorèse est réalisée sur l'appareil IEF (ProteanIIXL, BioRad) en 3 étapes : réhydratation active à 50V pour 24h favorisant l'entrée des protéines dans le gel, puis montage progressif du voltage à 6000V, et focalisation des protéines durant 10h à 6000V, l'ampérage étant limité à 50µA/gel.

Tableau 15: Programmation de l'appareil IEF

Etapes	Voltage à atteindre (V)	Montée	Durée	Température (°C)
1	250	rapide	15min	20
2	6000	progressive	5h	20
3	6000	rapide	60000V/h	20
4	500	rapide	12h et plus	20

Une étape d'équilibration des strips est réalisée avant la seconde dimension SDS PAGE. Il s'agit en fait d'une séquence de dénaturation des protéines dans un bain de SDS 2%, Urée 36%, 25% Tris HCl 1,5M pH8,8 , glycérol 20%, combinée à une réduction-alkylation des cystéines (130mM DTT pendant 15min et 135mM iodoacetamide pendant 20min à l'obscurité).

Les protéines, entourées de détergent SDS, peuvent maintenant être séparées en fonction de la mass moléculaire lors de la deuxième dimension d'électrophorèse.

Les gels 2D sont réalisés par positionnement de la bandelette (strip) au sommet d'un gel SDS-PAGE à 11%, maintenu par ajout d'agarose liquide (0,5% low melting point agarose). Le système de migration utilisé est le ProteanIIXL, BioRad, utilisé à ampérage constant. Les conditions de migration sont, par gel, de 16mA pendant 30min pour la pénétration des protéines

dans le gel SDS-PAGE et 24mA pendant 6h pour leur migration avec une tension limitée à 200 V pour les gels 2D associés à des IPG strips de 17cm.

Les gels ont été colorés au bleu colloïdal. Si la quantité de protéines s'est avérée trop faible pour être détectée au bleu, une coloration à l'argent a donc été envisageable dans un second temps. Le protocole de la coloration à l'argent consiste à incuber le gel dans 250 mL d'une solution de sensibilisation DTT (0,005g/L) puis dans une solution de coloration à l'argent (0,1% nitrate d'argent) pendant 20min et de le laver 3 fois à l'eau milliQ avant de le tremper dans la solution de révélation (15g/L de carbonate de sodium, 0,03% de formaldéhyde). La révélation est arrêtée par une solution d'acide acétique 10% après 10min, et le gel est de nouveau lavé 3 fois à l'eau milliQ.

Les gels ont été digitalisés sur un scanner (Bio-Rad, modèle "GS-800 Densitometer") et analysés sur le logiciel PD-Quest (8.0.1). L'information quantitative est obtenue par analyse densitométrique comparative du volume des spots protéiques.

4) Préparation des échantillons pour la spectrométrie de masse

4.1) Excision des spots ou bandes

Après séparation sur gels 2D, les tâches correspondantes aux protéines d'intérêt ont été découpées des gels. L'excision des spots d'intérêt s'est faite automatiquement par un robot (Gilson, modèle "Quad-Z 215"). Dans le cas des gels 1D, toute la piste a été découpée manuellement en 24 fractions consécutives (bandes).

Ces morceaux de gel d'acrylamide 1D ou 2D sont conditionnés en microplaque 96 puits.

4.2) Lavages et réduction-alkylation des cystéines

3 cycles de lavages de 15min alternant une solution de bicarbonate d'ammonium (NH_4CO_3) 25 mM et un mélange bicarbonate d'ammonium $\text{NH}_4\text{HCO}_3/\text{ACN}$ (1:1, v/v), puis un lavage de 10min dans l'acétonitrile 100% (ACN) sont réalisés. S'en suit une étape séquentielle de réduction (DTT 10mM, 56°C, 45min) et alkylation des cystéines (IA 55mM à l'obscurité 45min). Un autre cycle de lavage comme décrit précédemment est effectué avec au final l'obtention de morceaux de gel déshydratés.

4.3) Digestion trypsique

La digestion enzymatique ('Trypsin Porcine', V115A Promega) est effectuée pendant la nuit à température ambiante en ajoutant 1.8µl d'une solution à 12ng/µL, soit 21,5ng de trypsine par échantillon. Après 15min d'adsorption complète de la trypsine, 10-15µL de bicarbonate

d'ammonium 25mM sont ajoutés sur chaque bande (ou spot) pour éviter la déshydratation des gels lors de la digestion et favoriser la diffusion passive des peptides hors des mailles du gel.

4.4) Extraction des peptides

L'extraction est réalisée avec 10 μ l d'une solution comprenant 5% d'acide formique et 60% d'ACN. La microplaque est agitée sur un Thermomixer à une vitesse de 600 rpm pendant 30min à température ambiante.

4.5) Conditionnement MALDI ou nanoLC

Pour une analyse MALDI, 1 μ L du surnageant de digestion et 4 μ L de 0.1% TFA (Trifluoroacetic acid) sont déposés sur une cible Anchorchip (Format spots 96 de Bruker) présottée d'une matrice HCCA (acide α -cyano-4-hydroxycinnamique) et d'un calibrant standard (Bradykinin (1-7) $m/z = 757,39916 \rightarrow$ ACTH-clip (7-38) $m/z = 3617,92890$). Les peptides co-cristallisent avec la matrice resolubilisée par le 0,1% TFA.

Grâce au cœur hydrophobe des emplacements de la cible Anchorchip, l'échantillon est plus concentré au centre de chaque dépôt lors de l'analyse. Ce type de plaque est propice à l'identification des protéines.

L' HCCA, matrice utilisée en routine, est préconisée pour les analyses de peptides sur une assez large gamme de masse. Elle permet d'effectuer des mesures de masse sans discriminer les plus petits peptides des plus gros.

La cible est ensuite insérée dans le spectromètre de masse de type MALDI-TOF/TOF (UltraflexII, Bruker Daltonics).

Pour la nanoLC, les surnageants de digestion et d'extraction sont réunis dans des inserts en verre et lyophilisés. Le conditionnement final est obtenu en resolubilisant les peptides dans une solution de 0,1%FA (Formic Acid).

5) Acquisition des spectres de masse MALDI

Les spectres de masse des échantillons ont été obtenus par analyse MALDI-TOF/TOF (Autoflex III Smartbeam, Bruker) à l'aide du logiciel FlexControl (version 3.0) et puis exploités par le logiciel FlexAnalysis (version 3.0). Le logiciel FlexControl permet de parcourir les emplacements de la plaque selon deux modes. Le mode manuel a été utilisé afin d'ajuster au mieux les paramètres d'acquisition. Ces derniers sont décrits dans le tableau ci-dessous.

Tableau 16 : Paramètres d'acquisition des spectres de masse avec le logiciel FlexControl v3.0

Paramètres	Valeurs
Spectromètre	
Fréquence du laser	200Hz
Ion source 1	19,1kV
Ion source 2	16,54kV
Lentille	8,74kV
Réflecteur 1	21,05kV
Réflecteur 2	9,84kV
Suppression des ions de la matrice	700Da
Mode ionisation	Positif (réflecteur activé)
Détection	
Fenêtre de détection	700-4400Da
Gain du détecteur	7,1x
Gain électronique	100mV
Annotation des spectres	
Fenêtre d'annotation	750-4000Da
Algorithme de détection des pics	Snap (Averagine)
Rapport signal sur bruit	3
Soustraction de la ligne de base	TopHat
Seuil de qualité (Quality factor)	30
Lissage	SavitzkyGolay (4cycles, 0,15m/z)
Hauteur du pic	90%
Automatisation	
Puissance de tir initial du laser	22%
Addition des spectres de masse	2000

Le moteur de recherche MASCOT (MatrixScience, UK) permet d'identifier les protéines présentes dans l'échantillon analysé. Pour affiner la recherche dans la base de données, des paramètres de recherche y doivent être spécifiés. Voici les paramètres de recherche les plus influents dans l'identification : base de données (NCBIInr), taxonomie (*Homo sapiens*), enzyme (Trypsine/P), modifications (carbométhylation des cystéines, acétylation de l'extrémité N-terminale protéique, oxydation des méthionines), tolérance sur la masse du peptide (25ppm), nombres de sites de coupure tryptique manqués (2). Les paramètres de tolérance en masse permettent de jouer sur la précision entre les valeurs expérimentales (rapports m/z du spectre MS) et les valeurs théoriques (masses de peptides issues d'une digestion *in silico*) et de l'améliorer. Plus l'erreur est petite, plus la recherche est précise.

Le choix de l'enzyme est important pour la digestion : chaque protéase a une spécificité de clivage différent. La trypsine utilisée effectue une coupure spécifiquement après un résidu Arginine ou Lysine sauf s'il y a une proline. Comme il se peut que la digestion soit incomplète, un paramètre chiffré (nombre de clivage manqué) indiquant la tolérance à ce sujet doit être indiqué. Une valeur trop élevée pourrait diminuer la spécificité de la recherche.

6) Analyses nanoLC-ESI-MS/MS

6.1) MicroToF-QII

Pendant ma première année de thèse, les analyses ont été réalisées sur un spectromètre de masse ESI-QTOF de type MicroTOF-QII (Bruker) couplé à une nanoLC Ultimate 3000 (Dionex) (Figure 89). Pour le chargement des peptides sur la pré-colonne C18, le débit est fixé à 300 nL/mn tandis que la séparation des peptides sur la colonne analytique C18 se fait à un débit de 300 nL/mn (Figure 90). La phase mobile A correspond à 0,1% d'acide formique dans de l'eau et le solvant B correspond à 0,1% d'acide formique dans 80% d'ACN.

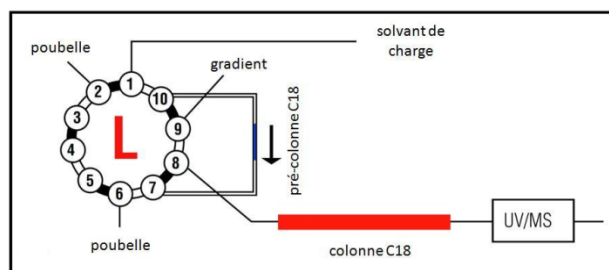


Figure 89 : Schéma du montage de la pré-colonne C18 sur les vannes de la nanoLC

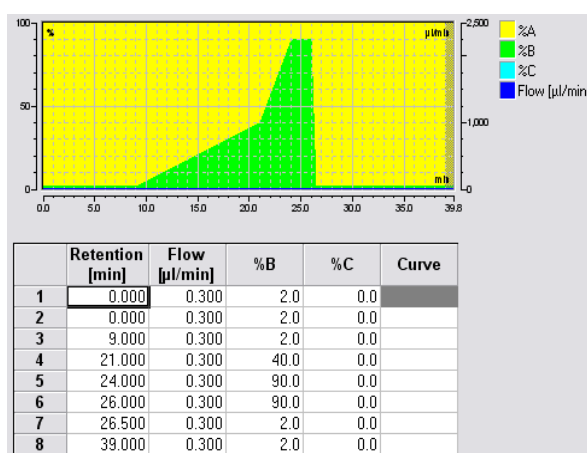


Figure 90 : Montage pour chromatographie en phase C18 : Gradients en solvants au cours du temps

6.2) Triple TOF 5600

6.2.1) Description technique du couplage nanoLC-ESI-MS/MS utilisé: système nano cHiP HPLC (Eksigent)

Comme souligné dans l'introduction de ce manuscrit, la technique séparative utilisée en amont du spectromètre de masse est primordiale pour obtenir une bonne décomplexification de l'échantillon peptidomique. Pour les analyses sur le sécrétome, nous avons utilisé un système Eksigent (AB Sciex) composé d'un module nanoLC-Ultra 2Dplus et un module cHiPLC Nanoflex

La nanoLC-Ultra 2Dplus est un système chromatographique à deux dimensions entièrement automatisées. Le système est composé de 2 pompes Dual Gradient 50-500 μ L/min pour 2 gradients binaires et d'une pompe de chargement de 1-30 μ L/min. Le compartiment colonne avec 2 vannes 10 ports est contrôlé en température. Les échantillons sont injectés par un module autosampler, maintenu à 4°C. L'ensemble est sous contrôle d'un logiciel de pilotage AS-2(Eksigent).

Le système cHiPLC est couplé au système nanoLC-ultra, sur un des 2 gradients. Il utilise une combinaison de Chips microfluidiques pour séparer les peptides issus d'une digestion et il existe trois connecteurs au niveau du système cHiPLC, chacun étant dédié à une cartouche miniaturisée « Chip » spécifique (Figure 91). L'avantage de ce système réside en 2 points : une facilité de connectique de montage associé à une très bonne reproductibilité de fabrication des colonnes, atout de poids pour l'homogénéité des temps de rétention des peptides sur des expériences à comparer entre elles, séparées dans le temps et sur différentes colonnes.

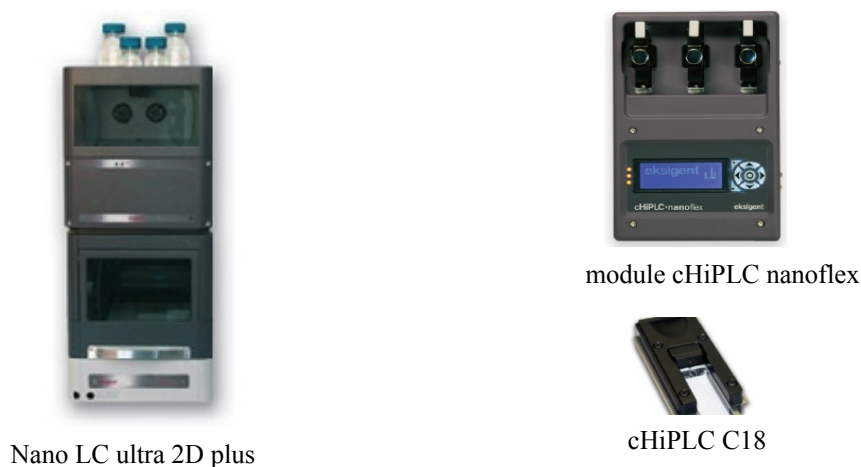


Figure 91 : système nano cHiP HPLC Eksigent

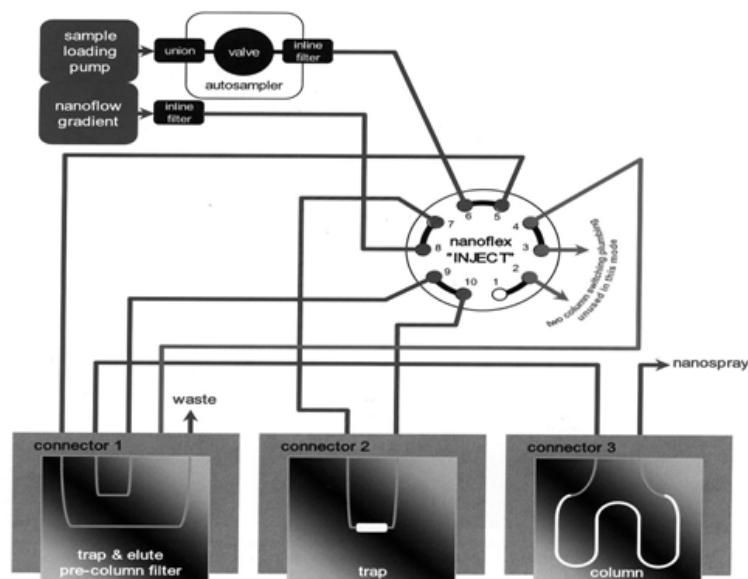


Figure 92 : Configuration Trap elute du système cHiPLC

Un tel système HPLC dispose de plusieurs modes de configuration possible suivant l'application qui sera utilisée : injection directe, plusieurs colonnes analytiques en série, etc. Dans notre cas, nous avons opté pour le mode « Trap Elute », communément utilisé dans le domaine de la protéomique. Une première étape de pré-concentration et dessalage est réalisée sur une pré-colonne (étape Trap) : pour notre étude, nous avons utilisé un modèle miniaturisé cHiP, avec une phase stationnaire Chrom XP C18 (billes de $3\mu\text{m}$, pores de 120\AA , de dimensions $200\mu\text{m} \times 0,5\text{mm}$). Le débit est de $2\mu\text{L}/\text{min}$ lors de la charge de l'échantillon dans la Trap colonne pendant 10 minutes. La deuxième étape consiste ensuite à séparer les peptides sur une colonne analytique: dans notre cas, nous avons de nouveau utilisé un modèle cHiP de même caractéristiques que la pré-colonne Trap. Les peptides sont ensuite séparés sur colonne cHiPLC C18 ($75\mu\text{m} \times 15\text{cm}$, Chrom XP C18-Cl $3\mu\text{m}$ 120\AA). Le débit est fixé à $300\text{nL}/\text{min}$ dans cette étape et le temps de séparation est ajusté selon la complexité de l'échantillon à analyser (plus l'échantillon est complexe, plus la durée du gradient utilisée doit être allongée). La température des 2 colonnes est maintenue à 45°C .

Une séparation de type nano-HPLC comme celle décrite précédemment utilise toujours 2 solvants. Dans notre cas, le solvant A correspond à 0.1% acide formique dans l'eau et le solvant B correspond à 0.1% acide formique dans ACN 99,99%. Le gradient est fractionné en 4 étapes:

- étape 1 : élution des peptides selon leur gradient d'hydrophobicité de 5 à 40% de B
- étape 2 : décrochage final des peptides de 40 à 80 (ou 95%) de B
- étape 3 : régénération de la colonne : 80-95% B

- étape 4 : équilibration de la colonne : 5% B

Une meilleure reproductibilité des résultats a été obtenue avec un lavage des colonnes effectué à 95% de B et des temps de régénération et d'équilibration plus longs (Figure 93).

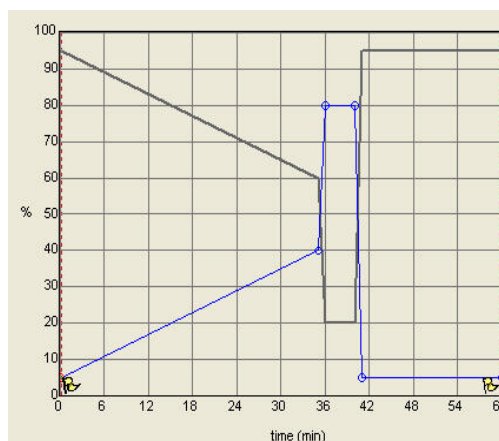


Figure 93 : Schéma du gradient chromatographique utilisé (pourcentage de solvant B en bleu, pourcentage du solvant A en noir)

6.2.2) Description technique du couplage nanoLC-ESI-MS/MS utilisé: système Triple TOF 5600 (AB Sciex)

L'instrument Triple TOF 5600 (AB SCIEX) est un spectromètre de masse en tandem : il combine un Quadripôle hybride et un accélérateur TOF (Figure 94) (Andrews, Simons et al. 2011). Sa vocation est de pouvoir être un instrument polyvalent (un appareil dit « Qual/Quant »), en permettant à la fois des mesures de masse précises et résolutes à l'aide de l'analyseur TOF (aspect qualitatif ou « Qual »), mais aussi une quantification avec le plus de justesse possible comme peut le faire un appareil de type Triple-Quad (aspect quantitatif ou « Quant »).

L'appareil est constitué de :

- (1) source d'ionisation
- (2) curtain plate avec rideau de gaz
- (3) Q-Jet (Quadripôle pour focaliser les ions)
- (4) Q0 (sélection d'ion précurseur)
- (5) Q1 (cellule de collision) suivie de Q2 (permet de garder tous les ions ou sélectionner seulement certains ions fils), suivi d'un hexapôle (transfert des ions dans le TOF)
- (6) pusher, réflectron et un détecteur.

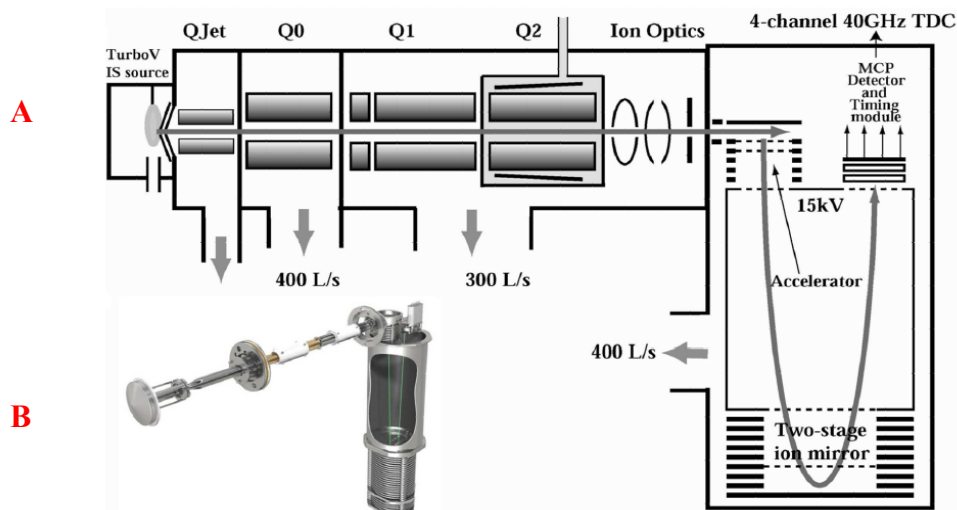


Figure 94 : (A) Illustration détaillée de l'instrument Triple TOF 5600 (AB SCIEX)
(B) Image de la machine

L'appareil est équipé d'une source d'ionisation nano-spray III (AB SCIEX), elle permet de travailler à des débits de 1nL/min à 50µL/min.

Le TripleTOF (AB Sciex) dispose de plus d'une gamme de masse utilisable en mode positif et négatif de 400-1250Da (m/z) en mode TOF-MS et 100-1500Da (m/z) en mode MS/MS. Quant à son pouvoir résolutif, nous avons constaté une résolution de 30000 en MS et de 25000 en MSMS (en mode high sensitivity).

La calibration de l'appareil est réalisée automatiquement par une injection de 25fmol d'un digestat trypsique de β -galactosidase (AB Sciex) toutes les 5h environ. Le fichier de calibration tient compte des masses des 5 peptides trypsiques répartis de manière homogène le long du gradient chromatographique. Un temps de rétention correspondant, déterminé par rapport à une injection de β -galactosidase réalisée au préalable, est associé à chacune de ces 5 masses. Une dérive de plus ou moins 60sec est tolérée sur ces temps de rétention de référence. L'instrument se sert de ces 5 masses pour se recalibrer en mode MS et 8 masses des fragments du peptide m/z=729 Da lui permettent une recalibration en mode MS/MS.

En ce qui concerne les logiciels associés au TripleTOF 5600, le logiciel de pilotage Analyst permet de régler la qualité du signal MS en mode « tuning », d'optimiser des paramètres de spray, de créer des méthodes d'acquisition et enfin de programmer des séries d'injection des échantillons (interface avec l'HPLC Eksigent intégrée). Les fichiers bruts obtenus par le spectromètre de masse AB Sciex sont des «.wiff» (windows interchange file format). Pour identifier les protéines à l'aide de l'algorithme Mascot disponible à partir du logiciel Proteinscape, ces «.wiff» doivent être ensuite convertis en «.mgf» (mascot generic

format). Finalement, nous avons également utilisé le logiciel Protein Pilot (ABSciex) qui propose une alternative pour l'identification et la validation des protéines à l'aide d'un second algorithme propre au fournisseur (Paragon). La quantification des protéines ainsi identifiées peut alors se poursuivre avec l'utilisation de la suite logicielle PeakView et MarkerView (ABS Sciex) qui est détaillée dans le chapitre 3 de la partie II.

6.2.3) Séquence d'acquisition MS et MSMS combinant identification et quantification : méthode dite « TOP 20 IDA »

L'objectif est d'être le plus exhaustif possible, donc fragmenter le plus de spectres possibles, limiter la redondance (chercher les peptides faiblement abondants) et obtenir des spectres de qualité. Ces aspects permettent d'obtenir de bonnes identifications sur la base des spectres MS/MS générés, sans toutefois être préjudiciable à l'aspect quantitatif. Pour ce dernier, le temps de cycle de l'instrument (temps nécessaire à réaliser un spectre MS complet préliminaire ainsi qu'un nombre défini de spectre de fragmentation MS/MS) est crucial : il doit être correctement défini de manière à avoir un nombre de points suffisants tout le long du profil d'un pic d'éluion peptidique. On tolère de manière générale un pic de type gaussien avec un minimum de 8 points répartis sur ce pic.

L'appellation « Top20 IDA » (Information Dependant Acquisition), correspond dans notre cas à un temps de cycle de 1.5sec comprenant 1 MS (0.25s) et 20 MSMS (20 x 0.06s): la sélection des 20 précurseurs qui seront fragmentés est effectuée selon l'abondance des peptides détectés sur le spectre MS préliminaire et sur la gestion du caractère d' « exclusion dynamique » d'un précurseur (un peptide est exclu de toute fragmentation après 2 spectres MS/MS et pendant une période de 30s) (Figure 95). Comme le temps moyen d'éluion d'un peptide est de 30s, un total de 20 points MS sera possible, ce qui définira correctement le profil d'éluion.

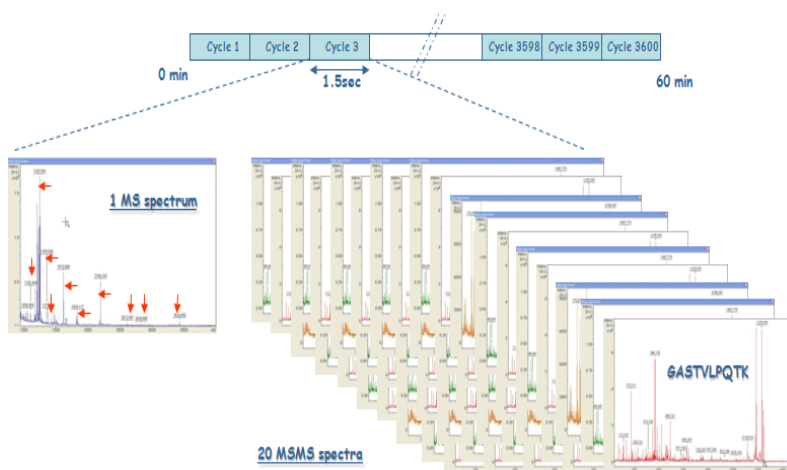


Figure 95 : Représentation schématique d'un cycle MS et MS/MS (Top 20) sur le triple TOF 5600

Cette approche permet de générer une grande quantité de données MS donnant accès aux intensités des peptides et à leur masse, ainsi que de données MS/MS permettant, grâce aux différents moteurs de recherche, l'identification de peptides et de protéines constituant

L'approche protéomique réalisée au cours de ce travail est schématisée comme suit :

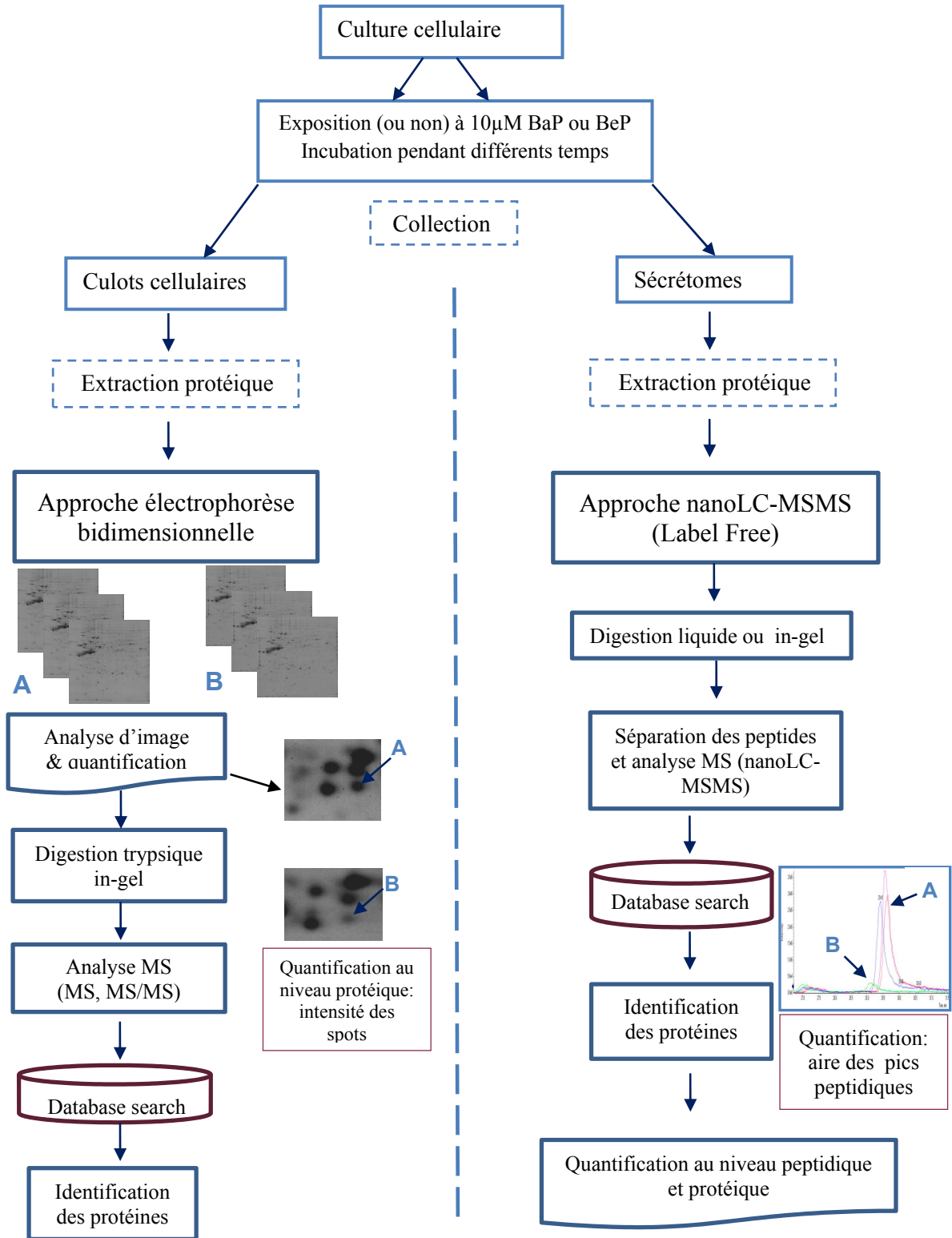


Figure 96 : Approches protéomiques développées

Annexe B
Supplemental data 3 - Article

Accession	Peptides identified (quantified)	Score Mascot	Average Normalised Abundances 2h				Average Normalised Abundances 72h			
			p-value	Control 2h	BeP 2h	BaP 2h	p-value	Control 72h	BeP 72h	BaP 72h
HS90A_HUMAN	42 (26)	2821,00	0,04	6,71e+004	5,88e+004	2,51e+005	2,37e-003	1,45e+005	1,47e+005	3,85e+005
HS90B_HUMAN	37 (18)	2566,20	0,02	1,14e+005	1,09e+005	2,52e+005	9,77e-003	1,42e+005	1,38e+005	2,50e+005
AGRIN_HUMAN	32	2299,47	0,15	1,38e+004	1,23e+004	1,17e+004	3,03e-003	8,45e+004	7,67e+004	1,04e+004
FLNA_HUMAN	35 (31)	2238,55	0,10	7,27e+004	6,55e+004	5,40e+004	0,69	7,58e+004	7,50e+004	7,80e+004
LAMA5_HUMAN	31	2086,09	0,45	5691,95	6802,82	5791,95	7,37e-005	4,08e+004	3,78e+004	5759,59
EF2_HUMAN	25	1967,45	0,03	1,39e+005	1,11e+005	2,53e+005	4,48e-004	1,14e+005	1,05e+005	1,77e+005
HSP71_HUMAN	23 (12)	1879,42	0,18	8,23e+004	6,65e+004	9,36e+004	0,01	5,37e+004	5,42e+004	7,83e+004
ACTB_HUMAN	26 (2)	1859,28	0,68	3131,78	2556,89	2776,97	0,04	1,43e+004	1,42e+004	9156,61
KPYM_HUMAN	28	1852,76	0,05	1,22e+005	1,48e+005	1,82e+005	0,47	1,60e+005	1,65e+005	1,67e+005
HSP7C_HUMAN	26 (21)	1846,41	0,29	3,43e+005	2,89e+005	3,20e+005	5,08e-003	1,85e+005	1,88e+005	2,31e+005
CH60_HUMAN	24 (23)	1821,26	0,11	7,37e+004	7,29e+004	1,17e+005	0,02	1,25e+005	1,31e+005	1,96e+005
UBA1_HUMAN	23	1819,28	0,03	4,62e+004	3,18e+004	8,79e+004	0,79	6,67e+004	6,54e+004	6,71e+004
ENOA_HUMAN	21 (14)	1787,48	8,31e-003	1,16e+005	9,49e+004	1,81e+005	0,04	2,06e+005	2,03e+005	2,63e+005
ACTG_HUMAN	25 (1)	1774,11	0,81	502,00	422,26	387,39	0,03	4440,80	4424,88	2185,96
TPIS_HUMAN	18	1624,16	0,40	3,77e+005	3,28e+005	3,51e+005	5,82e-003	3,79e+005	3,90e+005	5,98e+005
TKT_HUMAN	22 (21)	1506,76	0,41	1,13e+005	1,04e+005	9,90e+004	8,71e-003	9,92e+004	1,06e+005	1,81e+005
ACTN4_HUMAN	22 (12)	1468,09	0,06	7388,59	3881,10	1,89e+004	0,75	3,76e+004	3,86e+004	3,69e+004
TERA_HUMAN	19 (17)	1464,15	7,42e-003	2,89e+004	2,78e+004	4,71e+004	3,94e-004	3,66e+004	3,72e+004	8,54e+004
PGK1_HUMAN	16	1451,69	0,13	1,03e+005	7,98e+004	1,34e+005	0,06	1,54e+005	1,62e+005	2,05e+005

ALDOA_HUMAN	19	1419,40	0,85	5,08e+005	4,97e+005	4,95e+005	7,45e-003	4,75e+005	4,52e+005	7,38e+005
PTPRF_HUMAN	20	1401,82	0,31	1,59e+004	1,44e+004	1,29e+004	1,23e-003	1,11e+005	9,91e+004	2,30e+004
LAMB2_HUMAN	21 (20)	1368,81	0,08	2824,31	3912,68	4216,57	1,56e-003	2,92e+004	2,72e+004	5640,05
TBB4B_HUMAN	19 (1)	1363,22	0,02	6442,29	4554,07	1,56e+004	0,05	7342,31	8266,67	1,21e+004
TBB5_HUMAN	19 (2)	1341,18	0,05	3,43e+004	2,40e+004	5,97e+004	0,05	2,54e+004	2,72e+004	3,99e+004
GDIB_HUMAN	20 (15)	1333,54	0,08	2,10e+004	1,67e+004	3,69e+004	6,23e-003	4,32e+004	4,33e+004	7,58e+004
FAS_HUMAN	22 (21)	1315,24	0,02	5,06e+004	4,86e+004	7,53e+004	0,32	5,69e+004	5,75e+004	6,07e+004
TBA1B_HUMAN	16 (2)	1307,68	0,01	3117,97	2026,40	1,51e+004	0,02	8495,87	8624,24	1,41e+004
TBA1C_HUMAN	14 (0)	1234,48	---	---	---	---	---	---	---	---
TBA1A_HUMAN	14 (0)	1180,65	---	---	---	---	---	---	---	---
GRP78_HUMAN	16 (14)	1142,24	0,49	5,00e+004	3,93e+004	4,73e+004	0,04	7,04e+004	7,77e+004	8,60e+004
G3P_HUMAN	12	1067,17	0,04	1,52e+005	1,42e+005	3,52e+005	0,23	2,37e+005	2,67e+005	2,66e+005
B7TY16_HUMAN	17 (7)	1047,43	0,23	3598,03	4128,83	4919,38	0,25	8194,59	8017,64	1,04e+004
TBB4A_HUMAN	15 (0)	1047,14	---	---	---	---	---	---	---	---
QSOX1_HUMAN	15	1033,09	0,19	1,59e+004	1,19e+004	1,61e+004	1,05e-003	1,77e+005	2,00e+005	4,47e+004
G5EA52_HUMAN	14 (9)	1028,49	0,63	2,81e+004	2,47e+004	2,46e+004	4,17e-003	3,78e+004	4,10e+004	4,97e+004
LAMC1_HUMAN	16	1004,44	0,18	2170,53	1610,78	2490,31	3,22e-003	1,73e+004	1,70e+004	3098,53
PIIA_HUMAN	14	1003,64	0,03	2,35e+005	2,03e+005	1,77e+005	0,15	1,63e+005	1,63e+005	1,54e+005
CSTN1_HUMAN	16	997,82	0,23	2,60e+004	2,21e+004	2,02e+004	6,14e-005	1,21e+005	1,25e+005	3,03e+004
LDHA_HUMAN	16	995,25	0,04	6,30e+004	5,12e+004	1,33e+005	4,63e-003	1,01e+005	1,04e+005	1,92e+005
ROA2_HUMAN	14	973,51	0,01	1,08e+005	1,27e+005	4,62e+004	0,02	7,76e+004	8,15e+004	5,12e+004
1433Z_HUMAN	14 (13)	949,09	0,09	5,53e+004	3,78e+004	8,60e+004	1,19e-003	5,22e+004	5,79e+004	3,10e+004
HSPB1_HUMAN	12	906,87	0,29	2,21e+005	2,22e+005	1,83e+005	0,24	1,95e+005	2,12e+005	1,71e+005
TBB2A_HUMAN	15 (0)	889,66	---	---	---	---	---	---	---	---
G6PD_HUMAN	15 (14)	873,01	0,05	4,04e+004	3,68e+004	8,01e+004	0,10	3,96e+004	4,08e+004	5,16e+004
PLEC_HUMAN	17 (14)	866,83	0,04	9479,13	9505,27	6895,33	0,06	6728,44	5906,51	4824,90
TIMP1_HUMAN	7	847,51	0,12	3,81e+004	2,65e+004	2,28e+004	2,21e-003	3,93e+005	3,97e+005	5,98e+004

B5MCQ5_HUMAN	10	820,11	0,19	1,61e+004	1,55e+004	2,40e+004	0,02	4,16e+004	4,56e+004	6,73e+004
IQGA1_HUMAN	16	815,60	0,04	7122,98	4646,23	1,48e+004	0,25	2,32e+004	2,25e+004	1,96e+004
MDHM_HUMAN	11	814,19	0,20	2,95e+004	2,39e+004	4,49e+004	1,09e-003	4,79e+004	5,13e+004	7,68e+004
ATPB_HUMAN	13 (12)	808,48	0,10	9175,39	8423,84	1,70e+004	0,09	2,39e+004	2,80e+004	3,35e+004
Q53FA3_HUMAN	9 (0)	797,80	---	---	---	---	---	---	---	---
TBB3_HUMAN	11 (0)	794,51	---	---	---	---	---	---	---	---
ENPL_HUMAN	13 (12)	756,70	0,15	1,59e+004	1,45e+004	2,73e+004	0,06	2,82e+004	3,06e+004	4,12e+004
GDIA_HUMAN	11 (6)	755,68	0,14	6785,17	7740,69	9470,19	0,01	8832,28	7790,87	1,63e+004
LMNA_HUMAN	13 (11)	755,49	0,02	4,10e+004	3,85e+004	2,99e+004	0,13	2,89e+004	3,01e+004	3,39e+004
PROF1_HUMAN	10	751,45	0,20	6,60e+004	5,50e+004	9,17e+004	0,02	1,14e+005	1,09e+005	1,75e+005
PEBP1_HUMAN	9	747,21	0,24	5,96e+004	4,45e+004	6,20e+004	7,65e-003	7,81e+004	7,76e+004	1,18e+005
HSP76_HUMAN	8 (0)	726,20	---	---	---	---	---	---	---	---
PDIA1_HUMAN	10 (5)	724,34	0,12	7936,37	4233,09	1,06e+004	0,08	1,18e+004	1,38e+004	1,88e+004
PDIA4_HUMAN	10	723,30	0,53	3,28e+004	2,84e+004	2,91e+004	4,22e-003	4,20e+004	4,74e+004	5,70e+004
CLH1_HUMAN	14	716,06	0,55	1,65e+004	1,57e+004	1,71e+004	0,04	1,53e+004	1,44e+004	1,96e+004
IBP2_HUMAN	10	715,93	9,45e-003	5,99e+004	7,50e+004	9006,46	5,56e-003	9,59e+004	1,21e+005	1,77e+004
EF1G_HUMAN	11	705,75	0,03	1,24e+004	8897,61	3,61e+004	1,29e-003	2,17e+004	2,16e+004	4,22e+004
FINC_HUMAN	11	697,67	0,14	4041,20	4679,82	3805,09	0,01	1,49e+004	1,46e+004	5590,58
PXDN_HUMAN	11	690,79	0,88	822,24	995,33	798,33	2,28e-003	1,38e+004	1,40e+004	1497,34
CBPE_HUMAN	9	683,14	0,01	6809,40	7146,57	2685,51	5,11e-004	6,46e+004	5,52e+004	6043,21
NUCL_HUMAN	9	674,75	0,16	1,64e+004	1,39e+004	2,61e+004	0,04	2,37e+004	2,32e+004	3,85e+004
CLIC1_HUMAN	9	664,14	0,15	2,29e+004	2,01e+004	3,34e+004	0,19	2,97e+004	2,91e+004	2,36e+004
TBA8_HUMAN	7 (0)	663,13	---	---	---	---	---	---	---	---
FKBP4_HUMAN	9	655,31	0,27	1,70e+004	2,93e+004	2,98e+004	7,42e-003	1,57e+004	1,50e+004	3,64e+004
ENOB_HUMAN	5 (1)	645,31	0,08	2,41e+004	1,23e+004	3,39e+004	0,65	4,32e+004	4,31e+004	4,02e+004
CATD_HUMAN	11	638,68	0,11	3,92e+004	2,95e+004	3,68e+004	0,02	6,63e+004	6,86e+004	5,32e+004
TIF1B_HUMAN	9	630,45	0,33	2,57e+004	2,28e+004	2,33e+004	0,41	9362,83	9889,04	1,16e+004

MDHC_HUMAN	10	627,65	0,04	3,85e+004	3,51e+004	4,71e+004	0,02	4,45e+004	4,30e+004	6,77e+004
G6PI_HUMAN	7	621,39	0,14	1,62e+004	1,25e+004	2,72e+004	0,02	2,59e+004	2,67e+004	5,90e+004
EF1A1_HUMAN	11	610,26	6,90e-003	1,40e+005	1,51e+005	2,53e+005	0,04	1,11e+005	1,10e+005	1,60e+005
TPM4_HUMAN	8 (6)	607,42	0,20	4680,03	1862,97	6355,77	0,16	8607,99	8360,60	1,06e+004
IMB1_HUMAN	7 (6)	605,61	0,05	3805,73	3770,79	7060,02	0,01	4394,64	3982,64	8550,05
NPM_HUMAN	6	603,84	0,63	1,25e+004	1,22e+004	1,34e+004	0,08	3,03e+004	3,02e+004	5,04e+004
TCPH_HUMAN	8	603,63	0,01	3667,09	2938,30	1,30e+004	3,58e-003	5243,73	5706,87	1,25e+004
TAGL2_HUMAN	11	600,81	0,06	9844,27	6835,31	2,20e+004	0,42	1,66e+004	1,92e+004	1,91e+004
PGBM_HUMAN	10	593,39	0,05	153,43	134,13	393,08	4,87e-004	5697,66	5581,13	1182,47
AATC_HUMAN	12	591,65	0,07	1,24e+004	1,07e+004	1,39e+004	9,04e-004	1,66e+004	1,60e+004	3,07e+004
COF1_HUMAN	7 (6)	582,30	0,16	1,06e+005	6,36e+004	1,33e+005	0,04	1,04e+005	1,02e+005	1,36e+005
PRDX1_HUMAN	11 (8)	566,86	0,05	1,76e+004	1,62e+004	3,36e+004	0,01	3,89e+004	4,03e+004	7,29e+004
1433G_HUMAN	8 (7)	555,39	0,31	2,68e+004	2,86e+004	2,38e+004	0,03	1,47e+004	1,62e+004	2,33e+004
F16P1_HUMAN	6	544,27	0,39	2,30e+004	2,33e+004	2,08e+004	0,61	3,11e+004	3,26e+004	3,28e+004
E9PBU3_HUMAN	11	543,07	0,07	1,01e+004	8701,95	1,74e+004	0,43	1,10e+004	1,12e+004	1,19e+004
GRP75_HUMAN	8	538,86	0,64	1,54e+004	1,42e+004	1,72e+004	0,11	1,54e+004	1,71e+004	2,27e+004
A8K092_HUMAN	8 (7)	530,51	0,03	6041,87	4976,94	1,84e+004	0,14	2,22e+004	2,55e+004	3,07e+004
B4DKL4_HUMAN	7 (6)	528,03	9,95e-004	4387,56	3949,11	2732,98	4,33e-003	2,11e+004	2,26e+004	8832,58
PGAM1_HUMAN	10	527,21	0,94	4,08e+004	3,94e+004	4,11e+004	4,13e-003	3,90e+004	3,86e+004	5,78e+004
PARK7_HUMAN	8	519,21	0,41	2,22e+004	2,11e+004	2,37e+004	0,09	3,14e+004	3,29e+004	4,31e+004
ARP3_HUMAN	9	517,21	0,07	3674,26	4356,32	7895,63	0,43	1,18e+004	1,22e+004	1,43e+004
CLUS_HUMAN	8	514,86	0,03	3,26e+004	2,58e+004	5445,32	6,75e-005	2,52e+005	2,28e+005	1,89e+004
F8VRQ1_HUMAN	8	510,66	0,02	6,65e+004	9,21e+004	2,38e+004	0,13	1,67e+004	1,80e+004	1,40e+004
H7C0V9_HUMAN	7	510,06	0,11	2281,92	2024,24	1642,81	6,65e-004	2,95e+004	2,71e+004	4954,92
ALBU_HUMAN	6	508,70	3,58e-003	1,61e+006	2,37e+006	8,04e+004	0,04	3,66e+005	3,49e+005	1,14e+005
Q32Q12_HUMAN	8 (4)	505,62	0,03	1,04e+004	8258,49	1,88e+004	0,02	1,69e+004	1,72e+004	4,19e+004
HNRPK_HUMAN	8	503,37	0,09	1,41e+004	9384,29	2,34e+004	0,56	1,87e+004	1,91e+004	1,77e+004

NEO1_HUMAN	8	498,22	6,11e-003	2943,47	4034,99	927,32	0,01	2,20e+004	1,94e+004	2719,82
G5E9B2_HUMAN	8	493,30	0,02	1,42e+004	1,23e+004	2,39e+004	0,02	1,37e+004	1,22e+004	2,04e+004
VASN_HUMAN	6	491,35	0,26	3875,58	3380,50	2916,88	0,02	1,94e+004	1,85e+004	5182,78
VGF_HUMAN	8 (7)	490,68	0,01	3995,64	3457,75	1680,93	0,01	4251,61	4176,85	1214,93
SPIT1_HUMAN	8	488,96	0,51	5176,79	4698,32	5270,65	7,30e-003	2,69e+004	2,69e+004	7942,04
CBX3_HUMAN	6	468,14	0,36	1,79e+004	1,23e+004	1,91e+004	0,30	1,73e+004	1,81e+004	1,60e+004
IDHC_HUMAN	7 (6)	464,95	0,06	1288,94	650,13	4193,99	0,04	2713,94	3387,04	6970,05
ALDOC_HUMAN	4	456,83	0,09	1,32e+004	4651,28	4,64e+004	0,04	3,71e+004	3,94e+004	7,73e+004
HNRH1_HUMAN	5 (4)	451,22	0,15	7064,87	5570,87	9211,96	0,54	8054,71	8923,31	8988,59
F5H025_HUMAN	7	450,09	0,15	2717,12	2973,93	2526,95	0,05	9090,23	9555,71	6573,50
D6R9P3_HUMAN	8 (7)	447,01	0,70	1,63e+004	1,55e+004	1,66e+004	0,02	1,28e+004	1,54e+004	2,38e+004
H7BZJ3_HUMAN	6 (1)	446,39	0,04	1,11e+004	1,34e+004	3064,09	0,60	5727,93	6121,30	6090,32
TCPD_HUMAN	6	446,27	0,02	1661,04	968,14	9793,88	9,34e-003	3987,12	3789,99	9338,40
S10AB_HUMAN	5	446,18	4,06e-003	4168,92	2774,41	1,13e+004	0,10	1,08e+004	1,32e+004	2,40e+004
ILF3_HUMAN	7	444,53	0,13	3500,05	3041,43	5382,90	0,02	9943,82	9010,60	1,37e+004
FBLN3_HUMAN	7	443,24	0,01	4923,16	3093,91	435,58	1,30e-003	2,00e+004	1,72e+004	1761,08
E9PNQ6_HUMAN	7	438,87	0,02	879,20	535,31	3259,23	7,93e-003	4231,04	4950,76	1,70e+004
KCRU_HUMAN	6	438,86	0,01	1350,74	775,49	3668,76	0,05	3384,08	3427,73	6113,73
PTBP1_HUMAN	5	438,40	0,02	7056,11	6505,64	1,20e+004	0,04	8388,78	8203,22	1,38e+004
PRDX6_HUMAN	8	434,55	0,03	2,39e+004	2,08e+004	3,16e+004	0,02	1,90e+004	1,93e+004	3,77e+004
FUBP1_HUMAN	7	432,89	0,11	9823,17	1,26e+004	5918,90	0,02	6059,79	6491,87	8528,57
H14_HUMAN	7 (1)	430,19	0,14	196,60	85,89	1671,81	0,67	2331,93	2052,00	2445,04
LTBP1_HUMAN	7	430,10	0,03	850,68	783,84	160,96	7,83e-003	5627,88	5409,32	175,45
SYTC_HUMAN	9	428,78	0,73	2,17e+004	2,01e+004	1,98e+004	0,04	1,80e+004	1,58e+004	2,17e+004
EZRI_HUMAN	9 (7)	428,58	0,07	9498,88	6536,14	1,24e+004	0,04	1,41e+004	1,37e+004	1,79e+004
IF5A1_HUMAN	6	425,70	0,13	4428,25	1875,75	8452,74	0,03	7390,47	7470,67	1,61e+004
F8VTY8_HUMAN	7	422,76	0,10	1,22e+004	1,14e+004	1,78e+004	6,14e-003	8622,82	8595,00	1,43e+004

E9PCA1_HUMAN	6	420,27	0,08	1,11e+004	7870,65	1,90e+004	0,01	8892,46	6373,33	1,11e+004
TCPZ_HUMAN	9	418,31	1,22e-003	5634,34	4407,98	1,31e+004	8,43e-003	5463,97	5537,01	9728,31
GGCT_HUMAN	5	415,53	0,22	1,19e+004	1,01e+004	1,31e+004	5,54e-003	1,47e+004	1,36e+004	1,91e+004
F5GX14_HUMAN	7 (3)	412,94	0,22	3319,37	2826,78	2254,31	6,99e-003	7826,67	7288,56	2617,09
PCBP1_HUMAN	6 (4)	410,74	0,65	1,48e+004	1,36e+004	1,44e+004	0,05	1,38e+004	1,34e+004	1,90e+004
ENOG_HUMAN	4 (0)	408,67	---	---	---	---	---	---	---	---
PRDX2_HUMAN	6 (5)	406,37	0,10	1,41e+004	9259,06	3,01e+004	0,06	3,19e+004	3,15e+004	4,26e+004
B2RDM2_HUMAN	8	402,83	0,08	1,60e+004	1,45e+004	8178,65	0,02	1,43e+004	1,55e+004	2,07e+004
IF2G_HUMAN	6	402,41	0,07	2376,28	3262,49	4688,66	4,35e-003	2604,85	2411,27	6562,53
XRCC5_HUMAN	7	402,03	0,05	2916,35	2437,78	4540,33	0,05	5570,30	5801,91	3557,51
1433T_HUMAN	4 (3)	397,21	0,15	4978,56	3253,35	8050,80	0,60	7691,45	8542,78	6344,83
SYWC_HUMAN	5	395,71	0,04	2640,95	2589,90	6773,26	9,83e-003	5837,82	5157,22	1,18e+004
MUC5B_HUMAN	5	394,16	0,08	746,93	1094,51	133,58	0,01	7906,10	7371,94	372,17
DHX9_HUMAN	5	391,01	0,29	3318,23	5816,77	6867,01	0,02	4527,78	4987,53	8619,14
LKHA4_HUMAN	7	389,15	0,14	1468,06	1454,60	2314,48	1,87e-003	3019,06	2890,11	4669,47
APEX1_HUMAN	5	388,69	0,05	3,70e+004	3,35e+004	2,38e+004	5,46e-003	2,60e+004	2,74e+004	3,35e+004
SYAC_HUMAN	5	386,79	0,07	6087,69	4482,94	1,17e+004	0,17	9493,41	8858,16	6648,61
H90B4_HUMAN	7 (0)	386,57	---	---	---	---	---	---	---	---
2AAA_HUMAN	6	386,14	0,06	1506,01	753,42	6148,13	0,11	4451,05	3889,85	5082,36
BCAM_HUMAN	6	384,18	6,70e-003	8403,78	1,02e+004	3259,33	1,21e-003	1,75e+004	1,89e+004	4758,83
H12_HUMAN	7 (1)	382,63	0,26	644,57	979,58	1435,31	0,33	1444,77	1286,18	2003,90
SFPQ_HUMAN	6 (5)	380,43	0,32	1,66e+004	1,51e+004	1,44e+004	0,26	9349,34	9703,18	8813,58
XPO2_HUMAN	6 (5)	377,92	0,03	2345,05	2200,29	5110,74	0,06	3329,14	3478,75	4790,50
HSP74_HUMAN	7	373,45	0,02	3,00e+004	2,86e+004	1,20e+004	0,02	7687,82	8661,19	1,43e+004
DDB1_HUMAN	4	369,28	0,28	2102,56	1290,58	2429,79	0,02	4132,63	2311,26	4881,08
PLOC3_HUMAN	7	367,73	0,39	5090,74	4999,46	4359,85	2,78e-003	1,12e+004	1,18e+004	5838,00
B4DQJ8_HUMAN	6	366,10	0,08	6776,91	4864,63	1,27e+004	0,01	8157,14	8081,99	1,63e+004

1433E_HUMAN	6 (5)	365,96	0,04	3168,54	2066,99	5884,90	0,22	8121,32	8355,32	1,26e+004
CAZA1_HUMAN	6	365,94	0,06	8895,86	8002,20	1,48e+004	0,07	2,31e+004	2,26e+004	2,78e+004
KYNU_HUMAN	6	364,01	0,23	1,69e+004	1,13e+004	1,86e+004	0,02	2,95e+004	3,08e+004	4,45e+004
B4DUR8_HUMAN	7	362,04	0,09	6979,25	7033,58	1,17e+004	0,07	6227,98	6920,10	9388,80
RRBP1_HUMAN	6	361,02	0,48	5724,99	5043,34	5625,10	0,02	4405,95	4667,68	7127,04
PSA1_HUMAN	6	359,91	0,04	7330,66	6820,87	3648,40	2,14e-003	5566,38	5277,39	1,33e+004
ASSY_HUMAN	7	359,57	0,02	9925,54	7538,49	2,61e+004	0,37	2,03e+004	1,96e+004	1,82e+004
ANXA2_HUMAN	7	359,50	0,03	2275,70	2027,02	6991,43	0,01	1,08e+004	1,31e+004	1,11e+004
E9PRY8_HUMAN	5	357,76	0,19	1,55e+004	1,30e+004	1,73e+004	0,86	7692,81	7927,26	7979,27
TCPB_HUMAN	8 (6)	356,19	0,04	1421,24	722,42	4636,30	5,91e-003	2954,66	2399,57	6067,12
HNRPD_HUMAN	5 (4)	353,98	0,30	2,33e+004	1,95e+004	1,76e+004	0,62	1,77e+004	1,83e+004	1,72e+004
RLA0_HUMAN	5	351,78	8,30e-004	3231,38	3469,85	9097,85	0,03	3697,83	3422,32	6622,15
CALR_HUMAN	4	351,74	0,40	1,11e+004	9332,26	7142,21	0,11	1,16e+004	1,19e+004	2,31e+004
PSA_HUMAN	6 (5)	351,19	0,08	2681,99	1911,61	4961,79	0,03	7704,01	7092,56	1,19e+004
CAP1_HUMAN	5	350,89	0,07	4172,61	3016,15	9709,89	0,06	6204,57	5034,10	1,01e+004
G3V180_HUMAN	5	350,28	0,27	1,49e+004	1,02e+004	1,39e+004	0,56	1,25e+004	1,27e+004	1,35e+004
GFRA1_HUMAN	5	349,47	0,55	4668,91	4024,79	4999,97	8,58e-004	1,16e+004	1,20e+004	2486,08
PSA7_HUMAN	5	349,18	0,35	1,25e+004	1,07e+004	1,44e+004	0,34	2,57e+004	2,41e+004	2,94e+004
G3V5I3_HUMAN	5	348,98	0,25	1,08e+004	9203,35	9315,65	0,01	4,72e+004	5,42e+004	2,32e+004
HNRPC_HUMAN	6	348,78	0,01	3,60e+004	4,69e+004	1,88e+004	0,03	1,71e+004	1,52e+004	2,09e+004
PABP1_HUMAN	7 (5)	346,74	0,04	4376,71	2622,65	9672,87	1,80e-003	8256,46	9218,81	1,39e+004
HNRPF_HUMAN	4 (3)	346,71	0,07	2135,89	849,82	5585,65	0,04	3084,85	3121,22	4702,20
IF4B_HUMAN	7	345,28	0,50	7197,52	8235,65	7343,26	0,11	3923,74	4815,84	6256,34
CEL2_HUMAN	8	345,02	0,25	2193,36	1639,89	1568,75	1,85e-003	8501,31	8650,57	2283,36
RAN_HUMAN	6 (5)	342,98	0,03	2,25e+004	1,85e+004	3,69e+004	1,15e-003	2,05e+004	2,15e+004	2,66e+004
MAOX_HUMAN	5	342,69	0,12	1802,12	1529,07	3117,72	0,03	2275,97	2039,56	3702,14
PDC6I_HUMAN	5	337,67	0,36	381,27	588,01	447,17	0,04	997,14	906,61	223,86

JAM1_HUMAN	6	337,27	0,05	2614,67	2785,91	1721,20	0,13	2,83e+004	3,25e+004	1,43e+004
RSSA_HUMAN	5 (4)	336,91	0,73	2,41e+004	2,50e+004	2,45e+004	0,10	1,11e+004	1,20e+004	1,31e+004
BASP1_HUMAN	5	332,74	0,10	3334,00	1269,92	790,69	0,18	5485,58	5839,96	1721,73
XRCC6_HUMAN	6 (5)	332,10	0,06	5684,76	5744,51	1,06e+004	0,82	8034,33	8067,34	7877,14
BLVRB_HUMAN	5	331,81	0,70	4959,07	5430,65	4255,63	3,73e-004	4637,19	5489,82	1,07e+004
H0Y449_HUMAN	3	330,26	0,06	7188,43	6643,84	1868,48	0,25	749,72	717,76	864,67
1433B_HUMAN	4 (3)	327,78	0,26	8155,89	6564,00	9884,70	0,03	6351,95	7089,66	3464,83
B4DZF8_HUMAN	6	326,20	0,64	1297,36	1437,32	1477,83	0,02	1445,17	1322,67	6252,75
PTMA_HUMAN	5	325,41	0,06	8,99e+004	8,25e+004	4,43e+004	0,07	2,90e+004	2,63e+004	4,28e+004
CAPZB_HUMAN	5	324,40	0,20	3835,38	3616,33	7398,38	0,35	1,08e+004	1,15e+004	1,32e+004
ACTBL_HUMAN	6 (0)	321,01	---	---	---	---	---	---	---	---
COPG1_HUMAN	6	319,35	0,05	1321,17	868,47	5712,35	3,83e-003	2930,72	2717,79	5717,94
AP1B1_HUMAN	5 (4)	316,74	0,12	1722,71	1968,83	3416,83	0,01	1464,68	1202,78	3977,46
CNDP2_HUMAN	6	315,95	0,06	3760,38	2607,83	9025,04	7,65e-003	1,01e+004	1,12e+004	1,75e+004
Q5VU66_HUMAN	4 (1)	315,50	0,18	149,07	9,37	871,54	0,26	404,02	531,25	416,17
FETUA_HUMAN	4	314,83	0,04	2,06e+005	1,75e+005	2,07e+004	0,05	2,67e+004	1,61e+004	7835,01
NDKA_HUMAN	6 (2)	312,14	0,21	45,77	68,72	504,84	0,01	448,88	576,48	2776,81
PSA3_HUMAN	4	311,95	0,24	7969,86	5730,56	1,11e+004	0,03	9819,13	9284,69	1,39e+004
PSB5_HUMAN	4	310,75	0,08	6969,70	6438,47	9583,77	4,76e-003	6528,49	6060,65	1,10e+004
C1QBP_HUMAN	2	310,22	0,29	3689,95	1924,59	7443,47	6,67e-003	9207,87	1,00e+004	2,09e+004
IPYR_HUMAN	5	309,51	0,02	5502,50	5083,31	1,26e+004	6,10e-003	9123,95	9823,98	1,28e+004
SAHH_HUMAN	6	309,04	0,06	6084,70	3539,01	1,52e+004	0,02	1,27e+004	1,36e+004	1,85e+004
TBB6_HUMAN	6 (0)	307,94	---	---	---	---	---	---	---	---
PSME2_HUMAN	4	307,89	0,07	1521,31	620,24	8612,26	0,05	4848,33	4605,76	1,02e+004
TFF1_HUMAN	4	306,79	0,01	9,40e+004	1,13e+005	5595,16	4,87e-003	8,79e+004	7,58e+004	7967,67
APT_HUMAN	5	306,73	0,01	4157,28	4691,83	7517,75	0,61	5804,67	5988,92	6698,71
NHRF1_HUMAN	4	306,57	0,14	7768,38	4862,96	1,06e+004	0,03	5242,92	4738,82	2621,87

F5H2F4_HUMAN	6	306,26	0,08	3009,95	1843,58	7112,79	0,63	4091,78	4101,33	4498,72
CATA_HUMAN	5	305,44	0,69	3641,62	3985,64	3388,77	0,36	4956,50	5618,24	6225,18
HNRPU_HUMAN	4	305,13	0,35	1,29e+004	1,54e+004	1,51e+004	2,64e-003	1,38e+004	1,42e+004	2,50e+004
SBP1_HUMAN	6	303,49	0,22	1,66e+004	1,63e+004	1,28e+004	0,08	1,80e+004	1,92e+004	1,63e+004
HPRT_HUMAN	4	303,44	0,02	1690,73	1787,88	4963,16	2,27e-003	4600,36	3479,14	9963,36
H3BRU6_HUMAN	5 (3)	303,19	0,05	3885,81	3015,09	7653,22	0,68	5330,62	5042,45	5278,04
B4E0R6_HUMAN	4	302,60	0,12	1494,66	1228,79	2573,93	0,01	1226,49	1063,73	2834,64
E7ETN3_HUMAN	5	302,36	0,39	3714,28	3560,10	3288,16	0,09	9307,07	1,19e+004	4813,79
EPCAM_HUMAN	4 (3)	299,59	0,19	6858,69	5499,77	9020,10	0,05	1,88e+004	1,98e+004	1,29e+004
LYAG_HUMAN	4	298,85	0,20	822,44	273,02	889,58	0,04	3285,27	3669,37	2048,85
ANXA5_HUMAN	4	296,57	0,07	148,04	50,77	2336,56	0,02	2473,19	4246,38	7431,57
Q5STU3_HUMAN	5	296,01	0,09	1899,19	1679,64	3496,13	0,02	3572,34	3725,24	5445,38
CAND1_HUMAN	6	295,86	0,03	4715,21	5081,31	6920,35	0,02	6700,25	6292,71	1,17e+004
H2A1B_HUMAN	3 (0)	293,00	---	---	---	---	---	---	---	---
CH10_HUMAN	6	289,70	0,49	2,46e+004	2,16e+004	2,83e+004	0,04	2,91e+004	3,08e+004	4,04e+004
IF4A1_HUMAN	4 (1)	289,47	0,08	14,18	22,30	381,79	0,01	45,42	60,17	176,73
G3V1V0_HUMAN	6	288,39	0,27	9472,87	7236,68	6942,79	0,02	1,71e+004	1,55e+004	8938,04
HYOU1_HUMAN	5	287,69	0,45	815,45	977,23	1790,83	6,41e-003	823,88	2323,12	3648,42
H2A2A_HUMAN	3 (1)	287,66	0,33	1,04	1,04	6,69	0,65	186,79	230,25	186,90
B4DTU0_HUMAN	5	286,48	0,31	584,86	658,27	236,24	2,88e-004	6685,44	6639,48	750,10
ISOC1_HUMAN	4 (3)	284,44	0,07	4745,72	3963,75	6290,27	0,03	7472,05	7223,74	1,02e+004
AATM_HUMAN	5	283,58	0,07	5964,73	4621,50	8345,49	0,01	1,19e+004	1,21e+004	1,81e+004
PHB2_HUMAN	4	283,13	0,05	631,61	484,38	3033,48	0,04	2682,79	3015,13	5588,88
SODC_HUMAN	3	282,46	0,87	3,63e+004	4,14e+004	3,49e+004	0,01	5,25e+004	5,44e+004	1,04e+005
H0Y3Z3_HUMAN	5 (0)	281,97	---	---	---	---	---	---	---	---
CSTN3_HUMAN	5	277,45	0,02	427,71	422,33	113,84	7,35e-004	5429,95	4631,25	321,01
H2AX_HUMAN	4 (2)	276,19	0,16	2,66	1,43	381,27	0,69	395,37	447,05	327,00

SPTB2_HUMAN	4	276,04	0,09	3476,94	4594,45	1552,20	0,55	1649,38	1501,26	1356,40
NUDT5_HUMAN	4 (3)	275,78	0,10	2092,14	1514,53	4990,68	1,13e-003	5438,73	5351,82	9326,23
HOYN26_HUMAN	3	275,62	0,15	4289,07	3915,23	7202,19	0,07	7402,89	9354,37	1,46e+004
SPTN1_HUMAN	5	274,26	0,16	844,33	813,99	1360,07	0,16	1321,09	1329,44	2274,98
FREM2_HUMAN	5	273,40	0,64	1396,81	1312,69	1519,77	0,20	2406,17	2830,56	2838,48
B1AVU8_HUMAN	5	273,39	0,09	3177,21	1712,65	820,07	0,63	4406,65	4440,19	3734,82
FPRP_HUMAN	5	271,85	0,27	674,18	653,47	484,67	6,62e-003	3260,90	3060,95	647,52
SYK_HUMAN	5	271,23	0,54	1,00e+004	1,23e+004	1,04e+004	0,06	6491,66	6701,86	9528,62
Q5SZU1_HUMAN	4	270,96	7,25e-003	6355,65	6475,77	1,28e+004	0,94	6755,18	6904,09	6914,06
IF4A2_HUMAN	4 (1)	270,53	0,70	922,90	1179,21	924,45	0,74	1204,77	1177,43	1130,71
CADH1_HUMAN	4	269,86	0,03	7916,02	6354,78	4816,38	1,27e-003	2,57e+004	2,51e+004	1,01e+004
GLYM_HUMAN	6	264,46	0,08	2918,48	2849,91	4971,42	0,13	3231,42	3586,76	5528,62
GANAB_HUMAN	6	263,48	0,09	2928,28	2198,57	6225,77	0,40	9944,77	1,08e+004	1,06e+004
LAP2A_HUMAN	4 (3)	262,10	0,44	2347,55	1593,38	1861,95	0,08	708,09	667,77	219,29
FLNB_HUMAN	5 (3)	259,74	0,02	3320,47	3436,25	1287,96	0,02	2869,99	2698,69	1906,68
1433S_HUMAN	3 (2)	259,11	0,96	4085,13	4001,61	4155,57	0,25	1201,56	1156,73	969,20
F10A1_HUMAN	4	258,91	0,02	1867,25	2601,30	6504,93	0,17	3526,79	4171,95	5255,69
CNBP_HUMAN	4	258,42	0,03	2446,39	2721,05	675,95	0,15	515,37	505,00	631,53
B4DZI8_HUMAN	4	257,92	0,11	1120,56	1069,21	2497,63	0,44	4035,10	3712,15	4026,15
RLA1_HUMAN	4 (3)	255,47	0,19	6602,61	8693,12	1,12e+004	0,29	1,14e+004	1,06e+004	1,62e+004
HNRPM_HUMAN	4 (3)	255,21	0,34	901,07	1474,47	1230,25	0,54	1913,43	2182,03	1907,93
BOYIW6_HUMAN	5	254,83	0,07	6512,63	5690,23	1,10e+004	0,01	8899,83	8888,42	1,47e+004
LMNB1_HUMAN	5 (4)	254,34	0,71	1522,05	1633,29	1636,26	0,29	1183,02	1081,62	1712,92
LEG1_HUMAN	4	253,84	0,08	2,15e+004	2,46e+004	1,64e+004	0,10	2,10e+004	2,26e+004	2,00e+004
ECHM_HUMAN	3	253,84	0,06	145,16	18,93	2726,55	0,21	2702,15	2892,55	4414,16
GALT6_HUMAN	4	253,50	0,64	486,60	704,97	738,65	4,41e-003	4972,41	5108,78	2318,18
TPR_HUMAN	4	251,96	0,14	407,98	252,18	916,60	0,08	693,38	524,34	903,84

ANXA6_HUMAN	4	250,73	0,10	60,17	8,44	1409,73	0,26	3410,00	3845,18	3448,29
PRP19_HUMAN	4	248,22	0,67	4055,01	4157,94	5066,96	0,15	2387,84	2008,27	2947,13
RBBP4_HUMAN	3 (1)	248,18	0,05	4309,21	3741,01	2455,15	0,06	1032,22	849,54	1541,99
PRS7_HUMAN	5	247,58	0,16	2408,78	1812,97	4490,44	0,30	2429,23	2617,12	3502,54
LGUL_HUMAN	4	245,04	0,25	3444,79	2262,55	4516,02	0,15	5913,63	5273,51	8442,79
G3XAM7_HUMAN	3 (2)	244,87	0,27	155,86	85,99	459,04	0,33	1202,35	1259,19	727,81
Q5T5C7_HUMAN	3	244,37	0,05	629,94	442,39	3041,24	0,09	1223,70	1338,06	969,44
COMT_HUMAN	4	244,17	0,25	7749,79	6990,25	9436,71	0,58	5955,52	5732,82	5079,81
HMGB3_HUMAN	3	243,76	0,29	5266,59	3118,44	6123,55	0,02	5668,53	6058,87	1,30e+004
RLA2_HUMAN	4 (3)	242,95	0,13	543,03	67,35	1793,05	0,43	2122,65	2478,30	3138,36
GDF15_HUMAN	3	242,38	0,02	334,83	365,35	1218,35	8,20e-003	4222,84	5127,09	3,00e+004
ILF2_HUMAN	4	241,81	0,07	589,07	452,77	2172,66	0,02	3515,51	3273,06	8798,34
PCNA_HUMAN	3 (2)	241,01	0,30	2389,78	2479,43	3164,96	0,02	3038,69	3224,41	5330,18
HMGB1_HUMAN	5 (4)	240,03	3,52e-003	3,66e+004	4,60e+004	1,21e+004	0,06	1,30e+004	1,33e+004	1,78e+004
F5H897_HUMAN	3 (2)	239,79	0,08	9,18	51,23	532,34	0,07	470,06	566,34	944,33
PNPH_HUMAN	5	236,82	0,21	1511,55	1558,53	841,69	0,01	3119,78	3043,72	8373,66
NUCKS_HUMAN	4	236,44	2,49e-003	81,51	73,40	1255,54	0,03	1213,07	1074,65	2875,08
RBMX_HUMAN	5	234,04	0,05	8944,33	9519,18	5510,79	0,03	5110,39	4454,12	6194,02
TIMP2_HUMAN	4	232,99	0,02	3381,76	4342,91	1697,59	5,40e-003	2,78e+004	2,84e+004	8720,97
THIO_HUMAN	4	232,50	0,06	9630,09	9259,31	2,31e+004	0,05	2,21e+004	2,46e+004	5,98e+004
BZW2_HUMAN	4	231,58	0,16	380,04	184,22	2640,62	0,01	1884,41	1792,10	3088,75
HS105_HUMAN	3	231,53	0,35	4378,28	3209,30	3786,40	0,05	2054,31	1822,11	3490,28
J3KPF3_HUMAN	4	230,99	0,65	2946,75	3358,73	3754,85	0,08	7213,45	8652,54	1,03e+004
H31_HUMAN	3 (1)	230,95	0,27	9,71	25,56	261,12	0,22	2500,26	2917,25	5441,99
PSMD2_HUMAN	3	230,03	0,01	2823,46	3217,41	4658,19	7,22e-003	1740,50	1611,47	2829,91
NOLC1_HUMAN	4	229,33	2,36e-003	0,01	0,03	26,21	0,05	252,64	188,87	1455,47
UBP5_HUMAN	4	229,11	0,83	2894,27	2874,64	2928,03	0,03	1716,16	1843,97	2692,42

TRFL_HUMAN	3	228,04	0,01	1,18e+005	1,73e+005	1,01e+004	0,02	3,14e+004	3,40e+004	1,50e+004
H32_HUMAN	3 (1)	227,87	0,04	4,33	3,33	376,83	0,19	3359,10	4068,92	8983,68
F8W031_HUMAN	3	227,64	0,32	5785,65	5091,12	4058,51	0,06	7562,72	7511,51	1,26e+004
CYTC_HUMAN	4	227,62	3,63e-003	2,60e+004	2,44e+004	1,23e+004	9,69e-003	1,44e+005	1,44e+005	4,20e+004
B4DXJ1_HUMAN	3	226,49	0,06	2761,93	2234,92	3765,78	0,16	4042,99	3793,57	4853,43
SND1_HUMAN	5	225,07	0,18	2521,60	873,11	3946,39	0,57	2060,10	2229,21	2043,77
ESTD_HUMAN	3	224,63	0,30	3865,39	3524,67	5630,54	0,52	5091,19	4646,59	4419,67
HOYAE9_HUMAN	5	223,50	0,23	1875,69	1468,60	2438,60	8,08e-003	8307,57	9901,48	3633,85
GLU2B_HUMAN	4	223,29	0,12	5176,19	3496,93	8280,38	0,03	1,59e+004	1,64e+004	2,17e+004
GSLG1_HUMAN	3	223,08	0,29	549,96	419,42	269,57	0,02	1146,44	996,52	622,68
TCPA_HUMAN	4 (3)	222,28	0,11	1694,02	566,94	7496,15	3,79e-003	3477,89	3548,56	6504,60
J3QKR3_HUMAN	2	221,60	0,07	1066,92	290,21	2650,20	0,06	2708,30	2946,46	6112,76
RINI_HUMAN	3	221,52	0,08	239,64	186,51	1781,67	0,06	1329,07	1274,04	2143,99
A9R9N7_HUMAN	3	221,37	0,51	355,48	470,55	373,62	5,03e-003	6486,99	6764,86	1190,78
RABP1_HUMAN	3 (2)	220,56	0,15	359,68	53,21	1990,52	0,13	1587,67	1588,48	2005,54
J3KN67_HUMAN	3 (0)	220,41	---	---	---	---	---	---	---	---
PUR6_HUMAN	4	217,79	9,20e-003	8798,46	8843,84	1,38e+004	0,03	8702,17	8834,21	1,25e+004
F5GWP8_HUMAN	4 (3)	215,36	0,08	5,15e+004	4,14e+004	6,42e+004	0,34	5,13e+004	5,00e+004	5,71e+004
CRIP2_HUMAN	3	213,87	0,66	1589,38	1585,27	1908,91	0,12	1692,95	1558,42	1148,25
PHP14_HUMAN	2	208,96	0,10	1344,77	516,32	2714,05	0,32	1881,13	1719,13	2217,22
SULF2_HUMAN	4	208,60	0,26	675,80	316,56	324,81	0,07	2769,00	3159,59	1337,42
J3KTF8_HUMAN	3	208,51	0,11	8522,88	7481,48	1,14e+004	0,46	1,04e+004	1,11e+004	1,18e+004
G3V5Z7_HUMAN	3 (2)	205,71	0,17	4489,99	2821,11	7632,59	6,30e-003	6312,58	6844,51	1,40e+004
AT1A1_HUMAN	3 (2)	205,24	0,57	3711,68	2957,62	3078,13	0,18	1878,33	2330,84	1557,56
SRSF1_HUMAN	4	204,37	0,38	2714,17	2055,42	2435,88	0,29	2278,44	1655,91	1405,00
HOYDP9_HUMAN	4 (0)	203,78	---	---	---	---	---	---	---	---
TXD17_HUMAN	3	203,74	0,16	8678,48	7236,14	6981,50	5,88e-003	8510,74	8626,24	1,41e+004

NOTC3_HUMAN	4	202,68	0,49	341,24	379,04	465,10	2,49e-004	1982,25	1523,18	127,80
TWF1_HUMAN	3	202,57	0,07	2382,22	2054,38	3927,42	0,15	3096,69	3148,91	4203,11
H4_HUMAN	4	202,48	0,04	3576,85	3382,67	2,53e+004	0,26	3,93e+004	4,02e+004	7,50e+004
FLNC_HUMAN	4 (0)	201,27	---	---	---	---	---	---	---	---
ILEU_HUMAN	4	201,01	0,57	1172,53	1188,22	1488,14	0,03	2226,44	2513,48	3944,16
CDC37_HUMAN	3	200,54	0,09	3970,00	1546,52	1,10e+004	0,14	6274,75	6618,74	8980,26
RABP2_HUMAN	5 (4)	196,69	0,06	9697,63	7363,51	1,79e+004	0,09	3,45e+004	3,44e+004	4,05e+004
CO6A1_HUMAN	3	196,34	0,30	191,90	136,01	221,04	4,03e-004	2492,97	2456,45	623,35
FSTL4_HUMAN	4	195,68	0,73	251,23	243,38	200,22	1,22e-003	2596,97	2554,90	525,24
GELS_HUMAN	3	195,14	0,08	597,23	398,72	248,20	0,03	1829,82	2607,04	843,83
CSRP1_HUMAN	3	194,58	0,34	6951,40	6733,28	5766,09	0,26	8728,74	7816,88	6347,05
H2AY_HUMAN	3	194,18	0,11	120,49	83,05	747,97	0,12	1912,04	1892,64	7540,15
OLA1_HUMAN	3	193,86	0,09	789,40	636,14	3019,60	0,36	747,74	853,95	1243,33
E9PJK1_HUMAN	3	192,83	0,25	471,93	2340,57	482,50	1,31e-004	8080,86	7787,98	1047,69
STIP1_HUMAN	4	191,19	0,36	7526,61	8936,36	9398,81	0,29	6129,70	6741,93	7527,39
GBLP_HUMAN	4	190,47	0,07	4679,15	2752,75	1,19e+004	0,02	5320,25	6562,61	9512,21
S10AE_HUMAN	3	190,29	0,31	2161,97	1805,35	1718,70	0,03	3877,82	3710,71	2265,29
NONO_HUMAN	3 (2)	188,94	0,20	4377,77	3822,42	5412,29	0,26	3454,99	3540,21	4418,90
MRP_HUMAN	3	188,56	0,16	3393,79	3380,32	1281,73	0,07	1785,72	1241,44	636,48
A7MAP1_HUMAN	4	186,93	0,11	2112,36	1834,95	3231,39	0,58	6134,09	6356,20	5745,02
A8MTM1_HUMAN	2	184,30	0,15	501,31	65,19	2196,74	0,05	1223,76	1356,65	1738,65
RAP2C_HUMAN	3	184,11	0,05	5842,72	4222,00	9704,52	0,06	6694,42	6531,41	5052,17
E9PPV6_HUMAN	2	183,43	0,29	320,31	204,32	1938,61	0,09	1019,43	1414,53	3220,54
PEPL1_HUMAN	3 (2)	182,65	0,18	89,72	47,44	272,07	0,09	225,85	180,63	749,46
OAT_HUMAN	3	181,52	0,35	521,05	522,24	791,03	0,13	163,73	341,94	360,58
FUBP2_HUMAN	3	180,62	0,21	7145,89	6548,41	5491,89	0,12	4296,85	4299,33	3243,85
GRB2_HUMAN	3	180,33	0,45	1914,55	2129,95	2263,06	0,10	1555,44	1227,86	2387,10

EPHB4_HUMAN	5	180,26	0,17	4754,99	3058,84	1803,89	7,19e-003	4231,93	3903,67	1280,59
CDV3_HUMAN	3	180,21	0,03	7795,62	6510,81	2729,37	0,24	2105,24	1736,03	2221,30
MATN2_HUMAN	4	178,65	0,17	220,27	198,35	373,61	0,01	2172,23	2225,61	309,47
ACLY_HUMAN	3	177,36	0,17	1557,24	918,06	3213,00	0,03	1050,38	1167,89	2169,97
B4DR52_HUMAN	3 (2)	177,17	0,06	867,53	880,41	3734,81	1,11e-003	8503,75	7852,56	1,65e+004
TSP1_HUMAN	4	175,86	0,37	2865,59	2444,10	3090,60	4,33e-003	6064,44	5451,77	2833,33
HCD2_HUMAN	4	175,66	0,09	372,68	192,02	1328,16	0,17	2364,18	2497,19	3406,46
UB2V1_HUMAN	3	175,04	0,11	8416,85	1,05e+004	5676,82	0,05	4346,10	4010,19	7015,26
RAB10_HUMAN	3 (2)	174,67	7,70e-003	833,87	838,38	1714,08	0,84	1805,65	1709,71	1713,75
PHB_HUMAN	3	174,58	0,29	2618,18	2746,46	1312,56	0,20	1772,91	1997,85	2833,40
ERP29_HUMAN	3 (2)	174,23	0,19	3618,96	2184,87	5355,34	0,02	6709,71	7661,40	1,14e+004
THOC4_HUMAN	2	173,53	0,23	54,88	48,82	296,23	0,40	470,57	501,55	1161,82
COR1B_HUMAN	3	173,35	0,04	1076,63	582,99	3236,38	0,04	4074,92	3873,67	3081,48
HDGF_HUMAN	3	173,26	0,23	2,07e+004	1,65e+004	1,88e+004	6,62e-003	1,77e+004	1,81e+004	2,36e+004
G3V3X4_HUMAN	2	172,12	0,30	2890,90	2308,43	1651,03	0,10	3139,87	2926,89	4294,79
KTN1_HUMAN	5	171,90	0,03	6074,76	8905,12	1642,90	0,11	1484,19	1221,63	805,42
SAE1_HUMAN	2	171,89	0,07	15,66	22,56	169,05	0,02	47,08	55,33	328,61
PPT1_HUMAN	4	170,54	0,36	935,52	1410,88	1364,69	0,07	5811,60	5448,10	3256,79
PGP_HUMAN	2	170,06	0,25	41,80	49,57	411,15	0,04	383,31	364,20	922,05
K7EK07_HUMAN	3 (1)	168,94	0,42	1555,13	1396,39	2838,90	0,16	3483,99	4375,67	8089,82
SF3B3_HUMAN	4	168,13	0,07	690,72	431,62	2230,81	2,02e-003	1627,19	1857,21	4355,62
GGH_HUMAN	2	168,00	0,31	391,25	135,05	466,13	6,07e-003	3890,70	3590,25	1189,58
E9PNS7_HUMAN	3	167,84	0,04	1077,62	605,97	200,86	5,62e-003	3531,90	3554,21	556,64
EIF3B_HUMAN	3	167,69	0,03	3305,54	3949,10	1753,55	0,50	3085,14	3229,50	2796,94
PIIB_HUMAN	4	166,87	0,54	6233,82	6368,81	7531,83	0,41	2,84e+004	2,86e+004	3,11e+004
VDAC2_HUMAN	3	166,04	0,25	3557,88	4205,79	5631,53	0,04	4622,07	6012,59	8766,23
PSB1_HUMAN	4	165,77	0,55	5567,86	4937,41	4528,04	0,01	3761,67	4378,18	8316,65

DAG1_HUMAN	4	165,60	0,50	1109,47	1605,40	889,60	4,40e-003	3467,86	2835,51	494,85
B1ANR0_HUMAN	3 (1)	164,61	0,08	48,05	0,00	414,73	0,22	23,36	54,48	272,77
J3KN86_HUMAN	3 (2)	163,86	0,44	52,91	41,10	94,47	0,21	796,70	824,43	1024,31
H2AV_HUMAN	2 (1)	163,81	0,06	0,00	1,39	222,99	0,60	215,52	271,91	323,30
PVRL4_HUMAN	2	163,36	0,32	995,40	2449,33	770,74	7,66e-003	2230,82	2403,54	604,97
TALDO_HUMAN	3	162,94	0,25	1,74e+004	1,50e+004	1,93e+004	0,05	1,66e+004	1,77e+004	2,49e+004
COIA1_HUMAN	3	161,30	0,06	380,96	212,64	51,92	0,01	1577,16	1330,44	165,07
TCTP_HUMAN	2	160,52	0,23	1993,90	857,80	2919,47	0,01	2460,43	2652,25	7312,81
B0QYK0_HUMAN	3	159,84	0,13	2925,81	1567,89	2921,06	0,12	3268,35	4159,48	6349,70
PAIRB_HUMAN	3	156,71	3,93e-003	2537,62	2007,17	1155,36	0,03	906,04	876,00	543,67
GSHB_HUMAN	2	156,70	0,05	3881,43	3693,10	2211,11	5,50e-003	2754,07	2706,73	3686,21
PPIC_HUMAN	3	156,58	0,25	631,89	1052,08	495,14	3,82e-003	1,06e+004	1,06e+004	2673,61
A0AV56_HUMAN	3	155,92	0,31	2120,87	1721,43	2232,95	0,12	1867,19	1971,46	2191,22
PSA5_HUMAN	2 (1)	153,84	9,14e-003	1,12e+004	1,18e+004	6669,74	0,04	4158,10	4216,29	6788,48
PSA2_HUMAN	3	153,39	0,22	6105,92	5199,37	6981,51	0,08	4641,93	3778,55	8793,77
AIFM1_HUMAN	3	152,76	0,28	1665,41	3978,88	2433,86	0,02	1840,55	2316,75	4121,14
RL12_HUMAN	2	152,52	0,14	767,54	945,43	3014,75	0,01	953,70	932,65	1887,61
IF2A_HUMAN	2 (1)	152,51	0,55	3498,61	3416,94	3859,49	0,06	2336,08	2288,65	3027,84
AMPB_HUMAN	3 (2)	151,29	0,07	1348,83	1160,99	2629,88	0,03	2198,11	1873,42	3451,91
RL11_HUMAN	2	150,87	0,12	120,47	48,18	568,11	0,06	924,90	1035,85	1707,83
RBBP7_HUMAN	3 (1)	149,42	0,40	4996,42	4723,29	4201,17	0,06	2463,60	2231,29	3213,77
HNRPQ_HUMAN	3	148,47	9,15e-003	1,29e+004	1,33e+004	7645,22	0,93	7755,38	7924,41	8127,11
B4DDJ7_HUMAN	1	148,07	0,17	1041,72	217,68	2752,43	0,07	426,64	445,69	1360,15
E7EX60_HUMAN	3 (2)	148,06	0,16	39,57	14,53	152,90	0,08	386,56	627,54	252,50
ACL6A_HUMAN	2	148,00	0,06	748,71	905,66	1371,62	0,08	761,08	685,73	1116,40
RADI_HUMAN	3 (1)	147,16	0,14	89,61	2,21	861,70	0,27	2354,26	1925,09	2508,52
F5H6N3_HUMAN	1	146,99	0,02	2139,38	1735,96	366,77	0,11	163,77	130,63	11,50

H9KV70_HUMAN	1	146,44	0,11	853,45	444,67	273,61	0,01	7956,08	1,01e+004	3682,13
Q7Z451_HUMAN	2 (1)	145,91	0,37	629,65	790,86	705,62	0,74	864,53	890,69	956,74
DDX17_HUMAN	2	145,72	0,35	1411,12	1490,70	982,19	0,08	562,53	600,95	464,63
DCTP1_HUMAN	3	145,56	0,21	1293,03	938,79	1823,25	0,13	1113,44	1271,25	2121,63
F5H6I0_HUMAN	3	145,48	0,16	6008,99	6160,19	3276,34	0,03	3,42e+004	3,04e+004	8171,57
CUTA_HUMAN	2	144,29	0,11	1118,17	1472,66	290,83	0,46	2809,95	2671,52	1961,56
J3KP43_HUMAN	2	143,50	0,05	793,88	1059,16	2676,87	3,04e-003	1891,99	1805,52	2463,92
F5H6T1_HUMAN	2	143,29	0,12	62,63	126,89	391,59	0,41	643,84	939,81	1178,74
SC23B_HUMAN	4	142,28	0,39	1277,95	1229,99	1457,88	0,04	1910,65	2175,47	3619,88
APLP2_HUMAN	2	142,11	0,13	137,70	69,76	497,27	0,02	2270,03	2372,49	601,45
ECH1_HUMAN	2	141,92	0,26	52,40	43,94	467,22	0,05	1204,72	1193,75	1776,63
SDCB1_HUMAN	2	141,72	0,11	18,93	3,22	76,57	3,40e-003	813,67	944,61	103,54
IPSP_HUMAN	3	141,68	0,02	428,73	79,80	528,88	7,53e-003	4216,58	3807,18	602,68
GCSH_HUMAN	2	141,13	0,23	314,30	53,64	504,94	0,03	1564,35	1611,44	3924,97
H15_HUMAN	2	140,80	0,52	2289,05	2708,90	3612,36	0,10	4778,19	4450,19	6884,49
PRDX4_HUMAN	3 (1)	140,67	0,30	380,53	159,78	622,64	0,72	1322,71	1772,01	1676,13
AT1A2_HUMAN	2 (0)	140,58	---	---	---	---	---	---	---	---
ROA3_HUMAN	1	140,46	0,06	1082,41	1193,72	401,87	2,77e-004	284,37	322,53	6,82e-004
DDR1_HUMAN	3	140,01	0,25	3317,84	4545,41	3001,46	0,12	5913,82	5570,26	3273,87
LYPD3_HUMAN	2	139,36	0,03	1588,96	1250,03	259,17	6,41e-004	1,07e+004	1,08e+004	1586,26
E9PGI8_HUMAN	3	139,17	0,38	2343,92	1890,92	1612,01	0,36	1811,05	2029,81	2111,32
ANGL4_HUMAN	2	138,42	0,77	19,76	64,46	34,53	8,79e-003	1022,54	1420,28	21,01
PSB6_HUMAN	3	138,24	0,26	1,07e+004	8069,65	1,18e+004	0,01	8616,49	9382,97	1,95e+004
F5GYJ8_HUMAN	2 (1)	138,12	0,15	404,53	139,11	939,74	0,28	56,28	8,88	48,11
KI67_HUMAN	2	137,15	0,54	687,38	861,16	590,95	0,03	689,17	715,43	405,38
VTNC_HUMAN	2	136,56	1,15e-003	3837,09	4578,14	797,67	0,13	4486,62	4508,43	2551,19
PRS6B_HUMAN	3 (2)	136,27	0,01	321,03	253,87	1031,75	0,27	239,89	216,14	400,93

B4DDF9_HUMAN	2	135,09	0,36	152,22	98,98	554,63	0,51	1809,73	2029,01	1994,29
HCDH_HUMAN	2	134,77	6,91e-004	70,33	51,49	1304,97	0,19	636,74	1029,90	1280,06
NPNT_HUMAN	3 (2)	134,53	0,11	252,84	53,59	336,82	0,32	2320,29	3389,99	2942,47
RS5_HUMAN	1	133,71	0,03	791,67	685,33	3509,61	0,10	2413,93	2629,23	4175,80
B4E2C0_HUMAN	3	133,70	0,12	1630,55	1808,76	2520,36	0,03	5108,21	5028,10	1910,54
SIAS_HUMAN	3	133,54	0,12	460,24	61,41	1956,29	0,10	1629,28	1928,02	2131,89
NASP_HUMAN	3	133,51	0,41	3897,55	3714,41	4435,82	0,21	2416,96	2546,64	2847,13
DRG1_HUMAN	3	133,46	0,37	1329,97	1327,34	1626,88	0,29	1393,45	1627,17	1886,19
Q5JR95_HUMAN	2	133,46	0,05	39,20	7,68	1153,27	0,04	914,90	799,03	1392,77
SRSF3_HUMAN	3	133,21	0,04	6561,14	6490,01	1,05e+004	0,06	1,21e+004	1,29e+004	1,58e+004
VINC_HUMAN	2	132,94	0,16	238,21	80,04	938,30	0,45	752,66	627,96	1146,87
GRN_HUMAN	2	132,76	0,09	621,56	840,53	197,51	8,44e-003	1469,45	1644,84	589,22
TFPI1_HUMAN	2	132,30	0,31	5,68	15,13	19,77	0,03	282,43	429,85	110,16
B4DJV2_HUMAN	2	132,18	0,09	845,64	247,34	4025,20	0,11	4277,73	4169,40	5654,00
SYHC_HUMAN	3	131,85	0,12	1561,33	845,29	2651,00	0,08	2375,98	1921,58	3524,42
ADT2_HUMAN	3	130,99	0,11	758,53	490,44	3286,36	0,04	2231,81	2260,70	5459,23
FIS1_HUMAN	2	130,93	0,20	2,94	15,77	40,63	0,44	32,63	34,61	53,46
SF3B4_HUMAN	2	130,70	0,09	973,10	680,25	2255,07	0,01	825,75	575,17	1022,96
DKK1_HUMAN	2	130,62	0,02	267,47	435,05	11,82	8,11e-003	507,99	296,16	5,18
EFTU_HUMAN	2	129,93	0,36	2449,70	2688,71	3167,92	0,08	3103,14	3784,63	4086,29
HXK1_HUMAN	3	129,53	0,32	513,10	420,58	538,29	0,04	1055,12	937,40	1294,18
EFNA1_HUMAN	3	128,67	0,41	5326,45	4216,42	4850,34	0,01	1,37e+004	1,10e+004	5544,53
FUMH_HUMAN	2	127,37	0,41	451,65	349,83	620,03	5,28e-004	372,41	435,29	1096,63
CPSF6_HUMAN	2	126,97	0,22	1647,78	2116,05	1889,37	0,15	742,84	745,59	1160,98
SSBP_HUMAN	2	126,06	0,11	2913,47	2910,17	1710,77	0,09	1158,14	1309,24	1440,95
TFR1_HUMAN	3	125,78	0,19	231,41	366,31	173,89	0,14	444,74	532,45	730,03
H0YF32_HUMAN	2 (1)	125,50	0,14	77,74	9,35	824,33	0,19	190,84	160,54	380,46

PVRL2_HUMAN	2	125,38	0,13	766,51	1622,85	790,75	0,02	2549,43	2545,70	1388,51
A6NNI4_HUMAN	2	125,20	0,37	2061,57	1526,15	1424,12	7,89e-004	2,13e+004	2,05e+004	4617,00
GLRX3_HUMAN	1	125,10	0,28	907,39	943,24	1484,32	0,48	763,78	749,70	878,14
ERH_HUMAN	2	124,84	0,12	1346,24	535,30	2451,10	0,06	589,05	655,28	1902,93
IF4H_HUMAN	3	124,22	0,18	168,67	85,59	1169,06	0,34	289,73	417,47	456,82
PTMS_HUMAN	2	123,86	3,28e-003	3464,31	4304,32	174,12	0,10	653,69	620,44	141,03
6PGL_HUMAN	3	123,15	0,03	44,14	124,00	853,80	0,11	579,44	456,57	1444,77
C9J712_HUMAN	2	122,63	0,12	1098,67	929,30	1831,80	0,03	1067,59	1208,46	2528,54
SYG_HUMAN	2	122,49	0,11	189,04	58,33	2028,63	0,20	692,49	668,38	1064,79
PDXL2_HUMAN	1	122,08	0,33	4,00	7,64	2,01	0,05	357,67	395,14	22,44
K7EIV0_HUMAN	2	121,55	0,02	46,57	4,49	361,16	0,10	593,10	321,87	812,07
AMPM2_HUMAN	2	121,55	0,16	3,59	8,14	70,59	1,57e-003	1108,43	1032,54	334,25
DUT_HUMAN	2	120,82	0,60	368,70	367,42	472,27	0,12	17,35	10,05	59,95
K1967_HUMAN	2	120,75	0,23	402,20	134,67	718,55	0,50	392,61	287,49	284,88
1433F_HUMAN	2 (1)	120,27	0,37	19,24	14,66	157,70	0,87	439,31	408,97	401,29
MCM3_HUMAN	2	119,69	0,36	1728,42	984,71	922,92	0,36	50,26	37,13	48,40
EPIPL_HUMAN	2 (1)	116,92	0,16	282,43	57,94	850,55	0,49	560,20	513,47	930,89
GSHR_HUMAN	2	116,09	0,12	3080,70	2246,01	4055,46	0,04	3534,11	3581,62	6050,68
LG3BP_HUMAN	2	116,04	0,29	41,30	142,56	5,26	4,93e-003	1539,20	1865,63	290,62
PHS_HUMAN	2	116,03	0,28	2569,16	2170,99	2331,57	0,06	2567,43	2270,07	3136,78
ARPC3_HUMAN	2	115,69	0,10	1167,18	683,54	2096,50	0,01	1619,77	1159,69	2877,68
UB2V2_HUMAN	2	115,62	0,08	2841,87	3537,89	1414,56	0,19	1156,55	1368,41	1614,12
CREL2_HUMAN	2	114,51	0,13	53,68	14,40	65,01	0,03	291,25	204,57	501,53
CAPR1_HUMAN	2	114,11	0,03	365,50	571,95	1794,07	9,42e-003	591,92	716,32	1459,38
CPNE3_HUMAN	2	113,92	0,18	287,21	156,92	2295,25	0,60	1191,41	1414,78	1548,34
PSB4_HUMAN	2	113,52	0,06	3676,71	2445,46	6771,62	0,02	4079,65	3794,54	7605,75
MAP4_HUMAN	2	113,28	0,02	646,37	833,06	222,94	0,07	140,46	78,16	139,25

UK114_HUMAN	2	111,85	0,01	928,88	564,39	1856,17	0,51	1191,96	1302,72	1493,39
SPEE_HUMAN	3	111,74	5,50e-003	957,92	692,20	3034,11	0,01	1886,49	2272,72	3195,10
E9PLM6_HUMAN	2	110,84	0,02	1233,76	1175,33	107,53	0,01	2947,50	2912,15	408,26
TEBP_HUMAN	2	110,27	0,10	3179,20	1532,36	8056,14	0,11	5056,49	5451,65	6809,60
K7EM13_HUMAN	2	110,08	0,02	79,73	5,16	945,64	0,53	692,20	739,62	726,85
SEPT2_HUMAN	2	109,85	0,12	231,68	112,44	1088,86	0,21	506,11	649,41	306,40
MIF_HUMAN	3	109,81	0,04	9233,32	7199,97	1,90e+004	4,24e-003	1,68e+004	2,22e+004	3,23e+004
DESP_HUMAN	2 (1)	109,48	0,60	940,82	962,24	803,55	0,33	1193,48	1524,40	1491,57
AMPL_HUMAN	3 (2)	109,38	0,07	1470,03	1131,48	1675,77	0,14	1538,32	1581,51	2092,45
SMD1_HUMAN	1	109,18	0,20	1157,23	1428,63	3448,37	0,05	2007,44	1983,76	5877,81
GFAP_HUMAN	2	108,99	0,05	240,99	85,08	2892,77	0,43	3239,45	4917,77	8077,98
TADBP_HUMAN	2	108,94	0,02	26,83	2,50	314,84	0,42	112,91	244,00	278,79
RUVB1_HUMAN	2	108,93	0,05	24,29	2,03	582,38	0,20	245,01	203,09	318,81
BID_HUMAN	2	108,67	0,04	7,39	3,68	439,22	0,04	571,87	531,76	1082,76
CMBL_HUMAN	2	107,66	0,12	80,76	15,39	856,24	0,05	757,23	775,38	1528,06
PRDX5_HUMAN	2	107,59	0,33	245,92	281,13	659,73	0,30	683,25	329,65	404,40
DHISO_HUMAN	2	106,30	0,12	230,06	428,47	1742,75	2,83e-003	803,68	1022,19	3030,70
STC2_HUMAN	2	105,64	0,47	135,85	74,71	149,78	0,01	2728,62	2773,96	78,17
C9IZE4_HUMAN	2	105,54	0,13	3656,67	4156,11	3561,61	0,16	2597,36	2270,26	3425,79
QPCT_HUMAN	3	104,84	0,12	70,13	69,10	168,96	0,06	1059,10	924,85	294,51
E7ETB3_HUMAN	2	104,65	0,12	934,51	697,58	2241,72	0,05	2918,93	2875,98	4576,37
UBE2N_HUMAN	2	104,25	0,07	1538,43	1586,19	4349,14	0,02	2370,64	2419,95	6203,44
B4E283_HUMAN	2	103,94	3,18e-003	6,94	0,47	451,46	0,44	128,19	168,73	244,88
PP1B_HUMAN	2	103,52	0,04	8,52	0,00	418,26	0,17	110,91	168,82	390,03
NPM3_HUMAN	1	103,01	0,05	1119,24	1108,78	1436,09	0,23	2225,64	1255,68	1826,51
SCRN1_HUMAN	2	102,77	0,08	2746,77	2805,90	1742,25	0,26	1145,04	756,13	1007,43
H7C211_HUMAN	2	102,71	0,02	338,77	173,14	1648,07	0,02	850,42	975,70	2731,27

B4E1S6_HUMAN	2	102,33	0,11	690,43	427,84	918,09	4,33e-003	3985,53	4009,61	891,66
D6RHJ6_HUMAN	1	102,04	0,11	100,66	26,62	172,24	0,27	189,80	217,66	263,93
HNRDL_HUMAN	2 (1)	101,66	0,12	759,09	373,60	1489,75	0,70	1794,37	1981,43	1716,29
NRCAM_HUMAN	2	101,32	0,04	1,24e+004	1,33e+004	9887,66	0,05	8617,97	9649,63	1,20e+004
AAMP_HUMAN	2	101,31	0,32	1474,91	1184,56	1468,52	0,04	617,01	946,46	1876,93
RBM8A_HUMAN	1	100,85	0,11	19,80	45,96	0,19	0,01	370,30	360,32	773,91
LEG3_HUMAN	3	99,78	2,05e-003	3739,49	3505,44	6384,17	8,78e-004	7308,94	8798,92	1,35e+004
CSTF2_HUMAN	2	98,49	0,44	1255,22	1244,09	1140,74	0,07	514,31	535,68	908,75
F5H2Z3_HUMAN	2	98,24	0,02	1,86e+004	2,16e+004	1,07e+004	0,02	1881,02	1462,89	2312,26
H0Y8X4_HUMAN	2	98,24	0,10	502,13	224,70	1564,86	0,74	1,09e+004	1,09e+004	1,07e+004
F5HE57_HUMAN	1	97,92	0,42	0,00	0,00	7,82	0,52	41,96	26,94	84,47
AK1A1_HUMAN	2	97,71	0,12	3374,82	6344,41	1528,01	0,02	1227,34	1323,49	1680,14
H0YKT8_HUMAN	2	97,66	0,08	919,02	179,57	4693,60	0,02	5188,57	5165,85	1,14e+004
LAT1_HUMAN	1	96,94	0,66	369,95	4,67	86,75	0,26	410,12	522,56	664,91
DTD1_HUMAN	1	96,94	0,38	1270,73	1141,64	1482,78	0,13	222,31	201,95	481,61
PERI_HUMAN	2	96,50	0,12	6351,75	7674,48	1,07e+004	0,02	3825,38	3751,16	5383,61
NUMA1_HUMAN	2	96,44	0,59	160,70	170,93	207,57	0,70	148,69	144,63	138,46
ASPG_HUMAN	2	96,34	0,92	24,08	42,32	13,22	0,83	340,65	346,69	382,94
MCM2_HUMAN	2	96,33	0,08	550,59	386,28	751,70	0,55	124,24	163,39	179,66
CNPY3_HUMAN	1	96,08	0,22	0,00	0,00	84,29	0,28	549,25	612,45	764,01
CY1_HUMAN	1	95,66	0,19	51,31	11,31	262,29	0,12	280,60	368,22	795,36
HEXB_HUMAN	2	95,48	0,43	1108,01	561,37	705,59	0,67	3079,17	3255,67	3180,59
HNRH3_HUMAN	1	94,94	0,42	0,00	0,00	1,42	0,34	98,56	78,22	63,49
E9PD49_HUMAN	2	94,43	0,29	404,29	376,23	247,38	0,02	1349,30	1203,98	243,97
RS4X_HUMAN	2	94,31	0,04	1,00	4,99	824,49	0,05	639,11	694,77	1715,40
SYEP_HUMAN	3	94,26	0,56	2194,82	2382,58	1778,74	0,34	2002,69	1645,94	1941,73
CPSF5_HUMAN	2	94,15	0,05	113,37	102,26	851,32	0,05	112,89	200,79	1550,58

ERO1A_HUMAN	1	93,75	0,08	98,40	76,86	859,45	0,09	656,07	757,74	1041,74
BTF3_HUMAN	2	93,36	0,04	162,58	26,93	378,66	0,22	82,25	57,91	113,57
E9PGM1_HUMAN	2	93,25	0,24	35,97	39,03	475,35	0,09	238,92	218,30	625,80
IMA2_HUMAN	2	93,13	0,02	1492,14	1263,36	4233,27	0,01	1271,12	1369,45	2810,64
F2Z2W6_HUMAN	1	93,02	0,03	2610,68	3215,60	650,99	0,46	190,94	153,09	88,25
EIF3M_HUMAN	1	92,97	4,05e-004	0,00	0,00	278,10	0,67	90,43	109,41	241,65
PP14B_HUMAN	2	92,34	5,54e-003	2783,20	2921,12	948,72	0,05	672,90	815,98	1061,58
E7ETU3_HUMAN	2	92,25	0,05	1230,27	981,67	1955,55	0,12	2345,48	2411,72	1587,98
PP2AA_HUMAN	2	91,66	0,13	330,95	194,95	1010,34	0,10	270,74	214,93	412,79
EF1B_HUMAN	2	90,97	0,10	2,39e+004	3,25e+004	1,90e+004	0,06	8619,93	9427,91	1,13e+004
CYBP_HUMAN	2	90,94	0,08	738,82	299,17	2677,44	0,03	1899,27	2043,38	3734,85
QCR1_HUMAN	1	90,45	0,21	15,15	59,72	193,01	0,07	195,01	276,72	321,40
GLOD4_HUMAN	1	90,12	0,04	0,67	0,00	39,32	0,08	148,38	123,09	435,46
SYNC_HUMAN	2	89,80	0,14	764,43	372,39	1420,27	0,67	656,34	731,54	674,36
DYHC1_HUMAN	2	89,45	0,06	22,68	0,00	320,68	0,28	20,31	24,55	89,24
SYFB_HUMAN	2	89,12	0,42	1843,55	1921,87	2074,86	0,05	1747,60	1658,97	2307,97
PLAK_HUMAN	2 (1)	88,72	0,64	0,73	77,81	1,18	0,16	73,52	147,81	4,85
C9JIF9_HUMAN	1	88,41	0,60	433,70	306,97	388,77	0,63	343,60	177,24	248,28
RL3_HUMAN	1	87,76	0,22	12,22	47,26	127,35	0,10	139,01	168,04	302,67
K7EK06_HUMAN	2	87,70	0,08	944,75	642,60	1209,22	0,31	578,86	628,57	345,36
ENOPH_HUMAN	2	87,26	0,02	2,33	0,07	302,19	0,04	214,22	401,32	1177,25
TACD2_HUMAN	1	87,22	0,04	400,83	1319,63	83,24	1,05e-003	1605,46	1901,51	505,51
VASP_HUMAN	2	87,01	0,18	217,75	44,80	688,76	0,21	2046,19	2006,12	2366,56
ARRD1_HUMAN	2	86,28	0,59	207,94	245,52	317,79	0,25	24,12	17,03	74,96
RL1D1_HUMAN	1	86,28	0,26	73,67	1028,89	21,86	0,01	2188,11	1891,27	228,59
NAGK_HUMAN	2 (1)	86,17	0,51	46,26	31,06	80,55	0,15	138,21	88,74	14,44
SC61B_HUMAN	1	86,00	0,13	5,01	6,00	217,44	4,10e-003	216,02	256,42	605,25

ITI2_HUMAN	1	85,79	0,09	1063,46	735,10	25,44	0,07	726,93	778,81	212,00
PPM1G_HUMAN	2 (1)	85,35	0,11	31,95	9,16	280,12	0,06	110,47	71,95	181,93
HNRPL_HUMAN	2	85,34	0,09	297,04	252,24	736,46	0,24	934,20	1209,02	1008,19
PSB7_HUMAN	1	85,26	0,19	482,25	214,27	505,34	5,27e-003	3065,65	2531,98	5767,28
G3BP2_HUMAN	2	85,19	0,24	1246,20	1282,89	1118,37	8,76e-003	640,28	622,53	1527,67
IF1AX_HUMAN	2	84,90	0,10	76,45	69,22	933,39	0,47	465,98	364,04	757,17
BUB3_HUMAN	2	84,80	0,23	130,20	39,76	450,93	0,04	73,19	78,12	220,15
CF211_HUMAN	1	84,78	0,14	75,00	42,14	596,77	0,67	660,35	722,13	738,37
TCP4_HUMAN	1	84,38	0,74	2717,96	2675,24	3035,59	0,08	4292,04	4028,59	2559,87
AAAT_HUMAN	2	84,35	0,16	1022,43	557,99	2004,64	0,24	5449,00	5698,22	4396,65
PSME1_HUMAN	2	83,87	0,10	465,42	203,25	1863,89	0,52	1656,82	1851,94	2232,72
E7EWS7_HUMAN	2	83,52	0,04	78,43	103,21	894,58	0,05	542,15	726,90	2423,59
CYC_HUMAN	2	83,48	0,10	196,20	43,60	568,31	0,06	435,12	404,76	887,45
CD276_HUMAN	1	83,32	0,40	181,62	90,54	244,93	8,51e-003	2987,52	2380,03	430,44
RAB1A_HUMAN	2 (1)	83,11	0,11	434,74	212,97	1042,29	0,10	626,56	611,98	837,63
FKBP2_HUMAN	2	82,91	0,19	517,03	421,51	1345,94	0,06	2096,41	2127,05	3596,46
RGAP1_HUMAN	1 (0)	82,79	---	---	---	---	---	---	---	---
GSTO1_HUMAN	2	82,70	0,28	2192,79	1864,25	2500,08	8,67e-003	2449,38	2514,32	4226,56
G3BP1_HUMAN	1	82,55	0,24	63,50	65,20	690,76	0,37	47,79	22,45	70,75
PSDE_HUMAN	2	82,53	7,89e-003	0,37	1,06	181,68	0,05	63,93	100,20	246,55
TBCA_HUMAN	2	82,17	0,04	1,04e+004	7957,22	4717,27	0,18	4444,41	3775,74	5290,51
E9PN11_HUMAN	1	82,12	5,90e-004	0,00	0,00	108,65	0,24	9,38	12,61	155,73
KATL2_HUMAN	2 (0)	81,91	---	---	---	---	---	---	---	---
HOYCP8_HUMAN	2	81,30	0,07	24,94	10,98	126,37	0,01	612,40	649,90	1772,73
A8MUA9_HUMAN	1	80,39	0,14	3595,59	1220,24	6288,41	0,03	4732,04	4647,73	1,05e+004
K7EMH1_HUMAN	1	79,96	0,24	1265,71	1039,43	2139,85	0,18	972,33	1255,56	1410,87
NQO1_HUMAN	2	79,66	0,07	2225,32	2298,33	4999,24	3,23e-003	3277,30	4142,40	7073,20

UBQL1_HUMAN	2	78,88	0,37	3988,66	3546,51	4444,41	0,02	8328,16	8545,77	1,31e+004
E9PLG2_HUMAN	2	78,55	0,02	67,43	191,04	457,04	0,15	316,01	351,82	730,36
EIF3C_HUMAN	1 (0)	78,55	---	---	---	---	---	---	---	---
BZW1_HUMAN	1	77,80	0,11	471,70	247,09	2356,23	0,13	2033,88	1632,55	1357,82
PDXK_HUMAN	1	77,34	0,39	2988,61	2692,89	3285,77	0,80	1377,77	1500,93	1467,35
SMC1A_HUMAN	2 (1)	77,10	0,64	450,65	339,13	427,72	0,26	375,73	361,32	488,78
D6RER5_HUMAN	2	76,88	0,07	1060,32	840,46	2036,12	0,04	1233,79	1290,63	870,40
S10A8_HUMAN	1	76,43	0,11	3354,30	3650,74	723,25	0,71	2253,37	2743,47	2305,46
UBQL2_HUMAN	1	75,78	3,41e-004	0,01	0,09	72,24	0,16	4,92	12,91	66,33
AMGO2_HUMAN	1	75,77	0,42	0,00	0,00	0,49	0,17	380,84	115,25	1,73
DEST_HUMAN	2 (1)	75,69	0,21	3523,00	2289,50	4347,77	0,29	4611,07	4887,77	5223,31
CATL1_HUMAN	1	75,63	0,28	315,14	773,11	158,28	6,82e-003	960,37	939,54	295,21
PAK2_HUMAN	2	75,58	0,01	1838,01	1728,29	1041,31	7,85e-004	979,95	769,65	466,63
RAC3_HUMAN	2	75,47	3,94e-003	1498,17	1323,09	2908,76	0,08	2594,42	2887,24	2113,81
HN1_HUMAN	1	75,33	0,05	1810,37	2042,75	275,98	0,20	302,82	267,99	476,50
ECHA_HUMAN	2	75,25	0,17	49,37	80,39	176,79	0,39	1073,01	1107,00	1321,48
H0YLU7_HUMAN	2	75,21	0,07	13,99	1,99	420,36	0,35	294,90	529,47	859,89
PA1B2_HUMAN	2	75,10	0,18	291,32	95,62	791,61	0,09	21,92	1,99	19,73
LAD1_HUMAN	1	75,10	0,81	407,35	375,10	357,40	0,32	928,46	1091,71	631,34
E5RGU4_HUMAN	1	75,09	0,47	0,00	3,39	0,17	0,62	18,92	10,84	10,83
COPB_HUMAN	1	74,62	0,56	0,70	0,00	0,25	0,21	0,85	7,94	246,17
K7EJ88_HUMAN	1	74,54	0,03	181,59	64,66	203,41	0,04	116,46	149,37	37,03
RS18_HUMAN	2	74,49	0,05	1,84	5,98	1149,84	0,03	1497,47	1475,65	2334,25
E5RH77_HUMAN	1	74,36	0,04	478,40	432,19	1401,21	0,32	1312,62	1103,68	966,84
KINH_HUMAN	1	74,16	0,07	16,93	0,00	115,48	0,18	3,62	0,39	0,00
U2AF2_HUMAN	1	74,01	0,17	270,80	153,96	988,94	0,23	442,98	501,16	1839,95
MPCP_HUMAN	1	73,85	0,30	419,12	486,18	1035,01	0,18	1336,49	1466,56	2008,69

STC1_HUMAN	1	73,62	1,99e-003	2022,66	2675,95	493,16	7,58e-003	1949,16	1100,94	295,99
ELAV1_HUMAN	1 (0)	73,42	---	---	---	---	---	---	---	---
HMGB2_HUMAN	2 (1)	73,41	0,68	1983,38	1567,19	1772,76	0,09	1284,12	1256,79	1957,68
VDAC1_HUMAN	1	73,30	0,16	24,57	43,77	521,85	0,17	451,44	793,79	1416,05
RS7_HUMAN	2	72,98	0,05	62,66	14,07	897,88	0,47	916,41	1000,34	769,38
PSME3_HUMAN	2	72,97	0,05	62,76	1,86	589,93	4,42e-003	621,67	447,04	800,32
H0YCY6_HUMAN	1	72,60	0,05	55,44	45,40	490,71	0,05	307,14	262,12	465,96
IF6_HUMAN	2	72,03	0,08	3200,20	3527,73	4385,88	0,21	3464,95	3670,01	4320,24
ACON_HUMAN	2 (1)	71,98	0,91	748,47	741,89	748,14	0,27	755,55	609,09	560,70
LOXL2_HUMAN	1	71,94	0,41	35,01	13,66	8,41	3,80e-003	771,04	1033,91	100,14
SPRE_HUMAN	1	71,88	0,05	12,95	12,05	517,85	0,02	224,90	224,95	701,89
DDX3X_HUMAN	1	71,88	0,27	408,27	914,74	506,69	0,33	178,29	133,61	215,32
FUS_HUMAN	2	71,68	0,22	2942,05	1270,18	2883,79	0,03	4042,86	3916,70	5582,41
NT5D1_HUMAN	2	71,53	0,32	188,49	43,82	1332,84	0,14	232,92	588,29	1376,74
A8MVD5_HUMAN	1	71,33	0,28	591,65	746,64	749,89	0,08	348,16	379,89	511,64
IMDH2_HUMAN	1	71,20	0,16	263,69	121,00	1032,31	0,06	325,98	159,29	338,95
PUR2_HUMAN	1	70,81	0,11	649,67	662,05	286,49	0,40	175,64	221,58	137,20
RUVB2_HUMAN	1	70,76	0,01	123,41	115,23	905,54	0,07	356,28	410,85	562,30
VPS35_HUMAN	1	70,39	2,61e-004	0,00	0,00	148,34	0,20	73,90	38,01	122,11
AL9A1_HUMAN	1	70,38	0,36	1387,93	1529,70	1088,84	0,37	504,21	579,45	478,65
RL29_HUMAN	1	70,27	0,47	18,33	28,38	23,24	0,81	34,34	29,88	37,46
IDHP_HUMAN	2 (1)	70,22	0,27	0,00	0,00	0,51	0,33	420,59	773,64	434,40
E7ETI0_HUMAN	2	69,67	0,08	388,82	349,64	1337,60	0,05	876,68	1051,29	2476,53
FIBG_HUMAN	1	69,47	0,25	26,48	14,33	39,25	0,57	23,11	22,93	30,24
PXDC2_HUMAN	1	69,26	0,15	218,15	234,91	32,72	0,06	69,48	36,39	3,65
Q5T446_HUMAN	1	69,20	0,72	92,98	52,85	63,03	0,03	18,07	16,08	109,81
NIT2_HUMAN	1	68,86	0,21	88,60	116,33	637,09	0,26	462,50	557,82	614,63

U2AF1_HUMAN	2	68,62	0,43	144,06	33,96	170,13	0,12	434,92	419,20	1331,18
SORL_HUMAN	1	68,62	0,03	9,72	1,73	35,69	0,04	281,07	262,31	34,99
PGM1_HUMAN	1	68,55	0,14	427,45	152,50	1066,77	0,01	934,22	932,07	1874,48
SET_HUMAN	1	67,99	0,15	354,23	8,35	5636,52	0,04	4462,70	5197,64	1,21e+004
E5RI99_HUMAN	1	67,89	0,07	657,36	309,69	1662,44	0,03	998,47	1168,03	1609,28
K7ERQ8_HUMAN	1	67,77	0,02	3,64	0,15	154,09	0,05	155,27	185,05	537,98
UB2D2_HUMAN	1	67,74	0,17	333,31	158,43	1597,29	0,09	803,91	752,36	2052,98
CSN2_HUMAN	1	67,44	0,08	18,33	0,33	224,15	0,25	102,65	104,01	383,63
C9JJE2_HUMAN	1	67,43	0,04	709,34	556,42	321,86	0,16	277,45	112,85	106,86
LRRF1_HUMAN	1	67,36	8,98e-003	39,62	2,12	129,23	0,31	124,32	94,70	110,06
E5RJG7_HUMAN	1	67,31	0,02	4,85	1,05	133,07	0,02	42,28	4,61	0,35
MA1A1_HUMAN	2	67,18	0,04	7,60	6,42	174,22	0,16	1246,06	1378,98	655,20
FKB1A_HUMAN	1	67,06	0,32	4516,91	4551,24	3901,14	0,18	4301,98	5047,06	6406,48
HTRA1_HUMAN	1	66,80	0,26	21,41	11,28	70,80	6,89e-003	194,95	181,42	16,82
A8K3Z3_HUMAN	2	66,76	0,06	191,73	90,28	999,04	0,11	329,83	421,00	818,14
B4DZM8_HUMAN	1	66,16	0,02	2,55	0,00	87,19	0,72	112,95	129,73	139,45
PCTL_HUMAN	2	65,75	0,30	6212,22	6072,08	5412,12	0,72	4477,66	4215,00	4164,34
MOB1B_HUMAN	2	65,71	0,08	1327,33	1370,03	1835,97	0,15	1720,55	1498,38	1953,38
B4DWP2_HUMAN	1	65,42	0,17	2,96	0,29	28,34	0,02	918,80	960,61	43,30
PYRG2_HUMAN	1	65,37	0,31	68,99	60,03	209,40	0,09	384,88	372,63	624,40
PA1B3_HUMAN	1	65,31	0,23	497,96	519,54	666,72	0,66	118,68	86,99	100,66
LASP1_HUMAN	1	65,21	0,05	222,52	258,67	45,46	0,43	1,32	1,90	0,19
B4E351_HUMAN	2	65,08	0,56	13,33	32,30	27,65	0,03	58,76	36,61	3,79
RL6_HUMAN	1	65,07	0,37	45,17	53,63	0,00	0,02	0,00	2,42	104,59
E7EP00_HUMAN	2	64,22	0,04	0,00	0,00	3,53	0,17	28,98	12,32	486,11
SP16H_HUMAN	1	64,17	0,50	0,00	11,51	0,41	0,19	9,38	9,17	242,14
THUM1_HUMAN	1	64,13	0,08	353,41	509,54	249,25	0,61	70,96	42,42	64,04

ADRM1_HUMAN	1	63,84	0,13	849,16	1038,21	496,66	0,39	183,77	102,50	137,06
WIBG_HUMAN	1	63,82	0,31	88,77	5,39	0,00	1,00	0,00	0,00	0,00
B4GT1_HUMAN	2	63,80	0,10	58,13	0,00	142,22	0,09	1172,18	1233,08	501,51
PIN4_HUMAN	1	63,68	0,44	1592,56	1407,95	1248,07	0,17	1195,04	1506,45	1636,26
PIR_HUMAN	1	63,59	0,03	81,61	1,46	932,14	0,10	544,67	631,24	1468,90
CO3_HUMAN	1	63,48	0,08	300,66	473,71	124,32	0,04	72,82	26,07	8,41
H6ST1_HUMAN	1	63,45	0,45	292,17	215,54	159,81	0,02	460,36	473,81	93,40
HMCS1_HUMAN	1	63,30	0,02	163,98	151,38	911,90	0,27	649,37	613,45	816,70
GUAA_HUMAN	1	63,29	0,22	251,05	13,06	747,38	0,90	74,39	98,62	87,34
B4DXG7_HUMAN	1	62,84	0,01	280,54	395,29	99,37	8,75e-003	1288,20	1176,91	291,84
GFPT1_HUMAN	1	62,67	0,27	22,36	53,90	120,46	0,63	121,72	139,46	186,24
ARPC2_HUMAN	1	62,66	0,17	19,47	269,86	310,92	0,39	239,08	300,76	233,87
E9PJD9_HUMAN	1	62,47	0,23	81,96	23,44	912,20	0,17	401,12	507,12	1316,79
TOP2B_HUMAN	1	62,35	0,51	24,78	27,61	40,00	0,61	13,55	4,47	40,46
GLCM_HUMAN	1	62,24	0,08	64,46	6,66	243,59	0,15	422,14	362,94	535,35
ARHGQ_HUMAN	2 (0)	62,17	---	---	---	---	---	---	---	---
THSD4_HUMAN	1	62,15	0,09	20,34	41,81	0,19	0,01	103,92	54,58	1,12
HAS1_HUMAN	1 (0)	61,72	---	---	---	---	---	---	---	---
HOYJE9_HUMAN	1	61,57	7,42e-004	0,00	0,00	294,68	0,03	30,66	0,21	0,06
SH3L1_HUMAN	1	61,56	0,07	271,10	388,74	11,67	0,14	116,88	261,86	127,34
DDX6_HUMAN	1	61,48	0,08	14,49	9,54	255,55	0,81	24,42	22,06	72,09
B4DNJ6_HUMAN	2	61,32	0,21	1832,94	1429,91	2474,90	0,47	1287,41	1125,56	1090,60
CLIC4_HUMAN	2	61,09	0,11	2627,92	2872,90	1565,73	0,01	3980,92	2987,23	1687,78
ETFB_HUMAN	1	61,07	0,23	2672,41	1925,66	1927,78	0,41	2070,68	1949,44	1799,03
SAP3_HUMAN	2	61,07	0,79	149,86	145,72	144,19	0,03	4723,28	4020,73	1475,49
SH3L3_HUMAN	1	60,94	0,13	1666,96	1653,98	1297,31	0,03	708,36	352,01	333,36
SEM3F_HUMAN	1	60,80	0,15	339,99	439,06	135,89	3,15e-003	1689,50	1708,48	298,37

QORX_HUMAN	1	59,58	1,00	0,00	0,00	0,00	3,80e-003	0,00	0,00	130,27
EIF3G_HUMAN	1	59,35	0,12	331,88	444,87	158,30	0,05	39,45	52,32	112,51
CAN1_HUMAN	1	59,14	0,04	466,19	240,95	710,48	0,15	1155,52	1071,33	1334,68
S10AD_HUMAN	1	58,76	0,47	0,44	0,00	6,39	0,32	123,75	87,67	221,07
HP1B3_HUMAN	1	58,76	0,04	675,00	1307,11	112,77	0,01	252,25	232,82	777,11
RUXGL_HUMAN	1	58,65	2,47e-004	0,00	0,00	187,33	0,05	311,99	195,18	899,42
TXD12_HUMAN	1	58,05	0,20	95,87	8,21	300,08	0,02	615,83	703,23	1242,05
F5H4R6_HUMAN	2	58,02	5,54e-003	0,73	0,00	616,98	0,66	458,64	374,19	509,57
BASI_HUMAN	1	57,91	0,14	1057,69	1241,44	718,86	0,04	2660,60	3661,75	1558,43
EIF3I_HUMAN	1	57,68	0,34	605,40	775,23	992,87	0,62	440,67	444,18	505,87
SORT_HUMAN	1	57,67	0,12	98,33	7,22	107,09	0,06	1447,36	1688,72	693,20
MGN_HUMAN	1	56,80	0,03	5,16	11,41	1410,26	0,05	471,42	544,78	2254,23
GBB2_HUMAN	1	56,77	0,32	47,12	15,79	247,05	0,18	699,23	752,77	568,86
LHPP_HUMAN	1	56,72	0,08	163,20	46,11	1123,16	0,15	758,22	671,24	1182,60
MT1G_HUMAN	1	56,19	0,17	2322,86	2848,18	226,62	0,22	29,78	0,50	4,10
B3KSH1_HUMAN	1	56,00	0,05	2,57	2,27	278,93	0,24	66,66	81,52	170,99
ARPC5_HUMAN	1	55,92	0,22	167,81	41,38	682,92	0,09	938,36	1041,48	1312,31
SYYC_HUMAN	1	55,92	0,91	2080,25	2074,61	2093,50	0,61	1190,40	1119,60	1075,73
A6NKV8_HUMAN	1	55,91	0,42	0,24	0,00	0,12	0,42	0,07	0,00	0,00
PARP1_HUMAN	1	55,68	0,23	95,37	7,21	371,10	0,59	36,81	68,33	78,00
MER34_HUMAN	1	55,43	0,18	344,33	276,49	89,51	0,05	732,36	848,93	304,68
H0Y8L3_HUMAN	1	55,28	0,02	0,00	0,00	6,35	0,41	92,24	193,74	96,29
H10_HUMAN	1	54,71	0,35	125,99	207,60	339,69	0,60	186,37	33,27	53,62
HAT1_HUMAN	1	54,29	0,42	0,00	0,00	2,57	0,42	0,00	0,00	20,45
DNS2A_HUMAN	1	54,23	0,02	7057,09	7115,66	4292,37	0,07	5128,58	5642,06	6154,01
DDX1_HUMAN	1	54,11	0,02	2,49	0,00	331,68	0,36	10,58	5,11	14,47
H7BXW3_HUMAN	1	54,02	0,08	399,27	515,96	247,42	0,46	27,56	17,32	32,88

MCM6_HUMAN	1	53,78	0,50	990,71	917,38	920,01	0,29	571,00	676,93	514,77
RL14_HUMAN	1	53,59	0,11	206,88	225,65	1285,79	0,03	967,78	872,33	1683,70
BIEA_HUMAN	1	53,55	0,05	299,00	184,65	508,53	0,48	465,24	485,44	440,77
WDR1_HUMAN	1	53,32	0,15	128,16	5,56	1366,33	0,38	1590,18	1626,99	1736,13
IPO4_HUMAN	2	53,21	0,01	0,00	0,00	37,85	3,35e-003	0,00	0,01	36,16
RS16_HUMAN	1	53,07	2,93e-003	0,00	0,00	568,37	0,66	244,31	380,54	819,90
RL23_HUMAN	1	53,02	0,24	18,34	97,85	173,61	0,29	96,39	124,79	452,33
B4DP38_HUMAN	1	52,75	0,05	26,03	0,00	566,74	2,05e-003	470,92	520,48	1320,98
E5RI56_HUMAN	1	52,65	0,40	3,51	13,24	0,00	0,22	46,97	0,00	106,82
SPB6_HUMAN	1	52,41	0,08	82,34	9,45	829,72	0,33	714,65	783,94	1040,40
UGPA_HUMAN	1	52,24	0,13	87,83	15,06	669,55	0,04	818,27	956,73	2315,32
I3L3E9_HUMAN	1	52,07	2,07e-005	0,00	0,00	300,55	0,05	4,13	11,99	257,95
ITM2B_HUMAN	1	52,00	0,07	201,13	304,47	239,25	0,16	403,39	513,80	241,81
PCKGM_HUMAN	1	51,66	0,07	10,91	0,00	195,37	0,21	0,00	0,00	10,00
RAP1A_HUMAN	1	51,52	0,13	7,22	2,10	61,14	0,19	287,66	268,93	151,02
MARCS_HUMAN	1	51,25	0,30	11,29	8,57	53,96	0,22	75,03	150,90	133,78
GRHPR_HUMAN	1	51,11	0,09	390,48	164,01	660,80	0,48	375,57	374,18	460,23
B4DGP8_HUMAN	1	51,06	0,17	1122,37	93,87	3626,27	0,05	2012,10	2237,70	4539,33
B3KXI2_HUMAN	1	50,95	0,12	50,33	32,31	96,80	0,33	54,58	3,23	5,05
A8MZ56_HUMAN	1	50,75	0,02	0,66	13,33	104,85	0,64	113,26	91,03	114,78
BST2_HUMAN	1	50,73	0,34	84,65	64,65	61,00	0,09	326,01	219,32	106,67
CAH11_HUMAN	1	50,71	0,51	2,43	0,81	6,30	0,78	19,02	25,88	15,45
E9PCP3_HUMAN	1	50,66	0,15	215,24	140,77	81,38	0,33	4,91	1,36	0,26
PTK7_HUMAN	1	50,44	0,14	3,47	0,87	23,99	0,01	390,88	337,21	14,06
NR2CA_HUMAN	1	50,17	0,18	1019,52	994,09	724,65	9,17e-004	404,23	451,84	88,77
SMD2_HUMAN	1	50,08	0,25	4643,23	4251,70	5491,70	0,09	4107,51	4322,62	5296,10
H7C086_HUMAN	1	49,62	0,22	2111,38	2973,88	1819,91	0,10	1420,54	1486,41	1669,63

RL13_HUMAN	1	49,60	0,14	5,74	19,05	416,94	0,11	290,13	354,65	697,77
Q5JR04_HUMAN	1	49,56	0,11	90,86	68,81	645,86	0,08	270,48	143,42	511,17
ACINU_HUMAN	1	49,52	0,35	63,20	113,42	121,18	0,28	66,41	42,91	68,02
TOM70_HUMAN	1	49,41	0,04	13,24	0,00	611,01	0,50	610,16	719,02	790,31
RS9_HUMAN	1	49,19	2,30e-003	0,00	0,00	872,10	0,46	587,98	713,94	1123,74
RB11A_HUMAN	1	49,18	0,58	1,44	5,41	31,72	0,53	85,43	80,57	191,86
RHG01_HUMAN	1	49,09	0,05	21,65	0,00	459,12	4,04e-003	421,24	441,39	1202,04
C9JLV5_HUMAN	1	49,09	0,14	65,44	15,00	526,74	0,05	220,56	220,47	411,77
LA_HUMAN	1	49,09	0,12	816,06	301,90	2279,99	0,27	1126,76	1192,86	1073,76
NIBL1_HUMAN	1	48,61	0,18	62,46	44,26	391,31	0,68	590,67	529,06	436,63
SYCC_HUMAN	1	48,50	0,39	39,83	16,93	59,34	0,40	192,35	82,94	113,28
RS11_HUMAN	1	47,92	0,17	402,75	369,58	930,59	0,08	839,50	1044,26	1727,52
PFD2_HUMAN	1	47,87	0,14	980,60	844,20	477,82	0,48	159,23	129,16	45,99
G8JLC5_HUMAN	1 (0)	47,57	---	---	---	---	---	---	---	---
C9JP35_HUMAN	1	47,54	0,37	1256,21	1556,70	1522,33	0,72	1457,68	1259,15	1394,83
B0QYN7_HUMAN	1	47,49	0,04	819,00	255,42	2021,63	7,30e-003	1395,15	1606,09	3220,40
ACTY_HUMAN	1	47,15	0,07	7647,81	9089,00	5021,97	0,29	3928,65	4154,76	4819,79
F2Z2U8_HUMAN	1	46,94	0,34	3,09	6,76	23,74	0,45	94,66	47,92	55,72
CSTF3_HUMAN	1	46,66	4,16e-004	0,00	0,00	213,04	0,08	109,17	303,92	603,83
H0YAR3_HUMAN	1 (0)	46,65	---	---	---	---	---	---	---	---
D6RF35_HUMAN	1	46,54	0,16	683,85	170,30	64,48	0,49	14,46	17,07	27,70
PLMN_HUMAN	1	46,47	0,12	5325,69	3679,08	2053,67	0,05	2098,78	1142,93	1587,91
ODO2_HUMAN	1	46,37	0,18	0,00	0,00	26,84	0,05	38,10	58,21	219,41
STON2_HUMAN	1	46,22	1,00	0,00	0,00	0,00	0,17	8,55	27,48	534,62
NCAM2_HUMAN	1	46,13	0,16	85,77	97,92	63,93	5,76e-003	615,13	575,83	84,50
SCOT1_HUMAN	1	45,94	0,18	2050,85	2246,49	1114,34	0,19	1729,96	1999,95	1492,62
PEPL_HUMAN	1	45,90	---	167,35	41,94	0,00	0,42	---	---	---

E7EWZ6_HUMAN	1	45,06	0,03	1161,96	1182,08	170,23	5,70e-006	667,03	515,77	60,87
CSN8_HUMAN	1 (0)	44,88	---	---	---	---	---	---	---	---
GABT_HUMAN	1	44,80	0,09	3,99	0,00	22,00	0,71	91,58	116,41	72,94
PCDH1_HUMAN	1	44,60	0,11	209,35	132,64	406,86	0,15	983,99	901,06	542,65
BROX_HUMAN	1	44,60	0,02	1,65	1,17	111,06	0,59	565,70	496,03	672,15
DLDH_HUMAN	1	44,54	0,75	37,42	14,05	22,08	0,58	258,34	374,86	412,10
FUCO_HUMAN	1	44,50	0,13	9,98	34,22	43,81	0,35	406,38	267,28	282,04
MGP_HUMAN	1	44,33	0,43	942,03	698,90	807,11	0,47	648,53	585,28	620,69
PSD11_HUMAN	1	44,20	6,93e-003	708,22	451,67	1174,54	0,46	953,72	850,46	712,55
H7C3P4_HUMAN	1	44,03	0,63	3105,89	4110,48	3297,67	0,05	3222,26	3212,40	4591,85
CYFP1_HUMAN	1	44,02	0,43	0,00	0,01	0,74	0,22	159,29	81,28	52,42
MATR3_HUMAN	1	43,70	0,23	552,44	791,94	1158,36	0,69	736,31	769,32	846,99
ARC1B_HUMAN	1	43,65	0,02	3,05	0,00	526,18	0,21	615,82	682,39	874,34
PLSI_HUMAN	1	43,53	0,30	2,52	0,56	4,34	0,29	60,88	24,21	32,39
AHS1_HUMAN	1	43,47	0,06	799,51	455,63	2512,58	0,07	1426,83	1235,32	3189,36
GLGB_HUMAN	1	43,39	0,42	0,00	2,17	0,00	1,00	0,00	0,00	0,00
SCF_HUMAN	1	43,02	0,18	45,05	2,96	24,67	0,17	1155,42	1197,69	89,51
SEM4C_HUMAN	1	42,91	0,31	3,60	4,67	2,24	0,03	121,62	127,92	17,57
INO1_HUMAN	1	42,45	0,42	0,00	0,00	2,50	0,42	1,81e-003	0,00	3,71
VAPA_HUMAN	1	42,14	0,23	260,80	311,08	646,55	0,68	1001,19	1276,34	1277,00
LARP1_HUMAN	1	41,95	0,17	206,92	90,31	56,87	0,03	42,75	33,34	2,02
E7ETU9_HUMAN	1 (0)	41,93	---	---	---	---	---	---	---	---
F5GZN3_HUMAN	1	41,85	1,00	0,00	0,00	0,00	0,05	247,49	258,68	17,04
DHX15_HUMAN	1	41,68	0,11	396,76	244,18	906,41	0,21	290,39	227,53	561,98
ELOC_HUMAN	1	41,66	0,42	0,00	0,00	2,05	0,54	63,56	54,87	53,74
PYR1_HUMAN	1 (0)	41,42	---	---	---	---	---	---	---	---
A6NM98_HUMAN	1	41,34	0,81	108,66	87,94	137,63	0,13	54,79	17,79	18,90

PURB_HUMAN	1	41,28	0,22	368,66	365,35	280,34	0,21	127,77	68,59	115,39
F5H2A7_HUMAN	1	41,16	0,46	4,52	0,00	3,78	0,14	74,95	19,68	184,08
EHD4_HUMAN	1	41,09	0,16	189,17	607,49	620,00	0,17	1021,78	595,30	700,36
PPAC_HUMAN	1	41,02	0,02	3049,00	3411,19	1734,56	0,48	2233,18	2127,18	2157,20
AMD_HUMAN	1	40,95	0,42	0,00	0,00	0,84	0,05	575,78	517,15	6,25
RS12_HUMAN	1	40,95	0,33	7750,34	6686,81	6746,85	0,02	5923,00	6687,73	7916,00
DHE3_HUMAN	1	40,90	0,41	1277,32	852,36	1410,04	0,30	1749,87	1930,27	1706,73
SRC8_HUMAN	1	40,85	0,66	295,59	381,42	302,45	0,59	304,21	330,89	202,73
NOE1_HUMAN	1	40,61	0,36	6,78	6,86e-003	1,47	0,01	707,95	637,08	25,30
K7ER96_HUMAN	1	40,48	0,30	229,61	63,65	1154,81	0,02	369,26	380,09	884,63
F8VY04_HUMAN	1	40,25	0,33	296,53	185,24	383,13	0,22	448,32	328,44	112,72
F8WCB8_HUMAN	1 (0)	39,94	---	---	---	---	---	---	---	---
THOP1_HUMAN	1	39,91	0,06	29,11	0,00	368,64	0,08	270,75	468,78	615,74
Q5T8U2_HUMAN	1	39,66	0,11	212,85	66,48	1071,09	0,01	730,75	845,25	1653,26
DENR_HUMAN	1	39,62	0,72	1724,19	1676,00	2013,71	0,50	1606,12	1685,32	1336,96
J3KNP4_HUMAN	1	39,60	0,15	21,56	86,00	95,83	0,04	264,80	233,10	38,63
A6NEM2_HUMAN	1	39,51	0,14	359,50	219,95	739,28	0,12	1203,67	790,02	1189,04
K4DIA0_HUMAN	1	39,46	0,31	1517,33	929,19	1157,35	0,16	3116,60	3651,59	2763,26
B0QZ18_HUMAN	1	39,41	0,05	13,20	0,00	161,72	0,16	208,20	146,09	366,98
Q14UF6_HUMAN	1	39,27	0,27	429,89	363,80	327,20	0,04	429,76	246,33	133,56
ASNS_HUMAN	1	39,15	0,03	4,92	0,00	392,05	0,20	201,83	144,78	204,22
BSSP4_HUMAN	1	39,00	0,07	471,13	508,75	169,65	8,37e-003	1362,06	1435,12	390,24
E9PR17_HUMAN	1	38,91	9,21e-004	663,88	876,16	388,29	7,34e-003	1908,56	1919,87	868,40
PSB2_HUMAN	1	38,89	0,03	4,60	0,00	407,63	0,27	306,90	223,91	768,78
B4DLJ9_HUMAN	1	38,84	0,23	9,58	6,65	3,23	0,16	15,03	0,00	8,12
VTA1_HUMAN	1	38,64	0,06	43,58	0,00	539,30	0,22	497,82	661,82	741,65
GBRL2_HUMAN	1	38,52	3,77e-003	0,00	0,00	19,37	0,11	2,97	4,39	153,00

B0V043_HUMAN	1	38,46	0,03	186,45	335,92	115,55	0,56	281,66	337,49	351,07
H7BZ11_HUMAN	1	38,45	0,02	23,89	13,99	129,51	0,36	70,67	59,08	100,21
RD23B_HUMAN	1	38,35	0,08	51,78	1,24	1123,03	0,07	334,67	532,11	1408,91
H0Y8I3_HUMAN	1	38,09	0,14	7,67	22,24	342,24	0,33	59,28	0,67	4,29
SNUT2_HUMAN	1	37,85	0,03	1054,42	706,39	1449,85	0,14	1166,80	1232,27	1664,78
SMD3_HUMAN	1	37,67	0,04	0,35	0,00	38,39	0,79	289,73	339,48	933,88
TMED4_HUMAN	1	37,59	9,98e-003	0,00	0,00	15,84	0,27	46,20	99,23	256,42
T2FA_HUMAN	1	37,45	0,19	144,16	162,50	211,44	7,96e-003	43,00	23,97	120,92
FETA_HUMAN	1	37,38	0,10	1012,52	1245,58	452,72	0,12	675,49	787,77	1091,15
FSCN1_HUMAN	1	37,32	0,05	658,82	411,67	1733,27	0,57	793,45	848,51	735,14
DIAP1_HUMAN	1	37,26	1,61e-003	0,00	0,00	172,81	0,34	38,42	99,56	167,92
CN166_HUMAN	1	37,24	0,35	271,86	138,51	394,47	0,04	18,19	42,12	105,36
NAV1_HUMAN	1 (0)	37,01	---	---	---	---	---	---	---	---
CCD71_HUMAN	1	36,64	0,86	9837,97	9351,41	9542,63	0,09	3937,93	3752,16	3349,54
LRC47_HUMAN	1	36,59	0,33	951,98	571,94	596,75	0,29	354,46	517,18	321,14
DNJA2_HUMAN	1	36,47	0,04	30,56	4,31	1069,54	0,17	726,98	660,30	1265,36
GAN_HUMAN	1	36,38	0,37	128,47	180,43	209,81	0,09	111,98	103,51	127,61
F8W998_HUMAN	1	36,26	---	1,50	3,17	39,47	0,45	---	---	---

RECHERCHE DE BIOMARQUEURS D'EXPOSITION ET D'EFFET À DES CANCÉRIGÈNES DE L'ENVIRONNEMENT PAR SPECTROMÉTRIE DE MASSE

Résumé

Le Benzo(a)pyrène (BaP), appartenant à la famille des hydrocarbures aromatiques polycycliques (HAP) est cancérigène pour l'homme. Nous avons développé une approche protéomique quantitative nanoLC-MS/MS label-free pour identifier des biomarqueurs liés à l'exposition au BaP dans le sécrétome des cellules hépatiques humaines exposées au BaP *vs.* des cellules non exposées et exposées au Benzo(e)pyrène (BeP). Le BeP, agent non classifié comme cancérigène pour l'homme, est choisi comme contrôle négatif afin de distinguer les protéines spécifiques du BaP de celles des HAP. 847 protéines ont été identifiées et quantifiées, et 55 ont été fortement surexprimées avec un ratio supérieur à 5 : la plupart de ces protéines surexprimées sont précoces et liées au cancer. Une validation ultérieure de l'expression de ces protéines dans le plasma de la population exposée au BaP aidera dans le développement de biomarqueurs qui permettront d'améliorer la détection précoce, le pronostic et prévention.

Mots-clés : Benzo(a)pyrène, Benzo(e)pyrène, biomarqueurs, cancer, sécrétome, protéomique, quantification label-free, spectrométrie de masse.

Résumé en anglais

Benzo(a)pyrene (BaP) belongs to a class of polycyclic aromatic hydrocarbon and is reported as a potent human carcinogen. We performed a nanoLC-MS/MS-based label-free quantitative proteomics approach to identify potential biomarkers of exposure in the secretome of BaP-treated *vs.* non-treated and Benzo(e)pyrene (BeP)-treated human hepatoma cell line HepG2. BeP-treated cells were chosen as a negative control to distinguish the BaP-specific from the HAP-specific regulated proteins: BeP is not classifiable as to its carcinogenicity to humans. 847 proteins have been identified and quantified and 55 proteins were seen as being highly up-regulated with a fold change of at least 5. Most of these up-regulated proteins were focused in cancer-related activities. Further validation of expression of these proteins in the plasma of BaP-exposed population will assist in the development of biomarkers that will greatly improve early detection, prognosis, prediction of treatment response and prevention.

Keywords: Benzo(a)pyrene, Benzo(e)pyrene, biomarkers, cancer, secretome, proteomics, label-free quantification, mass spectrometry