

Détection sur cohorte, par approche métabolomique, des constituants de l'exposome chimique – Caractérisation structurale et mapping des disruptions métabolomiques induites.

Co-direction : Pr Alain Paris (Muséum national d'Histoire naturelle) et Dr Sandra Alves (Sorbonne Université)

1. Contexte et Objectifs

Les produits du métabolisme reflètent dans leur composition l'effet combiné des facteurs génétiques et environnementaux ; ils constituent ainsi, à un instant donné, une représentation fidèle du phénotype moléculaire, en correspondance avec un transcriptome exprimé à partir du génome, mais modulé par l'influence de l'environnement.¹ En conséquence, ces dernières années, la métabolomique, la plus récente des approches omiques de phénotypage large des organismes, a pris de plus en plus d'importance dans le domaine de la santé, particulièrement la nutrition et la physiopathologie, afin de mieux caractériser les sujets ou les patients dans une perspective de médecine personnalisée². Dans le domaine clinique, la métabolomique vise à révéler au travers d'une signature métabolique spécifique la présence d'un dysfonctionnement métabolique provoqué par une maladie, des facteurs environnementaux (alimentation, style de vie, etc.) ou une exposition à des composés xénobiotiques (polluants, médicaments, etc.) qui sont constitutifs de l'exposome chimique^{3,4}. La métabolomique conjugue des méthodes issues de différentes disciplines : la chimie analytique, l'analyse statistique ou la chimiométrie et, enfin, la biologie. Elle implique notamment l'utilisation de méthodes analytiques sensibles et robustes, le plus fréquemment la chromatographie (liquide ou gazeuse) couplée à la spectrométrie de masse (GC/MS ou LC/MS).⁵ Bien que chronophages, ces méthodes analytiques génératrices de variables permettent ainsi la caractérisation la plus exhaustive et la plus sensible possible des échantillons biologiques. Le recours à des méthodes de détection en spectrométrie de masse (MS) à très haut pouvoir de résolution ($R_p > 5 \times 10^5$), par utilisation d'instruments FT-ICR (Fourier Transform – Ion Cyclotron Resonance, transformée de Fourier et à résonance cyclotronique des ions)⁶⁻⁸ permet la séparation des composés de masses très proches avec une très bonne précision dans la mesure et, donc, de s'affranchir de l'étape préalable de séparation des analytes, augmentant ainsi par un facteur 10 les débits d'analyse sans altérer la qualité des données produites⁹ (perte d'informations).

Dans une seconde étape, les deux points « bloquants » dans l'exploitation des données spectrales restent i) l'identification non ambiguë des variables (analytes) permettant de construire une signature discriminante, c'est-à-dire porteuse de sens, pour chacun des individus d'un sous-groupe donné et ii) la classification optimale desdits individus. Ces deux points font l'objet de recherches innovantes en statistique multivariée ou en chimiométrie en introduisant les méthodes de régularisation des données ou des méthodes alternatives à l'analyse en composantes principales (ACP) comme l'analyse canonique régularisée (rCCA), l'analyse en

¹ Fiehn O. Metabolomics – the link between genotypes and phenotypes. *Plant Mol. Biol.*, 48(2002):155.

² Barberis E, Khoso S, Sica A, Falasca M, Gennari A, Dondero F, Afantitis A, Manfredi M. Precision medicine approaches with metabolomics and artificial intelligence. *Int J Mol Sci.*, 23(2022):11269.

³ Wild CP. (2005) Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev*, 14:1847-1850.

⁴ Balcells C, Xu Y, Gil-Solsona R, Maitre L, Gago-Ferrero P, Keun HC. (2024) Blurred lines: Crossing the boundaries between the chemical exposome and the metabolome. *Curr Opin Chem Biol*, 78:102407.

⁵ Wilson ID, Plumb R, Granger J et al. HPLC-MS-based methods for the study of metabolomics. *J. Chromatogr. B.*, 817(2005):67.

⁶ Alves S, Rathahao-Paris E, Tabet JC. Potential of Fourier transform mass spectrometry for high throughput metabolomics analysis. Pages 219-302. In *Advances in Botanical Research*, Vol. 67: *Metabolomics coming of age with its technological diversity* (2013). Ed. D Rolin.

⁷ Habchi B, Alves S, Paris A, et al. How to really perform high throughput metabolomics analyses efficiently? *TrAC Trends in Anal. Chem.*, 85(2016):128.

⁸ Rathahao-Paris E, Alves S, Paris A. Chap 2. High-Throughput Metabolomics Using Flow Injection Analysis and Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. In: Wood P.L. (eds) *Metabolomics. Neuromethods*, vol 159. Humana, New York, NY. 2021, p. 9-21 https://doi.org/10.1007/978-1-0716-0864-7_2 (ISBN 978-1-0716-0863-0).

⁹ Habchi B, Alves S, Jouan-Rimbaud Bouveresse D, et al. Potential of dynamically harmonized Fourier transform ion cyclotron resonance cell for high throughput metabolomics fingerprinting: control of data quality. *Anal. Bioanal. Chem.*, 410(2018):483.

composantes indépendantes (ICA) ou l'analyse en composantes communes (ComDim)¹⁰⁻¹². D'un positionnement différent, les méthodes fondées sur les réseaux de neurones artificiels (ANN) ou l'apprentissage profond (Deep Learning, DL) commencent à émerger dans le domaine de la métabolomique clinique mais restent, le plus souvent, utilisées à titre exploratoire¹³⁻¹⁵.

Les données métabolomiques englobent les signaux issus de métabolites endogènes et de composés caractéristiques de l'exposome chimique. Sur une étude préliminaire récemment réalisée à partir des données métabolomiques urinaires établies par détection à ultra haute résolution en FT-ICR sur une cohorte de près de 500 sujets recrutés dans l'étude NESCAV^{9,16}, cette étude épidémiologique étant destinée à évaluer le risque cardiovasculaire en fonction de facteurs anthropologiques, cliniques et environnementaux qui sont documentés comme métadonnées, nous avons mis en évidence la présence de médicaments utilisés pour traiter des cas de diabète de type 2 ou d'arythmie cardiaque, ce que les métadonnées mises ensuite à notre disposition nous ont confirmé¹⁷.

Cette première étude épidémiologique sur l'exposome « médicamenteux » effectuée par analyse directe en FT-ICR mérite d'être étendue sur deux volets : le premier pour détecter et caractériser d'autres composés de l'exposome chimique représentant la pharmacopée déclarée dans les métadonnées (*i.e.* les médicaments prescrits) et, ainsi, les variations corrélées du métabolome endogène, le second pour accéder aux données de l'exposome chimique putatif d'origine environnemental présent aussi bien chez les sujets sous prescription médicale que chez les autres, en s'appuyant sur la bibliographie récente et le développement d'outils bio-informatiques pour extraire sélectivement des signaux spécifiques..

2. Projet de recherche postdoctorale

L'objectif du projet de recherche décrit ci-après est de caractériser, dans l'urine de sujets de la cohorte NESCAV, déjà phénotypés par FT-ICR-MS, l'existence et la structure chimique de différents composés, marqueurs de l'exposome chimique, qui appartiennent à différentes familles chimiques pour mesurer au niveau épidémiologique l'état de multi-exposition à ces molécules et, ensuite, évaluer la réalité des disruptions métaboliques induites par ces multi expositions. Pour ce faire, trois voies d'exploration seront développées :

✓ Volet analytique

La méthode analytique privilégiera la production d'ions fragments en spectrométrie de masse en tandem (MS/MS) pour compléter les données déjà acquises en FIA-ESI-FTICR-MS. La méthode chimométrique utilisée s'appuiera sur l'analyse des corrélations fortes ($\rho > 0,95$) pour relier les ions parents et les ions fragments probables issus de chaque parent. Cette étape statistique vient d'être validée sur des extraits de mousse dont le métabolome endogène varie en fonction de l'exposome en métaux lourds (A. Paris *et al.*, en préparation).

✓ Volet statistique

Les techniques d'analyses de données multivariées envisageront dans un premier temps les ACP, les analyses en composantes indépendantes, en composantes communes ou les analyses canoniques en considérant alors

¹⁰ Rutledge DN. Comparison of principal components analysis, independent components analysis and common components analysis. *J Anal Testing*, 2(2018), doi: 10.1007/s41664-018-0065-5.

¹¹ Cariou V, Qannari EM, Rutledge DN, Vigneau E. (2018) ComDim: From multiblock data analysis to path modeling. *Food Quality and Preference*, 67:27–34.

¹² Cariou V, Qannari EM, Jouan-Rimbaud Bouveresse D, Rutledge DN. (2019) ComDim methods for the analysis of multiblock data in a data fusion perspective. *Data Fusion Methodology and Applications*, Chapter 7, 179-204, Elsevier BV.

¹³ Galal A, Talal M, Moustafa A. (2022) Applications of machine learning in metabolomics: Disease modeling and classification. *Front Genet*, 13:1017340.

¹⁴ Mendez KM, Broadhurst DI, Reinke SN. (2019) The application of artificial neural networks in metabolomics: a historical perspective. *Metabolomics*, 15:142.

¹⁵ Mendez KM, Reinke SN, Broadhurst DI. (2019) A comparative evaluation of the generalised predictive ability of eight machine learning algorithms across ten clinical metabolomics data sets for binary classification. *Metabolomics*, 15:150.

¹⁶ Alkerwi A, Guillaume M, Zannad F, Laufs U, Lair ML. (2010) Nutrition, environment and cardiovascular health (NESCAV): protocol of an inter-regional cross-sectional study. *BMC Public Health*, 10:698.

¹⁷ Habchi B, Alves S, Streeel S, Guillaume M, Donneau A-F, Appenzeller BMR, Rutledge DN, Paris A, Rathahao-Paris E. (2023) Chemical exposure highlighted without any a priori information in an epidemiological study by metabolomic FT-ICR-MS fingerprinting at high throughput and high resolution. *Chem Res Toxicol*, 36:1592.

la matrice d'information complémentaire (métadonnées anthropométriques et de prescription de médicaments) pour orienter les capacités de détection de variables de l'exposome chimique médicamenteux et celles du métabolome endogène influencé par cet exposome spécifique. Dans cette approche, les données corrélées MS et MS/MS seront associées. Une analyse complémentaire des signatures métaboliques qui se fondent sur les résultats de cette première analyse et qui sera produite par les techniques d'enchâssement (en anglais : embedding) comme tSNE ou PHATE, mais modifiées en tant que méthodes supervisées telles que nous les développons actuellement dans l'unité MCAM, permettra d'isoler des sous-groupes d'individus partageant une signature métabolique commune. Cette progression dans le repérage de biomarqueurs putatifs permettra d'inférer l'existence d'une possible influence d'un composé (ou plusieurs composés apparentés ou non) appartenant à l'exposome chimique, soit par caractérisation structurale directe des signaux caractéristiques de cette signature, soit par l'interprétation fonctionnelle des marqueurs métaboliques endogènes qui varient de façon concertée, vraisemblablement sous une influence exposomique.

✓ Volet bioinformatique

Une approche complémentaire, bioinformatique, consistera à repérer les variables du métabolisme qui attestent une fonctionnalisation des molécules de l'exposome subissant la détoxification, réalisée prioritairement par le foie, et l'élimination urinaire des métabolites. Les étapes de détoxification comprennent les voies d'hydroxylation, d'oxydation ou réduction et de conjugaison (glucuro-, glyco-, sulfo-conjugués, thioéthers, etc.).

La méthode bioinformatique utilisée pour documenter les hypothèses de voies de biotransformation s'appuiera sur l'approche qui consiste à repérer des ions parents et fragments correspondant à des contraintes déclarées pour l'existence de métabolites de xénobiotiques conjugués (dérivés glucuronide, glycoside, sulfate, glutathion, N-acétyl-cystéine, etc.), hydroxylés, oxydés et/ou réduits ou comprenant des atomes d'halogènes (Cl, Br) et pouvant être présents dans l'environnement. Cette méthode donne des résultats encourageants sur matrice urinaire et ne cible pas *a priori* une famille de composés spécifique, ce qui correspond pour l'instant à une limite apparaissant dans bon nombre d'études très récentes qui n'envisagent pour l'essentiel que quelques composés d'un exposome chimique ciblé¹⁸⁻²¹ sans considérer les autres xénobiotiques qui peuvent être présents simultanément et être ainsi à l'origine d'effets cocktail²². Des études plus rares mettent en avant la diversité moléculaire de l'exposome chimique.^{23,24}

L'efficacité de cette méthode bioinformatique est démontrée dans une étude précédente sur la détection des métabolites d'un xénobiotique.^{25,26} L'utilisation orientée filtres de défauts de masse et/ou de rapports isotopiques aidera à prédire *ab initio* ces nouvelles structures chimiques.

¹⁸ Berger S, Oesterle I, Ayeni KI, Ezekiel CN, Rompel A, Warth B. (2024) Polyphenol exposure of mothers and infants assessed by LC-MS/MS based biomonitoring in breast milk. *Anal Bioanal Chem*, 416:1759–1774.

¹⁹ Hyötyläinen T, Ghaffar zadegan T, Karthikeyan BS, Triplett E, Orešič M, Ludvigsson J. (2024) Impact of environmental exposures on human breast milk lipidome in future immune-mediated diseases. *Environ Science Technol*, 58:2214–2223.

²⁰ Hyötyläinen T, McGlinchey A, Salihovic S, Schubert A, Douglas A, Hay DC, O'Shaughnessy PJ, Iredale JP, Shaw S, Fowler PA, Orešič M. (2024) In utero exposures to perfluoroalkyl substances and the human fetal liver metabolome in Scotland: a cross-sectional study. *Lancet Planet Health*, 8:e5-e17.

²¹ Zhao Y, Meijer J, Walker DI, Kim J, Portengen L, Jones DP, Saberi Hosnijeh F, Vlaanderen J, Vermeulen R. (2024) Dioxin(-like)-Related Biological Effects through Integrated Chemical-wide and Metabolome-wide Analyses. *Environ Sci Technol*, 58:258-268.

²² Tkalec Ž, Codling G, Klánová J, Horvat M, Kosjek T. (2022) LC-HRMS based method for suspect/non-targeted screening for biomarkers of chemical exposure in human urine. *Chemosphere*, 300:134550.

²³ Bland GD, Abrahamsson D, Wang M, Zlatnik MG, Morello-Frosch R, Park JS, Sirota M, Woodruff TJ. (2024) Exploring applications of non-targeted analysis in the characterization of the prenatal exposome. *Sci Total Environ*, 912:169458.

²⁴ Manz KE, Feerick A, Braun JM, Feng YL, Hall A, Koelmel J, Manzano C, Newton SR, Pennell KD, Place BJ, Godri Pollitt KJ, Prasse C, Young JA. (2023) Non-targeted analysis (NTA) and suspect screening analysis (SSA): a review of examining the chemical exposome. *J Expo Sci Environ Epidemiol*, 33:524-536.

²⁵ Rathahao-Paris E., Paris A., Bursztyka J., Jaeg J.P., Cravedi J.P., Debrauwer L. (2014) Identification of xenobiotic metabolites from biological fluids using flow injection analysis high-resolution mass spectrometry and post-acquisition data filtering. *Rapid Commun Mass Spectrom*, 28:2713-2722.

²⁶ Rathahao-Paris E., Alves S., Debrauwer L., Cravedi J.P., Paris A. (2017) An efficient data filtering strategy for easy metabolite detection from the direct analysis of a biological fluid using Fourier transform mass spectrometry. *Rapid Commun Mass Spectrom*, 31:1–10.

Ces trois volets seront engagés simultanément lors de l'année de post-doctorat sur un certain nombre de molécules candidates. En cas de prolongation du financement, une extension de la caractérisation de l'exposome chimique complémentaire à celle obtenue lors de la 1^{ère} année sera effectuée durant une seconde année de post-doctorat.

3. Les équipes partenaires supports de l'interdisciplinarité

La recherche post-doctorale sera placée sous la responsabilité du **Pr Alain Paris du Muséum national d'Histoire naturelle** (MNHN, unité MCAM, UMR 7245, MNHN/CNRS), métabolomicien et toxicologue, qui assurera la tutelle en matière de « **modélisation des données** », et par le **Dr Sandra Alves, MC-HDR à Sorbonne Université** (IPCM, UMR CNRS 8232), chimiste-analyste spécialisée en analyse structurale et en métabolomique, qui assurera la tutelle en matière de **chimie-biologie**, en particulier en ce qui concerne la caractérisation structurale des composés identifiés de l'exposome et la reconstruction et la curation des réseaux métaboliques soumis à disruption. Le Pr émérite Douglas Rutledge^{9-12,17}, accueilli comme chercheur bénévole dans l'unité MCAM (MNHN) et spécialiste de chimiométrie, sera associé au travail de recherche post-doctorale sur le volet « modélisation des données ». Pour les volets « validation et identification structurale des biomarqueurs » et « bioinformatique », le Dr Estelle Rathahao-Paris (IR INRAE, HDR), spécialisée en analyse structurale et en métabolomique par FT-ICR et accueillie à Sorbonne Université, sera aussi associée au travail de recherche post-doctorale. Des réunions trimestrielles de l'équipe de recherche seront organisées au Muséum ou à Sorbonne Université pour le suivi des travaux.

4. Les moyens à disposition & faisabilité du projet

Des échantillons urinaires ont été collectés sous la responsabilité du Pr M. Guillaume de l'université de Liège et des épidémiologistes porteurs de l'étude NESCAV¹⁶. Nous avons pu accéder au bras wallon de cette étude : échantillons urinaires et métadonnées. Les données métabolomiques urinaires ont déjà été acquises par introduction directe des échantillons et détection FT-ICR-MS en modes d'ionisation ESI (Electro Spray Ionisation) positif et négatif.

Le projet de recherche post-doctorale proposé vise à exploiter de façon beaucoup plus poussée les données déjà acquises.

Des expériences analytiques complémentaires (spectrométrie de masse en tandem, MS/MS et séparation par mobilité ionique) seront effectuées pour l'identification et la quantification des médicaments et/ou leurs métabolites chez les individus concernés (variables d'intérêt).

5. Le planning prévisionnel

Ces trois volets méthodologiques seront engagés simultanément lors de la 1^{ère} année de post-doctorat sur un certain nombre de molécules candidates définies comme prioritaires en tant que composés de l'exposome. En cas de prolongation du financement, une extension de la caractérisation de l'exposome chimique complémentaire à celle obtenue lors de la 1^{ère} année de post-doctorat sera effectuée durant une seconde année.

6. Conclusion : Impact attendu pour l'ASU dans le cadre du projet SOUND

Le(la) chercheur(e) post-doctoral(e) sera accueilli(e) pour partie à l'IPCM (Institut Parisien de Chimie Moléculaire, UMR SU CNRS 8232) de Sorbonne Université dont l'équipe de spectrométrie de masse est aussi partenaire de l'infrastructure nationale **Infranalytics** puisque hébergeant un spectromètre de masse FT-ICR de 7 teslas ouvert à des collaborations extérieures. Ceci sera un gage d'ouverture scientifique pour les autres membres de cette infrastructure, en particulier ceux équipés en FT-ICR, et une chance pour le(la) chercheur(e) recruté(e) en post-doctorat pour étoffer son réseau scientifique dans le domaine de la spectrométrie de masse à très haute résolution.

Par ailleurs, le projet de recherche post-doctorale proposée dans le domaine de l'exploration sans *a priori* des molécules appartenant à l'exposome chimique qui associe les analyses par FT-ICR-MS (et -MS/MS) aux analyses statistiques multivariées et celles bioinformatiques, celles-ci permettant de déconvoluer les voies de

biotransformation des xénobiotiques constitutifs de l'exposome chimique par le biais d'hypothèses d'hydroxylations ou d'oxydations/réductions combinées à différents types de conjugaison est original dans la mesure où l'apport de la très haute résolution de détection des ions et du très haut débit d'analyse que permettent les analyses en FT-ICR ne sont pas couvertes pour l'instant par les différents laboratoires d'analyse intégrés dans l'infrastructure nationale de recherche **France Exposome**. Les méthodes développées pourront donc être mises à disposition des acteurs de France Exposome et le(la) chercheur(e) post-doctoral(e) pourra s'impliquer sur le dernier tiers de son post-doctorat dans cette mission complémentaire de formation et de communication sur cette démarche de recherche sur la caractérisation de l'exposome chimique.

Finalement, les deux infrastructures nationales de recherche qui sont impliquées dans ce projet, à savoir **Infranalytics** et **France Exposome**, seraient ainsi amenées à interagir dans un contexte d'exploration épidémiologique ciblée.