

La métabolomique : un nouvel outil en nutrition et toxicologie

RÉSUMÉ

La métabolomique : une nouvelle approche pour évaluer l'effet santé des aliments

La métabolomique est un outil visant à identifier les métabolites présents dans un échantillon biologique et à évaluer les réponses d'un système biologique à un stimulus. Cet article présente les atouts et limites de l'utilisation de cet outil dans le domaine de la nutrition.

Utilisation de la métabolomique pour étudier les risques d'origine toxique

La métabolomique combine analyses statistiques et profilages métaboliques pour déceler un ensemble de biomarqueurs impliqués préférentiellement dans la discrimination des différents sous-groupes de populations exposées à des composés chimiques présentant un potentiel toxique variable. Différents exemples illustrent l'approche utilisée et montrent comment une exploration toxicologique renouvelée peut être ainsi proposée.

Jean-Louis SÉBÉDIO

Plateforme d'exploration du métabolisme : des gènes aux métabolites, UMR 1019, INRA, Université d'Auvergne, 63122 Saint-Genès Champanelle

Directeur de recherche à l'INRA, Jean-Louis Sébédio a suivi une formation initiale en chimie (PhD en 1976) puis a passé 8 ans au Canada à travailler principalement sur l'hydrogénation des corps gras et la formation d'acides gras *trans*.

Il est entré à l'INRA en 1983 et a travaillé sur les conséquences physiologiques de la consommation d'acides gras modifiés formés au cours des traitements de friture (acides gras cyclisés, acides gras *trans*, etc.). A Clermont depuis fin 2004, Jean-Louis Sébédio travaille sur le développement de l'outil métabolomique pour les études en nutrition, dont le principal intérêt est la détermination de biomarqueurs liés à la prise de poids chez l'homme.

Alain PARIS

Unité Mét@risk, INRA, AgroParisTech, 16 rue Claude Bernard, 75231 Paris cedex 05

Alain Paris, ingénieur agronome de formation, est directeur de recherches à l'INRA. Il y développe des recherches dans le domaine de la toxicologie alimentaire. Les méthodes développées depuis une dizaine d'années dans le domaine de la métabolomique lui ont permis ainsi qu'à son équipe de mesurer les altérations du fonctionnement métabolique des organismes, que ce soit en lien avec des expositions à des toxiques, à des hormones ou en lien avec des régimes préparés à partir de plantes génétiquement modifiées. Les recherches conduites actuellement en collaboration avec des épidémiologistes s'intéressent à la détection et à la mesure des disruptions métaboliques sur des cohortes pour déceler des expositions à différents toxiques qui peuvent être retrouvés dans l'alimentation.

Conférence du 19 novembre 2009

La Lettre Scientifique de l'IFN engage la seule responsabilité de ses auteurs.



LA MÉTABOLOMIQUE : UNE NOUVELLE APPROCHE POUR ÉVALUER L'EFFET SANTÉ DES ALIMENTS

Jean-Louis Sébédio, Estelle Pujos-Guillot, Jean-François Martin

La métabolomique est un outil d'exploration des organismes vivants qui décrit de façon exhaustive les différents métabolites présents dans les fluides biologiques (sang, urine, salive), ou dans les tissus. L'utilisation de techniques analytiques très performantes comme la résonance magnétique nucléaire (RMN) ou la spectrométrie de masse couplée à la chromatographie en phase gazeuse (GC-MS) ou à la chromatographie liquide haute performance (LC-MS) et d'outils bioinformatiques ont été à l'origine de son développement. Elle a tout d'abord été utilisée dans le domaine végétal (Fiehn *et al.*, 2000) et en toxicologie (Nicholson *et al.*, 1999 ; Paris *et al.*, 2008) par exemple pour prédire les effets toxiques de médicaments dans les phases précoces de développement. Les études en nutrition sont encore récentes (Gibney *et al.*, 2005 ; Rezzi *et al.*, 2006). Il s'agit de mesurer les effets (ou conséquences métaboliques) d'une intervention nutritionnelle, par la mesure du changement de profil métabolique induit. L'objectif peut être d'identifier des biomarqueurs de maladies métaboliques ou bien des biomarqueurs d'exposition à des nutriments ou micronutriments.

Il existe deux approches métabolomiques : la première est une approche ciblée qui concerne l'étude d'une voie métabolique définie, comme par exemple le métabolisme des glucides ou celui des lipides (lipidomique). La deuxième est une approche globale qui consiste à définir une empreinte métabolique en caractérisant le plus grand nombre possible de métabolites afin d'identifier les diverses voies métaboliques perturbées suite à un stimulus. Une approche métabolomique se décompose en plusieurs étapes à savoir le prélèvement et la préparation de l'échantillon, l'analyse proprement dite souvent réalisée par RMN, GC-MS ou LC-MS, le traitement des données, l'analyse statistique de celles-ci, l'identification des métabolites ou biomarqueurs et l'étude des voies métaboliques perturbées. Dans cette courte revue, nous n'avons pas cherché à être exhaustif mais simplement essayé d'illustrer les points clés d'une analyse métabolomique par quelques exemples.

Le prélèvement et la préparation de l'échantillon

En métabolomique nutritionnelle, la variabilité des échantillons biologiques est souvent importante comparée à l'effet produit par une intervention nutritionnelle. Dans une étude récente, Maher *et al.* (2007) ont montré que cette variabilité résultait non seulement du prélèvement, du stockage et de la préparation de l'échantillon, mais également de la variabilité intra et inter-individuelle. De nombreuses études donnent des recommandations pour le prélèvement et la préparation des échantillons (Walsh *et al.*, 2006 ; Zelena *et al.*, 2009). Par exemple, nous avons récemment montré que le type d'anticoagulant choisi pour le prélèvement sanguin avait une influence sur les profils obtenus (Pereira *et al.*, 2009). Il serait donc nécessaire d'aller jusqu'à une standardisation des protocoles afin de pouvoir intégrer des données provenant de différentes plateformes.

Les techniques analytiques

Même si la RMN a été la technique qui a permis à la métabolomique de voir le jour, la spectrométrie de masse, de part les différentes techniques de couplage associées,

a joué un rôle essentiel dans son essor (Dettmer *et al.*, 2007 ; Werner *et al.*, 2008). La spectrométrie RMN est une technique robuste qui permet de scanner des centaines d'échantillons en un temps restreint (Lenz et Wilson, 2007). Il s'agit d'une méthode non destructive, ce qui permet l'utilisation du même échantillon pour d'autres analyses. Elle permet également d'analyser les tissus intacts sans extraction préalable. Cependant, de part sa faible sensibilité, elle ne permet que la détection de métabolites présents à forte concentration (de l'ordre de la μmole). La première approche de métabolomique nutritionnelle par RMN a concerné l'effet chez la femme d'un régime contenant des protéines de soja riche en isoflavones afin d'étudier chez l'humain les effets de phytoestrogènes administrés dans des conditions physiologiques (Solanky *et al.*, 2003). L'étude a montré que même si tous les sujets présentaient la même réponse au régime, à savoir une augmentation des composés indiquant une augmentation du métabolisme anaérobie et une inhibition de la gluconéogenèse, la variation interindividuelle était importante. Cette variation interindividuelle est un problème important en métabolomique nutritionnelle et il est important de bien contrôler les critères d'inclusion des sujets, les régimes et les modalités de prélèvements des échantillons biologiques comme mentionné auparavant. Il est également fondamental de minimiser la variabilité analytique.

La GC-MS permet d'obtenir des données très reproductibles et une identification d'un nombre important des constituants du métabolome par l'utilisation de standards et de bases de données (Wishart, 2008). En revanche, elle requiert généralement une étape de préparation de l'échantillon (réaction de dérivation par exemple) qui est parfois fastidieuse (Dettmer, 2007). L'utilisation de la chromatographie bidimensionnelle permet d'augmenter le nombre de métabolites séparés, cependant les traitements des données sont beaucoup plus complexes (Koek *et al.*, 2007). Parmi les différentes dérivations utilisées, celle par le chloroformate d'éthyle permet d'obtenir des spectres facilement interprétables (Gao *et al.*, 2009). Ainsi, dans le cadre des études que nous menons sur l'identification de biomarqueurs précoces de l'obésité (projet ANR métaprofile), il nous a été possible d'identifier et de quantifier avec une bonne répétabilité et une limite de détection dans une fourchette de 10 à 500 pg environ 70 composés (mélange d'acides aminés, d'acides carboxyliques, de composés phénoliques) dans les eaux fécales de sujets sains (Gao *et al.*, 2009).

Même si la GC-MS ou la LC-MS sont couramment employées en analyse ciblée comme par exemple la lipidomique, l'infusion d'un mélange de lipides extraits de plasma humain contenant plusieurs standards internes dans un spectromètre de masse haute résolution a permis de quantifier 95 espèces moléculaires des 10 classes lipidiques majeures (Graessler *et al.*, 2009). Il a été ainsi montré une diminution des plasmalogènes chez les sujets hypertendus, en particulier les plasmalogènes à phosphatidylcholine et à phosphatidyléthanolamine contenant du 20:4 ou du 22:5.

Dans le cadre d'une analyse globale en LC-MS, alors que les échantillons d'urine peuvent être injectés directement après une simple dilution, l'analyse du plasma requiert une précipitation des protéines au préalable (Wilson *et al.*, 2005). Le principal problème de la LC-MS est le phénomène de suppression d'ionisation qui est fréquent en particulier lorsque l'électrospray est utilisé comme méthode d'ionisation. Pour pallier à ce problème, l'utilisation de colonnes de particules inférieures à 2 μm permet d'augmenter significativement la résolution chromatographique (Plumb *et al.*, 2006). La combinaison de plusieurs plateformes est souvent nécessaire pour couvrir les différentes classes de métabolites et leur

large gamme de concentrations dans les matrices biologiques (Busher *et al.*, 2009). L'étude de Williams *et al.* (2009) portant sur la caractérisation de profils plasmatiques de rats obèses ou non illustre bien la plus value d'une approche multi plateformes pour la caractérisation de biomarqueurs de pathologies. En effet alors que la RMN et la GC-MS ont montré une diminution du glucose, une augmentation des lipides, du cholestérol et de la vitamine E, l'UPLC-MS a révélé une augmentation de l'acide taurocholique donc une possible modification dans la production des acides biliaires chez le rat obèse. De plus, la diminution de la quantité de taurine dans l'urine des rats obèses est certainement la conséquence de l'utilisation de cet acide aminé dans la production de l'acide taurocholique. De même, les résultats récemment publiés par Wishart *et al.* (2008) sur la composition du liquide céphalo-rachidien démontrent la puissance analytique lorsque 3 plateformes (RMN, LC-MS et GC-MS) sont utilisées conjointement.

Le traitement des données et les études statistiques

Quelle que soit la technique analytique utilisée, le traitement des données commence par l'extraction, l'alignement et la quantification des signaux issus des différents types d'appareil ; déplacements chimiques pour la RMN, ions pour la LC-MS ou la GC-MS. Les matrices obtenues contiennent plusieurs centaines à plusieurs milliers de variables. Cette phase est critique puisqu'elle conditionne la qualité de la matrice des données qui va faire l'objet des différentes analyses statistiques. Du fait de dérives instrumentales, on doit ensuite réaliser une étape de normalisation des données. En minimisant la variabilité analytique cette étape permet de mettre en évidence les effets souvent modérés induits par les études nutritionnelles.

Les données ainsi obtenues définissent une empreinte métabolique variant en fonction des facteurs nutritionnels étudiés. Les différentes techniques statistiques ont pour objet de mettre en évidence les modifications significatives des réseaux métaboliques affectés. Trois grands types d'analyses statistiques sont généralement utilisés.

Les analyses univariées : suivant la complexité du protocole biologique, le test de Student (T Test) ou l'analyse de la variance (ANOVA) sont réalisés pour chaque variable. Cela permet de définir la liste exhaustive des métabolites affectés par les facteurs de l'étude.

Les analyses multivariées non supervisées : du fait de la nature multivariée des données métabolomiques, les méthodes factorielles telle que l'analyse en composante principale (ACP) permettent de visualiser graphiquement les grandes sources de variabilité des données. Les techniques de classification permettent de visualiser la structuration des données en groupes homogènes, tant pour les métabolites que pour les échantillons.

Les analyses multivariées supervisées : ces techniques permettent de définir et valider des modèles permettant d'expliquer les modifications induites par le facteur nutritionnel étudié. Il s'agit des analyses discriminantes et notamment la « *partial least square discriminant analysis* » (PLS-DA).

L'identification des métabolites : une étape cruciale

En RMN, l'identification des métabolites est réalisée après mesure des déplacements chimiques. En GC-MS, l'utilisation de l'impact électronique permet d'obtenir des spectres standardisés facilitant l'étape d'identification dans les bases de données. Cependant, l'identification des métabolites est

certainement l'étape limitante de l'utilisation de la LC-MS en métabolomique. L'identification peut être réalisée à partir de requêtes automatisées dans les bases de données comme par exemple KEGG et/ou HMDB (Wishart, 2008) en utilisant les renseignements fournis par la spectrométrie de masse (masse exacte et spectres de fragmentation). Cependant, à la masse de l'ion moléculaire correspondent de nombreuses formules chimiques. De plus, même si la mesure de la masse exacte permet de déterminer la composition élémentaire d'un composé, plusieurs structures sont encore possibles. Il est alors possible de réduire le nombre de structures proposées en utilisant l'abondance isotopique comme filtre (Kind et Fiehn, 2006). Une autre solution consiste à avoir recours à la spectrométrie de masse en tandem (MSn). Ce procédé permet de sélectionner un ion stable (ion parent) de la source d'ions puis de le fragmenter et ensuite d'analyser les « ions fils » afin d'obtenir les informations structurales nécessaires à l'identification du composé d'intérêt (Werner *et al.*, 2008). Lorsque cette technique est réalisée à l'aide d'un spectromètre à haute résolution elle est d'une aide précieuse pour la recherche de candidats potentiels dans les bases de données (Loftus *et al.*, 2008).

Conclusion

La métabolomique est un outil d'exploration des organismes vivants qui décrit de façon exhaustive les différents métabolites présents dans les fluides biologiques. Même si la RMN a été la technique qui a permis à la métabolomique de voir le jour, la spectrométrie de masse de part les différentes techniques de couplage associées a joué un rôle essentiel dans son essor. Cet essor n'est pas seulement dû aux progrès analytiques mais également au développement de la bioinformatique. Les études en nutrition sont encore récentes mais l'outil métabolomique présente un fort potentiel pour la détection de biomarqueurs de maladies métaboliques ou bien de biomarqueurs d'exposition à des nutriments ou micronutriments. Cependant, l'identification des métabolites reste l'étape clé pour laquelle un effort important de construction de bases de données doit être mené.

Bibliographie

- Buscher JM, Czernik D, Ewald JC *et al.* Cross-Platform Comparison of Methods for Quantitative Metabolomics of Primary Metabolism. *Anal Chem* 2009; 81: 2135-2143.
- Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrom Rev* 2007; 26: 51-78.
- Fiehn O, Kopka J, Dormann P *et al.* Metabolite profiling for plant functional genomics. *Nature Biotech* 2000; 18: 1157-1161.
- Gao XF, Pujos-Guillot E, Martin JF *et al.* Metabolite analysis of human fecal water by gas chromatography/mass spectrometry with ethyl chloroformate derivatization. *Anal Chem* 2009; 393: 163-175.
- German JB, Roberts MA, Fay LB, Watkins SM. Metabolomics and individual metabolic assessment: the next great challenge for nutrition. *J Nutr* 2002; 139: 2486-2487.
- Gibney MJ, Walsh M, Brennan L *et al.* Metabolomics in human nutrition : opportunities and challenges. *Am J Clin Nutr* 2005; 82: 497-503
- Graessler J, Schwudke D, Schwarz PEH *et al.* Top-down lipidomics reveals ether lipid deficiency in blood plasma of hypertensive patients. *PLoS ONE* 2009; 4: e6261.
- Kind T, Fiehn O. Metabolomic database annotations *via* query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics* 2006; 7: 234.

Koek MM, Mulwijk B, van Stee LL, Hankemeier T. Higher mass loadability in comprehensive two-dimensional gas-chromatography-mass spectrometry for improved analytical performance in metabolomics analysis. *J Chromatogr A* 2008; 1186: 420-429.

Lenz EM, Wilson ID. Analytical strategies in metabolomics. *J Proteome Res* 2007; 6: 443-58.

Loftus N, Miseki K, Lida J *et al.* Profiling and biomarker identification in plasma from different Zucker rat strain via high mass accuracy multistage mass spectrometric analysis using liquid chromatography/mass spectrometry with a quadrupole ion trap-time of flight mass spectrometer. *Rapid Commun Mass Spectrom* 2008; 22: 2547-54.

Maher AD, Zirah, SFM, Holmes E, Nicholson JK. Experimental and Analytical Variation in Human Urine in 1H NMR Spectroscopy-Based Metabolic Phenotyping Studies. *Anal Chem* 2007; 79: 5204-5211.

Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* 1999; 29: 1181-1189.

Paris A, Domange C, Eveillard A *et al.* Métabonomique, métabolomique et les préoccupations en matière de sécurité alimentaire et de protection de la santé des consommateurs. *OCL* 2008 ; 15 : 300-4.

Pereira H, Martin JF, Joly C *et al.* Development and validation of a UPLC/MS method for a nutritional metabolomic study of human plasma. *Metabolomics*, sous presse.

Plumb RS, Johnson KA, Rainville P *et al.* The detection of phenotypic differences in the metabolic plasma profile of three strains of Zucker rats at 20 weeks of age using ultraperformance liquid chromatography/orthogonal acceleration time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 2006; 20:2800-6.

Rezzi S, Ramadan Z, Fay LB, Kochhar S. Nutritional metabolomics: Applications and perspectives. *J Proteome Res* 2007; 6: 513-525

Solanky KS, Bailey NJC, Beckwith-Hall *et al.* Application of biofluid 1H nuclear magnetic resonance-based metabolomic techniques for the analysis of the biochemical effects of dietary isoflavones on human plasma profile. *Anal Biochem* 2003; 323: 197-204.

Walsh MC, Brennan L, Malthouse JPG *et al.* Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. *Am J Clin Nutr* 2006; 84:531-539.

Werner E, Heilier JF, Ducruix C *et al.* Mass spectrometry for the identification of the discriminating signals from metabolomics: current status and future trends. *J Chromatogr B* 2008; 871: 143-163.

Williams R, Lenz EM, Wilson AJ *et al.* A multi-analytical platform approach to the metabolomic analysis of plasma from normal and Zucker (fa/fa) obese rats. *Mol Biosyst* 2006; 2:174-83.

Wilson ID, Plumb R, Granger J *et al.* HPLC-MS based methods of the study of metabolomics. *J Chromatogr B* 2005; 817: 67-76

Wishart DS. Metabolomics: applications to food science and nutrition research. *Trends Food Sc Tech* 2008; 19: 482-493.

Zelena E, Dunn W.B, Broadhurst D *et al.* Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum. *Anal Chem* 2009; 81: 1357-1364.

UTILISATION DE LA MÉTABOLOMIQUE POUR ÉTUDIER LES RISQUES D'ORIGINE TOXIQUE

Alain Paris

Introduction

Depuis une quinzaine d'années, en raison de l'essor des méthodes spectroscopiques mais aussi de la volonté des chercheurs de décrire le fonctionnement des systèmes biologiques de manière toujours plus exhaustive, une démarche analytique reposant sur la description de l'ensemble des voies de métabolisation présentes dans l'organisme a émergé. Elle fait suite aux descriptions holistiques s'intéressant aux gènes différentiellement exprimés, le transcriptome, et aux protéines différentiellement synthétisées, le protéome. Cette démarche est généralement appelée la métabolomique. Elle conjugue en fait deux approches complémentaires mais distinctes dans leurs objectifs ultimes. La première, la métabolomique *sensu stricto*, est la démarche analytique qui consiste à produire à haut débit et à quantifier les empreintes métaboliques ainsi réalisées. La seconde, dénommée métabonomique, adjoint à la première la statistique multivariée. Cette dernière consiste à traiter les données pour faire apparaître les facteurs structuraux qui permettent de résumer et d'expliquer l'information métabolomique, eu égard au questionnement scientifique à l'origine de l'exploration fonctionnelle réalisée. Ce second terme a été forgé par Nicholson *et al.* à la fin des années 1990 [1]. La métabolomique est adaptée au point de vue de l'analyste alors que la métabonomique recoupe les préoccupations du physiologiste qui s'intéresse d'abord aux éléments de régulation perçus au niveau de l'organisme et, par extension, au toxicologue qui analyse les éléments du dysfonctionnement physiologique dans le cas d'une exposition de l'organisme à des substances possédant un pouvoir toxique avéré.

L'objectif ultime de la métabolomique considérée dans son acception la plus large est ainsi de révéler un ensemble de biomarqueurs expliquant la perturbation engendrée par l'exposition de l'organisme à un nouvel environnement, nutritionnel, pharmacologique, toxicologique, ou pathologique. Plus largement, la variation d'origine génétique des organismes et sa répercussion sur le fonctionnement métabolomique se prête là-aussi très bien à une exploration métabolomique. Ceci montre tout le potentiel que cette démarche générique possède pour décrire au plus près le fonctionnement des organismes vivants. Elle recèle une limite en cela que, comme bon nombre des méthodes holistiques en vogue actuellement, elle est par essence relative. En effet, les biomarqueurs d'intérêt résultent toujours d'une démarche de comparaison des groupes d'individus étudiés dans un contexte d'analyse de la perturbation du fonctionnement « normal » des individus. Néanmoins, grâce aux différents ensembles de biomarqueurs qui émergent de ces analyses, un des atouts de cette approche est de produire différentes hypothèses de travail pouvant aider à interpréter les disruptions métaboliques induites par des perturbations, qu'elles soient d'origine toxique ou non, mais aussi à les hiérarchiser.

Les travaux fondateurs de la métabolomique sont ceux produits en spectroscopie de résonance magnétique nucléaire (RMN) par l'équipe de J. Nicholson à Londres par différentes études pharmacologiques ou toxicologiques [2-5]. Leurs résultats leur ont ainsi permis de dépasser les limites rencontrées jusqu'alors dans l'appréhension des effets toxiques, déjà bien décrits par le passé, de différentes substances modèles.

Les travaux présentés ci-après montrent comment l'analyse des disruptions métaboliques peuvent rendre compte de perturbations du système endocrinien induites par des stéroïdes, de perturbations induites par des toxiques environnementaux ou de métabolites secondaires présents chez les végétaux. Enfin, la généralité de la démarche sera illustrée dans le cas de l'évaluation de l'innocuité des aliments préparés à partir de produits végétaux issus de variétés génétiquement modifiées.

Les contraintes méthodologiques

Les données produites par les analyses métaboliques ne peuvent être réellement exploitables, c'est-à-dire produire des ensembles de biomarqueurs pertinents, que si certaines règles méthodologiques sont respectées. La démarche est par essence multidisciplinaire. Elle doit donc satisfaire aux exigences de chacune des disciplines « convoquées » pour faire émerger un résultat final fiable et, le plus souvent, original.

La démarche habituellement retenue est en fait triple.

Physiologique d'abord, elle doit mettre en œuvre les éléments propres à faire apparaître sans aucune ambiguïté le questionnement biologique. C'est au travers de cette mise en interrogation scientifique que les éléments du protocole expérimental seront disséqués, choisis et prévalidés. A cette étape, un des objectifs prioritaires y est d'identifier, pour les écarter, tous les facteurs confondants possibles, dont la seule présence dans l'étude pourra mettre réellement en défaut l'interprétation des biomarqueurs qui y seront générés.

Analytique ensuite. A partir des échantillons biologiques produits dans le cadre du protocole expérimental retenu à l'étape précédente, le choix de la méthode de production des données caractérisant le métabolome doit être effectué de façon à permettre une exploitation des données réaliste. Le plus souvent, les méthodes spectroscopiques retenues sont la spectrométrie de masse (SM) couplée ou non à la chromatographie liquide ou la RMN à haut champ. La première a le potentiel de produire de l'information quantitative sur un nombre de métabolites très important, mais ne peut soutenir des cadences suffisantes pour générer de l'information sur des effectifs très importants qui avoisinent plusieurs centaines d'individus. En revanche, la seconde est plus limitée dans le champ des biomarqueurs métaboliques potentiels qui pourront être caractérisés, mais peut produire de l'information pour des effectifs extrêmement nombreux, plusieurs centaines voire plusieurs milliers d'individus, sans risque de faire apparaître de dérive dans les signaux produits. L'information produite en RMN s'adresse d'abord aux éléments du métabolisme général et se révèle précieuse pour analyser les ajustements de la physiologie nutritionnelle.

Le dernier volet enfin concerne le champ de l'exploitation statistique des données. Les méthodes sont toutes multidimensionnelles et doivent s'adapter au statut des données produites à l'étape précédente. En effet, une redondance informationnelle existe aussi bien en RMN qu'en SM et doit rester limitée. Un « bruit » informationnel existe aussi qui ne peut être décrit par les facteurs étudiés qui ont été dûment décrits à la première étape et qui doit être corrigé en conséquence. Enfin, un dimensionnement suffisant du dispositif expérimental doit être de nature à éviter tout risque lié à un défaut de degré de liberté en comparaison du nombre de biomarqueurs potentiellement accessibles. De nombreuses négligences sur ce point précis mettent en défaut les critères de validation croisée qui eux seuls permettent de donner un statut plus sûr aux différents biomarqueurs

candidats retenus à cette étape. Différentes avancées très significatives dans ce domaine statistique ont été obtenues très récemment [6- 10].

A l'issue de cette analyse statistique, une caractérisation structurale approfondie des biomarqueurs candidats permet de proposer un ensemble de biomarqueurs qui deviendra alors une signature métabolique de la perturbation étudiée.

Pour l'instant, la démarche s'est avérée très prometteuse sur des dispositifs expérimentaux restreints quand il s'est agi de s'intéresser à des études conduites à l'échelle du laboratoire. En revanche, elle a plus de difficulté à montrer que la transposition à des cohortes suivies dans différentes études épidémiologiques est possible. Pourtant, dans le domaine d'une veille sanitaire de la population générale exposée par voie alimentaire au risque chimique présent à bas bruit, cette transposition devient indispensable pour évaluer au plus près les disruptions métaboliques induites par une exposition à différents résidus de pesticides, contaminants ou polluants environnementaux. Cet objectif ne sera atteint qu'au prix de progrès sensibles à réaliser dans les domaines analytique, informatique et statistique dont l'essor récent montre qu'il pourrait être atteint en moins d'une décennie.

Toxicologie et perturbation endocrinienne

La consolidation de l'outil métabolique tel qu'il existait au tout début des années 2000 a permis d'améliorer la détection des disruptions métaboliques et leur mesure. Ainsi, dans le cadre de l'emploi frauduleux des stéroïdes anabolisants en élevage, l'utilisation de cette démarche a montré dans un contexte expérimental comment il était possible d'accéder à des variations mesurables du métabolome qui pouvaient rendre compte d'un ajustement homéostatique du métabolisme général chez des bovins traités par des anabolisants stéroïdiens [11,12]. Jusqu'alors, les principes de l'analyse métabolique n'avaient été posés qu'avec des résultats obtenus en toxicologie aiguë et en pharmacologie. Ici, les perturbations du fonctionnement métabolique induites sont celles résultant d'une administration contrôlée d'hormones ayant un rôle dans le contrôle de la croissance des animaux. Ces variations sont nécessairement de moindre ampleur.

Les données expérimentales obtenues sur bovins mâles et femelles ont révélé des variations coordonnées du métabolome mesurée par RMN 2D sur urine qui s'expliquent en priorité par un effet lié au genre. Cependant, le facteur traitement anabolisant est pleinement détectable quel que soit le genre considéré. Il correspond à une réponse anabolique commune entre mâles et femelles traités aux anabolisants. Un dernier facteur soulignant la variation de la réponse métabolique spécifique à chaque genre est perceptible bien que mineur [11]. Certains biomarqueurs métaboliques ont été identifiés et validés au plan biochimique et physiologique grâce à une littérature nombreuse accumulée depuis le début des années 1960 [12]. Les observations obtenues confirment à nouveau que le traitement anabolisant associant un androgène et un œstrogène chez les mâles ou n'utilisant qu'un androgène chez les femelles se traduit par une modification du turnover des composés azotés chez tous ces animaux.

Fonctionnellement, la moindre excrétion des composés azotés (urée, créatinine, créatine) chez les animaux traités se traduit aussi par un bypass dans l'excrétion azotée via l'excrétion d'hippurate, un produit de conjugaison du benzoate à la glycine, qui compense très partiellement la réduction d'excrétion de composés azotés. Pour le métabolisme

énergétique, l'utilisation des anabolisants se traduit dans l'espèce bovine par une augmentation relative de l'excrétion de citrate mesuré dans l'urine et une diminution de celle du glucose urinaire. Ceci peut être mis en relation avec un ajustement coordonné qui engage vraisemblablement les activités lactate déshydrogénase, pyruvate déshydrogénase et isocitrate déshydrogénase [12].

Si la métabolomique permet de générer des ensembles de biomarqueurs pertinents au plan de l'interprétation fonctionnelle qui rendent compte des effets perçus au niveau du métabolisme général en réponse à différents types de perturbation, elle peut aussi être utilisée comme système de classification pour prédire la classe d'appartenance d'un individu donné au statut phénotypique *a priori* inconnu. C'est cette propriété particulière qui a fait l'objet d'un approfondissement méthodologique spécifique et qui s'inscrit dans la détection de profils endocriniens atypiques chez le sportif de haut niveau par l'analyse des disruptions métaboliques qui peuvent être induites par la prise de substances influençant les profils de sécrétion hormonale. Ces sportifs de haut niveau sont en effet soumis depuis quelques années au suivi médical longitudinal organisé avec l'appui des fédérations nationales. L'objectif était de montrer que des variations du métabolisme général peuvent être interprétées en fonction des variations concomitantes du pattern endocrinien en cortisol, testostérone et IGF-1, ce dernier reflétant bien les variations de la stimulation somatotrope au niveau hépatique.

Les modélisations statistiques actuellement en cours permettent d'indiquer l'appartenance probable à telle ou telle classe de disruption endocrinienne qui pourrait être expliquée par la prise de substances dopantes. Un travail spécifique de validation croisée nous indique dans cette étude quelle est la dimension de la base d'apprentissage nécessaire pour garantir un taux d'erreur négligeable. Il est vraisemblable qu'une cohorte de près d'un millier de sujets sera nécessaire. En parallèle, une étude plus qualitative a permis de montrer que plusieurs de ces ensembles de biomarqueurs rendant compte des effets liés à la variation de tel ou tel profil hormonal étaient relativement indépendants les uns des autres. Cette simplification en éléments unitaires de disruption asservis à tel ou tel disrupteur endocrinien pourrait représenter un premier pas vers l'analyse des réponses multiples induites par des cocktails de substances auxquels l'organisme peut être exposé simultanément.

Le problème de classification n'est pas en soi trivial. Il fait l'objet depuis de nombreuses années de recherches spécifiques sur les outils statistiques à mettre en œuvre [13-15]. Dans bon nombre de cas il fonctionne en système « boîte noire » car les biomarqueurs métaboliques restent insuffisamment caractérisés au plan structural. Une recherche spécifique sur leur caractérisation analytique ultime est menée actuellement en recourant à des analyses spectroscopiques à très haute résolution tant en SM qu'en RMN. Ces nouveaux résultats devraient aider à renforcer le jeu des biomarqueurs utiles à la prédiction des phénotypes métaboliques des sujets suivis dans le cadre d'études épidémiologiques.

Toxicologie des contaminants : les phtalates, le bisphénol A et les POP

Les phtalates et le bisphénol A sont largement utilisés en plasturgie car ils confèrent aux matériaux plastiques des propriétés rhéologiques particulières ; ils permettent d'en renforcer la souplesse. Les polluants organiques permanents (POP) regroupent des substances telles que les polychlorobiphényles (PCB), les polychlorodibenzo-p-dioxines

(PCDD), les polychlorodibenzofuranes (PCDF) et des hydrocarbures aromatiques polycycliques (HAP). Certains d'entre eux se concentrent dans la chaîne alimentaire et peuvent se retrouver à des quantités significatives dans les aliments. Des interrogations sur leur pouvoir de disruption endocrinienne sont apparues il y a une dizaine d'années sans que, depuis, une ligne claire en matière de gestion du risque ait été pleinement arrêtée. La métabolomique, de par son potentiel important à révéler des perturbations des systèmes biologiques, peut aider à faire progresser les connaissances en matière d'analyse des risques pour la santé du consommateur que posent ces substances. Des premières études ont été menées ces toute dernières années sur modèle animal pour tenter de cerner les effets présents aux faibles doses que peuvent induire ces substances. Mais c'est sur la première famille de ces composés que les résultats obtenus ont été les plus nombreux.

Les phtalates ne sont pas biologiquement neutres ; ce sont des inducteurs de peroxyosomes puissants chez le rat ou la souris. Ils sont également répertoriés comme de possibles anti-androgènes et possèderaient des propriétés œstrogéno-mimétiques. La réévaluation de leur profil d'altération fonctionnelle est menée actuellement en collaboration avec l'équipe de T. Pineau et P. Martin à Toulouse, en particulier en s'appuyant sur des données obtenues en analyse métabolomique. Il s'agit de déterminer plus précisément la dose qui est la plus probablement sans effet pour l'homme. Dans une étude récente, la variation de la réponse métabolique chez la souris est liée à la quantité de diéthylhexyl-phtalate (DEHP), un des phtalates les plus utilisés en plasturgie [16]. L'effet hypolipémiant de cette molécule, déjà bien documenté, est retrouvé dans l'analyse métabolomique des empreintes plasmatiques. Une modification du métabolisme énergétique est détectée par une diminution de la proportion des acides aminés cétoformateurs, une augmentation des acides aminés glucoformateurs et une baisse de la concentration en acides biliaires dans le cas d'une intoxication obtenue à la plus forte dose étudiée. En complément à cette étude métabolomique, une analyse du transcriptome a été réalisée sur les mêmes animaux. Ceci a permis de produire une information phénotypique globale encore plus structurée mais qui s'interprète dans le cas présent uniquement en situation de perturbation métabolique induite par une forte dose de DEHP. Différentes régulations sont impliquées dans l'augmentation de la synthèse d'hème. Il s'agit de l'augmentation du transport de la thréonine, de l'augmentation de l'expression du gène *Vanin1* qui est impliqué dans la biosynthèse de l'acide pantothénique et de la cystéamine, et de l'augmentation de l'expression du gène *Alas1* impliqué dans la biosynthèse de l'acide 5-aminolévulinique, un précurseur de l'hème [16]. Cette étude a permis en outre de découpler les ajustements physiologiques du métabolisme général en lien avec une exposition à de faibles doses de DEHP de celles, plus fortes, attestant un impact toxique sur l'organisme.

Le bisphénol A est un composé aux propriétés œstrogéniques attestées. Ses effets sur la physiologie de la reproduction sont détectés chez l'adulte mais induits *in utero* ou dans la période postnatale ou infantile [17]. C'est un œstrogène faible eu égard à son affinité pour les récepteurs ER α et ER β , mais il possède une forte affinité pour une nouvelle forme de récepteur membranaire (ERR γ) et interagit avec le récepteur AhR et le récepteur des hormones thyroïdiennes [18]. Il serait aussi un antagoniste du récepteur des androgènes [18]. Néanmoins, ses effets sur le métabolisme général n'ont pas été caractérisés. Des premiers résultats obtenus par une approche métabolomique indiquent un effet sur le métabolisme général détecté tant

au niveau plasmatique qu'hépatique sur une gamme de doses variant de 50 µg/kg PV/j à 5 mg/kg PV/j.

La troisième famille, celle composite des POP de type PCB et PCDD/F, a été étudiée, en collaboration avec l'équipe de G. Rychen et de H. Schroeder à l'université de Nancy. L'objectif était de savoir chez le rat jeune si, à des doses compatibles avec ce qui peut être retrouvé dans le lait de chèvres nourries avec des fourrages contaminés accidentellement par ces substances, un effet sur le métabolisme général et sur le métabolisme cérébral induit par une administration par voie orale sur une durée de 4 semaines de ces substances était perceptible. Cet exemple illustre clairement les exigences méthodologiques à respecter pour produire une modélisation statistique satisfaisante des données et éviter les interprétations abusives si l'analyse préalable de possibles facteurs confondants n'est pas réalisée et, enfin, pour renforcer les observations réalisées sur le métabolisme cérébral par des observations physiologiques complémentaires, ici celles obtenues sur des variations du comportement des animaux exposés à ces substances.

Toxicologie des métabolites secondaires végétaux

Il existe dans le règne végétal une quantité très importante de métabolites secondaires qui possèdent des propriétés biologiques plus ou moins bien caractérisées. Certaines plantes sont toxiques pour certaines espèces animales qui les consomment en trop grandes quantités, alors qu'elles sont inoffensives pour d'autres. Tel est le cas de la porcelle enracinée (*Hypochoeris radicata* L.) qui est toxique pour l'espèce équine. L'ingestion quasi exclusive en année sèche de cette plante est à l'origine de troubles neuromusculaires graves dans l'espèce équine (harper australien) pour lesquels aucun traitement thérapeutique n'existe.

Le modèle souris a été validé comme modèle de disruptions métaboliques orthologues de celles présentes chez le cheval intoxiqué en ciblant les tissus cibles de l'intoxication. Aucune connaissance en matière de cible toxicologique ou de mécanisme d'intoxication, ni de substances toxiques identifiées n'a été nécessaire pour progresser dans cette recherche [19]. Un biomarqueur métabolique candidat, le *scyllo*-inositol, a été caractérisé dans le foie et le cerveau d'animaux intoxiqués pendant trois semaines par de faibles concentrations (1-3 %) de poudre d'inflorescence incorporée à un aliment standard.

Ce marqueur a déjà été décrit dans d'autres troubles neurologiques comme l'alcoolisme chronique [20] ou la maladie d'Alzheimer [21]. Physiologiquement, le *scyllo*-inositol est présent à de très faibles concentrations au sein des systèmes nerveux périphérique et central mais ses fonctions précises sont encore peu connues. Le *scyllo*-inositol et son stéréoisomère, le *myo*-inositol, sont impliqués dans les mécanismes de conduction nerveuse au sein des nerfs périphériques qui, en cas de neuropathie diabétique aiguë, sont altérés [22].

En complément, une validation de la démarche de recherche toxico-métabolique a été obtenue grâce à des analyses fonctionnelles soit par imagerie cérébrale obtenue par IRM [23], soit par une étude approfondie du comportement. D'autres biomarqueurs métaboliques candidats ont ainsi été détectés, telle la choline dont l'augmentation indiquerait une possibilité d'inflammation et de démyélinisation. De même, le déséquilibre de la balance glutamine / glutamate serait en lien avec une excitotoxicité et une perturbation de la conduction neuronale qui pourraient expliquer les modifications comportementales observées chez la souris, en

particulier une diminution de l'état d'anxiété observée dans un contexte sécurisé.

Un autre domaine concerne celui des plantes génétiquement modifiées (PGM) pour lequel l'analyse métabolomique pourrait permettre de réaliser des avancées importantes pour la qualification de l'innocuité des aliments. En effet, la création de nouveaux événements génétiques peut s'accompagner d'une modification substantielle de la composition en métabolites secondaires naturels ou néosynthétisés. Pour l'instant, les méthodes d'expertise employées réglementairement dérivent pour la majeure partie d'entre elles de méthodes utilisées en recherche pharmaceutique. Elles ne sont guère appropriées dans le cas présent car il n'est pas possible de s'appuyer sur des expériences de toxicité aiguë quand il s'agit de tester un aliment. Une manière de détecter de possibles anomalies au niveau métabolomique est de mesurer la proportion de la variabilité métabolique imputable à l'événement de transgène et de la comparer à celle expliquée par d'autres facteurs étudiés en parallèle, tels les effets « cultivar », « pédologie-climat » ou « année » [24].

Conclusion

La métabolomique, qu'elle mette en œuvre des approches de production de données par SM ou par RMN, est très bien adaptée à la caractérisation des perturbations des organismes exposés de façon très variable à différents toxiques. Dans ses fondements, cette démarche est très apparentée à la thermodynamique ; la moindre variation de fonctionnement des systèmes biologiques est désormais mesurable grâce à ces outils dont la sensibilité et la résolution s'améliorent régulièrement [25]. Ces développements instrumentaux se sont accompagnés de développements bioinformatiques importants. Les approches actuellement en développement consistent à mieux décrire la part d'information des matrices d'information métabolomiques qui ne permet pas de prédire la réponse phénotypique objet de l'étude pour la retirer du tableau des données métabolomiques, que ce soit par des méthodes dites OSC-PLS (Orthogonal signal correction-partial least squares) ou par des méthodes à noyau [26], pour ensuite réaliser les prédictions de classifications ou de réponses. Un autre volet qui se développe en parallèle et qui concerne les traitements bioinformatiques des données a trait aux banques de données compilant les références structurales nécessaires à la caractérisation non ambiguë des biomarqueurs candidats retenus à l'étape de traitement statistique. Les plus grandes avancées à attendre sont celles qui seront permises par la production de références à très haute résolution en SM car elles ouvrent des perspectives prometteuses dans le domaine de l'analyse *ab initio* des réseaux métaboliques [27].

La versatilité de l'approche métabolomique permet de conduire des recherches visant à qualifier l'innocuité des systèmes alimentaires complexes qui contiennent des matières premières issues de PGM ou des extraits de plantes dont le risque toxicologique reste difficile à mesurer avec les méthodes usuelles. Au plan de la toxicologie fondamentale, la métabolomique est un outil qui devrait permettre de mieux cerner les effets additifs et de synergie en cas d'intoxication chronique qui sont induits par l'association de différents composés toxiques.

Bibliographie

1. Nicholson JK, Lindon JC, Holmes E (1999). 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* 29:1181-1189.

2. Holmes E, Foxall PJ, Nicholson JK, Neild GH, Brown SM, Beddell CR *et al.* (1994). Automatic data reduction and pattern recognition methods for analysis of ¹H nuclear magnetic resonance spectra of human urine from normal and pathological states. *Anal Biochem* 220:284–296.
3. Holmes E, Nicholls AW, Lindon JC, Ramos S, Spraul M, Neidig P *et al.* (1998). Development of a model for classification of toxin-induced lesions using ¹H NMR spectroscopy of urine combined with pattern recognition. *NMR Biomed* 11:235–244.
4. Coen M, Hong YS, Clayton TA, Rohde CM, Pearce JT, Reilly MD *et al.* (2007). The mechanism of galactosamine toxicity revisited; a metabolomic study. *J Proteome Res* 6: 2711–2719.
5. Yap IK, Clayton TA, Tang H, Everett JR, Hanton G, Provost JP *et al.* (2006). An integrated metabolomic approach to describe temporal metabolic dysregulation induced in the rat by the model hepatotoxin allyl formate. *J Proteome Res* 5:2675–2684.
6. Couto Alves A, Rantalainen M, Holmes E, Nicholson JK, Ebbels TM (2009). Analytic Properties of Statistical Total Correlation Spectroscopy Based Information Recovery in ¹H NMR Metabolic Data Sets. *Anal Chem* 81:2075–2084.
7. Veselkov KA, Lindon JC, Ebbels TM, Crockford D, Volynkin VV, Holmes E, Davies DB, Nicholson JK (2009). Recursive segment-wise peak alignment of biological ¹H NMR spectra for improved metabolic biomarker recovery. *Anal Chem* 81:56–66.
8. Cavill R, Keun HC, Holmes E, Lindon JC, Nicholson JK, Ebbels TM (2009). Genetic algorithms for simultaneous variable and sample selection in metabolomics. *Bioinformatics* 25:112–118.
9. Pearce JT, Athersuch TJ, Ebbels TM, Lindon JC, Nicholson JK, Keun HC (2008). Robust algorithms for automated chemical shift calibration of 1D ¹H NMR spectra of blood serum. *Anal Chem* 80:7158–7162.
10. Rantalainen M, Cloarec O, Ebbels TM, Lundstedt T, Nicholson JK, Holmes E, Trygg J (2008). Piecewise multivariate modeling of sequential metabolic profiling data. *BMC Bioinformatics* 9:105.
11. Dumas ME, Canlet C, André F, Vercauteren J, Paris A (2002). Metabolomic assessment of physiological disruptions using ¹H-¹³C HMBC-NMR spectroscopy combined with pattern recognition procedures performed on filtered variables. *Anal Chem* 74:2261–2273.
12. Dumas ME, Canlet C, Vercauteren J, André F, Paris A (2005). Homeostatic signature of anabolic steroids in cattle using ¹H-¹³C HMBC NMR metabolomics. *J Proteome Res* 4:1493–1502.
13. Ebbels TM, Keun HC, Beckonert OP, Bollard ME, Lindon JC, Holmes E, Nicholson JK (2007). Prediction and classification of drug toxicity using probabilistic modeling of temporal metabolic data: the consortium on metabolomic toxicology screening approach. *J Proteome Res* 6:4407–4422.
14. Zuber V, Strimmer K (2009). Gene ranking and biomarker discovery under correlation. *Bioinformatics* 25:2700–2707.
15. Ahdesmäki M, Strimmer K (2009). Feature selection in “omics” prediction problems using cat scores and false non-discovery rate control. *Ann Appl Stat* (in press, preprint, arXiv:0903.2003).
16. Eveillard A, Lasserre F, de Tayrac M, Polizzi A, Claus S, Canlet C, Mselli-Lakhal L, Gotardi G, Paris A, Guillou H, Martin PG, Pineau T (2009). Identification of potential mechanisms of toxicity after di-(2-ethylhexyl)-phthalate (DEHP) adult exposure in the liver using a systems biology approach. *Toxicol Appl Pharmacol* 236:282–292.
17. Hunt PA, Susiarjo M, Rubio C, Hassold TJ (2009). The bisphenol A experience: a primer for the analysis of environmental effects on mammalian reproduction. *Biol Reprod* 81:807–813.
18. Wetherill YB, Akingbemi BT, Kanno J, McLachlan JA, Nadal A, Sonnenschein C, Watson CS, Zoeller RT, Belcher SM (2007). *In vitro* molecular mechanisms of bisphenol A action. *Reprod Toxicol* 24:178–198.
19. Domange C (2008). Etude des disruptions métaboliques chez le rongeur suite à l'ingestion d'*Hypochaeris radicata* (L.), plante toxique pour l'espèce équine. Validation de l'approche métabonomique par des études comportementales et par imagerie cérébrale. Thèse de doctorat de l'Université de Toulouse. 5 novembre 2008.
20. Viola A, Nicoli F, Denis B, Confort-Gouny S, Le Fur Y, Ranjeva JP *et al.* (2004). High cerebral scyllo-inositol: a new marker of brain metabolism disturbances induced by chronic alcoholism. *MAGMA* 17:47–61.
21. Griffith HR, den Hollander JA, Stewart CC, Evanochko WT, Buchthal SD, Harrell LE *et al.* (2007). Elevated brain scyllo-inositol concentrations in patients with Alzheimer's disease. *NMR Biomed* 20:709–716.
22. Hirano F, Tanaka H, Okamoto K, Makino Y, Inaba M, Nomura Y *et al.* (1995). Natural course of diabetic peripheral neuropathy in spontaneous-onset diabetic Chinese hamsters. *Diabetes Res Clin Pract* 28:151–159.
23. Domange C, Canlet C, Traoré A, Biélicki G, Keller C, Paris A *et al.* (2008). Orthologous metabolomic qualification of a rodent model combined with magnetic resonance imaging for an integrated evaluation of the toxicity of *Hypochoeris radicata*. *Chem. Res Toxicol* 21:2082–2096.
24. Priymenko N, Canlet C, Gottardi G, Molina J, Delous C, Chauvin L *et al.* (2006). Etude métabonomique en RMN du ¹H de l'adaptation de rats en croissance à un régime préparé à partir de pommes de terre génétiquement modifiées. 2^{èmes} journées scientifiques du Réseau Français de Métabonomique et Fluxomique. 13–15 décembre 2006, Saint-Sauves d'Auvergne.
25. Coen M, Holmes E, Lindon JC, Nicholson JK (2008). NMR-based metabolic profiling and metabolomic approaches to problems in molecular toxicology. *Chem Res Toxicol* 21:9–27.
26. Bylesjö M, Rantalainen M, Nicholson JK, Holmes E, Trygg J (2008). K-OPLS package: kernel-based orthogonal projections to latent structures for prediction and interpretation in feature space. *BMC Bioinformatics* 9:106.
27. Bourqui R, Cottret L, Lacroix V, Auber D, Mary P, Sagot MF, Jourdan F (2007). Metabolic network visualization eliminating node redundancy and preserving metabolic pathways. *BMC Syst Biol* 1:29.

