Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Targeting the untargeted in molecular phenomics with structurally-selective ion mobility-mass spectrometry

Jody Christopher May, Randi Lee Gant-Branum and John Allen McLean

Systems-wide molecular phenomics is rapidly expanding through technological advances in instrumentation and bioinformatics. Strategies such as structural mass spectrometry, which utilizes size and shape measurements with molecular weight, serve to characterize the sum of molecular expression in biological contexts, where broad-scale measurements are made that are interpreted through big data statistical techniques to reveal underlying patterns corresponding to phenotype. The data density, data dimensionality, data projection, and data interrogation are all critical aspects of these approaches to turn data into salient information. Untargeted molecular phenomics is already having a dramatic impact in discovery science from drug discovery to synthetic biology. It is evident that these emerging techniques will integrate closely in broad efforts aimed at precision medicine.

Address
Department of Chemistry, Center for Innovative Technology, Vanderbilt Institute of Chemical Biology, Vanderbilt Institute for Integrative Biosystems Research and Education, Vanderbilt University, Nashville, TN 37235, USA

Corresponding author: McLean, John Allen (john.a.mclean@vanderbilt.edu)

Current Opinion in Biotechnology 2016, **39**:192–197

This review comes from a themed issue on **Systems biology**

Edited by **Mark P Styczynski** and **Fabian J Theis**

For a complete overview see the Issue and the Editorial

Available online 29th April 2016

http://dx.doi.org/10.1016/j.copbio.2016.04.013

Introduction

One of the emerging paradigms in systems, synthetic, and chemical biology is the ability to discern the entire molecular expression of a phenotypic sample, termed molecular phenomics, whereby the sum of factors influencing the biology (e.g. exposures, diseases, nutrition, social influences, etc.) are manifest in the consequential molecular landscape corresponding to that phenotype. Massive efforts in molecular phenomics are targeted in precision medicine to tailor prediction of medical outcome and treatment at an individual level [1,2]. To discern the systems-wide response requires massive numbers of measurements, and as such, approaches were largely developed around genome science, where parallel measurements are facilitated using gene-array technologies [3]. In medicine, recent advances include the development of coordinated data workflows correlating genomics data linked to the electronic medical records (EMR) of large cohorts of patients [4,5] and corresponding phenotype [6].

In large part, these efforts have centered on genomics data because of practical technological challenges associated with rapidly and broadly characterizing other molecular types (e.g. proteins, lipids, and metabolites) as a part of the overall diagnostic of biological status. This challenge is considerably exacerbated by the number of samples and/or sampling frequency necessary to characterize the systems-wide molecular complement, because it can vary dramatically depending on location (e.g. cell, tissue, or regional pathology) and time (e.g. health versus disease, disease stage of progression, acute verses chronic exposure). These broad scale measurement challenges are being addressed with advances in several key technologies, chief among them mass spectrometry (MS, [7–9]). These instrumental advances when tied to advances in bioinformatics have resulted in new strategies aimed at molecular phenomics using a breadth of molecular type [10–12], or the sum of molecular expression resulting in specific phenotype. Note that EMR in concert with molecular phenomics measurement is critical in the translation of molecular characterization with predication and interpretation of health status, outcomes, and the development of treatment regimens for improved health. Massive population scale efforts for integrating MS with individual patient phenotyping, in concert with genomics, have recently been described [13]. Molecular phenomics data are then interpreted through identifying the molecular changes that are important to phenotype and/or by mapping the molecular complement onto biological pathways, or networks, to understand alterations in phenotype. Importantly, specific phenotypic studies demonstrating the potential for integrating other data modalities with EMR motivate incorporation of MS data into precision medicine workflows [13]. This report highlights some recent advances in structural MS aimed at obtaining rapid and untargeted molecular information in both spatially-resolved and temporally-resolved analyses and emerging bioinformatics approaches for phenotypic interpretation.

Data density and data dimensionality: opportunities in space and time

To assess molecular phenotype using systems-wide approaches requires cataloging the signals and abundances for as many molecular species as possible, which is the purview of integrated omics studies (i.e. the near simultaneous measurement of proteomics, lipidomics, metabolomics, etc.) using a combination of liquid-chromatography or gas-chromatography coupled with MS-based detection (LC–MS and GC–MS, respectively). Contemporary approaches in omics measurements typically reduce the complexity of the sample to one biomolecular class, where steps for protein or lipid purification are often performed so that the resulting MS signals will correspond primarily to those species, which facilitates identification. However, when multiple classes of molecules are of interest this conventional approach requires the analysis of multiple samples in series and adds significant time to both the sample preparation steps and also to the sample analysis workflows. Structural separations on the basis of ion mobility-MS (IM-MS, [14]) can mitigate this limitation through performing the molecular class separations in the instrument itself with minimal sample pretreatment. This is accomplished in IM-MS through separating these species on the basis of structure where mobility-mass correlations emerge from the data as a result of the structural preferences for different classes of molecules (Figure 1a) [15,16,17••]. As illustrated in Figure 1a, lipids on average adopt larger structures than peptides and likewise peptides adopt larger structures than carbohydrates. Through inspection of collision cross section versus mass-to-charge plots, rapid identification of the different species can be performed in the instrument in parallel without purification of an individual class of molecules before analysis. Recent reports have demonstrated high interlaboratory reproducibility for IM-MS collision cross-section measurements (<5% RSD for metabolites [18•] and <3% RSD for lipids [19•]) and the utility of cataloging correlations for different molecular classes [17••]. The latter is particularly valuable in the cases of glycomics [20••,21•,22–25] and lipidomics [26••,27–29], where many isomeric species of the same elemental composition (i.e. same mass) can be differentiated on the basis of the structural IM separation. For example, fine structure separation and characterization was demonstrated for carbohydrate anomers in glycomics that are indistinguishable using conventional structural analysis [20••]. Moreover, additional MS fragmentation and IM structural separation facilitated on-line sequencing and stereoanalysis of glycans and to a lesser degree glycopeptides [20••,21•,22], demonstrating unexpected promiscuity of glycoenzymes toward sugar donors during glycan biosynthesis and suggesting the presence of glycan microheterogeneity not previously observed [21•]. In recognition of this enhanced information for carbohydrate characterization, recent efforts have focused on generating reference collision cross section datasets for carbohydrate structures to support glycomics and integrated omics analyses [17••,22–24]. In lipidomics, collisional cross section differences of less than 1% were baseline resolved in fine structure analysis and complex mixture analysis [29] providing the opportunity for eliminating multi-step purification for lipidomic analysis and have been demonstrated in biological contexts such as mouse uteri [26••], rat adipose tissue [28], among many others.

IM-MS can be particularly advantageous in imaging MS applications, where information regarding the heterogeneous spatial distribution of molecular species is desired [30]. To obtain spatially-resolved molecular characterization, imaging IM-MS [31] adds positional location information, for example where anatomical integrity is relevant to understanding the mechanism of action (e.g., the pharmacokinetics of a drug or toxin [32]) or physiologically-isolated changes within a living system (e.g., tumor formation [33]) are the overarching aims of the study. Where LC–MS approaches are conventionally used to reduce chemical complexity, it is challenging to perform on a pixel-by-pixel basis in imaging MS contexts. One approach is to use high mass resolving power and high mass accuracy imaging MS [34], whereby accurate mass information supports molecular identification. Chemical complexity can also be mitigated by combining the speed of IM separations with imaging MS without a concomitant increase in analysis time [35–37]. This configuration affords improved molecular selectivity with minimal sample pretreatment and has been demonstrated for tissue imaging applications [26••,38,39], and paraffin-embedded tumor sections [40,41]. The ability to use imaging IM-MS to distinguish two nominally same mass protein species is illustrated in Figure 1b, where without the capacity to separate these species before mass analysis would result in a single mass image comprised of a mixture of both species.

These same attributes in IM-MS, namely integrated omics and analysis speed, also directly facilitate fast time-resolution in longitudinal measurements. In this context, time may represent massive sample sets of patient samples [42], or real-time molecular phenotypic data in response to perturbation. Structural MS strategies have found great utility in performing near real-time read out through integrating 3D organotypic microfluidic devices in so called 'human-on-a-chip' studies [43–45]. These types of studies were initially developed through a recognition that a major inefficiency in pharmaceutical development is the paradigm of moving drug candidates from *in vitro* to animal studies to human trials. A significant fraction of drug candidates are deemed inappropriate to move to humans from the animal studies, because of differences in physiology. Likewise, a significant fraction of drug candidates that potentially could be efficacious in humans are also deemed inappropriate to move to humans, because they fail in the animal model, again because of differences in physiology. Thus, human organoids to replace the animal models in this workflow are of critical importance. In these strategies, the synthetic