Expert Review of Proteomics

ISSN: 1478-9450 (Print) 1744-8387 (Online) Journal homepage: www.tandfonline.com/journals/ieru20

Proteomics and the Human Metabolome Project

David S Wishart

To cite this article: David S Wishart (2007) Proteomics and the Human Metabolome Project, Expert Review of Proteomics, 4:3, 333-335, DOI: 10.1586/14789450.4.3.333

To link to this article: https://doi.org/10.1586/14789450.4.3.333

Published online: 09 Jan 2014.

Submit your article to this journal

Article views: 2166

View related articles

Citing articles: 5 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ieru20

For reprint orders, please contact:
reprints@future-drugs.com

Proteomics and the Human Metabolome Project

'While metabolomics tends to focus on small molecules and proteomics focuses on big molecules, these two fields actually share a remarkable number of similarities'

David S Wishart
University of Alberta, Departments of Computing Science & Biological Sciences, University of Alberta, Edmonton, AB, T6G, 2E8 Canada; National Institute for Nanotechnology, 11421 Saskatchewan Drive, Edmonton, AB, T6G 2M9, Canada
Tel.: +1 780 492 0383
Fax: +1 780 492 1071
david.wishart@ualberta.ca

Expert Rev. Proteomics 4(3), 333–335 (2007)

These days, it seems like a new omics word is being invented almost every week [101]. Many of these fanciful terms will never be used by anyone other than the individual who first coined them. However, one relatively new omics term that is likely to be around for quite some time is 'metabolomics'. Metabolomics is a newly emerging field of research concerned with the high-throughput identification and quantification of the small-molecule metabolites in the metabolome [1]. The metabolome (derived from the word metabolite and the Greek suffix 'ome', meaning collection or body) can be thought of as the complete complement of all small-molecule (<1500 Da) metabolites found in a specific cell, organ or organism. These small molecules can include chemical entities, such as peptides, amino acids, nucleic acids, carbohydrates, organic acids, vitamins, minerals and just about any other chemical that can be used, ingested or synthesized by a given cell or organism.

What is often forgotten by those of us involved in either genomics or proteomics is that metabolites are intimately connected to every part of a biological system. Metabolites serve as markers of both the proteome and genome. Indeed, just a single amino acid change in a protein or a single base change in a gene can lead to as much as a 10,000-fold change in the concentration of certain metabolites [2]. The importance of small molecules in medicine and biology cannot be overemphasized. Consider the following facts:

- Over 95% of all diagnostic clinical assays test for small molecules [3]
- 89% of known drugs are small molecules [102]
- 50% of all drugs are derived from pre-existing metabolites [4]
- 30% of identified genetic disorders involve diseases of small-molecule metabolism [5]

Furthermore, almost all of the leading (and most costly) causes of chronic disease and morbidity arise from adverse interaction of small molecules with our genome or proteome [6–8]. These include obesity (dietary sugars or fats), diabetes (dietary sugars), heart disease (dietary fats and cholesterol), cancer (pollutants or mutagens) and adverse drug reactions (drugs or drug metabolites).

Connection between metabolomics & proteomics

While metabolomics tends to focus on small molecules and proteomics focuses on big molecules, these two fields actually share a remarkable number of similarities. Both fields use similar kinds of equipment to detect or characterize their favorite molecules, including mass spectrometers, nuclear magnetic resonance (NMR) spectrometers, x-ray diffractometers and high-pressure liquid chromatography (LC) systems. Furthermore, both fields need comprehensive databases that can be searched by character strings, structure files, masses or spectral profiles. Additionally, both fields are aimed at identifying the interacting partners, networks and pathways for their respective classes of molecules. Likewise, both disciplines are concerned with accurately identifying and quantifying different molecules in different organs or tissues, under different conditions or for different diseases (i.e., biomarker identification). Finally, both fields

10.1586/14789450.4.3.333 © 2007 Future Drugs Ltd ISSN 1478-9450

are keenly focused on characterizing and cataloging the 2D and 3D structure of all their respective target molecules. Given these similarities, it is not surprising that many researchers in the field of metabolomics have also been quite active in the field of proteomics.

Beyond these methodological similarities, it is also important to remember that the biological connection between proteins and metabolites runs very deep. For instance, all proteins are composed of amino acids (i.e., metabolites). Likewise, most proteins (ingested or endogenous) are decomposed into metabolites (including essential amino acids, urea, trimethylamine *N*-oxide and creatinine). It is also worth noting that the source molecules for almost all protein post-translational modifications are naturally occurring metabolites (phosphates, sugars, alcohols, organic acids and aldehydes). Furthermore, the vast majority of proteins use metabolites as cofactors (NADH or flavin AD), signaling molecules (calcium, nitric oxide, zinc and prostaglandins), substrates or stabilizing reagents. Indeed, most proteins could not function, form complexes, recognize substrates or remain folded without these small-molecule partners. In other words, the proteome is designed to act on the metabolome and the metabolome is designed to act on the proteome.

'A particular challenge for any species-specific metabolomics effort is defining what should be included in the metabolome.'

The Human Metabolome Project

Over the past decade, a central goal in both proteomics and genomics has been to systematically catalogue the genomes and proteomes of selected model organisms. Not surprisingly, a central focus over the past few years in the field of metabolomics has been to do exactly the same thing for metabolites. The thinking behind these large-scale metabolomic projects is similar to the rationale for the Human Genome Project [9]. More specifically, a detailed catalog of all the genes/proteins/metabolites in a given organism would be expected to have a transformative effect on the molecular understanding of the biological processes or diseases associated with that organism. Currently, there are major efforts to characterize the metabolomes of *Arabidopsis* [10], *Escherichia coli* [11,12] and humans [13,14]. One of these efforts, the Human Metabolome Project (HMP), is attempting to identify and catalog all of the metabolites found in the human body. More specifically, this project has three major goals:

- To complete a metabolite inventory for human beings
- To generate resources that can facilitate metabolomics research across many different disciplines
- To provide detailed information about the linkage between human metabolites and the genes, proteins and pathways with which they are involved

The HMP combines both dry-lab (text mining and bioinformatics) and wet-lab (NMR, LC mass spectrometry [MS], gas chromatography MS) approaches to identify, quantify and validate known and previously unknown metabolites. To date, the HMP has identified 1100 drugs, 2600 endogenous metabolites and 2100 food additives or plant-derived products. This collection of compounds constitutes the first draft of the human metabolome. Information about these compounds and their structure, protein/enzyme targets, tissue/biofluid concentrations, disease associations and general biology is contained in three publicly available databases: the Human Metabolome Database (HMDB) [14], DrugBank [15] and FooDB [103]. In addition to these electronic databases, the HMP has also assembled (through purchases, acquisitions and syntheses) a large chemical library of nearly 1000 authenticated small-molecule compounds; ranging from 1 mg to 1 g of each sample. This collection, the Human Metabolome Library (HML), is the largest metabolite resource in the world.

A particular challenge for any species-specific metabolomics effort is defining what should be included in the metabolome:

- Should the metabolome be restricted to only endogenous compounds?
- Should it include exogenous drugs and drug metabolites?
- Should it include plant-derived food components or chemical food additives?
- Is it appropriate to include hypothetical compounds (i.e., compounds we think should be there but we have no proof of their presence)?
- What is the upper size or upper molecular weight limit for something to be called a metabolite (<1500 Da)?
- Should the metabolome be restricted to compounds that can be practically detected or detectable?

Unlike the genome, which is a clearly defined entity, the metabolome (like the proteome) has many definitions for many different people. This appears to be a source for both confusion and dispute within the metabolomics community [16]. Hopefully, the introduction of data standards and the establishment of dedicated bodies, such as the Metabolomics Society, will go a long way to resolving this issue.

What does the future hold?

The determination of a first draft of the human metabolome, along with the availability of the HML, will likely have a significant impact in a number of areas of proteomics and genomics. For instance, in the area of enzyme genomics, the HML should serve as an excellent source of substrates for large-scale enzymatic screens and for the discovery of new enzymes or enzyme functions [17]. Because so many metabolites are actually cofactors and ligands for proteins and enzymes, the HML could also find considerable utility in the field of structural proteomics (i.e., structural genomics). In particular, these compounds could be used in protein crystallization screens for x-ray crystallography or high-throughput protein stability surveys for NMR spectroscopy. Functional proteomics may also benefit from these ever-expanding metabolite libraries. In particular, it may be possible to perform rapid ligand-binding assays to identify the functions of unknown proteins using techniques, such as functional

annotation screening technology NMR [18]. Given that nearly 50% of known drugs are metabolites or derivatives of metabolites [4], it is likely that our expanded knowledge of metabolites and protein–metabolite interactions will lead to the discovery of many new drug leads – either through computational methods or standard metabolite screening approaches [19].

Because of the close technical and conceptual overlaps between proteomics and metabolomics, it is likely that the next generation of separation, detection and characterization technologies will greatly improve both proteomics and metabolomics. Certainly, technical developments in ultra-high-pressure LC, Fourier transform MS, Orbi-Trap™ MS, higher field NMR magnets and cold-probe NMR have already had a significant impact in both fields. Likewise, continuing developments in lab-on-a-chip technologies can be expected to make both metabolomics and proteomics much cheaper, significantly faster and far more accessible – especially in the area of clinical chemistry. Over the next few years, it is also likely that the blending of proteomics with metabolomics in clinical research will become much more common. Indeed, the use of metabolomic and proteomic data to identify important disease biomarkers will likely prove to be far more powerful than using either metabolites alone or proteins alone.

'Given that nearly 50% of known drugs are metabolites or derivatives of metabolites, it is likely that our expanded knowledge of metabolites and protein–metabolite interactions will lead to the discovery of many new drug leads.'

No doubt there are many other sanguine predictions about the prospects for metabolomics and the human metabolome [16]. However, it is important to remember that metabolomics will not be a panacea to solve all (or even most) biomedical problems. Indeed, it may be many years before metabolomics reaches the same level of acceptance and prevalence already achieved by genomics and proteomics. As with any new technology or any new 'ome' that is completed, there is a certain amount of hype that often creates an unreasonable expectation about its potential or promise. While the recent determination of a draft of the human metabolome clearly offers a number of exciting prospects, one must always remember that metabolites are only a small part of the picture, reflecting changes at many levels, which arise from innumerable interactions involving the genome, proteome and transcriptome. At the very least, the human metabolome offers a new perspective on biology that may provide a more complete picture of how our bodies really work [20].

References

1 German JB, Hammock BD, Watkins SM. Metabolomics: building on a century of biochemistry to guide human health. *Metabolomics* 1, 3–9 (2005).

2 Bory C, Boulieu R, Chantin C, Mathieu M. Diagnosis of alcaptonuria: rapid analysis of homogentisic acid by HPLC. *Clin. Chim. Acta* 189, 7–11 (1990).

3 Tietz NW. *Clinical Guide to Laboratory Tests, 3rd Edition.* WB Saunders Press, PA, USA (1995).

4 Mahido C, Ruchirawat S, Prawat H *et al.* Biodiversity and natural product drug discovery. *Pure Appl. Chem.* 70, 2065–2072 (1998).

5 Hamosh A, Scott AF, Amberger J *et al.* Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 30, 52–55 (2002).

6 Katzmarzyk PT, Janssen I. The economic costs associated with physical inactivity and obesity. *Can. J. Appl. Physiol.* 29, 90–115 (2004).

7 Johnson JA, Pohar SL, Majumdar SR. Health care use and costs in the decade after identification of Type 1 and Type 2 diabetes: a population-based study. *Diabetes Care* 29, 2403–2408 (2006).

8 Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA* 279, 1200–1205 (1998).

9 Watson JD. The human genome project: past, present, and future. *Science* 248, 44–49 (1990).

10 Kaplan F, Kopka J, Haskell DW *et al.* Exploring the temperature–stress metabolome of *Arabidopsis. Plant Physiol.* 136, 4159–4168 (2004).

11 Maharjan RP, Ferenci T. Global metabolite analysis: the influence of extraction methodology on metabolome profiles of *Escherichia coli*. *Anal. Biochem.* 313, 145–154 (2003).

12 Rabinowitz JD. Cellular metabolomics of *Escherchia coli*. *Expert Rev. Proteomics* 4, 187–198 (2007).

13 Smith CA, O'Maille G, Want EJ *et al.* METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* 27, 747–751 (2005).

14 Wishart DS, Knox C, Guo A *et al.* DrugBank: a comprehensive resource for *in silico* drug discovery and exploration. *Nucleic Acids Res.* 34(Database issue), D668–D672 (2006).

15 Wishart DS, Tzur D, Knox C *et al.* HMDB: the Human Metabolome Database. *Nucleic Acids Res.* 35(Database issue), D521–D526 (2007).

16 Pearson H. Meet the human metabolome. *Nature* 446, 8 (2007).

17 Kuznetsova E, Proudfoot M, Sanders SA *et al.* Enzyme genomics: application of general enzymatic screens to discover new enzymes. *FEMS Microbiol. Rev.* 29, 263–279 (2005).

18 Mercier KA, Baran M, Ramanathan V *et al.* FAST-NMR: functional annotation screening technology using NMR spectroscopy. *J. Am. Chem. Soc.* 28, 15292–15299 (2006).

19 Senisterra GA, Markin E, Yamazaki K *et al.* Screening for ligands using a generic and high-throughput light-scattering-based assay. *J. Biomol. Screen.* 11, 940–948 (2006).

20 Duarte ND, Becker SA, Jamshidi N *et al.* Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl Acad. Sci. USA* 104, 1777–1782 (2007).

Websites

101 Omes and Omics Glossary
www.genomicglossaries.com/content/omes.asp

102 DrugBank
www.drugbank.ca

103 FooDB
http://hmdb.med.ualberta.ca/foodb

Affiliation

- *David S Wishart*
University of Alberta, Departments of Computing Science & Biological Sciences, University of Alberta, Edmonton, AB, T6G, 2E8 Canada; National Institute for Nanotechnology, 11421 Saskatchewan Drive, Edmonton, AB, T6G 2M9, Canada
Tel.: +1 780 492 0383
Fax: +1 780 492 1071
david.wishart@ualberta.ca