Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

# Advances in computational metabolomics and databases deepen the understanding of metabolisms

Hiroshi Tsugawa[1,2]

Mass spectrometry (MS)-based metabolomics is the popular platform for metabolome analyses. Computational techniques for the processing of MS raw data, for example, feature detection, peak alignment, and the exclusion of false-positive peaks, have been established. The next stage of untargeted metabolomics would be to decipher the mass fragmentation of small molecules for the global identification of human-, animal-, plant-, and microbiota metabolomes, resulting in a deeper understanding of metabolisms. This review is an update on the latest computational metabolomics including known/expected structure databases, chemical ontology classifications, and mass spectrometry cheminformatics for the interpretation of mass fragmentations and for the elucidation of unknown metabolites. The importance of metabolome 'databases' and 'repositories' is also discussed because novel biological discoveries are often attributable to the accumulation of data, to relational databases, and to their statistics. Lastly, a practical guide for metabolite annotations is presented as the summary of this review.

**Addresses**
[1] RIKEN Center for Sustainable Resource Science, Japan
[2] RIKEN Center for Integrative Medical Sciences, Japan

Corresponding author: Tsugawa, Hiroshi (hiroshi.tsugawa@riken.jp)

**Current Opinion in Biotechnology** 2018, **54**:10–17

This review comes from a themed issue on **Analytical biotechnology**

Edited by **Hiroshi Shimizu** and **Fumio Matsuda**

https://doi.org/10.1016/j.copbio.2018.01.008

0958-1669/© 2018 Elsevier Ltd. All rights reserved.

## Why is untargeted metabolomics needed in biology?

Under the central dogma, the genome, transcriptome, and proteome are presented in terms of a 'signal flow' and the metabolome is considered the 'result' in metabolism. However, many studies have reported that the metabolites themselves are deeply involved in the physiological functions and homeostasis of living organisms. Examples are first, oxylipins [1], a oxidized fatty acids group that acts as bioactive metabolites in, for example, inflammatory responses and defense systems; second, oncometabolites [2–3], unexpected products from altered metabolism that are involved in tumorigenesis; third, damaged metabolites [4], chemically reactive compounds resulting from enzyme errors or spontaneous reactions that are normally regulated by damage-control systems; fourth, microbiota metabolites [5], metabolites secreted by gut microbiota affecting the host physiology; and finally phytochemicals [6], the plant specialized metabolites exerting various bioactivities on human metabolisms (Figure 1).

Mass spectrometry (MS)-based untargeted metabolomics has led to the discovery of these metabolites and updates on analytical chemistry and its informatics are essential for the elucidation of new physiological function and biological mechanisms.

## What is needed to improve untargeted metabolomics?

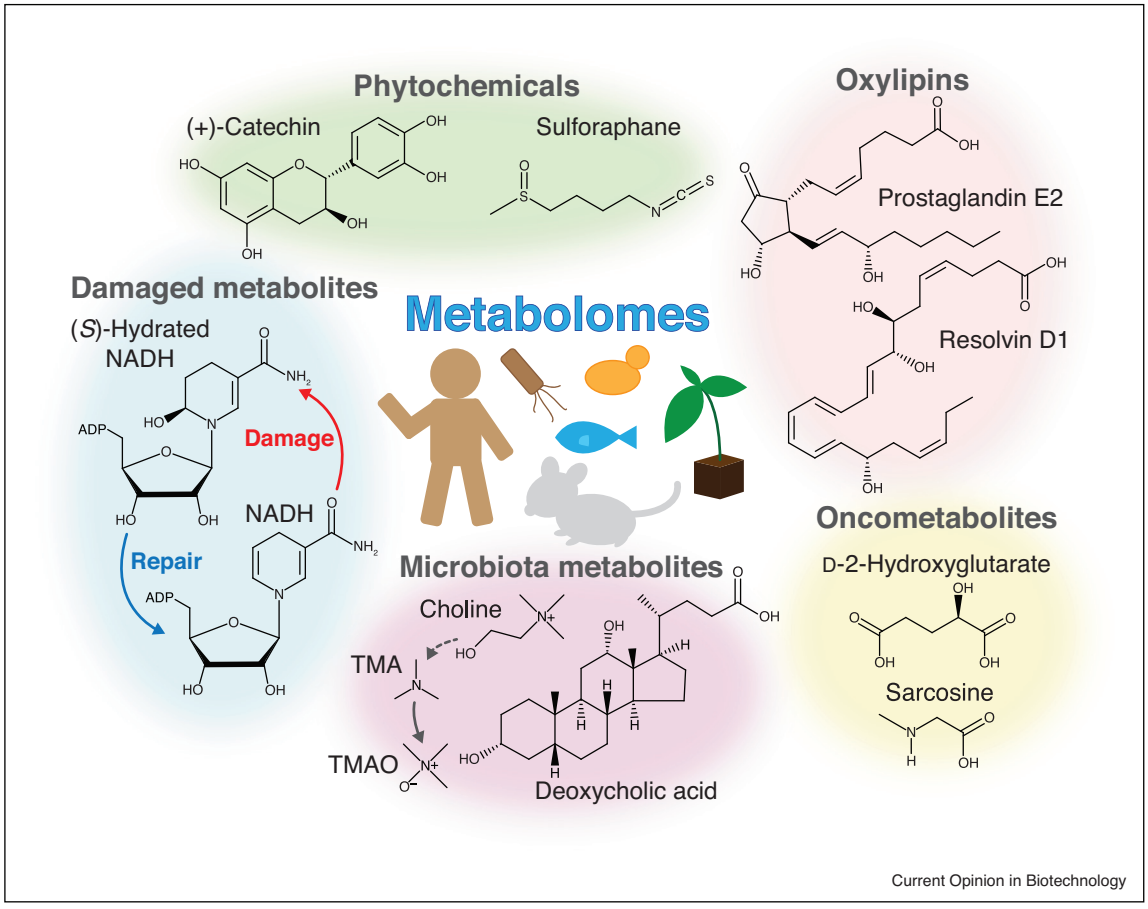
The handling of MS raw data, for example, feature detection, chromatogram deconvolution, isotope recognition, chromatogram alignment, and the exclusion of false-positive peaks is now a mature technique for untargeted metabolomics: of course, the advances also enhance the efficiency for biological discoveries. Software programs such as MS-DIAL [7], MZmine [8], XCMS [9], OpenMS [10], and other specialized programs for metabolomics and lipidomics are used as the pipeline of the metabolomics workflow [11–12]; the favorite program can be used while considering their advantages and disadvantages.

The biggest challenge is the decoding of physics/chemical phenomena of ionized metabolites such as ion interactions [13] (e.g. dimers, adduct ions) and mass fragmentations including in-source fragmentation and low-energy collision-induced dissociation-based fragmentations in mass spectrometers [14•]. Such knowledge will make ion feature detection more efficient and facilitate the global identification of metabolites in living organisms. To date, the 'computational mass fragmentation' using cheminformatics platforms like chemistry development kits [15] are the popular technique to assist the interpretation of mass fragmentations and to elucidate unknown structures with metabolome databases and repositories [16], which is presented below.

## Cheminformatics using spectral databases and structure databases

First, the current MS/MS spectral and biologically reported/expected structure databases were examined

**Figure 1**




Metabolomes linked to physiological functions. The screening of metabolomes is frequently performed by untargeted metabolomics. Bioactive metabolites are validated by targeted analysis for stereoisomer determinations in combination with other analytical platforms such as nucleic magnetic resonance (NMR) and X-ray. The abbreviations TMA and TMAO mean that trimethylamine and trimethylamine *N*-oxide, respectively.

for this review. The statistics was performed by RIKEN internal MS/MS spectral databases including our internal database, MassBank, GNPS, Metlin, ReSpect, and NIST14 (for spectrum count) and the structure databases of MS-FINDER version 2.24 [17••] that include 15 metabolome structure databases (for structure counts). As a result, 226,204 unique compounds were stored in the metabolome structure database whereas the MS/MS spectrum for 7195 compounds of these was recorded in the spectral database, where the first layer of InChIKey was used as the query. Computational metabolomics attempts to fill the large 'gap' between spectrum and structure counts. For a better understanding of the required technologies, the 'metabolome' is divided to four classes in this review, firstly, 'Known Structure-Known Spectrum (KS-KS)' where the reported structure is confirmed by the experimental MS/MS spectrum; secondly, 'Known Structure-Unknown Spectrum (KS-US)' where the biologically examined (or partially expected) structures for which the spectrum is not validated by standard compounds; thirdly, 'Unknown Structure-Known Spectrum (US-KS)' where the mass spectrum itself is frequently monitored in biological samples but the structure is not elucidated or reported in life-science papers; and finally, 'Unknown Structure-Unknown Spectrum (US-US)' where the putative dark matter of small molecules is unknown [18].

The identification of KS-KS metabolites is relatively easy with the aid of EI-MS and MS/MS matching algorithms [19,20•,21,22] combined with retention-time predictions [23,24,25•], and by means of the internal standards. Notably, study-dependent false discovery rate (FDR) estimations have recently been proposed in metabolomics [26••] while a platform-independent annotation rule of lipids has been proposed in lipidomics [27••]; they may facilitate the full automation of the metabolomics/lipidomics workflow.

A challenge in mass spectrometry cheminformatics is the annotation of KS-US and US-KS metabolites, and it has been met by three major computational approaches: the

Download English Version:

https://daneshyari.com/en/article/6487205

Download Persian Version:

https://daneshyari.com/article/6487205

Daneshyari.com