Metabolic Engineering 23 (2014) 1–8

Contents lists available at ScienceDirect

Metabolic Engineering

journal homepage: www.elsevier.com/locate/ymben

A computational framework for integration of lipidomics data into metabolic pathways

Noushin Hadadi [a,b], Keng Cher Soh [a,b], Marianne Seijo [a,b], Aikaterini Zisaki [a,b], Xueli Guan [c], Markus R. Wenk [c,d], Vassily Hatzimanikatis [a,b,*]

[a] Laboratory of Computational Systems Biotechnology, École Polytechnique Fédérale de Lausanne, CH 1015 Lausanne, Switzerland
[b] Swiss Institute of Bioinformatics, CH-1015 Lausanne, Switzerland
[c] Department of Biological Sciences, National University of Singapore, Singapore 117543, Singapore
[d] Swiss Tropical and Public Health Institute, 4002 Basel, Switzerland

ARTICLE INFO

Article history:
Received 19 September 2013
Received in revised form
26 November 2013
Accepted 24 December 2013
Available online 4 January 2014

Keywords:
Phospholipids
Database
Bioinformatics
Bis(monoacylglycero)phosphate (Bmp)

ABSTRACT

Lipids are important compounds for human physiology and as renewable resources for fuels and chemicals. In lipid research, there is a big gap between the currently available pathway-level representations of lipids and lipid structure databases in which the number of compounds is expanding rapidly with high-throughput mass spectrometry methods.

In this work, we introduce a computational approach to bridge this gap by making associations between metabolic pathways and the lipid structures discovered increasingly thorough lipidomics studies. Our approach, called NICELips (*N*etwork *I*ntegrated *C*omputational *E*xplorer for *Lip*idomics), is based on the formulation of generalized enzymatic reaction rules for lipid metabolism, and it employs the generalized rules to postulate novel pathways of lipid metabolism. It further integrates all discovered lipids in biological networks of enzymatic reactions that consist their biosynthesis and biodegradation pathways.

We illustrate the utility of our approach through a case study of bis(monoacylglycero)phosphate (BMP), a biologically important glycerophospholipid with immature synthesis and catabolic route(s). Using NICELips, we were able to propose various synthesis and degradation pathways for this compound and several other lipids with unknown metabolism like BMP, and in addition several alternative novel biosynthesis and biodegradation pathways for lipids with known metabolism. NICELips has potential applications in designing therapeutic interventions for lipid-associated disorders and in the metabolic engineering of model organisms for improving the biobased production of lipid-derived fuels and chemicals.

1. Introduction

Lipids are hydrophobic organic molecules that include some common compounds such as fats, oils, waxes, phospholipids and steroids (e.g., cholesterol). In addition, they include a large range of compounds with special structures and functions. Lipids have many key biological functions in living cells such as participating in signaling pathways, acting as energy storage sources and being the structural components of cell membrane. Moreover, they play an important role in various diseases such as obesity, diabetes and cancer (Murphy and Nicolaou, 2013). Lipid-associated disorders (or lipidoses) are a group of inherited metabolic disorders in which harmful amounts of lipids (fats) accumulate in some of the body's cells and tissues as a reason of defects in the biosynthesis or biodegradations of simple lipids (Albinet et al., 2013; Pralhada Rao et al., 2013). Moreover, lipids are important compounds as sources for biofuels and renewable energy (Karunanandaa et al., 2005; Runguphan and Keasling, 2013).

The term lipidome refers to the full lipid complement of cells, tissues and organisms. Lipidomics – the large-scale study of pathways and networks of cellular lipids in biological systems – aims to elucidate and characterize the lipidome (Watson, 2006; Wenk, 2005). Mass spectrometry (MS) is the most commonly used analytical method in lipidomics research (Hermansson et al., 2005; Lagarde et al., 2003; Zehethofer and Pinto, 2008). With the rapid

* Corresponding author at: Ecole polytechnique fédérale de Lausanne, Institut des sciences et ingénierie chimiques, EPFL SB ISIC LCSB, CH H4 625 (Bât. CH), Station 6, CH-1015 Lausanne, Switzerland.
E-mail addresses: noushin.hadadi@epfl.ch (N. Hadadi), kengcher.soh@epfl.ch (K. Cher Soh), marianne.seijo@epfl.ch (M. Seijo), aikaterini.zisaki@epfl.ch (A. Zisaki), xueli.guan@unibas.ch (X. Guan), bchmrw@nus.edu.sg (M.R. Wenk), vassily.hatzimanikatis@epfl.ch (V. Hatzimanikatis).

1096-7176/$ - see front matter © 2014 International Metabolic Engineering Society. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ymben.2013.12.007

growth in analytical technologies, in particular MS, vast amount of data is being generated for lipid structures (Sokol et al., 2013). Therefore, there is a need for developing comprehensive computational tools for data mining and system level identification of lipid species and organizing them into databases (Yetukuri et al., 2007).

Although there are a number of ongoing discoveries in lipidomics, assisted by functional genomics and biophysical studies, still a large number of structures and functions need to be discovered or clarified in lipid metabolism (Khalil et al., 2010; Oresic, 2009). Due to the diversity of possible chemical structures in lipids, bioinformatics is indispensable for accelerating the discovery of novel structures and their corresponding biosynthetic and catabolic pathways. One of the most diverse and ubiquitous classes of lipids is glycerophospholipids, and they constitute the majority of the common class called phospholipids. Phospholipids can be found in almost all organisms as they are the building blocks of cell membranes and are known to be biologically active (Nohturfft and Zhang, 2009; Pulfer and Murphy, 2003). Glycerophospholipids contain a polar head and a glycerol core with fatty acids attached to the glycerol moiety (Hanahan, 1997). They may be subdivided into distinct subclasses, based on the nature of the polar head group at the sn-3 position of the glycerol backbone (Fahy et al., 2005) (Fig. 1). The significant structural diversity of glycerophospholipids arises from (i) the variations in the head group, (ii) the length of the acyl chains and (iii) their degree of saturation in the acyl chains (Ejsing et al., 2006; Fahy et al., 2005; Wenk, 2010). As such, the lipidomics analysis of phospholipids is difficult due to their enormous structural diversity and also high hydrophobicity.

Until recently, there were few specialized databases focusing on lipids analysis and classifications. In 2007, LIPID MAPS structure Database (LMSD) became available (Fahy et al., 2007). LMSD is a comprehensive database for lipid structures that currently contains over 37,000 different classified structures with all the relevant physical and chemical information. However, a comprehensive database of lipid structure is not enough to fully understand their multiple biological roles in cell biology and pathology. To further clarify their functions and the enzymes related to their metabolism, it is essential to organize them in the context of biological pathways and derive their associations and interactions with enzymes and other lipids.

For the pathway level representation of lipids, KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa and Goto, 2000) is the most comprehensive database available for small molecules and contains biological pathway maps for different parts of lipid metabolism. However, lipid biological pathways in KEGG are limited to general lipid species and do not include all the lipid structures available through LMSD. Therefore, the growth rate of these two databases is different and this results in the creation of a big gap between them.

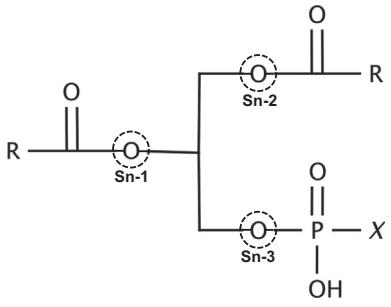

Fig. 1. General structure of glycerophospholipids. In most glycerophospholipids the phosphate (on sn-3 position) is one of the following polar head groups: serine, choline, ethanolamine, glycerol, or inositol (designated X at right). The acyl chains are shown with "R" in the sn-2 and sn-1 positions and can have different length and degrees of saturation.

The needs and limitations in lipids bioinformatics motivated us to develop a computational framework, NICELips, to generate associations between KEGG and LMSD databases and to enrich our knowledge of lipid metabolism. NICELips consists of several components integrated into a workflow and it is the first tool to provide an efficient and consistent procedure for linking lipid compound databases, such as LMSD, with pathway databases, such as KEGG. The central component involves the generation and reconstruction of lipid structures and metabolic reactions. Within this component we identify all the known enzymatic reactions of lipid metabolism in the KEGG database, and we formulate the generalized reaction rules for lipid reactions based on the molecular signatures of known enzymatic reactions. We then apply the generalized reaction rules through two different schemes. In the first scheme, we use a reaction mechanism generation algorithm and we apply the generalized reaction rules on a set of starting compounds. Using this scheme we reconstruct a comprehensive network of lipid metabolism that includes all the known reported structures and reactions in KEGG, but and in addition many novel reactions between KEGG compounds that have not been previously reported in databases. In the second scheme, we apply the retrosynthesis algorithm that uses the reaction rules against LMSD structures to find out their metabolic and catabolic reaction networks.

We illustrate the efficacy and usefulness of our approach in the study of bis(monoacylglycero)phosphate (BMP) metabolism. BMP has two glycerol subunits linked by a phosphodiester group and it is a structural isomer of and very similar to phosphatidylglycerol (PG). This phospholipid is highly enriched in the late endosomes, where it can amount up to 70% of the total phospholipids of the internal membrane (Goursot et al., 2010). BMP is assumed to be important for the structural and functional integrity of the late endosomes. Interestingly, BMP is also a unique lipid due to its stereochemical configuration different from that of other animal glycerophospholipids. Despite numerous studies, there is still missing essential knowledge concerning its properties, and biosynthetic and catabolic pathways. Based on the experimental evidence, BMP is synthesized from its structural isomer, (PG). After the removal of one fatty acid from the sn-2 position, lysophosphatidylglycerol is produced as next intermediate that then undergoes a transacylation reaction (Hullin-Matsuda et al., 2007). The results of the retrosynthesis experiment suggest various synthesis and degradation pathways for this compound.

Our studies here demonstrate how NICELips can provide a full overview of all lipid species in the cell, and particularly in the context of metabolic pathways that comprise all the chemical interactions and transformations between lipid compounds and enzymes. The results of this work have important implications for discovering novel therapeutic approaches for lipid-associated disorders, through proposing novel biosynthetic and biodegradation pathways to alternate the metabolism of genetically defective lipids. On the other hand, exploring the entire space of feasible reactions in lipid metabolism will open up opportunities for generating de novo reactions to design and engineer new metabolic capabilities for the bio-production of lipid-derived fuels and chemicals (Karunanandaa et al., 2005; Runguphan and Keasling, 2013).

2. Methods

The development of the NICELips framework is an extension of the earlier development of the pathway generation framework BNICE (Hatzimanikatis et al., 2005) tailored for the lipid metabolism.