Lipidomics informatics for life-science

D. Schwudke[a], A. Shevchenko[b], N. Hoffmann[c], R. Ahrends[c],[*]

[a] Research Center Borstel, Leibniz Center for Medicine and Biosciences, Borstel, Germany
[b] Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany
[c] Leibniz-Institut für Analytische Wissenschaften − ISAS − e.V., Dortmund, Germany

A B S T R A C T

Lipidomics encompasses analytical approaches that aim to identify and quantify the complete set of lipids, defined as lipidome in a given cell, tissue or organism as well as their interactions with other molecules. The majority of lipidomics workflows is based on mass spectrometry and has been proven as a powerful tool in system biology in concert with other Omics disciplines. Unfortunately, bioinformatics infrastructures for this relatively young discipline are limited only to some specialists. Search engines, quantification algorithms, visualization tools and databases developed by the 'Lipidomics Informatics for Life-Science' (LIFS) partners will be restructured and standardized to provide broad access to these specialized bioinformatics pipelines. There are many medical challenges related to lipid metabolic alterations that will be fostered by capacity building suggested by LIFS. LIFS as member of the 'German Network for Bioinformatics' (de.NBI) node for 'Bioinformatics for Proteomics' (BioInfra.Prot) and will provide access to the described software as well as to tutorials and consulting services via a unified web-portal.

1. Introduction

The aim of lipidomics studies is to establish the identity, quantity and time dependent distribution of lipophilic and amphiphilic metabolites in biological systems (Klose et al., 2013; Wenk, 2005). Lipids are involved in key biological mechanisms, and in recent years demands on analytical and informatics workflows have risen to study the influence of lipid metabolic regulation on the health status of organisms (Klose et al., 2012; Sampaio et al., 2011; Shevchenko and Simons, 2010). Lipidomics in concert with genomics, transcriptomics, and proteomics provides new avenues to study diseases within metabolic syndrome complex (Han, 2016), degenerative diseases (Wang and Han, 2016) and cancerogenesis (Beloribi-Djefaflia et al., 2016) to name just the most prominent fields. Unfortunately, search engines, quantification algorithms, quantification algorithms, visualization, validation and tools for lipidome comparisons exist but are neither streamlined, user friendly nor interconnected. Thus integration of a 'Lipidomics Informatics for Life-Science' unit (LIFS) into the 'German Network for Bioinformatics Infrastructure' (de.NBI) connected to the 'Bioinformatics for Proteomics' hub (BioInfra.Prot) will foster a system biology approaches for studying lipid metabolism. LIFS includes implementation, establishment and provision of bioinformatics services for lipidomics research within one webportal: i) We will provide our existing lipidomics software tools (LipidXplorer (Herzog et al., 2012; Herzog et al., 2011; Herzog et al., 2013), Skyline for

Lipidomics (Peng and Ahrends, 2016), LUX Score (Marella et al., 2015), LipidHome (Foster et al., 2013)). ii) We will extend and integrate our tools in user-friendly web interfaces to offer them to a broader public. iii) We will offer bioinformatics consulting services regarding large scale data handling and managing and iv) we will organize workshops for practitioners and bioinformaticians on lipidomics tools and data analysis and participate in the de.NBI-wide education activities.

2. Lipidomics software tools

2.1. Skyline for lipidomics: a comprehensive platform for targeted assays

Lipidomes comprise an extensive spectrum of chemical structures, which only mass spectrometry (MS)-based techniques provide the means to establish the identity and quantities of most lipids including sphingolipids (Bou Khalil et al., 2010; Lam and Shui, 2013; van Meer, 2005). For untargeted liquid chromatography (LC)-based lipidomics, software solutions such as LipidSearch (Taguchi et al., 2007) and LipidBlast (Kind et al., 2013) are available, and for shotgun lipidomics, direct infusion experiments, the software suite LipidXplorer can be utilized (Herzog et al., 2012). LipidSearch and LipidBlast depend on spectral libraries, whereas LipidXplorer uses de novo spectra interpretation for lipid identification. Besides this, other identification software packages, including LipidQA (Song et al., 2007), LIMSA
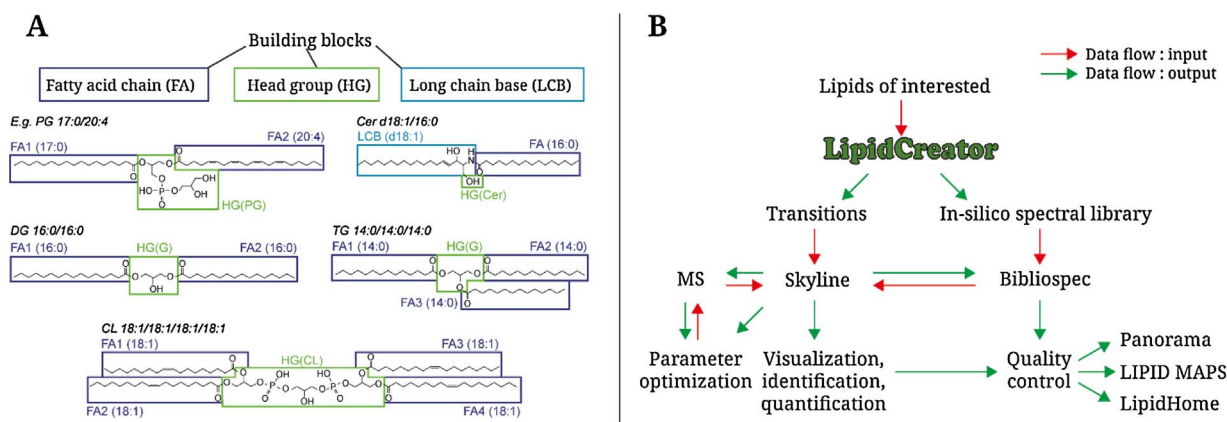
Fig. 1. Skyline for lipidomics: a comprehensive platform for targeted assays. A) Main lipid classes can be assembled from lipid building blocks such as fatty acid chain, long chain base and head group to create the precursor mass their fragment ions. B) Utilization of LipidCreator in the Skyline environment. Lipid information is transferred to Skyline by LipidCreator to compute transition lists and in-silico spectral libraries. Afterwards, Skyline can be used for visualization, optimization, data evaluation and quantification of targeted lipidomics results. Abbreviations: PG (Glycerophosphoglycerol), Cer (Ceramide), DG (Diradylglycerol), TG (Triradylglycerol), CL (Cardiolipin).

(Haimi et al., 2006), FAAT (Leavell and Leary, 2006), lipid (Hubner et al., 2009), LipidInspector (Schwudke et al., 2006b), ALEX (Husen et al., 2013) and Greasy (Kochen et al., 2016) can be applied for lipid identification. Unfortunately, there are currently no open source tools for targeted lipidomics. In response to the urgent need for an analysis software that is capable of handling data from targeted high-throughput lipidomics experiments, we developed a workflow for straightforward method design and analysis of selected and parallel reaction monitoring data. We used the Skyline platform, primarily designed for proteomics applications (MacLean et al., 2010), as a powerful basis to design a specific pipeline for lipid research (Peng and Ahrends, 2016). This extension offers the unique capability to assemble targeted mass spectrometry methods for complex lipids by making use of lipid fragmentation building blocks (Fig. 1A). With simple yet tailored modifications, targeted methods to analyze main lipid classes such as glycerophospholipids, sphingolipids, glycerolipids, cholesteryl-esters, and cholesterol can be quickly introduced into Skyline for easy application by end users without distinct bioinformatics skills. During the de.NBI funding period we will adapt the interim version, which is still working with amino acid based pseudo sequence tags to a stand-alone version (LipidCreator) which can be used as a plugin in Skyline (Fig. 1B). We will implement organism based pre-calculation for lipid species making it convenient for the user to obtain the transitions and target masses of interest. For a better reviewing process and to further accelerate the understanding of lipid fragmentation we will propose a lipid nomenclature connecting the MS1 with the MS2 fragment ion level, by introducing a standardized MS2 nomenclature. This will further help define molecular lipid species-specific fragments that provide information about the chemical composition of the fatty acyl chain of individual lipid molecules, such as long chain bases (LCBs) or fatty acids linked by ester, ether or vinyl-ether bond to their individual lipid backbone. We will seek feedback from and input by the scientific community, particularly from the LIPID MAPS consortium (Fahy et al., 2007; Sud et al., 2007) and HUPO PSI (Kaiser, 2002) initiative on the proposed nomenclature.

In summary, by the end of the funding period we will present a user-friendly "Plug and Play" workflow for lipidomics in response to the demand of a dedicated high throughput targeted software platform. This native cross-vendor tool for targeted lipidomics will allow to (i) create transitions and assays, (ii) optimize collision energies, (iii) visually review the obtained results, and (iv) quantify the lipids of interest. We envision that the lipid building block based Skyline application breaks ground not only for optimized method design, analysis, and data evaluation in targeted lipidomics, but also provide a gateway to spectral libraries and the sharing of experimental data. This will help

the research community to build up a comprehensive and vendor independent exchange platform to improve reproducibility and validation processes for lipidomics data.

2.2. LipidXplorer: a generic software tool for shotgun lipidomics

By definition, shotgun lipidomics relies on a direct infusion of total lipid extracts into a tandem mass spectrometer, however lipids can be identified in many ways. Historically, shotgun lipidomics was mainly associated with triple quadrupole mass spectrometers (reviewed in (Han et al., 2012)). Lipids were identified by a combination of precursor and neutral loss scans and, because of their low mass resolution, never relied on accurate masses of fragment precursors. In contrast, tandem mass spectrometers of the Orbitrap family and modern q-TOF deliver high mass resolution spectra along with low-ppm accuracy. Therefore masses of detected molecules can be associated with their elemental compositions and lipids consistently identified in the series of compositionally related samples by mapping their intact masses (Schwudke et al., 2007; Schwudke et al., 2011). The prerequisites for assigning correct elemental compositions of lipid precursors and fragments are still under debate (Bielow et al., 2017). Experimentally proven improvements for lipid identification were observed when the mass difference for certain elemental compositions can be resolved. The differentiation of 1-between alkyl-2-acyl glycerophospholipid and 1,2-diacyl glycerophospholipid can be achieved when the mass difference of 36.4 mDa is resolved. Furthermore, it is preferable to resolve the overlap of the second isotopic peak and lipid with exactly one less double bond which results in a mass difference of 9.0 mDa (Herzog et al., 2011; Schwudke et al., 2007). Specifically, for shotgun lipidomics, where the separation power is solely depending on the chosen MS instrument platform false assignments of major abundant glycerophospholipids can be minimized. From these estimates, one can state that the specificity of shotgun lipidomics increases for resolution above 100,000 (FWHM at m/z 750) for MS^1 and 30,000 (FWHM at m/z 200) for MS^2. High acquisition rate and sensitivity of any modern hybrid tandem mass spectrometers enable fragmentation of all candidate peaks independent of their intensities and produce a comprehensive dataset that comprise masses of all detectable precursors and all fragments generated by their collisional fragmentation (Schwudke et al., 2006a; Schwudke et al., 2011). Therefore, shotgun lipidomics software should support any data interpretation routine independent of instrumentation platform, and methods of collisional fragmentation. It should also be able to target any lipid class and species.

To this end, we developed a novel concept for the interpretation of large collections of shotgun spectra that does not rely on fixed