

Computational Lipidomics and Lipid Bioinformatics: Filling In the Blanks

Josch K. Pauling^{1,*}, Edda Klipp¹

¹Theoretical Biophysics, Institute of Biology, Humboldt Universität zu Berlin, Berlin, Germany

Summary

Lipids are highly diverse metabolites of pronounced importance in health and disease. While metabolomics is a broad field under the omics umbrella that may also relate to lipids, lipidomics is an emerging field which specializes in the identification, quantification and functional interpretation of complex lipidomes. Today, it is possible to identify and distinguish lipids in a high-resolution, high-throughput manner and simultaneously with a lot of structural detail. However, doing so may produce thousands of mass spectra in a single experiment which has created a high demand for specialized computational support to analyze these spectral libraries. The computational biology and bioinformatics community has so far established methodology in genomics, transcriptomics and proteomics but there are many (combinatorial) challenges when it comes to structural diversity of lipids and their identification, quantification and interpretation. This review gives an overview and outlook on lipidomics research and illustrates ongoing computational and bioinformatics efforts. These efforts are important and necessary steps to advance the lipidomics field alongside analytic, biochemistry, biomedical and biology communities and to close the gap in available computational methodology between lipidomics and other omics sub-branches.

1 Introduction

Over the last decade lipids have increasingly moved into the center of interest in biomedical research [1] with a European initiative started in 2005/2006 [2]. The term *cellular lipidome* first appeared in 2003 [3] and describes the full lipid complement of a cell. Lipidomics is commonly seen as a sub-branch of metabolomics and it deserves this separate niche underneath the omics umbrella because lipids are a group of highly diverse molecules with distinct chemical properties and important biological functions. Commonly known as fat for highly efficient energy storage, lipids serve many more functional roles. As an example, the plasma membrane is the frontier between a cell and its environment and all functional compartments of eukaryotic cells are separated from the cytosol by membranes [4]. These membranes are usually lipid bilayers built mainly of specific compositions of phospholipids, sphingolipids and sterols. This composition changes between cellular organelles. Furthermore, the inner and outer leaflets of a lipid bi-layer exhibit different lipid class and lipid species compositions providing the membrane with a particular structure, fluidity and function [5, 6]. Eicosanoids are yet another group of lipids that have been discovered as important mediators in inflammation [7], pain [8], and pregnancy [9]. A central precursor to eicosanoid synthesis is arachidonic acid (AA), an ω -6

*To whom correspondence should be addressed. Email: josch.pauling@hu-berlin.de

fatty acid with 20 carbons and four *cis*-double bonds. For example, in non-alcoholic fatty liver disease (NAFLD), which comprises the two different progression states non-alcoholic fatty liver (NAFL) and non-alcoholic steatohepatitis (NASH), specific fatty acids such as eicosapentaenoic acid (20:5 ω -3) and docosahexaenoic acid (22:6 ω -3) as well as eicosanoids such as certain hydroxyeicosatetraenoic acids (HETE) and dihydroxyeicosatrienoic acids (diHETrE) could be identified as potential markers for NAFL and NASH [10, 11, 12, 13] (lipid nomenclature is explained in section 2.2). This highlights the relationship between composition changes in lipid profiles and disease progression.

Single lipids display a high structural diversity since they can contain one or many fatty acyl moieties attached to a lipid class-specific molecular structure [14]. A single fatty acid can have varying numbers of carbon atoms, double bonds, double bond types, double bond positions and many other modifications such as hydroxyl groups attached at various carbon positions. Therefore, we observe a combinatorial explosion in theoretical molecular lipid entities. Cellular lipidomes and the sets of lipids they are composed of may be limited by a cell's metabolic and biosynthetic capabilities. Yeast, an organism that has been widely used for lipid research in eukaryotes due to its many advantages [15], synthesizes only a limited set of fatty acids such as palmitic acid (16:0) and stearic acid (18:0) and is only capable of desaturating in the Δ -9 position resulting in a limited amount of unsaturated fatty acids, e.g. palmitoleic acid (16:1 *cis*-9) and oleic acid (18:1 *cis*-9), [16]. This also greatly reduces the combinatorial space of possible lipid species in yeast which is advantageous for creating computational lipid models. However, this advantage is lost when investigating mammalian cells creating a high demand for computational aid. Hence, computational lipidomics research is an urgently needed field to close the gap that was created by rapid progress in lipid-analytical workflows involving high-resolution mass spectrometers allowing high-throughput identification and quantification of a plethora of molecular lipid entities.

2 Lipidomics: From a Sample to Functional Insights

This section provides a detailed overview of lipidomics from extracting lipids from a sample and measuring the lipidome using mass spectrometry to the analysis of mass spectra, their integration and interpretation.

2.1 The Lipidomics Workflow

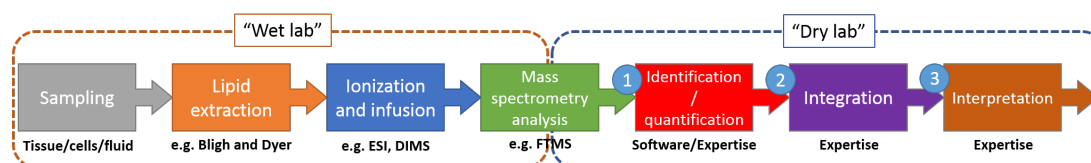


Figure 1: The lipidomics workflow from a sample to an interpretation in the context of integrated omics analyses. The numbers inside the "Dry lab" box indicate the different frontiers in computational lipidomics as presented in this section and throughout this review.

A typical lipidomics workflow comprises several necessary steps as shown in Figure 1. The first step in lipidomics is the extraction of lipids from a sample. Multiple extraction procedures have been published of which the methods proposed by Folch et al. [17] and by Bligh and Dyer [18] are the most commonly applied techniques with over 50000 and over 40000 citations, respectively. However, when probing for an even broader spectrum of lipids from various more known lipid classes, further chemical derivatization such as sulfation [19] or acetylation [20] may have to be applied because of varying ionization efficiency during electrospray ionization (ESI) [21, 22]. After the extraction step, lipids in the sample extract must be identified and quantified which is mostly done via mass spectrometry [23]. The careful addition of a specific set of lipid standards is crucial for accurate quantification [24]. In large-scale lipid experiments including several sample types, time points, scan ranges, ionization modes, etc., it is required to have software support capable of organizing, unifying, reading and annotating spectral libraries as well as saving a comprehensive report about found lipid compositions in a suitable data format. This format should be standardized to allow immediate further processing with third party software for subsequent statistical and functional analysis. With lipidome-scale molecular lipid species analysis moving into the center of attention, derived data-sets have become more and more comprehensive, complex, fragmented, and heterogeneous. Therefore, the **first frontier** (see Figure 1) of current computational challenges in the lipidomics realm is the development of software solutions for the automatic identification and quantification of lipids from large-scale, high-throughput lipidomics experiments. This task ranges from technical details of peak detection heuristics to lipid nomenclature and suitable data formats for saving and storing results as well as the formulation of standards throughout the entire workflow (where applicable). Currently, this is often solved by non-standardized, in-house solutions with very mixed outcomes and hardly comparable results.

After lipids in a data set have been characterized and a lipidome has been assembled, the next step in a lipidomics workflow is a statistical post-analysis for quality assessment and assurance as part of the identification and quantification scheme. Due to a lack of standardization, this step is mostly data set-specific and therefore universal methods would likely counteract its critical nature. However, if standard operating procedures were implemented to formulate a routine workflow beginning at lipid extraction and ending at statistical checks for quality control, software solutions could operate fully automatically to produce high-quality lipid data sets ready for functional analysis. This is **the second frontier** (see Figure 1) in computational lipidomics including not only the development of sophisticated statistical evaluation routines but also the collaboration with lipid analytics and biochemistry communities to formulate standard operating procedures and standardized interfaces facilitating a high-throughput start-to-end lipidomics routine.

The third and last step is functional analysis and the incorporation of lipid-mediated insights into a systems biology context, together with data sets from other omics fields. This includes various challenges for which solutions have already surfaced combining heterogeneous omics data sets, e.g. genomics and proteomics [25, 26]; but lipids often require dedicated lipid expertise, because of their distinct biochemical properties and their often tight relation between molecular structure and function. Proteins display direct physical and enzymatic interactions with lipids so that proteomics appears to be a suitable link. Developing methods dedicated to lipidomics data that integrate such findings with other omics data as well as platforms for

meta-analyses is the **third frontier** (see Figure 1) for computational lipidomics and lipid bioinformatics research.

2.2 Lipid Nomenclature and Ontology

Lipids are generally measured with different levels of structural detail. In a regular MS scan, an intact lipid species is expressed and annotated by its sum composition (e.g. 34:1) which is a combination of the sum of carbon atoms and the sum of double bonds on all fatty acyl moieties. Accordingly, the name "PC 34:1" denotes an intact phosphatidylcholine (PC) lipid molecule with a total sum of 34 carbon atoms and one double bond on its two fatty acyl moieties, while the carbon atoms incorporated in the glycerol and choline are included in the term "PC" and therefore not counted. A possible fatty acyl configuration for a "PC 34:1" would be one fatty acyl moiety consisting of 16 carbon atoms and without any double bonds (saturated) while the other contains 18 carbon atoms and a single double bond (monounsaturated). Once this fatty acid configuration is measured and known, the lipid may be expressed by its molecular composition, e.g. "PC 16:0-18:1". Another possible configuration for a "PC 34:1" is "PC 14:1-20:0" amongst others. Since there are many more possible modifications to a lipid molecule, the International Union of Pure and Applied Chemistry (IUPAC) developed specific nomenclature rules for the exact annotation of lipid molecules [27, 28, 29, 30, 31].

The invariant part of any phospholipid is its glycerophosphate. Often, two fatty acyl moieties are attached to the sn-1 and sn-2 positions of the glycerol backbone while a specific headgroup including the phosphate is attached at the sn-3 position. This determines the type of phospholipid. A PC consists of a choline headgroup, analogously a PE consists of an ethanolamine headgroup. Fatty acids can have many different molecular details. They differ in number of carbon atoms in the hydrocarbon chain, number and type (cis/trans [32]) of double bonds as well as type and position of other modifications such as hydroxylations.

Figure 2 shows an exemplary PC molecule. Following IUPAC rules its name is 1-Hexadecanoyl-

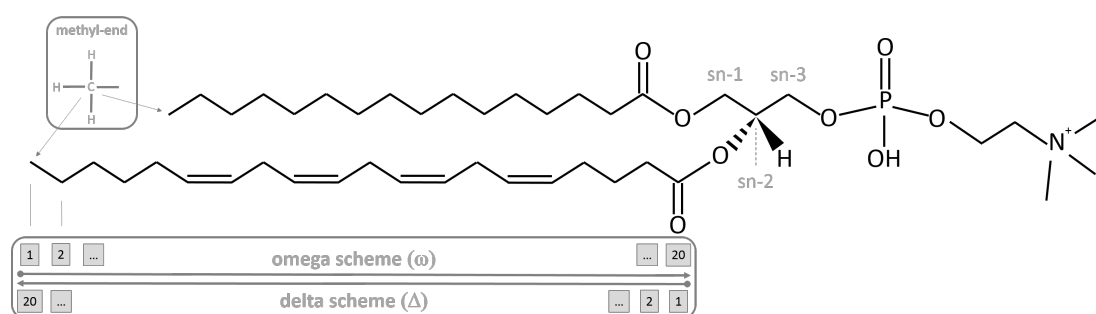


Figure 2: A phosphatidylcholine species which incorporates a fatty acyl 16:0 and a fatty acyl 20:4 (sum composition: PC 36:4; molecular composition: PC 16:0/20:4). The zig-zack structure resembles a hydrocarbon chain. Each elbow denotes a CH₂ hydrocarbon since carbon is tetravalent. Thus, the tip of a terminal edge denotes a CH₃ hydrocarbon (methyl end). Doubled lines indicate a double bond between two carbon atoms. The glycerol backbone consists of three carbon atoms each labeled by its stereospecific numbering (sn). Two schemes (ω and Δ) for counting carbon atoms of the hydrocarbon chain. They are used to denote double bond positioning.

2-([cis,cis,cis,cis]-5,8,11,14-eicosatetraenoyl)-sn-glycero-3-phosphocholine. "1-Hexadecanoyl" describes that a fatty acyl moiety with 16 carbon atoms and no double bonds deriving from the esterification of a hexadecaenoic acid (trivial name: palmitic acid) is attached at the sn-1 position of the glycerol. "2-([cis,cis,cis,cis]-5,8,11,14-eicosatetraenoyl" represents the fatty acyl moiety attached to the sn-2 position, in this case it consists of 20 carbon atoms and a total of four double bonds of the "cis" type at carbons 5, 8, 11, 14 when counting towards the terminal methyl end (CH₃) starting from the carboxyl end. More publicly known from food products is the omega (ω) counting scheme which starts from the opposite methyl end. Thus, the eicosatetraenoyl derives from an ω -6 eicosatetraenoic acid (trivial name: arachidonic acid). Finally, "sn-glycero" reflects the glycerol backbone and "3-phosphocholine" indicates the phosphocholine headgroup that is attached to the sn-3 position of the glycerol. Since the first version of lipid nomenclature rules, more lipids were identified and these nomenclature guidelines have been updated accordingly, e.g. for eicosanoids [31].

As the example illustrates, IUPAC nomenclature rules result in exact but complicated and lengthy names. Since lipidomics studies typically focus on full lipidome scans, many structural details are not contained in a resulting data set, such as double bond and fatty acyl stereochemistry. Hence, this information is often disregarded when lipidome components are annotated. Due to this lack of structural information, IUPAC rules, which provide unique names and full structural annotation by design, become unfeasible. Therefore, other alternatives have been proposed to annotate lipid molecules [33], most of which try to filter and abbreviate IUPAC conventional rules and to match them to a reasonable and simplified display of structural detail. As briefly stated above, full scans of intact lipid molecules can be annotated by their sum compositions, e.g. PC 34:1, reflecting the sum of carbon atoms and double bonds over both fatty acyl moieties. MS₂ analysis allows to resolve the fatty acyl moieties, so that lipids can be annotated by their molecular compositions, e.g. PC 16:0-18:1. In both cases spatial/positional information is missing and cannot be inferred from a molecule's m/z ratio (otherwise the shorthand notation was PC 16:0/18:1 indicating the exact positioning of the two fatty acyls). In addition, double bond positioning on the hydrocarbon chains and double bond type are also missing. Higher MS dimensions (MS_n, $n > 2$) may elucidate further structural details. Hence, the selection of annotation nomenclature is dictated by the type of MS analysis and the structural detail it can resolve.

In 2005 Fahy et al. [34] developed a classification system for lipids based on structural and biochemical considerations. An update was published in 2009 [35]. According to their proposal, lipids were separated into eight different lipid categories, namely fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, sterol lipids, prenol lipids, saccharolipids, and polyketides, each of which contain multiple lipid classes and possibly sub-classes. The above-mentioned PC molecule is classified as a glycerophospholipid because it is a lipid containing a glycerol and a phosphate.

2.3 Untargeted lipidomics

In untargeted lipidomics lipids are first extracted by a broad spectrum extraction method such as Bligh and Dyer [18]. The extract is then injected into a mass spectrometer, e.g. as in shotgun

lipidomics [36, 37, 38, 39, 40]. In shotgun lipidomics lipid molecules are ionized by an ESI source and directly injected into a high-resolution ion trap (ITMS) or Fourier transform mass spectrometer (FTMS). Finally, ions are detected and reported with their m/z (mass-over-charge) ratio (abscissa) and intensity (ordinate) creating the common mass spectrum. Detecting intact lipid molecules allows to map their m/z values to matching chemical compounds. Isomeric (e.g. positional isomers) as well as isobaric molecules cannot be resolved which may lead to erroneous identification and quantification and eventually to distorted lipidome compositions. In the case of lipid analysis this imposes a major limitation since the incorporation of particular fatty acids into lipid molecules is often a critical piece of information for functional analyses since many fatty acid compositions result in isomers, e.g. PC 16:0/18:1 and PC 16:1/18:0. In addition, lipid molecules across lipid classes may be isomeric as well (PC 34:1, PE 37:1 or isotopic isomers). In essence, untargeted lipidomics produces sufficient spectral lipidome profiles in a high-throughput manner but at the cost of sensitivity due to the lack of chromatographic separation and optimized extraction protocols.

2.4 Targeted lipidomics

Targeted lipidomics employs lipid category- and lipid class-optimized extraction protocols. Afterwards, the extracts are introduced to a liquid chromatography tandem mass spectrometry (LC-MSMS) system to achieve optimal separation of lipid species present in the extract prior to MSMS analysis [41, 23, 42, 43, 44, 45]. This strategy ensures best possible sensitivity by utilizing chromatographic separation and fragmentation producing complete and accurate lipidome profiles. However, this is achieved at the cost of simplicity and spectrum acquisition speed. Additionally, it imposes a bigger challenge for subsequent computational analysis and quality control in reassembling a complete lipidome from several independent, heterogeneous mass spectra and an additional time dimension.

2.5 Fragment dissociation and MS n analysis

To gain resolving power and to acquire more structural details, ions can be fragmented inside the collision cell of a mass spectrometer [46, 47, 48]. First, ions are isolated, e.g. by a multi-pole mass filter, and in the case of MS n ($n > 1$) subsequently fragmented $n - 1$ times. A fragment's m/z ratio may then reveal unique molecular details [49]. A typical fragment that derives from a PC molecule in positive ion mode has an m/z of 184.073321 which represents the phosphocholine that dissociated from the sn-3 position of the glycerol. However, SM (Sphingomyelin) precursors may show the same product ion since they also can contain a phosphocholine. This highlights that specific inference of molecular lipid species should be based on multiple fragments that serve as molecular evidence for the existence of a particular lipid. Fragmentation pathways have been well explored for many lipid classes [50] such as ceramides [51], phosphatidic acid [52], phosphatidylcholine [53], phosphatidylethanolamine [54], phosphatidylinositol [55, 56], and yeast sphingolipids [57]. The inference process can be performed by manual inspection of particular fragment-spectra when the data set of interest comprises only a few fragment-spectra. However, lipidomics technology focuses on

high-throughput, large-scale analyses of complete lipidomes and thus data sets deriving from complex lipid experiments can comprise thousands of fragment-spectra. In this case manual inspection is infeasible and computational inference is required.

3 Computational Lipidomics and Lipid Bioinformatics

The following section provides a summary of previous computational and bioinformatics methods in the field of lipidomics that have taken part in advancing the field and to connect it with other systems biology and medicine research areas. It should be mentioned that this is a selection of studies and therefore does not aim to be complete.

3.1 Lipid Databases and bioinformatics

There have already been several informatic efforts that have advanced the lipidomics field (third frontier). The first efforts were made to establish lipid databases that contain publicly available data on all currently known lipids. The LIPID MAPS Consortium created the lipid database LIPID MAPS (www.lipidmaps.org) [58], [59] as well as a collection of online tools [60]. LIPID MAPS not only provides information such as names, synonyms, exact mass, chemical formula, ontology and links to other OMICs databases such as KEGG (Kyoto Encyclopedia of Genes and Genomes) [61] and HMDB (Human Metabolome Database) [62] but it also provides schematics of molecular structures. Other databases that roughly cover a similar features list are LipidHome [63] and SwissLipids [64]. These are very useful databases for manual inspection and annotation of lipid mass spectra from regular MS analysis. The ALEX123 [65] database stands out since it contains a library of fragmentation patterns of lipid fragments released via CID/HCD (Collision-Induced dissociation; Higher energy collisional dissociation) which can be publicly queried on its web interface. In addition, ALEX123 [66] provides tools that utilize the database for automatic analysis of MSⁿ spectra ($n \leq 3$) which are currently not freely available. All these databases provide a substantial contribution to move lipid bioinformatics infrastructure closer to those concepts that are well established in the genomics and proteomics realms. Even though not a database, Skyline [67] is a software designed for proteomics applications which has recently been extended to offer the capability to assemble targeted mass spectrometry methods for the analysis of complex lipids such as glycerophospholipids, sphingolipids, glycerolipids, cholesteryl-esters, and cholesterol. Other bioinformatics contributions should be mentioned but are not further described in this review (in chronological order):

- A consensus yeast metabolic network reconstruction [68]. In this study the authors present not only the metabolic network model itself in a standardized SBML (Systems Biology Markup Language) format but they also emphasize on linking compounds to various public databases. An update was published in 2013 by Aung et al. with specific focus on essential lipid categories [69].
- Bioinformatics and computational methods for lipidomics [70]. This study dates back to 2009 and was one of the earlier studies highlighting the importance of computational support in areas for which over seven years later no general solutions have been provided.

- A computational framework for integrating lipidomics data into metabolic pathways [71]. This framework was named NICELips and can be used to investigate lipid metabolism. It allows reconstruction and discovery of biosynthetic and catabolic reactions based on thermodynamic considerations.
- A metric for lipidome homology [72]. This homology metric can quantify systematic differences in lipidome compositions. Hence, it is an early approach to lipidome meta-analyses.

3.2 Computer-assisted Lipid Identification and Quantification

A number of software solutions have been developed over the last decade that are capable of quickly annotating full MS spectra (first frontier). Among those are ALEX [73], ALEX123 [66], LipidSearch (Thermo Fisher Scientific), LipidView (AB SCIEX, Concord, ON), and LipidXplorer [74]. Usually, a peak detection is followed by mapping peak m/z values to a database entry to identify the underlying compound. This commonly produces multiple matches with isomeric and isobaric analytes which may be resolved by using isotope patterns. Another way is to scan the acquired spectral library for a specific set of theoretical lipids. In both environments spectral data allows the identification and quantification of intact lipid molecules which are accordingly annotated by their sum compositions.

Identification and possibly quantification of lipidomes resolved by their molecular lipid species composition (this provides the fatty acyl configuration on each lipid, e.g. PC 16:0/18:1) requires a lipid fragment database containing fragmentation patterns for each lipid species which differ between lipid classes and sub-classes. While identification based on a collection of measured fragment spectra has received some attention, accurate quantification of fragments-inferred molecular species proposes a challenge (see all kinds of isomers, subsection "Untargeted lipidomics") and sophisticated correction methods are required. The set of detected lipid ions as well as their corresponding fragment ions is finally reported. From this data (molecular) lipid species are first inferred before lipidome compositions are reassembled. Hereafter, statistical and functional analysis of the entire data set of the experiment can be conducted.

There is a necessity to advance towards automatic high-throughput, lipidome-scale solutions that identify lipids from fragmentation spectra on molecular species level with unique fragments as structural evidence that allow inference of a particular fatty acyl configuration for each lipid molecule. Since this is quite a revolutionary advancement in the field, so far only commercially available products such as ALEX123 [66], LipidSearch [75] and Lipotype [76] possess this ability. They are capable of identifying and annotating molecular lipid species from a library of fragmentation spectra through computational inference based on observable fragments. As briefly mentioned, this is a combinatorial challenge in the inference process since many observed fragments are not unique or specific for a single lipid molecule. All of these tools operate within the untargeted (shotgun) lipidomics framework where lipidomes are measured directly from biological lipid extracts. However, LipidSearch and Lipotype also support chromatographic separation. Greazy [77] is a freely available solution for MS/MS analysis but it is limited to phospholipid identification only. Phospholipids are generally simple in structure

containing one (lysophospholipids) or two fatty acids. Triacylglycerols (TAGs) consist of three and cardiolipins of four fatty acids greatly expanding the combinatorial space causing inference of molecular lipid species based on observed fragments to be vastly more complex. Overall, exact quantification of molecular lipidomes based on fragment spectra currently remains an unresolved problem.

3.3 Lipid Bioinformatics: Linking the Omics Landscape

With (semi-)automatic lipid identification and quantification software being applied more often, data of complex lipidomes from various biological samples will be more rapidly created. This reveals the next gap in computational support for lipid analysis: Algorithms and software that are capable of mining vast amounts of data that includes format and unit conversions, statistical processing and analysis, modeling, prediction and interpretation to form a systemic understanding of the underlying biological system (second and third frontiers). Partially, the aforementioned software tools provide pre-processed data sets with various levels of information and complexity but none of these can (semi-)automatically conduct data mining, statistical analysis, or data integration. While regular statistical software such as R is sufficient for conducting standard statistical assessments, comprehensive lipid-specific packages providing standardized routines have not yet been developed. Very recently Collins et al. published the R package LOBSTAHS for high-throughput annotation and putative identification of lipid, oxidized lipid, and oxylipin biomarkers in high-mass-accuracy HPLC-MS data [78]. Here, adduct ion formation patterns are exploited, a characteristic of mass spectrometry analyses using direct chemical ionization [79]. Similarly, mass spectrometry and lipidomics-specific analysis packages are needed, e.g. for quantitative peak deconvolution by isotopomers as well as by fragmentation patterns allowing a single peak to be mapped to several unresolved lipid molecules while maintaining accurate quantification. Data mining is usually done only for the particular investigated lipid data set. Online platforms that perform meta-analyses and data mining over multiple published data sets are missing entirely. Data integration is so far difficult and connecting lipid data with genomics, transcriptomics and proteomics is often based on individual solutions.

It is therefore a valuable advancement that computational modeling has already provided fruitful contributions that not only provide a way of integrating metabolic systems including various omics but also extend the analysis paradigm by temporal dynamics. A rather general method for kinetic modeling of biochemical networks has been proposed by Rao et al. [80]. A model for arachidonic acid metabolism in human polymorphic leukocytes based on ordinary differential equations (ODEs) has been developed by Yang et al. [81]. This model elucidated flux changes upon drug treatment and provided information for drug discovery. Kihara et al. [82] analyzed temporal and dynamic changes of the eicosanoid metabolic network in mouse bone marrow-derived macrophages (BMDM) upon stimulation of the Toll-like receptor 4 with Kdo2-Lipid A (KLA) and stimulation of the P2X7 purinergic receptor with adenosine 5-triphosphate. Subsequently, they developed a comprehensive kinetic model based on ODEs for the cyclooxygenase (COX) and lipoxygenase (LOX) mediated pathways involved in arachidonic acid metabolism. This approach provided evidence for functional couplings between involved enzymes which were experimentally validated. A third study by Schützhold et al. [83] models a compre-

hensive part of yeast lipid metabolism using an object-oriented stochastic approach instead of ODEs. This approach allowed to follow the dynamics of all lipid species with various fatty acids, different degrees of desaturation and multiple headgroups over time. Additionally, the model can be used to analyze the effects of parameter changes, potential mutations in the catalyzing enzymes or provision of different precursors and it allows to derive conclusions on the time- and location-dependent distributions of lipid species and their properties such as desaturation. In a fourth study by Gupta et al. [84] a kinetic model for eicosanoid metabolism in macrophage cells was developed based on an integrated network of eicosanoid metabolism and signaling. All of these studies demonstrate the efficacy of computational modeling approaches to create comprehensive, dynamic, and continuous models from discrete multi-omics data.

4 Summary and Conclusion

This review describes three major frontiers in lipidomics in which computational strategies are urgently needed not only to advance the field itself but also omics-driven research in general. These three frontiers are:

1. First frontier: analysis of large, heterogeneous mass spectral libraries including fragment spectra for identification and quantification
2. Second frontier: statistical routines for quality control and maintenance of comprehensive lipidomics datasets and start-to-end standardization
3. Third frontier: functional analysis, data-mining and integration into multi-omics settings

One of the major problems at the core of lipidomics is the exact quantification of molecular lipid species compositions based on fragment data for complete lipidomes. So far high-throughput solutions are not capable of exact quantification but approximations may be sufficiently accurate. Knowing the fatty acyl moieties is an important aspect in biomedical research determining metabolic fates and enabling quantitative modeling. Then, the development of workflows including computational as well as biochemical and analytical protocols from the sample to a readily processed data set is needed. This requires close interdisciplinary collaborations and common agreement to standard operating procedures to ensure high data set quality and comparability. In parallel, critical bioinformatics infrastructure must be gradually built up to support and enable large-scale and meta analyses on publicly available datasets. This infrastructure has been implemented for other omics fields which greatly fuels these branches. However, with the exception of a few dedicated lipid databases this publicly and freely accessible infrastructure is so far largely missing for lipidomics and has just recently gained attention of the bioinformatics community. With an increase in lipid-mediated investigations in health and disease and with the provision of computational solutions, standardized workflows and bioinformatics infrastructure, it is expected that lipid analysis will become a regular aspect in omics-driven studies over the next five years. This will contribute important insights into the various roles of lipids in biological networks and metabolic pathways.

5 Acknowledgements

This work was supported by the BMBF-funded program LiSyM (FKZ 031L0034).

References

- [1] M. R. Wenk. The emerging field of lipidomics. *Nature reviews. Drug discovery*, 4:594–610, 2005.
- [2] G. van Meer, B. R. Leeﬂang, G. Liebisch, G. Schmitz and F. M. Goi. The european lipidomics initiative: enabling technologies. *Methods in enzymology*, 432:213–232, 2007.
- [3] X. Han. Global analyses of cellular lipidomes directly from crude extracts of biological samples by ESI mass spectrometry: a bridge to lipidomics. *The Journal of Lipid Research*, 44(6):1071–1079, 2003.
- [4] M. Edidin. Lipids on the frontier: a century of cell-membrane bilayers. *Nature reviews. Molecular cell biology*, 4:414–418, 2003.
- [5] G. van Meer. Cellular lipidomics. *The EMBO journal*, 24:3159–3165, 2005.
- [6] G. van Meer, D. R. Voelker and G. W. Feigenson. Membrane lipids: where they are and how they behave. *Nature reviews. Molecular cell biology*, 9:112–124, 2008.
- [7] B. Samuelsson, S. Dahlen, J. Lindgren, C. Rouzer and C. Serhan. Leukotrienes and lipoxins: structures, biosynthesis, and biological effects. *Science*, 237(4819):1171–1176, 1987.
- [8] C. D. Funk. Prostaglandins and leukotrienes: Advances in eicosanoid biology. *Science*, 294(5548):1871–1875, 2001.
- [9] M. C. Wiltbank and J. S. Ottobre. Regulation of intraluteal production of prostaglandins. *Reprod Biol Endocrinol*, 1(1):91, 2003.
- [10] P. Puri, R. A. Baillie, M. M. Wiest, F. Mirshahi, J. Choudhury, O. Cheung, C. Sargeant, M. J. Contos and A. J. Sanyal. A lipidomic analysis of nonalcoholic fatty liver disease. *Hepatology*, 46(4):1081–1090, 2007.
- [11] P. Puri, M. M. Wiest, O. Cheung et al. The plasma lipidomic signature of nonalcoholic steatohepatitis. *Hepatology*, 50(6):1827–1838, 2009.
- [12] R. Loomba, O. Quehenberger, A. Armando and E. A. Dennis. Polyunsaturated fatty acid metabolites as novel lipidomic biomarkers for noninvasive diagnosis of nonalcoholic steatohepatitis. *Journal of Lipid Research*, 56(1):185–192, 2014.
- [13] D. W. L. Ma, B. M. Arendt, L. M. Hillyer, S. K. Fung, I. McGilvray, M. Guindi and J. P. Allard. Plasma phospholipids and fatty acid composition differ between liver biopsy-proven nonalcoholic fatty liver disease and healthy subjects. *Nutrition & Diabetes*, 6(7):e220, 2016.
- [14] R. W. Gross and X. Han. Lipidomics at the interface of structure and function in systems biology. *Chemistry & biology*, 18:284–291, 2011.

- [15] J. Nielsen. Systems biology of lipid metabolism: from yeast to human. *FEBS letters*, 583:3905–3913, 2009.
- [16] E. A. Dennis. Lipidomics joins the omics evolution. *Proceedings of the National Academy of Sciences*, 106(7):2089–2090, 2009.
- [17] J. Folch, M. Lees and G. H. Sloane Stanley. A simple method for the isolation and purification of total lipides from animal tissues. *The Journal of biological chemistry*, 226:497–509, 1957.
- [18] E. G. Bligh and W. J. Dyer. A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology*, 37(8):911–917, 1959.
- [19] R. Sandhoff, B. Brgger, D. Jeckel, W. D. Lehmann and F. T. Wieland. Determination of cholesterol at the low picomole level by nano-electrospray ionization tandem mass spectrometry. *Journal of lipid research*, 40:126–132, 1999.
- [20] G. Liebisch, M. Binder, R. Schifferer, T. Langmann, B. Schulz and G. Schmitz. High throughput quantification of cholesterol and cholesteryl ester by electrospray ionization tandem mass spectrometry (esi-ms/ms). *Biochimica et biophysica acta*, 1761:121–128, 2006.
- [21] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong and C. M. Whitehouse. Electrospray ionization for mass spectrometry of large biomolecules. *Science (New York, N.Y.)*, 246:64–71, 1989.
- [22] C. S. Ho, C. W. K. Lam, M. H. M. Chan, R. C. K. Cheung, L. K. Law, L. C. W. Lit, K. F. Ng, M. W. M. Suen and H. L. Tai. Electrospray ionisation mass spectrometry: principles and clinical applications. *The Clinical biochemist. Reviews*, 24:3–12, 2003.
- [23] P. T. Ivanova, S. B. Milne, M. O. Byrne, Y. Xiang and H. A. Brown. Glycerophospholipid identification and quantitation by electrospray ionization mass spectrometry. *Methods in enzymology*, 432:21–57, 2007.
- [24] J. D. Moore, W. V. Caufield and W. A. Shaw. Quantitation and standardization of lipid internal standards for mass spectroscopy. *Methods in enzymology*, 432:351–367, 2007.
- [25] N. Alcaraz, J. Pauling, R. Batra, E. Barbosa, A. Junge, A. G. Christensen, V. Azevedo, H. J. Ditzel and J. Baumbach. KeyPathwayMiner 4.0: condition-specific pathway analysis by combining multiple omics studies and networks with cytoscape. *BMC Systems Biology*, 8(1), 2014.
- [26] J. K. Pauling, A. G. Christensen, R. Batra et al. Elucidation of epithelial–mesenchymal transition-related pathways in a triple-negative breast cancer cell line model by multi-omics interactome analysis. *Integr. Biol.*, 6(11):1058–1068, 2014.
- [27] Iupac-iub commission on biochemical nomenclature (cbn). the nomenclature of lipids. *European journal of biochemistry*, 2:127–131, 1967.

- [28] The nomenclature of lipids. recommendations (1976) iupac-iub commission on biochemical nomenclature. *Lipids*, 12:455–468, 1977.
- [29] Prenol nomenclature. recommendations 1986. iupac-iub joint commission on biochemical nomenclature (jcbn). *European journal of biochemistry*, 167:181–184, 1987.
- [30] Iupac-iub joint commission on biochemical nomenclature (jcbn). the nomenclature of steroids. recommendations 1989. *European journal of biochemistry*, 186:429–458, 1989.
- [31] Guidelines on eicosanoid nomenclature. *Eicosanoids*, 2:65–68, 1989.
- [32] G. P. Moss. Basic terminology of stereochemistry (IUPAC recommendations 1996). *Pure and Applied Chemistry*, 68(12), 1996.
- [33] G. Liebisch, J. A. Vizcano, H. Kfeler, M. Trtzmüller, W. J. Griffiths, G. Schmitz, F. Spener and M. J. O. Wakelam. Shorthand notation for lipid structures derived from mass spectrometry. *Journal of lipid research*, 54:1523–1530, 2013.
- [34] E. Fahy, S. Subramaniam, H. A. Brown et al. A comprehensive classification system for lipids. *Journal of lipid research*, 46:839–861, 2005.
- [35] E. Fahy, S. Subramaniam, R. C. Murphy, M. Nishijima, C. R. H. Raetz, T. Shimizu, F. Spener, G. van Meer, M. J. O. Wakelam and E. A. Dennis. Update of the lipid maps comprehensive classification system for lipids. *Journal of lipid research*, 50 Suppl:S9–14, 2009.
- [36] X. Han and R. W. Gross. Shotgun lipidomics: multidimensional ms analysis of cellular lipidomes. *Expert review of proteomics*, 2:253–264, 2005.
- [37] D. Schwudke, G. Liebisch, R. Herzog, G. Schmitz and A. Shevchenko. Shotgun lipidomics by tandem mass spectrometry under data-dependent acquisition control. *Methods in enzymology*, 433:175–191, 2007.
- [38] D. Schwudke, K. Schuhmann, R. Herzog, S. R. Bornstein and A. Shevchenko. Shotgun lipidomics on high resolution mass spectrometers. *Cold Spring Harbor perspectives in biology*, 3:a004614, 2011.
- [39] K. Schuhmann, R. Herzog, D. Schwudke, W. Metelmann-Strupat, S. R. Bornstein and A. Shevchenko. Bottom-up shotgun lipidomics by higher energy collisional dissociation on ltq orbitrap mass spectrometers. *Analytical chemistry*, 83:5480–5487, 2011.
- [40] K. Schuhmann, R. Almeida, M. Baumert, R. Herzog, S. R. Bornstein and A. Shevchenko. Shotgun lipidomics on a ltq orbitrap mass spectrometer by successive switching between acquisition polarity modes. *Journal of mass spectrometry : JMS*, 47:96–104, 2012.
- [41] J. Krank, R. C. Murphy, R. M. Barkley, E. Duchoslav and A. McAnoy. Qualitative analysis and quantitative assessment of changes in neutral glycerol lipid molecular species within cells. *Methods in enzymology*, 432:1–20, 2007.

- [42] R. Deems, M. W. Buczynski, R. Bowers-Gentry, R. Harkewicz and E. A. Dennis. Detection and quantitation of eicosanoids via high performance liquid chromatography-electrospray ionization-mass spectrometry. *Methods in enzymology*, 432:59–82, 2007.
- [43] M. C. Sullards, J. C. Allegood, S. Kelly, E. Wang, C. A. Haynes, H. Park, Y. Chen and A. H. Merrill. Structure-specific, quantitative methods for analysis of sphingolipids by liquid chromatography-tandem mass spectrometry: "inside-out" sphingolipidomics. *Methods in enzymology*, 432:83–115, 2007.
- [44] T. A. Garrett, Z. Guan and C. R. H. Raetz. Analysis of ubiquinones, dolichols, and dolichol diphosphate-oligosaccharides by liquid chromatography-electrospray ionization-mass spectrometry. *Methods in enzymology*, 432:117–143, 2007.
- [45] J. G. McDonald, B. M. Thompson, E. C. McCrum and D. W. Russell. Extraction and analysis of sterols in biological matrices by high performance liquid chromatography electrospray ionization mass spectrometry. *Methods in enzymology*, 432:145–170, 2007.
- [46] J. M. Wells and S. A. McLuckey. Collision-induced dissociation (cid) of peptides and proteins. *Methods in enzymology*, 402:148–185, 2005.
- [47] L. Sleno and D. A. Volmer. Ion activation methods for tandem mass spectrometry. *Journal of mass spectrometry : JMS*, 39:1091–1112, 2004.
- [48] J. V. Olsen, B. Macek, O. Lange, A. Makarov, S. Horning and M. Mann. Higher-energy c-trap dissociation for peptide modification analysis. *Nature methods*, 4:709–712, 2007.
- [49] R. Almeida, J. K. Pauling, E. Sokol, H. K. Hannibal-Bach and C. S. Ejsing. Comprehensive lipidome analysis by shotgun lipidomics on a hybrid quadrupole-orbitrap-linear ion trap mass spectrometer. *Journal of the American Society for Mass Spectrometry*, 26:133–148, 2015.
- [50] F.-F. Hsu and J. Turk. Structural characterization of unsaturated glycerophospholipids by multiple-stage linear ion-trap mass spectrometry with electrospray ionization. *Journal of the American Society for Mass Spectrometry*, 19:1681–1691, 2008.
- [51] F.-F. Hsu and J. Turk. Characterization of ceramides by low energy collisional-activated dissociation tandem mass spectrometry with negative-ion electrospray ionization. *Journal of the American Society for Mass Spectrometry*, 13:558–570, 2002.
- [52] F. F. Hsu and J. Turk. Charge-driven fragmentation processes in diacyl glycerophosphatidic acids upon low-energy collisional activation. a mechanistic proposal. *Journal of the American Society for Mass Spectrometry*, 11:797–803, 2000.
- [53] F.-F. Hsu and J. Turk. Electrospray ionization/tandem quadrupole mass spectrometric studies on phosphatidylcholines: the fragmentation processes. *Journal of the American Society for Mass Spectrometry*, 14:352–363, 2003.
- [54] F. F. Hsu and J. Turk. Charge-remote and charge-driven fragmentation processes in diacyl glycerophosphoethanolamine upon low-energy collisional activation: a mechanistic proposal. *Journal of the American Society for Mass Spectrometry*, 11:892–899, 2000.

- [55] N. Zehethofer, T. Scior and B. Lindner. Elucidation of the fragmentation pathways of different phosphatidylinositol phosphate species (pipx) using irmpd implemented on a ft-icr ms. *Analytical and bioanalytical chemistry*, 398:2843–2851, 2010.
- [56] M. J. Rovillos, J. K. Pauling, H. K. Hannibal-Bach, C. Vionnet, A. Conzelmann and C. S. Ejsing. Structural characterization of suppressor lipids by high-resolution mass spectrometry. *Rapid communications in mass spectrometry : RCM*, 30:2215–2227, 2016.
- [57] C. S. Ejsing, T. Moehring, U. Bahr, E. Duchoslav, M. Karas, K. Simons and A. Shevchenko. Collision-induced dissociation pathways of yeast sphingolipids and their molecular profiling in total lipid extracts: a study by quadrupole tof and linear ion trap-orbitrap mass spectrometry. *Journal of mass spectrometry : JMS*, 41:372–389, 2006.
- [58] K. Schmelzer, E. Fahy, S. Subramaniam and E. A. Dennis. The lipid maps initiative in lipidomics. *Methods in enzymology*, 432:171–183, 2007.
- [59] M. Sud, E. Fahy, D. Cotter et al. Lmsd: Lipid maps structure database. *Nucleic acids research*, 35:D527–D532, 2007.
- [60] E. Fahy, M. Sud, D. Cotter and S. Subramaniam. Lipid maps online tools for lipid research. *Nucleic acids research*, 35:W606–W612, 2007.
- [61] M. Kanehisa and S. Goto. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28:27–30, 2000.
- [62] D. S. Wishart, D. Tzur, C. Knox et al. Hmdb: the human metabolome database. *Nucleic acids research*, 35:D521–D526, 2007.
- [63] J. M. Foster, P. Moreno, A. Fabregat, H. Hermjakob, C. Steinbeck, R. Apweiler, M. J. O. Wakelam and J. A. Vizcano. Lipidhome: a database of theoretical lipids optimized for high throughput mass spectrometry lipidomics. *PloS one*, 8:e61951, 2013.
- [64] L. Aimo, R. Liechti, N. Hyka-Nouspikel et al. The swisslipids knowledgebase for lipid biology. *Bioinformatics (Oxford, England)*, 31:2860–2866, 2015.
- [65] URL <http://mslipidomics.info/ALEX123/MS.php>.
- [66] URL <http://mslipidomics.info/ALEX123.html>.
- [67] B. Peng and R. Ahrends. Adaptation of skyline for targeted lipidomics. *Journal of proteome research*, 15:291–301, 2016.
- [68] M. J. Herrgrd, N. Swainston, P. Dobson et al. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nature biotechnology*, 26:1155–1160, 2008.
- [69] H. W. Aung, S. A. Henry and L. P. Walker. Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism. *Industrial Biotechnology*, 9(4):215–228, 2013.

- [70] P. S. Niemel, S. Castillo, M. Sysi-Aho and M. Oresic. Bioinformatics and computational methods for lipidomics. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences*, 877:2855–2862, 2009.
- [71] N. Hadadi, K. Cher Soh, M. Seijo, A. Zisaki, X. Guan, M. R. Wenk and V. Hatzimanikatis. A computational framework for integration of lipidomics data into metabolic pathways. *Metabolic engineering*, 23:1–8, 2014.
- [72] C. Marella, A. E. Torda and D. Schwudke. The lux score: A metric for lipidome homology. *PLoS computational biology*, 11:e1004511, 2015.
- [73] P. Husen, K. Tarasov, M. Katafiasz, E. Sokol, J. Vogt, J. Baumgart, R. Nitsch, K. Ekroos and C. S. Ejlsing. Analysis of lipid experiments (alex): a software framework for analysis of high-resolution shotgun lipidomics data. *PloS one*, 8:e79736, 2013.
- [74] R. Herzog, D. Schwudke and A. Shevchenko. Lipidexplorer: Software for quantitative shotgun lipidomics compatible with multiple mass spectrometry platforms. *Current protocols in bioinformatics*, 43:14.12.1–14.1230, 2013.
- [75] URL <https://www.thermofisher.com/order/catalog/product/IQLAAEGABSFAPCMBFK>.
- [76] URL www.lipotype.de.
- [77] M. A. Kochen, M. C. Chambers, J. D. Holman, A. I. Nesvizhskii, S. T. Weintraub, J. T. Belisle, M. N. Islam, J. Griss and D. L. Tabb. Greazy: Open-source software for automated phospholipid tandem mass spectrometry identification. *Analytical chemistry*, 88:5733–5741, 2016.
- [78] J. R. Collins, B. R. Edwards, H. F. Fredricks and B. A. S. V. Mooy. LOBSTAHS: An adduct-based lipidomics strategy for discovery and identification of oxidative stress biomarkers. *Analytical Chemistry*, 88(14):7154–7162, 2016.
- [79] D. I. Carroll, J. G. Nowlin, R. N. Stillwell and E. C. Horning. Adduct ion formation in chemical ionization mass spectrometry of nonvolatile organic compounds. *Analytical Chemistry*, 53(13):2007–2013, 1981. URL <http://dx.doi.org/10.1021/ac00236a015>.
- [80] S. Rao, A. van der Schaft, K. van Eunen, B. M. Bakker and B. Jayawardhana. A model reduction method for biochemical reaction networks. *BMC Systems Biology*, 8(1):52, 2014.
- [81] K. Yang, W. Ma, H. Liang, Q. Ouyang, C. Tang and L. Lai. Dynamic simulations on the arachidonic acid metabolic network. *PLoS computational biology*, 3:e55, 2007.
- [82] Y. Kihara, S. Gupta, M. R. Maurya, A. Armando, I. Shah, O. Quehenberger, C. K. Glass, E. A. Dennis and S. Subramaniam. Modeling of eicosanoid fluxes reveals functional coupling between cyclooxygenases and terminal synthases. *Biophysical journal*, 106:966–975, 2014.

- [83] V. Schützhöf, J. Hahn, K. Tummler and E. Klipp. Computational modeling of lipid metabolism in yeast. *Frontiers in Molecular Biosciences*, 3, 2016.
- [84] S. Gupta, M. R. Maurya, D. L. Stephens, E. A. Dennis and S. Subramaniam. An integrated model of eicosanoid metabolism and signaling based on lipidomics flux analysis. *Biophysical Journal*, 96(11):4542–4551, 2009.