Lipid Spectra Generator: *a simple script for the generation of accurate in-silico lipid fragmentation spectra*

David S Gertner [a], Jake P Violi [a], David P Bishop [b], Matthew P Padula [a*]

[a] School of Life Sciences and Proteomics Core Facility, Faculty of Science, University of Technology Sydney, Ultimo 2007, Australia

[b] School of Mathematical and Physical Sciences, Faculty of Science, University of Technology Sydney, Ultimo 2007, Australia

ABSTRACT: Due to the complexity of lipids in nature, the use of *in-silico* generated spectral libraries to identify lipid species from mass spectral data has become an integral part of many lipidomic workflows. However, many *in-silico* libraries are either limited in usability or their capacity to represent lipid species. Here we introduce Lipid Spectrum Generator, an open-source *in-silico* spectral library generator specifically designed to aid in the identification of lipids in LC-MS/MS analysis.

BACKGROUND

The complexity of lipids throughout nature is immense, where tens to hundreds of thousands of distinct lipid species may simultaneously be present within a single matrix [1-4]. The individual species which constitute this diversity are not themselves genetically encoded but arise in response to the dietary, physiological, and environmental factors which influence a given biological system [2-4]. Their dynamic nature in response to these factors, in conjunction with their biological importance, has led to their analysis being used to monitor cellular metabolic processes or even the development of pathologies [1, 3-8]. The discipline of Lipidomics, which concerns the identification and characterisation of these compounds, must therefore contend with this immense molecular diversity to reveal biomarkers of relevance. As a result, analytical technologies such as tandem mass spectrometric (MS/MS) analysis have championed the field due to their potential to structurally elucidate virtually all sample constituents [1-3, 9, 10].

Lipid identification employing MS/MS analysis follows the fragmentation of a precursor compound and the interpretation of its product ions (fragmentation spectra) [1, 2]. In place of manual interpretation, identification requires a reference against which fragmentation patterns can be compared. In ideal cases, molecular standards may be used to determine the fragmentation patterns for lipids of interest. However, the number of standards required to annotate a sample in this way is logistically infeasible for even the most well-resourced laboratories. As a result, the interpretation of fragmentation spectra, and thus the characterisation of lipid species present in a matrix, represents a severe bottleneck in lipidomic analyses.

In the absence of standards, lipidomic workflows have relied on computationally generated (*in-silico*) fragmentation spectra, as they facilitate rapid sample annotation and characterisation where it is otherwise impossible [9-13]. The advent of such spectra has resulted from an increased understanding of lipid fragmentation pathways [14, 15] and the often predictable manner in which lipids fragment. Hence, spectra of this kind are regularly produced through various rules-based [11, 12, 16], algorithmic [17-19], or machine learning [20] approaches, enabling an accurate representation of authentic lipid fragmentation patterns.

However, despite their broad adoption, many *in-silico* libraries are limited in their capacity to represent lipid species. Complications of this sort arise due to (i) limited consideration of lipid classes; (ii) shallow representation of lipid molecular species; (iii) incomplete fragmentation patterns which exclude characteristic ions or misrepresent the relative intensities for each fragment; (iv) incorrect annotations which assume structure not apparent within the spectra; (v) and the presence of duplicate fragmentation patterns which can obscure identifications. Whilst some recent publications have sought to remedy these shortcomings [11, 12, 20-22], many of these focus on replacing pre-existing lipidomic workflows and provide libraries that are otherwise externally incompatible, preventing their use elsewhere. Hence despite their utility, such cases represent a deviation from the convenience and universality of previous *in-silico* libraries [16].

In response to these limitations, we have developed an *in-silico* library generator for lipidomic analysis, Lipid Spectra Generator (LSG). The software is open-source, Python-based, and implements the cross-platform Qt5 graphical user interface to guide the user through the steps necessary to generate a highly

customisable *in-silico* library for lipidomic analysis. To date, LSG contains the most comprehensive set of templates available for a software of its type, enabling spectra generation for 87 lipid classes, using over 1000 potential fatty acid chains, allowing the generation of well over 1.2 million lipid species. Users may export MSP *in-silico* generated spectral libraries from these templates, enabling incorporation into any third-party or vendor-specific mass spectral library application and CSV precursor and transition lists that permit targeted lipidomic analysis and compatibility with the Skyline ecosystem [23].

EXPERIMENTAL

SOFTWARE DESIGN

Lipid Spectra Generator is an open-source, Python-based script that is freely available for download from GitHub (https://github.com/98104781/LSG/releases/tag/v1.0.0). A cross-platform Qt5-based GUI is implemented using PySide2, and it is designed to guide the user through the steps necessary to produce an MSP formatted spectral library or CSV formatted inclusion and transition lists. Lipid species are generated in the software through the specification of fatty acid tails and the selection of lipid classes. Fatty acids may be individually specified or generated as a range, permitting the creation of both highly specific and comprehensive spectral libraries. Lipid fragmentation spectra are generated from a set of pre-defined templates which can be customised in software, allowing for user modification of lipid fragmentation patterns and fragment intensities. The lipid fragmentation templates used have been collated from a set of peer-reviewed fragmentation studies of lipid classes, annotated mass spectra of lipid standards, and previously validated *in-silico* templates. The literature sources for these templates are provided in Supplementary Table 1.

MATERIALS

Methanol (MeOH), acetonitrile (ACN), and isopropanol (IPA) used were of LC/MS grade, Honeywell B&J Brand, purchased from ChemSupply Australia.

Ultra-pure water (UP) (18.2 MΩ-cm) was sourced from an in-lab water purification system (Sartorius).

PBS tablets, Avanti EquiSPLASH mix, HPLC grade chloroform, analytical reagent grade ammonium acetate and ammonium formate, and LC/MS grate (98% - 100% purity) Lichropur brand formic acid were all purchased from Sigma-Aldrich.

Glutamax, fetal bovine serum (FBS), Dulbecco's Modified Eagle Medium (DMEM) and eagle's minimal essential medium (EMEM) were purchased from Thermofisher scientific (Thermo Fisher Scientific, VIC, Scoresby).

SH-SY5Y SAMPLE PREPARATION

SH-SY5Y (human neuroblastoma) cells (passage 23) were cultured in a T175 flask using DMEM supplemented with 5% glutamax and 10% FBS. The cells were then seeded in a t75 flask at 50% confluency, using EMEM with 5% glutamax and 10% FBS. The cells were then left to adhere overnight. Following this, fresh media was applied, and cells were left for 48hrs. They were then harvested according to the following protocol:

First, media was removed and replaced with 5 mL of TrypLE. Flasks were then left in an incubator at 37 °C for 5 minutes. Cells suspended in the TrypLE were moved to a 15 mL tube and centrifuged at 300 g at room temperature for 5 minutes. The supernatant was then discarded, and 5 mL of PBS was added. The pellet was then gently broken up and again centrifuged under the same conditions. The previous two steps were conducted twice to ensure adequate washing. The remaining cell pellets were snap-frozen with liquid nitrogen and stored at -80 °C until extraction. Three replicate extracts were prepared according to the following protocol:

In a 15 mL falcon tube, 500 µL of chloroform was added to approximately 25 mg of cell mass, followed by 500 µL of methanol containing a mix of deuterium-labelled internal standards (50 ppb; EquiSPLASH Mix, Avanti Polar Lipids). Next, the samples were vortex-mixed for 30 seconds before adding 500 µL of ultra-pure water. Following this, the samples were then agitated for a period of 30 minutes at 4 °C on a rotary mixer before centrifugation at 3000 rpm for 10 minutes. After phase separation had occurred, 200 µL of the organic layer (bottom) was removed and dried under a continuous flow of dry N_2. Dry extracts were then reconstituted in ACN: IPA (100 µL; 1:1) and stored at –80 °C until analysis.

LC-MS/MS ANALYSIS

To demonstrate the utility of LSG, a set of LC-MS/MS lipidomic characterisations were conducted and processed against an LSG-generated spectral library. The number of identifications observed, along with their identification scores, were then compared to those achieved with a library generated using the LipiDex[11] Library Generator, and the previously published LipidBlast *in*-silico spectral library[16]. The LSG spectral library was generated using the default specified-tails list and contained only those classes expected to ionise under the selected ESI conditions. The LipiDex spectral library was generated using the LipiDex_HCD_Acetate library, with all default available fatty acids. Results were then batch processed and identified through use of MzMine 2, version 2.53 [24, 25]. For each library, the MzMine peak lists were first filtered using a precursor search, and then identified using the local spectral database search option. Identifications were accepted with a precursor tolerance of 5 ppm, spectral tolerance of 25 ppm, and minimum weighted dot-product cosine of 0.5, using two minimum matched signals. The 'crop spectra to m/z overlap' option was disabled to permit

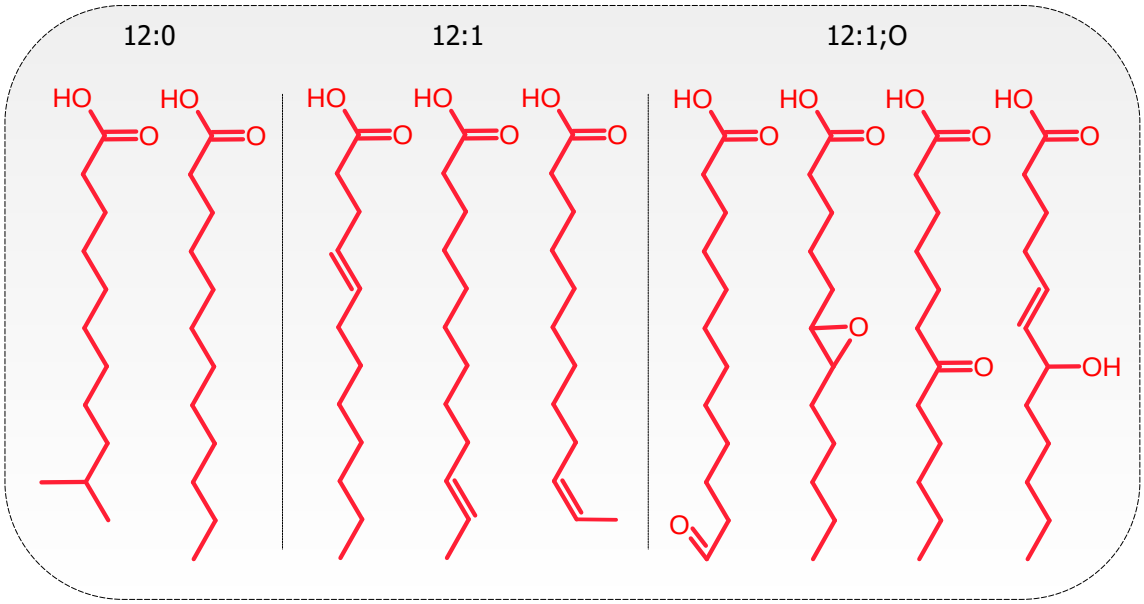


Figure 1 Example annotations assigned to isomeric fatty acids. Species on the left may be represented with 12:0, indicating a chain of twelve carbons with zero points of desaturation. Species in the center may be represented with the name 12:1, signifying a chain of twelve carbons, one point of desaturation. Species on the right may be represented with the name 12:1;O, signifying a chain of twelve carbons, one point of desaturation and one point of oxidation. Assuming no further fragmentation or sample derivatisation is conducted, the true structure of the fatty acids cannot be determined from the parent lipids' fragmentation pattern. Instead, the systematic names (e.g. 12:1:O and 12:0) are used to ambiguously designate the fatty acids' structure.

comparison, and results were weighted using the NIST11 (LC) weights.

During analysis, SH-SY5Y extracts were chromatographically separated on a Waters Acquity CSH column (1.7 µm, 135 Å, 150 mm x 2.1 mm) using an Agilent Infinity II 1290 UHPLC. The flow rate and column temperature were maintained at 0.5 mL/min and 65 °C, respectively. The mobile phases used differed in buffer composition between positive (10mM ammonium formate + 0.1% (v/v) formic acid to both A and B) and negative (10mM ammonium acetate to both A and B) ESI modes, although the solvent composition of mobile phases A (ACN: H_2O; 60: 40) and B (IPA: ACN; 90: 10) remained the same. The same chromatographic gradient was used for both analysis modes. The mobile phase was held initially at 30% B for 2 minutes and then raised linearly to 50% B from 2 to 2.5 minutes. Additional linear increases up to 85% B and 99% B occurred between 2.5 to 13 and 13 to 13.5 minutes, respectively. The final mix of 99% B was held until 15 minutes. The column was then re-equilibrated with an immediate drop to 30% B, which was held until 18 minutes. Mass spectrometric analysis was conducted using a Thermo Scientific Q Exactive Plus in Top-10 mode. MS1 spectra were accumulated between masses 200 to 1200 m/z for a max IT of 100 ms and AGC of 3e6 and analysed with a mass resolution of 70,000. MS2 spectra were accumulated for a max IT of 60 ms with an AGC of 1e5 and mass resolution of 17,500, and fragmented with a stepped collision energy at 10, 20 and 30 units.

ASSESSMENT OF TRUE-POSITIVE RATE

Validation followed the LC-MS/MS characterisations to determine the extent of true-positive identifications that could be achieved with each library. To conduct these tests, a previously evaluated reference matrix, the publicly available AdipoAtlas [26] samples, was used, which consisted of a database of 1,636 lipid identifications. 9 lipid classes from subcutaneous and visceral fat extracts assessed in positive ion mode only. These classes were selected based on their common presence throughout all lipidomic libraries and samples. Identification and scoring were conducted identically to the LC-MS/MS Analysis section, except where the spectral libraries were modified to contain only the expected fatty acids to limit analysis time. The resulting identifications were then compared to those present in reference 26 to determine the proportion of matching identifications. Any additional identifications were ignored for the purpose of the comparison.

RESULTS AND DISCUSSION

Many of the complications associated with previous *in-silico* libraries for lipidomics stem from restrictions related to their size and accuracy. Primarily, these arise due to the influence these aspects have on the efficacy of a library during analysis [27].

The size of a spectral library represents the total quantity of spectra it contains and, thus, the number of molecular species it accounts for. Any compound not present within a library cannot be identified when using

it. Hence, more extensive and comprehensive libraries are often preferred. When considering lipidomic spectral libraries, size may be further segmented into terms of lipid class and lipid molecular species. Yet, when library size is considered, classes are often the primary concern of lipidomic spectral libraries. As a result, many libraries provide a shallow representation of lipid molecular diversity and, thus, a limited domain of utility. However, spectral libraries may likewise become excessive in size and potentially burden analysis workflows due to the computational requirements necessary for their generation and use.

Beyond size, spectral accuracy is also critical in determining the effectiveness of an *in-silico* spectral library. The accuracy of an *in-silico* spectrum depends on various factors, including the accuracy of mass labels and the completion and specificity of fragmentation patterns. However, spectrum accuracy can also depend on the circumstances of analysis, e.g. the MS used. The presence and intensity of product ions can often vary along with analysis factors such as collision energy, causing otherwise accurate spectra may become invalid based on these circumstances [27].

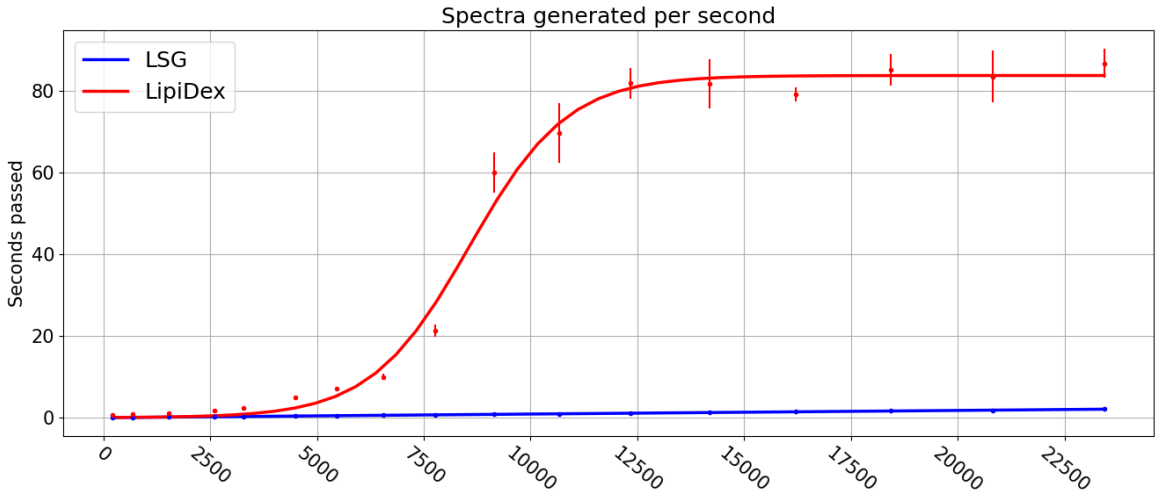


Figure 2 Libraries of varying size were generated to reveal the trend between processing time and the number of generated spectra. For these tests, TGs were selected due to the large number of potential species that could be generated. Although TG species may more frequently be observed in the form $[M+NH_4]+$, the $[M+Na]+$ adduct was selected here due to the far greater spectra complexity represented by LSG than in LipiDex. I.e. in a LipiDex library, a TG species with 3 unique fatty acids would produce a spectra of 4 fragments in a $[M+NH_4]+$ spectra and 3 in a $[M+Na]+$ spectra; whereas with LSG, the same species would produce 8 fragments in a $[M+NH_4]+$ spectra and 13 in a $[M+Na]+$ spectra. This indicates that LSG may produce far more complex spectra than competing software in a minute fraction of the time. Each test was conducted a total of 5 times, and the average duration plotted. Other spectra may take more or less time to generate, based on their complexity.

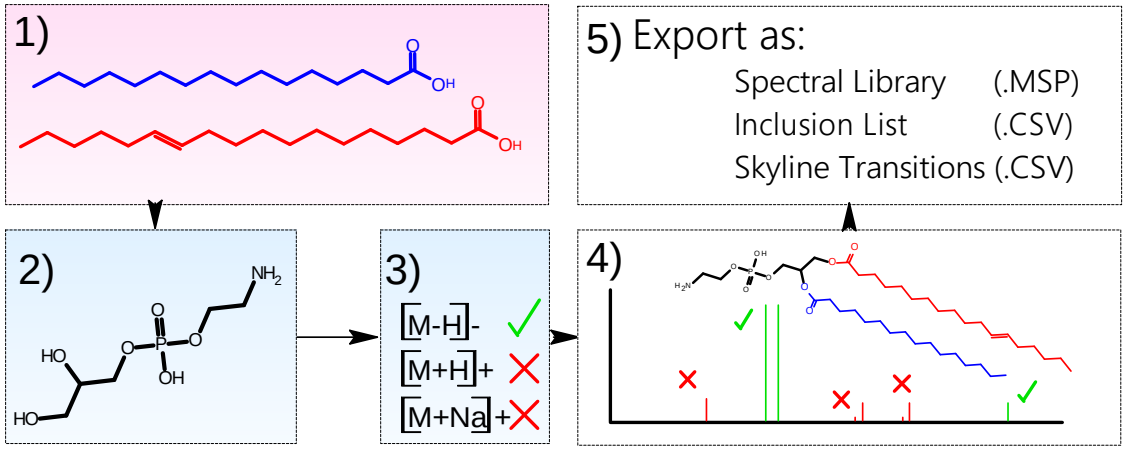


Figure 3 Flowchart outlining the process of spectral library generation using LSG. **1)** Fatty acid species are defined. Individual fatty acid species may be chosen or generated as a range. **2)** Lipid classes of interest are selected. **3)** Desired fragmentation spectra, based on adducts, are selected. **4)** Desired fragmentation spectra may be further modified to better represent the experimental fragmentation patterns observed. **5)** The export format is selected. Exporting as a .MSP file will produce a spectral library containing those classes selected. A .CSV inclusion list, formatted for a QE+ Orbitrap mass

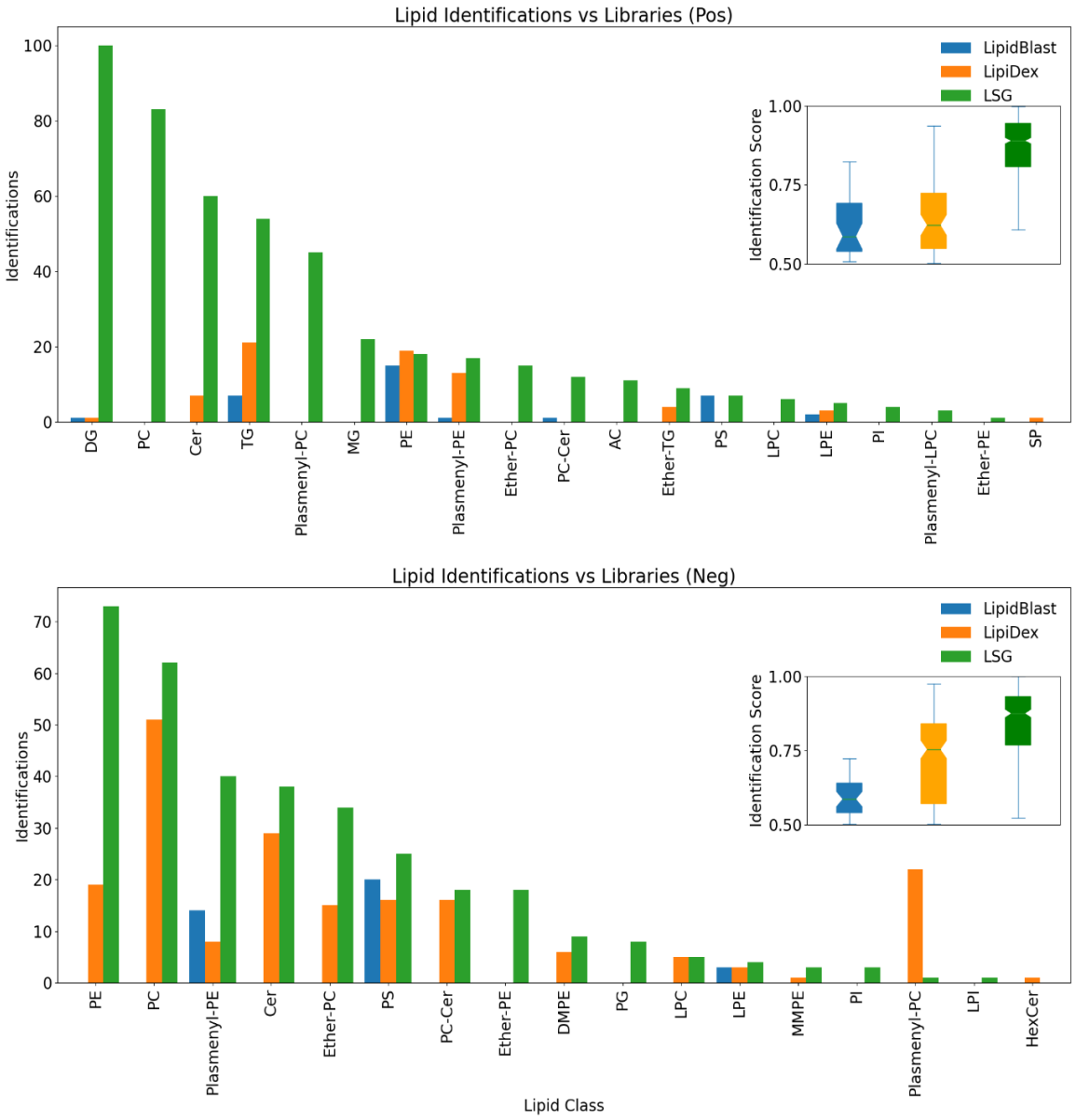


Figure 4 Quantity of candidate lipid identifications achieved with differing in-silico libraries organized by class, along with identification scores. The LSG library results are represented in green, LipiDex in orange and LipidBlast is in blue. The cut-off value for the identification score was reduced to 0.5 to permit all the libraries used to obtain candidates. Due to this decreased cut-off, some identifications are likely to be false positives, and thus remain as candidates rather than true identifications. 474 lipid candidates were achieved with LSG in positive analysis mode, and 342 in negative. 70 lipid candidates were achieved with LipiDex in positive mode, and 192 in negative. 34 lipid candidates were achieved with LipidBlast in positive mode, and 37 in negative. Candidates include multiple adduct states for individual lipid species. **AC** – Acylcarnitine, **CE** – Cholesterol Ester, **MG** – Monoacylglycerol, **DG** – Diacylglycerol, **TG** – Triacylglycerol, **PC** – Phosphatidylcholine, **LysoPC** – Lyso Phosphatidylcholine, **PE** – Phosphatidylethanolamine, **LysoPE** – Lyso Phosphatidylethanolamine, **MMPE –** Monomethyl Phosphatidylethanolamine, **DMPE** – Dimethyl Phosphatidylethanolamine, **PG** – Phosphatidylglycerol, **PI** – Phosphatidylinositol, **PS** – Phosphatidylserine, **SM** – Sphingomyelin, **HexCer –**ceramide-1-hexose.

Ideally, a spectral library must dynamically account for these variations, as they can influence the relative number of true-positive identifications made.

Correct spectra annotations are also of particular importance when considering spectra accuracy, as isomeric lipid species can often produce identical fragmentation patterns. As a result, spectral libraries can risk incorrectly assigning identifications to numerous candidates. Conventionally, conflicts of this sort can be resolved by assigning fragmentation patterns with structurally ambiguous annotations, such as those shown in Figure **1**, which prevent the over-interpretation of features that are not evident in the spectra [28, 29].

LSG was designed to directly address these concerns and permit the generation of spectral libraries suitable

for all analyses. As a result, many of the software's features are specifically designed to enable precise control over library size and spectral accuracy. Libraries are scaled in terms of classes and molecular species through the selection of fragmentation templates and the specification of fatty acids, whereas accuracy is controlled by adding, subtracting or manipulating the fragments in the templates. This enables dynamic modification of generated spectra through user input to account for variations that may arise due to instrumentation or method parameters. A simplified representation of the process of library generation is shown in Figure 3. The libraries produced are currently the most comprehensive of their kind, with a total of 136 fragmentation templates and 67 literature-sourced fatty acid species pre-defined in the software [30, 31], though over 1000 fatty acid species may be created to expand coverage over both well-documented and hypothetical lipid species. The generated spectra are likewise dynamic and can be modified through user input to account for variations that may arise due to instrumentation or method parameters.

Libraries are rapidly produced, as highlighted in Figure 2, with approximately 12,000 unique fragmentation spectra generated per second, representing almost two orders of magnitude increase in speed compared to the LipiDex Library Generator, which can take up to 80 seconds to create a library of equal size. Furthermore, library sizes are efficiently reduced to prevent computational burden through the use of ambiguous fatty acid notation. The fatty acid species represented in software are intentionally ambiguous as to the arrangement of certain isomeric features, such as the isomerism of double bonds or the sites and types of oxidation. Maintaining this level of ambiguity permits a single fatty acid annotation to represent a plethora of isomeric molecular species, such as those in Figure **1**. The resulting compression prevents molecular structure over-interpretation and considerably reduces library size without compromising the number of molecular species accounted for, as a single fragmentation spectra may apply to numerous molecular species. Structural ambiguity of this type is further utilised when considering the sn-arrangement for the fatty acids attached to a lipid's backbone. Typically, the fragmentation spectra produced from these types of isomers differ only in the ratios of certain fragments but are otherwise identical. Thus, by disregarding this variability and maintaining an equal ratio between these variable fragments, library size is further compressed by a proportion equal to the number of potential isomers at the expense of the specificity of the identification. However, for some specific lipids, such as cardiolipins, unique fragments are indeed produced based on this arrangement. For these classes, some sn-specificity is considered by default; however, this option may be disabled by the user if maintaining a smaller library is of more utility than the added specificity. In cases where the generated spectra are devoid of information regarding lipid fatty acid composition, whether by default or after user manipulation, lipid notation is changed to species-level notation [32], and the libraries are further compressed.

Minor fragments produced by some oxidised fatty acid species are ignored to maintain the benefits of this compression. While such fragments may provide evidence for the precise site of fatty acid oxidation, they are not necessary for determining the degree of oxidation nor on which fatty acid the oxidation occurs. Rather, with their removal, the degree of structural information contained within the fragmentation spectra remains consistent with that regarding the types and sites of desaturation. A caveat of this simplification is that the generated fragmentation spectra are not representative of any particular species in which the type and site of oxidation interfere with the generalised

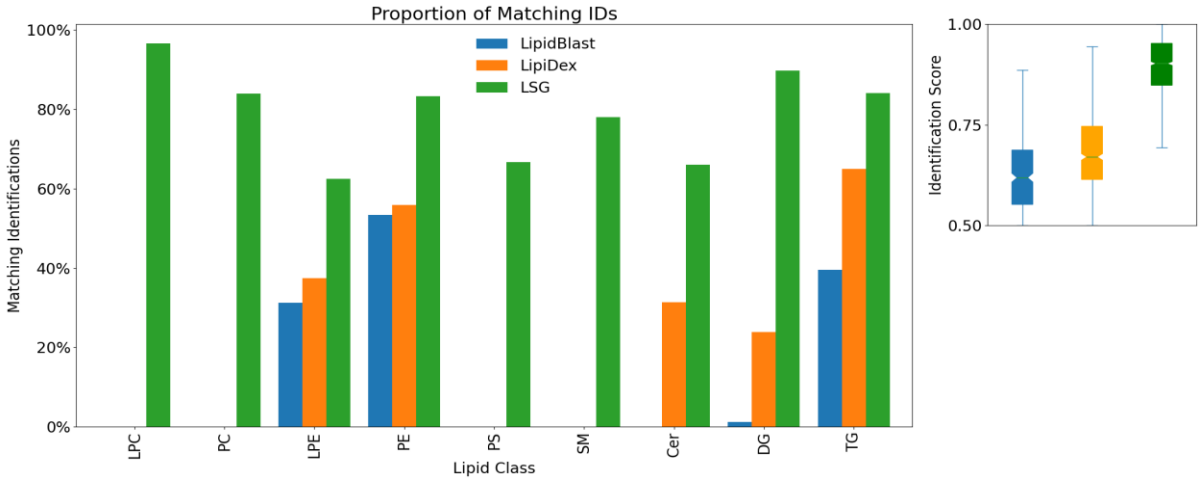


Figure 5 Quantity of candidate identifications matching the literature contents of the AdipoAtlas reference samples. Sample processing and scoring was conducted in an identical manner to Figure 4. Analysis was conducted in positive mode only, as several of the evaluated classes do not ionize under negative esi conditions.

fragmentation pathways, which serve as the basis for the templates. Fortunately, such deviations are limited to specific molecular arrangements, such as with ceramides containing α-hydroxy fatty acids[33]. However, it is essential to understand the limitations of the templates regarding these molecular species. Other factors which may also influence the relative intensity of certain fatty acid fragments, such as poly-unsaturation, are likewise not considered.

To verify whether the modifications provided by LSG translated to the more thorough annotation of a sample's lipidome, an LSG-generated spectral library was applied in analysis, and the number of features observed compared to those achieved with a LipiDex generated library, and the LipidBlast spectral library. For this comparison, triplicate injections of an SH-SY5Y extract were chromatographically separated in both positive and negative esi modes and processed according to the aforementioned conditions. The results of this comparison are provided in Figure 4. LSG achieves a 3.5 to 4-fold increase in lipid identifications over the LipiDex and LipidBlast libraries, while identification scores also increase markedly. It is suggested here that the greater quantity of annotations observed results from the greater coverage of molecular species provided by the LSG library. In contrast, the increase in identification score is suggested to be the result of its improved spectral accuracy. However, verifying identification quality is also necessary to conclude the differing performances of the libraries. The publicly available AdipoAtlas [26] samples were used as a reference to benchmark the extent of true-positive identifications that could be achieved with the libraries used. The results of this evaluation are presented in Figure 5. For the 9 lipid classes examined, the LSG-generated spectral library consistently achieved a near-complete level of true-positive lipid identifications, while each identification similarly matched with greater score.

The expanded coverage of molecular species is evident where LSG contains a total of 136 pre-defined fragmentation templates encompassing a total of 87 classes, approximately 2-fold greater than the 54 fragmentation templates and 29 classes provided by LipidBlast. However, this disparity is also exacerbated due to LSG's more thorough representation of lipid species. Whilst the LipidBlast library is purported to contain over 200,000 fragmentation spectra, the library does not utilise many of the compression techniques featured in LSG. Thus the LipidBlast library contains multiple duplicate spectra, which specify structural features that cannot be determined from the fragmentation pattern. This may be observed where LipidBlast provides 5476 spectra per phospholipid class to represent a total of 1176 unique fragmentation patterns. Such duplicates are elsewhere known to be burdensome, requiring additional data processing to eliminate duplicate identifications [34].

In contrast, the LipidDex Library Generator provides 72 fragmentation templates covering 46 classes in their HCD_Acetate Library, whilst utilising many of the same compression techniques used in LSG. However, the templates used to maintain the same simplified patterns used throughout LipidBlast, which often ignore the presence of several major characteristic ions. As a result, the LipiDex-generated library cannot achieve the same quantity of identifications or scores as LSG.

CONCLUSION

LSG is an open-source *in-silico* spectral library generator specifically designed to aid in the identification of lipids in LC-MS/MS analysis. LSG-generated libraries are exportable in the universal .MSP format, permitting their use in any third-party or vendor-specific mass spectral library application. Furthermore, the features included in this software allow for the production of user-catered spectral libraries by selecting individual lipid classes and modifying their fragmentation spectra. With a pre-set 136 lipid templates covering 87 lipid classes, and over 1000 fatty acids, LSG is the most comprehensive software of its type. In direct comparison to previously published in-silico libraries used for lipidomic analysis, the greater comprehensivity of LSG had enabled a 3.5 to 4-fold increase in lipid identifications, greater coverage of true-positive identifications, along with a marked increase in identification score.

ASSOCIATED CONTENT

Supporting Information

Supplementary Table 1 – Lipid Classes included in the LSG software. References are provided as the source of the fragmentation templates.

Supplementary Data: LSG Candidate identifications in Pos and Neg modes for the SH-SY5Y analysis. Pos and Neg mode .raw datafiles used in SH-SY5Y analysis.

AUTHOR INFORMATION

Corresponding Author

* Correspondence: matthew.padula@uts.edu.au

Author Contributions

DG: Conceptualisation, experimental design, software development, performed experiments, data analysis, wrote original draft and edited manuscript.
JV: Sample preparation, performed experiments
DB: Reviewed and edited manuscript
MP: Reviewed and edited manuscript

Acknowledgments

The Authors would like to thank Associate Professor Ken Rodgers for supplying the SH-SY5Y neuroblastoma cells used in the lipidomic analysis; in addition to Johnny Gertner and Thomas Lockwood for their support and Python experience.
DG is supported by a UTS Doctoral scholarship.

References

1. Wenk, M. R., The emerging field of lipidomics. *Nature reviews Drug discovery* **2005,** *4* (7), 594-610.

2. Wenk, M. R., Lipidomics: New Tools and Applications. *Cell* **2010,** *143* (6), 888-895.

3. Han, X., Lipidomics for studying metabolism. *Nature Reviews Endocrinology* **2016,** *12* (11), 668-679.

4. Wang, M.; Wang, C.; Han, R. H.; Han, X., Novel advances in shotgun lipidomics for biology and medicine. *Progress in lipid research* **2016,** *61*, 83-108.

5. Pietiläinen, K. H.; Róg, T.; Seppänen-Laakso, T.; Virtue, S.; Gopalacharyulu, P.; Tang, J.; Rodriguez-Cuenca, S.; Maciejewski, A.; Naukkarinen, J.; Ruskeepää, A.-L.; Niemelä, P. S.; Yetukuri, L.; Tan, C. Y.; Velagapudi, V.; Castillo, S.; Nygren, H.; Hyötyläinen, T.; Rissanen, A.; Kaprio, J.; Yki-Järvinen, H.; Vattulainen, I.; Vidal-Puig, A.; Orešič, M., Association of lipidome remodeling in the adipocyte membrane with acquired obesity in humans. *PLoS Biol* **2011,** *9* (6), e1000623-e1000623.

6. Haoula, Z.; Ravipati, S.; Stekel, D. J.; Ortori, C. A.; Hodgman, C.; Daykin, C.; Raine-Fenning, N.; Barrett, D. A.; Atiomo, W., Lipidomic analysis of plasma samples from women with polycystic ovary syndrome. *Metabolomics* **2015,** *11* (3), 657-666.

7. Gross, R. W.; Han, X., Lipidomics in Diabetes and the Metabolic Syndrome. In *Methods in Enzymology*, Academic Press: 2007; Vol. 433, pp 73-90.

8. Hu, T.; Zhang, J. L., Mass-spectrometry-based lipidomics. *J Sep Sci* **2018,** *41* (1), 351-372.

9. Köfeler, H. C.; Fauland, A.; Rechberger, G. N.; Trötzmüller, M., Mass Spectrometry Based Lipidomics: An Overview of Technological Platforms. *Metabolites* **2012,** *2* (1), 19-38.

10. Cajka, T.; Fiehn, O., Comprehensive analysis of lipids in biological systems by liquid chromatography-mass spectrometry. *TrAC Trends in Analytical Chemistry* **2014,** *61*, 192-206.

11. Hutchins, P. D.; Russell, J. D.; Coon, J. J., LipiDex: An Integrated Software Package for High-Confidence Lipid Identification. *Cell Systems* **2018,** *6* (5), 621-625.e5.

12. Koelmel, J. P.; Kroeger, N. M.; Ulmer, C. Z.; Bowden, J. A.; Patterson, R. E.; Cochran, J. A.; Beecher, C. W. W.; Garrett, T. J.; Yost, R. A., LipidMatch: an automated workflow for rule-based lipid identification using untargeted high-resolution tandem mass spectrometry data. *BMC Bioinformatics* **2017,** *18* (1), 331.

13. Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M., MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature Methods* **2015,** *12* (6), 523-526.

14. Pi, J.; Wu, X.; Feng, Y., Fragmentation patterns of five types of phospholipids by ultra-high-performance liquid chromatography electrospray ionization quadrupole time-of-flight tandem mass spectrometry. *Analytical Methods* **2016,** *8* (6), 1319-1332.

15. Hsu, F.-F.; Turk, J., Electrospray ionization with low-energy collisionally activated dissociation tandem mass spectrometry of glycerophospholipids: mechanisms of fragmentation and structural characterization. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2009,** *877* (26), 2673-2695.

16. Kind, T.; Liu, K.-H.; Lee, D. Y.; DeFelice, B.; Meissen, J. K.; Fiehn, O., LipidBlast in silico tandem mass spectrometry database for lipid identification. *Nature methods* **2013,** *10* (8), 755-758.

17. Wolf, S.; Schmidt, S.; Müller-Hannemann, M.; Neumann, S., In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinformatics* **2010,** *11* (1), 148.

18. Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S., MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *J Cheminform* **2016,** *8*, 3.

19. Heinonen, M.; Rantanen, A.; Mielikäinen, T.; Pitkänen, E.; Kokkonen, J.; Rousu, J. In *Ab initio prediction of molecular fragments from tandem mass spectrometry data*, German Conference on Bioinformatics, Gesellschaft für Informatik eV: 2006.

20. Kangas, L. J.; Metz, T. O.; Isaac, G.; Schrom, B. T.; Ginovska-Pangovska, B.; Wang, L.; Tan, L.; Lewis, R. R.; Miller, J. H., In silico identification software (ISIS): a machine learning approach to tandem mass spectral identification of lipids. *Bioinformatics* **2012,** *28* (13), 1705-1713.

21. Kind, T.; Okazaki, Y.; Saito, K.; Fiehn, O., LipidBlast templates as flexible tools for creating new in-silico tandem mass spectral libraries. *Anal. Chem.* **2014,** *86* (22), 11024-11027.

22. Ni, Z.; Angelidou, G.; Lange, M.; Hoffmann, R.; Fedorova, M., LipidHunter Identifies Phospholipids by High-Throughput Processing of LC-MS and Shotgun Lipidomics Datasets. *Anal. Chem.* **2017,** *89* (17), 8800-8807.

23. Pino, L. K.; Searle, B. C.; Bollinger, J. G.; Nunn, B.; MacLean, B.; MacCoss, M. J., The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass spectrometry reviews* **2020,** *39* (3), 229-244.

24. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M., MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **2010,** *11*, 395.

25. Myers, O. D.; Sumner, S. J.; Li, S.; Barnes, S.; Du, X., One Step Forward for Reducing False Positive and False Negative Compound Identifications from Mass Spectrometry Metabolomics Data: New Algorithms for Constructing Extracted Ion Chromatograms and Detecting Chromatographic Peaks. *Anal. Chem.* **2017,** *89* (17), 8696-8703.

26. Lange, M.; Angelidou, G.; Ni, Z.; Criscuolo, A.; Schiller, J.; Blüher, M.; Fedorova, M., AdipoAtlas: A reference lipidome for human white adipose tissue. *Cell Rep Med* **2021,** *2* (10), 100407.

27. Stein, S., Mass Spectral Reference Libraries: An Ever-Expanding Resource for Chemical Identification. *Anal. Chem.* **2012,** *84* (17), 7274-7282.

28. Liebisch, G.; Vizcaíno, J. A.; Köfeler, H.; Trötzmüller, M.; Griffiths, W. J.; Schmitz, G.; Spener, F.; Wakelam, M. J. O., Shorthand notation for lipid structures derived from mass spectrometry. *Journal of lipid research* **2013,** *54* (6), 1523-1530.

29. Züllig, T.; Köfeler, H. C., HIGH RESOLUTION MASS SPECTROMETRY IN LIPIDOMICS. *Mass Spectrometry Reviews* **2021,** *40* (3), 162-176.

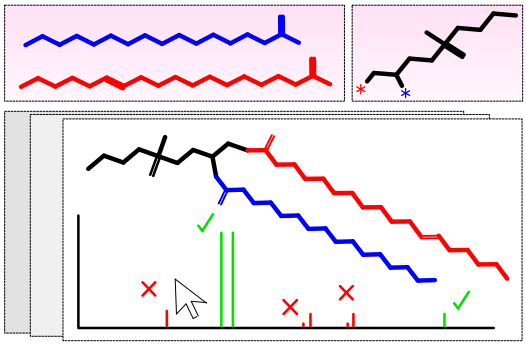
30. Kenar, J. A.; Moser, B. R.; List, G. R., Chapter 2 - Naturally Occurring Fatty Acids: Source, Chemistry, and Uses. In *Fatty Acids*, Ahmad, M. U., Ed. AOCS Press: 2017; pp 23-82.

31. Riecan, M.; Paluchova, V.; Lopes, M.; Brejchova, K.; Kuda, O., Branched and linear fatty acid esters of hydroxy fatty acids (FAHFA) relevant to human health. *Pharmacology & Therapeutics* **2022,** *231*, 107972.

32. Liebisch, G.; Fahy, E.; Aoki, J.; Dennis, E. A.; Durand, T.; Ejsing, C. S.; Fedorova, M.; Feussner, I.; Griffiths, W. J.; Köfeler, H.; Merrill, A. H., Jr.; Murphy, R. C.; O'Donnell, V. B.; Oskolkova, O.; Subramaniam, S.; Wakelam, M. J. O.; Spener, F., Update on LIPID MAPS classification, nomenclature, and shorthand notation for MS-derived lipid structures. *J Lipid Res* **2020,** *61* (12), 1539-1555.

33. Hsu, F.-F., Electrospray ionization with higher-energy collision dissociation tandem mass spectrometry toward characterization of ceramides as [M + Li]+ ions: Mechanisms of fragmentation and structural identification. *Anal Chim Acta* **2021,** *1142*, 221-234.

34. Dynamics, N. Progenesis QI Support for LipidBlast. http://nonlinear.com/progenesis/qi/v2.4/faq/compound-search-lipid-blast.aspx.



For Table of Contents Only