

WaxAlly: NEW SOFTWARE TO IDENTIFY ACYCLIC LIPIDS FROM GAS CHROMATOGRAPHY COUPLED TO MASS SPECTROMETRY DATALucas Paradizo Roma^{a,*} and Déborah Yara A. C. Santos^a^aDepartamento de Botânica, Instituto de Biociências (IB), Universidade de São Paulo (USP), 05508-090 São Paulo – SP, Brasil

Received: 12/06/2023; accepted: 06/27/2024; published online: 08/16/2024

Gas chromatography coupled to mass spectrometry (GC-MS) has been widely used, and along with other methodologies such as derivatization, is a powerful tool to analyze phytocompounds, including cuticular wax components. The loss of structural information upon fragmentation of waxes makes it difficult to interpret and identify the GC-MS mass spectra of these compounds with the traditional methods, as digital libraries. Therefore, as a new complementary alternative to traditional methods of identifying aliphatic acyclic lipids based on mass spectra, we present the new software WaxAlly in this technical note. WaxAlly is a new software created based on simple algorithms, enabling the rapid recognition by the user of eight aliphatic acyclic lipid classes, including alkanes, alkenes, aldehydes, ketones, esters, and TMS derivatives of free fatty acids, primary and secondary alcohols, and their homologues with carbon chains varying between 10 and 100 carbons. Additionally, the WaxAlly software provides a section for data organization, internet comparison to NIST and PubChem databases, and academic information about mass fragmentation of acyclic lipids. The software has proven to be a very useful complementary technique in identifying plant wax lipid homologues, and new fragmentation patterns of lipid classes can be added in the future to improve the program.

Keywords: cuticular waxes; derivatized compounds; standard digital library; fatty acids; alcohols.

INTRODUCTION

Lipids are organic compounds characterized by their hydrophobic nature, often soluble in non-polar organic solvents. Examples of lipids are fatty acids, steroids, and waxes. These compounds play crucial roles in maintaining life. For instance, they serve as elementary constituents of lipoprotein membranes,¹ source of energy,² chemical signalers via peroxidation,³ and are a crucial part of hydrophobic coatings on plant surfaces.⁴ Moreover, the understanding of lipids holds great importance for the fields of archeology⁵ and health.⁶

The lipid coatings that cover plant surfaces, known as cuticular waxes, stand out primarily for their role as a barrier, limiting water loss in aerial tissues with no secondary thickening, such as flowers, leaves, stems, and fruits.⁷ Waxes exhibit a diverse array of lipid classes, predominantly alkanes, primary and secondary alcohols, aldehydes, ketones, and esters. These compounds are mainly acyclic and aliphatic, often unbranched, and long-chained, typically ranging in size from 16 to 36 carbons, with esters reaching up to 60 carbons, arranged in homologous series.⁸

Waxes, along with other lipids, are traditionally analyzed through gas chromatography coupled to mass spectrometry (GC-MS).^{9,10} The electron impact ionization (EII) technique at 70 eV is widely employed alongside GC-MS, facilitating the separation of complex mixtures and providing a robust method for compound identification.¹¹ The efficacy of MS is enhanced when combined with additional methodologies, such as derivatization, which modifies the analyte properties to increase molecular thermal and catalytic stability, as well as reduce volatilization point.¹²

Although mass spectrometry (MS) has significantly advanced lipid research,¹³ interpreting mass spectra and identifying compounds remains a challenge hindering lipid study. Among the best-known ways to identify surface lipids using GC-MS are: (i) comparing mass spectra and retention times of the sample with those obtained by

standards injected in the same methodology; (ii) applying theoretical knowledge of specific fragmentation patterns for each lipid class; and (iii) utilizing virtual libraries (i.e., NIST). Even though identification via standards comparison is robust, obtaining a diverse range of standards can be challenging and quite expensive. Interpreting fragmentation patterns based on theoretical knowledge enables the identification of different compounds but requires expertise in mass spectrometry, potentially consuming time for inexperienced researchers or studies with numerous samples. Conversely, virtual libraries like NIST can swiftly and reliably identify many plant compounds based on spectra.¹⁴ However, for compounds whose diagnostic fragments are not significant or multiple and constant within a class, homologue identification may be compromised, as is often observed in lipids.¹⁵ Hence, combining these strategies is essential to enhance the efficacy of lipid mass spectra identification.

Considering saturated acyclic lipids, characteristic of waxes, two sets of MS fragments are essential for their identification. One set is responsible for recognizing the lipid class (such as alkanes, ketones, and others), while the other set is necessary for homologue identification. The fragments related to organic function (called diagnostic class ions or fragments, DCF) are repetitive and identical for most of the homologous series, thus enabling the recognition of the compound class. On the other hand, fragments related to the original molecule mass (for instance molecular ion (M^+) and $[M - 15]^+$), which depend on the total number of carbons in the molecule, enable the homologue identification (called diagnostic homologue ions or fragments, DHF).¹⁶

Although commercial digital libraries are accurate and extensively used in identifying plant compounds, they can lead to incorrect identifications as false positives for these acyclic lipids.¹⁵ The identification process of these type of compounds faces two problems. The first one occurs when the DCF are much more abundant than the DHF (Figure 1a), as observed for alkanes, alkenes and aldehydes. For these compounds, during the automatic search using commercial digital libraries, the DHF are underestimated, resulting in the correct identification of the class but not the correct

*e-mail: lucas.p.roma@ib.usp.br

Associate Editor handled this article: Boniek G. Vaz

homologue. The second problem arises with compounds that contain two or more DHF for correct homologue identification, sometimes with short M^+ or $[M - 15]^+$ ions (i.e., ketones, esters and secondary alcohols) (Figure 1b). The multiple potential combinations between the DHF and lower abundances of M^+ or $[M - 15]^+$ hinder digital identification, leading to incorrect identification.

In this context, the software WaxAlly, registered at INPI (Instituto Nacional da Propriedade Industrial) under process number BR512022002375-0, was developed. to complement traditional methods for identifying common acyclic lipids found in cuticular waxes using mass spectrum observation and interacting with the software. Thus, the program aids in identifying acyclic homologues of alkanes, alkenes, aldehydes, ketones, esters, and trimethylsilyl (TMS)-derivatized free fatty acids, primary and secondary alcohols, with chains lengths ranging from 10 and 100 carbons. This is achieved through algorithms utilizing mass spectra DHF values obtained comparing the GC-MS mass spectra values with the software. The main goal of the software is to assist both new and experienced users in identifying a broad range of acyclic aliphatic lipids using mass spectra.

EXPERIMENTAL

The software was developed using the IDE (Integrated Development Environment) Qt Creator v.7.0.0 in C++ programming language,¹⁷ designed for use in Windows and is entirely in English. The Qt Creator is a simple and robust IDE for creating software, providing features that help users develop both the code and visual elements.

The WaxAlly comprises four main windows: (I) Home - containing basic information and main tutorials; (II) Calculator - allowing users to enter the DHF values into specific boxes within a sketch of a mass spectrum of a typical lipid class by the comparison between the mass spectrum of the sample and the sketch; (III) Prediction - allowing users enter a value to determine the carbon chain length of a lipid class to generate a mass spectrum sketch with the main DHF mass values; and (IV) Report - where all the identifications can be organized in a text file. The software interface was designed to be simple and accessible for both new and

experienced users for lipid identification using mass spectrometry.

The Home window displays the main information of the program, along with basic tutorials featuring step-by-step instructions on software operation, the appropriate method for inserting data, and how to utilize the software tools.

In the Calculator window, users can access each tab of lipid classes, visually compare the mass spectrum of their sample with the sketch of the program, and then identify which class the mass spectrum of the sample corresponds to. After the user chooses the correct Calculator tab, the next step is entering the specific DHF mass value into the corresponding box. If the mass values entered by the user in the boxes are correct, the software will identify the homologue and provide important information, such as the name, molecular mass, molecular formula, and the mass of some important DHF for the identification of the chosen class. Currently, this window is divided into seven main compound class tabs, including one called Fast. Each of the main tabs may contain sub-tabs. For example, the alcohol class tab (Alcohols) is subdivided into derivatized primary alcohols (Primary TMS), derivatized secondary alcohols (long-chain secondary TMS), derivatized isomeric secondary alcohols (isomeric secondary TMS) and secondary alcohols with derivatized methyl and ethyl ends (methyl secondary TMS and ethylic secondary TMS respectively; Figure 2). Each class tab or sub-tab contains a sketch of a mass spectrum corresponding to that class, with grey bars corresponding to the main DCF that characterize the mass spectrum of the selected class and black bars, with empty boxes on top, corresponding to the DHF needed to identify the homologue. These empty boxes should be filled in with proper mass values obtained observing and comparing the sample mass spectrum with the sketch. On the left side and above the mass spectrum sketch, there are seven auxiliary buttons to assist the user with complementary information for the correct identification of the analyzed sample mass spectrum (Table 1).

The Report window, designed to save the identified compound class and name, is structured into three sections (Figure 3). The first section (Report Assistant, Figure 3, green area) allows the construction of a basic report including identified compounds, chemical class, and sample name. The second section, called Report Text, consists of a text editing area where the previous report can

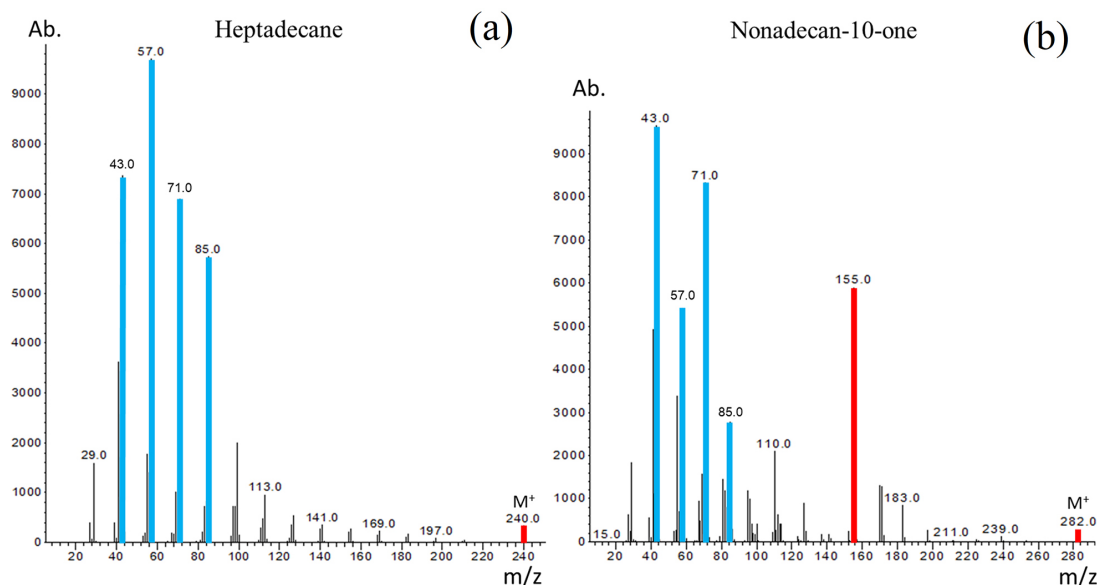


Figure 1. Examples of different types of mass spectra found in aliphatic acyclic lipids of waxes. The mass spectra of heptadecane (alkane) (a) and nonadecan-10-one (ketone) (b) exhibit less abundant or multiple DHF (diagnostic homologue ions or fragments) of the homologue (red bars) when compared to the DCF (diagnostic class ions or fragments) (blue bars)

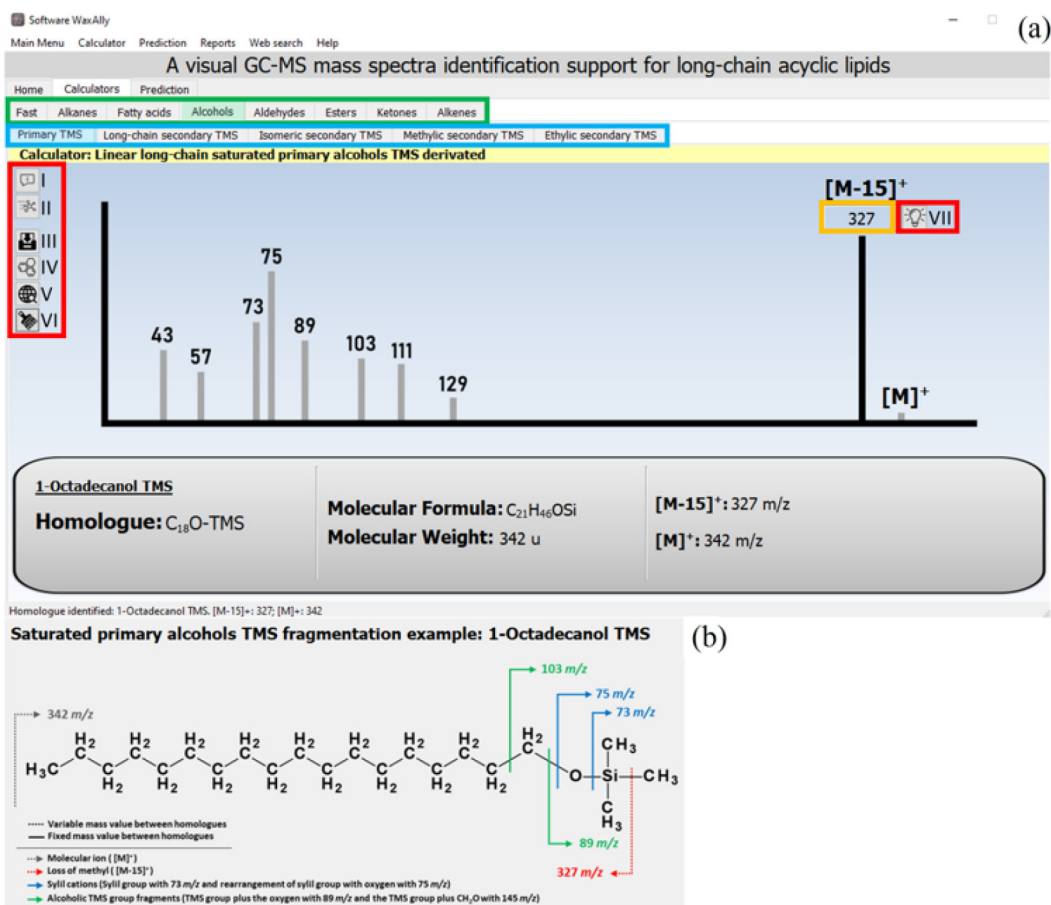


Figure 2. Calculator window exemplified by the Alcohols tab with Primary alcohols TMS sub-tab selected to illustrate the identification (a) and the fragmentation pattern of the selected compound class (b) using as example the homologue 1-octadecanol TMS derivative. In (a), the Class tabs and Fast tab are highlighted in green, and the specific sub-tabs of each class are in blue. The designated DHF mass value input box is highlighted in orange, with the value of $m/z = 327$, and the auxiliary buttons, described in Table 1, are highlighted in red

Table 1. Auxiliary buttons of the Calculator function. Position numbers correspond to Figure 3a

Icon	Position	Name	Function
	I	How to use	Basic information about the fragmentation of the selected class and some tips on how the selected calculator works
	I	Fragmentation example	An example of the fragmentation pattern of the selected calculator
	III	Open WaxAlly Report	Open the Report Assitant window
	IV	Molecule structure	A representation of the identified molecule
	VI	Web Search	Open Internet Search window
	VI	Clear	Clear identification results
	VII	DHF tips	Show some tips to fill in the DHF inputs

DHF: diagnostic homolog ions or fragments.

be saved or loaded as a text document (.txt; Figure 3, blue area). The last section, the Internet Search window, enables the user to carry out an online search using compound names or molecular formula using two databases, NIST Webbook (<https://webbook.nist.gov/>) and PubChem (<https://pubchem.ncbi.nlm.nih.gov/>; Figure 3, yellow area).

The design of the Prediction window (Figure 4) resembles that of its Calculator counterpart and includes nearly all the elements and tools found in the latter. This window allows the user, after selecting a class, to enter the carbon number of the DHF in order to obtain information about the homologue, including the mass of important fragments and the molecular formula.

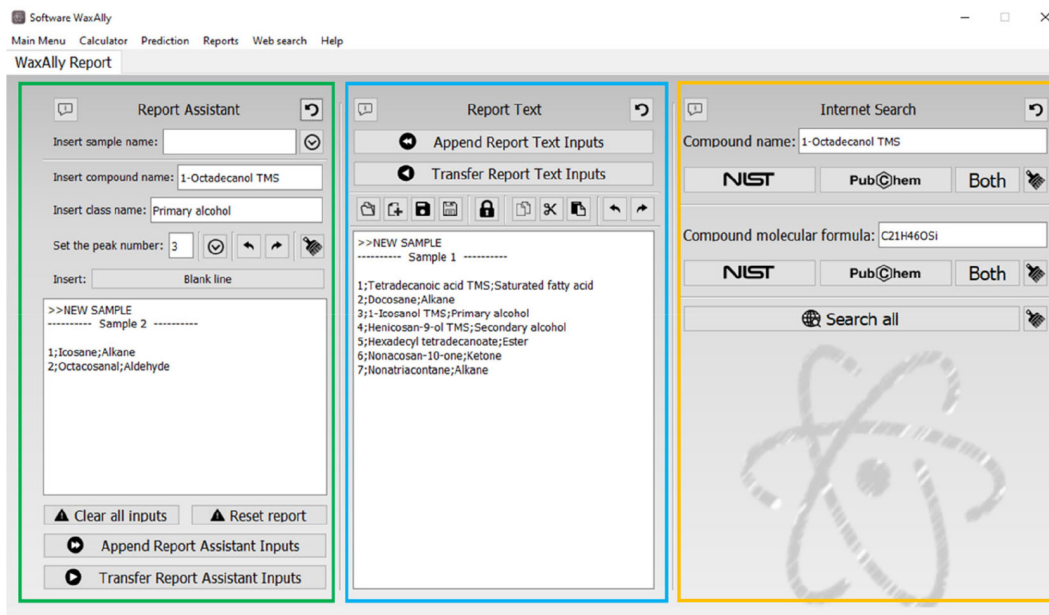


Figure 3. Report window screenshot. Outlined in green is the region of the window where recently identified compounds can be added to the current report. In blue is the area where the text can be edited and the reports saved or loaded in .txt format. In orange is the section where searches can be conducted on the NIST and PubChem sites using both the compound name and the molecular formula

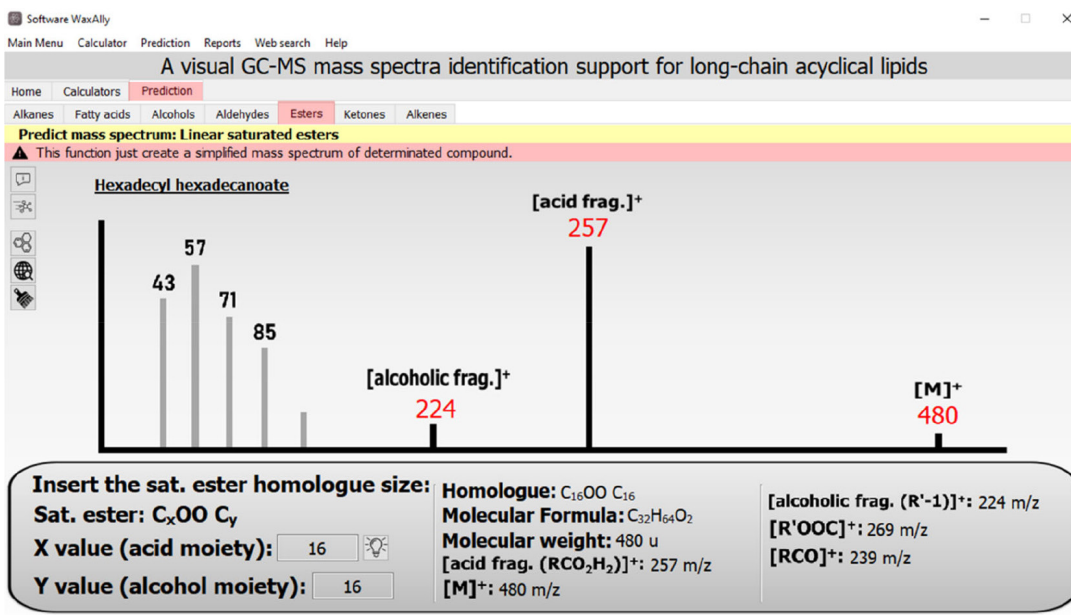


Figure 4. Prediction window exemplifying the prediction function by generating a sketch of a mass spectrum of an ester. In this example, the mass spectrum of hexadecyl hexadecanoate, an ester containing 16 carbons in its acid and alcohol parts, is displayed. Information of diagnostic ion masses, as well as the molecular formula and molecular weight of the compound, is shown

RESULTS AND DISCUSSION

Software workflow

The software provides users with several information to assist the identification process of acyclic lipids through the observation of the sample mass spectra and the interaction with the program. To start a simple analysis, the users can follow these basic steps:

- (1) Choose a mass spectrum from a homologue to identify (Figure 5a);
- (2) If the user has already identified the class by the DCFs, choose the corresponding Calculator tab (Figure 5b). If not, the user can explore some support tools available on each Calculator tab (☐,

- ☒ and ☒; Table 1) to help recognize the lipid class based on DCF and then choose the correct tab (Figure 5c);
- (3) Fill the input boxes above the DHF with the corresponding mass value obtained from the sample (Figure 5d) and run the calculator. With the correct DHF value, the software will identify the homologue, and important information will be displayed below the mass spectrum sketch (Figure 5e).
- (4) After the homologue identification, the user can begin creating a basic report by clicking on ☒ (Open WaxAlly Report tab; Figure 5f).
- (5) In the Report Assistant section of the Report window (Figure 5f), press ☒ to add the identified compound to the text area.
- (6) The user can proceed steps 1-5 for all desired homologues.

- (7) After the identification of all sample compounds, the results can be transferred to the Report Text section by pressing “Transfer Report Assistant Inputs” and saved in a .txt file (Figure 5g). More information about the identified compounds can be accessed on the NIST and PubChem websites, which are quickly accessible through the Internet Search section of the Report window (Figure 5h).

Calculator and prediction algorithms

Identifying each homologue involves transforming DHF into mathematical equations, which are based on isolating their CH₂ units removing the mass values that do not repeat along the chain of each fragment (for example, the CH₃ units of the molecular poles or carbons linked to organic functions). Subsequently, the mass corresponding to this set of CH₂ units is calculated, the number of carbon atoms in the fragment is determined and, consequently, the corresponding homologue is identified.

A simple example is the calculation used to identify primary alcohols TMS-derivatized (Figure 6). The main DCFs of this class are $m/z = 73, 75, 89, 103, 111, \text{ and } 129$, which appear in mass spectra of all homologues of saturated primary alcohols TMS derivatives. For this class, the main DHF corresponds to $[M - 15]^+$, which is abundant and easily distinguishable.¹⁸ This fragment consists of a CH₃ terminal with a mass of 15u, CH₂ units with a mass of 14u which repeat n times, and an end OSi(CH₃)₂ with total mass of 74u, corresponding to the organic function with the silyl derivative group less one methyl. After adding the DHF mass value, the software calculates the number of CH₂ repetitive units of the molecule subtracting the mass of the two poles (CH₃ = 15u and OSi(CH₃)₂ = 74u) from the $[M - 15]^+$ and dividing the remaining mass by 14, which corresponds to the CH₂ repetitive units of the molecule. To obtain the total carbon number of the corresponding primary alcohol, the next step is to add the carbon from the CH₃ pole that was subtracted before. In other words, one carbon should be added to the number of CH₂ repetitive units. Considering the 1-octadecanol TMS, the DHF is $[M - 15]^+ = 327$.

Subtracting the mass values corresponding to the two poles leaves 238, which corresponds to 17 CH₂ units. Adding the pole carbon, the primary alcohol chain contains 18 carbons, corresponding to 1-octadecanol TMS.

The algorithms created for all classes of acyclic lipids were constructed based on two premises: (i) establishing an interval of mass values that the user can input into the DHF boxes, controlling the maximum and minimum size of each homologue to guarantee a reliable result; (ii) ensuring that the mass value of the total CH₂ repetitive units is multiple of 14, allowing a trustworthy chain length. It is important to note that the identifications produced by the program should, whenever possible, receive a second confirmation by obtaining the compound mass spectrum from an article or site, as well as compare retention time with internal standards.

The algorithms for the Prediction function are diametrically opposite to those used for the calculators. Here, the number of carbons that determine chain size is used to generate all the mass values of the DHF.

CONCLUSIONS

The necessity of analyzing lipids extends to many different study areas, ranging from cellular membranes to archeology, including the examination of cuticular waxes, a group of compounds extremely important to plant survival in the terrestrial environment. In this technical note, we present the software we created and emphasize its proven effectiveness in identifying the components of plant cuticular waxes, whose classes typically consist of compounds following a homologous series, as well as other types of samples rich in unbranched acyclic long-chain lipids. This effectiveness primarily stems from the directly interaction of the user with the software to identify the compounds, differing from other traditional methods, for instance the digital libraries. Additionally, the didactic nature of the software in the mass spectrometry of different lipid classes make it a promising tool for both new and experienced researchers, even

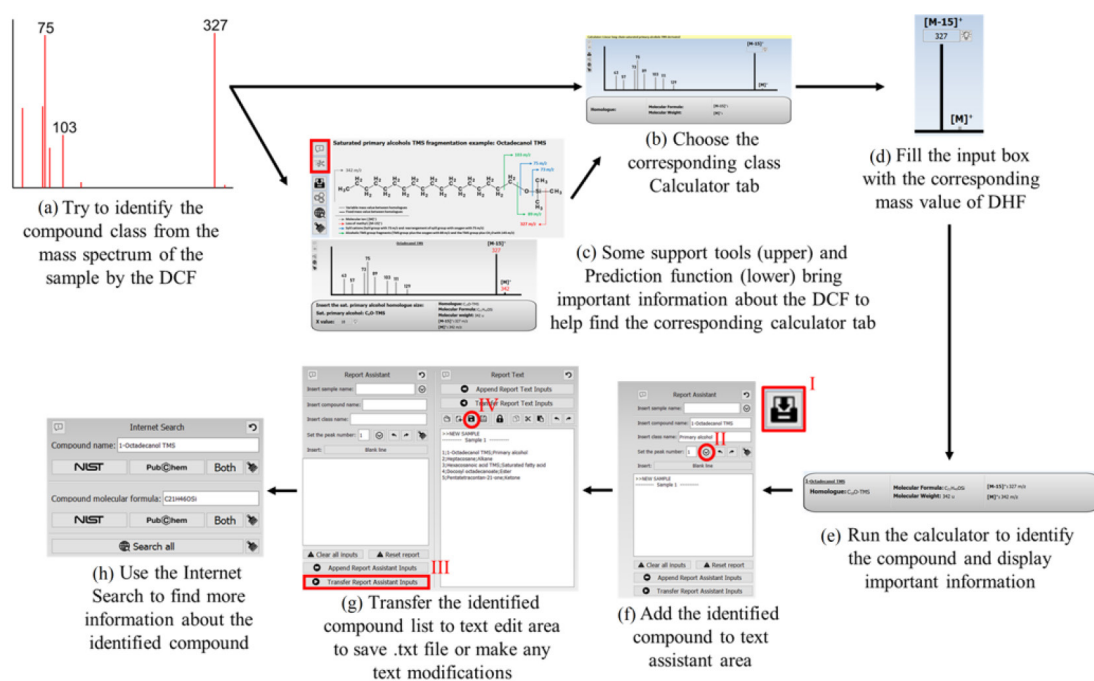


Figure 5. Diagram of a suggested workflow to carry out the identification and organization of long-chain acyclic lipids using WaxAlly software. The red squares in (c) indicate the buttons to open the auxiliary windows of each calculator; in (f), the buttons to open the Report Window (I) and save the compound identification temporarily in the Report Assistant section (II); and in (g), the button to transfer the temporary identifications of Report Assistant section to the Report Text section (III) and the button to save the report to a .txt file (IV)

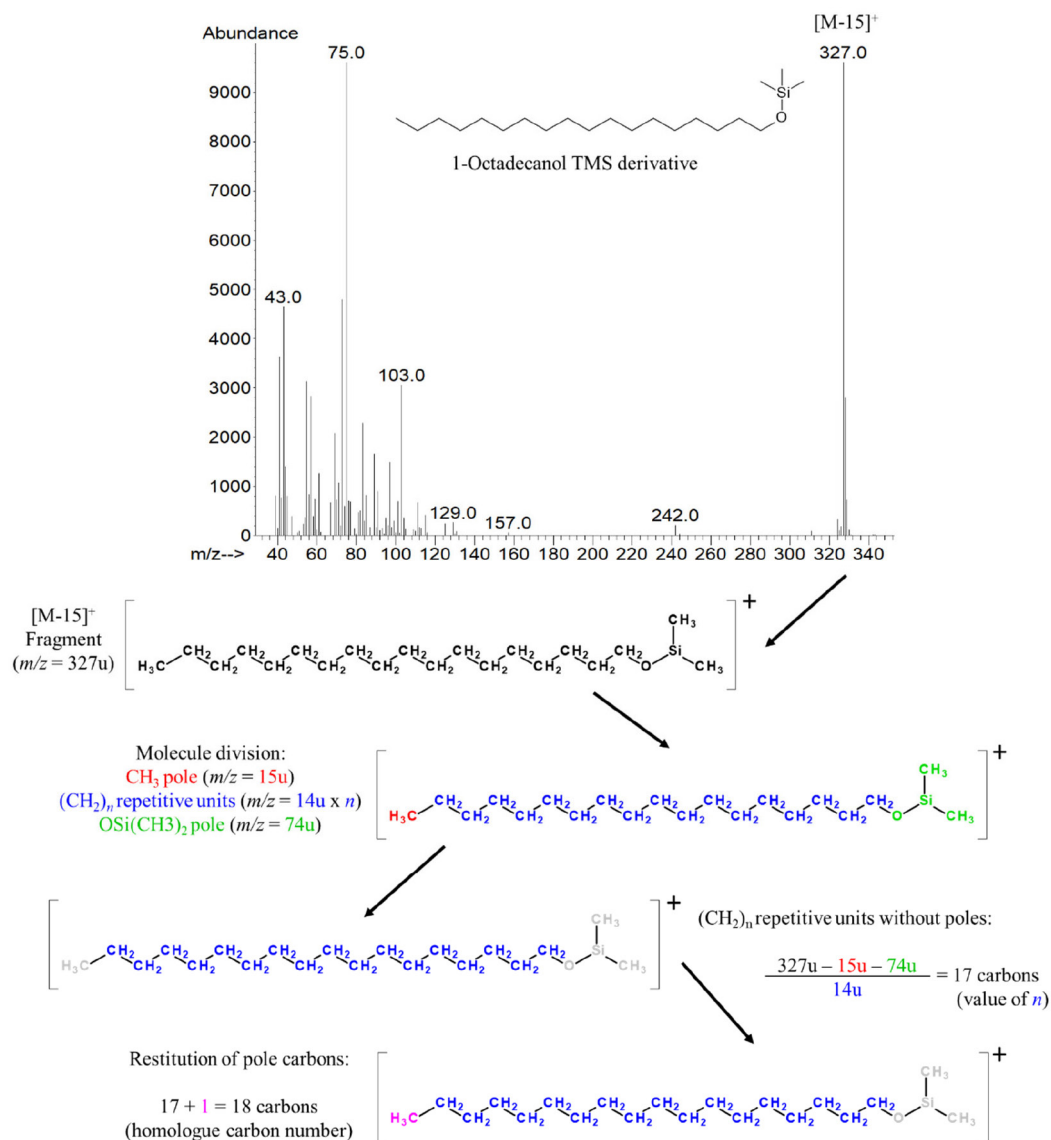


Figure 6. Flowchart of each calculation steps involved in Calculator function exemplified by 1-octadecanol TMS. The Prediction function works by reversing these algorithms to construct the fragments

those whose basic training does not include mass spectrometry. In the future, due to the mathematical premise of how the software was constructed, several new calculators (identification and prediction) can be added as needed. This is because any class of compounds with a fragmentation pattern in MS can be reduced and transformed into a mathematical algorithm, enabling its inclusion in the software.

The acquisition of the software can be made through direct contact with the authors (waxallysoftware@gmail.com) or with Agência USP de Inovação (AUSPIN; transtec@usp.br), which holds the software registration rights.

ACKNOWLEDGMENTS

The authors would like to thank the National Council for Scientific and Technological Development (CNPq, process numbers: 140085/2019-0 and 311543/2021-9) for the scientific research grants. Coordination for the Improvement of Higher Education Personnel (CAPES) for funding this study (funding code 001). Agência USP de Inovação (AUSPIN) for assisting in the registration process of the software (BR512022002375-0).

AUTHORS CONTRIBUTION

L. P. R. and D. Y. A. C. S. designed the study and wrote the article. L. P. R. built the software.

REFERENCES

- Levental, I.; Lyman, E.; *Nat. Rev. Mol. Cell Biol.* **2023**, *24*, 107. [Crossref]
- Asadollahi, E.; Trevisiol, A.; Saab, A. S.; Looser, Z. J.; Dibaj, P.; Kusch, K.; Ruhwedel, T.; Möbius, W.; Jahn, O.; Baes, M.; Weber, B.; Abel, E. D.; Balabio, A.; Popko, B.; Kassmann, C. M.; Ehrenreich, H.; Hirrlinger, J.; Nave, K.-A.; *bioRxiv* 2022. [Crossref]
- Liu, J.; Kang, R.; Tang, D.; *FEBS J.* **2022**, *289*, 7038. [Crossref]
- García-Coronado, H.; Tafolla-Arellano, J. C.; Hernández-Oñate, M. Á.; Burgara-Estrella, A. J.; Robles-Parra, J. M.; Tiznado-Hernández, M. E.; *Plants* **2022**, *11*, 1133. [Crossref]
- García-Granero, J. J.; Suryanarayan, A.; Cubas, M.; Craig, O. E.; Cárdenas, M.; Ajithprasad, P.; Madella, M.; *Front. Ecol. Evol.* **2022**, *10*, 840199. [Crossref]
- Parthasarathy, S.; Soundararajan, P.; Krishnan, N.; Karuppiah, K. M.;

- Devadasan, V.; Prabhu, D.; Rajamanikandan, S.; Velusamy, P.; Gopinath, S. C. B.; Raman, P.; *Biomass Convers. Biorefin.* **2023**, *13*, 15543. [Crossref]
7. Müller, C.; Riederer, M.; *J. Chem. Ecol.* **2005**, *31*, 2621. [Crossref]
8. Bernard, A.; Joubès, J.; *Prog. Lipid Res.* **2013**, *52*, 110. [Crossref]
9. Cardona, J. B.; Grover, S.; Busta, L.; Sattler, S. E.; Louis, J.; *Planta* **2023**, *257*, 22. [Crossref]
10. Zhao, Z.; Zhao, J.; Peng, C.; Duan, X.; Deng, M.; Wen, J.; *Sci. Hortic.* **2023**, *311*, 111805. [Crossref]
11. Grayson, M. A. In *The Encyclopedia of Mass Spectrometry, vol. 9, Part A*; Gross, M. L.; Caprioli, R. M., eds.; Elsevier: Amsterdam, 2016, ch. 4. [Crossref]
12. Halket, J. M.; Zaikin, V. G.; *Eur. J. Mass Spectrom.* **2004**, *10*, 1. [Crossref]
13. Subramaniam, S.; Fahy, E.; Gupta, S.; Sud, M.; Byrnes, R. W.; Cotter, D.; Dinasarapu, A. R.; Maurya, M. R.; *Chem. Rev.* **2011**, *111*, 6452. [Crossref]
14. Heller, S.; *Today's Chemist at Work* **1999**, *8*, 45. [Link] accessed in August 2024
15. Stein, S.; *Anal. Chem.* **2012**, *84*, 7274. [Crossref]
16. Silverstein, M. R.; Webster, F. X.; Kiemle, D. J.; *Spectrometric Identification of Organic Compounds*, 7th ed.; John Wiley and Son: New Jersey, 2005.
17. Rischpater, R.; *Application Development with Qt Creator*; Packt Publishing Ltd.: Birmingham, 2014.
18. Taghizadeh, T.; *Produtos Naturais no Controle de Insetos*; EdUFSCar: São Carlos, 2001.