

SMoIESTY_platform

USER'S GUIDE

Dr. Panteleimon G. Takis^{1,2}

¹National Phenome Centre and Imperial Clinical Phenotyping Centre, Department of Metabolism, Digestion and Reproduction, IRDB Building, Imperial College London, Hammersmith Campus, London, W12 0NN, United Kingdom

²Section of Bioanalytical Chemistry, Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, South Kensington Campus, London, SW7 2AZ, United Kingdom

Table of Contents	Pages
1. INTRO - ABOUT THE SMOLESY_PLATFORM	2
2. PREPARING SPECTRA FOLDER	3
3. PREPARING OUTPUT FOLDER	4
4. INSTALLING/RUNNING SMOLESY_PLATFORM.....	4
I) INSTALLING/RUNNING SMOLESY_PLATFORM BY MATLAB.....	4
II) INSTALLING/RUNNING SMOLESY_PLATFORM BY THE COMPILED FILE FOR MACOS.....	5
III) INSTALLING/RUNNING SMOLESY_PLATFORM BY THE COMPILED FILE FOR WINDOWS.....	6
5. LOADING ^1H -NMR SPECTRA.....	7
→NOTE 1	7
6. DEFINE OUTPUT FOLDER	8
7. CONVERTING ^1H -NMR IMAGINARY DATA INTO SMOLESY	9
→NOTE 1	9
8. EXPORTING SMOLESY DATA (OPTION: ALIGNED TO GLUCOSE)	10
→NOTE 1	10
OPTION: Align SMoESY data to Glucose.....	10
→NOTE 2	10
9. PLOTTING ^1H -NMR VS SMOLESY DATA (OPTIONS: SYNCHRONIZE PLOTS – USE CURSOR).....	11
OPTION_1: ACTIVATE SYNCHRONIZING ZOOM.....	11
OPTION_2: ACTIVATE DATA CURSOR	11
10. OPTIONAL FUNCTIONALITIES OF SMOLESY_PLATFORM.....	12
10.1 MANUAL PEAK PICKING.....	12
THE EXAMPLE OF ACETATE METABOLITE IN 5 PLASMA ^1H -NMR SPECTRA.....	13
→Step 1.....	13
→Step 2.....	13
→Step 3.....	13
→Step 4.....	14
→Step 5.....	15
THE EXAMPLE OF GLUCOSE METABOLITE IN 5 PLASMA ^1H -NMR SPECTRA.....	16
THE EXAMPLE OF A SPECTRAL REGION 0.9-1.25 PPM IN 5 PLASMA ^1H -NMR SPECTRA	17
10.2 SEMI-AUTOMATED PEAK PICKING	18
THE EXAMPLE OF L-ALANINE AND L-VALINE ^1H -NMR SIGNALS SEMI-AUTOMATED CALIBRATION	19
→Step 1.....	19
→Step 2.....	19
→Step 3.....	19
→Step 4.....	20
→Step 5.....	20
11. EXAMPLES-TEMPLATES.....	21
12. TECHNICAL REQUIREMENTS - SUPPORT.....	22
13. COPYRIGHT-LICENSE	23
14. REFERENCES.....	24

1. Intro - about the SMoESY_platform

The **SMoESY_plarform** is a free, user-interactive toolbox, created in MATLAB programming suite (2019b update 5, MathWorks). The purpose of this toolbox is to facilitate the implementation of the SMoESY (**S**mall **M**olecules **E**nhan**c**ement **S**pectroscop**Y**) method into NMR metabolomics pipeline and in general NMR analytical platforms^{1,2}. It provides the transformation of conventional ¹H-NMR spectra into SMoESY data and it incorporates various plotting functions-tools for the detailed visual examination of the data. Moreover, SMoESY_platform offers the opportunity of manual and/or semi-automated calibration/alignment of multiple SMoESY peaks from unlimited number of NMR spectra. Another feature of the toolbox is the direct integration of already highly resolved and baseline-free SMoESY signals of metabolites, aiming at their fast quantification by low computational cost integration. Except for the semi-automated targeted profiling, SMoESY_platform provides the option of SMoESY spectra-binning (i.e. bucketing) as well as automated variable shaped bucketing, suitable for untargeted NMR-based metabolomics (e.g. diseases fingerprinting) and multivariate statistical analyses (MVA). It is noteworthy that all results from the spectral analyses performed by the toolbox, are directly exported into user-friendly files format (i.e. tab-limited .txt files), which could be straightforwardly loaded into existing metabolomics planforms-toolboxes for further analyses. Finally, SMoESY_platform provides direct feedback/instructions to the user for each step of its computational processes, therefore is a user-friendly toolbox which can be easily operated by non-NMR experts.

2. Preparing spectra folder

1D- ^1H -NMR spectra (e.g. 1D-NOESY) should be in one folder as indicated in Fig. 1. Under each spectrum title (e.g. Spectrum 1, Spectrum 2 in Fig. 1), the experiment folder should be numerical (e.g. 1 in Fig. 1). Please **ensure that only 1 experiment folder exists**.

Please note that **no non-spectral folders should be inside the parent NMR data folder**.

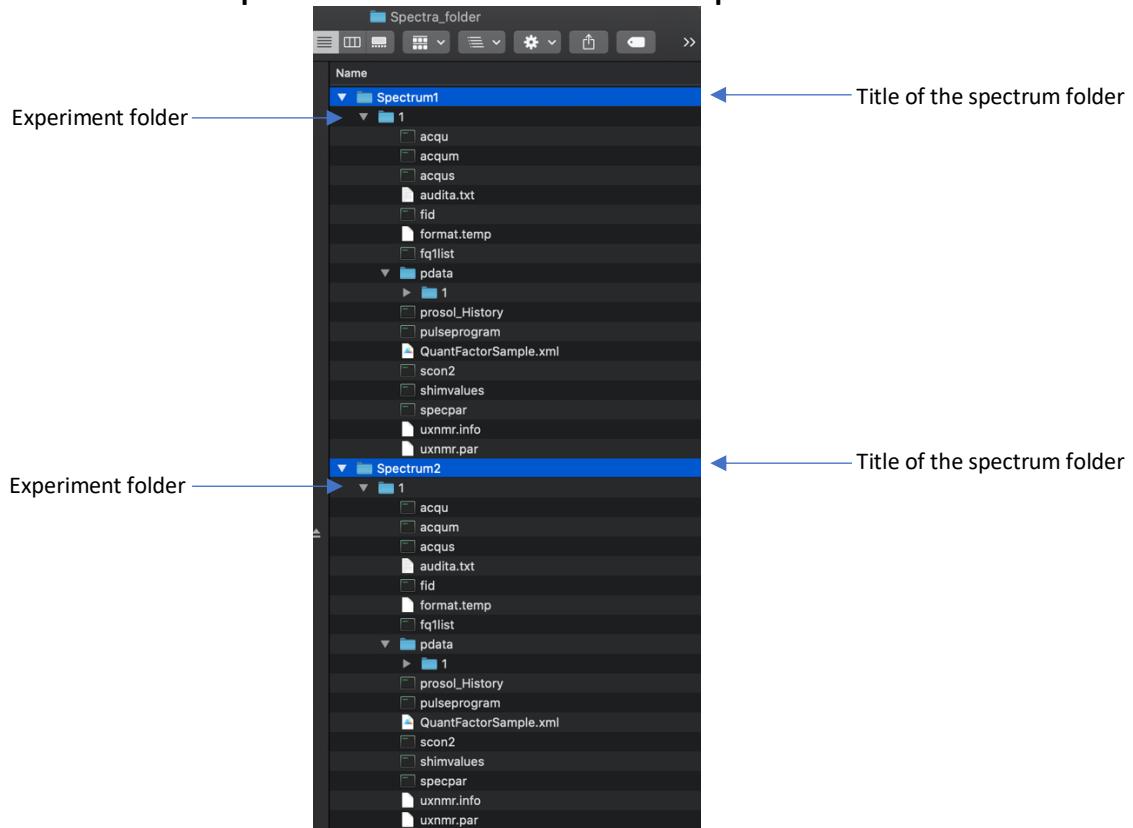


Figure 1. Spectra containing folder; In this example the spectra folder contains 2 1D-NOESY plasma spectra with the depicted structure.

3. Preparing output folder

Before running the GUI, the user should create a folder, where all results will be exported by the GUI.

4. Installing/Running SMoESY_platform

i) Installing/running SMoESY_platform by MATLAB

After downloading the toolbox folder from https://github.com/pantakis/SMoESY_platform, open the “Set Path” dialog window from MATLAB (Fig. 2):



Figure 2.

and in the popped-up window choose the option “Add with Subfolders...” (Fig. 3) and select the parent folder of the toolbox, which contains the MATLAB code of the GUI.

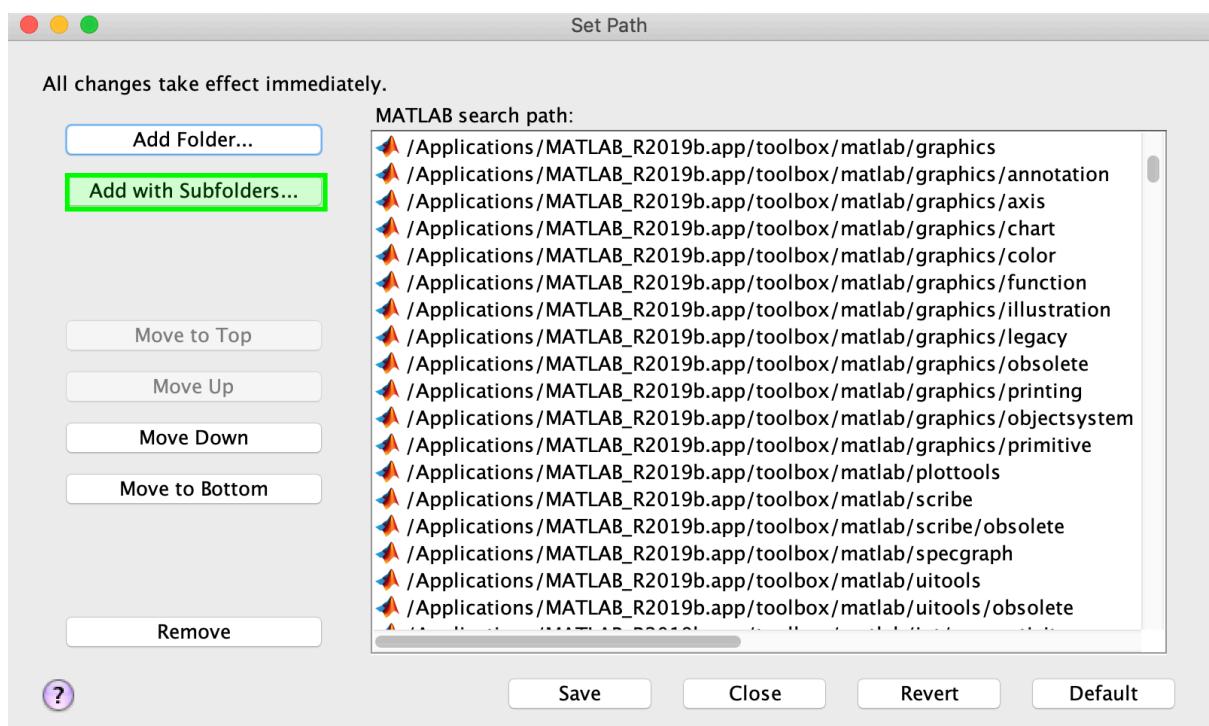


Figure 3.

Finally, from the MATLAB command window type: “SMoESY_platform” and press “Enter” so as to launch the GUI (Fig. 4).

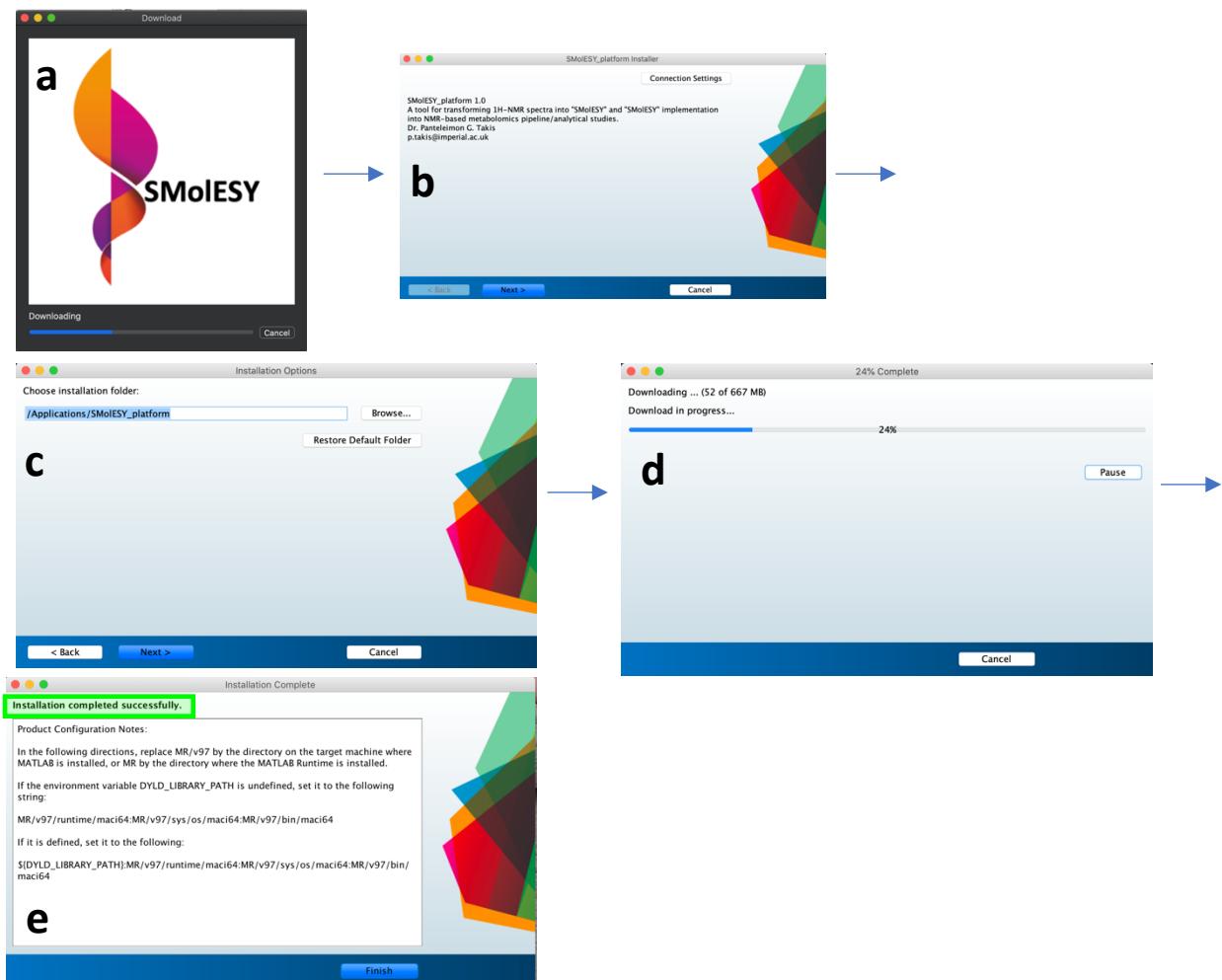


Figure 4.

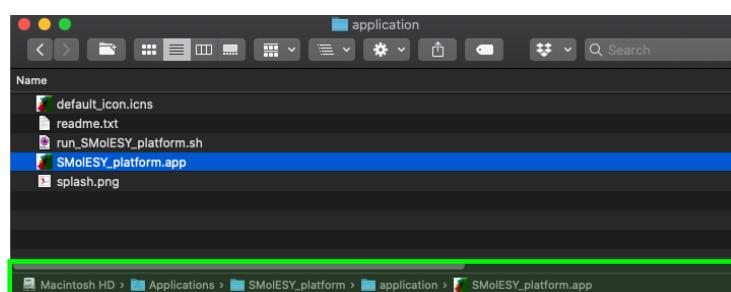
User's Guide SMoESY_platform
https://github.com/pantakis/SMoESY_platform

ii) Installing/running SMoESY_platform by the compiled file for MacOS

After downloading the toolbox folder from https://github.com/pantakis/SMoESY_platform, in the folder “Compiled_for_MAC_OS”, double click the file “SMoESY_platform_installer.app”. Then follow the on-screen instructions (internet connection is required) for downloading/installing the MATLAB RUNTIME modules in order to run the toolbox. The following dialog windows (a-e) should appear sequentially on screen:

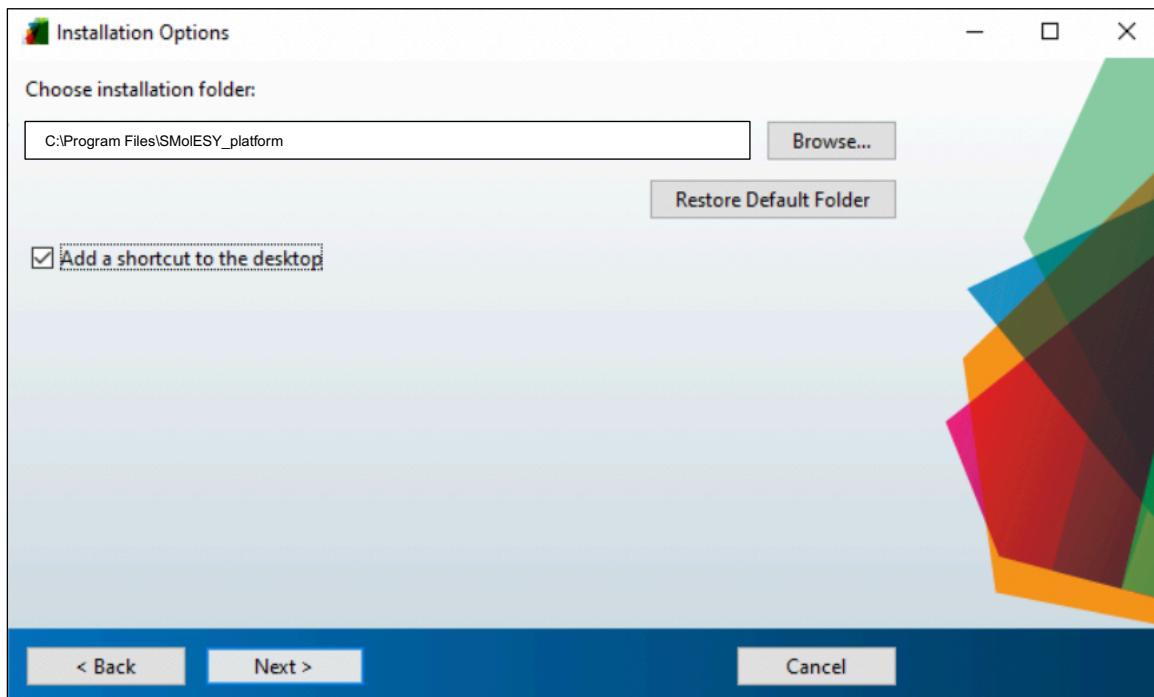


- In panel c, please ensure that the installation folder is: [/Applications/SMoESY_platform](#)
- In panel e, ignore the Product Configuration Notes, once the installation is completed successfully.
- To launch the application please double-click the “SMoESY_platform.app”, which should be located in the folder: “[..//Applications/SMoESY_platform/application](#)”, as you could see below:



iii) Installing/running SMoESY_platform by the compiled file for Windows

After downloading the toolbox folder from https://github.com/pantakis/SMoESY_platform, in the folder “Compiled_for_Windows”, double click the file “SMoESY_platform_installer.exe”. Then follow the on-screen instructions (internet connection is required) for downloading/installing the MATLAB RUNTIME modules in order to run the toolbox as indicated above (i.e. for Mac OS). Please ensure that you create a desktop shortcut of the toolbox in order to launch it as shown below:



5. Loading ^1H -NMR spectra

Pressing the button “”, a dialog window opens (Fig. 5) in order to define the parent folder containing all NMR spectra to be loaded into the SMoIESY_platform.

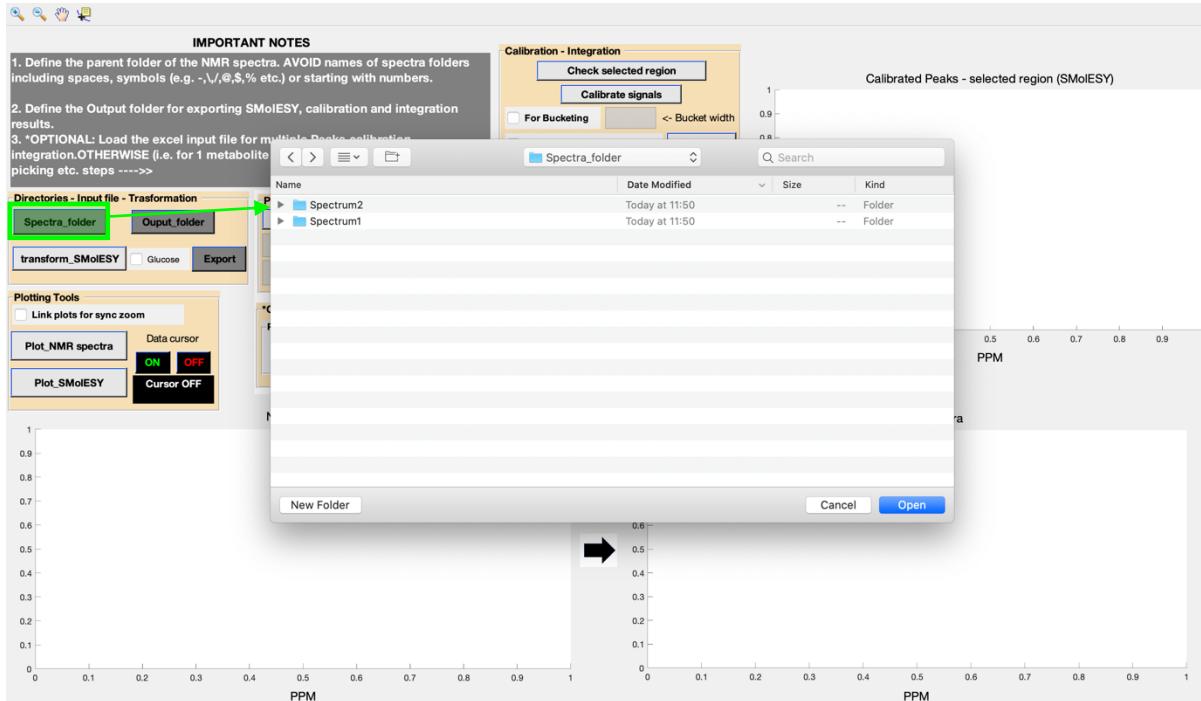


Figure 5.

→NOTE 1

The “IMPORTANT NOTES” notifications box, informs the user for any errors (Fig. 6a) or the successful (Fig. 6b) reading/loading of the NMR data by the toolbox.

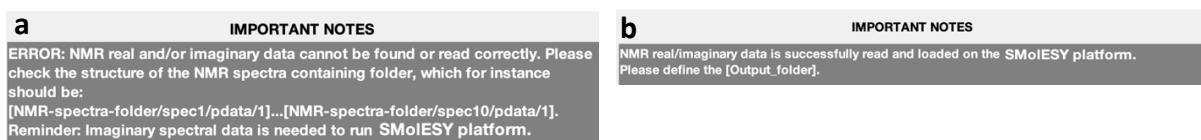


Figure 6.

6. Define output folder

Pressing the button “**Output_folder**”, a dialog window opens (Fig. 7) in order to define the output folder where all processing, SMoESY data etc. will be exported by the toolbox.

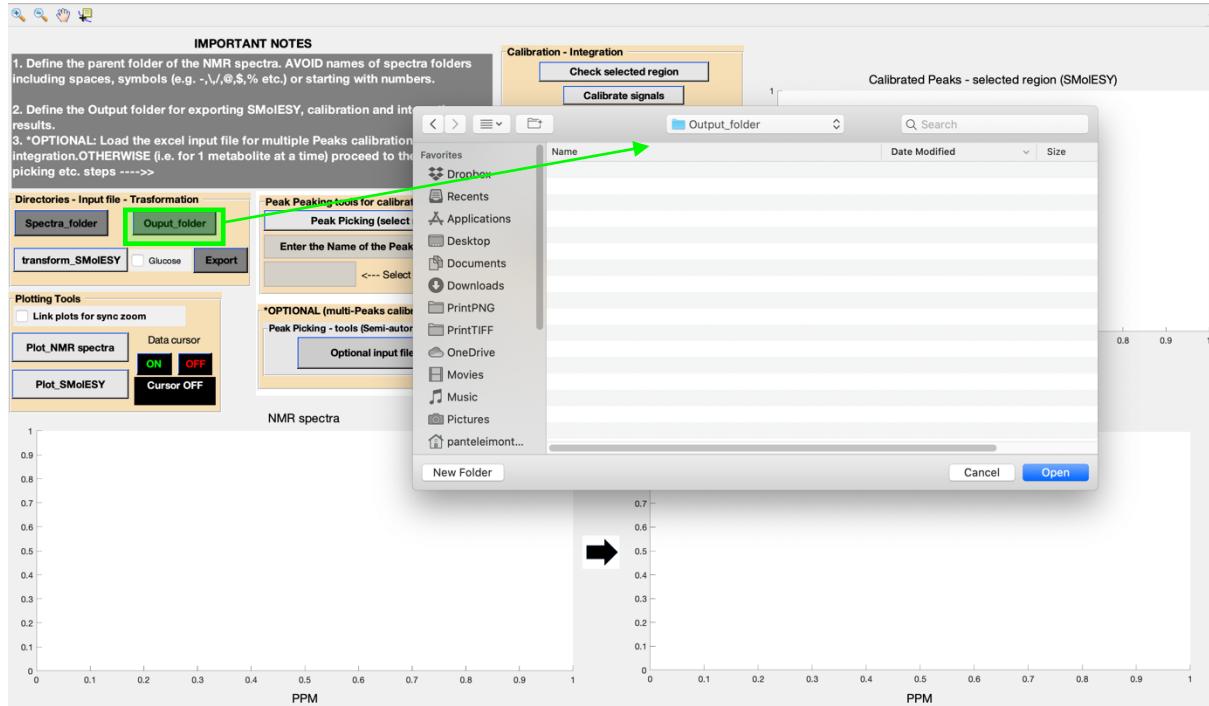


Figure 7.

7. Converting ^1H -NMR imaginary data into SMoESY

Pressing the button “**transform_SMoESY**” (Fig. 8), the imaginary spectral data of the loaded ^1H -NMR (e.g. 1D-NOESY) spectra will be transformed into the SMoESY.

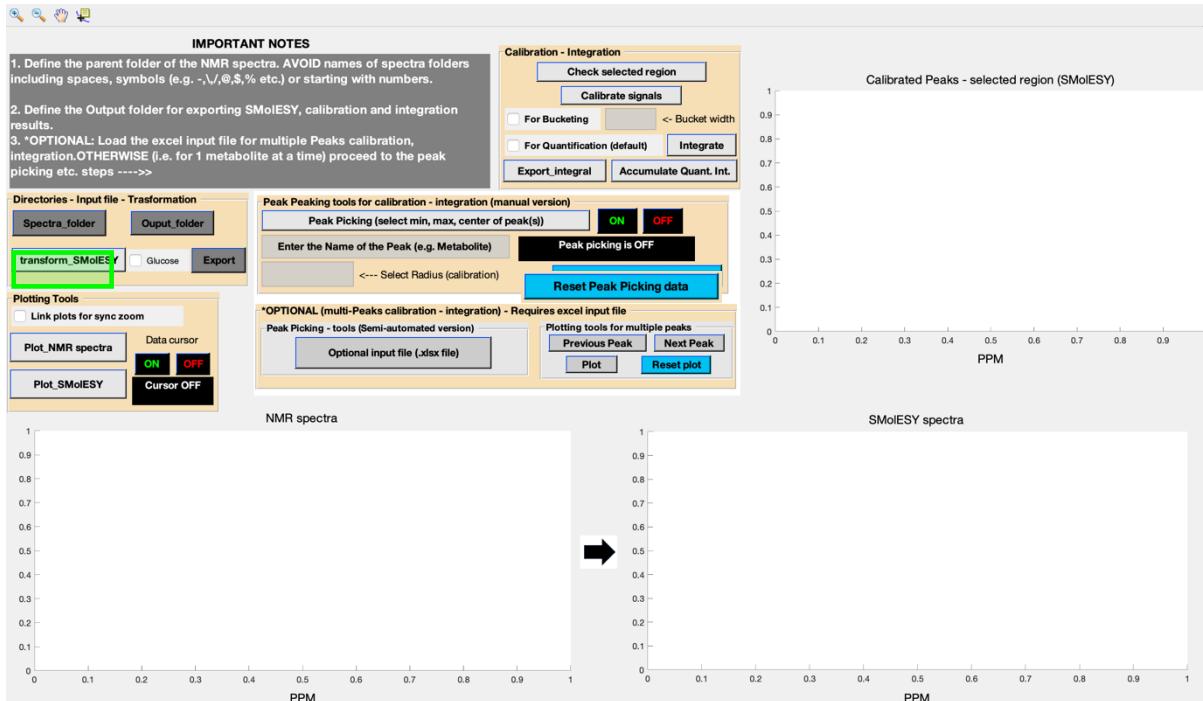


Figure 8.

→NOTE 1

The “IMPORTANT NOTES” notifications box, informs the user for any errors (Fig. 9a) or the successful (Fig. 9b) transformation of the ^1H -NMR data into SMoESY.



Figure 9.

8. Exporting SMoESY data (OPTION: aligned to Glucose)

Pressing the button “” (Fig. 10a), the SMoESY data is exported to the previously defined output folder as a tab_limited .txt file named as: “SMoESY_spectra.txt”.

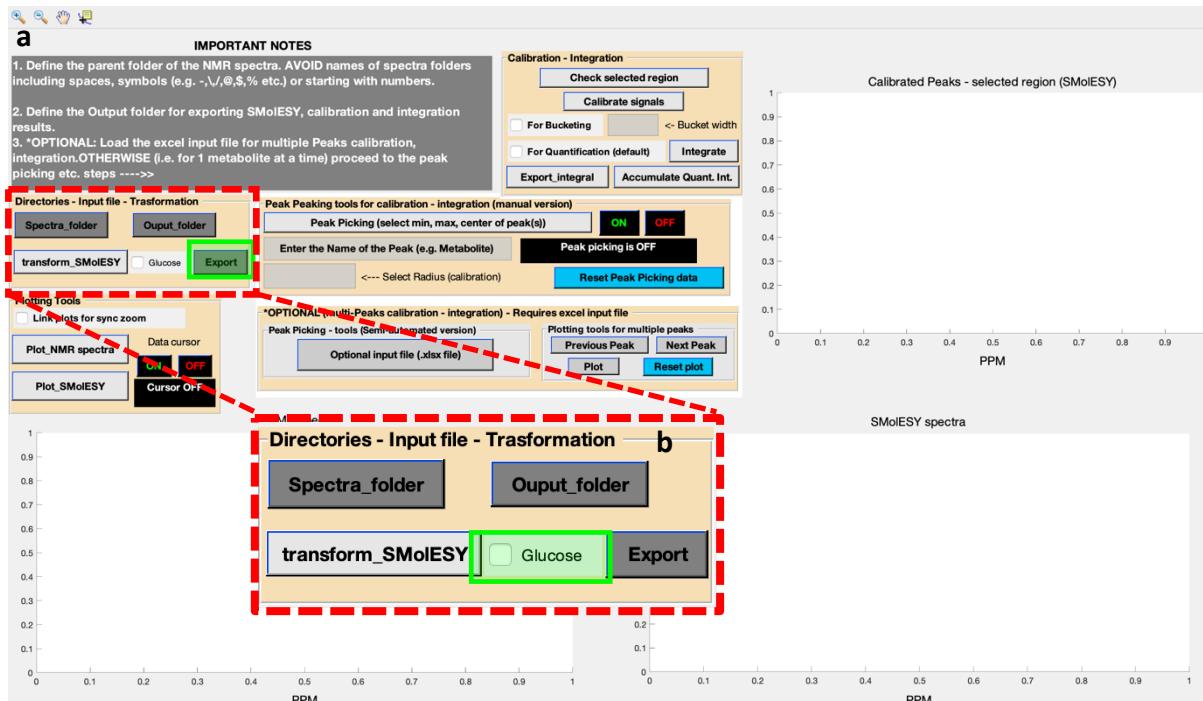


Figure 10.

→NOTE 1

The “IMPORTANT NOTES” notifications box, informs the user for any errors (Fig. 11a) or the successful (Fig. 11b) exportation of the SMoESY data.



Figure 11.

OPTION: Align SMoESY data to Glucose (Fig. 10b) (“”) doublet and export data to the output folder as a tab_limited .txt file named as: “SMoESY_ALIGNED_to_GLUCOSE_spectra.txt”.

→NOTE 2

i) The “IMPORTANT NOTES” notifications box, informs the user for any errors or the successful (Fig. 12) exportation of the aligned to Glucose SMoESY data.

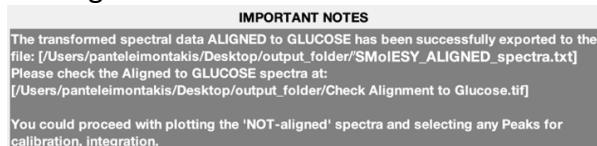


Figure 12.

ii) It is noted that after aligning to Glucose, user could export aligned to Glucose SMoESY data, however, SMoESY_platform employs only the initial (NOT-aligned) SMoESY data for any following analyses.

9. Plotting ^1H -NMR vs SMoESY data (OPTIONS: synchronize plots – use cursor)

Pressing the buttons “” and “” (Fig. 13a), NMR spectra and SMoESY are both plotted in two panels, respectively (Fig. 13a).

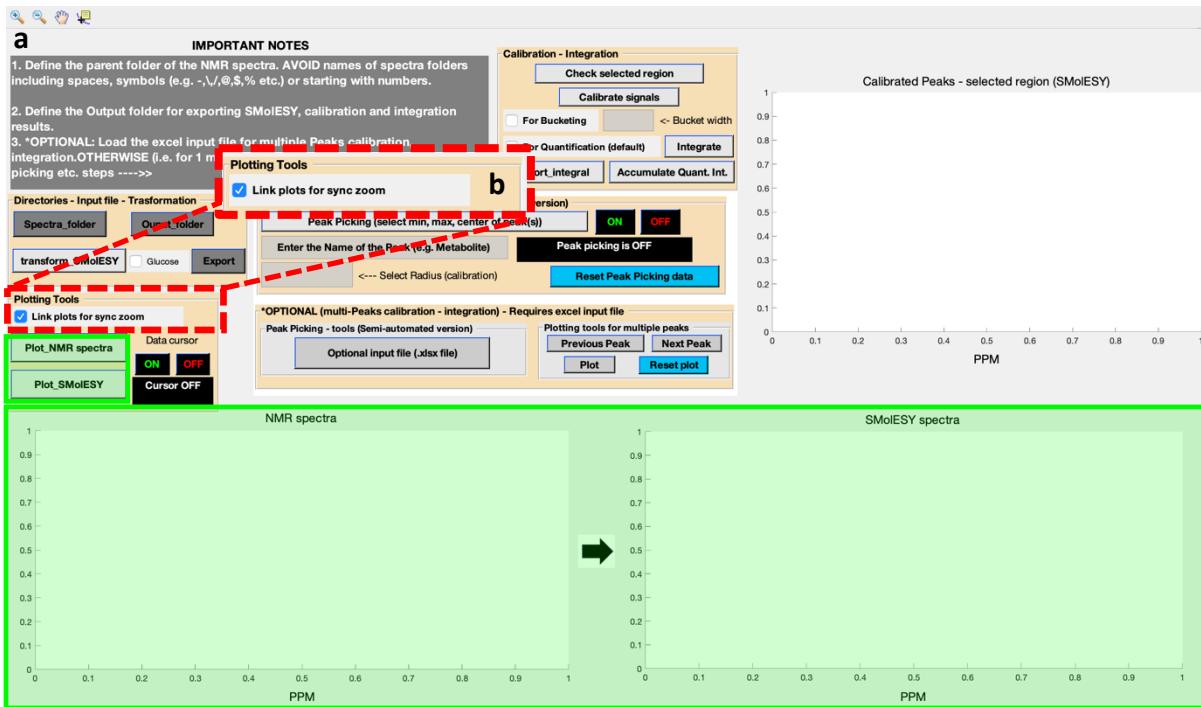


Figure 13.

OPTION_1: Activate synchronizing zoom by the checkbox “” (Fig. 13b) for () for both plots.

OPTION_2: Activate data cursor “” by pressing “ON” button, for extracting information of peak positions for both panels independently. Spectral features position is automatically printed in the legend of each panel (Fig. 14) by mouse-over in the spectral area of each panel.

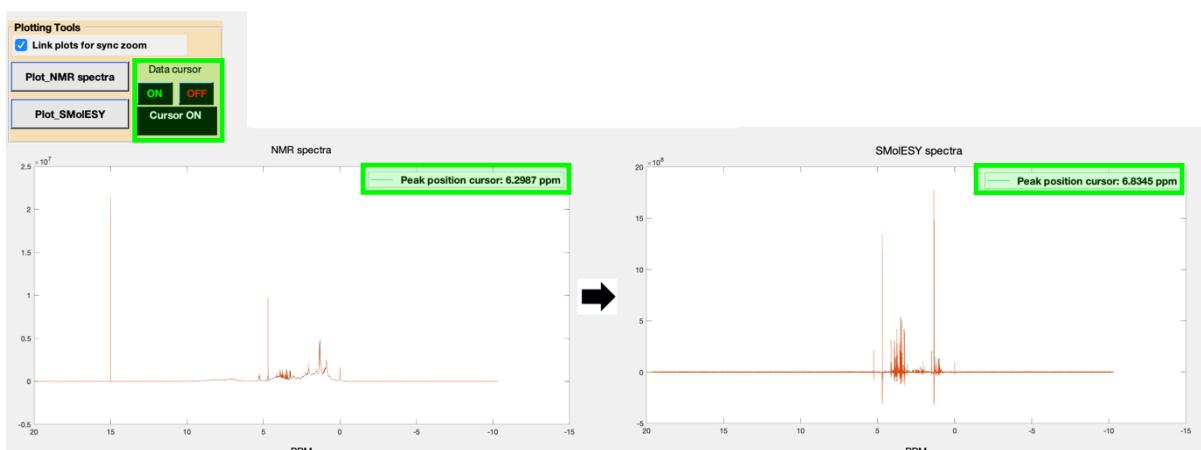


Figure 14.

10. Optional functionalities of SMoESY_platform

10.1 Manual peak picking in the SMoESY data – Calibration of a defined region from multiple spectra to a specific peak – Integration

Activating peak picking by pressing “” button and pressing “” button (Fig. 15a), the user is able to manually define 3 values in the spectral area in order to focus on a specific spectral region (Fig. 15b).

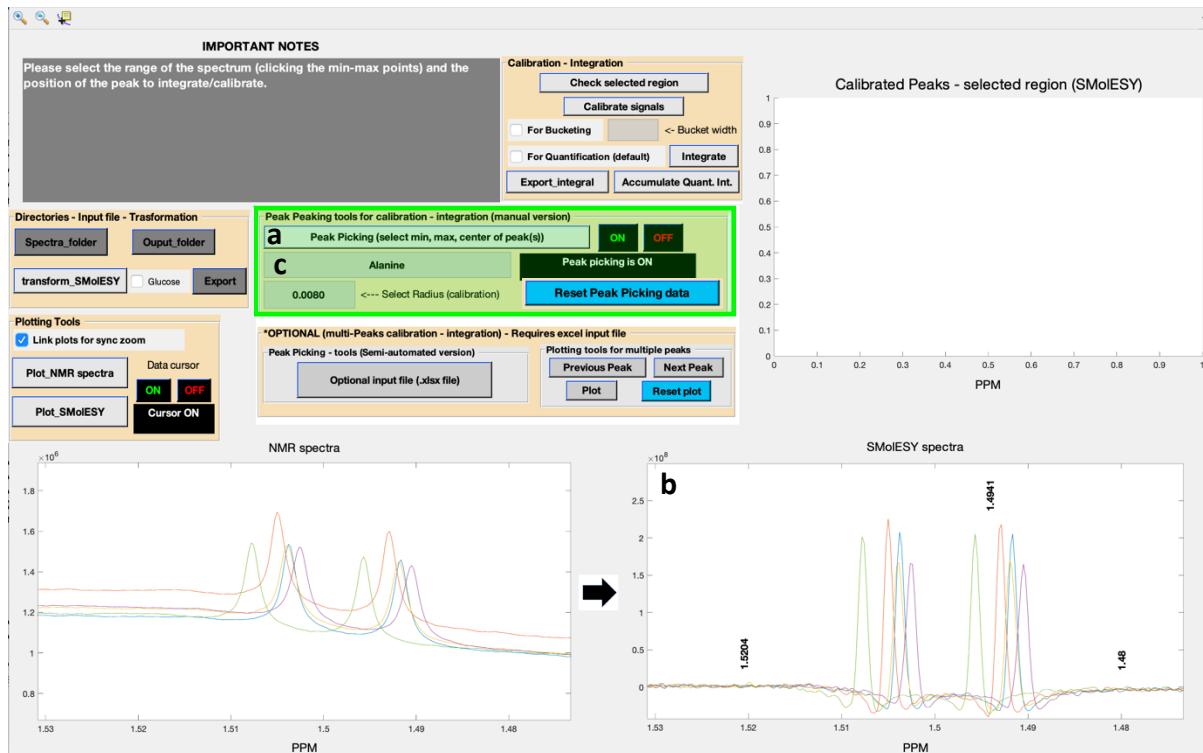


Figure 15.

In particular by clicking on the SMoESY spectra panel (Fig. 15b), user could select:

- i) 2 values to define the min-max ppm of the selected spectral region (see Fig. 15b, values: 1.5204 ppm, 1.48 ppm).
- ii) 1 Ref. value to define the reference ppm value where the selected spectral region will be aligned for all spectra (see Fig. 15b, value: 1.4911 ppm).

User could define any metabolite(s) (or compound(s)) enclosed in the selected region by inserting a unique word (i.e. valine-isoleucine) in the dialog box: “” (see Fig. 15c, name: Alanine).

A 4th value should be inserted in the dialog box: “” (see Fig. 15d, value: 0.0080). This value, termed as radius, drives the algorithm to search/select the spectral feature (i.e. peak) with the highest intensity in the spectral region for all spectra;

[Ref. value – Radius] < highest peak < [Ref. value + Radius],
and calibrates/aligns it at the defined reference value.

The example of acetate metabolite in 5 plasma ^1H -NMR spectra, explaining the step-by-step followed procedure is pointed below:

→ Step 1

Assigning the ^1H -NMR singlet of acetate -CH₃ group in the 5 plasma spectra see asterisks in Fig. 16)

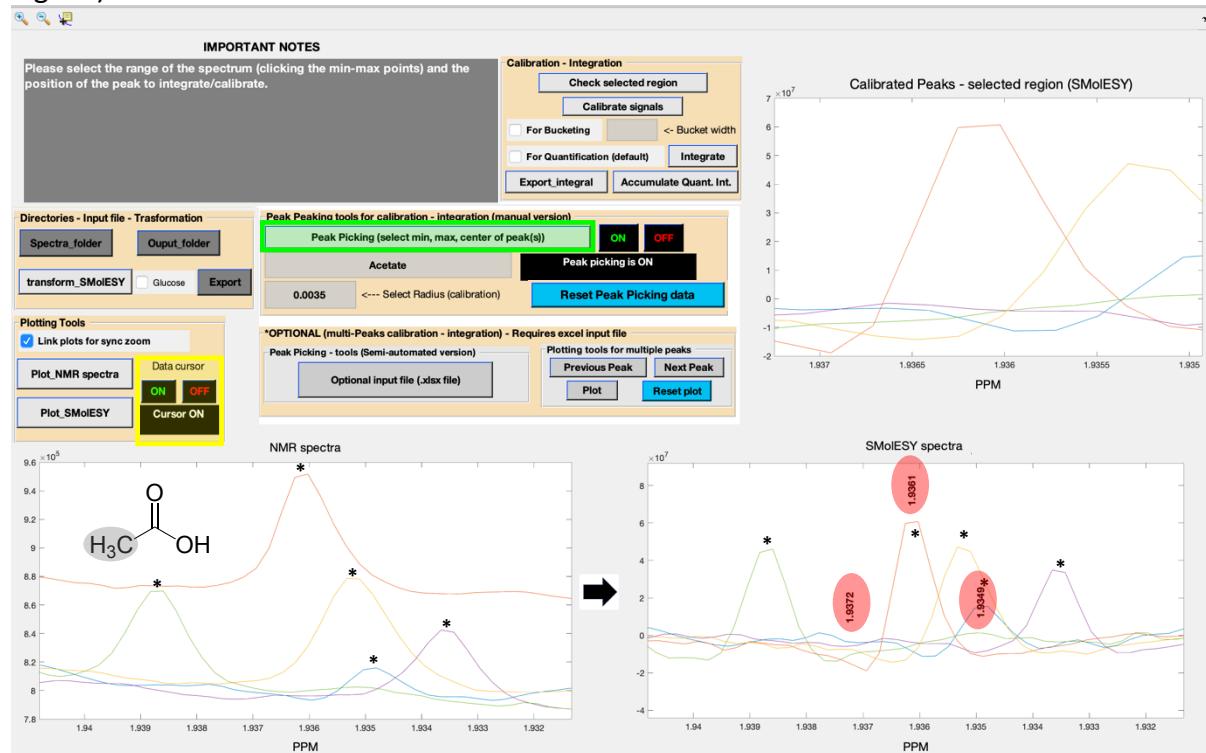


Figure 16.

→ Step 2

3 values (Fig. 16 red circles) manually defined by firstly clicking the activated peak picking function (Fig. 16, green highlighted) and afterwards clicking on the spectral area of the SMoESY panel (**TIP:** the plot zoom functions must be deselected and you need to click above/below the lines in the plot). The min and max values define the spectral region where all 5 singlets of acetate will be aligned according to the reference peak pointed by the 3rd value (i.e. 1.9361 ppm).

→ Step 3

To define radius, user could activate “ ” (Fig. 16 yellow highlighted) and pass the mouse over the peaks of acetate in order to see the positions of the most remotest signals of acetate from the reference value (i.e. 1.9361 ppm), and each ppm point of the spectral window will appear at the legend (Fig. 17 green highlighted) of the SMoESY spectra panel.

User's Guide SMoESY_platform
https://github.com/pantakis/SMoESY_platform

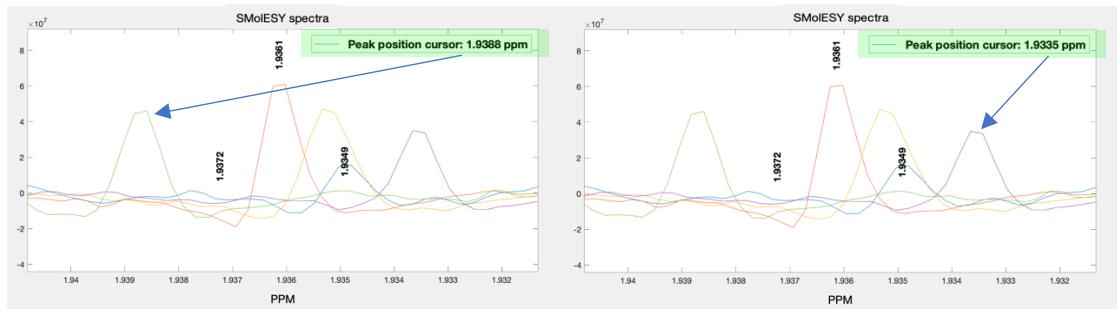


Figure 17.

In this example, the most remoted signals of acetate resonate at 1.9388 and 1.9335 ppm (Fig. 17 green highlighted) and their distances from the reference value (i.e. 1.9361 ppm) are 0.0027 and 0.0026 ppm, respectively.

Consequently, the radius should be at least 0.0027 ppm. For this case, we put as radius the value of 0.0035 ppm, since no any other signal from other metabolites is present at high intensity. In more crowded spectral regions, the radius should be as accurate as possible.

→ Step 4

After defining all parameters for the calibration of acetate signals from the 5 spectra, user could see the results of the defined spectral region for calibration in the 3rd panel of the GUI, “Calibrated peaks – selected region” (Fig. 18a), by pressing the “Check selected region” button (see Fig. 18 green highlighted).

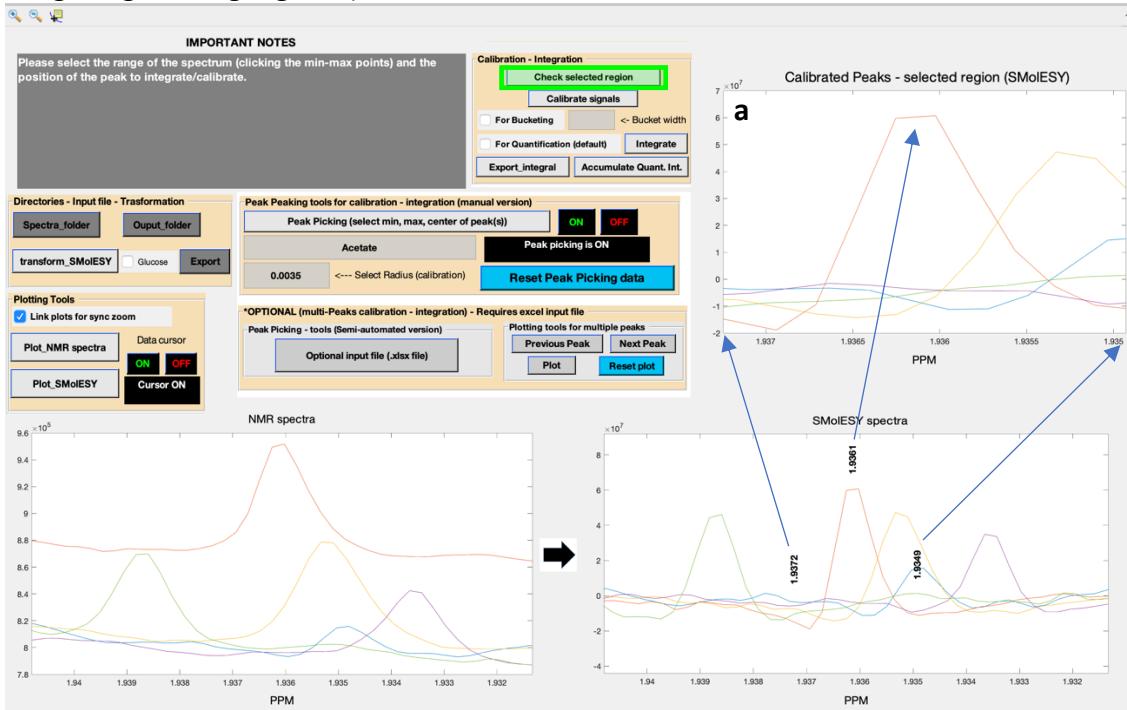


Figure 18.

If the user is satisfied with the selected region, the calibration of the signals will be completed by pressing the “Calibrate signals” button (Fig. 19 highlighted in green). Otherwise, user could redo the process after pressing “Reset Peak Picking data” button (Fig. 19 highlighted in red box).

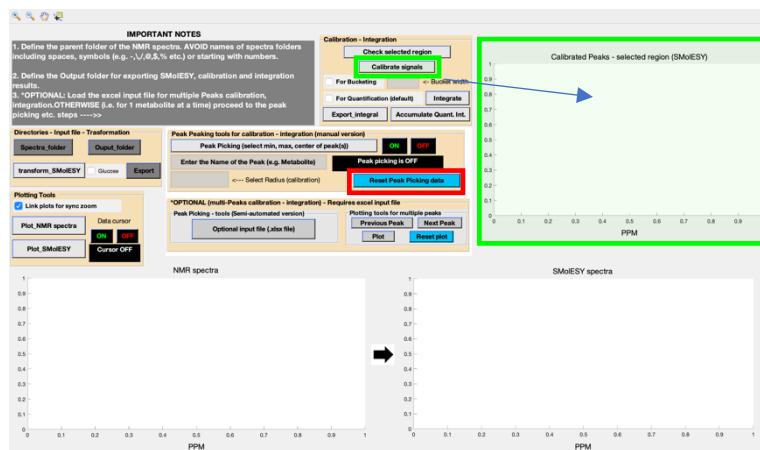


Figure 19

→TIP: If no calibration is needed, put 0.000 (ppm) value for the “radius” 0.000 <-- Select Radius (calibration)

→Step 5

If calibration of SMoESY peaks of interest is completed (i.e. the singlet of acetate for that example), the integrals (OPTION_1: “ For Quantification (default)”) or the binning (i.e. bucketing) (OPTION_2: “ For Bucketing <- Bucket width ”) of the selected spectral region from the 5 plasma SMoESY spectra could be calculated by pressing “ Integrate ” button and exported to the subfolder of the output folder named as acetate (e.g. in the folder: .../output_folder/acetate/) by pressing “ Export_integral ” button.

OPTION_1_explained: Targeted analyses – Integration for Quantification

By selecting the option of “ For Quantification (default) ”, user activates the function of integrating SMoESY signals for any specified metabolite. Integration process takes into account the negative part of the SMoESY signals and provides accurate integration values. **It is very important** i) that all signals should be integrated by the same way (Fig. 20), as well as the reference signal (if any) for absolute quantification or univariate statistical analyses (e.g. t-test etc.). **For absolute quantification, this function works for 1 defined signal per time. For any other use, user could include for integration multiple signals.**

After pressing “ Integrate ” button, user will be informed when integration process is finished and be ready for exporting “integral.txt” tab limited file in the subfolder of the output folder, namely in our case: “.../output_folder/acetate/acetate-integrals.txt”. The exported integrals file could be easily loaded to any platform for further employment. **It is highly noted that the algorithm integrates the WHOLE window defined by the MIN-MAX values pointed by the user (e.g. see yellow area of Fig. 20).** In addition, the accumulation of all integrals from multiple metabolites signals could be merged into one file named as “Cummulative_integrals.txt” by pressing “ Accumulate Quant. Int. ” button. **Please note that this function is NOT available for bucketing results.**

OPTION_2_explained: Untargeted analyses – Bucketing

By selecting the option of “ For Bucketing <- Bucket width ”, user should specify the bucket size, that should be NOT be < 0.001 ppm. After pressing “ Integrate ” button, user will be informed when bucketing process is finished and be ready for exporting “bucket_table.txt” tab limited file in the subfolder of the output folder, namely in our case:

User's Guide SMoESY_platform
https://github.com/pantakis/SMoESY_platform

".../output folder/acetate/bucket table.txt". The exported bucket table could be easily loaded to any multivariate (MVA) statistics toolbox for performing MVA statistics.

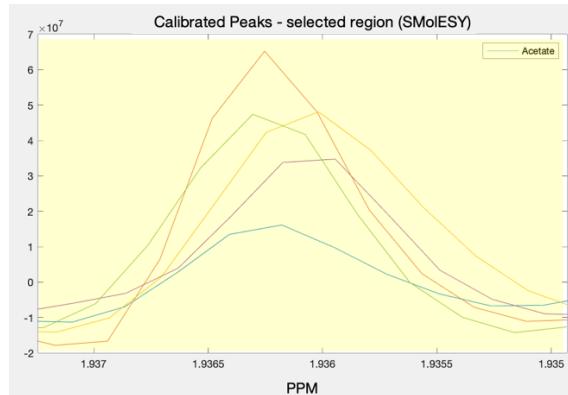


Figure 20. Yellow area highlights the integrated spectral area for calibrated acetate signals in the 5 plasma SMoESY spectral data.

The example of Glucose metabolite in 5 plasma ^1H -NMR spectra

Following the same procedure as acetate, the user would choose the values as indicated in Fig. 21:

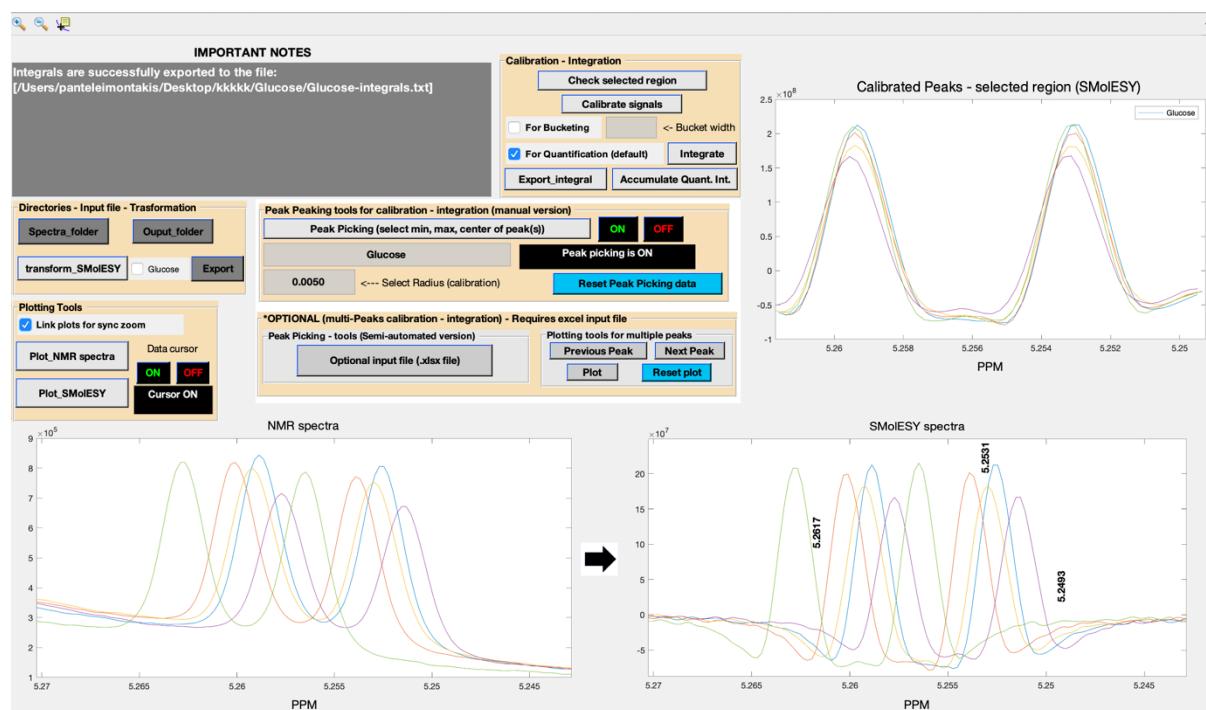


Figure 21.

IMPORTANT NOTE: For a comparative study e.g. of Glucose quantity between the spectra, user could use the integration of the Glucose doublet as calibrated in Fig. 21. However, for absolute quantification, the user should calibrate/integrate each signal of the doublet (see below the examples of semi-automated alignment-integration of L-alanine and L-valine) and sum the integrals.

User's Guide SMoESY_platform
https://github.com/pantakis/SMoESY_platform

The example of a spectral region 0.9-1.25 ppm in 5 plasma ^1H -NMR spectra

Following the same procedure as above, the user would choose the values as indicated in Fig. 22 and perform the binning of the aligned region at 1.0658 ppm with bin width 0.0050 ppm:

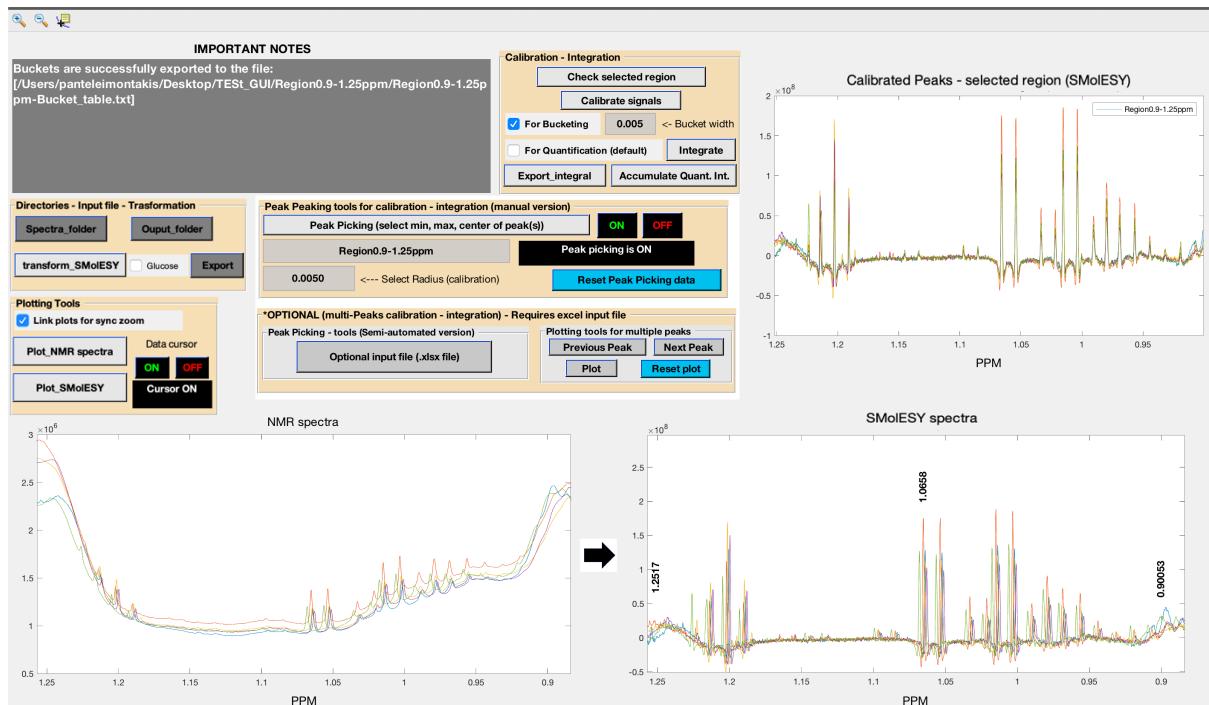


Figure 22

10.2 Semi-automated peak picking for calibration in the SMoESY data for multiple spectral regions – Integration

In the section 10.1 of the user's guide, the manual procedure for calibration/integration regarding 1 spectral region (i.e. acetate, glucose signals for 5 spectra) was shown. However, SMoESY_platform offers the option of a semi-automated defining of multiple spectral regions (i.e. signals) (Fig. 23 green highlighted) and their calibration/integration/bucketing.

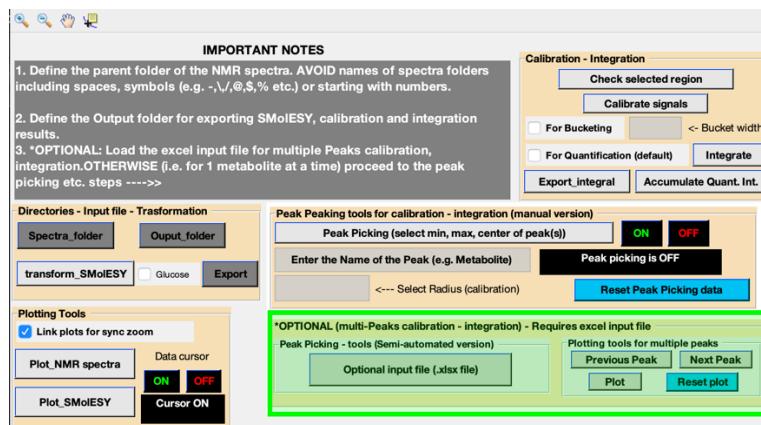


Figure 23.

Namely, the user can skip steps 2-4 of the section 10.1, and upload an excel file (.xlsx) (Fig. 24) by pressing the “” button.

A template of the excel file structure is provided at the Github folder named as “Multiple_Peaks_input_test.xlsx”.

Metabolites	Min_window	Max_window	Calibration_Center	Calibration_Center_Radius
1				
2				
3				
4				
5				

Figure 24.

An example of this process is described below:

The example of L-alanine and L-valine ^1H -NMR signals semi-automated calibration-integration in 5 plasma ^1H -NMR spectra, explaining the step-by-step followed procedure is pointed below:

→Step 1

The assignment of the ^1H -NMR doublet of L-alanine -CH₃ (Fig. 25a) and the one doublet of L-valine -CH₃ (Fig. 25b) groups in the SMoESY spectra. The spectral areas where the 2 components of each doublet in the 5 spectra are highlighted by blue boxes in Fig. 25.

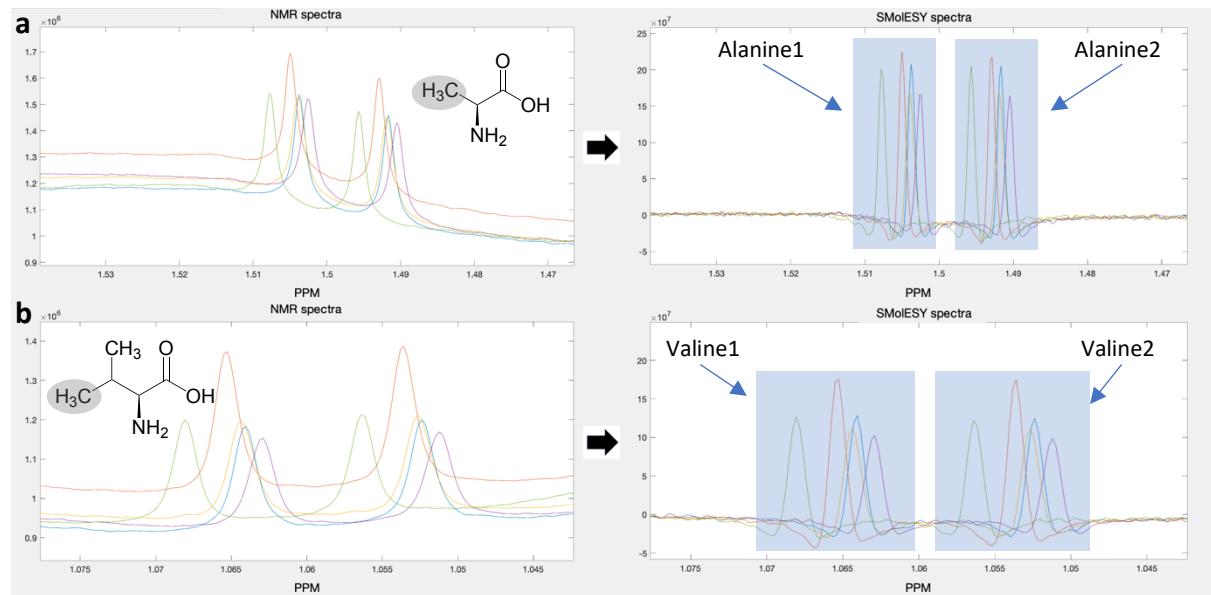


Figure 25.

→Step 2

Employing the SMoESY_platform plotting tools (e.g. data cursor etc.), user could immediately calculate all required values for calibrating each component of the two doublets (Fig. 25), as previously explained in the steps 2-3 at 10.1 section, and include them in the excel file as depicted in Fig. 26.

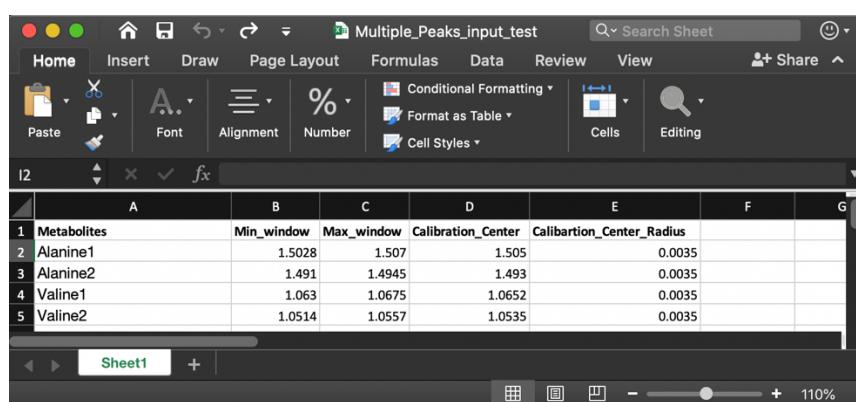


Figure 26.

→Step 3

After loading the excel file ("Optional input file (.xlsx file)") in the toolbox, user could immediately calibrate all peaks to the reference values reported in the excel file (i.e. Calibration_Center), by pressing the "Calibrate signals" button.

→Step 4

After being informed by the toolbox that the calibration of the peaks is finished, the user can navigate from calibrated peak to peak by pressing the “previous” or “next” buttons in the

“Plotting tools for multiple peaks” panel. Each time, user could press “Next Peak” → “Plot” → “Next Peak”

etc. buttons. When pressing “Plot” the calibrated peaks (e.g. Alanine1, Alanine2 etc.) for all spectra are plotted in the “Calibrated peaks” panel of the GUI (Fig. 27).

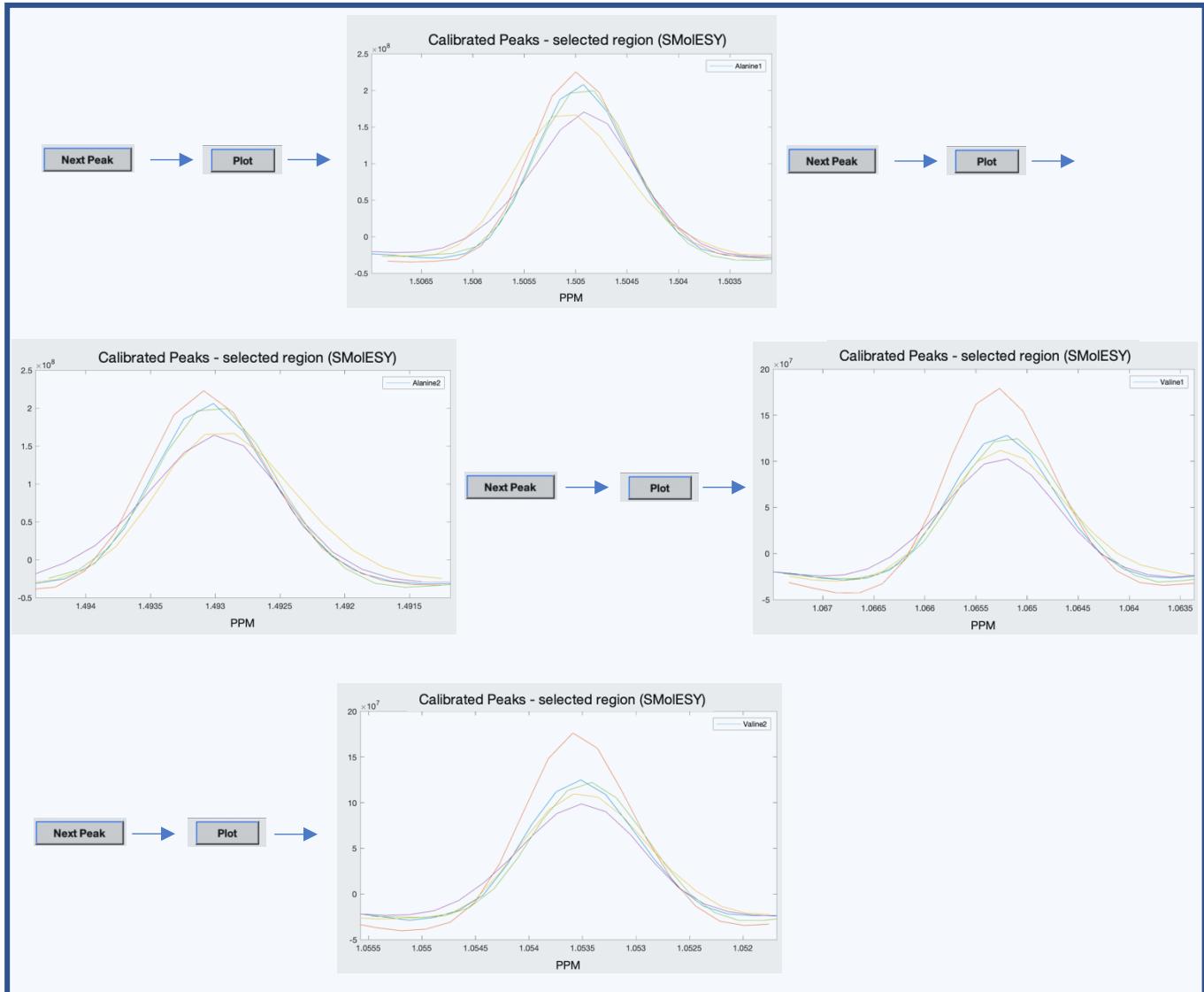


Figure 27.

→Step 5

For integration of peaks or bucketing of multiple spectral regions (i.e. variable shape buckets), the procedure is the same as in [step 5 of the section 10.1](#).

→TIP: If no calibration is needed, put 0.000 (ppm) value for the “radius” either in the excel file at the “Calibration_Center_Radius” cells or in the

11. Examples-templates

After downloading the toolbox folder from https://github.com/pantakis/SMoESY_platform, in the folder “Examples-Templates”, user could find:

- A folder containing 5 plasma samples ^1H -NMR 1D-NOESY spectra: [Five_plasma_spectra].
- The output folder containing all results from the workflow analyses described in sections 10.1 and 10.2 of the user's guide: [Output_five_plasma_spectra].
- The excel file template for multiple peaks alignment-integration which could be also used for variable shaped bucketing: [Multiple_Peaks_input_test].

12. Technical requirements - support

→ SMoESY_platform toolbox was created in MATLAB, 2019b (MathWorks). The software code is compatible with MATLAB 2017b and above.

→ The compiled versions of SMoESY_platform for MAC_OS and WINDOWS, were created by MATLAB, 2019b Update 5 (MathWorks) and they were tested in macOS Catalina (10.15.2) – macOS Mojave (10.14.6 - 18G2022) and WINDOWS 10, respectively, where they are fully functional. For their installation, both compiled versions of SMoESY_platform need R2019b (9.7) 64-bit MATLAB Runtime, which is freely available by MathWorks: (<https://uk.mathworks.com/products/compiler/matlab-runtime.html>).

MATLAB Runtime will be automatically downloaded/installed during the installation of both compiled versions of SMoESY_platform.

→ For troubleshooting and/or other technical information for the SMoESY_platform, contact Dr. Panteleimon G. Takis *via* e-mail: p.takis@imperial.ac.uk, with title of the e-mail: “Query for SMoESY_platform”

13. Copyright-License

Copyright (C) 2020 Panteleimon Takis, Ph.D.

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see <<https://www.gnu.org/licenses/>>.

Email: p.takis@imperial.ac.uk

GNU GENERAL PUBLIC LICENSE
Version 3, 29 June 2007

Copyright (C) 2007 Free Software Foundation, Inc. <<https://fsf.org/>>
Everyone is permitted to copy and distribute verbatim copies
of this license document, but changing it is not allowed.

14. References

1. Emwas, A.-H. *et al.* NMR Spectroscopy for Metabolomics Research. *Metabolites* **9**, 123 (2019).
2. Vignoli, A. *et al.* High-Throughput Metabolomics by 1D NMR. *Angew. Chemie - Int. Ed.* **58**, 968–994 (2019).