[](https://www.biochem.mpg.de/en)[](https://www.biochem.mpg.de/mann)[](https://github.com/MannLabs/alphamap) py\_diAID User Guide

Developed by: Patricia Skowronek, Eugenia Voytik, Sander Willems, and Matthias Mann.

This step-by-step guide presents instructions for the installation and usage of py\_diAID.

Table of Contents

[Description 1](#_Toc154074393)

[Windows installation 2](#_Toc154074394)

[How to use py\_diAID 5](#_Toc154074395)

[Optimal dia-PASEF methods 5](#_Toc154074396)

[Load a library 5](#_Toc154074397)

[Analysis software instructions 6](#_Toc154074398)

[Specify method parameters 7](#_Toc154074399)

[Optimization 8](#_Toc154074400)

[Create method 9](#_Toc154074401)

[Evaluate method 10](#_Toc154074402)

[Synchro-PASEF methods 11](#_Toc154074403)

[Load a library 12](#_Toc154074404)

[Define a scan area 12](#_Toc154074405)

[Generate a synchro-PASEF method 13](#_Toc154074406)

[Load a synchro-PASEF method 14](#_Toc154074407)

[Evaluate a synchro-PASEF method 15](#_Toc154074408)

[Make a GIF for a synchro-PASEF method 16](#_Toc154074409)

# Description

Over the last few years, Data-Independent Acquisition (DIA) coupled with Parallel Accumulation – Serial Fragmentation (PASEF) has been gaining interest among proteomics researchers. The scan mode dia-PASEF offers comprehensive proteome coverage, a high degree of reproducibility, and quantitative accuracy while utilizing a much larger ion beam proportion than conventional DIA methods 1,2. Its successor, synchro-PASEF, enables the creation of methods with even shorter cycle times, improving quantitative accuracy while also being highly specific due to the linking of fragment signals with precursor masses 3. Existing tools generate dia-PASEF and synchro-PASEF methods with equidistant isolation widths and necessitate manual adjustment of the window design to the precursor density cloud.

We present py\_diAID, a Python-based package for Data-Independent Acquisition providing Automated Isolation Design. py\_diAID optimally generates dia-PASEF and synchro-PASEF methods and places them optimally within the m/z – ion mobility (IM) plane. Additionally, it can generate variable isolation widths aligned to the precursor density in m/z, facilitating short acquisition cycles while covering virtually the entire m/z-IM-range. Our findings indicate that methods created with py\_diAID are advantageous for studying deep proteomes from cell lines, clinical samples with regular and very low sample input, as well as for exploring post-translational modifications such as phosphorylation 2, 4, 5.

We provide py\_diAID as a Python module, as well as a Graphical User Interface (GUI) for the Windows operating system, under an Apache 2.0 license. py\_diAID generates dia-PASEF and synchro-PASEF methods with an optimal window design. It also facilitates quality control of the precursor distribution in the m/z-IM plane and allows the evaluation of existing dia-PASEF and synchro-PASEF methods.

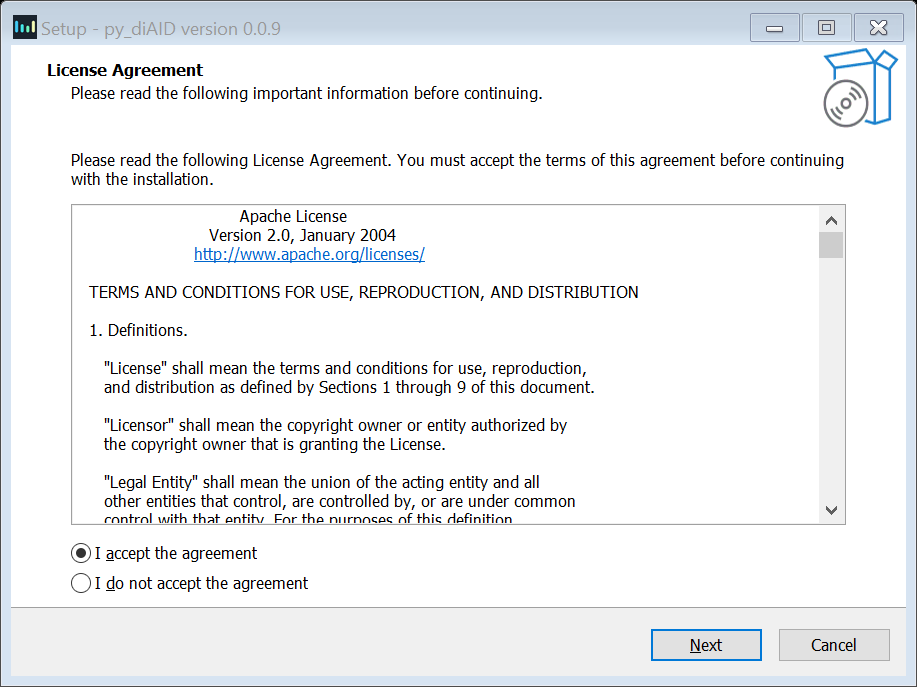
1. Meier, F. *et al.* diaPASEF: parallel accumulation–serial fragmentation combined with data independent acquisition. *Nat. Methods* **17**, 1229–1236 (2020).
2. Skowronek, P. *et al.* Rapid and In-Depth Coverage of the (Phospho-)Proteome With Deep Libraries and Optimal Window Design for dia-PASEF. *Mol. Cell. Proteomics* **21**, 100279 (2022).
3. Skowronek, P. *et al.* Synchro-PASEF Allows Precursor-Specific Fragment Ion Extraction and Interference Removal in Data-Independent Acquisition. *Mol. Cell. Proteomics* **22**, 100489 (2023).
4. Rosenberger, F. A. *et al.* Spatial single-cell mass spectrometry defines zonation of the hepatocyte proteome. *Nat. Methods* **20**, 1530–1536 (2023).
5. Thielert, M. *et al.* Robust dimethyl‐based multiplex‐DIA doubles single‐cell proteome depth via a reference channel. *Mol. Syst. Biol.* **19**, 2022.12.02.518917 (2023).

# Windows installation

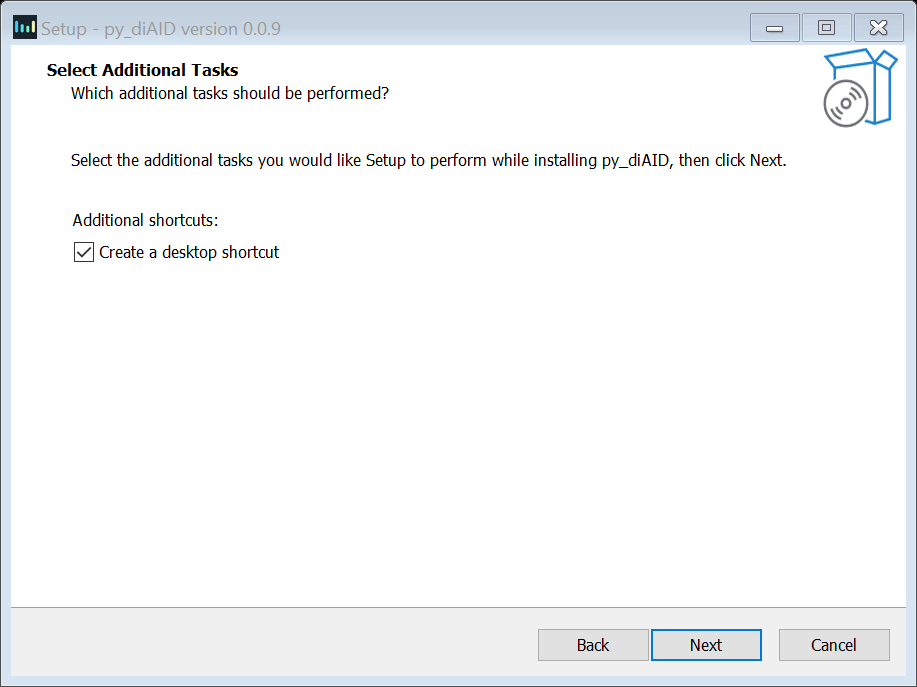
We offer py\_diAID as a browser-based GUI and a Python package. All modes are also described on the GitHub repository (<https://github.com/MannLabs/pydiaid>).

Windows 10 is required to install py\_diAID, and you need admin privileges if you want to run py\_diAID for all users (right-click on the py\_diAID logo on your desktop and select “Run as administrator”). We recommend uninstalling any previous py\_diAID versions before updating py\_diAID to prevent installation errors.

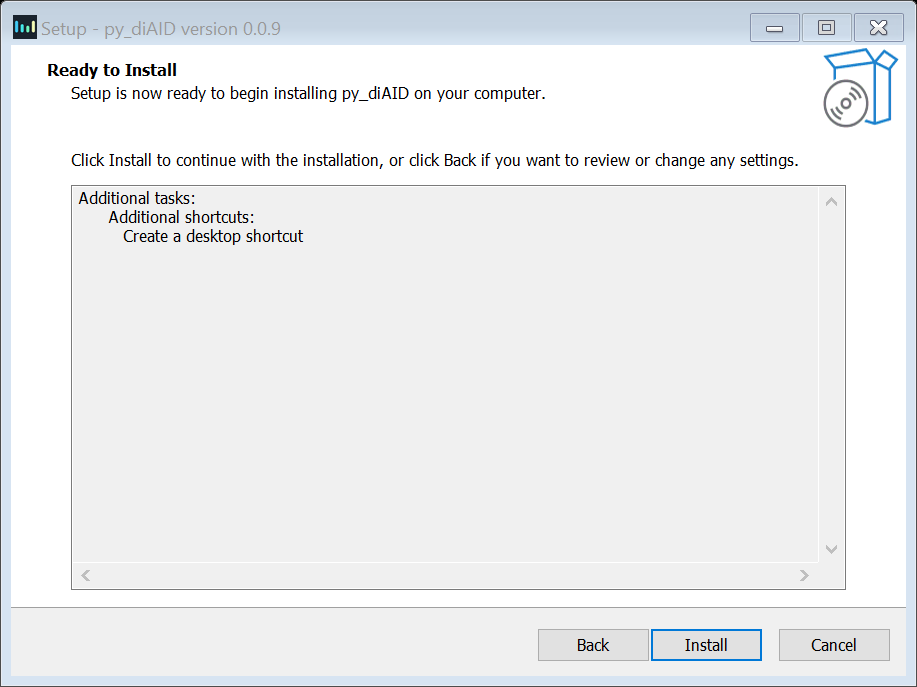
1. Download [the latest release](https://github.com/MannLabs/diaid_pasef/releases) of the package for Windows (pydiaid\_gui\_installer\_windows.exe) from the GitHub repository and execute the .exe file.
2. If the “User Account Control” dialog asks you for permission for the app to make changes to your device, please press the “Yes” button.
3. We recommend selecting “Install for me only (recommended)” if a “Select Setup Install Mode” dialog window shows up.
4. Please accept the License Agreement and click “Next” in the “Setup – py\_diAID version X.X.X” dialog window.



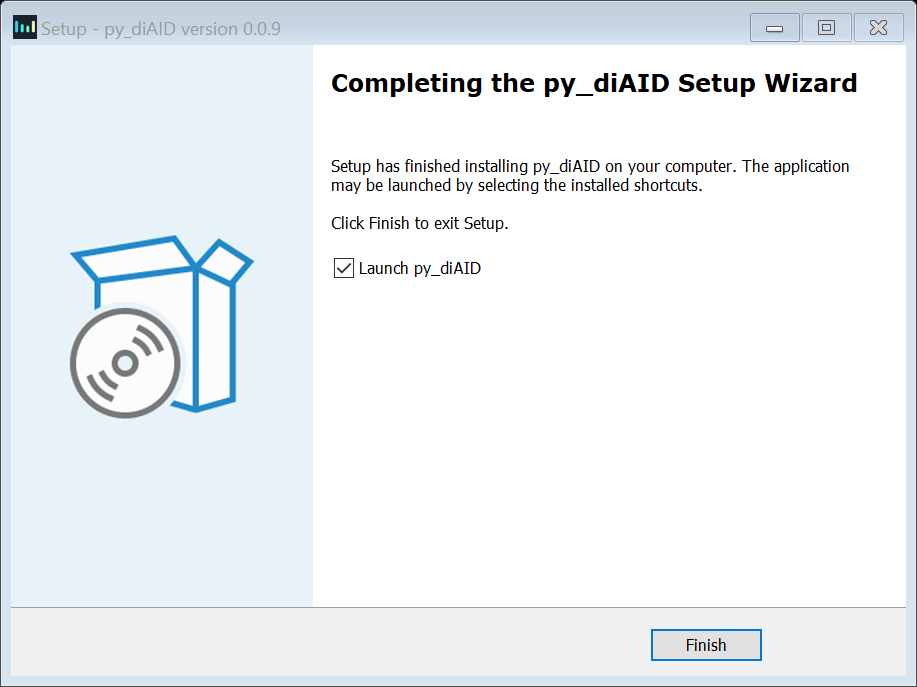
1. In the “Select Additional Tasks” dialog window, activate the “Create a desktop shortcut” check box and click the “Next” button.



1. Check the settings and press “Install”. You can change the setting with the “Back” or “Cancel” button prior to the installation.

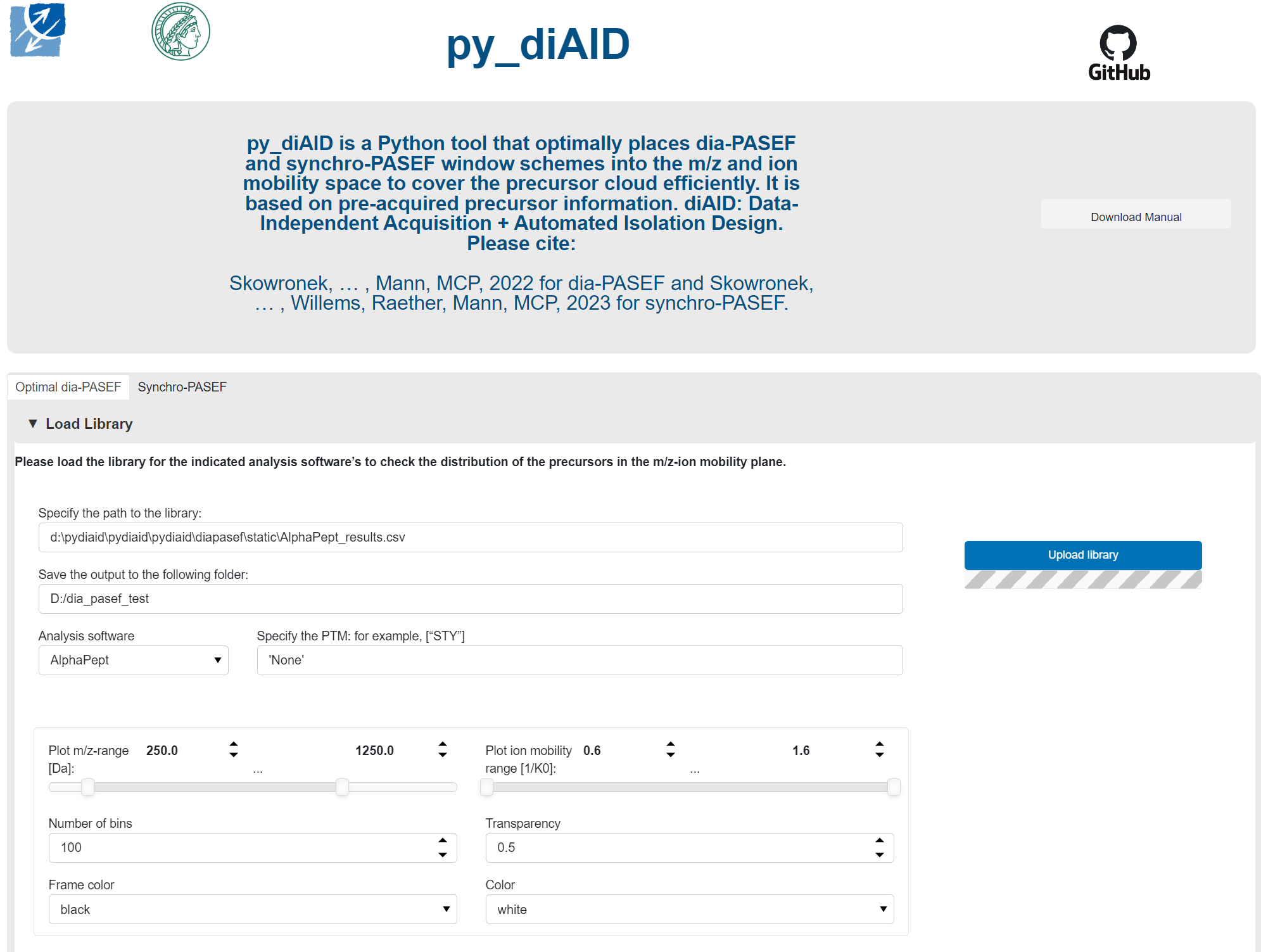


1. Please wait until the installation is finished, activate the “Launch py\_diAID” box, and click “Finish”.



1. If a “Window Security Alert” window appears, press “Allow access” to prevent the Windows Defender Firewall from blocking py\_diAID on your PC.

# How to use py\_diAID

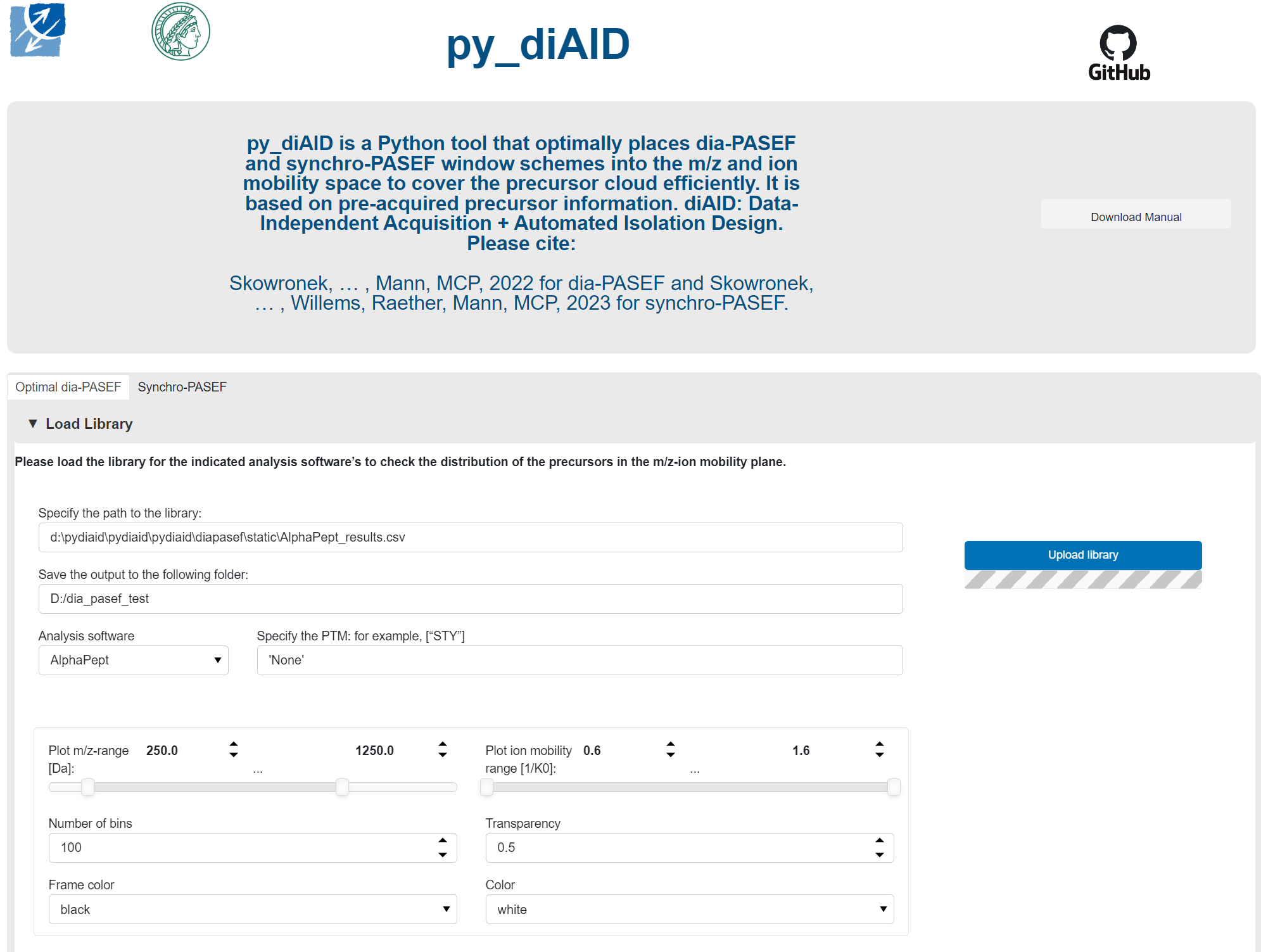


When you launch py\_diAID, a terminal window opens, displaying all the background information about py\_diAID, and a new browser tab titled “py\_diAID X.X.X” opens in your default browser. You can terminate py\_diAID either by closing that tab or by pressing “Ctrl+C” in the active terminal window. We recommend using either Google Chrome or Mozilla Firefox for the best performance. However, py\_diAID does not require an internet connection as it runs directly on your local system.

Py\_diAID facilitates the generation of optimal dia-PASEF and synchro-PASEF methods, indicated by two separate tabs. Firstly, we will explain the steps for generating an optimal DIA-PASEF method.

## Optimal dia-PASEF methods

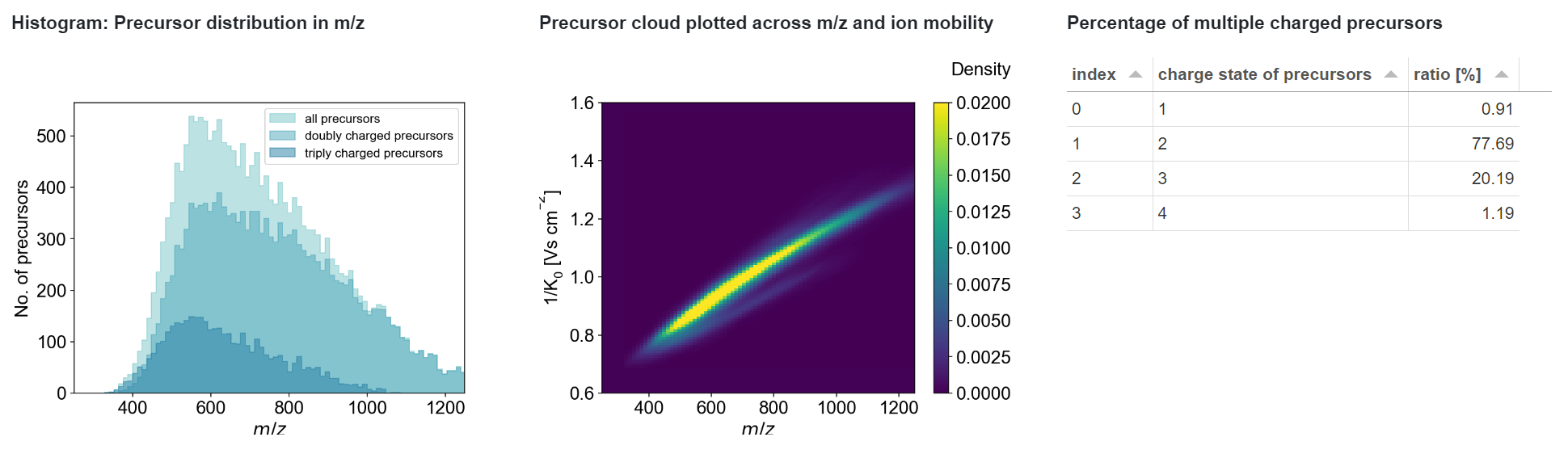
### Load a library



* Provide the filepath to the library (i.e., proteomics dataset) analyzed by AlphaPept, AlphaPeptDeep, MaxQuant, Spectronaut, FragPipe or DIA-NN library. Important: Indicate the software used in the drop-down menu labeled “Analysis software”.
* Specify the unique characters related to the post-translational modification of interest using a list. This function filters the column for modified peptide names based on the specified strings in the list. For Spectronaut, [“STY”] would be suitable for filtering phosphopeptide precursors and [“UniMod:28”] for FragPipe.
* The following settings apply to all plots generated by py\_diAID: “Plot m/z-range” and “Plot ion mobility-range” establish the limits of the m/z and ion mobility axes in all plots. py\_diAID calculates and plots kernel density estimates to represent the precursor distribution. The “Number of bins” influences the resolution of these plots. We recommend 100 bins for efficiency and up to 300 bins for sharper figures; however, this might take up to 30 minutes. “Transparency,” “Frame color,” and “Color” are properties of the isolation windows that are plotted to visualize the dia-PASEF methods.

A test library can be found at the following path: <*py\_diAID installation directory>\ pydiaid\diapasef\static\AlphaPept\_results.csv*. It represents a single-run from 200 ng tryptic HeLa digest acquired with 21-minute Evosep gradients and dda-PASEF.

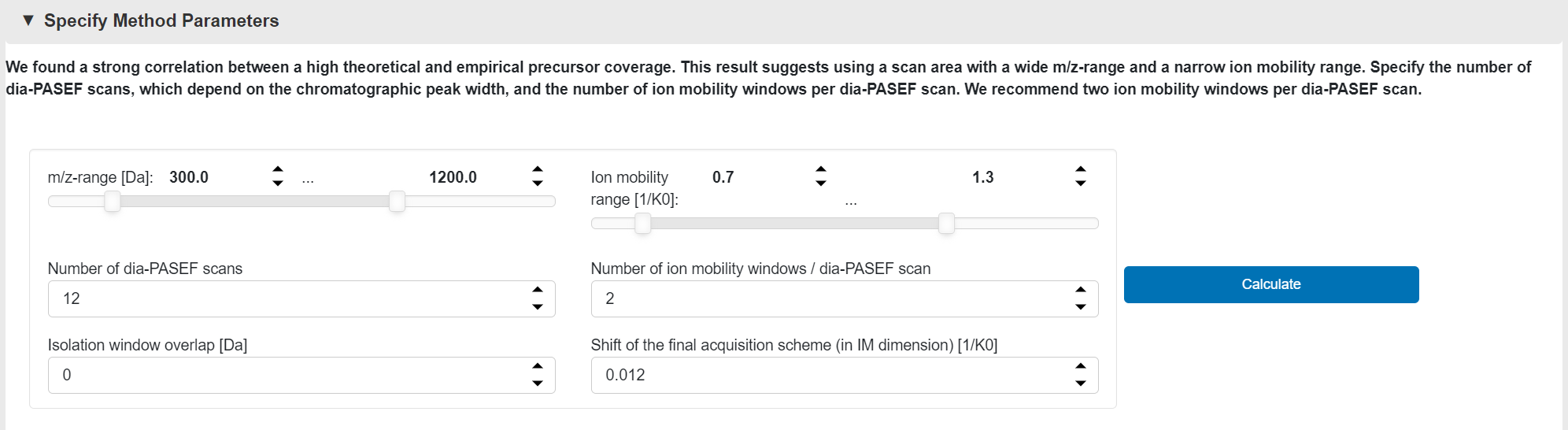
Press the “Upload Library” button to visualize the precursor distribution of the library/dataset with a histogram over m/z, a kernel density plot, and the charge state ratios. You can use these plots to check the ion mobility distribution of your samples. This is the only required step to activate the “create” and “evaluate” buttons.



### Analysis software instructions

* *AlphaPept:* Please use the **results.csv** file. py\_diAID uses the columns “q\_value”, “decoy”, “mz”, “mobility”, “charge”, “protein”, and “precursor”.
* *AlphaPeptDeep:* Please use the library as a **.tsv** **file**. py\_diAID uses the columns 'PrecursorMz', 'IonMobility', 'PrecursorCharge', 'ProteinID' and 'ModifiedPeptide'.
* *MaxQuant:* Please use the **evidence.txt** file. py\_diAID uses the columns “Reverse”, “m/z”, “1/K0”, “Charge”, “Proteins”, and “Modified sequence”.
* *Spectronaut single-run:* Please export the dataset in the **normal long** format already pre-filtered with “Quantification Data Filtering” and “No Decoy” as .tsv or .csv file. py\_diAID requires the columns “'FG.PrecMzCalibrated”, “FG.ApexIonMobility”, “FG.Charge”, “PG.ProteinGroups”, and “EG.PrecursorId”.
* *Spectronaut library:* Please export the library as an **.xls file**. py\_diAID requires the columns “PrecursorMz”, “IonMobility”, “PrecursorCharge”, “UniProtIds”, and “ModifiedPeptide”.
* *FragPipe:* The dataset should be analyzed while spectral library generation is active. Upload the **library.tsv** file. py\_diAID will use the columns “PrecursorMz”, “PrecursorIonMobility”, “PrecursorCharge”, “ProteinId”, and “ModifiedPeptideSequence”.
* DIA-NN library: Please use the DIA-NN library as a **.csv** **file**. py\_diAID uses the columns 'PrecursorMz', 'IonMobility', 'PrecursorCharge', 'ProteinID' and 'ModifiedPeptide'. Please note, regular DIA-NN output files that lack m/z information cannot serve as input for py\_diAID.

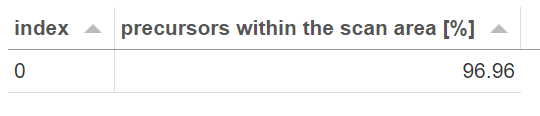
### Specify method parameters



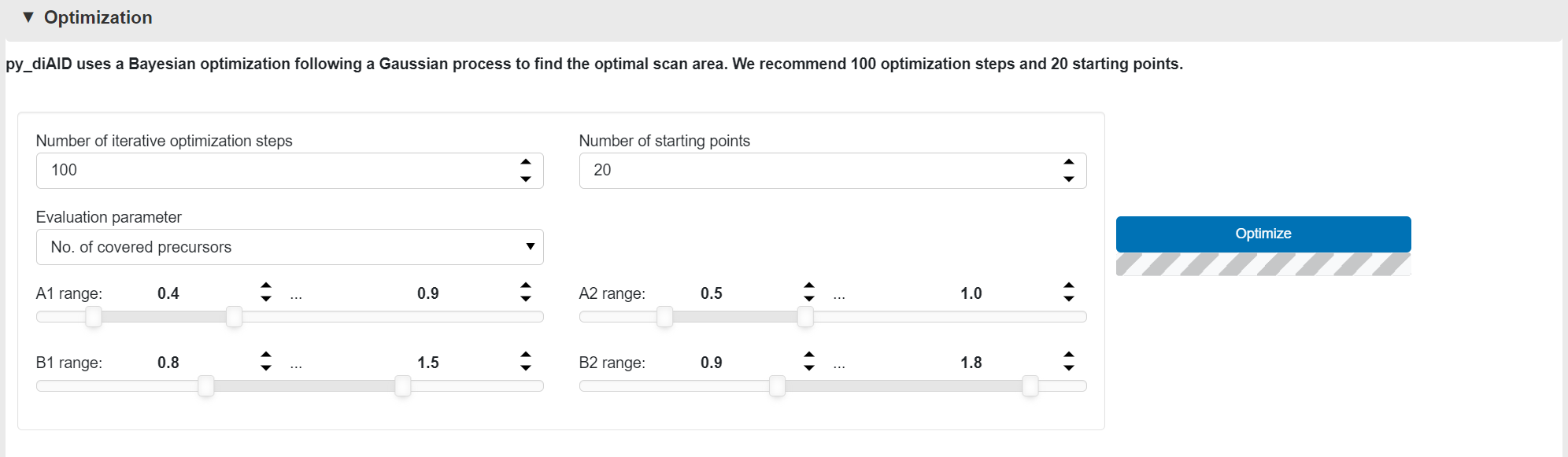
The following parameters define the optimal dia-PASEF method:

* ***m/z*-range:** Variable isolation windows allow for the coverage of a broad m/z-range while maintaining a short cycle time. Additionally, we found that methods with a high theoretical precursor coverage also result in the acquisition of more precursors in single runs. Therefore, we recommend using wide m/z-ranges when generating dia-PASEF methods.
* **Ion mobility range:** We recommend a reduced ion mobility range. For proteomics experiments, we use 0.7-1.3 1/K0.
* **Number of dia-PASEF scans:** The number of dia-PASEF scans influences the cycle time. Since the ramp and accumulation time is set to 100 ms by default, you can calculate the cycle time as follows: (MS1 scan + fragment scans) x (ramp time + transfer time) = (1+4) x (100 ms + 6.6 ms) = 533 ms. We recommend 8 dia-PASEF scans for 11 min Evosep gradients, 12 dia-PASEF scans for 21 min Evosep gradients, and 20 dia-PASEF scans for 44 min Evosep gradients (see [“Rapid and in-depth coverage of the (phospho-)proteome with deep libraries and optimal window design for dia-PASEF”](https://doi.org/10.1016/j.mcpro.2022.100279) Experimental procedures).
* **Number of ion mobility windows / dia-PASEF scans:** One dia-PASEF scan can be separated into multiple ion mobility windows. The number of ion mobility windows defines how often the quadrupole changes its position. For two ion mobility windows per dia-PASEF scan, the quadrupole occupies two isolation positions during a single dia-PASEF scan. We recommend two ion mobility windows per dia-PASEF scan since this results in the highest precursor coverage based on simulations and in higher peptide identifications (see [“Rapid and in-depth coverage of the (phospho-)proteome with deep libraries and optimal window design for dia-PASEF”](https://doi.org/10.1016/j.mcpro.2022.100279) Suppl. Figure S5).
* **Isolation window overlap:** This value defines the overlap in Da. We recommend no isolation window overlap for short cycle times.
* **Shift of the final acquisition scheme (in IM dimension**)**:** The dia-PASEF acquisition scheme will be optimized for a maximized theoretical precursor coverage. However, in practice, the quadrupole has a blind spot and does not record signal while changing to the isolation position. This blind spot results in missed precursors in the top ion mobility window region. The blind spot has an ion mobility height of approximately 0.022 1/K0. Therefore, we recommend to shift the final acquisition scheme upwards to not miss signal in the densest doubly charged precursor cloud.

Press “calculate” to calculate the precursors, which fall within the selected scan area defined by the *m/z*- and ion mobility ranges.



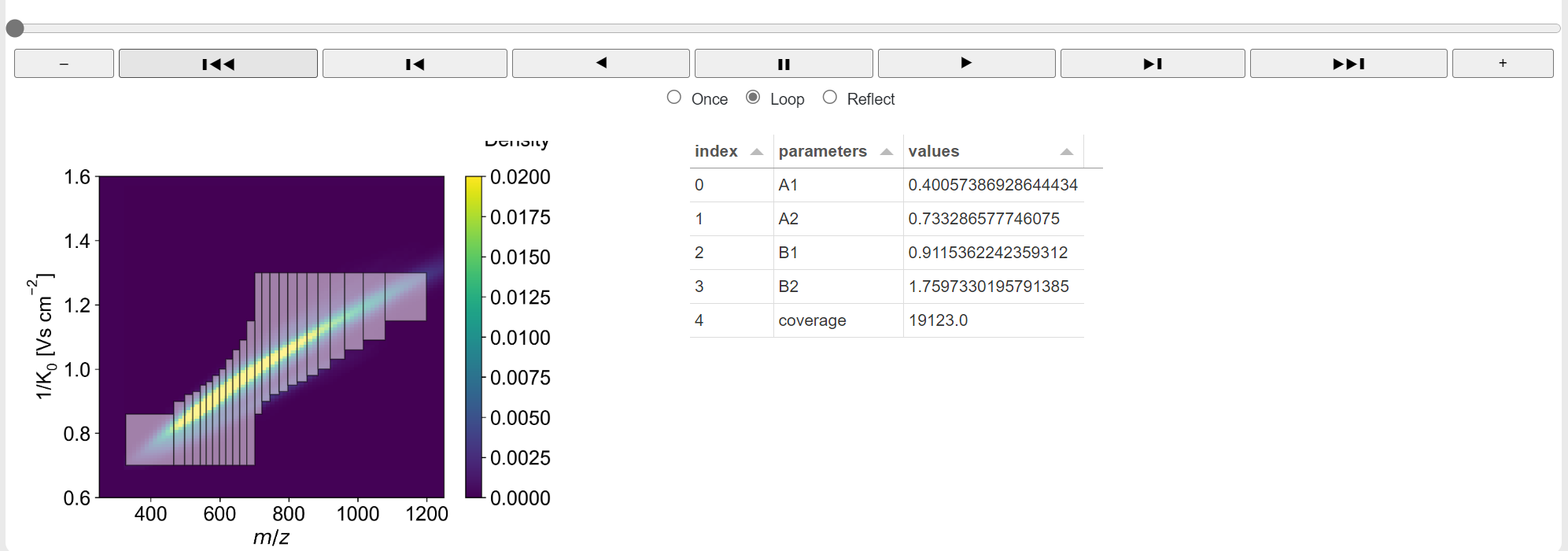
### Optimization



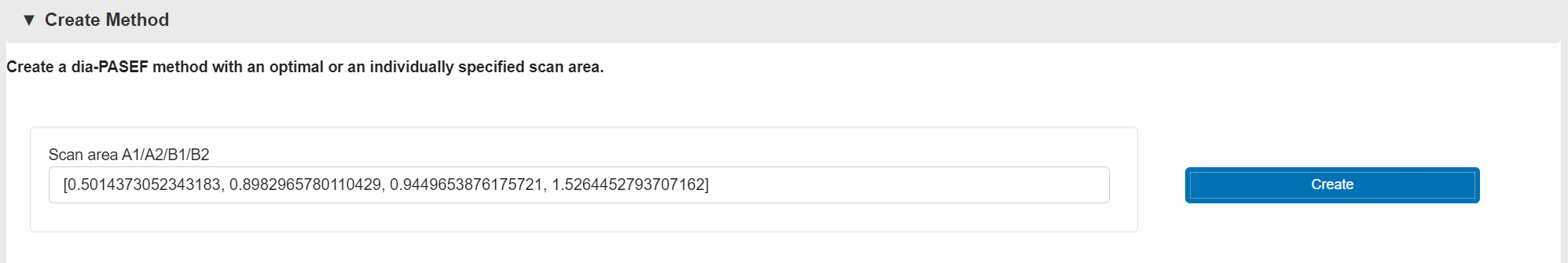
py\_diAID calculates the isolation window width based on the precursor distribution in *m/z.* The position of the isolation windows in the *m/z* and ion mobility plane is dependent on a trapezoid that reflects the scan area. py\_diAID selects a set of random A1, A2, B1, and B2 coordinates that define the corners of the trapezoid. A1, A2, B1, and B2 coordinates must lie within the respective ranges. Next, it generates a dia-PASEF method based on the placement of the trapezoid and the specified method parameters. This step is followed by the evaluation of the developed dia-PASEF method. Depending on the evaluation result, py\_diAID decides which trapezoid coordinates should be attempted next, to find the optimal precursor coverage. The decision is based on a Bayesian optimization following a Gaussian process offered by the skopt package. A more detailed description of the py\_diAID algorithm can be found in our publication: [“Rapid and in-depth coverage of the (phospho-)proteome with deep libraries and optimal window design for dia-PASEF”](https://doi.org/10.1016/j.mcpro.2022.100279) Figure 2 and Suppl. Figure S6.

* **Number of iterative optimization steps:** This value defines the number of iterations to find the optimal position of the dia-PASEF acquisition scheme.
* **Number of starting points**: The algorithm initially tries out multiple random sets of coordinates until it considers the evaluation results to decide which parameters are reasonable to try next. **“**Number of starting points” defines the number of this initial random set. We recommend 200 iterative optimization steps and 20 starting points for a result that provides 99.2% reproducibility (see [“Rapid and in-depth coverage of the (phospho-)proteome with deep libraries and optimal window design for dia-PASEF”](https://doi.org/10.1016/j.mcpro.2022.100279) Suppl. Figure S7).
* **Evaluation parameter:** py\_diAID can use different factors to evaluate the coverage of the dia-PASEF methods. “Number of covered precursors” includes all precursors and is recommended.
* **A1/A2/B1/B2 range:** The coordinates A1, A2, B1, B2 of the trapezoid are selected from within these ranges. (Please find a more detailed description in Figure 2 and supplementary Figure S6 in our [publication](https://doi.org/10.1016/j.mcpro.2022.100279).)

Press the “Optimize” button to start the optimization process. py\_diAID needs approximately 15 minutes for 200 iterative optimization steps. You can supervise all optimization steps after the optimization process is finished using the built-in player.

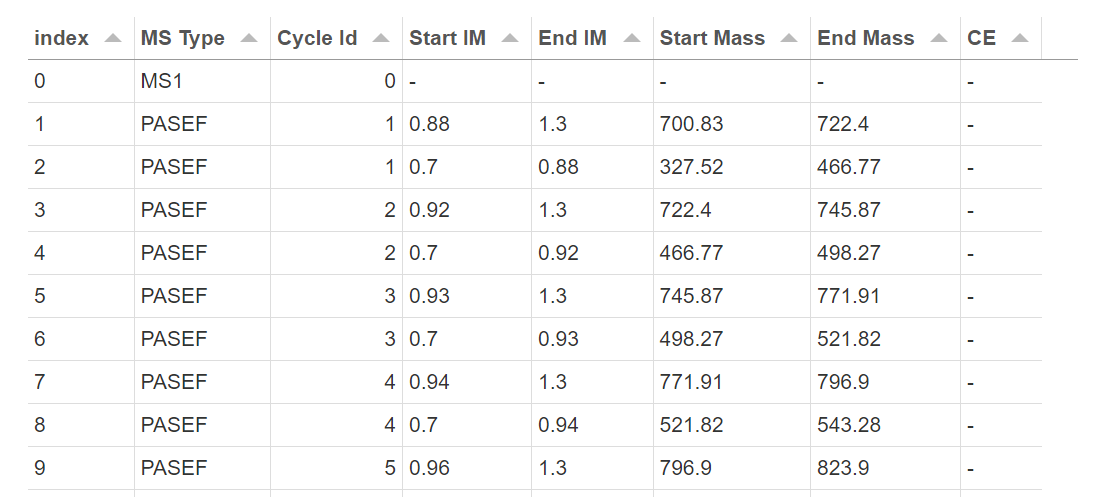


### Create method

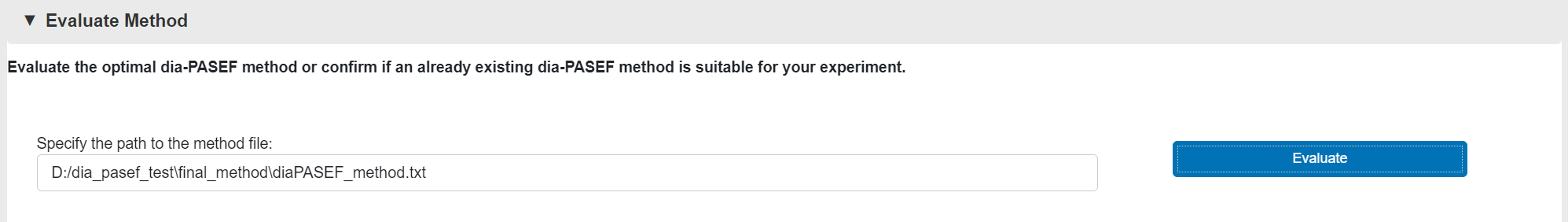


py\_diAID records the optimal trapezoid coordinates in the file “Scan area A1/A2/B1/B2”. If you already have coordinates from a previous optimization process, you can bypass the optimization phase and replace these coordinates with the previously obtained ones.

Press the “Create” button to generate the final dia-PASEF method based on the specified trapezoid coordinates. py\_diAID shifts this method upwards in the ion mobility dimension by the value specified in “Shift of the final acquisition scheme”.

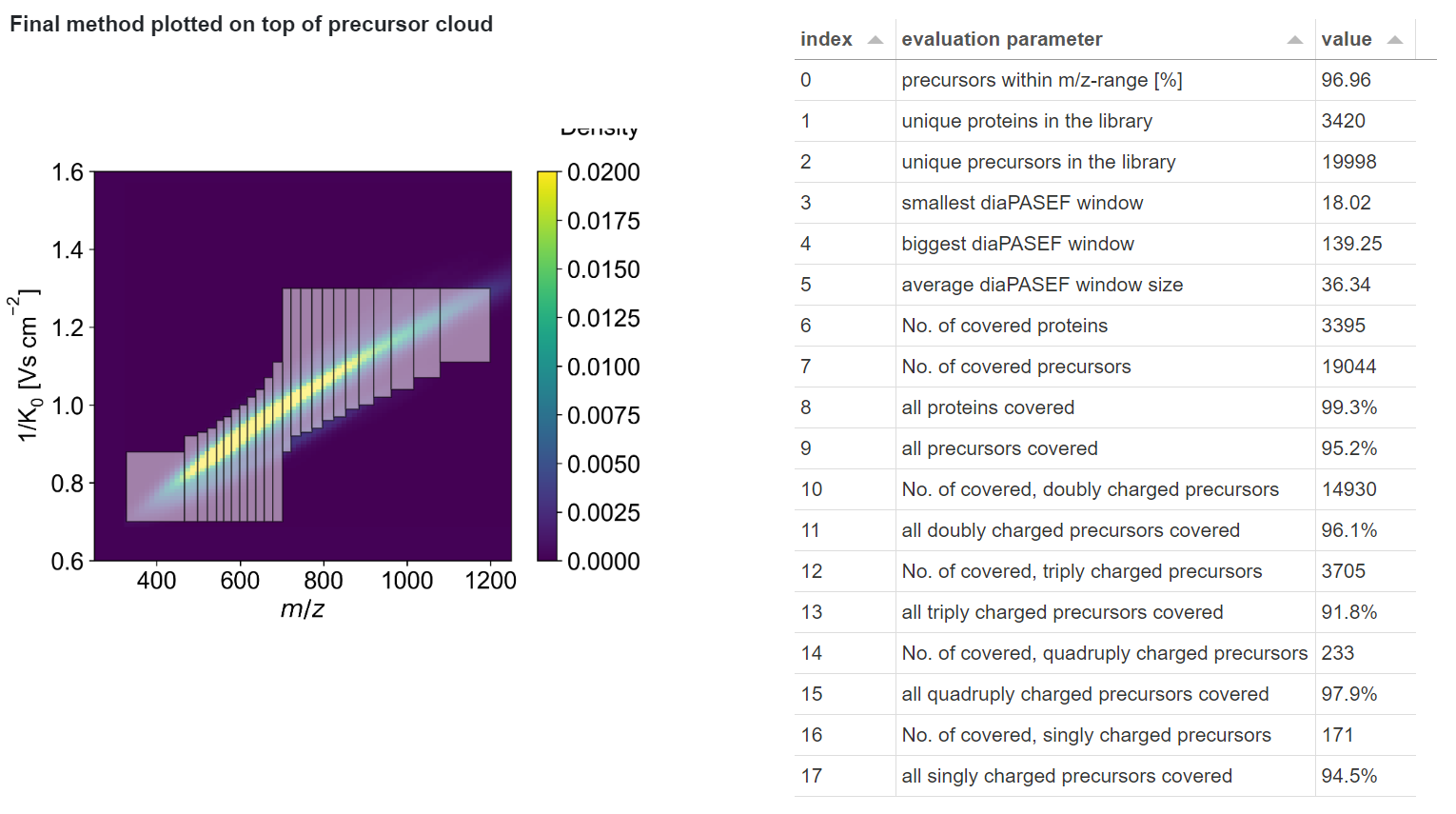


### Evaluate method



py\_diAID automatically writes the file path of the final acquisition scheme in the field labeled “Specify the path to the method file.” Alternatively, you can designate a custom path to evaluate whether an existing dia-PASEF method is suitable for your experiment, based on the previously generated library.

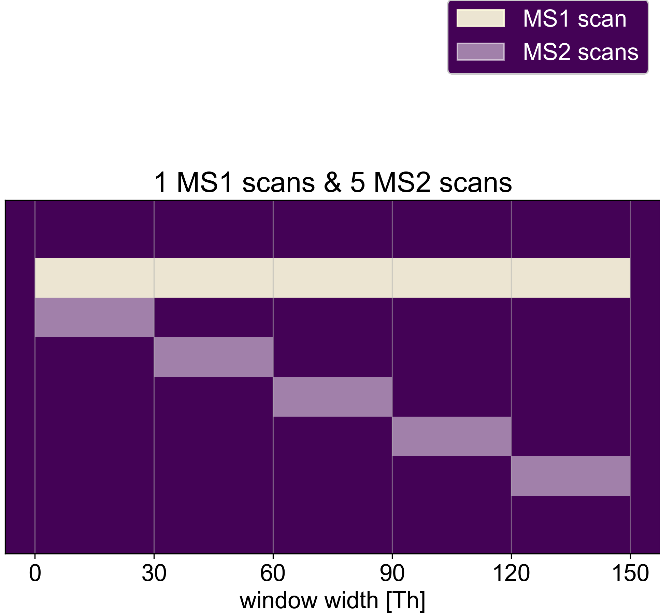
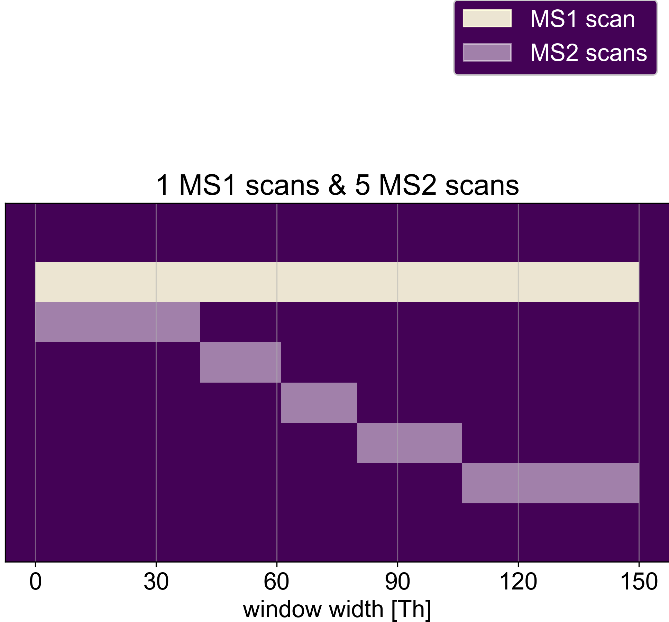
Press the “Evaluate” button to display the result of the evaluation



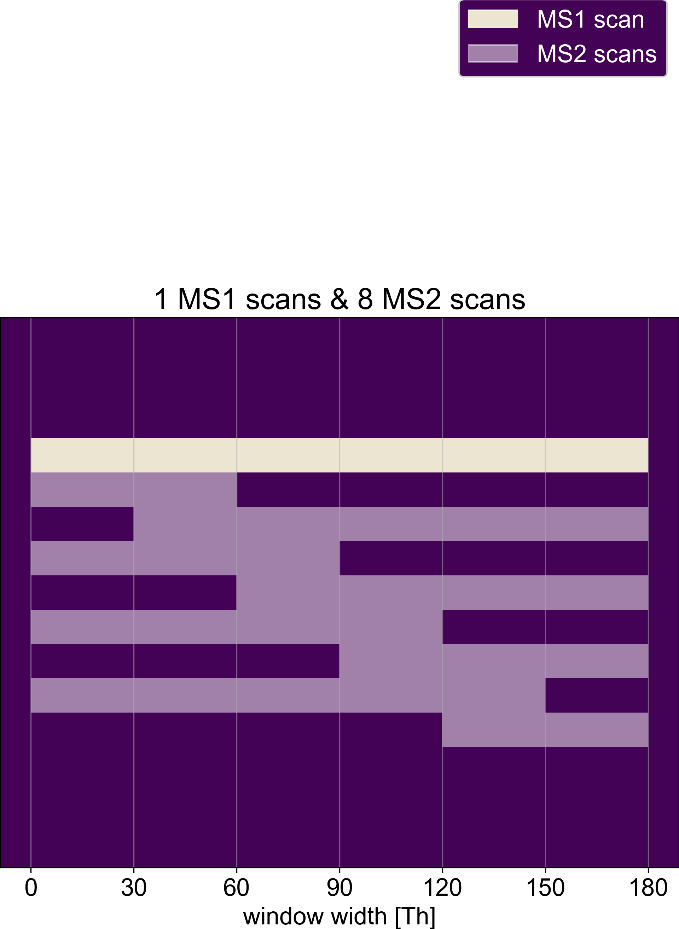
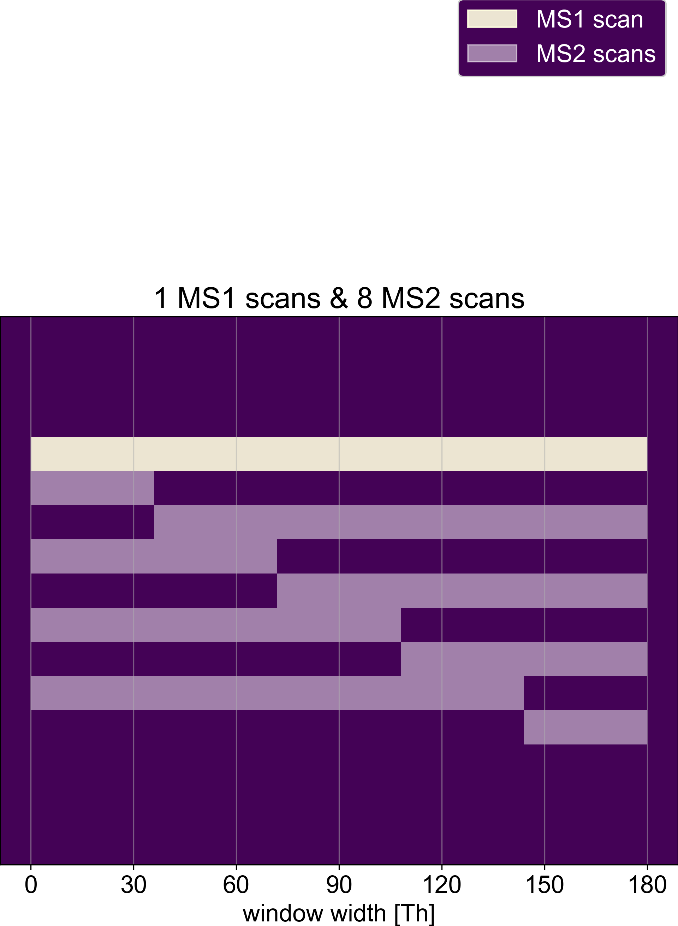
## Synchro-PASEF methods

Py\_diAID enables the generation of three synchro-PASEF acquisition schemes. Please note that only methods with equidistant isolation windows are currently supported by timsControl 5.0. Py\_diAID allows the generation of a wider variety of method schemes, which we believe will be relevant for future developments in this scan mode - hence py\_diAID already supports these. If your generated acquisition scheme is not compatible with timsControl 5.0, py\_diAID will issue a warning.

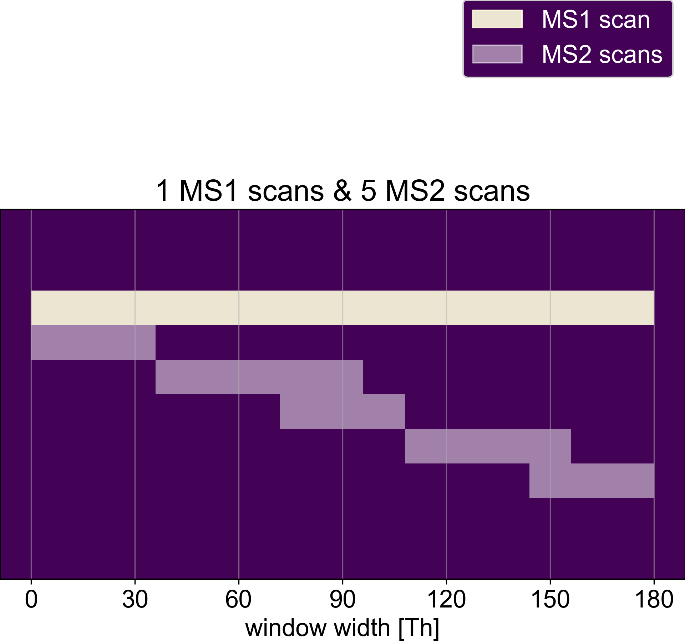
* **Classical synchro-PASEF** refers to an acquisition scheme where the isolation windows are positioned next to each other. The isolation windows can have fixed/equidistant (left) or variable width (right).

* **Highly accurate synchro-PASEF**, on the other hand, describes a method where two scans are always paired together. The first scan covers the left half of the scan range, and the second scan covers the right half. The division point where the scans are separated shifts to the right. This scheme is akin to the [second slice-PASEF scheme](https://doi.org/10.1101/2022.10.31.514544) developed by the groups of Markus Ralser and Vadim Demichev.



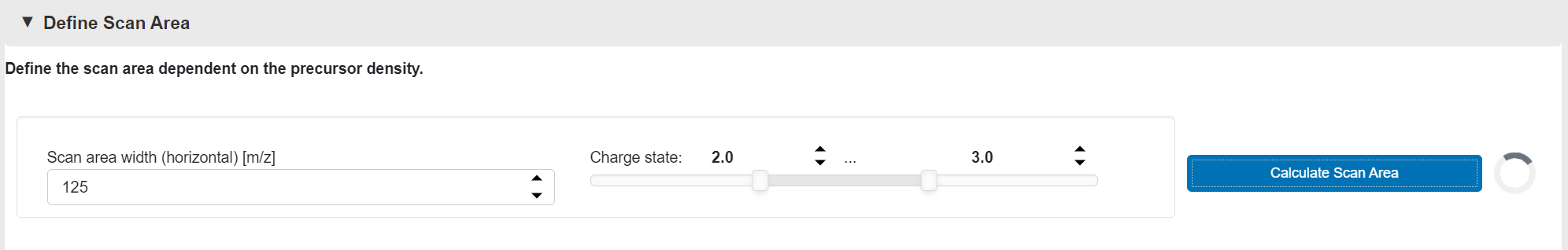
* Finally, the “**individual synchro-PASEF**” option allows complete freedom in defining the acquisition scheme pattern.



### Load a library

To load a peptide library, follow the instructions for optimal dia-PASEF methods on page 5. A test library can be found at <py\_diAID installation directory>\pydiaid\synchropasef\static \evidence\_MaxQuant\_270223.txt.

### Define a scan area

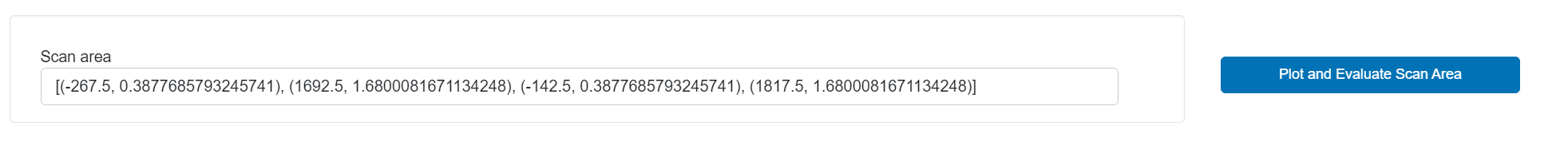


This card allows you to define the scan area that the synchro-PASEF method should cover.

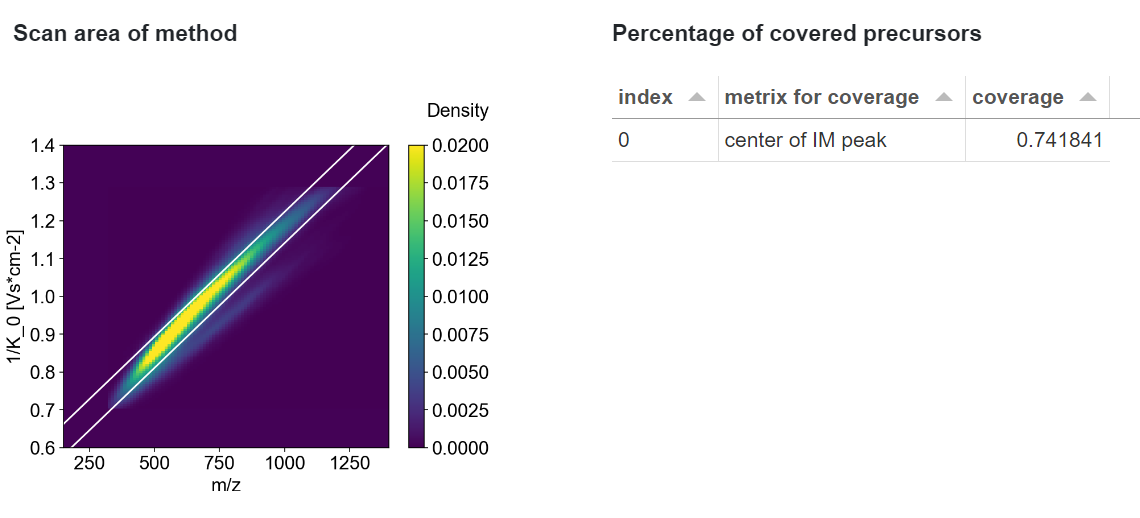
* **Scan area width (horizontal) [m/z]:** The scan area is a parallelogram, the width of which can be specified in m/z terms. If the isolation windows have an fixed/equidistant width, this scan area width also determines the width of the isolation window. For instance, having a scan area width of 125 Th and 4 synchro scans results in 25 Th for each isolation window.
* **Charge state:** The peptide library is filtered based on the specified charge states. Subsequently, the scan area is optimally positioned by considering only these specific peptide populations.

To determine the scan area, py\_diAID begins by filtering the peptide library for the specified charge states. It then calculates a linear regression based on the precursor population and subsequently moves the scan area in small steps both left and right. The purpose of this step-by-step movement is to find the ideal positioning so that the maximum number of precursors fall within the scan area.

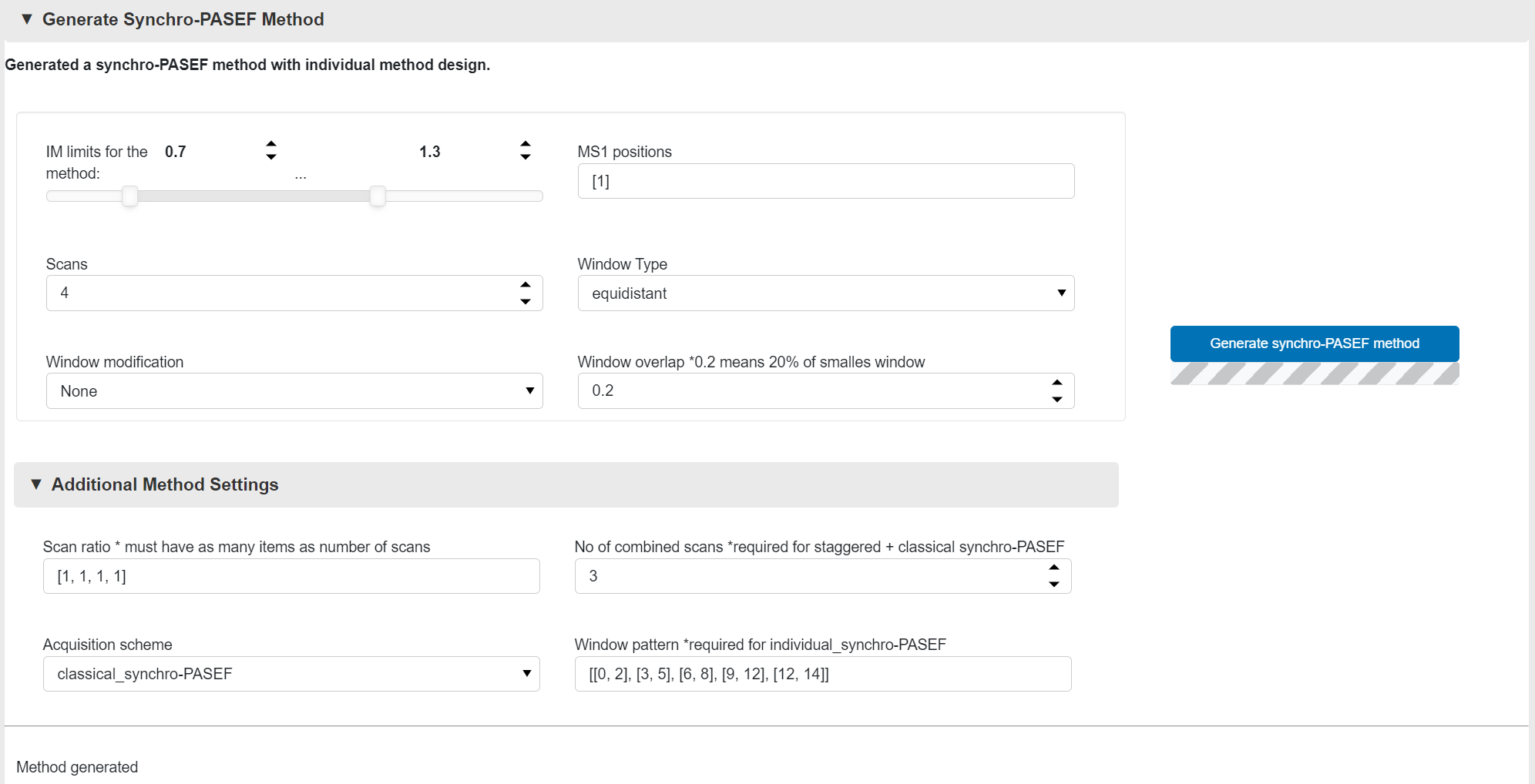
By pressing “Calculate scan area”, you can initiate this process. Once the process is complete, the final scan area coordinates will be displayed in the “Scan area” field.



To visualize the scan area and evaluate its coverage, simply click on “Plot and evaluate scan area”. This will generate a graph with the scan area superimposed over the precursor distribution, along with a percentage representing the number of covered precursors.



### Generate a synchro-PASEF method

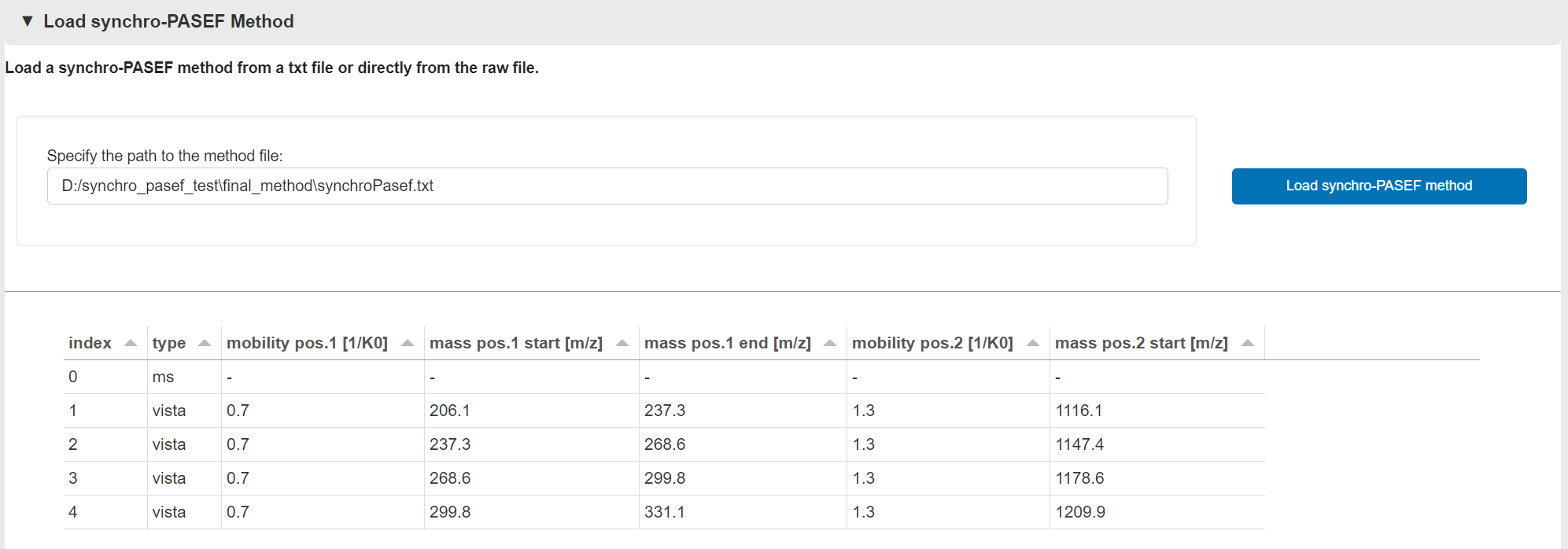


In this card, you can define the method parameters for the synchro-PASEF method. Only methods with equidistant isolation window widths can be calibrated using timsControl 5.0. If your generated acquisition scheme is not compatible with timsControl 5.0, py\_diAID will issue a warning.

* **Ion Mobility (IM) limits for the method:** This range defines the ion mobility covered by the synchro-PASEF method. For proteomics experiments, we typically use 0.7-1.3 1/K0, as most precursors are present within this range.
* **MS1 positions:** Here, you can specify the positions at which py\_diAID places MS1 scans. For instance, [1, 4] would result in a synchro-PASEF method with an MS1 scan as the first scan, followed by three MS2 scans before the next MS1 scan is scheduled.
* **Scans:** This setting defines the number of synchro scans (= isolation windows). We recommend a low scan count to create synchro-PASEF methods with shorter cycle times to maximize the efficiency of this scan mode.
* **Window type:** Py\_diAID can generate isolation windows with fixed/equidistant width, or with variable width where each synchro scan covers approximately the same number of precursors, or a pre-defined width. If you select the pre-defined setting, specify the desired window ratios in ‘additional method settings” and “scan ratio”. If not, py\_diAID will use the last used scan ratio as a default. Please note, the number of scans and items in the scan ratio list must match to use the pre-defined setting.
* **Window modification:** This setting allows you to choose between no window modification, window overlap, or a staggered isolation window scheme. For “overlap”, define a “window overlap”, or py\_diAID will use a default of 0.2 (20% window overlap). If you choose “staggered,” py\_diAID will combine adjacent isolation windows into one. For the “classical synchro-PASEF” acquisition scheme, it combines three scans by default. If you want a different number of combined scans, specify it under “Additional method settings” and “No. of combined scans.” For the “highly accurate synchro-PASEF” method, py\_diAID will always have an overlap of one scan/bin length which equals a 50% window overlap. For an individual Synchro-PASEF method, this setting will be ignored.
* **Window overlap:** This defines the extent of overlap between isolation windows. “Overlap” must be defined under “Window modification” for this setting to be necessary.
* **Scan ratio:** This parameter sets the ratio of scan widths. For example, using [1, 5, 1] results in a middle isolation window that is five times wider. The “pre-defined” option must be selected under “Window type” for this setting to be applicable.
* **Acquisition scheme:** Py\_diAID offers three types of acquisition schemes: classical synchro-PASEF, highly accurate synchro-PASEF, and the customizable individual synchro-PASEF. For highly accurate synchro-PASEF, the step width is defined by the “scans” and “window type” settings. For the “individual synchro-PASEF” option the acquisition scheme is designed according to the window distributions defined under “window pattern”.
* **Window pattern:** This setting is only in use for the individual synchro-PASEF acquisition scheme. Here, the scan lengths are defined by ranges. For instance, [0, 2] indicates that the first three scans/bins are combined into a single isolation window, starting at the m/z position 0. Py\_diAID automatically updates the ‘scans’ and ‘scan\_ratio’ fields.

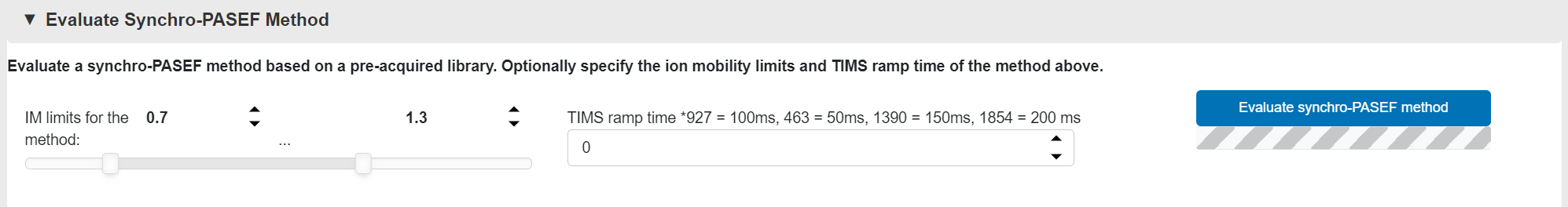
Press “Generate synchro-PASEF method” to create the synchroPasef.txt file. You will find it in the “final\_method” folder as defined under “Load library”, “Save the output at the following folder”. This file also serves as input for timsControl.

### Load a synchro-PASEF method



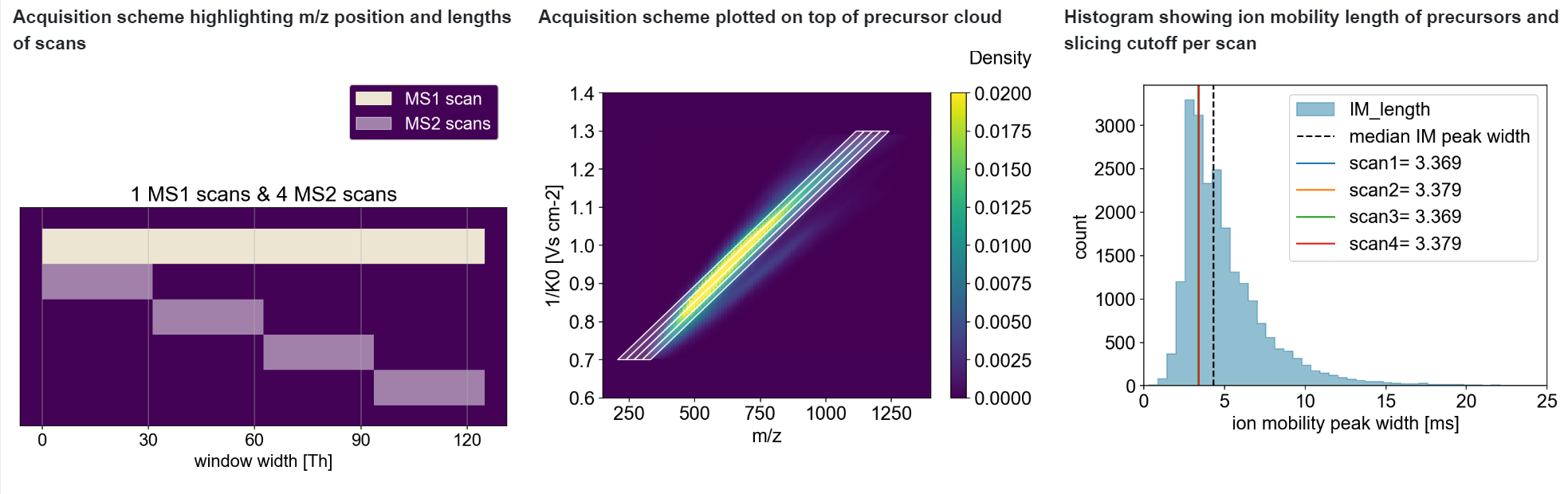
In this card, you have the option to load the newly generated synchro-PASEF method and review its parameters. The current method path is automatically updated in the “Specify the path to the method file” field. Alternatively, you can also load a pre-existing method. Press “Load synchro-PASEF method” to display the method parameters in a table format.

### Evaluate a synchro-PASEF method

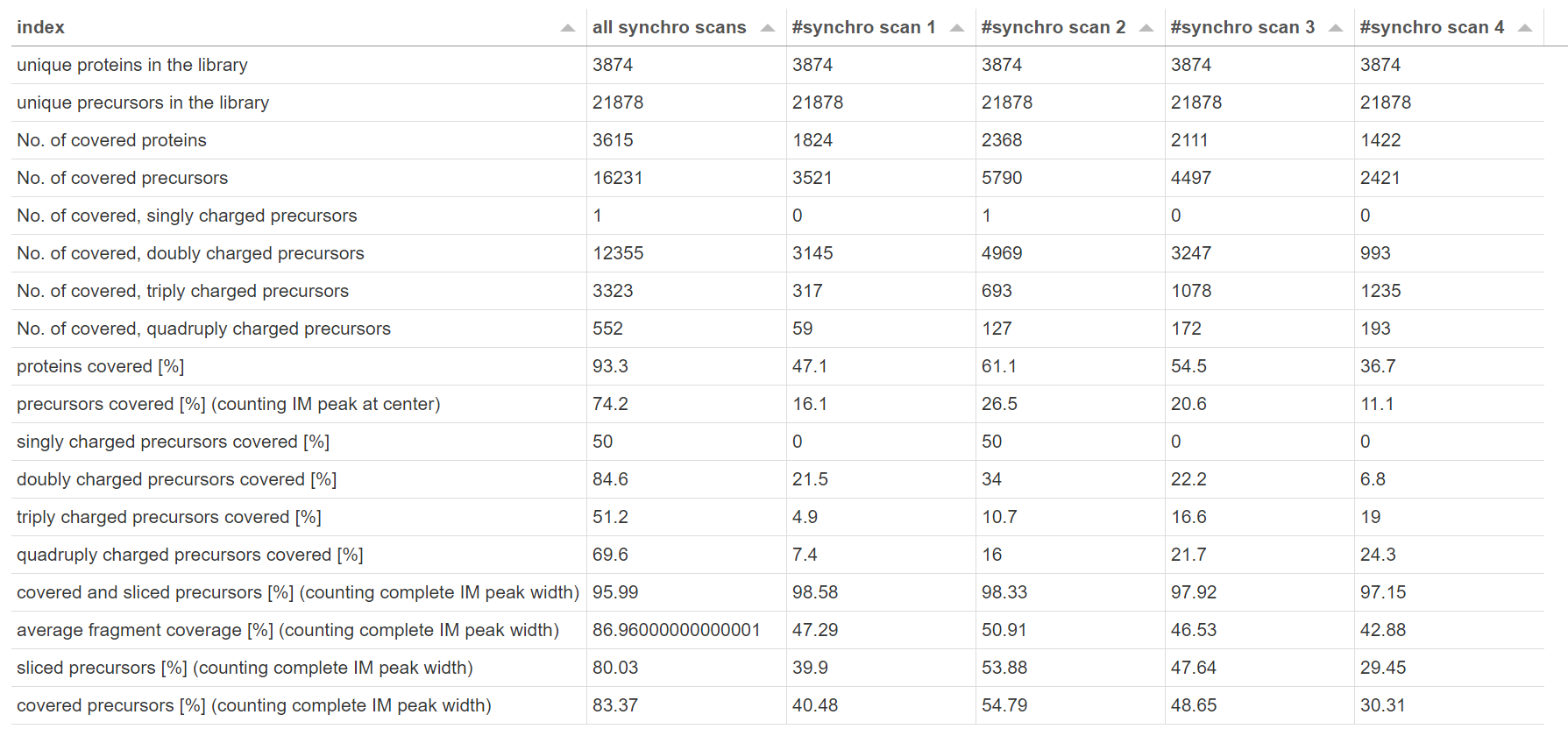
 In this card, you have the option to evaluate either the newly generated synchro-PASEF method or any pre-existing methods. This is only possible if you have already loaded a library in the “load library” card and a method in the “load synchro-PASEF method” card. If your peptide library includes information about the precursors’ ion mobility length, you can further define two settings: the ion mobility range of the method, which defaults to the one defined while generating the method, and the TIMS ramp time for raw data acquisition. Based on these values, py\_diAID can calculate and display the synchro scan width in milliseconds along the ion mobility dimension or in other words the slicing cutoff per scan.

Press “Evaluate synchro-PASEF method" for py\_diAID to generate up to three method evaluation plots and characteristics for benchmarking the method’s suitability.

Plots to benchmark synchro-PASEF methods:



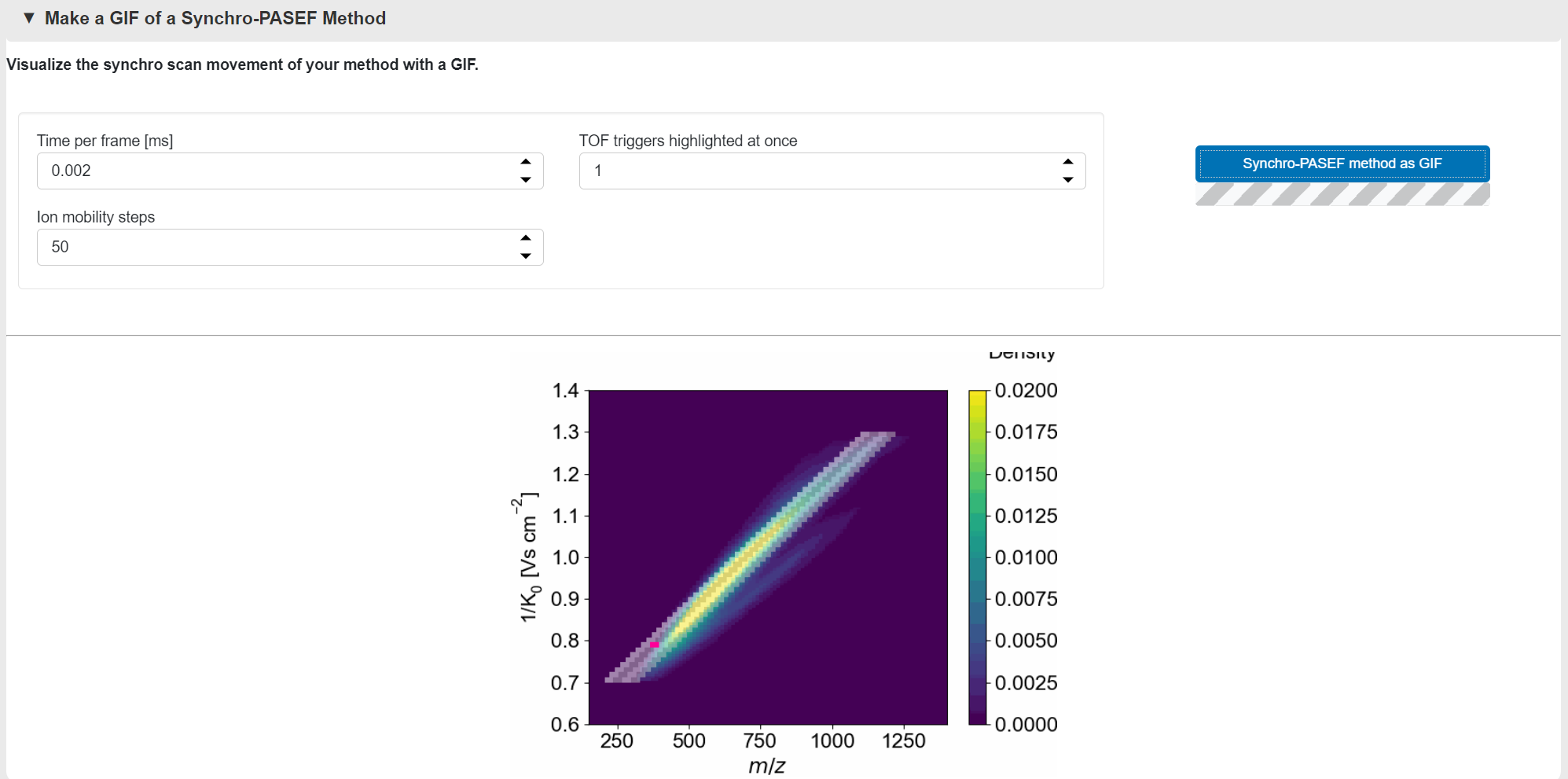
The left plot shows the widths of the isolation windows in the acquisition scheme and their order, regardless of their placement in the m/z and ion mobility plane. This plot aids in better understanding the placement of overlapping isolation windows. The middle plot overlays the isolation window on top of the kernel density distribution of precursors, helping to estimate precursor coverage. The right histogram displays the ion mobility peak width of peptides in the library, overlaid with the slicing cutoff of each synchro scan. Partitioning the fragment signal into adjacent synchro scans along the ion mobility peak enables linking fragment signal to their precursor signal, resulting in high specificity (refer to [our publication](https://doi.org/10.1016/j.mcpro.2022.100489) for further details). Precursors with an ion mobility peak width wider than the slicing cutoff of a synchro scan will be partitioned, and those with a smaller width will only be sliced if they are not positioned in the center of the synchro scan. We suggest using methods where the median ion mobility peak width is wider than the slicing cutoff for the majority of isolation windows. The last plot is only generated if the library contains ion mobility peak width information and if the TIMS ramp time exceeds zero.



Moreover, py\_diAID calculates the total precursor coverage as well as coverage per charge state, for all synchro scans and individual synchro scans. If the peptide library contains ion mobility peak width information, the displayed table will also show slicing characteristics such as the percentage of covered and sliced precursors, and the average percentage of covered fragments.

Together, these plots and the table should assist you in evaluating the suitability of the generated method. We recommend using methods with high precursor coverage and slicing percentage. All this information will be saved in the earlier specified output table at “final\_method”.

### Make a GIF for a synchro-PASEF method



In this final card, you can create a GIF of your own method to visualize the diagonal and ultra-fast movement of the quadrupole through the m/z and ion mobility plane. The GIF will be saved in the previously specified output table under “gif”.

* **Time per frame [ms]:** This setting specifies the duration of each frame in the GIF. A detailed representation of the diagonal synchro scan movement requires a short frame time, such as 0.002 ms.
* **TOF trigger highlighted at once:** If you select many ion mobility steps, therefore depicting the synchro scan movement with high resolution, a single moving TOF trigger might be visually difficult to follow. This setting allows you to highlight multiple TOF triggers at once in pink.
* **Ion mobility step:** This setting determines the resolution of the synchro scan movement. Py\_diAID will generate one snapshot per ion mobility step per scan. Therefore, GIFs with numerous ion mobility steps will take longer to generate and will result in longer, but more detailed GIFs. The maximum resolution is at 200 ion mobility steps.