Automated Library Building Using AlphaPeptDeep and Carafe AI Tools in Skyline

David Shteynberg[1] ; Bo Wen[1] ; Nicholas Shulman[1] ; Matthew Chambers[1] ; Brian Pratt[1] ; William S. Noble[1,2] ; Michael J. MacCoss[1] ; Brendan X. MacLean[1s]

[1]University of Washington, Seattle, WA [2]Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA

Introduction

Skyline is widely-used and open-source Windows application for analyzing mass spectrometry results. Skyline has automated tools for generating and managing spectral libraries in .blib format.

Previously, inputs for building .blib spectral libraries with Skyline were limited to using files containing matched peptide results or to modifying existing spectral libraries also contained in other .blib files. Alternatively, Skyline can be used with Koina for generating AI-predicted libraries. However, this approach requires sending the peptide sequences to a second party over the internet where the peptides can be seen by someone outside the organization.

The next version of Skyline will present new features enabling the use of AI on the local machine for predicting MS/MS spectra, retention times and collisional cross-sections using the tools AlphaPeptDeep and Carafe.

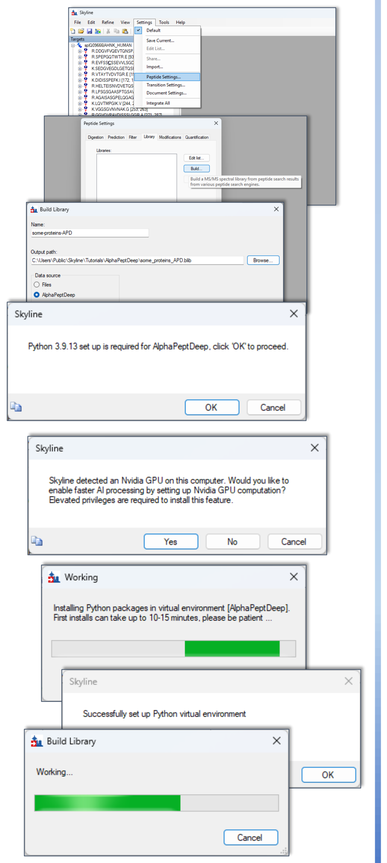
Skyline was modified to make local use of AlphaPeptDeep and also Carafe software. This required modifications such that all processing would occur on the local computer running Skyline. To achieve this, Skyline handles all installation of the necessary software and dependencies for processing of the data locally using either AlphaPeptDeep or Carafe.

For example, if the computer is equipped with an Nvidia GPU, Skyline will configure the local machine to take advantage of the additional computational power, thus, significantly speeding up processing.

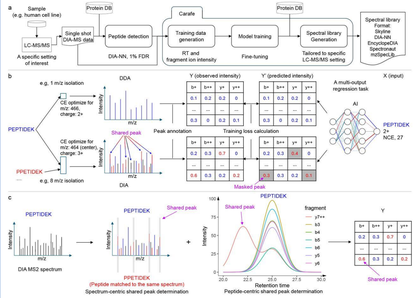
AlphaPeptDeep downloads and runs a pre-trained model to build predicted spectral libraries, while Carafe further improves the performance by allowing fine-tuning of the predictive models using transfer learning based on a second Skyline document or results of DIA-NN provided as input.

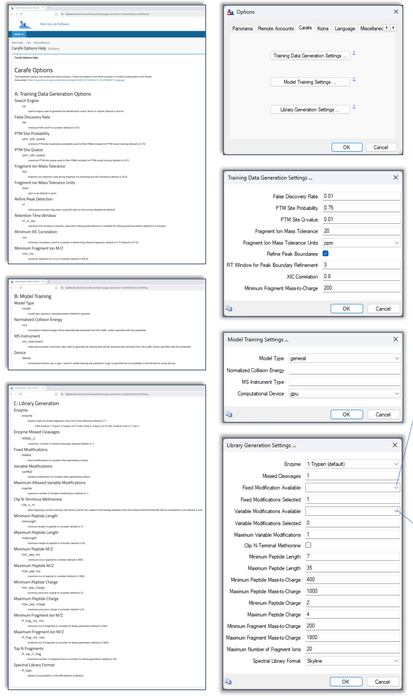
Setup

On a clean slate computer running Skyline for the first time, it is possible to set up the software environment for AI analysis in less than 10 to 15 minutes, and this task only has to occur once per Skyline installation. Setting up Nvidia GPU software adds a nominal amount of time to the process.

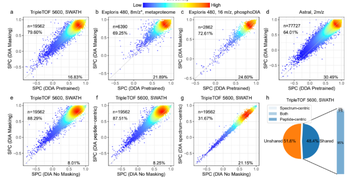


Carafe Options

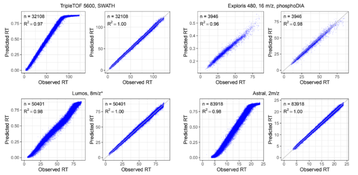




Carafe Fine-Tuning

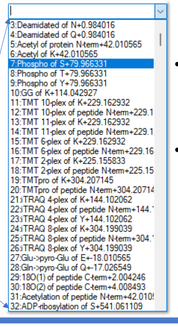


- Fine-tuning with DIA data improves fragment ion intensity predictions



- Fine-tuning with DIA data improves retention time predictions

Unimod PTM Support

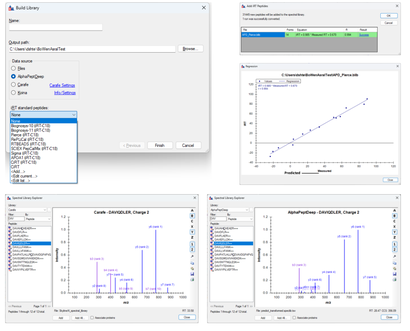


- Provides support for 1,069 different peptide modifications as defined by Unimod
- Interface supports filtering and search by text or modification mass to facilitate rapid selection of relevant PTMs

Conflict of Interest Statement

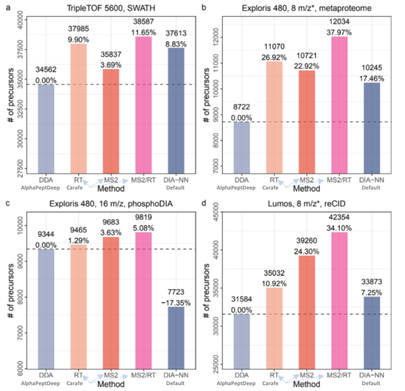
Five major mass spectrometry vendors (Agilent, Bruker, Shimadzu, ThermoFisher, and Waters) provide financial support for the development and maintenance of Skyline. Michael J. MacCoss is a paid consultant for Thermo Fisher Scientific.

Skyline Results



Carafe and AlphaPeptDeep are used to predict AI spectral libraries inside Skyline. Spectral Library Explorer visualizes the predicted fragment spectra. Predictions include Indexed Retention Times (iRT) and Collisional Cross Sections (CCS) for peptides.

Predictive Model Comparisons



Fine-tuning improves peptide detection in DIA data. DDA: AlphaPeptDeep pretrained model; MS2: fine-tuned fragment ion model; RT: fine-tuned RT model; MS2/RT: fine-tuned fragment ion and RT models; DIA-NN: library generated using DIA-NN on DDA data

Conclusions

- Skyline locally integrates support for building AI spectral libraries.
- AlphaPeptDeep and Carafe are supported locally.
- AlphaPeptDeep is currently available
  - Skyline-daily version 25.1.1.147
- Carafe support is available as a “preview” with full support to be provided soon after ASMS.
- Peptides for which spectra are generated stay on the local computer and are not exposed to another computer on the internet.
- Setup of software dependencies and components necessary for analysis is done automatically with minimal effort on the part of the end-user.
- Skyline local AI spectral library prediction can be performed utilizing the GPU on computers equipped with Nvidia GPU components.
- On computers not equipped with Nvidia GPU, processing will occur on the CPU automatically.
- Depending on the hardware, CPU processing takes about 10x longer than GPU processing.
- Carafe applies fine-tuning by accepting as input a second Skyline document.
- Carafe via Skyline can also fine-tune based on input from a report generated by the DIA-NN search engine.
- Carafe user-facing options are exposed with a simple to understand interface having preset defaults.
- Skyline AI spectral library predictions provide support for Unimod post-translational modifications.
- AlphaPeptDeep peptide predictions in Skyline include fragment ion intensities, retention times and collisional-cross sections.
- Carafe peptide predictions allow fine tuning of fragment ion intensities and retention times based on user’s previously collected data.

References

Wen B, Hsu C, Zeng WF, Riffle M, Chang A, Mudge M, Nunn B, Berg MD, Villén J, MacCoss MJ, Noble WS. Carafe enables high quality in silico spectral library generation for data-independent acquisition proteomics. bioRxiv [Preprint]. 2024 Oct 18:2024.10.15.618504. doi: 10.1101/2024.10.15.618504. PMID: 39463980; PMCID: PMC11507862.

Zeng WF, Zhou XX, Willems S, Ammar C, Wahle M, Bludau I, Voytik E, Strauss MT, Mann M. AlphaPeptDeep: a modular deep learning framework to predict peptide properties for proteomics. Nat Commun. 2022 Nov 24;13(1):7238. doi: 10.1038/s41467-022-34904-3. PMID: 36433986; PMCID: PMC9700817.