
SpecGlobTool documentation

version: 0.1

Olivier Langella

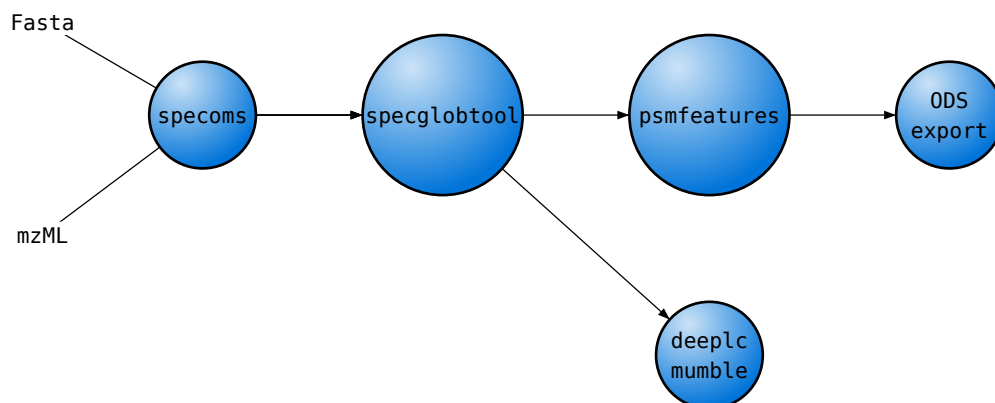
Contents

Introduction	3
PSM CBOR file workflow	3
Command line interface	3
JSON parameter file	3
Tests for SpecGlobTool	3
SpecGlobTool PSM score using TIDD	4
Bibliography	4

Introduction

SpecGlobTool aligns MS2 fragmentation spectra and a peptide sequence. It finds the best explanation of experimental peaks given a peptide sequence and is able to deal with precursor mass delta. Algorithm is described in (Prunier et al., 2023).

PSM CBOR file workflow



Command line interface

SpecGlobTool works on a PSM CBOR file as input and produces a new PSM CBOR file as output.

```
specglobtool -p params.json -i input.cbor -o output.cbor
```

SpecGlobTool works also using UNIX pipes

```
cat input.cbor | specglobtool -p params.json > output.cbor
```

JSON parameter file

SpecGlobTool parameters are specified in a JSON file

JSON parameter file

```
{
  "fragment_tolerance_unit": "dalton",
  "fragment_tolerance": 0.02,
  "spectrum": {
    "deisotope": true,
    "minimum_mz": 150,
    "n_most_intense": 120
  }
}
```

Tests for SpecGlobTool

First you need a PSM CBOR file. Currently, you have to use SpecOMS

```
specoms -p /gorgone/pappso/versions_logiciels_pappso/deepprot/v0.0.11/B73/
ion_enhancer_params.ods -m /gorgone/pappso/moulon/
raw/20250424_deepprot_langella_mzml/20200218_RD_mais_deepprot_1.mzML /gorgone/pappso/moulon/
users/archives_users/Luciana/deepprot_project/luciana/maize_genotypes/no_redundant_db/
```

```
panAnnotation.B73.v1.0.d12.19.fasta.transdecoder.pep.renamed.NR.fa /gorgone/pappso/moulon/  
database/contaminants_standarts.fasta -o /gorgone/pappso/moulon/users/  
Olivier/20250402_RD_mais_deepprot/specoms/specoms_1_F2_vs_B73.cbor -a
```

Run specglobtool on existing cbor PSM file :

```
specglobtool -i /gorgone/pappso/moulon/users/Olivier/20250402_RD_mais_deepprot/specoms/  
specoms_1_F2_vs_B73.cbor -o new_specoms_1_F2_vs_B73.cbor -c 10 -p ../doc/typst/parameters.json
```

Export results to JSON file :

```
specglobtool-export -i new_specoms_1_F2_vs_B73.cbor --json new_specoms_1_F2_vs_B73.json
```

Export results to ODS file :

```
specglobtool-export -i new_specoms_1_F2_vs_B73.cbor --ods new_specoms_1_F2_vs_B73.ods
```

SpecGlobTool PSM score using TIDD

Computing PSM features to filter specglobtool results

```
psmfeatures -i new_specoms_1_F2_vs_B73.cbor -o new_specoms_1_F2_vs_B73_with_features.cbor -c  
10
```

Parameters to compute features can be described in a JSON file :

JSON parameter file to compute features

```
{  
  "fragment_tolerance": 0.02,  
  "fragment_tolerance_unit": "dalton",  
  "ion_list": "y b",  
  "minimum_mz": 150,  
  "hyperscore": {  
    "dynamic_range": 100,  
    "minimum_mz": 150,  
    "n_most_intense": 100  
  }  
}
```

Export PSM features for SVM process

```
psmfeatures-export -i new_specoms_1_F2_vs_B73_with_features.cbor --tsvd svm_F2_vs_B73
```

This command creates a directory called "svm_F2_vs_B73" containing data and Rscript to run the Support Vector Machine algorithm.

Bibliography

Prunier, G., Cherkaoui, M., Lysiak, A., Langella, O., Blein-Nicolas, M., Lollier, V., et al. (2023). Fast alignment of mass spectra in large proteomics datasets, capturing dissimilarities arising from multiple complex modifications of peptides. *BMC bioinformatics* 24, 421. doi: 10.1186/s12859-023-05555-y