

---

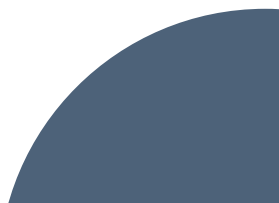
# PeptidOMS documentation

## PeptidOMS-ANR-24-CE45-3296

11/07/2025 - 25/03/2026

---

Olivier Langella  
PAPPSO  
2025-2026  
*technical documentation*



## Table of content

1	Introduction .....	3
2	Source code .....	3
3	PSM CBOR file workflow .....	3
4	Command line interface .....	3
5	JSON parameter file .....	3
6	CBOR peptidoms eval PSM structure .....	4
7	Tests for Spoms .....	4
7.1	Smaller dataset .....	4
8	Tests for PeptidOMS .....	4
8.1	Smaller dataset .....	4
9	Tests for ProteinMatcher .....	5
9.1	Smaller dataset .....	5
9.2	C++ vs Java issue .....	5
10	Close look on AIADGSLLDLLR .....	8
11	Benchmark specpetidoms vs spoms fast alignment .....	8
12	debug Spoms .....	9
	Bibliography .....	9

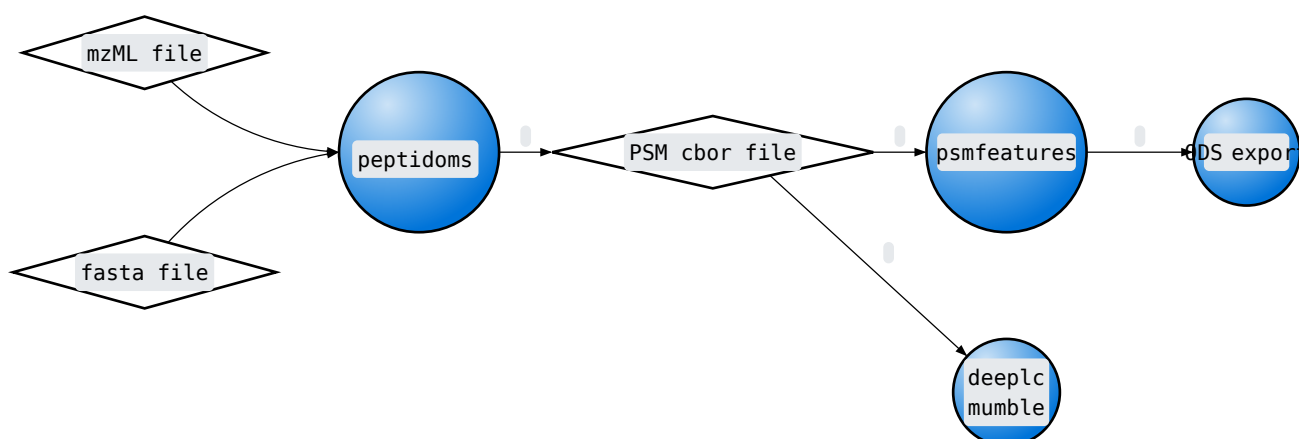
## 1 Introduction

PeptidOMS aligns MS2 fragmentation spectra directly to protein sequences. Algorithm is described in (Benoist et al., 2025).

## 2 Source code

Current source code is available at <https://forge.inrae.fr/peptidoms/specpeptidoms>.

## 3 PSM CBOR file workflow



## 4 Command line interface

PeptidOMS works on a collection of MS/MS spectra (mzML file typically) and a protein sequence database (FASTA file) and produces a PSM CBOR file as output.

```
peptidoms -p params.json -m mzdata.mzml -f protein_sequences.fasta -o results.cbor
```

To get a readable file results, use “peptidoms-export” to produce JSON or ODS files:

```
peptidoms-export --ods results.ods -i results.cbor
```

```
peptidoms-export --json results.json -i results.cbor
```

## 5 JSON parameter file

PeptidOMS 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
  },
  "max_interpretations_per_spectrum": 10,
  "decoy_prefix": "rev_"
}
```

## 6 CBOR peptidoms eval PSM structure

PeptidOMS writes results in a PSM CBOR file. For each PSM, PeptidOMS uses the “eval” section to store its scores :

peptidoms PSM eval structure

```
{
  "proforma": "AIADGSLLDLLR",
  "protein_list": [
    {
      "accession": "GRMZM2G083841_P01",
      "position": [429]
    }
  ],
  "eval": {
    "peptidoms": {
      "bracket": "[A][I]ADGSLLDLLR",
      "spc": 12,
      "score": 76,
      "nam": 0
    }
  }
}
```

## 7 Tests for Spoms

```
./src/spoms -
p ../doc/typst/json/spoms_parameters_test.json -f /gorgone/pappso/moulon/database/Genome_Z_mays_5a.fasta -m /
gorgone/pappso/data_extraction_pappso/mzML/20120906_balliau_extract_1_A01_urnb-1.mzML -o /tmp/test.cbor
```

```
./src/spoms -
p ../doc/typst/json/spoms_parameters_test.json -f /gorgone/pappso/moulon/database/Genome_Z_mays_5a.fasta -m /
gorgone/pappso/data_extraction_pappso/mzcbor/20120906_balliau_extract_1_A01_urnb-1.mzcbor -o /tmp/test.cbor
```

Process for spectrum index 2406 finished. Slowest one is 2287 2290  
total 236

### 7.1 Smaller dataset

```
./src/spoms -p ../doc/typst/json/spoms_parameters_test.json -f ../tests/data/fasta/small_zea_mays.fasta -m ../
tests/data/peaklists/peaklist_15046.mgf -o /tmp/test.cbor
```

```
./src/spoms -p ../tests/data/parameters/DATNVGDEGGFAPNIIENK_parameters.json -f ../tests/
data/fasta/databankThreeProteinsHuman.fasta -m ../tests/data/peaklists/DATNVGDEGGFAPNIIENK.mgf -o
DATNVGDEGGFAPNIIENK.cbor -c 10
```

## 8 Tests for PeptidOMS

```
./src/peptidoms -
p ../doc/typst/json/peptidoms_parameters.json -f /gorgone/pappso/moulon/database/Genome_Z_mays_5a.fasta -m /
gorgone/pappso/data_extraction_pappso/mzML/20120906_balliau_extract_1_A01_urnb-1.mzML -o /tmp/test.cbor
```

### 8.1 Smaller dataset

```
./src/peptidoms -p ../doc/typst/json/peptidoms_parameters.json -f ../tests/data/fasta/small_zea_mays.fasta -
m ../tests/data/peaklists/peaklist_15046.mgf -o /tmp/test.cbor
```

```
./src/peptidoms -p ../tests/data/parameters/DATNVGDEGGFAPNIIENK_parameters.json -f ../tests/data/fasta/databankThreeProteinsHuman.fasta -m ../tests/data/peaklists/DATNVGDEGGFAPNIIENK.mgf -o DATNVGDEGGFAPNIIENK.cbor -c 10
```

## 9 Tests for ProteinMatcher

```
./src/proteinmatcher -p ../doc/typst/json/proteinmatcher_parameters.json -f /gorgone/pappso/moulon/database/Genome_Z_mays_5a.fasta -m /gorgone/pappso/data_extraction_pappso/mzML/20120906_balliau_extract_1_A01_urnb-1.mzML -o /tmp/test.cbor
```

good match “index”: 2605, “native\_id”: “controllerType=0 controllerNumber=1 scan=2606” “proforma”: “MESGT-GNGDSEVQR”,

TPVDNALR on index 4503 should be a good match

QQVM[MOD:00425]VGYSDSGK on index 4291

“accession”: “GRMZM2G326111\_P01”, “positions”: [ “eval”: { “matcher”: { “score”: 107855 }, “peptidoms”: { “bracket”: “[G][N]GTGGESIYG EK”, “nam”: 0, “score”: 69, “spc”: 11 } }

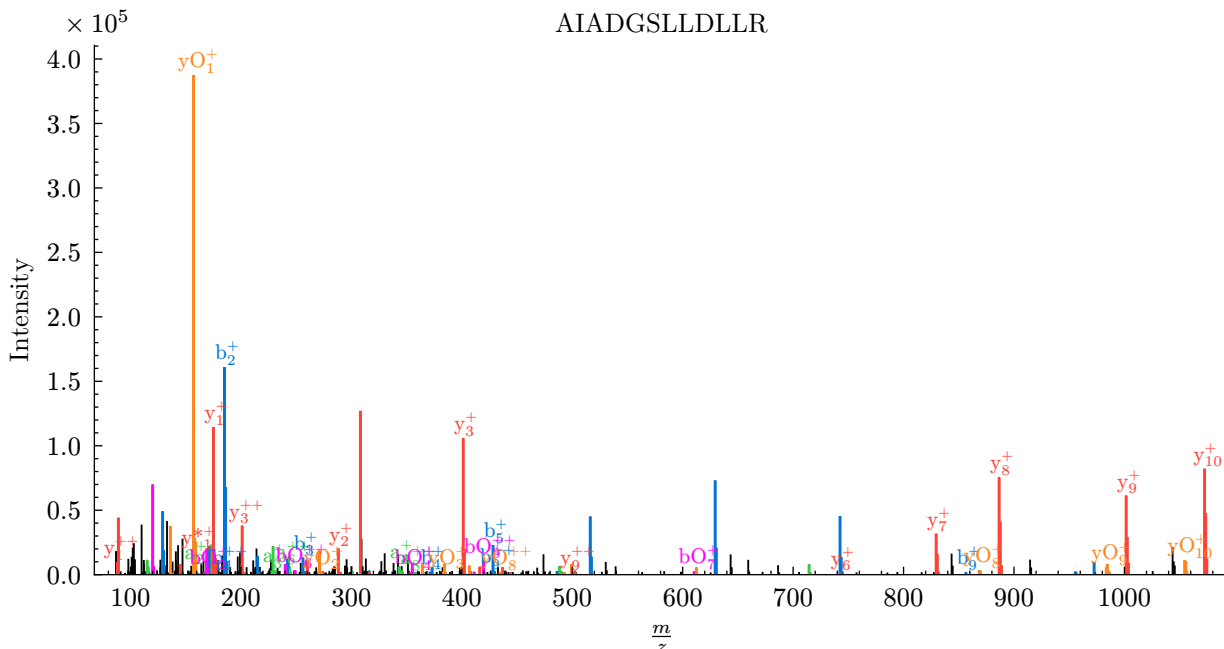
### 9.1 Smaller dataset

```
./src/proteinmatcher -p ../doc/typst/json/proteinmatcher_parameters.json -f ../tests/data/fasta/small_zea_mays.fasta -m ../tests/data/peaklists/peaklist_15046.mgf -o /tmp/test.cbor
```

### 9.2 C++ vs Java issue

Some important issue remains in the C++ version.

For example on a small Zea mays database (tests/data/fasta/small\_ze\_mays.fasta) and a single spectra (test/data/peaklists/peaklist\_15046.mgf), the java version can successfully identify AIADGSLLDLLR from GRMZM2G083841\_P01



JAVA results :

```
spectrum title;spectrum scan;spectrum id;peptide;protein positions;alignment;score;number of shared
peaks;post-processed non-aligned mass;post-processed peptide;post-processed protein positions;post-processed
alignment;post-processed score;post-processed number of common peaks
Elution from: 0.03 to 0.03 period: 0 experiment: 1 cycles: 1 AIADGSLLDLLR
GRMZM2G083841_P01;-1;0;AIADGSLLDLLR;[429,440];[A][I]ADGSLLDLLR;76;14;;;;;;;;;
```

---

Elution from: 0.03 to 0.03    period: 0    experiment: 1 cycles: 1 AIADGSLLDLR  
GRMZM2G083841\_P01;-1;0;GSSEALIR;[188,195];[182,11]GSSEAL[242,13]IR;44;11;;;;;;;;;

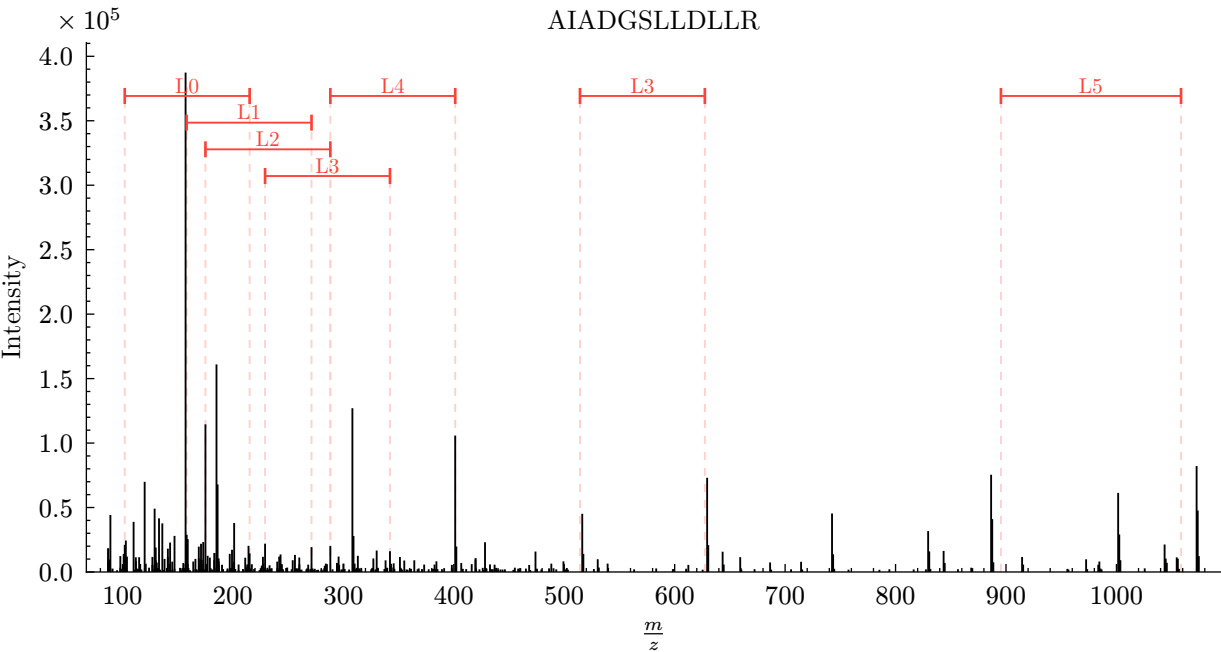
C++ results

```

{
  "psm_list": [
    {
      "proforma": "[+469.2425]?RTELLR",
      "protein_list": [
        { "accession": "GRMZM2G060167_P01", "positions": [235] }
      ],
      "eval": {
        "peptidoms": {
          "bracket": "[R]TELLR",
          "nam": 0,
          "score": 35,
          "spc": 6
        }
      }
    },
    {
      "proforma": "[+165.0446]-FLRLLR",
      "protein_list": [
        { "accession": "GRMZM2G322129_P01", "positions": [561] }
      ],
      "eval": {
        "peptidoms": {
          "bracket": "[165.045]FLRLLR",
          "nam": 165.0445575331208,
          "score": 34,
          "spc": 7
        }
      }
    },
    {
      "proforma": "[+241.0690]-ALRLLR",
      "protein_list": [
        { "accession": "GRMZM2G121981_P01", "positions": [402] }
      ],
      "eval": {
        "peptidoms": {
          "bracket": "[241.069]ALRLLR",
          "nam": 241.06904753312082,
          "score": 34,
          "spc": 7
        }
      }
    },
    {
      "proforma": "[+168.0898]-GTTWL[+242.1290]LR",
      "protein_list": [{ "accession": "GRMZM2G322468_P01", "positions": [60] }],
      "eval": {
        "peptidoms": {
          "bracket": "[168.09]GTTWL[242.129]LR",
          "nam": 410.0898455331213,
          "score": 26,
          "spc": 7
        }
      }
    },
    {
      "proforma": "[+349.1297]-GSL[+99.0689]ELIR",
      "protein_list": [
        { "accession": "GRMZM2G119921_P01", "positions": [160] }
      ],
      "eval": {
        "peptidoms": {
          "bracket": "[349.1297]GSL[99.0689]ELIR",
          "nam": 349.12965000000002,
          "score": 26,
          "spc": 7
        }
      }
    }
  ]
}

```

10 Close look on AIADGSLLDLLR



11 Benchmark specpetidoms vs spoms fast alignment

2026-02-20

```
/tests/data/fasta/very_small_zea_mays.fasta

langella@milano:~/developpement/git/peptidoms/specpeptidoms/build$ ./tests/benchmark/catch2-benchmark-tests
[specpeptidoms] -s --benchmark-samples 1
Filters: [specpeptidoms]
Randomness seeded to: 226359403
/home/langella/developpement/git/peptidoms/specpeptidoms/tests/common.cpp readMgf 58 1

~~~~~
catch2-benchmark-tests is a Catch2 v3.7.1 host application.
Run with -? for options

-----
cbor psm map test suite.
...: PSM cbor features:...

-----
/home/langella/developpement/git/peptidoms/specpeptidoms/tests/benchmark/benchmark_alignment.cpp:80
.....

benchmark name                samples      iterations    est run time
                                mean          low mean      high mean
                                std dev      low std dev   high std dev
-----
fastAlign on all protein      1              1             6.03099 s
                                15.7746 s    15.7746 s    15.7746 s
                                0 ns         0 ns         0 ns

/home/langella/developpement/git/peptidoms/specpeptidoms/tests/benchmark/benchmark_alignment.cpp:153:
PASSED:
  REQUIRE( locations.size() == 5 )
with expansion:
  5 == 5

benchmark name                samples      iterations    est run time
```



	mean std dev	low mean low std dev	high mean high std dev
locations	1 3.46705 s 0 ns	1 3.46705 s 0 ns	3.46423 s 3.46705 s 0 ns

```
=====
All tests passed (1 assertion in 1 test case)
langella@milano:~/developpement/git/peptidoms/specpeptidoms/build$ ./tests/benchmark/catch2-benchmark-tests
[spoms] -s --benchmark-samples 1
Filters: [spoms]
Randomness seeded to: 3013200688
/home/langella/developpement/git/peptidoms/specpeptidoms/tests/common.cpp readMgf 58 1
```

~~~~~
catch2-benchmark-tests is a Catch2 v3.7.1 host application.
Run with -? for options

-----
cbor psm map test suite.
...: PSM cbor features:...

-----
/home/langella/developpement/git/peptidoms/specpeptidoms/tests/benchmark/benchmark\_alignment\_ng.cpp:84
.....

| benchmark name           | samples<br>mean<br>std dev | iterations<br>low mean<br>low std dev | est run time<br>high mean<br>high std dev |
|--------------------------|----------------------------|---------------------------------------|-------------------------------------------|
| fastAlign on all protein | 1<br>17.2535 ms<br>0 ns    | 1<br>17.2535 ms<br>0 ns               | 18.1479 ms<br>17.2535 ms<br>0 ns          |

=====
test cases: 1 | 1 passed
assertions: - none -

12 debug Spoms

```
langella@piccolo:~/developpement/git/peptidoms/specpeptidoms/build$ ./src/spoms -p ../doc/typst/json/
spoms_parameters_test.json -f /gorgone/pappso/moulon/database/Genome_Z_mays_5a.fasta -m /gorgone/pappso/
data_extraction_pappso/mzML/20120906_balliau_extract_1_A01_urnb-1.mzML -o /tmp/test.cbor
...
unknown@0, unknown(): start spectrum index: 504    accession: "GRMZM2G466358_P01"
Erreur de segmentation
```

Bibliography

Benoist, É., Jean, G., Rogniaux, H., Fertin, G., and Tessier, D. (2025). SpecPeptidOMS Directly and Rapidly Aligns Mass Spectra on Whole Proteomes and Identifies Peptides That Are Not Necessarily Tryptic: Implications for Peptidomics. Journal of Proteome Research 24, 2159–2172. doi: [10.1021/acs.jproteome.4c00870](https://doi.org/10.1021/acs.jproteome.4c00870)