

<https://msda.unistra.fr>

MSDA tutorial: How to read *de novo* results

The *de novo* pipeline will generate for each peaklist one MSBlast html result file, plus one Excel file containing

- An overview of the search: settings, date, related informations...
- The PepNovo+ results
- The MSBlast results exported from the html file

All these files will be packaged in a single zipped file.

Results from PepNovo+

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|----|--------------------------------|----------------------|----------|---------|--------|------------|---------------|---------|--------|-------------------------------------|--------------------------|-------------------------|---------------------------------------|
| | Spectrum name | PepNovo sequence tag | [M+H] | M/z | Charge | Rank score | PepNovo Score | N-Gap | C-Gap | PepNovo extended sequence for Blast | PepNovo+ 'MSBlast score' | Will be sent to MSBlast | Comments |
| 1 | | | | | | | | | | | | | |
| 2 | Cmpd 1, +MSn(519.62), 0.0 min | | | | | | | | | | | | #could not read spectrum correctly... |
| 3 | Cmpd 2, +MSn(311.02), 0.1 min | | | | | | | | | | | | #could not read spectrum correctly... |
| 4 | Cmpd 3, +MSn(387.32), 0.1 min | LVR | 387,252 | 387,252 | 1 | -1,06 | 32 | 0 | 0 | BLVR | -1000000000 | no | |
| 5 | Cmpd 4, +MSn(701.70), 0.2 min | | | | | | | | | | | | #could not read spectrum correctly... |
| 6 | Cmpd 5, +MSn(650.23), 0.2 min | | | | | | | | | | | | #could not read spectrum correctly... |
| 7 | Cmpd 6, +MSn(530.09), 0.3 min | QGLQQESLR | 1058,549 | 529,778 | 2 | 5,014 | 35,918 | 0 | 0 | BANXXXXXELSR | 3,02774 | no | |
| 8 | Cmpd 7, +MSn(553.77), 0.3 min | FLSW | 552,282 | 552,282 | 1 | 0,361 | -29,002 | 0 | 0 | BLFLC | -1000000000 | no | |
| 9 | Cmpd 8, +MSn(832.31), 0.3 min | | | | | | | | | | | | #could not read spectrum correctly... |
| 10 | Cmpd 9, +MSn(521.44), 0.4 min | GAQEAEER | 1041,539 | 521,273 | 2 | 7,574 | 43,339 | 281,126 | 0 | BXXAQQEAEER | 4,66799 | yes | |
| 11 | Cmpd 10, +MSn(680.09), 0.4 min | QALGGPES | 1358,683 | 679,845 | 2 | 0,731 | 61,968 | 0 | 619,33 | BAQLGGPESAXXXXXXX | 5,91309 | yes | |
| 12 | Cmpd 11, +MSn(673.66), 0.4 min | PASYAQR | 1345,675 | 673,341 | 2 | -0,121 | -0,614 | 553,262 | 0 | BXXXXXXVPASSFKR | 3,19684 | no | |
| 13 | Cmpd 12, +MSn(397.27), 0.5 min | EEKMK | 793,415 | 397,211 | 2 | 3,016 | 31,53 | 0 | 0 | BEEKMK | 1,23434 | no | |
| 14 | Cmpd 13, +MSn(595.64), 0.5 min | LNTLAFK | 1189,599 | 595,303 | 2 | 1,564 | 26,046 | 383,179 | 0 | BXXLQNVDMK | 3,45963 | no | |

Every PepNovo+ results are stored into an Excel spectrum report. The columns are the following:

| | |
|-------------------------------------|--|
| Spectrum name | The title of the spectrum used by PepNovo+ to determined the sequence tag |
| PepNovo sequence tag | The tag determined by PepNovo+ with PTMs included |
| [M+H] | The mass in Dalton |
| M/z | The mass divided by the charge |
| Charge | The charge |
| Rank score | Relative score used by PepNovo+ to order the computed solutions |
| PepNovo Score | Score representing the quality of the spectrum |
| N-Gap | Mass of unknown amino acids at the left of the sequence tag (0 when peptide is N-terminal) |
| C-Gap | Mass of unknown amino acids at the right of the sequence tag (0 when peptide is C-terminal) |
| PepNovo extended sequence for Blast | Sequence formatted for a MSBlast process with X's instead of missing amino acids, tags for N-terminal and/or C-terminal peptides, etc. The selected sequence may be different than the PepNovo sequence tag ! |
| PepNovo+ 'MSBlast score' | Score calculated by PepNovo+, a high score will indicate that the sequence has a good probability to match something in the following MSBlast process |
| Will be sent to MSBlast | Yes or No |
| Comments | Output message when PepNovo+ is unable to get something from a spectrum |

<https://msda.unistra.fr>

Results from MSBlast

MSBlast returns an html file, where positive hits are exported by MSDA into an Excel peptide report.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q |
|----|--------------------------|--------------|-----------------------|-------------|--------------|-----------------------|-------------------------------------|---|---------------|-----------------|-----------------------|-----------------|---------------------------|--------------------|-------------------|-----------------------------|--------------------------------------|
| | Protein accession number | Protein name | Description | Gene symbol | Organism | MSBlast protein score | PepNovo extended sequence for Blast | PepNovo sequence found in these spectra | MSBlast query | MSBlast subject | MSBlast peptide score | Sequence length | Query compared to Subject | MSBlast Identities | MSBlast Positives | Max consecutive amino acids | Max consecutive amino acids (strict) |
| 2 | sp P35232 PHB_HUMAN | PHB_HUMAN | PHB_HUMAN Prohibi PHB | | Homo sapiens | 746 | BXFISLGEDYEER | Cmpd 99, +MSn(738.11), 5.5 min | BXFISLGEDYEER | RIFTSIGEDYDER | 82 | 13 | 13 + FTS+GEDY+ER | 9 | 12 | 11 | 4 |
| 3 | sp P35232 PHB_HUMAN | PHB_HUMAN | PHB_HUMAN Prohibi PHB | | Homo sapiens | 746 | BAALLSAEGDSK | Cmpd 18, +MSn(1061.97), 0.8 min | BAALLSAEGDSK | KAAHSAEGDSK | 74 | 12 | 12 +AA+SAEGDSK | 9 | 12 | 12 | 7 |
| 4 | sp P35232 PHB_HUMAN | PHB_HUMAN | PHB_HUMAN Prohibi PHB | | Homo sapiens | 746 | RQLAAEDSLAYXXXXR | Cmpd 23, +MSn(804.35), 6.4 min | RQLAAEDSLAY | RKLEAAEDSLAY | 64 | 11 | 11 +LEAAEDSLAY | 8 | 10 | 9 | 6 |
| 5 | sp P35232 PHB_HUMAN | PHB_HUMAN | PHB_HUMAN Prohibi PHB | | Homo sapiens | 746 | BPLFLFDCR | Cmpd 96, +MSn(460.88), 5.3 min | BPLFLFDCR | KPHIFDCR | 58 | 8 | 8 +P++FDCR | 5 | 8 | 8 | 4 |
| 6 | sp P35232 PHB_HUMAN | PHB_HUMAN | PHB_HUMAN Prohibi PHB | | Homo sapiens | 746 | BXKAQQEAEER | Cmpd 9, +MSn(521.44), 0.4 min | BXKAQQEAEER | KQVAQQEAEER | 57 | 10 | 10 + AQQEAEER | 7 | 8 | 7 | 7 |
| 7 | sp P35232 PHB_HUMAN | PHB_HUMAN | PHB_HUMAN Prohibi PHB | | Homo sapiens | 746 | BNNQNVNLTIR | Cmpd 112, +MSn(593.61), 6.0 min | BNNQNVNLTIR | KDLQNVNLTIR | 57 | 11 | 11 ++ QNVN+LTIR | 7 | 10 | 8 | 4 |
| 8 | sp P35232 PHB_HUMAN | PHB_HUMAN | PHB_HUMAN Prohibi PHB | | Homo sapiens | 746 | BOFAGELLTQR | Cmpd 116, +MSn(575.81), 6.1 min | AGELLTQR | AGELTQR | 53 | 8 | 8 AGEL+TQR | 7 | 8 | 8 | 4 |
| 9 | sp P35232 PHB_HUMAN | PHB_HUMAN | PHB_HUMAN Prohibi PHB | | Homo sapiens | 746 | BXVASQLPIR | Cmpd 19, +MSn(434.30), 6.8 min | BXVASQLPIR | RPVASQLPIR | 53 | 9 | 9 + VASQLPIR | 7 | 8 | 7 | 7 |
| 10 | sp P35232 PHB_HUMAN | PHB_HUMAN | PHB_HUMAN Prohibi PHB | | Homo sapiens | 746 | BVFESGSK | Cmpd 29, +MSn(390.60), 1.4 min | BVFESGSK | KVFESGSK | 53 | 8 | 8 +VFESGSK | 6 | 8 | 8 | 4 |
| 11 | sp P35232 PHB_HUMAN | PHB_HUMAN | PHB_HUMAN Prohibi PHB | | Homo sapiens | 746 | BXXXXXXGDGLLEIR | Cmpd 175, +MSn(667.51), 8.1 min | GDGLLEIR | GDGLLEIR | 51 | 8 | 8 GDGL+LEIR | 7 | 8 | 8 | 4 |

The columns are the following:

| | |
|---|---|
| Protein accession number | Accession number of the protein retrieved in the fasta file |
| Protein name | Protein name corresponding to the accession number |
| Description | Full description of the protein |
| Gene symbol | The corresponding gene symbol |
| Organism | Organism where the protein belong |
| MSBlast protein score | Corresponds to the protein "Total Score" value |
| PepNovo extended sequence for Blast | Input sequence tag, it can be retrieved in the PepNovo+ sheet |
| PepNovo sequence found in these spectra | List of spectra corresponding to this sequence, according to PepNovo+ results |
| MSBlast query | Corresponds to the "Query" sequence |
| MSBlast subject | Corresponds to the "Subject" sequence |
| MSBlast peptide score | Corresponds to the peptide "Score" value |
| Sequence length | Number of amino acids in the subject |
| Query compared to Subject | Corresponds to the line between Query and Subject |
| MSBlast Identities | Corresponds to the "Identities" numerator: number of strict amino acids matches (ie. I and L won't match) |
| MSBlast Positives | Corresponds to the "Positives" numerator: number of amino acids matches (ie. I and L will match) |
| Max consecutive amino acids | Size of the longest subsequence made of common matches |
| Max consecutive amino acids (strict) | Size of the longest subsequence made of strict common matches |

This is an example of the html output, most of the information is stored in the Excel file.

```

^ = sp|P35232|PHB\_HUMAN Prohibitin OS=Homo sapiens GN=PHB PE=1 SV=1
      Length = 272

Total Score: 746

      0          60          120          180          240
      |          |          |          |          |
sp|P35232|PHB_HUMAN |-----|-----|-----|-----|-----|272
Local hits (HSPs)  |-----|-----|-----|-----|-----|

Score = 82 (43.0 bits)
Identities = 9/13 (69%), Positives = 12/13 (92%)

Query:   14 BXFISLGEDYEER 26
         + FTS+GEDY+ER
Sbjct:  105 RIFTSIGEDYDER 117
    
```