MS Amanda Standalone
User Manual

February 2014
Viktoria Dorfer, Peter Pichler, Stephan Winkler, and Karl Mechtler

1. Installation of MS Amanda Standalone

In order to install MS Amanda Standalone please download and extract necessary files:

1. Download the MS Amanda Standalone .zip archive from http://ms.imp.ac.at/?goto=msamanda
2. Right click on the downloaded .zip file and select the menu item "Properties" in the context menu
3. If visible, click "Unblock" at the bottom right of the Properties dialog
4. Click "OK" to close the Properties dialog
5. Extract the .zip file

MS Amanda Standalone is now ready for use!

2. Usage of MS Amanda Standalone

MS Amanda Standalone can be used from the command line or called from any already established proteomics pipelines. The syntax of the command line call of MS Amanda is defined as:

```
MSAmanda.exe spectrum.mgf|folder database.fasta settings.xml [output]
```

The parameters that have to be defined are in detail:

spectrum.mgf | folder

The first parameter has to be a single spectrum file in .mgf format or a folder containing .mgf files. If a folder is specified, then every .mgf file in the folder is processed.

database.fasta

The second parameter requires a protein database in .fasta format.

settings.xml

The third parameter must be an .xml file containing all other relevant parameters, see section "3. Settings.xml" for more details.

output

The fourth parameter is optional and specifies the name of the produced output file. If the parameter is not defined, then the output filename will be the name of the spectrum file or the folder name with suffix "_output".

In this output file, all data are given as tab separated values. Additionally, all relevant information about the executed search will be saved in the file "outputfilename_settings.xml".

The output file contains one peptide spectrum match in each line. Reported values comprise the scan number, peptide sequence, protein accessions, MS Amanda score, modifications, rank, and input filename.

3. Settings.xml

The settings.xml file contains all relevant search parameters that can be changed for MS Amanda. The shipped settings.xml file contains the following information:

```
<?xml version="1.0" encoding="utf-8" ?>
<settings>
  <search_settings>
    <enzyme>Trypsin</enzyme>
    <missed_cleavages>2</missed_cleavages>
    <modifications>
      <!-- <modification [fix="true|false"] [nterm="true|false"] [cterm="true|false"]
              [protein="true|false"] [delta_mass="+XX.XX"]>Name[(AA,AA)]</modification>-->
      <modification fix="true" nterm="false"
              cterm="false">Carbamidomethyl(C)</modification>
      <modification nterm="false" protein="false">Oxidation(M)</modification>
      <!--modification delta_mass="17.231">MyOwnModif(K)</modification-->
    </modifications>
    <instrument>b,y</instrument>
    <!-- <ms1_tol unit="Da|ppm|mmu">VALUE</ms1_tol>  default: "5 ppm"-->
    <ms1_tol unit="ppm">5</ms1_tol>
    <!-- <ms2_tol unit="Da|ppm|mmu">VALUE</ms1_tol>  default: "0.02 Da"-->
    <ms2_tol unit="Da">0.02</ms2_tol>
    <max_rank>5</max_rank>
    <generate_decoy>false</generate_decoy>
  </search_settings>

  <basic_settings>
    <instruments_file>instruments.xml</instruments_file>
    <unimod_file>unimod.xml</unimod_file>
    <enzyme_file>enzymes.xml</enzyme_file>
    <monoisotopic>true</monoisotopic>
  </basic_settings>
</settings>
```

In detail, these parameters are used in the following way:

<enzyme>

The enzyme that should be used for theoretically digesting the protein database. The name stated here is searched for in the enzyme file. Several enzymes are provided by default, user defined enzymes can be added (see description on the parameter **<enzyme_file>**).

<missed_cleavages>

The maximum number of considered missed cleavages. Valid values range from 0 to 5.

<modifications>

All considered modifications. Both fixed and variable modifications can be specified. Predefined modifications from the unimod file can be selected by specifying the modification name or title, a user defined modification can be used by specifying a user defined name and a delta mass. The syntax of a modification is

```
<modification [fix="true|FALSE"] [nterm="true|FALSE"] [cterm="true|FALSE"]
    [protein="true|FALSE"] [delta_mass="+00.00"]>Name[(AA1,AA2)]</modification>
```

The attributes *fix*, *nterm*, *cterm*, *protein*, and *delta_mass* are – with some exceptions - optional. Default values are shown in capital letters.

- fix: fixed modification

- nterm: n-terminal modification
- cterm: c-terminal modification
- protein: nterm on protein level
- delta_mass: mass change for self-defined modifications; will be ignored if modification name and position can be found in the unimod file

Amino acids for amino acid specific modifications must be declared next to the modification name with amino acids in brackets, separated by a comma (e.g., Phospho(S,T,Y), Oxidation(M), Methyl(K,R), …).

User defined modifications must contain a delta_mass, a name, and a position (nterm, cterm, or amino acids).

Examples for **valid** modifications:

```
<modification fix="true" nterm="false" cterm="false">Carbamidomethyl(C)</modification>

<modification fix="true" nterm="false">Carbamidomethyl(C)</modification>

<modification fix="true">Carbamidomethyl(C)</modification>

<modification fix="true" nterm="false" cterm="false">Carbamidomethyl(C,E)</modification>

<modification>Oxidation(M)</modification>

<modification fix="false">Oxidation(M)</modification>

<modification protein="true" nterm="true">Acetyl</modification>

<modification delta_mass="+33.45">MyModification(K)</modification>

<modification delta_mass="+33.45">MyModification(K,R)</modification>

<modification delta_mass="+33.45" fix="true">MyModification(K)</modification>

<modification delta_mass="+33.45" nterm="true">MyModification</modification>
```

Modifications cannot be both n-terminal and c-terminal. Protein level modifications are only valid in combination with n-terminal modifications; either an amino acid or a terminal position has to be specified.

Examples for **invalid** modifications:

```
<modification nterm="true" cterm="true">Carbamidomethyl(C)</modification>
<!-- invalid because both nterm and cterm are activated-->

<modification protein="true" nterm="false">Acetyl</modification>
<!-- invalid because protein level must be accompanied by positive nterm-->

<modification protein="true">Acetyl</modification>
<!-- invalid because protein level must be accompanied by positive nterm-->

<modification nterm="true">MyModification</modification>
<!-- invalid because self-defined modification lacks delta_mass-->
```

<instrument>

Fragment ions to be considered in the search. The name stated here is searched for in the instruments file. Several fragment ion combinations are provided by default, user defined combinations can be added (see description of the parameter **<instruments_file>**).

<ms1_tol>

Precursor mass tolerance. Possible units are “Da” (Dalton), “mmu”, and “ppm”, which can be defined in the “unit” attribute. Valid ranges are:

- Da: 0.0 Da to 3.0 Da
- mmu: 0.0 mmu to 3000.0 mmu
- ppm: 0.0 ppm to 3000.0 ppm

<ms2_tol>

Fragment ion mass tolerance. Again, possible units are “Da” (Dalton), “mmu”, and “ppm”, which can be defined in the “unit” attribute. Valid ranges are:

- Da: 0.0 Da to 2.0 Da
- mmu: 0.0 mmu to 2000.0 mmu
- ppm: 0.0 ppm to 2000.0 ppm

<max_rank>

Highest peptide rank reported by MS Amanda. Valid values range from 1 to 10.

<generate_decoy>

Defines whether a decoy database shall be created and searched against. Decoy fasta files are generated by reverting protein sequences, accessions are marked with the prefix “REV_”.

<instruments_file>

File containing all possible fragment ion combinations. Additional combinations can be added as XML nodes:

```
<setting name="b, y">
  <series>b</series>
  <series>y</series>
</setting>
<setting name="b, y, -H2O, -NH3">
  <series>b</series>
  <series>y</series>
  <series>H2O</series>
  <series>NH3</series>
</setting>
```

<unimod_file>

File containing all possible modifications. This file is used for identifying specified modifications. A full list of possible modifications can be downloaded from http://www.unimod.org/.

<enzyme_file>

File containing all possible enzymes. New enzymes can be specified by adding a new xml node:

```
<enzyme>
    <name>Trypsin</name>
    <cleavage_sites>KR</cleavage_sites>
    <inhibitors>P</inhibitors>
    <!-- <position>after|before</position> -->
    <position>after</position>
  </enzyme>
  <enzyme>
    <name>Trypsin/P</name>
    <cleavage_sites>KR</cleavage_sites>
    <position>after</position>
  </enzyme>
```

The **<name>** tag specifies the name of the enzyme, in **<cleavage_sites>** all amino acids indicating a cleavage position shall be named. **<inhibitors>** entitle all cleavage inhibitors, whereas the tag **<position>** specifies whether before or after the site cleavage is triggered.

<monoisotopic>

Defines whether monoisotopic mass values shall be used (in contrast to average mass values).

4. License

5. Contact

This research project is a collaboration of the Protein Chemistry Group at IMP and the Bioinformatics Research Group at FH OÖ, Campus Hagenberg.
For any further questions, bug reports, or ideas please contact Viktoria Dorfer (viktoria.dorfer@fh-hagenberg.at), Peter Pichler (peter.pichler@imp.ac.at), Stephan Winkler (stephan.winkler@fh-hagenberg.at), or Karl Mechtler (karl.mechtler@imp.ac.at) or post your comment in the MS Amanda Google Group (https://groups.google.com/d/forum/msamanda).