Dig2MZ

Table of contents

1. Overview

This application digests fasta proteins and reports all peak's mz, charge and other infos.

2. Usage

```
usage: Dig2MZ <fasta> [-a] [--cyscam] [-d <arg>] [-e <arg>] [-f <arg>] [-h]
[-i <arg>]
       [-L <arg>] [-l <arg>] [-m <arg>] [--non-verbose] [-o <arg>] [-p
<arg>] [-q <arg>]
       [-u <arg>] [-v]
 -a,--average                     set the average mass mode for peptide
mass
                                  calculation
                                  by default: MONOISOTOPIC.
    --cyscam                      modification of all protein's cysteins
                                  by S-carboxamidomethyl cysteines (CysCAM,
+57 Da).
 -d,--delimiter <arg>             define the field delimiter to display
                                  by default: \t.
 -e,--enzymes <arg>               define enzymes that digest proteins
separately with:
                                  enzyme name among 'Caspase-1, Caspase-10,
Glu-C_bicarbonate, Caspase-3,
                                  Thermolysine, Lys-C, Pepsin_pH1.3,
Caspase-8, Glu-C_phosphate, Caspase-9,
                                  Caspase-5, CNBr, ChymoTrypsin_lowspec,
Trypsin, ChymoTrypsin_FYL,
                                  Proteinase-K, Caspase-7, Pepsin_pHgt2,
ChymoTrypsin_highspec,
                                  BNPS-Skatole, Enterokinase, Caspase-6,
Arg-C, Asp-N, ChymoTrypsin_FYLW,
                                  Caspase-4, Caspase-2'.
                                  or custom motifs respecting the following
grammar:
                                  <pre-cut> <cut-token> <post-cut>
                                  <cut-token> := '|'
                                  <pre-cut> := (<AA> or <AA-class>)+
                                  <post-cut> := (<AA> or <AA-class>)+
                                  <AA> := [A-Z]
                                  <AA-class> := '[' AA+ ']'
                                  by default: Trypsin.
 -f,--fields <arg>                define the fields to display (1:MZ,
2:Charge,
                                  3:Enzyme, 4:Seq, 5:MC)
                                  by default: [1, 2, 3, 4, 5].
 -h,--help                        print this message.
 -i,--setting-file <arg>          give a property file with all input
settings.
 -L,--pept-len-filter <arg>       define a filter over length of digested
peptides
```

```
                                    by default: 6.
 -l,--pept-mz-lower-filter <arg>    define the lower mz bound (included) of
                                    digested peptides
                                    by default: 400.
 -m,--mc-max-num <arg>              define the number of maximum missed
cleavages
                                    (for digestion)
                                    by default: 1.
    --non-verbose                   non verbose mode (no header for
settings).
 -o,--oximet-max-num <arg>          define the number of maximum oxidated
                                    methionines
                                    by default: 0.
 -p,--precision <arg>               define the decimal precision for any
                                    mass-to-charge ratio
                                    by default: 6.
 -q,--pept-charge-filter <arg>      define a filter over charges on digested
                                    peptides as a sequence of integers and/or
intervals like in 1, 2, 3:5,
                                    10:7
                                    by default: [2, 3].
 -u,--pept-mz-upper-filter <arg>    define the upper mz bound (included) of
                                    digested peptides
                                    by default: 2000.
 -v,--version                       print the version info.
```

3. Example

```
# a first way to execute the application with lots of options
# the results are redirected in file digests.out and errors in digests.log
$ java -jar Dig2MZ-1.1.jar -a -e Lys-C -f 1,2,4,5 -l 1 -p4 -q 2:4
uniprot-human.fasta > digests.out 2> digests.log

# .. or the more compact way with all options defined in a setting file
$ java -jar Dig2MZ-1.1.jar -i settings.properties uniprot-human.fasta >
digests.out 2> digests.log

$ more digests.out
# ================================================================
# Generated by Dig2MZ v.1.1
#
# Input Proteins ------------------------------------------------
# read from file                uniprot-human_swissprot.fasta
# modified (fixed) with         CYS_CAM
# digested by enzyme            [Lys-C, pattern: K|X, mc#=1]
# Digested Peptides ---------------------------------------------
# filtered with charges         [2, 3, 4]
# filtered with length          >= 1
# filtered over mzs in interval  [400.0, 2000.0]
# with masses computed in mode   AVERAGE
# Output Fields -------------------------------------------------
```

```
# all                             [MZ, Charge, Enzyme, Seq, MC]
# selected indices                [1, 2, 4, 5]
# Real
# Decimal precision format        4
# ===================================================================
MZ Charge Seq MC
1192.3452 3 NDDNAITSPIAGKTSVLRAIPVEVLANSYDISTK 1
894.5108 4 NDDNAITSPIAGKTSVLRAIPVEVLANSYDISTK 1
1355.1842 2 LILSFSLC(C2H3NO)LMVLSC(C2H3NO)SAQLLPWQK 0
658.7105 2 NDDNAITSPIAGK 0
1578.9621 2 MSTKLILSFSLC(C2H3NO)LMVLSC(C2H3NO)SAQLLPWQK 1
1052.9772 3 MSTKLILSFSLC(C2H3NO)LMVLSC(C2H3NO)SAQLLPWQK 1
694.2949 2 AGREGLEWVELK 0
463.1991 3 AGREGLEWVELK 0
651.3932 3 RGGSGRSNGLEQAFC(C2H3NO)NLK 0
488.7968 4 RGGSGRSNGLEQAFC(C2H3NO)NLK 0
556.3634 4 NGRQEVEVFRPFQSRDEK 0
518.5931 2 ERERFSIV 0
895.6627 3 AGREGLEWVELKNDDNAITSPIAGK 1
671.9989 4 AGREGLEWVELKNDDNAITSPIAGK 1
1139.3187 2 TSVLRAIPVEVLANSYDISTK 0
...

# get digested peaks and sort by mz
$ tail -n+19 digests.out | sort -k1 -n | uniq > digests_sorted.out

$ more digests_sorted.out
MZ Charge Seq MC
400.0003 4 ILLEGRRLISDALK 0
400.0006 5 HMEDPLEMERSPQLRK 0
400.0032 4 TAIQQLRSVIRALK 0
400.0057 5 IRQFEEQFERERNSK 0
400.0065 5 AFVYNSSLVSHQEIHHK 0
400.0080 5 RDATHDYRQALATHVNK 0
400.0101 5 ELVERRRTMMEDFRK 0
400.0101 5 PMVNHAEASRLNIERMK 0
400.0117 5 NSPRLRMRTETPSHWK 0
400.0124 5 PQLHSMVARSLC(C2H3NO)RNAAGK 0
400.0138 5 LC(C2H3NO)RLSMQC(C2H3NO)LRDFRIK 0
400.0145 5 HVIIGFSIENSHDRIMK 0
400.0152 5 RLMADELERFTSMRIK 0
400.0161 5 RVTRTGFEDGLFAGWRK 0
400.0167 5 VEQLFGLGLRPRGEC(C2H3NO)HK 0
400.0210 5 DHLTLGTGVAGIDMRRGVK 0
400.0212 5 RAALC(C2H3NO)FRRNLGTYNRK 0
400.0233 5 SIQISHFNPPPPHLRQK 0
400.0240 5 FYASVRC(C2H3NO)DIRRIQALK 0
...
```

4. Releases

The latest version is v1.21 - Download app and the default properties file.

4.1. Rel1.21

Type	Changes
Bug Fix	In missed cleavages mode, modified peptides had weird position shifts.
New	New --cyscam option. The CysCAM fixed modification is now optional.
New	New output field "Mods" that give the number of modifications of the peptide digests.

4.2. Rel1.1

Type	Changes
New	Add two new options to define an interval over mzs *pep-mz-lower-filter* (-l) and *pep-mz-upper-filter* (-u)
Change	Change option name for peptide length filter '-l' -> '-L'.