

RMassBank: Automatic Recalibration and Processing of Tandem HR-MS Spectra for MassBank

Interactive Demonstration

Emma Schymanski,

Michael Stravs, Heinz Singer & Juliane Hollender

Eawag – Department of Environmental Chemistry

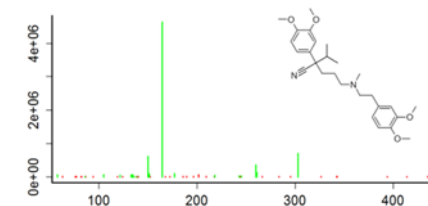
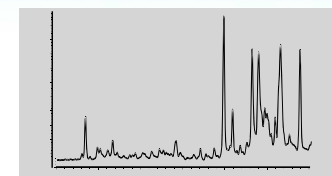
Dübendorf (Zurich), Switzerland



RMassBank Questions: massbank@eawag.ch

Demonstration Overview

- Install (and update) RMassBank and associated programs
- Run RMassBank for trial data
- Get a basic understanding for the workflow
- Understand the manual checking required
 - “Fail peaks” – checking these in raw data
 - Automatic annotation with CTS – editing this data
- Generate records with RMassBank
- Upload these records to MassBank



MassBank Record: EA067410

Home | Spectrum | Quick | Detail | Substructure | Browser | Download | Index | MassBank ID

Verapamil; LC-ESI-ITFT; MS2; 45%; R=15000; [M+H]⁺

Mass Spectrum Chemical Structure

Mass Spectrum: A plot showing a single sharp peak at m/z 354.1. The y-axis is relative intensity (0 to 100) and the x-axis is m/z (0 to 400).
 Chemical Structure: A skeletal structure of Verapamil, a calcium channel blocker, with a central carbon atom bonded to a methyl group, a propyl group, and two 2-methoxyphenyl groups.

ACCESSION: EA067410
 RECORD_TITLE: Verapamil; LC-ESI-ITFT; MS2; 45%; R=15000; [M+H]⁺
 DATE: 2012-09-09
 AUTHOR: Steffy M, Schymanski E, Singer H, Department of Environmental Chemistry, Eawag
 LICENSE: http://massbank.org/MSBank/File/License.html
 COPYRIGHT: Copyright (C) 2011 Eawag, Dübendorf, Switzerland
 COMMENT: CONFIDENCE: standard compound
 COMMENT: RANKO_SCORE_ID: 474

Starting RMassBank

Open up RStudio (or R console)

- o And open up “RMassBank_Demo.R”

The screenshot shows the RStudio interface with the RMassBank_Demo.R script open in the editor and the Package list pane on the right.

Script Content (RMassBank_Demo.R):

```

1 # Interactive RMassBank Demonstration
2 # E. Schymanski, Eawag, 27/11/2012
3
4 # First, check that you have the latest stable RMassBank installed
5 source("http://bioconductor.org/biocLite.R")
6 # Identify packages requiring update
7 old.packages(repos=biocinstallRepos())
8 # update all packages that are out of date
9 biocLite(character(), ask=FALSE)
10 # or just install RMassBank (does the rest for you)
11 source("http://bioconductor.org/biocLite.R")
12 biocLite("RMassBank")
13 biocLite("RMassBankData")
14 # Careful, if you haven't updated for a while this can
15 # require update of several related packages (several minutes)
16

```

Console Output:

```

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Loading required package: rcom
Loading required package: rscproxy
[Workspace loaded from ~/.RData]

Warning message:
Class "C++Object" is defined (with package slot 'Rcpp') but no metadata object found
to revise subclass information---not exported? Making a copy in package '.GlobalEnv'
>

```

Package List:

Package Name	Description
abind	Combine multi-dimensional arrays
acepack	ace() and avas() for selecting regression transformations
actuar	Actuarial functions
ada	ada: an R package for stochastic boosting
adaptMCMC	Implementation of a generic adaptive Monte Carlo Markov Chain sampler
AER	Applied Econometrics with R
akima	Interpolation of irregularly spaced data
AlgDesign	Algorithmic Experimental Design
ALL	A data package
ALL	A data package
alr2	Data to accompany Applied Linear Regression 3rd edition
amap	Another Multidimensional Analysis Package
anchors	Statistical analysis of surveys with anchoring vignettes
animation	A gallery of animations in statistics and utilities to create animations
ape	Analyses of Phylogenetics and Evolution

Installing / Updating RMassBank

To check if you need to update RMassBank:

```

1 # Interactive RMassBank Demonstration
2 # E. Schymanski, Eawag, 27/11/2012
3
4 # First, check that you have the latest stable RMassBank installed
5 source("http://bioconductor.org/biocLite.R")
6 # Identify packages requiring update
7 old.packages(repos=biocinstallRepos())
8 # update all packages that are out of date
9 biocLite(character(), ask=FALSE)
10 # or just install RMassBank (does the rest for you)
11 source("http://bioconductor.org/biocLite.R")
12 biocLite("RMassBank")
13 biocLite("RMassBank")
14 # Careful, if you get an error, you may need to update the RMassBank
15 # require updating the RMassBank
16 # if you get an error, you may need to update the RMassBank

```

Console ~/ ↻

```

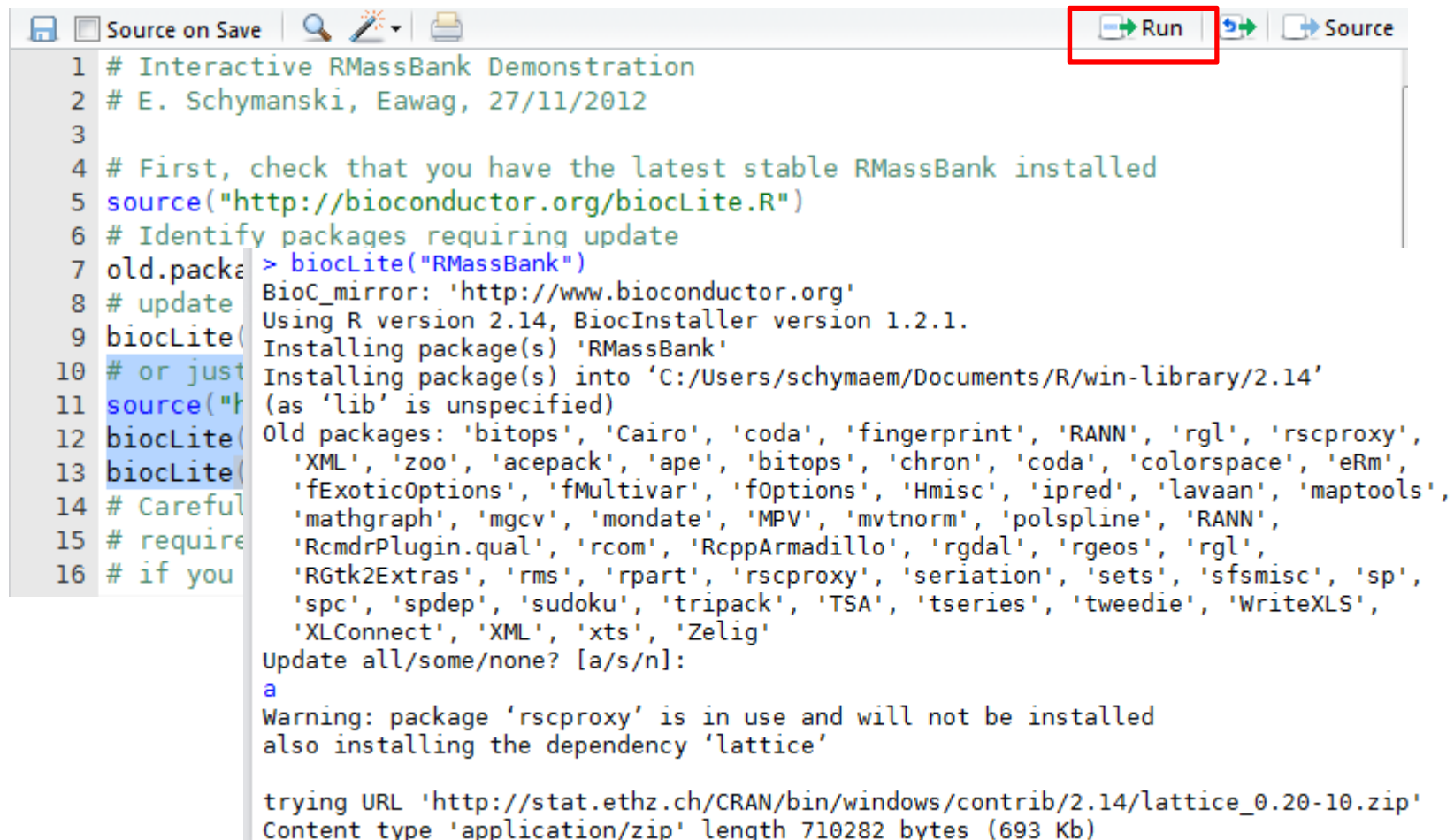
RM> source("http://bioconductor.org/biocLite.R")
BiocInstaller version 1.2.1, ?biocLite for help
RM> old.packages(repos=biocinstallRepos())

```

Package	LibPath
"bitops"	"C:/Users/schymaem/Documents/R/win-library/2.14"
"Cairo"	"C:/Users/schymaem/Documents/R/win-library/2.14"
"coda"	"C:/Users/schymaem/Documents/R/win-library/2.14"
"fingerprint"	"C:/Users/schymaem/Documents/R/win-library/2.14"
"RANN"	"C:/Users/schymaem/Documents/R/win-library/2.14"
"rgl"	"C:/Users/schymaem/Documents/R/win-library/2.14"
"rscproxy"	"C:/Users/schymaem/Documents/R/win-library/2.14"
"XML"	"C:/Users/schymaem/Documents/R/win-library/2.14"
"zoo"	"C:/Users/schymaem/Documents/R/win-library/2.14"
"acepack"	"C:/Program Files/R/R-2.14.1/library"
"ape"	"C:/Program Files/R/R-2.14.1/library"
"bitops"	"C:/Program Files/R/R-2.14.1/library"

Installing / Updating RMassBank

To install RMassBank:



```

Source on Save  Run  Source
1 # Interactive RMassBank Demonstration
2 # E. Schymanski, Eawag, 27/11/2012
3
4 # First, check that you have the latest stable RMassBank installed
5 source("http://bioconductor.org/biocLite.R")
6 # Identify packages requiring update
7 old.packages > biocLite("RMassBank")
8 # update BioC_mirror: 'http://www.bioconductor.org'
9 biocLite( Using R version 2.14, BiocInstaller version 1.2.1.
10 # or just Installing package(s) 'RMassBank'
11 source(" Installing package(s) into 'C:/Users/schymaem/Documents/R/win-library/2.14'
12 biocLite( (as 'lib' is unspecified)
13 biocLite( Old packages: 'bitops', 'Cairo', 'coda', 'fingerprint', 'RANN', 'rgl', 'rscproxy',
14 # Careful 'XML', 'zoo', 'acepack', 'ape', 'bitops', 'chron', 'coda', 'colorspace', 'eRm',
15 # require 'fExoticOptions', 'fMultivar', 'fOptions', 'Hmisc', 'ipred', 'lavaan', 'maptools',
16 # if you 'mathgraph', 'mgcv', 'mondate', 'MPV', 'mvtnorm', 'polspline', 'RANN',
'RCmdrPlugin.qual', 'rcom', 'RcppArmadillo', 'rgdal', 'rgeos', 'rgl',
'RGtk2Extras', 'rms', 'rpart', 'rscproxy', 'seriation', 'sets', 'sfsmisc', 'sp',
'spc', 'spdep', 'sudoku', 'tripack', 'TSA', 'tseries', 'tweedie', 'WriteXLS',
'XLConnect', 'XML', 'xts', 'Zelig'
Update all/some/none? [a/s/n]:
a
Warning: package 'rscproxy' is in use and will not be installed
also installing the dependency 'lattice'

trying URL 'http://stat.ethz.ch/CRAN/bin/windows/contrib/2.14/lattice_0.20-10.zip'
Content type 'application/zip' length 710282 bytes (693 Kb)

```

Installing from source – if all else fails

(this means hand-installing other packages, avoid if possible!)

The screenshot shows a web browser window with the address bar displaying `www.bioconductor.org/packages/release/bioc/html/RMassBank.html`. The page content is as follows:

Details

biocViews	Bioinformatics , MassSpectrometry , Metabolomics , Software
Version	1.0.0
In Bioconductor since	BioC 2.11 (R-2.15)
License	Artistic-2.0
Depends	rcdk,yaml,mzR,methods
Imports	XML,RCurl
Suggests	gplots,RMassBankData,xcms
System Requirements	OpenBabel
URL	
Depends On Me	
Imports Me	
Suggests Me	RMassBankData

Package Downloads

Package Source	RMassBank 1.0.0.tar.gz
Windows Binary	RMassBank 1.0.0.zip (32- & 64-bit)
MacOS 10.5 (Leopard) binary	RMassBank 1.0.0.tgz
Package Downloads Report	Download Stats

Contact us: webmaster@bioconductor.org
 Hosting provided by Fred Hutchinson Cancer Research Center
 Copyright © 2003 - 2012

FRED HUTCHINSON
CANCER RESEARCH CENTER
A LIFE OF SCIENCE

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Installing from source – if all else fails

(this means hand-installing other packages, avoid if possible!)

```
#If this doesn't work, try installing from source
# http://www.bioconductor.org/packages/release/bioc/html/RMassBank.html
# download Package : > install.packages("C:/DATA/RMassBank/RMassBank_Versions/RMassBank_1.0.0.tar.gz",
# http://www.bioconductor.org/packages/release/bioc/html/RMassBank.html repos=NULL, type="source")
install.packages("C:/DATA/RMassBank/RMassBank_Versions/RMassBank_1.0.0.tar.gz",
                  repos=NULL, type="source")
Installing package(s) into 'C:/Users/schymaem/Documents/R/win-library/2.14'
(as 'lib' is unspecified)
* installing *source* package 'RMassBank' ...
** R
** inst
** preparing package for lazy loading
Warning: package 'rcdk' was built under R version 2.14.2
Warning: package 'rcdklibs' was built under R version 2.14.2
Warning: package 'fingerprint' was built under R version 2.14.2
Warning: package 'iterators' was built under R version 2.14.2
Warning: package 'yaml' was built under R version 2.14.2
Warning: package 'Rcpp' was built under R version 2.14.2
** help
*** installing help indices
** building package indices ...
*** tangling vignette sources ...
'RMassBank.Rnw'
** testing if installed package can be loaded
Warning messages:
1: package 'rcdk' was built under R version 2.14.2
2: package 'rcdklibs' was built under R version 2.14.2
3: package 'fingerprint' was built under R version 2.14.2
4: package 'iterators' was built under R version 2.14.2
5: package 'yaml' was built under R version 2.14.2
6: package 'Rcpp' was built under R version 2.14.2

* DONE (RMassBank)
> |
```

Loading RMassBank

```
30 # Once you have got through all that, you need to load the packages:
31 library("RMassBank")
32 library("RMassBankData")
33
```

32:25 (Top Level) ↕

R Script ↕

Console ~/ ↕

```
> # Once you have got through all that, you need to load the packages:
> library("RMassBank")
Loading required package: mzR
Loading required package: Rcpp
Loading required package: XML
Loading required package: RCurl
Loading required package: bitops
Loading required package: zoo

Attaching package: 'zoo'

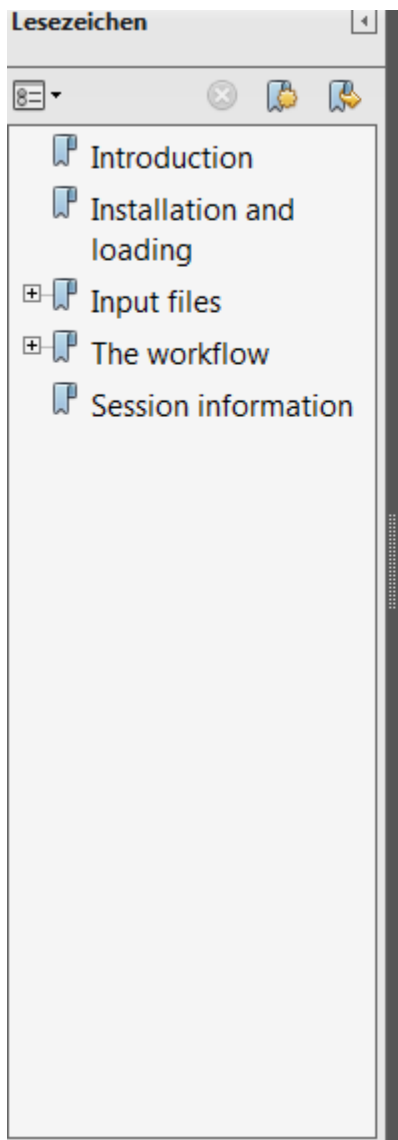
The following object(s) are masked from 'package:base':

  as.Date, as.Date.numeric

Loading required package: iterators
Warning messages:
1: package 'Rcpp' was built under R version 2.14.2
2: package 'XML' was built under R version 2.14.2
3: package 'bitops' was built under R version 2.14.2
4: package 'zoo' was built under R version 2.14.2
5: package 'rcdk' was built under R version 2.14.2
6: package 'rcdklibs' was built under R version 2.14.2
7: package 'fingerprint' was built under R version 2.14.2
8: package 'iterators' was built under R version 2.14.2
> library("RMassBankData")
Warning message:
package 'RMassBankData' was built under R version 2.14.2
>
```


Finding out more about RMassBank

vignette("RMassBank")



RMassBank: The workflow by example

Michael Stravs

October 2, 2012

We'll kind-of work through this today...

1 Introduction	2
2 Installation and loading	2
3 Input files	3
3.1 LC/MS data	3
3.2 Compound list	4
3.3 Settings	4
4 The workflow	7
4.1 Mass spectrometry workflow	7
4.2 MassBank record workflow	14
5 Session information	16

Before getting started:

Install / Check OpenBabel

If you think you already have this, check e.g. (for Windows....)

- All Programs => OpenBabel 2.X.X
- C:\Program Files\OpenBabel-2.3.0

If you don't have this, please download (not *required* but it is *recommended*)

- <http://openbabel.org/>

openbabel.org/wiki/Main_Page

Log in / create account

Page Discussion Read View source View history Go Search

Open Babel: The Open Source Chemistry Toolbox

Open Babel is a chemical toolbox designed to speak the many languages of chemical data. It's an open, collaborative project allowing anyone to search, convert, analyze, or store data from molecular modeling, chemistry, solid-state materials, biochemistry, or related areas.

- **Ready-to-use programs, and complete programmer's toolkit**
- Read, write and convert over **110 chemical file formats**
- Filter and search molecular files using **SMARTS** and other methods
- Supports molecular modeling, cheminformatics, bioinformatics
- Organic chemistry, inorganic chemistry, solid-state materials, nuclear chemistry
- Downloaded over **164,000 times** and used by over **40 related projects**
- [More about Open Babel](#)
- [Open Babel on SourceForge](#)

Download

Official User Guide
Open Babel
or How I learned to love the chemical file format

[Online docs](#)
[PDF](#)
[Buy Book](#)
[Read Paper](#)

[Browse the API](#)

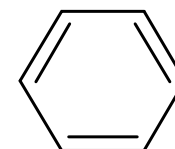
To support Open Babel, please cite J. Cheminf. 2011, 3:33

Before getting started:

Install / Check OpenBabel

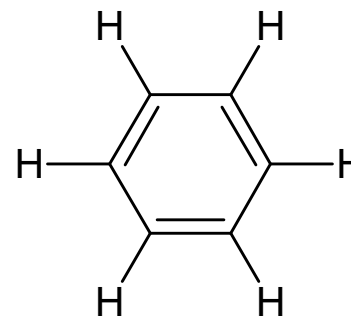
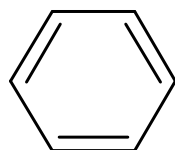
Why OpenBabel?

- Convert SMILES (c1ccccc1) to structure
- C:\Program Files\OpenBabel-2.3.0



NOTE:

- By default, RMassBank will work without OpenBabel and has a back-up solution using CACTUS
- But, structures are less visually appealing



Before getting started:

Install / Check Proteowizard

Why Proteowizard?

- Convert raw data files to an open format (we chose mzML)
- Is one of most established and supported converters

**NOTE: If you will only ever use RMassBank once (i.e. today)
Don't bother installing! The test data is already converted!**

Check to see if you have it

- E.g. for me: C:\DATA\Program_Info\pwiz
- You want "MSConvertGUI.exe"

Name	Date modified	Type	Size
MassLynxRaw.dll	27.01.2012 03:00	Application extens...	125 KB
MassSpecDataReader.dll	27.01.2012 03:00	Application extens...	44 KB
MathNet.Iridium.dll	26.04.2011 23:15	Application extens...	240 KB
msaccess.exe	27.01.2012 03:01	Application	5'792 KB
msbenchmark.exe	27.01.2012 03:03	Application	4'044 KB
mscat.exe	27.01.2012 03:02	Application	4'025 KB
msconvert.exe	27.01.2012 03:01	Application	5'755 KB
MSConvertGUI.exe	27.01.2012 03:01	Application	54 KB
MSConvertGUI.exe.manifest	09.01.2012 22:44	MANIFEST File	3 KB
msdiff.exe	27.01.2012 03:01	Application	4'206 KB
msdir.exe	27.01.2012 03:01	Application	4'166 KB
MSFileReader.XRawfile2.dll	13.12.2011 06:22	Application extens...	237 KB

Before getting started:

Install / Check Proteowizard

No Proteowizard (pwiz)?: <http://proteowizard.sourceforge.net/downloads.shtml>

proteowizard.sourceforge.net/downloads.shtml

info download user docs dev docs contact

ProteoWizard Downloads

[Get ProteoWizard \(3 steps\)](#)

Step 1, Choose Download Type:

Windows installer (includes vendor reader support)

Step 2, Read License Agreements

Pwiz Core AB Sciex WiffReader Agilent MHDAC Bruker CompassXtract Thermo msFileReader Waters WRDAC

ADDENDUM TO APACHE LICENSE

To the best of our ability we deliver this software to you under the Apache 2.0 License listed below (the source code is available in the ProteoWizard project). This software does, however, depend on other software libraries which place further restrictions on its use and redistribution. By accepting the license terms for this software, you agree to comply with the restrictions imposed on you by the license agreements of the software libraries on which it depends:

Agilent Mass Hunter Data Access Component Library
Waters Raw Data Access Component Library
Bruker CompassXtract
Thermo-Scientific MSFileReader Library
Waters Raw Data Access Component Library
AB Sciex WIFF Reader Library

NOTE: If you do not plan to redistribute this software yourself, then you are the "end-user" in the above agreements.

Apache License - Covers Core ProteoWizard Tools and Library

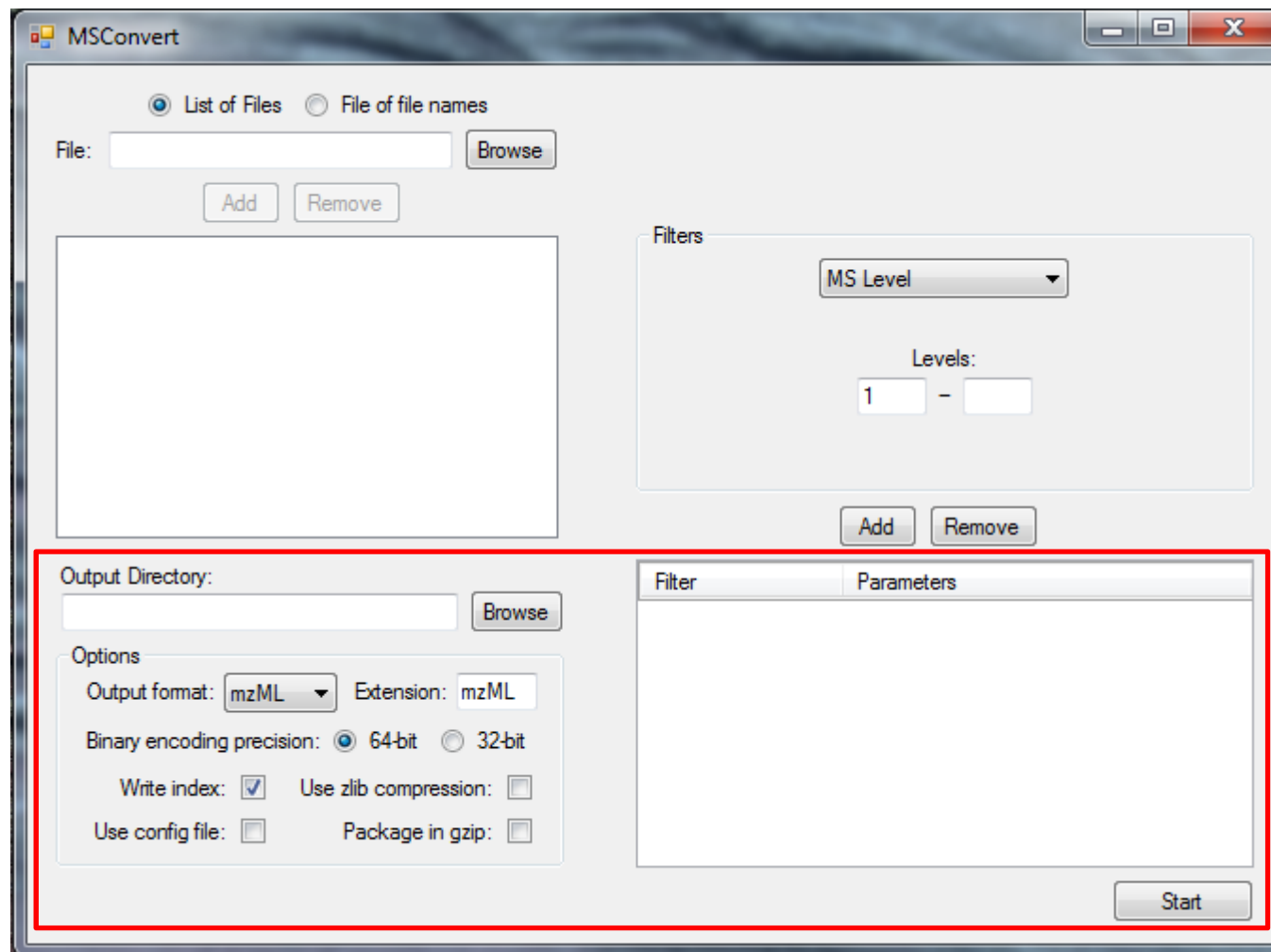
Step 3, Click below to agree to Licenses and get ProteoWizard

I agree to the licensing terms, download ProteoWizard

Before getting started:

Install / Check Proteowizard

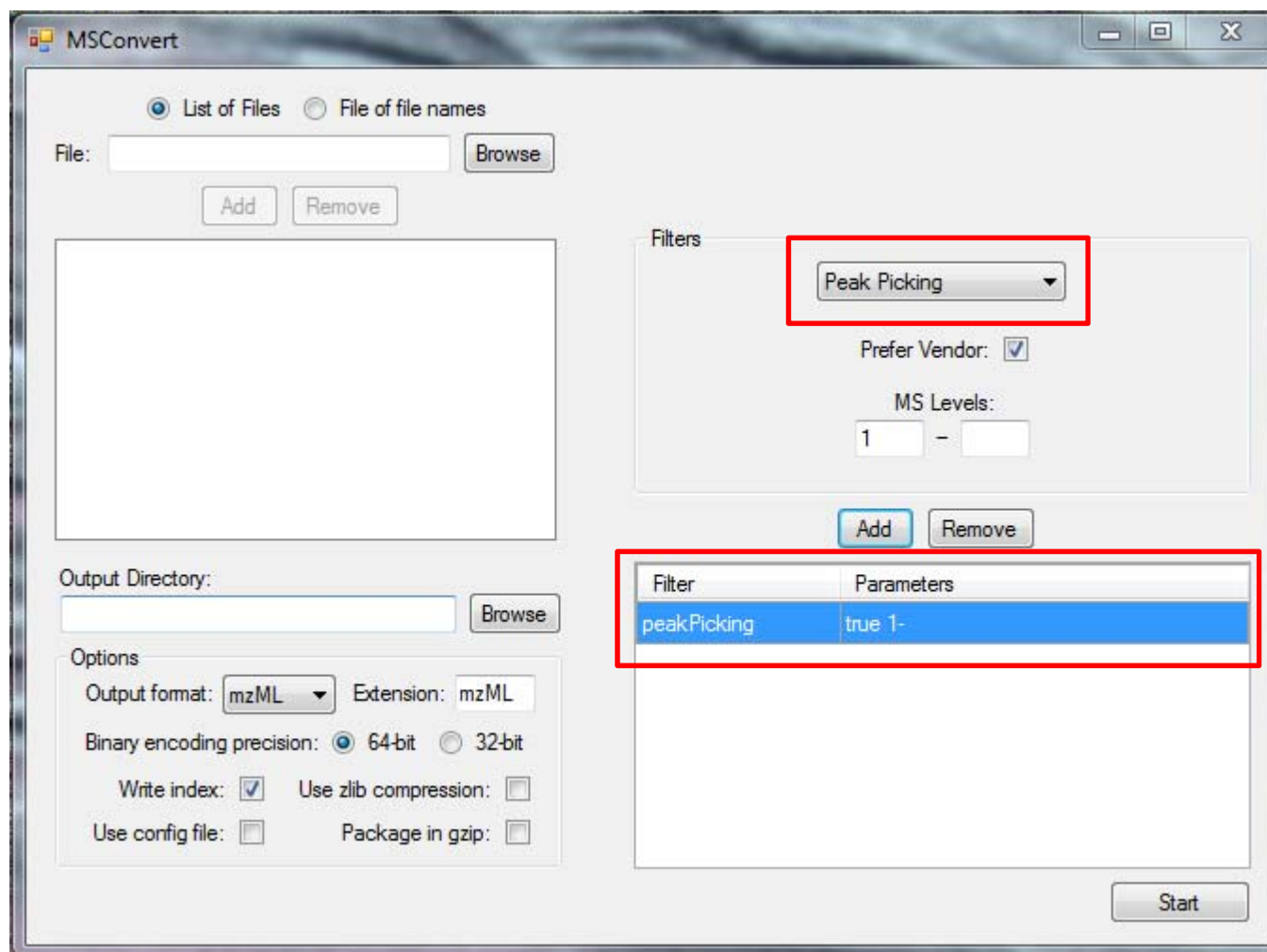
To use profile data (RMassBank deprofiler):



Before getting started:

Install / Check Proteowizard

To use centroid data (Pwiz deprofiler – example data prepared this way):



Before getting started

File names...

The files are used to identify a compound in the workflow

- Currently, we require a format like:
xxxxxxxxxxxxx_1234_.mzML
- _1234_ is the compound ID which is also in the compound list
- mzML is the format currently read by RMassBank

- Advanced / alternative formats possible
 - Will require adjustment to the RMassBank code
 - Check out functions `msms_workflow` and `findMsMsHR`

Before getting started:

The compound list

Go back to the R / RStudio window

- Choose a working directory (change this!!! and create a new one if you like)
- Copy the compound list into this directory:

```
39 # Save a copy of the compound list|
40 # First, set a working directory (I make sure this exists)
41 setwd("C:/DATA/RMassBank/RMassBank_Demo")
42 # then copy the file there
43 file.copy(system.file("list/NarcoticsDataset.csv",
44                       package="RMassBankData"), "./Compoundlist.csv")
45
```

```
> setwd("C:/DATA/RMassBank/RMassBank_Demo")
> # then copy the file there
> file.copy(system.file("list/NarcoticsDataset.csv",
+                       package="RMassBankData"), "./Compoundlist.csv")
[1] TRUE
>
```


Before getting started:

The settings file (lucky last bit!!)

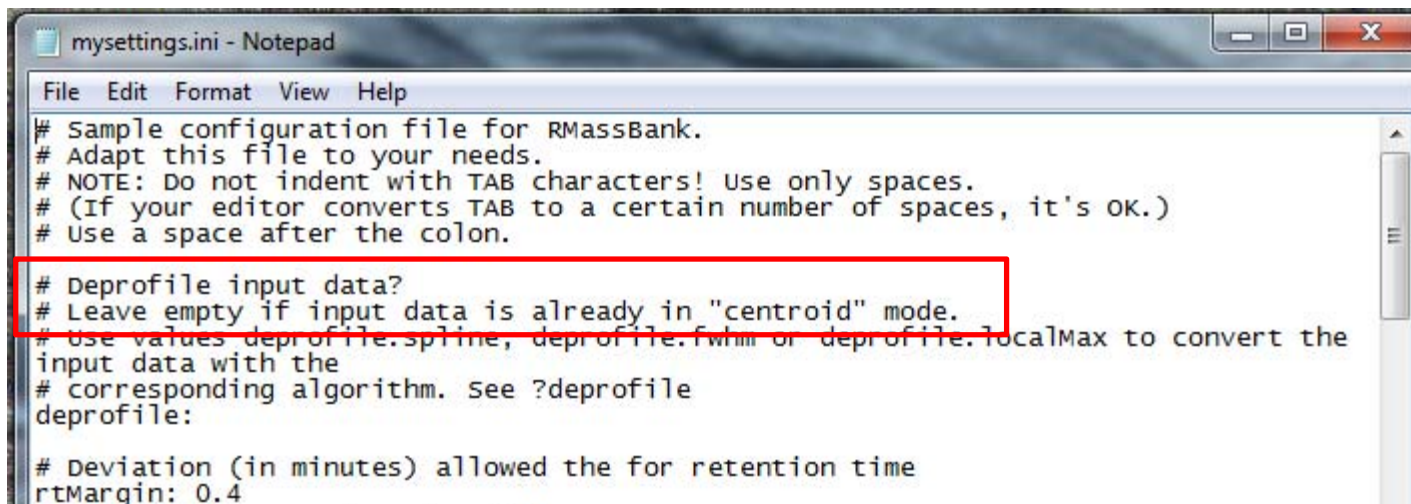
A number of settings are necessary for successful generation:

- Data processing
- Correct annotation of spectra

Now, generate a settings file:

```
RmbSettingsTemplate("mysettings.ini")
```

- Open up in any text editor and edit...



```
mysettings.ini - Notepad
File Edit Format View Help
# Sample configuration file for RMassBank.
# Adapt this file to your needs.
# NOTE: Do not indent with TAB characters! Use only spaces.
# (If your editor converts TAB to a certain number of spaces, it's OK.)
# Use a space after the colon.
# Deprofile input data?
# Leave empty if input data is already in "centroid" mode.
# Use values deprofile.spline, deprofile.fwhm or deprofile.localMax to convert the
input data with the
# corresponding algorithm. see ?deprofile
deprofile:
# Deviation (in minutes) allowed the for retention time
rtMargin: 0.4
```

Before getting started:

The settings file (lucky last bit!!)

```

mysettings.ini - Notepad
File Edit Format View Help
# Deviation (in minutes) allowed the for retention time
rtMargin: 0.4
# Systematic retention time shift
rtShift: -0.3

# Directory to openBabel. Required for creating molfiles for MassBank export.
# If no openBabel directory is given, RMassBank will attempt to use the CACTUS
webservice
# for SDF generation. You really should install openBabel though; the CACTUS
structures
# have explicit hydrogen atoms...
# Points to the directory where babel.exe (or the Linux "babel" binary) is located.
babeldir: |
# Example:
# babeldir: "C:\Program Files (x86)\openBabel-2.3.1\"

# which MassBank record version to use; version 2 is advised.
use_version: 2

# Include reanalyzed peaks?
use_rean_peaks: TRUE

# annotate the spectra files with (putative) molecular formulas for fragments?
add_annotation: TRUE

# Annotations for the spectrum:
annotations:
# Author etc. annotation
authors: Nomen Nescio, The Unseen University
copyright: Copyright (C) xxx
license: CC BY-SA

instrument: LTQ Orbitrap XL Thermo scientific
instrument_type: LC-ESI-ITFT
confidence_comment: standard compound

```

These settings are OK for test data

Add the OpenBabel location here
CAREFUL – ‘ and “ as in example

Define record format (keep 2)

These are either TRUE or FALSE

Reanalyzed = formula + N₂O
Annotation adds formula to records

Change author and copyright to
something relevant for you...

Before getting started:

The settings file (lucky last bit!!)

```

mysettings.ini - Notepad
File Edit Format View Help
annotations:
  # Author etc. annotation
  authors: Nomen Nescio, The Unseen University
  copyright: Copyright (c) XXX
  license: CC BY-SA
  instrument: LTQ Orbitrap XL Thermo Scientific
  instrument_type: LC-ESI-ITFT
  confidence_comment: standard compound
  compound_class: N/A; Environmental Standard
  internal_id_fieldname: INTERNAL_ID
#
# HPLC annotations:
#
# example: lc_gradient: 90/10 at 0 min, 50/50 at 4 min, 50/50 at 25
25 min, 90/10 at 25.1 min, 90/10 at 30 min
lc_gradient:
# example: lc_flow: 200 uL/min
lc_flow:
# example: lc_solvent_a: water with 0.1% formic acid
lc_solvent_a:
lc_solvent_b:
# example: lc_column: XBridge C18 3.5um, 2.1x50mm, Waters
lc_column:
# Prefix for MassBank accession IDs
entry prefix: XX
ms_type: MS2
ionization: ESI
ms_dataprocessing:
  RECALIBRATE: loess on assigned fragments and MS1

# List of data-dependent scans in their order (relative to the parent scan), for
annotation of the MassBank records
# For every data-dependent scan event, specify an element with:
# mode: fragmentation mode, e.g. CID
# ces: "short" format collision energy (for record title)

```

License: change for own records?
These settings are OK for test data

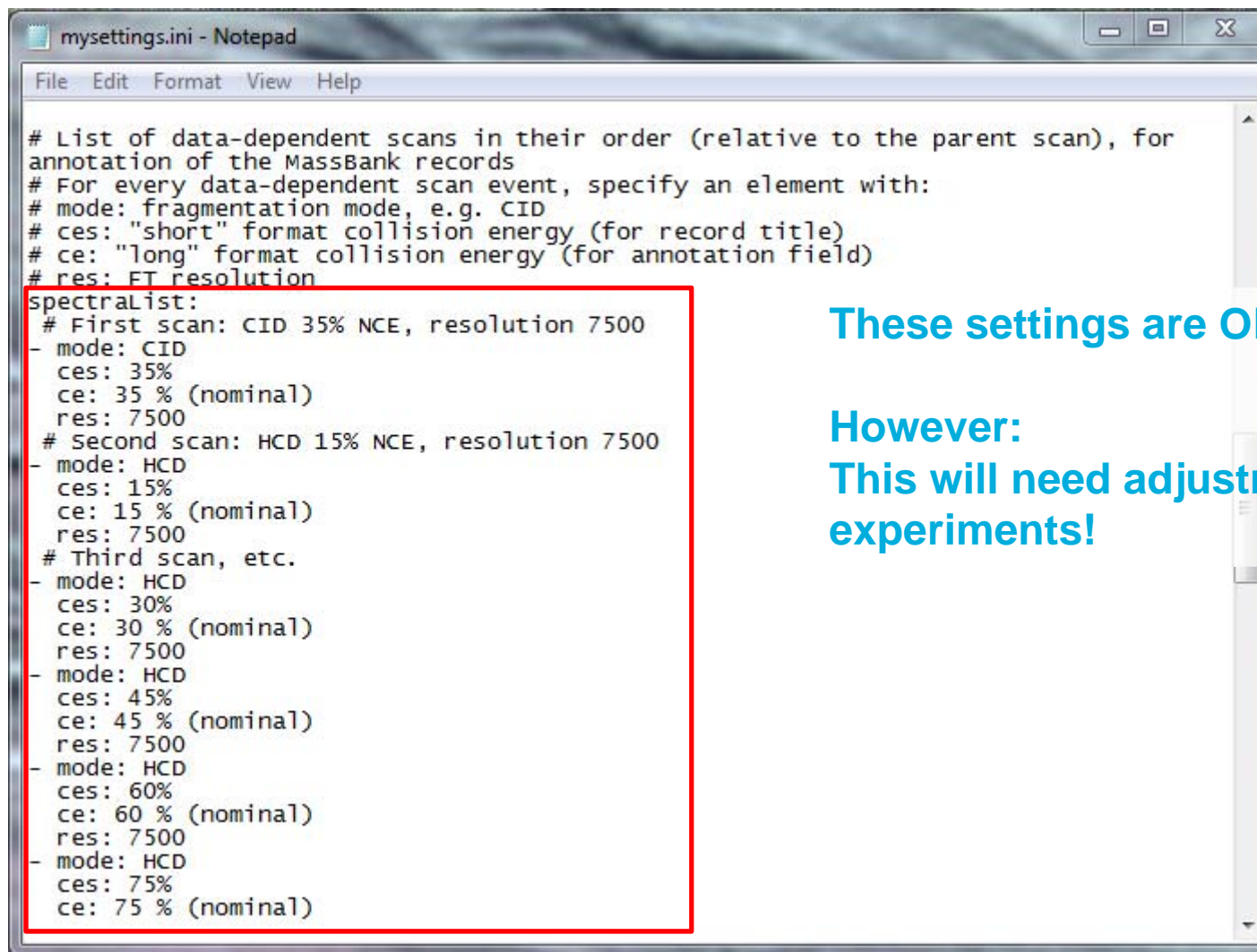
These are currently empty
You can use the examples for the example data. Solvent B was methanol!
These values will need adjusting for own chromatographic program

Choose your record letters!
E.g. Eawag = EA

These are OK for test data

Before getting started:

The settings file (lucky last bit!!)



```

mysettings.ini - Notepad
File Edit Format View Help

# List of data-dependent scans in their order (relative to the parent scan), for
# annotation of the MassBank records
# For every data-dependent scan event, specify an element with:
# mode: fragmentation mode, e.g. CID
# ces: "short" format collision energy (for record title)
# ce: "long" format collision energy (for annotation field)
# res: FT resolution
spectralList:
# First scan: CID 35% NCE, resolution 7500
- mode: CID
  ces: 35%
  ce: 35 % (nominal)
  res: 7500
# Second scan: HCD 15% NCE, resolution 7500
- mode: HCD
  ces: 15%
  ce: 15 % (nominal)
  res: 7500
# Third scan, etc.
- mode: HCD
  ces: 30%
  ce: 30 % (nominal)
  res: 7500
- mode: HCD
  ces: 45%
  ce: 45 % (nominal)
  res: 7500
- mode: HCD
  ces: 60%
  ce: 60 % (nominal)
  res: 7500
- mode: HCD
  ces: 75%
  ce: 75 % (nominal)
  
```

These settings are OK for test data

However:
This will need adjustment for own
experiments!

Before getting started:

The settings file (lucky last bit!!)

Finally: Save this before you go on!!!

```

mysettings.ini - Notepad
File Edit Format View Help
# Shifts of the starting points for RMassBank accession numbers.
# Change these if you measure different adducts
accessionNumberShifts:
  pH: 0 # [M+H]+: Accession numbers 1-14
  pM: 16 # [M]+: 17-30
  pNa: 32 # [M+Na]+: 33-46
  mH: 50 # [M-H]-: 51-64
  mFA: 66 # [M+FA]-: 67-80

# A list of known electronic noise peaks
electronicNoise:
- 189.825
- 201.725
- 196.875
# Exclusion width of electronic noise peaks (from unmatched peaks, or prior to
# reanalysis)
electronicNoiseWidth: 0.3

# recalibration settings:
# recalibrate by: dppm or dmz
recalibrateBy: dppm

# recalibrate MS1:
# separately (separate)
# with common curve (common)
# do not recalibrate (none)
recalibrateMS1: common

# Custom recalibration function: You can overwrite the recal function by
# making any function which takes rcdData$recalField ~ rcdData$mzFound.
# The settings define which recal function is used.
# Note: if recalibrateMS1 is "common", the setting "recalibrator:" is
# meaningless
# because the MS1 points will be recalibrated together with the MS2 points
# the MS2 recalibration function.
recalibrator:
  MS1: recalibrate.loess
  MS2: recalibrate.loess
  
```

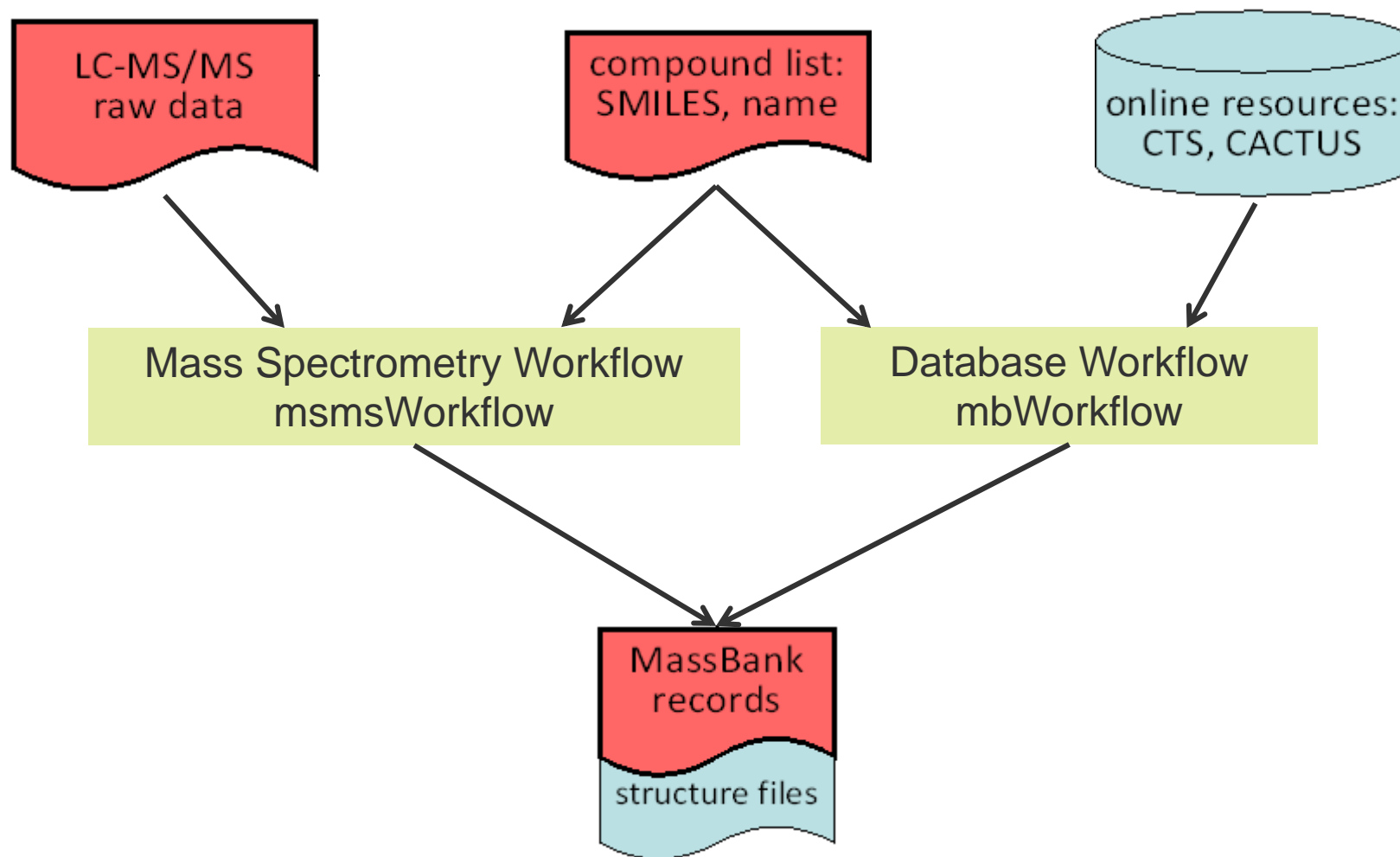
Fixed accession number: **XXYYYYZZ**;
XX = letters, **YYYY** = compound ID
ZZ = "shifts" – for us, these adducts

Remove consistent electronic noise
 Add the middle mass of the noise (often broad peaks, or slightly varying)
 Then add exclusion width –not too generous!

Recalibrate settings:
 Mass accuracy unit
 MS & MS/MS

These are OK for test data
 Changes necessary if recalibration
 needs adjusting

RMassBank Workflow – Simple Form

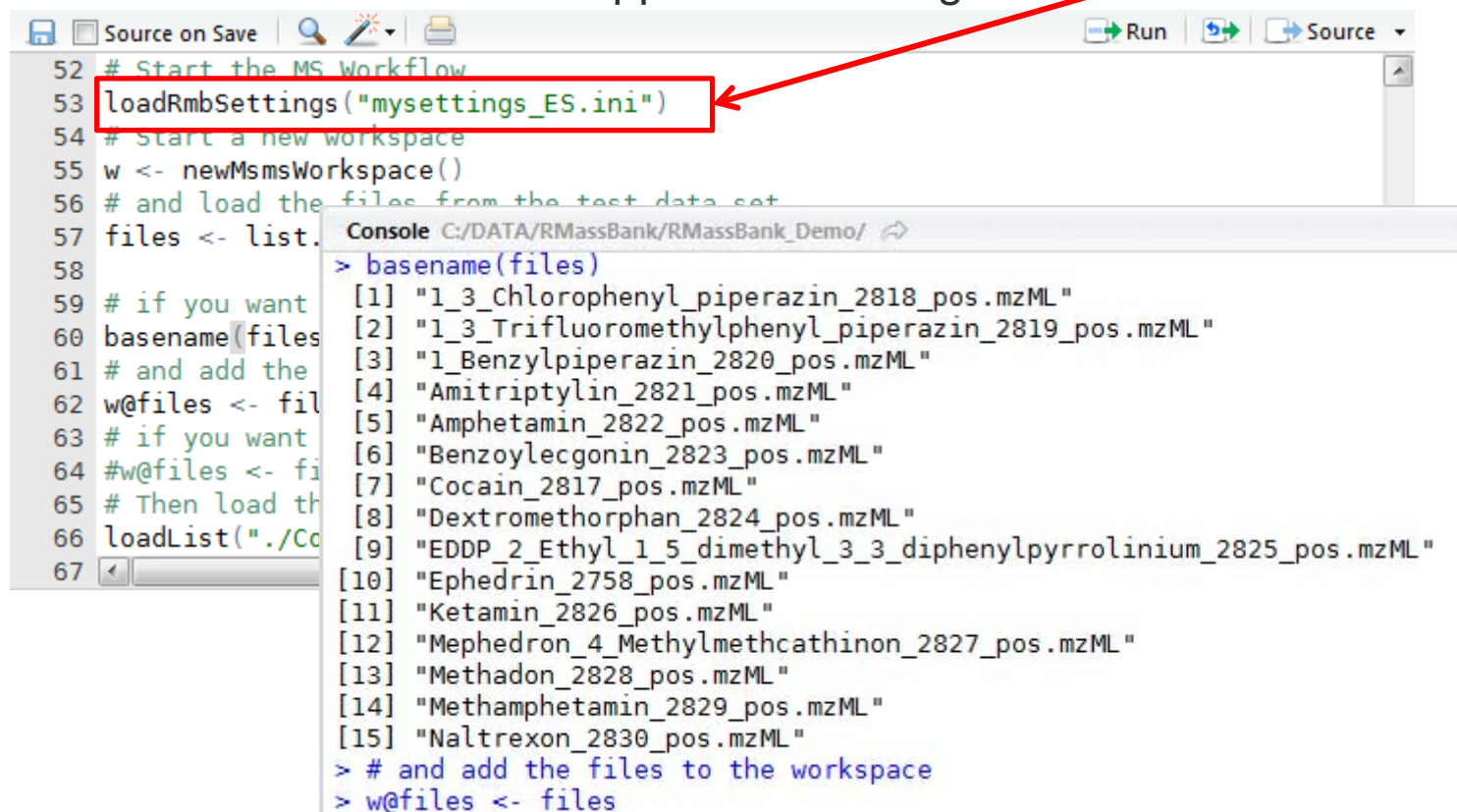


Mass Spectrometry Workflow

Now we can get started!

Go back to your R session and “RMassBank_Demo.R”

- Select and run this whole snippet – but change the file name first!



```
52 # Start the MS Workflow
53 loadRmbSettings("mysettings_ES.ini")
54 # Start a new workspace
55 w <- newMsmsWorkspace()
56 # and load the files from the test data set
57 files <- list.files("C:/DATA/RMassBank/TestData", pattern = "*.mzML")
58
59 # if you want to use the workspace
60 basename(files)
61 # and add the files to the workspace
62 w@files <- files
63 # if you want to use the workspace
64 #w@files <- files
65 # Then load the files
66 loadList("./C:/DATA/RMassBank/TestData", files)
67
```

```
> basename(files)
[1] "1_3_Chlorophenyl_piperazin_2818_pos.mzML "
[2] "1_3-Trifluoromethylphenyl_piperazin_2819_pos.mzML "
[3] "1_Benzylpiperazin_2820_pos.mzML "
[4] "Amitriptylin_2821_pos.mzML "
[5] "Amphetamin_2822_pos.mzML "
[6] "Benzoyllecgonin_2823_pos.mzML "
[7] "Cocain_2817_pos.mzML "
[8] "Dextromethorphan_2824_pos.mzML "
[9] "EDDP_2_Ethyl_1_5_dimethyl_3_3_diphenylpyrrolinium_2825_pos.mzML "
[10] "Ephedrin_2758_pos.mzML "
[11] "Ketamin_2826_pos.mzML "
[12] "Mephedron_4_Methylmethcathinon_2827_pos.mzML "
[13] "Methadon_2828_pos.mzML "
[14] "Methamphetamin_2829_pos.mzML "
[15] "Naltrexon_2830_pos.mzML "
> # and add the files to the workspace
> w@files <- files
```

Mass Spectrometry Workflow

msmsWorkflow: 8 Steps in summary

- Step 1: Search all `files` for MS/MS spectra of compound
- Step 2: Molecular formula fitting for all peaks (large tolerance)
- Step 3: Analyzed spectra aggregated into a list
- Step 4: Recalibration curve is calculated; all spectra are recalibrated
- Step 5: Recalibrated spectra are reanalyzed (new formula fit post calibration)
- Step 6: Reanalyzed spectra aggregated. Electronic noise removed.
 - Back-up saved as `archiveName.RData`
- Step 7: Fail peaks are reanalyzed (formula + 2N, O)
 - Back-up saved as `archiveName_RA.RData`
- Step 8: Multiplicity Filtering: peaks occurring once only are removed.
 - Back-up saved as `archiveName_RF.RData`
 - High Intensity Fail Peaks: `archiveName_Failpeaks.csv`

Mass Spectrometry Workflow

8 steps in RMassBank code...

- Steps 1:4 – up to recalibration

```
# Start the workflow with steps 1 to 4:  
w <- msmsWorkflow(w, mode="pH", steps=c(1:4), archivename = "pH_narcotics")
```

- Warning: this can take a while....

- Progress:

```
> # Start the workflow with steps 1 to 4:  
> w <- msmsWorkflow(w, mode="pH", steps=c(1:4), archivename = "pH_narcotics")  
msmsWorkflow: Step 1  
|  
| 0%3: >C:/Users/schymaem/Documents/R/win-  
library/2.14/RMassBankData/spectra/1_3_Chlorophenyl_piperazin_2818_pos.mzML  
|  
| 7%3: >C:/Users/schymaem/Documents/R/win-  
library/2.14/RMassBankData/spectra/1_3-Trifluoromethylphenyl_piperazin_2819_pos.mzML  
|  
| 13%3: >C:/Users/schymaem/Documents/R/win-  
library/2.14/RMassBankData/spectra/1_Benzylpiperazin_2820_pos.mzML  
|  
|=====| 20%3:  
>C:/Users/schymaem/Documents/R/win-library/2.14/RMassBankData/spectra/Amitriptylin_2821
```

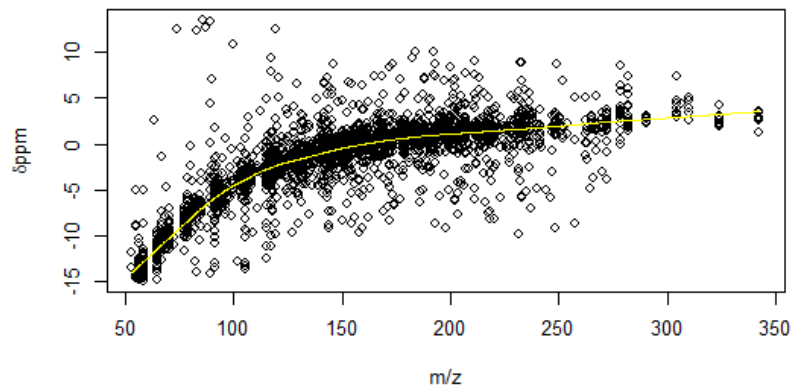
- P.S. don't worry if lots of "3"s appear, this is (relatively) meaningless

Mass Spectrometry Workflow

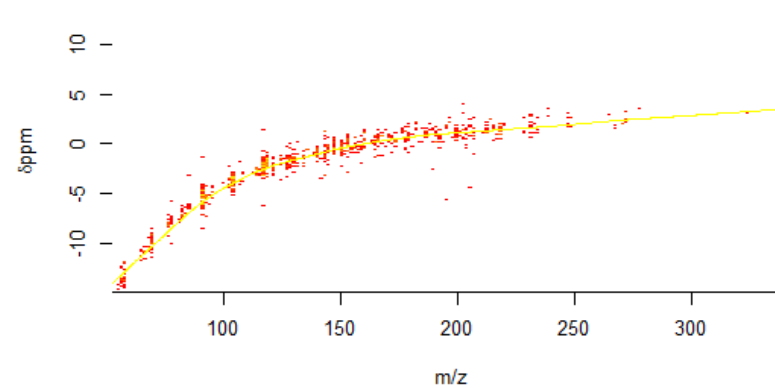
Steps 1:4 of RMassBank code...

- Recalibration plot at the end

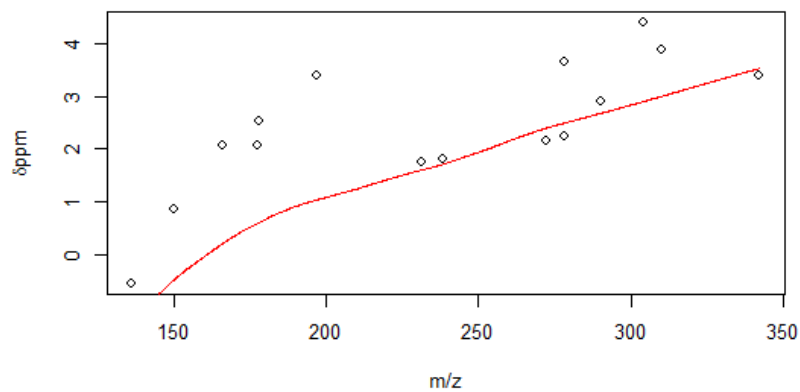
MS2 scatterplot



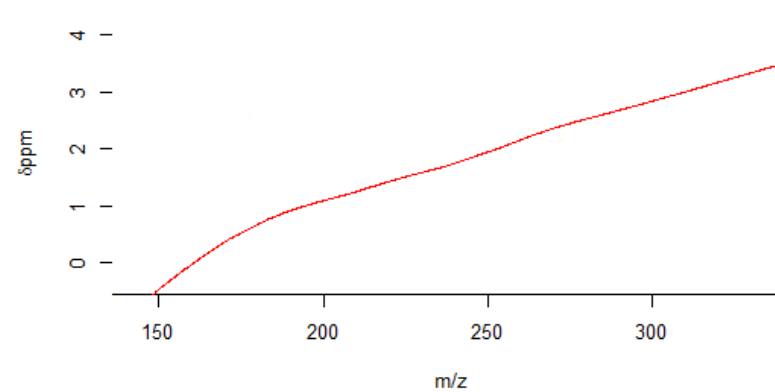
MS2 density



MS1 scatterplot

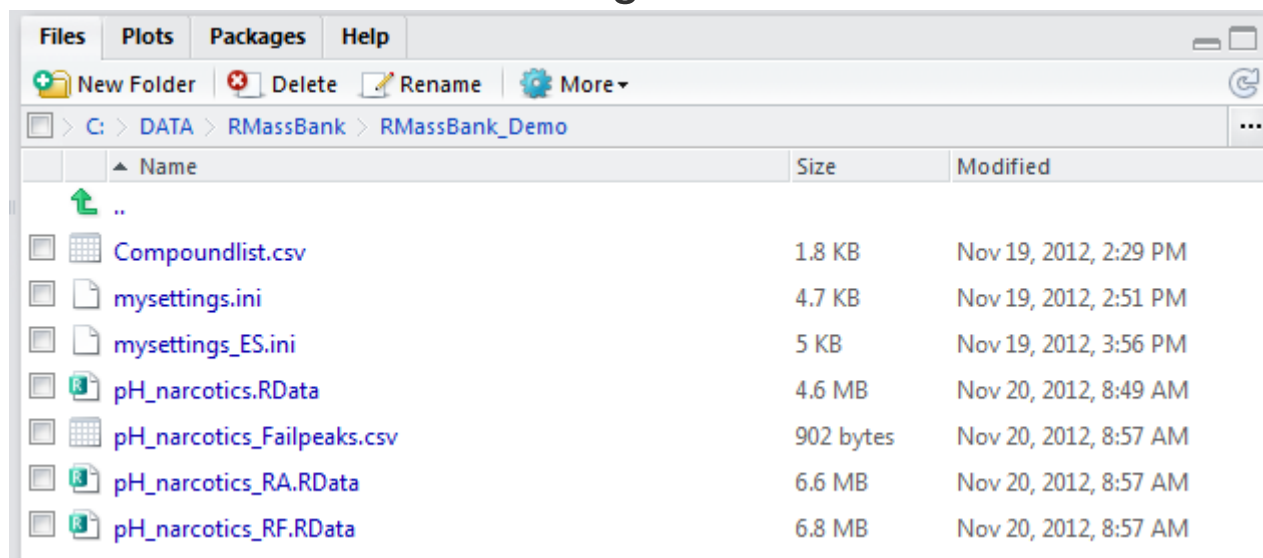


MS1 density



Mass Spectrometry Workflow

File listing should now look something like:



- o Open up the fail peaks (outside R) and have a look

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	OK	name	cpdID	scan	mzFound	formula	mzCalc	dppm	dbe	mz	int	formulaC	parentSca	aMax	mzCenter
2		407	2817	558	182.1191	NA	NA	NA	NA	182.1193	117786.4	0	556	9902212	304.1543
3		2231	2821	963	233.1351	NA	NA	NA	NA	233.1355	255822.6	0	961	16383558	278.1903
4		3491	2824	723	215.1468	NA	NA	NA	NA	215.1471	242559.4	0	721	15875343	272.2009
5		3988	2825	725	249.1548	NA	NA	NA	NA	249.1553	516037.1	0	723	23457888	278.1903
6		4110	2825	730	193.0993	NA	NA	NA	NA	193.0995	138038.3	0	723	5673618	278.1903
7		5543	2828	938	265.1612	NA	NA	NA	NA	265.1618	146493.5	0	933	4158767	310.2165
8		6503	2830	352	211.0738	C12H9N3O	211.074	-0.9633	13	211.0741	24055.95	1	346	195149.4	342.17

Mass Spectrometry Workflow

Fail Peaks

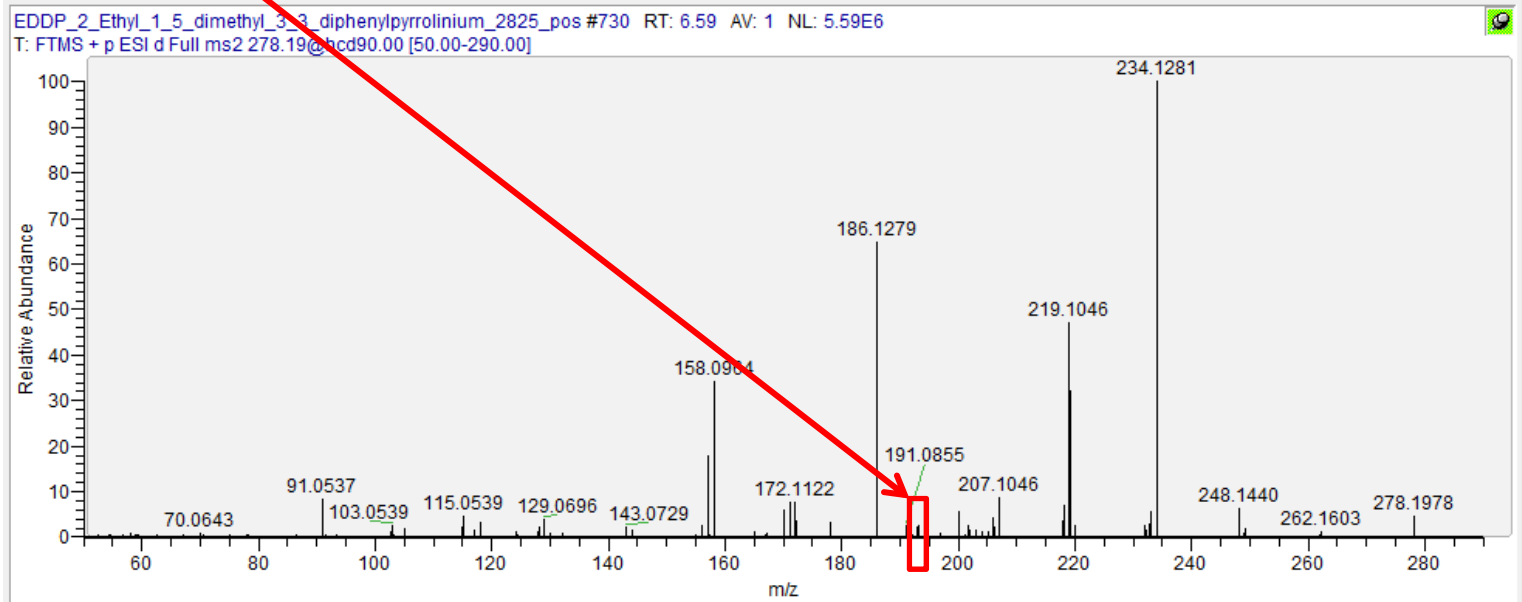
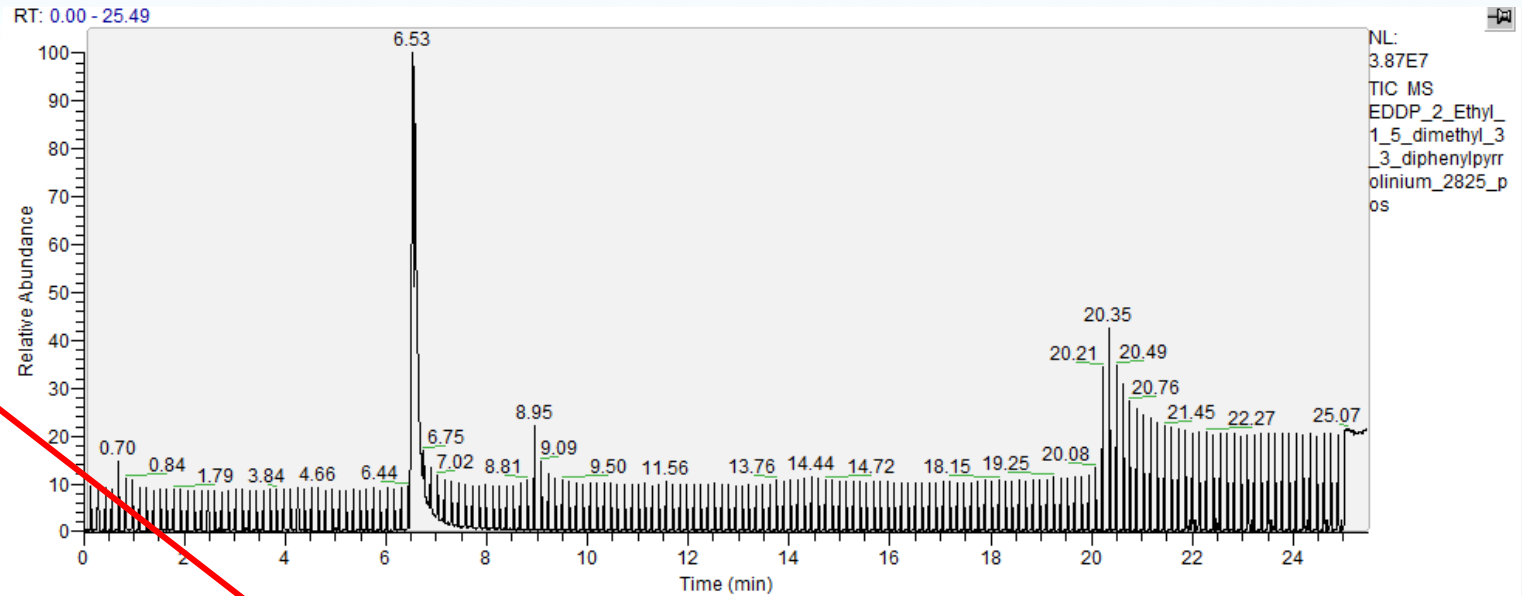
Compound ID MS/MS Scan m/z & I of MS/MS peak I, m/z of precursor

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	OK	name	cpdID	scan	mzFound	formula	mzCalc	dppm	dbe	mz	int	formulaC	parentScan	aMax	mzCenter
2		407	2817	558	182.1191	NA	NA	NA	NA	182.1193	117786.4	0	556	9902212	304.1543
3		2231	2821	963	233.1351	NA	NA	NA	NA	233.1355	255822.6	0	961	16383558	278.1903
4		3491	2824	723	215.1468	NA	NA	NA	NA	215.1471	242559.4	0	721	15875343	272.2009
5		3988	2825	725	249.1549	NA	NA	NA	NA	249.1553	516837.1	0	723	23457888	278.1903
6		4110	2825	730	193.0993	NA	NA	NA	NA	193.0995	138038.3	0	723	5673618	278.1903
7		5543	2828	938	265.1611	NA	NA	NA	NA	265.1618	146493.5	0	933	4158767	310.2165
8		6503	2830	352	211.0738	C12H9N3O	211.074	-0.9633	13	211.0741	24055.95	1	346	195149.4	342.17

No formula calculated; not replicated (only present once) and low I compared with parent => "true" fail peaks
Look at raw data of "2825"

Mass Spectrometry Workflow

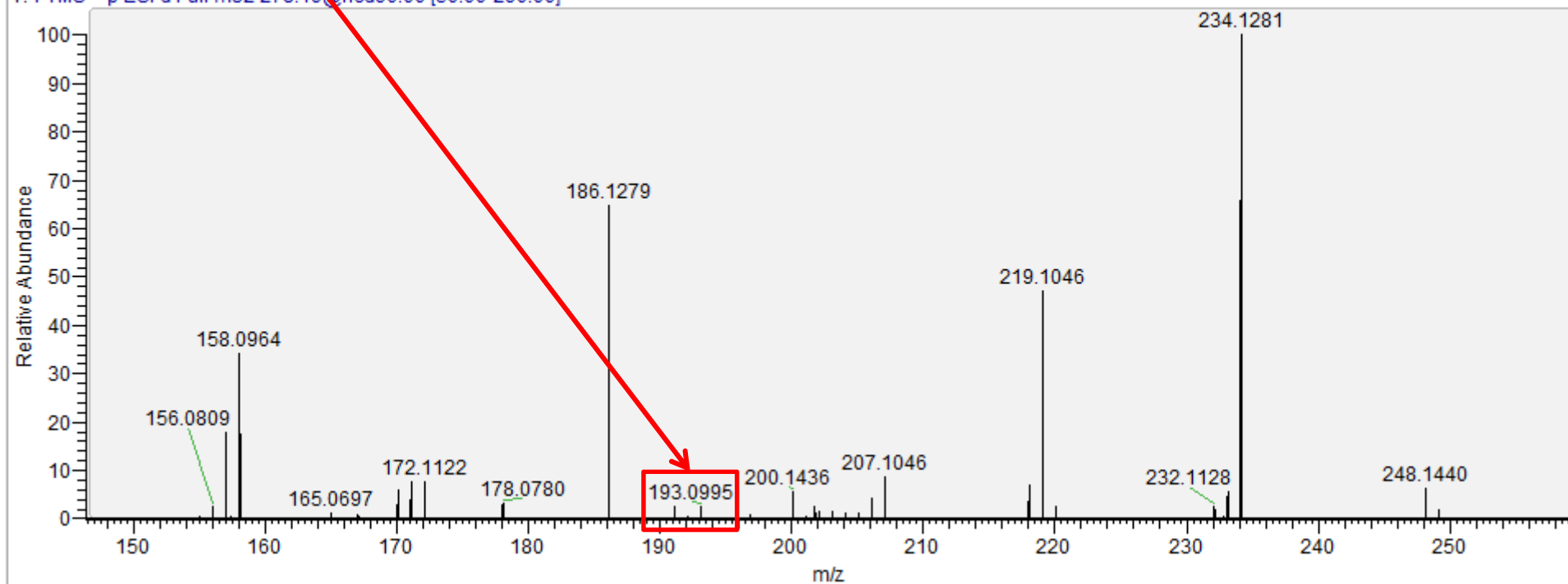
Fail Peak



Mass Spectrometry Workflow

Fail Peaks

EDDP_2_Ethyl_1_5_dimethyl_3_3_diphenylpyrrolinium_2825_pos #730 RT: 6.59 AV: 1 NL: 5.59E6
T: FTMS + p ESI d Full ms2 278.19@hcd90.00 [50.00-290.00]



“Fail peak” is very minor part of spectrum, already within noise levels

Mass Spectrometry Workflow

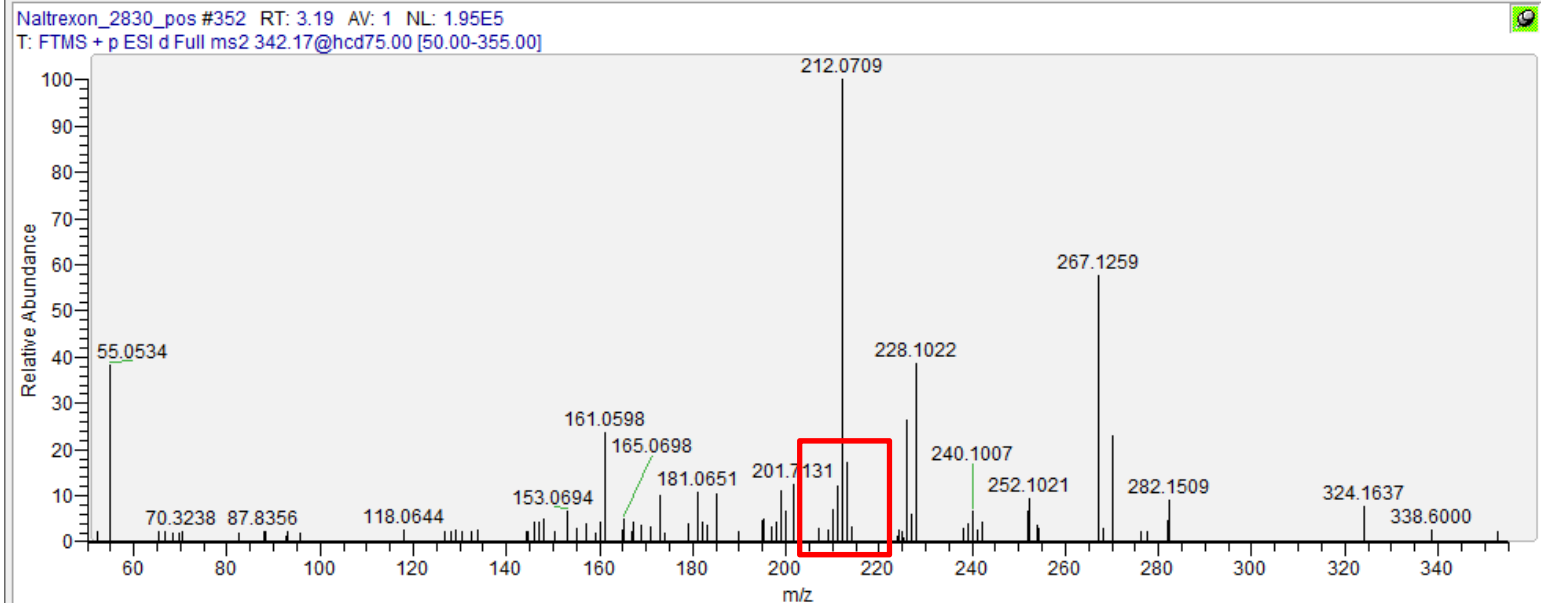
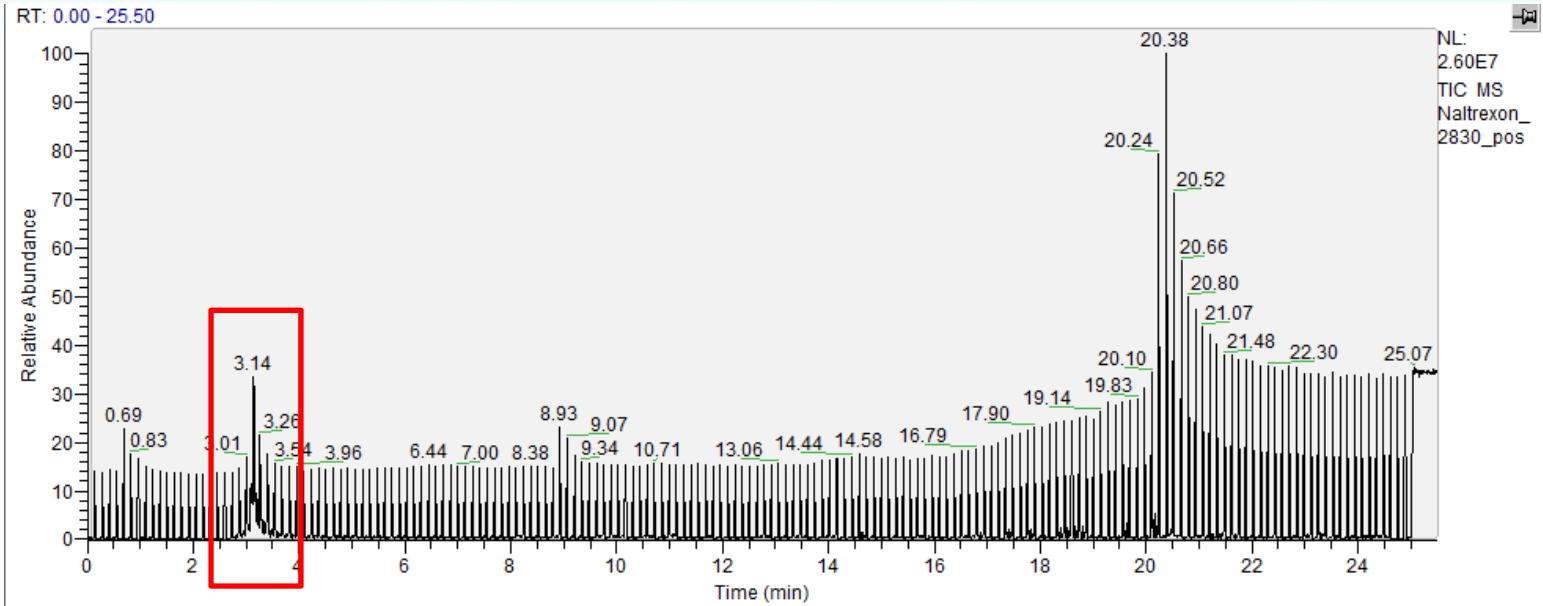
Fail Peaks

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	OK	name	cpdID	scan	mzFound	formula	mzCalc	dppm	dbe	mz	int	formulaC	parentSca	aMax	mzCenter
2		407	2817	558	182.1191	NA	NA	NA	NA	182.1193	117786.4	0	556	9902212	304.1543
3		2231	2821	963	233.1351	NA	NA	NA	NA	233.1355	255822.6	0	961	16383558	278.1903
4		3491	2824	723	215.1468	NA	NA	NA	NA	215.1471	242559.4	0	721	15875343	272.2009
5		3988	2825	725	249.1548	NA	NA	NA	NA	249.1553	516037.1	0	723	23457888	278.1903
6		4110	2825	730	193.0993	NA	NA	NA	NA	193.0995	138038.3	0	723	5673618	278.1903
7		5543	2828	938	265.1612	NA	NA	NA	NA	265.1618	146498.5	0	933	4158767	310.2165
8		6503	2830	352	211.0738	C12H9N3O	211.074	-0.9633	13	211.0741	24055.95	1	346	195149.4	342.17

Formula calculated, but also low intensity compared with precursor
Go to raw data...

Mass Spectrometry Workflow

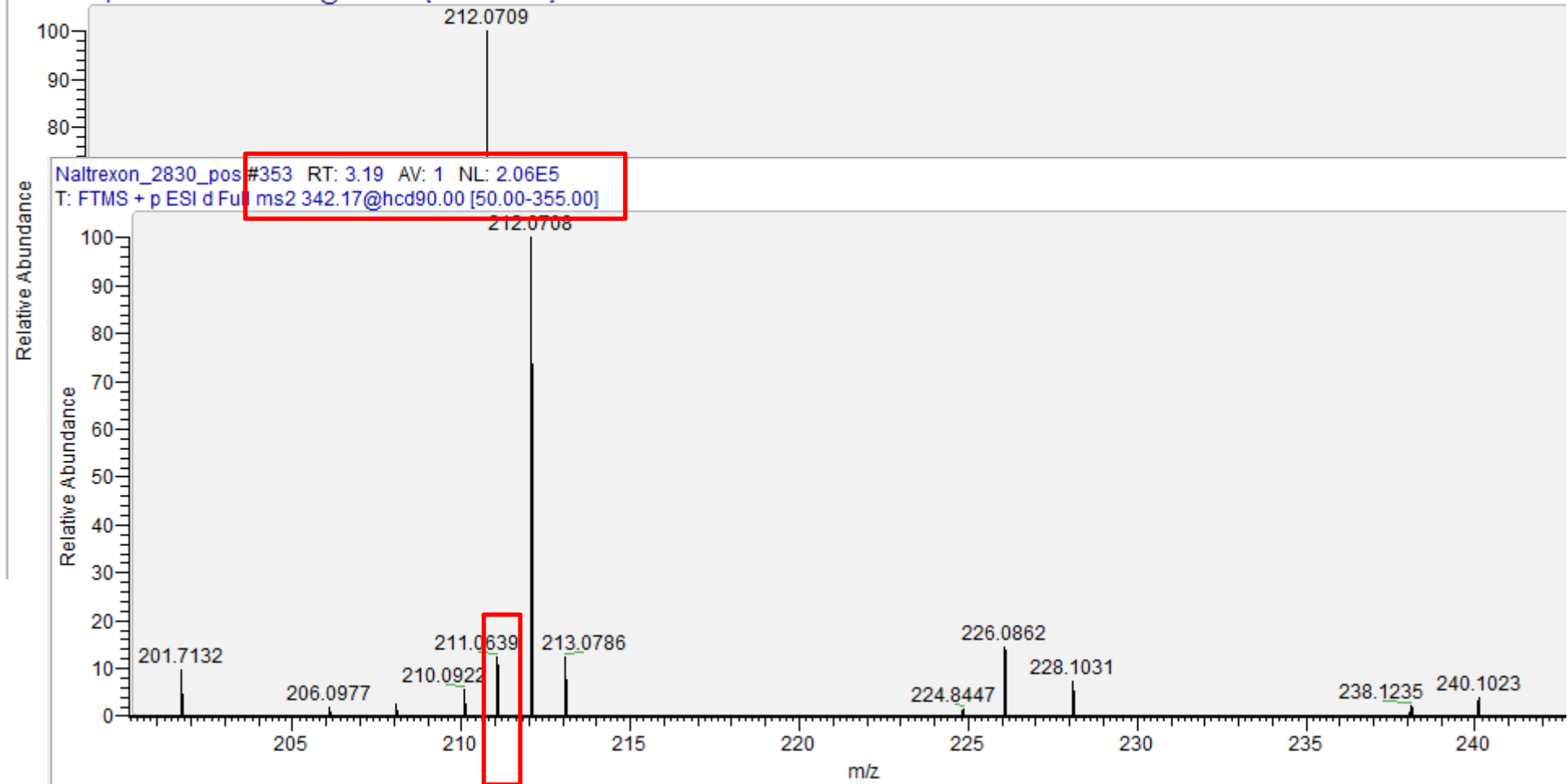
Fail Peak
(with formula)



Mass Spectrometry Workflow

Fail peak (with formula)

Naltrexon_2830_pos #352 RT: 3.19 AV: 1 NL: 1.95E5
T: FTMS + p ESI d Full ms2 342.17@hcd75.00 [50.00-355.00]



Mass Spectrometry Workflow

Fail Peaks

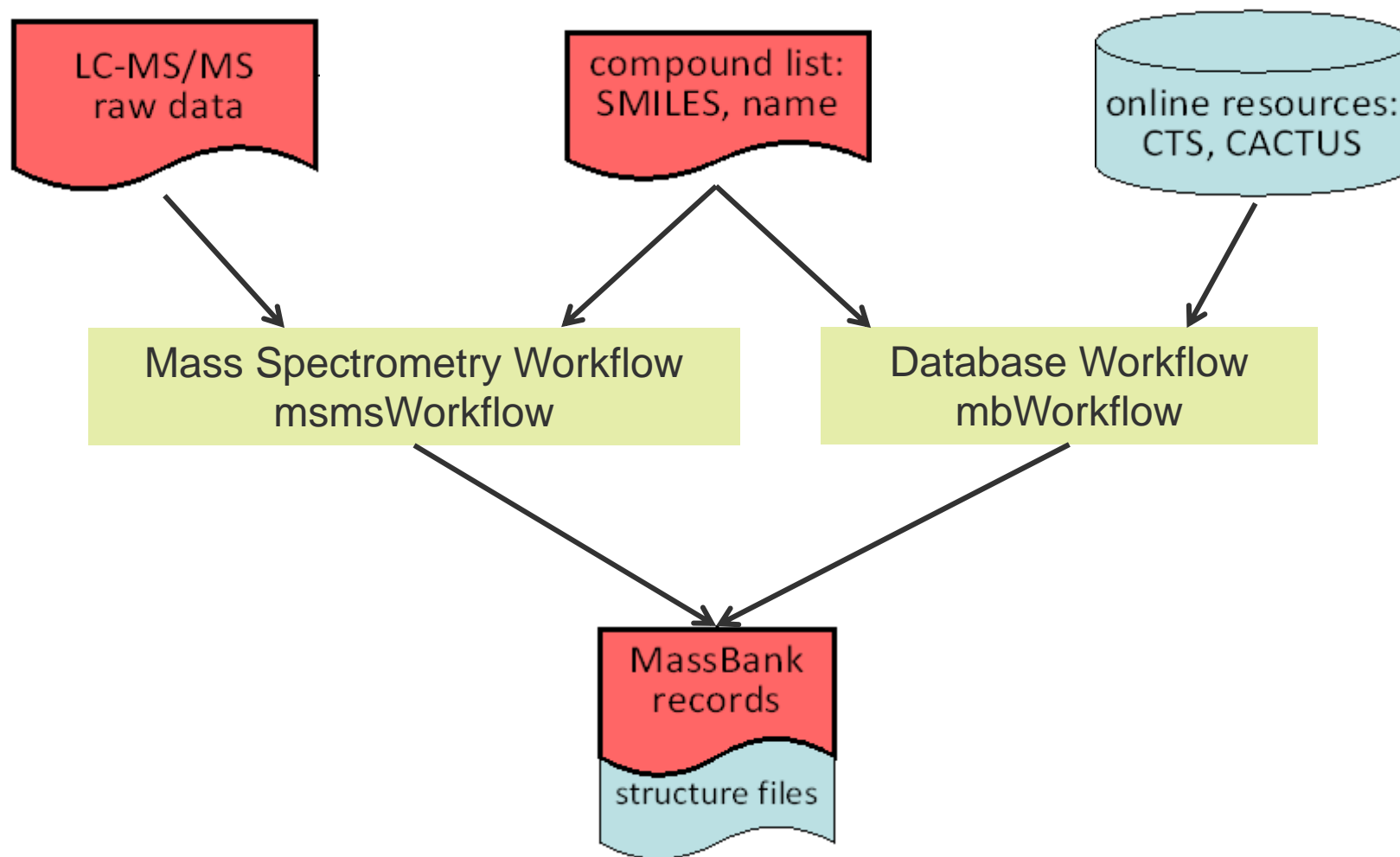
Include this “fail peak” by entering a 1 in the OK column

- Save under a new name...
 - e.g. “pH_narcotics_Failpeaks_wOKs.csv”

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	OK	name	cpdID	scan	mzFound	formula	mzCalc	dppm	dbe	mz	int	formulaC	parentSca	aMax	mzCe
2	0	407	2817	558	182.1191	NA	NA	NA	NA	182.1193	117786.4	0	556	9902212	304.:
3	0	2231	2821	963	233.1351	NA	NA	NA	NA	233.1355	255822.6	0	961	16383558	278.:
4	0	3491	2824	723	215.1468	NA	NA	NA	NA	215.1471	242559.4	0	721	15875343	272.:
5	0	3988	2825	725	249.1548	NA	NA	NA	NA	249.1553	516037.1	0	723	23457888	278.:
6	0	4110	2825	730	193.0993	NA	NA	NA	NA	193.0995	138038.3	0	723	5673618	278.:
7	0	5543	2828	938	265.1612	NA	NA	NA	NA	265.1618	146493.5	0	933	4158767	310.:
8	1	6503	2830	352	211.0738	C12H9N3C	211.074	-0.9633	13	211.0741	24055.95	1	346	195149.4	34
9															

We are now ready to go onto the Record Creation part of the workflow...

RMassBank Workflow – Simple Form



MassBank Record Workflow

Start a new MB workspace and reset the “Infolists” (compound information)

- Reset and load “Infolists” (compound information) – we will use a part-filled one

```
79
80 # Next, start the MassBank Record Workflow:
81 mb <- newMbWorkspace(w)
82 mb <- resetInfolists(mb)
83 # To speed up this example, we have provided a partially complete list
84 mb <- loadInfolists(mb, system.file("infolists_incomplete",
85                                   package="RMassBankData"))
```

- Add fail peaks, if applicable, and start the workflow:

```
90 # now add the fail peaks from above, if you have included any
91 mb <- addPeaks(mb, "pH_narcotics_Failpeaks_wOKs.csv")
92
93 # Start the record generation workflow
94 mb <- mbWorkflow(mb, infolist_path="./Narcotics_infolist.csv")
95 |
```

MassBank Record Workflow

Progress:

```
Console C:/DATA/RMassBank/RMassBank_Demo/ ↵
3
3
2824: smiles
3
3
2825: dbname
3
3
3
3
2758: smiles
3
3
2826: smiles
mbWorkflow: Step 2
The file ./Narcotics_infolist.csv was generated with new compound information. Please
check and edit the table, and add it to your infolist folder.
```

This shows the source of the data for querying CACTUS and CTS: Most come from SMILES, but the back-up of a “name” (dbname) is useful in some cases

There is new information to check manually
Please open the table in Excel/OpenOffice...

MassBank Record Workflow

Infolist Editing (1) - Names

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		id	dbcas	dbname	dataused	COMMENT	COMMENT	CH\$NAME	CH\$NAME	CH\$NAME	CH\$COMP	CH\$FORM	CH\$EXACT
2	1	2819	15532-75-1	1-(3-Triflu	smiles	standard c	2819	1-[3-(triflu	1-(3-(Trifluoromethyl		N/A; Envir	C11H13F3	230.1031
3	2	2821	50-48-6	Amitriptyl	smiles	standard c	2821				N/A; Envir	C20H23N	277.183
4	3	2822	300-62-9	Amphetar	smiles	standard c	2822	Amphetar	Amfetami (1-methyl		N/A; Envir	C9H13N	135.1048
5	4	2823	519-09-5	Benzoylec	dbname	standard c	2823	Benzoylec	None	3-benzoy	N/A; Envir	C16H19NC	289.1314
6	5	2817	50-36-2	Cocaine	dbname	standard c	2817	Cocaine	Unknown	(1S,3S,4R,	N/A; Envir	C17H21NC	303.1471
7	6	2824	125-71-3	Dextrome	smiles	standard c	2824				N/A; Envir	C18H25NC	271.1936
8	7	2825	30223-73-	EDDP (2-E	dbname	standard c	2825				N/A; Envir	C20H23N	277.183
9	8	2758	299-42-3	Ephedrine	smiles	standard c	2758	Ephedrine (1R,2S)-2-(methylam			N/A; Envir	C10H15NC	165.1154
10	9	2826	6740-88-1	Ketamine	smiles	standard c	2826				N/A; Envir	C13H16Cl	237.092
11													

Fix up the names here

- “None”, “Unknown” unnecessary
- At least one name should be present (copy from dbname)

H	I	J	K
CH\$NAME	CH\$NAME	CH\$NAME	CH\$COMP
1-[3-(triflu	1-(3-(Trifluoromethyl		N/A; Envir
Amitriptyline			N/A; Envir
Amphetar	Amfetami (1-methyl		N/A; Envir
Benzoylec	3-benzoyloxy-8-me		N/A; Envir
Cocaine	(1S,3S,4R,5R)-3-ben		N/A; Envir
Dextromethorphan			N/A; Envir
EDDP (2-Ethylidene-1,5-dimeth			N/A; Envir
Ephedrine (1R,2S)-2-(methylam			N/A; Envir
Ketamine			N/A; Envir

MassBank Record Workflow

Infolist Editing (2) – double-check CAS Numbers (if they were in original csv)

C	D	E	F	G	H	I	J	K	L	M	N	O	P
dbcas	dbname	dataused	COMMENT	COMMENT	CH\$NAME	CH\$NAME	CH\$NAME	CH\$COMP	CH\$FORM	CH\$EXACT	CH\$SMILE	CH\$IUPAC	CH\$LINK.C
9 15532-75-9	1-(3-Trifluoromethylphenyl)ethan-1-ol	smiles	standard c	2819	1-[3-(trifluoromethyl)phenyl]ethan-1-ol	1-(3-(Trifluoromethyl)phenyl)ethan-1-ol		N/A; Envir	C11H13F3O	230.1031	c1c(C(F)(F)F)c(C(F)(F)F)c1	InChI=1S/230.1031	15532-75-9
1 50-48-6	Amitriptyline	smiles	standard c	2821	Amitriptyline			N/A; Envir	C20H23N	277.183	C1(\c2c(Cc1ccccc1)ccn2)	InChI=1S/277.183	50-48-6
2 300-62-9	Amphetamine	smiles	standard c	2822	Amphetamine	Amfetami (1-methylamino)ethan-2-ol		N/A; Envir	C9H13N	135.1048	c1(ccccc1)CN	InChI=1S/135.1048	300-62-9
3 519-09-5	Benzoylcocaine	dbname	standard c	2823	Benzoylcocaine	3-benzoyloxy-8-methyl-8-azabicyclo[3.2.1]octane		N/A; Envir	C16H19NO	289.1314	CN1C2CCC2C1OC(=O)c3ccccc3	InChI=1S/289.1314	519-09-5
7 50-36-2	Cocaine	dbname	standard c	2817	Cocaine	(1S,3S,4R,5R)-3-benzoyloxy-N,1,2,3,4-tetrahydroisoquinoline		N/A; Envir	C17H21NO	303.1471	CN1C2CCC2C1OC(=O)c3ccccc3	InChI=1S/303.1471	50-36-2
4 125-71-3	Dextromethorphan	smiles	standard c	2824	Dextromethorphan			N/A; Envir	C18H25NO	271.1936	c12[C@]34CN5C=CC(=O)N(C5)C43	InChI=1S/271.1936	125-71-3
5 30223-73-8	EDDP (2-Ethylidene-1,5-dimethylpiperazine)	dbname	standard c	2825	EDDP (2-Ethylidene-1,5-dimethylpiperazine)			N/A; Envir	C20H29N	277.183	None	None	None
8 299-42-3	Ephedrine	smiles	standard c	2758	Ephedrine	(1R,2S)-2-(methylamino)propan-1-ol		N/A; Envir	C10H15NO	165.1154	c1([C@H](O)CNC)CC1	InChI=1S/165.1154	299-42-3
5 6740-88-1	Ketamine	smiles	standard c	2826	Ketamine			N/A; Envir	C13H16ClN	237.092	c1([C@H](O)CNC)CC1	InChI=1S/237.092	33643-46-8

MassBank Record Workflow

Infolist Editing (3) – reality-checking entries & delete “rubbish”

M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
CH\$EXACT	CH\$SMILE	CH\$IUPAC	CH\$LINK.C	CH\$LINK.C	CH\$LINK.F	CH\$LINK.F	CH\$LINK.I	CH\$LINK.I	CH\$LINK.I	CH\$LINK.I	CH\$LINK.I	CH\$LINK.I	CHEMSPIDER
230.1031	c1c(C(F)(F	InChI=1S/	15532-75-9						CID:4295	KKIMDKM	4145		
277.183	C1(\c2c(C	InChI=1S/	50-48-6	2666		D07448			CID:2160	KRMDCWI	2075		
135.1048	c1(ccccc1)	InChI=1S/	300-62-9	CHEBI	Delete this				CID:3007	KWTSXDU	13852819		
289.1314	CN1C2CCC	InChI=1S/	519-09-5						CID:2340	GVGYEFKI	2250		
303.1471	CN1C2CCC	InChI=1S/	50-36-2	27958		C01416			CID:44622	ZPUCINDJ	10194104		
271.1936	c12[C@]3	InChI=1S/	125-71-3			D03742			CID:69161	MKXZASY	13109865		
277.183	None	None		None	None	None	None		CID:None	None			
165.1154	c1([C@H]	InChI=1S/	299-42-3	15407		C01575			CID:9294	KWGRBVC	8935		
237.092	c1([C@@]	InChI=1S/	33643-46-8							YQEZLKZA	158414		

Check this - ChemSpider & PubChem ID usually similar

MassBank Workflow

Infolist Editing (4) – checking “Amphetamine” in ChemSpider

www.chemspider.com/Chemical-Structure.13852819.html?rid=ef48bb23-7628-4c07-96a9-f63f7aca5c18

RSC Synt

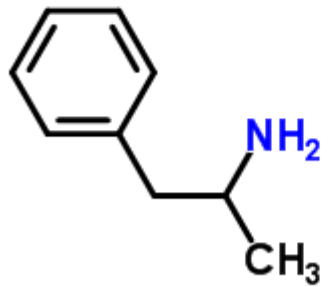
ChemSpider
The free chemical database

RSC


About | More Searches | Web APIs | Help

Retrieved ChemSpider ID is correct, so no changes necessary eg. Pyrid

Search term: **Amphetamine** (Found by approved synonym) ?



? 2D 3D Save Zoom

 - 0 of 1 defined stereocentres

Amphetamine

ChemSpider ID: **13852819**

Molecular Formula: C₉H₁₃N

Average mass: 135.206207 Da

Monoisotopic mass: 135.104797 Da

▼ Systematic name
1-Phenyl-2-propanamine

▶ SMILES and InChIs

▶ Cite this record

Want to comment on this record?

Leave Feedback

MassBank Record Workflow

Infolist Editing (5) – fill in missing entries for EDDP - search by CAS

M	N	O	P	Q	R	S	T	U	V	W	X
CH\$EXACT	CH\$SMILE	CH\$IUPAC	CH\$LINK.C	CH\$LINK.C	CH\$LINK.F	CH\$LINK.F	CH\$LINK.L	CH\$LINK.F	CH\$LINK.I	CH\$LINK.C	CHEMSPIDER
230.1031	c1c(C(F)(F	InChI=1S/	15532-75-9					CID:4296	KKIMDKM	4145	
277.183	C1(\c2c(C	InChI=1S/	50-48-6	2666		D07448		CID:2160	KRMDCWI	2075	
135.1048	c1(ccccc1)	InChI=1S/	306-63-8			D07445		CID:3007	KWTSXDU	13852819	
289.1314	CN1C2CCC	InChI=1S/	1519-09-5					CID:2340	GVGYEFKI	2250	
303.1471	CN1C2CCC	InChI=1S/	50-36-2	27958		C01416		CID:44622	ZPUCINDJ	10194104	
271.1936	c12[C@]3	InChI=1S/	125-71-3			D03742		CID:69161	MKXZASY	13109865	
277.183	None	None		None	None	None	None	CID:None	None		
165.1154	c1([C@H]	InChI=1S/	299-42-3	15407		C01575		CID:9294	KWGRBVC	8935	
237.092	c1([C@@]	InChI=1S/	33643-46-8						YQEZLKZA	158414	

Search ChemSpider and PubChem with CAS 30223-73-5 (in csv)




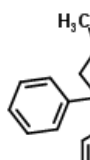

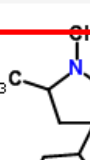

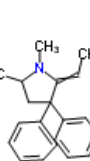
MassBank Record Workflow

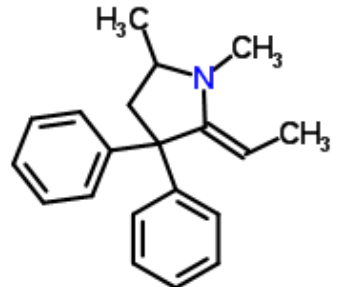
Infolist Editing (5) – EDDP - search ChemSpider by CAS: 30223-73-5



www.chemsp
About | More Searches | Web APIs | Help
eg. Pyridi

3 hits found.
Search terms: 30223-
(Found by synonym)

Grid Tile Tabl

ID	Structure
4526936  - 0/1 defined	
23254962  - 0/1 defined	
23621123  - 0/1 defined	



[2D](#) [3D](#) Save Zoom
 - Double-bond stereo
 - 0 of 1 defined stereocentres

Pyrrolidine, 1,5-dimethyl-2-ethylidene-3,3-diphenyl

ChemSpider ID: 4526936
 Molecular Formula: C₂₀H₂₃N
 Average mass: 277.40329 Da
 Monoisotopic mass: 277.1830

[Systematic name](#)
 (2Z)-2-Ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine

[SMILES and InChIs](#)

[Cite this record](#)

Want to comment on this record?

Leave Feedback

[SMILES and InChIs](#)
SMILES:
C/C=C\1/C(CC(N1C)C)(c2ccccc2)c3ccccc3

Std. InChI:
InChI=1S/C20H23N/c1-4-19-20(15-16(2)21(19)3,17-11-7-5-8-12-13-9-6-10-14-18/h4-14,16H,15H2,1-3H3/b19-4-

Std. InChIKey:
 AJRJPORIQGYFMT-PVOVUMCXSA-N

[Cite this record](#)

Names and Identifiers

Names and Synonyms [?](#)

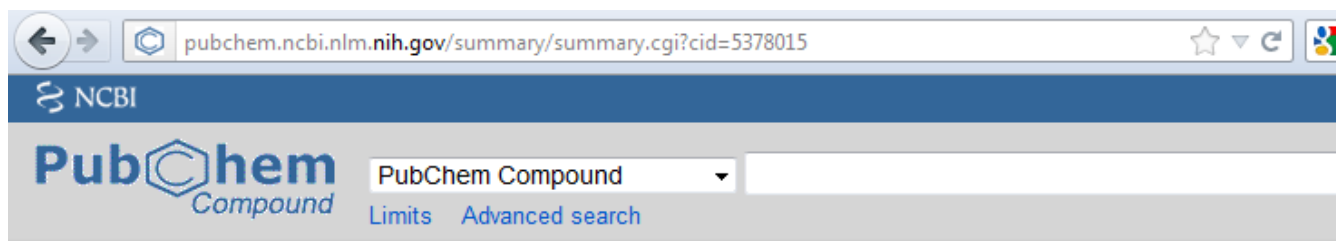
[Validated by Experts](#), [Validated by Users](#), [Non-Validated](#), [Removed by Experts](#)

EDDP	(2Z)-2-Ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine
	Pyrrolidine, 1,5-dimethyl-2-ethylidene-3,3-diphenyl

MassBank Record Workflow

=> <http://pubchem.ncbi.nlm.nih.gov/search/search.cgi#>

Infolist Editing (5) – EDDP - search PubChem by CAS: 30223-73-5



2-ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine - Compound



Summary (CID 5378015)

You can copy this straight from the URL!

Also known as: (2Z)-2-Ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine, 30223-73-5, EDDP, Eddp-3,3, 2-Et-1,5-Dime-3,3-DPP, AC1NTC6G

Molecular Formula: C₂₀H₂₃N Molecular Weight: 277.40332 InChIKey: AJRJPORIQGYFMT-PVOVUMCXSA-N

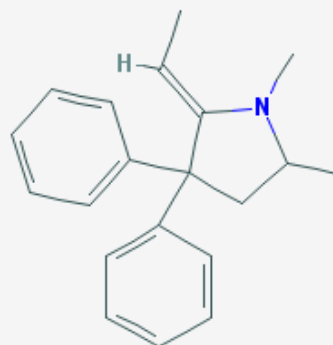
Table of Contents [Show subcontent titles](#)

- [Identification](#)
- [Related Records](#)
- [Biomedical Effects and Toxicity](#)
- [Literature](#)
- [Classification](#)
- [Chemical and Physical Properties](#)

[Expand all sub-sections](#)

2D Structure

3D Conformer



MassBank Record Workflow

Infolist Editing (5) – fill in missing entries for EDDP – now filled in?

135.1048	c1(ccccc1)	InChI=1S/	300-62-9	CHEBI		D07445		CID:3007	KWTSXDU	13852819	
289.1314	CN1C2CCC	InChI=1S/	519-09-5					CID:2340	GVGYEFKI	2250	
303.1471	CN1C2CCC	InChI=1S/	50-36-2	27958		C01416		CID:44622	ZPUCINDJ	10194104	
271.1936	c12[C@]3	InChI=1S/	125-71-3			D03742		CID:69161	MKXZASY	13109865	
277.183	None	None		None	None	None	None	CID:None	None		
165.1154	c1([C@H])	InChI=1S/	299-42-3	15407		C01575		CID:9294	KWGRBVC	8935	
237.092	c1([C@@])	InChI=1S/	33643-46-8						YQEZLKZA	158414	

M	N	O	P	Q	R	S	T	U	V	W	X
CH\$EXACT	CH\$SMILE	CH\$IUPAC	CH\$LINK.C	CH\$LINK.C	CH\$LINK.F	CH\$LINK.F	CH\$LINK.L	CH\$LINK.F	CH\$LINK.I	CH\$LINK.CHEMSPIDER	
230.1031	c1c(C(F)(F	InChI=1S/	15532-75-9					CID:4296	KKIMDKM	4145	
277.183	C1(\c2c(C	InChI=1S/	50-48-6	2666		D07448		CID:2160	KRMDCWI	2075	
135.1048	c1(ccccc1)	InChI=1S/	300-62-9			D07445		CID:3007	KWTSXDU	13852819	
289.1314	CN1C2CCC	InChI=1S/	519-09-5					CID:2340	GVGYEFKI	2250	
303.1471	CN1C2CCC	InChI=1S/	50-36-2	27958		C01416		CID:44622	ZPUCINDJ	10194104	
271.1936	c12[C@]3	InChI=1S/	125-71-3			D03742		CID:69161	MKXZASY	13109865	
277.183	C/C=C\1/C	InChI=1S/	30223-73-5					CID:53780	AJRJPORIC	4526936	
165.1154	c1([C@H])	InChI=1S/	299-42-3	15407		C01575		CID:9294	KWGRBVC	8935	
237.092	c1([C@@])	InChI=1S/	33643-46-8						YQEZLKZA	158414	

MassBank Record Workflow

Infolist Editing (6) – Check missing PubChem entry?

M	N	O	P	Q	R	S	T	U	V	W	X
CH\$EXACT	CH\$SMILE	CH\$IUPAC	CH\$LINK.C	CH\$LINK.C	CH\$LINK.F	CH\$LINK.F	CH\$LINK.L	CH\$LINK.F	CH\$LINK.I	CH\$LINK.CHEMSPIDER	
230.1031	c1c(C(F)(F	InChI=1S/	15532-75-9					CID:4296	KKIMDKM	4145	
277.183	C1(\c2c(C	InChI=1S/	50-48-6	2666		D07448		CID:2160	KRMDCWI	2075	
135.1048	c1(ccccc1)	InChI=1S/	300-62-9			D07445		CID:3007	KWTSXDU	13852819	
289.1314	CN1C2CC	InChI=1S/	519-09-5					CID:2340	GVGYEFKI	2250	
303.1471	CN1C2CC	InChI=1S/	50-36-2	27958		C01416		CID:44622	ZPUCINDJ	10194104	
271.1936	c12[C@]3	InChI=1S/	125-71-3			D03742		CID:69161	MKXZASY	13109865	
277.183	C/C=C\1/C	InChI=1S/	30223-73-5					CID:53780	AJRJPORIC	4526936	
165.1154	c1([C@H]	InChI=1S/	299-42-3	15407		C01575		CID:9294	KWGRBVC	8935	
237.092	c1([C@@]	InChI=1S/	33643-46-8						YQEZLKZA	158414	

pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=182137

NCBI
PubChem Compound

PubChem Compound
Limits Advanced search

(S)-Ketamine - Compound Summary (CID 182137)

Also known as: Esketamine, (-)-Ketamine, (S)-(-)-Ketamine, (-)-Ketamine, CHEBI:60799, cyclohexanone, 2-(2-chlorophenyl)-2-(methylamino)-, (2S)-, 33643-46-8

Molecular Formula: C₁₃H₁₆ClNO Molecular Weight: 237.72524 InChIKey: YQEZLKZALYSWHR-ZDUSSCGKSA-N

Links and Relations
Properties
Compound ID: 182137

MassBank Record Workflow

Infolist editing finished! Save under a new name and restart workflow

- Moving Infolists into a separate Infolist folder is recommended!

```

100
101 mb <- resetInfolists(mb)
102 mb <- loadInfolists(mb, "C:/DATA/RMassBank/InfoLists_demo")
103 # NOTE: if you didn't want to correct the whole list, use this one:
104 #mb <- loadInfolists(mb, system.file("infolists", package="RMassBankData"))
105 # and start the workflow again - "no new data added" is a good sign!
106 mb <- mk
107
> mb <- resetInfolists(mb)
> mb <- loadInfolists(mb, "C:/DATA/RMassBank/InfoLists_demo")
> # NOTE: if you didn't want to correct the whole list, use this one:
> #mb <- loadInfolists(mb, system.file("infolists", package="RMassBankData"))
> # and start the workflow again - "no new data added" is a good sign!
> mb <- mbWorkflow(mb)
mbWorkflow: Step 1
mbWorkflow: Step 2
No new data added.
mbWorkflow: Step 3
mbWorkflow: Step 4
Compiling: 1_3_Chlorophenyl_piperazin_2818_pos.mzML
Compiling: 1_3_Trifluoromethylphenyl_piperazin_2819_pos
Compiling: 1_Benzylpiperazin_2820_pos.mzML
Compiling: Amitriptylin_2821_pos.mzML
Compiling: Amphetamin_2822_pos.mzML
Compiling: Benzoylecgonin_2823_pos.mzML
3
3
3
mbWorkflow: Step 5
mbWorkflow: Step 6
mbWorkflow: Step 7
mbWorkflow: Step 8
>

```

And that's it – DONE!

MassBank Record Workflow

Take a look at your new records!

- You can open these in any text editor

The screenshot illustrates the MassBank record workflow. It shows a Windows File Explorer window with the following structure:

- Root: C:\DATA\RMassBank\RMassBank_Demo
- Subfolders:
 - EA (highlighted with a red box)
 - moldata
 - recdata (highlighted with a red box)

The 'recdata' folder contains a list of text files:

Name	Date modified	Type
EA275801.txt	20.11.2012 17:05	Text Document
EA275802.txt	20.11.2012 17:05	Text Document
EA275803.txt	20.11.2012 17:05	Text Document
EA275804.txt	20.11.2012 17:05	Text Document
EA275805.txt	20.11.2012 17:05	Text Document
EA275806.txt	20.11.2012 17:05	Text Document
EA275807.txt	20.11.2012 17:05	Text Document
EA275808.txt	20.11.2012 17:05	Text Document
EA275809.txt	20.11.2012 17:05	Text Document
EA275810.txt	20.11.2012 17:05	Text Document
EA275811.txt	20.11.2012 17:05	Text Document
EA275812.txt	20.11.2012 17:05	Text Document
EA275813.txt	20.11.2012 17:05	Text Document

MassBank Records “recdata”

```

EA275805.txt - Notepad
File Edit Format View Help
ACCESSION: EA275805
RECORD_TITLE: Ephedrine; LC-ESI-ITFT; MS2; 60%; R=7500; [M+H]+
DATE: 2012.11.20
AUTHORS: M. Stravs, E. Schymanski, H. Singer, Eawag
LICENSE: CC BY-SA
COPYRIGHT: Copyright (c) Eawag, 2012
COMMENT: CONFIDENCE standard compound
COMMENT: INTERNAL_ID 2758
CH$NAME: Ephedrine
CH$NAME: (1R,2S)-2-(methylamino)-1-phenyl-1-propanol
CH$COMPOUND_CLASS: N/A; Environmental standard
CH$FORMULA: C10H15NO
CH$EXACT_MASS: 165.1154
CH$SMILES: c1([C@H]([C@@H](N)C)C)C1=CC=CC=C1
CH$IUPAC: InChI=1S/C10H15NO/
-/m0/s1
CH$LINK: CAS 299-42-3
CH$LINK: CHEBI 15407
CH$LINK: KEGG C01575
CH$LINK: PUBCHEM CID:9294
CH$LINK: INCHIKEY KWGRBVOPPL
CH$LINK: CHEMSPIDER 8935
AC$INSTRUMENT: LTQ Orbitrap
AC$INSTRUMENT_TYPE: LC-ESI-IT
AC$MASS_SPECTROMETRY: MS_TY
AC$MASS_SPECTROMETRY: IONIZA
AC$MASS_SPECTROMETRY: ION_MO
AC$MASS_SPECTROMETRY: FRAGME
AC$MASS_SPECTROMETRY: COLLIS
AC$MASS_SPECTROMETRY: RESOLU
AC$CHROMATOGRAPHY: COLUMN_NA
AC$CHROMATOGRAPHY: FLOW_GRAD
5/95 at 25 min, 90/10 at 25.
AC$CHROMATOGRAPHY: FLOW_RATE
AC$CHROMATOGRAPHY: RETENTION
AC$CHROMATOGRAPHY: SOLVENT A
AC$CHROMATOGRAPHY: SOLVENT B
MS$FOCUSED_ION: BASE_PEAK 16
MS$FOCUSED_ION: PRECURSOR_M/Z 166.1226
MS$FOCUSED_ION: PRECURSOR_TYPE [M+H]+
MS$DATA_PROCESSING: RECALIBRATE loess on assigned fragments and MS1
MS$DATA_PROCESSING: WHOLE RmassBank
PK$ANNOTATION: m/z num {formula mass error(ppm)}
56.0495 1 C3H6N+ 56.0495 -0.46
70.0652 1 C4H8N+ 70.0651 0.49
91.0542 1 C7H7+ 91.0542 -0.51
93.07 1 C7H9+ 93.0699 1.22
115.0542 1 C9H7+ 115.0542 -0.58
117.0699 1 C9H9+ 117.0699 0.11
118.0647 1 C8H8N+ 118.0651 -3.86
132.081 1 C9H10N+ 132.0808 1.55
133.0886 1 C9H11N+ 133.0886 0.14
135.0804 1 C9H11O+ 135.0804 -0.68
148.1121 1 C10H14N+ 148.1121 -0.04
PK$NUM_PEAK: 11
PK$PEAK: m/z int. rel.int.
56.0495 151421.4 24
70.0652 152358.5 24
91.0542 150051.6 24
93.07 24399.9 3
115.0542 229068.4 37
117.0699 1548461.5 251
118.0647 12712.5 2
132.081 35161.8 5
133.0886 1423545.6 231
135.0804 94627.2 15
148.1121 6145796.5 999
//

```

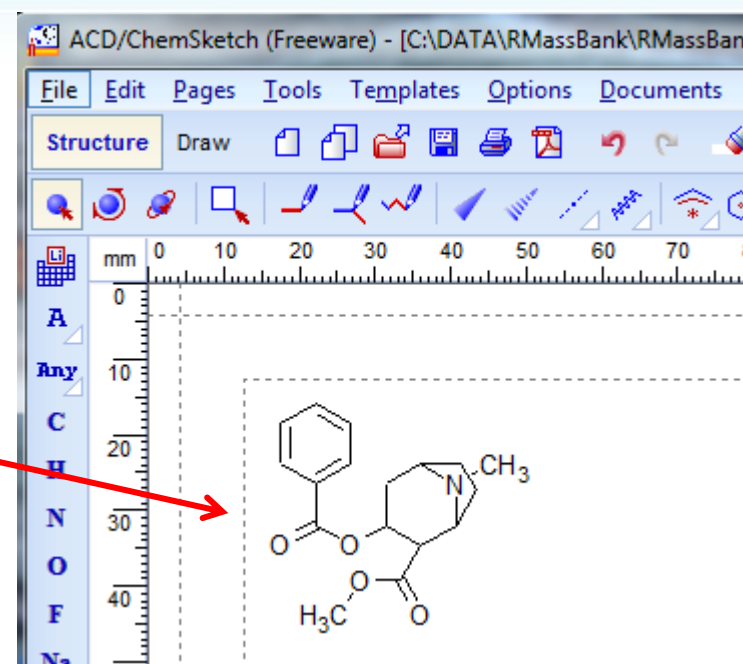
MassBank Record Structures “moldata”

RMassBank_Demo ▶ EA ▶ moldata

Search moldata

Share with ▾ Burn New folder

Name	Date modified	Type
2758.mol	20.11.2012 17:05	MOL File
2817.mol	20.11.2012 17:05	MOL File
2818.mol	20.11.2012 17:05	MOL File
2819.mol	20.11.2012 17:05	MOL File
2820.mol	20.11.2012 17:05	MOL File
2821.mol	20.11.2012 17:05	MOL File
2822.mol	20.11.2012 17:05	MOL File
2823.mol	20.11.2012 17:05	MOL File
2824.mol	20.11.2012 17:05	MOL File
2825.mol	20.11.2012 17:05	MOL File
2826.mol	20.11.2012 17:05	MOL File
2827.mol	20.11.2012 17:05	MOL File
2828.mol	20.11.2012 17:05	MOL File
2829.mol	20.11.2012 17:05	MOL File
2830.mol	20.11.2012 17:05	MOL File
list.tsv	20.11.2012 17:05	TSV



list.tsv - Notepad

File Edit Format View Help

```

1-(3-Chlorophenyl)piperazine 2818.mol
1-[3-(trifluoromethyl)phenyl]piperazine 2819.mol
1-Benzylpiperazine 2820.mol
Amitriptyline 2821.mol
Amphetamine 2822.mol
Benzoyllecgonine 2823.mol
Cocaine 2817.mol
Dextromethorphan 2824.mol
EDDP (2-Ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidin) 2825.mol
Ephedrine 2758.mol
Ketamine 2826.mol
Mephedrone 2827.mol
Methadone 2828.mol
Methamphetamine 2829.mol
Naltrexone 2830.mol

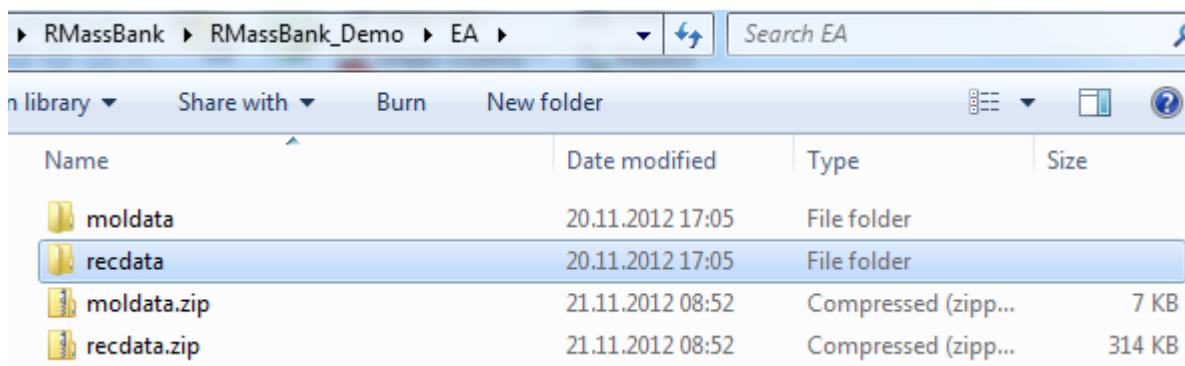
```

This is the list that connects structure and record

Final Tasks for MassBank records:

Create “recdata.zip” and “moldata.zip” (e.g. with 7Zip)

- This wasn't automated as this requires additional packages for Windows



These zip files are now ready for upload to MassBank

- If these are new records for NORMAN MassBank
 - Email these to massbank@normandata.eu
- If these are private records, you can upload to your own MassBank
 - I'll give a few screenshots in the next few slides

Uploading MassBank Records

Example with “uchem-massbank” screenshots

Eawag Uchem-MassBank

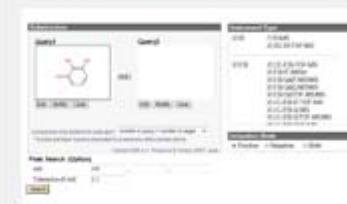
➔ Spectrum Search



➔ Quick Search



➔ Substructure Search



➔ Browse Page



➔ Peak Search



➔ Spectral Browser



➔ Record Index



- [MassBank Administration Tool](#)

Uploading MassBank Records

Example with “uchem-massbank” screenshots



MassBank Administration Tool

Main Menu

[Record Validator](#)

[Instrument Editor](#)

[Record Registration](#)

[Record List](#)

[Structure Registration](#)

[Structure List](#)

[Validator](#)

[SQL File Generator](#)

[File Upload](#)

[Record List Generator](#)

[Version Information](#)

[Database Manager](#)

[HOME](#)

Uploading MassBank Records

Record validation

Database : Record Version : 2 1 (old record version)

info: validation archive is [reodata.zip] or [*.msbk].

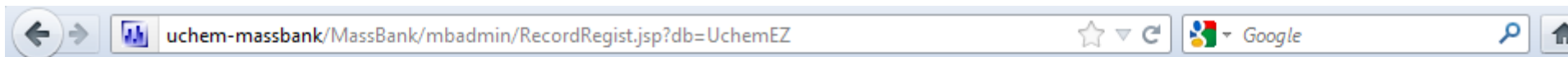
0 ok, 210 warn, 0 error / 210 files

Name	Status	Details
EA275801.txt	warn	value of required item [CH\$COMPOUND_CLASS:] is not compound class format.
EA275802.txt	warn	value of required item [CH\$COMPOUND_CLASS:] is not compound class format.
EA275803.txt	warn	value of required item [CH\$COMPOUND_CLASS:] is not compound class format.
EA275804.txt	warn	value of required item [CH\$COMPOUND_CLASS:] is not compound class format.
EA275805.txt	warn	value of required item [CH\$COMPOUND_CLASS:] is not compound class format.
EA275806.txt	warn	value of required item [CH\$COMPOUND_CLASS:] is not compound class format.
EA275807.txt	warn	value of required item [CH\$COMPOUND_CLASS:] is not compound class format.
EA275808.txt	warn	value of required item [CH\$COMPOUND_CLASS:] is not compound class format.
EA275809.txt	warn	value of required item [CH\$COMPOUND_CLASS:] is not compound class format.
EA275810.txt	warn	value of required item [CH\$COMPOUND_CLASS:] is not compound class format.
EA275811.txt	warn	value of required item [CH\$COMPOUND_CLASS:] is not compound class format.

Technically, this should define if the compound is a natural product or not; there is no way to automate this (unambiguously) – hence this warning
We print “N/A: Environmental Standard

Uploading MassBank Records

Record Registration



ADMIN	Record Validator	Record Registration	Structure Registration	File Upload	Sql File Generator	Version Information	HOI
MENU	Instrument Editor	Record List	Structure List	Validator	Record List Generator	Dababase Manager	

ADMIN	Record Validator	Record Registration	Structure Registration	File Upload	Sql File Generator	
MENU	Instrument Editor	Record List	Structure List	Validator	Record List Generator	

Databas

Record Registration

Record /

Database : Record Version : 2 1 (old record version)

Record Archive :

** please specify your [recdata.zip] or [*msbk].*

info : 210 record registered.

info : Done.

Uploading MassBank Records

Structure Registration

uchem-massbank/MassBank/mbadmin/StructureRegist.jsp?db=UchemEZ

ADMIN MENU

Record Validator | Record Registration | **Structure Registration** | File Upload | Sql File Generator

ADMIN MENU | Instrument Editor | Record List | Structure List | Validator | Record List Generator

Datab

Structure Registration

Database : UchemEZ

Structure Archive : Browse... Registration

** please specify your [moldata.zip] or [*msbk].*

info : 15 molfile registered.

info : Done.

Records Registered


Browse MassBank to see what they look like!

Record Index

[mass calculator](#)

[Home](#) | [Spectrum](#) | [Quick](#) | [Peak](#) | [Substructure](#) | [Browser](#) | [Browse](#) | [Index](#) | MassBank ID:

Contributor : [CASMI \(42\)](#) [Chubu Univ. \(2,628\)](#) [EA Uchem Orbi \(5,132\)](#)
[EA Uchem Orbi Test \(641\)](#) [EQ Uchem Q Ex \(1,262\)](#) [Eawag Uchem Adducts \(153\)](#)
[Fukuyama Univ. \(340\)](#) [IMM, CAMS & PUMC, China \(192\)](#) [IPB Halle \(528\)](#)

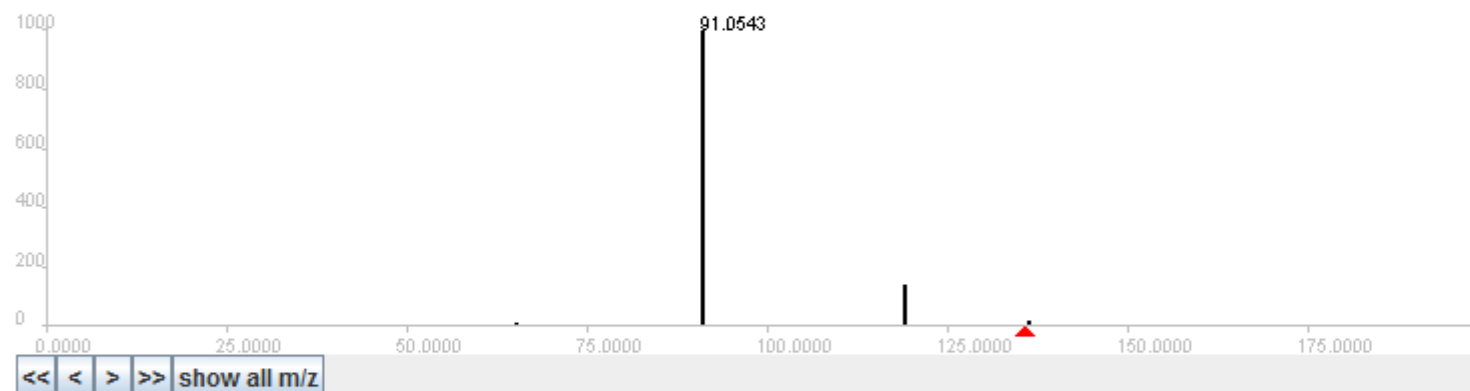
<input type="checkbox"/>	<input checked="" type="checkbox"/> Amphetamine	14 spectra	C9H13N	 135.10480	
<input type="checkbox"/>	LC-ESI-ITFT; MS2; 15%; R=15000; [M+H]⁺				EA282208
<input type="checkbox"/>	LC-ESI-ITFT; MS2; 15%; R=7500; [M+H]⁺				EA282202
<input type="checkbox"/>	LC-ESI-ITFT; MS2; 30%; R=15000; [M+H]⁺				EA282209
<input type="checkbox"/>	LC-ESI-ITFT; MS2; 30%; R=7500; [M+H]⁺				EA282203
<input type="checkbox"/>	LC-ESI-ITFT; MS2; 35%; R=15000; [M+H]⁺				EA282214
<input type="checkbox"/>	LC-ESI-ITFT; MS2; 35%; R=7500; [M+H]⁺				EA282201
<input type="checkbox"/>	LC-ESI-ITFT; MS2; 45%; R=15000; [M+H]⁺				EA282210
<input type="checkbox"/>	LC-ESI-ITFT; MS2; 45%; R=7500; [M+H]⁺				EA282204
<input type="checkbox"/>	LC-ESI-ITFT; MS2; 60%; R=15000; [M+H]⁺				EA282211
<input type="checkbox"/>	LC-ESI-ITFT; MS2; 60%; R=7500; [M+H]⁺				EA282205
<input type="checkbox"/>	LC-ESI-ITFT; MS2; 75%; R=15000; [M+H]⁺				EA282212
<input type="checkbox"/>	LC-ESI-ITFT; MS2; 75%; R=7500; [M+H]⁺				EA282206
<input type="checkbox"/>	LC-ESI-ITFT; MS2; 90%; R=15000; [M+H]⁺				EA282213
<input type="checkbox"/>	LC-ESI-ITFT; MS2; 90%; R=7500; [M+H]⁺				EA282207

MassBank Record: EA282210

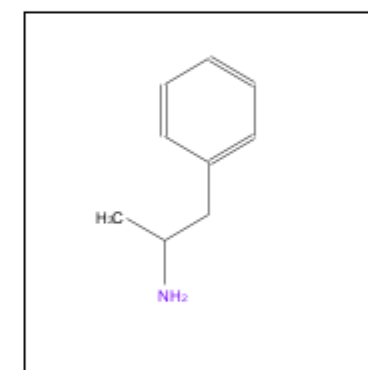
[Home](#) | [Spectrum](#) | [Quick](#) | [Peak](#) | [Substructure](#) | [Browser](#) | [Browse](#) | [Index](#) | MassBank ID:

Amphetamine; LC-ESI-ITFT; MS2; 45%; R=15000; [M+H]⁺

Mass Spectrum



Chemical Structure



ACCESSION: EA282210
 RECORD_TITLE: Amphetamine; LC-ESI-ITFT; MS2; 45%; R=15000; [M+H]⁺
 DATE: 2012.11.20
 AUTHORS: M. Stravs, E. Schymanski, H. Singer, Eawag
 LICENSE: CC BY-SA
 COPYRIGHT: Copyright (C) Eawag, 2012
 COMMENT: CONFIDENCE standard compound
 COMMENT: INTERNAL_ID 2822

CH\$NAME: Amphetamine
 CH\$NAME: Amfetamine

Take-Home Messages: RMassBank Demo

That was a 61-slide introduction to RMassBank

- Install (and update) RMassBank and necessary programs
- Run RMassBank for trial data
 - Get a basic understanding for the workflow
- Understand the manual checking required
 - “Fail peaks” – checking these in raw data
 - Automatic annotation with CTS – searching/editing this data
- Generate records with RMassBank
- A quick impression how to upload these records

It will be impossible to remember everything!

- vignette(“RMassBank”) – this contains an extensive explanation
- Stravs et al. 2012, *J. Mass Spectrom.*, DOI: 10.1002/jms.3131

Acknowledgements

- Marie-Curie Post Doctoral Fellowship (E. Schymanski),
- Eawag Discretionary Funds (M. Stravs)
- Coauthors: **Michael Stravs**, Heinz Singer, Juliane Hollender
- R & MassBank help: Steffen Neumann, Michael Gerlich, Carsten Kuhl, (IPB)
- Discussions: Matthias Ruff, Martin Loos (Eawag);
Tobias Schulze, Martin Krauss, Werner Brack (UFZ)
- MassBank & Naming Rights: Prof. Takaaki Nishioka
- NORMAN Association
- And thank you all for listening!



Any Questions?

massbank@normandata.eu (Tobias – record upload)

massbank@eawag.ch (RMassBank: E. Schymanski, M. Stravs)

<http://metabolomics-forum.com/viewforum.php?f=29>

emma.schymanski@eawag.ch

DOI: 10.1002/jms.3131

All Details Contained Within!