Curatr: a web application for creating, curating, and sharing a mass spectral library

Andrew Palmer (1), Prasad Phapale (1), Dominik Fay (1), Theodore Alexandrov (1,2)
(1) European Molecular Biology Laboratory, Heidelberg, Germany
(2) Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA, USA
Overview

This documentation accompanies the paper and details

- Webapp Overview
- Building a molecular library
- Recording authentic standard provenance
- Curating a dataset
- Export
Installation

The source code for curatr is available at [github.com/alexandrovteam/curatr](http://github.com/alexandrovteam/curatr)

Instructions for running an instance of curatr using the built in django server can be found along with the code.

This document contains a description of using curatr to build a spectral library, using the example of the European Molecular Biology Metabolomics Core Facility.

This library can be found at the base url [curatr.mcf.embl.de](http://curatr.mcf.embl.de) and all example links use this base url. If a personal instance is configured, this base url can be replaced.
Curatr Main Page

- Login to access authenticated areas
- Access spectral library
- Build spectral library
- Customisable institute name

Welcome

to the EMBL - Metabolomics Core Facility Spectral Library

Molecular Library
Explore the molecules and spectra contained in the EMBL - Metabolomics Core Facility spectral library

- Explore Molecules »

Curate Spectra
Add additional datasets and spectra to the library

- Start Curating »
Explore the library molecule-by-molecule

Explore the library spectrum-by-spectrum
Curatr Main Page

Data-set submission and curation

Standard provenance recording
Authentication
During installation, an ‘admin’ user is configured. They have the power to create new users through the admin link (e.g. for a local install http://localhost:8080/admin/ )

Users can log in through the link in the top right corner of the website

Successfully login is indicated by the welcome message
Building a molecular library
Steps in building a molecular library

Any authenticated user can add standards to the library.

- A standard is a purchased example of a particular molecule. Curatr supports multiple standards (e.g. from different manufacturers) for the same molecule.

Choose the ‘Add Standard’ option from the ‘Curate’ menu.

Standards can be added individually or as a batch.
Adding a single standard

Select link to switch to batch mode

Choose the molecule that corresponds to this standard

Record standard provenance information

Save when done
Choosing a molecule

- If this is the first standard of a particular molecule the molecular information must also be submitted.
- Otherwise, existing molecular information should be used.
Adding a single standard

Submitting molecular information

The name and sum molecular formula are **required**

We recommend that molecular information is completed as fully as possible.

Custom tags (e.g. molecule class, or protocol compatibility) can also be added.
Adding a single standard

Once a molecule is added it can be selected from the list

Add Standard

Curatr associates fragmentation spectra with a specific chemical source, this is usually a standard of a particular molecule purchased from a supplier.

You are currently adding a single standard, there is also the option to
Add Batch

Instructions
Complete all field in the form:

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Vendor</th>
<th>Lot num</th>
<th>Location</th>
<th>Purchase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tricetin (C2:0)</td>
<td>Tricaprylin (C8:0)</td>
<td>D-Sphingosine</td>
<td>1,3-Diolein 1,2-Dilinoleoyl-sn-glycerol-3-phosphoethanolamine</td>
<td>Adenosine monophosphate</td>
</tr>
<tr>
<td>Triacylglycerin (C6:0)</td>
<td>Tricaprin (C10:0)</td>
<td>Glyceryl trilaurate</td>
<td>1,2-Dilinoleoyl-sn-glycerol</td>
<td>Adenosine monophosphate</td>
</tr>
<tr>
<td>Tricaprylin (C8:0)</td>
<td>Tricaprin (C10:0)</td>
<td>Glyceryl trimyristate</td>
<td>1,3-Diolein</td>
<td>Adenosine monophosphate</td>
</tr>
<tr>
<td>Tricaprin (C10:0)</td>
<td>Glyceryl tricaprylin</td>
<td>Glyceryl tristearate</td>
<td>1,2-Dilinoleoyl-sn-glycerol</td>
<td>Adenosine monophosphate</td>
</tr>
<tr>
<td>Tricaprin (C10:0)</td>
<td>Glyceryl tricaprylin</td>
<td>Glyceryl tristearate</td>
<td>1,2-Dilinoleoyl-sn-glycerol</td>
<td>Adenosine monophosphate</td>
</tr>
<tr>
<td>Tricaprin (C10:0)</td>
<td>Glyceryl tricaprylin</td>
<td>Glyceryl tristearate</td>
<td>1,2-Dilinoleoyl-sn-glycerol</td>
<td>Adenosine monophosphate</td>
</tr>
<tr>
<td>Tricaprin (C10:0)</td>
<td>Glyceryl tricaprylin</td>
<td>Glyceryl tristearate</td>
<td>1,2-Dilinoleoyl-sn-glycerol</td>
<td>Adenosine monophosphate</td>
</tr>
</tbody>
</table>

Molecule not in the list? add it
Batch standards addition

- Batch addition is performed using a tab-separated-spreadsheet.
- There is an example supplied with curatr.
- Full molecular information must be supplied along with each standard.
  - Existing standards and molecules may be updated/overwritten.
  - Standards are linked to molecules first by Pubchem ID then by name.

Add Standards

Curatr associates fragmentation spectra with a specific chemical source, this is usually a standard of a particular molecule purchased from a supplier.

You are currently adding a batch standard, there is also the option to Add Single.

Instructions

The standards to add should be formatted as a tab delimited text file. An example file is here. This file must contain the fields: "id", "name", "formula", "inchi", "solubility", "vendor", "vendor_id", "hmdb_id", "chebi_id", "lipidmaps_id", "cas_id", "pubchem_id" but the values can be blank.

WARNING: This will overwrite any existing standards with identical IDs!

Tab delimited file:

Choose File: No file chosen

Save
Browsing and Searching by Molecule
Spectral Library -> Browse Molecules

All molecules added to the library are shown in one searchable table. Click on a name for full details and to see associated fragmentation spectra.

Filter the table by keywords or m/z

Click on a column title to sort

m/z for all adducts

<table>
<thead>
<tr>
<th>Name</th>
<th>Sum Formula</th>
<th>Exact Mass</th>
<th>Pubchem ID</th>
<th>Spectra Count</th>
<th>Tags</th>
<th>[M+H]</th>
<th>[M+Cl]</th>
<th>[M-H2O]</th>
<th>[M-H2O+H]</th>
<th>[M+H]*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tricaprin (C10:0)</td>
<td>C33H62O6</td>
<td>554.4546397</td>
<td>69310</td>
<td>18</td>
<td>553.44736325</td>
<td>589.42404098</td>
<td>535.43679857</td>
<td>537.45135147</td>
<td>555.46191616</td>
<td></td>
</tr>
<tr>
<td>Tricaprylin (C8:0)</td>
<td>C27H50O6</td>
<td>470.36073932</td>
<td>10850</td>
<td>18</td>
<td>469.35348267</td>
<td>506.3301406</td>
<td>451.34289818</td>
<td>453.35745109</td>
<td>471.36801577</td>
<td></td>
</tr>
<tr>
<td>INOSINE 5'-TRIPHOSPHATE</td>
<td>C10H15N4O14P3</td>
<td>507.97976216</td>
<td>8583</td>
<td>12</td>
<td>506.97248571</td>
<td>542.94916344</td>
<td>488.96192102</td>
<td>490.97647393</td>
<td>508.98703861</td>
<td></td>
</tr>
<tr>
<td>O-ACETYL-L-SERINE</td>
<td>C5H9NO4</td>
<td>147.0531777</td>
<td>99478</td>
<td>10</td>
<td>146.04888132</td>
<td>182.02255905</td>
<td>128.03531663</td>
<td>130.04986954</td>
<td>148.06043422</td>
<td></td>
</tr>
<tr>
<td>POPC</td>
<td>C42H82NO8P</td>
<td>759.57799462</td>
<td>99134</td>
<td>10</td>
<td>758.57053204</td>
<td>794.54720872</td>
<td>740.55996736</td>
<td>742.57452026</td>
<td>760.58508495</td>
<td></td>
</tr>
<tr>
<td>D-Sphingosine</td>
<td>C18H37NO2</td>
<td>299.28242942</td>
<td>5353955</td>
<td>9</td>
<td>298.27515297</td>
<td>334.2518307</td>
<td>280.26458829</td>
<td>282.27914119</td>
<td>300.28970588</td>
<td></td>
</tr>
<tr>
<td>Glyceryl trimristate</td>
<td>C45H86O6</td>
<td>722.84244047</td>
<td>11148</td>
<td>9</td>
<td>721.63516402</td>
<td>757.61184175</td>
<td>703.62459334</td>
<td>705.63915224</td>
<td>723.64971699</td>
<td></td>
</tr>
<tr>
<td>Hyodeoxycholic acid</td>
<td>C24H40O4</td>
<td>392.29265976</td>
<td>9963687</td>
<td>9</td>
<td>391.28538331</td>
<td>427.26206104</td>
<td>373.27481862</td>
<td>375.28937153</td>
<td>393.29993621</td>
<td></td>
</tr>
<tr>
<td>NICOTINAMIDE MONONUCLEOTIDE</td>
<td>C11H15N2O8P</td>
<td>334.05692044</td>
<td>14180</td>
<td>9</td>
<td>333.04932999</td>
<td>369.02600372</td>
<td>315.0387613</td>
<td>317.05331421</td>
<td>335.06387889</td>
<td></td>
</tr>
<tr>
<td>TAG</td>
<td>C53H100O6</td>
<td>832.75199388</td>
<td>9</td>
<td>8</td>
<td>831.74471738</td>
<td>867.721400015</td>
<td>813.73415269</td>
<td>815.7487056</td>
<td>833.75927028</td>
<td></td>
</tr>
</tbody>
</table>
Configure adducts
Configure Adducts

Adduct m/zs are precomputed for all molecules, so possible adducts must be configured in advance.

Adducts are supplied using the \([nM+D]^c\) nomenclature.

e.g.

- \([\text{M}+\text{H}-\text{H}_2\text{O}]^+\)
  - \(\text{NM} = 1\)
  - Delta formula = +H-H2O
  - Charge = 1

- \([\text{M}+\text{Cl}]^-\)
  - \(\text{NM} = 1\)
  - Delta formula = +Cl
  - Charge = -1

- \([2\text{M} + \text{H}]^+\)
  - \(\text{NM} = 2\)
  - Delta formula = +H
  - Charge = 1

Add adduct

| NM: | 1 |
| Delta formula: |
| Charge: | 1 |

Save | Cancel
Curating a dataset
Overview of curation steps

Select data file to upload

Select adducts and standards expected in the dataset
Hold ‘Ctrl’ or ⌘ to select multiple

Metadata fields: Optionally record LC and MS protocol IDs

Click ‘Upload’ to start processing

---

Upload Dataset

Process of submitting data:

1. Ensure the standards in the dataset are in the Standard Library
2. Convert the MS dataset to .mzML
3. Select the datafile (.mzml) and standards from the list below
4. Click ‘Upload’

Mzml file: [Choose File] No file chosen

Adducts:
1. [1M-H2O+H]+1
2. [1M-H]+1
3. [1M+H]+1
4. [1M+Na]+1

Standards:
1. L-Leucine
2. L-Arginine
3. L-HISTIDINE
4. L-ALANINE

MS1 mass accuracy (ppm): 10.0
Precursor Window (m/z): 1.0
LC Info: 
MS Info: 
Instrument Info (obsolete): 
Ionization Method: ESI
Ion Analyzer: QFT
Upload
Once processing has finished - click 'view' to start curating

Alternative route to submit a dataset
Dataset standards are shown in order of the number of fragmentation spectra.

Data-set details are shown, logged in users are also able to delete.
Extracted ion chromatogram with MS/MS
Additional information to help with selection

Fragmentation spectrum

Scan Number 831
Precursor m/z: 368.2568
Retention time: 1.50
Precursor Fraction: 1.00
collision energy 20.0 electronvolt

Accept: add to library
Several export formats are supported by default.

These can then be submitted to external spectral repositories.

<table>
<thead>
<tr>
<th>Export Type</th>
<th>Polarity to Export</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
</tr>
<tr>
<td>massbank</td>
<td>▼</td>
</tr>
<tr>
<td>metabolights</td>
<td>▼</td>
</tr>
<tr>
<td>tsv</td>
<td>▼</td>
</tr>
<tr>
<td>mgf</td>
<td>▼</td>
</tr>
<tr>
<td>msp</td>
<td>▼</td>
</tr>
</tbody>
</table>
Acknowledgments

Alexandrov Team (EMBL)

Dominik Fay
Andrew Palmer
Theodore Alexandrov

Metabolomics Core Facility (EMBL)

Prasad Phapale

Funding

This project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 634402.

Code Availability

The source code for curatr available from https://github.com/alexandrovteam/curatr and is distributed under the Apache 2.0 licence.