Guide for the contribution of spectra to MACE and the .mace-format

S. Schulz, March 8, 2022

Contribution to MACE is very welcome. The database is conceptually an add-on database with spectra not present in widely distributed mass spectral libraries.

The MACE database consists of EI-MS spectra obtained at 70 eV by GC/MS. Only spectra not present in the Wiley 7 or the NIST 17 databases [1] should be contributed. The contributed spectra should be preferably from compounds proven by synthesis, isolation or other methods that unequivocally confirm their structure, as well as their derivatives. The spectra will be checked for consistency before including them into the database to have a curated compound list, not filled with already known compounds.

The MACE library is provided as a simple text file using the NIST data structure [2]. The user can incorporate it into their local databases. A problem is that the ending of NIST data files is unfortunately .msp. This can lead to some difficulties in Windows. We therefore changed the ending to .mace to make the file distinct. Nevertheless, being a text file, the ending is not important.

A sample entry is shown below. The spectrum starts always with the name and is ended by an empty line. Preformatted fields are always on one line. At the end of the text after : the actual data input follows. These data can have different formats, for details see the NIST manual [2]. It is important to exactly following this document structure. Each field with a : has a single line. This is also true for the comments line. It is one long line. Line breaks in the comment line example in this document are only here for readability.

```
Name:  Dehydrojasmone
Synonyms:  4-Methylene-5-((Z)-2-penten-1-yl)-2-cyclopenten-1-one
Synonyms:  (Z)-4-Methylene-5-(pent-2-en-1-yl)cyclopent-2-en-1-one
InChIKey:  PRLXPOSTWLLTTJ-PLNGDYQASA-N
Formula:  C11H14O
MW: 162
ExactMass:  162.1045
Comments:  Contributor=P.Stamm;_S.Schulz;_TU-Braunschweig Spectrum_id=SC-37 RI=1300
   Phase=DB5-MS Reference=https://doi.org/10.1021/acs.joc.1c00145
   Smiles=O=C(C=C1)C(C/C=CCC)C1=C Mode=EI-quadrupole;_Agilent_MSD License=CC_BY-SA
   Compound_class=jasmone_derivative floral scent of Araceae CAS=2344-123-45
   Source=Scan_1137_SD467P.D
Num Peaks:  90
37 3; 38 17; 39 176; 40 30; 41 198; 42 12; 43 7; 45 1; 49 2; 50 39; 51 113; 52
96; 53 93; 54 14; 55 136; 56 6; 57 5; 58 4; 61 3; 62 15; 63 53; 64 18; 65 125;
66 102; 67 71; 68 44; 69 75; 70 5; 72 1; 73 1; 74 8; 75 7; 76 8; 77 238; 78 99;
79 154; 80 35; 81 25; 82 10; 83 2; 86 1; 87 3; 88 1; 89 19; 90 8; 91 363; 92 86;
93 23; 94 999; 95 84; 96 6; 98 1; 101 1; 102 11; 103 72; 104 24; 105 255; 106
48; 107 113; 108 45; 109 4; 115 86; 116 16; 117 42; 118 12; 119 175; 120 65; 121
19; 122 2; 127 7; 128 20; 129 25; 130 4; 131 25; 132 20; 133 503; 134 98; 135
```

9; 141 1; 143 3; 144 7; 145 10; 146 4; 147 99; 148 11; 149 1; 161 16; 162 172; 163 22; 164 2;

The different fields will be explained now and should be filled by the contributor:

Name: Always begins the entry. Use trivial or systematic name here for ease of use when appropriate, but not abbreviations.

Synonyms: Add systematic name here when trivial name is used. You can have several synonyms included, each one in one line. Please remove any stereodescriptors that can be associated with one enantiomer because MS cannot differentiate enantiomers. Please use the star convention if necessary to assign the relative configuration of compounds with more than one stereogenic center according to IUPAC nomenclature, e. g. (3*R**,5*S**) [3]. This procedure will avoid claims of identified enantiomers by GC/MS by inexperienced users.

InChIKey: Draw the compound in Chemdraw or other drawing package and select InChIKey from the *Copy as* menu and insert here. This key is used in many web and other applications. It allows direct connection e. g. to PubChem in MSSearch.

Formula, MW and ExactMass: Add data here, e. g. copied from Chemdraw Analysis window. MsSearch will not work when these data are missing.

Comments: **Only one line!** The Comments field includes tags. These tags are followed by a = and must not have spaces. Any text outside tags will occur in the output under Comments, the rest in the appropriate fields. The data are best visible when the database is configured to show these fields. The comments line is one long line of characters in the input file. Spaces in tag entries must be connected by a _ character.

The tags should be replaced with appropriate entries in the sample file.
`Contributor=` Replace P.Stamm with your own name.
`RI=` Retention index of compound.
`Phase=` Specify GC-phase used for RI calculation.
`Mode=` Specify whether quadrupol or TOF data, etc, and brand of used instrument.
`Reference=` The doi reference of the publication of the original spectrum, preferably from an authentic sample.
`Smiles=` Obtained e. g. from chemdraw as described for the InChikey. Allows easy drawing of the structure by copying it into drawing programs.
`Licence=` stays as it is. Allows usage by others, but not use in a commercial product.
`Class=` specify compound class. The next, unconnected words occur in the comments line with no tags, origin of the compound.
`Spectrum=` is an internal number given by me, one can leave this out. It will be given when the compound is added to MACE.
`CAS=` add CAS number if known. The number can be easily obtained by importing the Smiles code into Scifinder and performing a structure search. Please be aware that some double bond formatting might be necessary in Scifinder.
`Source=` Add origin of the data from your local files, if you want.

The data are better when all tags are filled. Nevertheless, one can sent the data also with some fields open.

Num Peaks: Number of peaks to follow. Usually already in the data when mass spectrum is exported into the text file. Absolutely necessary for correct function of the file.

Then the actual data are following, which do not need to be on one line. They can also have slightly different format [3].

Copy all your compounds into one document. Every compound entry begins with Name, preceded by at least one blank line. Save as .mace file or other text format. Send to **mace@tu-braunschweig.de** via email.

All the compounds will be added to the MACE library for public use in the open access data repository of TU Braunschweig after a quality check. MACE: Mass Spectra for Chemical Ecology.

Generating .mace files via NIST MSSearch

The data for submitting a mass spectrum can be obtained via the Librarian tool of MSSearch in the NIST program. The peak is selected from a GC/MS-run. A library search is then performed using the NIST program. A good spectrum should be selected, maybe performing baseline subtraction first. A right click on the spectrum opens MSSearch. Select the `Librarian` tool. Go to the small export icon and press export. Create a file in .msp format with the name of the compound as title. Select overwrite.

Now open the .msp file in an editor that allows long lines, e. g. notepad. The file locks like this:

```
Name:  Scan 33763 (261.393 min):  AMN037.D...
DB: 15
Num Peaks:  224
38 1; 39 33; 40 6; 41 200; 42 38;
```

Replace the scan... part with the name of the compound. You can save the scan data for later use in the `Source` tag. Then replace the `DB:` line with the header obtained from the sample file and fill out the respective information.

Copy all single spectrum files into one document with the .mace ending.

I am sure other programs allow similar transformations as well. Send me routines to be added to this document. JCAMP-DX files and other formats can be converted into NIST format by the Lib2Nist program [4]. Therefore, one can export a spectrum via JCAMP-DX and convert it via Lib2Nist.

References and Notes

[1] We opted not to use the newest library versions of NIST and Wiley, because they are likely less widespread in labs than older versions.

[2] *NIST MS Search User Guide*, section "NIST Text Format of Individual Spectra": http://chemdata.nist.gov/mass-spc/ms-search/docs/Ver20Man_11.pdf

[3] https://www.acdlabs.com/iupac/nomenclature/93/r93$_6$35.htm according to IUPAC, Commission on Nomenclature of Organic Chemistry. A Guide to IUPAC Nomenclature of Organic Compounds (Recommendations 1993), 1993, Blackwell Scientific publications

[4] https://chemdata.nist.gov/mass-spc/ms-search/Library_conversion_tool.html

Stefan Schulz, Anton Möllerke March 8, 2022