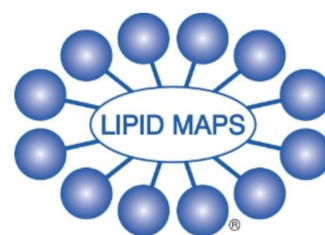2023

LIPID MAPS®
Curation of lipids into LMSD

Lipids of biological relevance are curated via the dedicated LMSD curation interface. The primary data item is the molfile structure of the lipid drawn in the Ketcher software. From this, other data are derived (formula, mass, SMILES etc). It follows therefore that lipids which can be represented correctly as a molfile are eligible for inclusion. Those not characterized fully (eg unknown location of a double bond, or ambiguous distribution of acyl chains) are ineligible for inclusion in LMSD.

When curating from the scientific literature, a check is made that the experiments conducted are sufficient to characterize the molecule to the level of detail claimed. For instance, there is NMR, ECD, synthetic standard, or chemical evidence used to determine stereochemistry. Untargeted mass spectroscopy, with just a mass value, is not sufficient to fully characterize a molecule. Detailed review (for example, to verify chemical shift values are correct for the chemistry of each atom) is not practicable.

Lipids are drawn in a standard orientation. This is automatically generated if created from the abbreviation via the structure drawing application in the curation interface, for more common types of lipid. For other types, refer to existing lipids in LMSD and maintain consistency.

Classification
Usually, the classification of a lipid is clear, however when a lipid can go in multiple classifications, we observe a rule where the one with highest functionality should be the primary class. For example, Testosterone glucuronide would have a primary classification of a glucuronide and a secondary classification as a C19 steroid. In the case of functionalized fatty acids, primary classification goes to the most descriptive if applicable (eg octadecanoid, resolvin, prostaglandin) otherwise the primary class should be that of the highest priority group according to IUPAC (oxo > hydroxy > unsaturation). Remaining functional groups are assigned to secondary classification.

In rare cases when a lipid could exist in more than one of the eight main categories, (eg LMST05050031) then which is the main one is potentially arbitrary. To combat this, we assign the highest numerical category as the primary (eg polyketide >sterol >glycerolipid)

The abbreviation (sum composition), if applicable, is calculated by the system after selecting the correct headgroup from the pulldown menu.

Using the UniChem API (www.ebi.ac.uk/unichem/), crosslinks are added to PubChem, ChEBI, HMDB, SwissLipids, KEGG and PDBe at the time of curation if the molecule is already extant in any these resources. Periodically, crosslinks to these resources are updated.

Names and systematic names are input manually.

It is important to note the lipid classification follows the desires of the lipid community, and where lipids are classified or grouped, the LM classification is true to the community definition.

Provenance
The paper from which the lipid structure has been curated is recorded. The curation interface can populate these data from the PubMed ID. The NCBI taxonomy number of the species in which the lipid is found is also recorded. Should a lipid be found in other species, these citations can also be added. As a general rule, this would only be if the taxonomic class was different.

User deposition
Users can submit lipid structures at lipidmaps.org/new/reg/ drawing the structure (submitted subsequently as a molfile) along with the classification and provenance. Curation staff will review the submission, checking accuracy (eg do the structure and name agree) and modifying as necessary to standardise the orientation, classification etc. The deposition will go live in LMSD once the paper describing it is published.