
ChemTastesDB - version 2.1

ChemTastesDB is a database that includes curated information of 4075 molecular tastants. ChemTastesDB is distributed to the scientific community to expand the information of molecular tastants, which could assist the analysis of the relationships between molecular structure and taste, as well as *in silico* (QSAR/QSPR) studies for taste prediction. Examples of QSPR approaches for the prediction of molecular taste are given in the following publication: Rojas, C., Abril-González, M., Ballabio, D. & García, F. (2025). *ChemTastesPredictor*: An ensemble of machine learning classifiers to predict the taste of molecular tastants. *Chemometrics and Intelligent Laboratory Systems*. 261, 105380. <https://doi.org/10.1016/j.chemolab.2025.105380>.

The 4075 molecular tastants are categorized into one of the five basic tastes (sweet, bitter, umami sour and salty), as well as to other classes related to non-basic tastes (tasteless, non-sweet, non-bitter, multitaste and miscellaneous). The molecules are categorized into following ten classes: sweet (1313), bitter (1615), umami (220), sour (49), salty (16), multitaste (179), tasteless (232), non-sweet (304), non-bitter (28), and miscellaneous (119).

ChemTastesDB provides the following information for each molecule: name, PubChem CID, CAS registry number, canonical SMILES string, class taste and the reference to the scientific sources from where data were retrieved. In addition, the molecular structure in the HyperChem (.hin) format of each compound is provided.

This is version 2.1 of the ChemTastesDB. In this new version, 1131 newly curated compounds were added. These new molecules were retrieved from 52 new bibliographic references.

CONTENT

Four files are provided:

1. "ChemTastesDB_readme.pdf" [datafile], containing a complete description of the ChemTastesDB.
2. "ChemTastesDB_database.xlsx" [datafile], containing information on the individual molecular tastant, that is, molecular ID, name, PubChem CID, CAS registry number, canonical SMILES string, class taste and the reference to the scientific sources from where records were retrieved.
3. "ChemTastesDB_references.xlsx" [datafile], containing the comprehensive list of all scientific references with their extended details. This information is provided as a code including first author surname and year of the reference. For instance, Charoenkwan2020 refers to the following citation: "Charoenkwan, P., Yana, J., Nantasenamat, C., Hasan, M. M., & Shoombuatong, W. (2020). iUmami-SCM: A novel sequence-based predictor for prediction and analysis of umami peptides using a scoring card method with propensity scores of dipeptides. *Journal of Chemical Information and Modeling*, 60(12), 6666-

6678". For molecules retrieved from books and book chapters, the page number is indicated in brackets [pp], while specific codes found in scientific references are shown in parentheses (cod). For example:

- a) The AMLEQVAMTDK amino acid sequence is referred to Wang2022_(AK11), that is, this compound is labelled as AK11 in the paper "Wang, W., Yang, L., Ning, M., Liu, Z., & Liu, Y. (2022). *A rational tool for the umami evaluation of peptides based on multi-techniques. Food Chemistry, 371, 131105*".
- b) The 3-O-caffeoyl- γ -quinic acid lactone compound is referred as Belitz2009_[948], that is, this molecule is available in page 948 of the book "Belitz, H.-D., Grosch, W., & Schieberle, P. (2009). *Food chemistry (Fourth ed.): Springer*".
- c) The 3-Acetoxy-5,7-dihydroxy-4'-methoxyflavanone molecular tastant is referred as Kinghorn2010_(89)/[278], that is, this compound could be found as code number 89 in page 278 of the chapter book "Kinghorn, A. D., Chin, Y. W., Pan, L., & Jia, Z. (2010). *Natural products as sweeteners and sweetness modifiers. In H.-W. Liu & L. Mander (Eds.), Comprehensive natural products II: Chemistry and biology, vol. 3 (pp. 269-315): Elsevier*".

4. "ChemTastesDB_molecules.rar" [datafile], containing the .hin Hyperchem file of each molecular tastant optimized by the mechanics force field (MM+). Files are named as the molecular IDs of the ChemTastesDB_database excel file.

Additional details on the content and curation pipeline can be found in the following papers:

- a) Rojas, C., Ballabio, D., Pacheco Sarmiento, K., Pacheco Jaramillo, E., Mendoza, M., & García, F. (2022). *ChemTastesDB: A curated database of molecular tastants. Food Chemistry: Molecular Sciences, 4, 100090*. <https://doi.org/10.1016/j.fochms.2022.100090>.
- b) Rojas, C., Abril-González, M., Ballabio, D. & García, F. (2025). *ChemTastesPredictor: An ensemble of machine learning classifiers to predict the taste of molecular tastants. Chemometrics and Intelligent Laboratory Systems. 261, 105380*. <https://doi.org/10.1016/j.chemolab.2025.105380>.

CONDITIONS

The database is freeware and may be used if proper reference is given to the authors. Preferably refer to the following papers:

1. Rojas, C., Ballabio, D., Pacheco Sarmiento, K., Pacheco Jaramillo, E., Mendoza, M., & García, F. (2022). *ChemTastesDB: A curated database of molecular tastants. Food Chemistry: Molecular Sciences, 4, 100090*. <https://doi.org/10.1016/j.fochms.2022.100090>.
2. Rojas, C., Abril-González, M., Ballabio, D. & García, F. (2025). *ChemTastesPredictor: An ensemble of machine learning classifiers to predict the taste of molecular tastants. Chemometrics and Intelligent Laboratory Systems. 261, 105380*. <https://doi.org/10.1016/j.chemolab.2025.105380>.

LICENCE

The ChemTasteDB is distributed with a Creative Commons Attribution 4.0 licence. The Creative Commons Attribution license allows re-distribution and re-use of a licensed work on the condition that the creator is appropriately credited. Read more here: <https://creativecommons.org/licenses/by/4.0/legalcode>

You are free to share - copy and redistribute the material in any medium or format. The licensor cannot revoke these freedoms as long as you adhere the following license terms:

Attribution - You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

NonCommercial - You may not use the material for commercial purposes.

NoDerivatives - If you remix, transform, or build upon the material, you may not distribute the modified material.

WEB CONTACTS

Please contact us with comments or questions, or if you have synthesized or identified new molecular tastants to be included in future versions of the ChemTastesDB. Also, let us know if you find any bugs we may have missed!

Contact information: Cristian Rojas, Grupo de Investigación en Quimiometría y QSAR, Facultad de Ciencia y Tecnología, Universidad del Azuay, crojasvilla@gmail.com.

HAVE FUN!!!
