

new: “CAS Common Chemistry in 2021: Expanding Access to Trusted Chemical Information for the Scientific Community”

Egon Willighagen 

Published May 22, 2022

Citation

Willighagen, E. (2022). new: “CAS Common Chemistry in 2021: Expanding Access to Trusted Chemical Information for the Scientific Community”. In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/wnect-mj679>

Keywords

Curation, Chemistry, Cas, Bioclipse, Cdk

Abstract

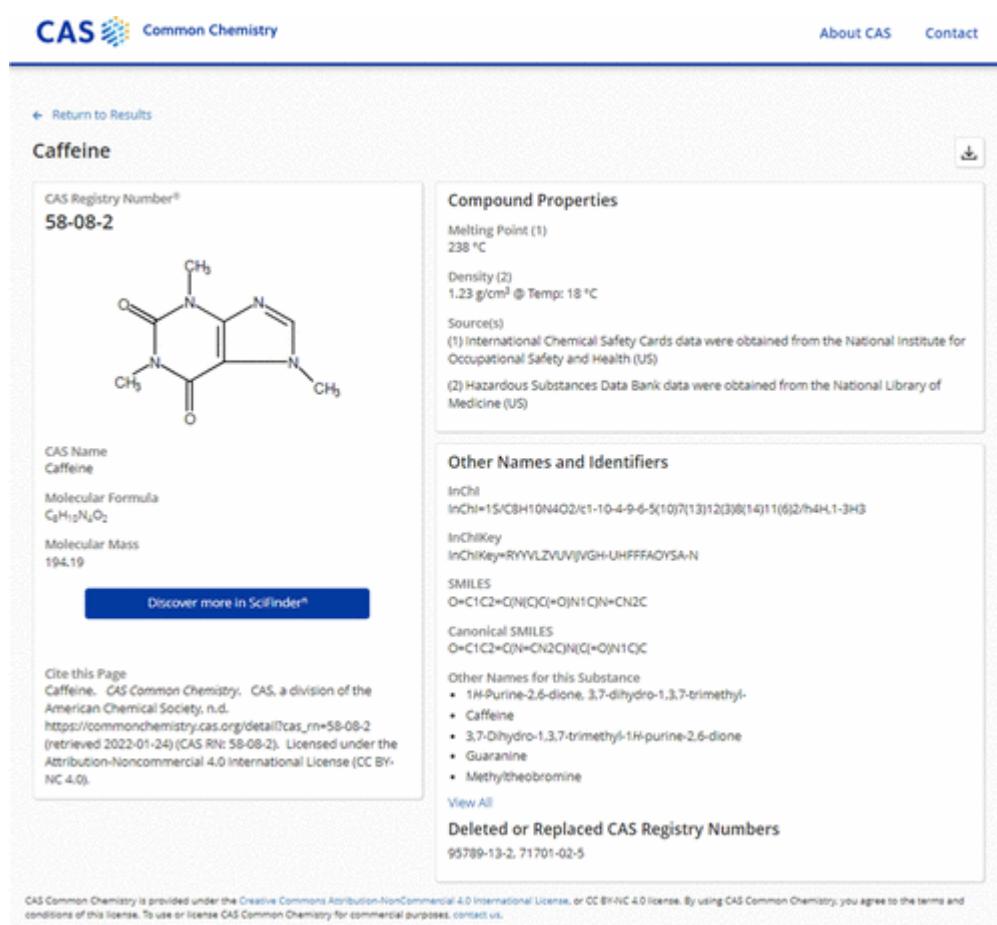
Open Science is happening. The merits are no longer theoretical or idealistic but tangible. Research is faster than ever, more vetted than ever (think PubPeer), more cited than ever. Fairly, not just because of Open Science, but open access causes readership causes impact causes citations. When new people and organizations start adopting Open Science this warms my hearth.

Copyright

Copyright © Egon Willighagen 2022. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Open Science is happening. The merits are no longer theoretical or idealistic but tangible. Research is faster than ever, more vetted than ever (think PubPeer), more cited than ever. Fairly, not just because of Open Science, but open access causes readership causes impact causes citations. When new people and organizations start adopting Open Science this warms my hearth.

So, when I was asked to work with [Chemical Abstracts Service](#) (CAS) on a new, bigger than ever version of [Common Chemistry](#) (which started as a project between CAS and Wikipedia), I welcomed the project. I don't quite remember the first meetings, but roughly my task became to work with the new content and match this against Wikidata and Wikipedia. It aligned well with [BridgeDb](#), [Scholia](#), and our metabolomics research, so I even could find sufficient research time for it. This work is now published in the [JChem: CAS Common Chemistry in 2021: Expanding Access to Trusted Chemical Information for the Scientific Community](#) (doi:[10.1021/acs.jcim.2c00268](https://doi.org/10.1021/acs.jcim.2c00268)).



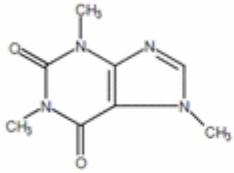
CAS Common Chemistry

About CAS Contact

Return to Results

Caffeine

CAS Registry Number[®]
58-08-2



Compound Properties

Melting Point (1)
238 °C

Density (2)
1.23 g/cm³ @ Temp: 18 °C

Source(s)
(1) International Chemical Safety Cards data were obtained from the National Institute for Occupational Safety and Health (US)
(2) Hazardous Substances Data Bank data were obtained from the National Library of Medicine (US)

CAS Name
Caffeine

Molecular Formula
C₈H₁₀N₄O₂

Molecular Mass
194.19

Discover more in SciFinder®

Cite this Page
Caffeine. CAS Common Chemistry. CAS, a division of the American Chemical Society, n.d.
https://commonchemistry.cas.org/detail/cas_rn=58-08-2 (retrieved 2022-01-24) (CAS RN: 58-08-2). Licensed under the Attribution-Noncommercial 4.0 International License (CC BY-NC 4.0).

Other Names and Identifiers

InChI
InChI=1S/C8H10N4O2/c1-10-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3

InChIKey
InChIKey=RYYVLZUVVJVGH-UHFFFAOYSA-N

SMILES
O=C1C2=C(N(C)C(=O)N1C)N=C2N

Canonical SMILES
O=C1C2=C(N(CN2)C)N(C(=O)N1)C

Other Names for this Substance

- 1H-Purine-2,6-dione, 3,7-dihydro-1,3,7-trimethyl-
- Caffeine
- 3,7-Dihydro-1,3,7-trimethyl-1H-purine-2,6-dione
- Guaranine
- Methyltheobromine

[View All](#)

Deleted or Replaced CAS Registry Numbers

95789-13-2, 71701-02-5

CAS Common Chemistry is provided under the Creative Commons Attribution-NonCommercial 4.0 International License, or CC BY-NC 4.0 license. By using CAS Common Chemistry, you agree to the terms and conditions of this license. To use or license CAS Common Chemistry for commercial purposes, contact us.

Figure 2 from the article. Detailed record for caffeine in CAS Common Chemistry (image: CC-BY).

About Wikidata, the paper writes (CC-BY):

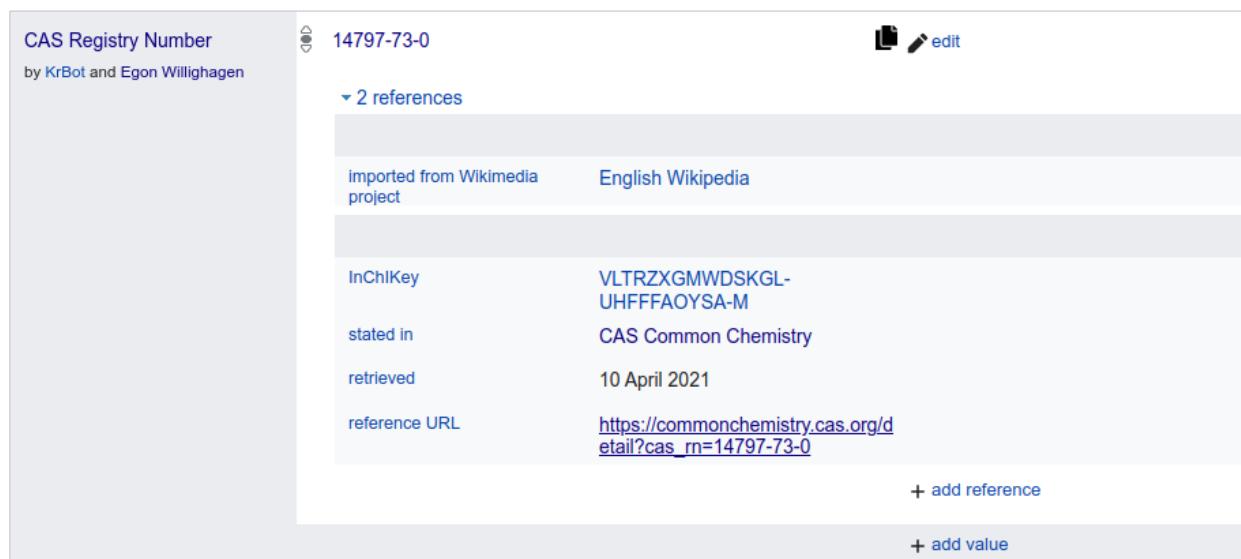
The latest release of CAS Common Chemistry has also supported updates and corrections to CAS RNs in Wikidata and Wikipedia. (22) InChIKeys were calculated from CAS SMILES using Bacting 0.0.31 (23) with the Chemistry Development Kit 2.7.1 (24) and were matched with content in Wikidata. The CAS RNs were then compared. References to CAS Common Chemistry were added for CAS RNs that matched. Mismatches have been shared with the Wikidata and

chem-bla-ics

Wikipedia communities so that they can manually review and correct the misleading entries using CAS Common Chemistry as a reference. Because Wikidata also curates identifiers from other data sources, validated CAS RNs in Wikidata may also be used to cross-reference with other resources. Scripts are provided in the Supporting Information.

The alignment is a continuous process, as new chemical compounds get added to Wikidata on a weekly basis. The comparison of Common Chemistry with Wikidata and Wikipedia resulted in a wealth of curation data, e.g. inconsistent CAS numbers linked to InChIKeys, where Common Chemistry had a different match than Wikidata or Wikipedia.

CAS registry numbers were not added to Wikidata in this process, only confirmed or reported as different. The latter allowed manual curation by the community, which it did. Reports [look like this](#). When a InChIKey-CAS RN combination in Wikidata was confirmed, it was recorded as a reference, like this:



The screenshot shows a Wikidata item page for the CAS Registry Number 14797-73-0. The page is titled "14797-73-0" and is contributed by KrBot and Egon Willighagen. It contains two references:

- Imported from Wikimedia project**: English Wikipedia
- InChIKey**: VLTRZXGMWDSKGL-UHFFFAOYSA-M
- stated in**: CAS Common Chemistry
- retrieved**: 10 April 2021
- reference URL**: https://commonchemistry.cas.org/detail?cas_rn=14797-73-0

At the bottom of the page, there are buttons for "+ add reference" and "+ add value".

Screenshot of Wikidata with two references, one reflecting a confirmation by the English Wikipedia (potentially the result of the original Common Chemistry project) and the second as outcome of the now published project.

Thanks to everyone on this project and [Andrea Jacobs](#) particularly for leading this open science project.