Chemical Data

Other compound identifiers

NSC_CAS_Sept2013.csv NSC to CAS number. We only have CAS number for 130,893 NSCs.

NSC_PubChemSID.csv NSC to PubChem SID. This is the SID from the DTP upload to PubChem.

divii_mlsmr.csv NSC to PubChem SID for the Diversity Set.

2D structures

All Open (June 2016 Release) 284176 compounds. 81 MB compressed, uncompresses to 710 MB

All Open (Sept 2014 Release) 280816 compounds. 78 MB compressed, uncompresses to 732 MB

All Open (March 2012 Release) 273885 compounds. 64 MB compressed, uncompresses to 648 MB

Mechanistic Set

3D structures

Mechanistic Set

SMILES strings

SMILES strings - 237,771 structures in SMILES format. This database contains essentially all open structures in the NCI database up until about June, 1995. It includes metal-containing compounds and other 'weird stuff'. It is therefore up to the user to ascertain the usefulness of any of these SMILES strings for the intended purpose. Because different conversion programs produce different output, two versions of the SMILES database are provided.

Converted using Babel 4.2 MB compressed using standard Unix compress, uncompresses to ca. 15 MB. (zip compressed) The program Babel v. 1.6 was used to convert 3D coordinates, which had been generated by the program Corina v. 1.7 from the connection tables. (Babel needs 3D coordinates when reading SD files.) The resulting Babel output was modified by simple string substitution to solve the problem of nitro groups lacking formal charges, which leads many SMILES readers to create an -N-O-H group. Thus, N(=O)O was replaced by [N+](=O)[O-], and N(=O)(O) was replaced by [N+](=O)([O-]).

Converted using CACTVS 4.4 MB compressed using standard Unix compress, uncompresses to ca. 15 MB. (zip compressed) The program CACTVS v. 3.2 was used to convert the connection tables to SMILES strings. Thanks to Wolf-Dietrich Ihlenfeldt for providing us with the conversion scripts handling the formal charge problem and other 'unusual stuff' in the NCI database.

Chemical Names

chemnames_Aug2013.zip All chemical names available. First field is NSC number, second filed is the name and the third field is the name type (most just generic "Chemical Name"). The field separator is a "|". Note that we do not have chemical names for most of the compounds and many of the names that are there are systematic names that might not be very useful for searching.

Name types

- USAN - United States Adopted Name
- VAN - CAS valid ambiguous name
- DOT Name - Department of Transportation Name
- Brand Name
- 6th C.I.
- 7th C.I.
- 8th C.I.
- 9th C.I.
- Belgian Name
- Czech Name
- Dutch Name
- French Name
- German Name
- Hindustan Name
- INN:BAN
- Italian Name
- Japanese Name
- Persian Name
- Polish Name
- Portuguese Name
- Roumanian Name
- Russian Name
- Spanish Name
- Swedish Name

- Chemical Name - not in one of the other categories or unknown type