Enamine REAL Database

Release: 2026.1

Date of release: April 2026

Number of molecules: 13.6 billion

About the REAL Database

The REAL Database is a drug-like subset of the REAL Space. It is provided as CXSMILES. (Extended SMILES to represent special features of molecules, i.e., enhanced stereochemical representation) or as SDF. The Database is split into parts. Molecules are sorted based on Heavy atom count.

Enamine REAL Space, Enumerated

Release: 2026.1

Date of release: April 2026

Number of molecules: 94.5 billion

About the REAL Space, Enumerated

The REAL Space, Enumerated replicates the fragment-based version of the REAL Space. It is provided as CXSMILES in tranches. Each tranche is stored as a bz2 file and contains information on the type (*s*- or *m*-), HAC, and s logP. For example, SH07P180.cxsmiles.bz2 contains the *s*-REAL Compounds with 7 heavy atoms and 1.8 ≥s LogP > 1.9.

Access_1 REAL_DB_2026.01/CXSMILES

Enamine REAL Database provided in CXSMILES format.

Access_2 REAL_DB_2026.01/REAL_compounds_by_chemical_classes

Selected chemical classes of compounds from the REAL Database:

- Amino acids (25.8M)
- Carboxylic acids (228.9M)
- Hydroxamates (5.7M)
- Lead-like aliphatic carboxylic acids (132.8M)
- Lead-like aliphatic primary amines (162.1M)
- Lead-like aromatic carboxylic acids (65.5M)
- Lead-like aromatic primary amines (330.4M)
- Secondary amines, 8-21 heavy atoms (191.1M)
- Terminal acetylenes (503.4M)

Access_3 REAL_DB_2026.01/SDF

Enamine REAL Database provided in SD format. Split into 9 parts by Heavy Atom Count (HAC) ranges.

Access_4 REAL_DB_2026.01/Samples and REAL_DB_2026.01/Samples_SDF

Diverse sample sets of the REAL Database (13.6M, 136M, and 1.4B compounds). Provided in CXSMILES and SDF formats.

Access_5 REAL_DB_2026.01/Sets

Curated subsets of the REAL Database. Provided in CXSMILES format:

- REAL lead-like (11.3B)

- REAL 350/3 lead-like (3.1B)

- REAL fragments (57.5M)

- REAL strict fragments (2.0M)

- REAL natural product-like compounds (842.4M)

- REAL PPI modulators (1.5B)

Access_6 REAL_Space_2026.01_Covalents_CXSMILES

Enamine REAL Compounds bearing covalent warheads. The data is organized by the warhead:

- 2-Chloropropanamides (186K)

- 4-(Dimethylamino)but-2-enamides (10.7M)

- Acrylamides (93.9M)

- Activated ureas (72.9M)

- Acylaminonitriles (71M)

- Aldehydes (109M)

- Arylators (800M)

- Azetidinones (290M)

- Boronic acids and esters (31.7M)

- But-2-ynamides (1.4M)

- ChloroFluoroAcetamides (46.4M)

- Chloroacetamides (54.6M)

- Chloromethyl ketones (3)

- Cyanamides (9.3M)

- Cyanoacrylamides (36.7M)

- Cyanoacrylic acids and esters (610K)
- Disulfides (14.2M)
- Fluorosulfates (1.8M)
- MaleImides (4.4M)
- Methyl fumarate amides (9.8M)
- Nitro alkanes (61.2M)
- Oxetanones (258K)
- Polyfluorophenyl esters (42.7K)
- Propargyl amines (83.3M)
- Propiolamides (51.6M)
- Sulfamoyl fluorides (48.3M)
- Sulfonyl fluorides (9.4M)
- Vinyl sulfones (105M)

Access_7 REAL_Space_2026.01_Covalents_SDF

Same as user6 but in SD format.

Access_8 REAL_Space_2026.01_tranches

Enamine REAL Space, Enumerated provided as tranches in CXSMILES format.

Access_9 REAL_DB_2026.01/REAL_compounds_by_reaction_type

Enamine REAL Database separated by reaction (chemistry) type into two files: sREAL (3.4B compounds) and mREAL (10.2B compounds). Provided in CXSMILES format.