

# Simplified molecular-input line-entry system

The **simplified molecular-input line-entry system** (**SMILES**) is a specification in the form of a line notation for describing the structure of chemical species using short ASCII strings. SMILES strings can be imported by most molecule editors for conversion back into two-dimensional drawings or three-dimensional models of the molecules.

The original SMILES specification was initiated in the 1980s. It has since been modified and extended. In 2007, an open standard called OpenSMILES was developed in the open-source chemistry community. Other linear notations include the Wiswesser line notation (WLN), ROSDAL, and SYBYL Line Notation (SLN).

## Contents

### History

### Terminology

### Graph-based definition

### SMILES definition as strings of a context-free language

### Description

Atoms

Bonds

Rings

Aromaticity

Branching

Stereochemistry

Isotopes

Examples

Other examples of SMILES

### Extensions

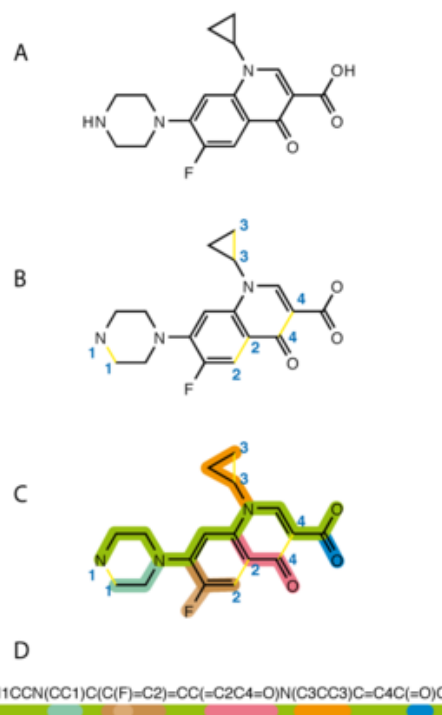
### Conversion

### See also

### References

## SMILES

<b>Filename extension</b>	.smi
<b>Internet media type</b>	chemical/x-daylight-smiles
<b>Type of format</b>	<u>chemical file format</u>



SMILES generation algorithm for Ciprofloxacin:  
break cycles, then write as branches off a main backbone

## History

The original SMILES specification was initiated by David Weininger at the USEPA Mid-Continent Ecology Division Laboratory in Duluth in the 1980s.<sup>[1][2][3][4]</sup> Acknowledged for their parts in the early development were "Gilman Veith and Rose Russo (USEPA) and Albert Leo and Corwin Hansch (Pomona College) for supporting the work, and Arthur Weininger (Pomona; Daylight CIS) and Jeremy Scofield (Cedar River Software, Renton, WA) for assistance in programming the system."<sup>[5]</sup> The Environmental Protection Agency funded the initial project to develop SMILES.<sup>[6][7]</sup>

It has since been modified and extended by others, most notably by Daylight Chemical Information Systems. In 2007, an open standard called "OpenSMILES" was developed by the Blue Obelisk open-source chemistry community. Other 'linear' notations include the Wiswesser Line Notation (WLN), ROSDAL and SLN (Tripos Inc).

In July 2006, the IUPAC introduced the InChI as a standard for formula representation. SMILES is generally considered to have the advantage of being slightly more human-readable than InChI; it also has a wide base of software support with extensive theoretical backing (such as graph theory).

## Terminology

---

The term SMILES refers to a line notation for encoding molecular structures and specific instances should strictly be called SMILES strings. However, the term SMILES is also commonly used to refer to both a single SMILES string and a number of SMILES strings; the exact meaning is usually apparent from the context. The terms "canonical" and "isomeric" can lead to some confusion when applied to SMILES. The terms describe different attributes of SMILES strings and are not mutually exclusive.

Typically, a number of equally valid SMILES strings can be written for a molecule. For example, CCO, OCC and C(O)C all specify the structure of ethanol. Algorithms have been developed to generate the same SMILES string for a given molecule; of the many possible strings, these algorithms choose only one of them. This SMILES is unique for each structure, although dependent on the canonicalization algorithm used to generate it, and is termed the canonical SMILES. These algorithms first convert the SMILES to an internal representation of the molecular structure; an algorithm then examines that structure and produces a unique SMILES string. Various algorithms for generating canonical SMILES have been developed and include those by Daylight Chemical Information Systems, OpenEye Scientific Software, MEDIT, Chemical Computing Group, MolSoft LLC, and the Chemistry Development Kit. A common application of canonical SMILES is indexing and ensuring uniqueness of molecules in a database.

The original paper that described the CANGEN<sup>[2]</sup> algorithm claimed to generate unique SMILES strings for graphs representing molecules, but the algorithm fails for a number of simple cases (e.g. cuneane, 1,2-dicyclopropylethane) and cannot be considered a correct method for representing a graph canonically.<sup>[8]</sup> There is currently no systematic comparison across commercial software to test if such flaws exist in those packages.

SMILES notation allows the specification of configuration at tetrahedral centers, and double bond geometry. These are structural features that cannot be specified by connectivity alone, and therefore SMILES which encode this information are termed isomeric SMILES. A notable feature of these rules is that they allow rigorous partial specification of chirality. The term isomeric SMILES is also applied to SMILES in which isomers are specified.

## Graph-based definition

---

In terms of a graph-based computational procedure, SMILES is a string obtained by printing the symbol nodes encountered in a depth-first tree traversal of a chemical graph. The chemical graph is first trimmed to remove hydrogen atoms and cycles are broken to turn it into a spanning tree. Where cycles have been broken, numeric suffix labels are included to indicate the connected nodes. Parentheses are used to indicate points of branching on the tree.

The resultant SMILES form depends on the choices:

- of the bonds chosen to break cycles,
- of the starting atom used for the depth-first traversal, and
- of the order in which branches are listed when encountered.

## SMILES definition as strings of a context-free language

---

From the view point of a formal language theory, SMILES is a word. A SMILES is parsable with a context-free parser. The use of this representation has been in the prediction of biochemical properties (incl. toxicity and biodegradability) based on the main principle of chemoinformatics that similar molecules have similar properties. The predictive models implemented a syntactic pattern recognition approach (which involved defining a molecular distance)<sup>[9]</sup> as well as a more robust scheme based on statistical pattern recognition.<sup>[10]</sup>

## Description

---

### Atoms

Atoms are represented by the standard abbreviation of the chemical elements, in square brackets, such as [Au] for gold. Brackets may be omitted in the common case of atoms which:

1. are in the "organic subset" of B, C, N, O, P, S, F, Cl, Br, or I, and

2. have no formal charge, and
3. have the number of hydrogens attached implied by the SMILES valence model (typically their normal valence, but for N and P it is 3 or 5, and for S it is 2, 4 or 6), and
4. are the normal isotopes, and
5. are not chiral centers.

All other elements must be enclosed in brackets, and have charges and hydrogens shown explicitly. For instance, the SMILES for water may be written as either O or [OH2]. Hydrogen may also be written as a separate atom; water may also be written as [H]O[H].

When brackets are used, the symbol H is added if the atom in brackets is bonded to one or more hydrogen, followed by the number of hydrogen atoms if greater than 1, then by the sign + for a positive charge or by - for a negative charge. For example, [NH4+] for ammonium ( $\text{NH}_4^+$ ). If there is more than one charge, it is normally written as digit; however, it is also possible to repeat the sign as many times as the ion has charges: one may write either [Ti+4] or [Ti++++] for titanium(IV)  $\text{Ti}^{4+}$ . Thus, the hydroxide anion ( $\text{OH}^-$ ) is represented by [OH-], the hydronium cation ( $\text{H}_3\text{O}^+$ ) is [OH3+] and the cobalt(III) cation ( $\text{Co}^{3+}$ ) is either [Co+3] or [Co+++].

## Bonds

A bond is represented using one of the symbols . - = # \$ : / \.

Bonds between aliphatic atoms are assumed to be single unless specified otherwise and are implied by adjacency in the SMILES string. Although single bonds may be written as -, this is usually omitted. For example, the SMILES for ethanol may be written as C-C-O, CC-O or C-CO, but is usually written CCO.

Double, triple, and quadruple bonds are represented by the symbols =, #, and \$ respectively as illustrated by the SMILES O=C=O (carbon dioxide  $\text{CO}_2$ ), C#N (hydrogen cyanide  $\text{HCN}$ ) and [Ga+]\$[As-] (gallium arsenide).

An additional type of bond is a "non-bond", indicated with ., to indicate that two parts are not bonded together. For example, aqueous sodium chloride may be written as [Na+] . [Cl-] to show the dissociation.

An aromatic "one and a half" bond may be indicated with :; see § Aromaticity below.

Single bonds adjacent to double bonds may be represented using / or \ to indicate stereochemical configuration; see § Stereochemistry below.

## Rings

Ring structures are written by breaking each ring at an arbitrary point (although some choices will lead to a more legible SMILES than others) to make an acyclic structure and adding numerical ring closure labels to show connectivity between non-adjacent atoms.

For example, cyclohexane and dioxane may be written as C1CCCCC1 and O1CCOCC1 respectively. For a second ring, the label will be 2. For example, decalin (decahydronaphthalene) may be written as C1CCCC2C1CCCC2.

SMILES does not require that ring numbers be used in any particular order, and permits ring number zero, although this is rarely used. Also, it is permitted to reuse ring numbers after the first ring has closed, although this usually makes formulae harder to read. For example, bicyclohexyl is usually written as C1CCCC1C2CCCCC2, but it may also be written as C0CCCCC0C0CCCCC0.

Multiple digits after a single atom indicate multiple ring-closing bonds. For example, an alternative SMILES notation for decalin is C1CCCC2CCCC12, where the final carbon participates in both ring-closing bonds 1 and 2. If two-digit ring numbers are required, the label is preceded by %, so C%12 is a single ring-closing bond of ring 12.

Either or both of the digits may be preceded by a bond type to indicate the type of the ring-closing bond. For example, cyclopropene is usually written C1=CC1, but if the double bond is chosen as the ring-closing bond, it may be written as C=1CC1, C1CC=1, or C=1CC=1. (The first form is preferred.) C=1CC-1 is illegal, as it explicitly specifies conflicting types

for the ring-closing bond.

Ring-closing bonds may not be used to denote multiple bonds. For example, C1C1 is not a valid alternative to C=C for ethylene. However, they may be used with non-bonds; C1 . C2 . C12 is a peculiar but legal alternative way to write propane, more commonly written CCC.

Choosing a ring-break point adjacent to attached groups can lead to a simpler SMILES form by avoiding branches. For example, cyclohexane-1,2-diol is most simply written as OC1CCCCC1O; choosing a different ring-break location produces a branched structure that requires parentheses to write.

## Aromaticity

Aromatic rings such as benzene may be written in one of three forms:

1. In Kekulé form with alternating single and double bonds, e.g. C1=CC=CC=C1,
2. Using the aromatic bond symbol `:`, e.g. C1:C:C:C:C:C1, or
3. Most commonly, by writing the constituent B, C, N, O, P and S atoms in lower-case forms b, c, n, o, p and s, respectively.

In the latter case, bonds between two aromatic atoms are assumed (if not explicitly shown) to be aromatic bonds. Thus, benzene, pyridine and furan can be represented respectively by the SMILES c1ccccc1, n1ccccc1 and o1ccccc1.

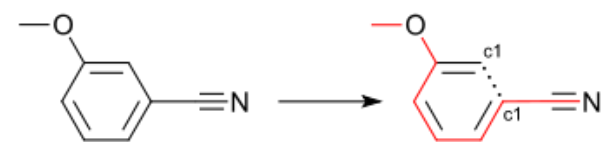
Aromatic nitrogen bonded to hydrogen, as found in pyrrole must be represented as [nH]; thus imidazole is written in SMILES notation as n1c[nH]cc1.

When aromatic atoms are singly bonded to each other, such as in biphenyl, a single bond must be shown explicitly: c1ccccc1-c2ccccc2. This is one of the few cases where the single bond symbol `-` is required. (In fact, most SMILES software can correctly infer that the bond between the two rings cannot be aromatic and so will accept the nonstandard form c1ccccc1c2ccccc2.)

The Daylight and OpenEye algorithms for generating canonical SMILES differ in their treatment of aromaticity.

## Branching

Branches are described with parentheses, as in CCC(=O)O for propionic acid and FC(F)F for fluoroform. The first atom within the parentheses, and the first atom after the parenthesized group, are both bonded to the same branch point atom. The bond symbol must appear inside the parentheses; outside (E.g.: CCC=(O)O) is invalid.



Visualization of 3-cyanoanisole as C0c(c1)cccc1C#N.

Substituted rings can be written with the branching point in the ring as illustrated by the SMILES C0c(c1)cccc1C#N (see depiction (<https://web.archive.org/web/20130522091354/http://www.daylight.com/daycgi/depict?434f63286331296363633143234e>)) and C0c(cc1)ccc1C#N (see depiction (<https://web.archive.org/web/20130522074308/http://www.daylight.com/daycgi/depict?434f6328636331296363633143234e>)) which encode the 3 and 4-cyanoanisole isomers. Writing SMILES for substituted rings in this way can make them more human-readable.

Branches may be written in any order. For example, bromochlorodifluoromethane may be written as FC(Br)(Cl)F, BrC(F)(F)Cl, C(F)(Cl)(F)Br, or the like. Generally, a SMILES form is easiest to read if the simpler branch comes first, with the final, unparenthesized portion being the most complex. The only caveats to such rearrangements are:

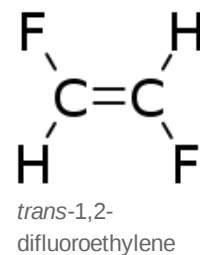
- If ring numbers are reused, they are paired according to their order of appearance in the SMILES string. Some adjustments may be required to preserve the correct pairing.
- If stereochemistry is specified, adjustments must be made; see Stereochemistry § Notes below.

The one form of branch which does *not* require parentheses are ring-closing bonds. Choosing ring-closing bonds appropriately can reduce the number of parentheses required. For example, toluene is normally written as Cc1ccccc1 or c1ccccc1C, avoiding the parentheses required if written as c1cc(C)ccc1 or c1cc(ccc1)C.

## Stereochemistry

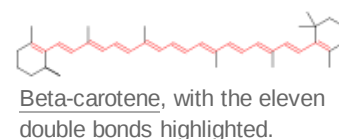
SMILES permits, but does not require, specification of stereoisomers.

Configuration around double bonds is specified using the characters / and \ to show directional single bonds adjacent to a double bond. For example, F/C=C/F (see depiction (<https://web.archive.org/web/20130522072357/http://www.daylight.com/daycgi/depict?462f433d432f46>)) is one representation of *trans*-1,2-difluoroethylene, in which the fluorine atoms are on opposite sides of the double bond (as shown in the figure), whereas F/C=C\F (see depiction (<https://web.archive.org/web/20130522074206/http://www.daylight.com/daycgi/depict?462f433d435c46>)) is one possible representation of *cis*-1,2-difluoroethylene, in which the fluorines are on the same side of the double bond.



Bond direction symbols always come in groups of at least two, of which the first is arbitrary. That is, F\C=C\F is the same as F/C=C/F. When alternating single-double bonds are present, the groups are larger than two, with the middle directional symbols being adjacent to two double bonds. For example, the common form of (2,4)-hexadiene is written C/C=C/C=C/C.

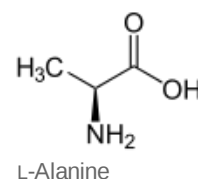
As a more complex example, beta-carotene has a very long backbone of alternating single and double bonds, which may be written



CC1CCC/C(C)=C1/C=C/C(C)=C/C=C/C(C)=C/C=C/C=C(C)/C=C/C=C(C)/C=C/C2=C(C)/CCCC2(C)C.

Configuration at tetrahedral carbon is specified by @ or @@. Consider the four bonds in the order in which they appear, left to right, in the SMILES form. Looking toward the central carbon from the perspective of the first bond, the other three are either clockwise or counter-clockwise. These cases are indicated with @@ and @, respectively (because the @ symbol itself is a counter-clockwise spiral).

For example, consider the amino acid alanine. One of its SMILES forms is NC(C)C(=O)O, more fully written as N[CH](C)C(=O)O. L-Alanine, the more common enantiomer, is written as N[C@@H](C)C(=O)O (see depiction (<https://web.archive.org/web/20130704043108/http://www.daylight.com/daycgi/depict?4e5b434040485d28432943283d4f294f>)). Looking from the nitrogen-carbon bond, the hydrogen (H), methyl (C), and carboxylate (C(=O)O) groups appear clockwise. D-Alanine can be written as N[C@H](C)C(=O)O (see depiction (<https://web.archive.org/web/20130522072012/http://www.daylight.com/daycgi/depict?4e5b434040485d28432943283d4f294f>)).



While the order in which branches are specified in SMILES is normally unimportant, in this case it matters; swapping any two groups requires reversing the chirality indicator. If the branches are reversed so alanine is written as NC(C(=O)O)C, then the configuration also reverses; L-alanine is written as N[C@H](C(=O)O)C (see depiction (<https://web.archive.org/web/20130522073747/http://www.daylight.com/daycgi/depict?4e5b434040485d2843283d4f294f2943>)). Other ways of writing it include C[C@H](N)C(=O)O, OC(=O)[C@@H](N)C and OC(=O)[C@H](C)N.

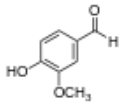
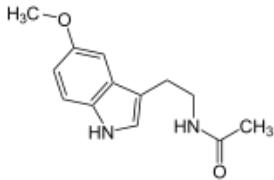
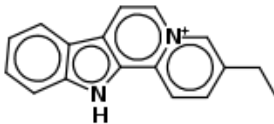
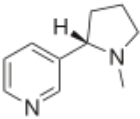
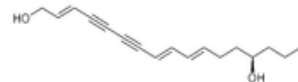
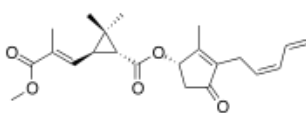
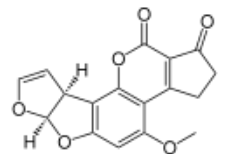
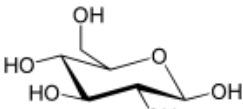
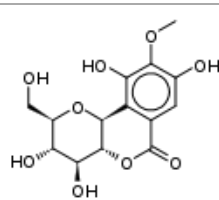
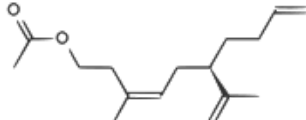
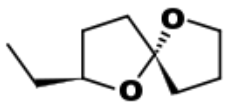
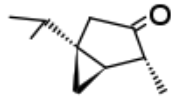
Normally, the first of the four bonds appears to the left of the carbon atom, but if the SMILES is written beginning with the chiral carbon, such as C(C)(N)C(=O)O, then all four are to the right, but the first to appear (the [CH] bond in this case) is used as the reference to order the following three: L-alanine may also be written [C@@H](C)(N)C(=O)O.

The SMILES specification includes elaborations on the @ symbol to indicate stereochemistry around more complex chiral centers, such as trigonal bipyramidal molecular geometry.

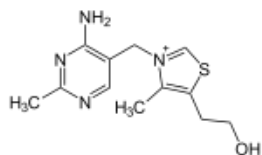
## Isotopes

Isotopes are specified with a number equal to the integer isotopic mass preceding the atomic symbol. Benzene in which one atom is carbon-14 is written as [14C]1CCCC1 and deuteriochloroform is [2H]C(Cl)(Cl)Cl.

## Examples

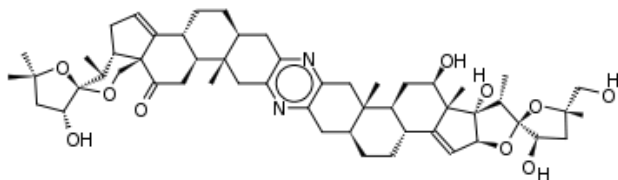
Molecule	Structure	SMILES formula
<u>Dinitrogen</u>	$N\equiv N$	<chem>N#N</chem>
<u>Methyl isocyanate (MIC)</u>	$CH_3-N=C=O$	<chem>CN=C=O</chem>
<u>Copper(II) sulfate</u>	$Cu^{2+}SO_4^{2-}$	<chem>[Cu+2].[O-]S(=O)(=O)[O-]</chem>
<u>Vanillin</u>		<chem>O=Cc1ccc(OC)c(O)c1</chem> <chem>O=Cc1ccc(OC)cc1O</chem>
<u>Melatonin</u> ( $C_{13}H_{16}N_2O_2$ )		<chem>CC(=O)NCCC1=CNC2c1cc(OC)cc2</chem> <chem>CC(=O)NCCC1c[nH]c2ccc(OC)cc12</chem>
<u>Flavopereirin</u> ( $C_{17}H_{15}N_2$ )		<chem>CCc(c1)ccc2[n+]1ccc3c2[nH]c4c3cccc4</chem> <chem>CCc1c[n+]2ccc3c4cccc4[nH]c3c2cc1</chem>
<u>Nicotine</u> ( $C_{10}H_{14}N_2$ )		<chem>CN1CCC[C@H]1c2cccnc2</chem>
<u>Oenanthotoxin</u> ( $C_{17}H_{22}O_2$ )		<chem>CCC[C@@H](O)CC\C=C\C=C\C=C\C#C#C\C=C\C=C\CO</chem> <chem>CCC[C@@H](O)CC/C=C/C=C/C/C#C#C/C=C/CO</chem>
<u>Pyrethrin II</u> ( $C_{22}H_{28}O_5$ )		<chem>CC1=C(C(=O)C[C@@H]1OC(=O)[C@@H]2[C@H](C2(C)C)/C=C(\C)/C(=O)OC)C/C=C\C=C</chem>
<u>Aflatoxin B<sub>1</sub></u> ( $C_{17}H_{12}O_6$ )		<chem>O1C=C[C@H]( [C@H]1O2)c3c2cc(OC)c4c3OC(=O)C5=C4CCC(=O)5</chem>
<u>Glucose</u> ( $\beta$ -D-glucopyranose) ( $C_6H_{12}O_6$ )		<chem>OC[C@H](O1)[C@@H](O)[C@H](O)[C@@H](O)[C@H](O)1</chem>
<u>Bergenin</u> (cuscutin, a resin) ( $C_{14}H_{16}O_9$ )		<chem>OC[C@H](O1)[C@@H](O)[C@H](O)[C@@H]2[C@@H]1c3c(O)c(OC)c(O)cc3C(=O)O2</chem>
<u>A pheromone of the Californian scale insect</u>		<chem>CC(=O)OCCC(/C)=C\C[C@H](C(C)=C)CCC=C</chem>
<u>(2S,5R)-Chalcogran: a pheromone of the bark beetle <i>Pityogenes chalcographus</i></u> <sup>[11]</sup>		<chem>CC[C@H](O1)CC[C@@]12CCCCO2</chem>
<u><math>\alpha</math>-Thujone</u> ( $C_{10}H_{16}O$ )		<chem>CC(C)[C@@]12C[C@@H]1[C@H](C)C(=O)C2</chem>

Thiamine (vitamin B<sub>1</sub>, C<sub>12</sub>H<sub>17</sub>N<sub>4</sub>OS<sup>+</sup>)



OCCc1c(C)[n+](cs1)Cc2cnc(C)nc2N

To illustrate a molecule with more than 9 rings, consider cephalostatin-1,<sup>[12]</sup> a steroidal 13-ringed pyrazine with the empirical formula C<sub>54</sub>H<sub>74</sub>N<sub>2</sub>O<sub>10</sub> isolated from the Indian Ocean hemichordate Cephalodiscus gilchristi:



Starting with the left-most methyl group in the figure:

```
CC(C)(O1)C[C@@H](O)[C@@]1(O2)[C@@H](C)
[C@@H]3CC=C4[C@]3(C2)C(=O)C[C@H]5[C@H]4CC[C@@H](C6)
[C@]5(C)Cc(n7)c6nc(C[C@@]89(C))c7C[C@@H]8CC[C@@H]%10[C@@H]9C[C@@H](O)
[C@@]%11(C)C%10=C[C@H](O%12)[C@]%11(O)[C@H](C)[C@]%12(O%13)[C@H]
(O)C[C@@]%13(C)CO
```

Note that % appears in front of the index of ring closure labels above 9; see § Rings above.

## Other examples of SMILES

The SMILES notation is described extensively in the SMILES theory manual provided by Daylight Chemical Information Systems and a number of illustrative examples are presented. Daylight's depict utility provides users with the means to check their own examples of SMILES and is a valuable educational tool.

## Extensions

SMARTS is a line notation for specification of substructural patterns in molecules. While it uses many of the same symbols as SMILES, it also allows specification of wildcard atoms and bonds, which can be used to define substructural queries for chemical database searching. One common misconception is that SMARTS-based substructural searching involves matching of SMILES and SMARTS strings. In fact, both SMILES and SMARTS strings are first converted to internal graph representations which are searched for subgraph isomorphism.

SMIRKS, a superset of "reaction SMILES" and a subset of "reaction SMARTS", is a line notation for specifying reaction transforms. The general syntax for the reaction extensions is REACTANT>AGENT>PRODUCT (without spaces), where any of the fields can either be left blank or filled with multiple molecules delimited with a dot (.), and other descriptions dependent on the base language. Atoms can additionally be identified with a number (e.g. [C:1]) for mapping,<sup>[13]</sup> for example in [CH2:1]=[CH:2][CH:3]=[CH:4][CH2:5][H:6]>>[H:6][CH2:1][CH:2]=[CH:3][CH:4]=[CH2:5] (<http://cdb.ics.uci.edu/cgi-bin/Smi2DepictWeb.py?smiles=%5BCH2%3A1%5D%3D%5BCH%3A2%5D%5BCH%3A3%5D%3D%5BCH%3A4%5D%5BCH2%3A5%5D%5BH%3A6%5D%3E%3E%5BH%3A6%5D%5BCH2%3A1%5D%5BCH%3A2%5D%3D%5BCH%3A3%5D%5BCH%3A4%5D%3D%5BCH2%3A5%5D>).<sup>[14]</sup>

## Conversion

SMILES can be converted back to two-dimensional representations using structure diagram generation (SDG) algorithms.<sup>[15]</sup> This conversion is not always unambiguous. Conversion to three-dimensional representation is achieved by energy-minimization approaches. There are many downloadable and web-based conversion utilities.

## See also

- SMILES arbitrary target specification (SMARTS), an extension of SMILES for specification of substructural queries
- SYBYL Line Notation, another line notation
- International Chemical Identifier (InChI), the IUPAC's alternative to SMILES
- Molecular Query Language, a query language allowing also numerical properties, e.g. physicochemical values or distances
- Chemistry Development Kit, 2D layout and conversion software
- OpenBabel, JOELib, OELib (conversion)

## References

1. Weininger, David (February 1988). "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules". *Journal of Chemical Information and Computer Sciences*. **28** (1): 31–6. doi:10.1021/ci00057a005 (<https://doi.org/10.1021%2Fci00057a005>).
2. Weininger, David; Weininger, Arthur; Weininger, Joseph L. (May 1989). "SMILES. 2. Algorithm for generation of unique SMILES notation". *Journal of Chemical Information and Modeling*. **29** (2): 97–101. doi:10.1021/ci00062a008 (<https://doi.org/10.1021%2Fci00062a008>).
3. Weininger, David (August 1990). "SMILES. 3. DEPICT. Graphical depiction of chemical structures". *Journal of Chemical Information and Modeling*. **30** (3): 237–43. doi:10.1021/ci00067a005 (<https://doi.org/10.1021%2Fci00067a005>).
4. Swanson, Richard Pommier (2004). "The Entrance of Informatics into Combinatorial Chemistry" (<https://wayback.archive-it.org/2118/20100925010036/http://64.251.202.97/pubs/asist2002/17-swanson.pdf>) (PDF). In Rayward, W. [Warden] Boyd; Bowden, Mary Ellen (eds.). *The History and Heritage of Scientific and Technological Information Systems: Proceedings of the 2002 Conference of the American Society of Information Science and Technology and the Chemical Heritage Foundation* (<https://books.google.com/books?id=76OOQannpBgC&pg=PA205>). Medford, NJ: Information Today. p. 205. ISBN 9781573872294.
5. Weininger, Dave (1998). "Acknowledgements on Daylight Tutorial smiles-etc page" (<http://www.daylight.com/meetings/summerschool98/course/dave/smiles-etc.html>). Retrieved June 24, 2013.
6. Anderson, E.; Veith, G. D.; Weininger, D. (1987). *SMILES: A line notation and computerized interpreter for chemical structures* (<https://nepis.epa.gov/Exe/ZyPDF.cgi/2000CAUR.PDF?Dockey=2000CAUR.PDF>) (PDF). Duluth, MN: U.S. EPA, Environmental Research Laboratory-Duluth. Report No. EPA/600/M-87/021.
7. "SMILES Tutorial: What is SMILES?" ([http://www.epa.gov/med/Prods\\_Pubs/smiles.htm](http://www.epa.gov/med/Prods_Pubs/smiles.htm)). U.S. EPA. Retrieved September 23, 2012.
8. Hutchison D, Kanade T, Kittler J, Klienberg JM, Mattern F, Mitchell JC, Naor M, Nierstrasz O, Rangan CP, Steffen B, Sudan M, Terzopoulos D, Tygar D, Vardi MY, Weikum G, Raschid L, Neglur G, Grossman RL, Liu B (2005). "Assigning Unique Keys to Chemical Compounds for Data Integration: Some Interesting Counter Examples" ([https://doi.org/10.1007%2F11530084\\_13](https://doi.org/10.1007%2F11530084_13)). In Ludäscher B (ed.). *Data Integration in the Life Sciences*. Lecture Notes in Computer Science. **3615**. Berlin: Springer. pp. 145–157. doi:10.1007/11530084\_13 ([https://doi.org/10.1007%2F11530084\\_13](https://doi.org/10.1007%2F11530084_13)). ISBN 978-3-540-27967-9. Retrieved February 12, 2013.
9. Sidorova, J. Anisimova M, 'NLP-inspired pattern recognition in chemical application', Pattern Recognition Letters, 45 (2014) 11-16.
10. Sidorova, J, Garcia, J, 'Bridging from syntactic to statistical methods: Classification with automatically segmented features from sequences', Pattern Recognition, 48 (11), 3749-3756
11. Byers, JA; Birgersson, G; Löfqvist, J; Appelgren, M; Bergström, G (March 1990). "Isolation of pheromone synergists of bark beetle, *Pityogenes chalcographus*, from complex insect-plant odors by fractionation and subtractive-combination bioassay" (<http://www.chemical-ecology.net/pdf/Byersetal1990a.pdf>) (PDF). *Journal of Chemical Ecology*. **16** (3): 861–76. doi:10.1007/BF01016496 (<https://doi.org/10.1007%2FBF01016496>). PMID 24263601 (<https://pubmed.ncbi.nlm.nih.gov/24263601>). S2CID 226090 (<https://api.semanticscholar.org/CorpusID:226090>).
12. "CID 183413" (<https://pubchem.ncbi.nlm.nih.gov/compound/183413>). PubChem. Retrieved May 12, 2012.
13. "SMIRKS Tutorial" ([http://daylight.com/dayhtml\\_tutorials/languages/smirks/](http://daylight.com/dayhtml_tutorials/languages/smirks/)). Daylight. Retrieved October 29, 2018.
14. "Reaction SMILES and SMIRKS" (<http://www.daylight.com/meetings/summerschool01/course/basics/smirks.html>). Retrieved October 29, 2018.
15. Helson, H. E. (1999). "Structure Diagram Generation". In Lipkowitz, K. B.; Boyd, D. B. (eds.). *Rev. Comput. Chem. Reviews in Computational Chemistry*. **13**. New York: Wiley-VCH. pp. 313–398. doi:10.1002/9780470125908.ch6 (<https://doi.org/10.1002%2F9780470125908.ch6>). ISBN 9780470125908.



---

Retrieved from "[https://en.wikipedia.org/w/index.php?title=Simplified\\_molecular-input\\_line-entry\\_system&oldid=1005522521](https://en.wikipedia.org/w/index.php?title=Simplified_molecular-input_line-entry_system&oldid=1005522521)"

---

**This page was last edited on 8 February 2021, at 03:07 (UTC).**

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.