SMILES Notations

Dhananjay

September 22, 2025

1 Introduction

SMILES: **S**implified **M**olecular **I**nput **L**ine **E**ntry **S**ystem

- It's a chemical notation that allows a user to represent a chemical structure in a way that can be perceived by the computers.
- Five basic syntax rules.
- If the basic rules of chemistry are ignored in the SMILES entry, then the system will issue a warning and ask the structure to be edited.

Let's go through with the rules

1. Atoms and Bonds: SMILES support all the elements in the periodic table. Atom is represented using its respective symbol. Upper case letters refer to non aromatic atoms, lower case refers to aromatic atoms. Bonds are denoted as shown below:
   - - Single Bond (Default and need not be entered)
   - = Double Bond
   - # Triple Bond
   - * Aromatic Bond
   - . Disconnected Structure
2. Simple Chains: Combining atomic symbols and bond symbols, simple chains structures can be represented. The structures that are entered using SMILES are hydrogen suppressed. Some examples are:

SMILES Notation	**Formula**	**Name**
CC	CH_3CH_3	Ethane
$C=C$	CH_2CH_2	Ethene
$C[Br]$	CH_3Br	Bromomethane
$C\#N$	CN	Cyanide
$[Na].[Cl]$	$NaCl$	Sodium Chloride

3. Branches: A branch from a chain is specified by placing the SMILES symbol(s) for the branch between paranthesis. The string in parantheses is placed directly after the symbol for the atom to which it is connected. If it is connected by a double or triple bond, the bond symbol immediately follows the left paranthesis. Some examples are as mentioned below:

SMILES Notation	**Name**
$CC(O)C$	$2-Propanol$
$CC(=O)C$	$2-Propanone$
$CC(CC)C$	$2-Methylbutane$
$CC(C)CC(=O)$	$2-Methylbutanal$
$c1c(N(=O)=O)cccc1$	$Nitrobenzene$
$CC(C)(C)CC$	$2,2-Dimethylbutane4$

4. Rings: SMILES allow a user to identify ring structures by using numbers to identify the opening and closing ring structure. For example, in the structure $C1CCCCC1$, the first carbon has a number '1' which connects via a single bond to the last carbon, also having number '1'. The resulting structure is Cyclohexane. Chemicals having multiple rings can be identified by using different numbers for each ring. If a double, triple or aromatic bond is used for the ring closure, the bond symbol is placed before the ring closure number. Let us understand this with some examples:

SMILES Notation	**Name**
C=1CCCCC1	Cyclohexene
C*1C*C*C*C*C1	Cyclohexene
C1CCCCC1	Benzene
C1OC1CC	Ethyloxirane
c1cc2ccccc2cc1	Naphthalene

5. Charged Atoms: Charges on an atom can be used to override the knowledge regarding valence that is built into SMILES software. The format for identifying a charged atom consists of the atom followed by brackets which enclose the charge on atom. The magnitude of charge may be explicitly stated by (-1) not (-)

SMILES Notation	**Name**
$CCC(=O)[O-]$	Ionized form of Propanoic Acid
$[n+]1ccccc1CC(=O)$	1-Carboxylmethyl pyridinium