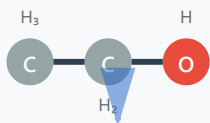


# SMILES Notation

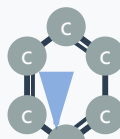
Ethanol



CCO

SMILES String

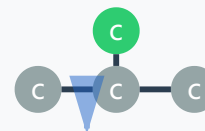
Benzene



c1ccccc1

Ring Closure

Branched



CC(C)C

Branching ()

## Syntax rules

String-based molecular encoding

## Canonical SMILES

Unique molecular representation

## SMARTS patterns

Substructure search patterns

## Tokenization

Breaking into meaningful units

## Augmentation strategies

Data augmentation techniques