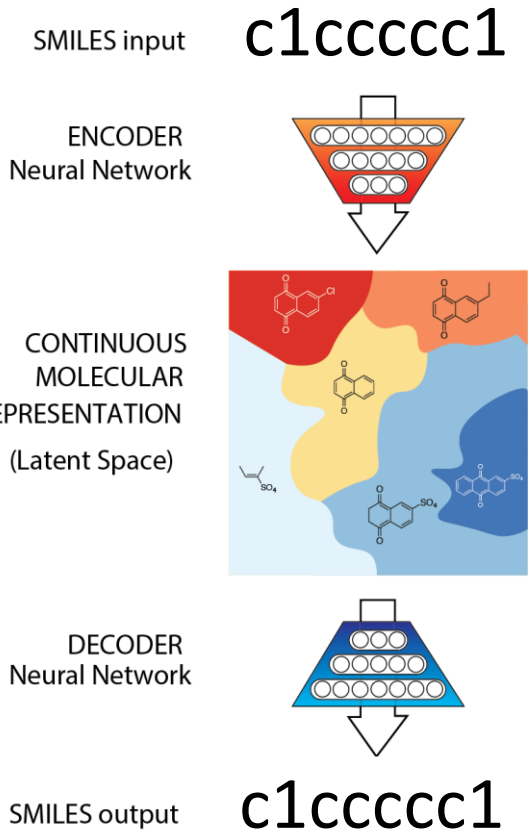
DEEPSMILES

Noel O'Boyle

Andrew Dalke



Gómez-Bombarelli et al. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Central Science* **2018**, *4*, 268–276.

DEEPSMILES

Noel O'Boyle

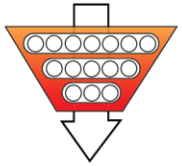
Andrew Dalke

NextMove Software

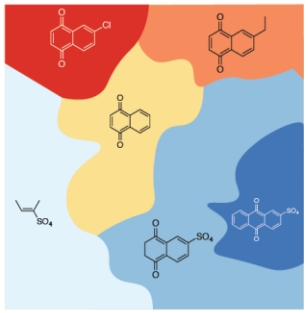
dalke scientific
More science. Less time.

SMILES input c1ccccc1

ENCODER
Neural Network

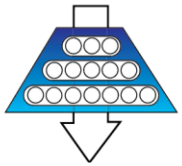


CONTINUOUS
MOLECULAR
REPRESENTATION
(Latent Space)



Gómez-Bombarelli et al. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Central Science* **2018**, *4*, 268–276.

DECODER
Neural Network



Invalid syntax

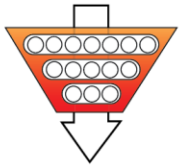
SMILES output c1ccccc1 c1ccccc1(C)C c1ccc(cc1 c1ccccc2

SMILES input

c1ccccc1

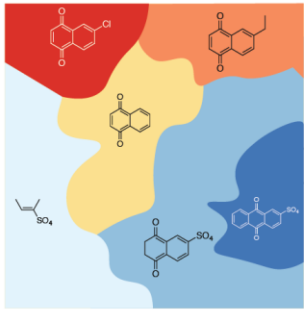
DeepSMILES

ENCODER
Neural Network

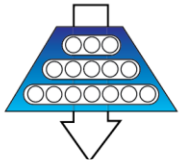


CONTINUOUS
MOLECULAR
REPRESENTATION

(Latent Space)



DECODER
Neural Network



DeepSMILES

c1ccccc1 c1ccccc1(C)C c1ccc(cc1 c1ccccc2

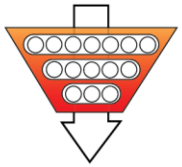
Invalid syntax

SMILES input

c1ccccc1

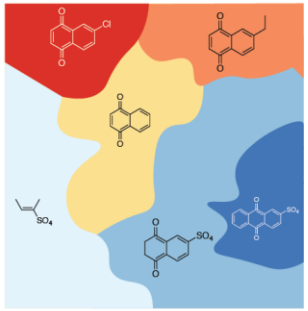
DeepSMILES

ENCODER
Neural Network

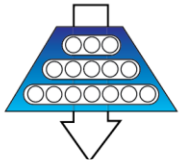


CONTINUOUS
MOLECULAR
REPRESENTATION

(Latent Space)



DECODER
Neural Network



DeepSMILES

c1ccccc1 c1ccccc1(C)C

RING CLOSURE NOTATION

SMILES …c**1**ccccc**1** ….c**2**ccccc**2** ….c**3**ccccc**3**

RING CLOSURE NOTATION

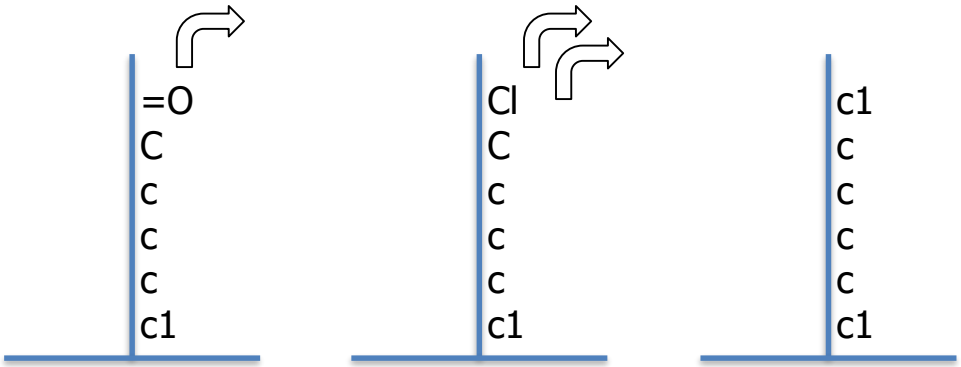
SMILES …c**1**ccccc**1** ….c**2**ccccc**2** ….c**3**ccccc**3**

DeepSMILES …ccccccc**6**

BRANCH NOTATION

SMILES c1ccc(C(=O)Cl)cc1

DeepSMILES c1cccC=O)Cl))cc1



SH Eisman. A Polish-type notation for chemical structures. *J. Chem. Doc.* **1964**, *4*, 186.
H Hiz. A linearization of chemical graphs. *J. Chem. Doc.* **1964**, *4*, 173.

DEEPSMILES

Noel O'Boyle

Andrew Dalke

See preprint on ChemRxiv for more info

https://github.com/nextmovesoftware/deepsmiles

```
>>> import deepsmiles as ds
>>> converter = ds.Converter(rings=True,
...                          branches=True)
>>> converter.encode("c1ccccc1")
'cccccc6'
>>> converter.decode("cccccc6")
'c1ccccc1'
```