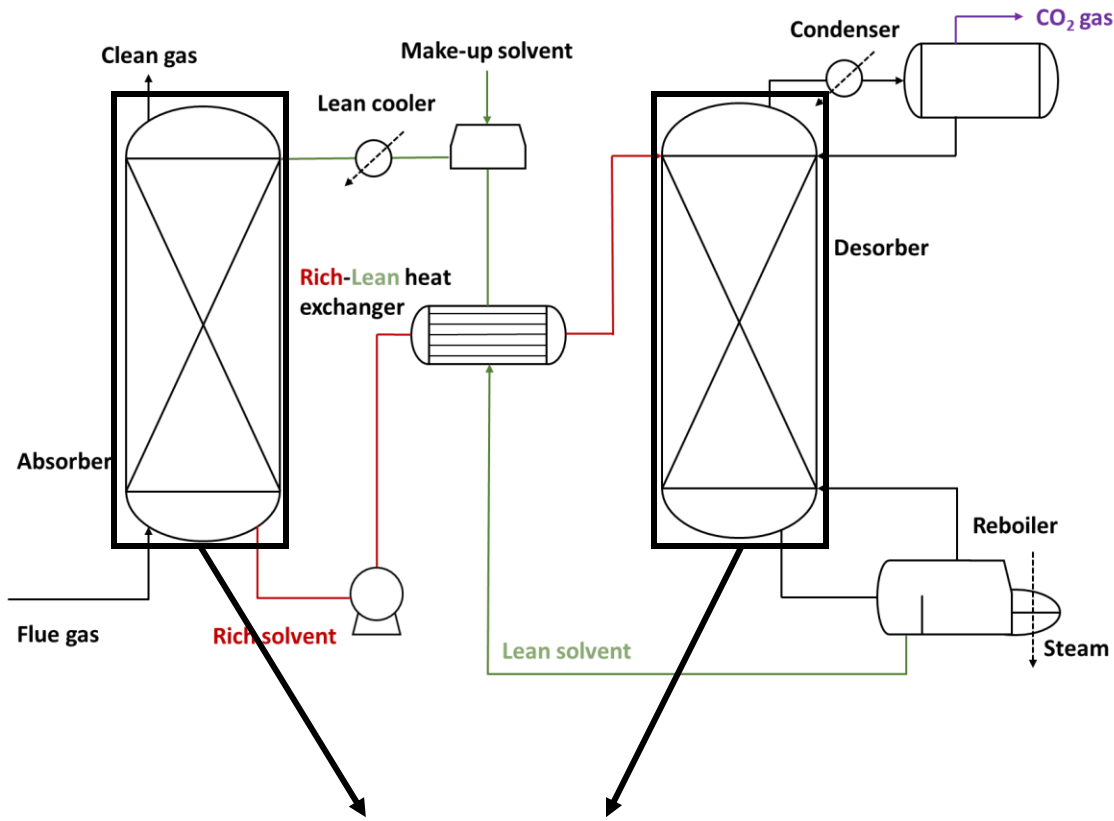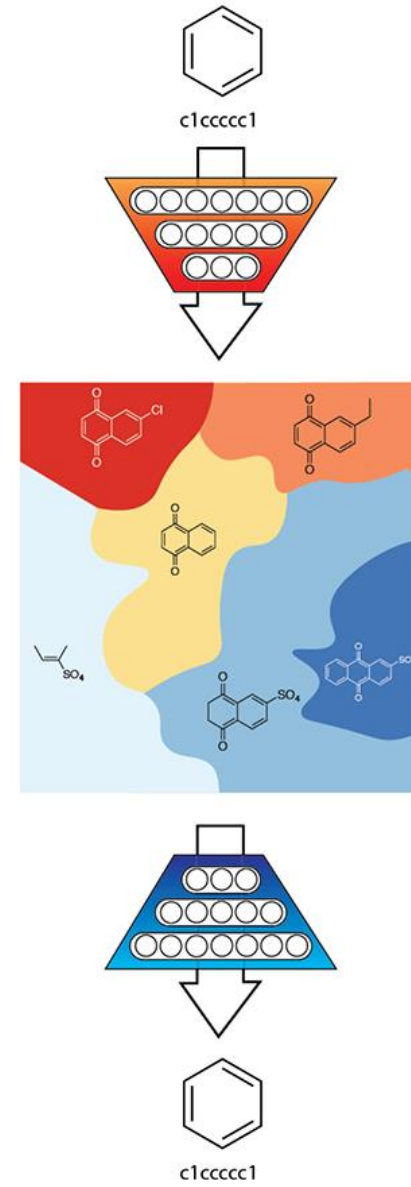Lingfeng (Griffin) Gui, Alhasan Abdellatif , Yongliang (Harry) Yan, Florian Baakes

51 000 000 000 t$_{CO_2 \, eq}$

$\Theta_{CO_2}(\mathrm{T}, \mathrm{p}), \; \gamma(T, p), \; \mu(T, p)$

c1ccccc1

c1ccccc1

Validity

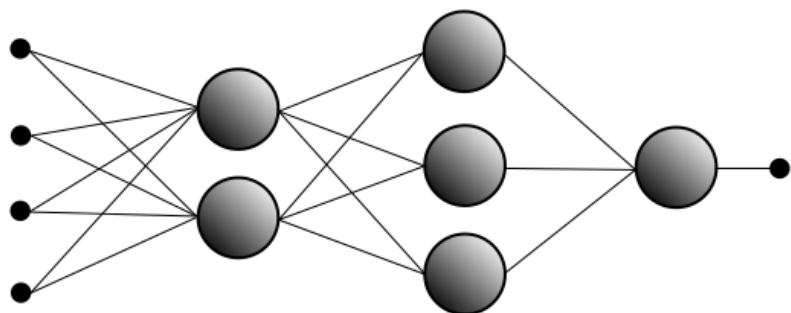Given a sample size x, how many of those x sampled molecules are valid molecules?

Reconstruction

Given an encoded molecule x and the sampled molecule y, to what percentage are they the same?
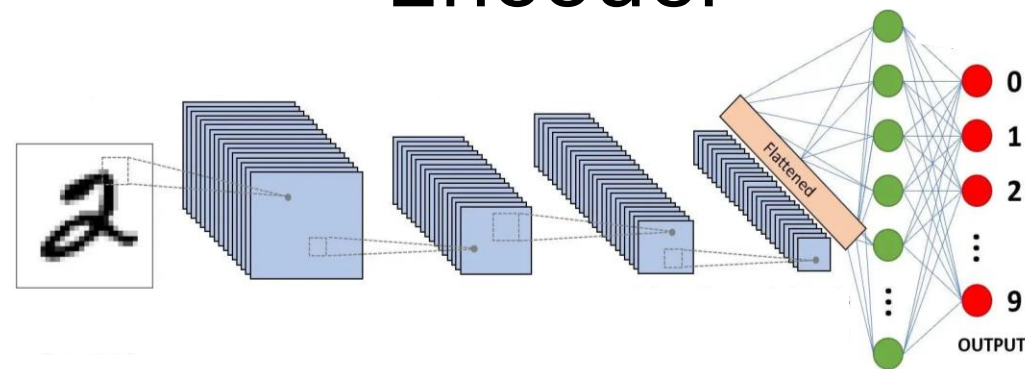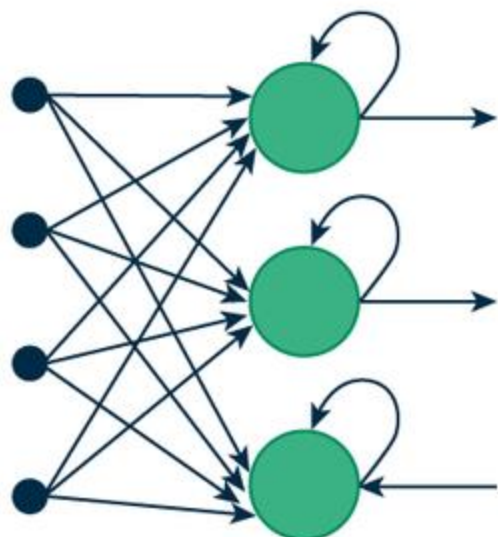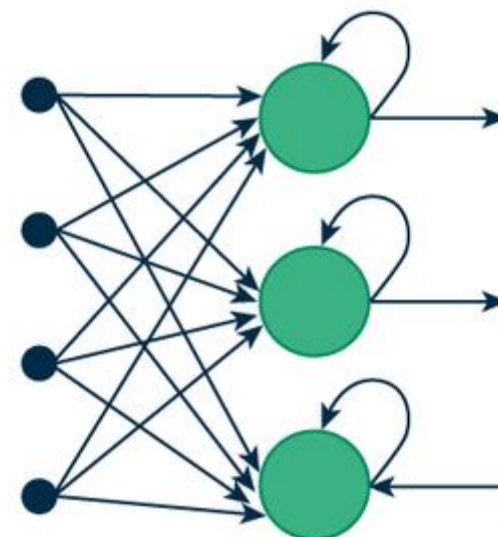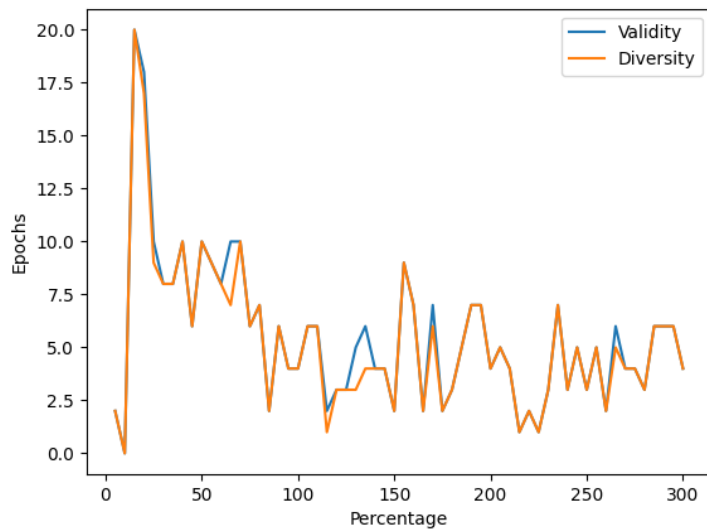
or

Are they the same or not?

Diversity

Given a sample size x, how many of those x sampled molecules are not identical?

Encoder

Encoder

Decoder

Decoder

SMILES
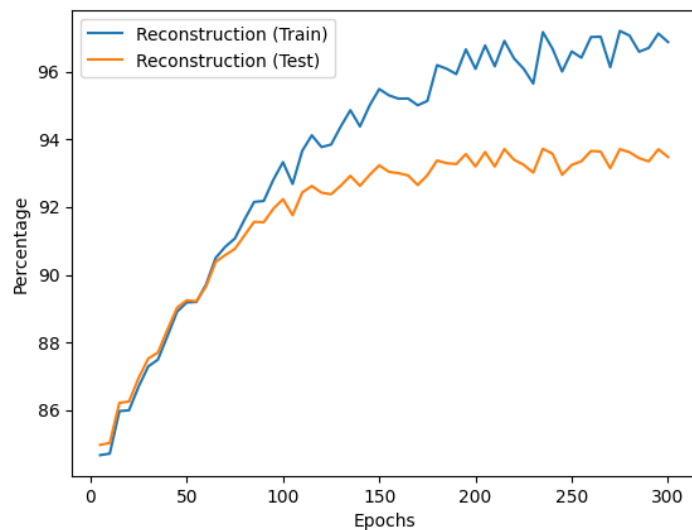
SELFIES
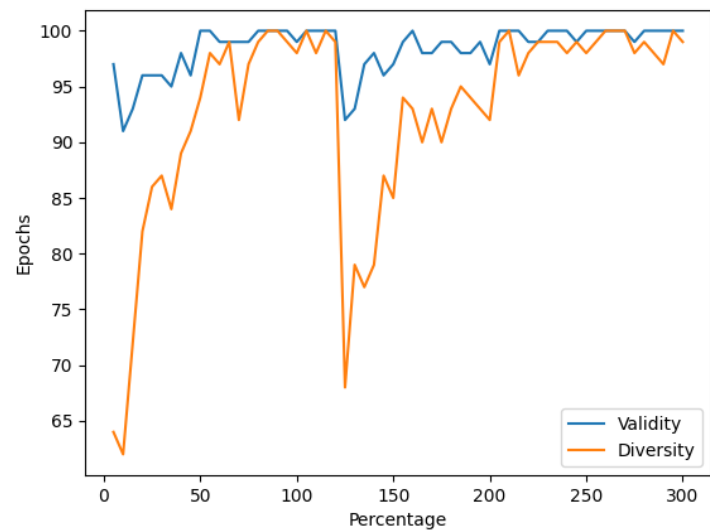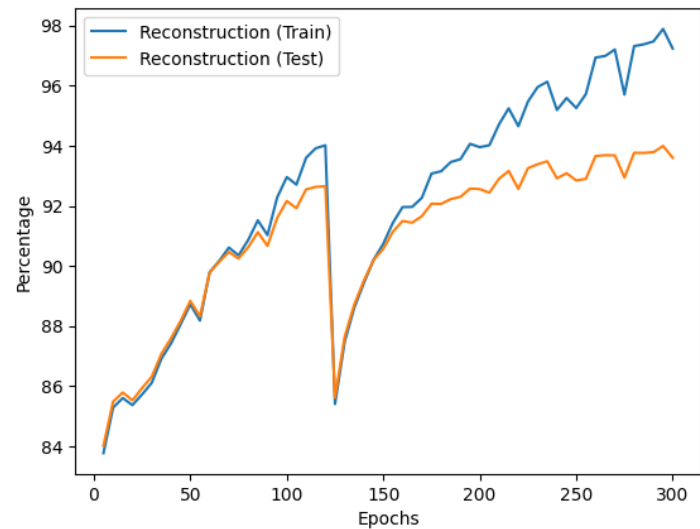
Training Data

Sampled Data

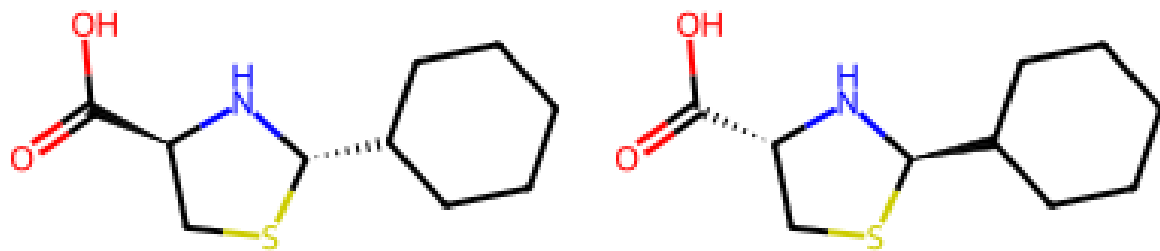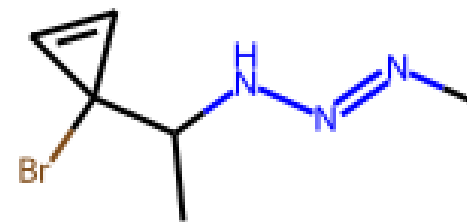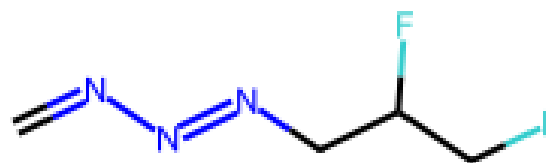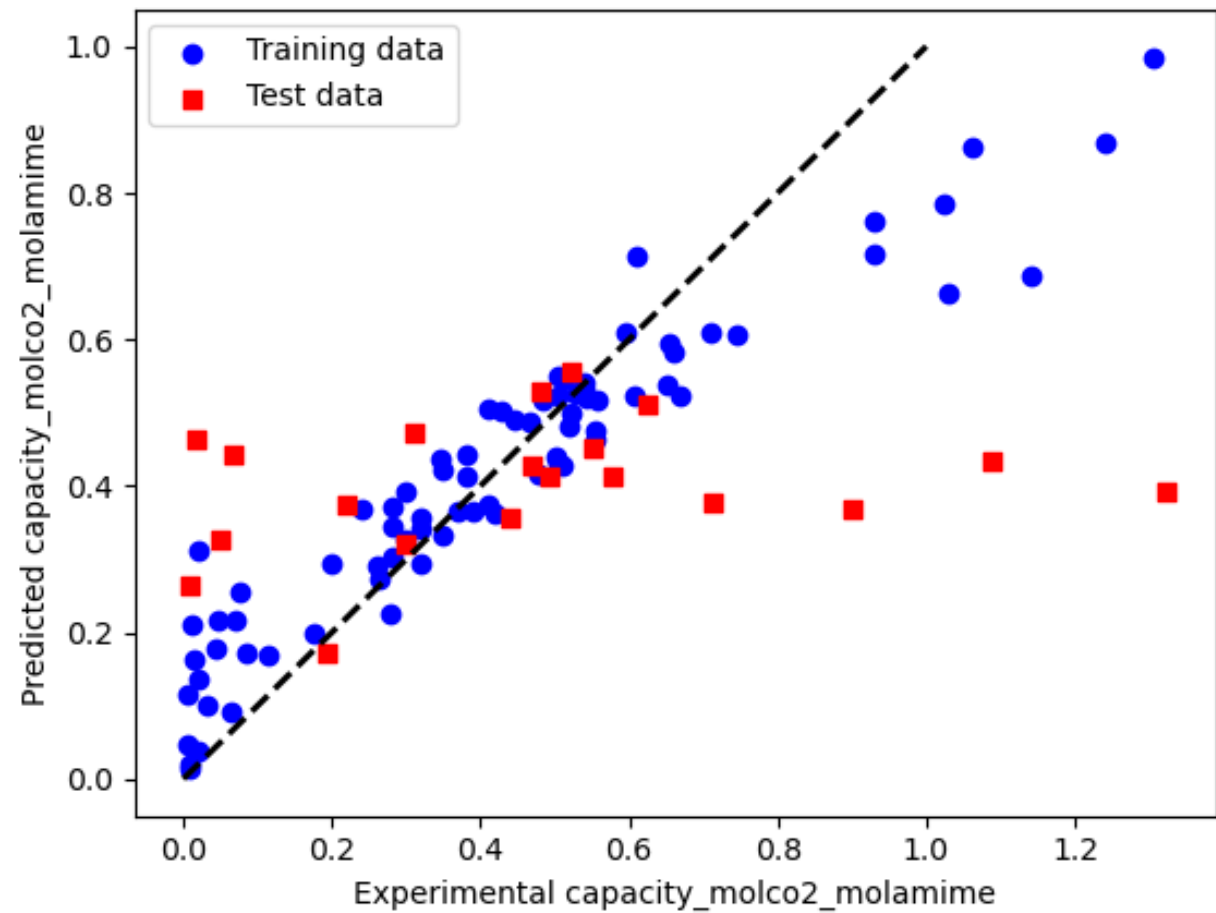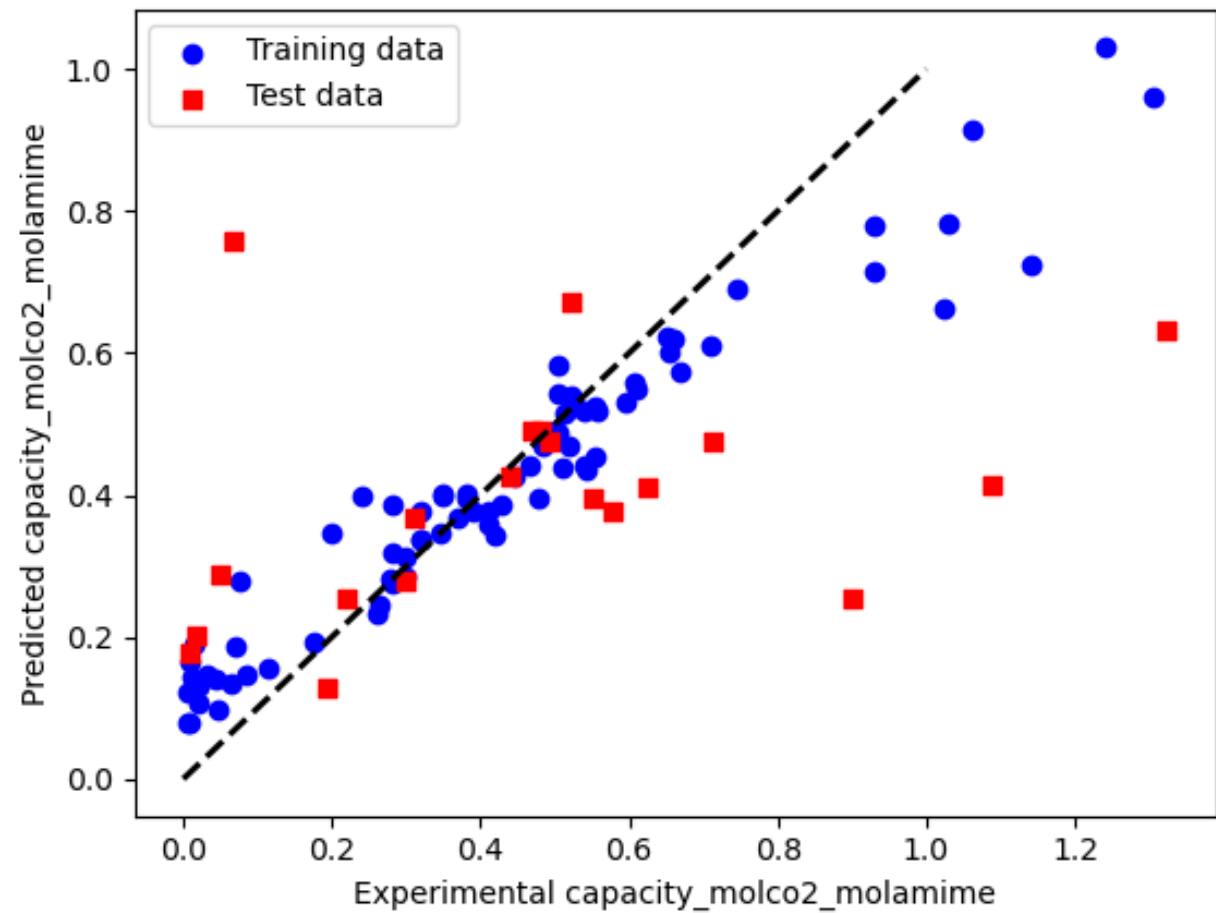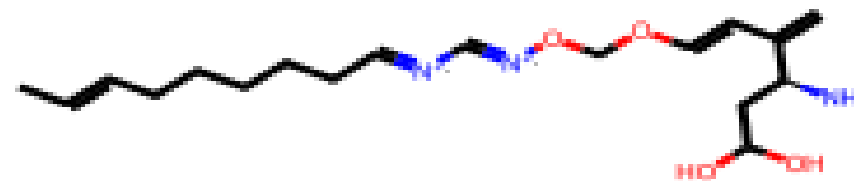# SMILES

SELFIES

Molecules with high CO$_2$ loading

0.86765

0.7521

Pred: 0.396325, Exp: 0.554

What have we learned?

Alhasan:

- Molecular representation SMILES/SELFIES
- Metrics

Griffin:

- Molecular representation SMILES/SELFIES
- VAE in Pytorch practice

Harry:

Convert SMILES to one hot encoding

VAE for molecules design

Florian:

- Molecular representation SMILES/SELFIES
- Metrics

Further improvements

- Models should go beyond black-boxes and can be interpretable and explainable

- Models need to be validated and tested experimentally