Path-Aware and Structure-Preserving Generation of Synthetically Accessible Molecules

Juhwan Noh
KAIST-CBE

Dae-Woong Jeong
LG AI Research

Kiyoung Kim
LG AI Research

Se-Hui Han
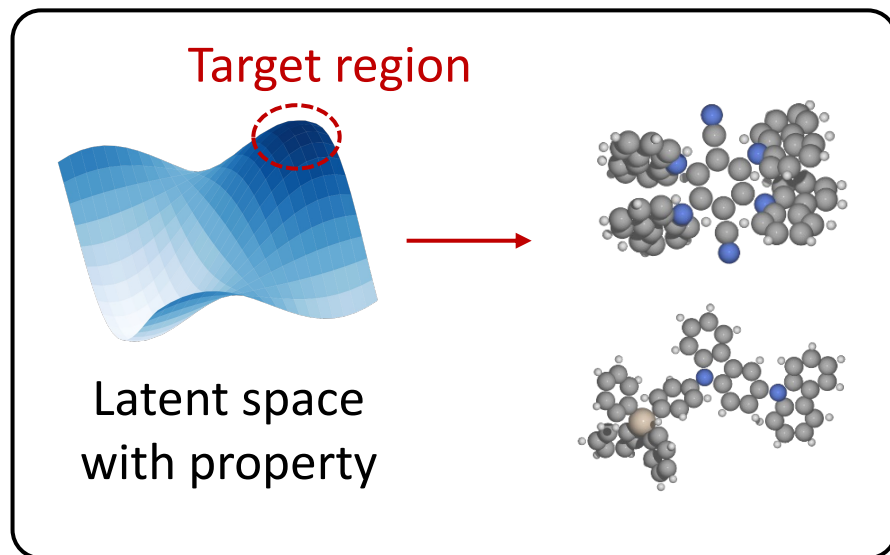LG AI Research

Moontae Lee
LG AI Research, UIC

Honglak Lee
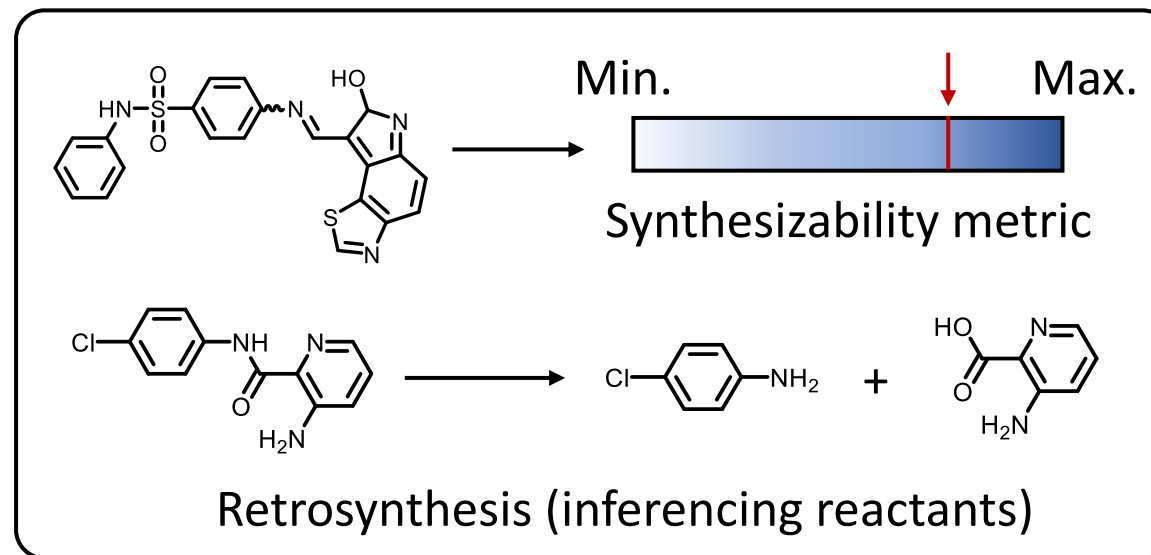LG AI Research

Yousung Jung*
KAIST-CBE/AI

Motivation: Design of New Molecules and Synthesizability

Generative molecular design

Target region

Latent space
with property

Identification of synthesizability

Min.   Max.

Synthesizability metric

Retrosynthesis (inferencing reactants)

- Direct optimization within latent space
- Unknown synthesizability

- Synthesizability metric & retrosynthetic planning
- Sequential design pipeline with generative model
- Iterative & ineffective post-hoc process

Ertl and Schuffenhauer *J. Cheminf.* (2009)

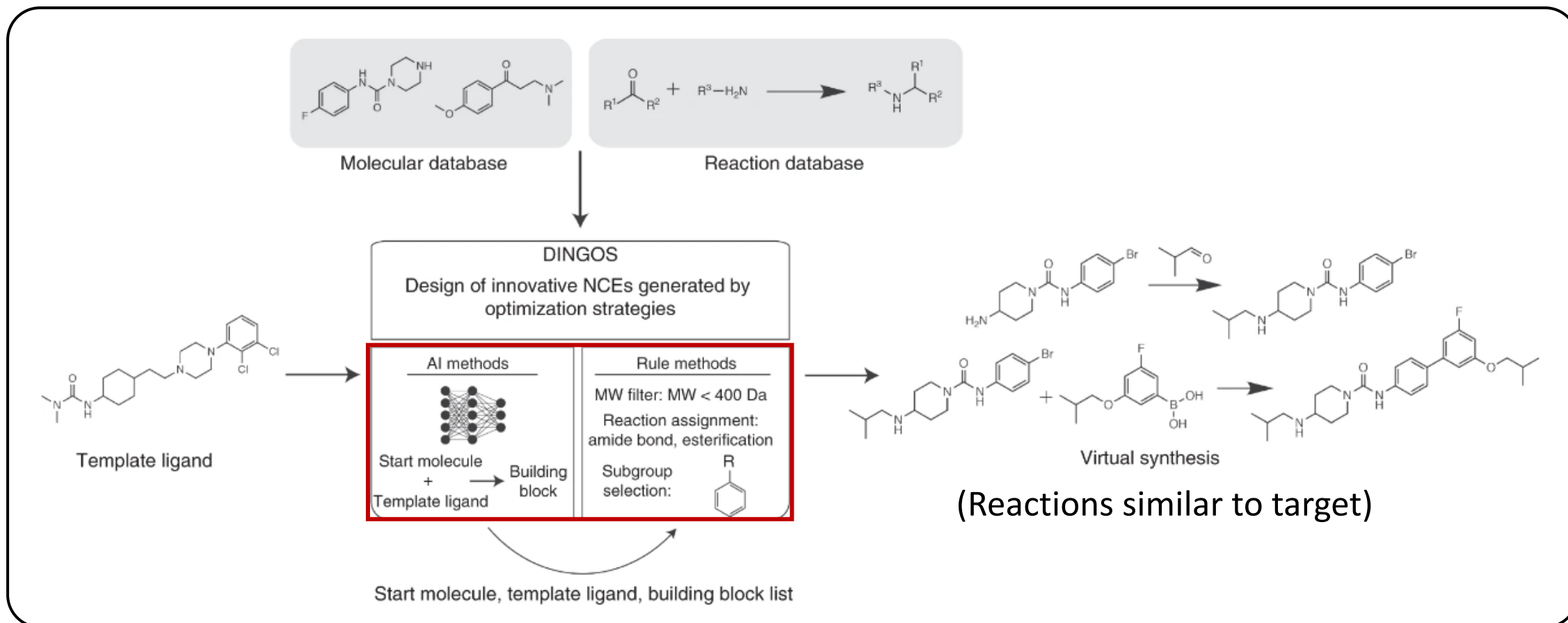Sanchez-Lengeling & Aspuru-Guzik *Science* (2018)   Segler et al. *Nature* (2018)

Scope of This Work

Q1. Can we combine generation and synthesizability?

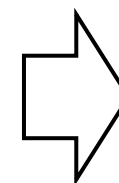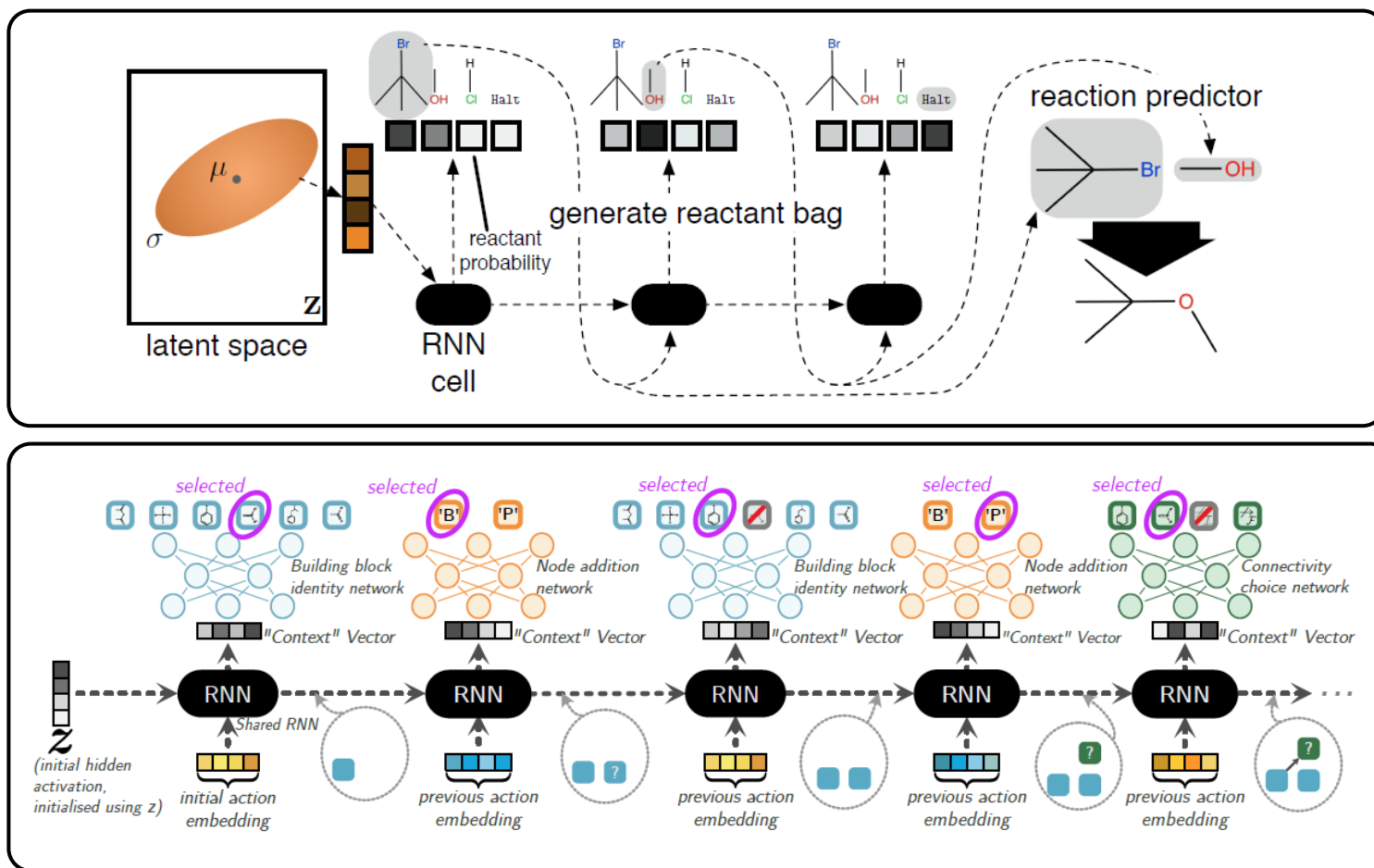Q2. Can we also impose structural constraints during generation?

Related Works: Design of Synthesizable Analogs

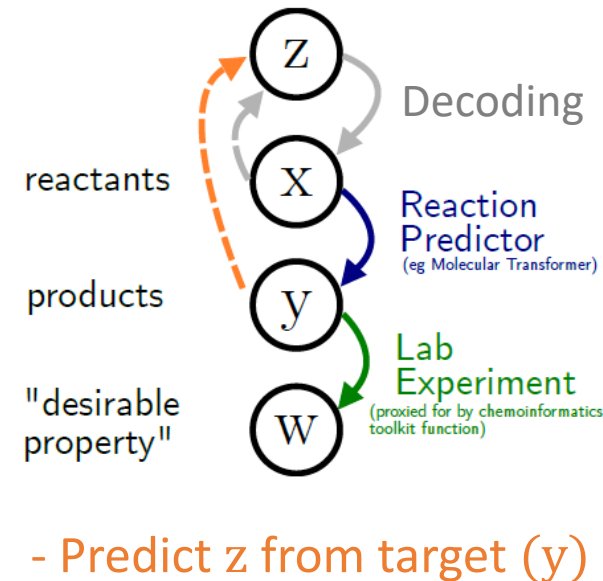Machine learning and rule-based design (DINGOS)

(Reactions similar to target)

Related Works: Design of Synthesizable Analogs

Generative model (WAE) for chemical reactions

Workflow of synthesizable analog design

- Predict z from target (y)

Bradshaw et al. NIPS (2019,2020)
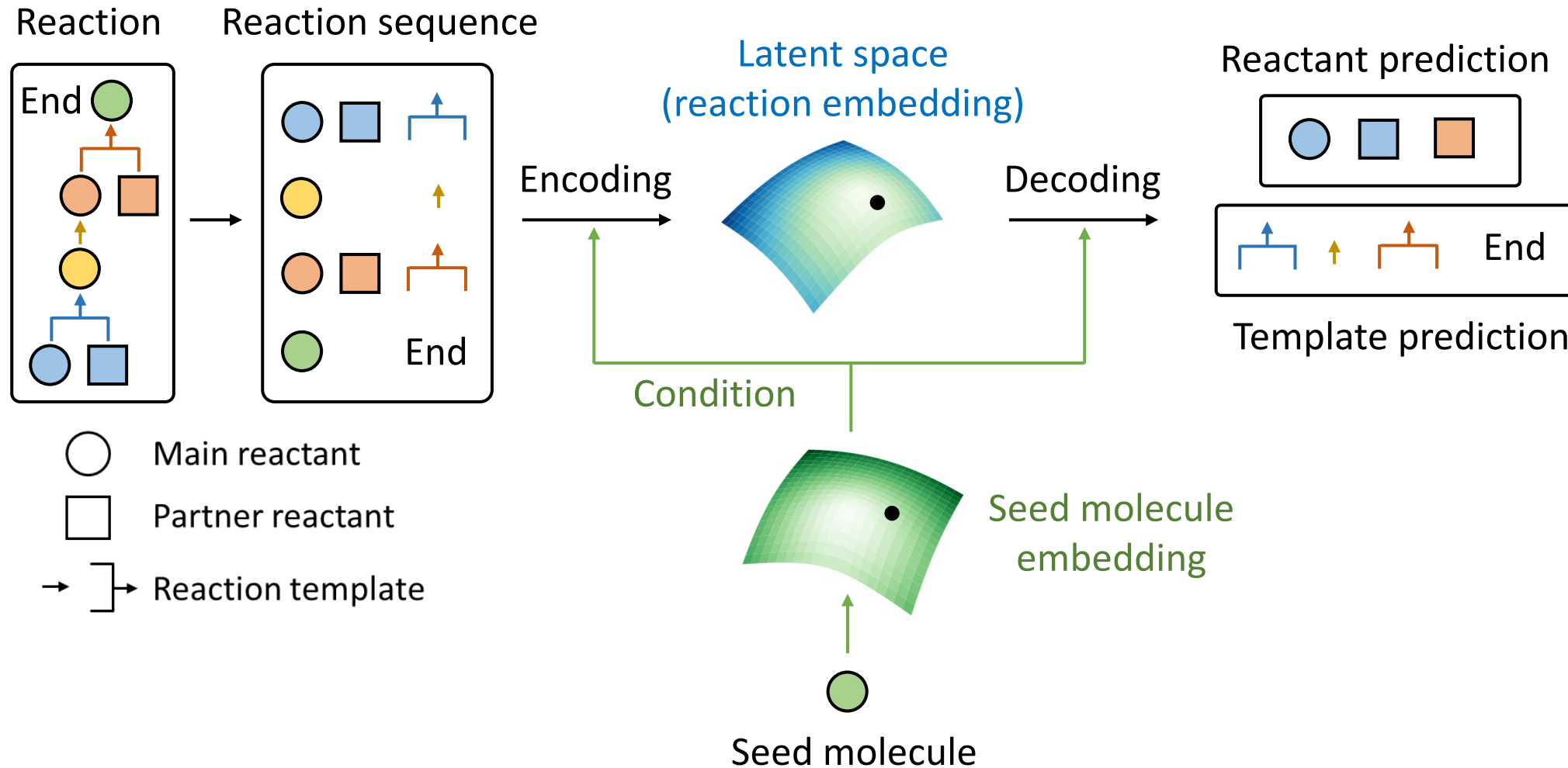
Related Works: Design of Synthesizable Analogs

SynNet: Direct conditional generation using Markov Decision Process (MDP)
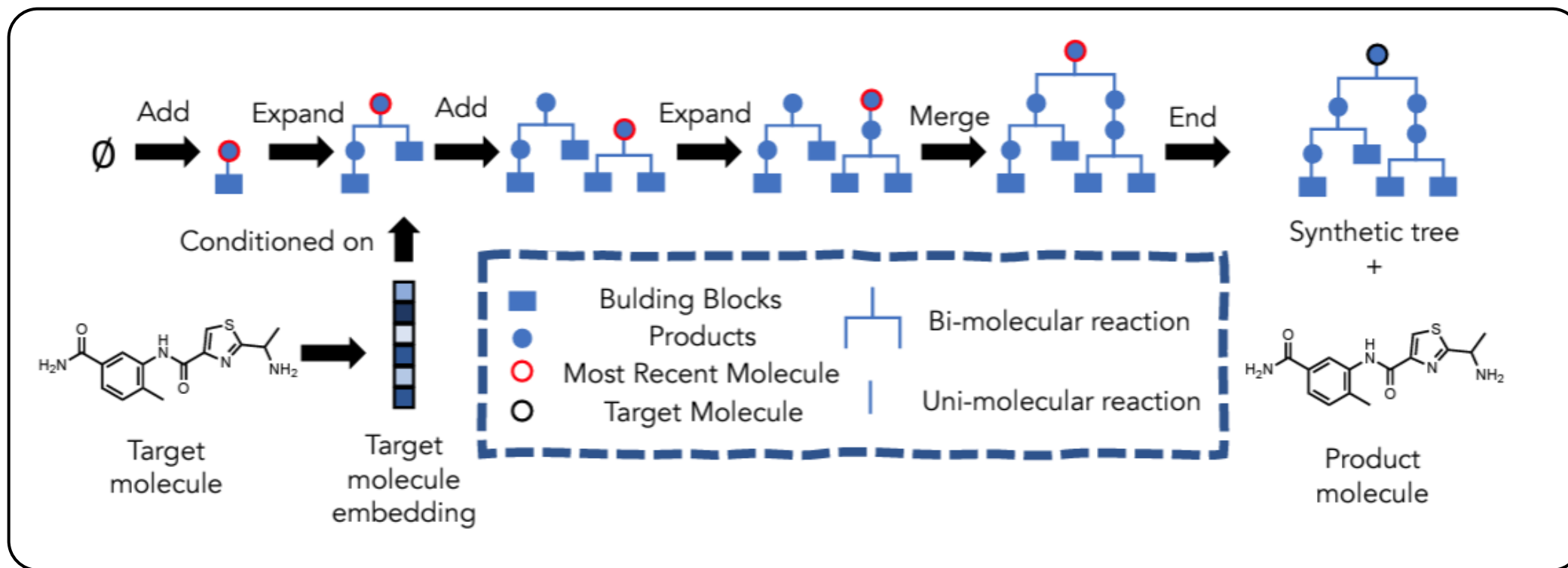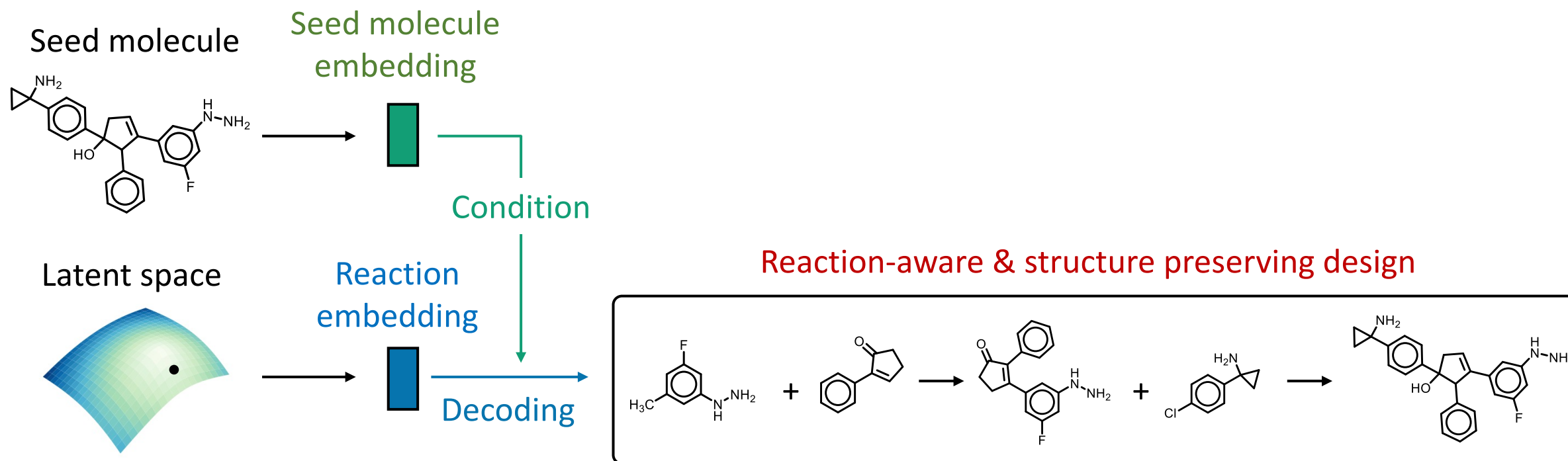

Gao et al. ICLR (2022)

Joint Learning Chemical Reactions and Molecules via Latent Variables

- Conditional Reaction Sequence Variational Autoencoder (C-RSVAE)
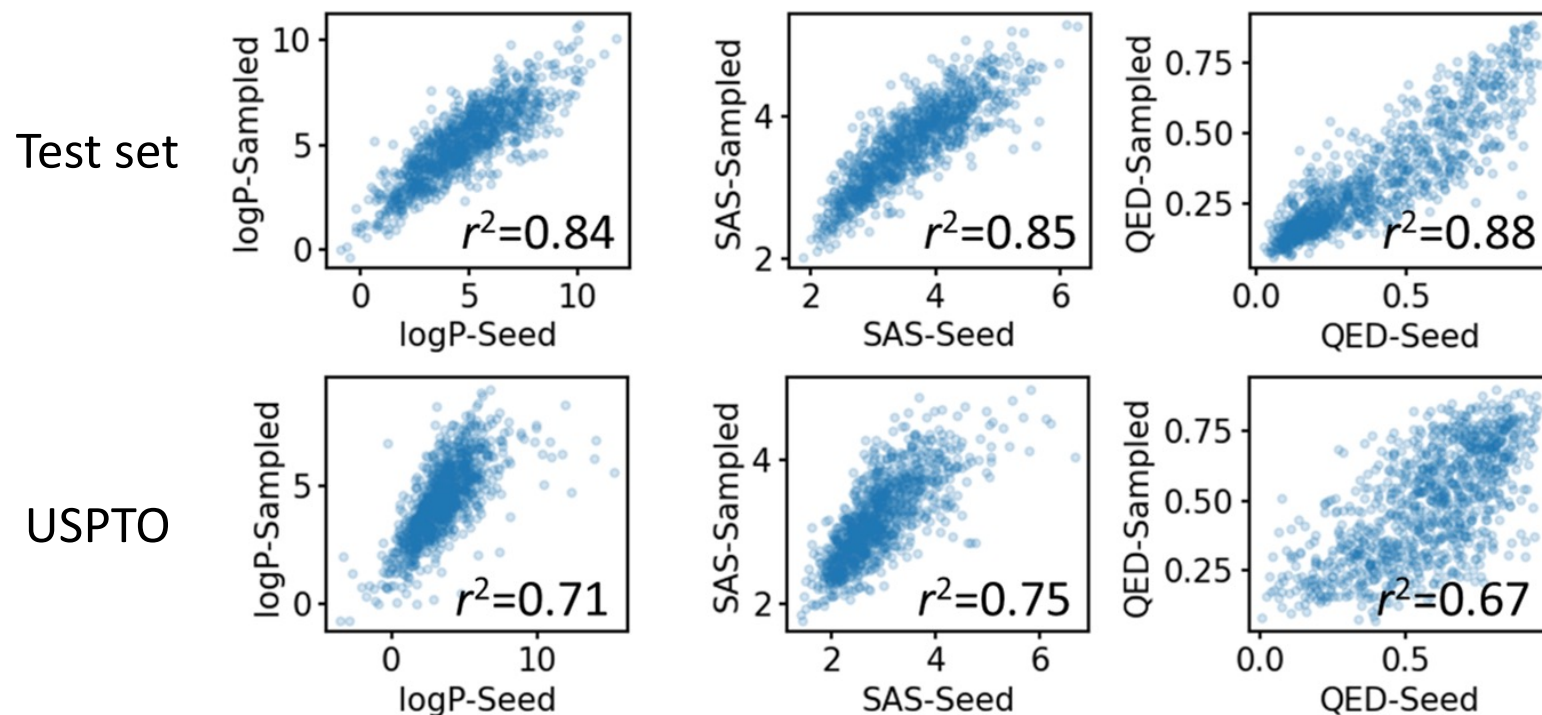
Joint Learning Chemical Reactions and Molecules via Latent Variables

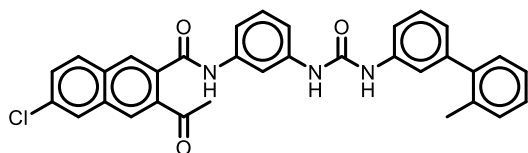- Conditional Reaction Sequence Variational Autoencoder (C-RSVAE)

Experiments

Correlation between properties from generated & seed molecules

Test set

USPTO

- 1,000 seed molecules in test set/USPTO

- Properties averaged with top 20 similar generated molecules are used

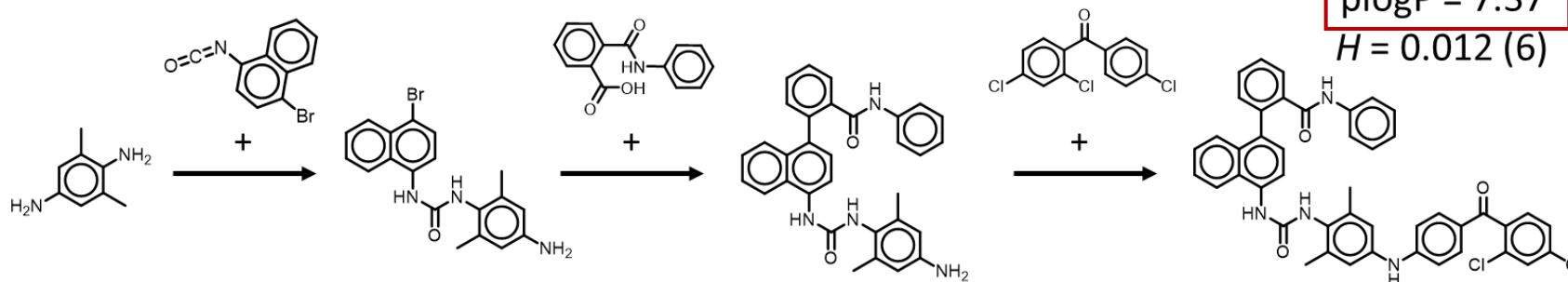- High correlation coefficient to the properties of seed compounds
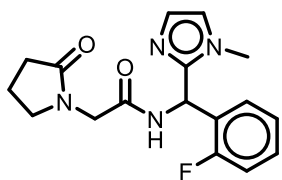
Experiments

Seed from JT-VAE

Proposed reactions by this work

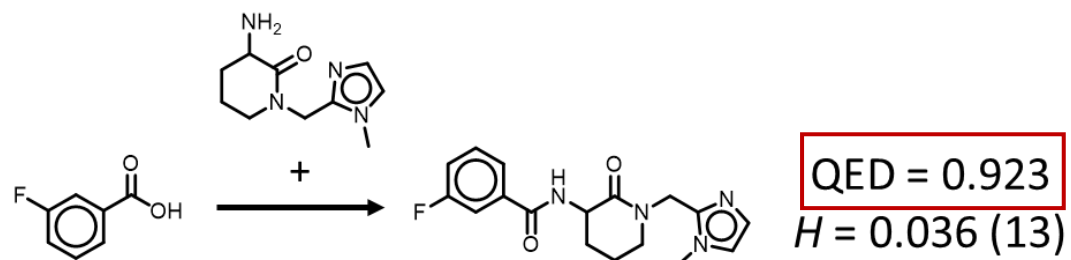plogP = 5.30

plogP = 7.37
H = 0.012 (6)

Seed from SSVAE

Proposed reactions by this work

QED = 0.901

QED = 0.923
H = 0.036 (13)

Jin et al. NIPS (2018)
Kang and Cho, *J. Chem. Inf. Model* (2018)

Experiments

DRD3 (PDB ID: 3PBL)

Proposed reaction of 1st in *this work*

Vina score: -11.9 kcal/mol
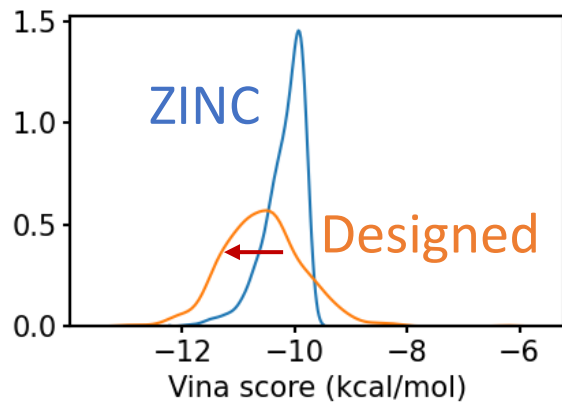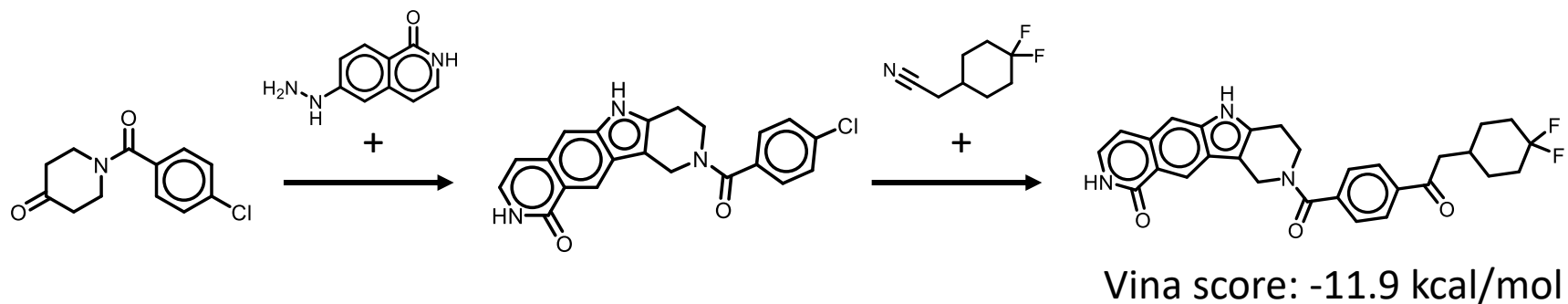
SARS-Cov-2 (PDB ID: 7L11)

Proposed reaction of 1st in *this work*

Vina score: -10.3 kcal/mol

TDC Interface (https://tdcommons.ai/)
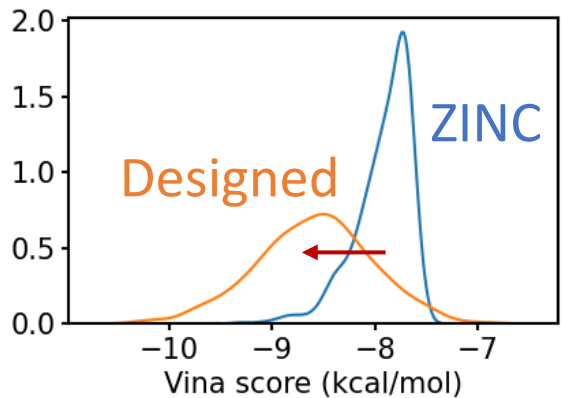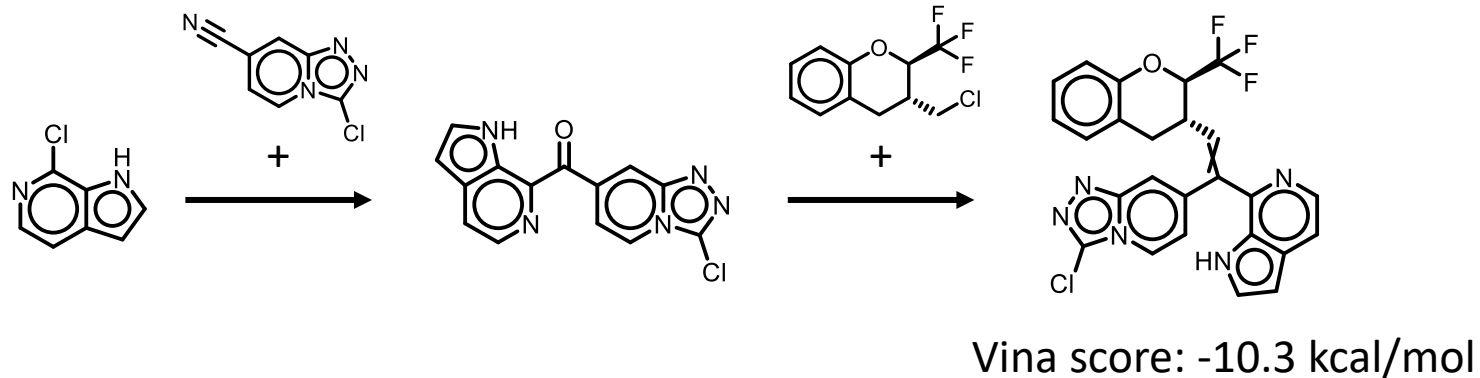
Conclusions

- Generative framework for synthesizable molecules preserving structural similarity to target compound

- We found that high property correlation derived from imposing structural similarity during generation

- We show that model could effectively design synthesizable molecules with improved property

Supported by