

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

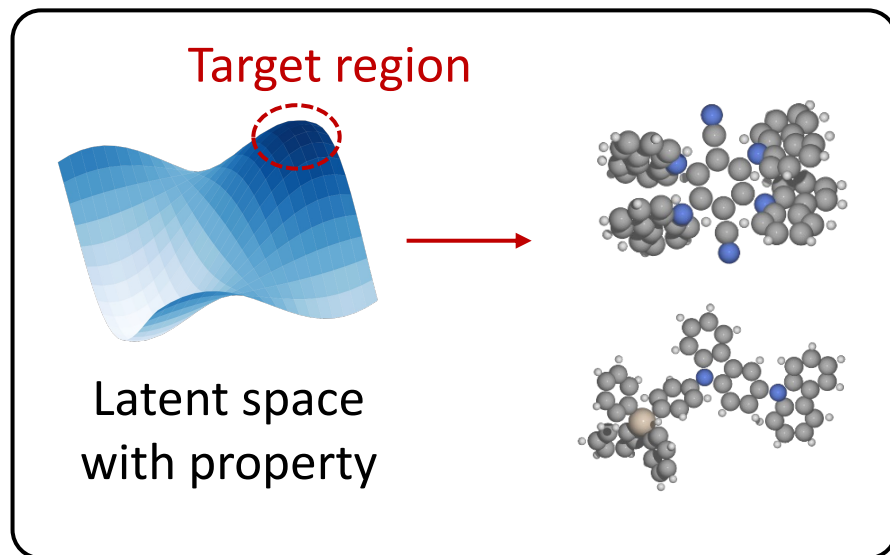
Honglak Lee
LG AI Research

Yousung Jung*
KAIST-CBE/AI



Motivation: Design of New Molecules and Synthesizability

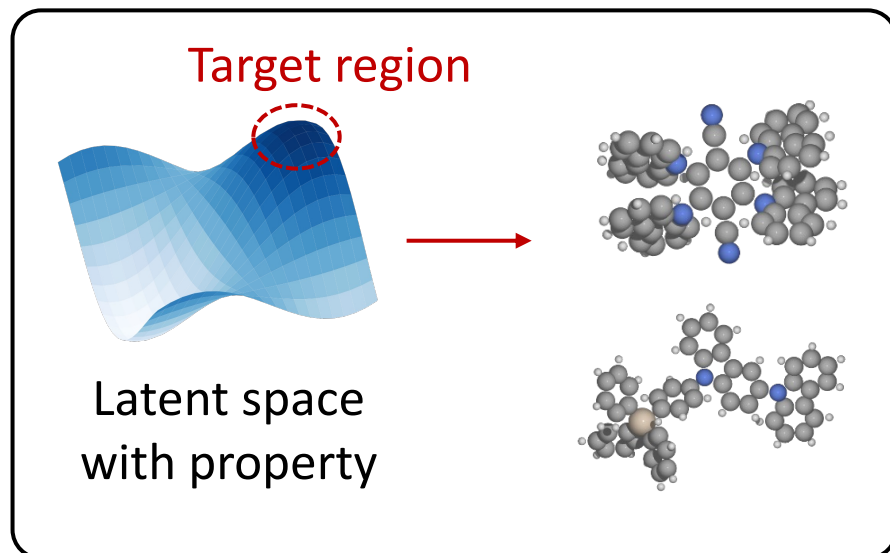
Generative molecular design



- Direct optimization within latent space
- **Unknown synthesizability**

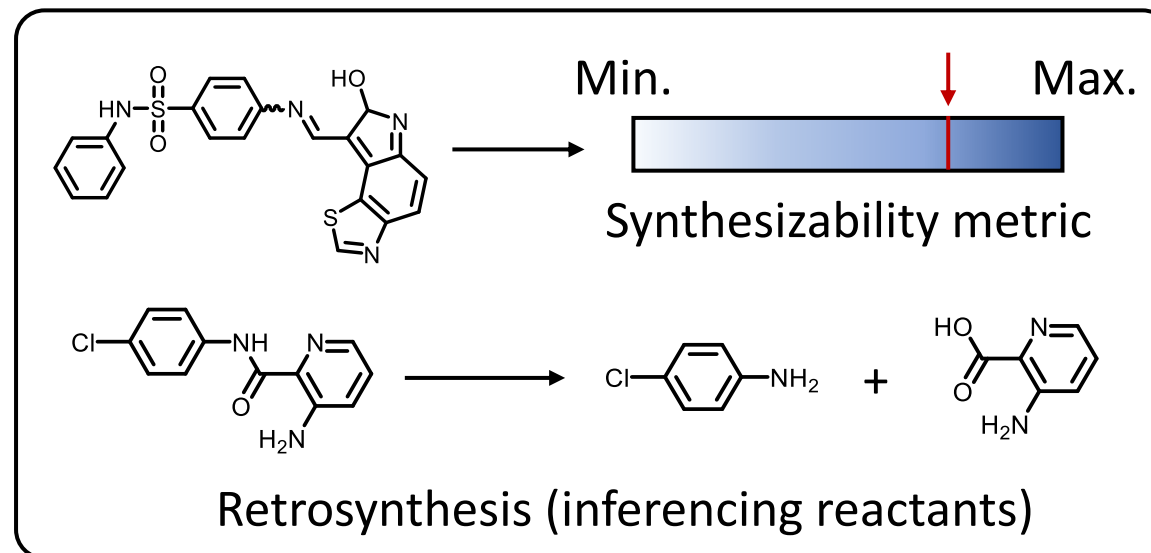
Motivation: Design of New Molecules and Synthesizability

Generative molecular design



- Direct optimization within latent space
- **Unknown synthesizability**

Identification of synthesizability



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

Scope of This Work

Q1. Can we combine generation and synthesizability?

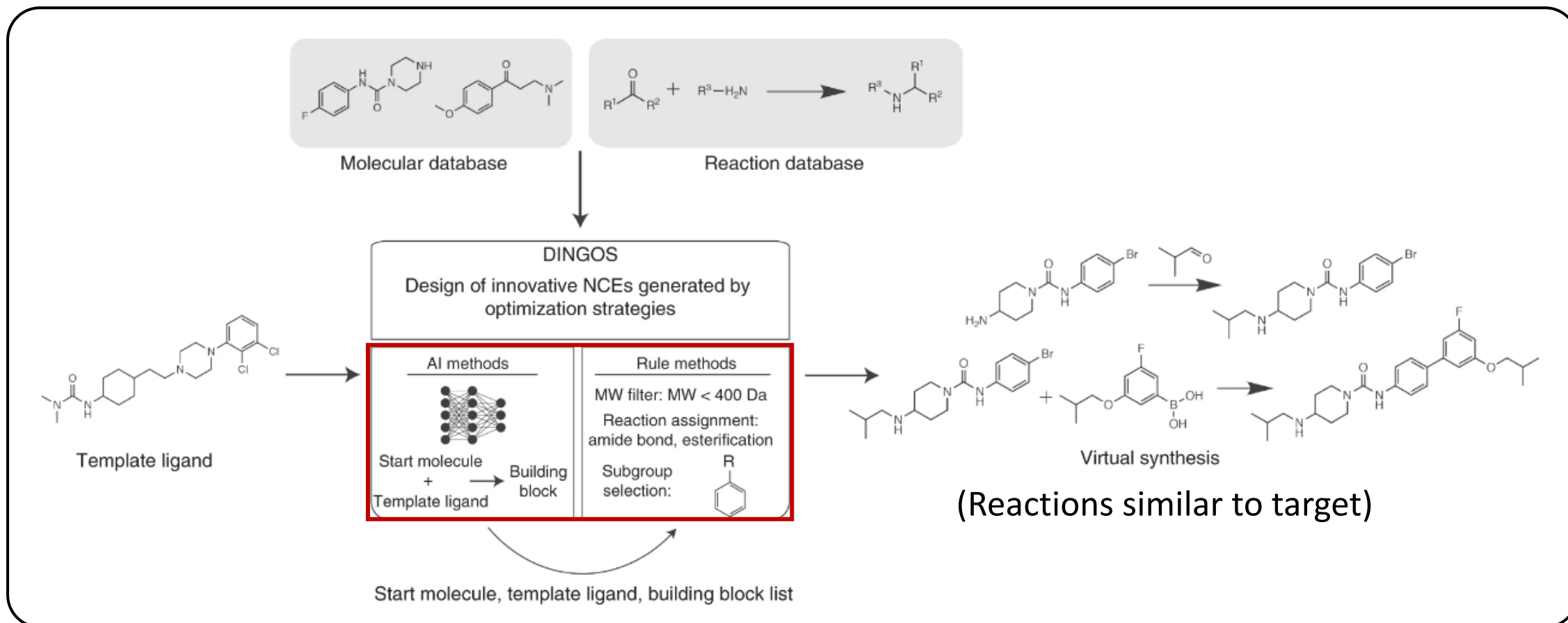
Scope of This Work

Q1. Can we combine generation and synthesizability?

Q2. Can we also impose structural constraints during generation?

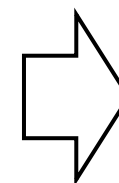
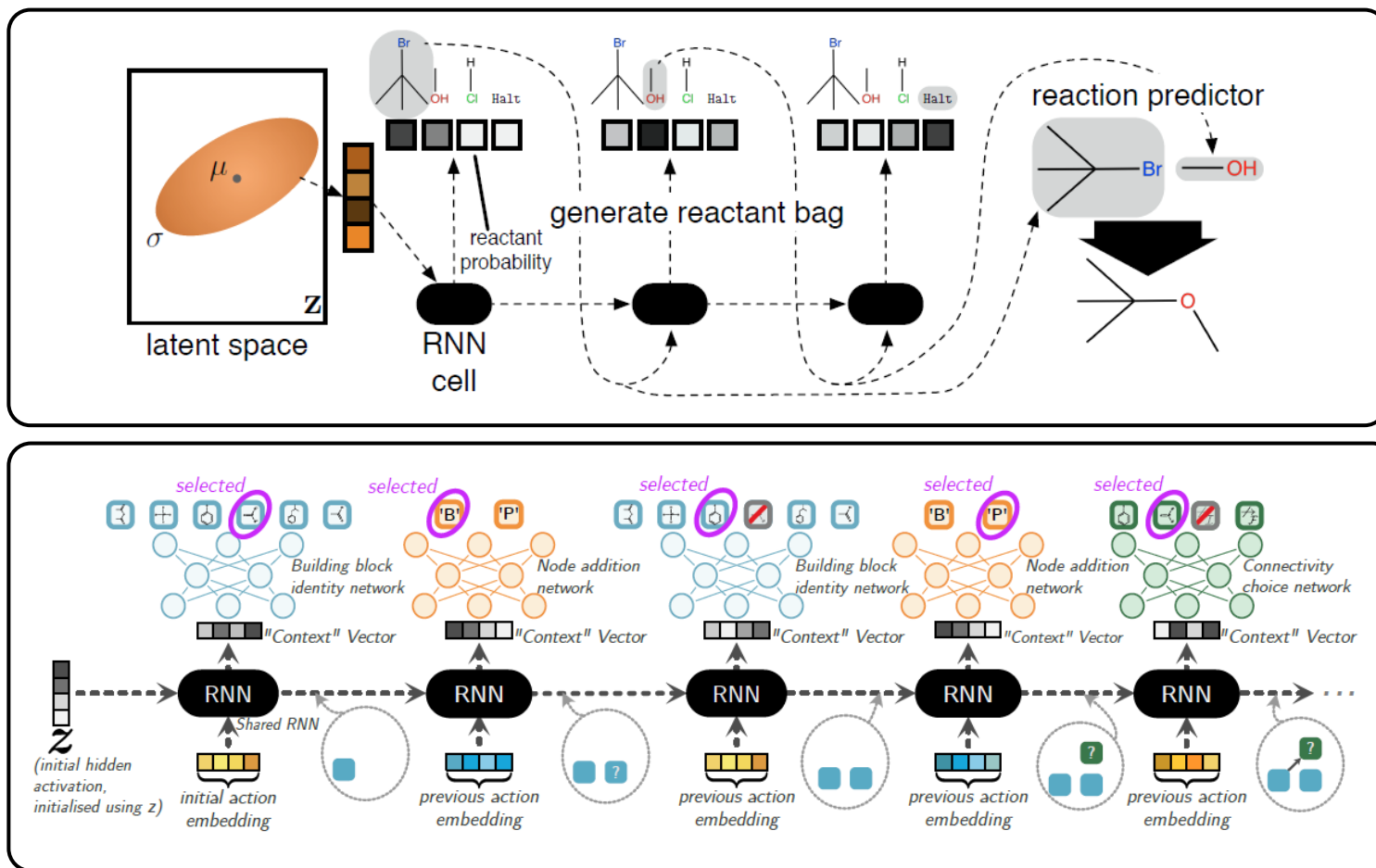
Related Works: Design of Synthesizable Analogs

Machine learning and rule-based design (DINGOS)

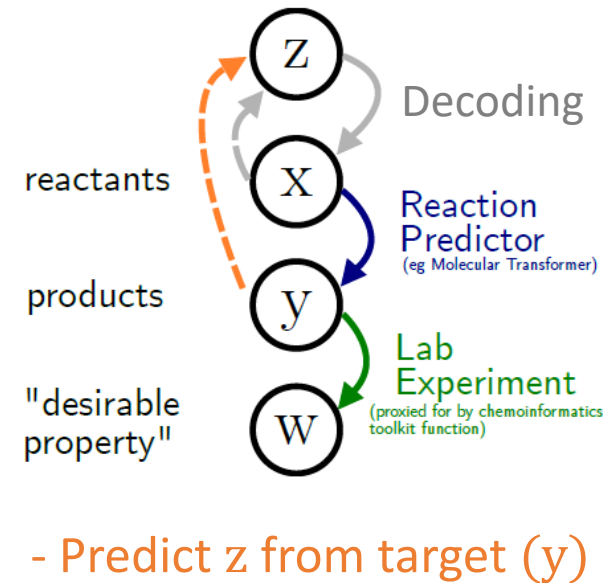


Related Works: Design of Synthesizable Analogs

Generative model (WAE) for chemical reactions

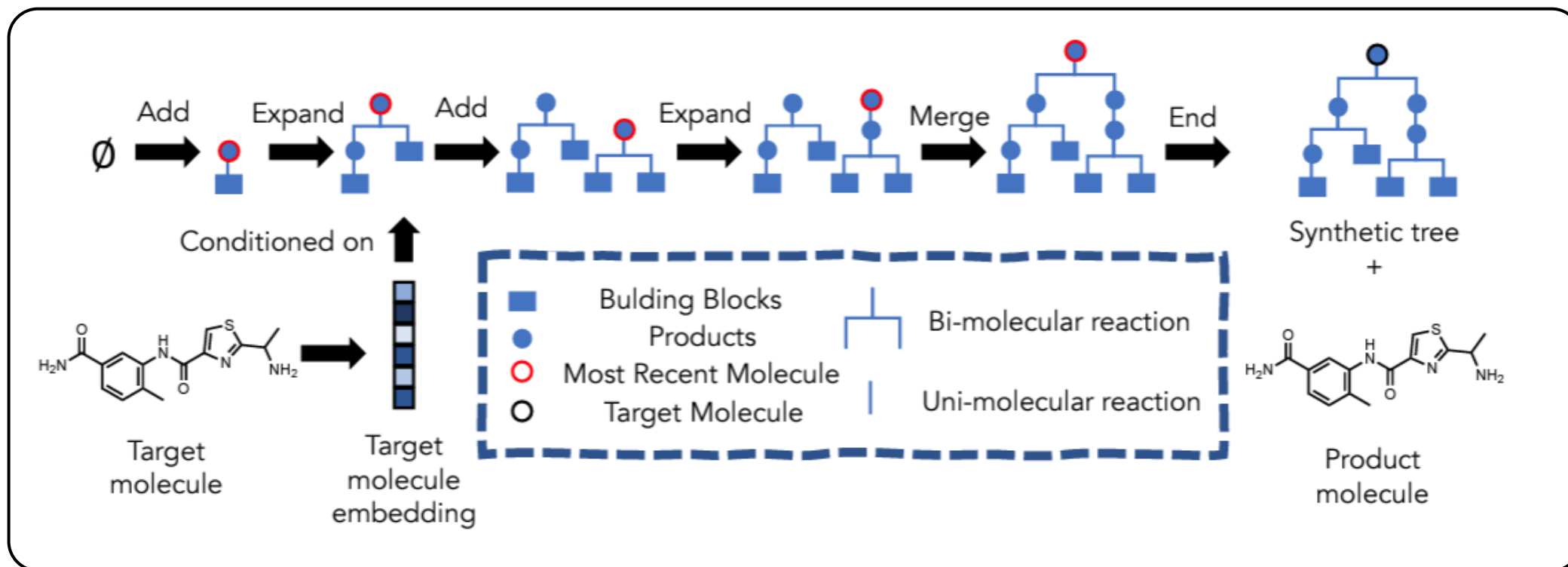


Workflow of synthesizable analog design



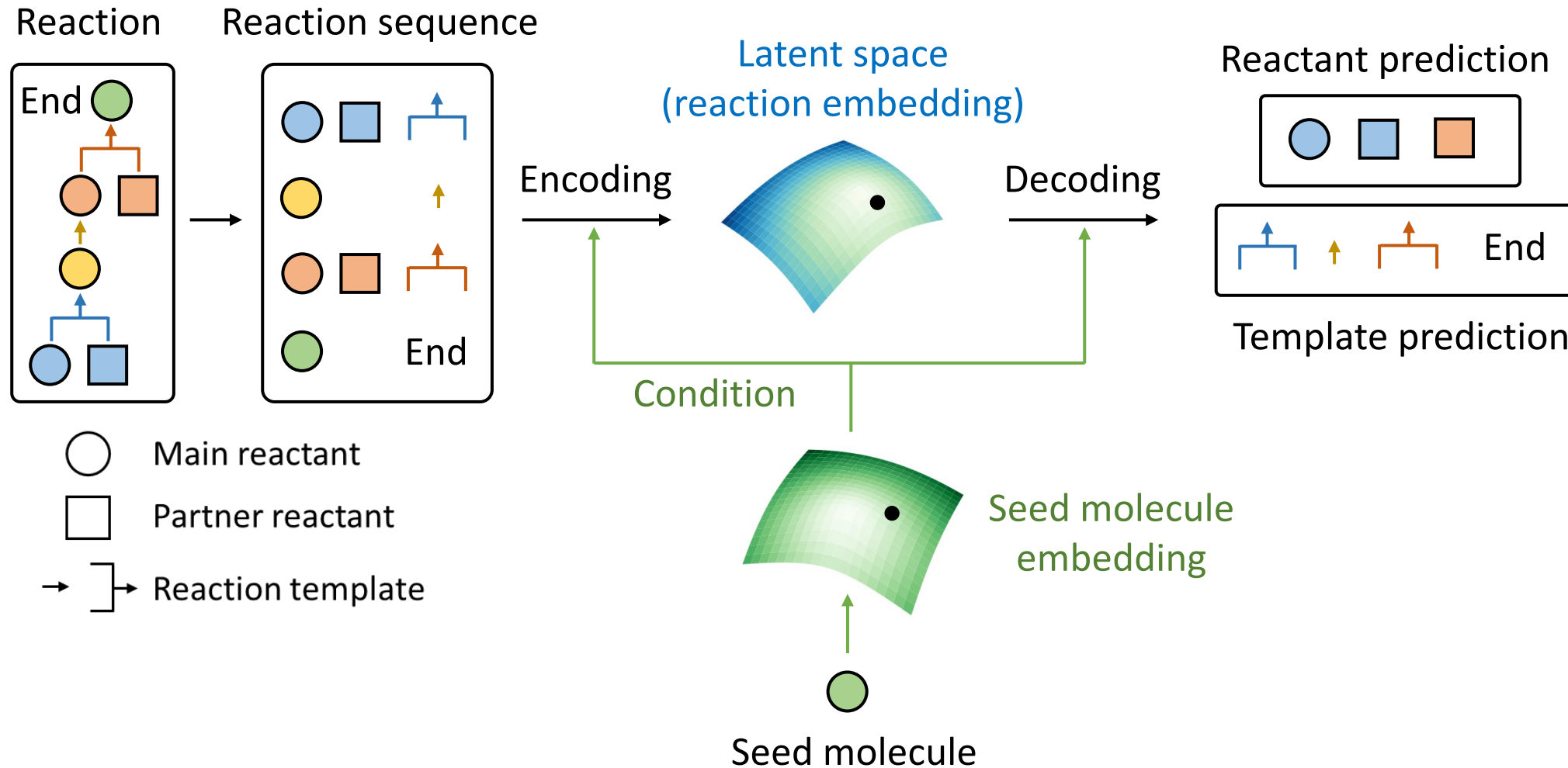
Related Works: Design of Synthesizable Analogs

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



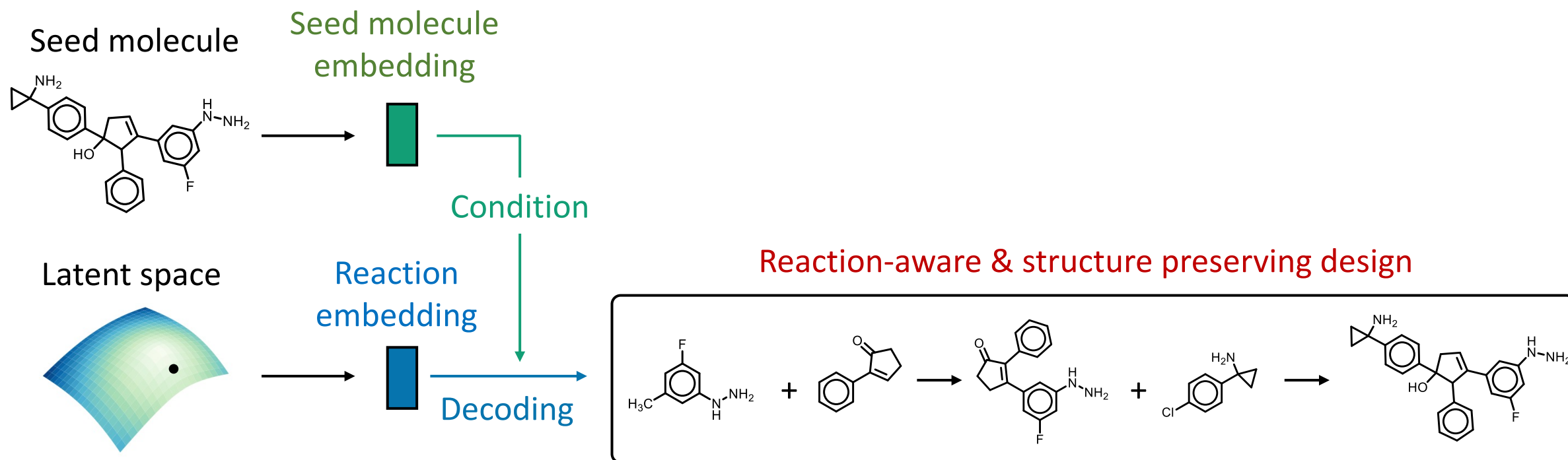
Joint Learning Chemical Reactions and Molecules via Latent Variables

- Conditional Reaction Sequence Variational Autoencoder (C-RSVAE)



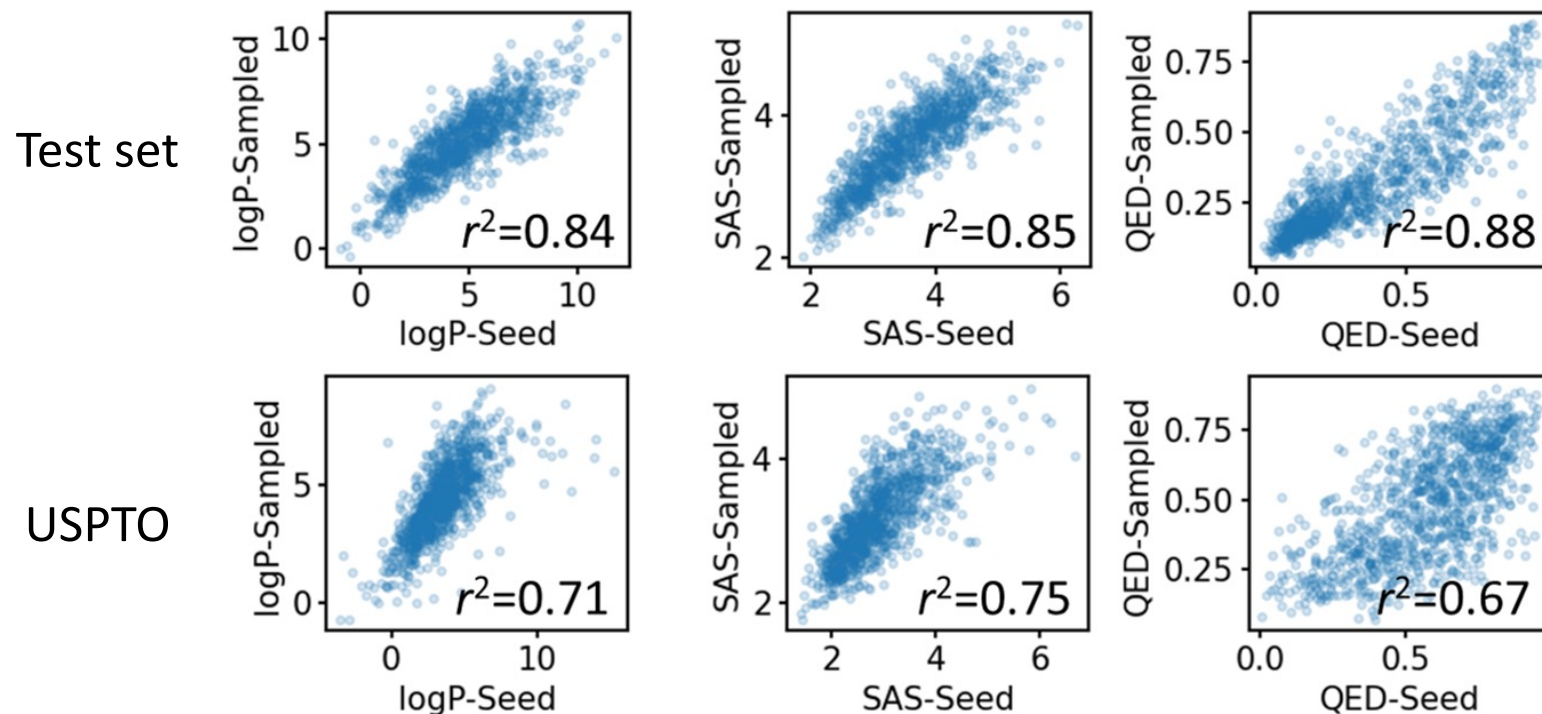
Joint Learning Chemical Reactions and Molecules via Latent Variables

- Conditional Reaction Sequence Variational Autoencoder (C-RSVAE)



Experiments

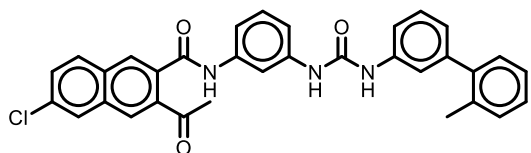
Correlation between properties from generated & seed molecules



- 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

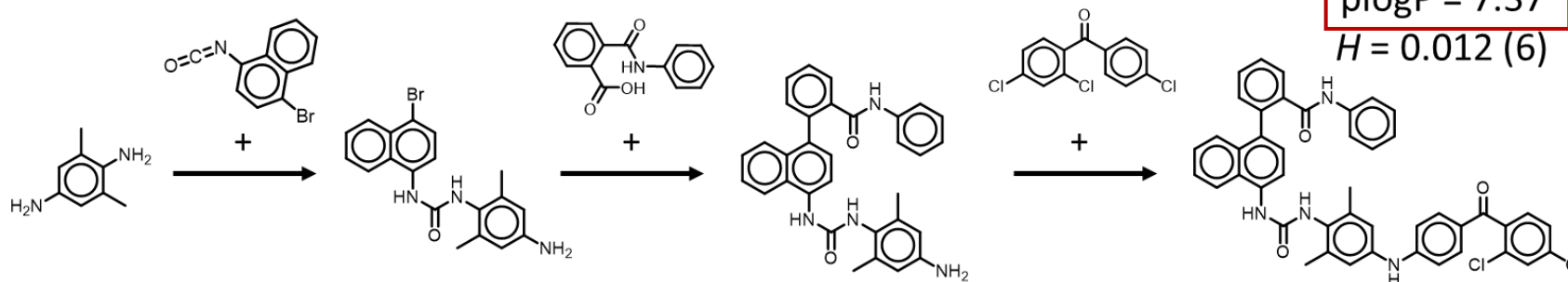
Experiments

Seed from JT-VAE

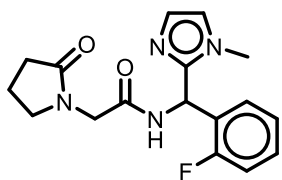


plogP = 5.30

Proposed reactions by this work

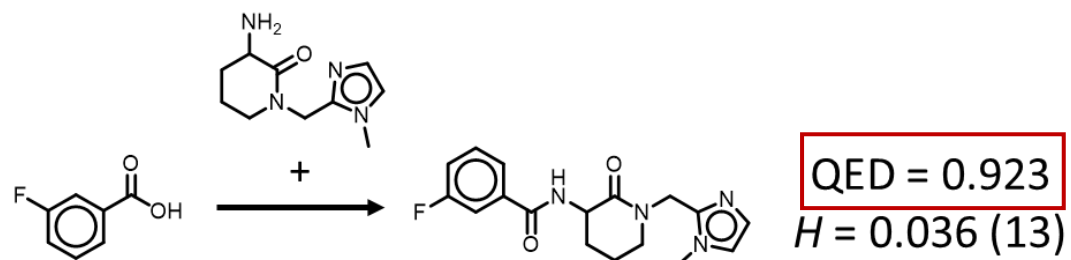


Seed from SSSVAE



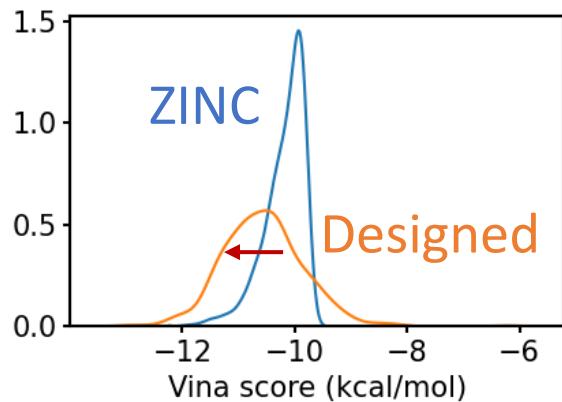
QED = 0.901

Proposed reactions by this work

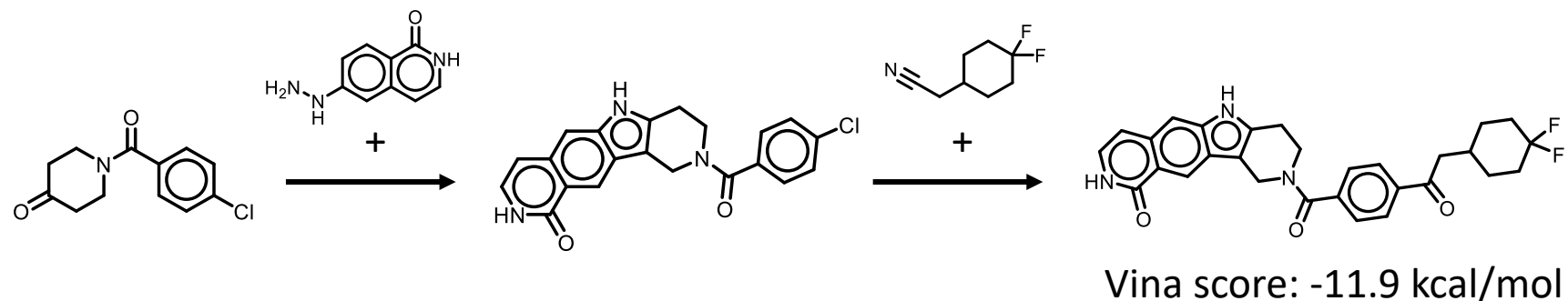


Experiments

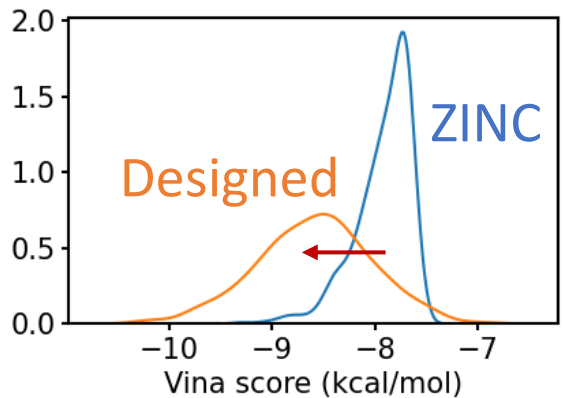
DRD3 (PDB ID: 3PBL)



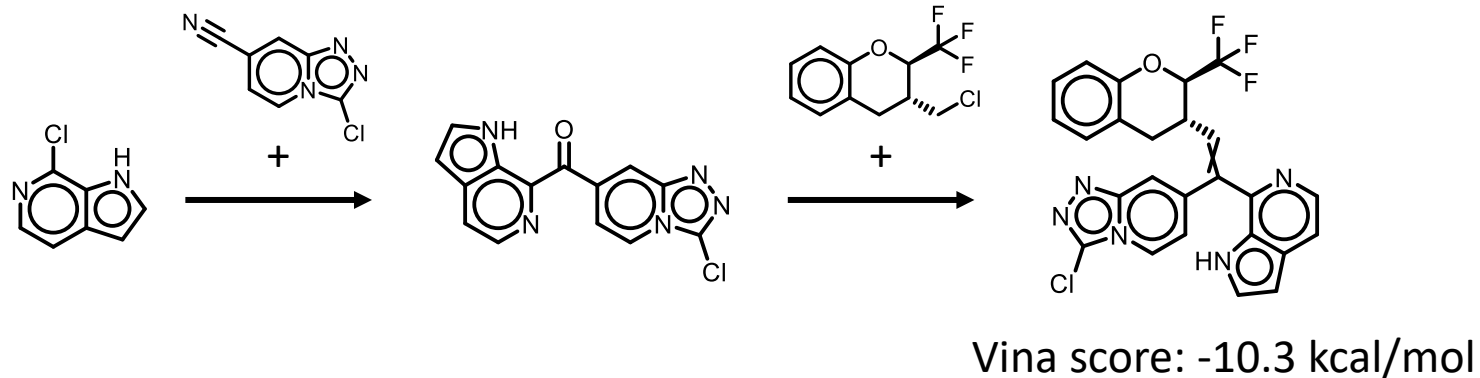
Proposed reaction of 1st in *this work*



SARS-Cov-2 (PDB ID: 7L11)



Proposed reaction of 1st in *this work*



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



Institute of Information & Communications
Technology Planning & Evaluation