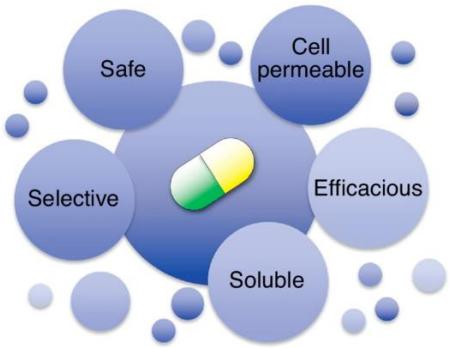
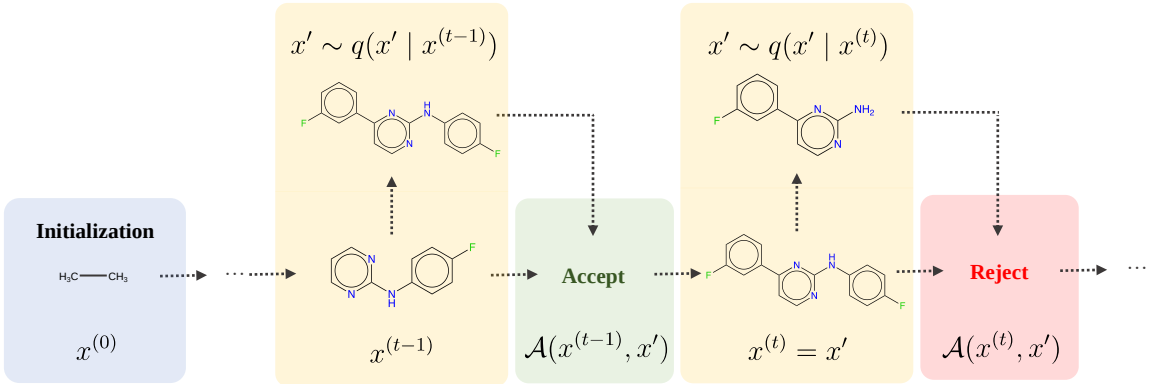
MARS: Markov Molecular Sampling for Multi-objective Drug Discovery

Yutong Xie[1,2], Chence Shi[1,3], Hao Zhou[1], Yuwei Yang[1], Weinan Zhang[4], Yong Yu[4], Lei Li[1]

[1]ByteDance AI Lab, China
[2]University of Michigan, USA
[3]Quebec AI Institute (Mila), Canada
[4]Shanghai Jiao Tong University, China

Preview





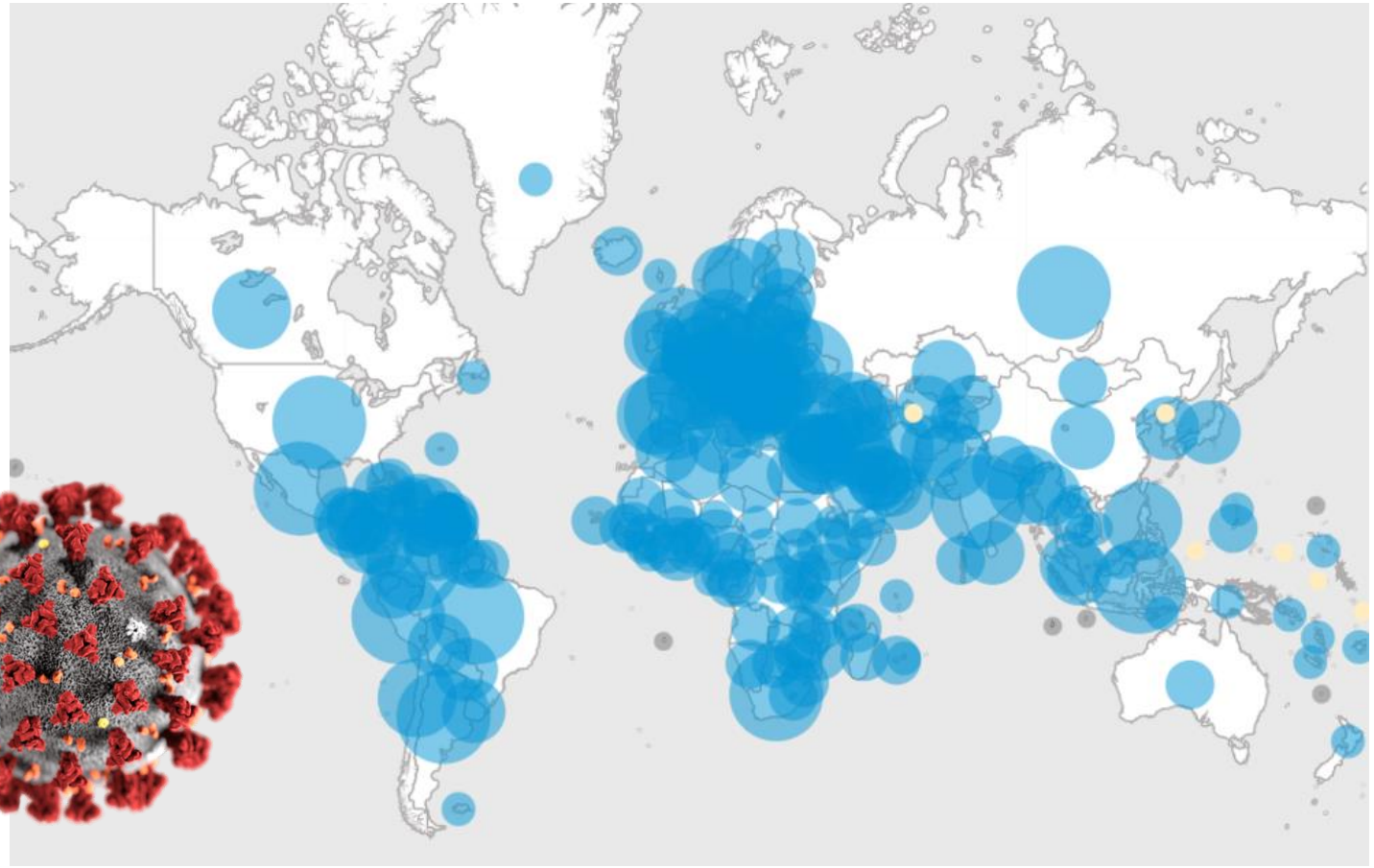
Multi-objective Drug Discovery

MARS Framework

1 Background
2 Challenges

3 Proposed Method
4 Experiment Results

COVID-19 Pandemic



- In 2020, COVID-19 pandemic rapidly spread over the world.
- More than 123 million confirmed cases including over 2.7 million deaths.

https://covid19.who.int/

Drug Discovery Phases

Successful applications in drug discovery

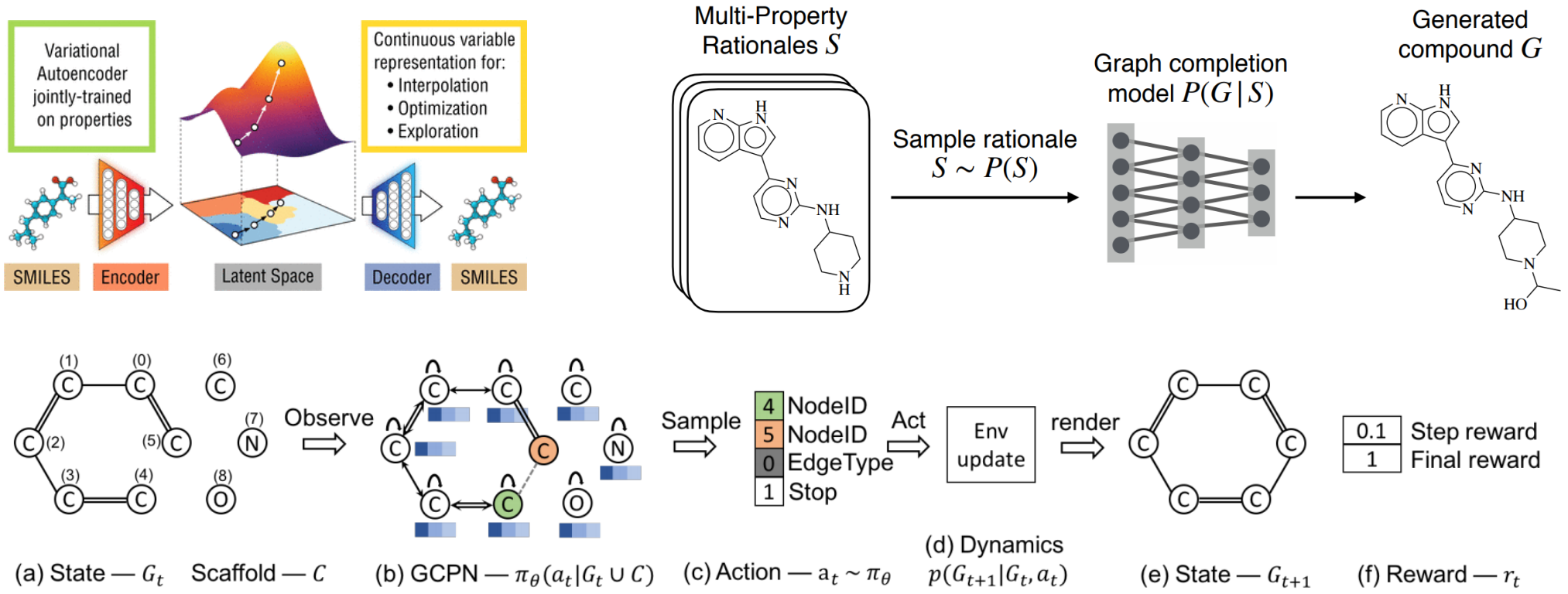
- Target identification and prioritization based on gene–disease associations
- Target druggability predictions
- Identification of alternative targets (splice variants)

- Compound design with desirable properties
- Compound synthesis reaction plans
- Ligand-based compound screening

- Tissue-specific biomarker identification
- Classification of cancer drug–response signatures
- Prediction of biomarkers of clinical end points

- Determination of drug response by cellular phenotyping in oncology
- Precise measurements of the tumour microenvironment in immuno-oncology

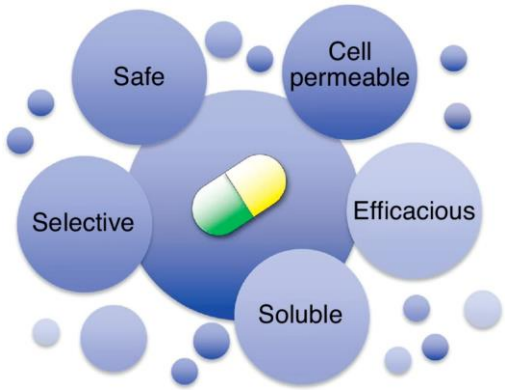
AI Powered Drug Discovery



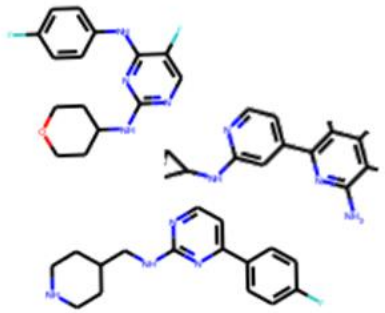
Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules, Gomez-Bombarelli et al., ACS Central Science 2016
Graph Convolutional Policy Network for Goal-Directed Molecular Graph Generation, You et al., NeurIPS 2018
Multi-Objective Molecule Generation using Interpretable Substructures, Jin et al., ICML 2020

Challenges

Multiple properties



Diverse and novel



Lack of annotated data

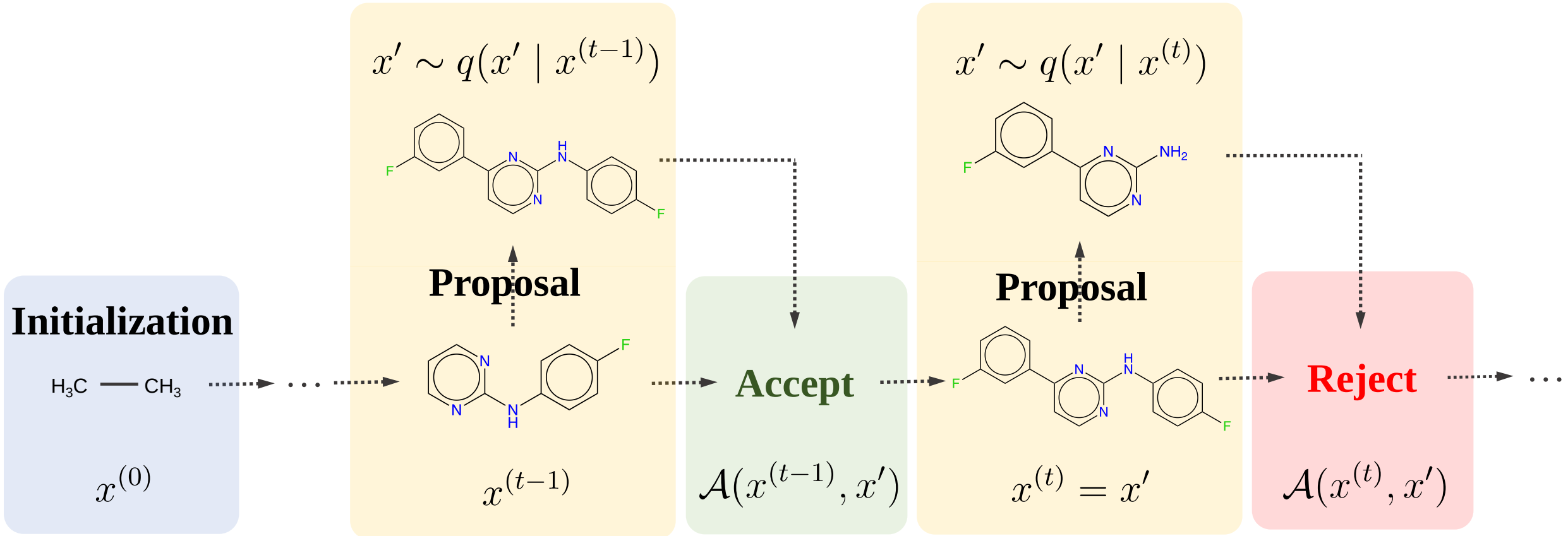
MArkov moleculaR Sampling

- **A combination of multiple objectives: can be complex!**

$$\pi(x) = \underbrace{s_1(x) \circ s_2(x) \circ s_3(x) \circ \cdots \circ s_K(x)}_{\text{desired properties}}$$

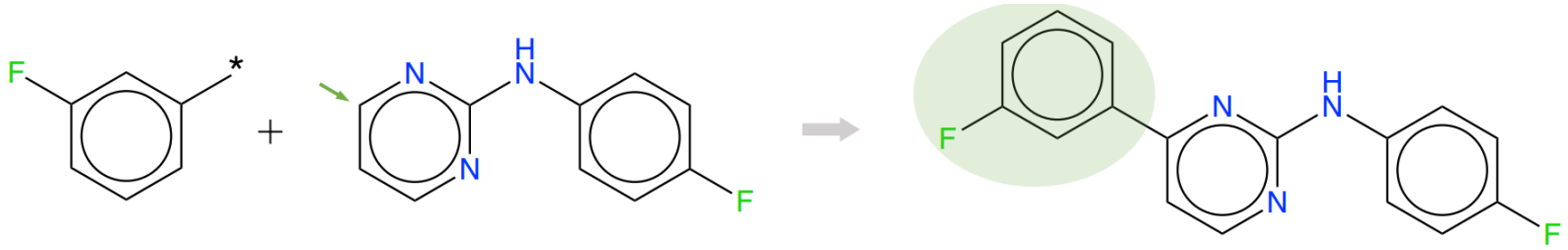
- **Markov-chain Monte Carlo (MCMC) sampling**
- **Adaptive molecular graph editing proposal**

MARS Framework

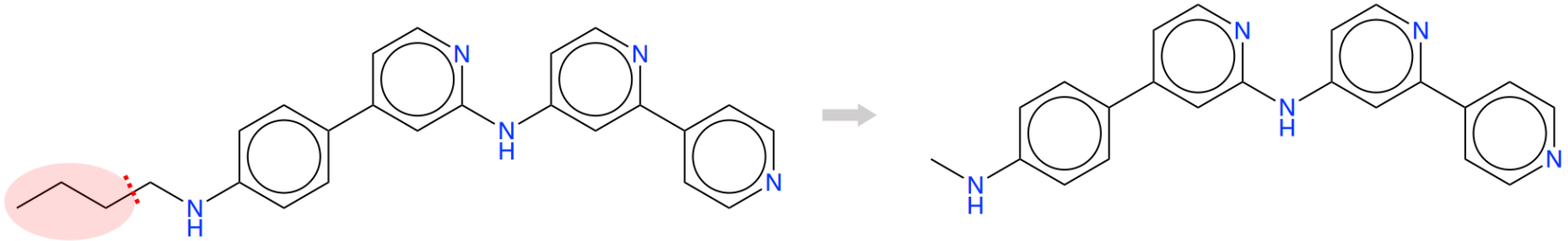


Molecular Graph Editing with MPNNs

Adding Fragment



Deleting Fragment



Parameterized with MPNNs: choosing atoms, fragments, and bonds

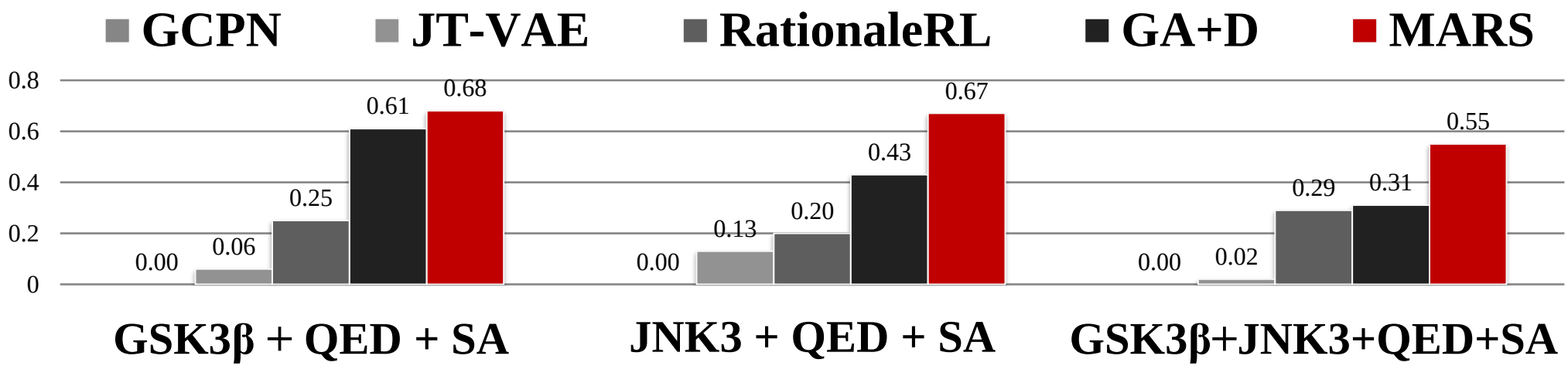
Fragment based drug design: from experimental to computational approaches, Kumar et al., Current Medicinal Chemistry, 2012
Neural message passing for quantum chemistry, Gilmer er al., ICML 2017

Adaptive Self-training

Algorithm 1: MARS

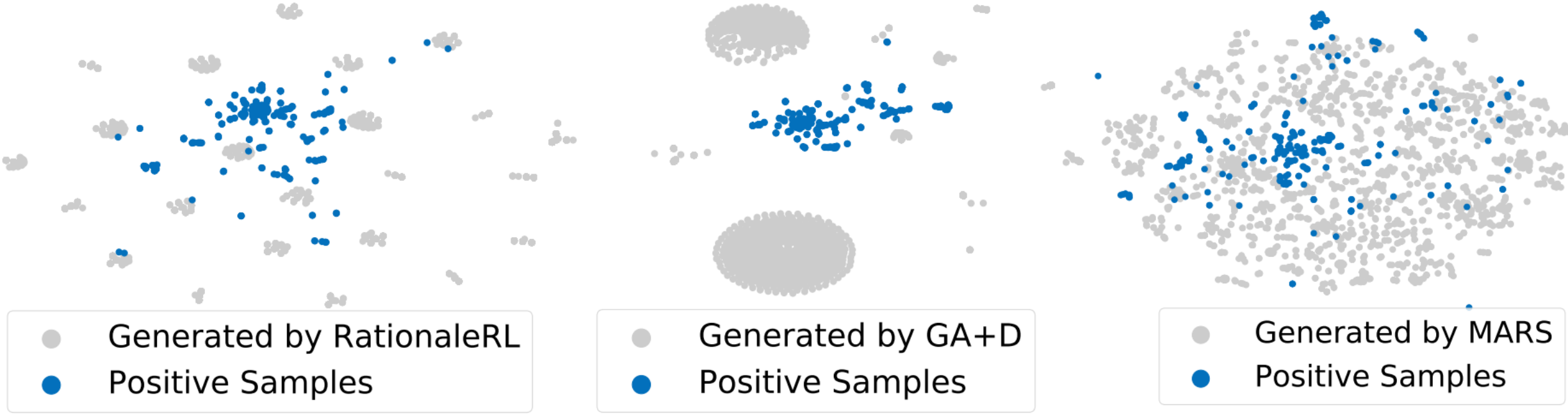
Set N initial molecules $\{x_i^{(0)}\}_{i=1}^{N}$ and initialize the molecular graph editing model $\mathcal{M}_\theta$
Create an empty editing model training dataset $\mathcal{D} = \{\}$
for $t = 1, 2, \ldots$ **do**
    for $i = 1, 2, \ldots, N$ **do**
        Compute probability distributions $(p_{\text{add}}, p_{\text{frag}}, p_{\text{del}}) = \mathcal{M}_\theta(x_i^{(t-1)})$ as Equations 7-9
        Sample a candidate molecule x' from the proposal distribution $q(x' \mid x_i^{(t-1)})$ defined with probability distributions $p_{\text{add}}, p_{\text{frag}}, p_{\text{del}}$ as Equations 3-4
        if $u < \mathcal{A}(x_i^{(t-1)}, x')$ *where* $u \sim \mathcal{U}_{[0,1]}$ **then**
            Accept the candidate molecule $x_i^{(t)} = x'$
        else
            Refuse the candidate molecule $x_i^{(t)} = x_i^{(t-1)}$
        if *The candidate improves the objectives, i.e.* $\pi(x') > \pi(x_i^{(t-1)})$ **then**
            Adding the editing record $(x_i^{(t-1)}, x')$ into the dataset $\mathcal{D}$
    $\theta^{new} \longleftarrow \arg\max \log M_\theta(\mathcal{D})$

MARS generates better molecules!



- **Success rate**: percentage of molecules satisfying all properties
- **Novelty**: compared with positive ones in the database
- **Diversity**: differences within generated molecules
- **Product of metrics**: SR × Novelty × Diversity

MARS explores larger chemical space!



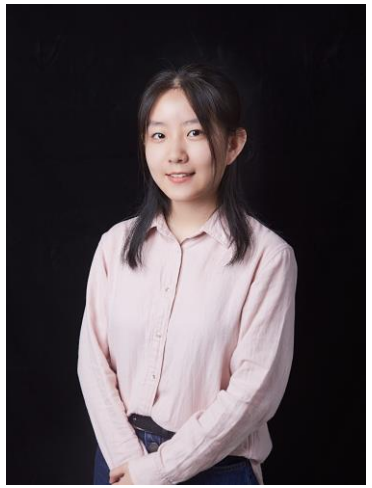
(a) RationaleRL (b) GA+D (c) MARS

Key Takeaways

- **Drug discovery and development is long and risky**
- **Challenges: multi-objective, novel and diverse, lack of data**

- **We propose MARS, a simple yet flexible framework**
  - Alternative to existing deep generative models
  - Based on MCMC sampling => multi-objective
  - Self-adaptive proposal trained on the fly => no need for data
  - Generates better molecules and explores lager chemical space! => can discover novel and diverse drug-like molecules

Thank you for listening!



Yutong Xie

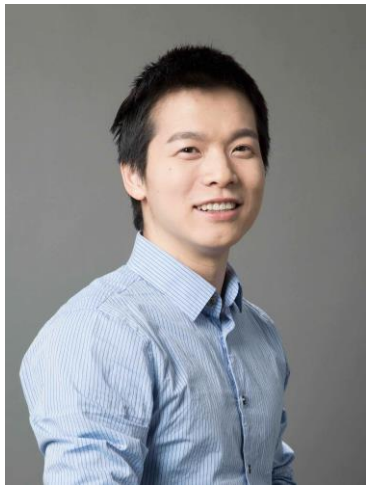
Chence Shi

Hao Zhou

Yuwei Yang

Weinan Zhang

Yong Yu



Lei Li