Diversity-Aware Reinforcement Learning for *de novo* Drug Design

Hampus Gummesson Svensson, Ind. PhD*†
Supervisors: Morteza Haghir Chehreghani*, Ola Engkvist†*, Christian Tyrchan‡ and Alexander Schliep*

Abstract

Fine-tuning a pre-trained generative model has demonstrated good performance in **generating promising drug molecules**. Nevertheless, without an adaptive update mechanism for the reward function, the optimization process can **become stuck in local optima**. The efficacy of the optimal molecule in a local optimization may **not translate to usefulness** in the subsequent drug optimization process or as a potential standalone clinical candidate. Therefore, it is important to generate a **diverse set** of promising molecules. Prior work has modified the reward function by **penalizing structurally similar molecules**, primarily focusing on finding molecules with **higher rewards**. In this work, we investigate a wide range of **intrinsic motivation** methods and strategies to **penalize the extrinsic reward**, and how they **affect the diversity** of the generated molecules.

Diversity-Aware Reinforcement Learning

We investigate two approaches to encourage diversity [1]:

- **penalty on** the extrinsic reward
- provide **intrinsic reward** (intrinsic motivation).

Given an extrinsic reward $R(A)$ for generated molecule A, the agent will receive a reward **at the end of the generation sequence** in the form of

$$\hat{R}(A) = f(A) \times R(A) + R_I(A).$$

We present a diversity-aware reinforcement learning framework to encourage diversity while keeping the quality high.

Diversity-Aware Reinforcement Learning Framework

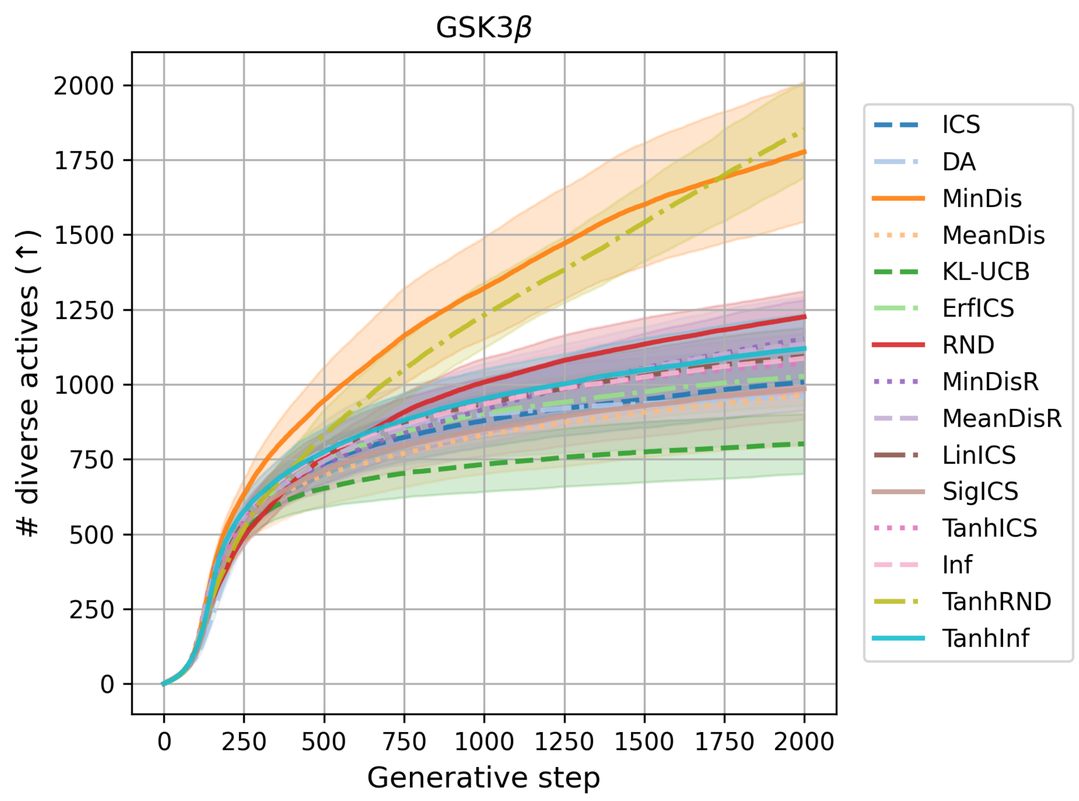
input: $I, B, \theta_{\text{prior}}, h$
$\mathcal{M} \leftarrow \emptyset$ ▷ Initialize memory
$\theta \leftarrow \theta_{\text{prior}}$ ▷ The pre-trained policy is fine-tuned
for i=1,...,I **do** ▷ Generative steps
  $L(\theta) \leftarrow 0$
  $\mathcal{B} \leftarrow \emptyset$
  for b=1,...,B **do** ▷ Generate batch of molecules
    $t \leftarrow 0$
    $a_t \leftarrow a^{(\text{start})}$ ▷ Start token is always initial action
    $s_{t+1} \leftarrow a_t$
    while s_{t+1} is not terminal **do**
      $t \leftarrow t + 1$
      $a_t \sim \pi_\theta(s_t)$ ▷ Sample next token in sequence
      $s_{t+1} \leftarrow a_{0:t}$ ▷ Subsequence defines next state
    end while
    $\mathcal{B} \leftarrow \mathcal{B} \cup s_{t+1}$ ▷ Final states represents molecule
    Observe property score $r(s_{t+1})$ ▷ Extrinsic reward
    if $r(s_{t+1}) \geq h$ **then** ▷ Memory of active molecules
      $\mathcal{M} \leftarrow \mathcal{M} \cup \{s_{t+1}\}$
    end if
    Compute and store penalty $f(s_{t+1})$
  end for
  for $A \in \mathcal{B}$ **do**
    Compute intrinsic reward $R_I(A)$ ▷ Exploration bonus
    Compute diversity-aware reward $\hat{R}(A)$
    Compute loss $L_A(\theta)$ wrt $\hat{R}(A)$
    $L(\theta) \leftarrow L(\theta) + L_A(\theta)$ ▷ Accumulate loss
  end for
  Update θ by one gradient step minimizing $L(\theta)$
end for
output: $\mathcal{M}$ ▷ Memory of active molecules

Selected Results

Here we show experiments on one extrinsic reward function, namely activity on the **Glycogen Synthase Kinase 3 Beta** (GSK3β) protein. It is a well-established classification task to optimize the activity against the GSK3b protein. We display the number of diverse actives per generative step. Given a set $\mathcal{H}$ of molecules such that $\forall A \in \mathcal{H}, R(A) \geq h$, the number of diverse actives is defined by

$$\mu(\mathcal{H}; D) = \max_{\mathcal{C} \in \mathcal{P}(\mathcal{H})} |\mathcal{C}| \text{ s.t. } \forall x \neq y \in \mathcal{C} : d(x, y) \geq D,$$

where $\mathcal{P}$ is the power set and $d(x, y)$ is a distance metric.



Conclusions [1]

- By integrating both structure-based and prediction-based methods, we facilitate a more explorative and comprehensive search of the chemical space.
- The combination of random network distillation (RND) with a tanh-based penalty (TanhICS) yields the most substantial improvements in molecular diversity.

*Department of Computer Science and Engineering, Chalmers University of Technology and University of Gothenburg, Sweden
†Molecular AI, Discovery Sciences, R&D, AstraZeneca, Gothenburg, Sweden
‡Medicinal Chemistry, Research and Early Development, Respiratory and Immunology (R&I), BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

References

[1] Hampus Gummesson Svensson et al. *Diversity-Aware Reinforcement Learning for de novo Drug Design*. 2024. arXiv: 2410.10431 [cs.LG].

WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM