

# Introduction of KIT & DC14

Hyun Kil Shin

Department of predictive toxicology

Korea Institute of Toxicology





# CONTENTS

- I Introduction
- II Molecule design with generative AI



# CHAPTER I

*QSAR & Generative AI*

## Introduction



# KIT & UST

- **Korea Institute of Toxicology (KIT)**
  - Research institute under Ministry of Science
  - Internal funding program: 3 + 3 + 3 (1,700,000 EUR / year)
  - External funding program
  - Students make employment contract, which includes insurance, paid vacation, and monthly income.
  - Dormitory is available if there is empty room.
- **UST (University of Science and Technology)**
  - KIT is one of campus in UST.
  - Students apply to UST specifying campus in UST.
  - UST has its own dormitory.



# PI

## • Education

- Bachelor: Biology (1<sup>st</sup>), Computer science (2<sup>nd</sup>)
- Master & Doctor: Cheminformatics
- Major experiences: ADMET prediction (machine learning)

## • Research Experiences

- Korea Institute of Toxicology: in silico model development
- University of Science & Technology (UST): AI toxicity prediction (lecture)
- AiFrenz (board member): Advanced machine learning

## • Global Collaboration (Horizon 2020, Horizon Europe)

- AI drug design: AiChemist
- Nanotoxicity: Gov4Nano & SUNSHINE



# ToxSTAR webpage

<https://www.kitox.re.kr/toxstar>



## NOTICE


 27  
2021.12

### Prediction models were added NEW

Drug-induced cholestasis, cirrhosis, hepatitis, and steatosis models were implemented in ToxSTAR. You can u...

Periodic Table and Drawing Structure Information

2019-09-07

This is a notice.

2019-09-06

## What is ToxSTAR?

ToxSTAR is a platform to predict human toxicity caused by drugs and chemicals through integration of biotechnology (BT) and information technology (IT). Currently, we are working on Drug-Induced Liver Injury (DILI) prediction with alternative testing methods to support 3Rs.

[Learn More](#)



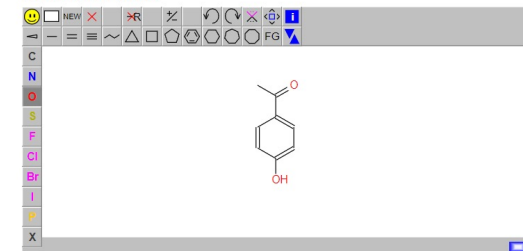
# ToxSTAR

## PREDICTION

### How to use

1. The models implemented in this page were developed and published. You can find details of the models here (<https://pubmed.ncbi.nlm.nih.gov/32116729/>)
2. Summary of the models are described [here](#).
3. Input molecule preparation
  - a. You can draw it in the Molecule Editor (drawing tutorial: <https://youtu.be/bJ4wjuPcVlg>)
  - b. You can find Smiles Code of your molecule from PubChem (<https://pubchem.ncbi.nlm.nih.gov/>)
4. Select molecule type as 'Drug' if the input molecule is a drug molecule or 'Drug metabolite' if it is a drug metabolite. This option will select prediction models to use.
5. Click submit button. When prediction outcomes are ready, this page will automatically show the result.

### Molecule Editor



### Smiles Code

[Example](#)

### Molecule Type

Drug  Drug metabolite

Submit

B. Bienfait and P. Ertl, [JSME: a free molecule editor in JavaScript](#), J. Cheminformatics 5:24 (2013)



# ToxBBB

<https://www.kitox.re.kr/toxbbb>

## Prediction

### Prediction



Organic Matter  
**유기물**

예측하기

Mineral  
**무기물**

예측하기

투과율

분자구조

logBB

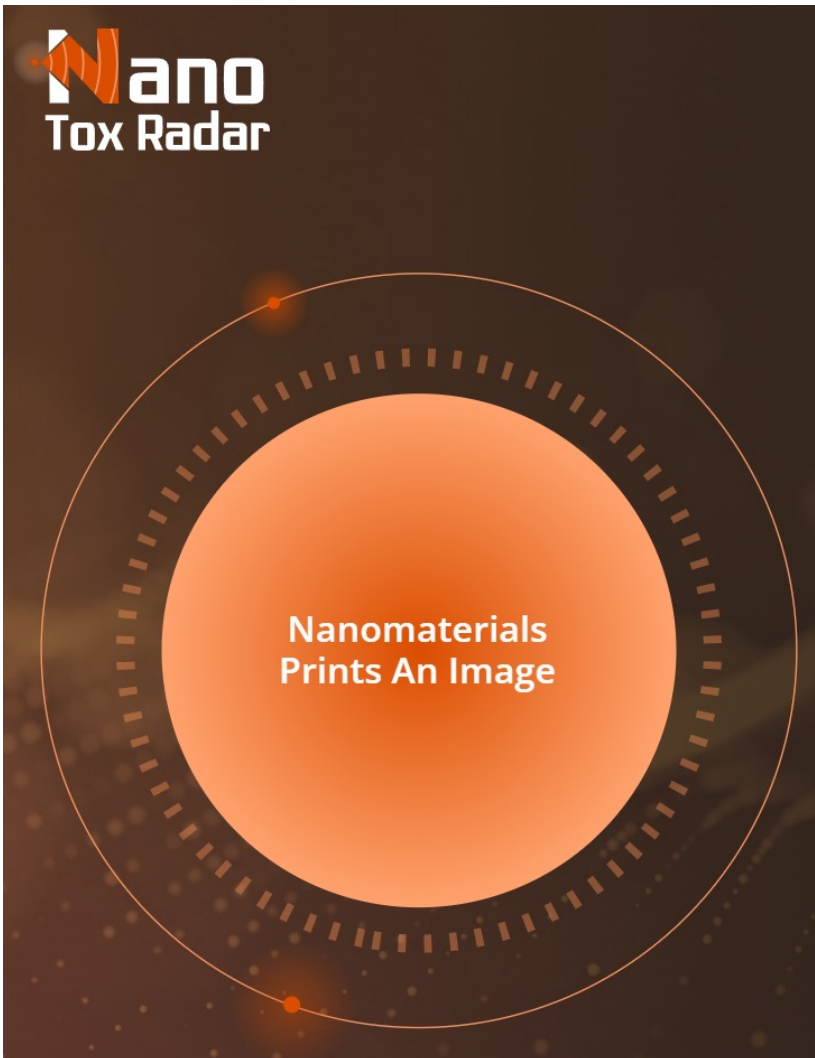
$$\log BB = \log \left( \frac{[\text{Brain : 뇌에서의 물질 양}]}{[\text{Blood : 혈액에서의 물질 양}]} \right)$$

logBB 예측 결과

• logBB 점수 : 0.48

# NanoTox Radar

<https://www.kitox.re.kr/nanotoxradar>



The screenshot shows the NanoTox Radar interface. On the left, a 3D model of a spherical nanomaterial is shown with a diameter of 50nm. The model is composed of three concentric layers: an innermost red core labeled "Fe (5%)", a middle orange layer labeled "Cr2O3", and an outermost white layer labeled "Carboxyl (lipoic acid)". On the right, the "Nanotoxicity Prediction Program" configuration panel is visible. It includes a "Reset" button, a "Shape" dropdown menu with "Sphere" selected, a "Core" dropdown menu with "Cr2O3" selected, a "Doping" dropdown menu with "Fe" selected, a "Coating" dropdown menu with "Carboxyl (lipoic acid)" selected, and a "Diameter" input field set to "50 nm". A "Completion" button is located at the bottom of the configuration panel.

The screenshot shows the "Prediction Output" section of the NanoTox Radar interface. It features a semi-circular gauge with three segments: a blue segment labeled "Safe" (0 to 0.3), a grey segment labeled "Uncertain" (0.3 to 0.7), and a red segment labeled "Warning" (0.7 to 1). The gauge is labeled "Cytotoxicity probability (0-1)". To the right of the gauge, the following data is displayed: "Cytotoxicity Probability(A549 cell death > 20% at 100 µg/mL) : 0.1952", "Cytotoxicity benchmark dose 5% (A549, BMD5, µg/mL) : 5.4572", and "Zeta Potential (Ev) : -8.0367". At the bottom, there are three buttons: "Reset", a download icon, and a print icon.





THE  
2024

**LUSH  
PRIZE**



**SUPPORTING  
ANIMAL-FREE  
TESTING**

### **Shortlist: Young Researcher Prize**

**Jonathan Blum**, University of Konstanz, Germany

**Lauren Coelho**, Universidade Federal de Goias, Brazil

**Martina Iulini**, Università degli Studi di Milano, Italy

**Dr Barbara Jozef**, Eawag, ETH domain, Switzerland

**Byunggik Kim**, Johns Hopkins University, USA

**Dr Molly Kozminsky**, Iowa State University, USA

**Dr Chao Ma**, New York University, USA

**Dr Kirsty Meldrum**, Swansea University, UK

**Aina Kehinde Oluwasegun**, Friedrich Schiller University, Germany

**Dr Nuria Roldan**, PETA Science Consortium International e.V., Germany

**Dr Maren Schenke**, Johns Hopkins University, USA

**Dr Hyun Kil Shin**, Korea Institute of Toxicology, South Korea

**Dr Nicoleta Spinu**, esqLABS GmbH, Netherlands

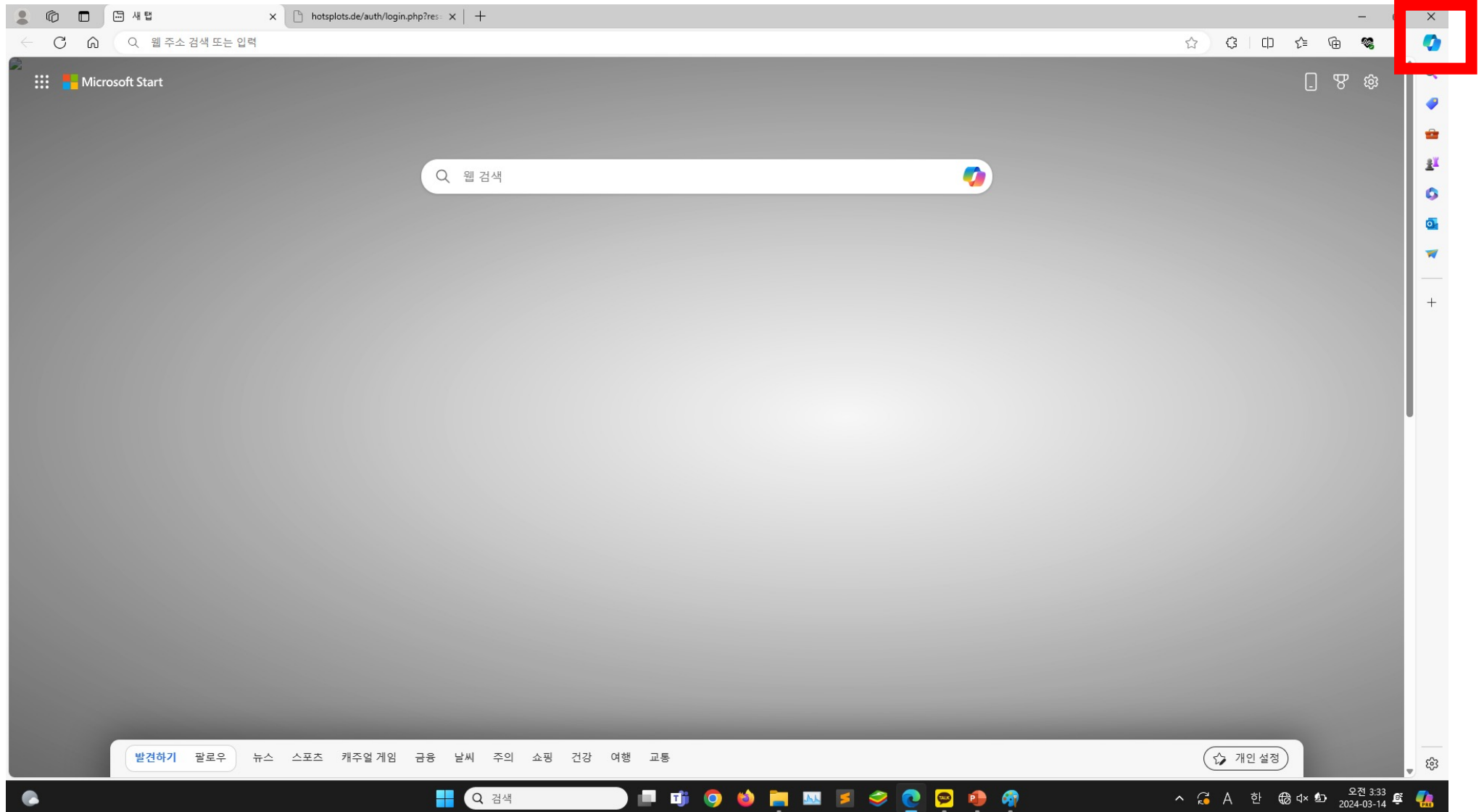
**Zheng Tan**, University of British Columbia, Canada

**Sam Youmans**, XCellR8, UK

[lushprize.org/2024-prize/2024-prize-shortlist](https://lushprize.org/2024-prize/2024-prize-shortlist)

# ChatGPT

- President of South Korea wrote his speech using ChatGPT, and strongly encourages to use LLM.





# Copilot

Cornell University

We gratefully acknowledge support from the Simons Foundation, member institutions, and all contributors. [Donate](#)

arXiv > cs > arXiv:1711.07478

Search... All fields Search

Help | Advanced Search

Computer Science > Machine Learning

[Submitted on 20 Nov 2017]

## Implementing the Deep Q-Network

Melrose Roderick, James MacGlashan, Stefanie Tellex

The Deep Q-Network proposed by Mnih et al. [2015] has become a benchmark and building point for much deep reinforcement learning research. However, replicating results for complex systems is often challenging since original scientific publications are not always able to describe in detail every important parameter setting and software engineering solution. In this paper, we present results from our work reproducing the results of the DQN paper. We highlight key areas in the implementation that were not covered in great detail in the original paper to make it easier for researchers to replicate these results, including termination conditions and gradient descent algorithms. Finally, we discuss methods for improving the computational performance and provide our own implementation that is designed to work with a range of domains, and not just the original Arcade Learning Environment [Bellemare et al., 2013].

Subjects: **Machine Learning (cs.LG)**; Artificial Intelligence (cs.AI)

Cite as: arXiv:1711.07478 [cs.LG] (or arXiv:1711.07478v1 [cs.LG] for this version) <https://doi.org/10.48550/arXiv.1711.07478>

**Submission history**

From: Melrose Roderick [\[view email\]](#)

[v1] Mon, 20 Nov 2017 16:40:33 UTC (276 KB)

**Download:**

- PDF
- Other formats (license)

Current browse context: **cs.LG**

< prev | next > new | recent | 1711

Change to browse by: cs, cs.AI

**References & Citations**

- NASA ADS
- Google Scholar
- Semantic Scholar

**DBLP - CS Bibliography**

[listing](#) | [bibtex](#)

Melrose Roderick  
James MacGlashan  
Stefanie Tellex

**Export BibTeX Citation**

**Bookmark**

Bibliographic Tools Code, Data, Media Demos Related Papers About arXivLabs

### Bibliographic and Citation Tools

Bibliographic Explorer (What is the Explorer?)

Litmaps (What is Litmaps?)

scite Smart Citations (What are Smart Citations?)

Which authors of this paper are endorsers? | [Disable MathJax](#) (What is MathJax?)

채팅 작성 미리 파악

돌아온 것을 환영합니다! 어떤 내용에 대해 채팅하려고 하나요?

**Deep-Q-Network** 을 처음 발표한 논문 알려줘

- ✓ "Deep-Q-Network 을 처음 발표한 논문" 검색 중
- ✓ "Deep-Q-Network first paper" 검색 중
- ✓ 답변을 생성하는 중...

Deep Q-Network (DQN)은 Mnih et al. [2015]에 의해 제안된 인공지능 기법 중 하나로, 강화학습의 발전된 형태입니다<sup>1</sup>. 이 논문은 많은 딥 강화학습 연구의 벤치마크 및 구축 지점이 되었습니다<sup>1</sup>. 이 논문의 주요 기여 중 하나는 Q-함수 근사를 위한 깊은 컨볼루션 신경망 아키텍처, 단일 단계 업데이트가 아닌 무작위 훈련 데이터의 미니 배치 사용, 다음 상태의 Q-값을 추정하기 위해 이전 네트워크 매개 변수 사용 등이 있습니다<sup>2</sup>. 이 논문의 정확한 제목은 "Human-level control through deep reinforcement learning"입니다. 이 논문을 참조하시려면<sup>1</sup>을 참조하십시오. 다른 질문이 있으신가요? 😊

자세한 정보: 1 / 30

1. arxiv.org 2. arxiv.org 3. scienceon.kisti.re.kr

4. scienceon.kisti.re.kr 5. scienceon.kisti.re.kr +2 더 보기

? DQN이 어떻게 작동하는지 설명해줘.

다른 강화학습 알고리즘이 있나요?

Mnih et al. [2015]의 다른 논문을 추천해주세요

무엇이든 물어보세요

<https://typeset.io/>

# Do hours worth of reading in minutes

Upload PDF & get answers backed by

Results From:  SciSpace papers  My library

🔍 Get insights from top papers directly

Try asking or searching for:

- 🔍 How does climate change impact biodiversity?
- 🔍 Why are aging Covid patients more susceptible to severe complications?
- 🔍 How does social media affect the college selection process?
- 🔍 What are the interesting theories about dark matter and dark energy?
- 🔍 What is the significance of higher-dimensional algebra?



Large language models have been shown to be beneficial for chemists in various ways. They can help in understanding the language of chemistry, such as molecular models and structural formulas, which are essential for meaningful learning of the subject <sup>[1]</sup>. Additionally, these models can be parameterized through in-context examples, allowing them to perform novel tasks and follow instructions with just a few examples <sup>[2]</sup> <sup>[3]</sup>. Furthermore, the use of large language models can aid in estimating language models using large datasets by breaking down the estimation process into sub-processes and parallelizing computation <sup>[4]</sup>. Overall, these models have the potential to enhance the capabilities of chemists and improve their understanding and performance in the field.

Was this helpful?



🗨️ Ask a follow up

<span>Add columns (2) ▼</span> <input type="checkbox"/> PDF <input type="checkbox"/> Open Access <a href="#">More filters</a> <span>Sort by: Relevance ▼</span> <span>📄 CSV</span> <span>⋮</span>	
Papers (10)	Insights
<p> Open access • Journal Article • DOI <a href="#">🔗</a></p> <p><input type="checkbox"/> <a href="#">A linguagem química e o ensino da química orgânica</a></p> <p> Nidia F. Roque +1 more</p> <p>01 Jan 2008 • Química Nova</p> <p>24 Citations  PDF    Ask Copilot</p>	<p>The paper discusses the importance of chemical language and molecular models in the development of chemistry.</p>
<p> Open access • Posted Content • DOI <a href="#">🔗</a></p> <p><input type="checkbox"/> <a href="#">Large Language Model Programs</a></p> <p> 09 May 2023</p> <p> PDF    Ask Copilot</p>	<p>The paper discusses the benefits of embedding large pre-trained language models within algorithms or programs, but does not specifically mention their use in chemistry.</p>

# Large Language Model Programs

09 May 2023 -

**TL;DR:** In this paper, the authors extend this line of reasoning and present a method which further expands the capabilities of an LLM by embedding it within an algorithm or program, and obtain a 6.4% improvement over the chain-of-thought baseline.

[View 8 related papers](#)

**Abstract:** In recent years, large pre-trained language models (LLMs) have demonstrated the ability to follow instructions and perform novel tasks from a few examples. The possibility to parameterise an LLM through such in-context examples widens their capability at a much lower cost than finetuning. We extend this line of reasoning and present a method which further expands the capabilities of an LLM by embedding it within an algorithm or program. To demonstrate the benefits of this approach, we present an illustrative example of evic ...[read more](#)

**Ask Copilot: Save time, read 10X faster with AI**

Explain Abstract of this paper in 2 lines

Conclusions from the paper

Results of the paper

Methods used in this paper

Summarise introduction of this paper

What are the contributions of this paper

[Show more questions](#)

**Related Papers (5)**

Ask Copilot

8

Related papers

6

Related topics

Copilot

English (en)

Highlight text to get explanation

or

Methods used

Summarise introduction

Literature survey

Practical Implications

Future works

Dataset used

Contributions

Explain Abstract of this paper in 2 lines, Cc

+12 more

Write your question...

NEW





# CHAPTER

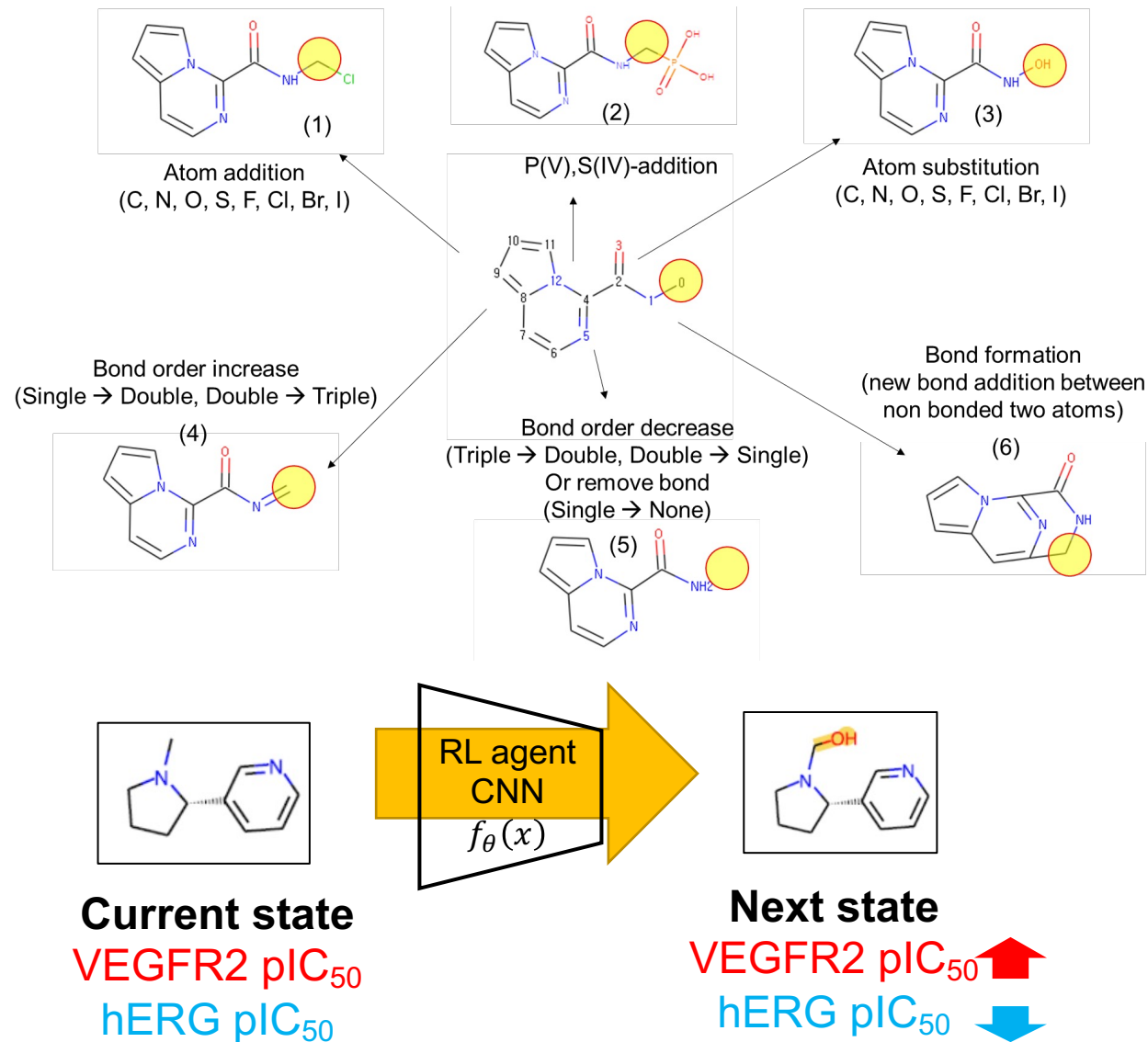
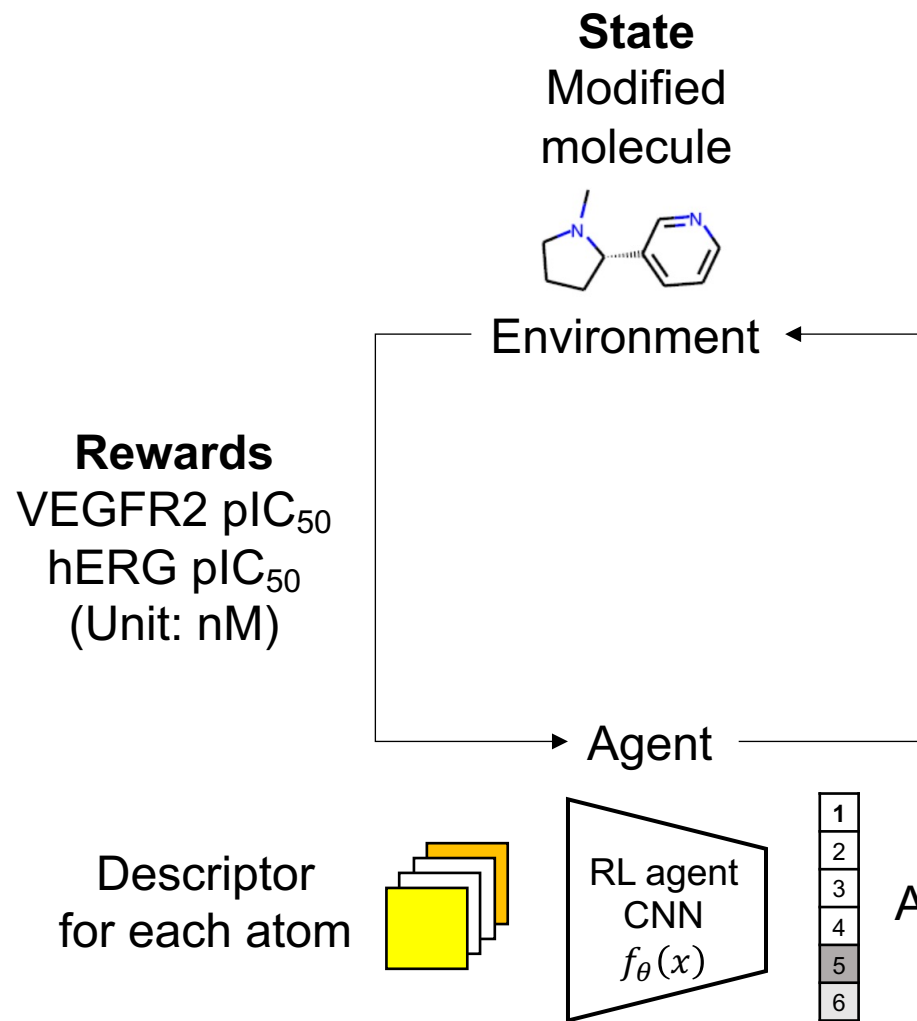


*QSAR & Generative AI*

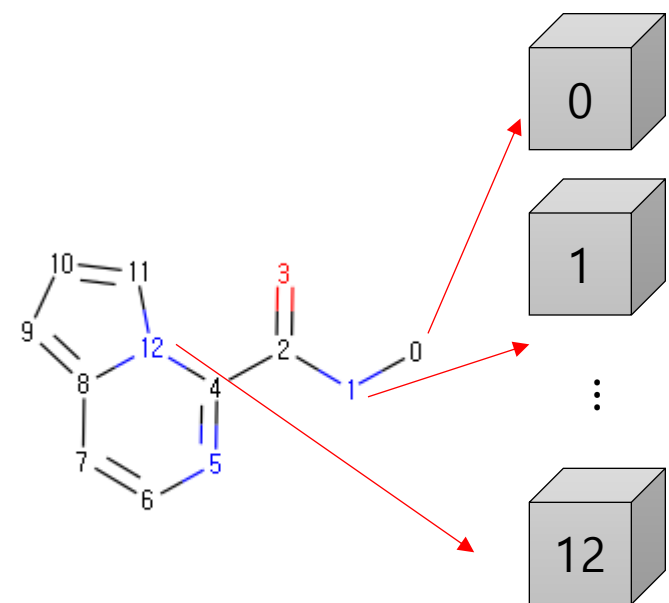
# Molecule design with Generative AI



# DRL for de novo drug design



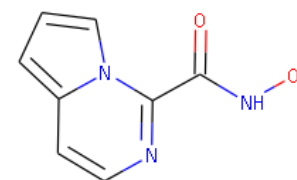
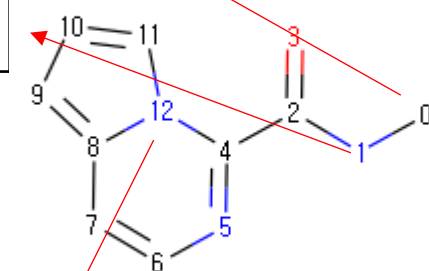
# Molecule structure modification



RL model  
CNN  
 $f_{\theta}(x)$



	1	2	3	4	5	6
0	0.1	-0.3	<b>9.9</b>	1.2	1.0	4.0
1	0.2	-1.3	8.4	2.9	0.1	4.9
			⋮			
12	1.1	-3.0	2.1	1.2	-1.0	2.0

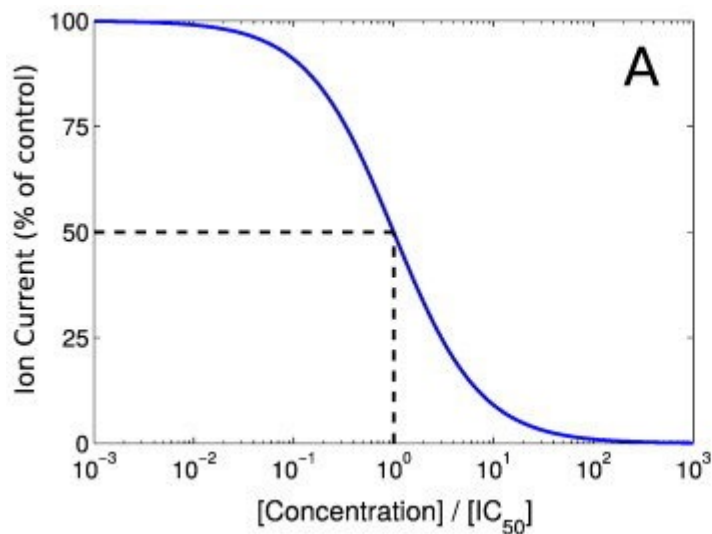


Apply action 3 on atom 0



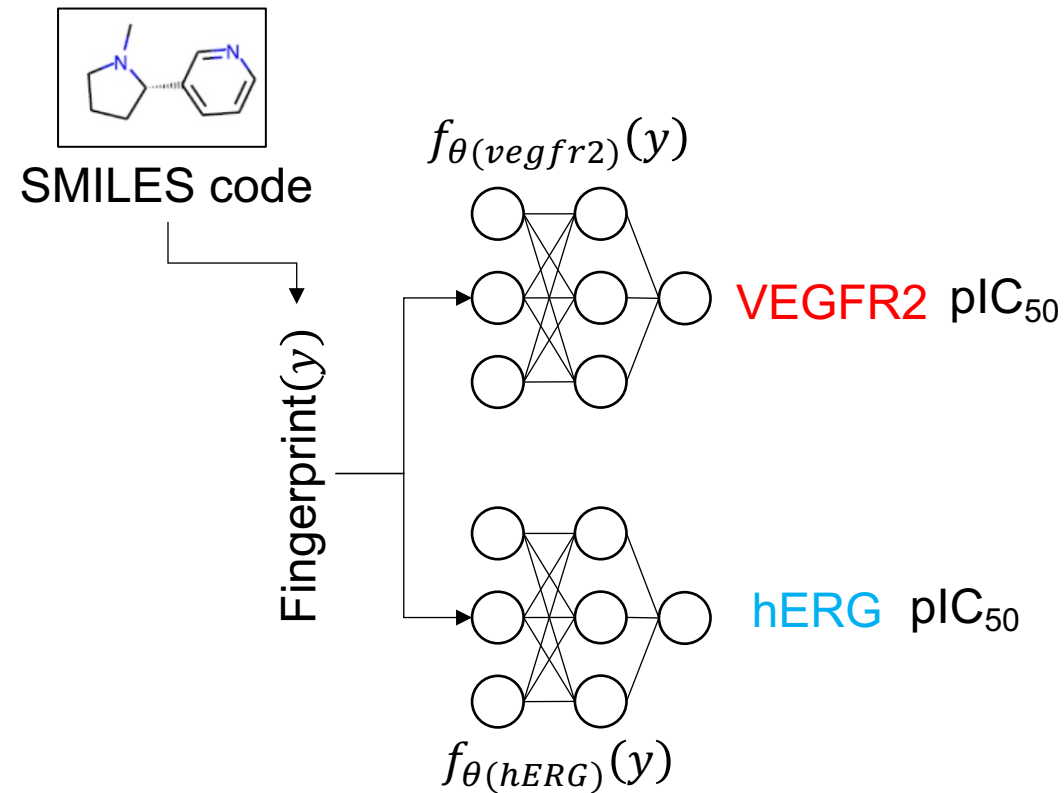
# Reward

- $IC_{50}$

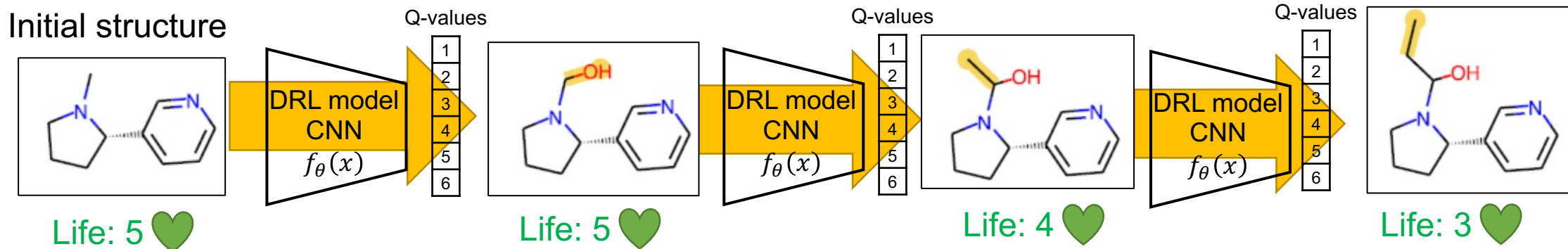
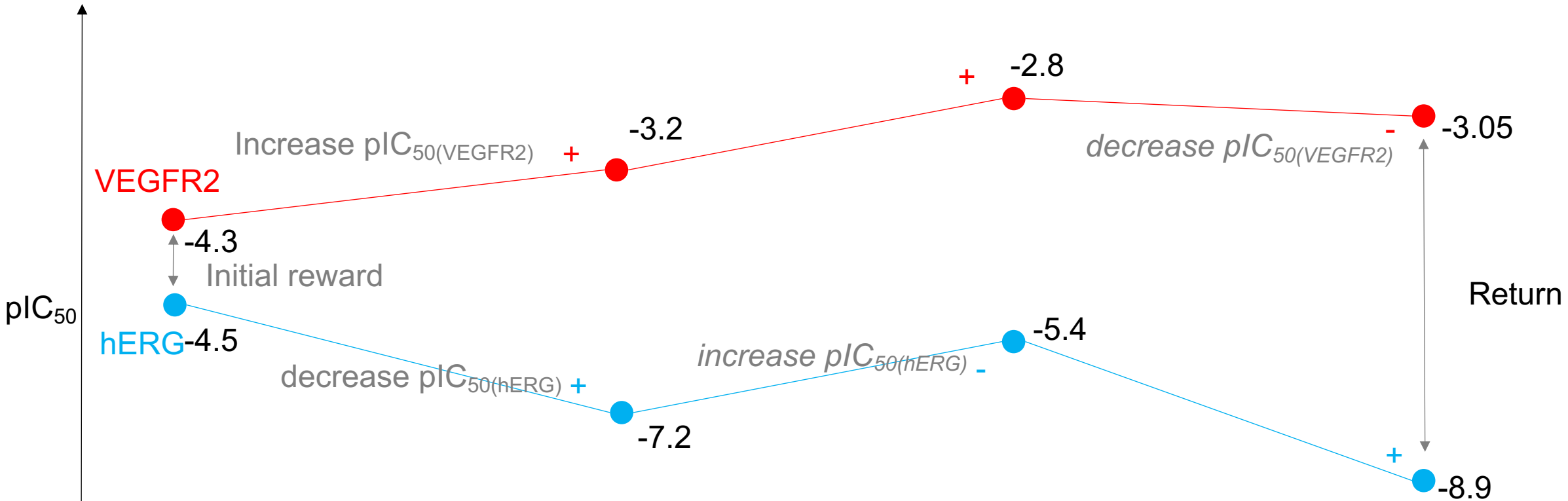


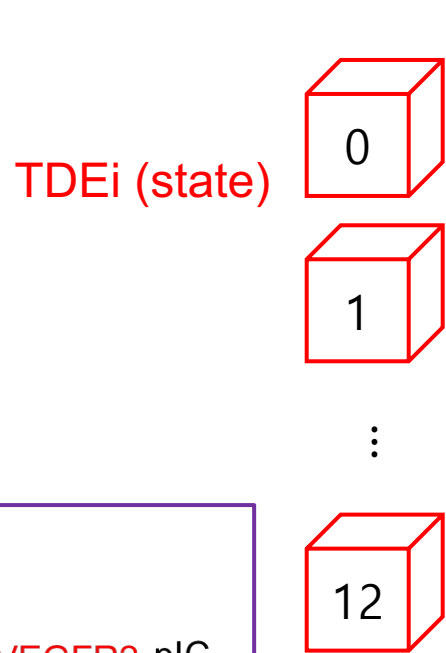
- $pIC_{50}$  ( $-\log_{10} IC_{50}$ )

- $VEGFR2(pIC_{50}) - hERG(pIC_{50})$  - penalty

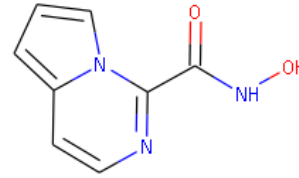
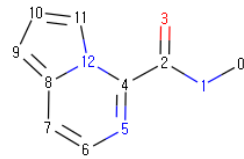


# Efficacy & toxicity both matter



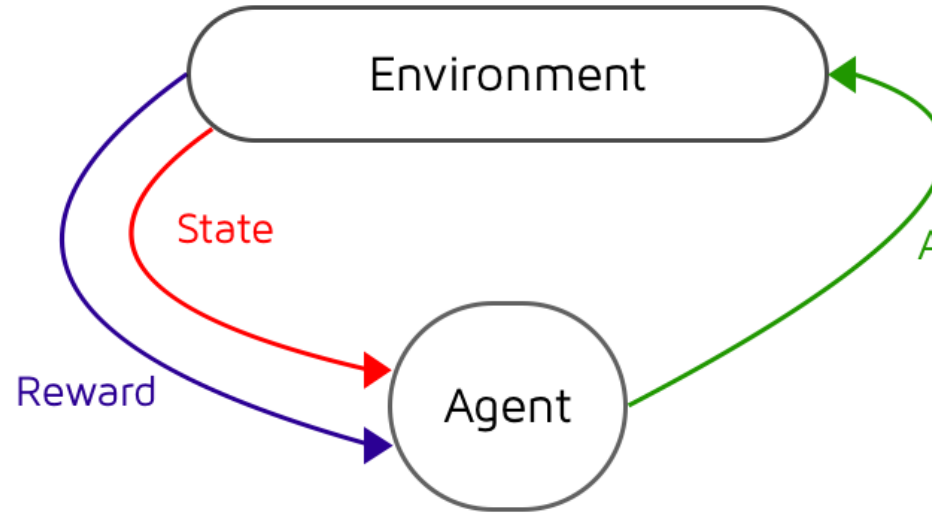


Initial structure

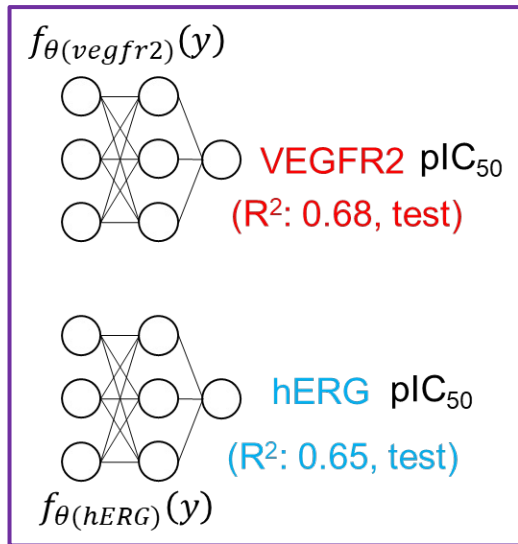


Take action

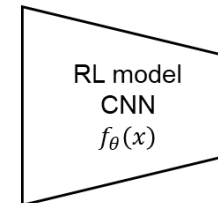
- 1) Apply action
- 2) Update pIC<sub>50</sub> (VEGFR2 & hERG)
- 3) TDEi calculation
- 4) Life update



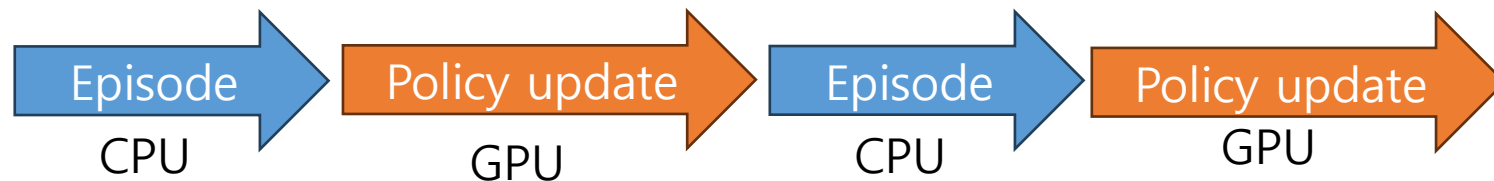
Action (atom 0, action 3)



- When episode comes to an end
- Calculate loss (11 algorithms)
    1. Cross entropy
    2. Value optimization
    3. Policy optimization
    4. MCTS



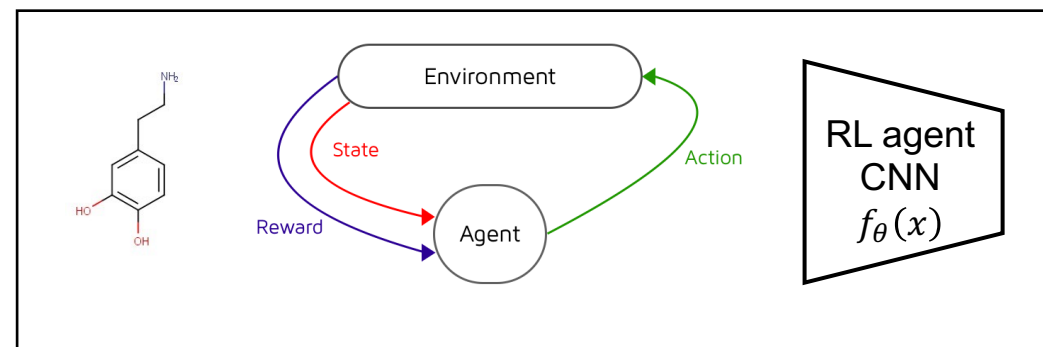
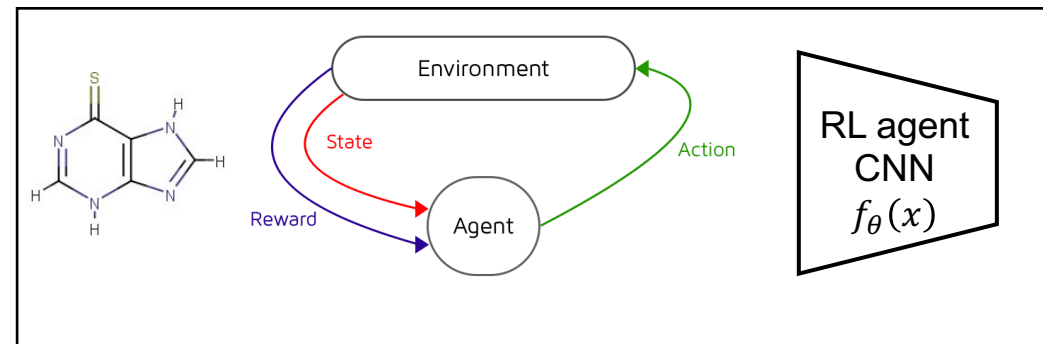
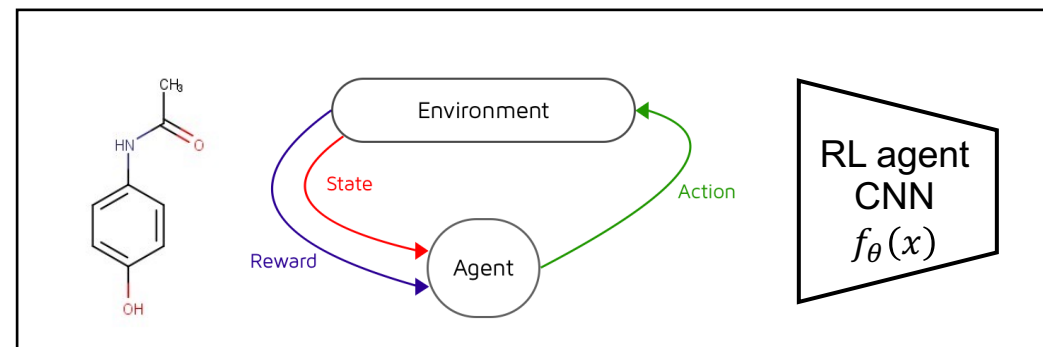
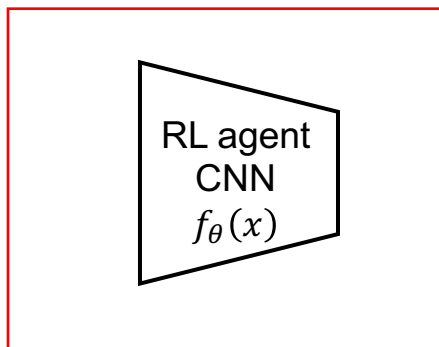
	1	2	3	4	5	6
0	0.1	-0.3	<b>9.9</b>	1.2	1.0	4.0
1	0.2	-1.3	8.4	2.9	0.1	4.9
12	1.1	-3.0	2.1	1.2	-1.0	2.0





# Asynchronous DRL

Parallelize molecule generation

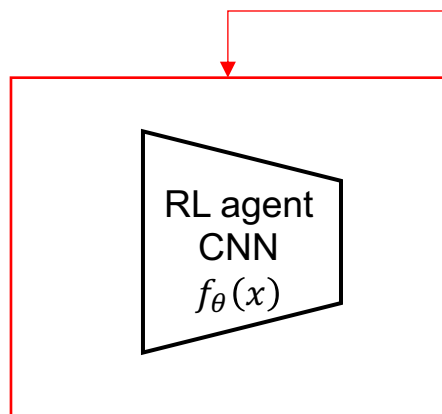


RL agents running episode in each node do not have identical weights.

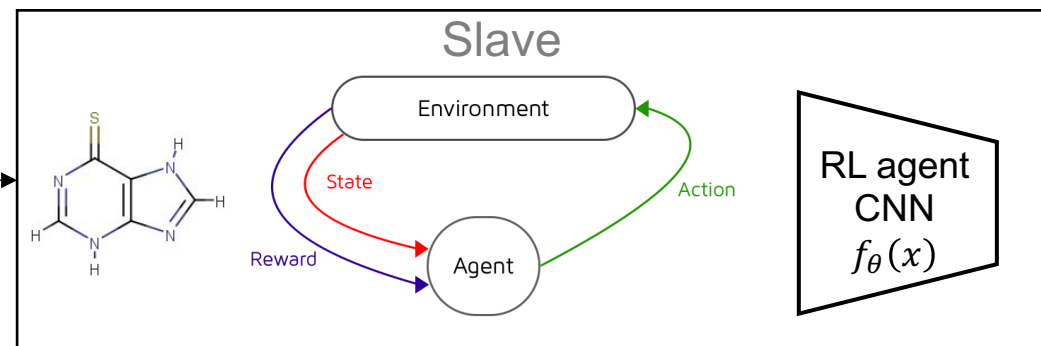
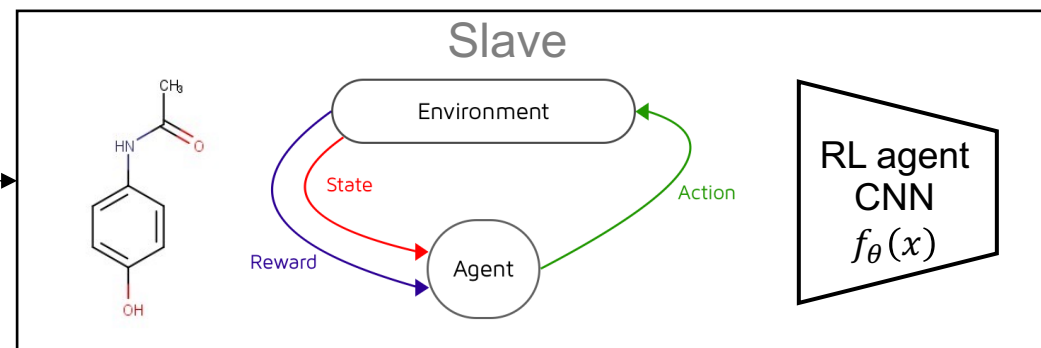
# Asynchronous DRL: Master-slave

Queue: First-In-First-Out

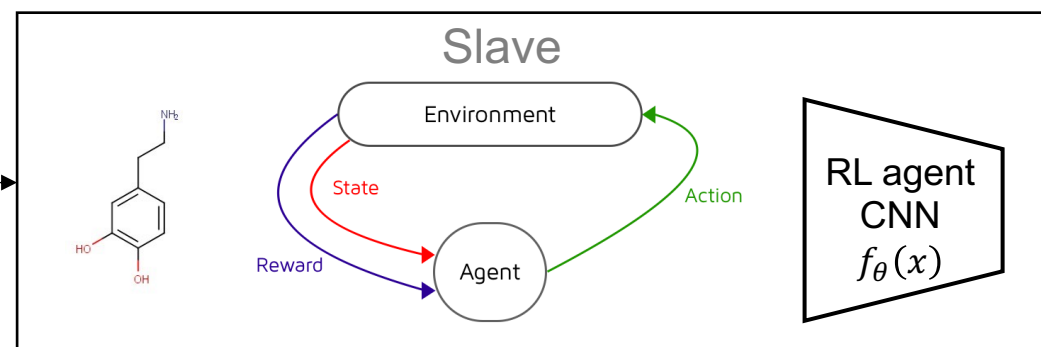
Library: mpi4py, mpi\_master\_slave



Master



⋮



# DC14

<https://aichemist.eu/>

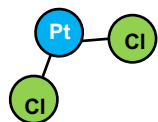
## Rule of five

1. Hydrogen bond donors < 5
2. Hydrogen bond acceptors < 10
3. Molecular weight < 500
4. ClogP < 5

## Beyond the rule of five

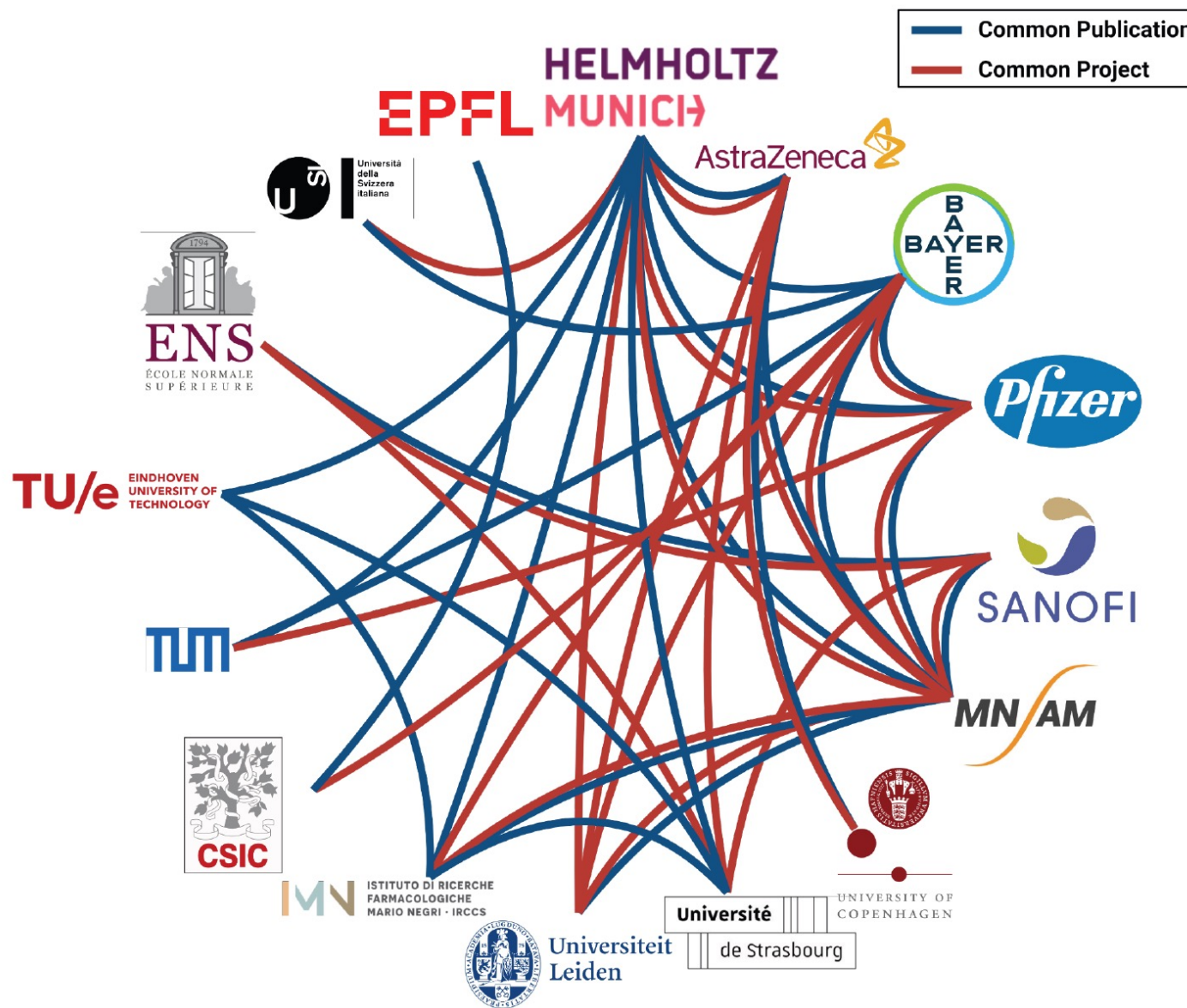


Natural products



Inorganic pharmaceuticals

Covalent inhibitor



**Looking for a PhD student for the project  
(or master-phD program candidate)**



# Acknowledgements



## I Funding

KIT (1711159817)

NRF(MSIT)-2019R1F1A1061955

NRF(MSIT)-2015M3A7B6027948

KEITI(MOE)-2019002490006

KIAT(MOTIE)-P0019147



National Research  
Foundation of Korea



Ministry of Science, ICT  
and Future Planning



Ministry of Environment



한국산업기술진흥원  
Korea Institute for Advancement of Technology



Ministry of Trade,  
Industry and Energy

## II KIT members

Jung-Hwa Oh (Ph.D.), Hyoung-Yun Han (Ph.D.), Se-Myo Park,  
Soojin Kim, Woo-Keun Kim (Ph.D.), Hang-Suk Chun (Ph.D.),  
Sangwoo Lee (Ph.D.), Myung-Gyun Kang (Ph.D.), Daeui Park (Ph.D.),  
Sungbo Hwang (Ph.D.), Seokjoo Yoon (Ph.D.)

## Others

Seungsang Oh (Prof.) Korea University

Oh-Kyoung Kwon (Ph.D.) Korea Supercomputing Center (KISTI)



안전성평가연구소  
Korea Institute of Toxicology



# Thank you

Q&A