**Stage M2 EI2D**

**Titre du stage**: Generative Reinforcement Learning for *de novo* Drug Design

**Institution**: LIPN – Université Sorbonne Paris Nord

**Information générales**:

**Encardant**: Thomas Papastergiou, Céline Rouveirol

**Contacts**: papastergiou@lipn.univ)paris13.fr; rouveirol@lipn.univ-paris13.fr

**Localisation**: LIPN

**Durée**: 6- mois – stage rénumeré

**Mots clés**: Machine Learning; Deep Reinforcement Larning; De novo drug design

# Introduction

The discovery of penicillin in 1928 revolutionized the treatment of bacterial infections, but the overuse of antibiotics has led to rising resistance, causing over 33,000 deaths annually in Europe [1]. Gram-negative bacteria, especially those producing β-lactamases such as NDM-1, can deactivate even last-resort antibiotics, posing a major global threat [2]. Developing effective inhibitors is crucial but costly and time-consuming, so in silico approaches with AI and AI generative models are increasingly explored to accelerate drug discovery [3].

## Generative Models using Reinforcement Learning

Different approaches leveraging generative models have been used to conditionally generate molecules (i.e. generate molecules with desired properties): e.g. in [4] a conditional variational autoencoder has been employed for generating molecules with specific physiochemical properties, in [5] entangled conditional adversarial autoencoders have been used, in [6] a conditional recursive neural network (cRNN) has been used for generating molecules with desired physico-chemical properties as well as another model targeting biological activities. Another approach is to use a generative model within a Reinforcement Learning loop for guiding the generation process through the chemical space for acquiring desired properties as in ReLeaSE [7]. In this work, a Stack-RNN employing Gated Recurrent Units (GRU) is used as a generative model, a Long Short Term Memory (LSTM) model is used as a predictive model for predicting the physico-chemical or biological properties that will used to shape the reward function. The Generative model is treaded like a policy approximation model and thus the policy gradient are computed using the REINFORCE algorithm [8] for maximizing the policy towards obtaining higher rewards.
Although ReLeaSE can be very effective for generating general purpose drug-like molecules with desired properties, ReLeaSE showed when adapted to the generation of NDM-1 inhibitors pure efficacy in generating strong active NDM-1 inhibitors. Furthermore, ReLeaSE

showed low efficacy in producing valid molecules (after 100 epochs of reinforcement learning, the percentage of valid molecules was about 20%).

## Objectives and Challenges

The overall objective of this intership is the design of an efficient generative model for generating compounds potentially active against the NDM-1 enzyme using an Reinforcement Learning (LR) approach. An efficient generative model needs i) to produce high rates of both valid molecules and active molecules against NDM-1 ii) to generate **novel** molecules (i.e. the generated molecules should not belong to the training datasets) and finally iii) to show **high diversity** in the generated molecules (i.e. the generative models should not to produce similar molecules).

**Work description Task1:** We will design, implement and evaluate a reinforcement learning generative model for the generation of NDM inhibitors based on the **Cross Entropy Method (CME** in the following), a sophisticated **RRN-like deep network as policy** and a simple reward function based on an **activity classifier** and a **molecule validity parser**. In the CME the *elites*, i.e. the generated molecules with highest rewards, are used for the evolution of the policy. The novelty in our approach will rely on the fact that we will use clustering techniques for selecting at each epoch high diversity elites' representatives for the evolution of the policy, aiming at acquiring a generation model that will generate molecules having high diversity. This will give the generative model the ability to escape from suboptimal local minima of the policy function.

**Task 2:** In this Task, we will design a **sophisticated reward function** based on the knowledge acquired by exploiting an elaborate analysis of the activity classification model via Explainable AI (XAI) methodologies. The rationale behind this is that the analysis will provide us with insights on why a molecule has a certain biological activity. These insights will then be exploited to shape reward functions that could guide the generation towards molecules with expected properties.

**Each of the tasks will be carried out by a different intern.**

## Student's profile

We search for two (2) highly motivated M2 in Informatics intern students, with good competencies in supervised and unsupervised Machine Learning methods and good programming skills in Python. **No background in Chemoinfomatics is required**

## For applying

The applications for the internship can be done via the email: thomas.papastergiou@univ-paris13.fr where the following documents are asked:

1. A CV
2. a letter of motivation
3. the grades of the M1 and M2 years

4. contacts (position, mail, telephone) of one or two persons that could recommend you.

**References**


[1] *Global Antimicrobial Resistance and Use Surveillance System (GLASS) Report 2021*, 1st ed. Geneva: World Health Organization, 2021.

[2] C. Shi, J. Chen, X. Kang, X. Shen, X. Lao, and H. Zheng, "Approaches for the discovery of metallo-β-lactamase inhibitors: A review," *Chem. Biol. Drug Des.*, p. cbdd.13526, May 2019, doi: 10.1111/cbdd.13526.

[3] V. D. Mouchlis *et al.*, "Advances in De Novo Drug Design: From Conventional to Machine Learning Methods," *Int. J. Mol. Sci.*, vol. 22, no. 4, p. 1676, Feb. 2021, doi: 10.3390/ijms22041676.

[4] J. Lim, S. Ryu, J. W. Kim, and W. Y. Kim, "Molecular generative model based on conditional variational autoencoder for de novo molecular design," *J. Cheminformatics*, vol. 10, no. 1, p. 31, Dec. 2018, doi: 10.1186/s13321-018-0286-7.

[5] D. Polykovskiy *et al.*, "Entangled Conditional Adversarial Autoencoder for de Novo Drug Discovery," *Mol. Pharm.*, vol. 15, no. 10, pp. 4398–4405, Oct. 2018, doi: 10.1021/acs.molpharmaceut.8b00839.

[6] P.-C. Kotsias, J. Arús-Pous, H. Chen, O. Engkvist, C. Tyrchan, and E. J. Bjerrum, "Direct steering of de novo molecular generation with descriptor conditional recurrent neural networks," *Nat. Mach. Intell.*, vol. 2, no. 5, pp. 254–265, May 2020, doi: 10.1038/s42256-020-0174-5.

[7] S. R. Atance, J. V. Diez, O. Engkvist, S. Olsson, and R. Mercado, "De novo drug design using reinforcement learning with graph-based deep generative models," Jul. 2021, doi: 10.26434/chemrxiv-2021-9w3tc.

[8] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Mach. Learn.*, vol. 8, no. 3–4, pp. 229–256, May 1992, doi: 10.1007/BF00992696.