

# KNOWLEDGE-AUGMENTED GRAPH MACHINE LEARNING FOR DRUG DISCOVERY: FROM PRECISION TO INTERPRETABILITY

---

Zhiqiang Zhong & Davide Mottin  
Aarhus University



# OUTLINE

---

- I. Introduction and Motivation
- II. Background of Drug Discovery
- III. Graph Machine Learning (GML) and Knowledge Graph (KG) in Drug Discovery
- IV. Knowledge-augmented Graph Machine Learning (KaGML) for Drug Discovery
- V. Practical Resources
- VI. Open Challenges and Future Directions

# FUNDAMENTALS OF DRUG DISCOVERY

# DRUG DISCOVERY PRINCIPLES

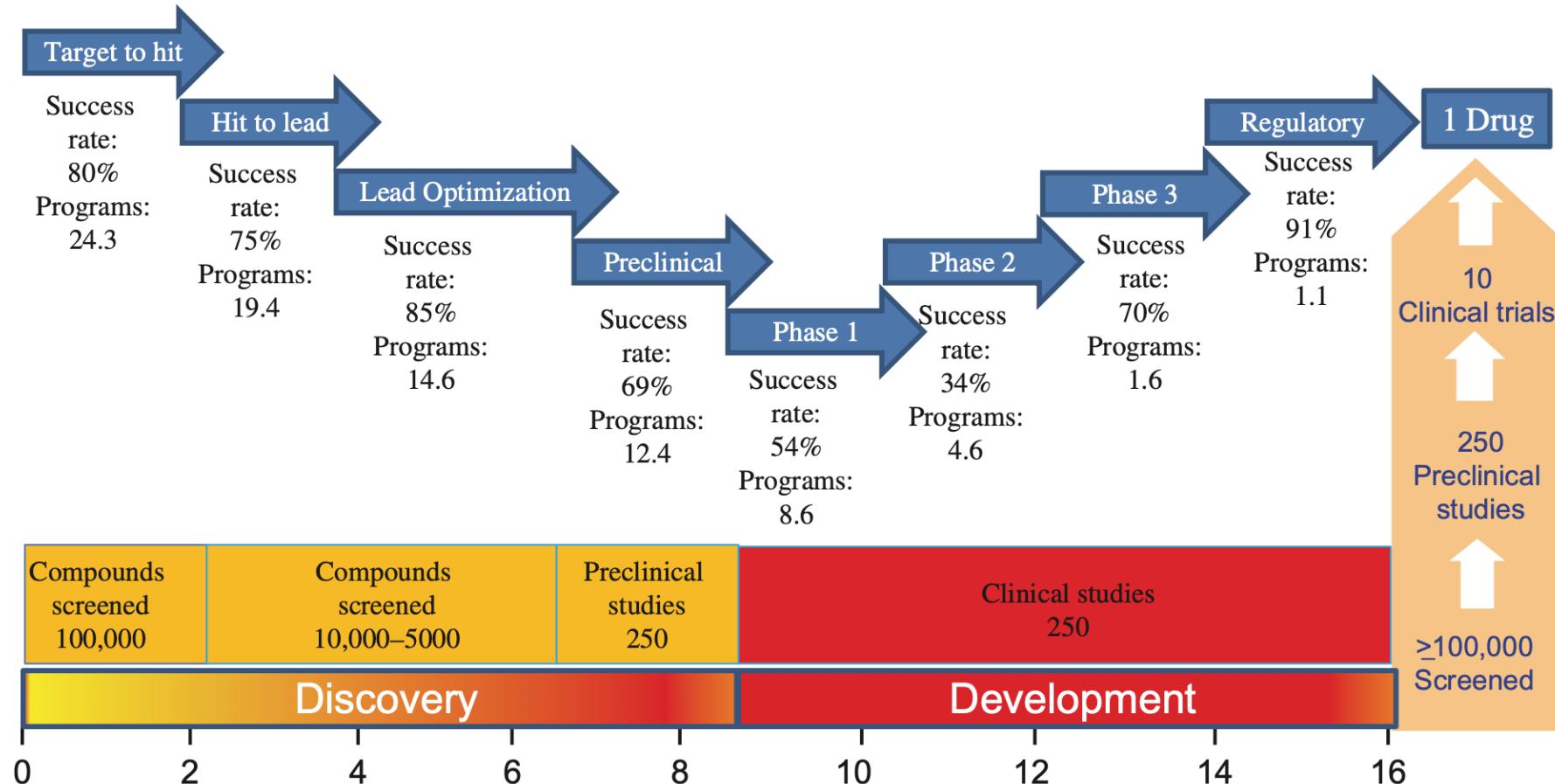


Figure Source: Benjamin E. Blass, Basic Principles of Drug Discovery and Development, 2022

# TECHNOLOGIES FOR DRUG DISCOVERY

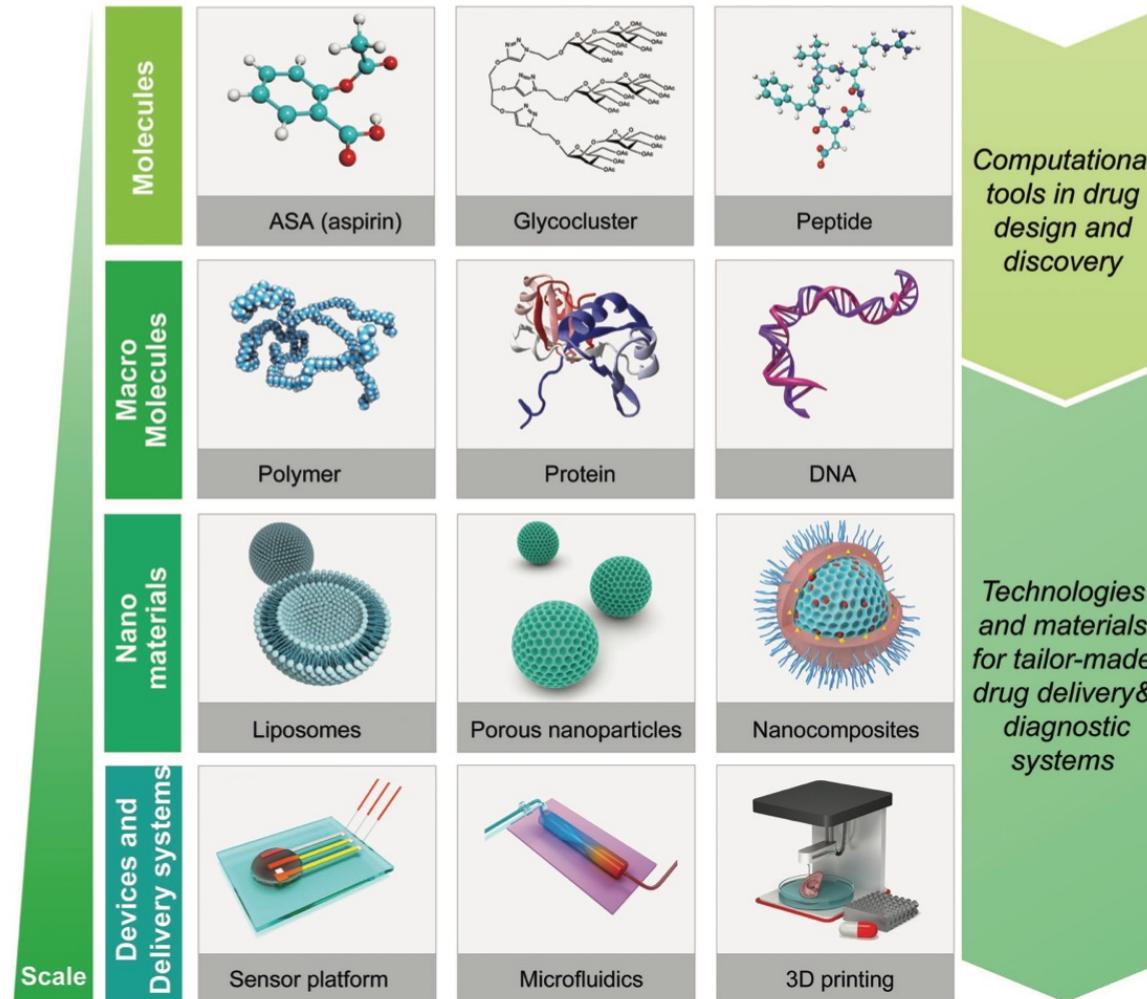


Figure Source: Sahlgren et al., Advanced Healthcare Materials, 2017

# EXAMPLE TOP/DOWN PROTEOMICS

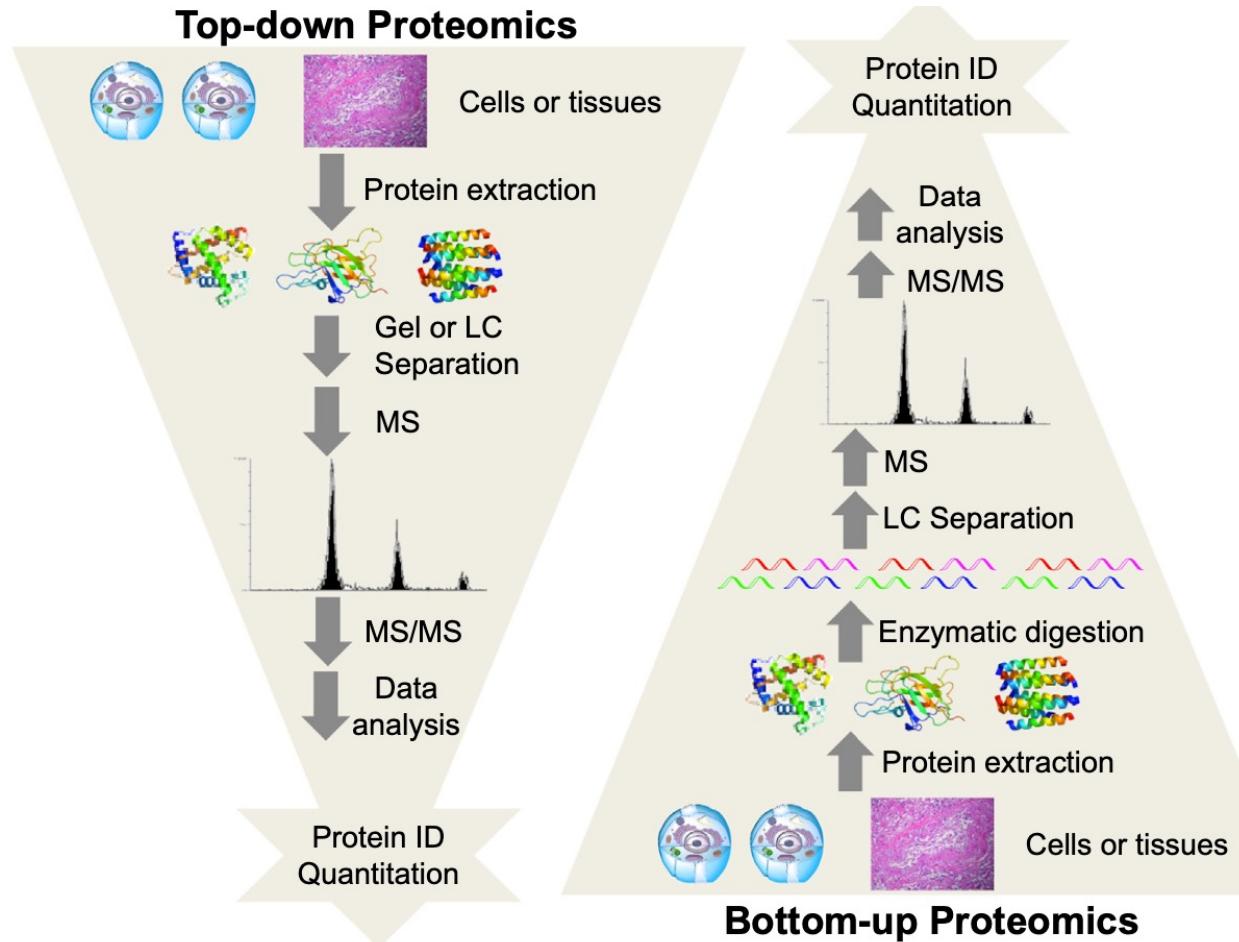


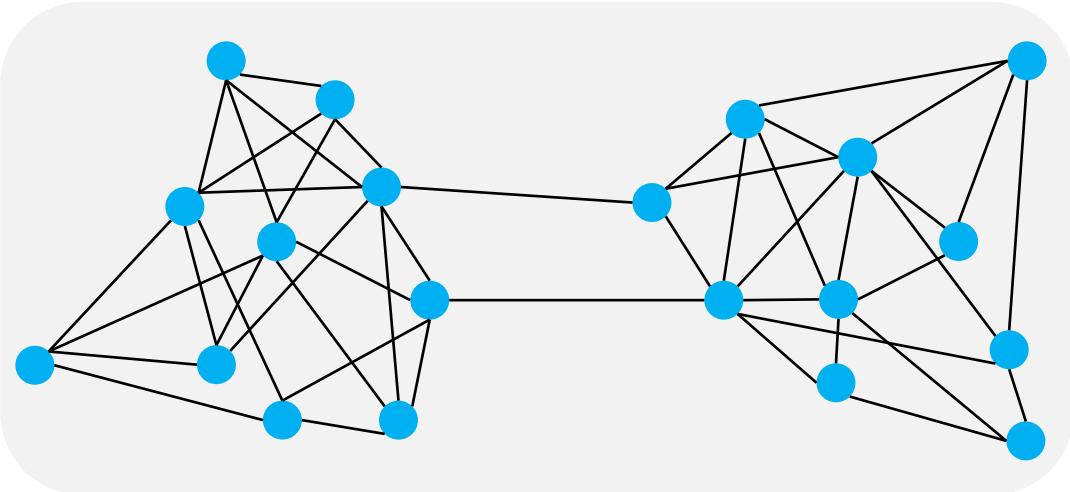
Figure Source: Benjamin E. Blass, Basic Principles of Drug Discovery and Development 2022

# UBIQUITOUS GRAPH-STRUCTURED DATA IN DRUG DISCOVERY

# WHAT IS GRAPH?

---

Graph  
 $\mathcal{G} = (V, E, X)$



Node attribute matrix

$$\mathbf{X} = \begin{bmatrix} 0.4 & 4.4 & \cdots & 3 \\ 0.3 & 9.1 & \cdots & 6 \\ \vdots & \vdots & \ddots & \vdots \\ 0.1 & 0 & \cdots & 6 \\ 0.7 & 1.8 & \cdots & 1 \end{bmatrix}$$

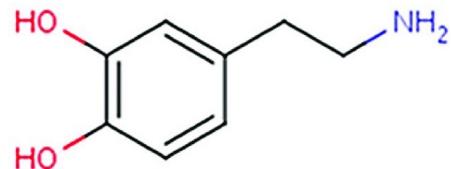
Adjacency matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & \cdots & 0 & 1 \\ 0 & 1 & \cdots & 1 & 0 \end{bmatrix}$$

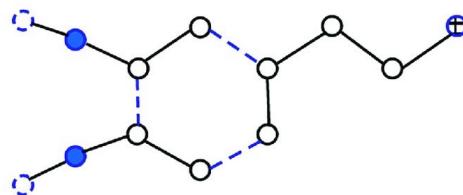
# BIO-DATA ARE GRAPH-STRUCTURED

---

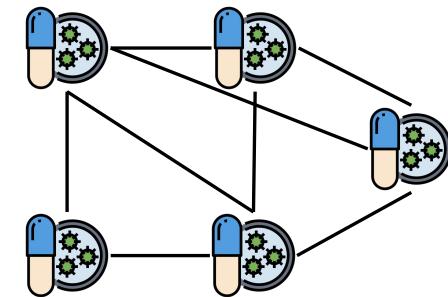
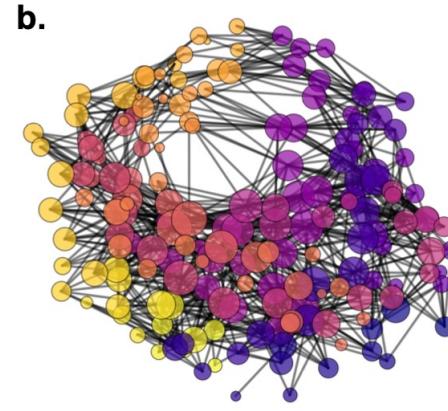
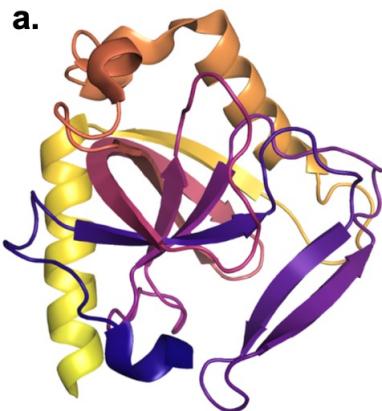
Dopamine



Molecular graph



Small Molecules



**Figure 8:** Illustration of **a.** a protein (PDB accession: 3EIY) and **b.** its graph representation derived based on intramolecular distance with cut-off threshold set at 10Å.

Proteins

Figure Source: *Gaudet et al.*, Brief. Bioinformatics 2022

Drug-Drug Interaction

# WHAT IS KNOWLEDGE DATABASE?

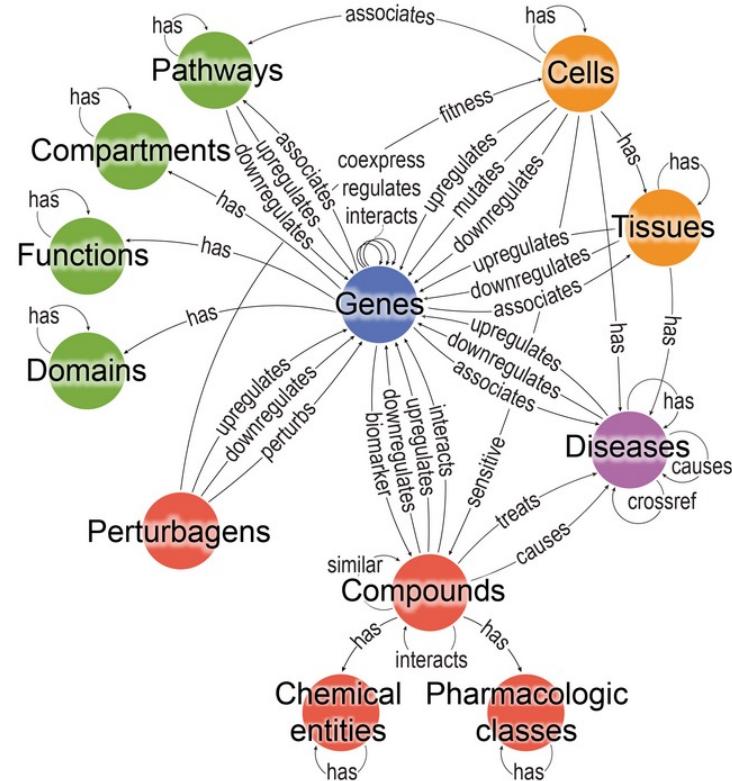
---

“**Knowledge** is the awareness of facts or as practical skills, the understanding of the world, and may also refer to familiarity with objects or situations” [1]

- A knowledge database  $\mathcal{D}$  represents the identified knowledge in a well-organised and structured format.
- A knowledge graph (KG) is a database represented as a directed heterogeneous graph  $\mathcal{KG} = (\mathcal{V}, \mathcal{E}, \phi, \psi)$  with an entity type mapping function  $\phi: \mathcal{V} \rightarrow Z^A$  and an edge type mapping function  $\psi: \mathcal{E} \rightarrow Z^R$ .

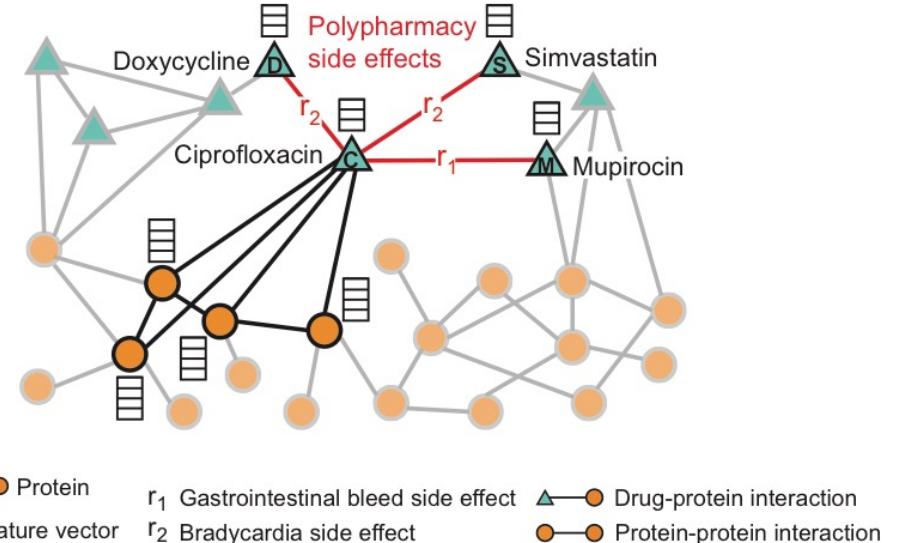
[1] <https://dbpedia.org/page/Knowledge>

# GRAPH-STRUCTURED BIO-KNOWLEDGE



Bioteque

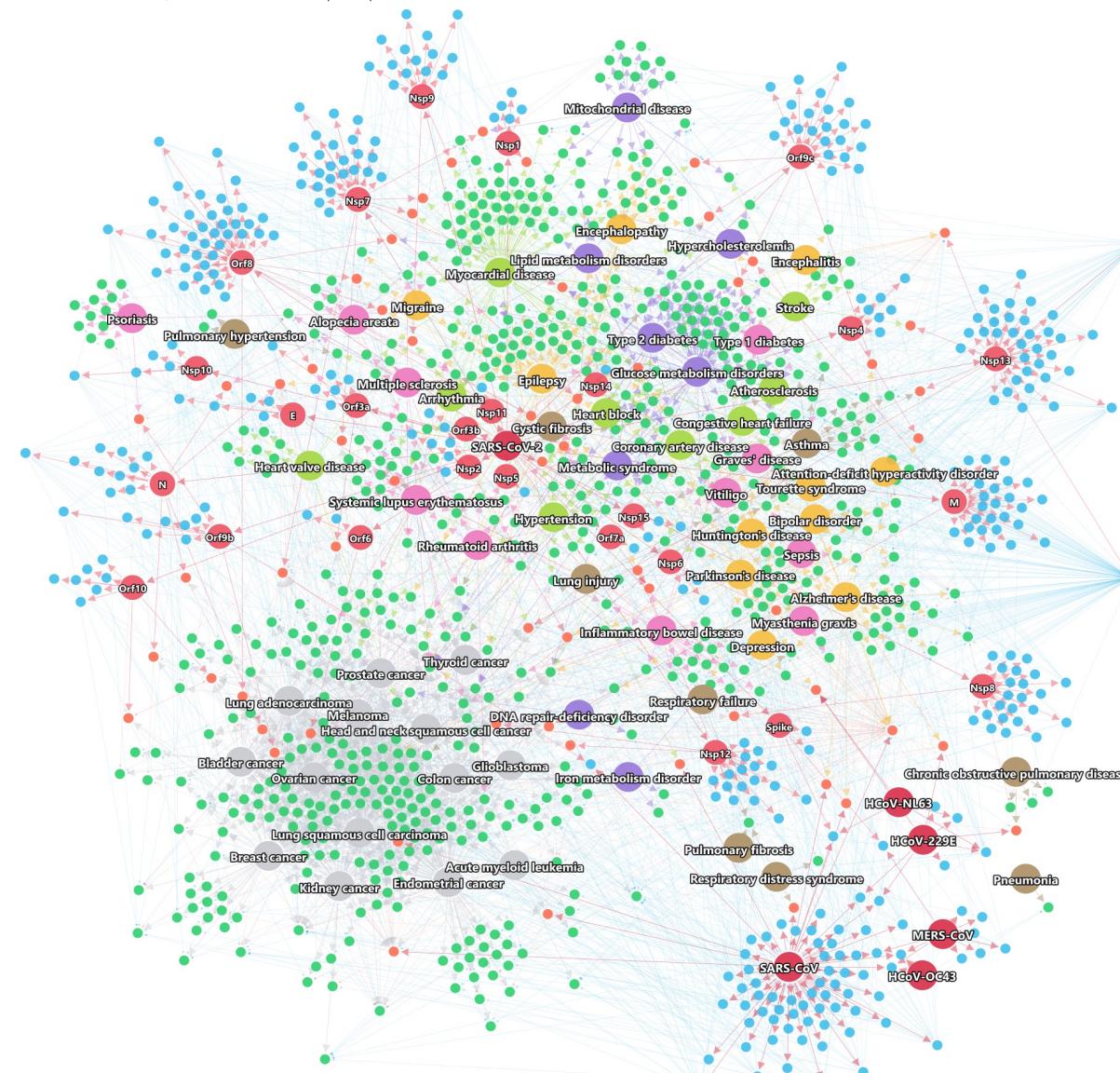
Figure Source: Fernández-Torra *et al.*, Nature Communication 2022



Decagon

Figure Source: Zitnik, *et al.*, Bioinformatics 2022

# DISEASE NETWORK MAP FOR COVID-19



- Autoimmune
- Cancer
- Cardiovascular
- Metabolic
- Neurological
- Pulmonary
- Human coronaviruses (HCoVs)
- SARS-CoV-2 proteins

- Target host proteins of HCoVs
- Disease associated proteins
- Overlap of ● and ●

- Human PPIs
- Virus-human PPIs

- Disease-protein associations

Figure Source: Zhou et al., PLOS Biology 2020

# INTELLIGENT DRUG DISCOVERY RESEARCH QUESTIONS AND TASKS

# INTEGRATING AI IN DRUG DISCOVERY

---

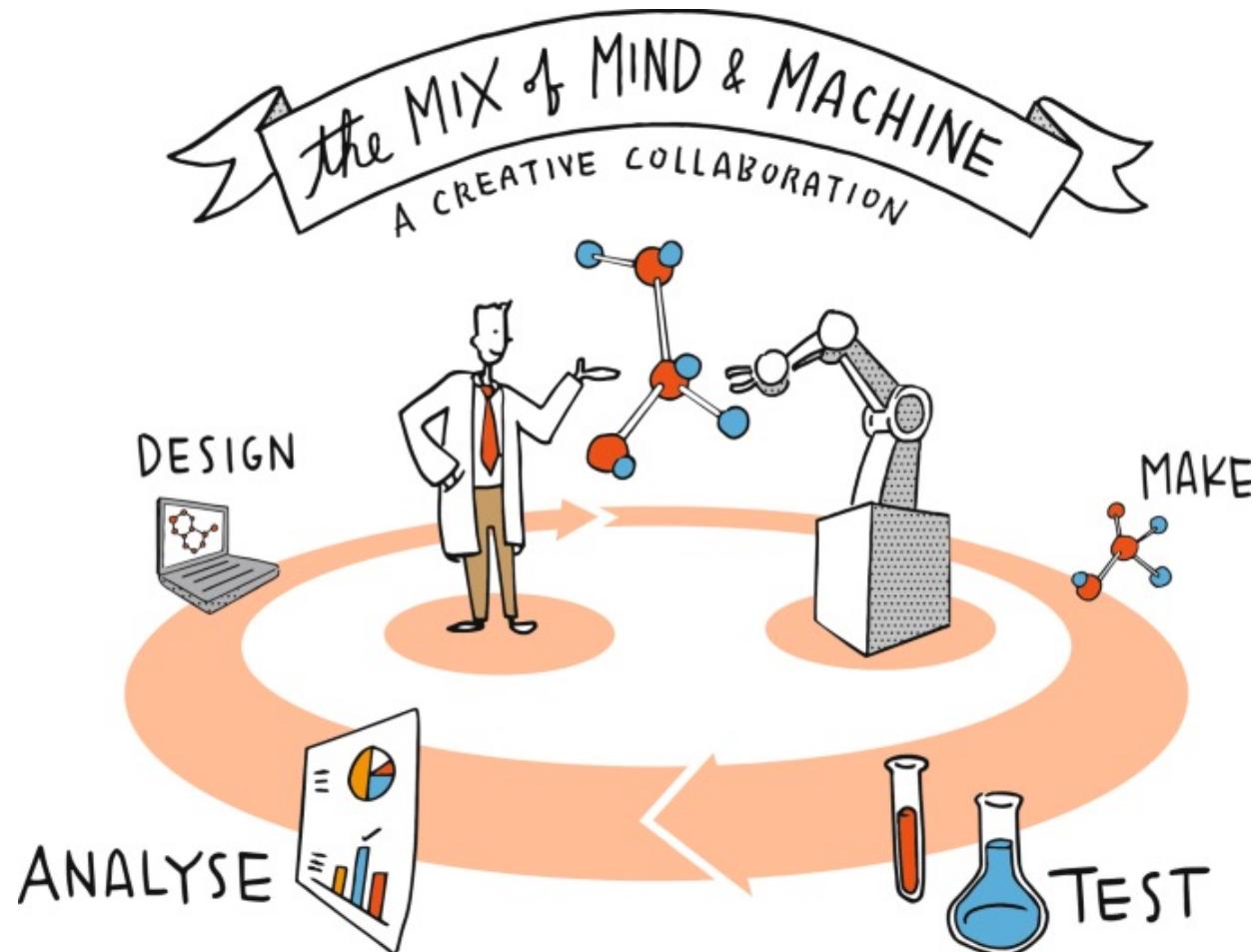
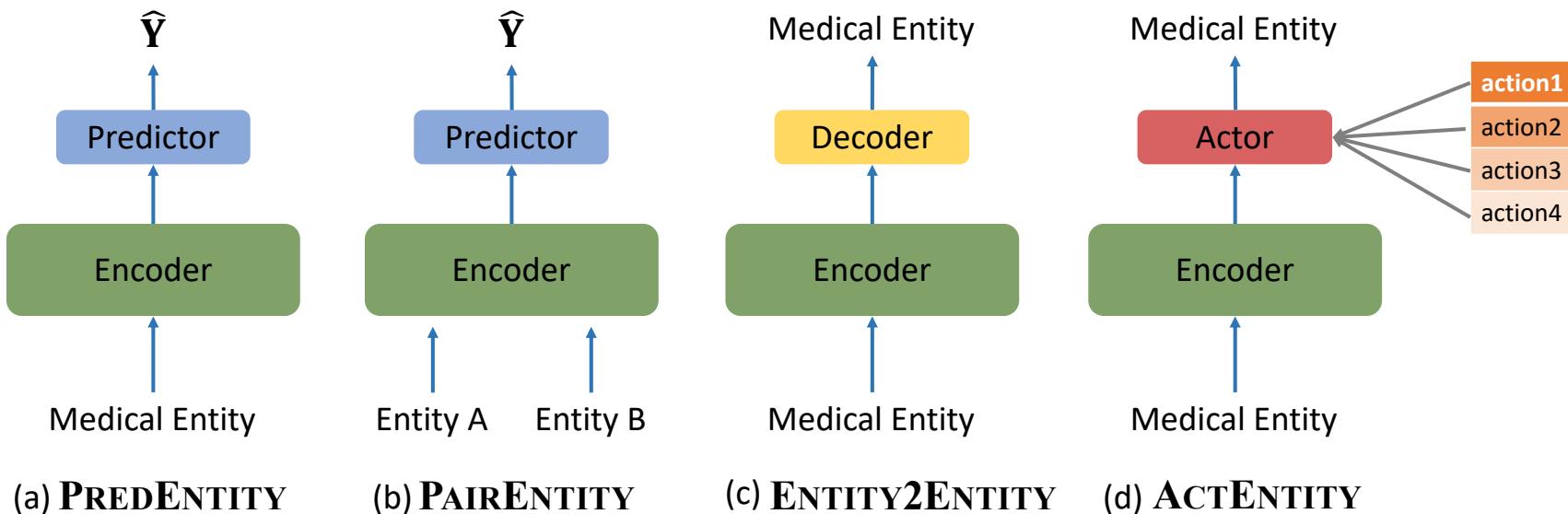


Figure Source: Schneider, et al., Nature Review Drug Discovery 2019

# INTELLIGENT DRUG DISCOVERY TASKS

---

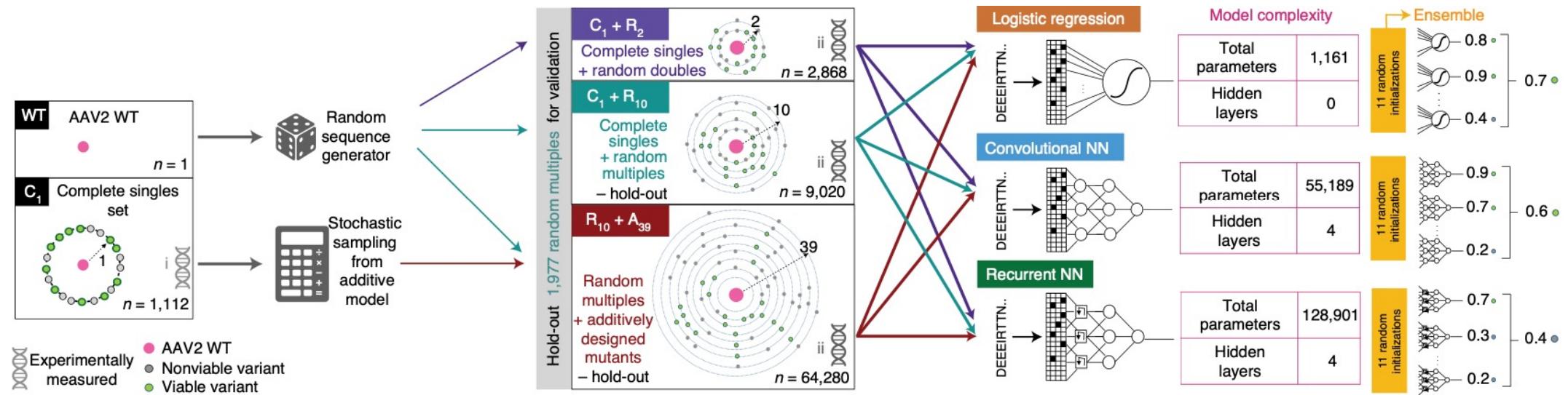
- Four **schemes** to realise intelligent drug discovery



# INTELLIGENT ENTITY-PREDICTION TASKS

## PREDENTITY

Bryant *et al.*, (Nat. Biotechnol.'21)

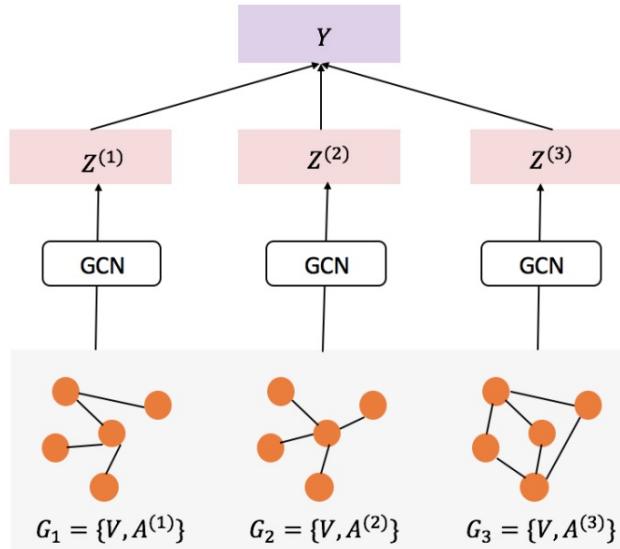


- AI algorithms take AAV proteins as input and make **predictions** on whether they form a valid 3D structure.

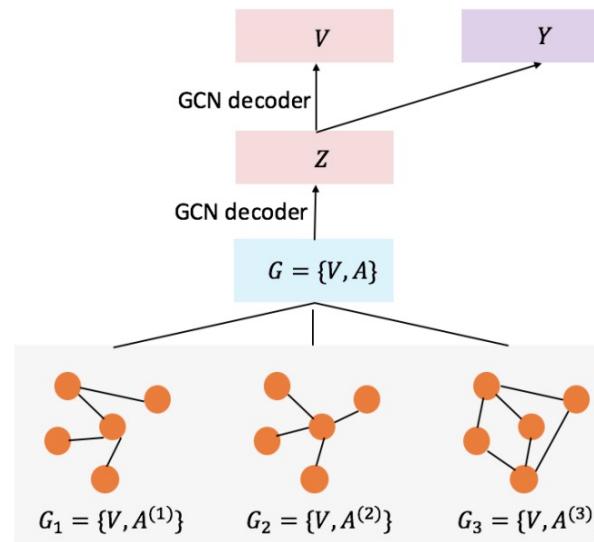
# INTELLIGENT ENTITY-PAIRING TASKS

## PAIRENTITY

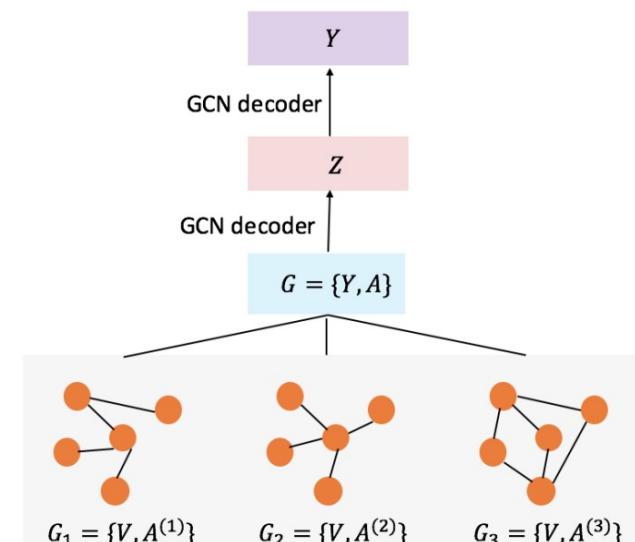
Ma et al., (IJCAI'18)



(a)



(b)



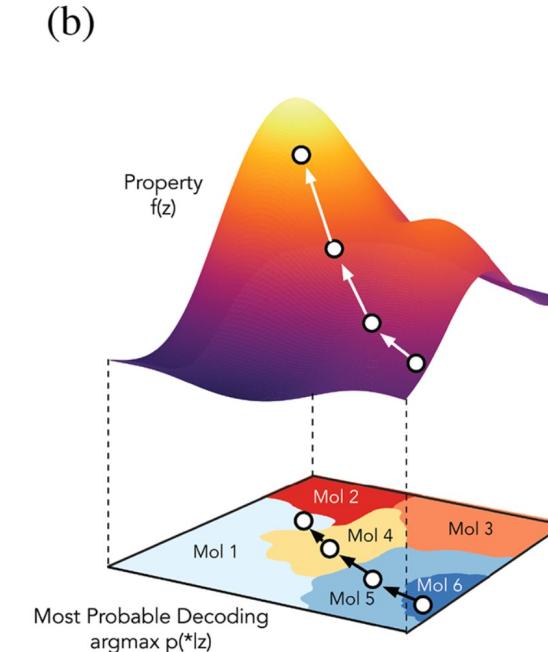
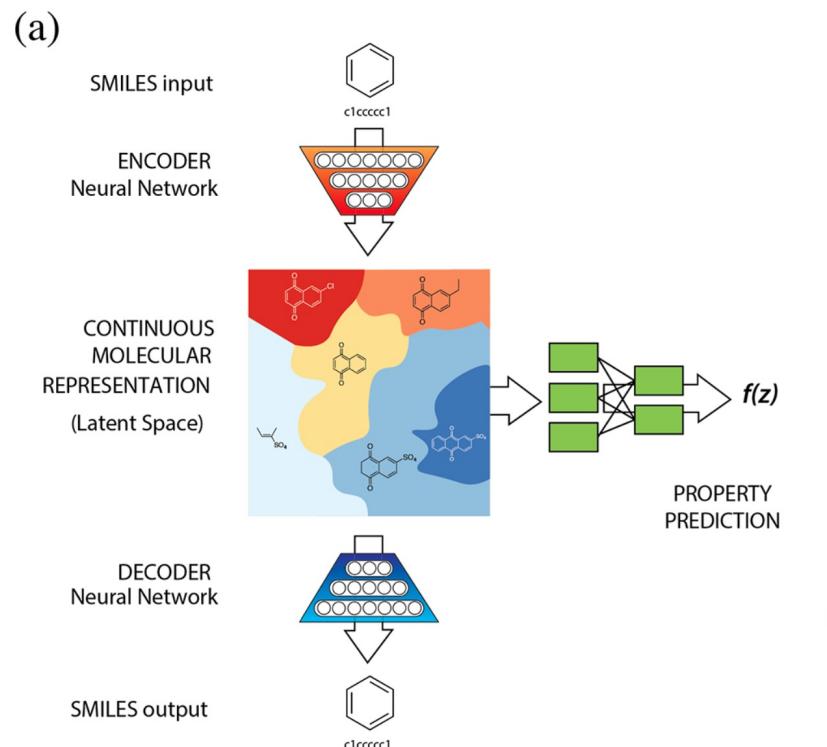
(c)

- Attentive multi-view graph auto-encoders to predict the **similarity** between different drug graphs.

# INTELLIGENT ENTITY-TO-ENTITY TASKS

## ENTITY2ENTITY

Gómez-Bombarelli *et al.*, (ACS Cent. Sci.'18)

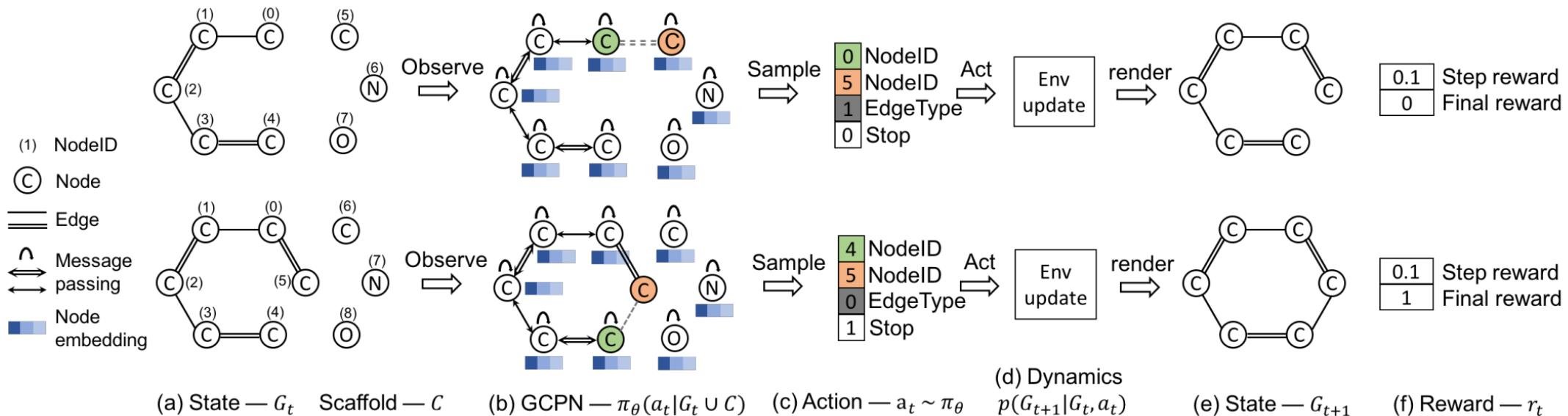


- Generate new molecules with desired new properties based on existing molecules.

# INTELLIGENT ACTION-PREDICTION TASKS

## ACTENTITY

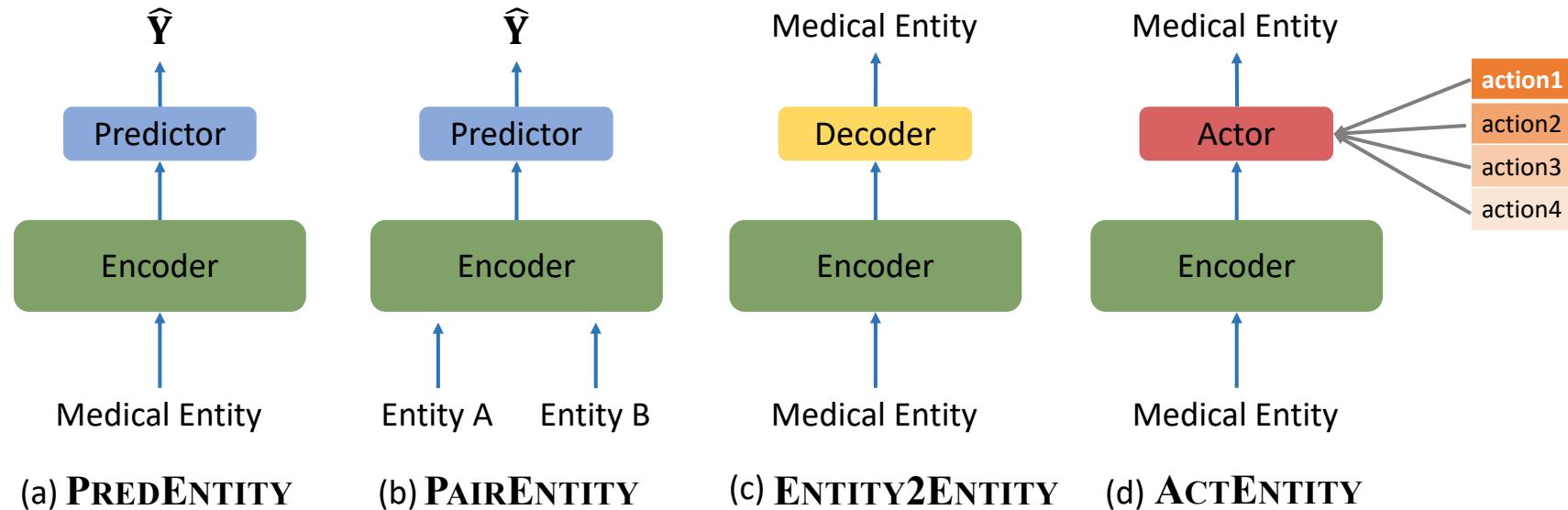
You et al., (NeurIPS'18)



- A set of **actions** are predicted continuously to **modify** existing molecules for desired states.

# INTELLIGENT DRUG DISCOVERY TASKS

- Four **schemes** to realise intelligent drug discovery



- Different biomedical **Entity** can be classified at different levels, including molecular (e.g., **Gene, Molecule**), macro-molecular (e.g., **Protein, Antibodies, Enzymes, Receptors, Compound**, and organism levels (e.g., **Cell**).

# ACKNOWLEDGEMENT

---



Zhiqiang Zhong  
Postdoc  
Aarhus University



Davide Mottin  
Asst. Prof.  
Aarhus University



Our work is supported by the Horizon Europe and Danmarks Innovationsfond under the Eureka, Eurostar grant no E115712.

# Thank you!

## Questions?

[zzhong@cs.au.dk](mailto:zzhong@cs.au.dk)  
<https://zhiqiangzhongddu.github.io/>





AARHUS  
UNIVERSITY