

Tutorial at ECML-PKDD, September 17th /18th 2021



Machine Learning and Reasoning for Drug Discovery

Part A: Introduction to drug discovery pipeline & ML tasks

Truyen Tran, Thin Nguyen and Tri Nguyen
{truyen.tran, thin.nguyen, minhtri}@deakin.edu.au

<https://bit.ly/3Edqgz4>

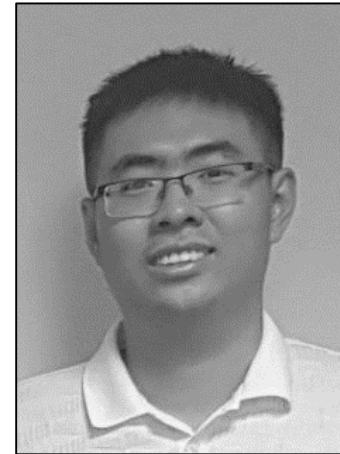
Logistics



Truyen Tran



Thin Nguyen



Tri Nguyen

<https://bit.ly/3Edqgz4>

Agenda

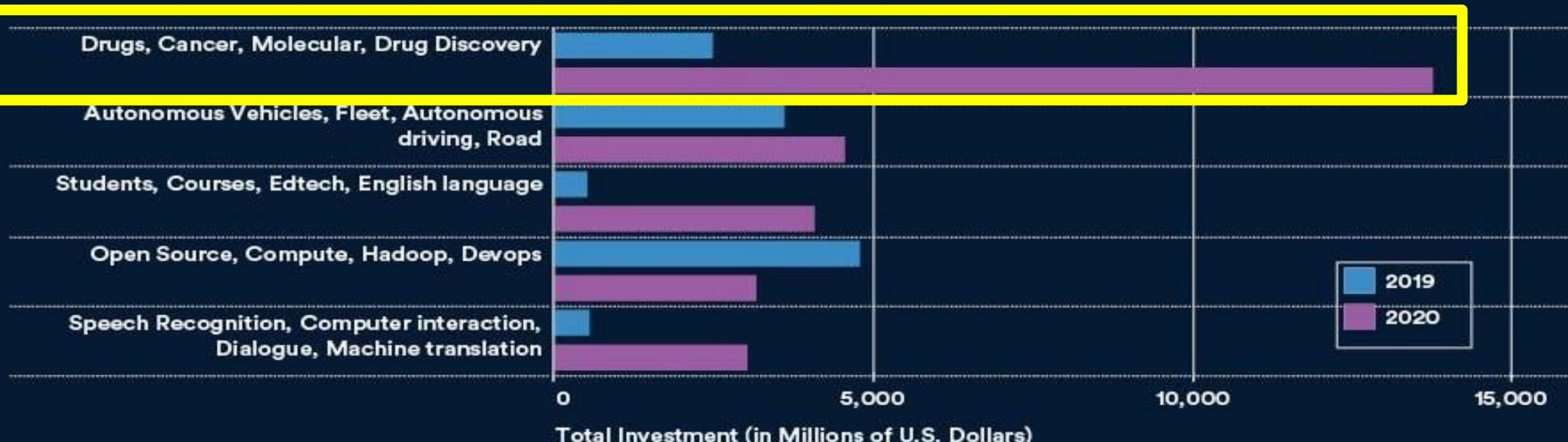
- Intro to drug discovery pipeline & ML tasks
- Molecular representation and property prediction
- Protein representation and protein-drug binding
- Molecular optimisation & generation
- Knowledge graph reasoning & Drug synthesis
- Wrapping up

3

AI Industry Shifts Focus to Drug Discovery

Global investment spikes in reaction to pandemic.

GLOBAL PRIVATE INVESTMENT IN AI BY FOCUS AREA, 2019 VS 2020



Source: AI Index 2021, HAI Stanford

Data Source: CAPIQ, Crunchbase, and NetBase Quid
Adapted from the AI Index 2021 Annual Report

Eroom's law (inverse of Moore's)

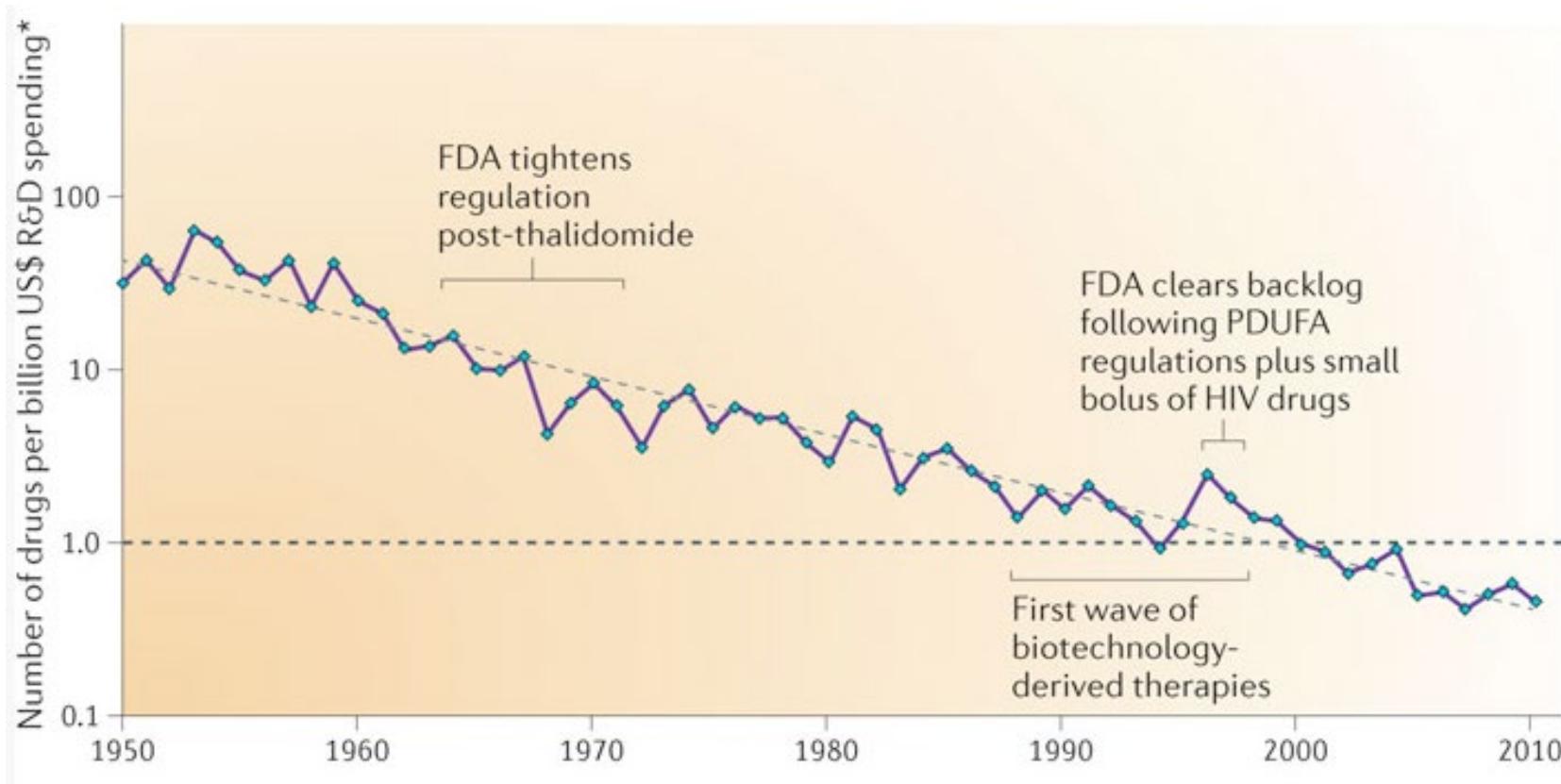
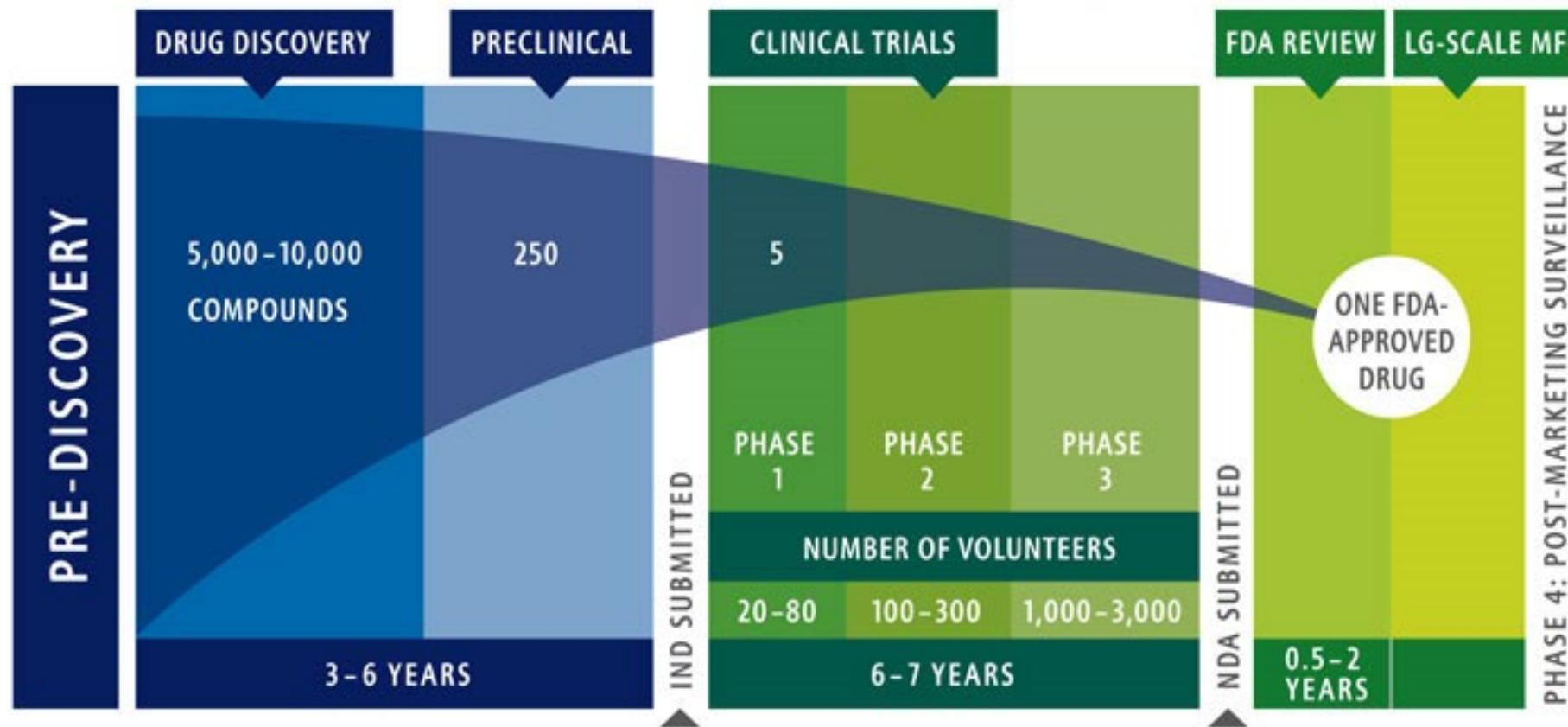


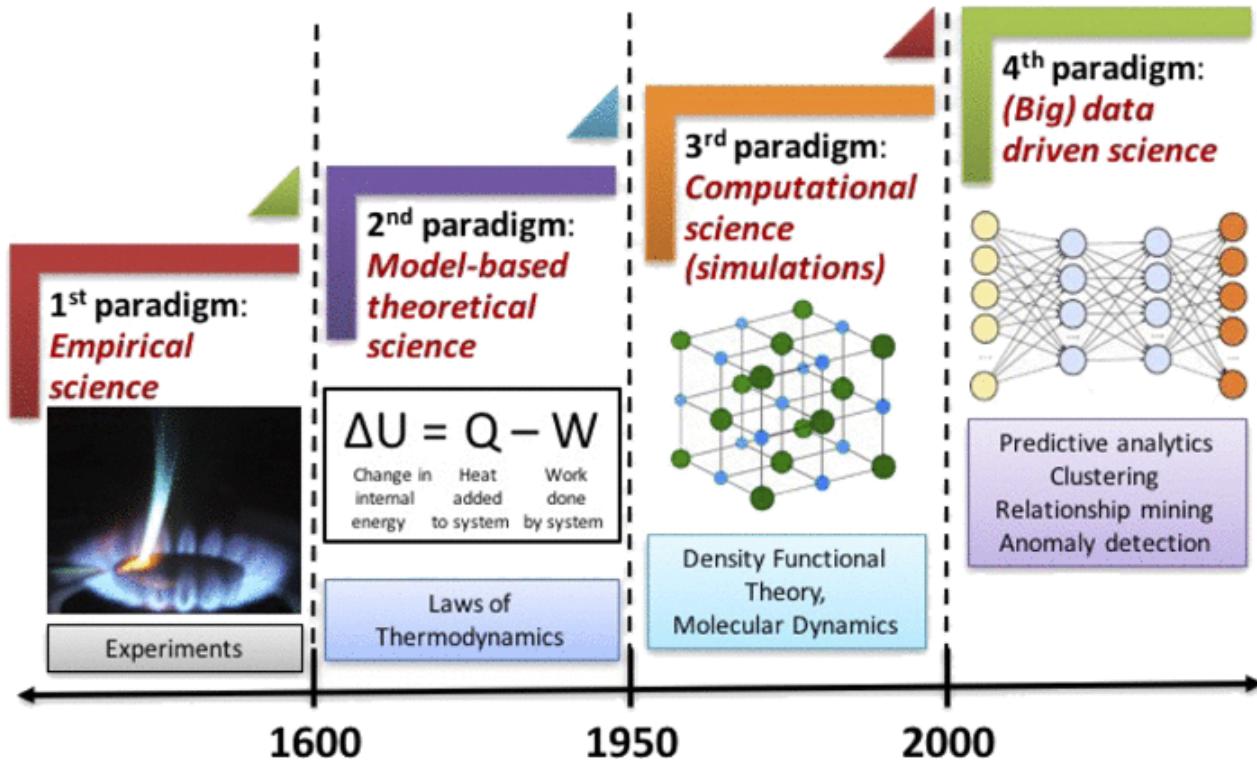
Image source: <https://www.science.org/content/blog-post/erom-s-law>

Drug discovery and development

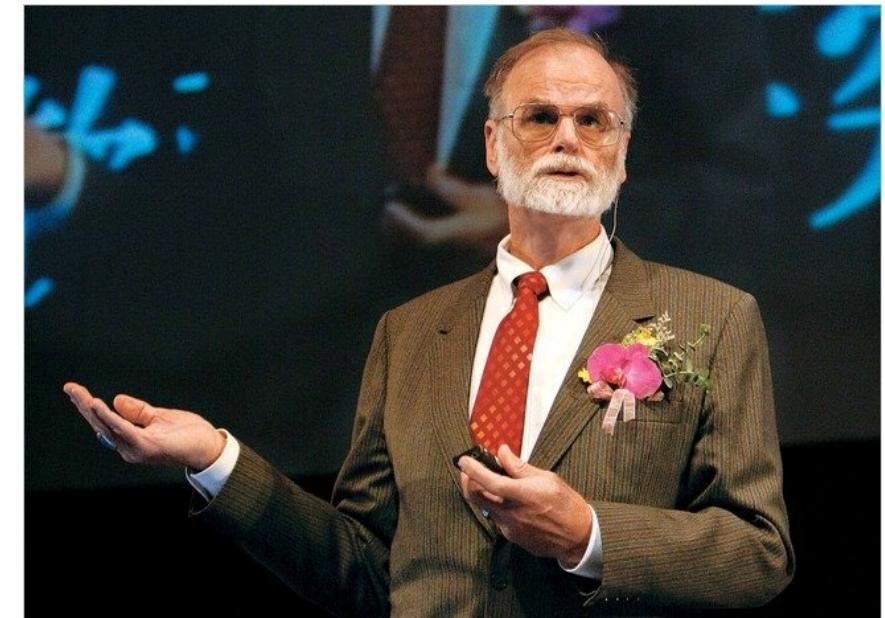


Source: Pharmaceutical Research and Manufacturers of America

The fourth paradigm of science

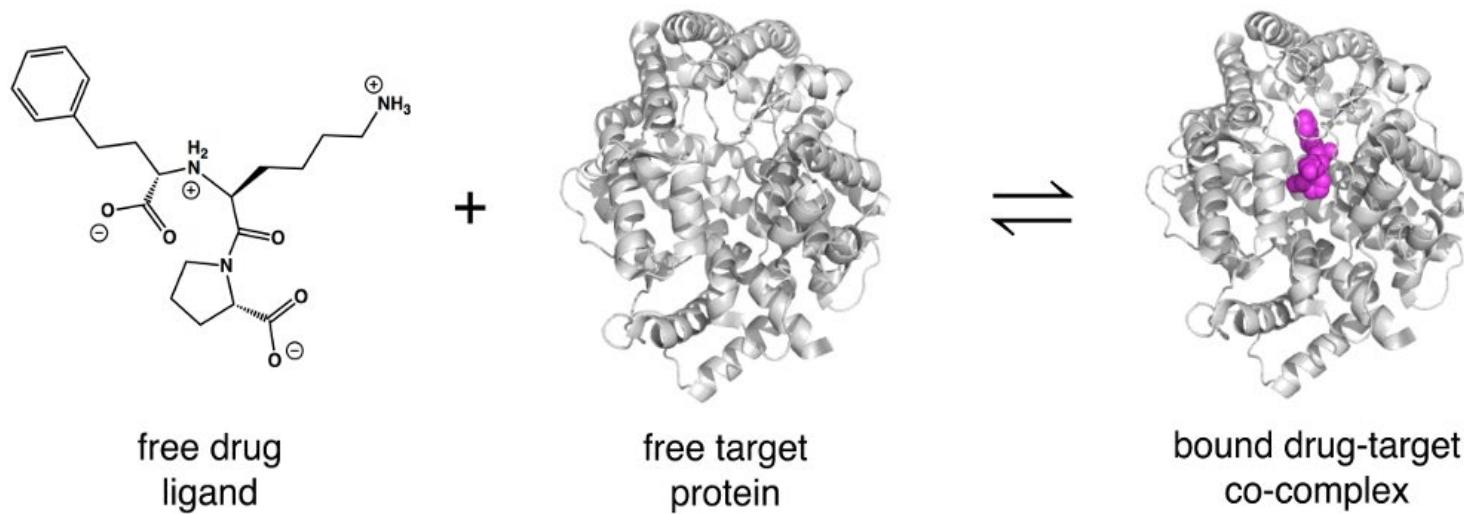


Agrawal, A., & Choudhary, A. (2016). Perspective: Materials informatics and big data: Realization of the “fourth paradigm” of science in materials science. *Appl Materials*, 4(5), 053208.



Jim Gray, Turing Award 1998 (1944-2007)
Honoured as father of **The 4th Paradigm**
Image source: Wikipedia

- **Drug discovery** is the process through which potential new medicines are identified. It involves a wide range of scientific disciplines, including **biology**, **chemistry** and **pharmacology** (*Nature*, 2019).



Source: c4xdiscovery.com

- Drug is a small molecule **that binds to a bio target** (e.g., protein) and **modifies its functions** to produce useful physiological or mental effects.
 - Proteins are **large biomolecules** consisting of chains of amino acid residues.

Drug-likeness

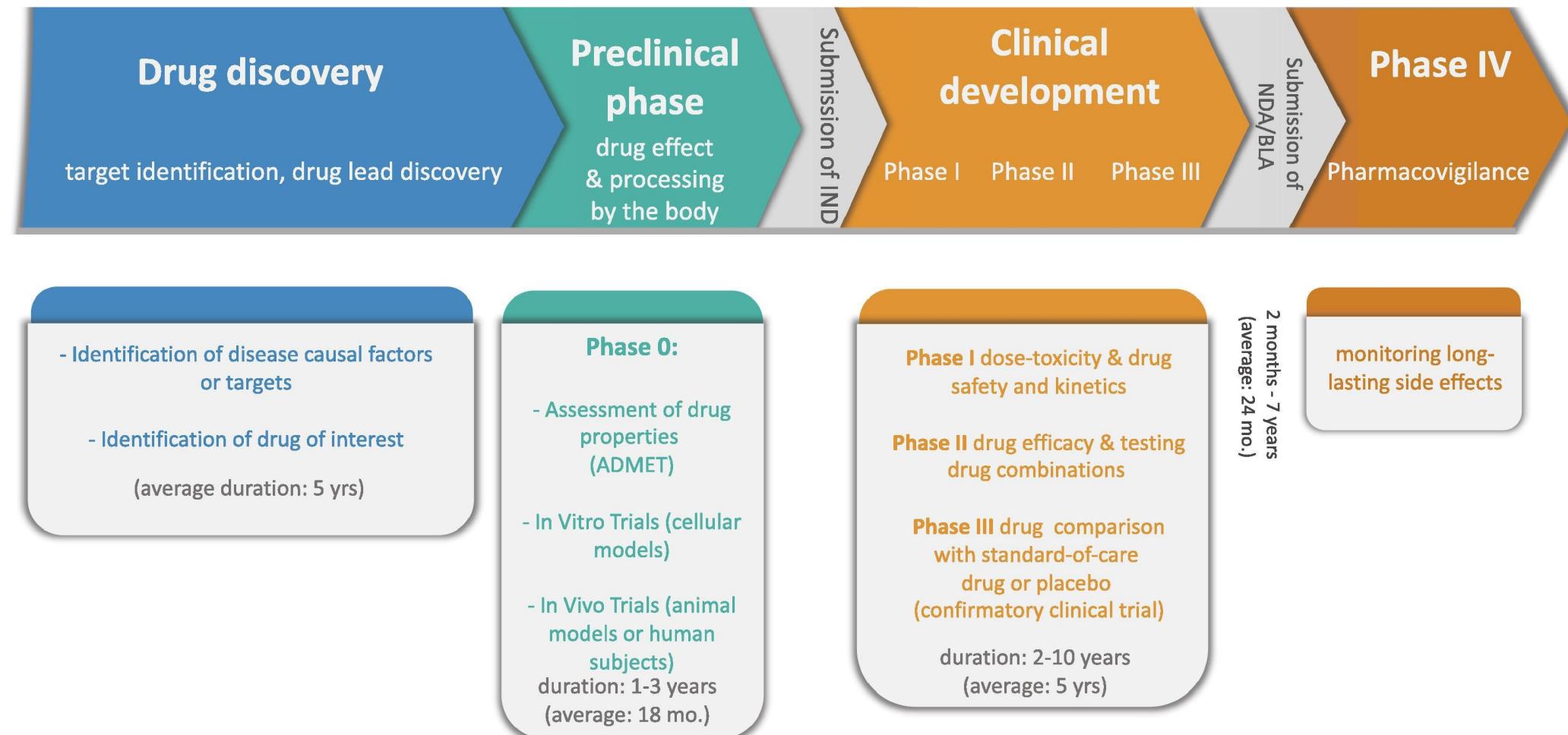
- Solubility in water and fat, e.g., measured by LogP. Most drugs are admitted orally → pass through membrane.
- Potency at the bio target → target-specific binding.
- Ligand efficiency (low energy binding) and lipophilic efficiency.
- Small molecular weight → affect diffusion

Lipinski's Rule of Five (RO5)

1. No more than 5 hydrogen bond donors (the total number of nitrogen–hydrogen and oxygen–hydrogen bonds)
2. No more than 10 hydrogen bond acceptors (all nitrogen or oxygen atoms)
3. A molecular mass less than 500 daltons
4. An octanol–water partition coefficient ($\log P$) that does not exceed 5

https://en.wikipedia.org/wiki/Lipinski%27s_rule_of_five

Lipinski, Christopher A., et al. "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings." *Advanced drug delivery reviews* 23.1-3 (1997): 3-25.



Réda, Clémence, Emilie Kaufmann, and Andrée Delahaye-Duriez. "Machine learning applications in drug development." *Computational and structural biotechnology journal* 18 (2020): 241-252.

Drug discovery as reasoning

Reasoning is to deduce new knowledge from previously acquired knowledge in response to a query (or a cues)

Practical setting: (query, database, answer) triplets

- Classification: Query = *Is this a drug?* Database = *atomic structure of drug*.
- Regression: Query = *how toxic is this drug?* Database = *drug*.
- QA: Query = *NLP question*. Database = *context/image/text*.
- Multi-task learning: Query = *task ID*. Database = *drug/protein*.
- Zero-shot learning: Query = *task description*. Database = *data*.
- Drug-protein binding: Query = *drug*. Database = *protein*.
- Recommender system: Query = *Target (drug)*. Database = {*CCI, PPI, Drug-target, gene-diseases*};

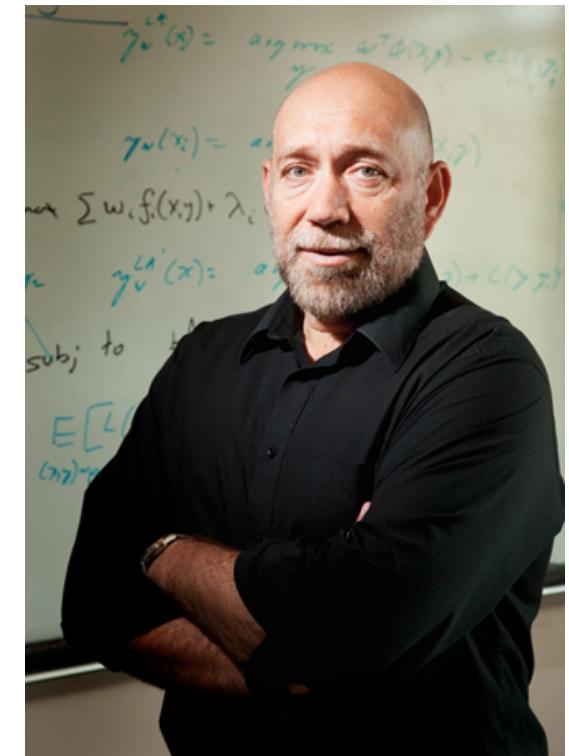
Drug discovery as learning to reason

Learning is to improve itself by experiencing ~
acquiring knowledge & skills

Learning to reason is to improve the ability to
decide if a knowledge base entails a predicate.

E.g., given a disease and a knowledge base,
determines if a drug will have treatment effect.

See our IJCAI'21 tutorial for more detail:
<https://neuralreasoning.github.io>



(Dan Roth; ACM Fellow; IJCAI
John McCarthy Award)

Khardon, Roni, and Dan Roth. "Learning to reason." *Journal of the ACM* (JACM) 44.5 (1997): 697-725.

The three basic questions in drug discovery

Given a molecule, is this drug? Aka properties/targets/effects prediction.

- Drug-likeness
- Targets it can modulate and how much
- Its dynamics/kinetics/effects/metabolism if administered orally or via injection

Given a target, what are molecules?

- If the list of molecules is given, pick the good one. If evaluation is expensive, need to search, e.g., using Bayesian Optimization.
- If no molecule is found, need to generate from scratch → generative models + Bayesian Optimization, or Reinforcement Learning.
- How does the drug-like space look like?

Given a molecular graph, what are the steps to make the molecule?

- Synthetic tractability
- Reaction planning, or retrosynthesis

We need powerful machine learning for drug discovery

Expressiveness

- Can represent the complexity of the biomedical world
- Ideally, can represent all the chemical space
- Can compute anything computable

Learnability

- Have mechanism to learn from the training signals (or lack of)

Generalizability

- Work on unseen data

Will neural networks be suitable for drug-discovery reasoning?

Reasoning is not necessarily achieved by making logical inferences

There is a continuity between [algebraically rich inference] and [connecting together trainable learning systems]

Central to reasoning is composition rules to guide the combinations of modules to address new tasks



"When we observe a visual scene, when we hear a complex sentence, we are able to explain in formal terms the relation of the objects in the scene, or the precise meaning of the sentence components. However, there is no evidence that such a formal analysis necessarily takes place: we see a scene, we hear a sentence, and we just know what they mean. This suggests the existence of a middle layer, already a form of reasoning, but not yet formal or logical."

On suitability of deep learning for drug discovery

Theoretical

- **Expressiveness:** Neural nets can approximate any function.
- **Learnability:** Neural nets are trained easily.
- **Generalisability:** Neural nets generalize surprisingly well to unseen data.

Practical

- **Generality:** Applicable to many domains.
- **Competitive:** DL is hard to beat as long as there are data to train.
- **Scalability:** DL is better with more data, and it is very scalable.

Tutorial at ECML-PKDD, September 17th /18th 2021



Machine Learning and Reasoning for Drug Discovery

Part B1: Molecular representation and property prediction

Truyen Tran, Thin Nguyen and Tri Nguyen
{truyen.tran, thin.nguyen, minhtri}@deakin.edu.au

<https://bit.ly/3Edqgz4>

Agenda

Molecular representation learning

- Fingerprints
- String representation
- Graph representation
- Self-supervised learning

Molecular property prediction

- Approximating quantum chemistry computation
- Graph regression and classification
- Graph multitask learning
- Explanation
- Data efficient learning

Neural representation of the world

Vector → Embedding, MLP

Sequence & Tree → RNN (LSTM, GRU), Tree-RNN

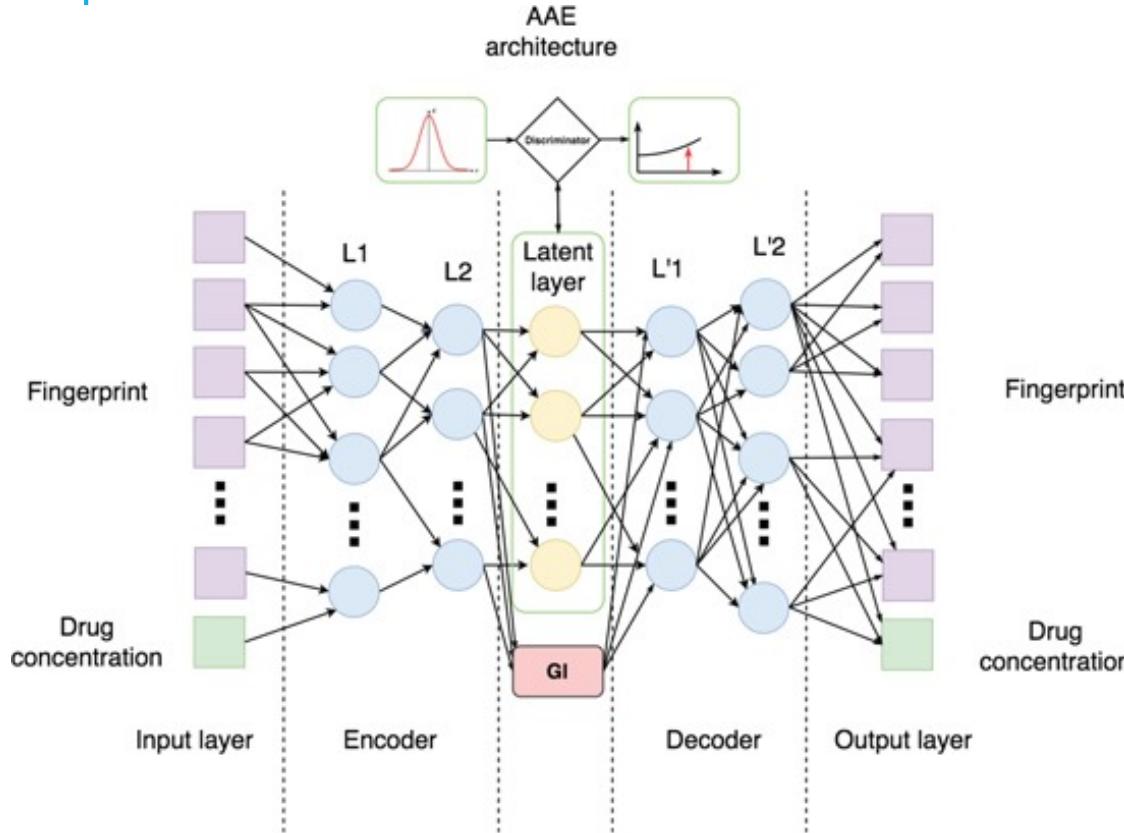
Unordered set → Word2vec, Attention, Transformer

Graph → GNN (node2vec, DeepWalk, GCN, Graph

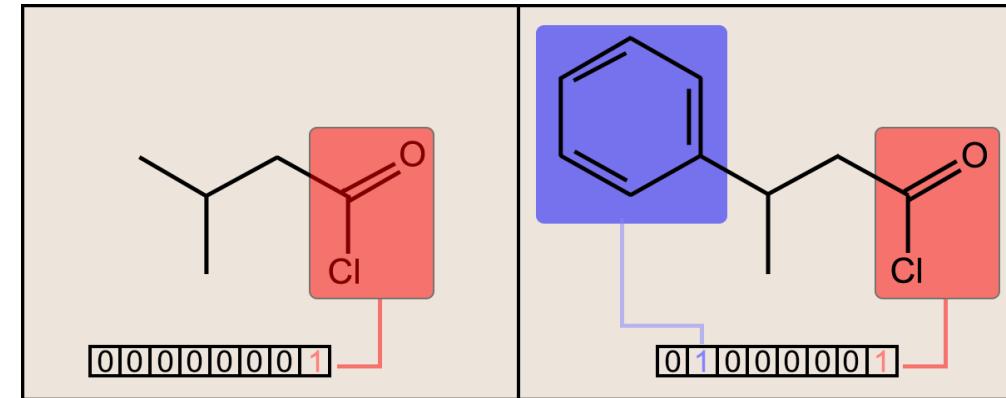
Attention Net, Column Net, MPNN etc)

- Grid is a special case → CNN (AlexNet, VGG, ResNet, EfficientNet, etc)
- Transformer is a special case of GNN on fully connected graph.

Molecule → fingerprints



Kadurin, Artur, et al. "The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology." *Oncotarget* 8.7 (2017): 10883.



Molecular Descriptors. (2019, October 26). Retrieved September 14, 2021, from <https://chem.libretexts.org/@go/page/192626>

Graph → hashing → vector. Mostly discrete. Substructures coded.

Vectors are easy to retrieve & manipulate.

- Ready for use in classical ML algorithms (e.g., SVM, RF, kNN)

Very difficult to reconstruct the graphs from fingerprints.

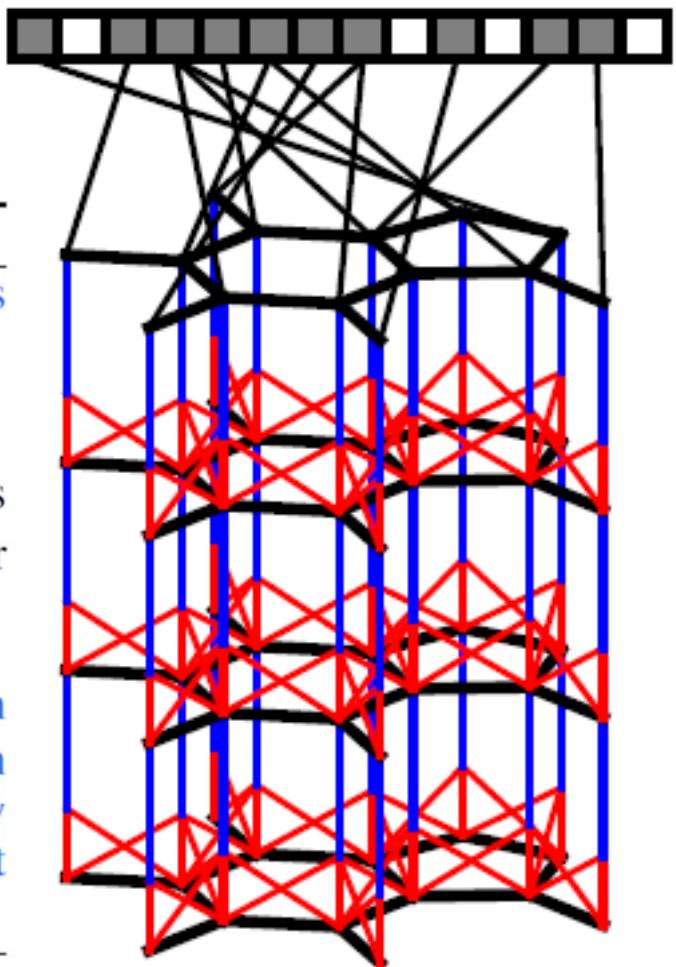
(Circular) fingerprints can be learnt

Algorithm 1 Circular fingerprints

```
1: Input: molecule, radius  $R$ , fingerprint length  $S$ 
2: Initialize: fingerprint vector  $\mathbf{f} \leftarrow \mathbf{0}_S$ 
3: for each atom  $a$  in molecule
4:    $\mathbf{r}_a \leftarrow g(a)$             $\triangleright$  lookup atom features
5: for  $L = 1$  to  $R$             $\triangleright$  for each layer
6:   for each atom  $a$  in molecule
7:      $\mathbf{r}_1 \dots \mathbf{r}_N = \text{neighbors}(a)$ 
8:      $\mathbf{v} \leftarrow [\mathbf{r}_a, \mathbf{r}_1, \dots, \mathbf{r}_N]$      $\triangleright$  concatenate
9:      $\mathbf{r}_a \leftarrow \text{hash}(\mathbf{v})$             $\triangleright$  hash function
10:     $i \leftarrow \text{mod}(r_a, S)$        $\triangleright$  convert to index
11:     $\mathbf{f}_i \leftarrow 1$                   $\triangleright$  Write 1 at index
12: Return: binary vector  $\mathbf{f}$ 
```

Algorithm 2 Neural graph fingerprints

```
1: Input: molecule, radius  $R$ , hidden weights  $H_1^1 \dots H_R^5$ , output weights  $W_1 \dots W_R$ 
2: Initialize: fingerprint vector  $\mathbf{f} \leftarrow \mathbf{0}_S$ 
3: for each atom  $a$  in molecule
4:    $\mathbf{r}_a \leftarrow g(a)$             $\triangleright$  lookup atom features
5: for  $L = 1$  to  $R$             $\triangleright$  for each layer
6:   for each atom  $a$  in molecule
7:      $\mathbf{r}_1 \dots \mathbf{r}_N = \text{neighbors}(a)$ 
8:      $\mathbf{v} \leftarrow \mathbf{r}_a + \sum_{i=1}^N \mathbf{r}_i$             $\triangleright$  sum
9:      $\mathbf{r}_a \leftarrow \sigma(\mathbf{v} H_L^N)$             $\triangleright$  smooth function
10:     $\mathbf{i} \leftarrow \text{softmax}(\mathbf{r}_a W_L)$        $\triangleright$  sparsify
11:     $\mathbf{f} \leftarrow \mathbf{f} + \mathbf{i}$             $\triangleright$  add to fingerprint
12: Return: real-valued vector  $\mathbf{f}$ 
```



#REF: Duvenaud, David K., et al. "Convolutional networks on graphs for learning molecular fingerprints." *Advances in neural information processing systems*. 2015.

Molecule → string

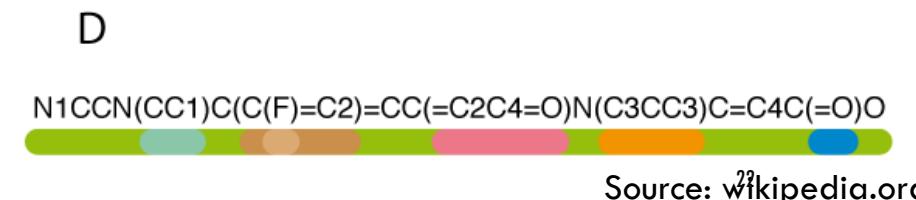
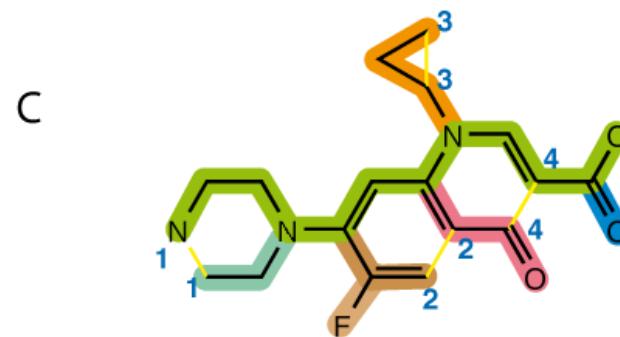
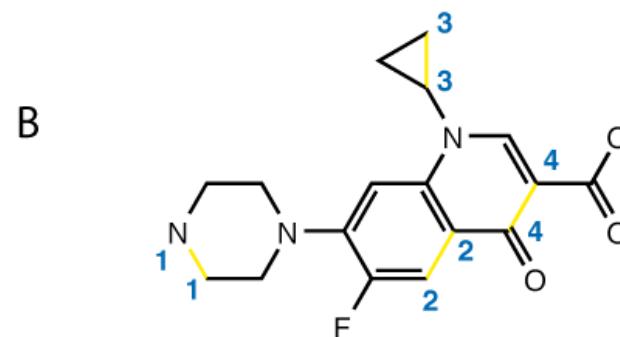
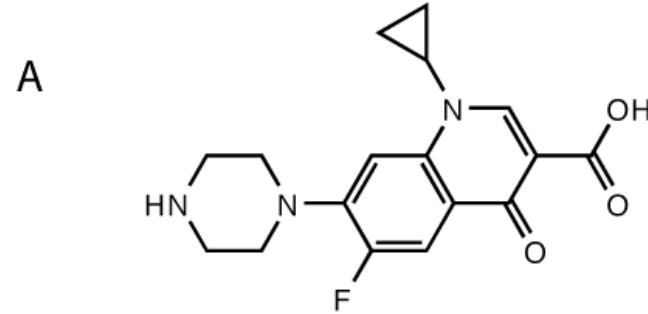
SMILES = Simplified Molecular-Input Line-Entry System

Ready for encoding/decoding with sequential models (seq2seq, MANN, RL).

BUT ...

- String → graphs is not unique!
- Lots of string are invalid
- Precise 3D information is lost
- Short range dependencies in graph may become long range in string

#REF: Gómez-Bombarelli, Rafael, et al. "Automatic chemical design using a data-driven continuous representation of molecules." *arXiv preprint arXiv:1610.02415* (2016).



Molecule → graphs

No regular, fixed-size structures

Graphs are ***permutation invariant***:

- #permutations are exponential function of #nodes
- The probability of a generated graph G need to be marginalized over all possible permutations

Multiple objectives:

- **Diversity** of generated graphs
- **Smoothness** of latent space
- Agreement with or optimization of multiple “**drug-like**” objectives

Graphs are natural representation

Molecule as graph: atoms as nodes, chemical bonds as edges

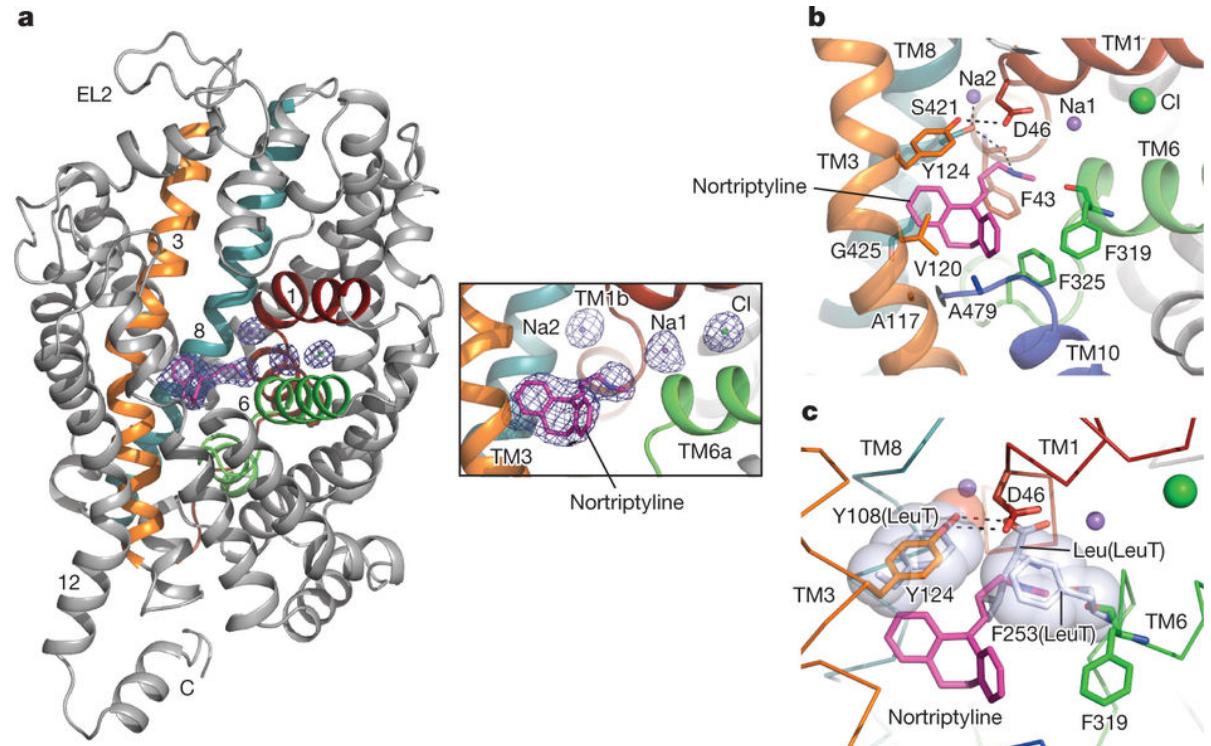
Computing molecular properties as **graph classification/regression**

Drug-target binding as **graph-in-graph interaction**

Chemical-chemical interaction as **graph-graph relation**

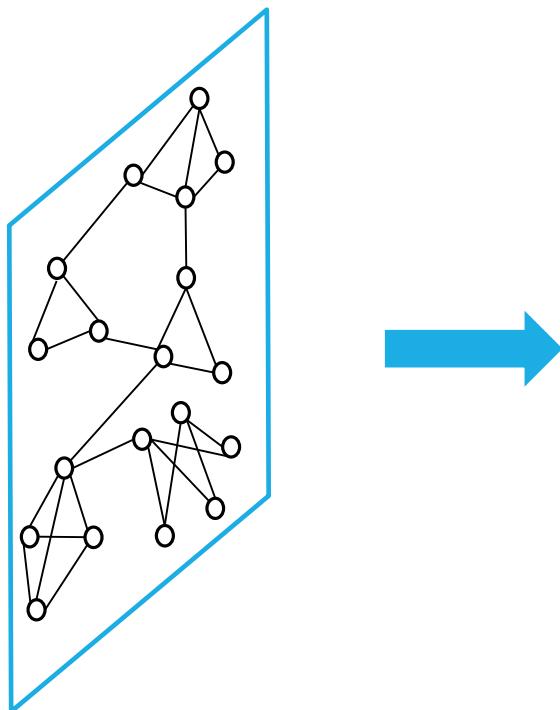
Molecular optimisation as **graph edit/translation**

Chemical reaction as **graph morphism**

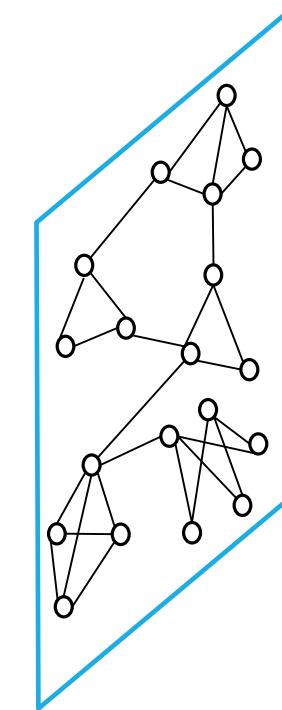


#REF: Penmatsa, Aravind, Kevin H. Wang, and Eric Gouaux. "X-ray structure of dopamine transporter elucidates antidepressant mechanism." *Nature* 503.7474 (2013): 85-90.

Graph filtering: Refining node embedding



Graph Filtering



$$\mathbf{A} \in \{0, 1\}^{n \times n}, \mathbf{X} \in \mathbb{R}^{n \times d}$$

$$\mathbf{A} \in \{0, 1\}^{n \times n}, \mathbf{X}_f \in \mathbb{R}^{n \times d_{new}}$$

Prelim: Skip-gram

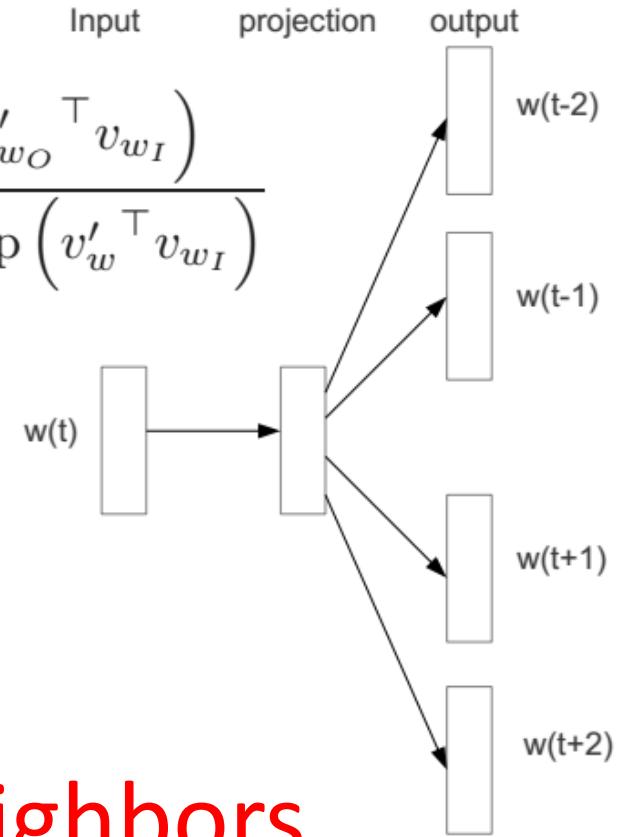
Loss function

$$\frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(w_{t+j} | w_t)$$

Negative sampling

$$\log \sigma({v'_{w_O}}^\top v_{w_I}) + \sum_{i=1}^k \mathbb{E}_{w_i \sim P_n(w)} [\log \sigma(-{v'_{w_i}}^\top v_{w_I})]$$

$$p(w_O | w_I) = \frac{\exp({v'_{w_O}}^\top v_{w_I})}{\sum_{w=1}^W \exp({v'_{w}}^\top v_{w_I})}$$



Predicting neighbors

DeepWalk

Algorithm 1 DEEPWALK(G, w, d, γ, t)

Input: graph $G(V, E)$

window size w

embedding size d

walks per vertex γ

walk length t

Considered as #epochs

Embedding matrix

Output: matrix of vertex representations $\Phi \in \mathbb{R}^{|V| \times d}$

1: Initialization: Sample Φ from $\mathcal{U}^{|V| \times d}$

2: Build a binary Tree T from V For Hierarchical Softmax

Iterate over each epoch >

3: **for** $i = 0$ to γ **do**

4: $\mathcal{O} = \text{Shuffle}(V)$

5: **for each** $v_i \in \mathcal{O}$ **do**

6: $\mathcal{W}_{v_i} = \text{RandomWalk}(G, v_i, t)$

7: SkipGram($\Phi, \mathcal{W}_{v_i}, w$) Finding neighbours of each node

8: **end for**

9: **end for**

Neighbour nodes Window size

Node2Vec

Grover, Aditya, and Jure Leskovec. "node2vec: Scalable feature learning for networks." *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. 2016.

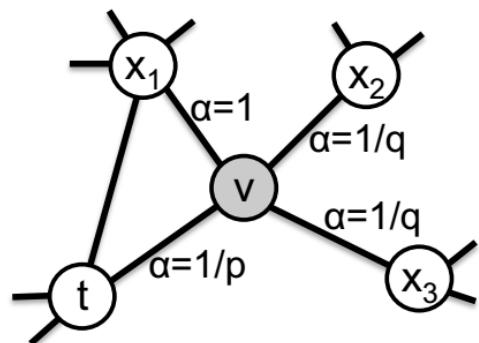
Similar to DeepWalk in using Skip-gram model for unsupervised learning.

Only modifies the search for neighboring nodes that balance between BFS and DFS.

Defines edge embedding based on node embedding

- Can solve link prediction problem

2nd order Random Walk



Consider random walk that just travelled edge (t, v) . The walk will decide which is the next node x that it should go from v by computing π_{vx}

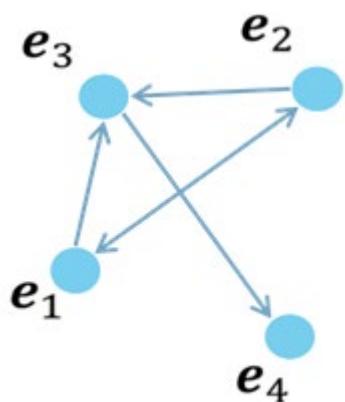
$$\pi_{vx} = \alpha_{pq}(t, x) \cdot w_{vx}$$

$$\alpha_{pq}(t, x) = \begin{cases} \frac{1}{p} & \text{if } d_{tx} = 0 \\ 1 & \text{if } d_{tx} = 1 \\ \frac{1}{q} & \text{if } d_{tx} = 2 \end{cases}$$

p and q are hyper-parameters

Graph convolutional nets

Data graph



GCN update rule, vector form

$$h_{v_i}^{(l+1)} = \sigma \left(\sum_j \frac{1}{c_{ij}} h_{v_j}^{(l)} W^{(l)} \right)$$

GCN update rule, matrix form

$$f(H^{(l)}, A) = \sigma \left(\hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right)$$

Do, Kien, Truyen Tran, and Svetha Venkatesh. "Learning deep matrix representations." *arXiv preprint arXiv:1703.01454* (2017).

Graph attention

Veličković, Petar, et al. "Graph Attention Networks." *International Conference on Learning Representations*. 2018.

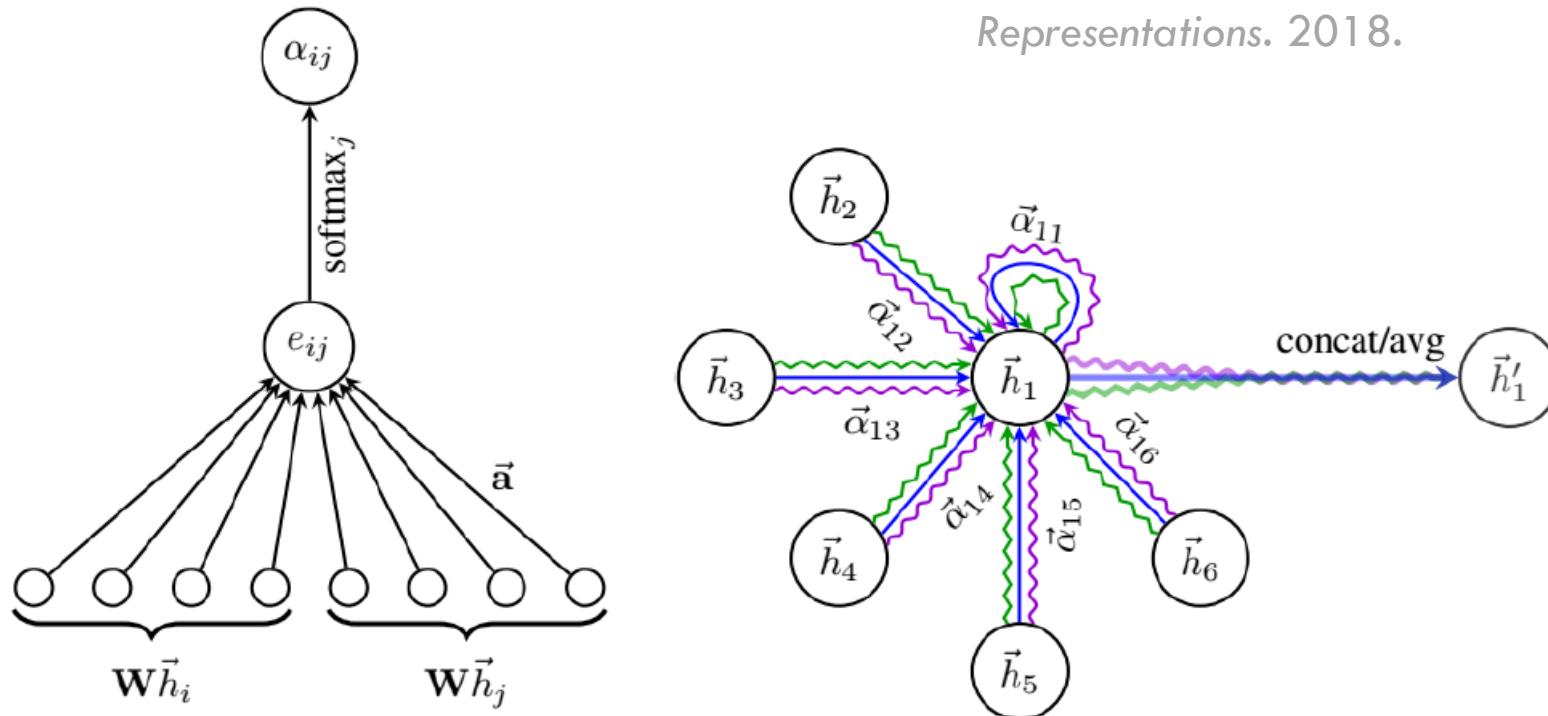
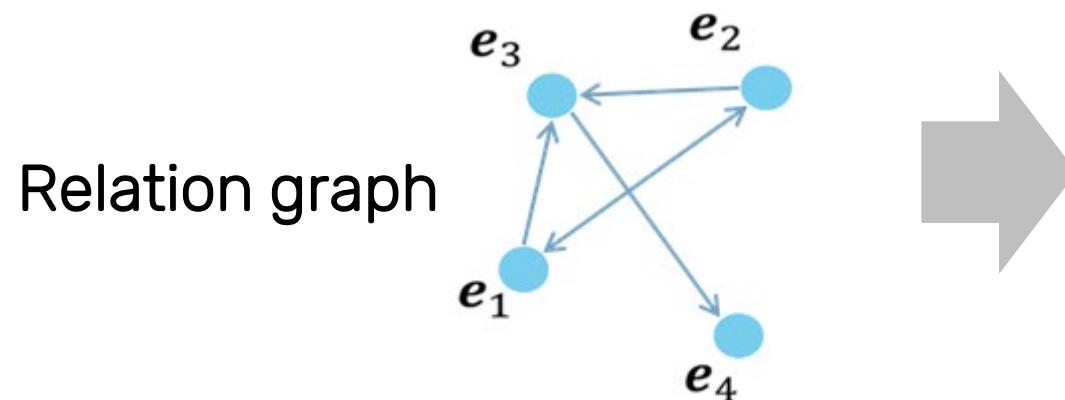


Figure 1: **Left:** The attention mechanism $a(\mathbf{W}\vec{h}_i, \mathbf{W}\vec{h}_j)$ employed by our model, parametrized by a weight vector $\vec{\mathbf{a}} \in \mathbb{R}^{2F'}$. **Right:** An illustration of multi-head attention (with $K = 3$ heads) by node 1 on its neighborhood. Different arrow styles and colors denote independent attention computations. The aggregated features from each head are concatenated or average to obtain \vec{h}'_1 .

Message passing GNN

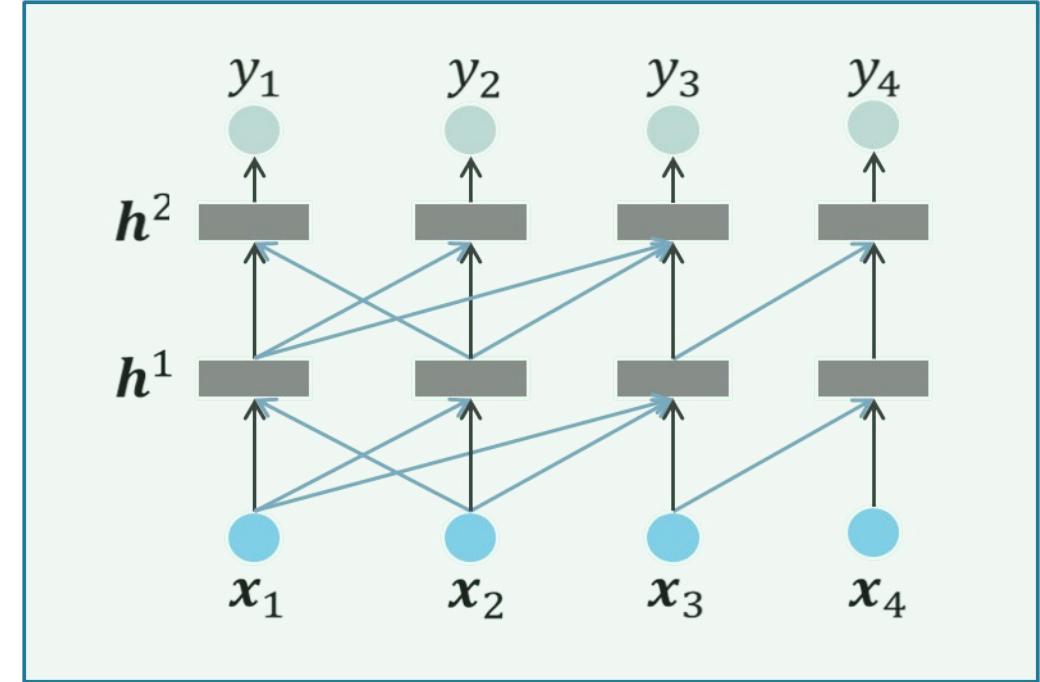


Collecting messages

$$m_i^{(k+1)} = \sum_{v_j \in N(v_i)} M_k \left(h_i^{(k)}, h_j^{(k)}, e_{ij} \right)$$

#REF: Pham, Trang, et al. "Column Networks for Collective Classification." AAAI. 2017.

17/09/2021



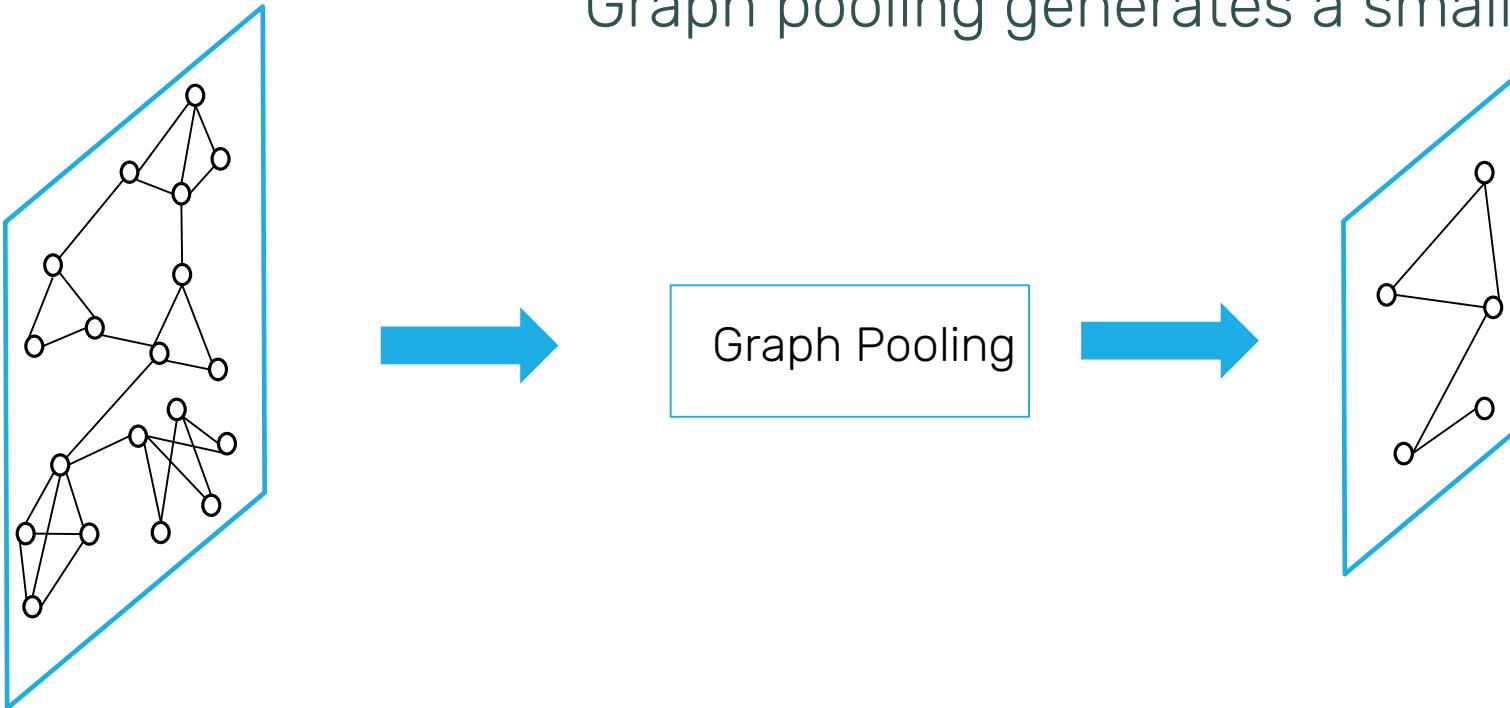
Refining node embedding

$$h_i^{(k+1)} = U_k \left(h_i^{(k)}, m_i^{(k+1)} \right)$$

Gilmer, Justin, et al. "Neural message passing for quantum chemistry." *arXiv preprint arXiv:1704.01212* (2017) 31

Graph pooling & readout

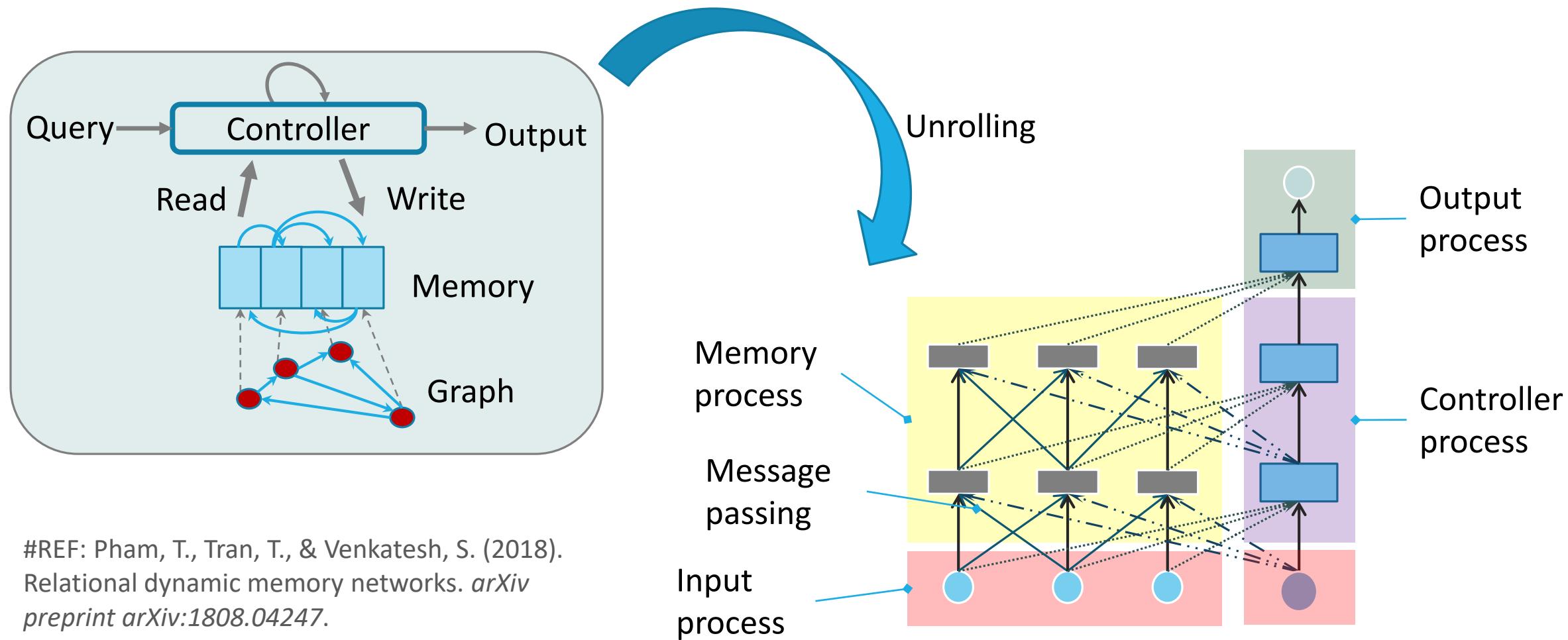
Graph pooling generates a smaller graph



$$\mathbf{A} \in \{0, 1\}^{n \times n}, \mathbf{X} \in \mathbb{R}^{n \times d}$$

$$\mathbf{A}_p \in \{0, 1\}^{n_p \times n_p}, \mathbf{X}_p \in \mathbb{R}^{n_p \times d_{new}}, n_p < n$$

RDMN: A graph processing machine



Self-supervised representation learning for drugs

Early works: DeepWalk, Node2vec

BERT-like through masking and reconstruction of parts:

- SMILES sequence: ChemBERTa
- Molecule graph: GROVER

Contrastive learning (local manifold smoothness):

- Graph contrastive learning: GCC

BERT-like self-supervised learning

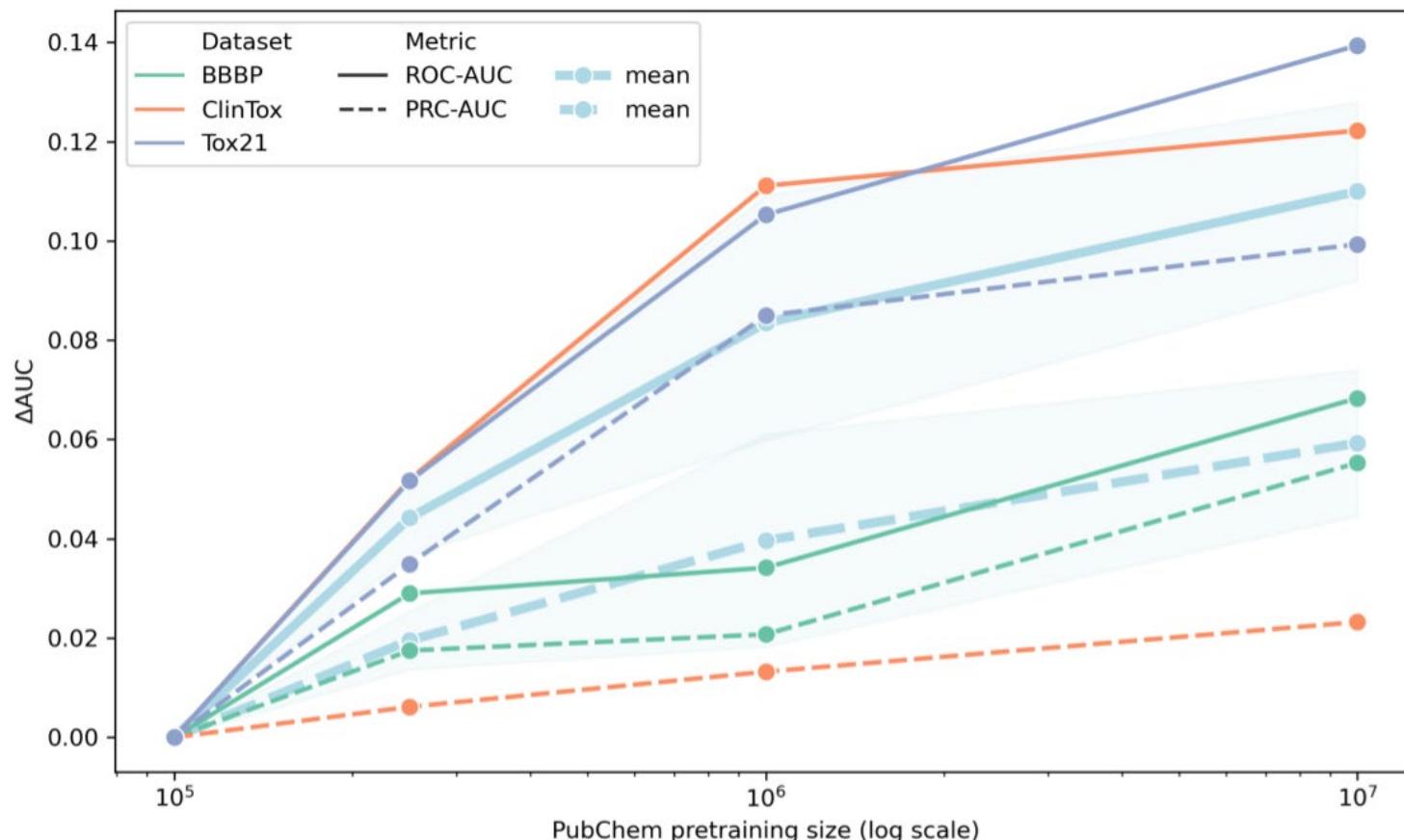
ChemBERTa (Chithrananda, S et.al) uses pretraining procedure from RoBERTa with 10M unique SMILES from PubChem.

	BBBP 2,039		ClinTox (CT_TOX) 1,478		HIV 41,127		Tox21 (SR-p53) 7,831	
	ROC	PRC	ROC	PRC	ROC	PRC	ROC	PRC
ChemBERTa 10M	0.643	0.620	0.733	0.975	0.622	0.119	0.728	0.207
D-MPNN	0.708	0.697	0.906	0.993	0.752	0.152	0.688	0.429
RF	0.681	0.692	0.693	0.968	0.780	0.383	0.724	0.335
SVM	0.702	0.724	0.833	0.986	0.763	0.364	0.708	0.345

#REF: Chithrananda, S., Grand, G., & Ramsundar, B. (2020). Chemberta: Large-scale self-supervised pretraining for molecular property prediction. Machine Learning for Molecules Workshop NeurIPS 2020

BERT-like self-supervised learning

ChemBERTa (Chithrananda, S et.al) uses pretraining procedure from RoBERTa with 10M unique SMILES from PubChem.



#REF: Chithrananda, S., Grand, G., & Ramsundar, B. (2020). Chemberta: Large-scale self-supervised pretraining for molecular property prediction. Machine Learning for Molecules Workshop NeurIPS 2020

Sub-graph masking self-supervised learning

GROVER (Rong et.al) pretrains with contextual property prediction and graph level motif prediction using GNNTransformer architecture

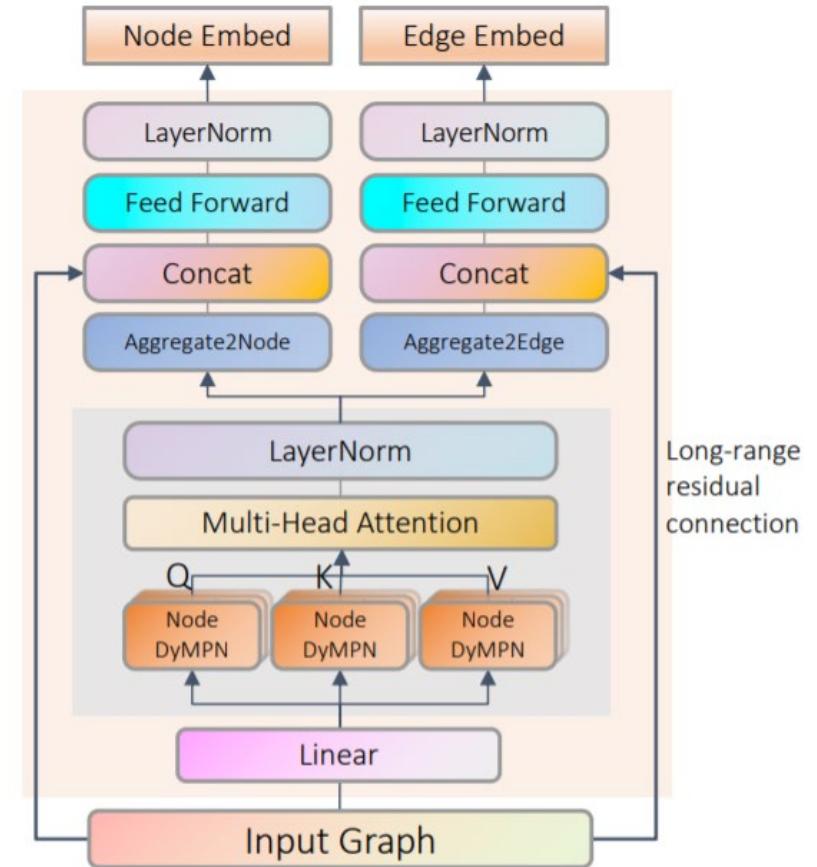
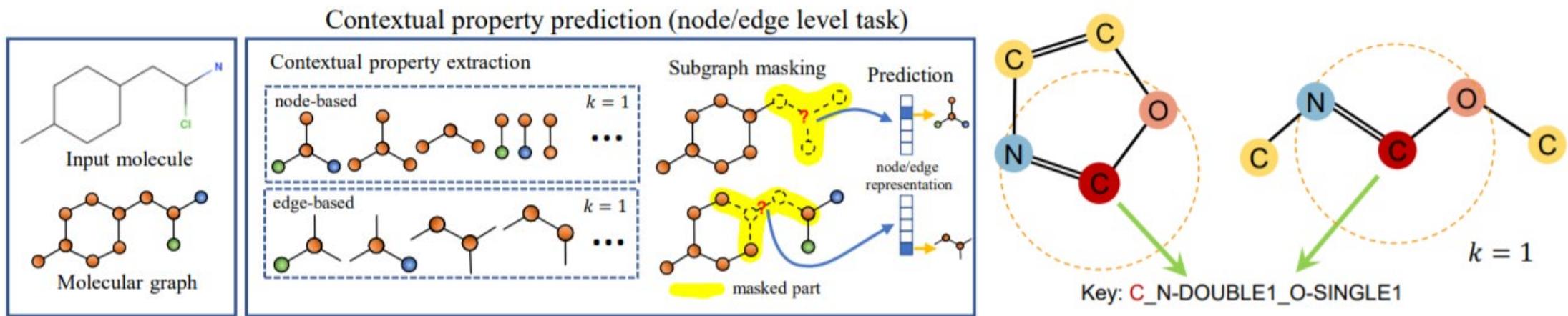


Figure 1: Overview of GTransformer.

#REF: Rong, Y., Bian, Y., Xu, T., Xie, W., Wei, Y., Huang, W., & Huang, J. (2020). Self-supervised graph transformer on large-scale molecular data

Sub-graph masking self-supervised learning

GROVER (Rong et.al) pretrains with contextual property prediction and graph level motif prediction using GNNTransformer architecture

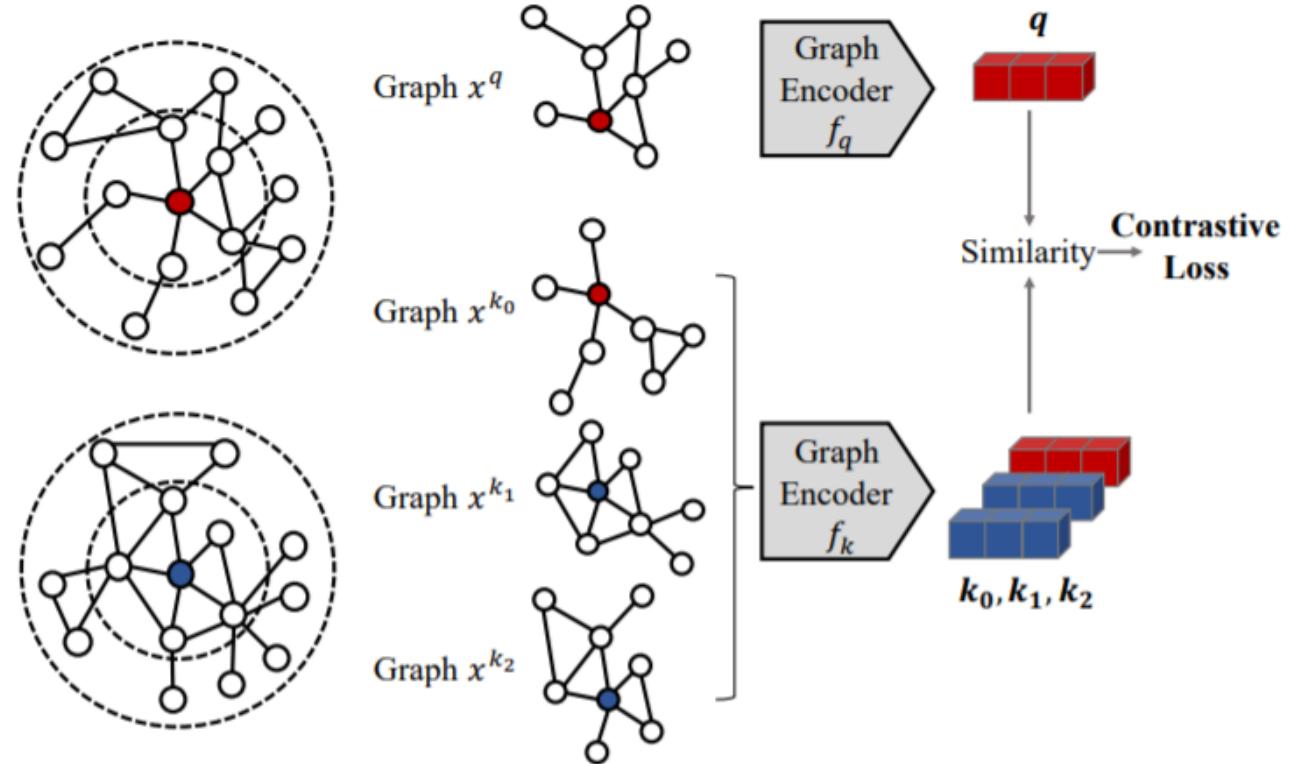


#REF: Rong, Y., Bian, Y., Xu, T., Xie, W., Wei, Y., Huang, W., & Huang, J. (2020). Self-supervised graph transformer on large-scale molecular data

Contrastive learning self-supervised

GCC (Qiu, Jiezhong, et al.) use subgraph instance discrimination (SID) as pretraining task and InfoNCE as training objective.

SID performs two random walks with restart to sampling from a k-neighbour subgraph to use as positive pair and samples from other k-neighbour subgraph as negative pair.



#REF: Qiu, Jiezhong, et al. "GCC: Graph contrastive coding for graph neural network pre-training." *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2020.

GCC (Qiu, Jiezhong, et al)

Table 2: Node classification.

Datasets	US-Airport	H-index
$ V $	1,190	5,000
$ E $	13,599	44,020
ProNE	62.3	69.1
GraphWave	60.2	70.3
Struc2vec	66.2	> 1 Day
GCC (E2E, freeze)	64.8	78.3
GCC (MoCo, freeze)	65.6	75.2
GCC (rand, full)	64.2	76.9
GCC (E2E, full)	68.3	80.5
GCC (MoCo, full)	67.2	80.6

Table 3: Graph classification.

Datasets	IMDB-B	IMDB-M	COLLAB	RDT-B	RDT-M
# graphs	1,000	1,500	5,000	2,000	5,000
# classes	2	3	3	2	5
Avg. # nodes	19.8	13.0	74.5	429.6	508.5
DGK	67.0	44.6	73.1	78.0	41.3
graph2vec	71.1	50.4	–	75.8	47.9
InfoGraph	73.0	49.7	–	82.5	53.5
GCC (E2E, freeze)	71.7	49.3	74.7	87.5	52.6
GCC (MoCo, freeze)	72.0	49.4	78.9	89.8	53.7
DGCNN	70.0	47.8	73.7	–	–
GIN	75.6	51.5	80.2	89.4	54.5
GCC (rand, full)	75.6	50.9	79.4	87.8	52.1
GCC (E2E, full)	70.8	48.5	79.0	86.4	47.4
GCC (MoCo, full)	73.8	50.3	81.1	87.6	53.0

#REF: Qiu, Jiezhong, et al. "GCC: Graph contrastive coding for graph neural network pre-training." *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2020.

Agenda

Molecular representation learning

- Fingerprints
- String representation
- Graph representation
- Self-supervised learning

Molecular property prediction

- Approximating quantum chemistry computation
- Graph regression and classification
- Graph multitask learning
- Explanation
- Data efficient learning

The three basic questions in drug discovery

Given a molecule, is this drug? Aka properties/targets/effects prediction.

- Drug-likeness
- Targets it can modulate and how much
- Its dynamics/kinetics/effects/metabolism if administered orally or via injection

Given a target, what are molecules?

- If the list of molecules is given, pick the good one. If evaluation is expensive, need to search, e.g., using Bayesian Optimization.
- If no molecule is found, need to generate from scratch → generative models + Bayesian Optimization, or Reinforcement Learning.
- How does the drug-like space look like?

Given a molecular graph, what are the steps to make the molecule?

- Synthetic tractability
- Reaction planning, or retrosynthesis

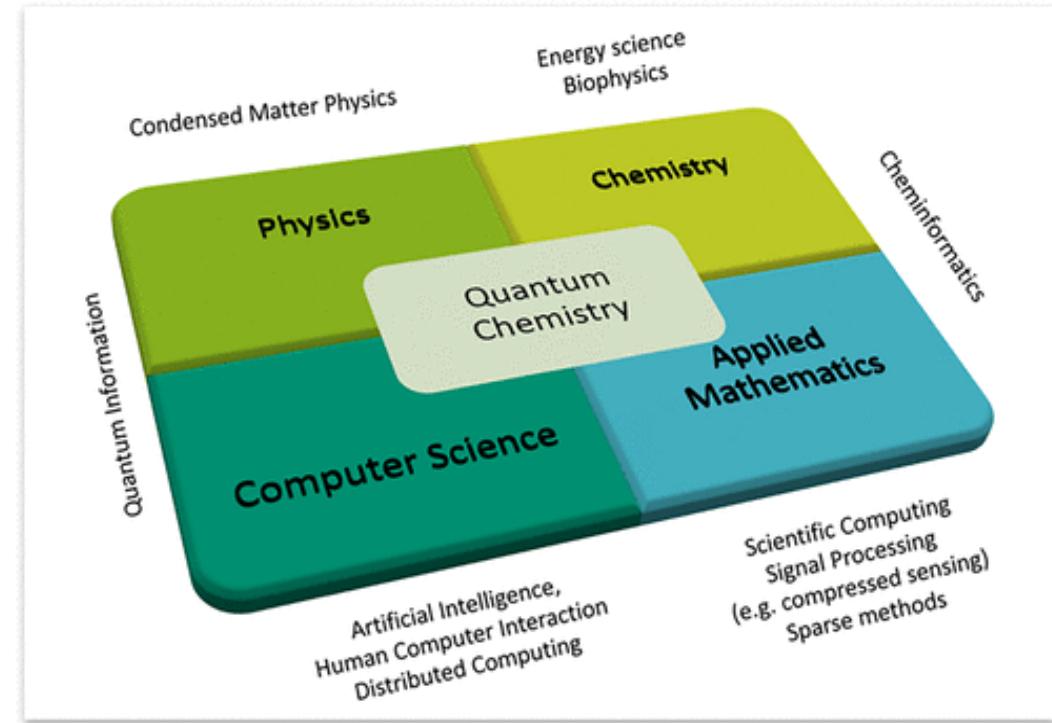
Quantum chemistry

In chemistry we mostly need to know about electronic structure (e.g., electron density and electronic energy).

The density can be inferred from the wave function. But solving wave equation is very difficult.

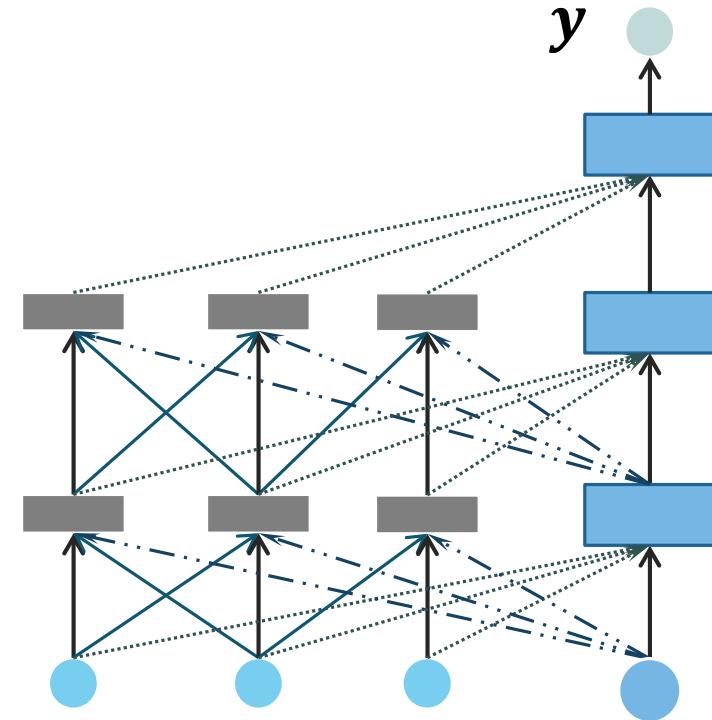
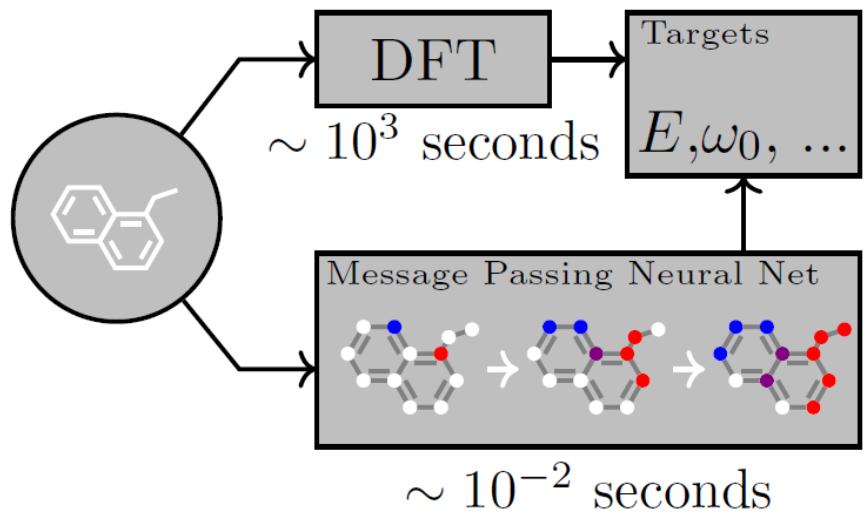
Density Functional Theory (DFT): electron density is a function of space and time.

- Hohenberg-Kohn theorem: the density of any system determines its ground-state properties.
- Electron density functional → total energy of our system.



Aspuru-Guzik, Alán, Roland Lindh, and Markus Reiher. "The matter simulation (r) evolution." ACS central science 4.2 (2018): 144-152.

Approximating DFT



Gilmer, Justin, et al. "Neural message passing for quantum chemistry." *arXiv preprint arXiv:1704.01212* (2017).

Molecule property prediction

- A fundamental task in many stages of drug discovery
 - E.g., virtual screening and lead optimisation
- Molecule properties
 - Binding affinity
 - Have effects on cell expression
 - Toxicity, ADME property (Absorption, Distribution, Metabolism and Excretion)
 - Interacting with other molecules
 - Easy to synthesize

Molecular activity prediction

Collecting messages

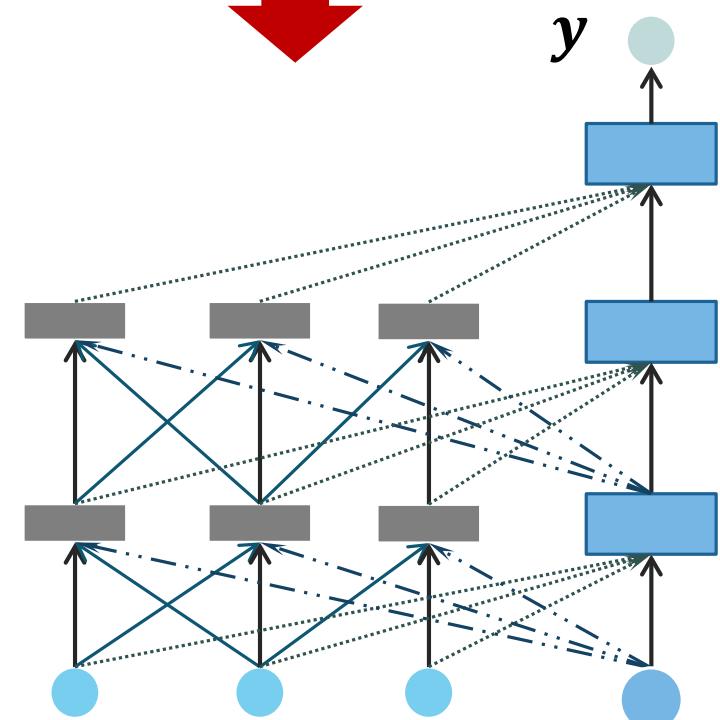
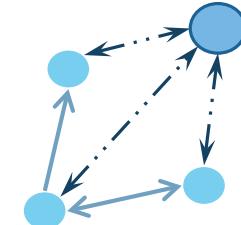
$$m_i^{(k+1)} = \sum_{v_j \in N(v_i)} M_k \left(h_i^{(k)}, h_j^{(k)}, e_{ij} \right)$$

Refining node embedding

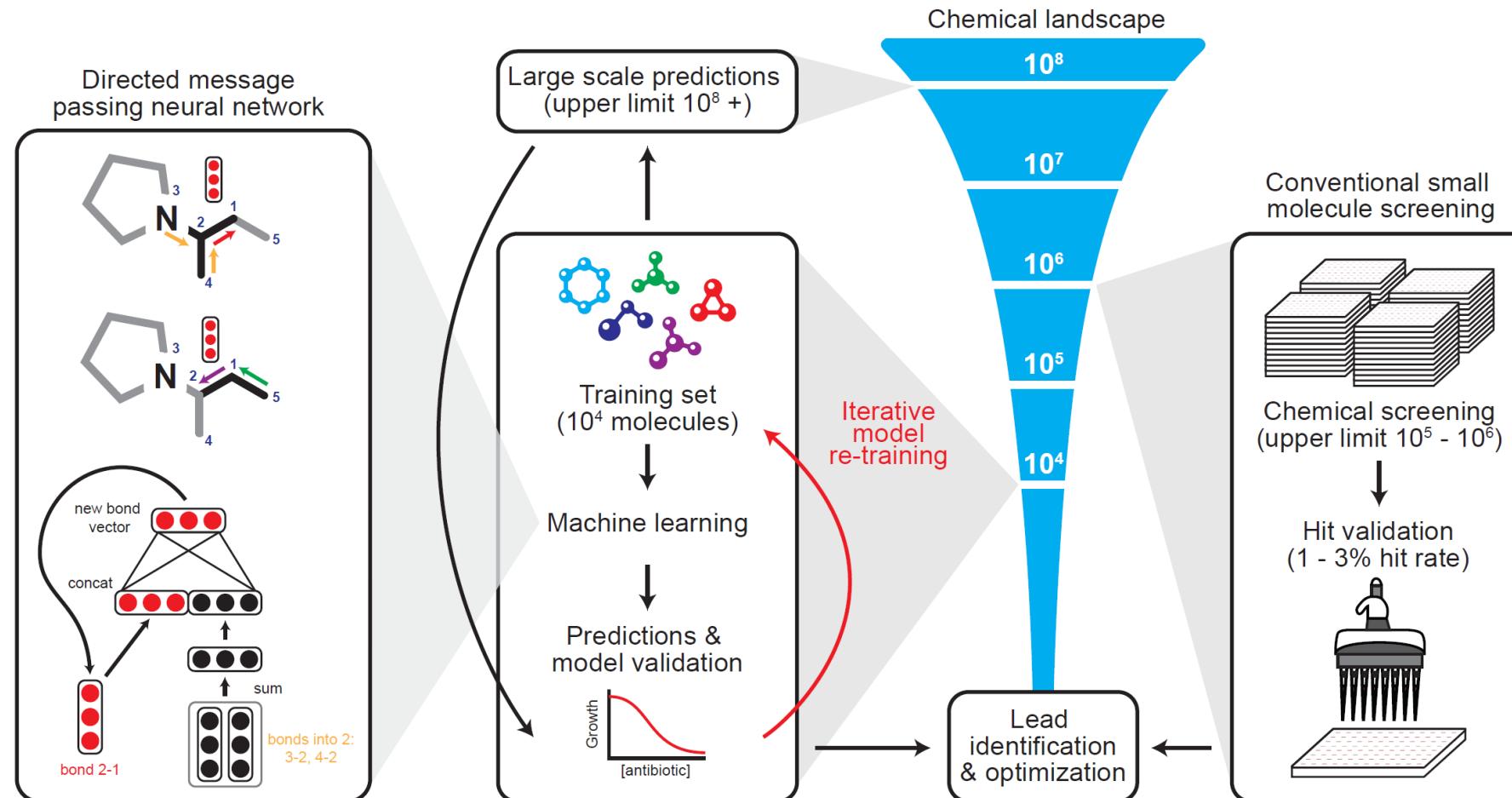
$$h_i^{(k+1)} = U_k \left(h_i^{(k)}, m_i^{(k+1)} \right)$$

#Ref: Pham, Trang, Truyen Tran, and Svetha Venkatesh. "Graph Memory Networks for Molecular Activity Prediction." ICPR'18.

Activity as query

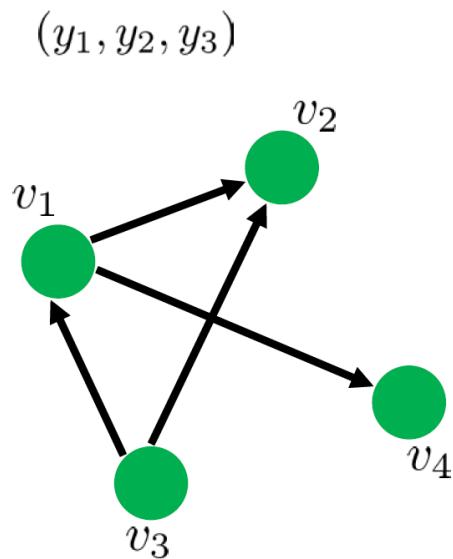


Example: Antibiotic discovery of halicin



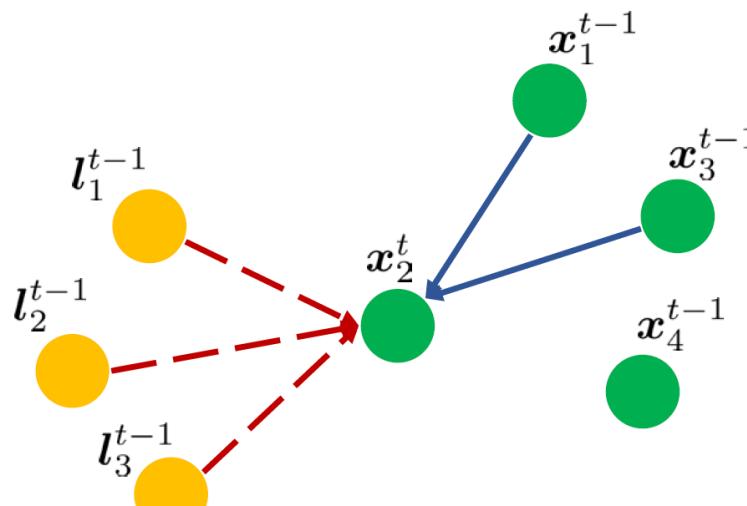
GAML for drug multi-target prediction

Scale linearly with number of targets + efficient processing through message factoring.

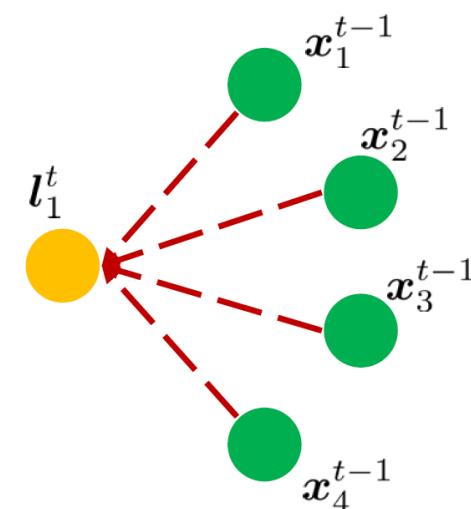


(a) A input graph with 4 nodes and 3 labels

#REF: Do, Kien, et al. "Attentional Multilabel Learning over Graphs-A message passing approach." *Machine Learning*, 2019.



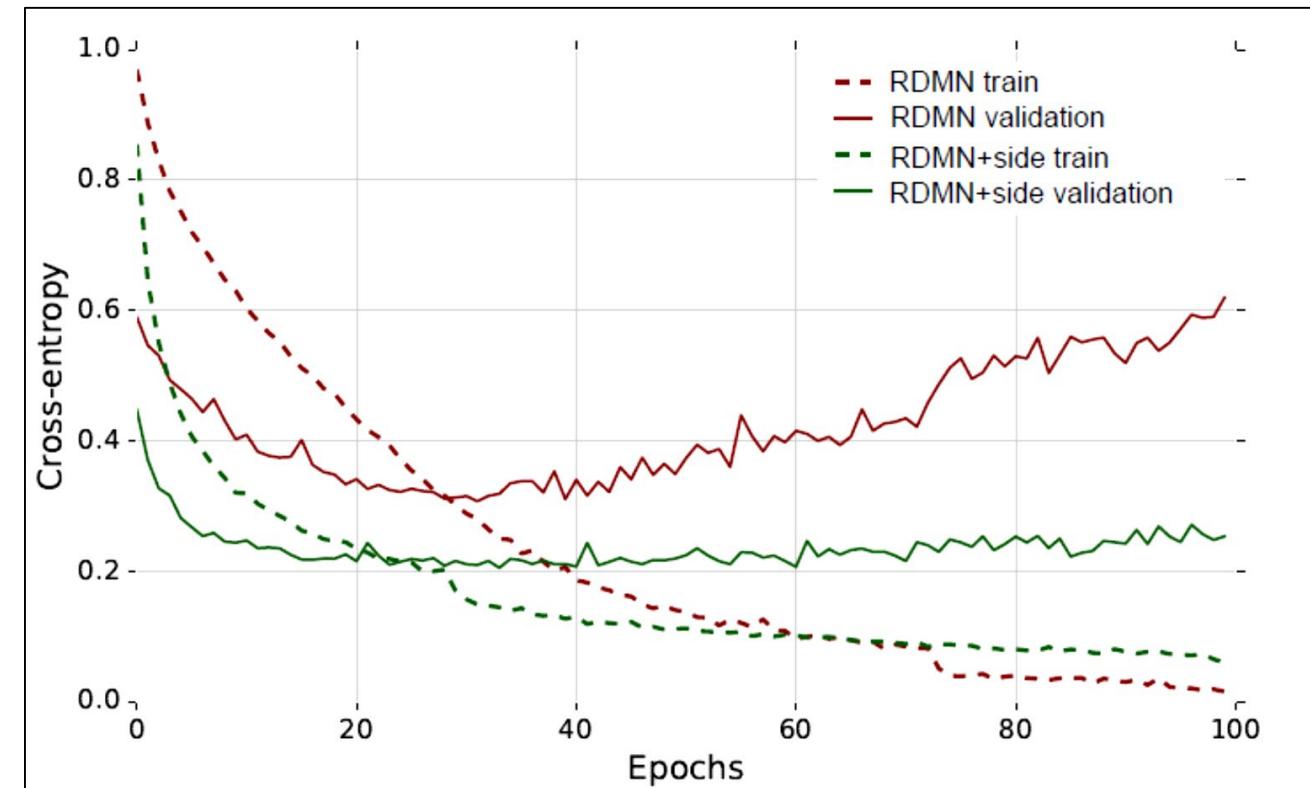
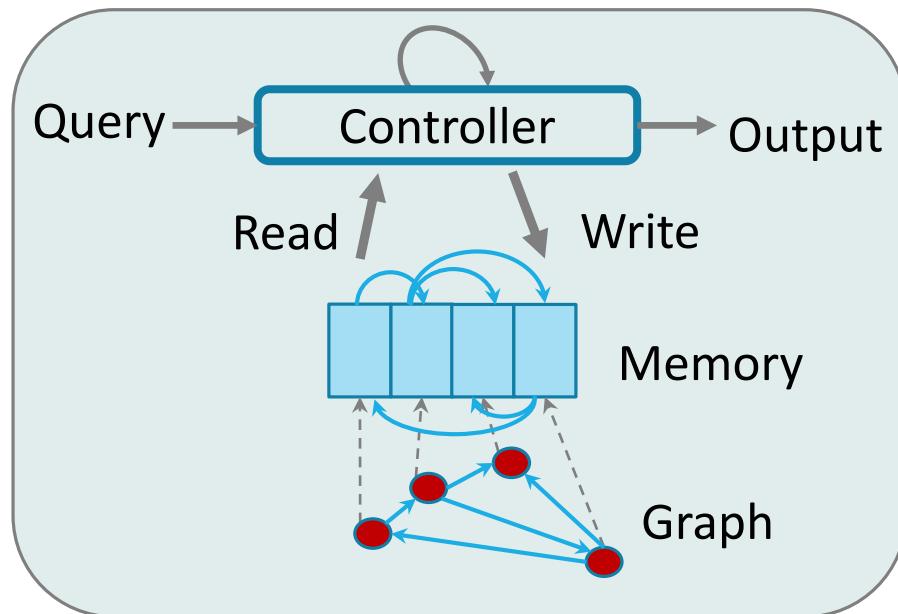
(b) Input node update



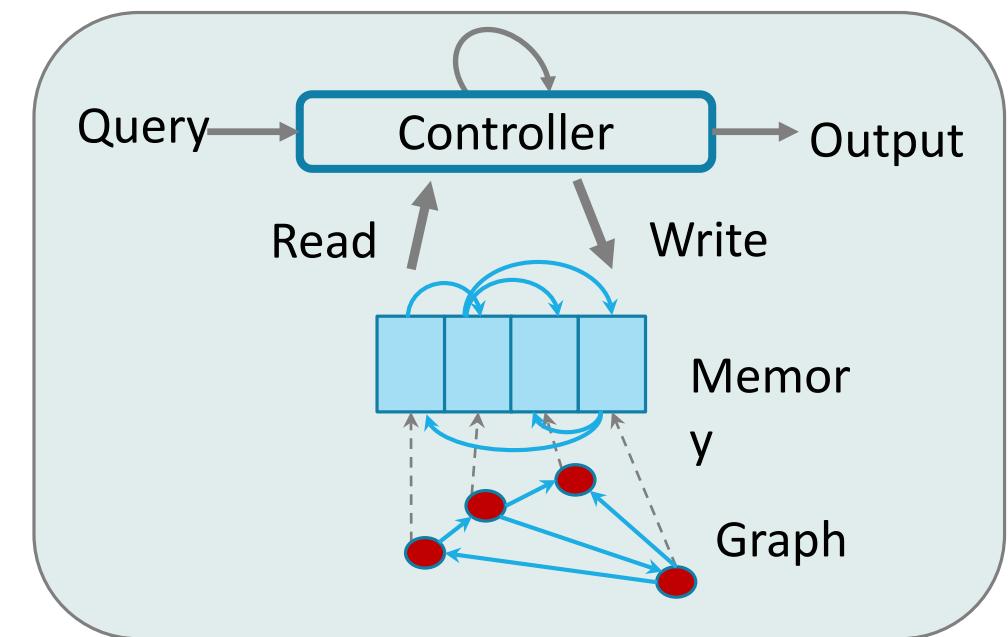
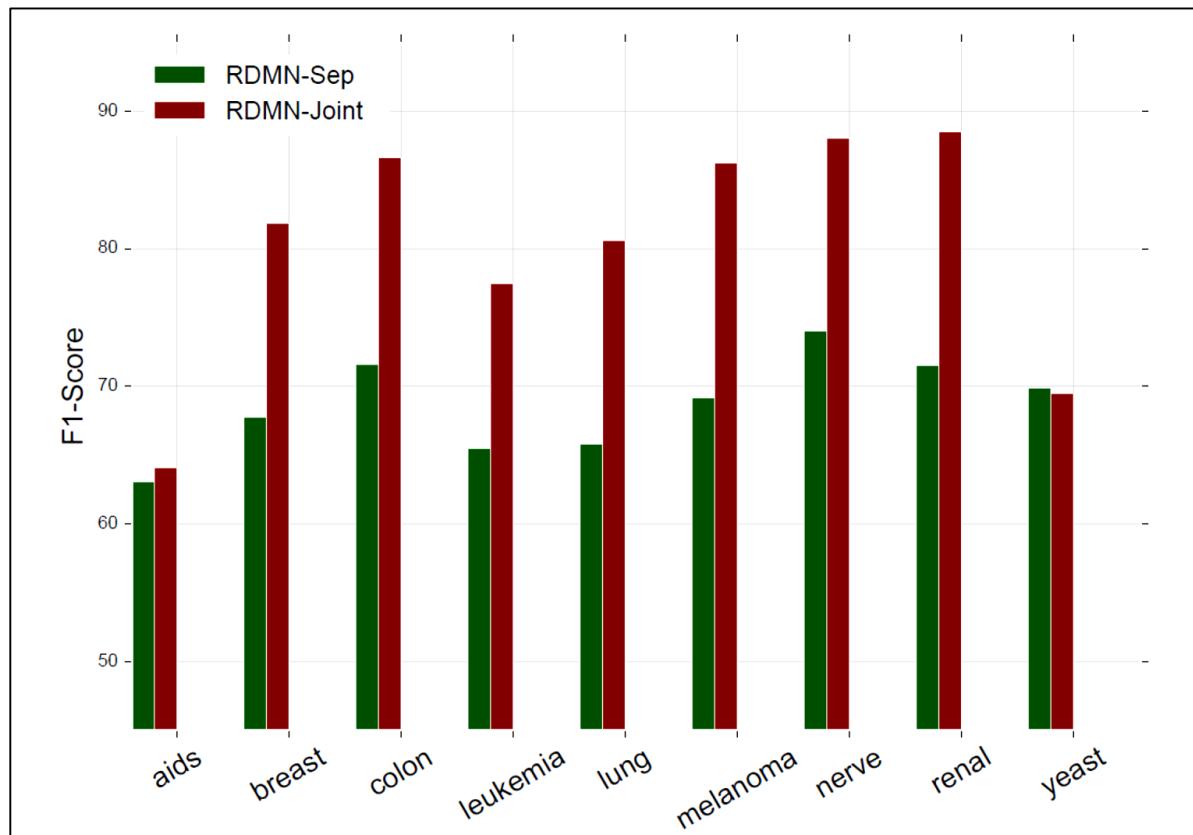
(c) Label node update

More flexible drug-disease response with RDMN

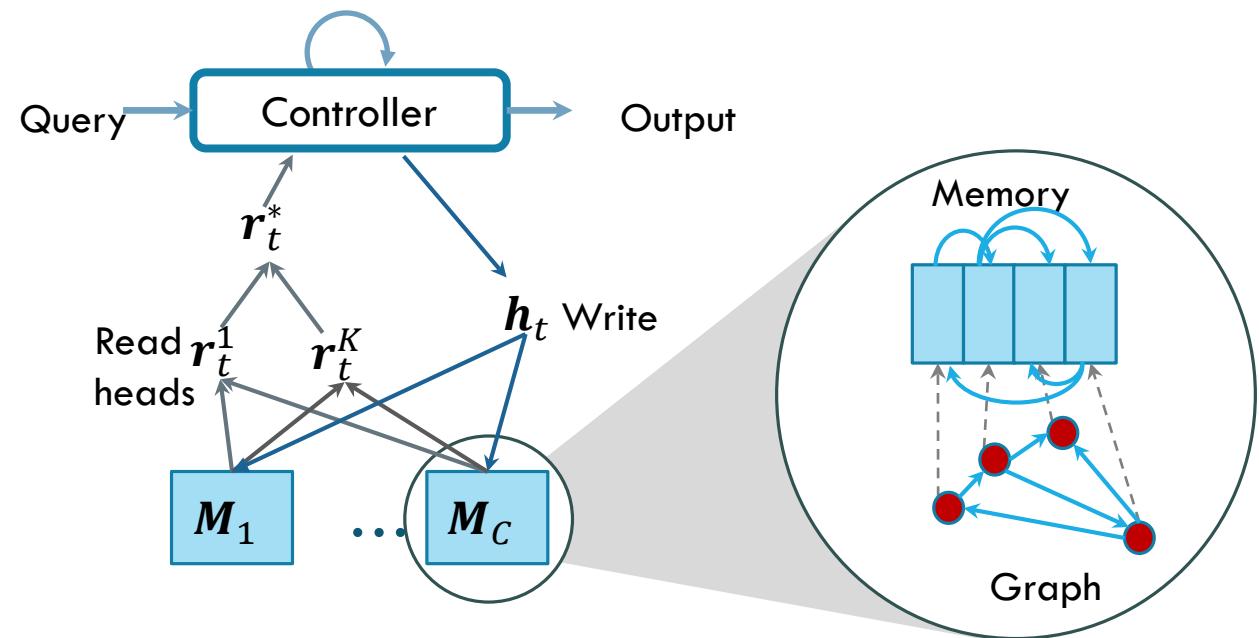
Model	MicroF1	MacroF1	Average AUC
SVM	66.4	67.9	85.1
RF	65.6	66.4	84.7
GB	65.8	66.9	83.7
NeuralFP [19]	68.2	67.6	85.9
MT-NN [51]	75.5	78.6	90.4
RDMN	77.8	80.3	92.1



Tying param helps multiple diseases response with RDMN



Drug-drug interaction as graph-graph reasoning



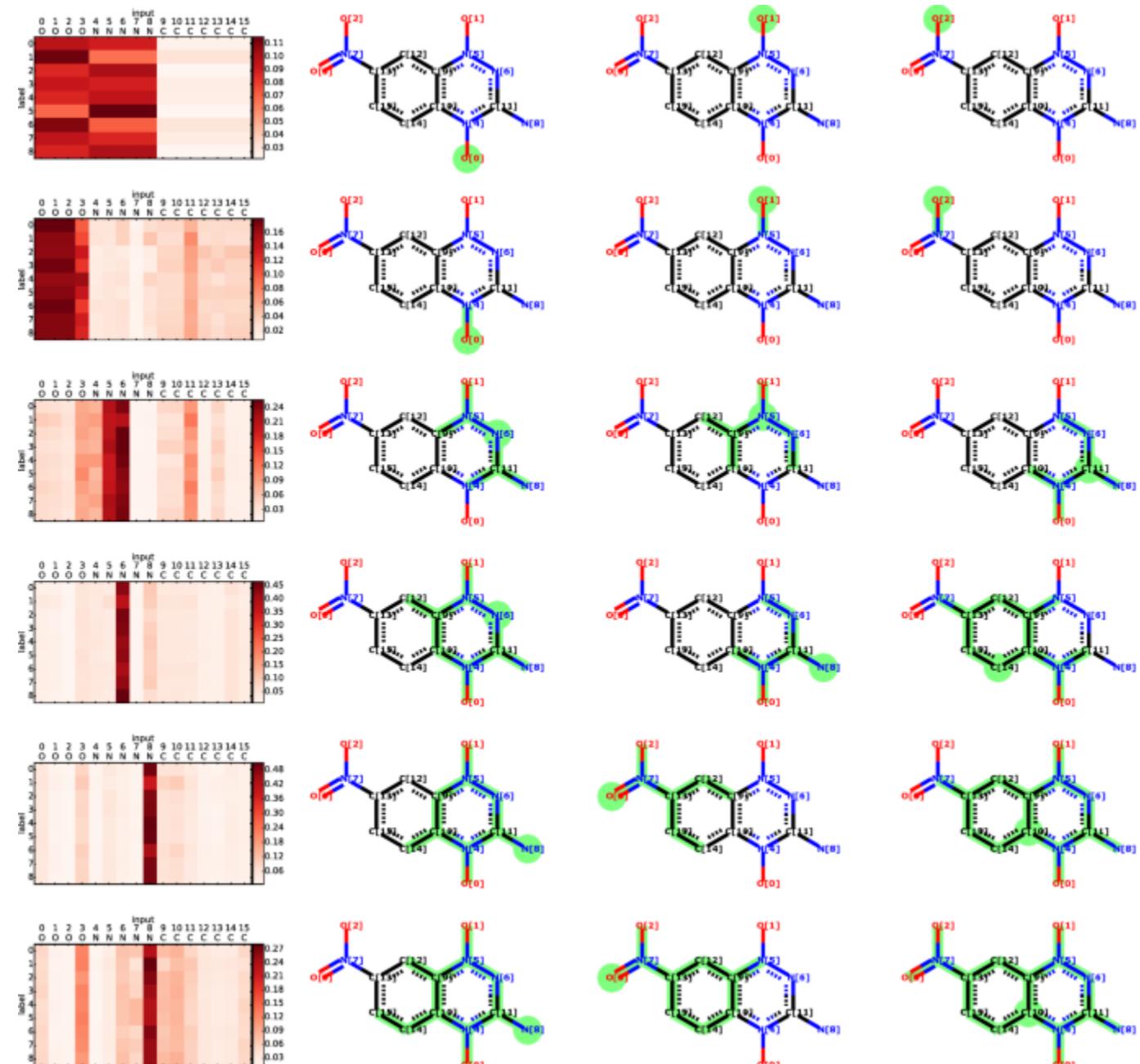
#REF: Pham, Trang, Truyen Tran, and Svetha Venkatesh. "Relational dynamic memory networks." *arXiv preprint arXiv:1808.04247*(2018).

Attention-based explanation

Internal attention can provide certain capability of explanation

External counter-factual methods can be more precise by generating small changes to the molecules and see the effect.

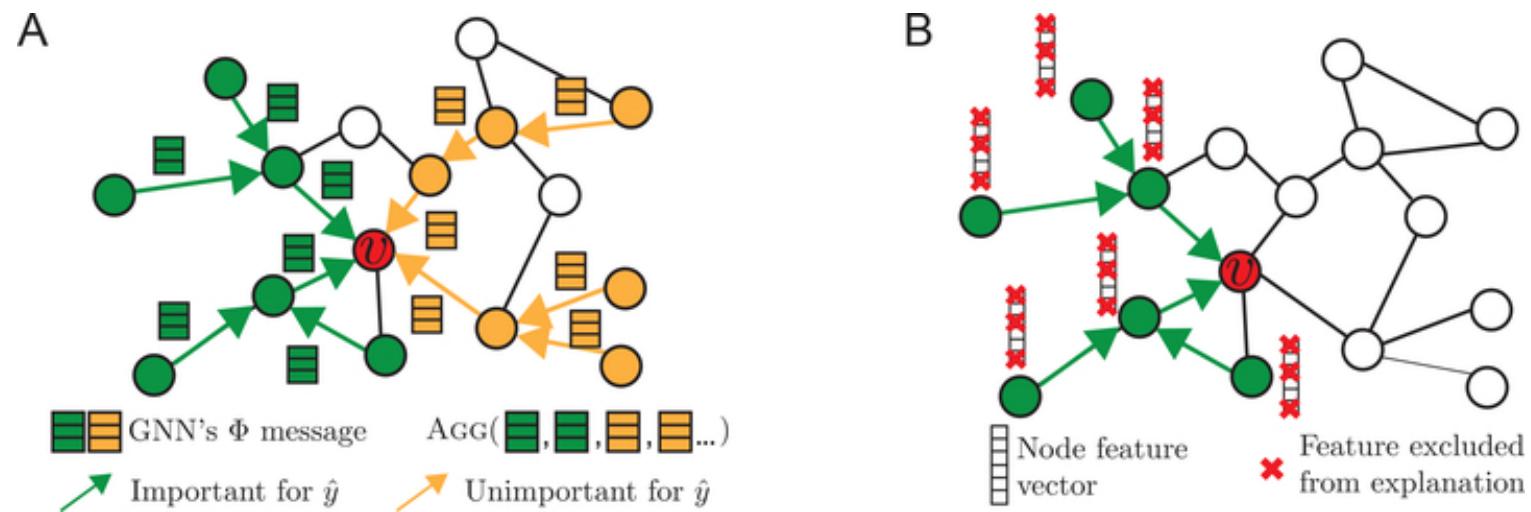
More on this later.



#REF: Do, Kien, et al. "Attentional Multilabel Learning over Graphs-A message passing approach." *Machine Learning*, 2019.

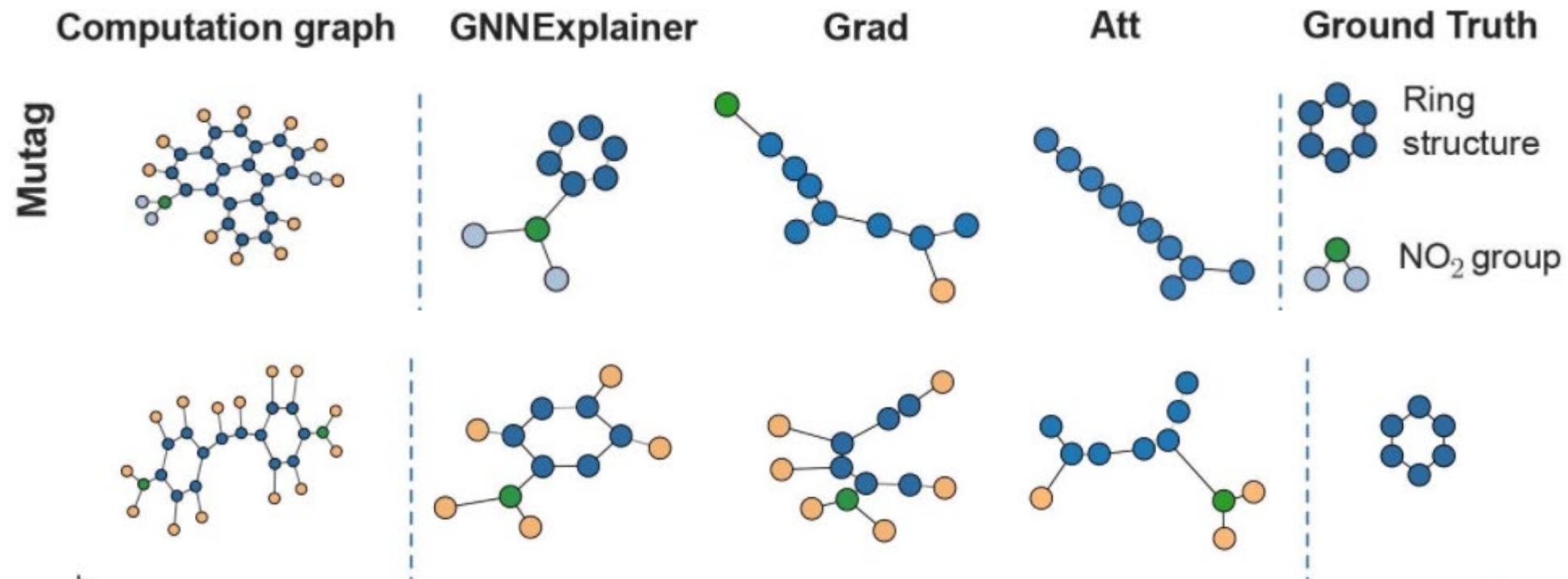
GNNExplainer

GNNExplainer explains DL model by providing a small subgraph of the input graph together with a small subset of node features that are most influential for the prediction.



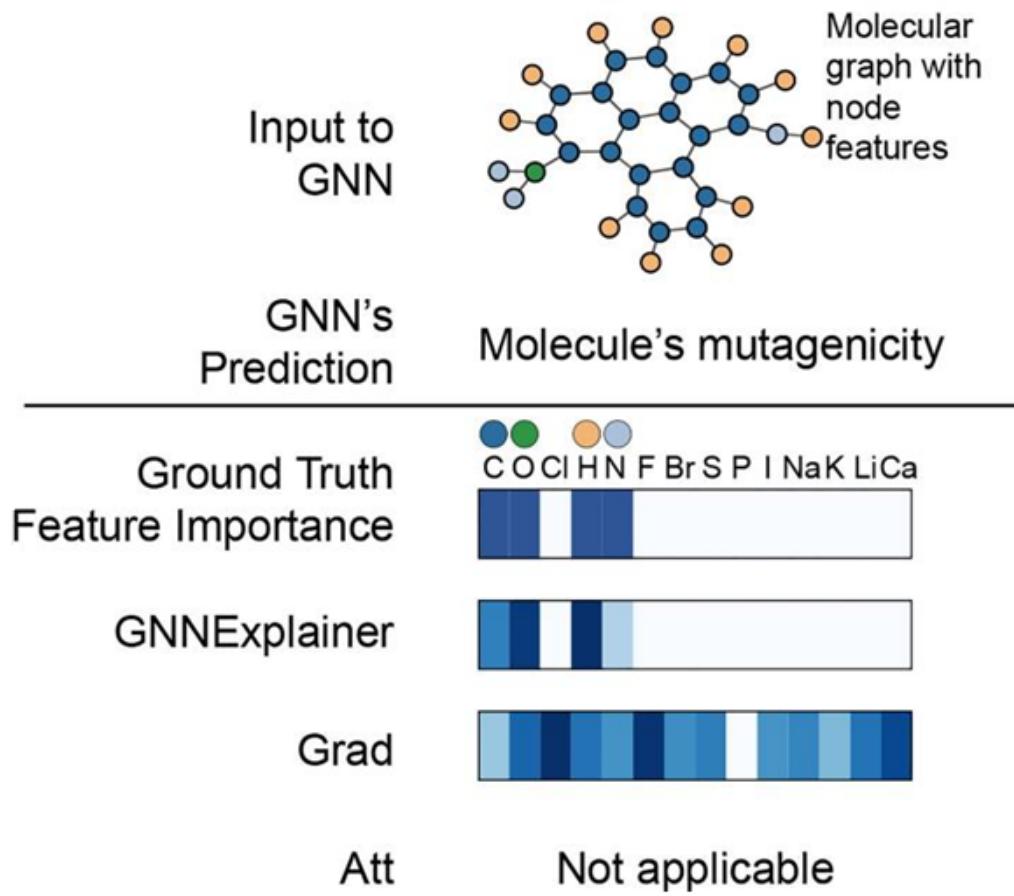
$$\max_{G_S} MI(Y, (G_S, X_S)) = H(Y) - H(Y|G = G_S, X = X_S)$$

GNNExplainer (2)



Ying, Rex, et al. "Gnnexplainer: Generating explanations for graph neural networks." *Advances in neural information processing systems* 32 (2019): 9240.

GNNExplainer (3)

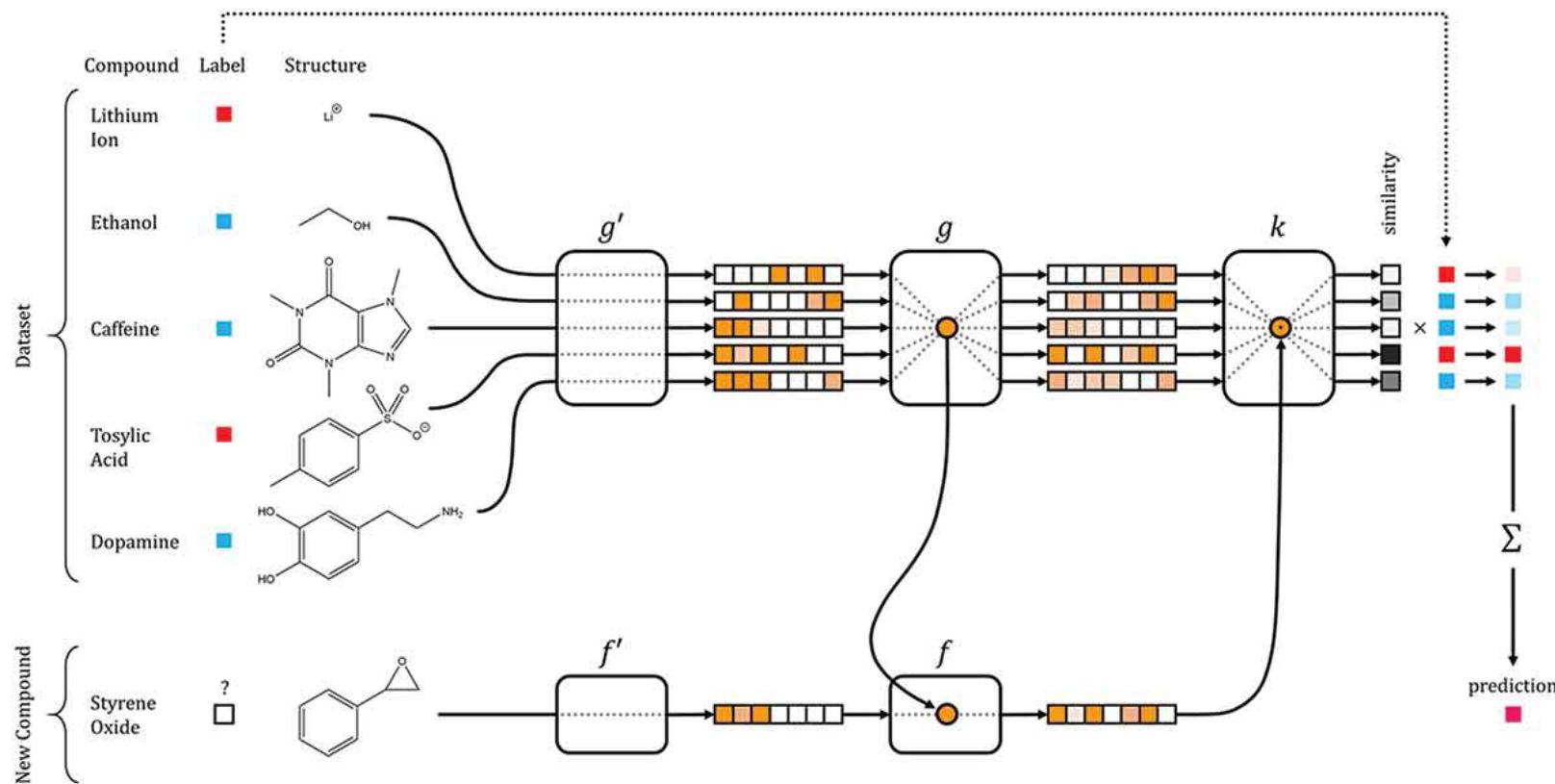


Few-shot learning for drug property prediction

Some ideas are in place:

Prototype, distance metric learning.

Joint feature learning of many small tasks, then fine-tuning on new task.



Altae-Tran, Han, et al. "Low data drug discovery with one-shot learning." ACS central science 3.4 (2017): 283-293.

Tutorial at ECML-PKDD, September 17th /18th 2021



Machine Learning and Reasoning for Drug Discovery

Part B2: Protein representation and protein-drug binding

Truyen Tran, Thin Nguyen and Tri Nguyen
{truyen.tran, thin.nguyen, minhtri}@deakin.edu.au

<https://bit.ly/3Edqgz4>

Agenda

Protein representation learning

- Embedding, BERT
- 2D contact map
- 3D structure

Drug-target binding prediction

- Multi-target prediction
- Drug-protein binding as graph-graph interaction
- Cold-start problem
- Explanation

Background on protein

Proteins are large biomolecules.

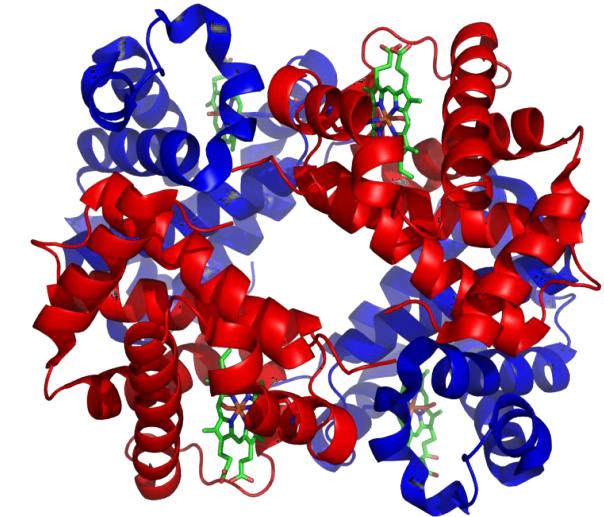
Long chains of amino acids (residue). There are 20 types of amino acids.

Residues are attracted to each other by physical and chemical forces.

Residue chain folds to form the 3D structure.

Proteins 3D structure determine their function.

Performing many different functions in the organisms such as transportation (hemoglobin), hormonal (insulin), protection (immunoglobulin), etc.



3D structure of hemoglobin

Background on protein (cont.)

- Amino acid structure
 - Amine ($-NH_2$)
 - C_α - R side chain
 - Carboxyl ($-COOH$)

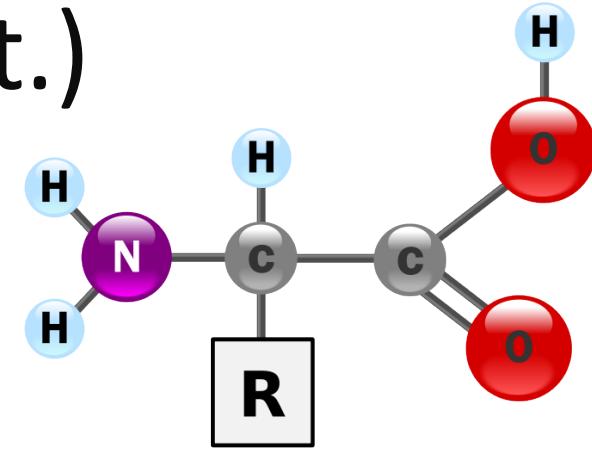
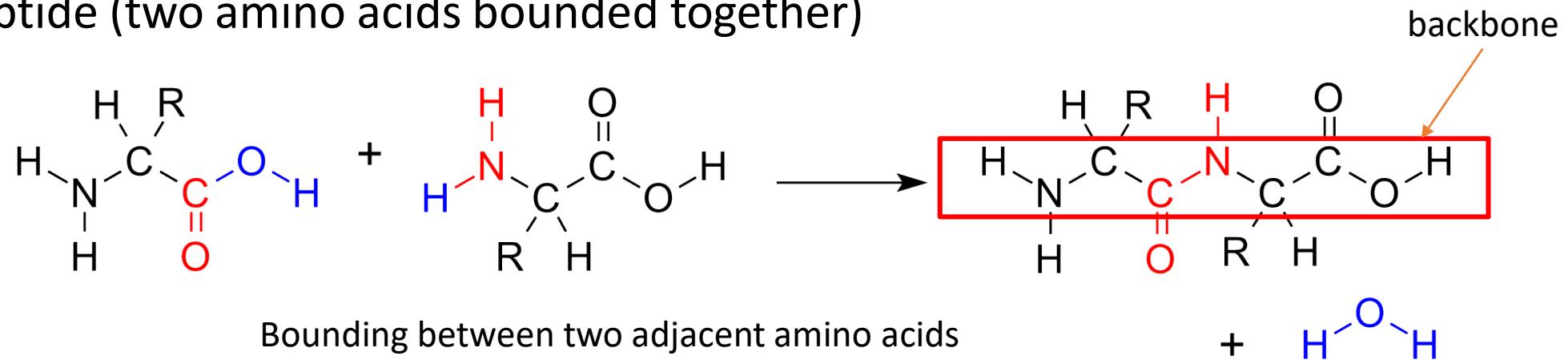


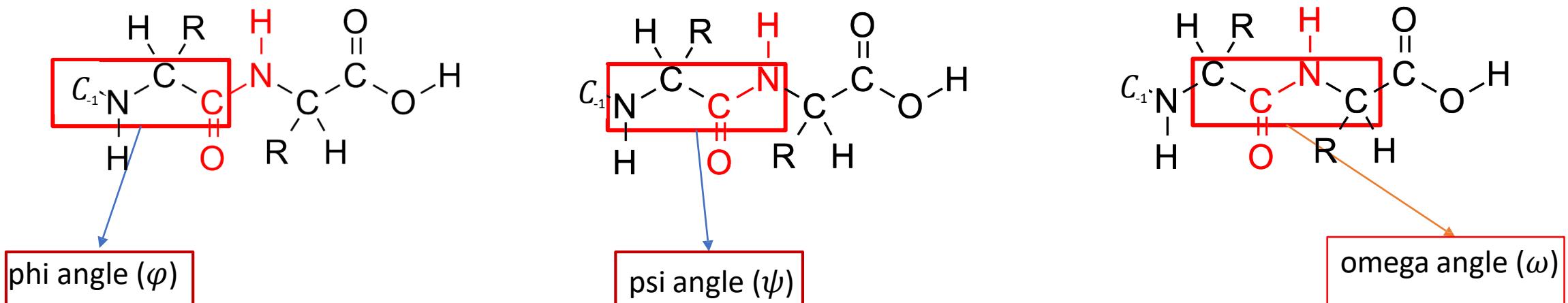
Fig 2. General structure of amino acid [2]

- Dipeptide (two amino acids bounded together)



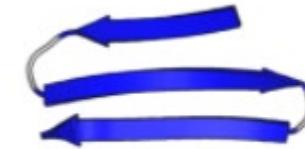
Background on protein (cont.)

- Local structure: torsion angle



Background on protein (cont.)

- Local structure: secondary structure:
 - Local folded structure due to the interaction between atoms in the backbone chain.
 - Eight types of secondary structure. α helix and the β pleated sheet are two most common secondary structure.
- Tertiary structure:
 - The overall 3D structure of protein sequence
 - R group interactions between residues also contribute to form the tertiary structure
- Quaternary structure:
 - The arrangement of multiple polypeptide chains

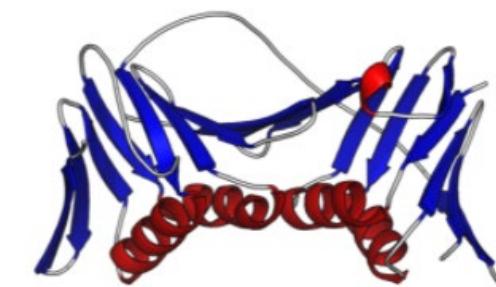


β -Sheet (3 strands)

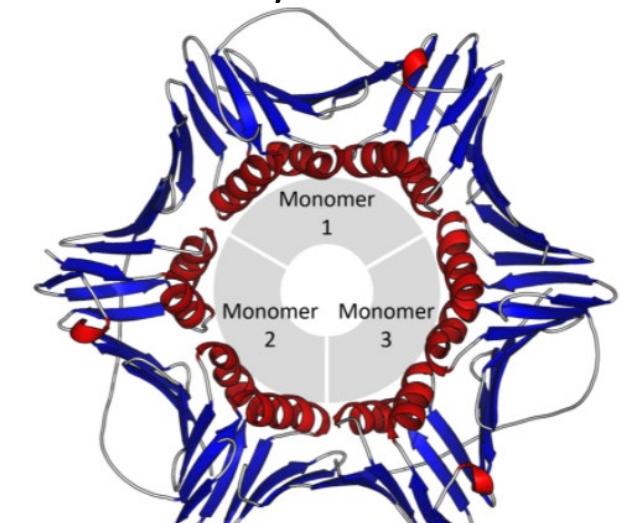


α -helix

Secondary structure



Tertiary structure



Quaternary structure

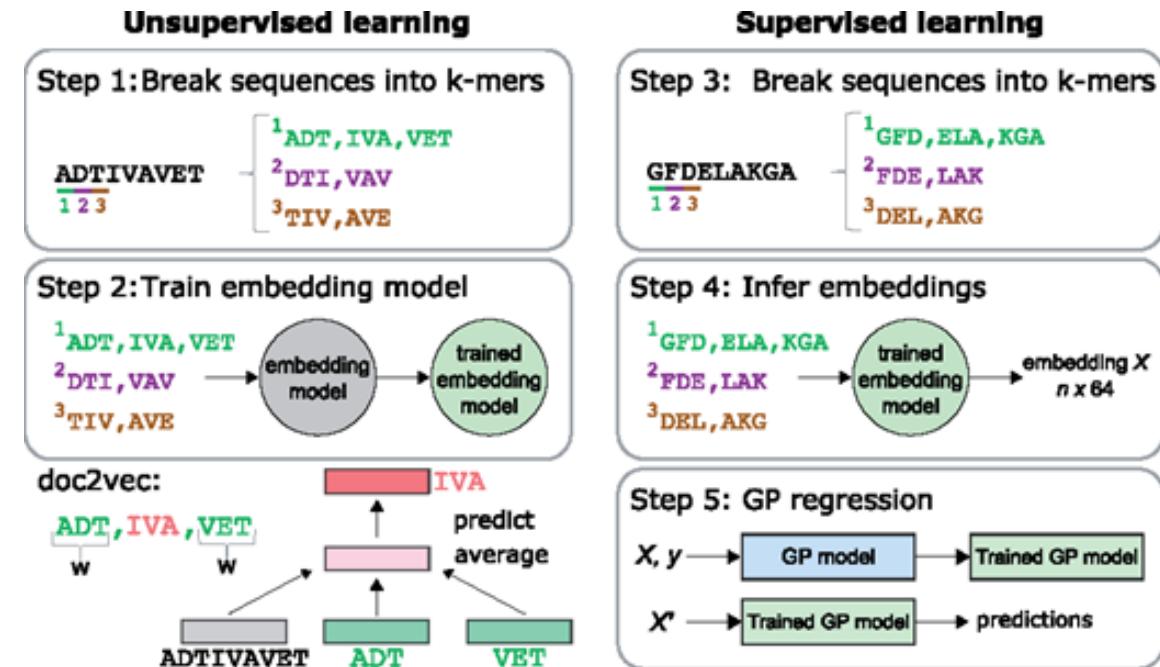
Representing proteins

1D sequence (vocab of size 20) – hundreds to thousands in length

2D contact map – requires prediction

3D structure – requires folding information, either observed or predicted. Only a limited number of 3D structures are known.

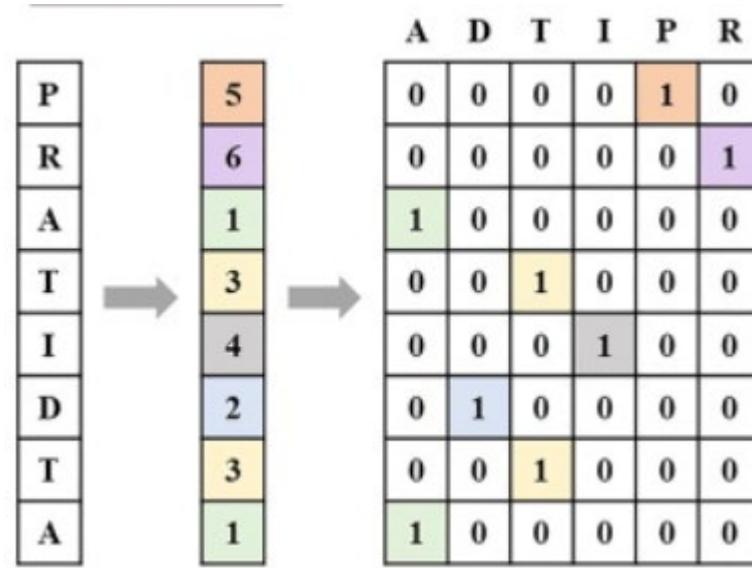
NLP-inspired embedding (word2vec, doc2vec, glove, seq2vec, ELMo, BERT, etc).



#REF: Yang, K. K., Wu, Z., Bedbrook, C. N., & Arnold, F. H. (2018). Learned protein embeddings for machine learning. *Bioinformatics*, 34(15), 2642-2648.

Sequential representation

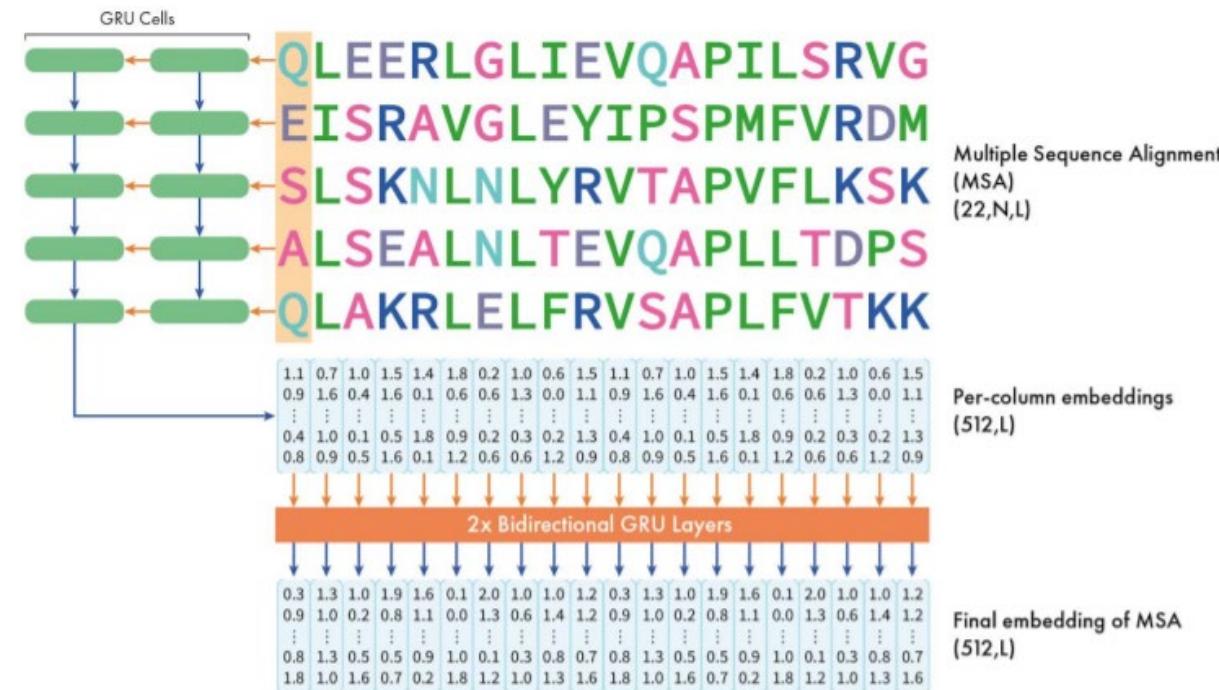
- **One-hot encoding:** simple, but inherently sparse, memory-inefficient, contains no prior knowledge and contextual information.



#REF: Lim, S., Lu, Y., Cho, C. Y., Sung, I., Kim, J., Kim, Y., ... & Kim, S. (2021). A review on compound-protein interaction prediction methods: Data, format, representation and model. Computational and Structural Biotechnology Journal, 19, 1541.

Sequential representation (cont.)

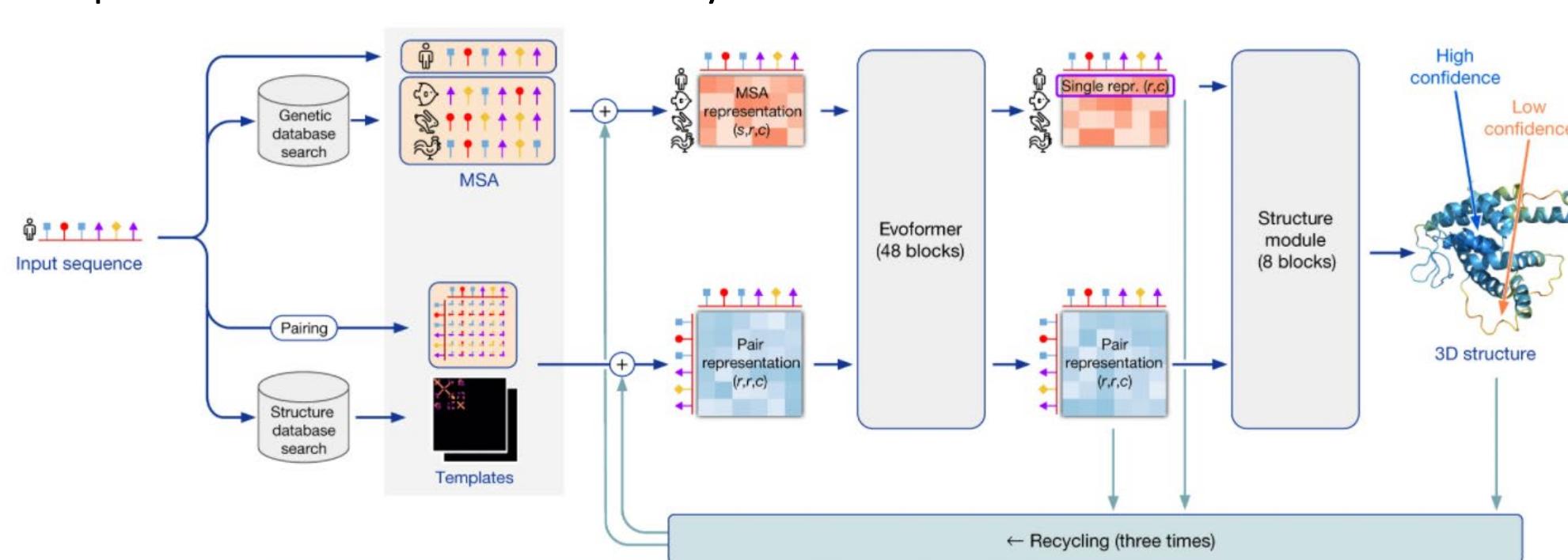
- **One-hot encoding:** simple, but inherently sparse, memory-inefficient, contains no prior knowledge and contextual information.
- **Evolutionary information:** search for related proteins to for multiple sequence alignment (MSA) and extract evolutionary information. Can be effective (AlphaFold2) but computational costly and requires sufficient data and diversity.



#REF: Kandathil, S. M., Greener, J. G., Lau, A. M., & Jones, D. T. (2021). Ultrafast end-to-end protein structure prediction enables high-throughput exploration of uncharacterised proteins.

Sequential representation (cont.)

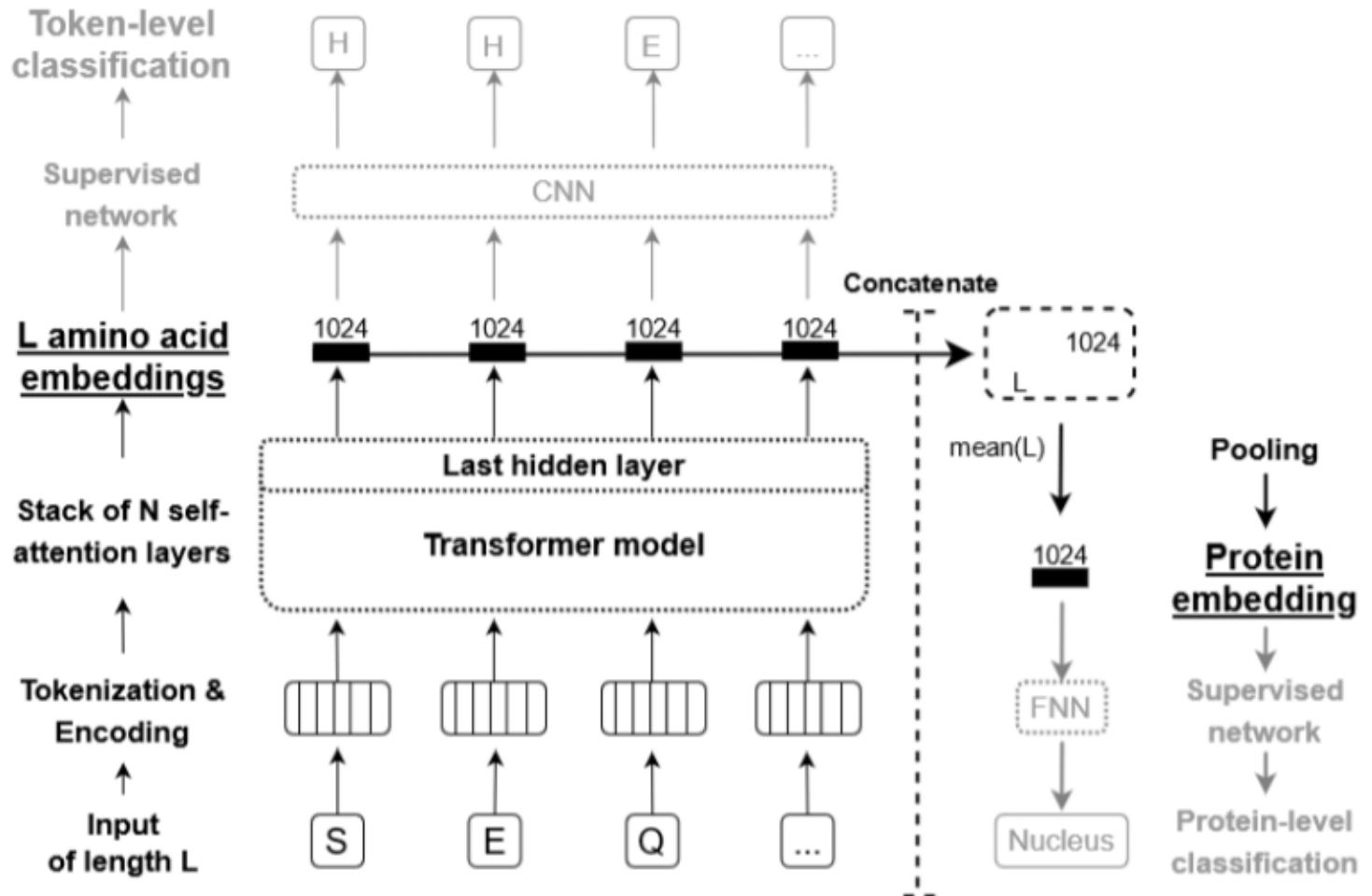
- **Evolutionary information:** search for related proteins to for multiple sequence alignment (MSA) and extract evolutionary information. Can be effective (AlphaFold2) but computational costly and requires sufficient data and diversity.



#REF:Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589.

NLP-inspired embedding

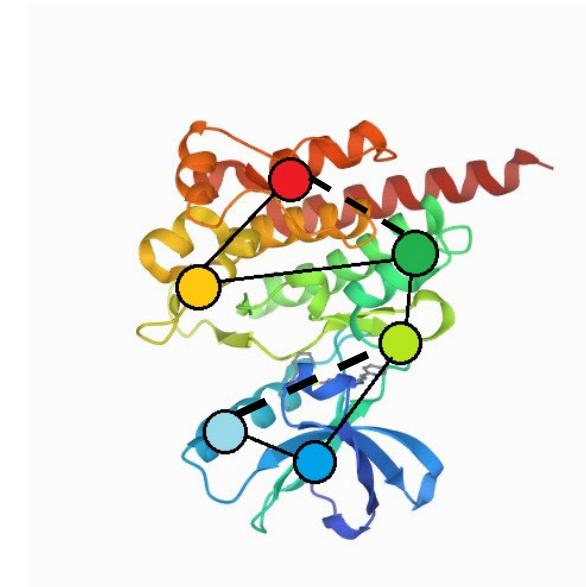
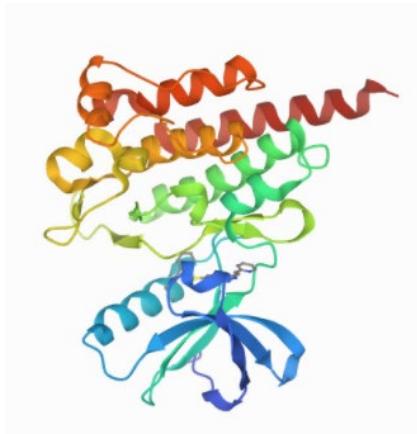
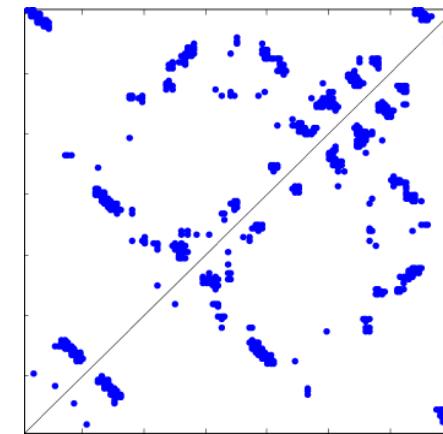
- **NLP-inspired embedding:** protein sequence as a sentence and residues as tokens then apply language modelling (word2vec, doc2vec, glove, seq2vec, ELMo, BERT, etc).



#REF: Elnaggar, A., Heinzinger, M., Dallago, C., Rihawi, G., Wang, Y., Jones, L., ... & Rost, B. (2020). ProtTrans: towards cracking the language of Life's code through self-supervised deep learning and high performance computing

Spatial representation

- **2D contact/distance map**: distance between residues pairs of 3D protein structure.
- **3D structure**: coordinate of residues in 3D space.
- **Graph representation**: residues as nodes and distance as edges. Can combine with sequential representation via attributed graph.



Protein sequence

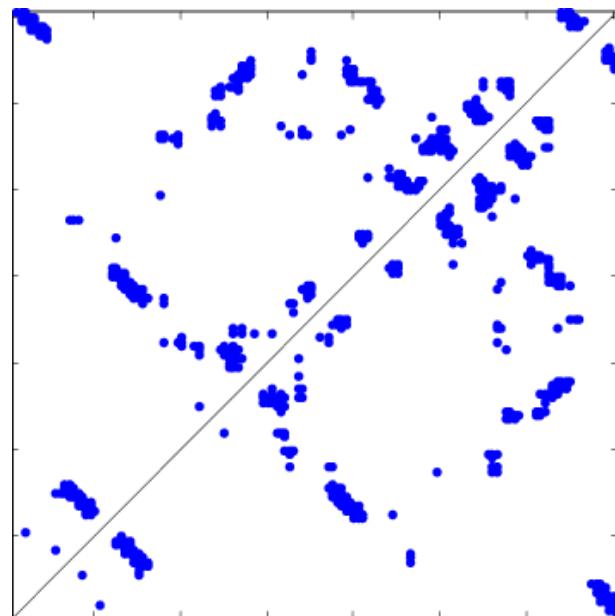
...AKLMMAATTAGVVVTTH...

↓
Pretraining

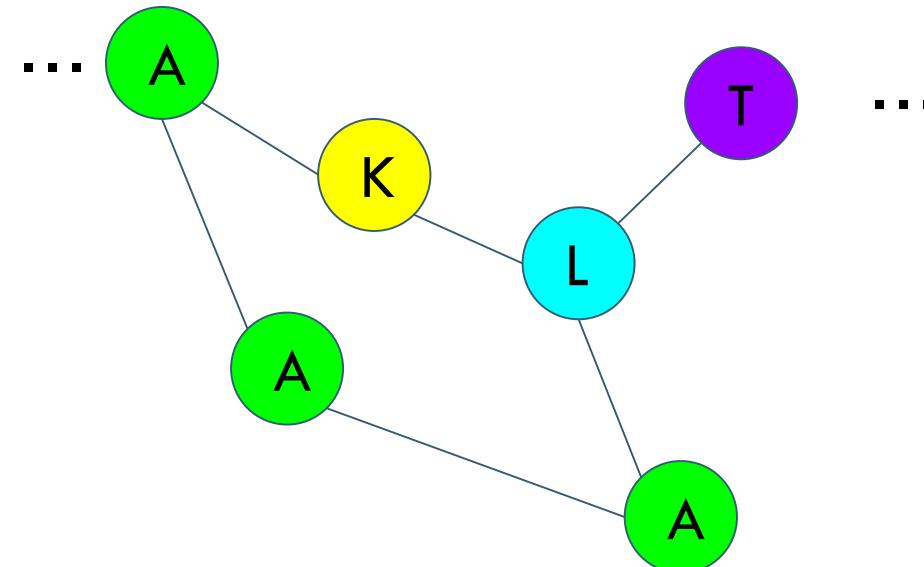
Protein sequence
embedding



Protein contact map 



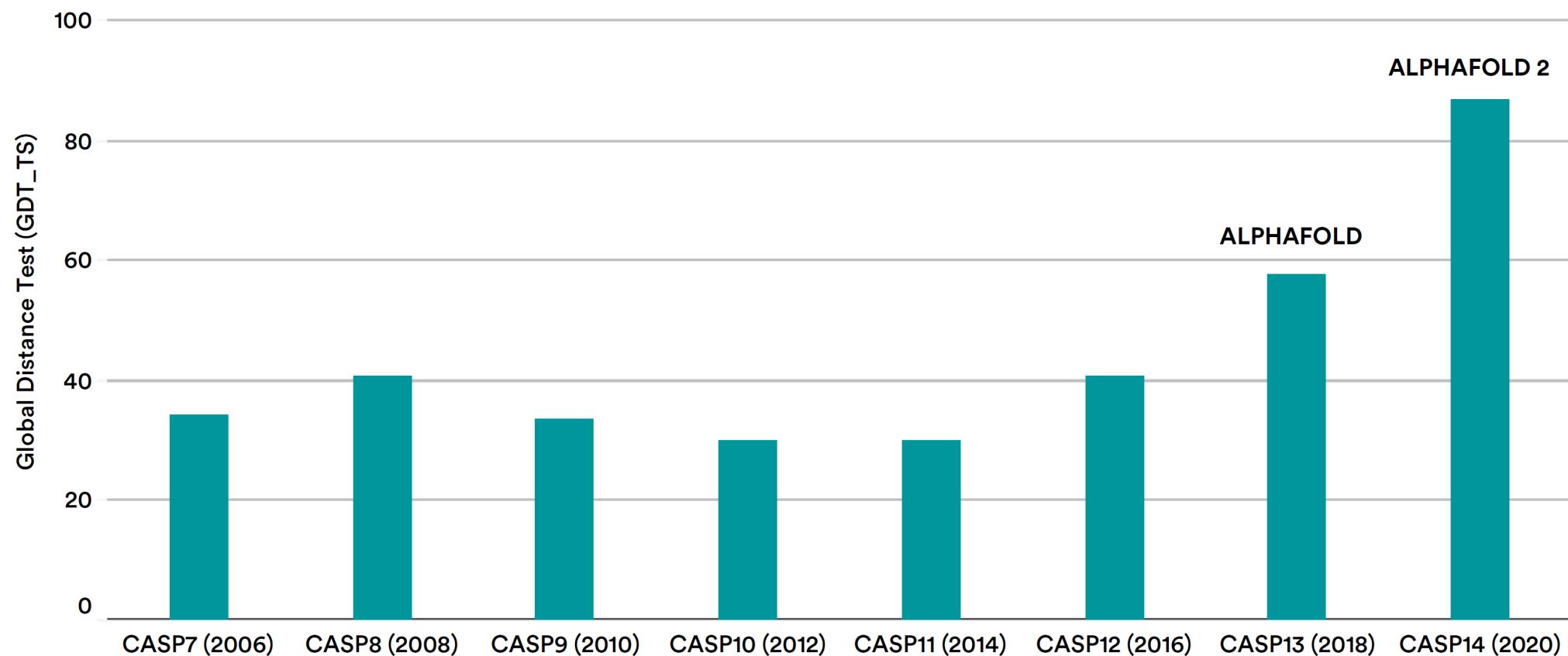
Protein graph



Protein folding prediction progress

CASP: MEDIAN ACCURACY of PREDICTIONS in FREE-MODELING by THE BEST TEAM, 2006-20

Source: DeepMind, 2020 | Chart: 2021 AI Index Report

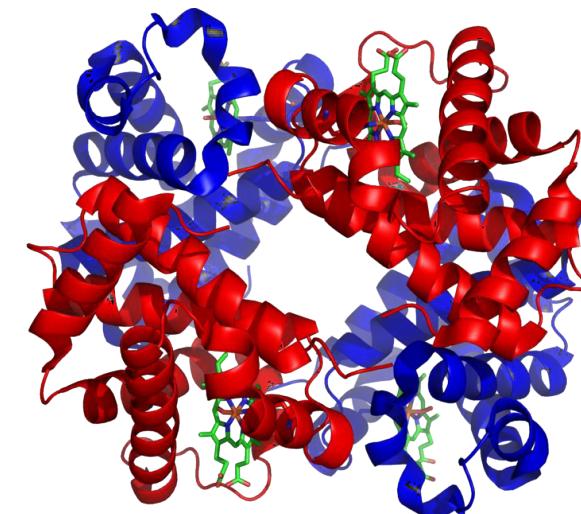


Protein 3D structure prediction

- Input: protein amino acid sequence $X = (x_1 x_2 \dots x_L)$ where L is protein sequence length, x_i is residue type at position i .
- Output: the 3D coordinate of residues in protein sequence.

...MEFMWKR...

Amino acid sequence



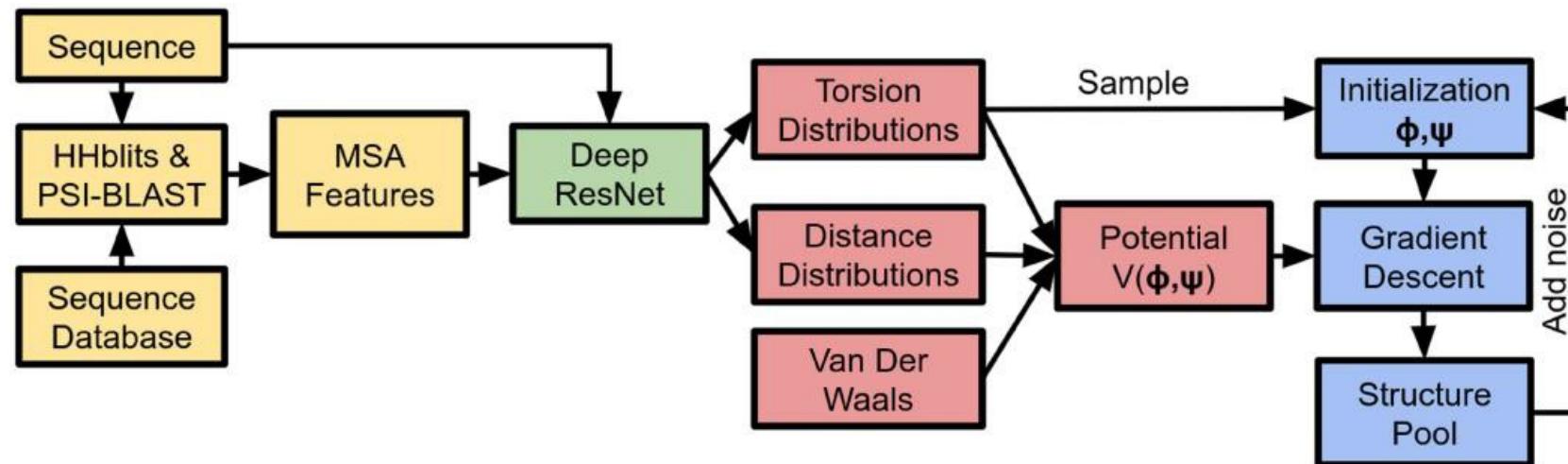
#REF: "Hemoglobin", En.wikipedia.org, 2020. [Online]. Available: <https://en.wikipedia.org/wiki/Hemoglobin>.

Approaches

- Template-based
 - Using known structure of proteins that have high sequence similarity with target protein as the initial structure template.
 - From the initial structure, protein fragments are inserted or deleted to minimize the global free energy.
 - An effective method if the target protein sequence has at least 30% sequence identity with the template protein.
- Template-free
 - Without using solved protein structure
 - Main goal is to find a conformation that has minimum free energy
 - Require a vast computational resource such as powerful super computer or distributed computing projects (Rosetta@Home, Folding@Home)

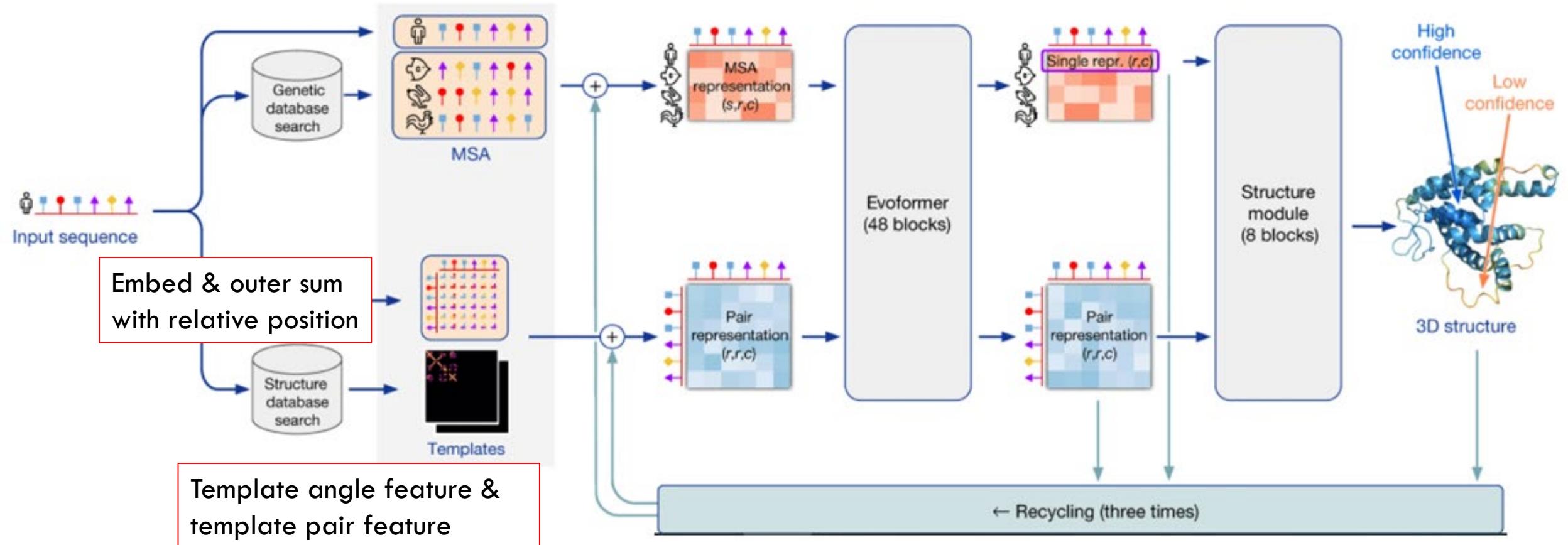
AlphaFold

- State-of-the-art in protein structure prediction
- Template-free approach
- Construct a potential mean force which can accurately describe the protein 3D structure



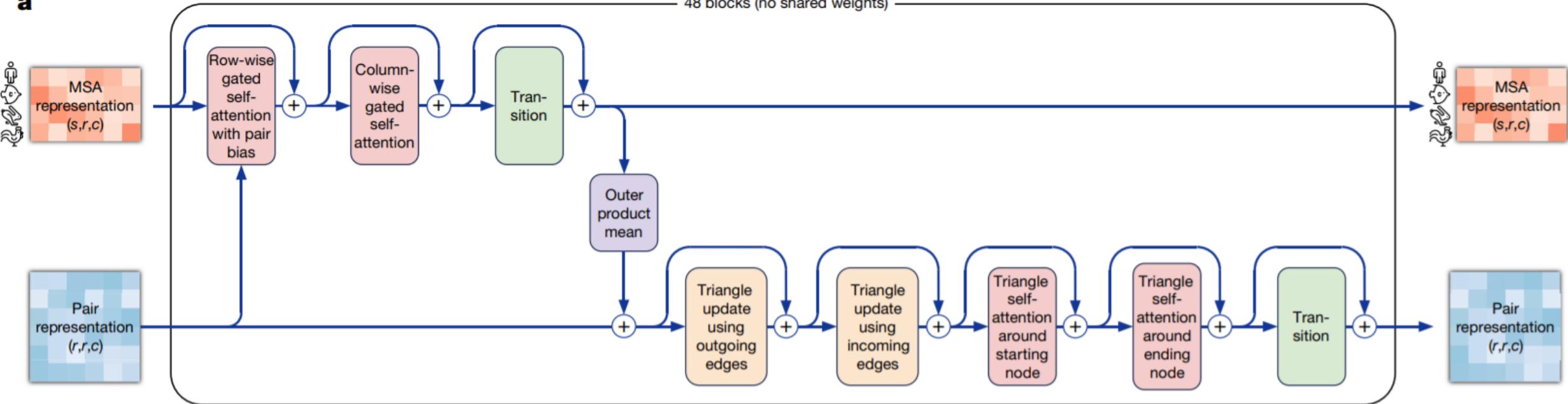
The pipeline of AlphaFold to predict the 3D structure of a protein sequence

AlphaFold 2



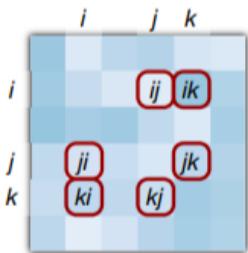
AlphaFold 2 – Evoformer block

a

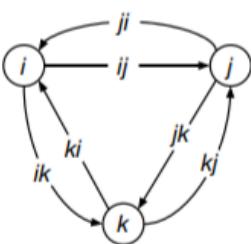


b

Pair representation
(r, r, c)

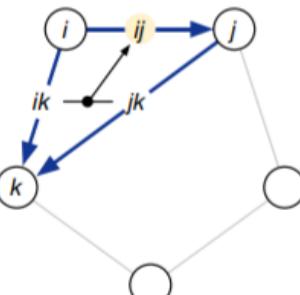


Corresponding edges
in a graph

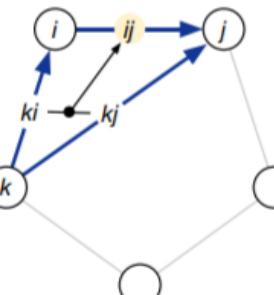


c

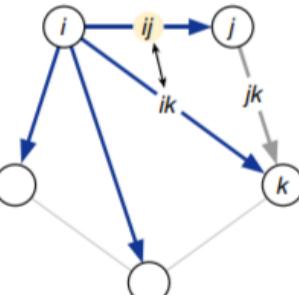
Triangle multiplicative update
using 'outgoing' edges



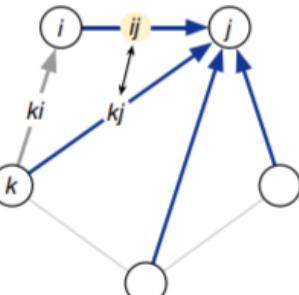
Triangle multiplicative update
using 'incoming' edges



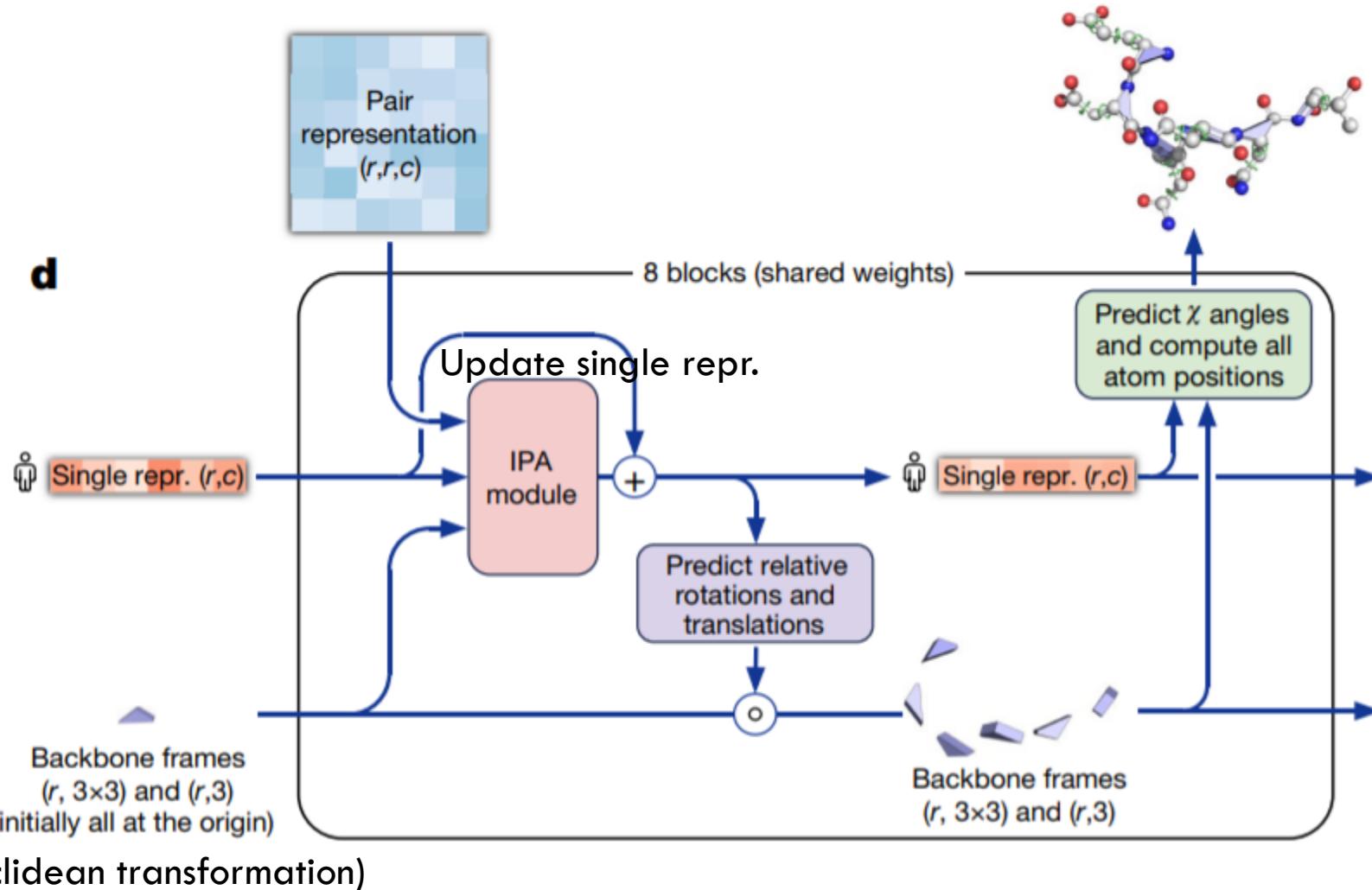
Triangle self-attention around
starting node



Triangle self-attention around
ending node



AlphaFold 2 – Structure module



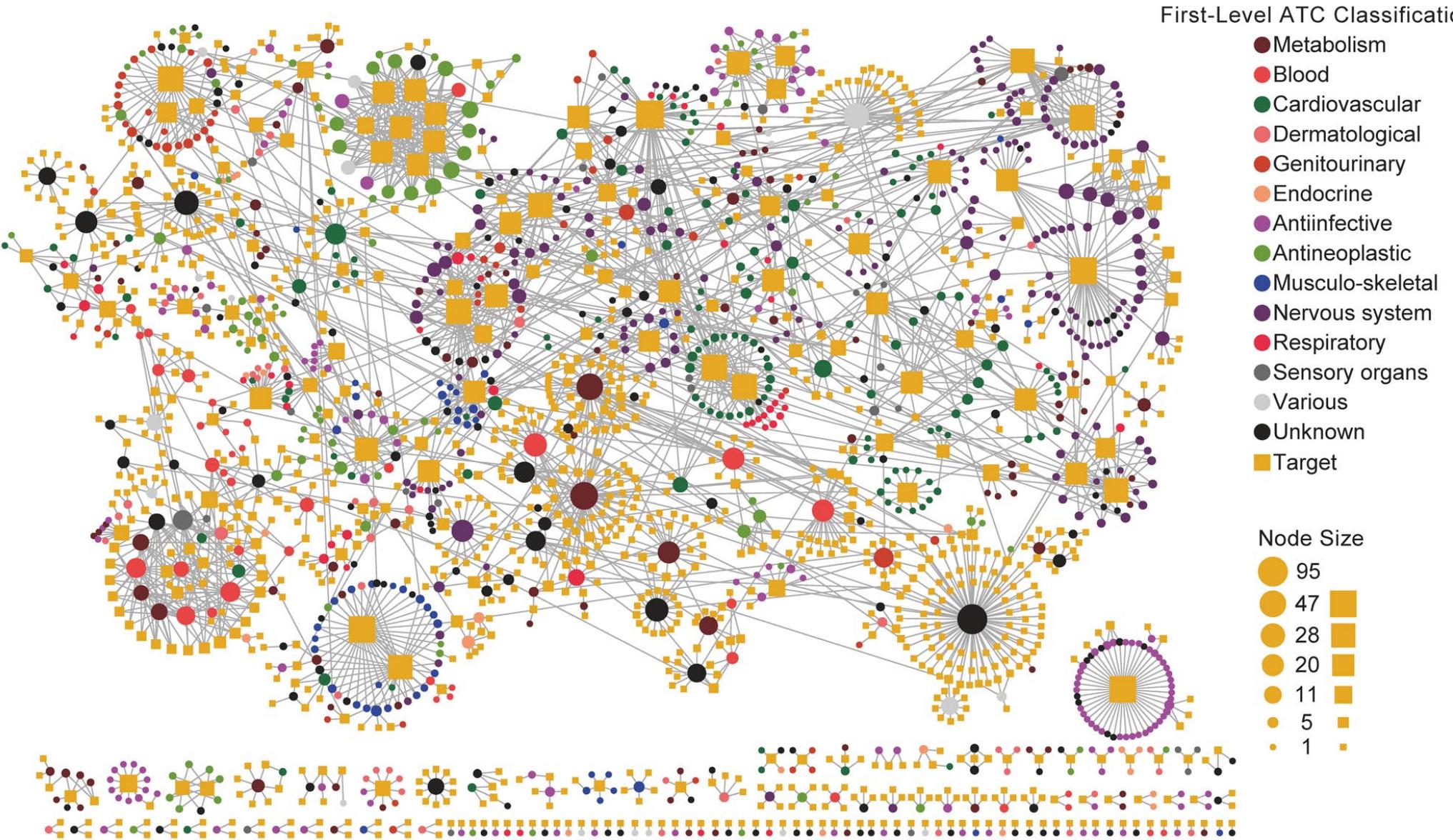
Agenda

Protein representation learning

- Embedding, BERT
- 2D contact map
- 3D structure

Drug-target binding prediction

- Multi-target prediction
- Drug-protein binding as graph-graph interaction
- Cold-start problem
- Explanation



Drug-target bipartite network

Cheng, Feixiong, et al. "Prediction of drug-target interactions and drug repositioning via network-based inference." *PLoS computational biology* 8.5 (2012): e1002503.

Bipartite network operations

Link prediction

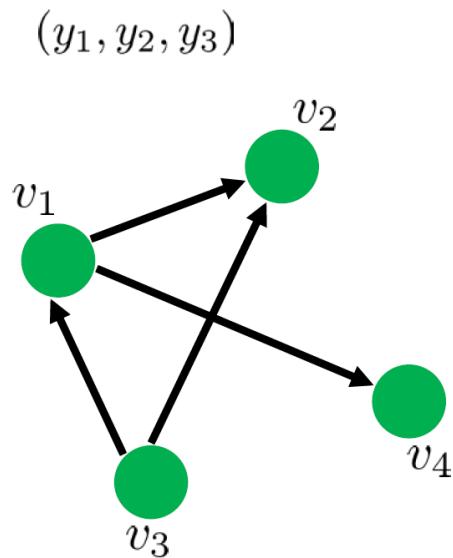
Recommendation techniques, e.g., SVD, random walks, nearest neighbours.

Knowledge graph completion techniques, e.g., TransE

More on this later in Part C2.

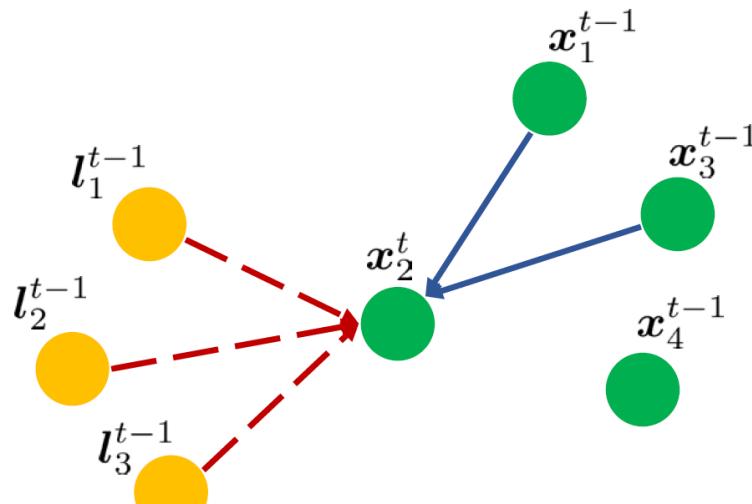
Multi-target binding

Scale linearly with number of targets + efficient processing through message factoring.

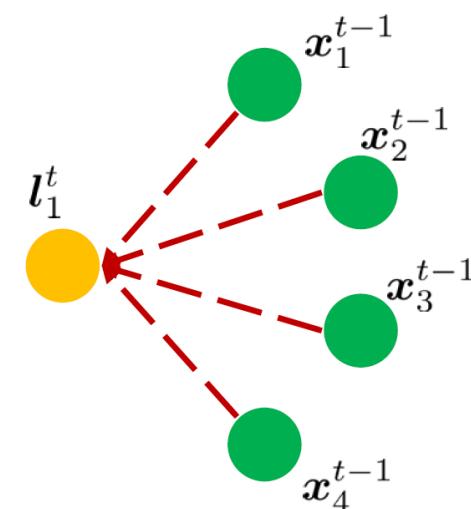


(a) A input graph with 4 nodes and 3 labels

#REF: Do, Kien, et al. "Attentional Multilabel Learning over Graphs-A message passing approach." *Machine Learning*, 2019.

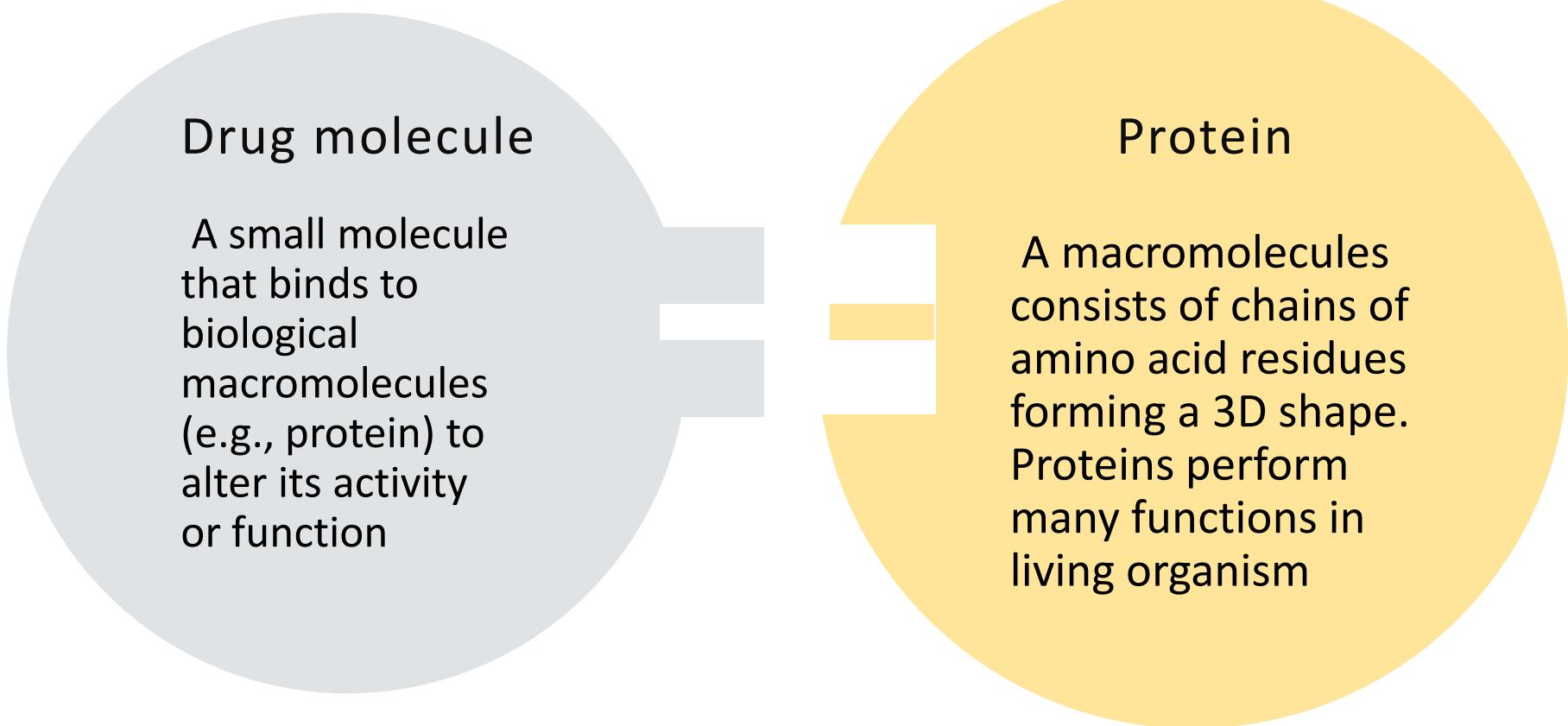


(b) Input node update

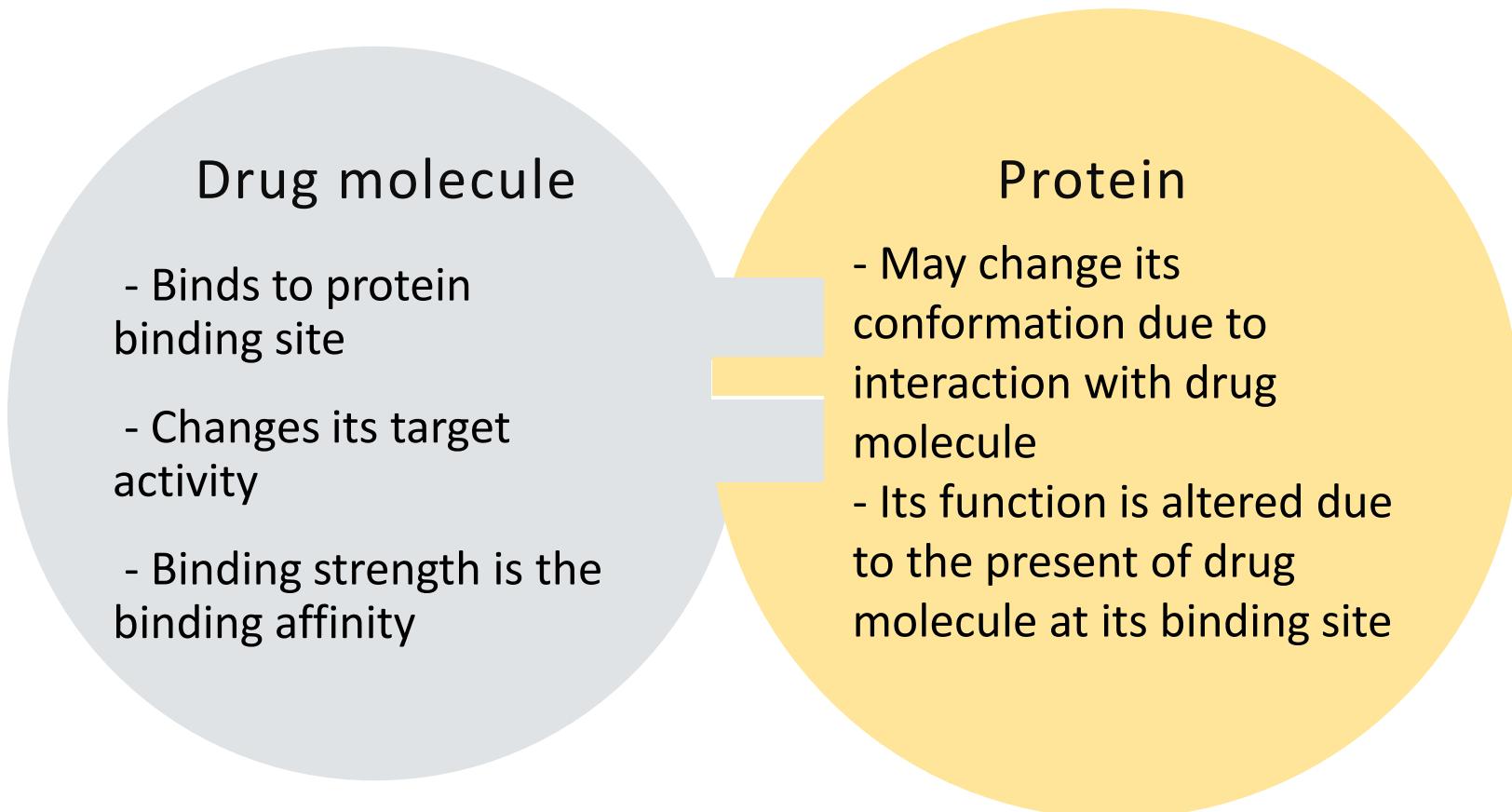


(c) Label node update

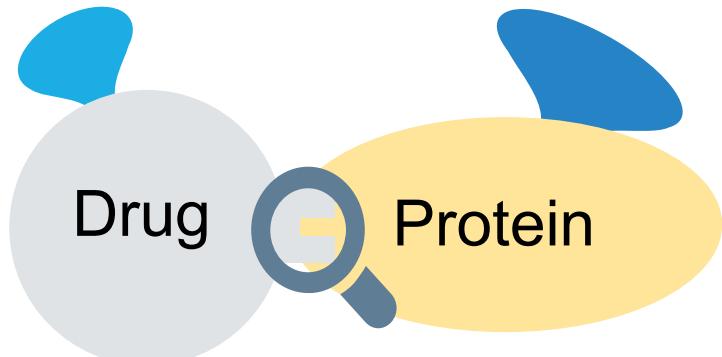
Drug and protein



Drug and protein (2)



Drug and protein (3)



We need to understand drug-target interaction because:

- Fast and safe drug repurposing and discovery/repurposing process for swift pandemic reaction
- Find solutions for challenging diseases

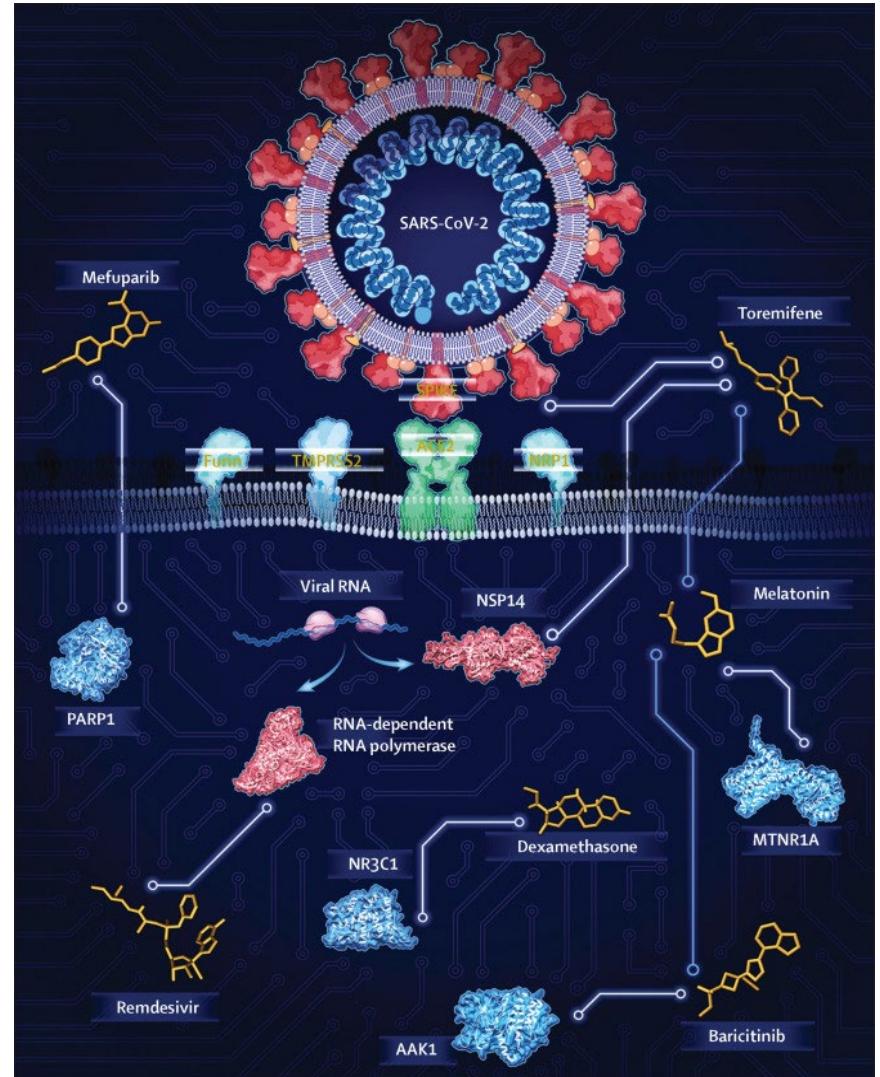


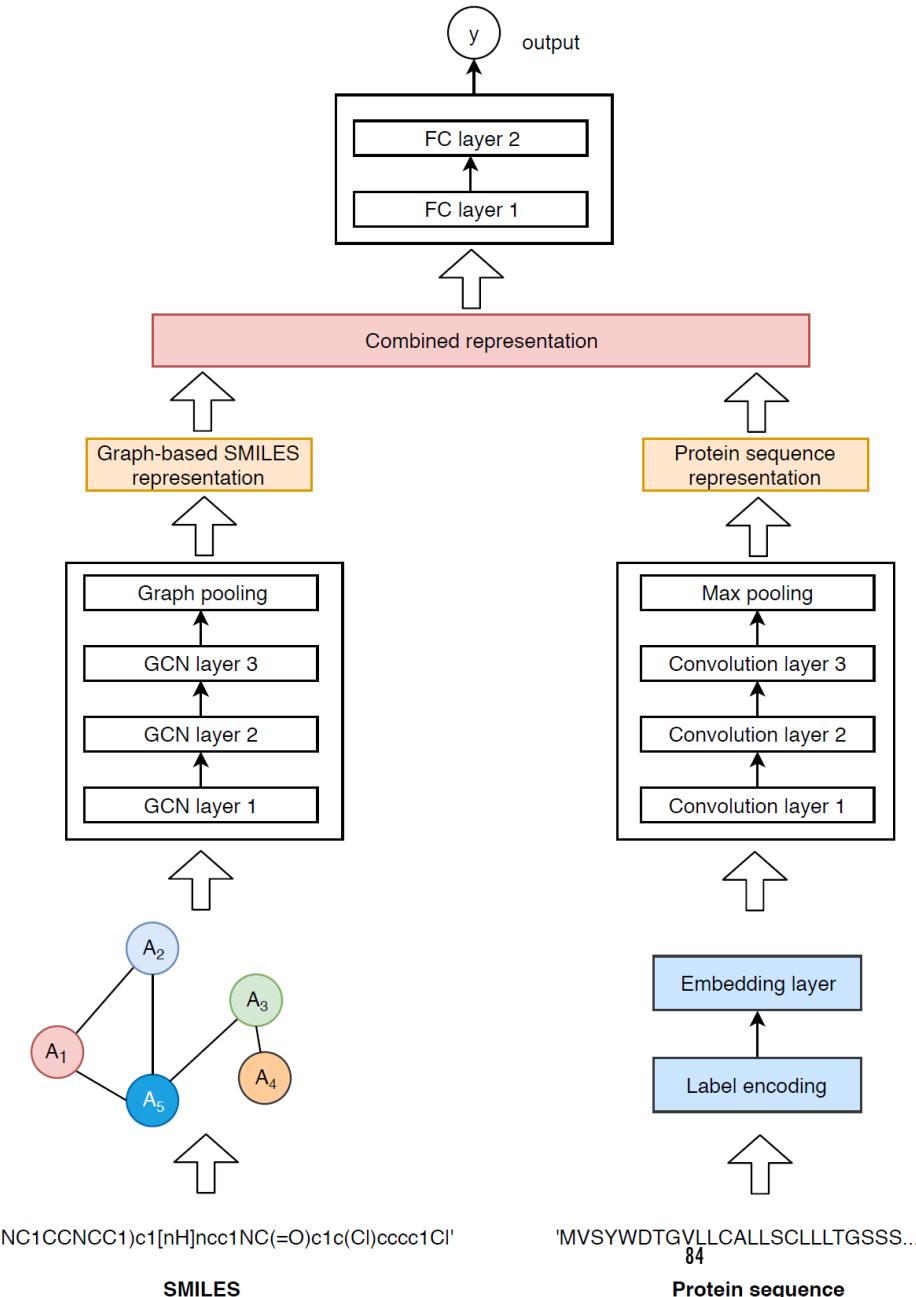
Image credit: Lancet

Drug-target binding as question-answering

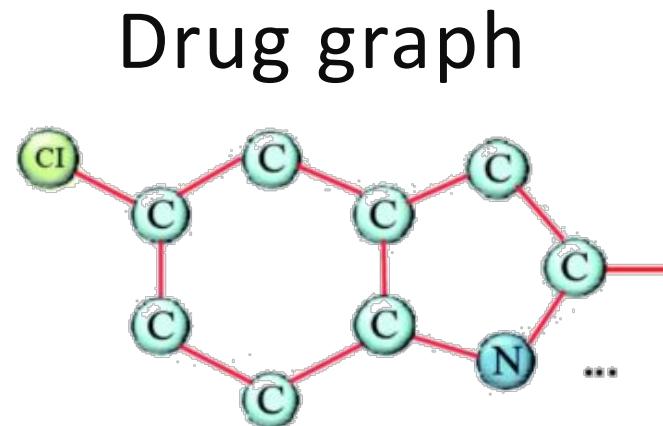
- **Context/database:** Binding targets (e.g., RNA/protein sequence, or 3D structures), as a set, sequence, or graph + existing binding, interaction databases.
- **Query:** Drug (e.g., SMILES string, or molecular graph)
- **Answer:** Affinity, binding sites, modulating effects, conformation changes.

17/09/2021

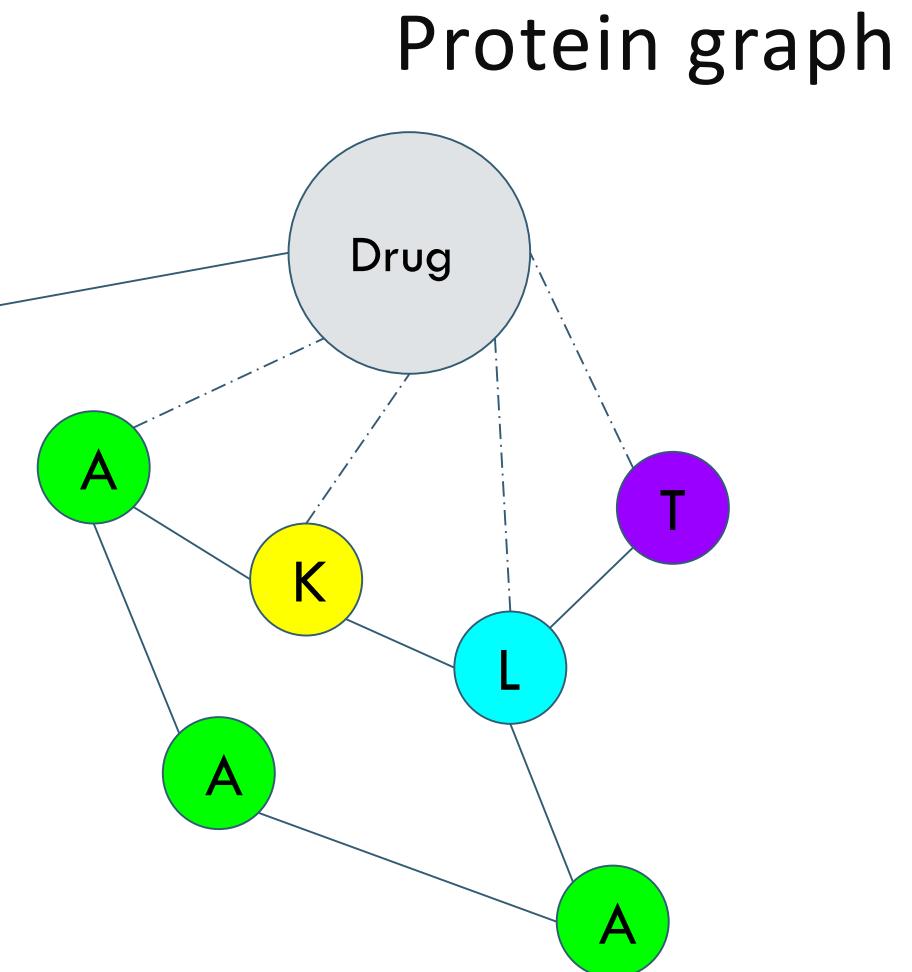
#REF: Nguyen, T., Le, H., & Venkatesh, S. (2019). GraphDTA: prediction of drug–target binding affinity using graph convolutional networks. *BioRxiv*, 684662.



GEFA: Drug-protein binding as **graph-in-graph** interaction



Graph-in-Graph
interaction

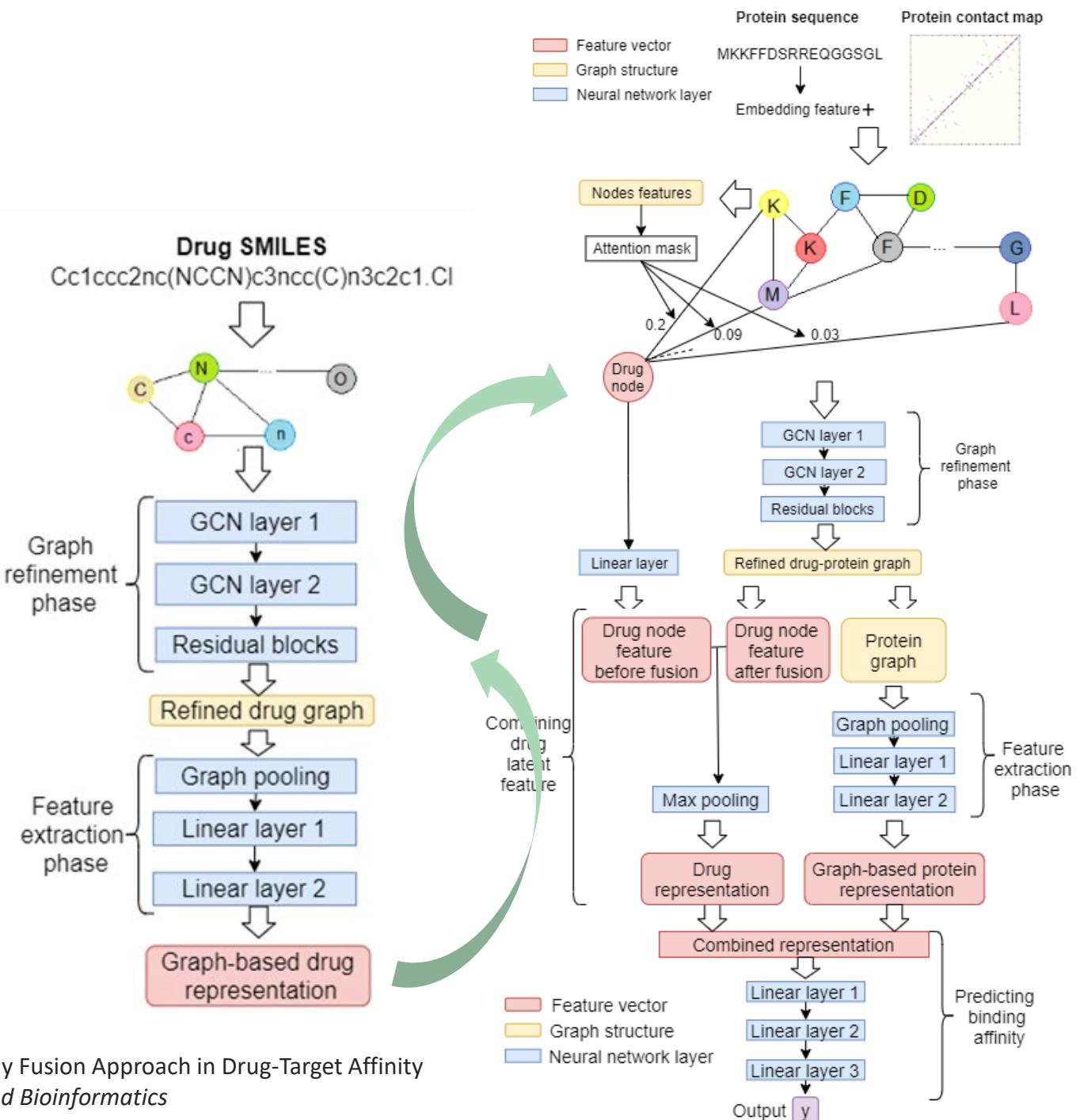


GEFA (cont.)

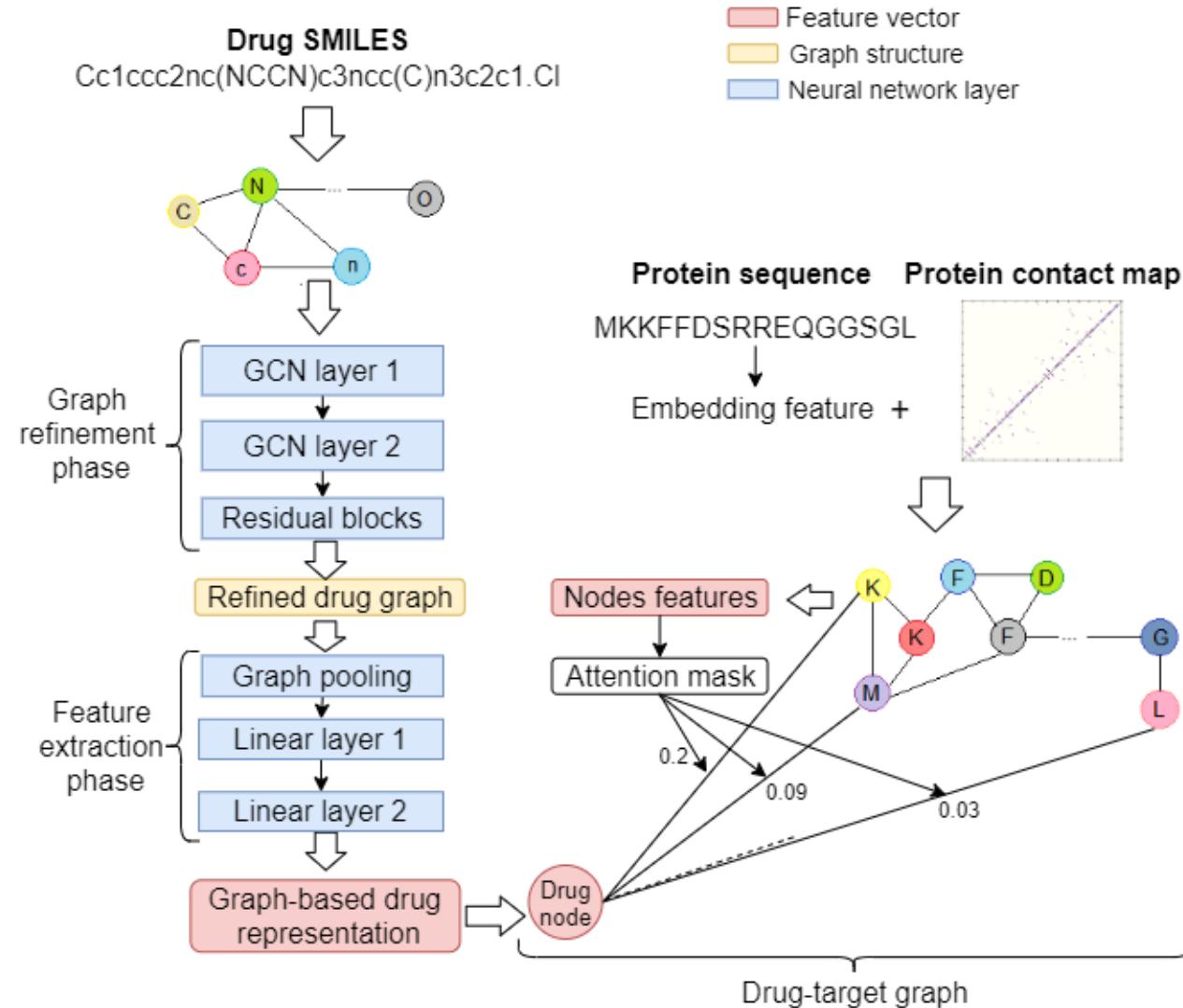
We designed a model for detailed interaction between drug and protein residues.

The architecture is a new *graph-in-graph*.

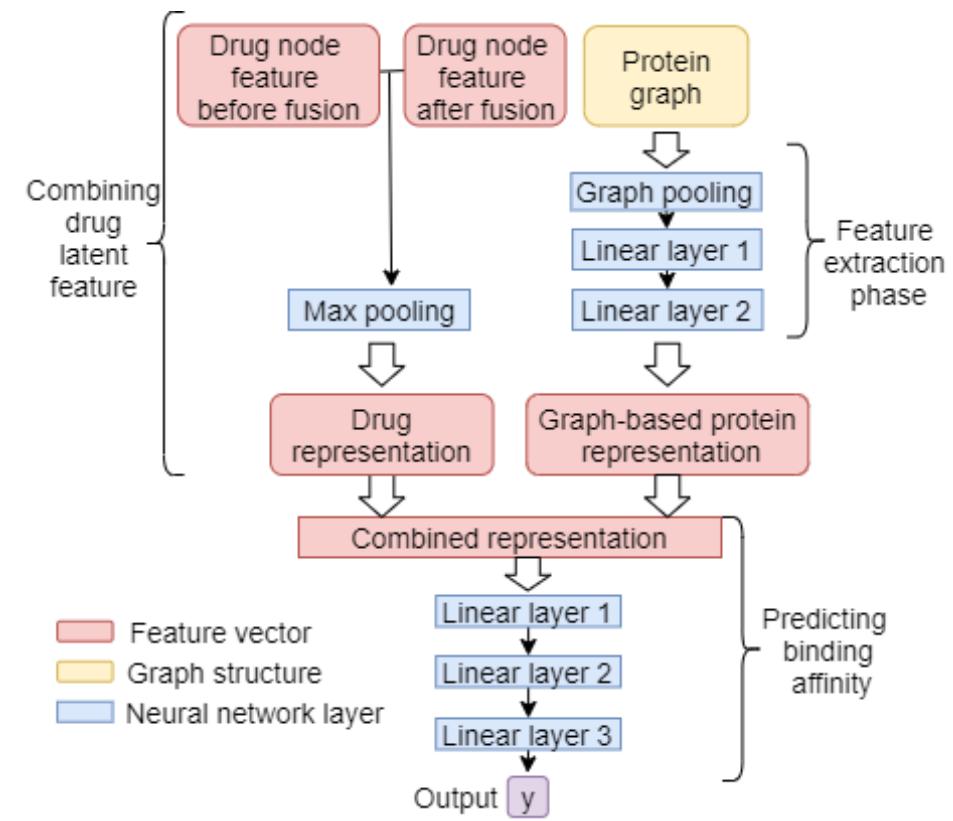
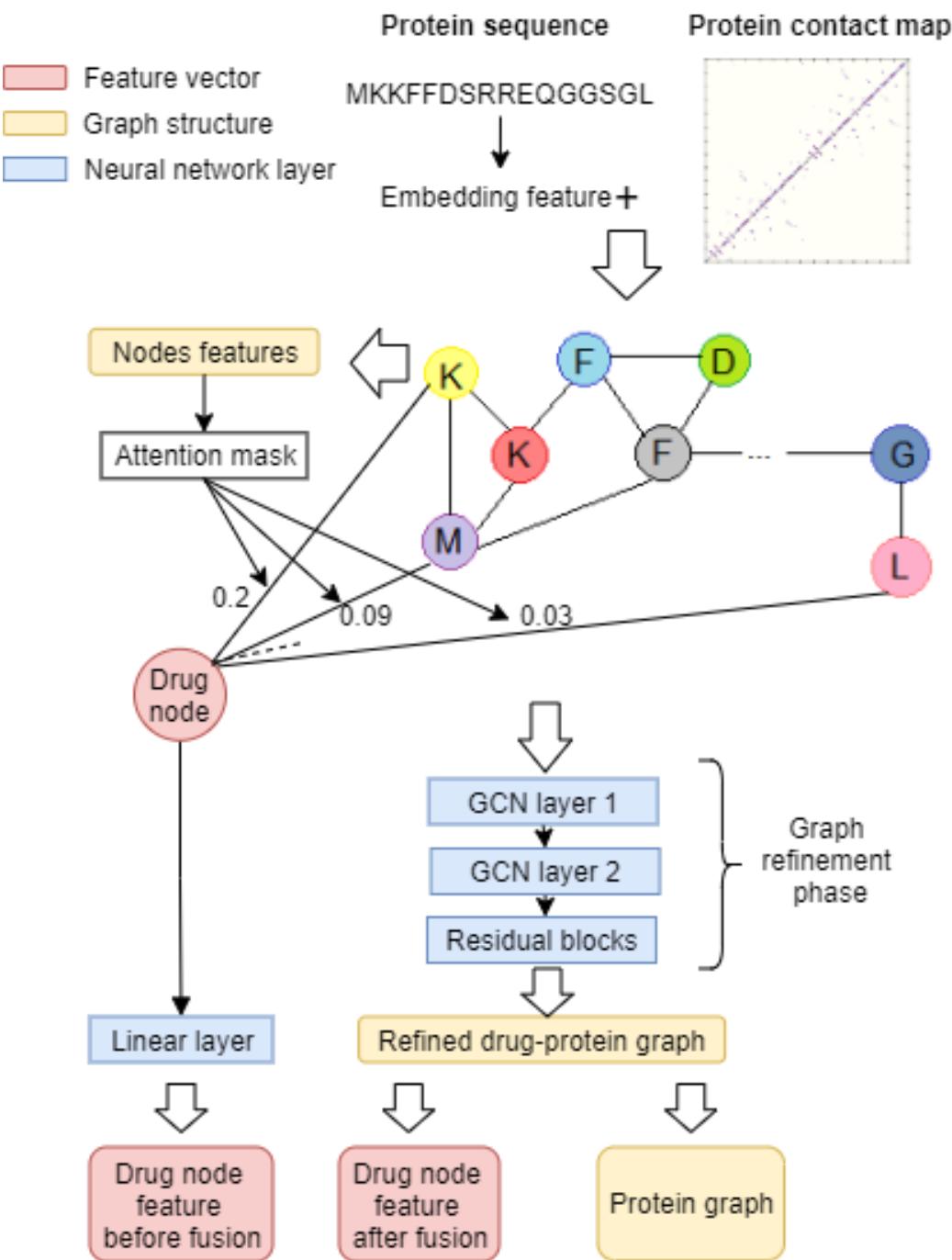
This results in more accurate and precise prediction of binding site and strength.



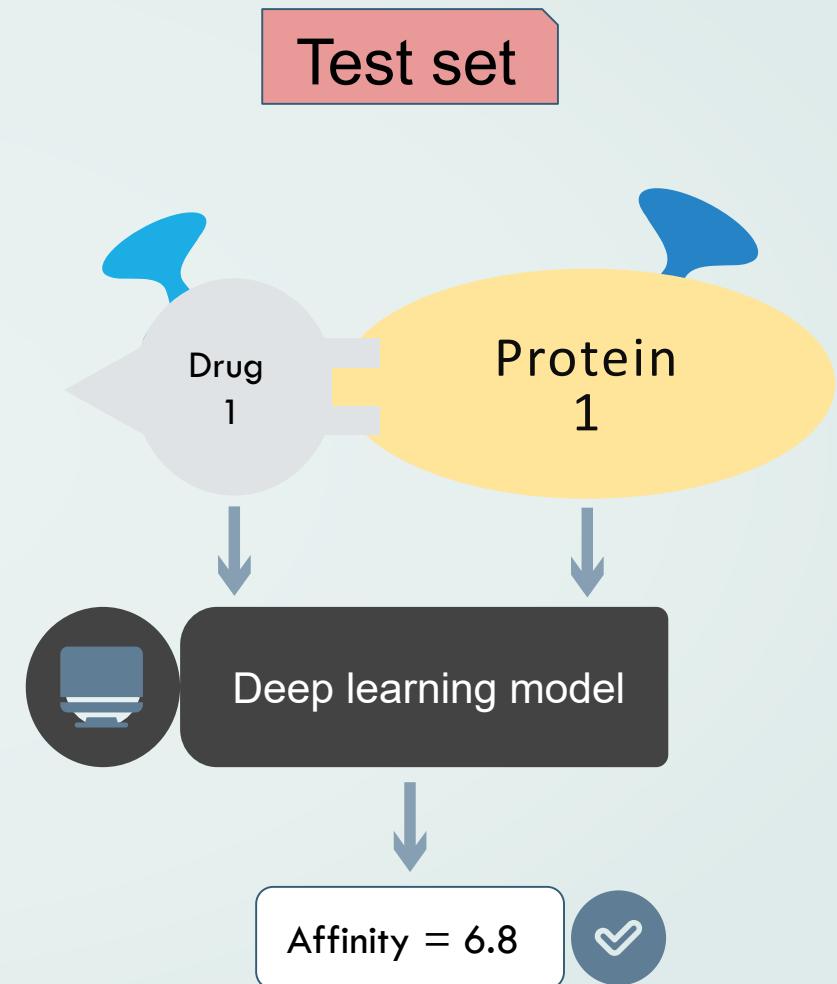
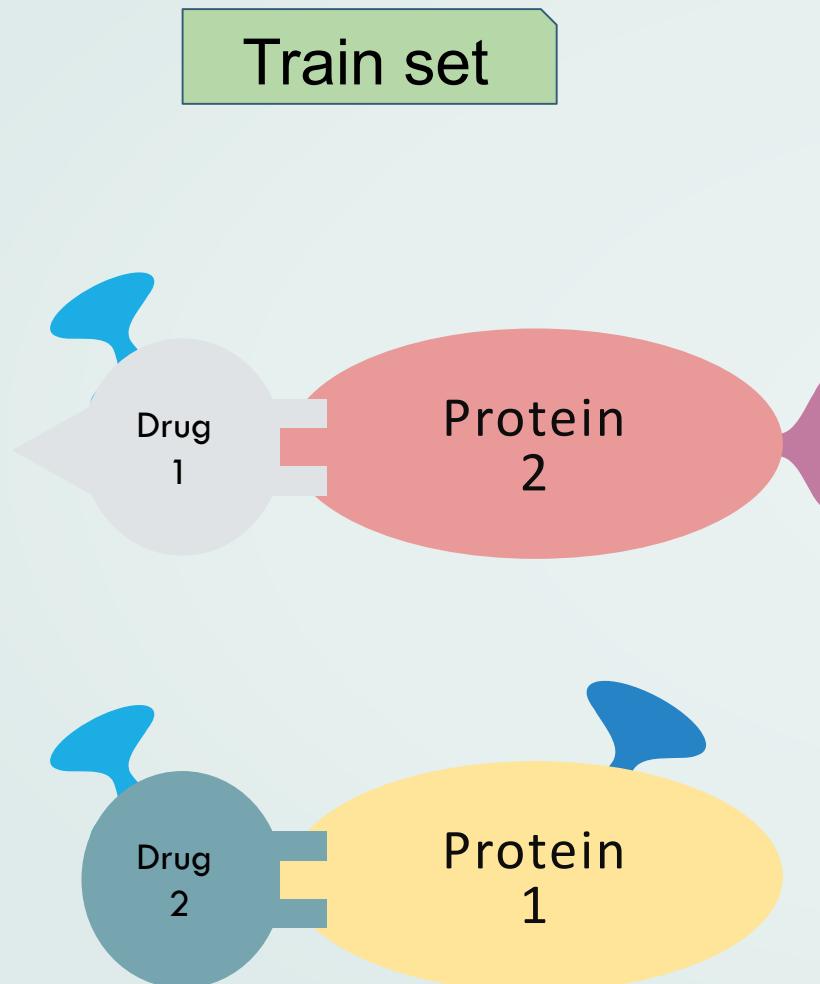
GEFA (cont.): Drug-target graph



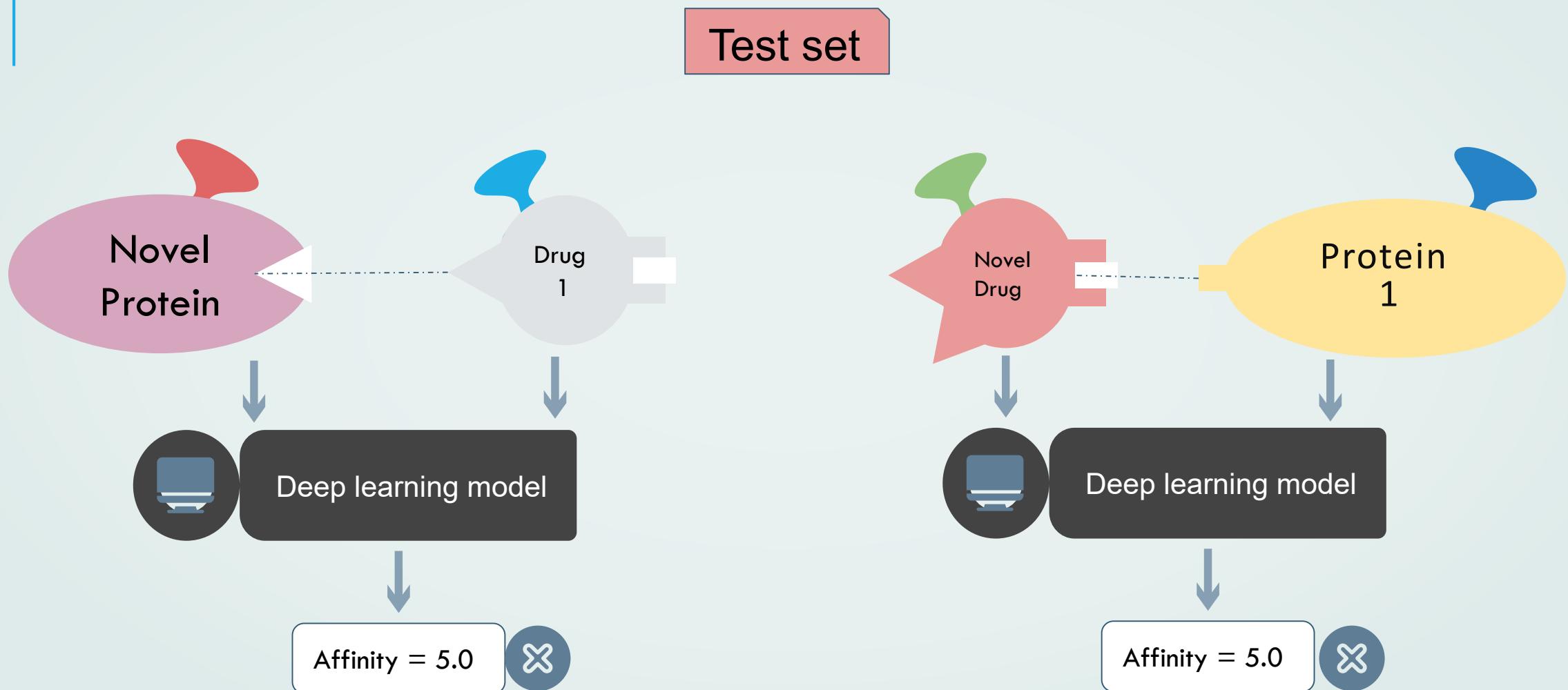
GEFA (cont.): GCN + fusion



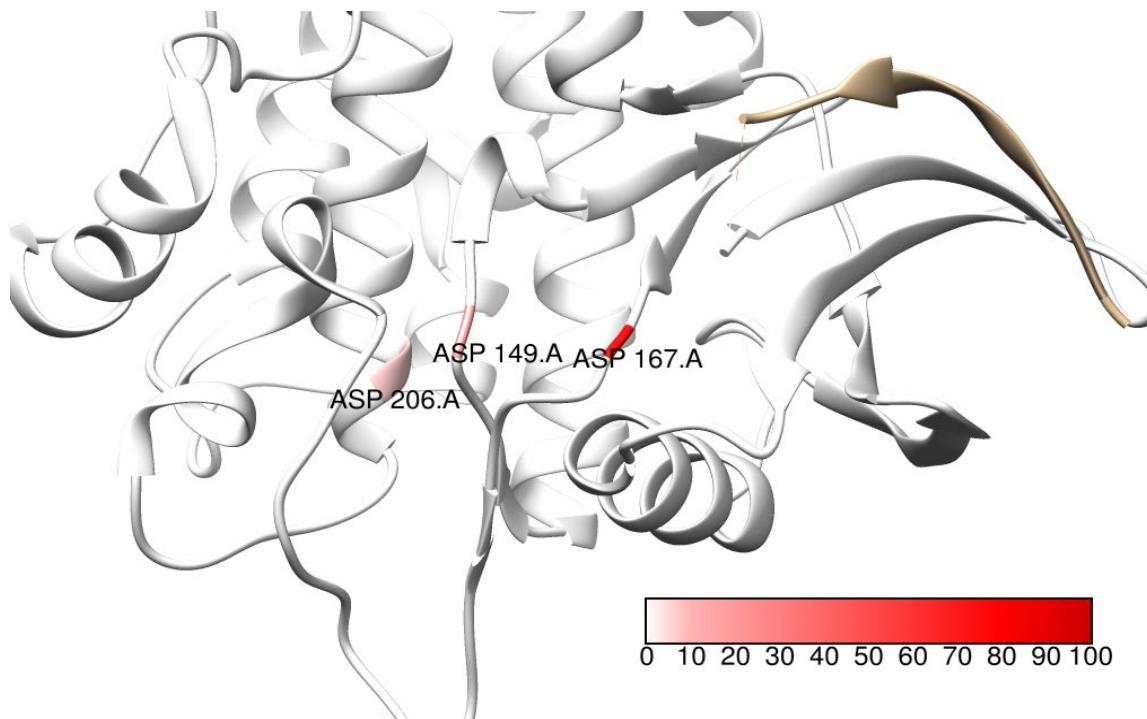
Cold-start problem



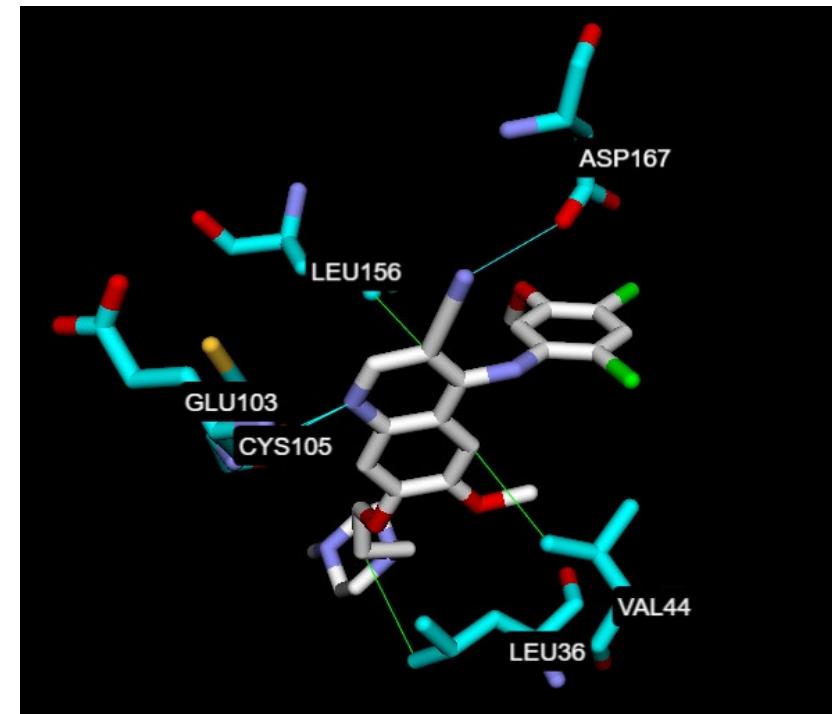
Cold-start problem



Results visualization

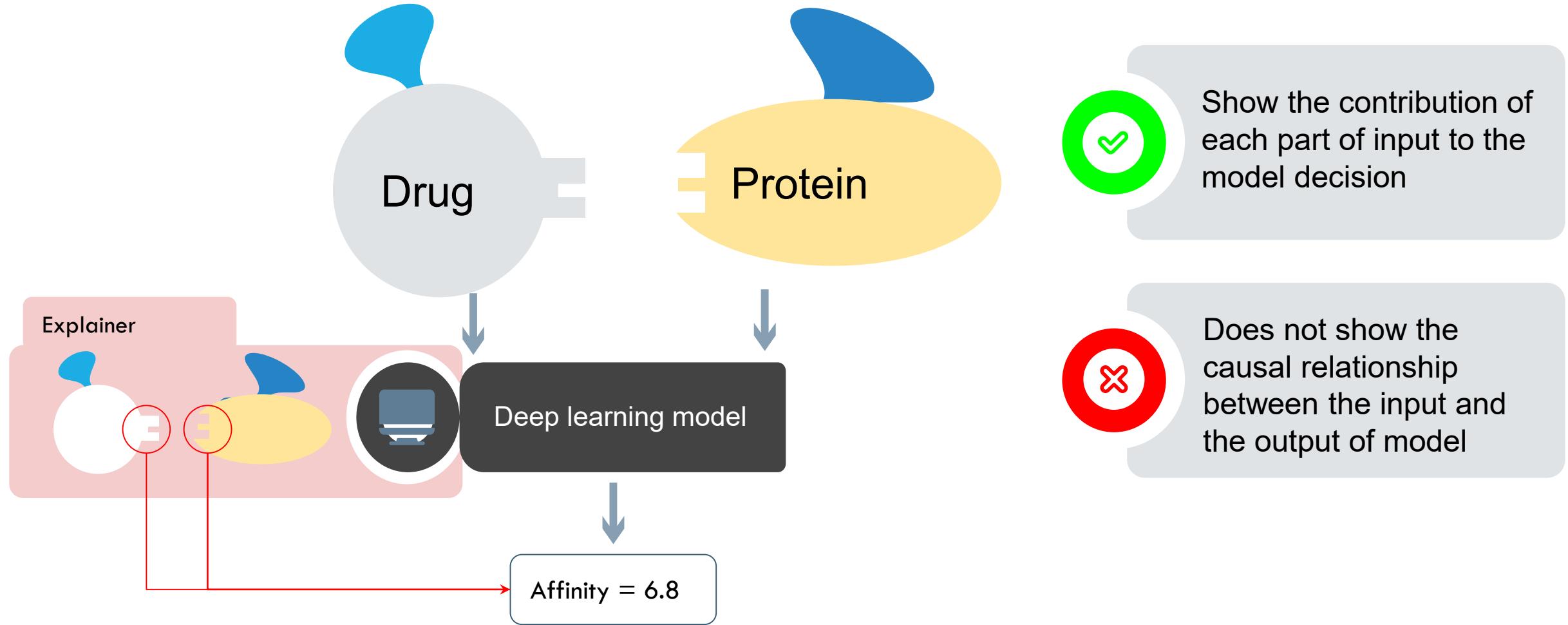


Attention values at predicted binding sites of MST1 target. Residues ASP 167.A has highest attention score.

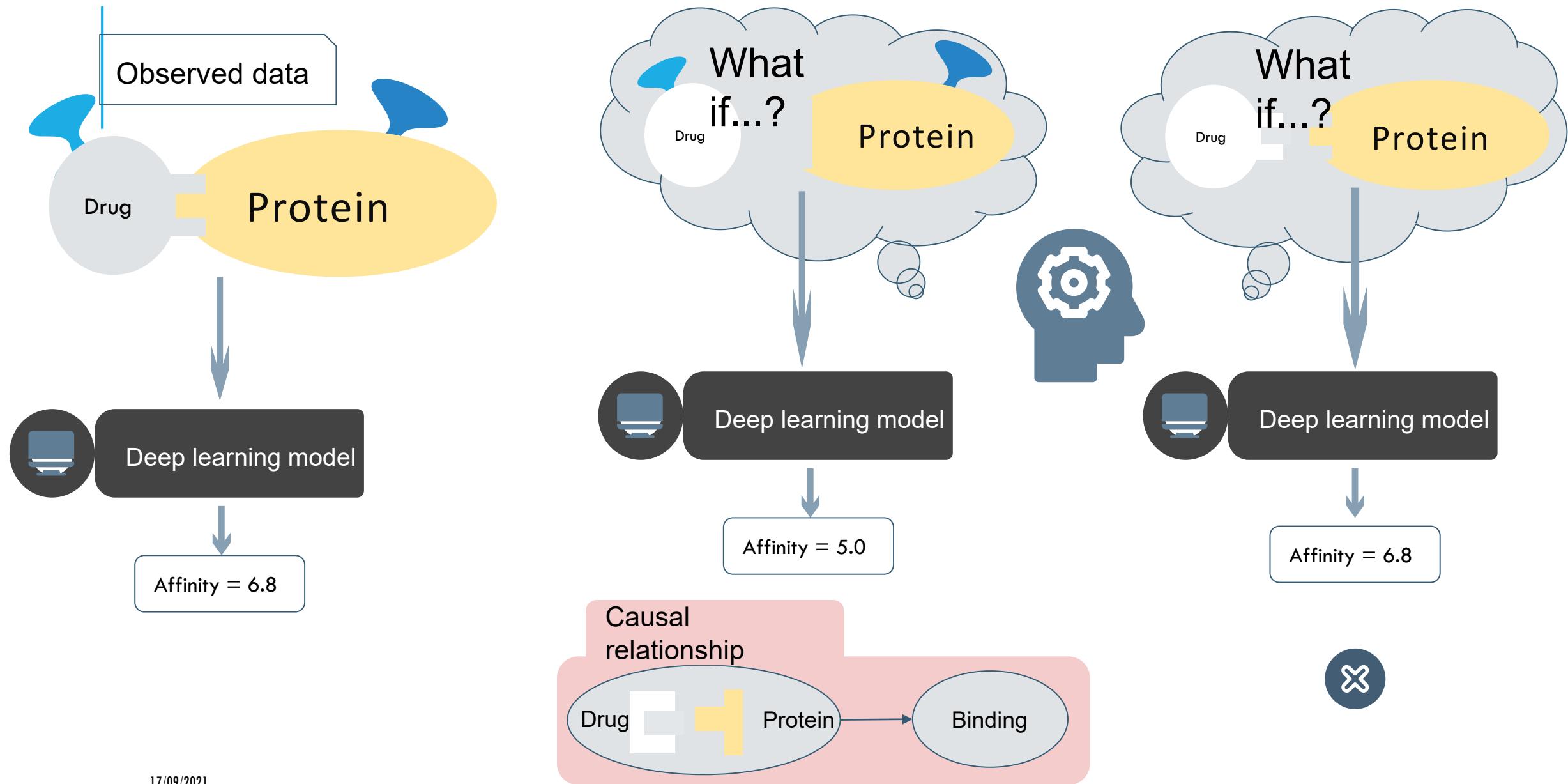


Residue-ligand interaction predicted by simulation. Residue ASP 167.A is one of binding site

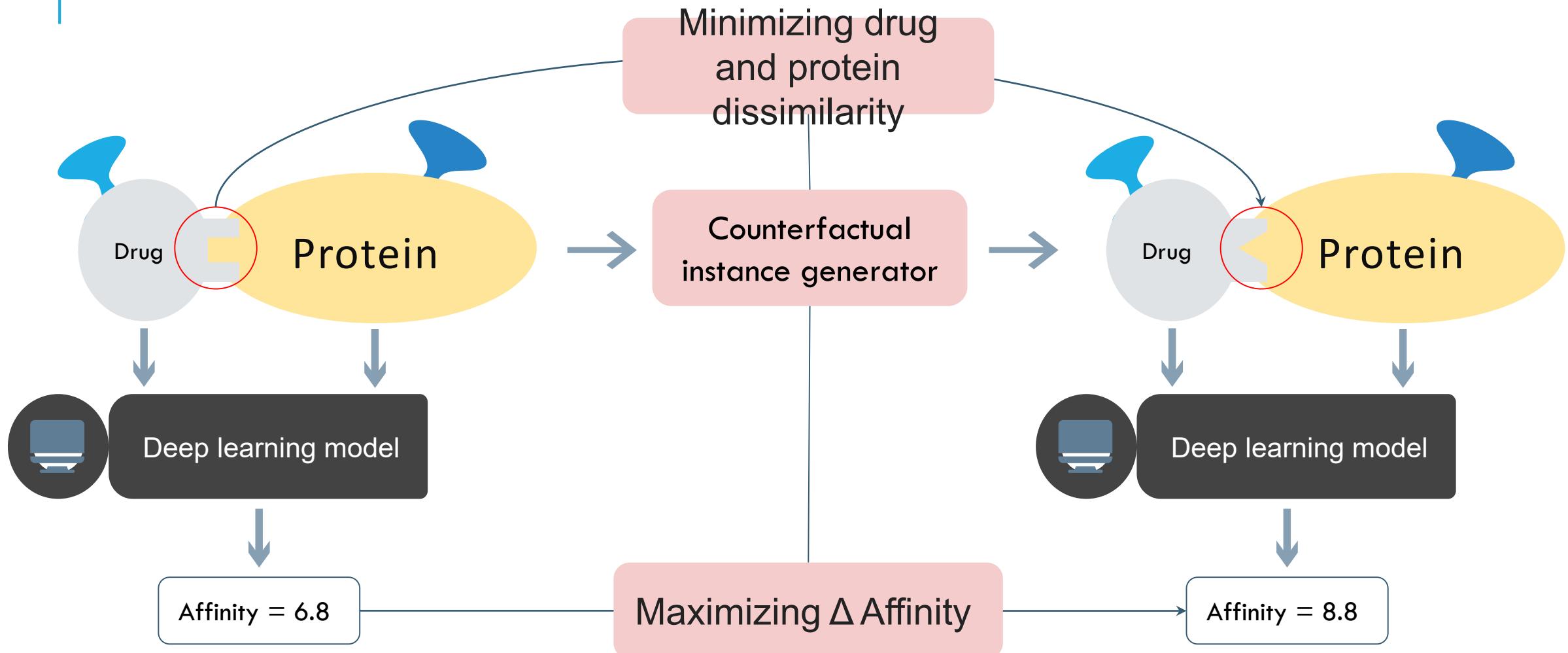
Explaining DTA deep learning model: feature attribution



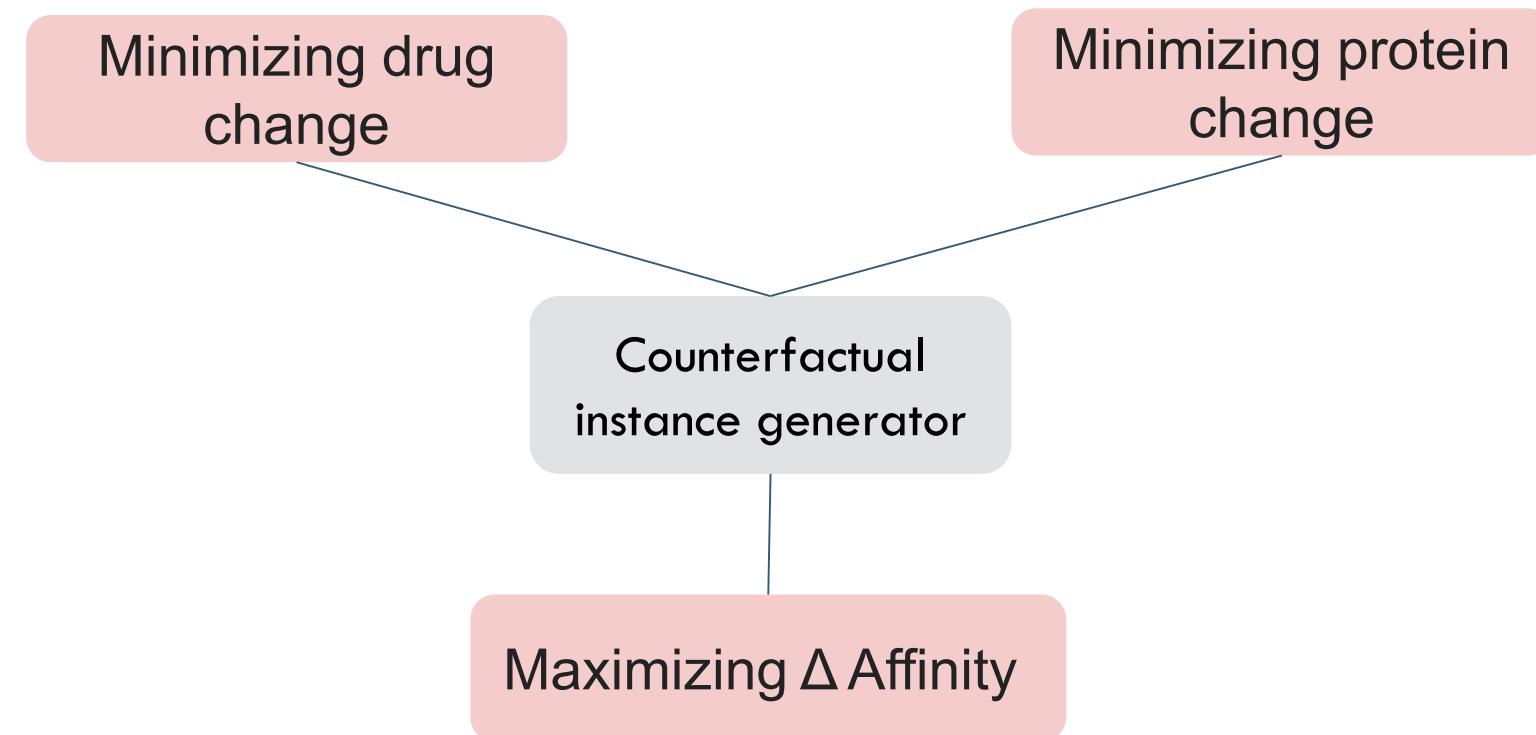
Explaining DTA deep learning model: counterfactual



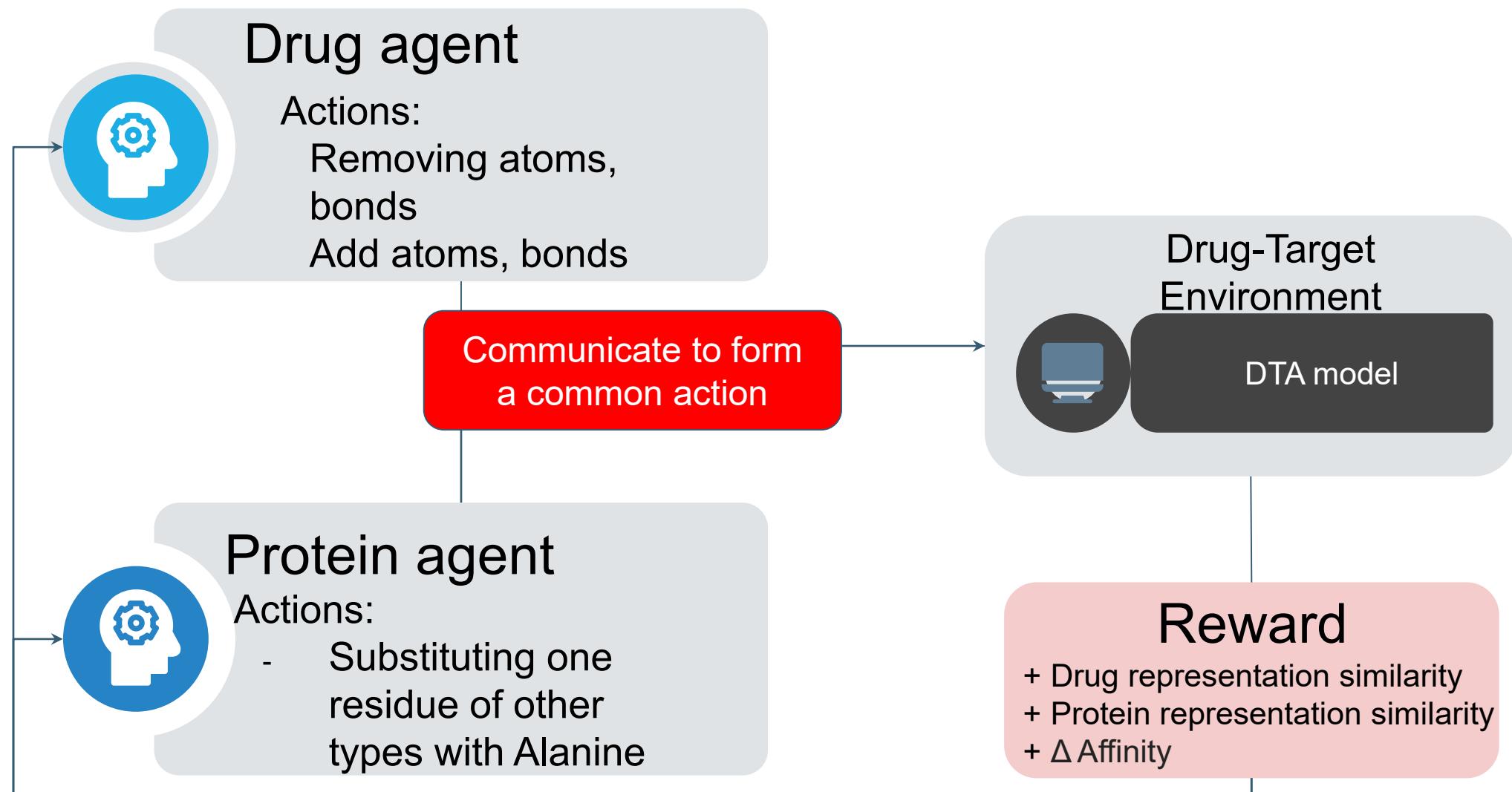
Counterfactual explanation generation



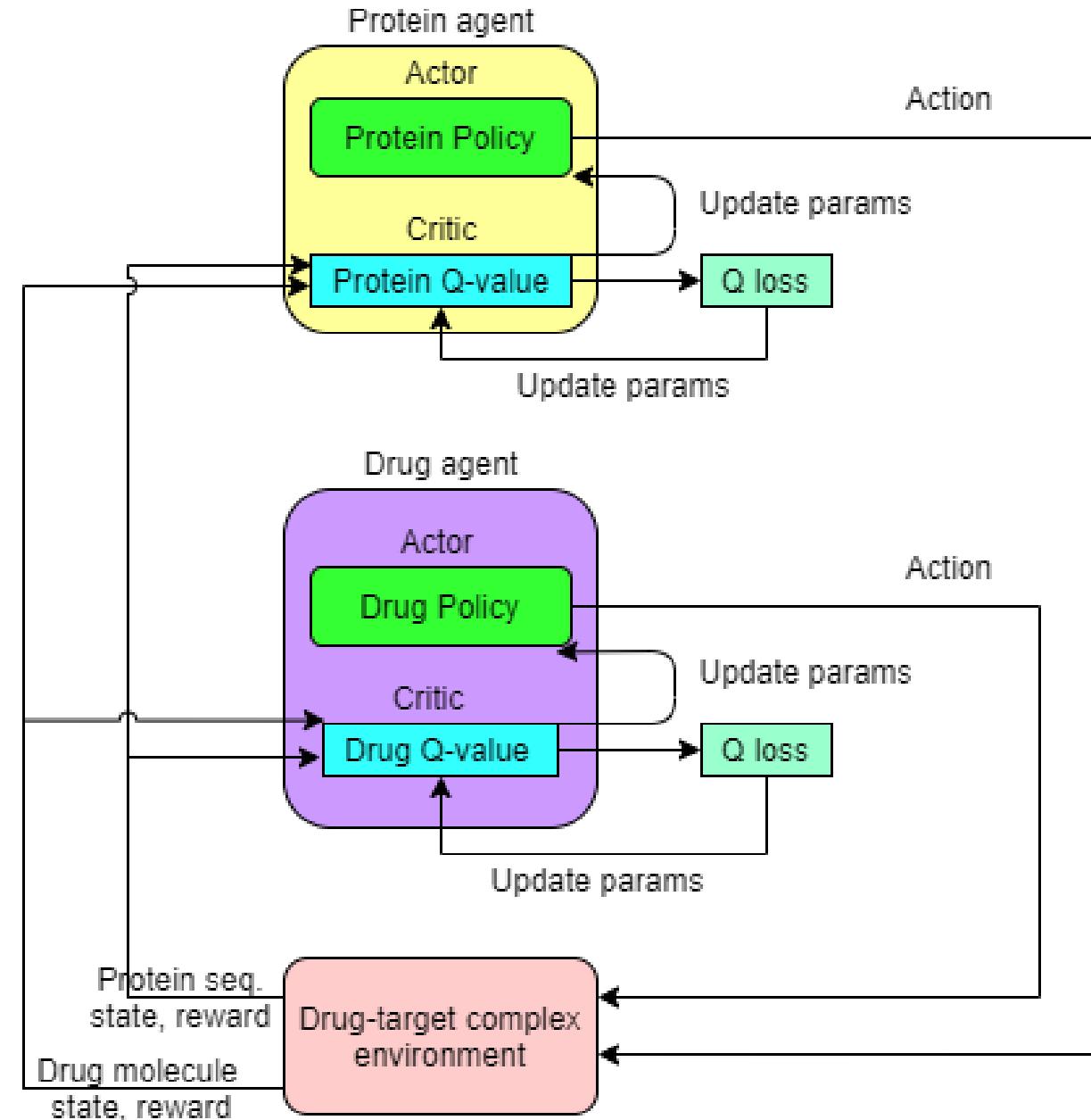
Counterfactual explanation generation objective



MACDA: MultiAgent Counterfactual Drug-target Affinity framework



MACDA: MultiAgent Counterfactual Drug-target Affinity framework



Tutorial at ECML-PKDD, September 17th /18th 2021



Machine Learning and Reasoning for Drug Discovery

Part C1: Molecular optimisation & generation

Truyen Tran, Thin Nguyen and Tri Nguyen
{truyen.tran, thin.nguyen, minhtri}@deakin.edu.au

<https://bit.ly/3Edqgz4>

Agenda

Molecular optimisation

- Bayesian optimisation in latent space
- Goal-directed reinforcement learning

Generative molecular generation

- Deep generative models for molecules
- Recurrent models for molecules

The three basic questions in drug discovery

Given a molecule, is this drug? Aka properties/targets/effects prediction.

- Drug-likeness
- Targets it can modulate and how much
- Its dynamics/kinetics/effects/metabolism if administered orally or via injection

Given a target, what are molecules?

- If the list of molecules is given, pick the good one. If evaluation is expensive, need to search, e.g., using Bayesian Optimization.
- If no molecule is found, need to generate from scratch → generative models + Bayesian Optimization, or Reinforcement Learning.
- How does the drug-like space look like?

Given a molecular graph, what are the steps to make the molecule?

- Synthetic tractability
- Reaction planning, or retrosynthesis

Traditional combinatorial chemistry

Generate variations on a template

Returns a list of molecules from this template that

- Bind to the pocket with good pharmacodynamics?
- Have good pharmacokinetics?
- Are synthetically accessible?

#REF: Talk by Chloé-Agathe Azencott titled “Machine learning for therapeutic research”, 12/10/2017

Exploring the space of drugs

The space of drugs is estimated to be $1e+23$ to $1e+60$

- Only $1e+8$ substances synthesized thus far.
- It is impossible to model this space fully.

The current technologies for graph generations are constantly refined.

- Search-based: Start from somewhere, search for better graphs (need no data, but need reliable graph evaluator)
- Generative models: Build an ambitious model of the chemical space (needs lot of data).
- Combination of both.

Drug design as structured machine translation, aka conditional generation

Can be formulated as structured machine translation:

- Inverse mapping of (knowledge base + binding properties) to (query) → One to many relationship.

Representing graph as string (e.g., SMILES), and use sequence VAEs or GANs.

Generative graph models

- Model nodes & interactions
- Model cliques

Sequences

- Iterative methods

Reinforcement learning

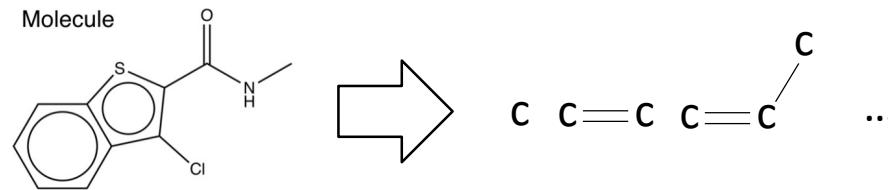
- Discrete objectives

Any combination of these + memory.

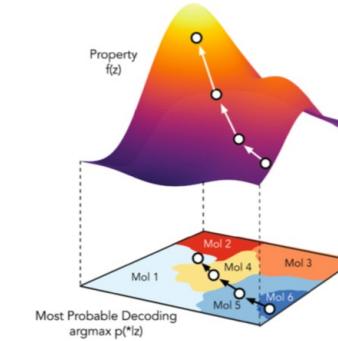
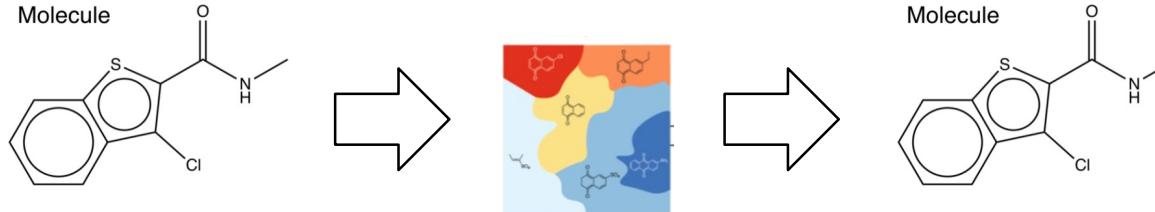
Molecular optimisation

We optimize a starting molecule towards desirable properties. Often we need to balance among multiple objectives, including similarity to the original molecule.

Strategy 1: Sequentially move in the discrete chemical space (e.g., atom & bond addition/deletion), making sure the results are chemically valid.



Strategy 2: Mapping the discrete structure into continuous space, search in the latent space, then map back to the discrete space.



Black-box optimisation

This applies to any process that we do not have detailed knowledge but can be sure of objective function.

Typical methods:

- Bayesian optimisation
- Reinforcement learning
- Active learning
- Evolutionary algorithms

Reinforcement learning for molecular optimisation

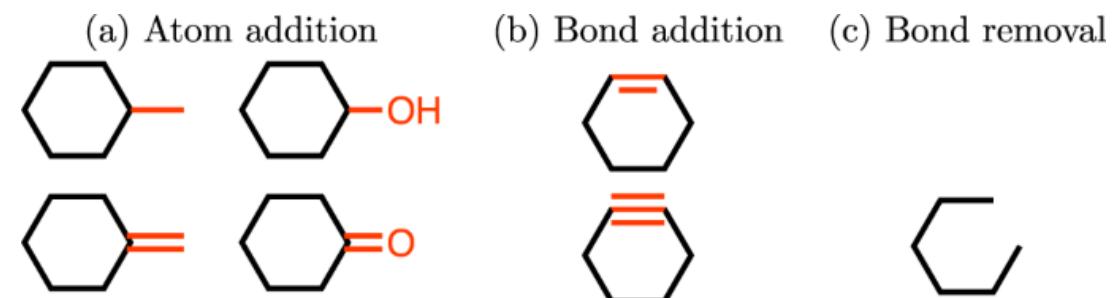
State: Current molecule

Actions: Atom/Bond addition/removal

Rewards: the properties of molecules
(final reward) and chemical validity
(intermediate and final reward)

Learning:

- Policy gradient
- Q-value function with TD learning



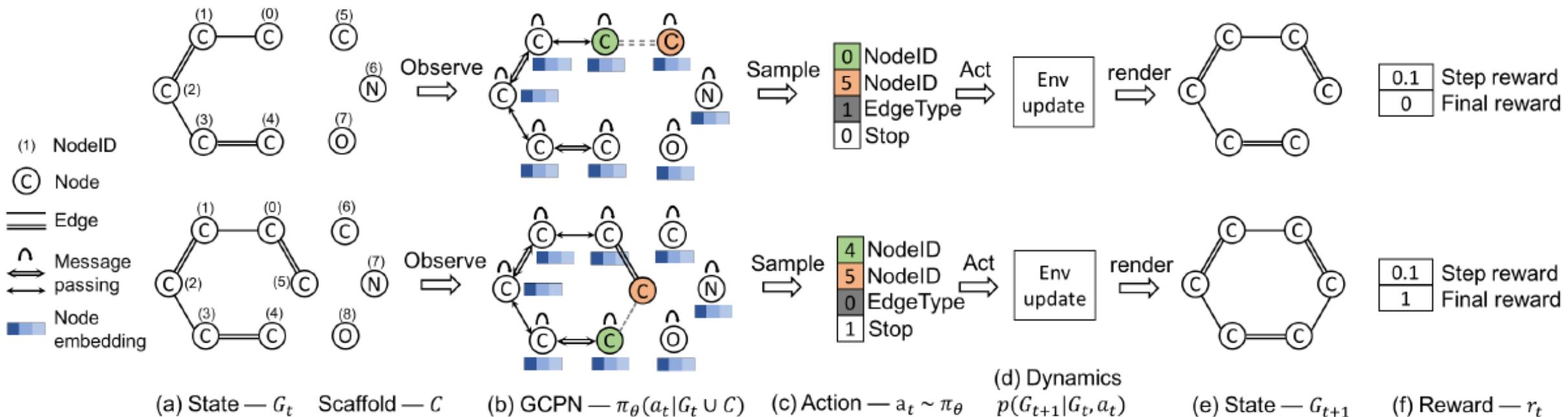
Zhou, Zhenpeng, et al. "Optimization of molecules via deep reinforcement learning." *Scientific reports* 9.1 (2019): 1-10.

Graphs + Reinforcement learning

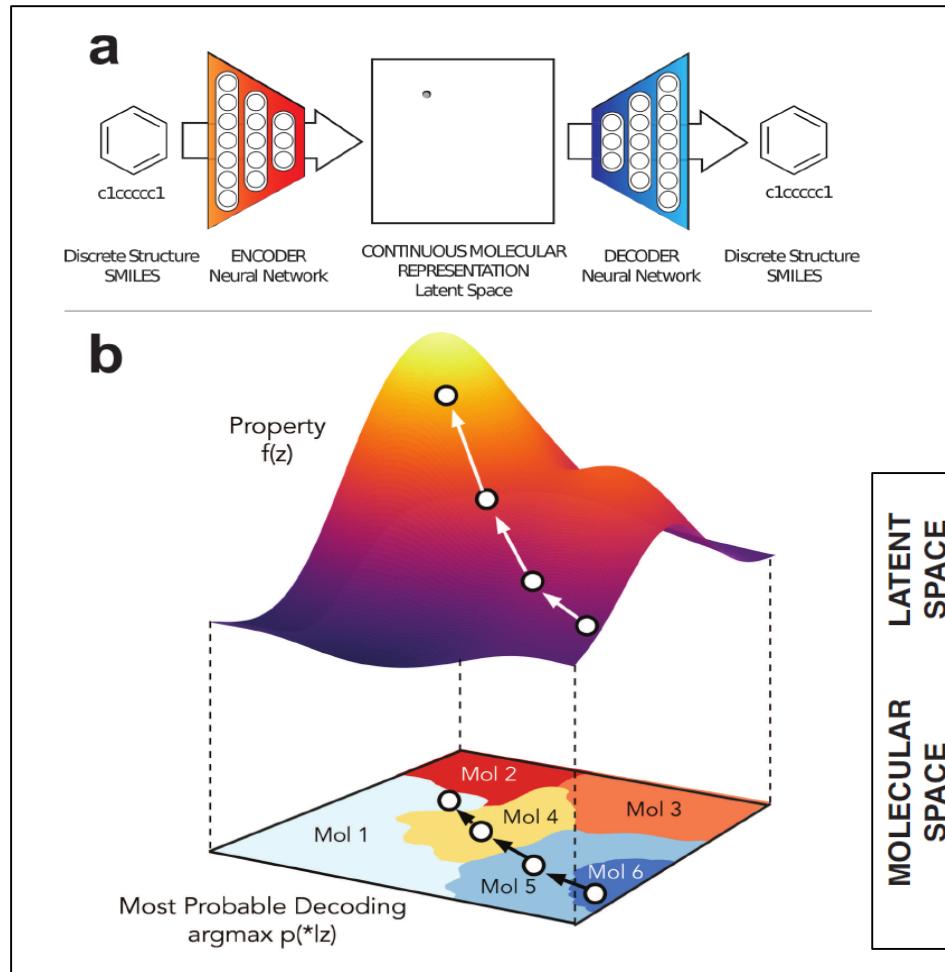
Generative graphs are very hard to get it right: The space is too large!

Reinforcement learning offers step-wise construction: one piece at a time

- A.k.a. Markov decision processes
- As before: Graphs offer properties estimation

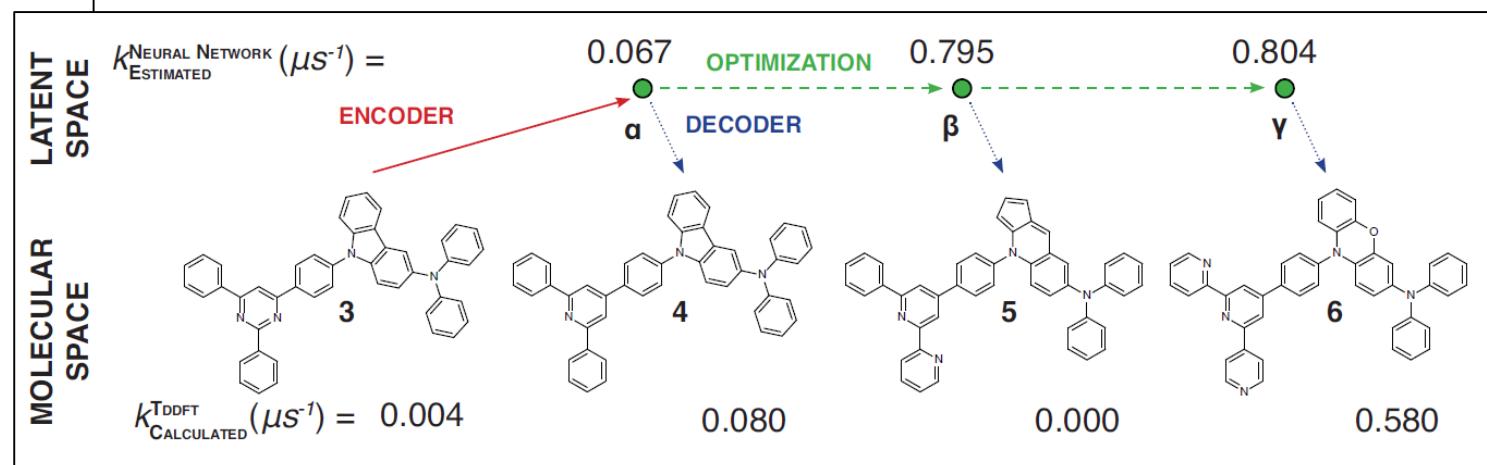


Searching in the latent space

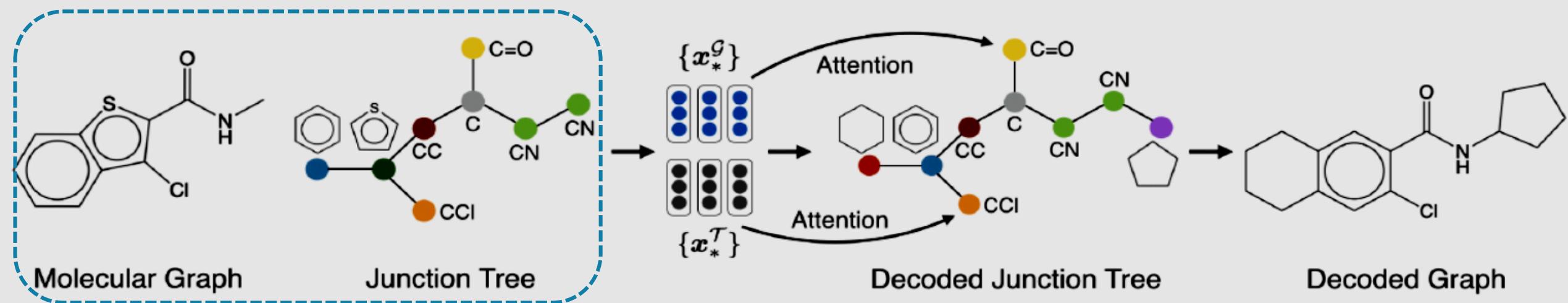


Model: SMILES → VAE+RNN

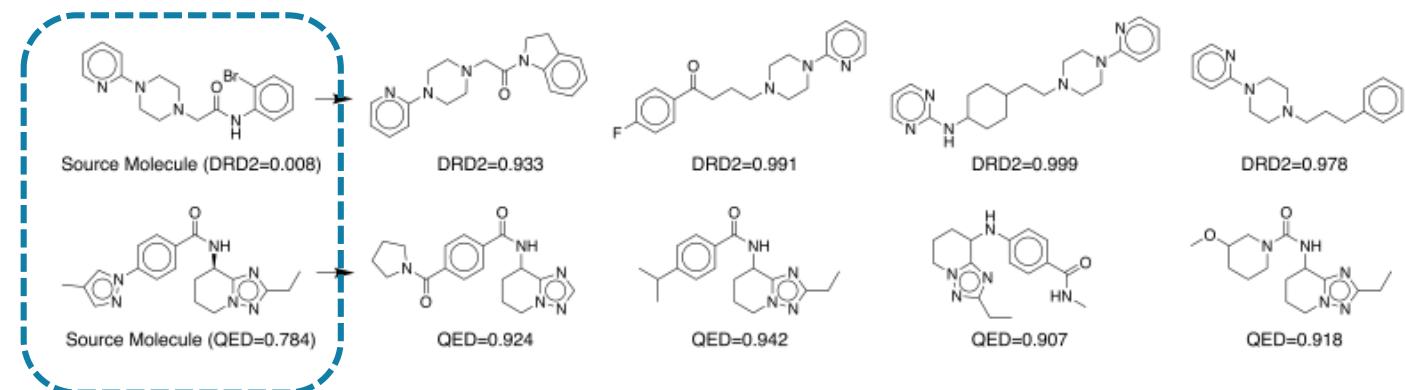
#REF: Gómez-Bombarelli, Rafael, et al.
"Automatic chemical design using a data-driven continuous representation of molecules." *ACS Central Science* (2016).



Molecular optimization as machine translation



- It is easier to modify existing molecules, aka “*molecular paraphrases*”
- Molecular optimization as graph-to-graph translation



#REF: Jin, W., Yang, K., Barzilay, R., & Jaakkola, T. (2019). Learning multimodal graph-to-graph translation for molecular optimization. *ICLR*.

Agenda

Molecular optimisation

- Bayesian optimisation in latent space
- Goal-directed reinforcement learning

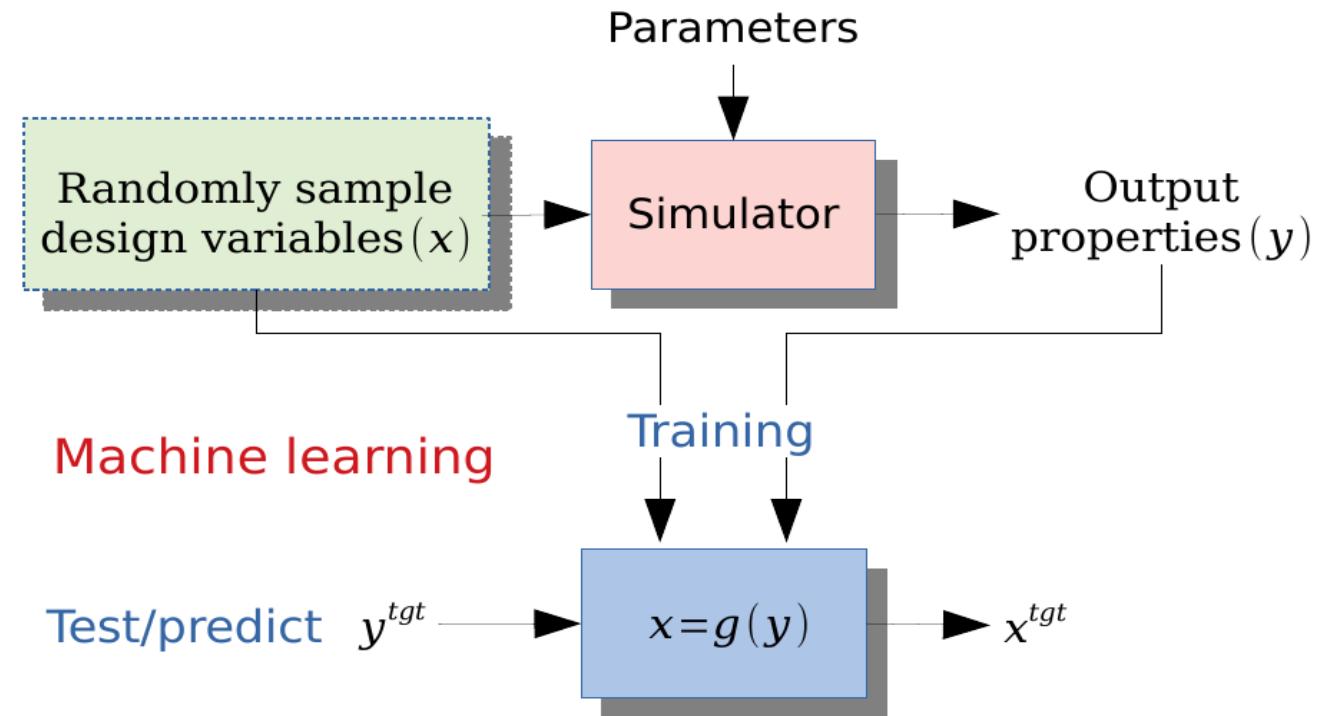
Generative molecular generation

- Deep generative models for molecules
- Recurrent models for molecules

Inverse design

- Leverage the existing data and query the simulators in an offline mode
- Avoid the global optimization by learning the inverse design function $f^{-1}(y)$
- Predict design variables in a single step

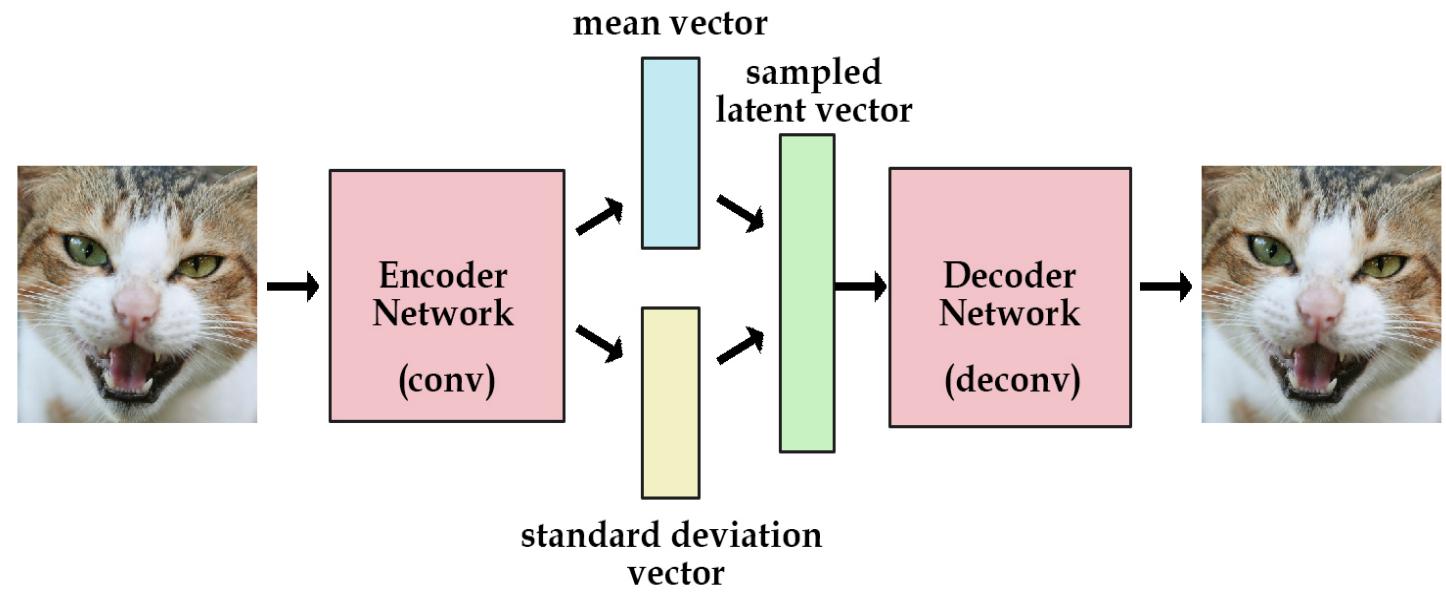
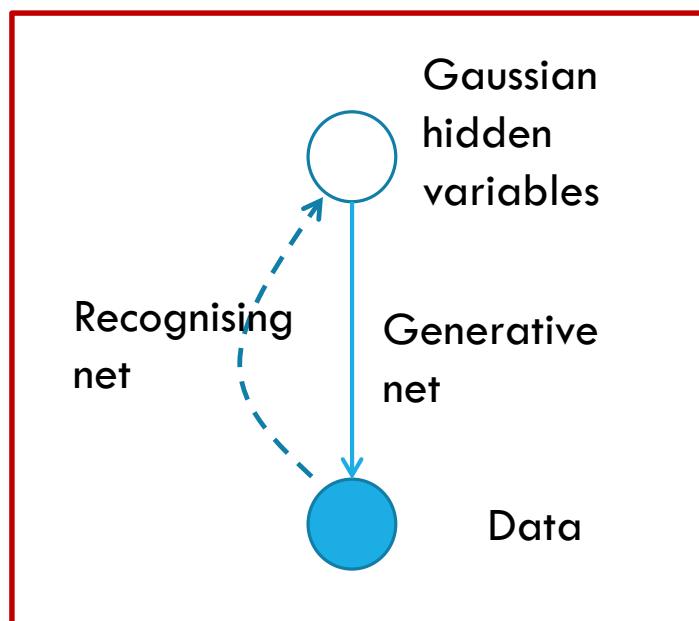
Create dataset offline



$$\mathbf{x}^{\text{target}} = g(\mathbf{y}^{\text{target}})$$

Variational Autoencoder

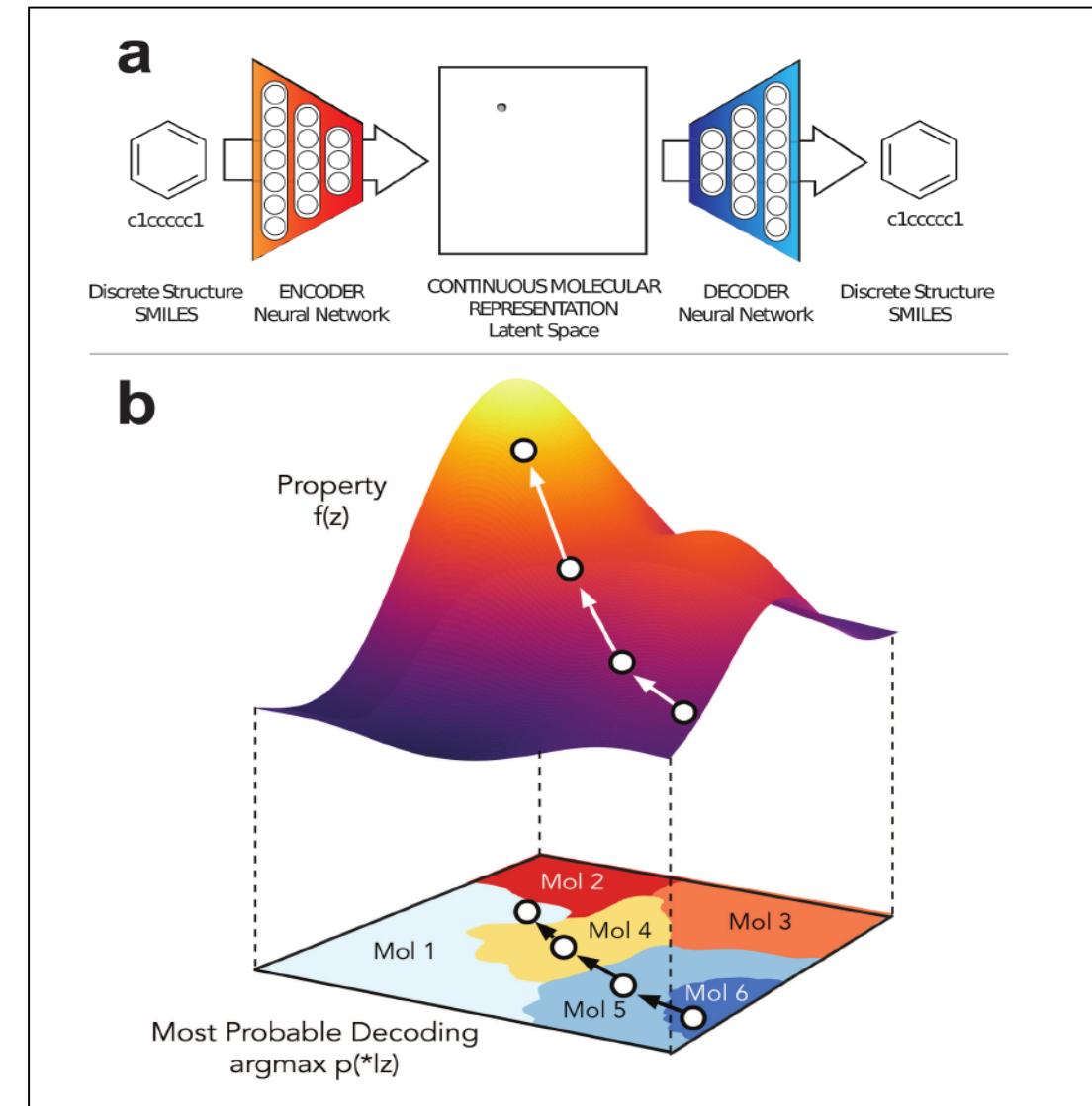
Two separate processes: generative (hidden → visible) versus recognition (visible → hidden)



VAE for drug space modelling

Model: SMILES → VAE+RNN

#REF: Gómez-Bombarelli, Rafael, et al.
"Automatic chemical design using a data-driven continuous representation of molecules." *ACS Central Science* (2016).



GraphVAE

Handles irregular structures

- Predict the whole adjacency matrix, node types and edge types

Deals with variable size graph

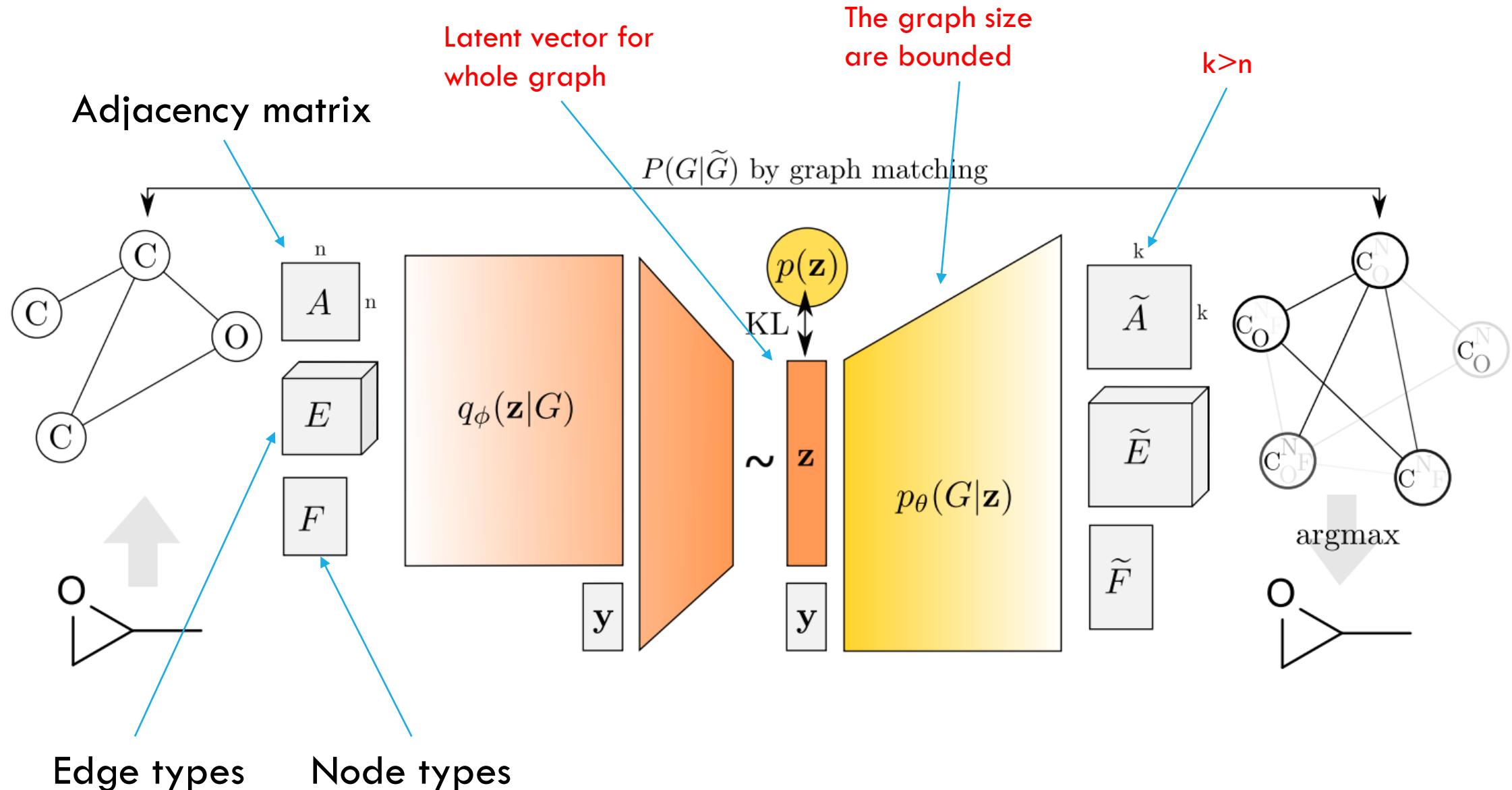
- Bounded by the size of the largest graph in training data.

Handles permutation invariance

- Matching every pair of nodes in 2 graphs

Partially promotes diversity

#REF: Simonovsky, M., & Komodakis, N. (2018). GraphVAE: Towards Generation of Small Graphs Using Variational Autoencoders. *arXiv preprint arXiv:1802.03480*.





Junction tree VAE

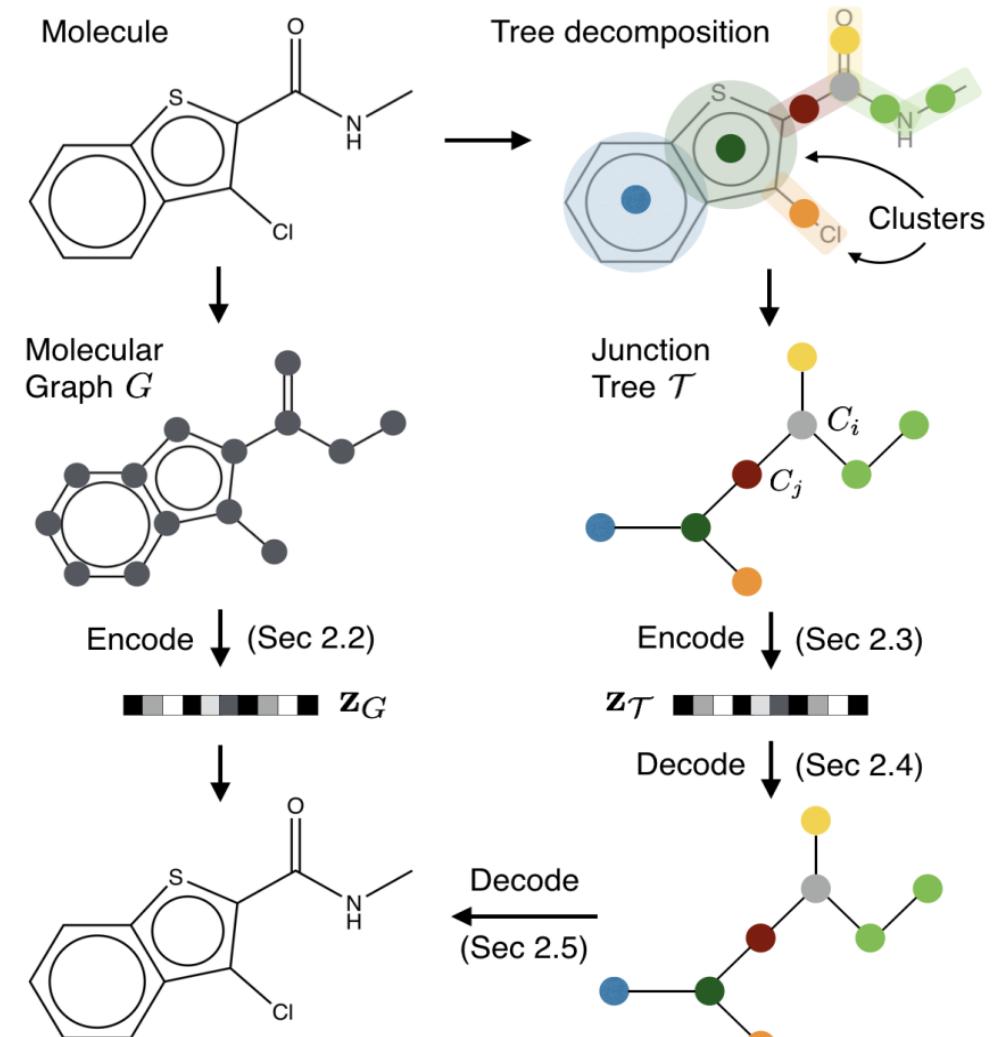
Solve many problems with
SMILES

Junction tree is a way to build
a “thick-tree” out of a graph

Cluster vocab:

- rings
- bonds
- atoms

Jin, W., Barzilay, R., & Jaakkola, T. (2018). Junction Tree
Variational Autoencoder for Molecular Graph
Generation. ICML'18.



Algorithm 2 Tree decomposition of molecule $G = (V, E)$

$V_1 \leftarrow$ the set of bonds $(u, v) \in E$ that do not belong to any rings.

$V_2 \leftarrow$ the set of simple rings of G .

for r_1, r_2 **in** V_2 **do**

Merge rings r_1, r_2 into one ring if they share more than two atoms (bridged rings).

end for

$V_0 \leftarrow$ atoms being the intersection of three or more clusters in $V_1 \cup V_2$.

$\mathcal{V} \leftarrow V_0 \cup V_1 \cup V_2$

$\mathcal{E} \leftarrow \{(i, j, c) \in \mathcal{V} \times \mathcal{V} \times \mathbb{R} \mid |i \cap j| > 0\}$. Set $c = \infty$ if $i \in V_0$ or $j \in V_0$, and $c = 1$ otherwise.

Return The maximum spanning tree over cluster graph $(\mathcal{V}, \mathcal{E})$.

Jin, W., Barzilay, R., & Jaakkola, T. (2018). Junction Tree Variational Autoencoder for Molecular Graph Generation. *ICML'18*.

Method	Reconstruction	Validity
CVAE	44.6%	0.7%
GVAE	53.7%	7.2%
SD-VAE ²	76.2%	43.5%
GraphVAE	-	13.5%
JT-VAE	76.7%	100.0%

GraphRNN

A case of graph dynamics: nodes and edges are added sequentially.

Solve tractability using BFS

You, Jiaxuan, et al.

"GraphRNN: Generating realistic graphs with deep auto-regressive models." ICML (2018).

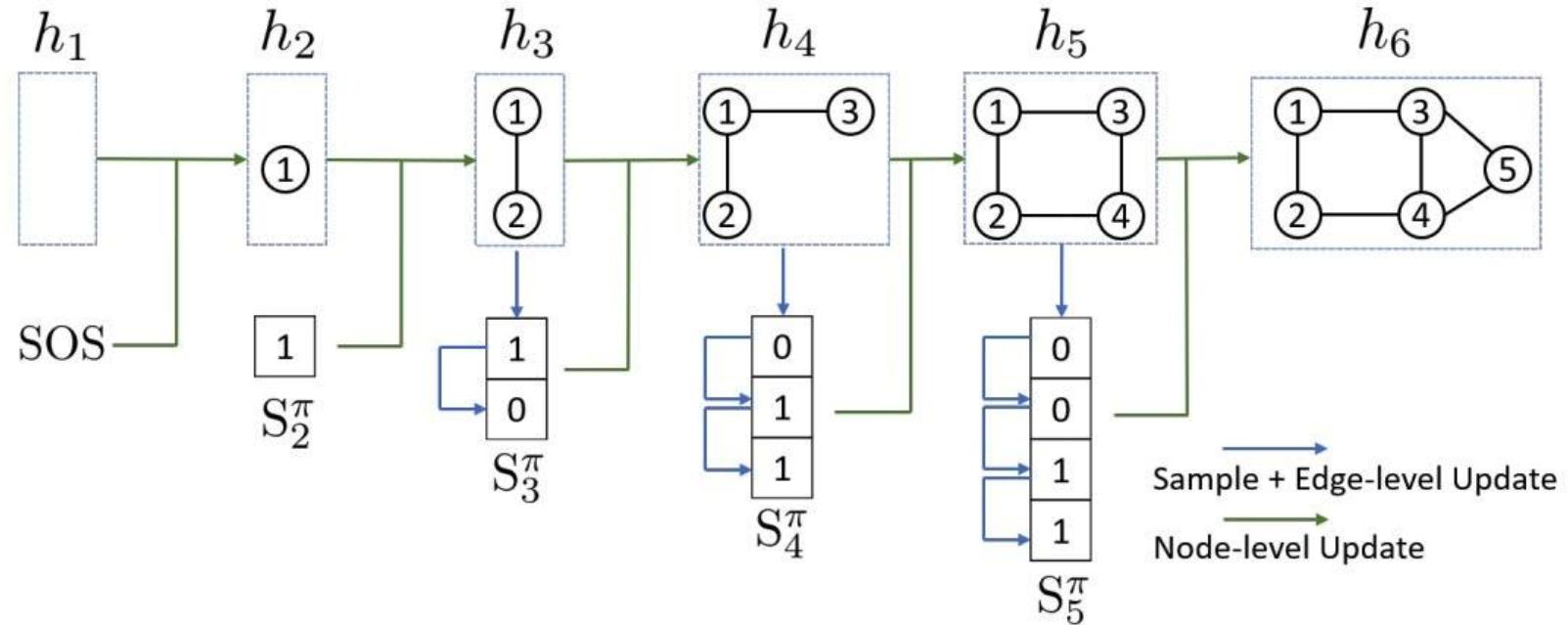


Figure 1. GraphRNN at inference time. Green arrows denote the graph-level RNN that encodes the “graph state” vector h_i in its hidden state, updated by the predicted adjacency vector S_i^π for node $\pi(v_i)$. Blue arrows represent the edge-level RNN, whose hidden state is initialized by the graph-level RNN, that is used to predict the adjacency vector S_i^π for node $\pi(v_i)$.

Problems with VAE + BO style

It is still an interpolation problem

Searching beyond the high density region in the training data will result really bad generation.

We need a more intrinsic exploration strategy

- Compositionality
- Grammar/syntax
- Network that generates generative networks (on going)

Grammar VAE

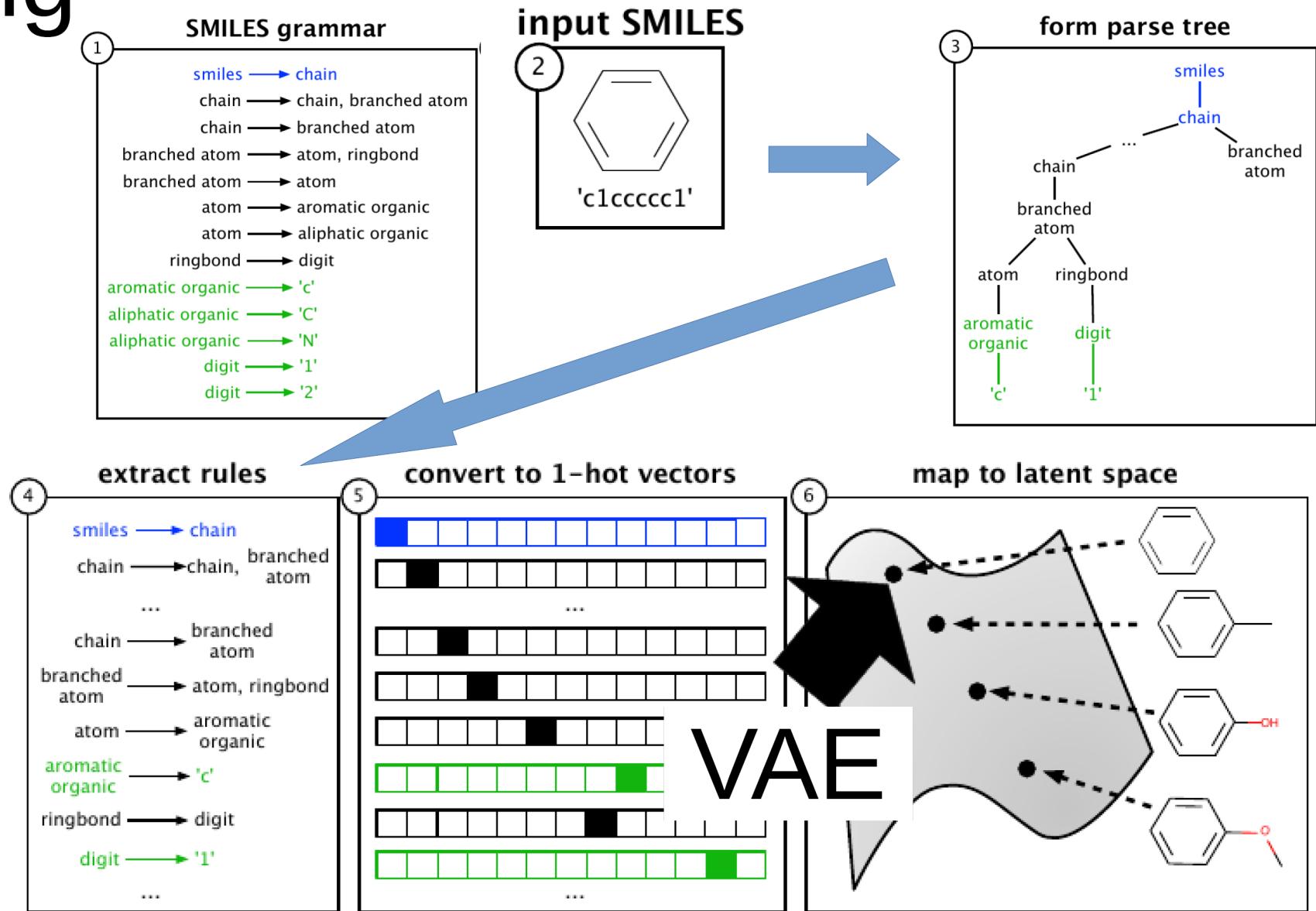
Proposed method

- 1) Define a grammar
- 2) Take a valid sequence and parse it into a sequence of production rules
- 3) Learn a VAE that produces sequences of grammar production rules
- 4) Use this VAE to generate valid sequences of production rules
- 5) Applying these rules in order will yield the original sequence

Kusner, Matt J., Brooks Paige, and José Miguel Hernández-Lobato. "Grammar variational autoencoder." *International Conference on Machine Learning*. PMLR, 2017.

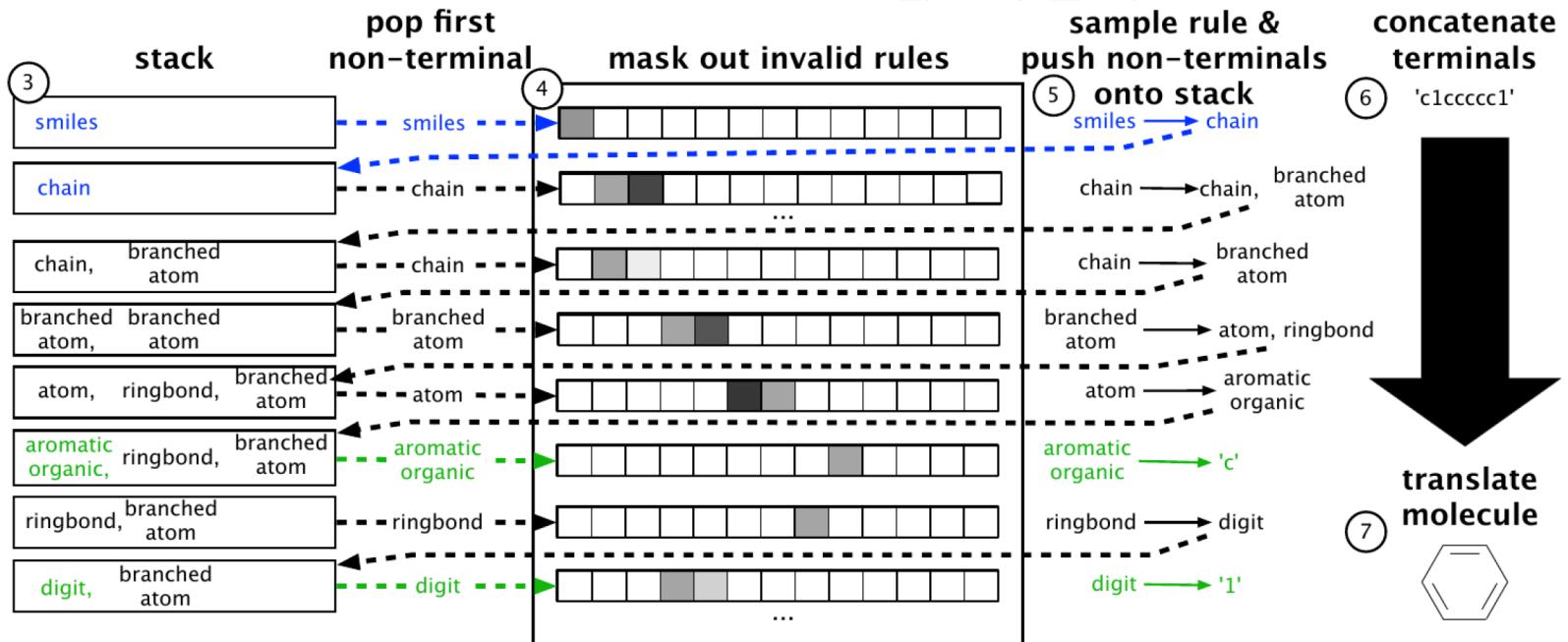
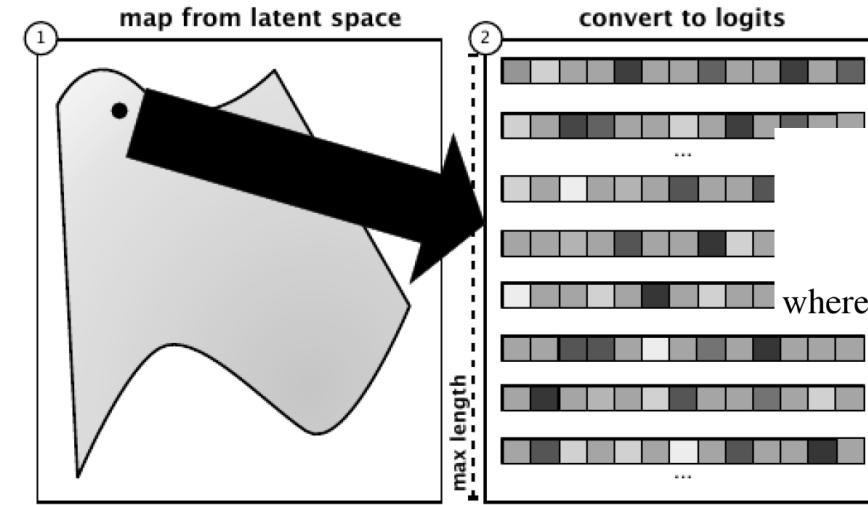
Encoding

Grammar VAE (2)



Decoding

Grammar VAE (3)



Other works

Shi, Chence, et al. "GraphAF: a Flow-based Autoregressive Model for Molecular Graph Generation." *International Conference on Learning Representations*. 2019.

Mahmood, Omar, et al. "Masked graph modeling for molecule generation." *Nature communications* 12.1 (2021): 1-12.

Tutorial at ECML-PKDD, September 17th /18th 2021



Machine Learning and Reasoning for Drug Discovery

Part C2: Knowledge graph reasoning & Drug synthesis

Truyen Tran, Thin Nguyen and Tri Nguyen
{truyen.tran, thin.nguyen, minhtri}@deakin.edu.au

<https://bit.ly/3Edqgz4>

Agenda

Reasoning on biomedical knowledge graphs

- Recommendation
- Drug repurposing

Retrosynthesis

- Chemical planning
- Chemical reaction as graph morphism

Wrapping up

Drug repurposing

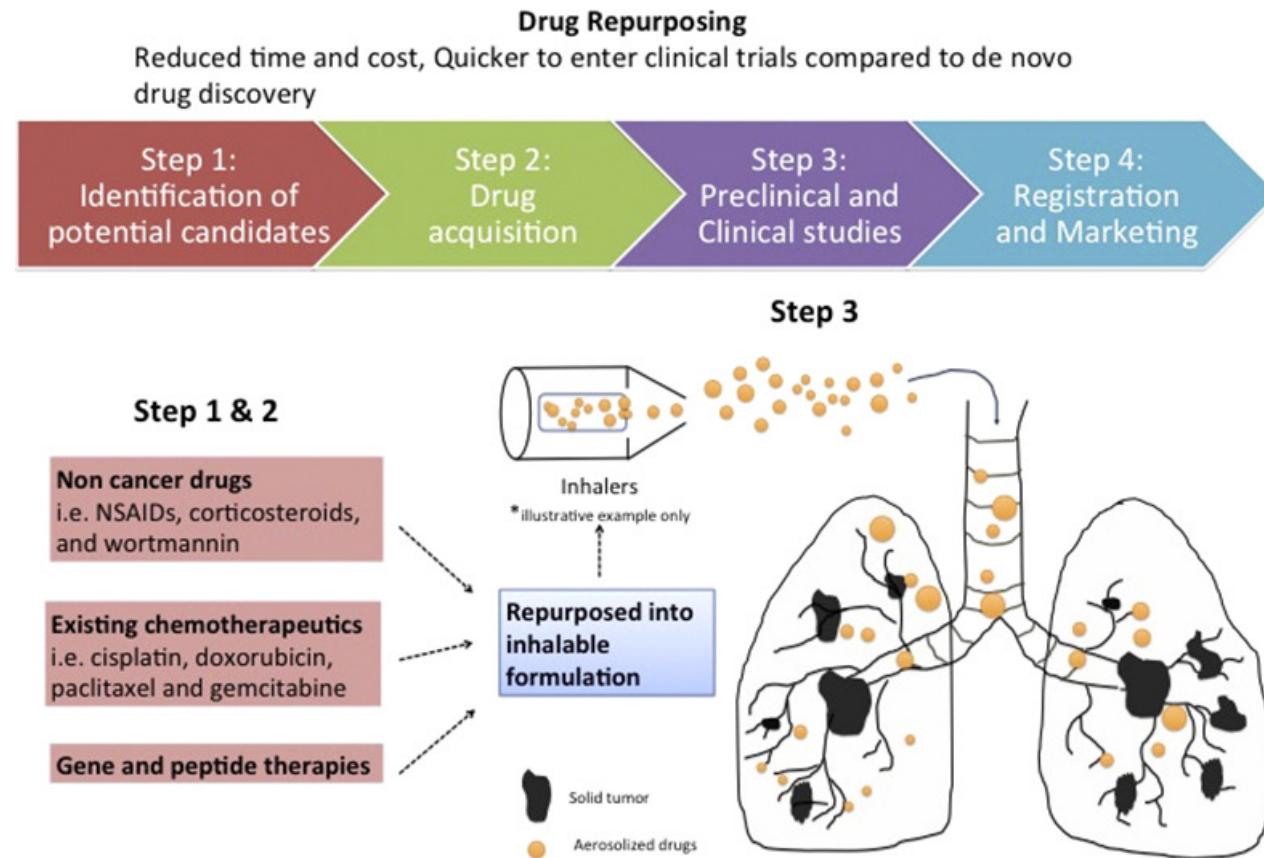
De novo drug discovery is costly, takes long time without guarantee of success

Market for rare diseases is too small to warrant commercial development

Urgent, new diseases like COVID-19 can't wait

Drug repurposing (aka drug repositioning, reprofiling or re-tasking) is one of the best ways to go.

- Finding new uses for approved or investigational drugs designed for other purposes.



Lee, Wing-Hin, et al. "The potential to treat lung cancer via inhalation of repurposed drugs." *Advanced drug delivery reviews* 133 (2018): 107-130.

Approaches

Signature matching, similarity-based (drug-disease, drug-drug, adverse effect profile)

Detailed drug-protein binding prediction

Pathway/network mapping → Knowledge graph inference, e.g., link prediction

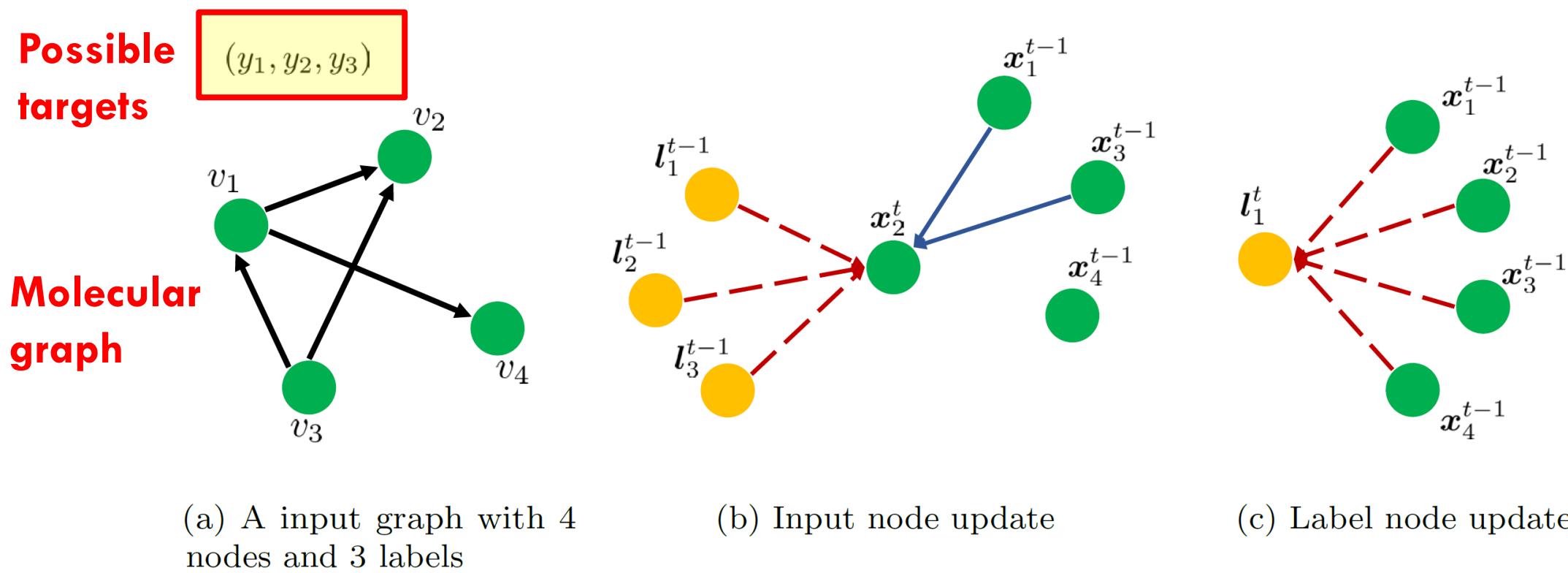
Retrospective clinical analysis from electronic health records

Pushpakom, Sudeep, et al. "Drug repurposing: progress, challenges and recommendations." *Nature reviews Drug discovery* 18.1 (2019): 41-58.

"The general genomic layout and the general replication kinetics and the biology of the MERS, SARS and [SARS-CoV-2] viruses are very similar, so testing drugs which target relatively generic parts of these coronaviruses is a logical step".

(Vincent Munster, Chief o US National Institutes of Health Viral Ecology Unit, as of Feb 2020)

Repurposing as multi-target prediction over molecular graph



(a) A input graph with 4 nodes and 3 labels

(b) Input node update

(c) Label node update

#REF: Do, Kien, et al. "Attentional Multilabel Learning over Graphs-A message passing approach." *Machine Learning*, 2019.

Repurposing as knowledge graph completion

TransE

$$E(h, r, t) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_{\ell_{1/2}}$$

TransF

$$\mathbf{M}_{r,h} = \sum_{i=1}^s \alpha_r^{(i)} \mathbf{U}^{(i)} + \mathbf{I}$$

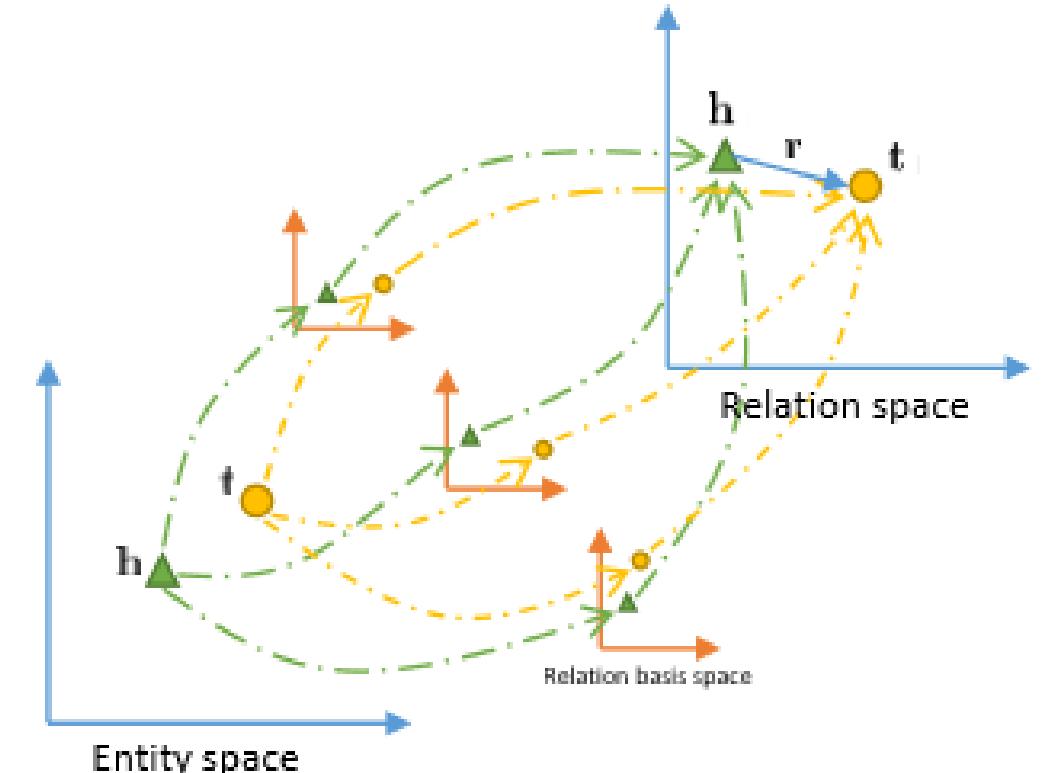
$$\mathbf{M}_{r,t} = \sum_{i=1}^s \beta_r^{(i)} \mathbf{V}^{(i)} + \mathbf{I}$$

$$\mathbf{h}_\perp = \mathbf{M}_{r,h} \mathbf{h}, \quad \mathbf{t}_\perp = \mathbf{M}_{r,t} \mathbf{t}$$

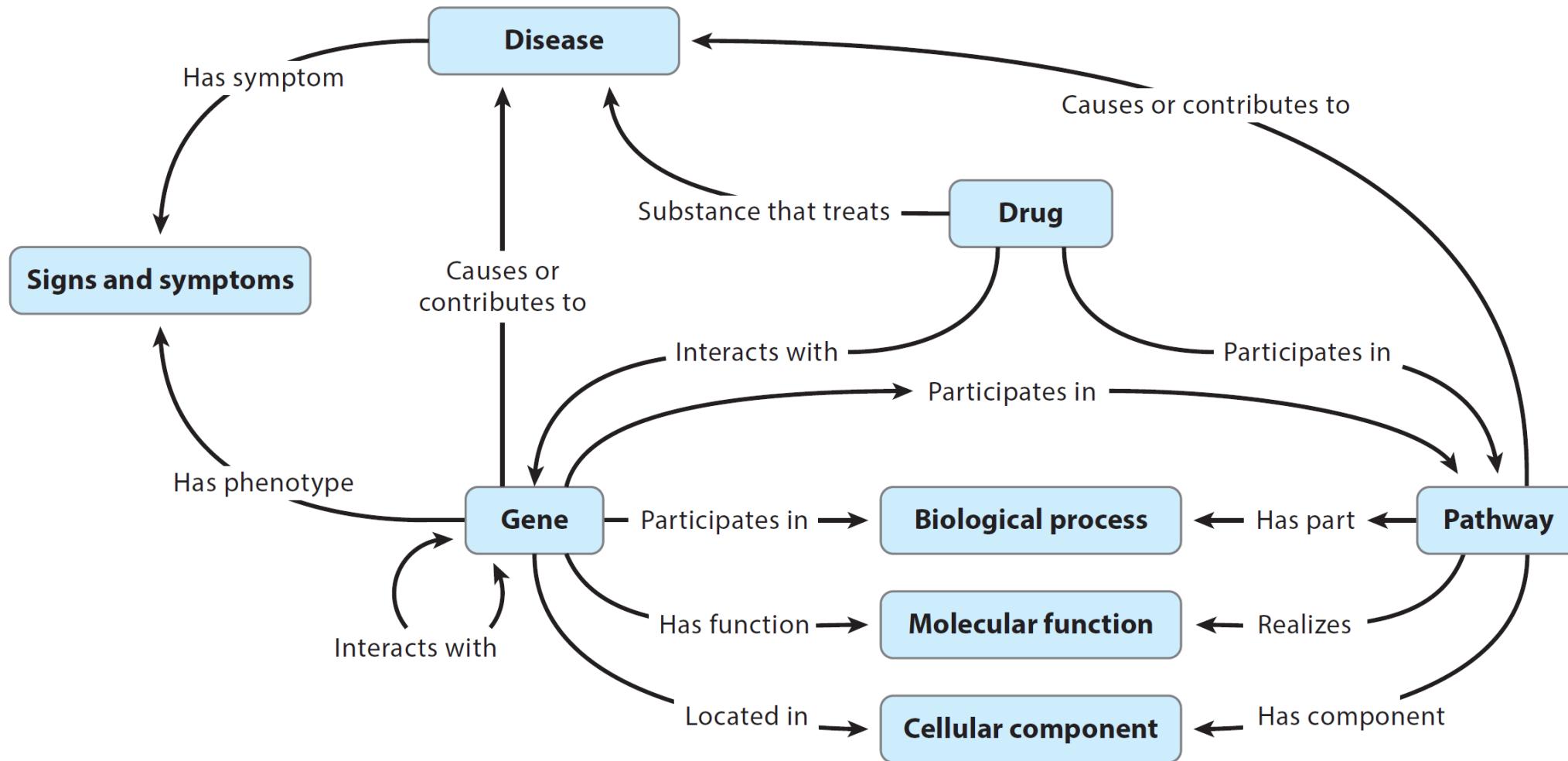
$$E(h, r, t) = \|\mathbf{h}_\perp + \mathbf{r} - \mathbf{t}_\perp\|_{\ell_{1/2}}$$

Loss function $\mathcal{L} = \sum_{(h,r,t) \in \mathcal{T}} \sum_{(h',r',t') \in \mathcal{T}'} [E(h, r, t) + \gamma - E(h', r', t')]_+$

Do, Kien, Truyen Tran, and Svetha Venkatesh. "Knowledge graph embedding with multiple relation projections." 2018 24th International Conference on Pattern Recognition (ICPR). IEEE, 2018.



Biomedical knowledge graphs



Some biomedical knowledge graphs

KG Dataset	Link	Entities	Triples	Entity Types	Relation Types	Contains Features	Constituent Datasets	Version Info	Last Update
Hetionet [56]	https://het.io/	47K	2.2M	11	24	X	29	X	2017
DRKG [65]	https://github.com/gnn4dr/DRKG	97K	5.7M	13	107	X	34	X	2020
BioKG [151]	https://github.com/dsi-bdi/biokg	105K	2M	10	17	categorical	13	X	2020
PharmKG [164]	https://github.com/MindRank-Biotech/PharmKG	7.6K	500K	3	29	continuous	7	X	2020
OpenBioLink [14]	https://zenodo.org/record/3834052	184K	4.7M	7	30	X	17	X	2020
Clinical Knowledge Graph [124]	https://data.mendeley.com/datasets/mrcf7f4tc2/1	16M	220M	35	57	X	35	X	2020

Examples of ontologies suitable for drug discovery

Ontology Name	Entities Covered	Classes	Average # of children	Classes with no definition	Number of Properties	Max Depth	License
Monarch Disease Ontology (MonDO)	Diseases	24K	5	8K	25	16	Creative Commons
Experimental Factor Ontology (EFO)	Diseases	28K	6	7K	66	20	Apache 2.0
Orphanet Rare Disease Ontology (ORDO)	Rare Diseases	15K	17	8.5K	24	11	Creative Commons
Medical Subject Headings (MeSH)	Medical Terms	300K	4	270K	38	15	UMLS License
Human Phenotype Ontology (HPO)	Disease Phenotype	19K	3	6.5K	0	16	HPO License
Disease Ontology (DO)	Diseases	19K	4	8K	89	33	Creative Commons
Drug Target Ontology (DTO)	Drug Targets	10K	4	3K	43	11	Creative Commons
Gene Ontology (GO)	Genes	44K	-	-	11	-	Creative Commons

Biomedical knowledge graph construction

- Nodes: Terms within biological ontologies
- Edge: Relationship between terms. Some notable entities relationship resources:

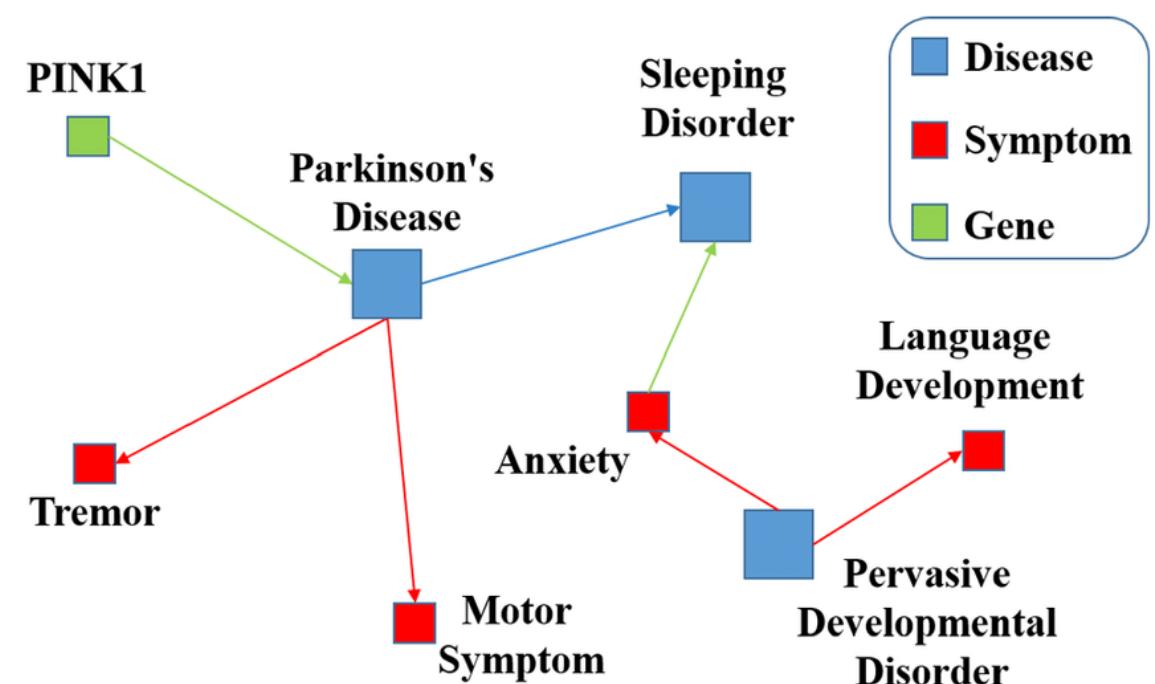
Dataset	First Released	Update Frequency	Updated < 1 Year Ago	Curation Method	Primary Domain	Summary
STRING	2003	Monthly	✓	Expert & Automated	Protein/Gene Interactions	One of the most commonly used sources for physical and functional protein-protein interactions in existing KGs.
BioGRID	2003	Monthly	✓	Expert	Biological Interactions	Contains interactions between gene, protein and chemical entities with could be included directly in a KG.
IntAct	2003	Monthly	✓	Expert	Molecular Interactions	Contains molecular reactions between gene, protein and chemical entities. Uses UniProt for identifiers.
OmniPath	2016	> Annually	✓	Expert	Pathways	An integrator of interaction resources that could be included in a KG via its RDF version.
Pathway Commons	2010	Biannually	✓	Expert & Automated	Pathways	A collection of many resources, including the others discussed in this table.

Table 4: Primary data sources relating to interactions.

#REF: Bonner, S., Barrett, I. P., Ye, C., Swiers, R., Engkvist, O., Bender, A., ... & Hamilton, W. (2021). A review of biomedical datasets relating to drug discovery: A knowledge graph perspective. arXiv preprint arXiv:2102.10062.

Reasoning with KGs

- Biomedical knowledge graph for drug discovery
- Use case of biomedical knowledge in:
 - Polypharmacy prediction
 - Drug-target interaction prediction
 - Gene-Disease prioritisation



#REF: Yuan, J., Jin, Z., Guo, H., Jin, H., Zhang, X., Smith, T., & Luo, J. (2020). Constructing biomedical domain-specific knowledge graph with minimum supervision. *Knowledge and Information Systems*, 62(1), 317-336.

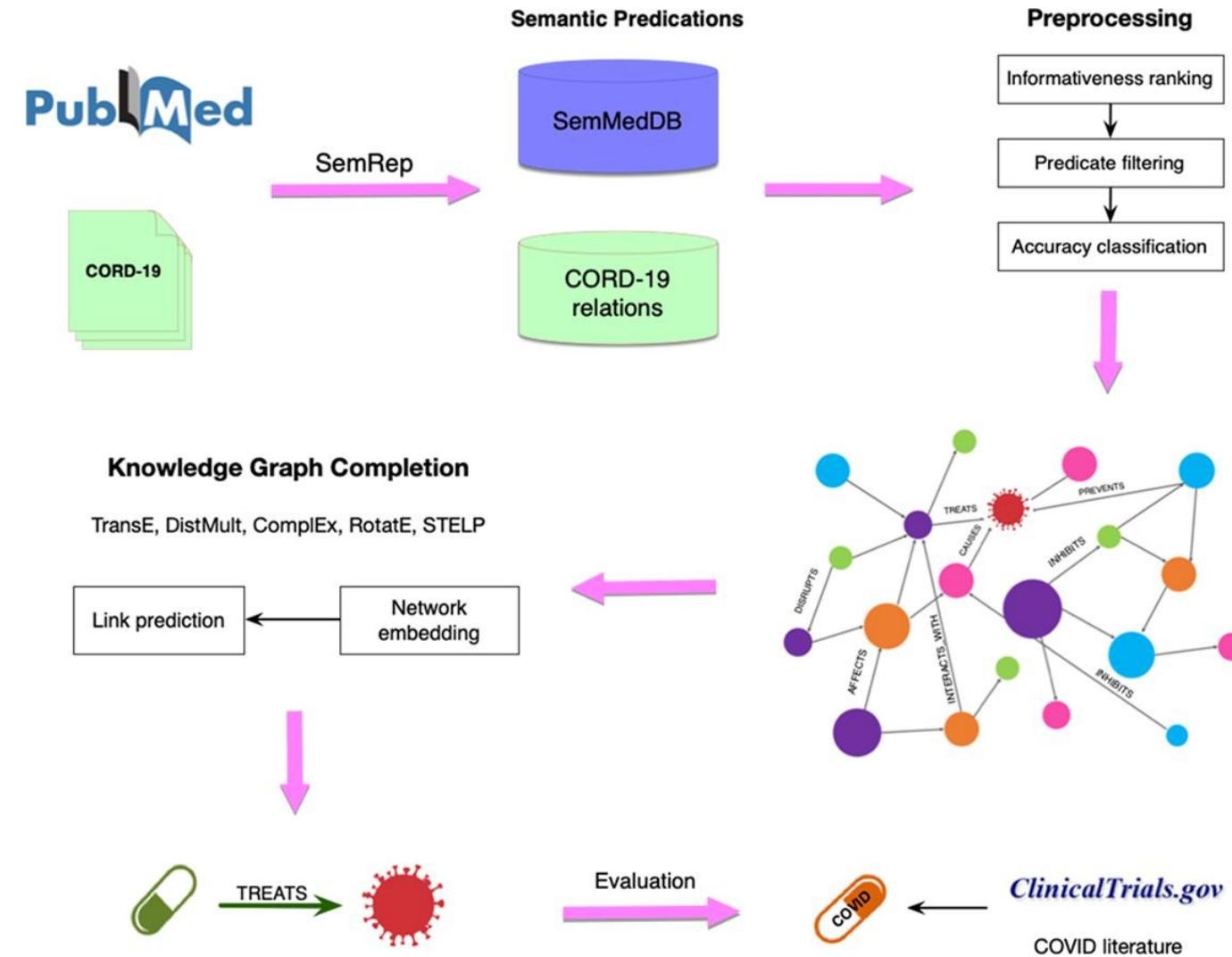
Example: COVID-19 drug repurposing using link prediction

Biomedical knowledge graph built from PubMed and COVID-19 research literature.

Knowledge graph completion methods for drug repurposing.

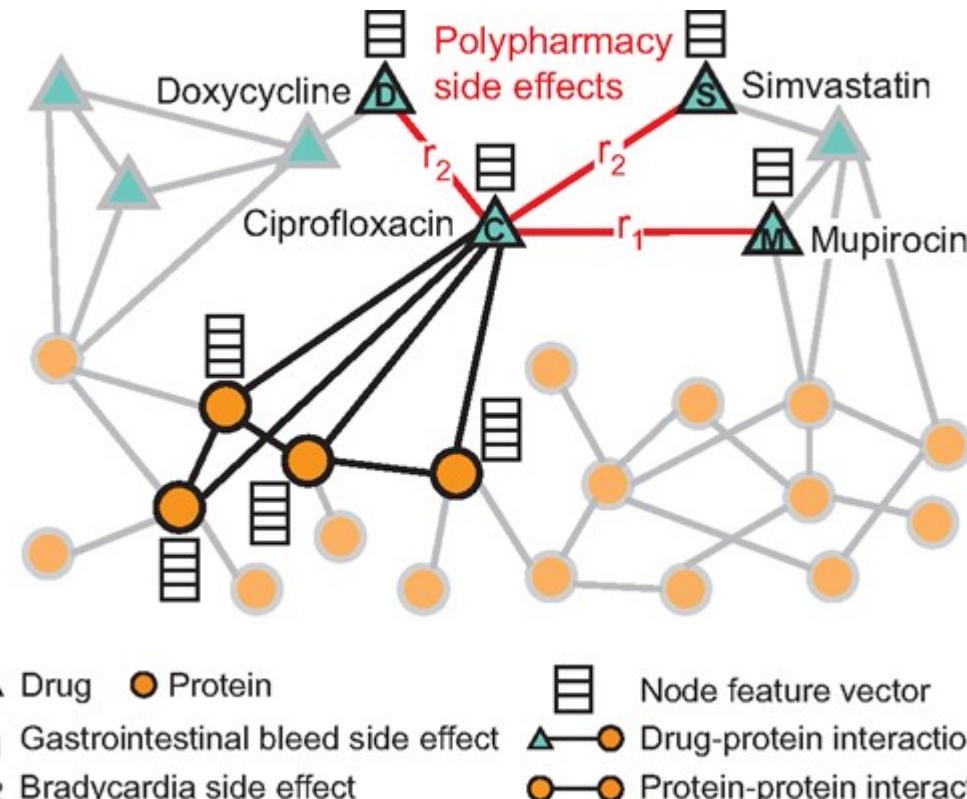
Aim for novel drug recommendation.

Zhang, Rui, et al. "Drug repurposing for COVID-19 via knowledge graph completion." *Journal of biomedical informatics* 115 (2021): 103696.



Polypharmacy prediction

Predicting the adverse side effect when using one or more drugs simultaneously.

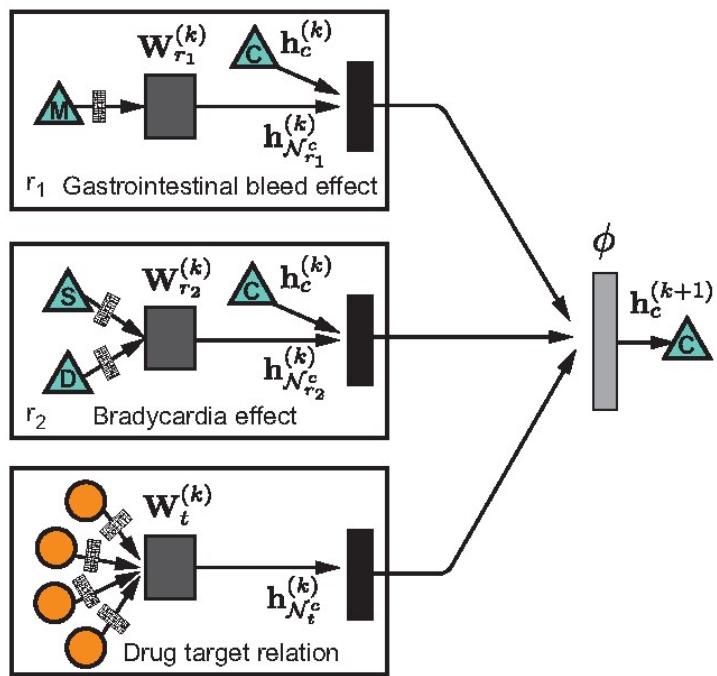


#REF: Zitnik, M., Agrawal, M., & Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics, 34(13), i457-i466.

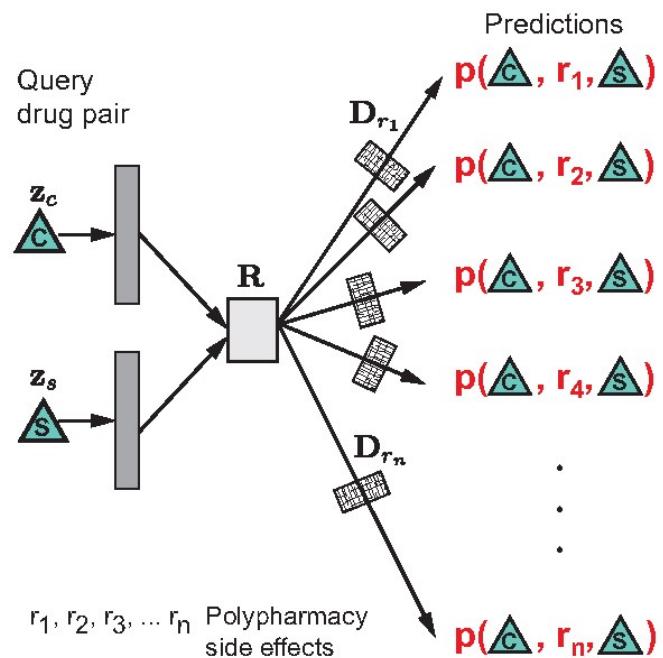
Polypharmacy prediction

Decagon (Zitnik,M et.al)

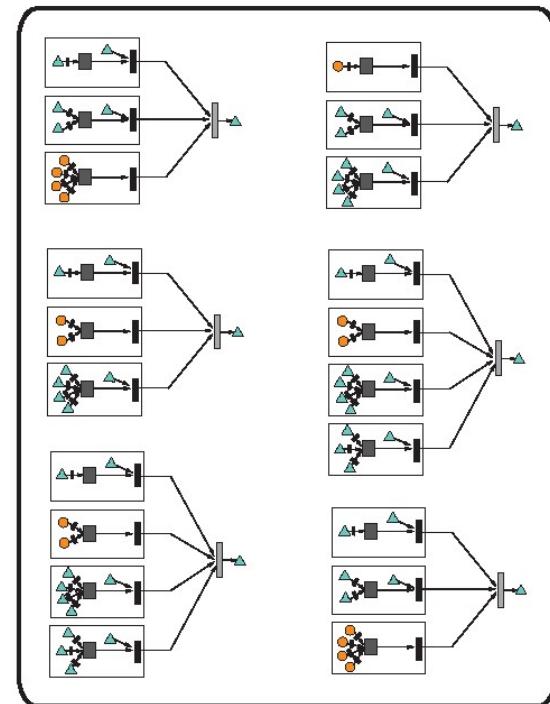
A GCN per-layer update for a single drug node (in blue)



B Polypharmacy side effect prediction



C A batch of networks for six drugs

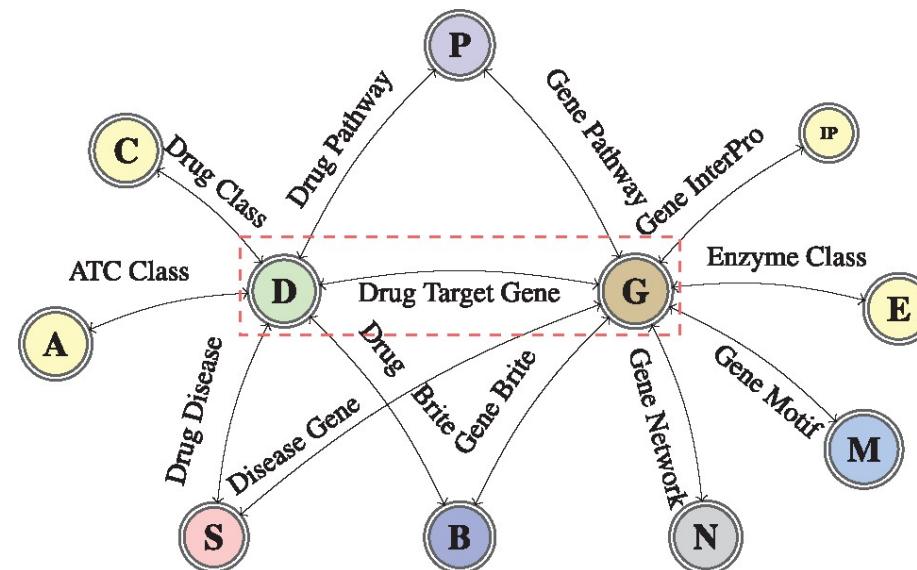


#REF: Zitnik, M., Agrawal, M., & Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics, 34(13), i457-i466.

Drug-target interaction prediction

Predicting the unknown interaction between drug and target

TriModel (Mohamed, S. K. et.al) learns a low rank vector representation of knowledge entities and relations.

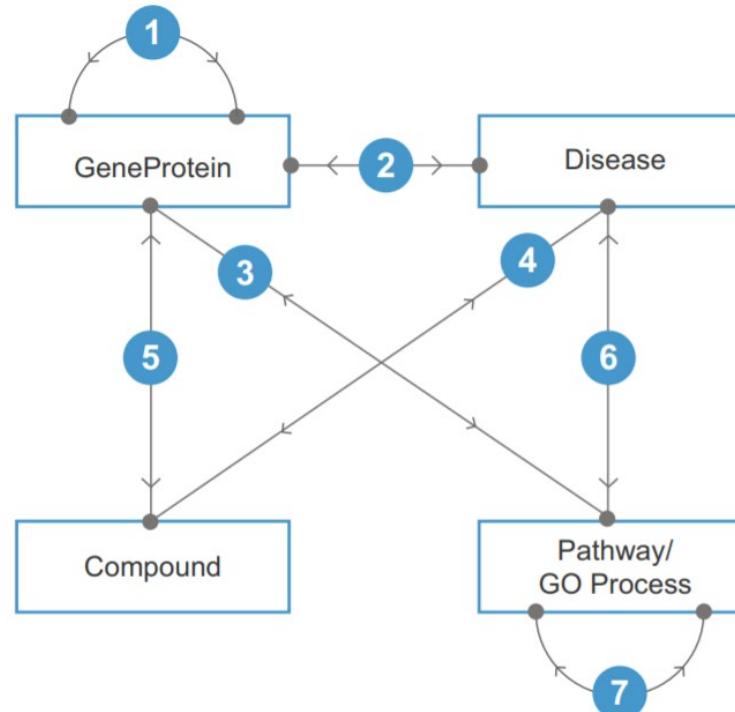


#REF: Mohamed, S. K., Nováček, V., & Nounou, A. (2020). Discovering protein drug targets using knowledge graph embeddings. *Bioinformatics*, 36(2), 603-610.

Gene-disease prioritisation

Predicting the relationship between diseases and molecular entities (proteins and genes).

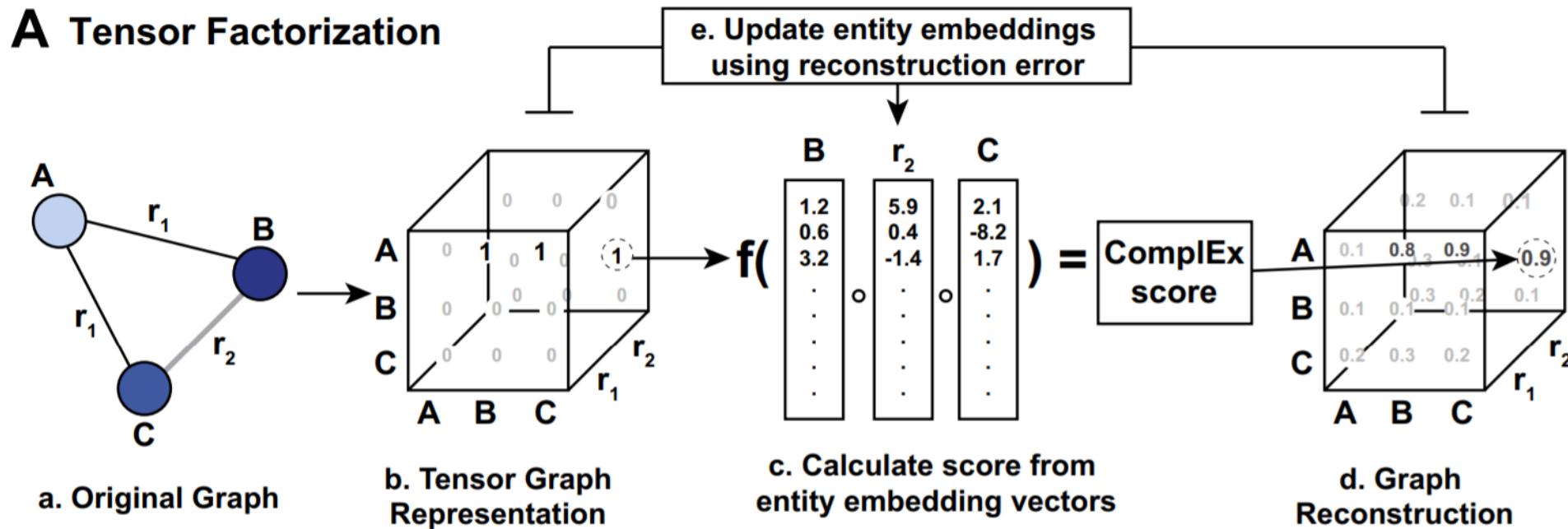
Rosalind (Paliwal et.al) solve the Gene-disease prioritisation as link prediction problem.



#REF: Paliwal, S., de Giorgio, A., Neil, D., Michel, J. B., & Lacoste, A. M. (2020). Preclinical validation of therapeutic targets predicted by tensor factorization on heterogeneous graphs. *Scientific reports*, 10(1), 1-19.

Gene-disease prioritisation

Predicting the relationship between diseases and molecular entities (proteins and genes).



#REF: Paliwal, S., de Giorgio, A., Neil, D., Michel, J. B., & Lacoste, A. M. (2020). Preclinical validation of therapeutic targets predicted by tensor factorization on heterogeneous graphs. *Scientific reports*, 10(1), 1-19.

Agenda

Reasoning on biomedical knowledge graphs

- Recommendation
- Drug repurposing

Retrosynthesis

- Chemical planning
- Chemical reaction as graph morphism

Wrapping up

The three basic questions in drug discovery

Given a molecule, is this drug? Aka properties/targets/effects prediction.

- Drug-likeness
- Targets it can modulate and how much
- Its dynamics/kinetics/effects/metabolism if administered orally or via injection

Given a target, what are molecules?

- If the list of molecules is given, pick the good one. If evaluation is expensive, need to search, e.g., using Bayesian Optimization.
- If no molecule is found, need to generate from scratch → generative models + Bayesian Optimization, or Reinforcement Learning.
- How does the drug-like space look like?

Given a molecular graph, what are the steps to make the molecule?

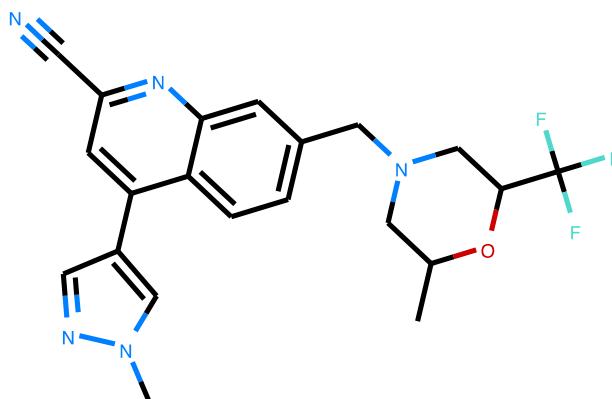
- Synthetic tractability
- Reaction planning, or retrosynthesis

Retrosynthesis prediction

Once a molecular structure is designed, how to synthesize it?

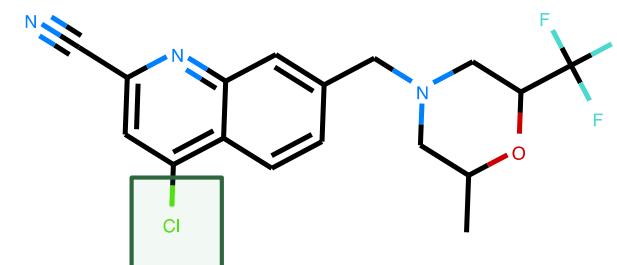
Retrosynthesis planning/prediction

- Identify a set of reactants to synthesize a target molecule

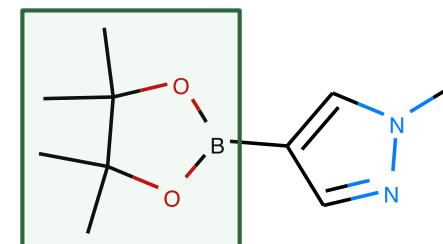


Product (Given)

Predict Reactants
→
Reaction Type
(optional)



Reactant A



Reactant B

...

...

G2Gs Framework for retrosynthesis prediction

Each molecule is represented as a molecular graph

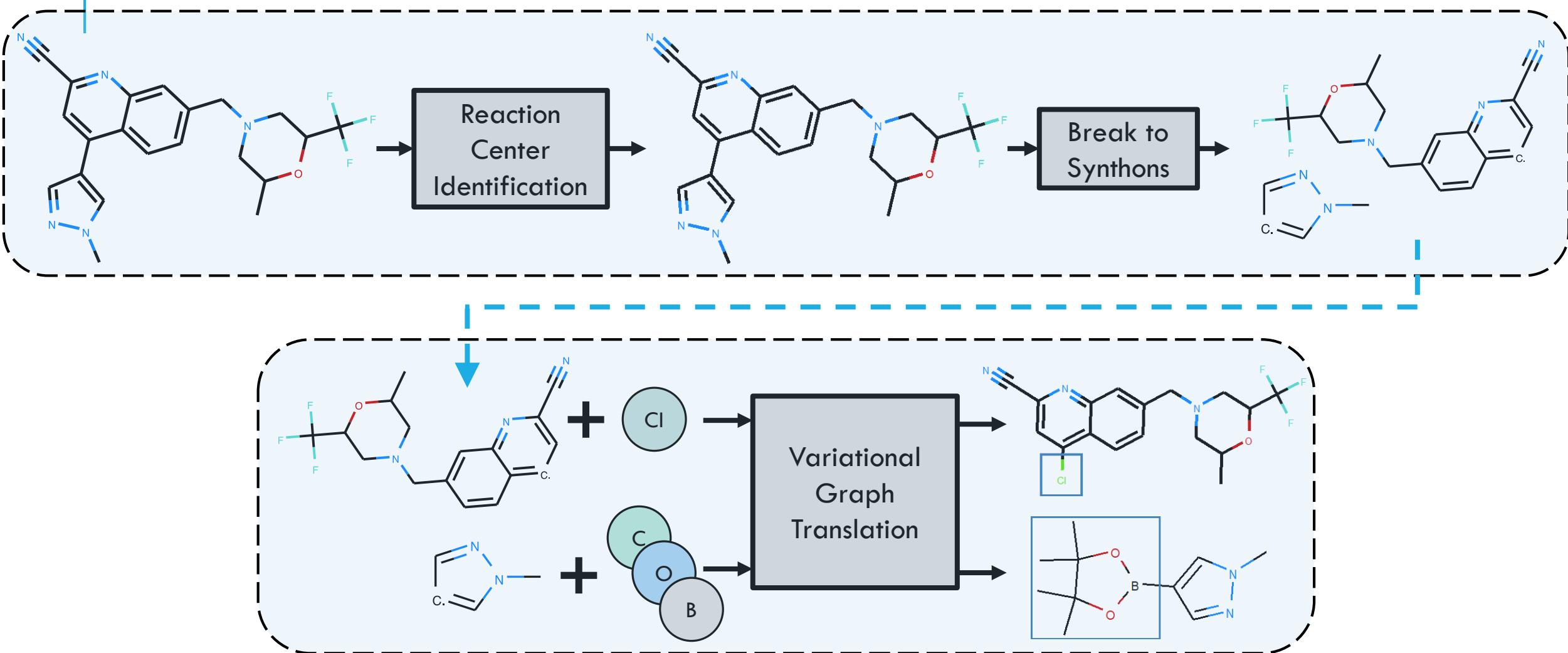
Formulate the problem as a graph (**product molecule**) to a set of graphs (**reactants**)

The whole framework are divided into two stages

- Reaction center identification
- Graph translation

Shi, Chence, et al. "A graph to graphs framework for retrosynthesis prediction." *International Conference on Machine Learning*. PMLR, 2020.

G2Gs Framework for retrosynthesis prediction (cont.)



Reaction prediction as neural graph morphism

Input: A set of graphs = a single big graph with disconnected components

Output: A new set of graphs. Same nodes, different edges.

Model: Graph morphism

Method: Graph transformation policy network (GTPN)

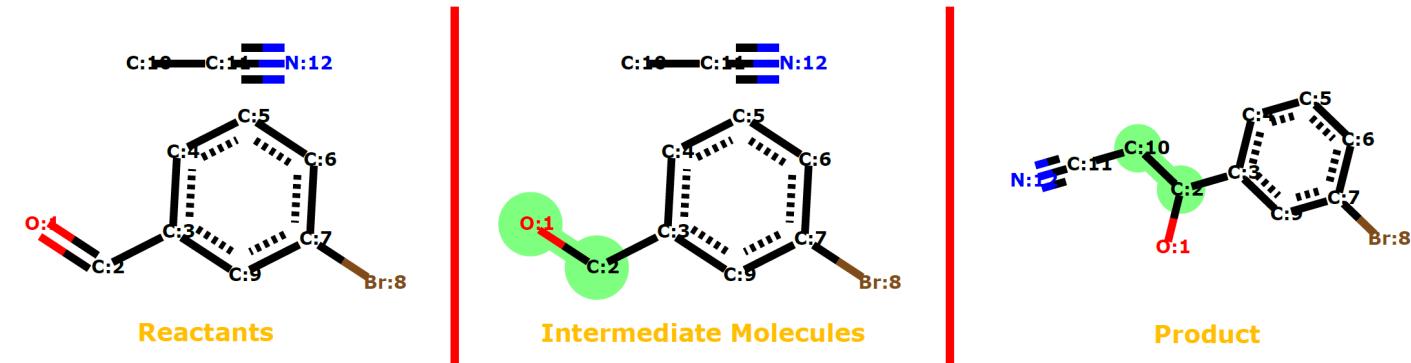
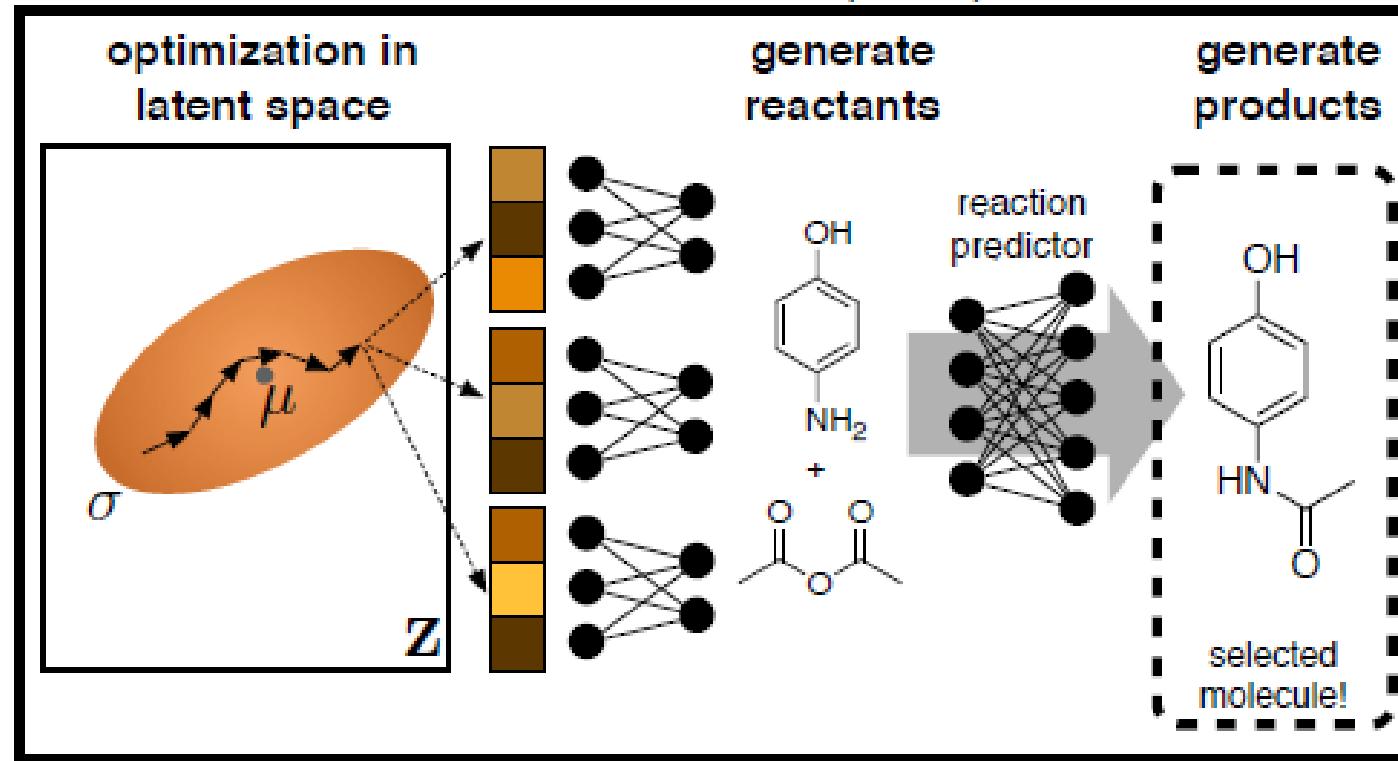


Figure 1: A sample reaction represented as a set of graph transformations from reactants (leftmost) to products (rightmost). Atoms are labeled with their type (Carbon, Oxygen,...) and their index (1, 2,...) in the molecular graph. The atom pairs that change connectivity and their new bonds (if existed) are highlighted in green. There are two bond changes in this case: 1) The double bond between O:1 and C:2 becomes single. 2) A new single bond between C:2 and C:10 is added.

Kien Do, Truyen Tran, and Svetha Venkatesh. "Graph Transformation Policy Network for Chemical Reaction Prediction." *KDD'19*.

Searching for synthesizable molecules



Bradshaw, J., et al. "A model to search for synthesizable molecules." *Advances in Neural Information Processing Systems 32* (2019).

Agenda

Reasoning on biomedical knowledge graphs

- Recommendation
- Drug repurposing

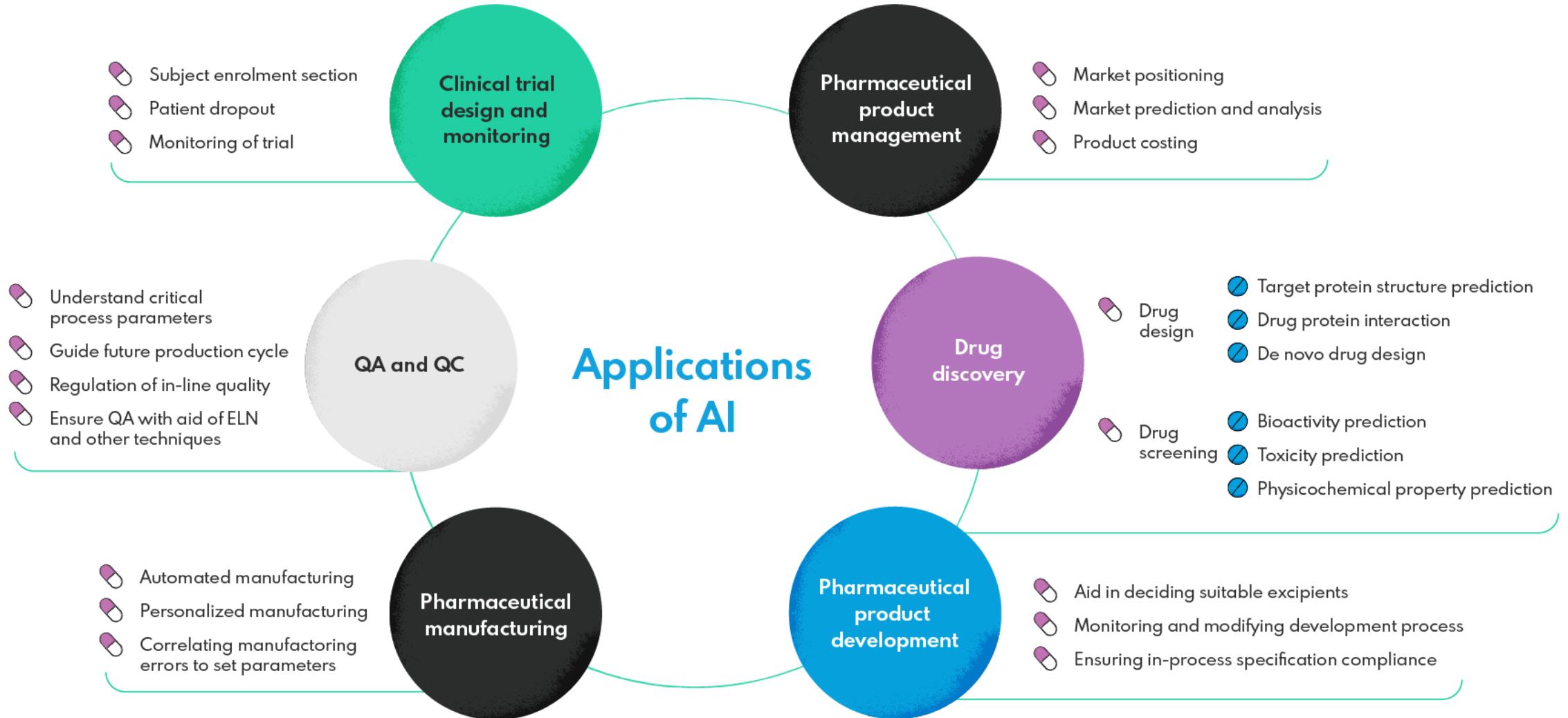
Retrosynthesis

- Chemical planning
- Chemical reaction as graph morphism

Wrapping up

Topics covered

Intro to drug discovery pipeline & ML tasks
Molecular representation and property prediction
Protein representation and protein-drug binding
Molecular optimisation & generation
Drug synthesis & machine reasoning



Source: DARIUSZ JACOSZEK, 2021

<https://nexocode.com/blog/posts/artificial-intelligence-in-drug-discovery-and-development/>

Drug discovery enjoys SOTA ML tools

Attention, transformers & graphs

Deep generative models

Reinforcement learning

Self-supervised learning

Advances in NLP

Reasoning over biomedical knowledge graphs

Much more to be done

Biomed complexity + huge chemical space

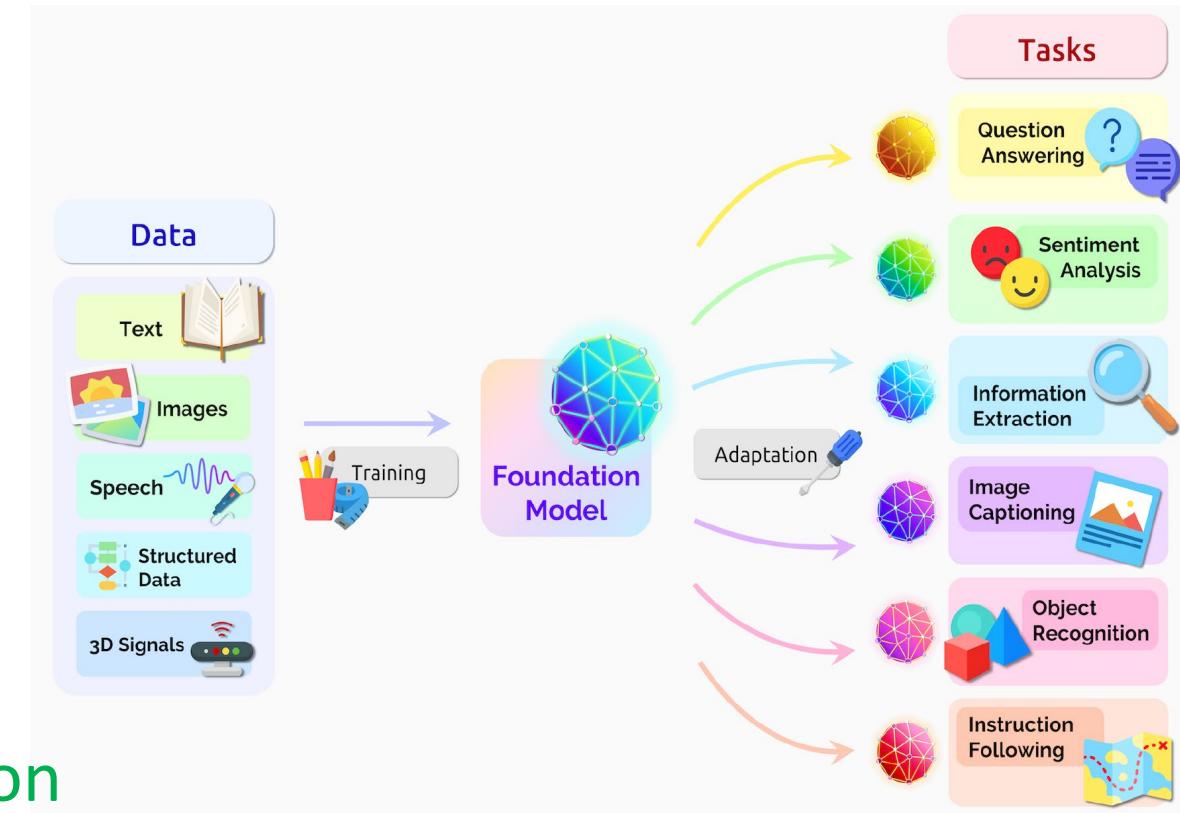
Data quality issues + biases + incompleteness

Huge computational investment

Uncertainty handling

More efficient human-machine co-creation.

Q: Is there any chance for “foundation model” as found in Internet data?



Picture taken from (Bommasani et al, 2021)