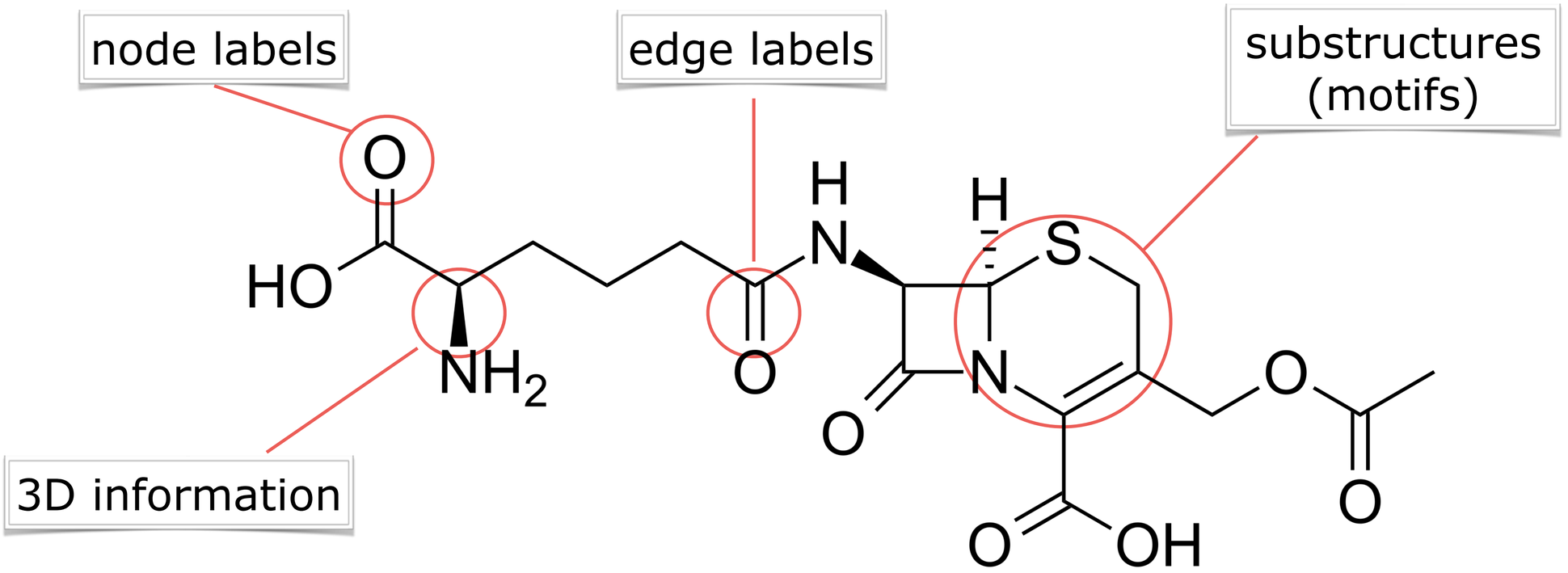
Representation and Generation of Molecular Graphs

Wengong Jin
MIT CSAIL

in collaboration with
Tommi Jaakkola, Regina Barzilay, Kevin Yang, Kyle Swanson

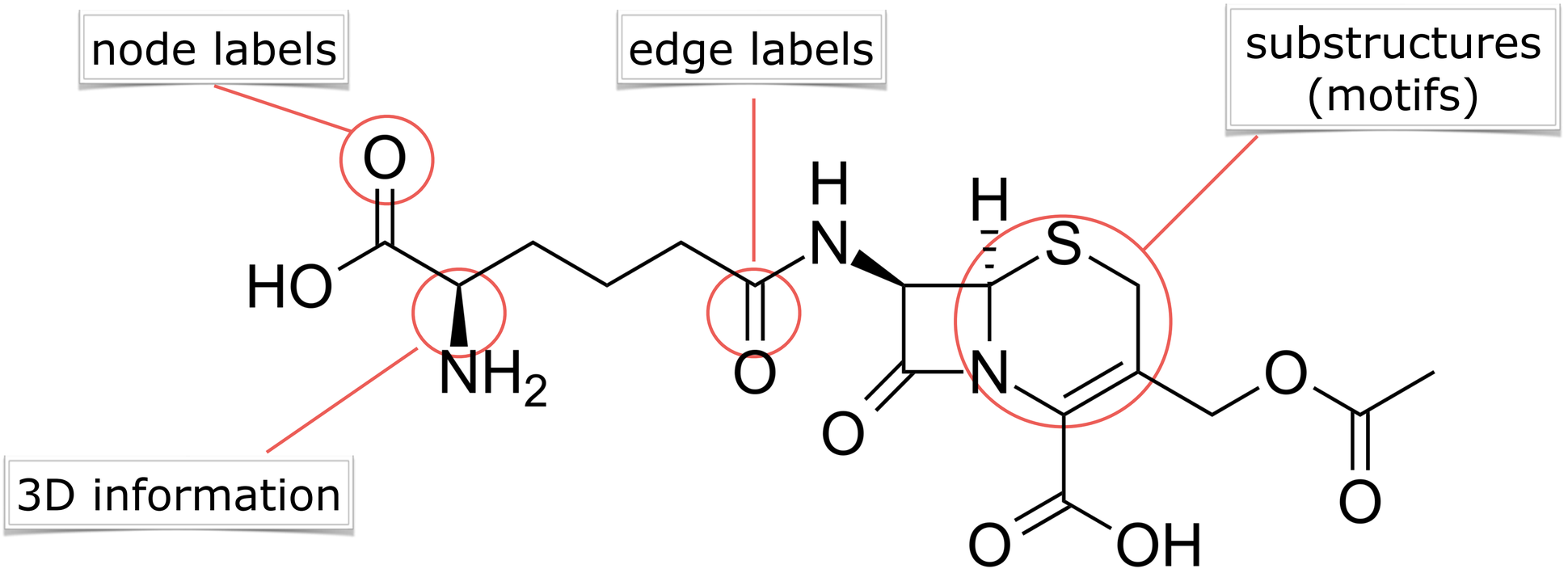
Why are molecules interesting for ML?

- E.g., antibiotic (cephalosporin)



Why are molecules interesting for ML?

- E.g., antibiotic (cephalosporin)

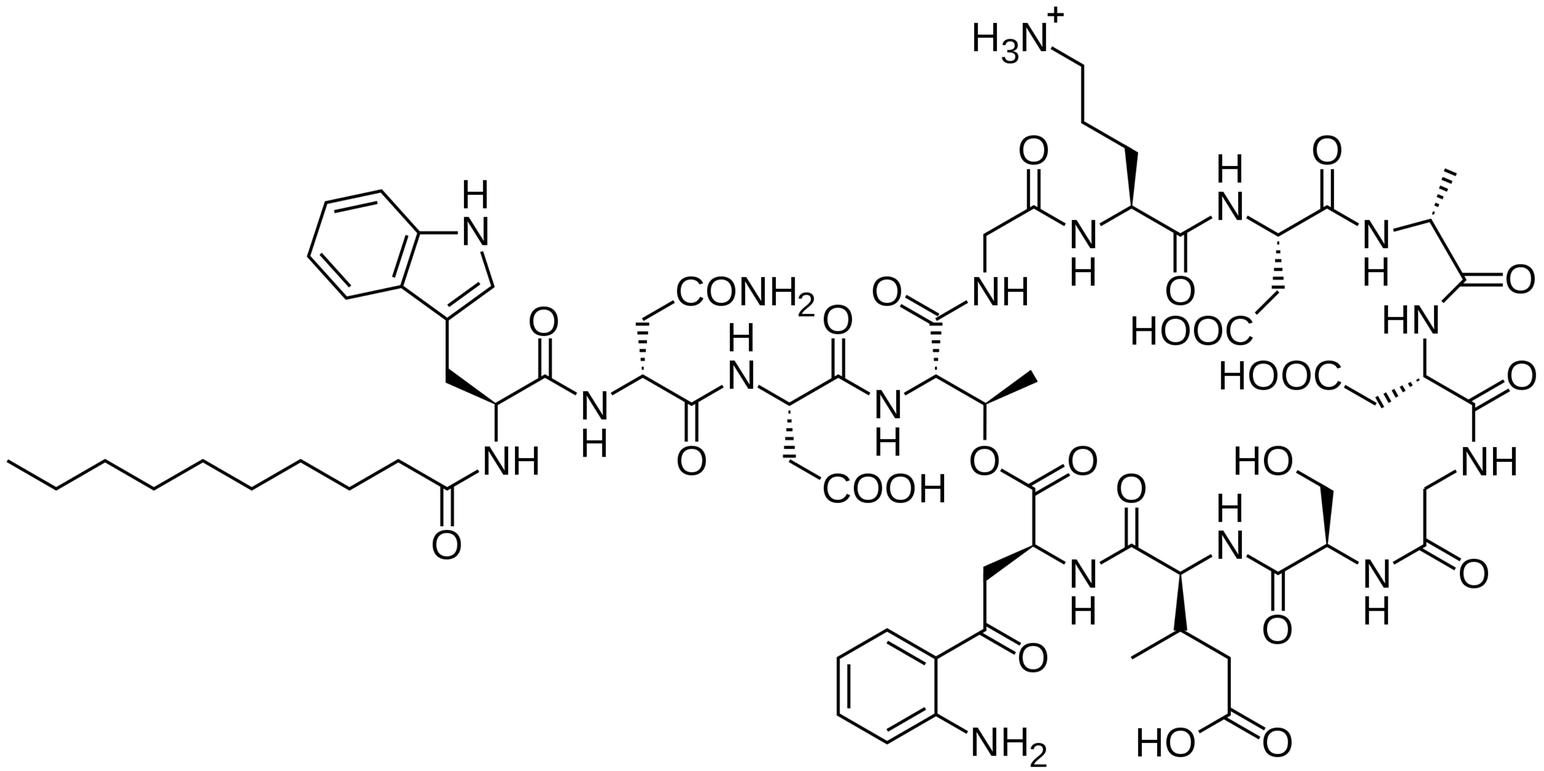


Together give rise to various chemical properties

(e.g., solubility, toxicity, …)

Why are molecules interesting for ML?

- Properties may depend on intricate structures;
- The key challenges are to automatically predict chemical properties and to generate molecules with desirable characteristics



(Daptomycin antibiotic)

Interesting ML Problems

- **Deeper into known chemistry**
  - extract chemical knowledge from journals, notebooks (NLP)

- **Deeper into drug design**
  - molecular property prediction (graph representation)
  - (multi-criteria) lead optimization (graph generation)

- **Deeper into reactions**
  - forward reaction prediction (structured prediction)
  - forward reaction optimization (combinatorial optimization)

- **Deeper into synthesis**
  - retrosynthesis planning (reinforcement learning)

Interesting ML Problems

- **Deeper into known chemistry**
  - extract chemical knowledge from journals, notebooks (NLP)

- **Deeper into drug design**
  - molecular property prediction (graph representation)
  - (multi-criteria) lead optimization (graph generation)

- **Deeper into reactions**
  - forward reaction prediction (structured prediction)
  - forward reaction optimization (combinatorial optimization)

- **Deeper into synthesis**
  - retrosynthesis planning (reinforcement learning)

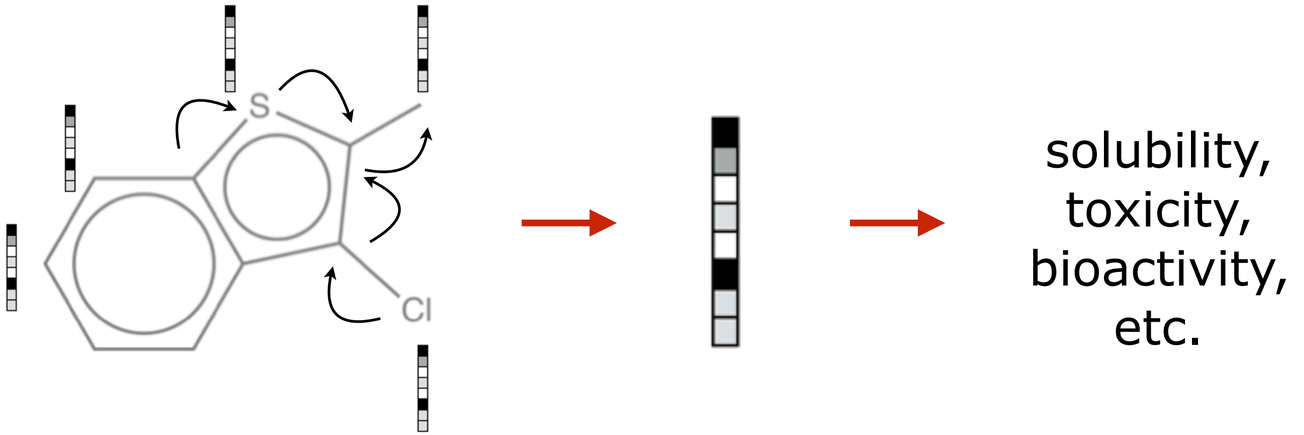
Automating Drug design

- **Key challenges:**

1. **representation and prediction:** learn to predict molecular properties
2. **generation and optimization:** realize target molecules with better properties programmatically
3. **understanding:** uncover principles (or diagnose errors) underlying complex predictions

GNNs for property prediction?

- Are GNN models operating on molecular graphs sufficiently expressive for predicting molecular properties (in the presence of “property cliffs”)?



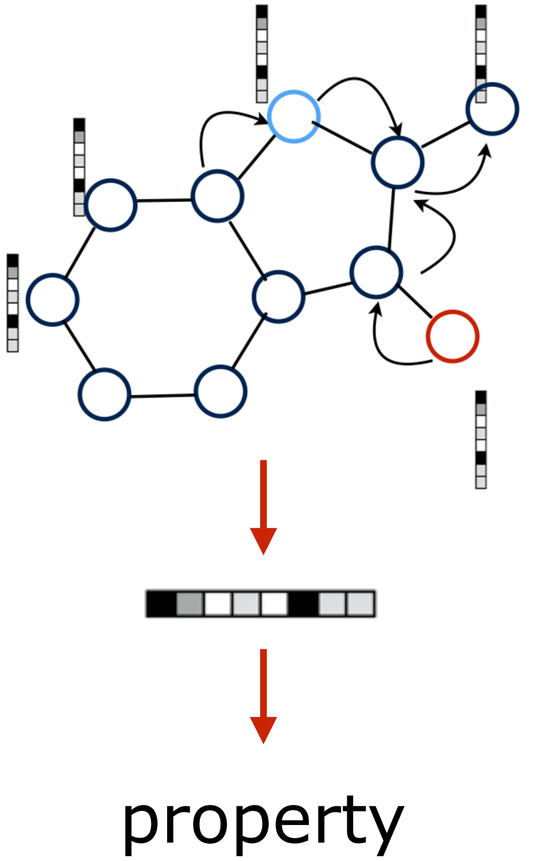
GNN embedding aggregation prediction

- A number of recent results pertaining to the power of GNNs (e.g., Xu et al. 2018, Sato et al. 2019, Maron et al., 2019, …);

Are basic GNNs sufficiently expressive?

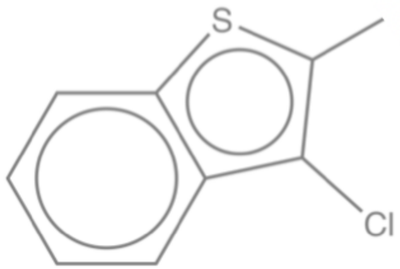
- **Theorem [Garg et al., 2019]:** GNNs with permutation invariant readout functions cannot "decide"
  - girth (length of the shortest cycle)
  - circumference (length of the longest cycle)
  - diameter, radius
  - presence of conjoint cycle
  - total number of cycles
  - presence of c-clique
  - etc. (?)

- (most results also apply to MPNNs)



Beyond simple GNNs: sub-structures

- Learning to view molecules at multiple levels [Jin et al., 2019]

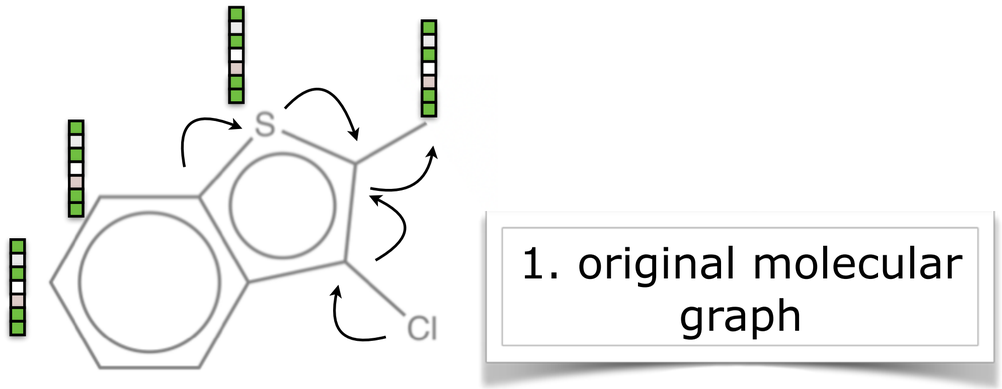


1. original molecular graph

Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Beyond simple GNNs: sub-structures

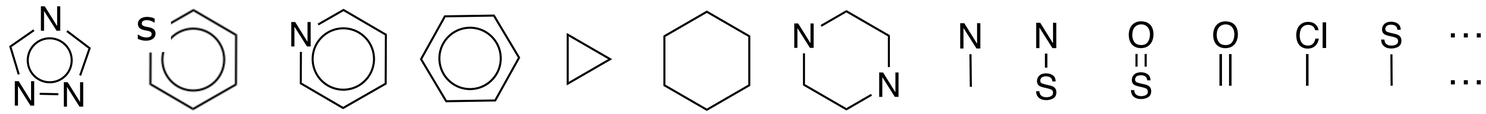
- Learning to view molecules at multiple levels



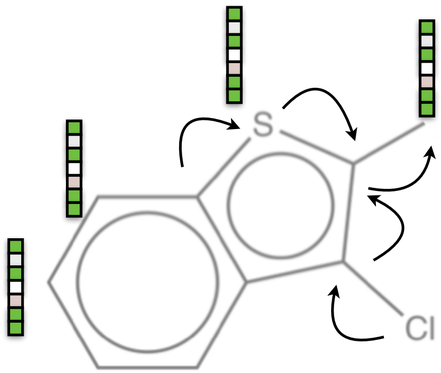
Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Beyond simple GNNs: sub-structures

- Learning to view molecules at multiple levels



a dictionary of substructures

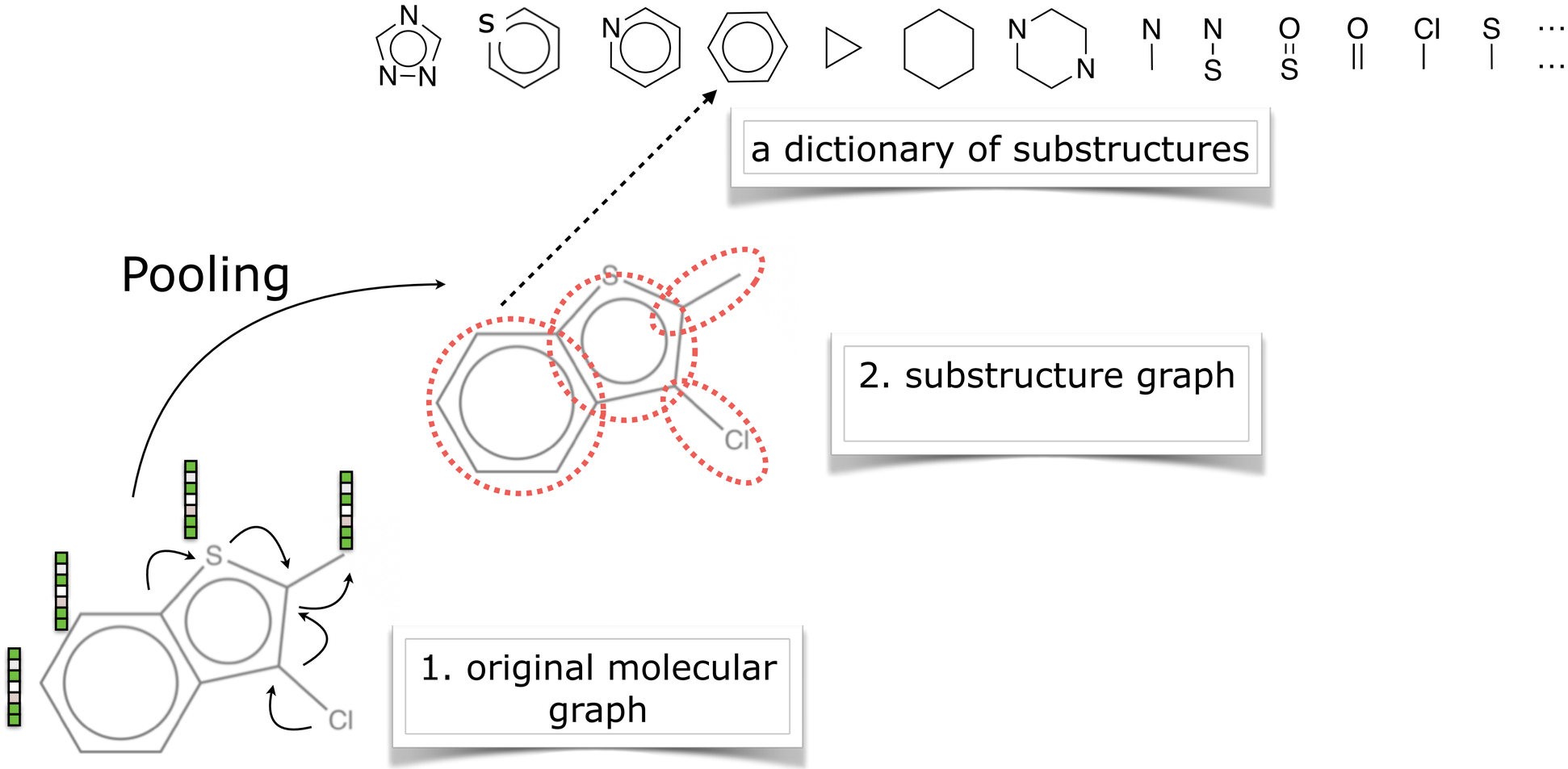


1. original molecular graph

Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Beyond simple GNNs: sub-structures

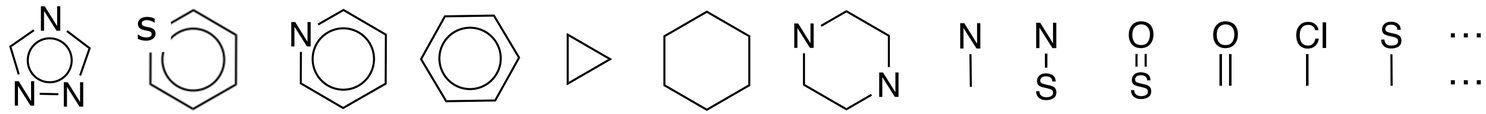
- Learning to view molecules at multiple levels



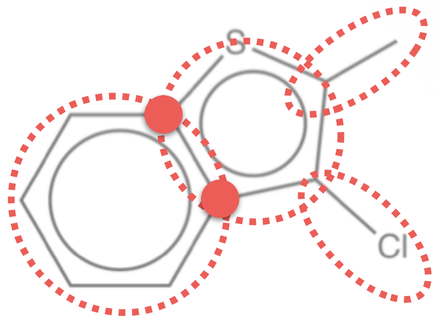
Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Beyond simple GNNs: sub-structures

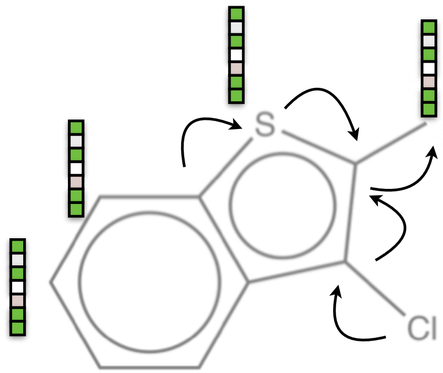
- Learning to view molecules at multiple levels



a dictionary of substructures



2. substructure graph with attachments

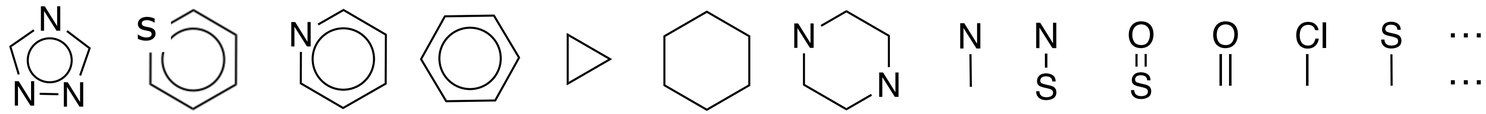


1. original molecular graph

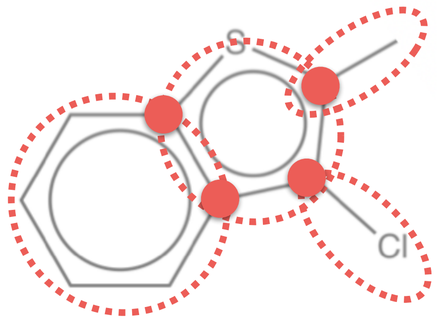
Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Beyond simple GNNs: sub-structures

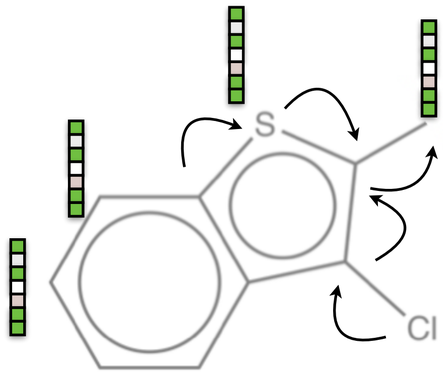
- Learning to view molecules at multiple levels



a dictionary of substructures



2. substructure graph with attachments

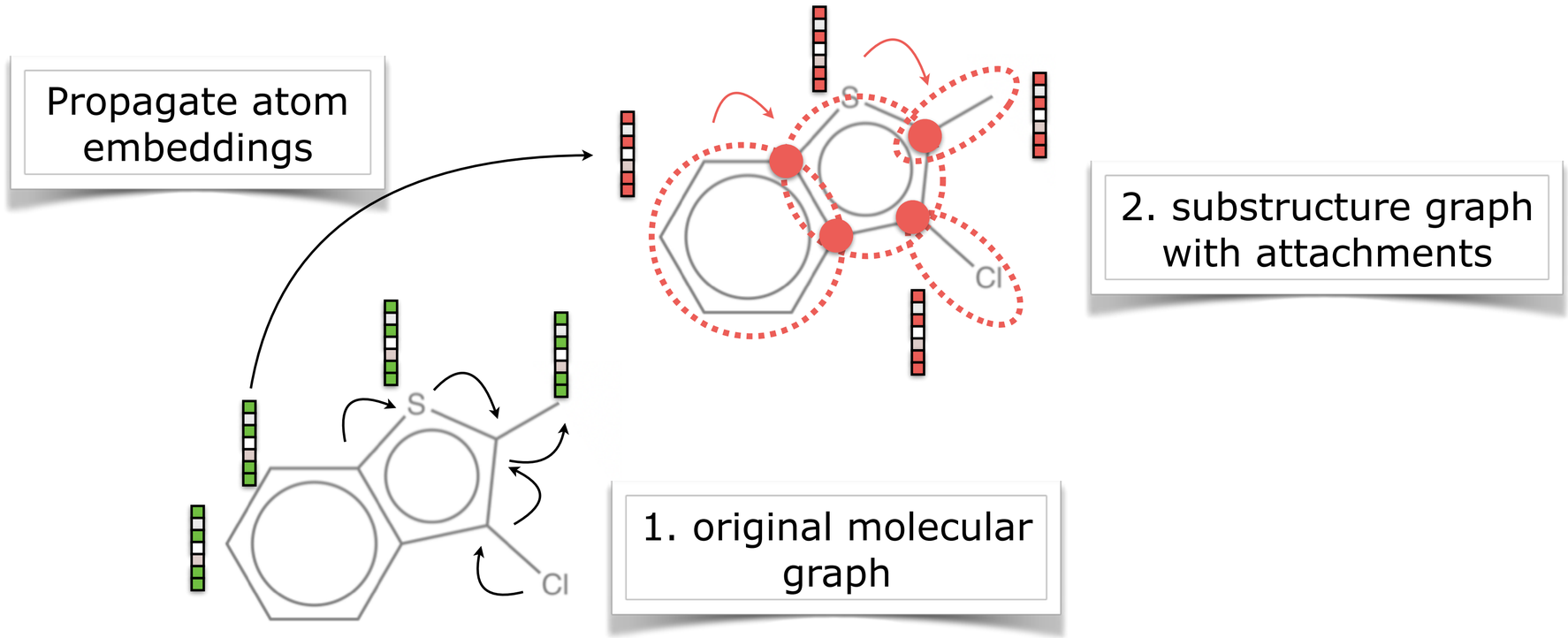


1. original molecular graph

Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Beyond simple GNNs: sub-structures

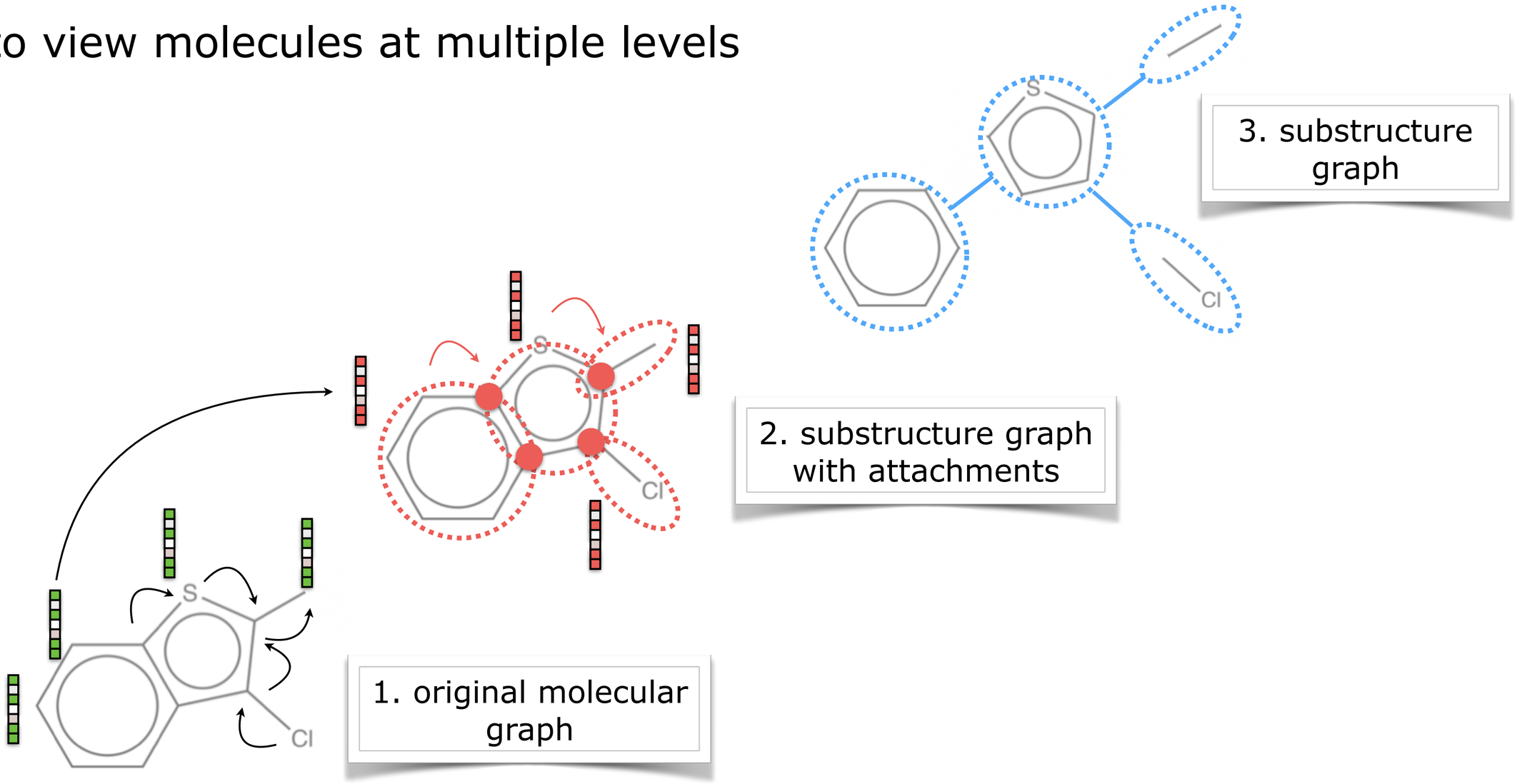
- Learning to view molecules at multiple levels



Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Beyond simple GNNs: sub-structures

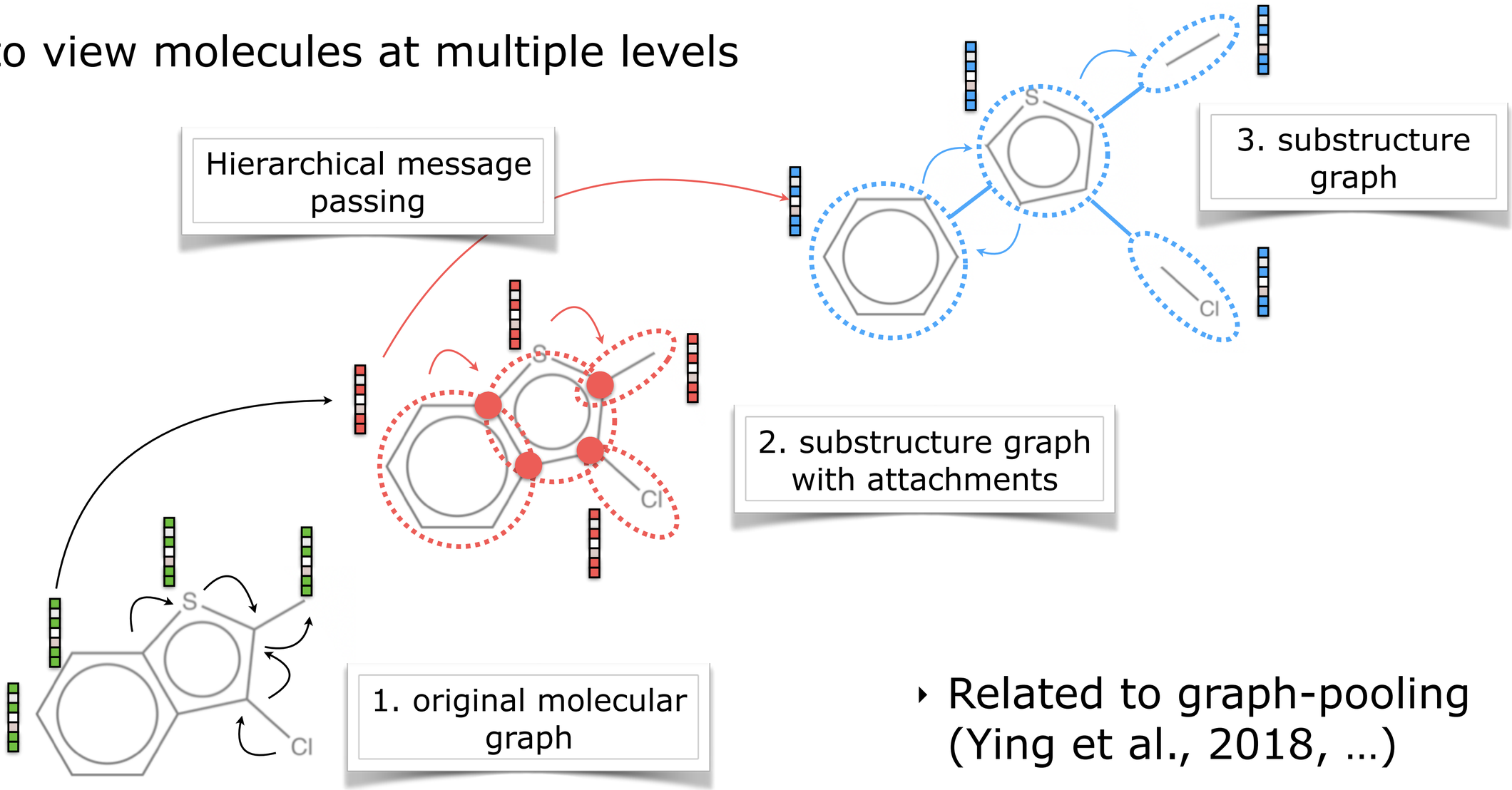
- Learning to view molecules at multiple levels



Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Multi-resolution representations

- Learning to view molecules at multiple levels

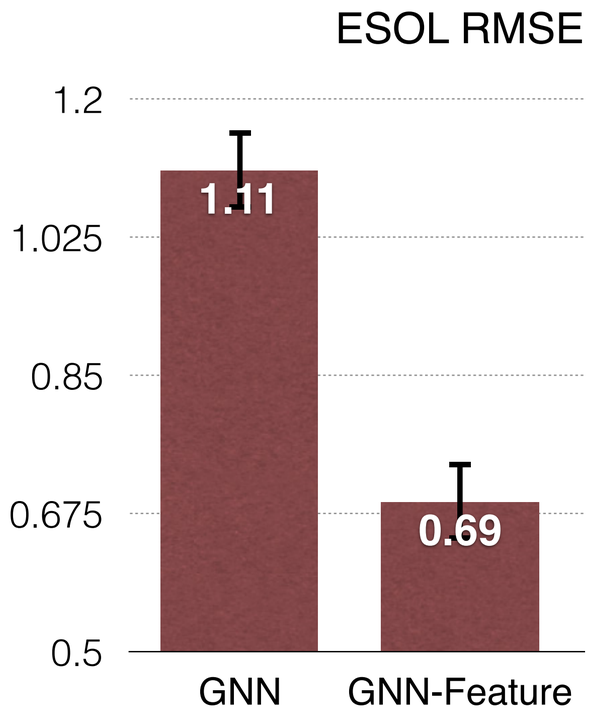


- Related to graph-pooling (Ying et al., 2018, …)

Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

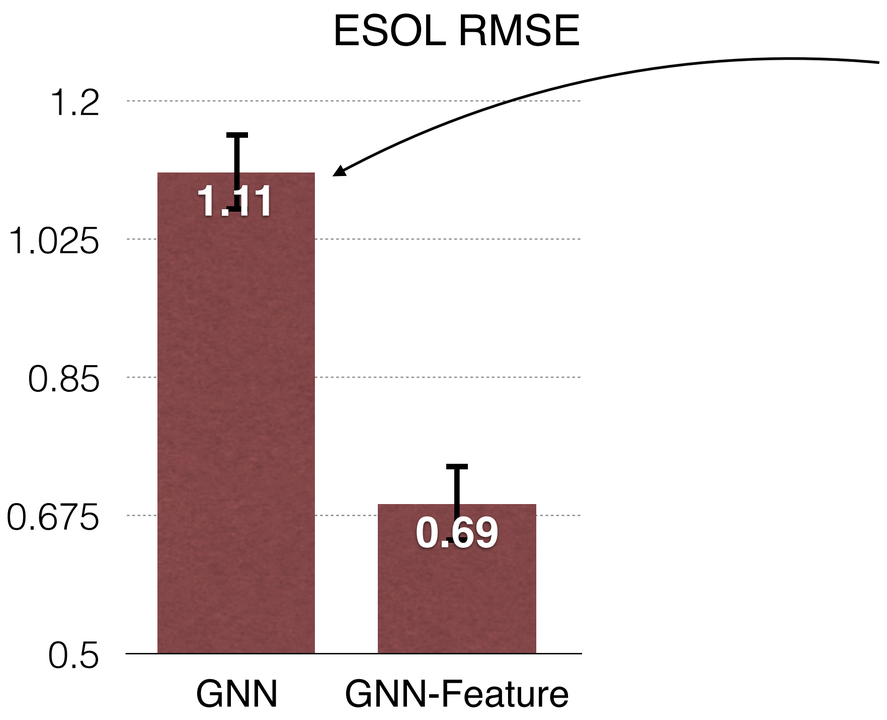
Experiments on solubility

- ESOL dataset (averaged over 5 folds)



Experiments on solubility

- ESOL dataset (averaged over 5 folds)

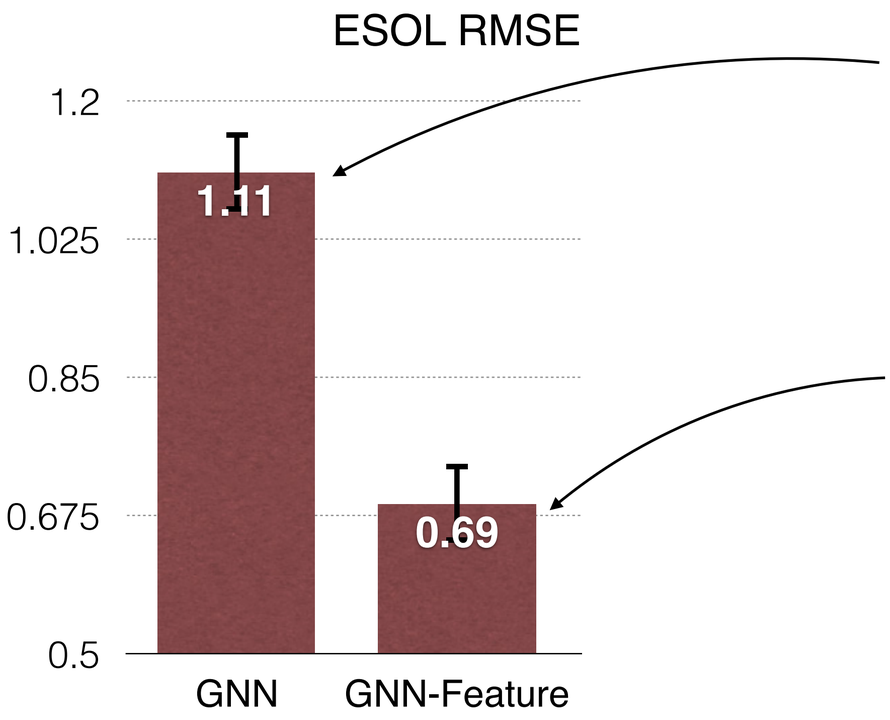


Raw GNN

- atom feature: only atom type label

Experiments on solubility

- ESOL dataset (averaged over 5 folds)



Raw GNN

- atom feature: only atom type label

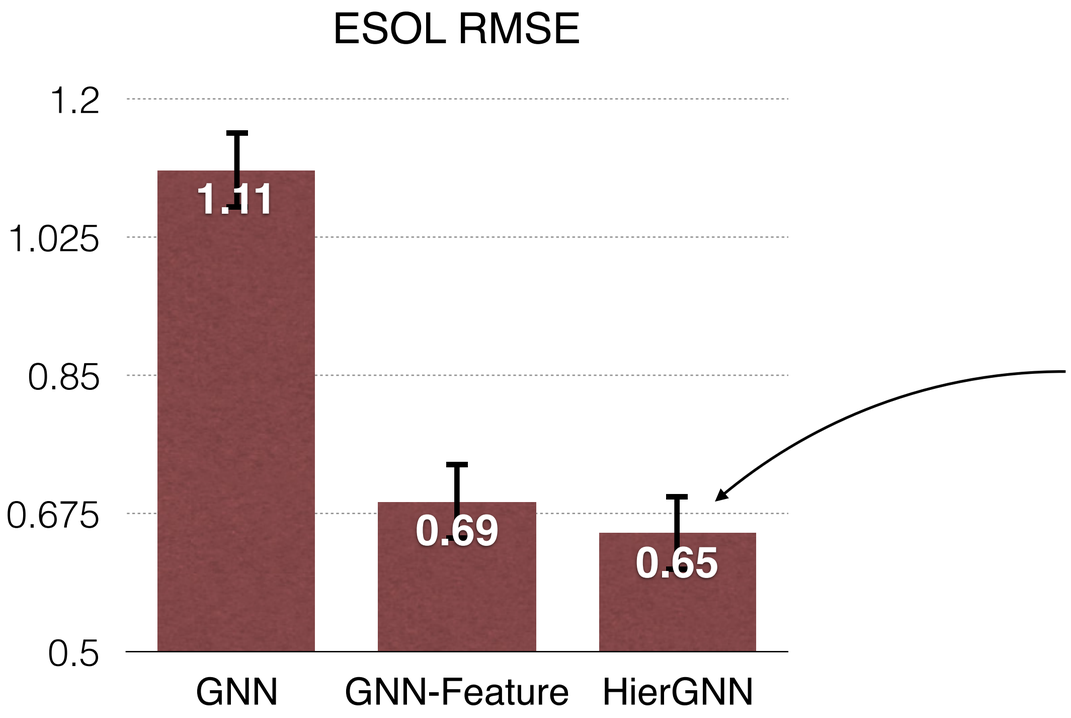
GNN with **features**

- atom type label
- degree
- valence
- **whether an atom is in a cycle**
- **whether an atom is in an aromatic ring**
- ……

Cycle information

Experiments on solubility

- ESOL dataset (averaged over 5 folds)

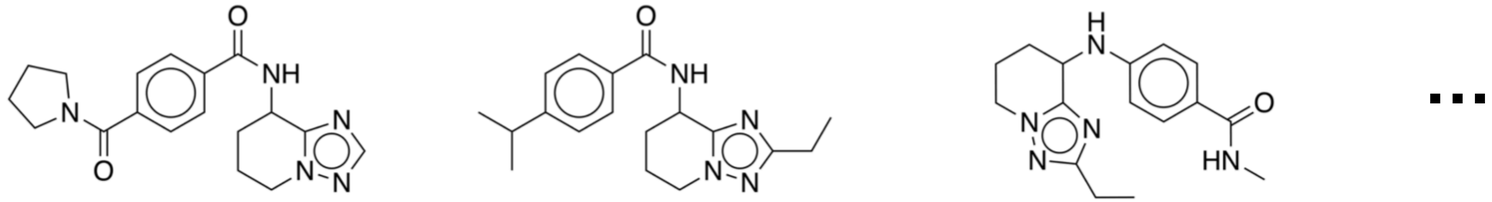


Hierarchical GNN

- Atom features: still just atom type
- But has extra substructure information built into the architecture

New Antibiotic Discovery

- If we can accurately predict molecular properties, we can screen (select and repurpose) molecules from a large candidate set



...

- Antibiotic Discovery [Stokes et al., 2019]
  - Trained a model to predict the inhibition against E. Coli (some bacteria…)
  - Data: ~2000 measured compounds from Broad Institute at MIT
  - Screened in total ~100 million compounds
  - Biologists tested 15 molecules (top prediction, structurally diverse) in the lab
  - 7 of them are validated to be inhibitive in-vitro
  - 1 of them demonstrate strong inhibition against other bacteria (e.g., A. baumannii)
  - All of them are <u>new</u> antibiotics distinct from existing ones!

Learning to Discover Novel Antibiotics from Vast Chemical Spaces (2019), J. Stokes, K. Yang, K. Swanson, **W. Jin**, R. Barzilay, T. Jaakkola et al.

Automating Drug design

- **Key challenges:**

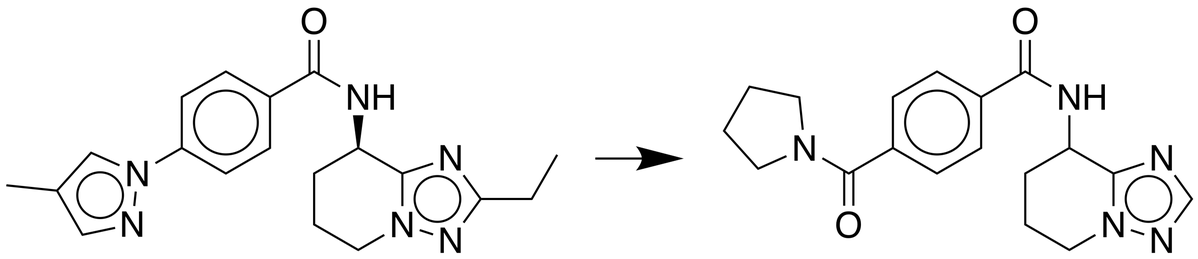
1. **representation and prediction:** learn to predict molecular properties
2. **generation and optimization:** realize target molecules with better properties programmatically
3. **understanding:** uncover principles (or diagnose errors) underlying complex predictions

De novo molecule optimization

- **Goal:** We aim to programmatically turn precursor molecules into molecules that satisfy given design specifications

De novo molecule optimization

- **Goal:** We aim to programmatically turn precursor molecules into molecules that satisfy given design specifications



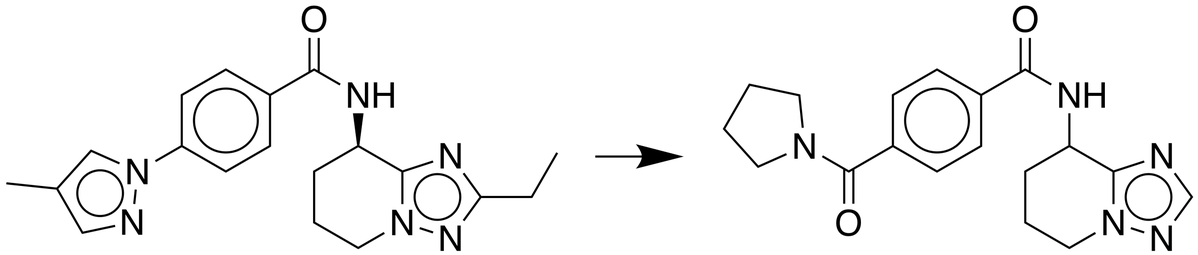
Source Molecule (QED=0.784)

QED=0.924

- Similar but …
- Better drug-likeness

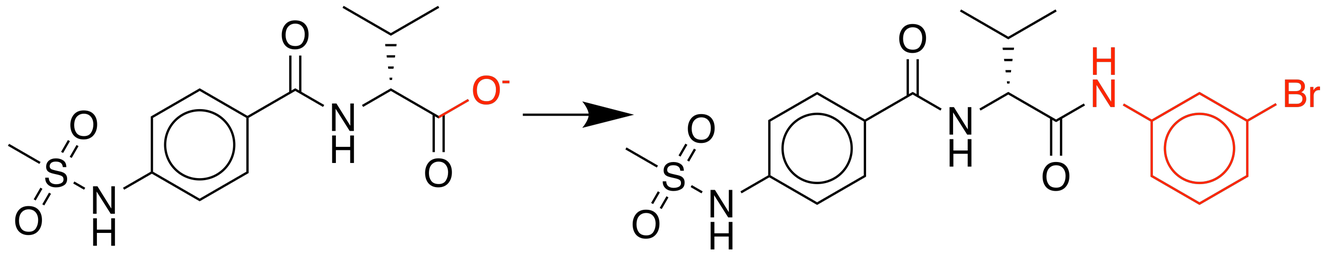
De novo molecule optimization

- **Goal:** We aim to programmatically turn precursor molecules into molecules that satisfy given design specifications



Source Molecule (QED=0.784) QED=0.924

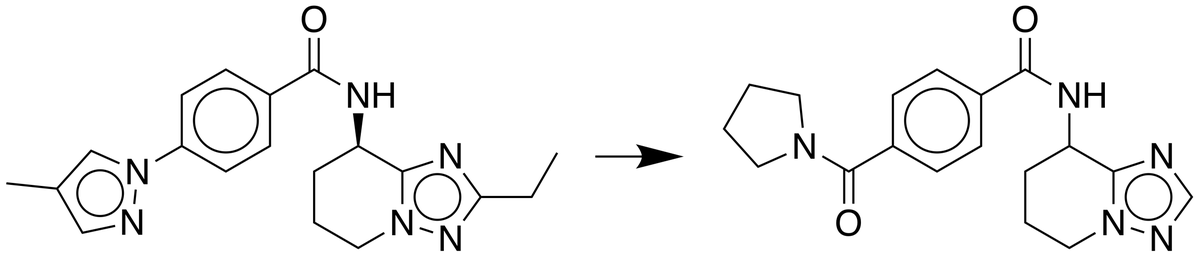
- Similar but …
- Better drug-likeness



- Similar but …
- Better solubility

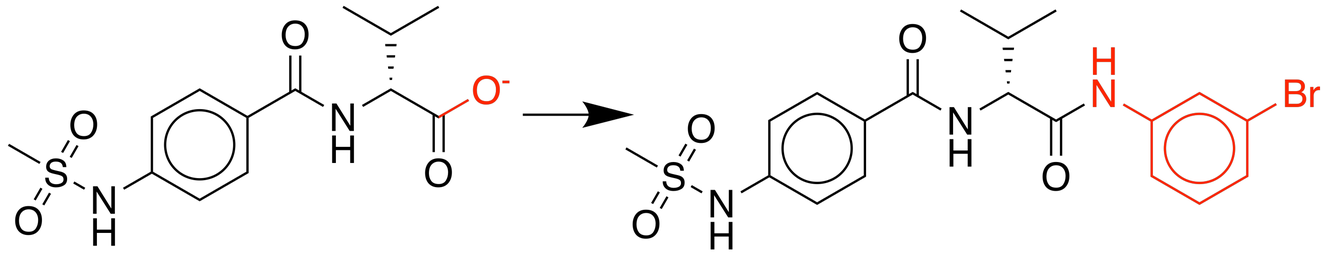
De novo molecule optimization

- **Goal:** We aim to programmatically turn precursor molecules into molecules that satisfy given design specifications



Source Molecule (QED=0.784) QED=0.924

- Similar but …
- Better drug-likeness

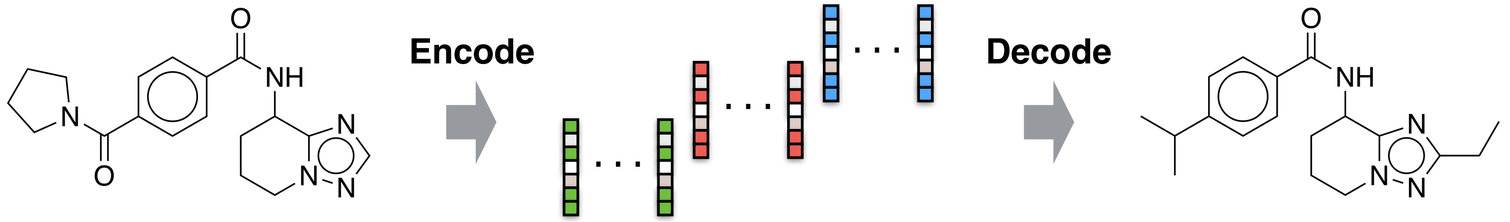


- Similar but …
- Better solubility

- <u>Need to learn a molecule-to-molecule mapping (i.e., graph-to-graph)</u>

Molecule optimization as Graph Translation

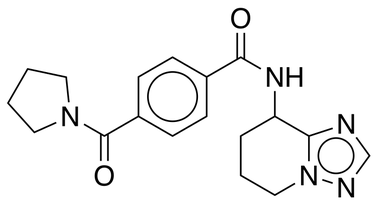
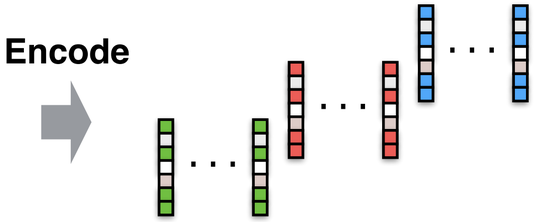
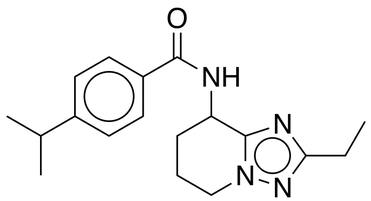
- **Goal:** We aim to programmatically turn precursor molecules into molecules that satisfy given design specifications

Source X 

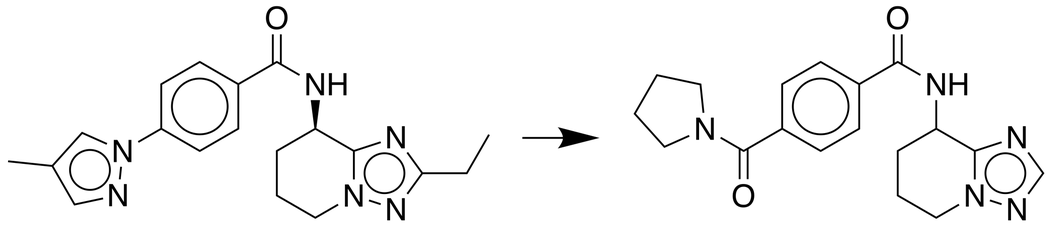
Y Target

Molecule optimization as Graph Translation

- **Goal:** We aim to programmatically turn precursor molecules into molecules that satisfy given design specifications

Source X    Y Target

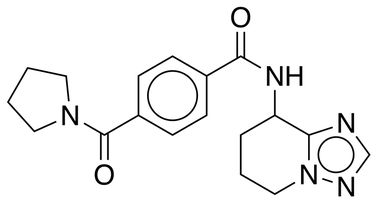
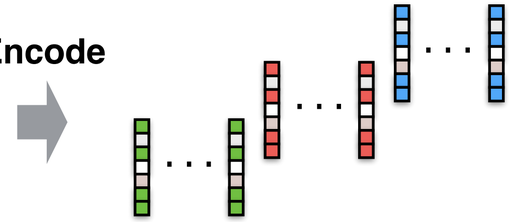
- The training set consists of (source, target) molecular pairs, e.g.,

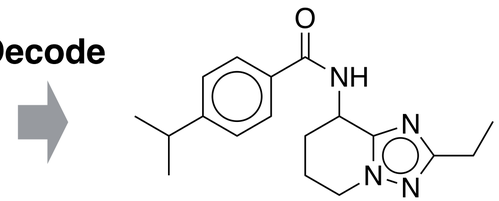
Source  Target

... ...

Molecule optimization as Graph Translation

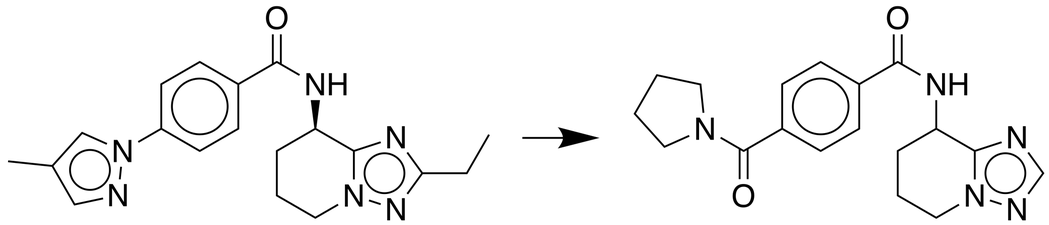
- **Goal:** We aim to programmatically turn precursor molecules into molecules that satisfy given design specifications

Source X  



Y Target

- The training set consists of (source, target) molecular pairs, e.g.,

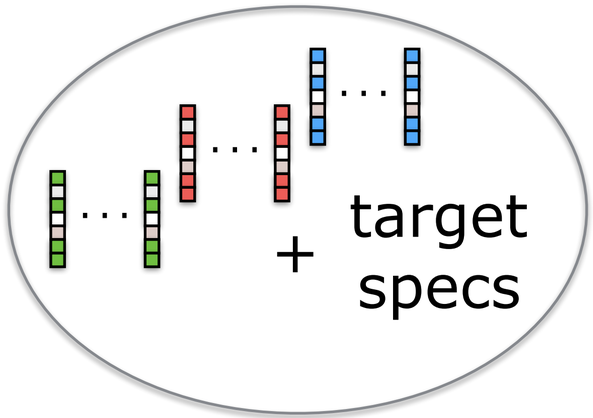
Source  Target

... ...

- **Key challenges:** graph generation, diversity, multi-criteria optimization

Graph-to-Graph Translation (Decoder)

- Modifying a pre-cursor to meet target specifications

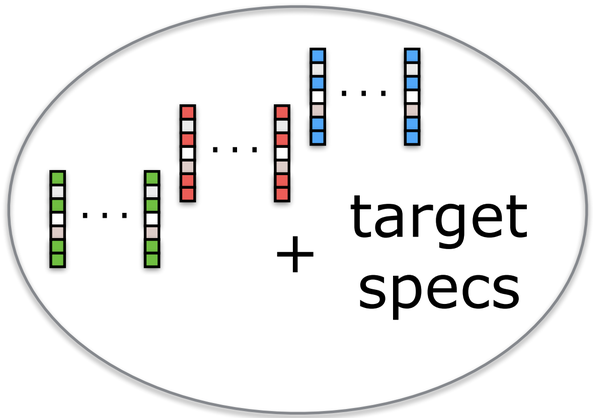


Source vectors ← Hierarchical GNN Encoder (more expressive power)

Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Graph-to-Graph Translation (Decoder)

- Modifying a pre-cursor to meet target specifications



Source vectors

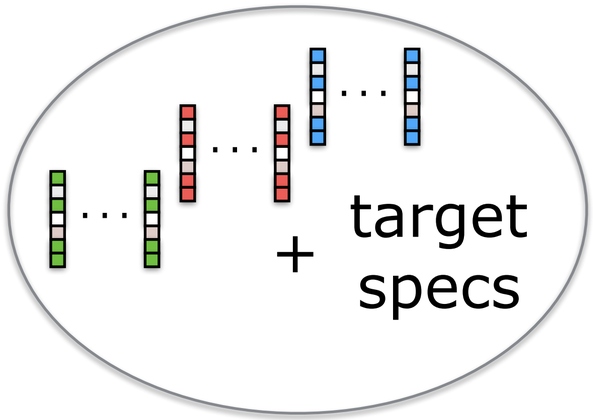
Hierarchical GNN Encoder
(more expressive power)

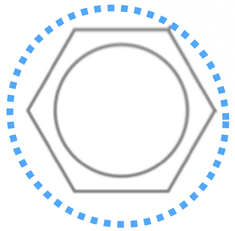
Hierarchical Graph Decoder
(reverse of encoding process)

Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

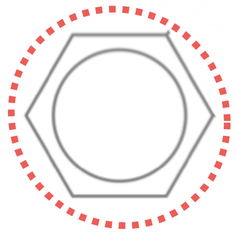
Graph-to-Graph Translation (Decoder)

- Modifying a pre-cursor to meet target specifications





Substructure graph



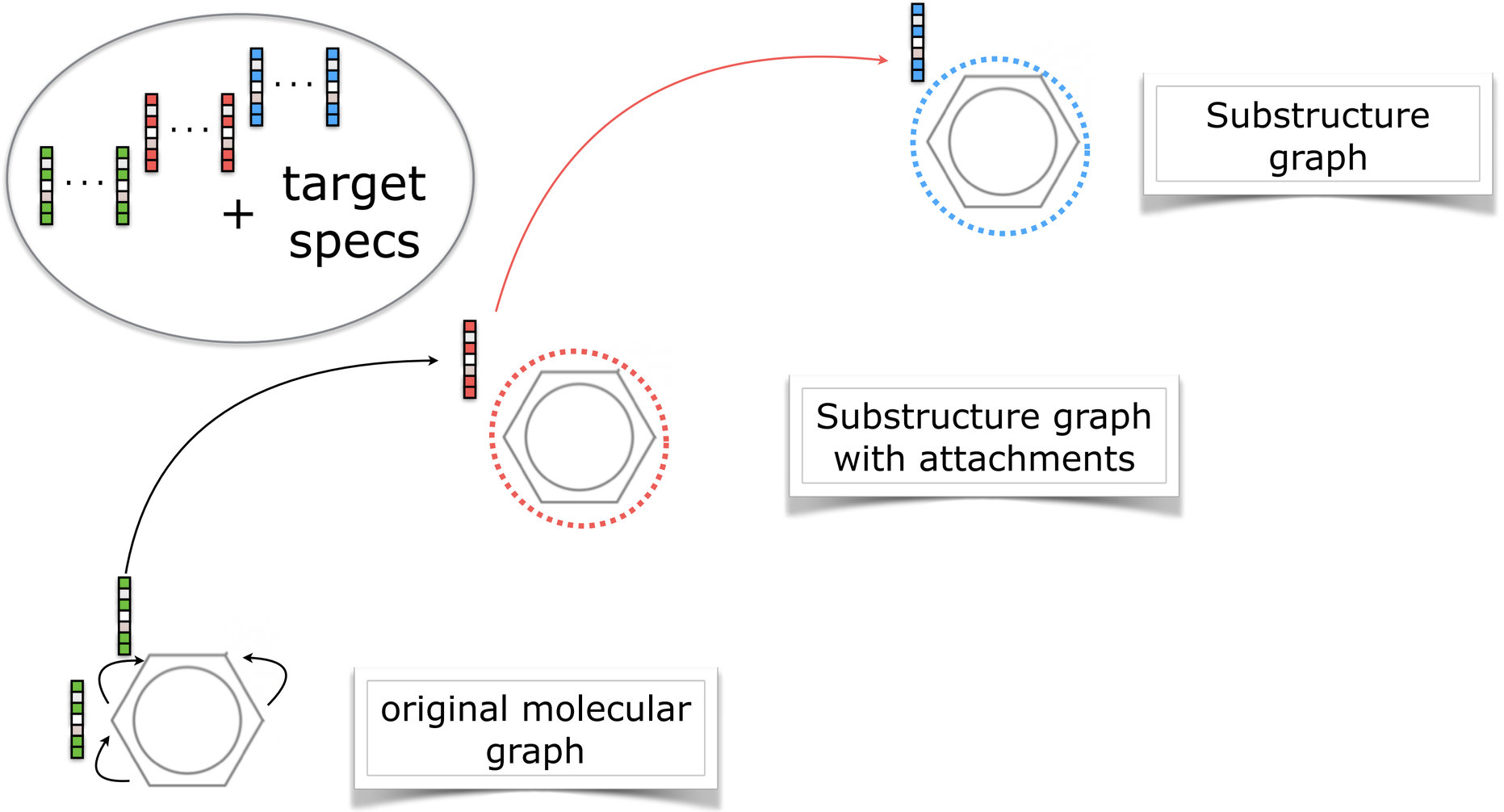
Substructure graph with attachments

Original molecular graph

Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Graph-to-Graph Translation (Decoder)

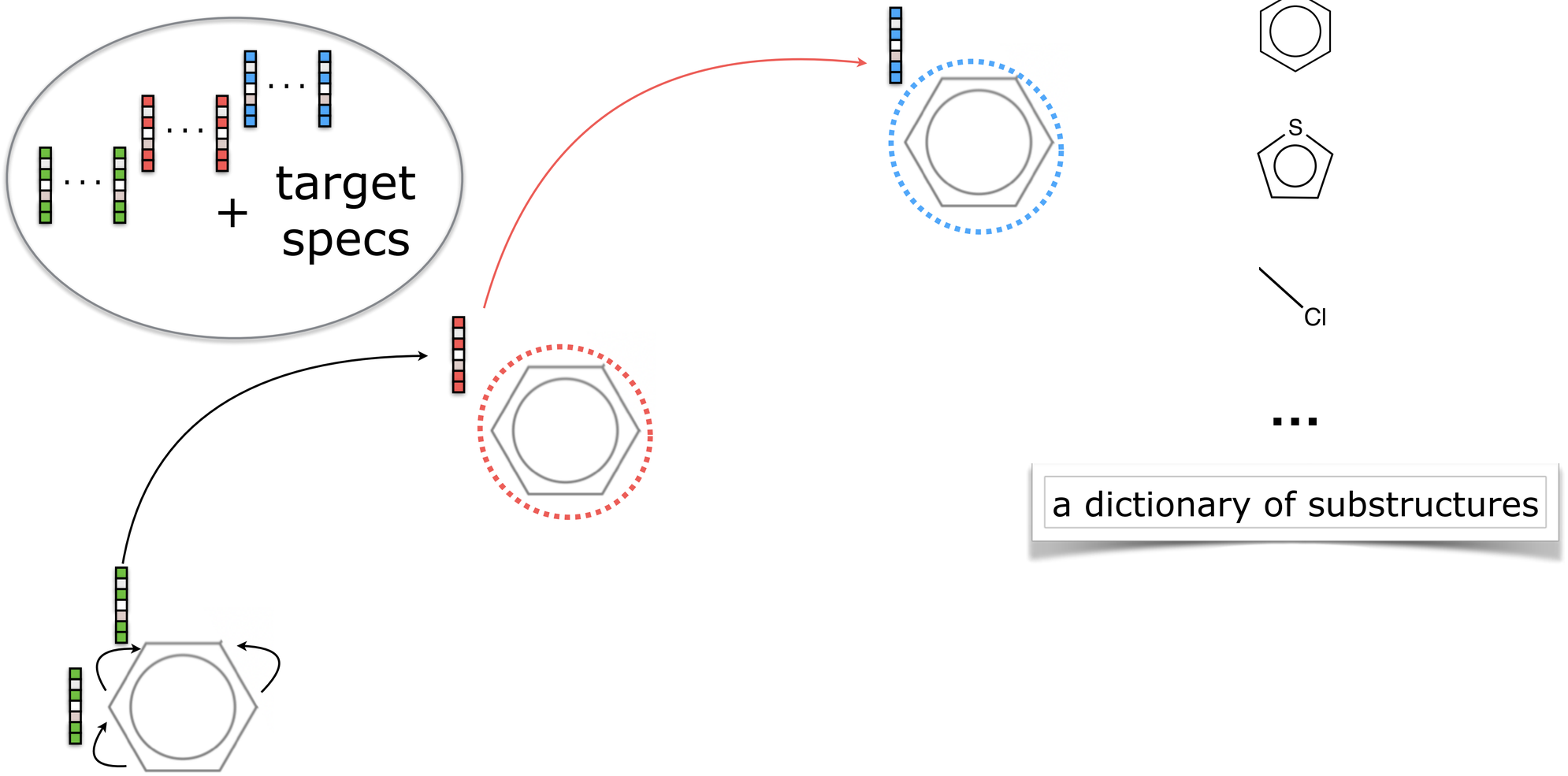
- Modifying a pre-cursor to meet target specifications



Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Graph-to-Graph Translation (Decoder)

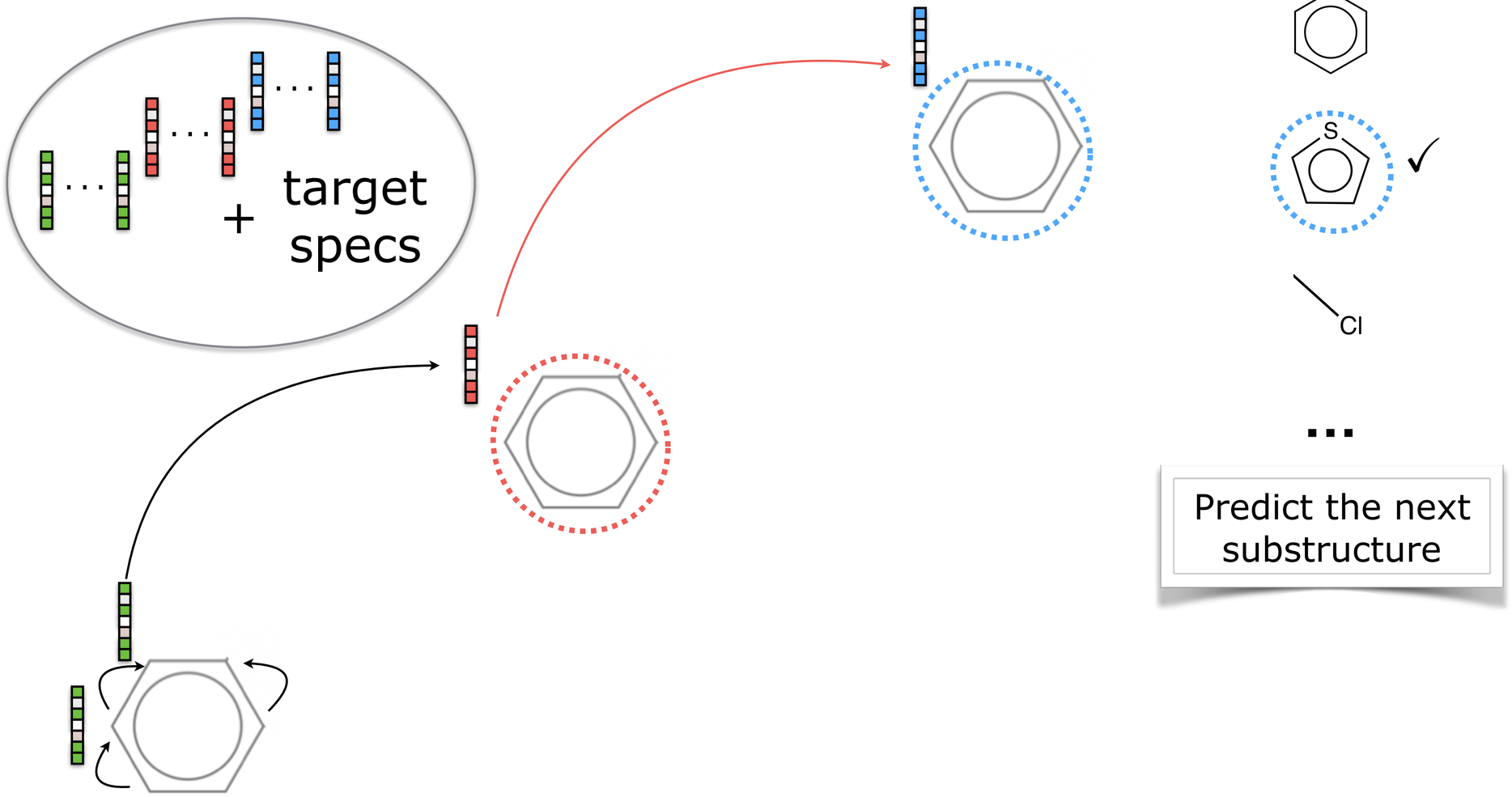
- Modifying a pre-cursor to meet target specifications



Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Graph-to-Graph Translation (Decoder)

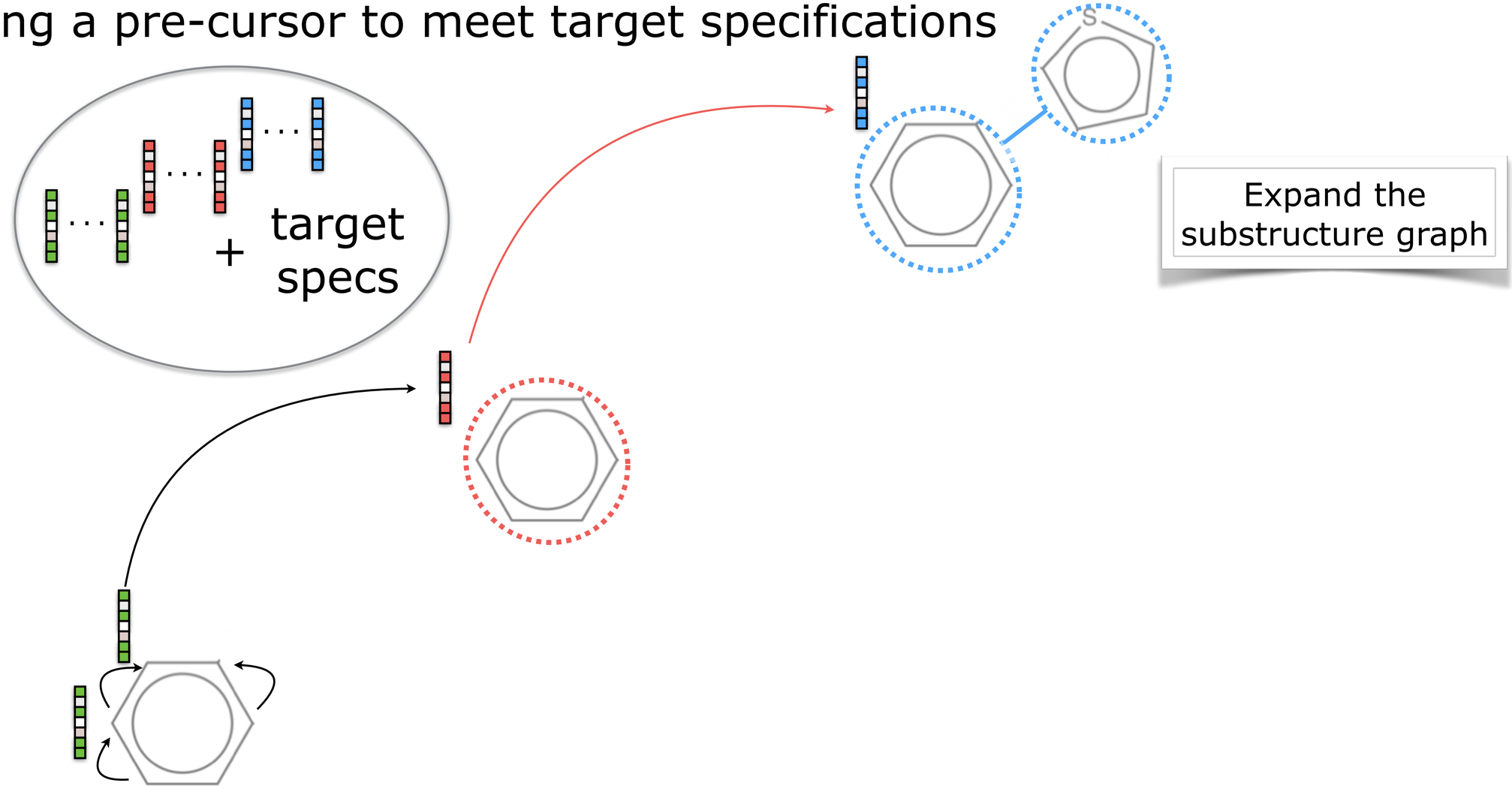
- Modifying a pre-cursor to meet target specifications



Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Graph-to-Graph Translation (Decoder)

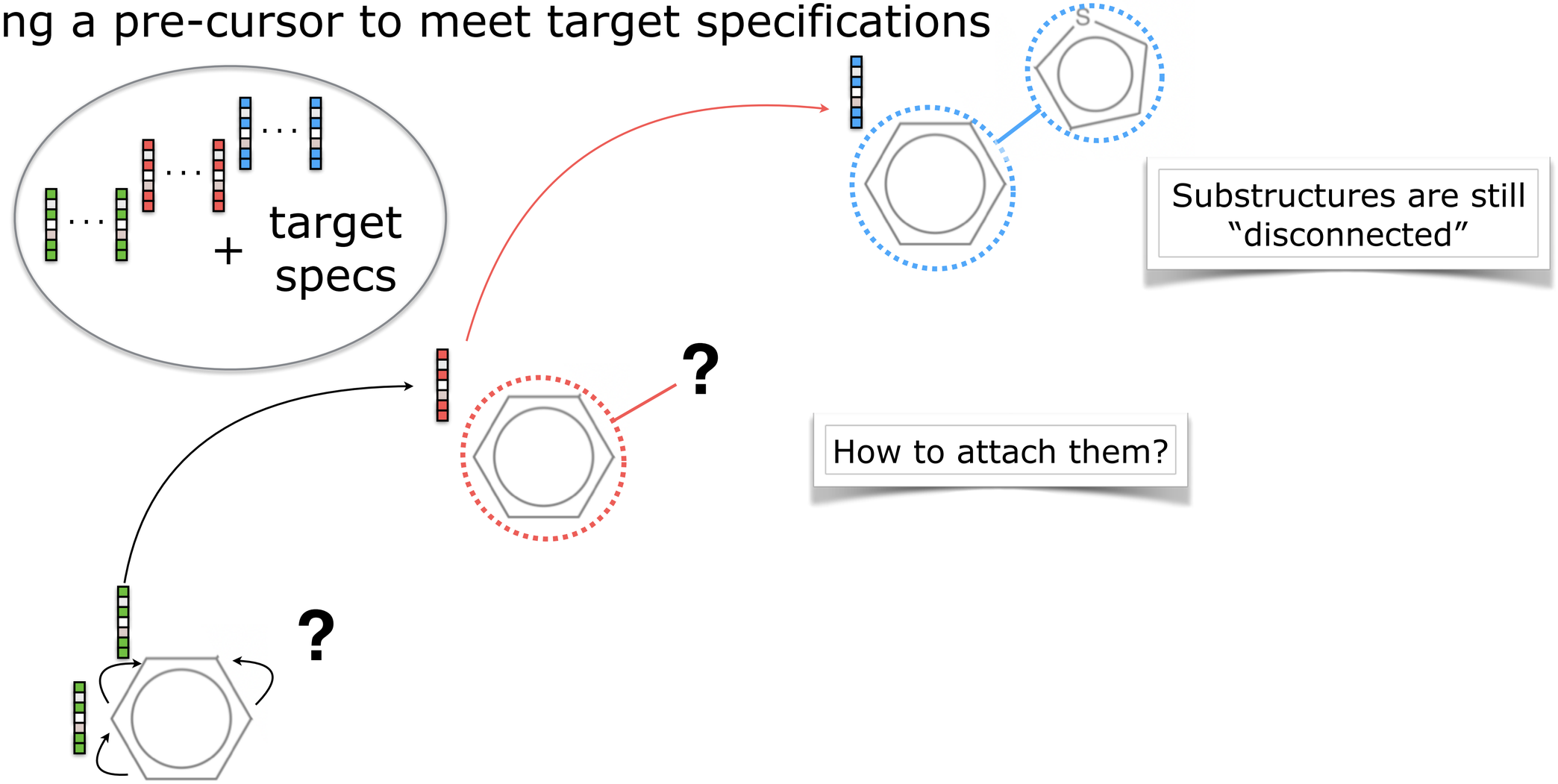
- Modifying a pre-cursor to meet target specifications



Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Graph-to-Graph Translation (Decoder)

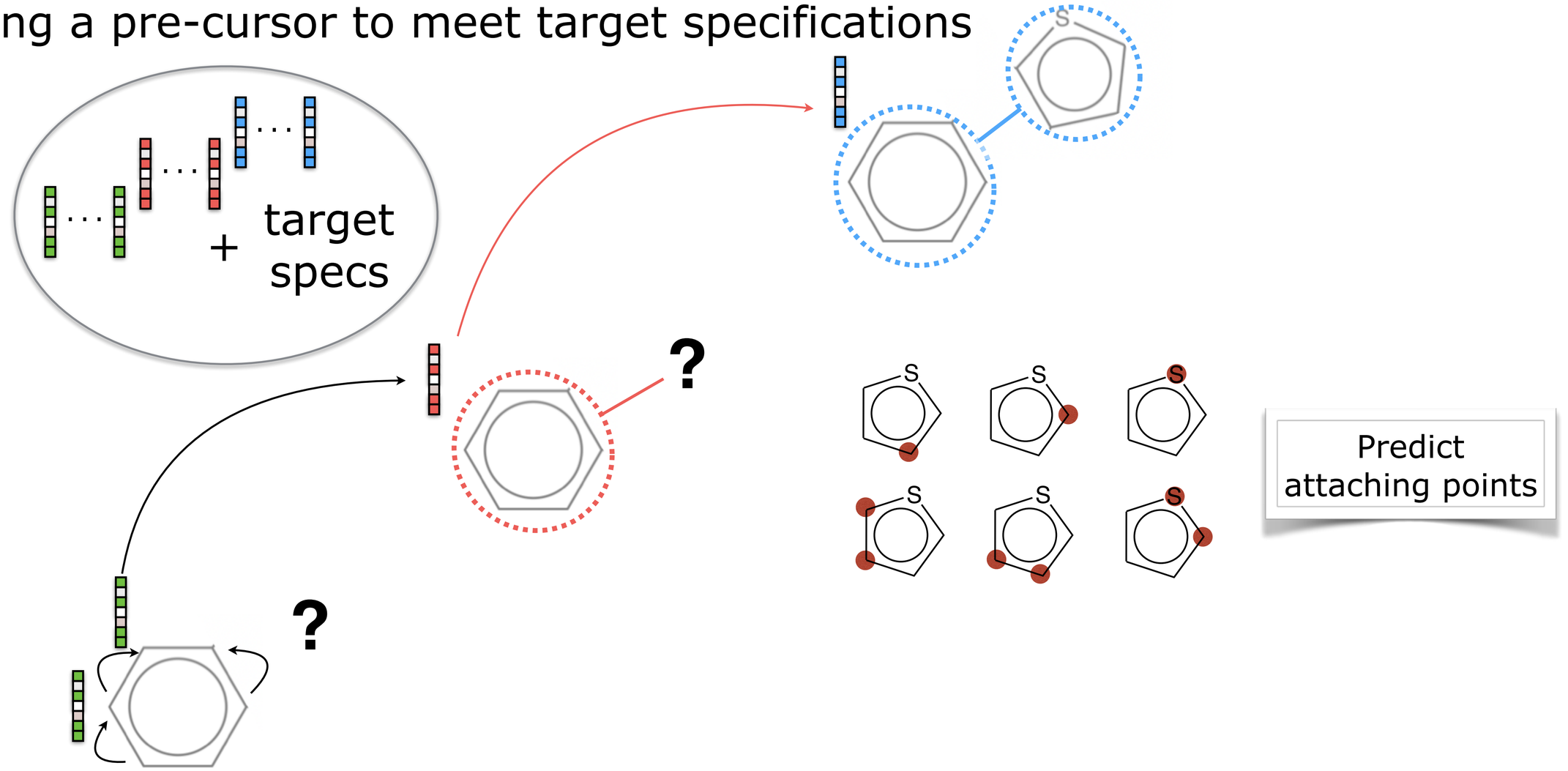
- Modifying a pre-cursor to meet target specifications



Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Graph-to-Graph Translation (Decoder)

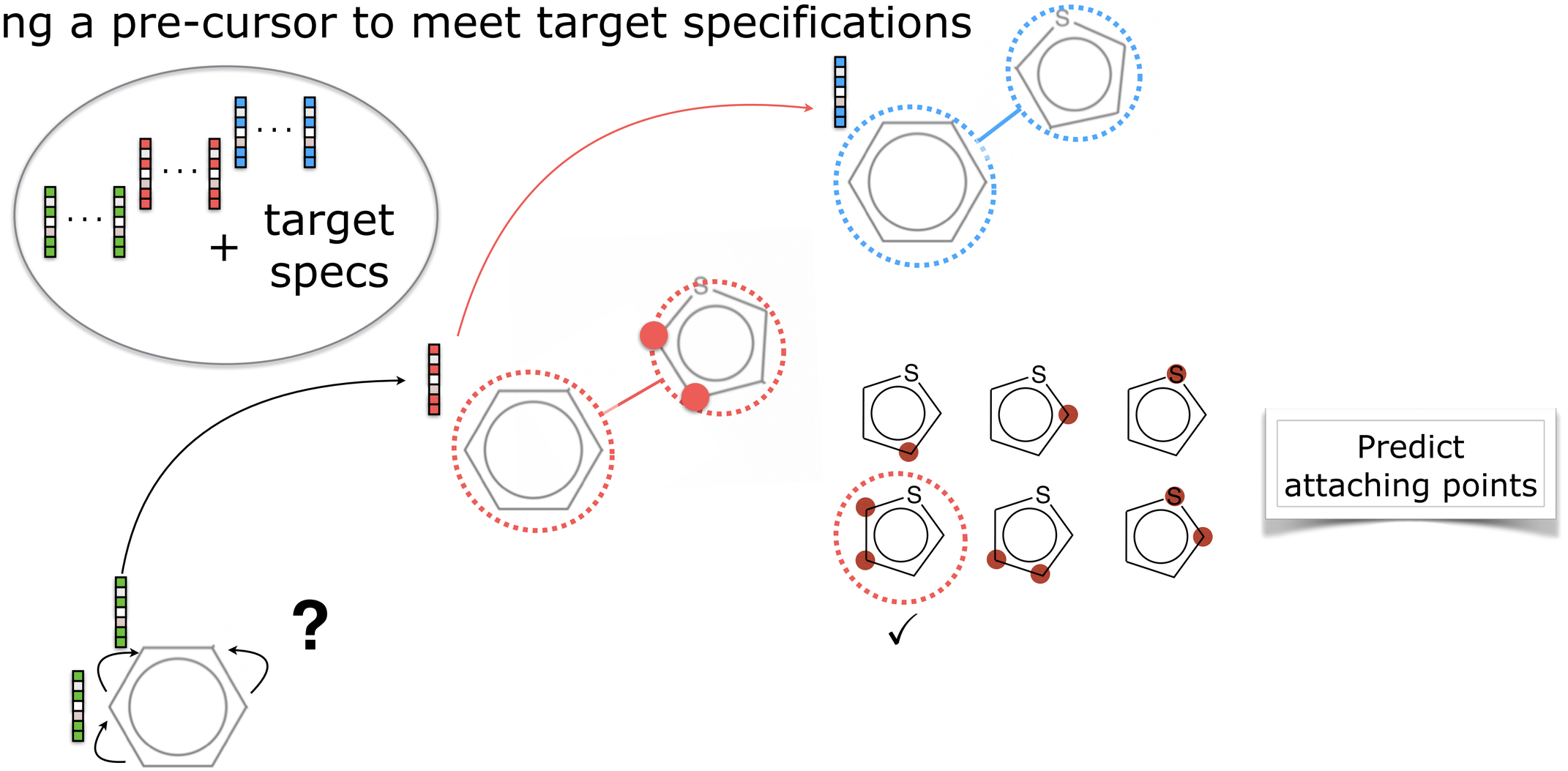
- Modifying a pre-cursor to meet target specifications



Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Graph-to-Graph Translation (Decoder)

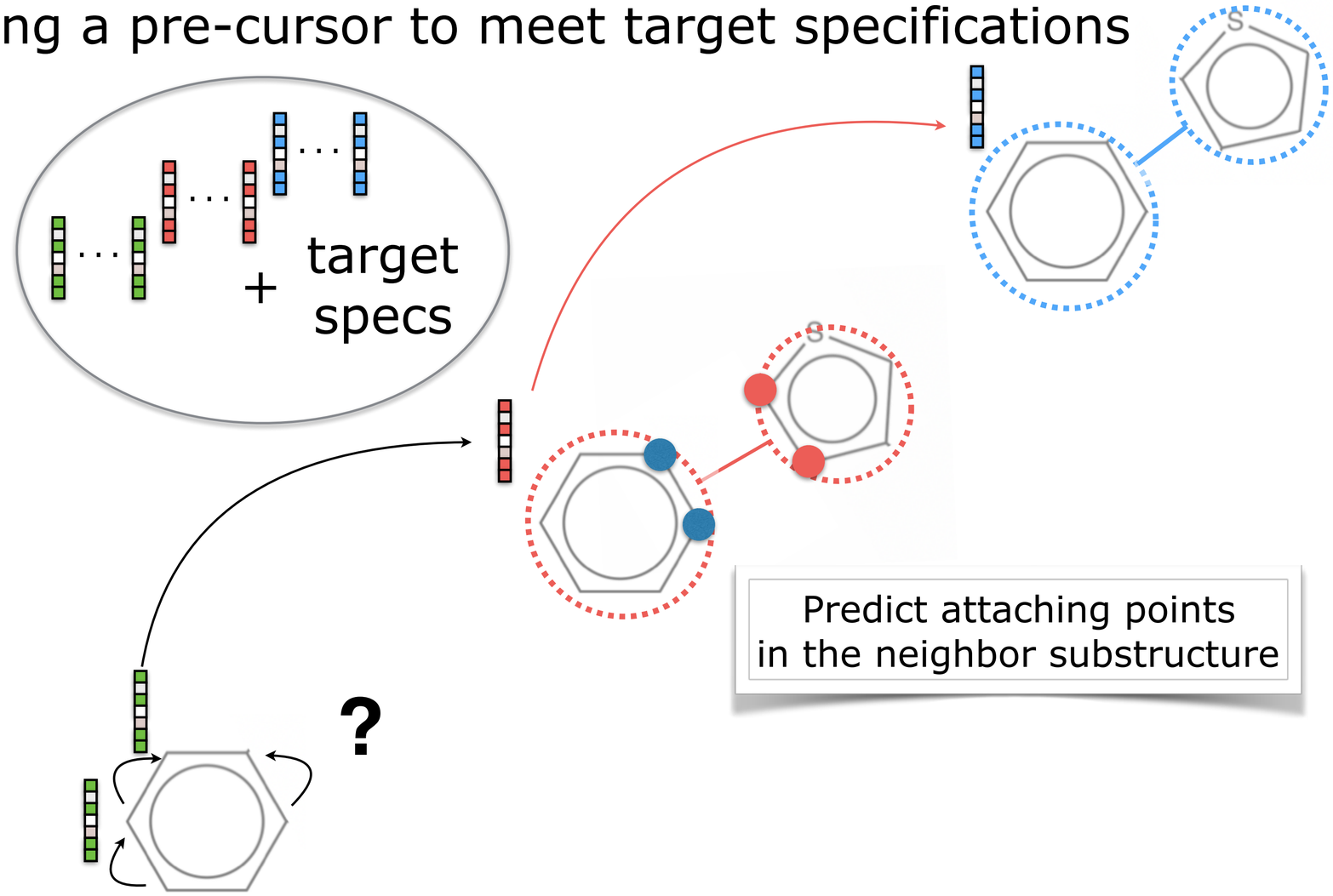
- Modifying a pre-cursor to meet target specifications



Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Graph-to-Graph Translation (Decoder)

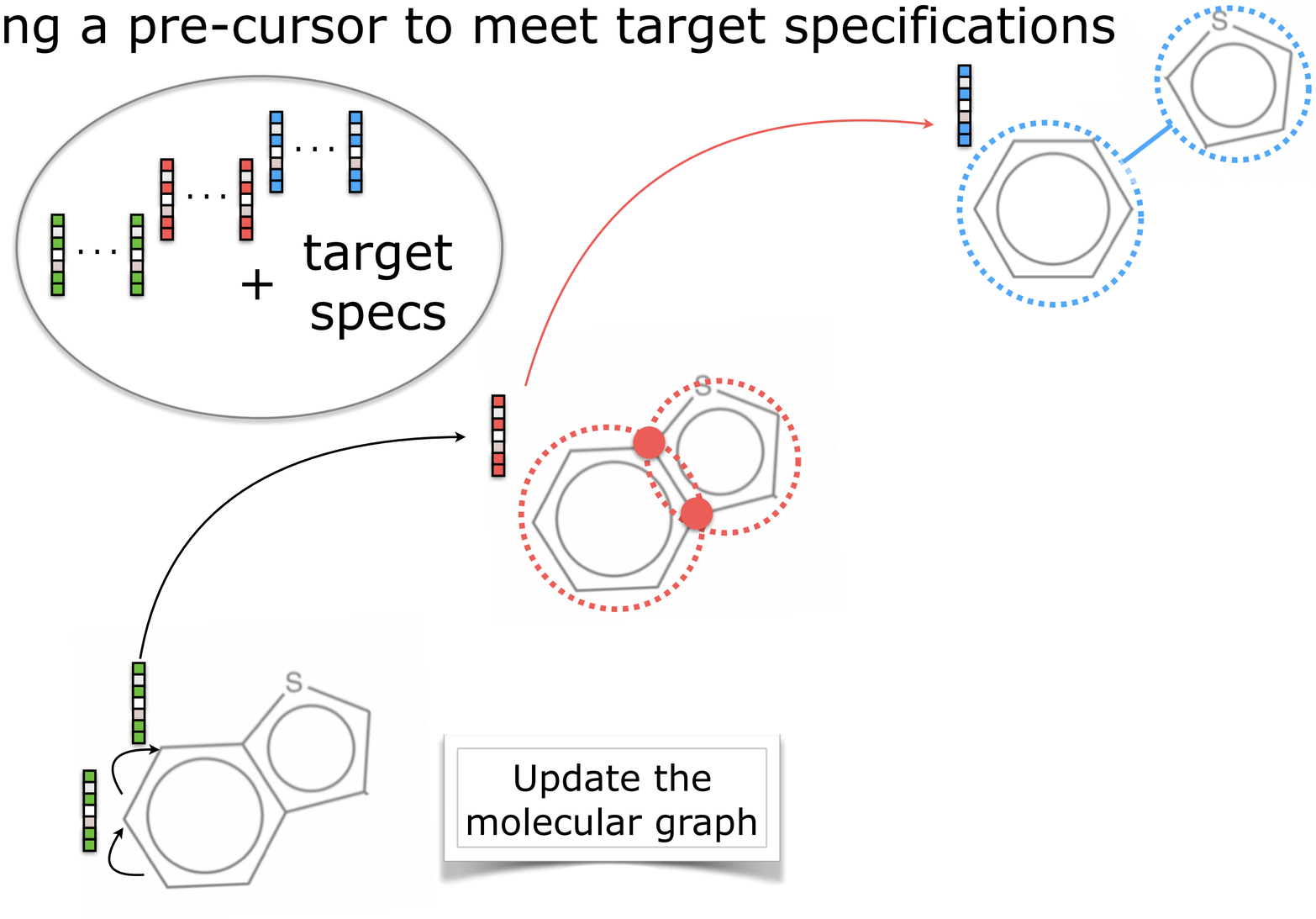
- Modifying a pre-cursor to meet target specifications



Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Graph-to-Graph Translation (Decoder)

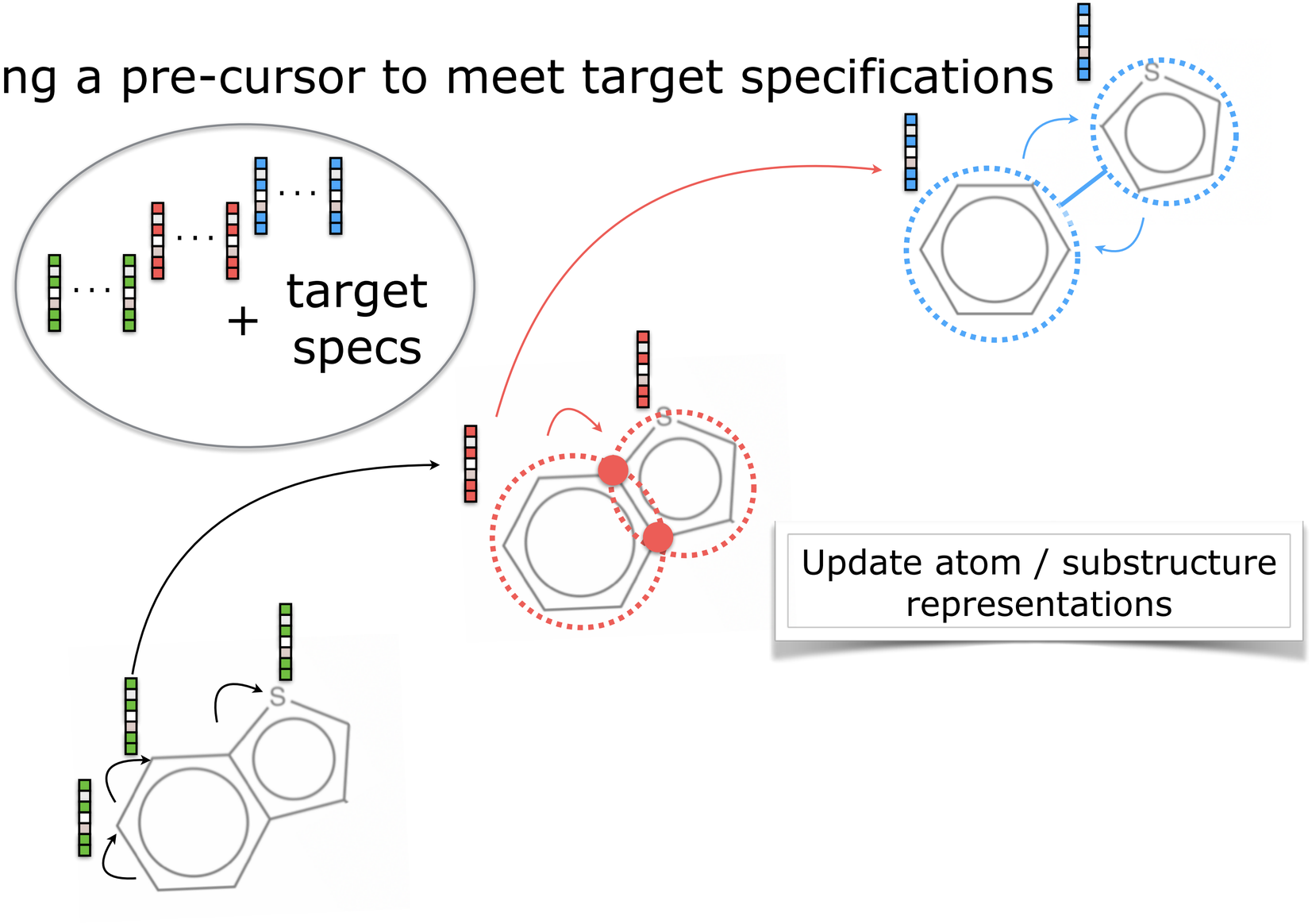
- Modifying a pre-cursor to meet target specifications



Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Graph-to-Graph Translation (Decoder)

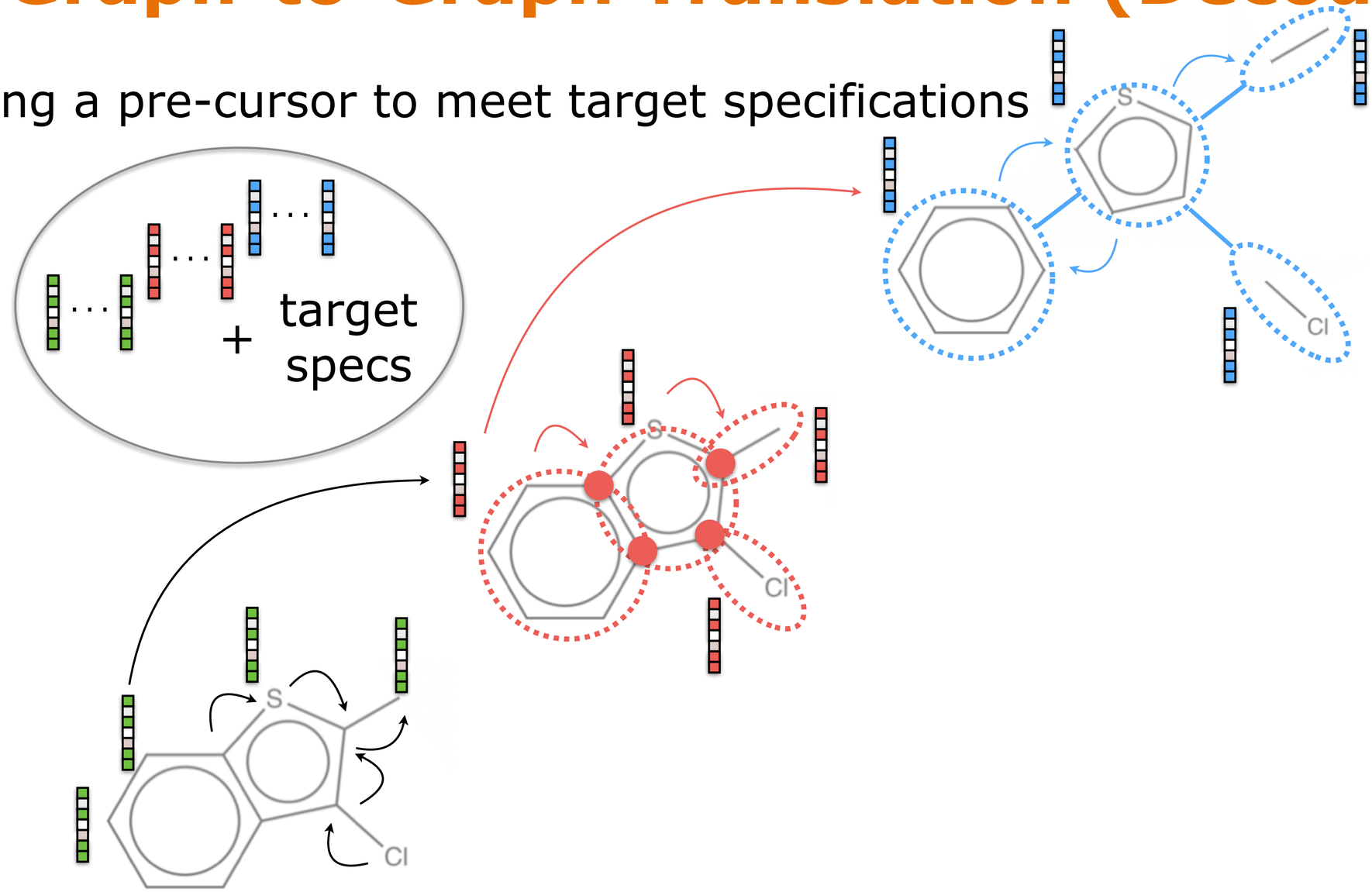
- Modifying a pre-cursor to meet target specifications



Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Graph-to-Graph Translation (Decoder)

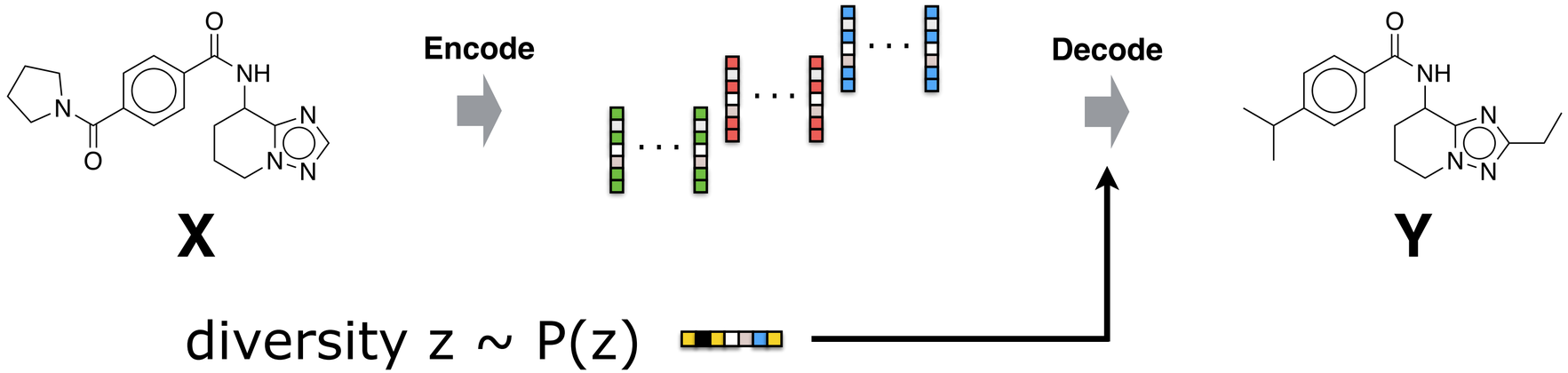
- Modifying a pre-cursor to meet target specifications

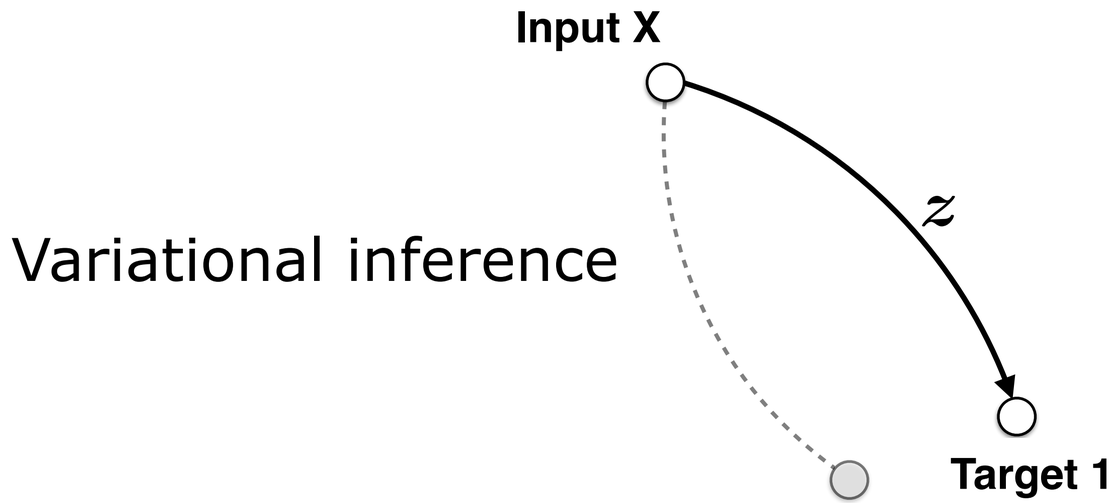


Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

De novo molecule optimization: diversity

- **Goal:** We aim to programmatically turn precursor molecules into versions that satisfy given design specifications



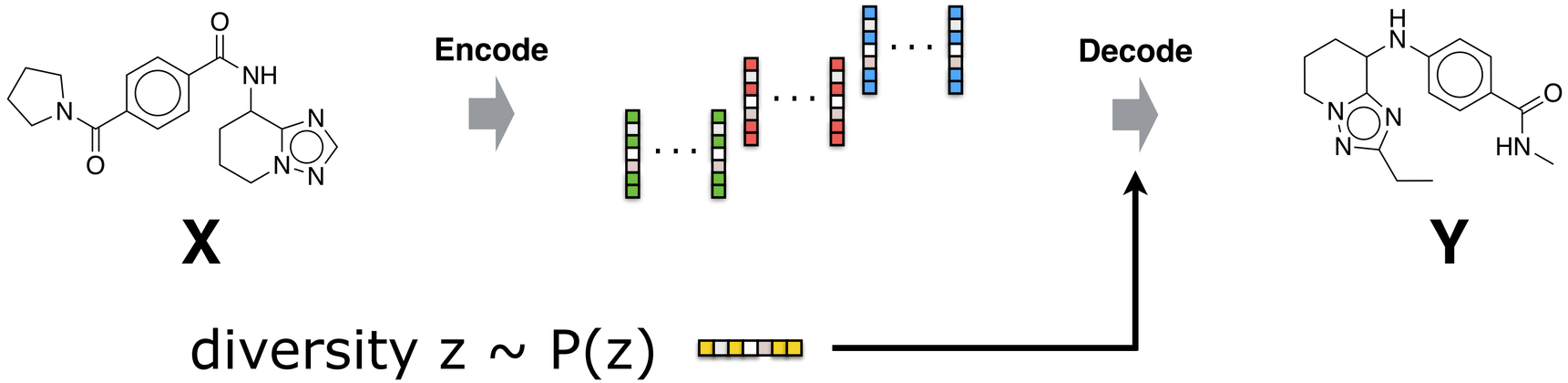


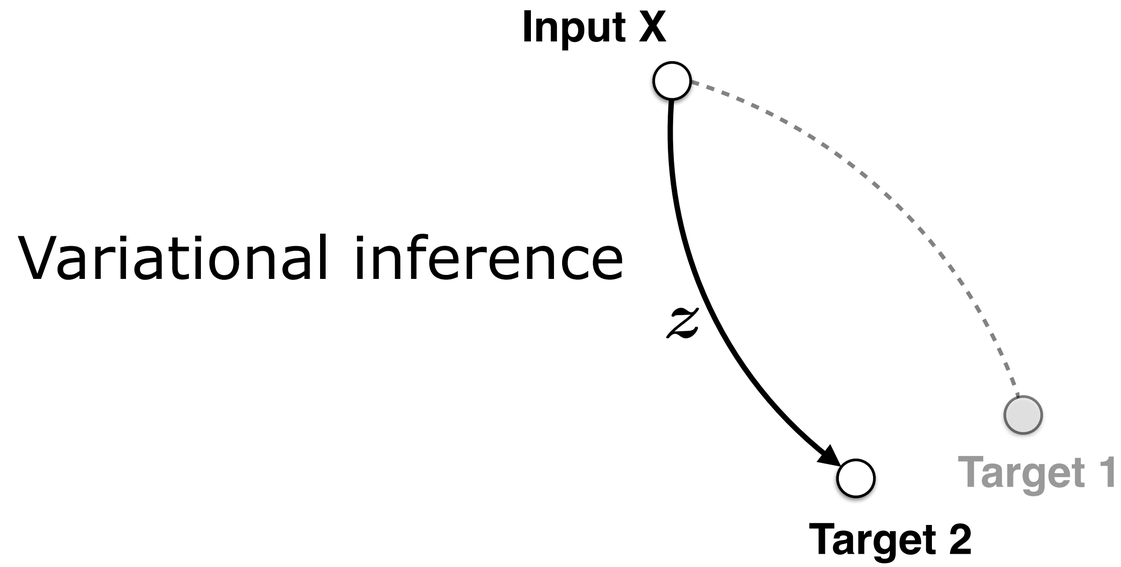
Variational inference

Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

De novo molecule optimization: diversity

- **Goal:** We aim to programmatically turn precursor molecules into versions that satisfy given design specifications

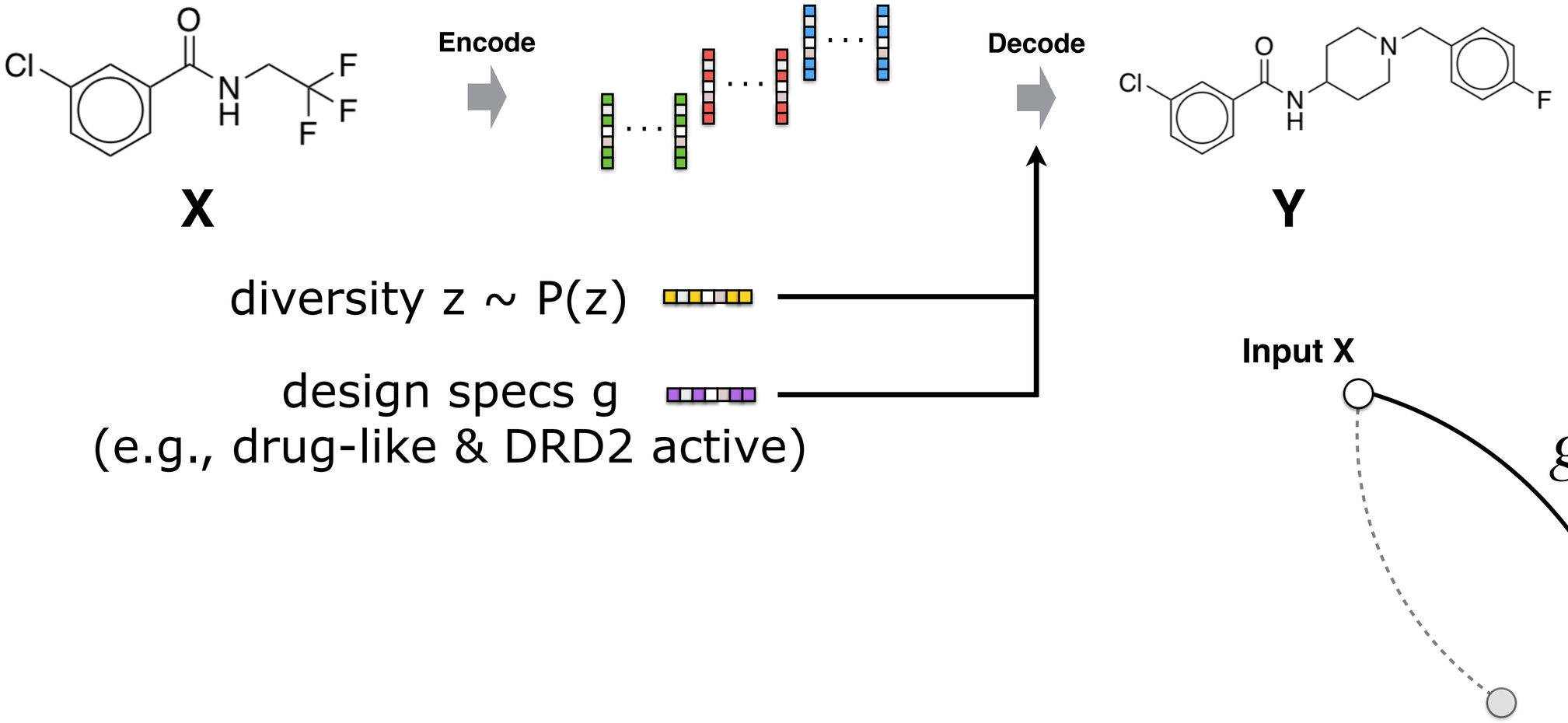




Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

De novo molecule optimization: specs

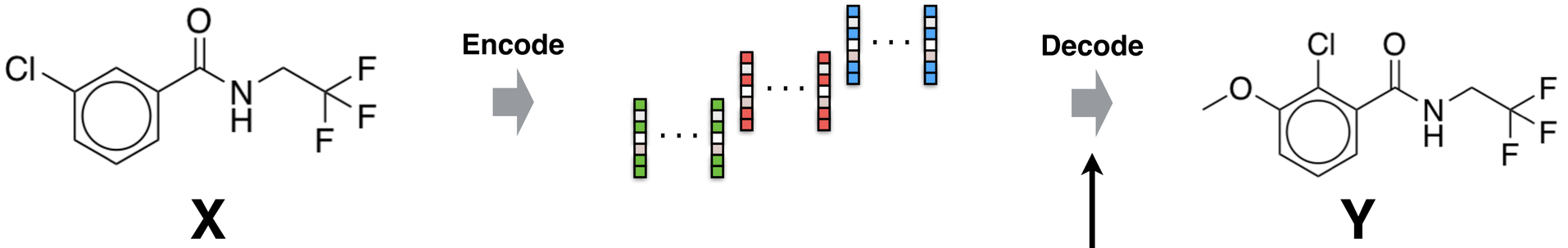
- **Goal:** We aim to programmatically turn precursor molecules into versions that satisfy given design specifications

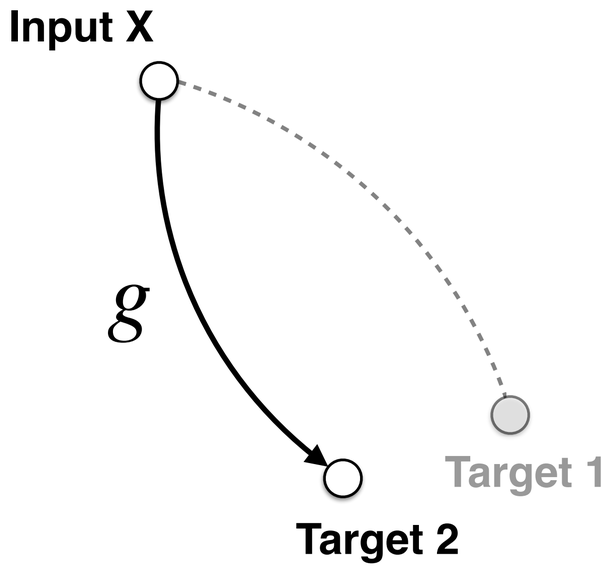


Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

De novo molecule optimization: specs

- **Goal:** We aim to programmatically turn precursor molecules into versions that satisfy given design specifications

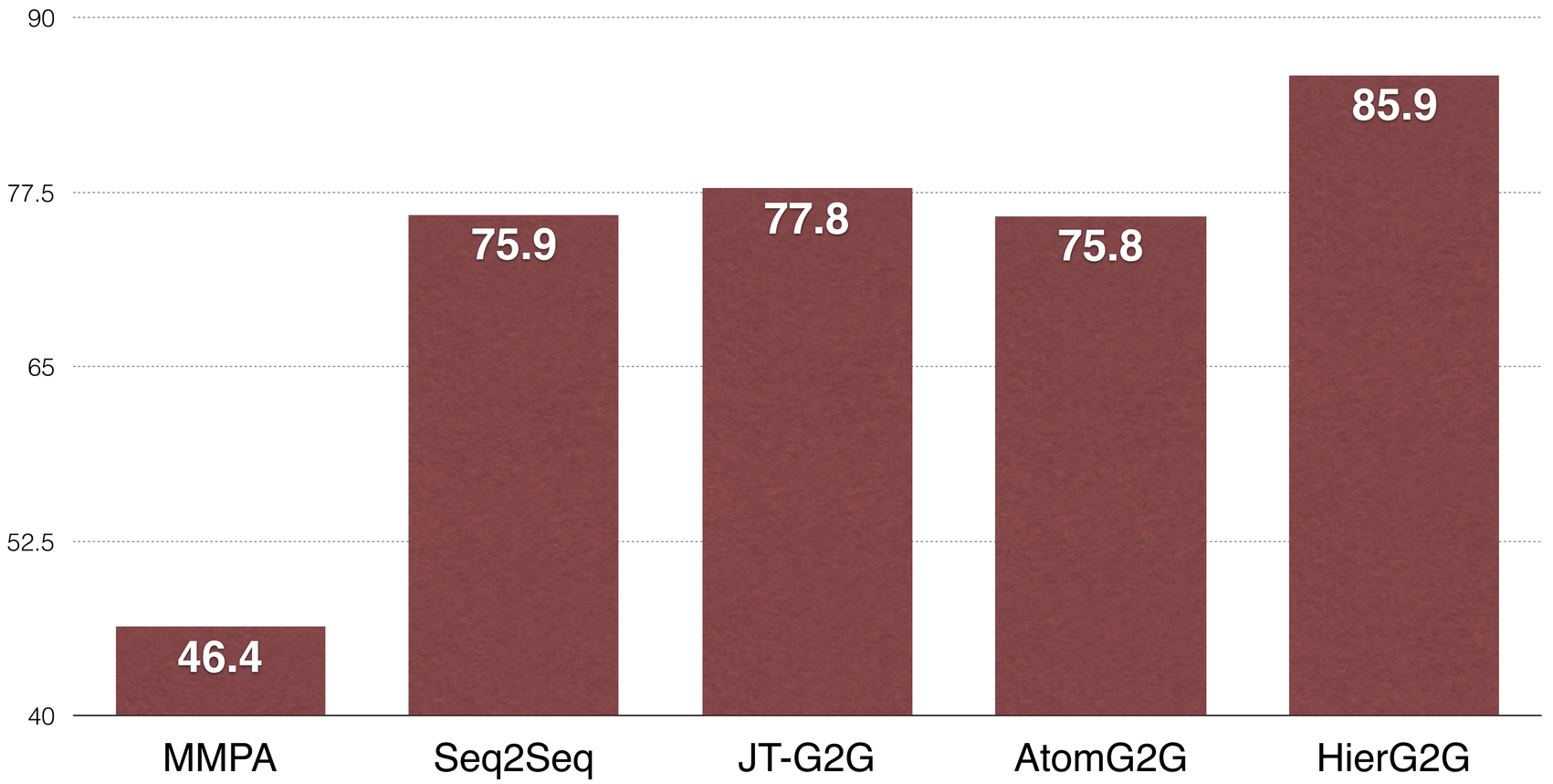




Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Example results (DRD2)

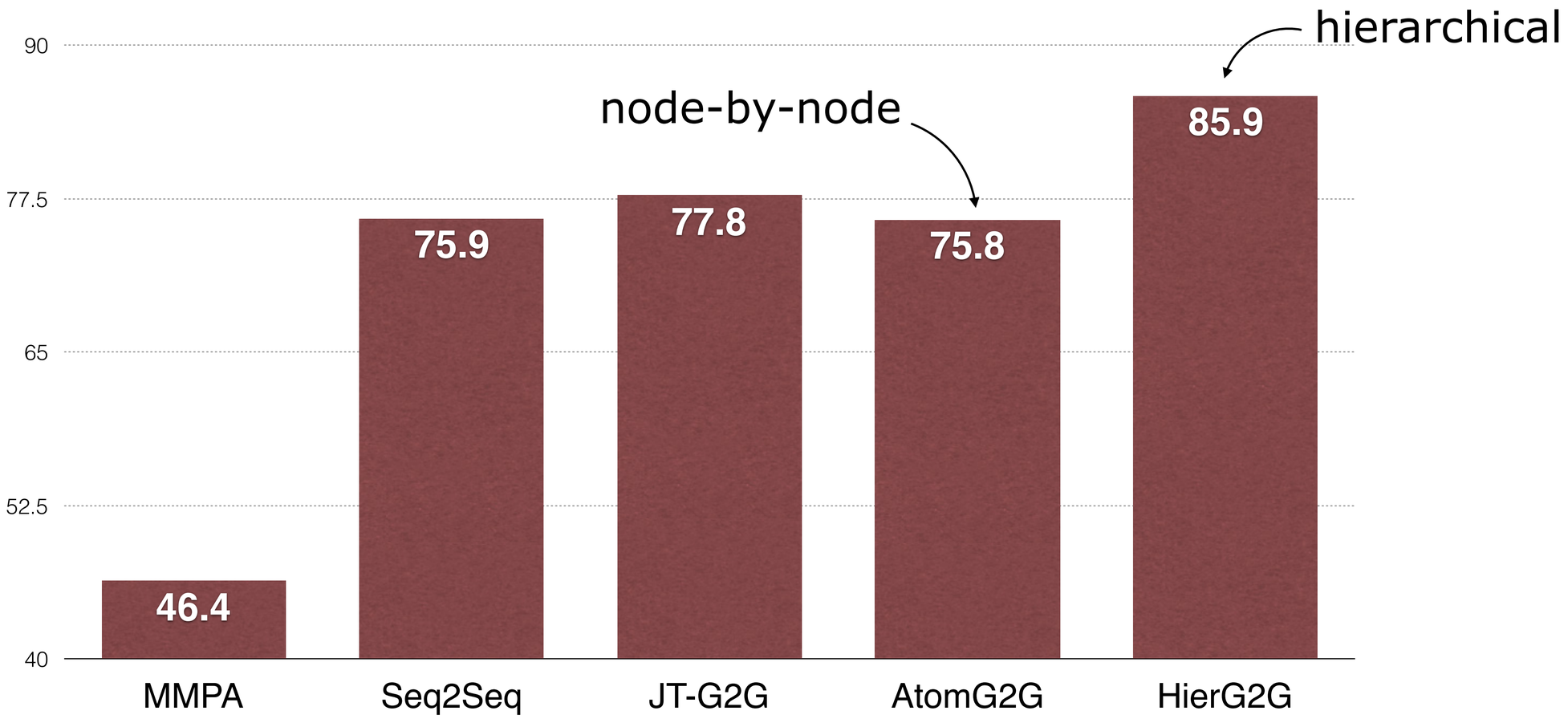
- Single property optimization: DRD2 success % (from inactive to active)



Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Example results (DRD2)

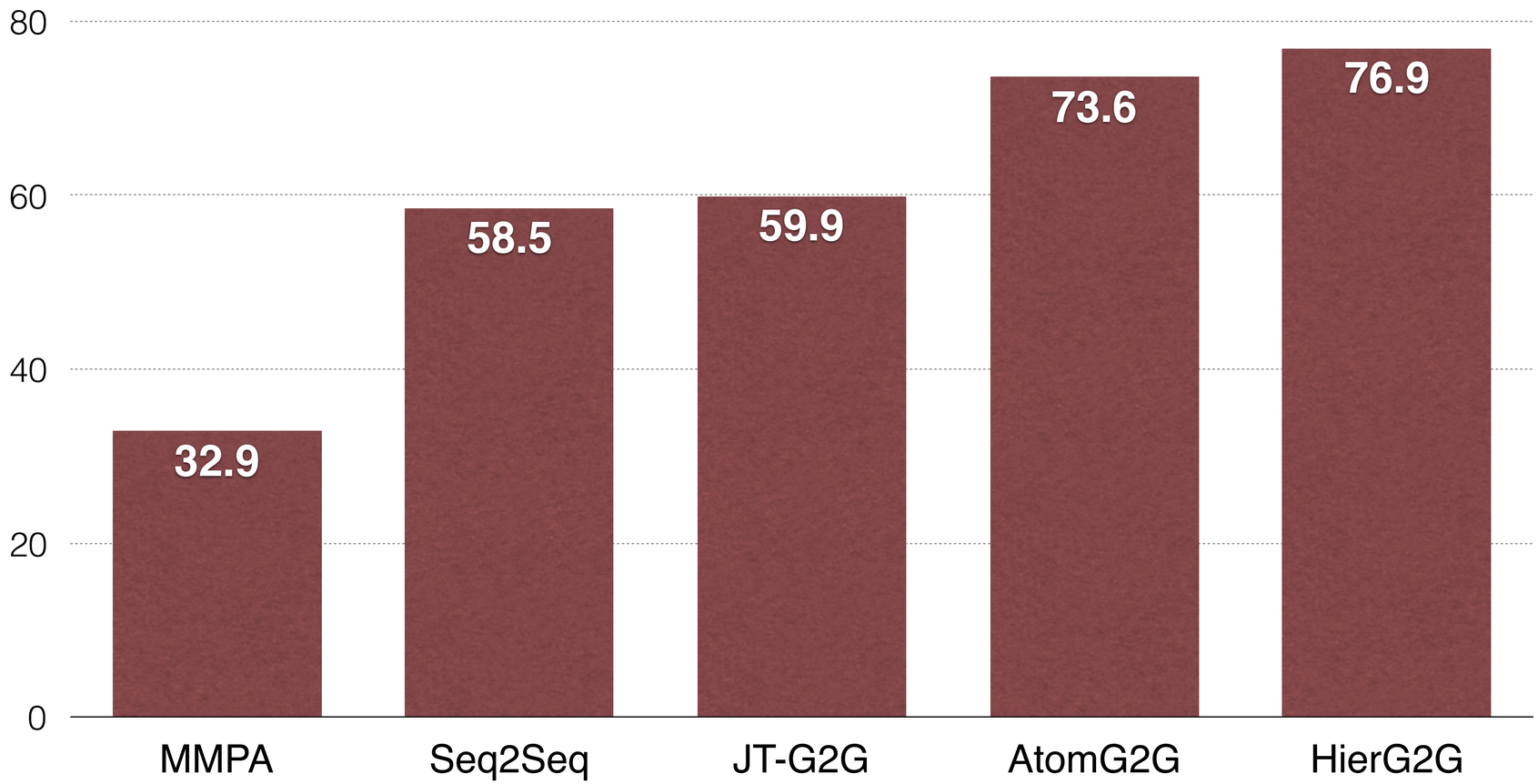
- Single property optimization: DRD2 success % (from inactive to active)



Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Example results (drug-likeness)

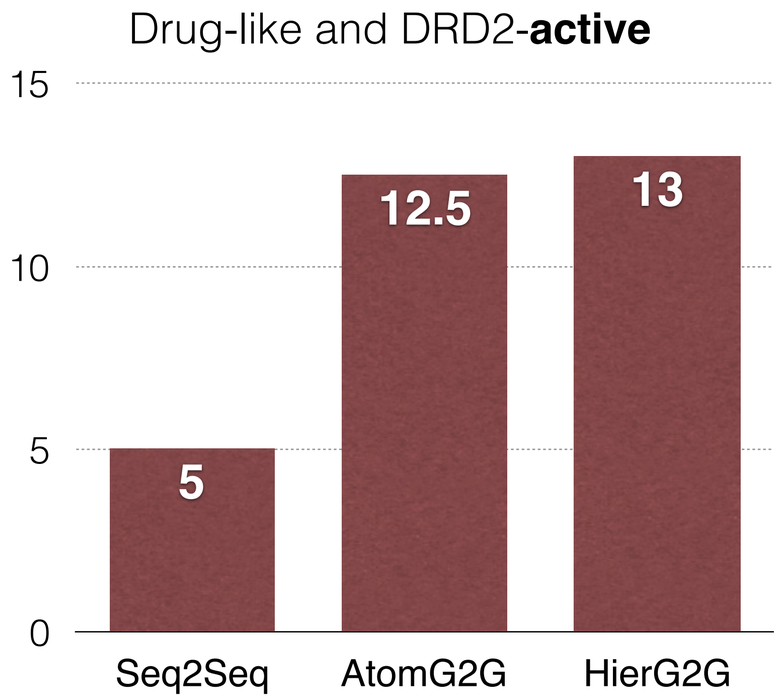
- Single property optimization: drug-likeness (QED) success % (QED > 0.9)

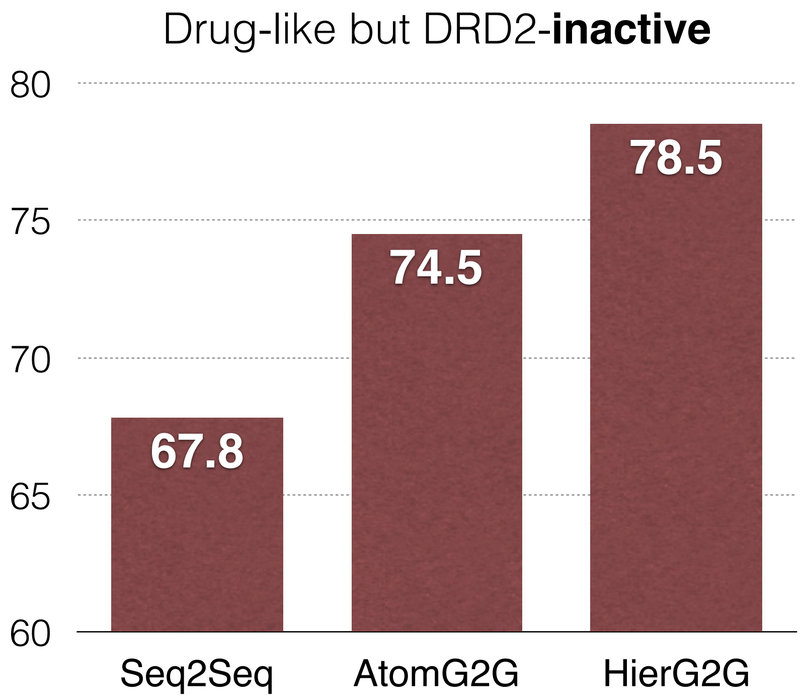


Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Example results (multiple design specs)

- Multi-criteria success % (design specs driven generation)





- **Challenge:** only 1.6% training pairs are both drug-like and DRD2-active

---

Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Disentangling what's important

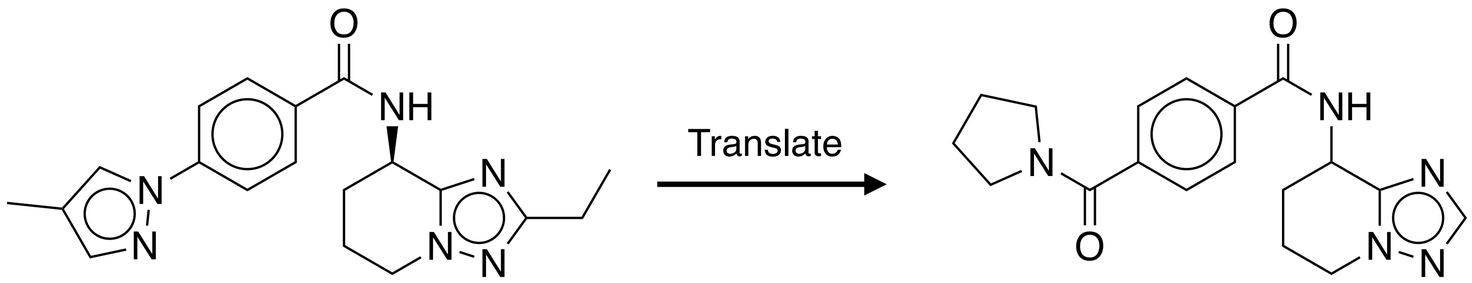
- Models are complicated, important to assess how individual parts contribute to performance

Method	QED	DRD2
HierG2G	**76.9%**	**85.9%**
• atom-based decoder	76.1%	75.0%
• two-layer encoder	75.8%	83.5%
• one-layer encoder	67.8%	74.1%
• GRU MPN	72.6%	83.7%

Hierarchical Graph-to-Graph Translation for Molecules (2019). W. Jin, R. Barzilay, and T. Jaakkola

Still many ways to improve

- Generating complex objects (e.g., molecules) is hard
- Assessing the quality of the object (property prediction) is substantially easier



hard to realize

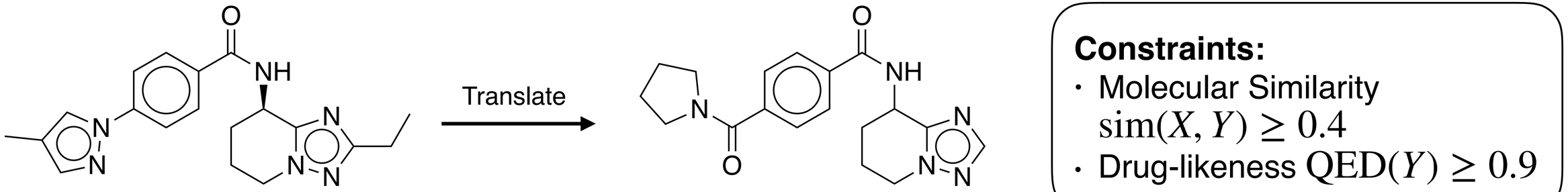
Constraints:

- Molecular Similarity $\mathrm{sim}(X, Y) \geq 0.4$
- Drug-likeness $\mathrm{QED}(Y) \geq 0.9$

easier to check/predict

Property-guided generation

- Generating complex objects (e.g., molecules) is hard
- Assessing the quality of the object (property prediction) is substantially easier



Constraints:
- Molecular Similarity $\mathrm{sim}(X, Y) \geq 0.4$
- Drug-likeness $\mathrm{QED}(Y) \geq 0.9$

hard to realize

easier to check/predict

- **Target augmentation:** we can use property predictor to generate additional (self-supervised) data for the generative model

---

Iterative Target Augmentation for Effective Conditional Generation (2019). K. Yang, W. Jin, K. Swanson, R. Barzilay, and T. Jaakkola

Target augmentation = stochastic EM

- **Objective:** maximize the log-probability that generated candidates satisfy the properties of interest (structure is now a latent variable)

$$\sum_{X \in \text{source set}} \log \left[\sum_{Y} P(\text{target specs}|Y) P(Y|X;\theta) \right]$$

- **E-step:** generate candidates from the current model; filter/reweight by property predictor (~ posterior samples)

- **M-step:** maximize the log-probability of new (weighted) targets

---

Iterative Target Augmentation for Effective Conditional Generation (2019). K. Yang, W. Jin, K. Swanson, R. Barzilay, and T. Jaakkola

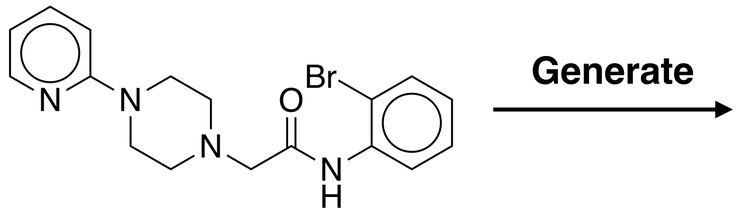
Target augmentation = stochastic EM

- **Objective:** maximize the log-probability that generated candidates satisfy the properties of interest (structure is now a latent variable)

$$\sum_{X \in \text{source set}} \log \left[\sum_{Y} P(\text{target specs}|Y) P(Y|X;\theta) \right]$$

- **E-step:** generate candidates from the current model; filter/reweight by property predictor (~ posterior samples)

- **M-step:** maximize the log-probability of new (weighted) targets



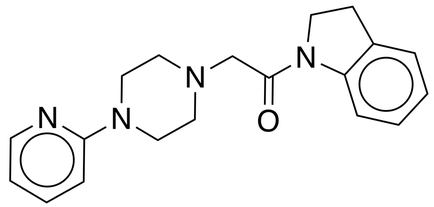
Iterative Target Augmentation for Effective Conditional Generation (2019). K. Yang, W. Jin, K. Swanson, R. Barzilay, and T. Jaakkola

Target augmentation = stochastic EM

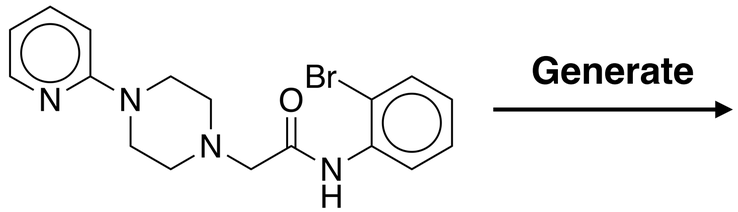
- **Objective:** maximize the log-probability that generated candidates satisfy the properties of interest (structure is now a latent variable)

$$\sum_{X \in \text{source set}} \log \left[\sum_{Y} P(\text{target specs}|Y) P(Y|X;\theta) \right]$$

- **E-step:** generate candidates from the current model; filter/reweight by property predictor (~ posterior samples)



- **M-step:** maximize the log-probability of new (weighted) targets



Generate →

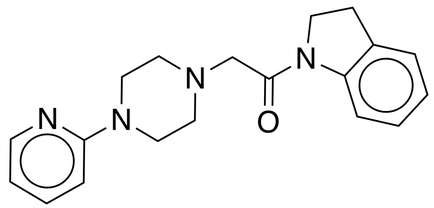
Iterative Target Augmentation for Effective Conditional Generation (2019). K. Yang, W. Jin, K. Swanson, R. Barzilay, and T. Jaakkola

Target augmentation = stochastic EM

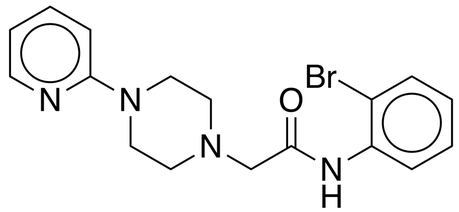
- **Objective:** maximize the log-probability that generated candidates satisfy the properties of interest (structure is now a latent variable)

$$\sum_{X \in \text{source set}} \log \left[\sum_{Y} P(\text{target specs}|Y) P(Y|X;\theta) \right]$$

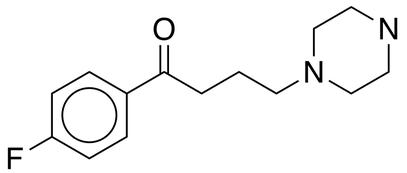
- **E-step:** generate candidates from the current model; filter/reweight by property predictor (~ posterior samples)



- **M-step:** maximize the log-probability of new (weighted) targets



Generate



Iterative Target Augmentation for Effective Conditional Generation (2019). K. Yang, W. Jin, K. Swanson, R. Barzilay, and T. Jaakkola

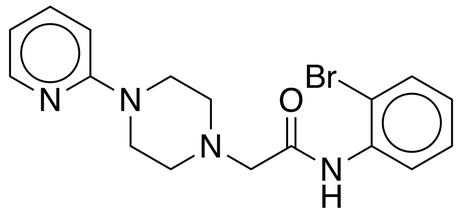
Target augmentation = stochastic EM

- **Objective:** maximize the log-probability that generated candidates satisfy the properties of interest (structure is now a latent variable)

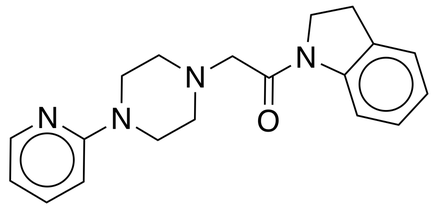
$$\sum_{X \in \text{source set}} \log \left[\sum_{Y} P(\text{target specs}|Y) P(Y|X;\theta) \right]$$

- **E-step:** generate candidates from the current model; filter/reweight by property predictor (~ posterior samples)

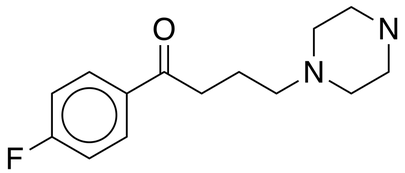
- **M-step:** maximize the log-probability of new (weighted) targets



Generate



DRD2=0.933

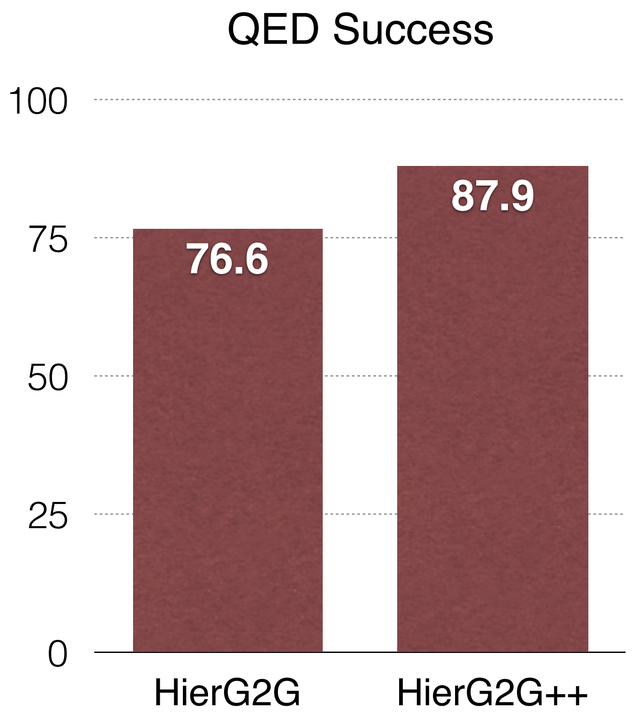


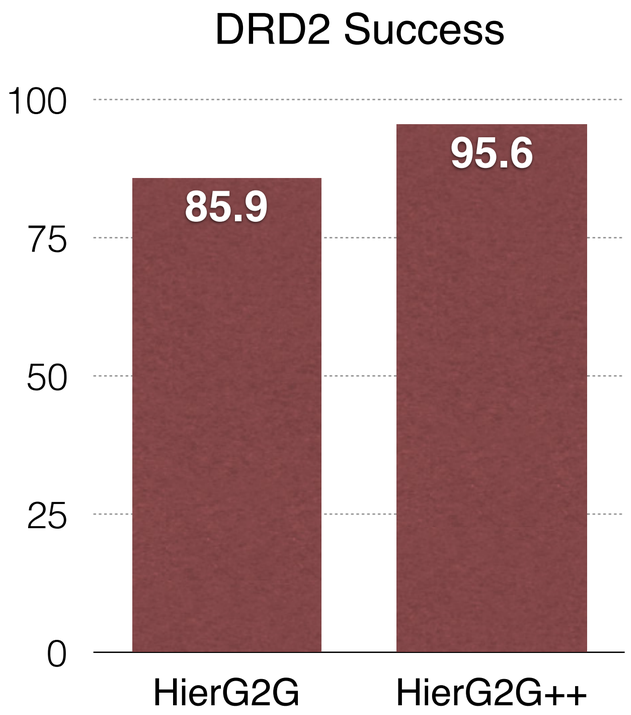
DRD2=0.008

Iterative Target Augmentation for Effective Conditional Generation (2019). K. Yang, W. Jin, K. Swanson, R. Barzilay, and T. Jaakkola

Example results: gains

- Substantial gains in translation/optimization success %

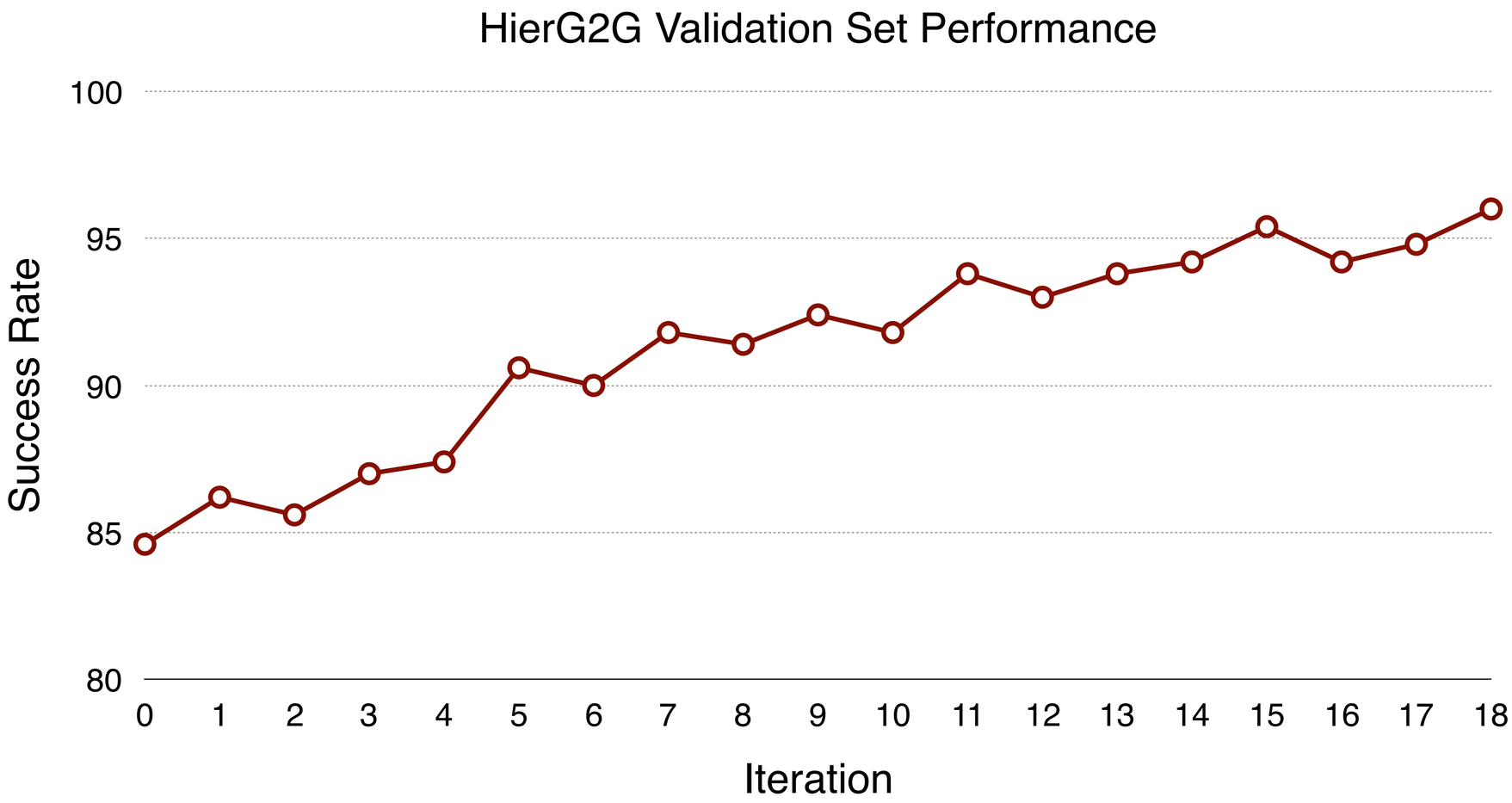




Iterative Target Augmentation for Effective Conditional Generation (2019). K. Yang, W. Jin, K. Swanson, R. Barzilay, and T. Jaakkola

Example results: gains

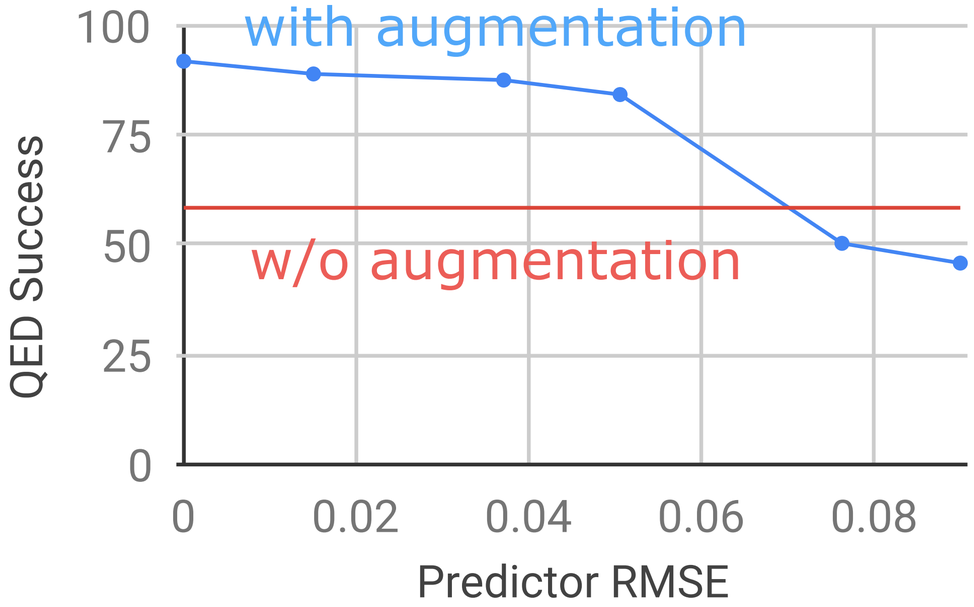
- Consistently improving …

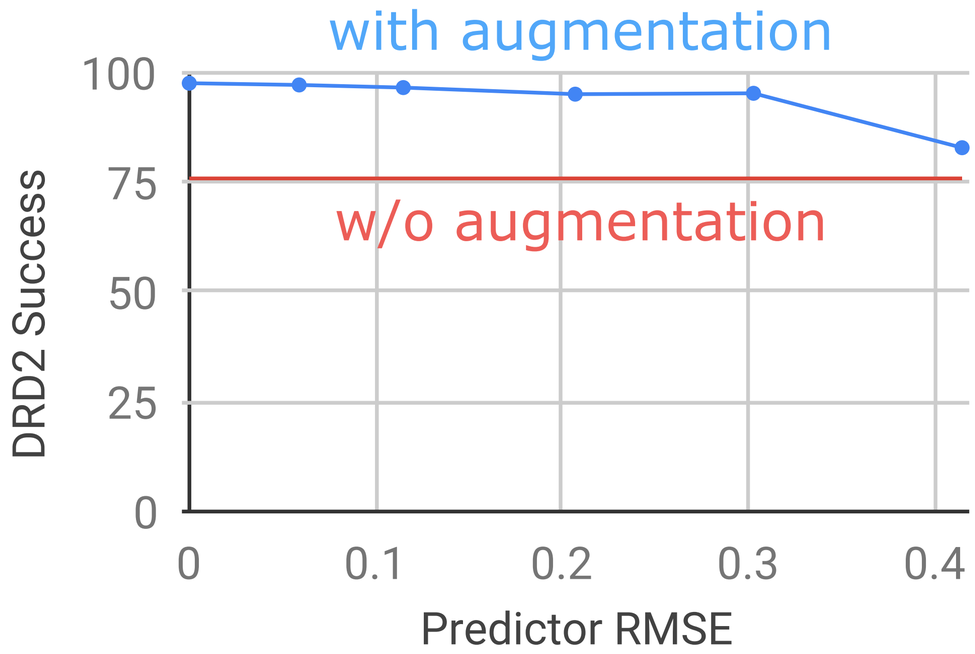


Iterative Target Augmentation for Effective Conditional Generation (2019). K. Yang, W. Jin, K. Swanson, R. Barzilay, and T. Jaakkola

Example results: robustness

- The gains are robust against errors in the property predictor





- Note: curves are for a weaker seq2seq model; baseline performance is much lower, but final performance with augmentation comparable to hierG2G.

Iterative Target Augmentation for Effective Conditional Generation (2019). K. Yang, W. Jin, K. Swanson, R. Barzilay, and T. Jaakkola

Summary

- Molecules as structured objects provide a rich domain for developing ML tools; key underlying challenges shared with other areas involving generation/ manipulation of diverse objects

- ML molecular design methods are rapidly becoming viable tools for drug discovery

- Several key challenges remain, however:
  - effective multi-criteria optimization
  - incorporating 3D features, physical constraints
  - generalizing to new, unexplored chemical spaces (domain transfer)
  - explainability, etc.