Molecule Optimization with Deep Generative Models

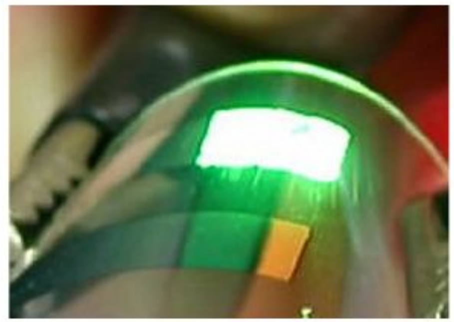
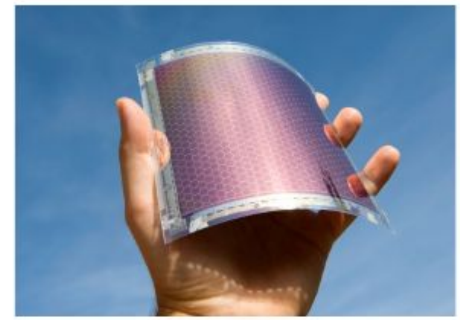
(LIKE22)

José Miguel Hernández–Lobato
Department of Engineering
University of Cambridge

http://jmhl.org, jmh233@cam.ac.uk

Drug and material design

Goal: find novel molecules that optimally fulfill various metrics.

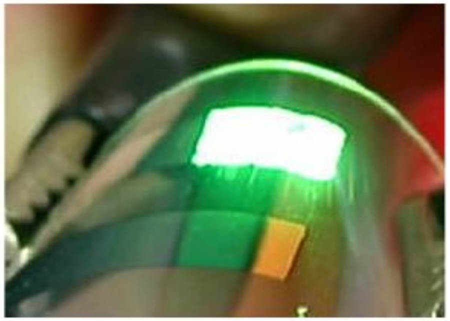
About 10^8 compounds in databases, potential ones: $10^{20} - 10^{60}$.

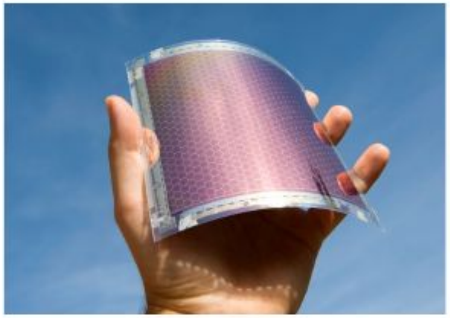
Challenges:

- Evaluating molecular properties is slow and expensive.
- Chemical space is huge.

Drug and material design

Goal: find novel molecules that optimally fulfill various metrics.





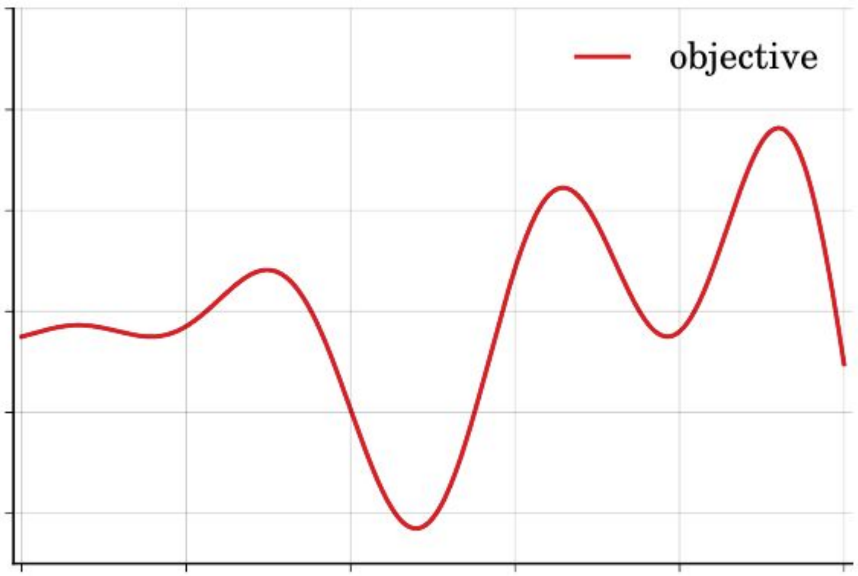
About 10^8 compounds in databases, potential ones: $10^{20} - 10^{60}$.

Challenges:

- Evaluating molecular properties is slow and expensive.
- Chemical space is huge.

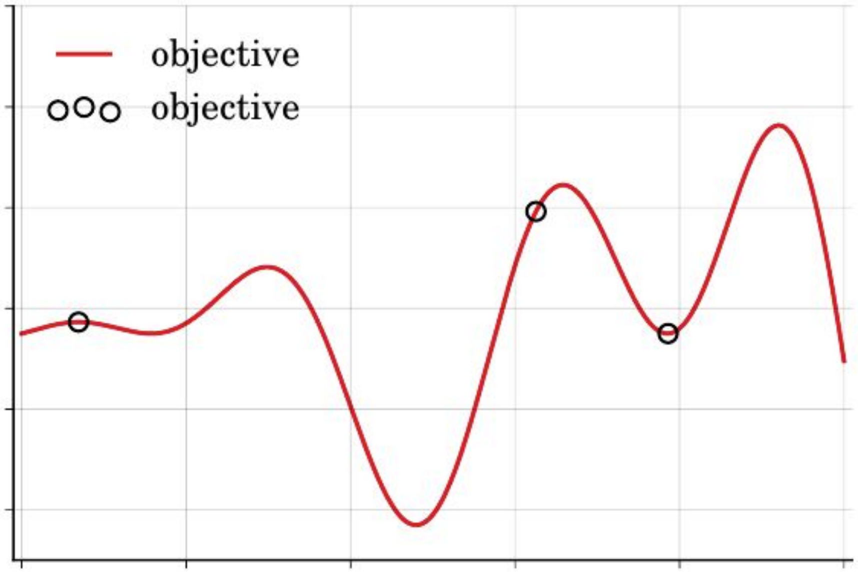
Bayesian optimization can accelerate the search.

Bayesian optimization



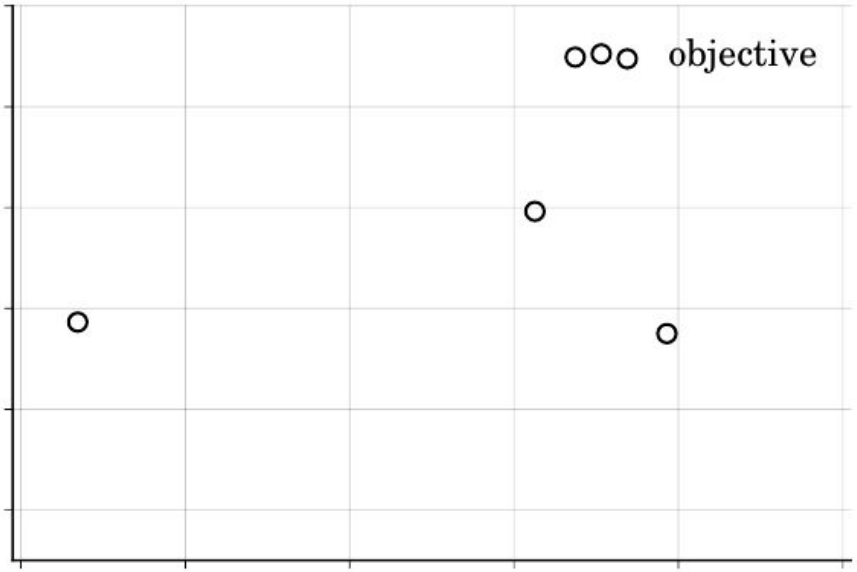
1. Get initial sample.

Bayesian optimization



❶ **Get initial sample.**

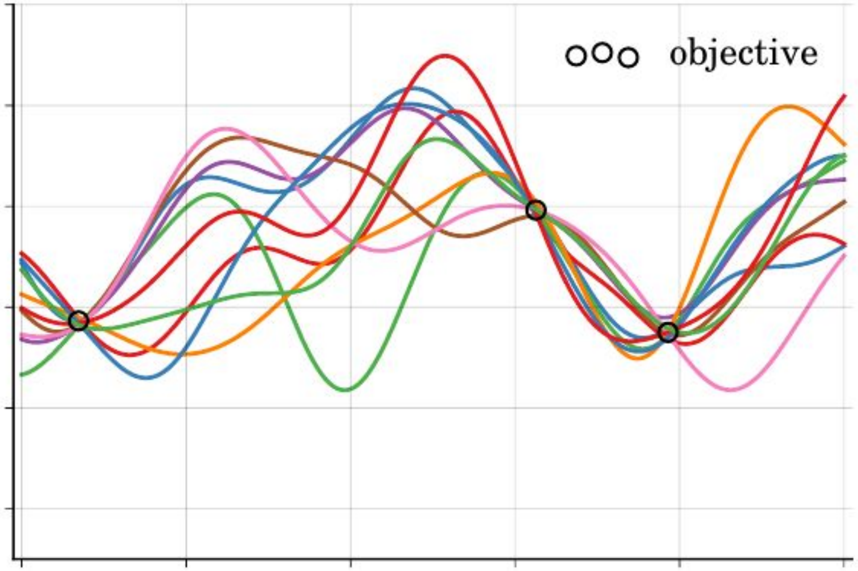
Bayesian optimization



1. **Get initial sample.**
2. Fit a model to the data:

$$p(y|\mathrm{x}, \mathcal{D}_n)\,.$$

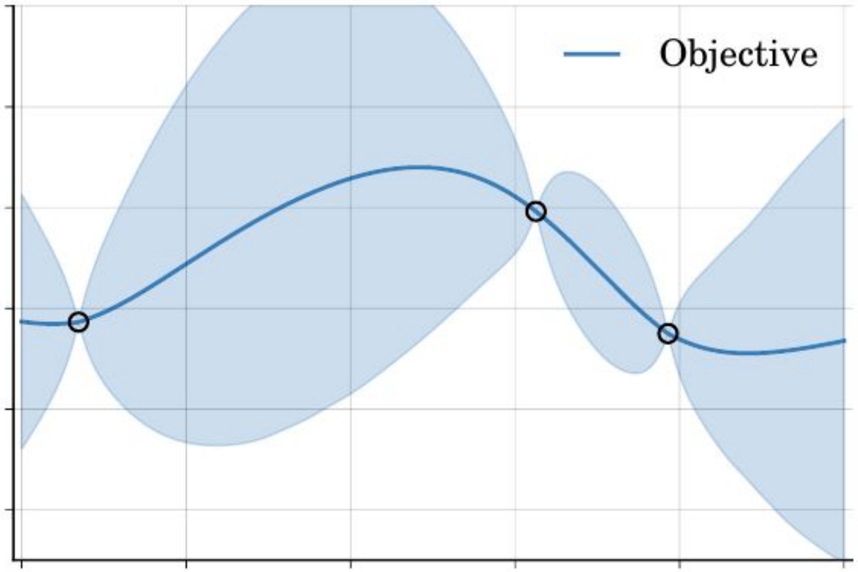
Bayesian optimization



1. Get initial sample.
2. **Fit a model to the data:**

$$p(y|x, \mathcal{D}_n)\,.$$

Bayesian optimization



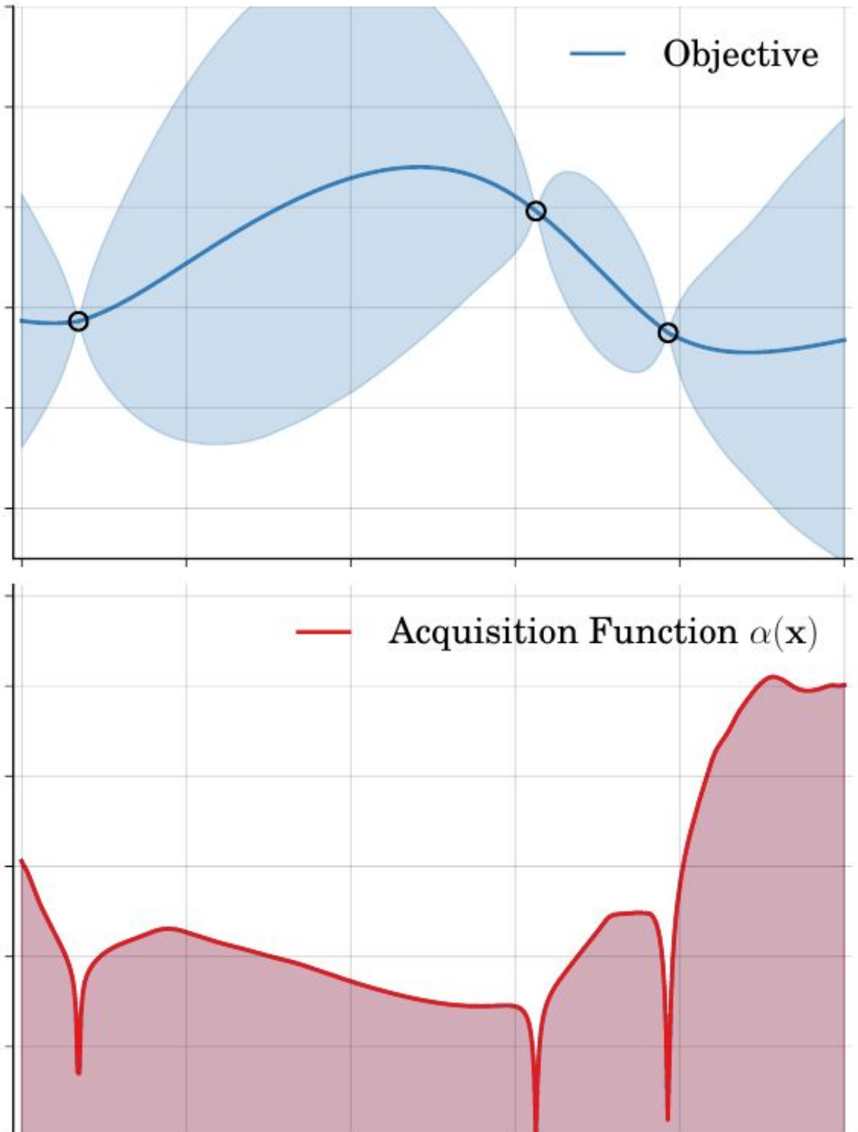
1. Get initial sample.
2. **Fit a model to the data:**

$$p(y|x, \mathcal{D}_n) .$$

3. Select data collection strategy:

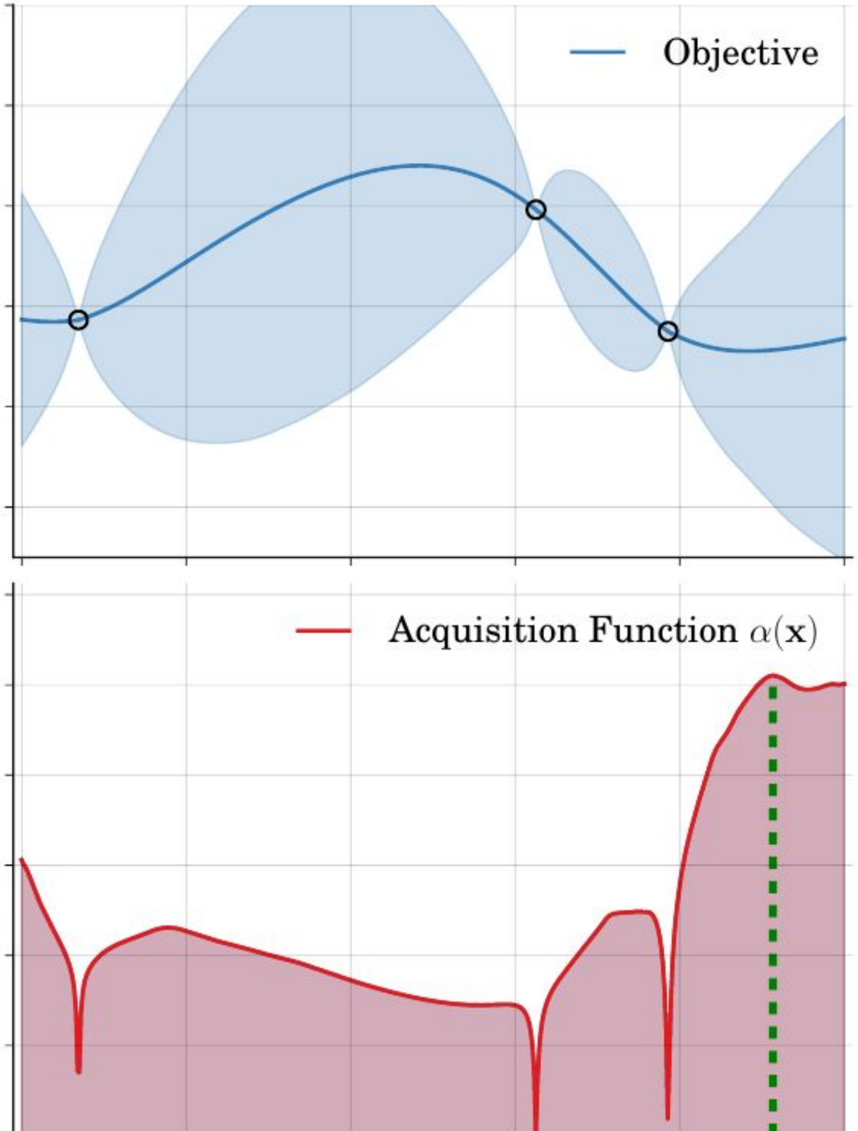
$$\alpha(x) = \mathrm{E}_{p(y|x,\mathcal{D}_n)}[U(y|x, \mathcal{D}_n)] .$$

Bayesian optimization



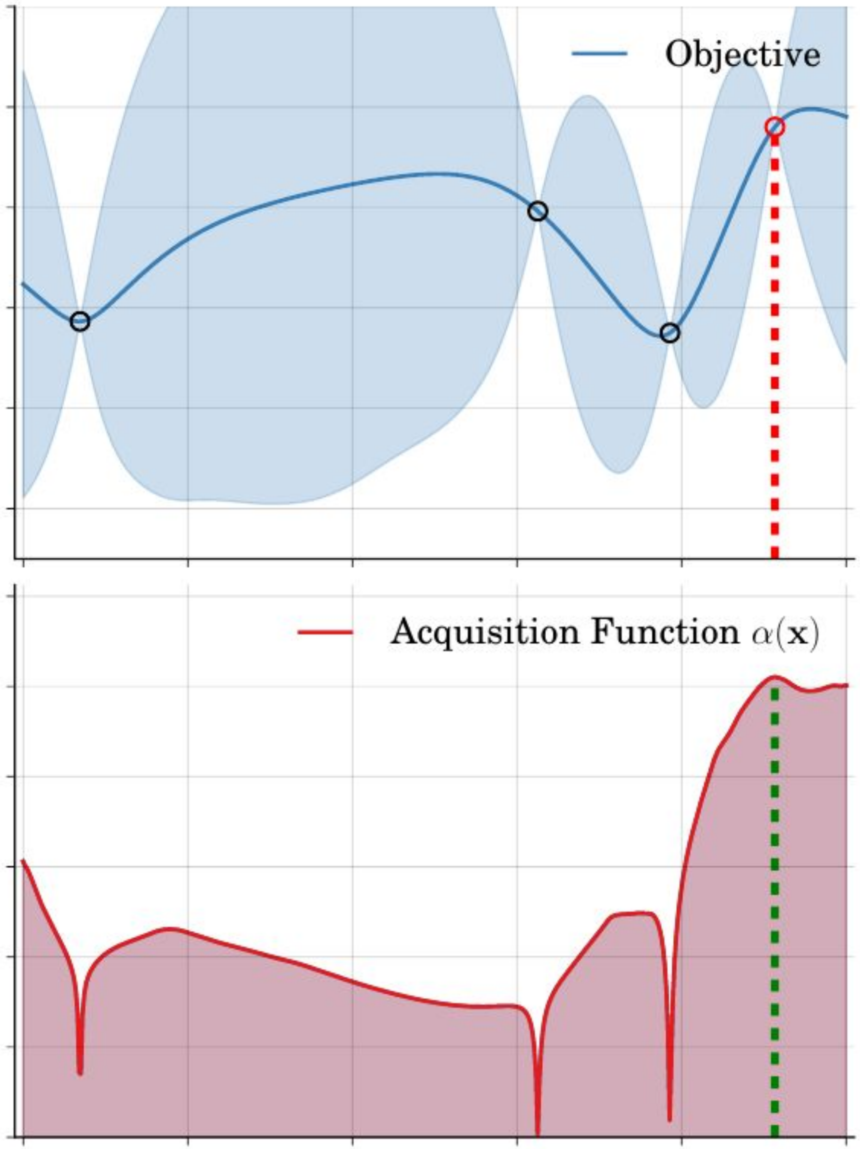
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathsf{x}, \mathcal{D}_n)\,.$$
3. **Select data collection strategy:**
$$\alpha(\mathsf{x}) = \mathsf{E}_{p(y|\mathsf{x}, \mathcal{D}_n)}[U(y|\mathsf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathsf{x})$.

Bayesian optimization



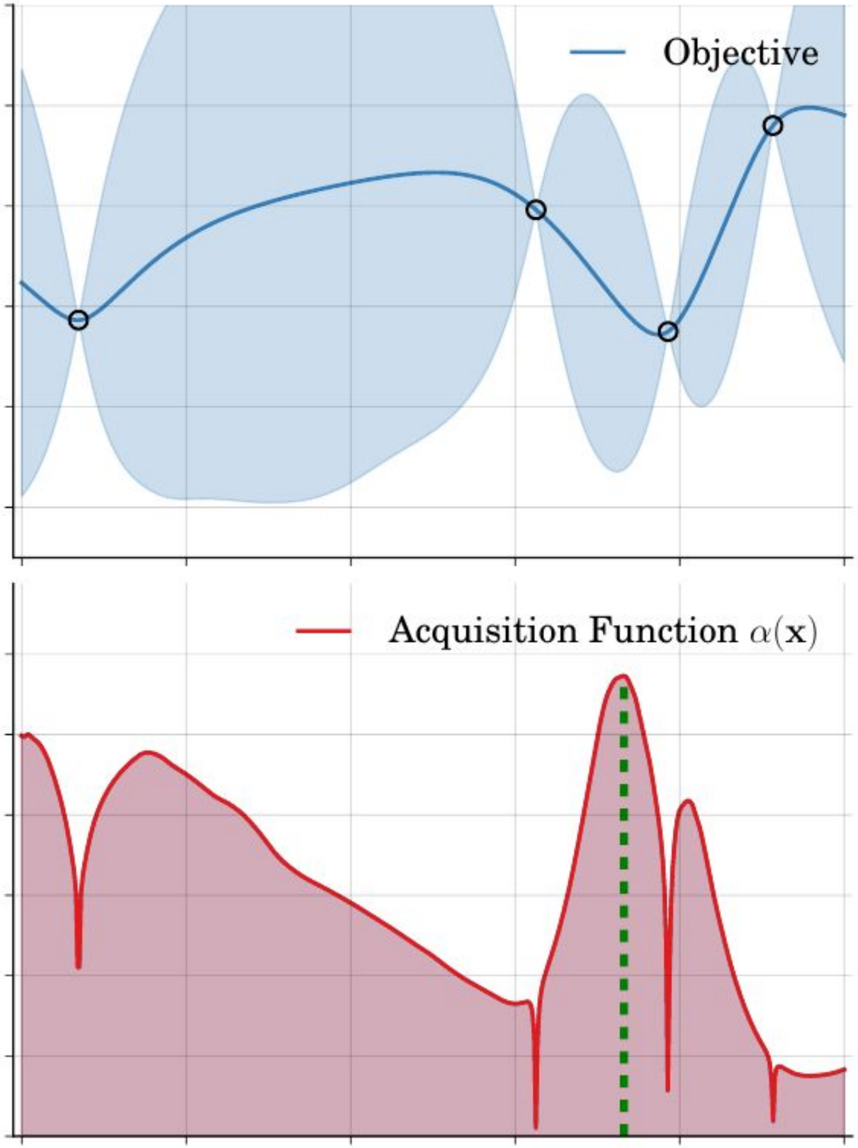
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathrm{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathrm{x}) = \mathrm{E}_{p(y|\mathrm{x}, \mathcal{D}_n)}[U(y|\mathrm{x}, \mathcal{D}_n)]\,.$$
4. **Optimize acquisition function $\alpha(\mathrm{x})$.**
5. Collect data and update model.

Bayesian optimization



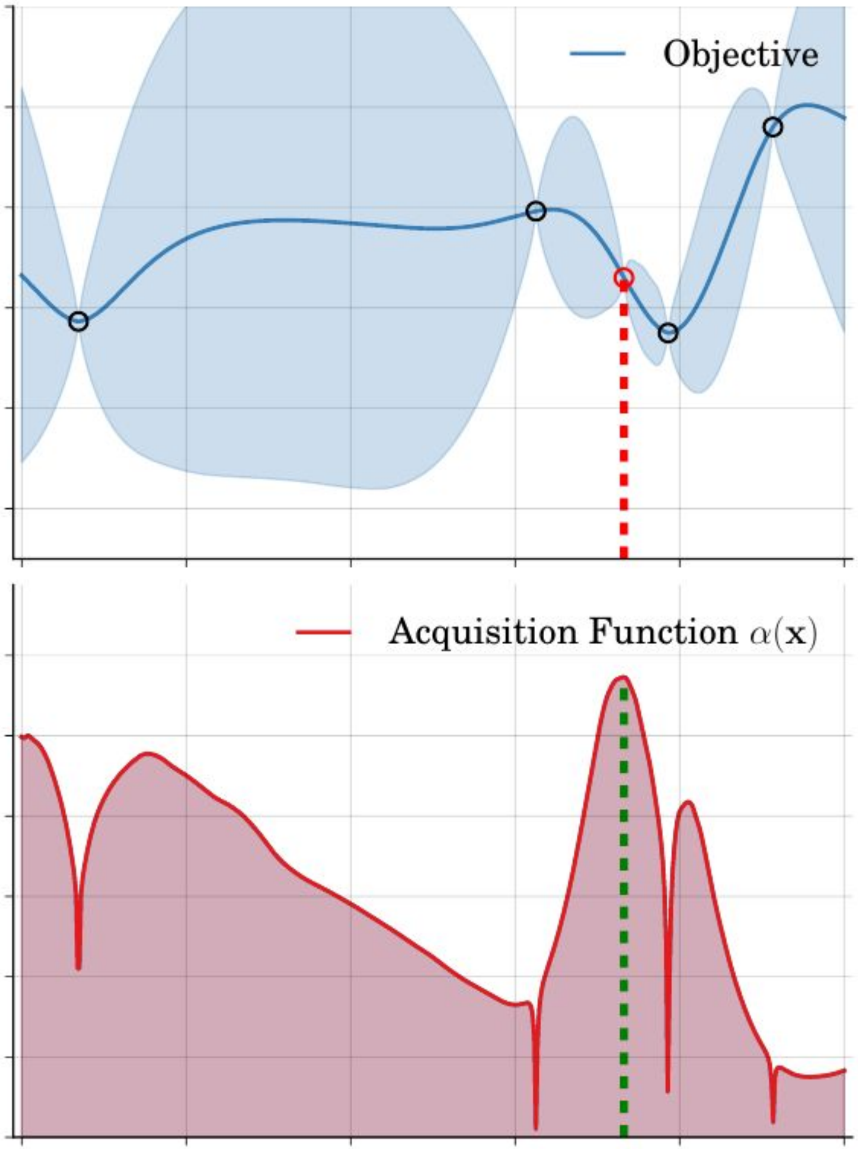
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathsf{x}, \mathcal{D}_n)\,.$$
3. Select data collection strategy:
$$\alpha(\mathsf{x}) = \mathsf{E}_{p(y|\mathsf{x},\mathcal{D}_n)}[U(y|\mathsf{x}, \mathcal{D}_n)]\,.$$
4. Optimize acquisition function $\alpha(\mathsf{x})$.
5. **Collect data and update model.**
6. Repeat!

Bayesian optimization



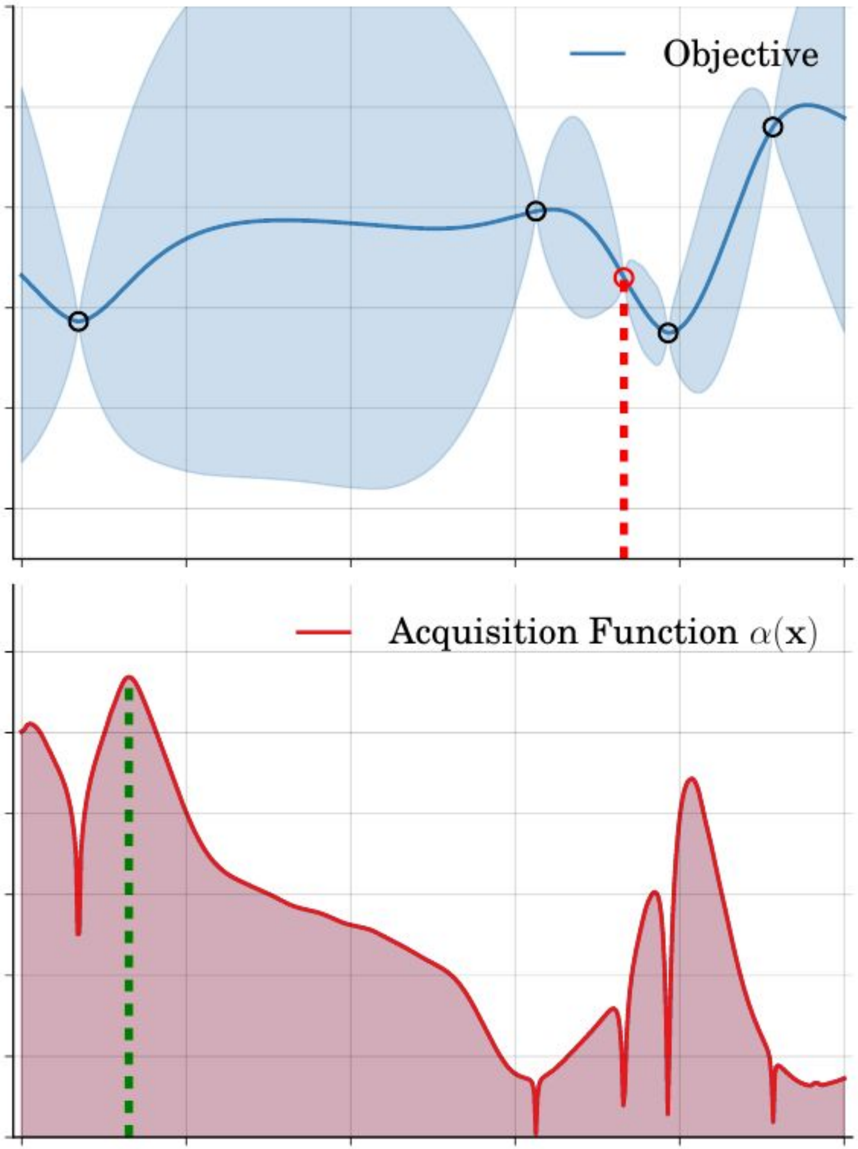
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathrm{x}, \mathcal{D}_n)\,.$$
3. **Select data collection strategy:**
$$\alpha(\mathrm{x}) = \mathrm{E}_{p(y|\mathrm{x},\mathcal{D}_n)}[U(y|\mathrm{x}, \mathcal{D}_n)]\,.$$
4. **Optimize acquisition function $\alpha(\mathrm{x})$.**
5. **Collect data and update model.**
6. **Repeat!**

Bayesian optimization



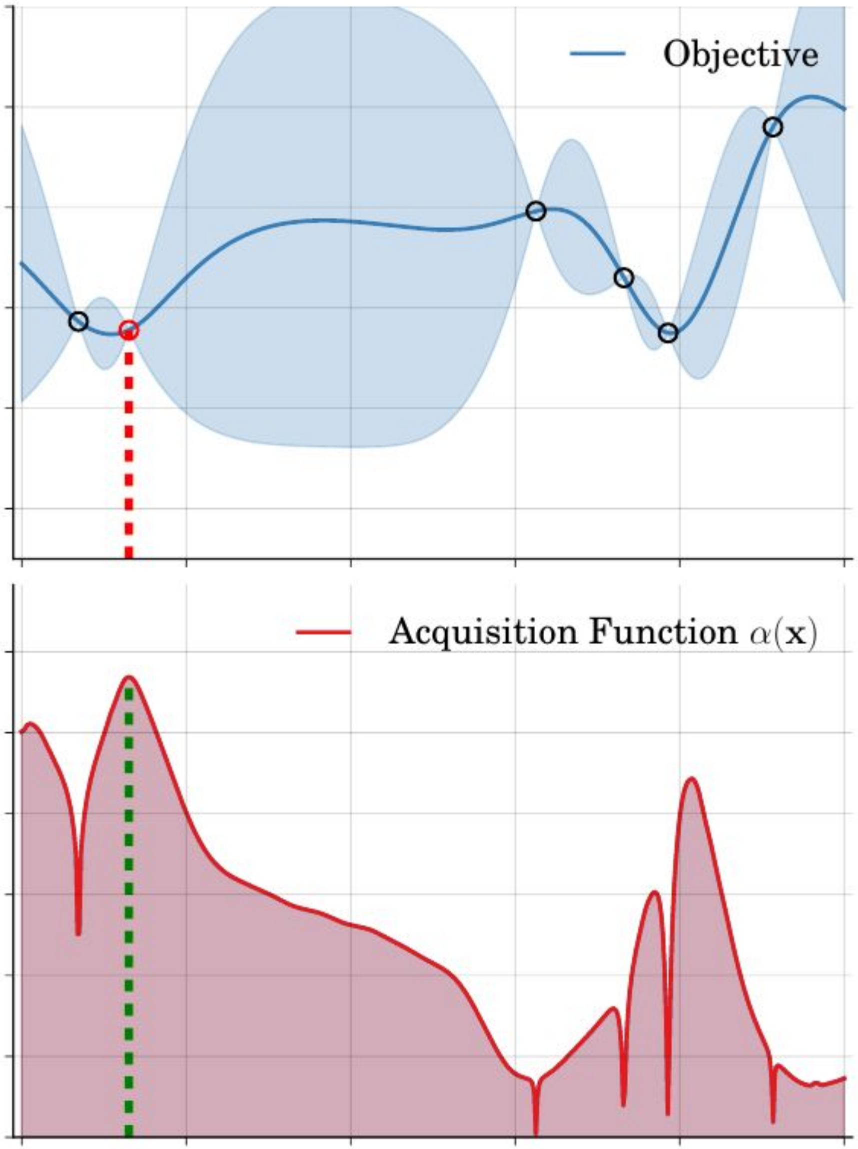
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathrm{x}, \mathcal{D}_n)\,.$$
3. **Select data collection strategy:**
$$\alpha(\mathrm{x}) = \mathrm{E}_{p(y|\mathrm{x},\mathcal{D}_n)}[U(y|\mathrm{x}, \mathcal{D}_n)]\,.$$
4. **Optimize acquisition function $\alpha(\mathrm{x})$.**
5. **Collect data and update model.**
6. **Repeat!**

Bayesian optimization



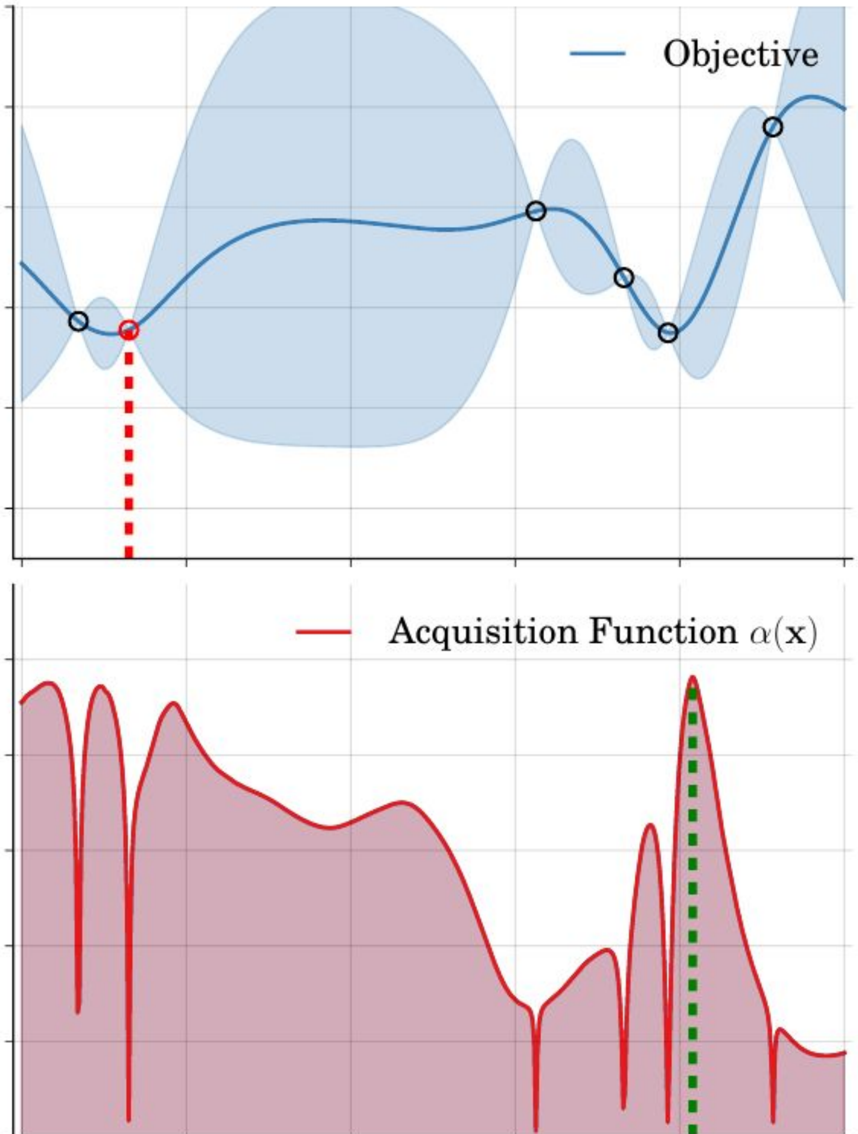
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathrm{x}, \mathcal{D}_n)\,.$$
3. **Select data collection strategy:**
$$\alpha(\mathrm{x}) = \mathrm{E}_{p(y|\mathrm{x},\mathcal{D}_n)}[U(y|\mathrm{x}, \mathcal{D}_n)]\,.$$
4. **Optimize acquisition function $\alpha(\mathrm{x})$.**
5. **Collect data and update model.**
6. **Repeat!**

Bayesian optimization



1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathrm{x}, \mathcal{D}_n)\,.$$
3. **Select data collection strategy:**
$$\alpha(\mathrm{x}) = \mathrm{E}_{p(y|\mathrm{x},\mathcal{D}_n)}[U(y|\mathrm{x}, \mathcal{D}_n)]\,.$$
4. **Optimize acquisition function $\alpha(\mathrm{x})$.**
5. **Collect data and update model.**
6. **Repeat!**

Bayesian optimization



❶ Get initial sample.

❷ Fit a model to the data:

$$p(y|\mathsf{x}, \mathcal{D}_n)\,.$$

❸ **Select data collection strategy:**

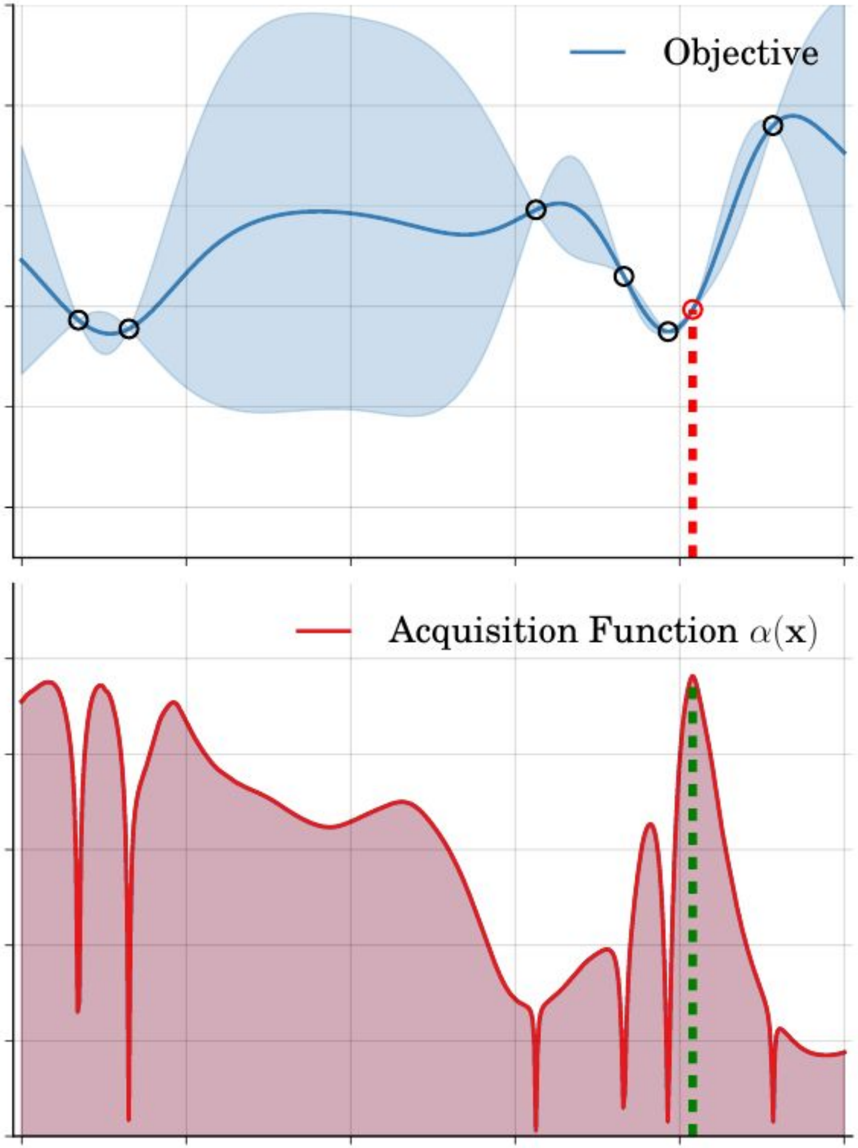
$$\alpha(\mathsf{x}) = \mathsf{E}_{p(y|\mathsf{x},\mathcal{D}_n)}[U(y|\mathsf{x},\mathcal{D}_n)]\,.$$

❹ **Optimize acquisition function $\alpha(\mathsf{x})$.**

❺ **Collect data and update model.**

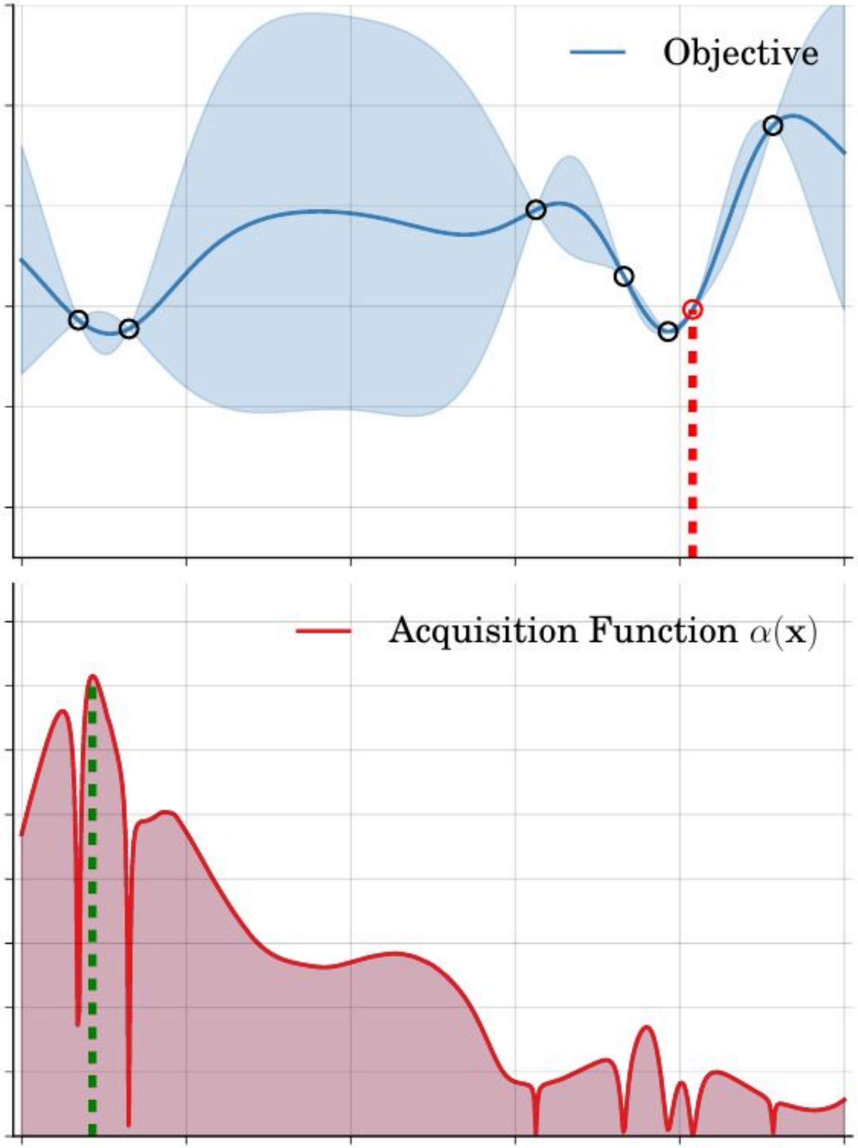
❻ **Repeat!**

Bayesian optimization



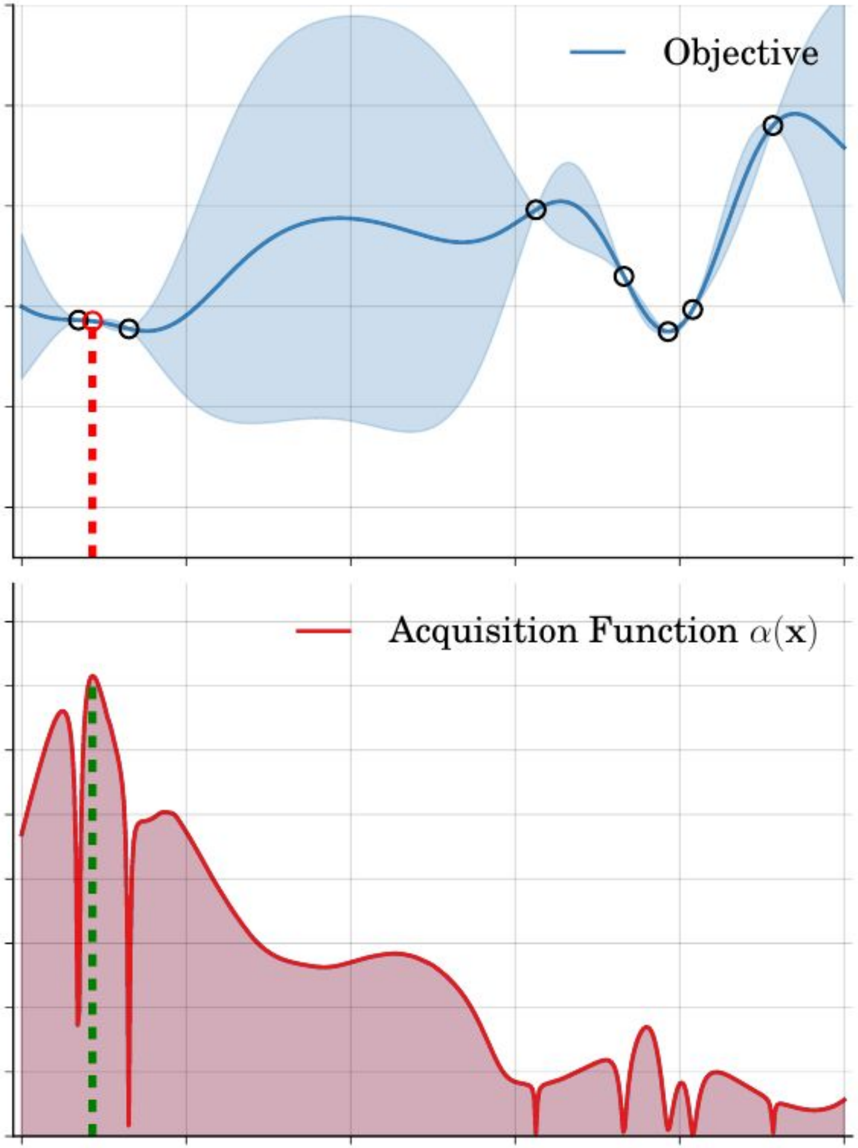
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathrm{x}, \mathcal{D}_n)\,.$$
3. **Select data collection strategy:**
$$\alpha(\mathrm{x}) = \mathrm{E}_{p(y|\mathrm{x},\mathcal{D}_n)}[U(y|\mathrm{x}, \mathcal{D}_n)]\,.$$
4. **Optimize acquisition function $\alpha(\mathrm{x})$.**
5. **Collect data and update model.**
6. **Repeat!**

Bayesian optimization



1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathrm{x}, \mathcal{D}_n)\,.$$
3. **Select data collection strategy:**
$$\alpha(\mathrm{x}) = \mathrm{E}_{p(y|\mathrm{x},\mathcal{D}_n)}[U(y|\mathrm{x}, \mathcal{D}_n)]\,.$$
4. **Optimize acquisition function $\alpha(\mathrm{x})$.**
5. **Collect data and update model.**
6. **Repeat!**

Bayesian optimization



1. Get initial sample.
2. Fit a model to the data:

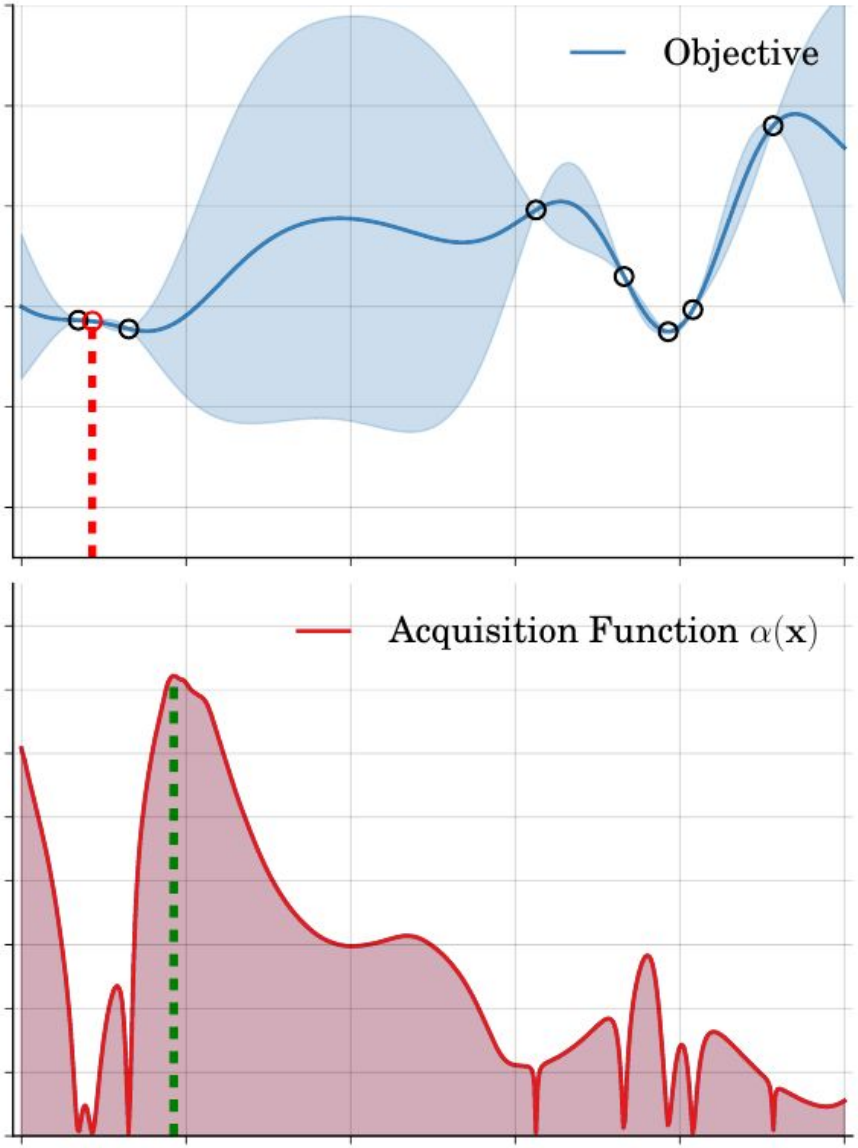
$$p(y|\mathrm{x}, \mathcal{D}_n)\,.$$

3. **Select data collection strategy:**

$$\alpha(\mathrm{x}) = \mathrm{E}_{p(y|\mathrm{x},\mathcal{D}_n)}[U(y|\mathrm{x}, \mathcal{D}_n)]\,.$$

4. **Optimize acquisition function $\alpha(\mathrm{x})$.**
5. **Collect data and update model.**
6. **Repeat!**

Bayesian optimization



1. Get initial sample.
2. Fit a model to the data:

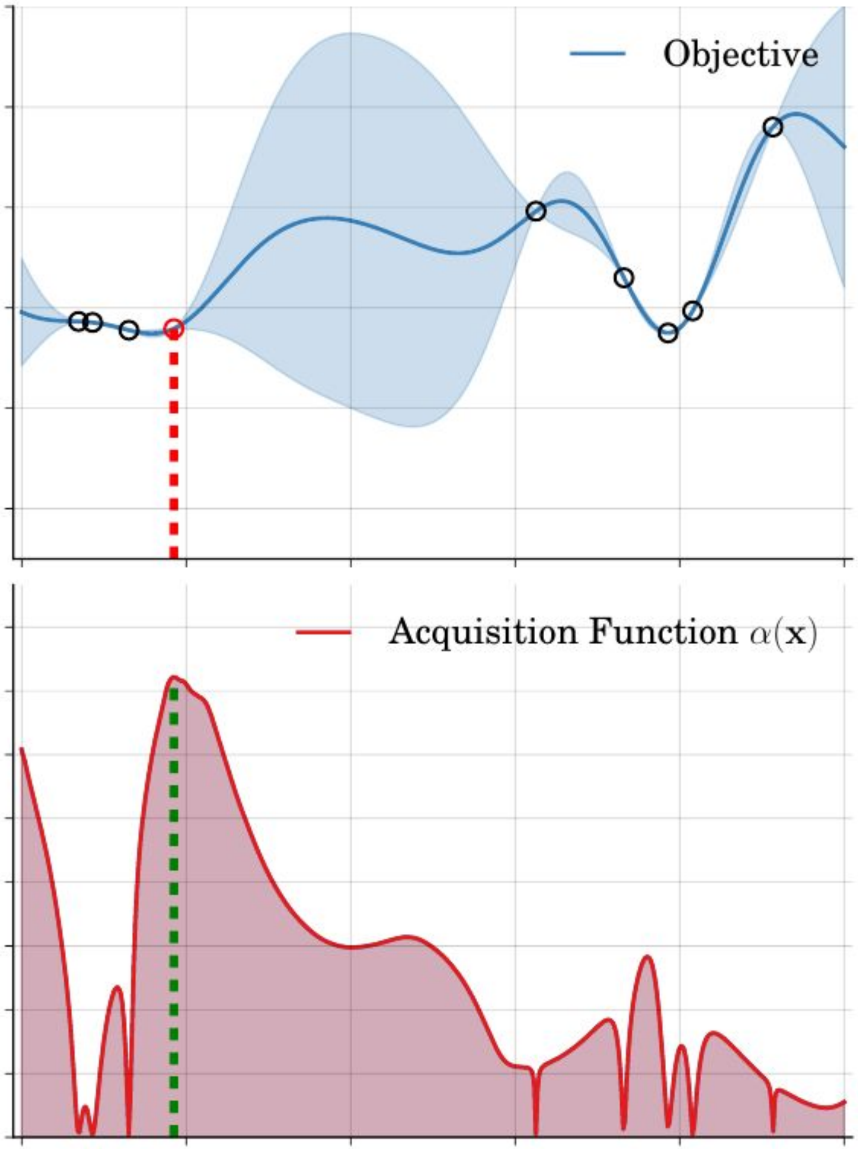
$$p(y|\mathrm{x}, \mathcal{D}_n)\,.$$

3. **Select data collection strategy:**

$$\alpha(\mathrm{x}) = \mathrm{E}_{p(y|\mathrm{x},\mathcal{D}_n)}[U(y|\mathrm{x}, \mathcal{D}_n)]\,.$$

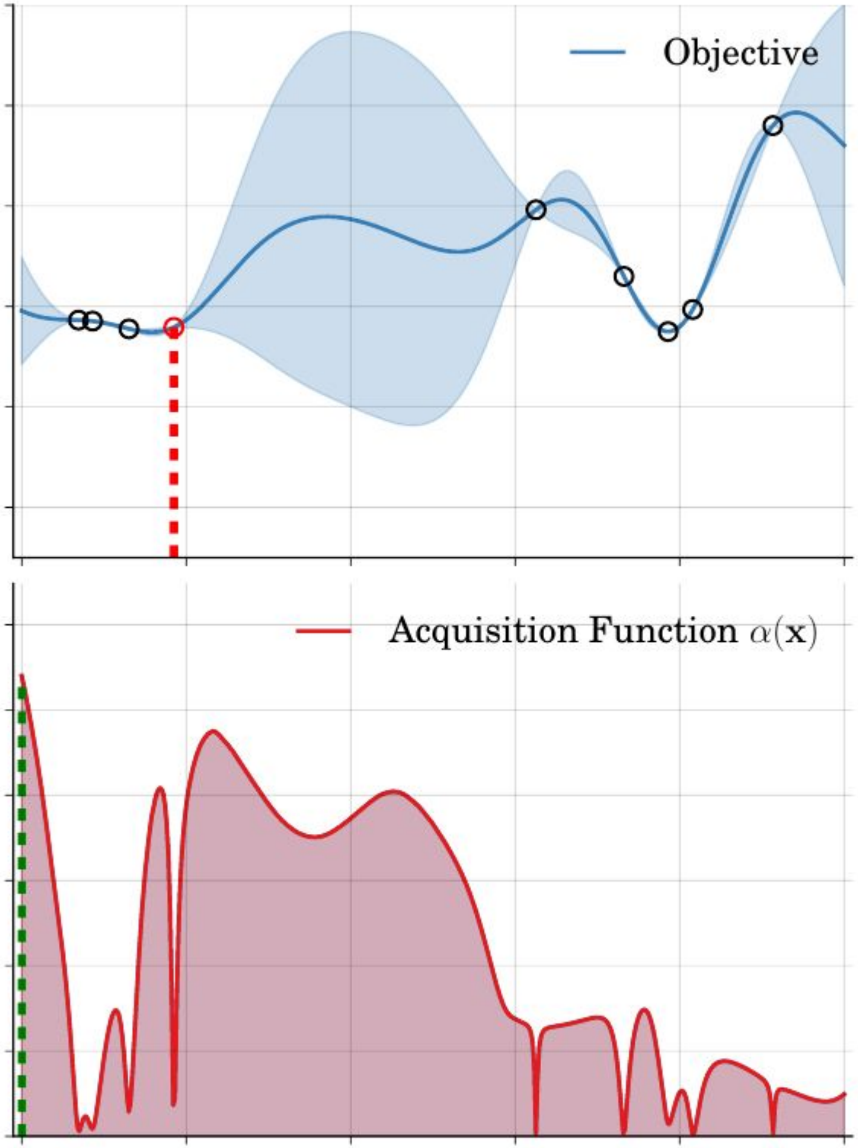
4. **Optimize acquisition function $\alpha(\mathrm{x})$.**
5. **Collect data and update model.**
6. **Repeat!**

Bayesian optimization



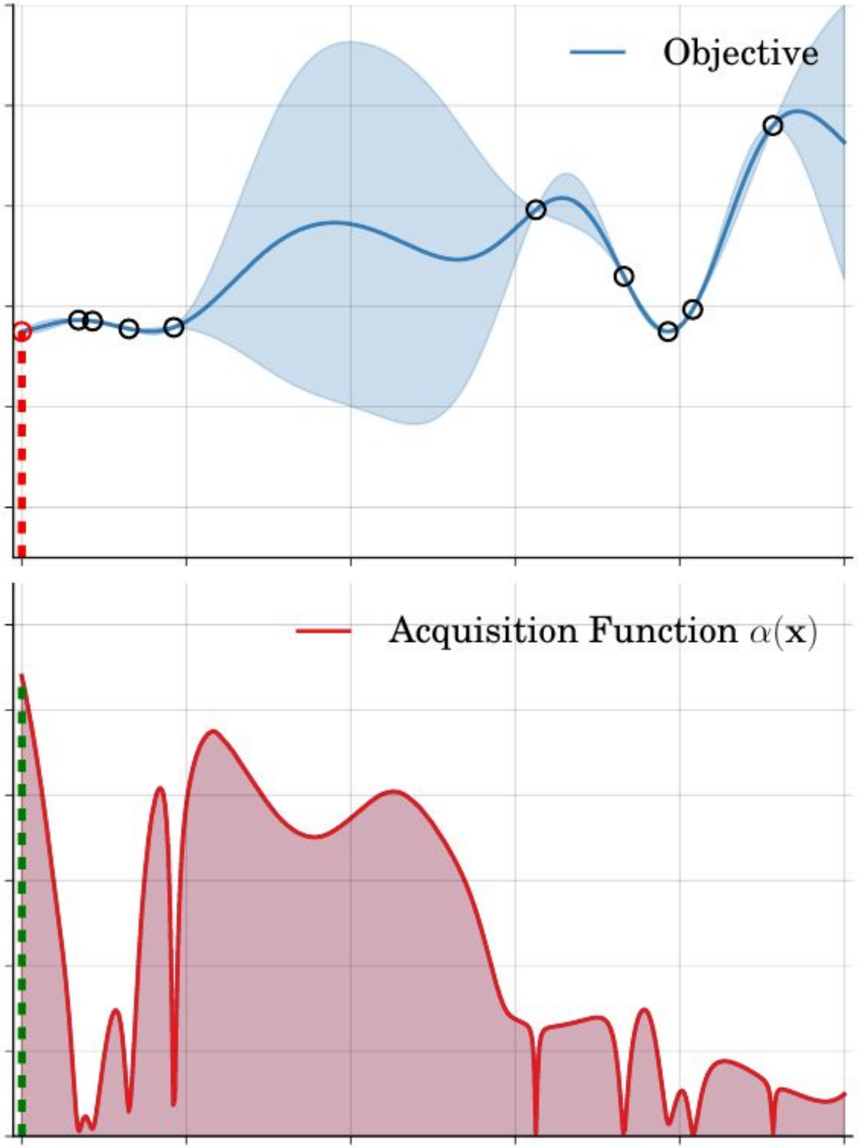
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathrm{x}, \mathcal{D}_n)\,.$$
3. **Select data collection strategy:**
$$\alpha(\mathrm{x}) = \mathrm{E}_{p(y|\mathrm{x},\mathcal{D}_n)}[U(y|\mathrm{x}, \mathcal{D}_n)]\,.$$
4. **Optimize acquisition function $\alpha(\mathrm{x})$.**
5. **Collect data and update model.**
6. **Repeat!**

Bayesian optimization



1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathrm{x}, \mathcal{D}_n)\,.$$
3. **Select data collection strategy:**
$$\alpha(\mathrm{x}) = \mathrm{E}_{p(y|\mathrm{x},\mathcal{D}_n)}[U(y|\mathrm{x}, \mathcal{D}_n)]\,.$$
4. **Optimize acquisition function $\alpha(\mathrm{x})$.**
5. **Collect data and update model.**
6. **Repeat!**

Bayesian optimization



❶ Get initial sample.

❷ Fit a model to the data:

$$p(y|\mathsf{x}, \mathcal{D}_n) \,.$$

❸ Select data collection strategy:

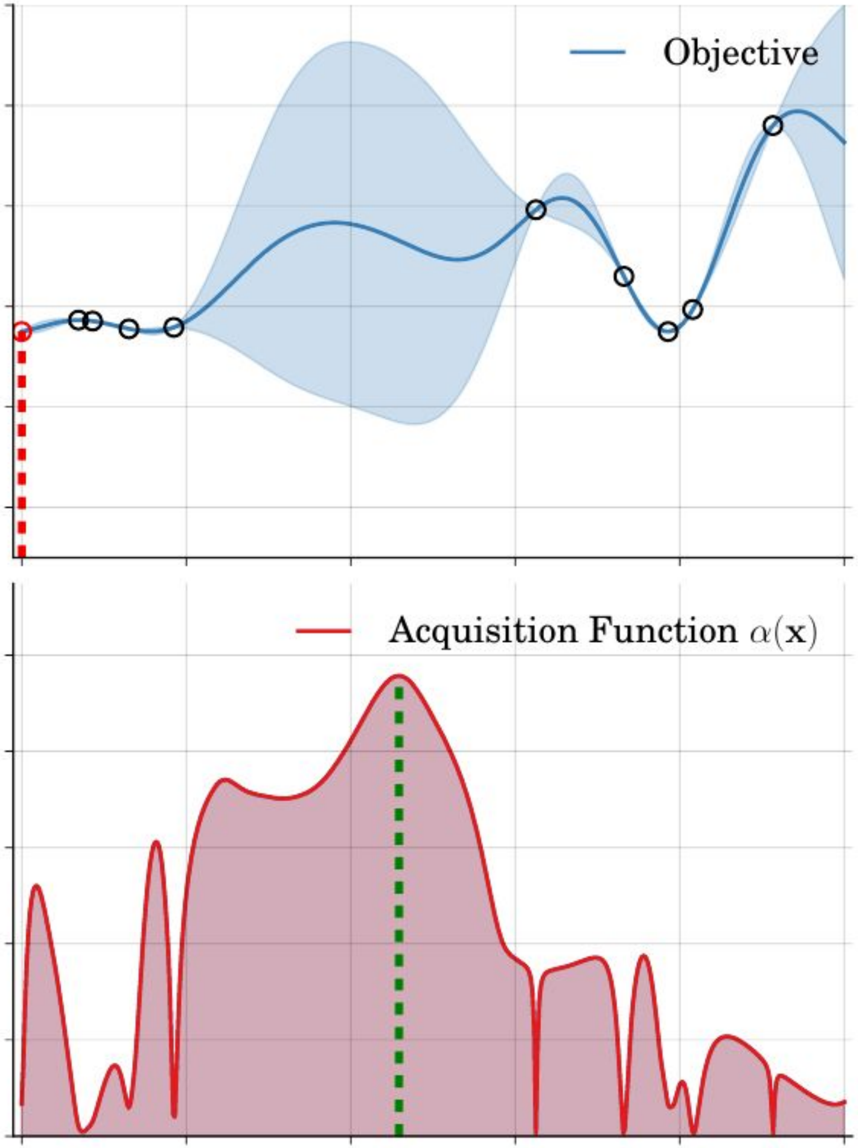
$$\alpha(\mathsf{x}) = \mathsf{E}_{p(y|\mathsf{x},\mathcal{D}_n)}[U(y|\mathsf{x}, \mathcal{D}_n)] \,.$$

❹ Optimize acquisition function $\alpha(\mathsf{x})$.

❺ Collect data and update model.

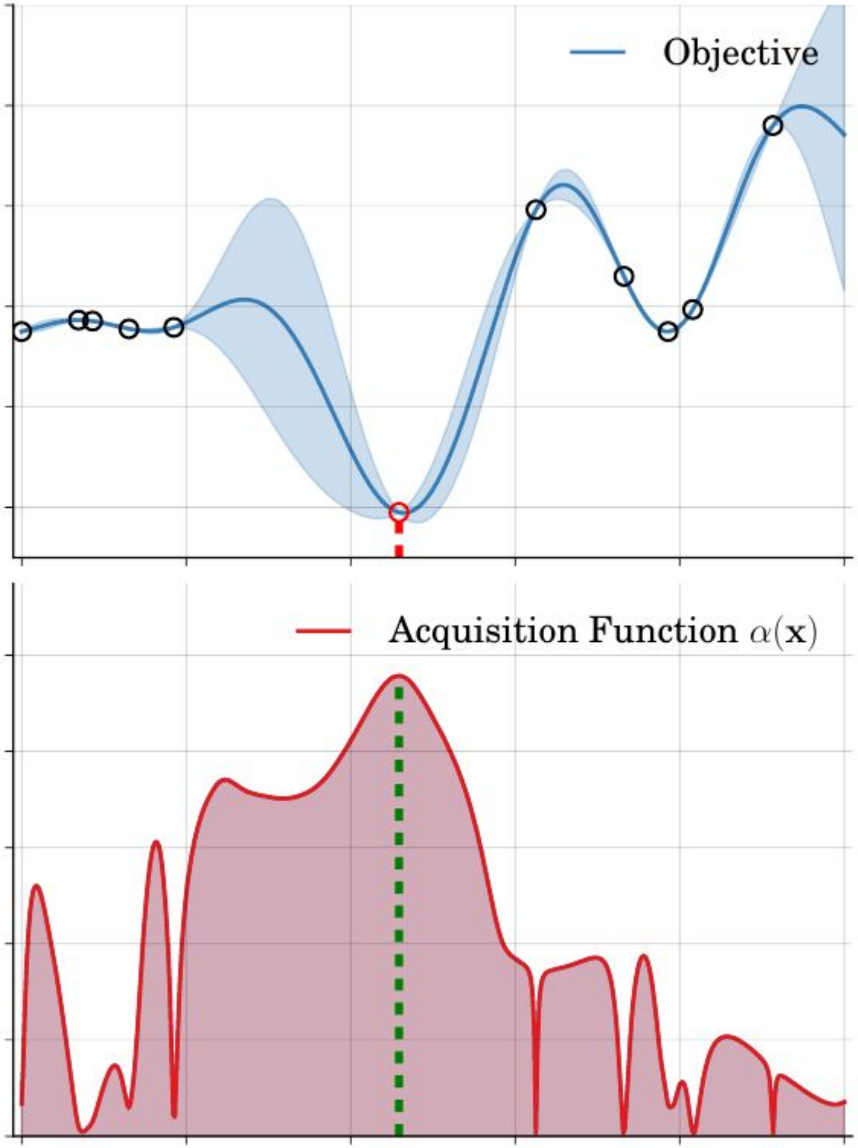
❻ Repeat!

Bayesian optimization



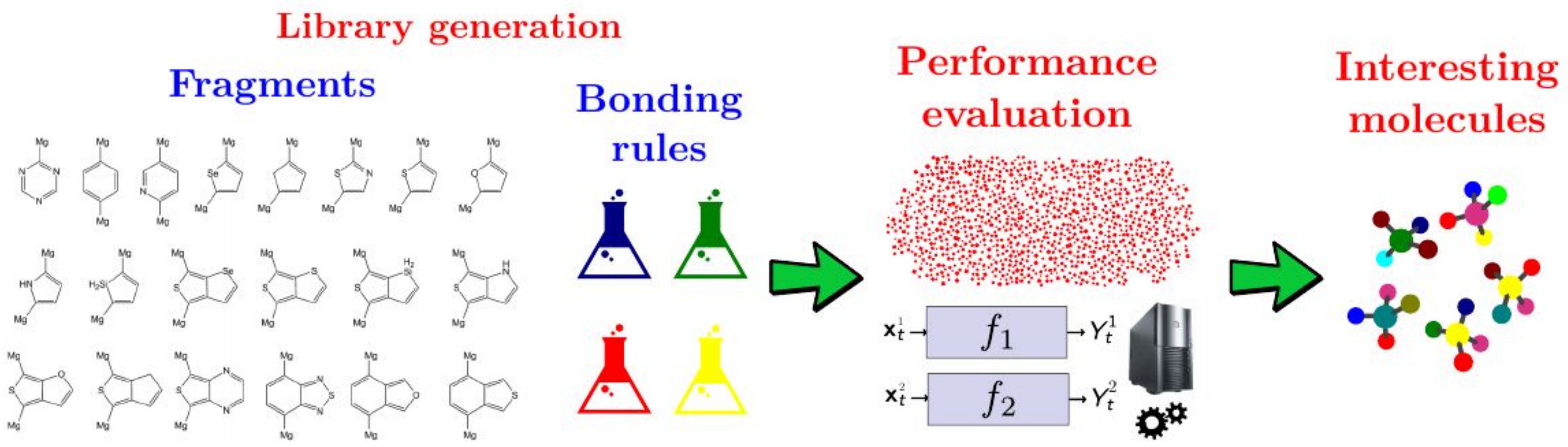
1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathsf{x}, \mathcal{D}_n)\,.$$
3. **Select data collection strategy:**
$$\alpha(\mathsf{x}) = \mathrm{E}_{p(y|\mathsf{x},\mathcal{D}_n)}[U(y|\mathsf{x},\mathcal{D}_n)]\,.$$
4. **Optimize acquisition function $\alpha(\mathsf{x})$.**
5. **Collect data and update model.**
6. **Repeat!**

Bayesian optimization

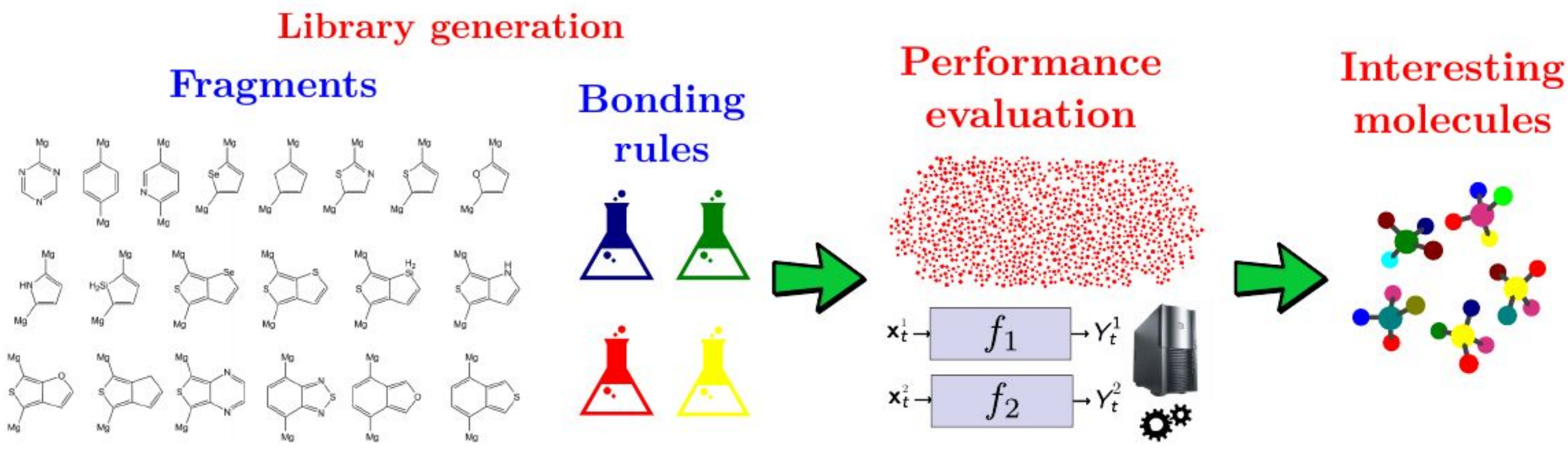


1. Get initial sample.
2. Fit a model to the data:
$$p(y|\mathrm{x}, \mathcal{D}_n)\,.$$
3. **Select data collection strategy:**
$$\alpha(\mathrm{x}) = \mathrm{E}_{p(y|\mathrm{x},\mathcal{D}_n)}[U(y|\mathrm{x}, \mathcal{D}_n)]\,.$$
4. **Optimize acquisition function $\alpha(\mathrm{x})$.**
5. **Collect data and update model.**
6. **Repeat!**

Discovering new optimal molecules

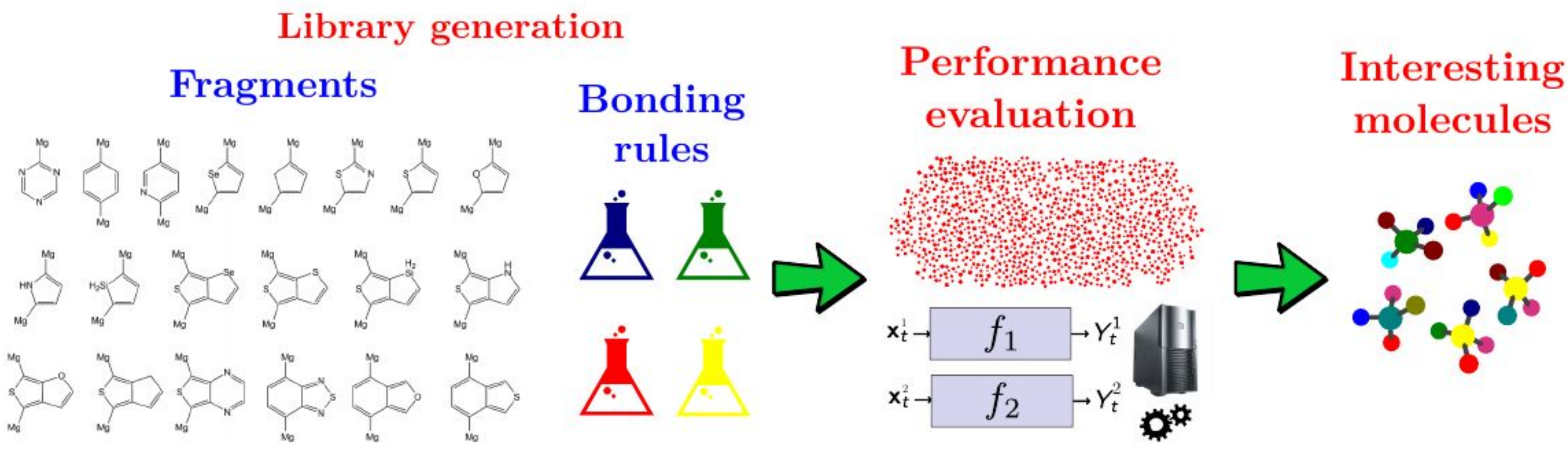


Discovering new optimal molecules



Bayesian optimization (BO) can **accelerate** the search!

Discovering new optimal molecules

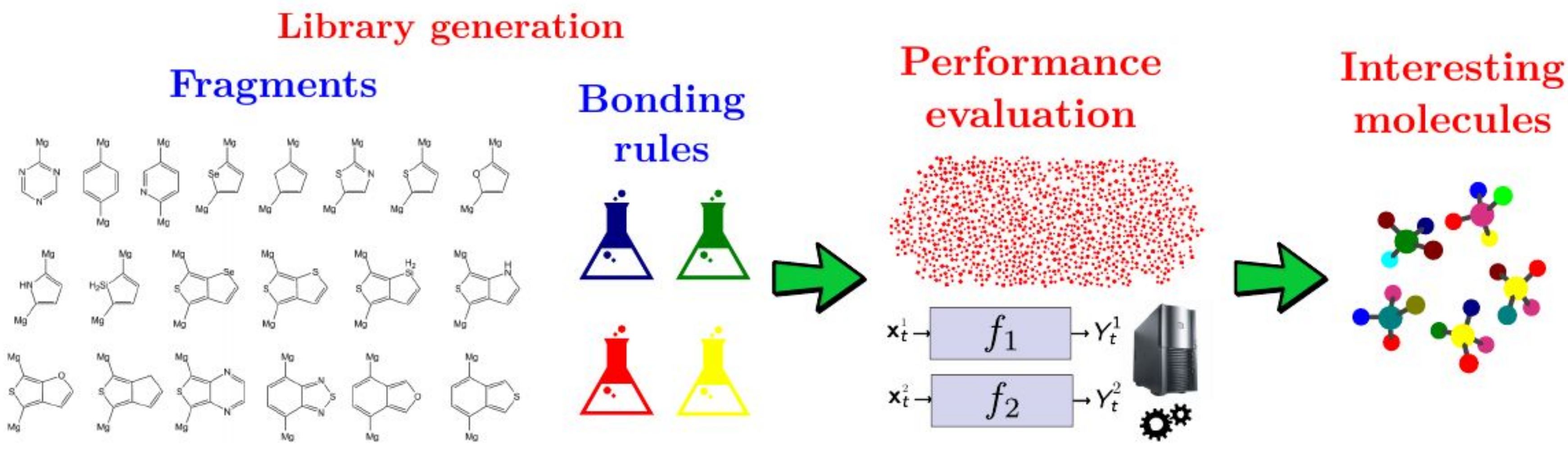


Bayesian optimization (BO) can **accelerate** the search!

Challenges:

- The search space is discrete and structured.
- BO output should exhibit regularities found in real-world molecules.

Discovering new optimal molecules



Bayesian optimization (BO) can **accelerate** the search!

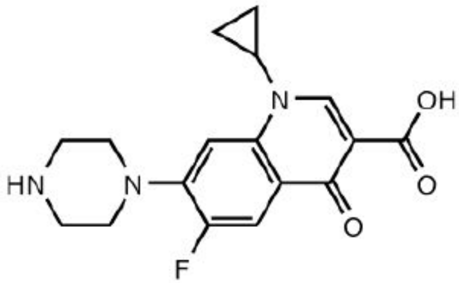
Challenges:

- The search space is discrete and structured.
- BO output should exhibit regularities found in real-world molecules.

Solution: combine BO methods with **generative models** of molecules.

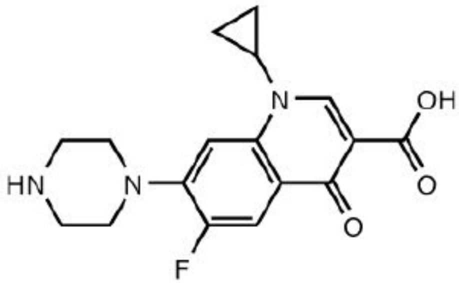
Gómez-Bombarelli⋆, Wei⋆, Duvenaud⋆, Hernández-Lobato⋆, Sánchez-Lengeling, Sheberla, Aguilera-Iparraguirre, Hirzel, Adams and Aspuru-Guzik, 2018. (⋆ equal contributors).

Molecules encoded as **character strings** using SMILES language.



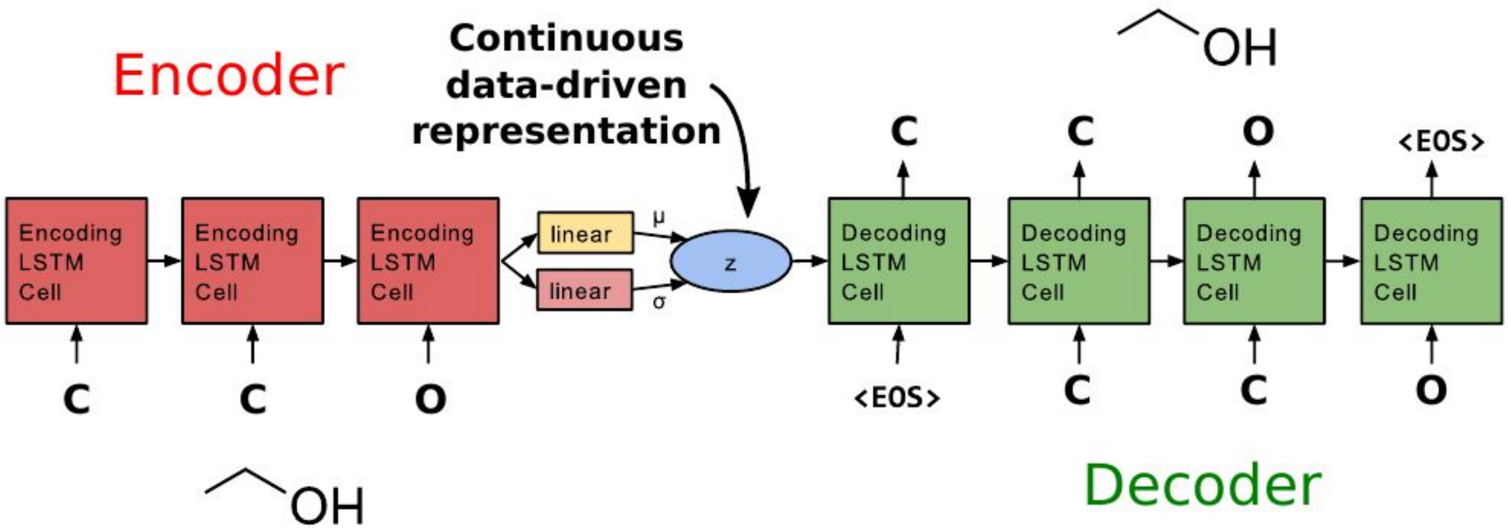
→ N1CCN(CC1)C(C(F)=C2)=CC(=C2C4=O)N(C3CC3)C=C4C(=O)O

Molecules encoded as **character strings** using SMILES language.



N1CCN(CC1)C(C(F)=C2)=CC(=C2C4=O)N(C3CC3)C=C4C(=O)O

A **generative model** on SMILES is obtained using a **seq2seq** **VAE**.



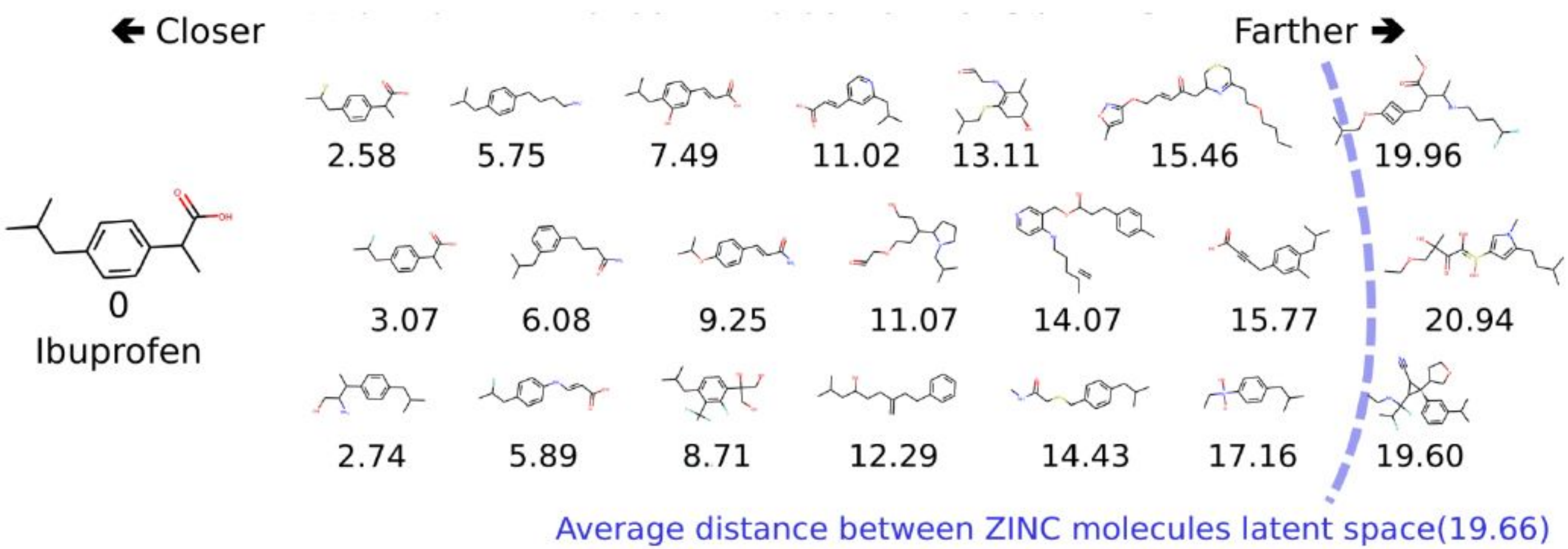
Sampled molecules have statistics similar to those of real molecules:

source	data set	logP	SAS	QED
Data	ZINC	2.46 (1.43)	3.05 (0.83)	0.73 (0.14)
GA	ZINC	2.84 (1.86)	3.80 (1.01)	0.57 (0.20)
VAE	ZINC	2.67 (1.46)	3.18 (0.86)	0.70 (0.14)
Data	QM9	0.30 (1.00)	4.25 (0.94)	0.48 (0.07)
GA	QM9	0.96 (1.53)	4.47 (1.01)	0.53 (0.13)
VAE	QM9	0.30 (0.97)	4.34 (0.98)	0.47 (0.08)

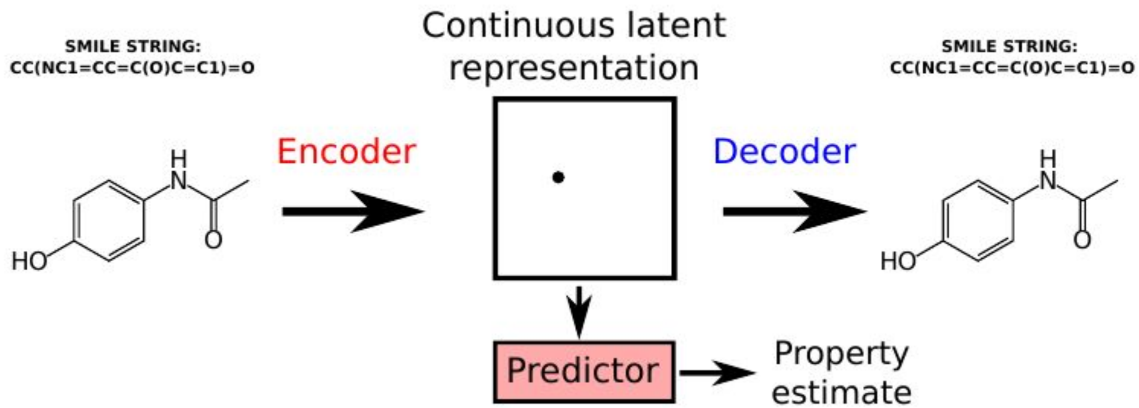
Sampled molecules have statistics similar to those of real molecules:

source	data set	logP	SAS	QED
Data	ZINC	2.46 (1.43)	3.05 (0.83)	0.73 (0.14)
GA	ZINC	2.84 (1.86)	3.80 (1.01)	0.57 (0.20)
VAE	ZINC	2.67 (1.46)	3.18 (0.86)	0.70 (0.14)
Data	QM9	0.30 (1.00)	4.25 (0.94)	0.48 (0.07)
GA	QM9	0.96 (1.53)	4.47 (1.01)	0.53 (0.13)
VAE	QM9	0.30 (0.97)	4.34 (0.98)	0.47 (0.08)

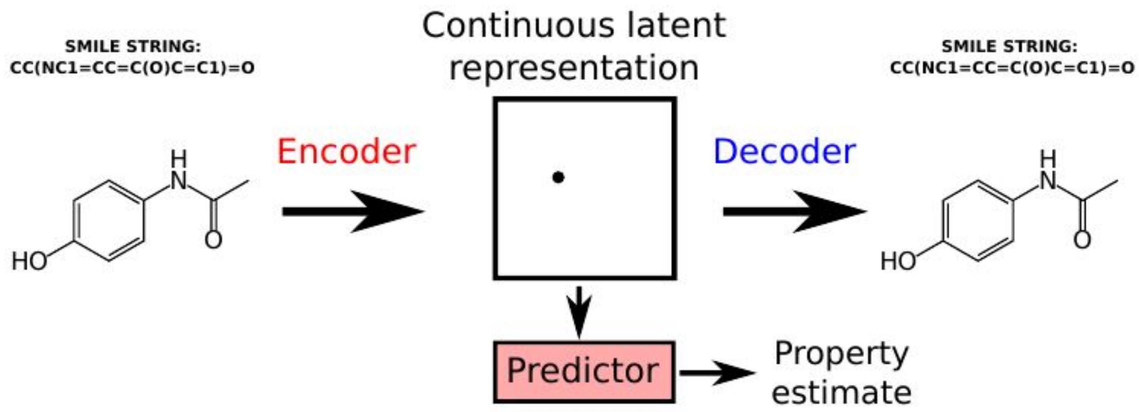
Nearby latent representations decode into similar molecules:



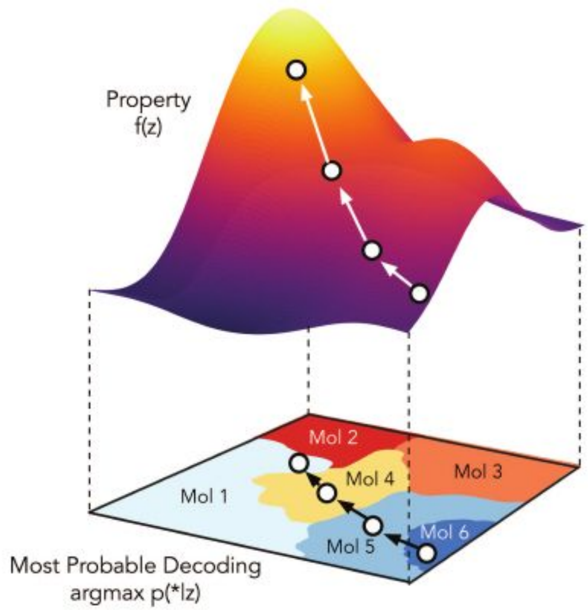
Easy to add a **surrogate model** from latent space to property.



Easy to add a **surrogate model** from latent space to property.

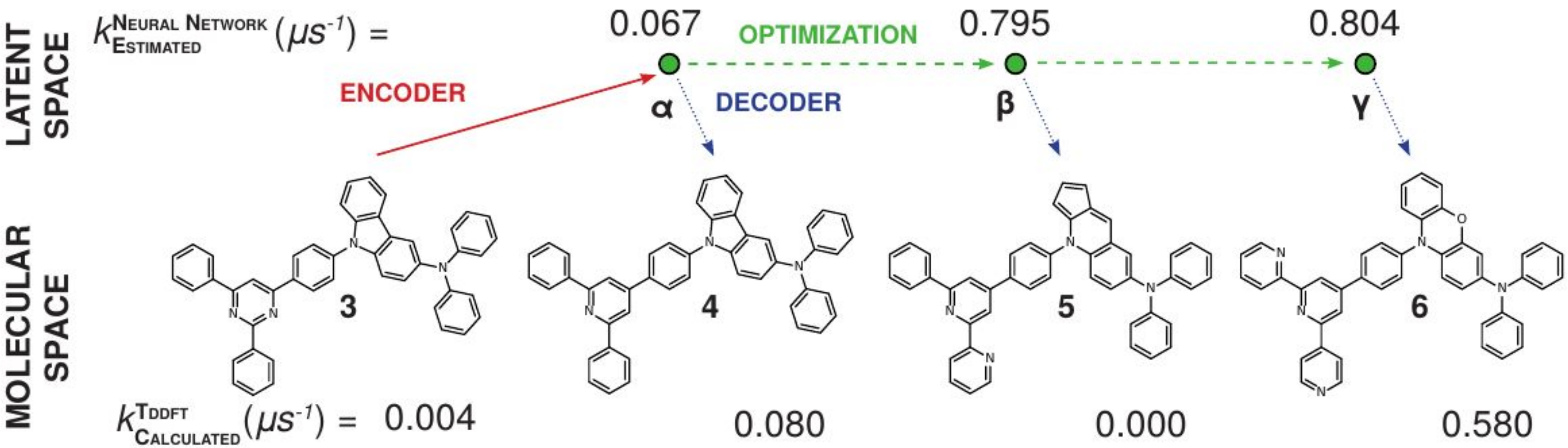


Gradient-based optimization can be used in latent space.



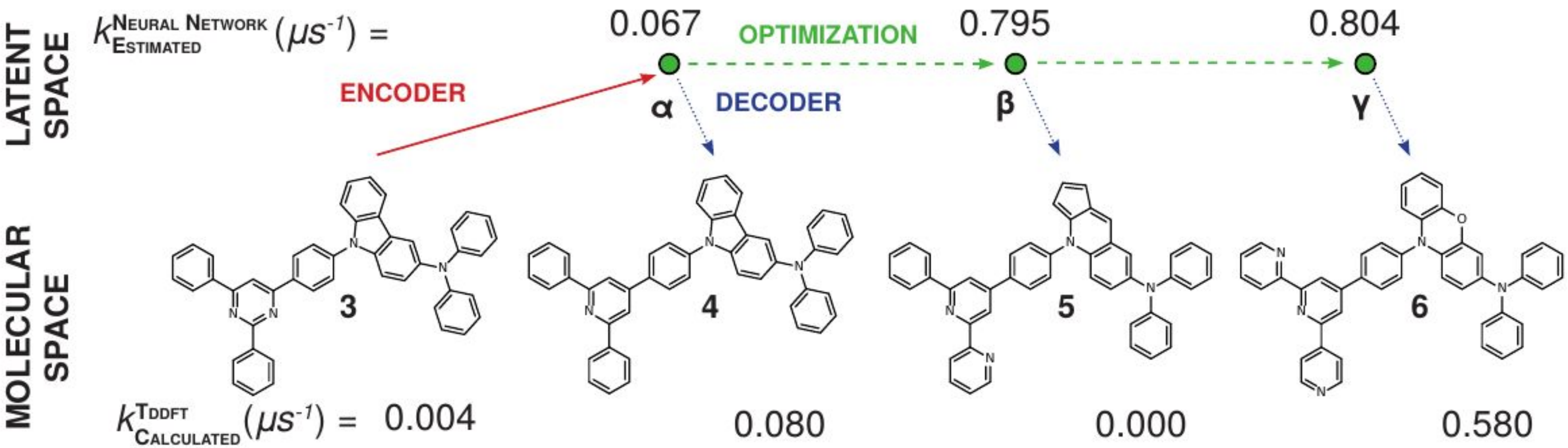
Local optimization of OLED molecules

We optimize the delayed fluorescence decay rate κ_{TADF}, as estimated from TDDFT computations on 150,000 molecules.



Local optimization of OLED molecules

We optimize the delayed fluorescence decay rate κ_{TADF}, as estimated from TDDFT computations on 150,000 molecules.

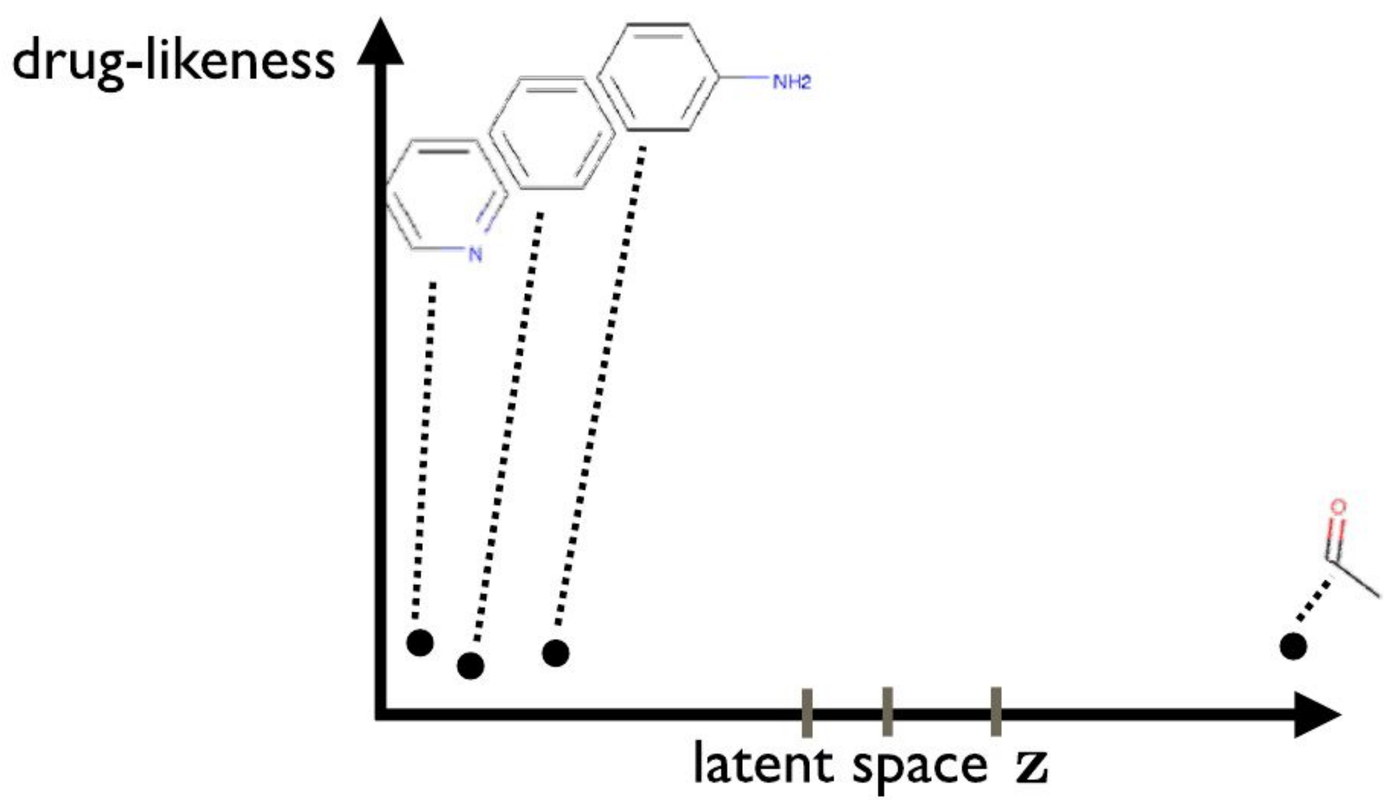


However, many of the sampled SMILES strings are not valid molecules.

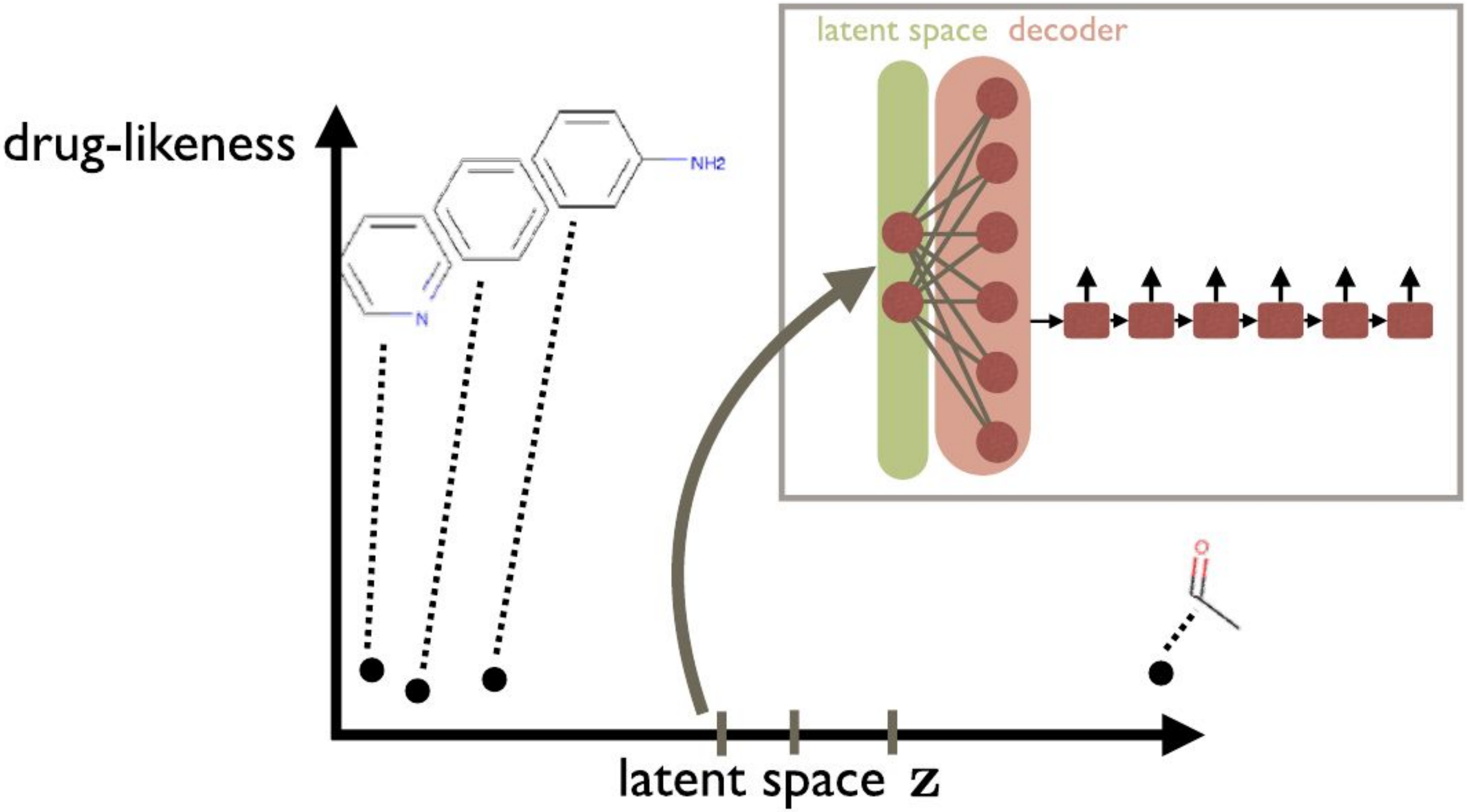
Grammar Variational Autoencoder

Kusner M. J., Paige B. and Hernández-Lobato J. M.
Grammar Variational Autoencoder,
In ICML, 2017.

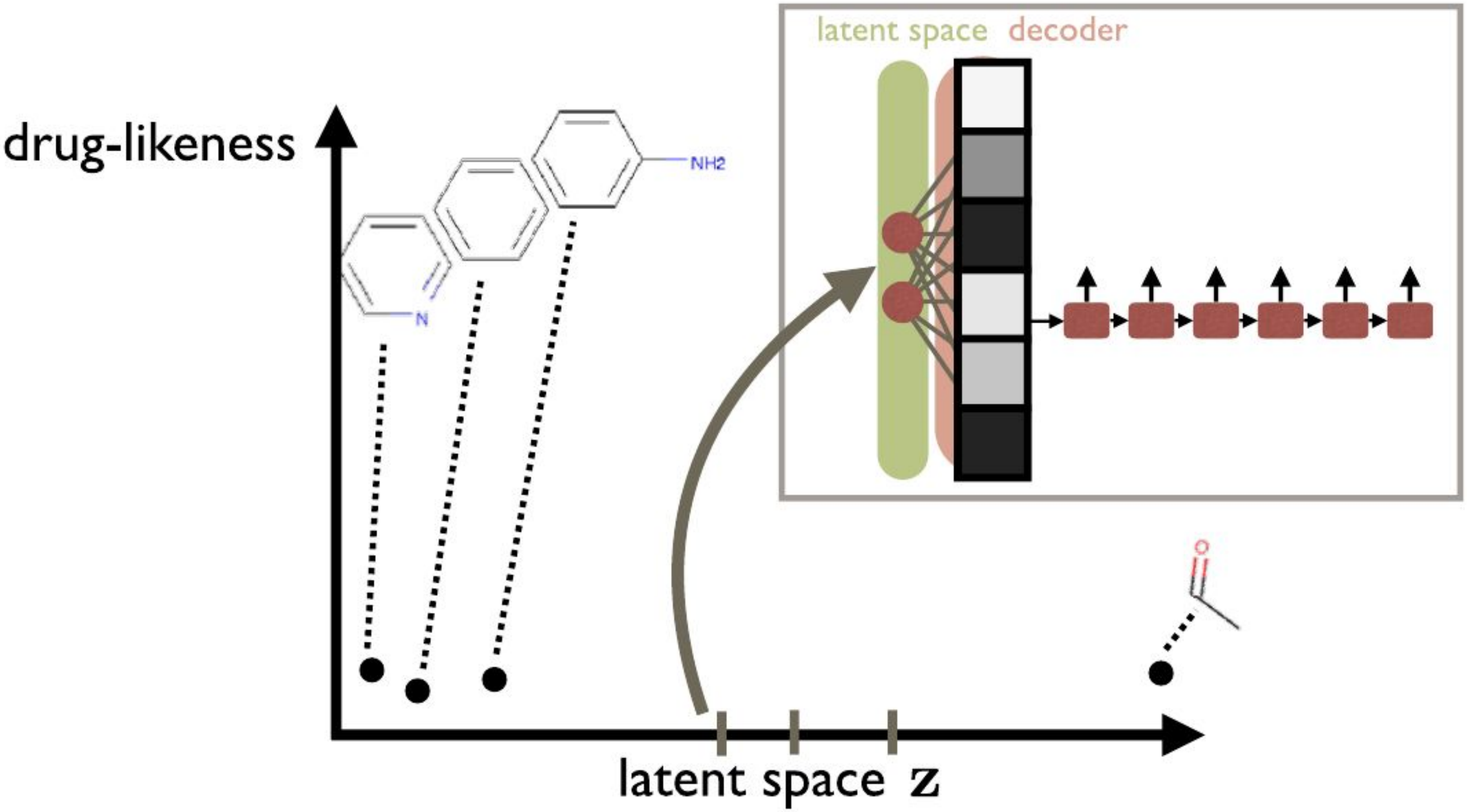
THE PROBLEM



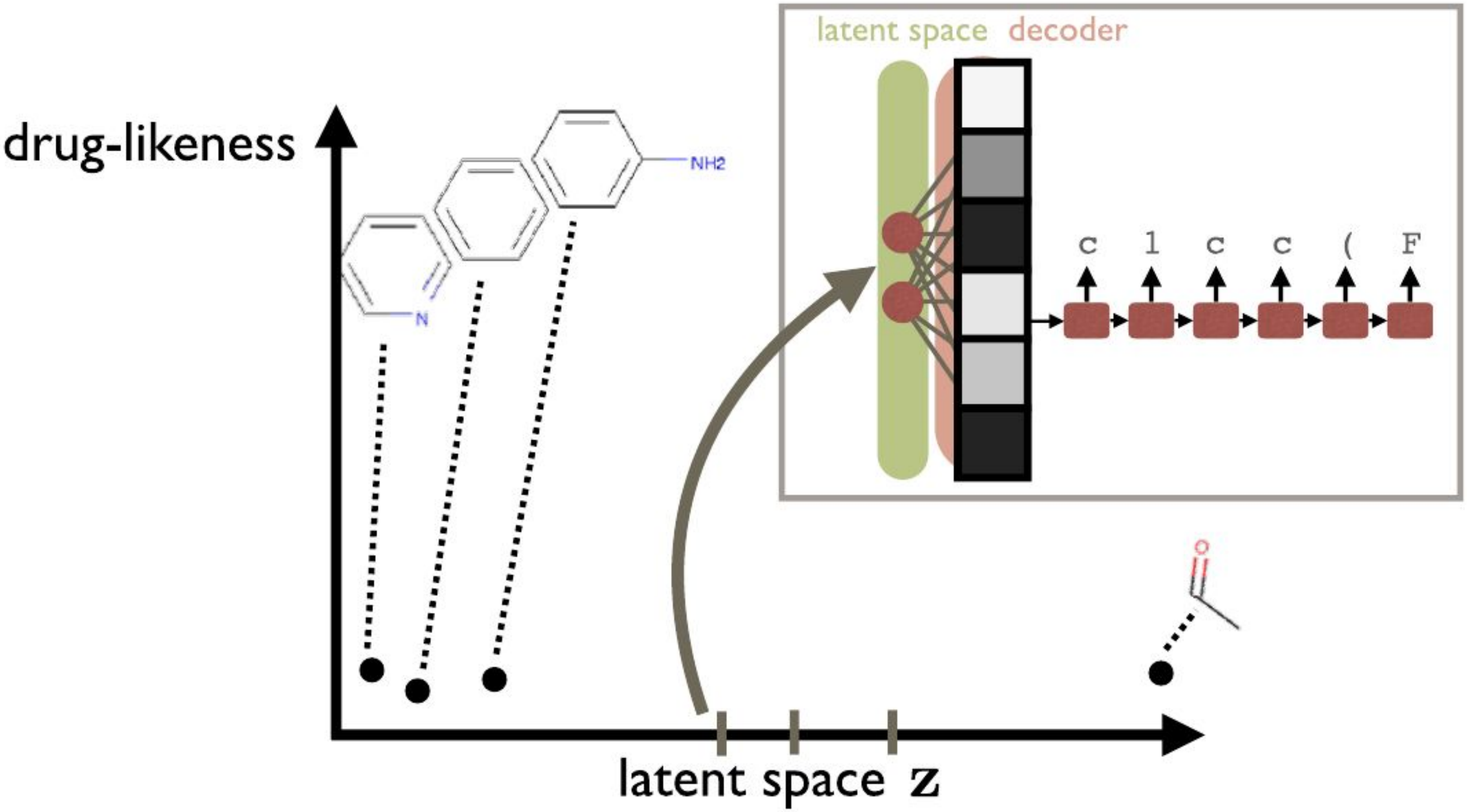
THE PROBLEM



THE PROBLEM

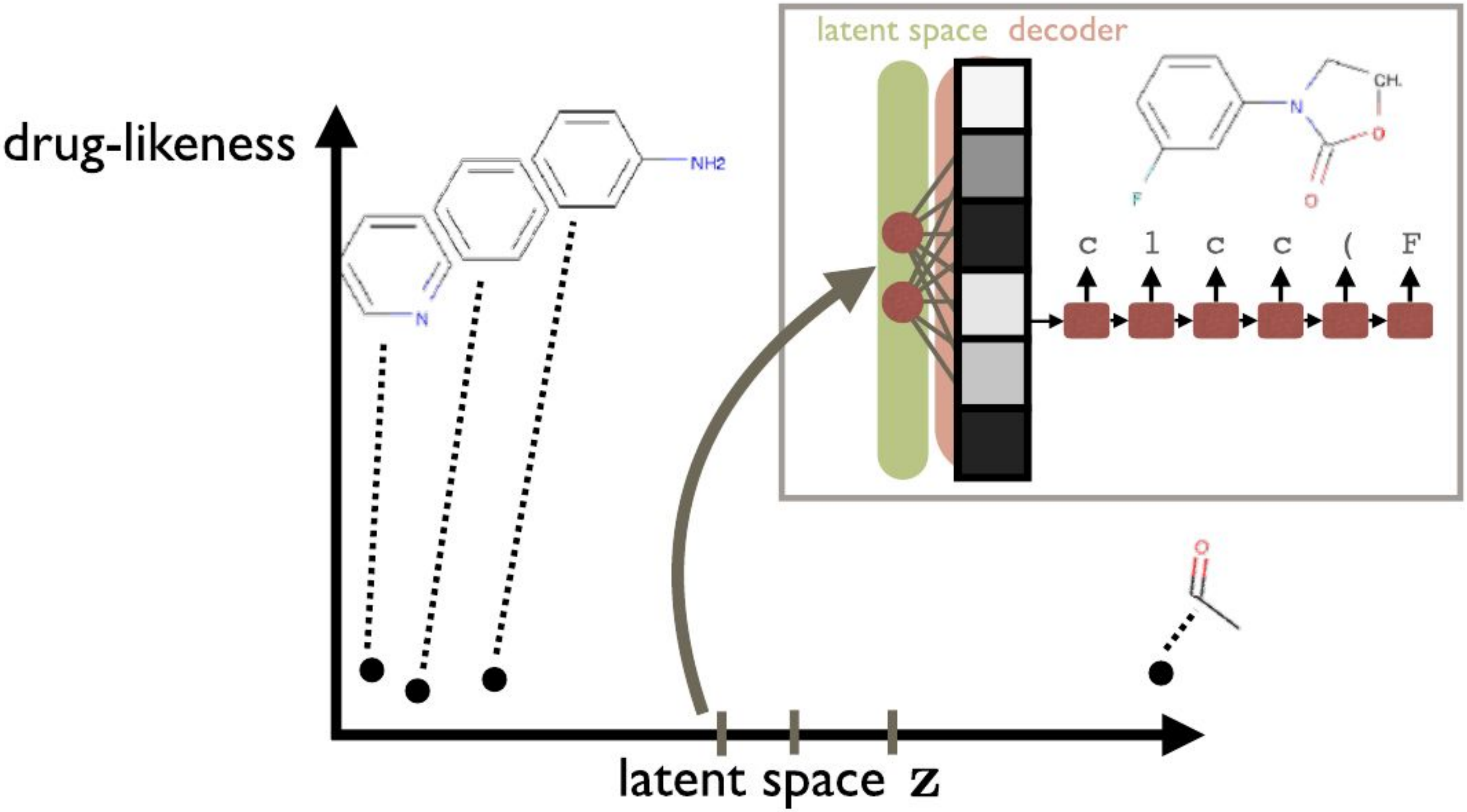


THE PROBLEM



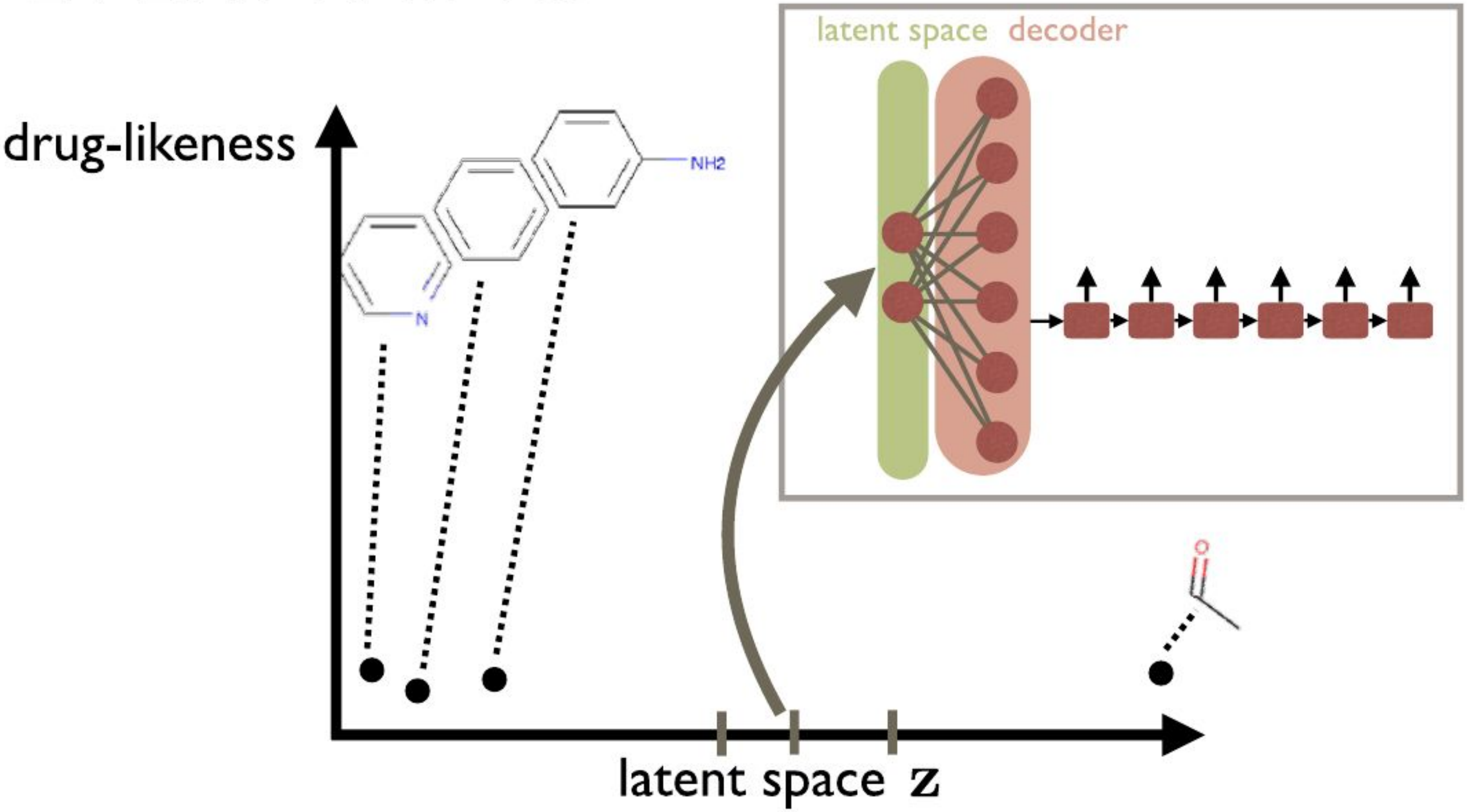
drug-likeness
NH2
N
O
latent space
decoder
c
1
c
c
(
F
latent space z

THE PROBLEM

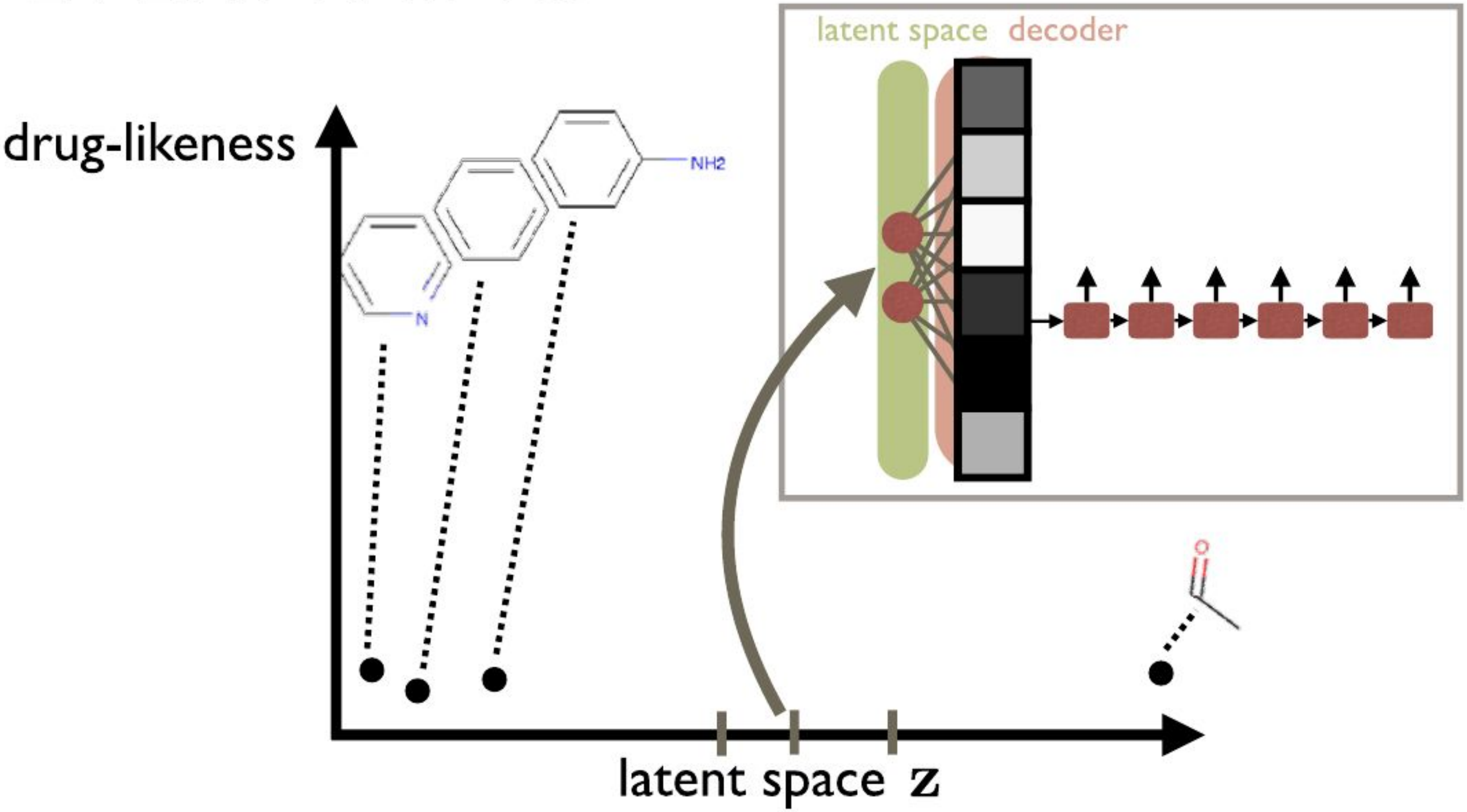


latent space
decoder
drug-likeness
NH2
N
CH
N
O
O
F
c
1
c
c
(
F
latent space z

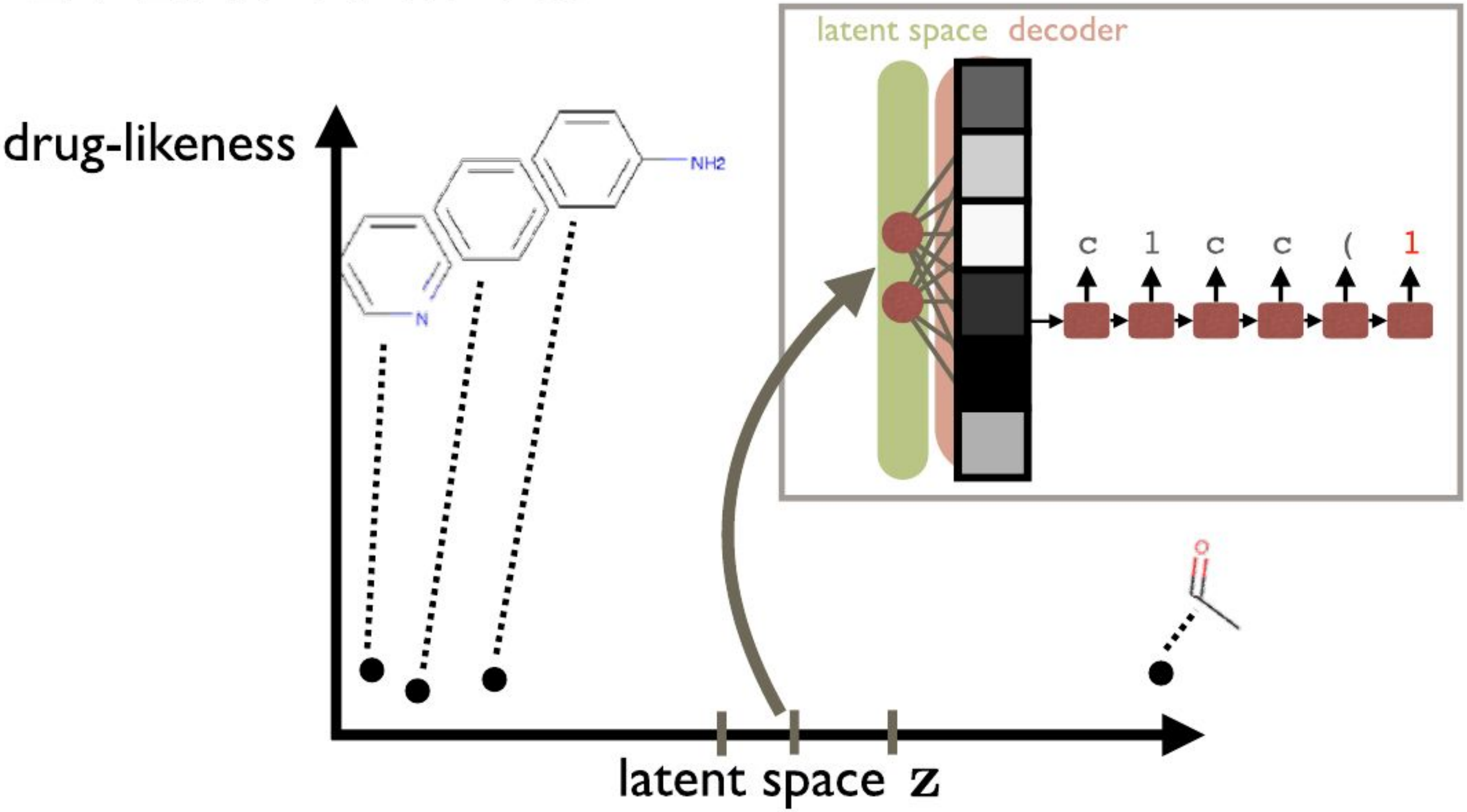
SMALL CHANGES RUIN EVERYTHING



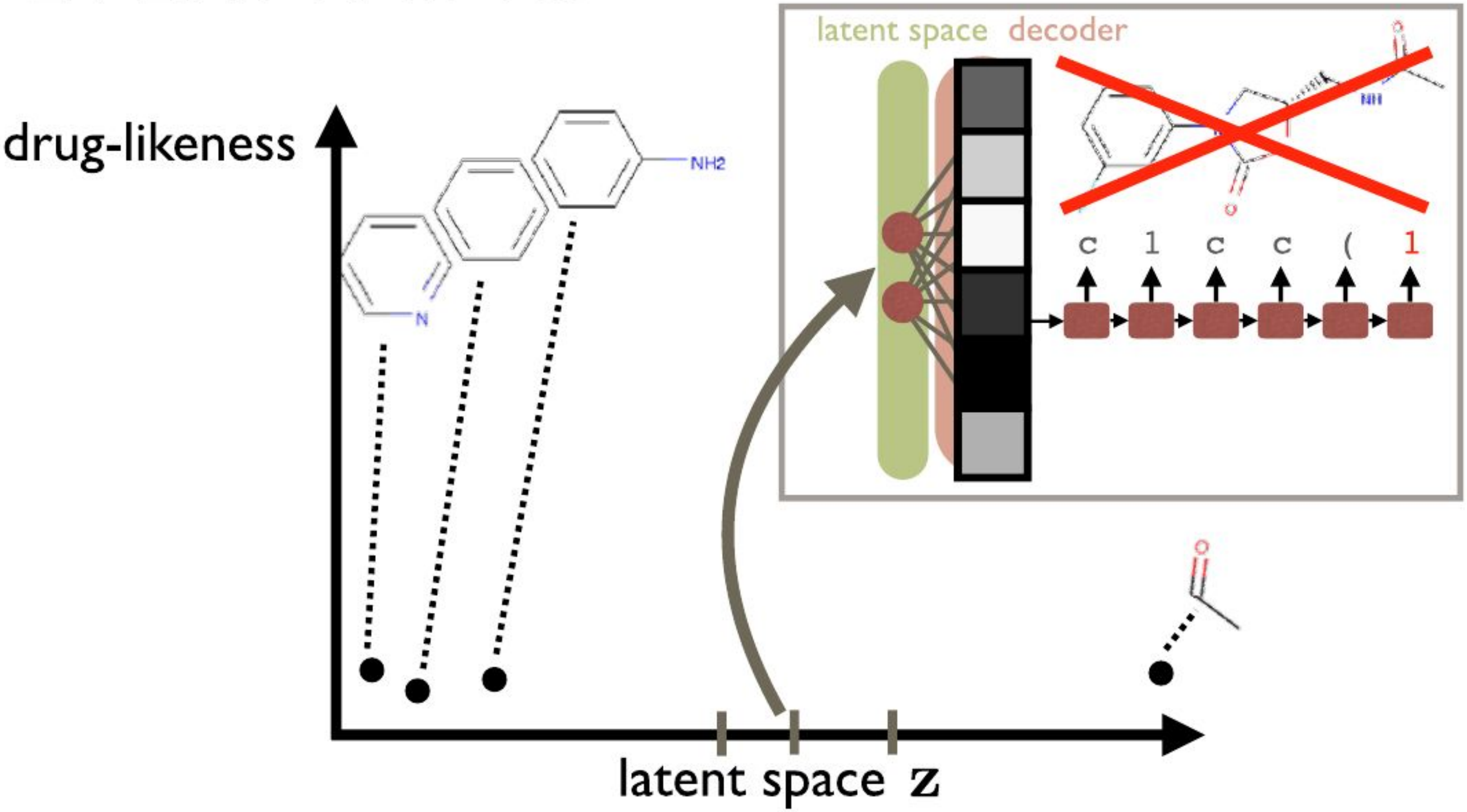
SMALL CHANGES RUIN EVERYTHING



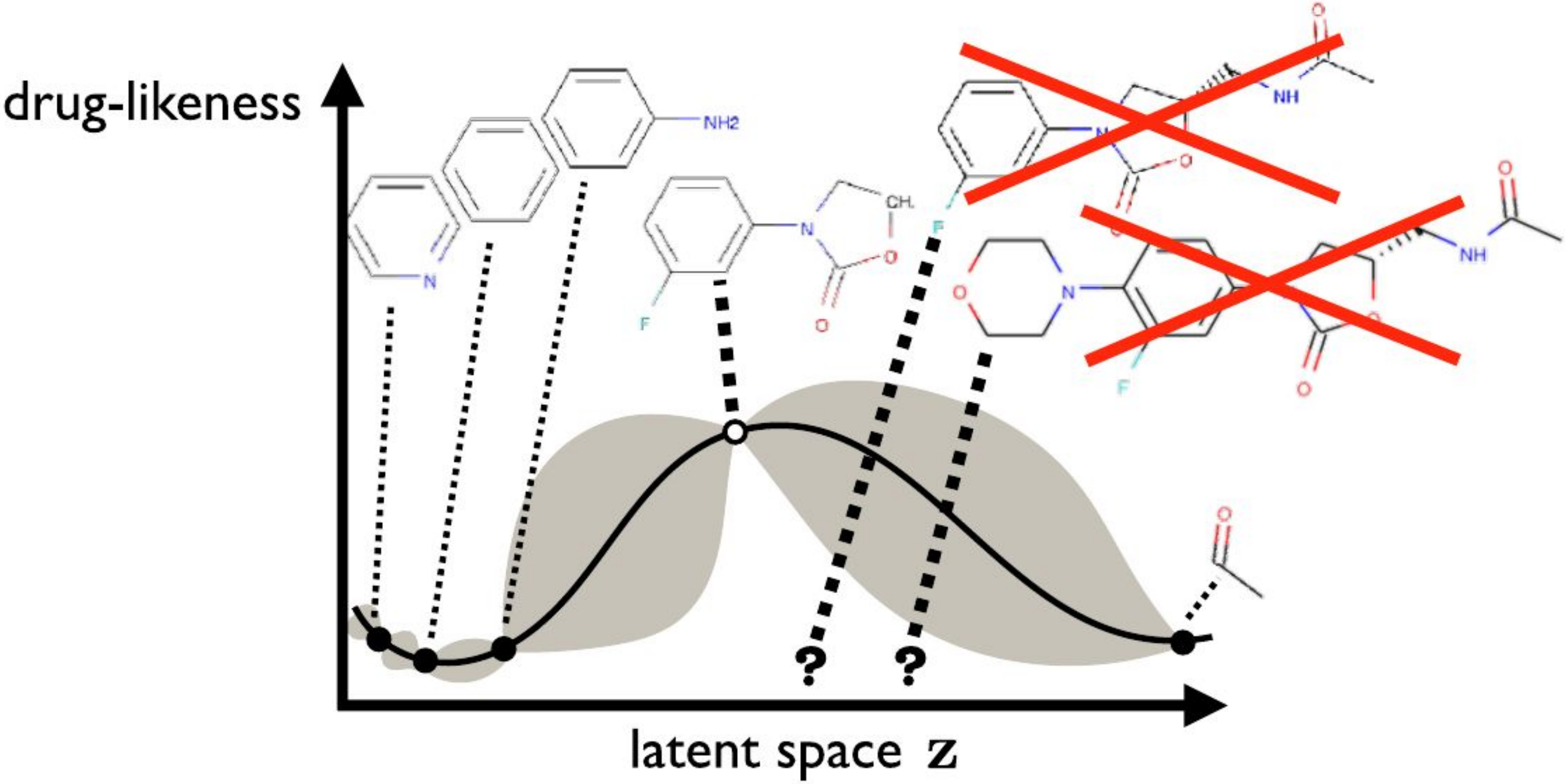
SMALL CHANGES RUIN EVERYTHING



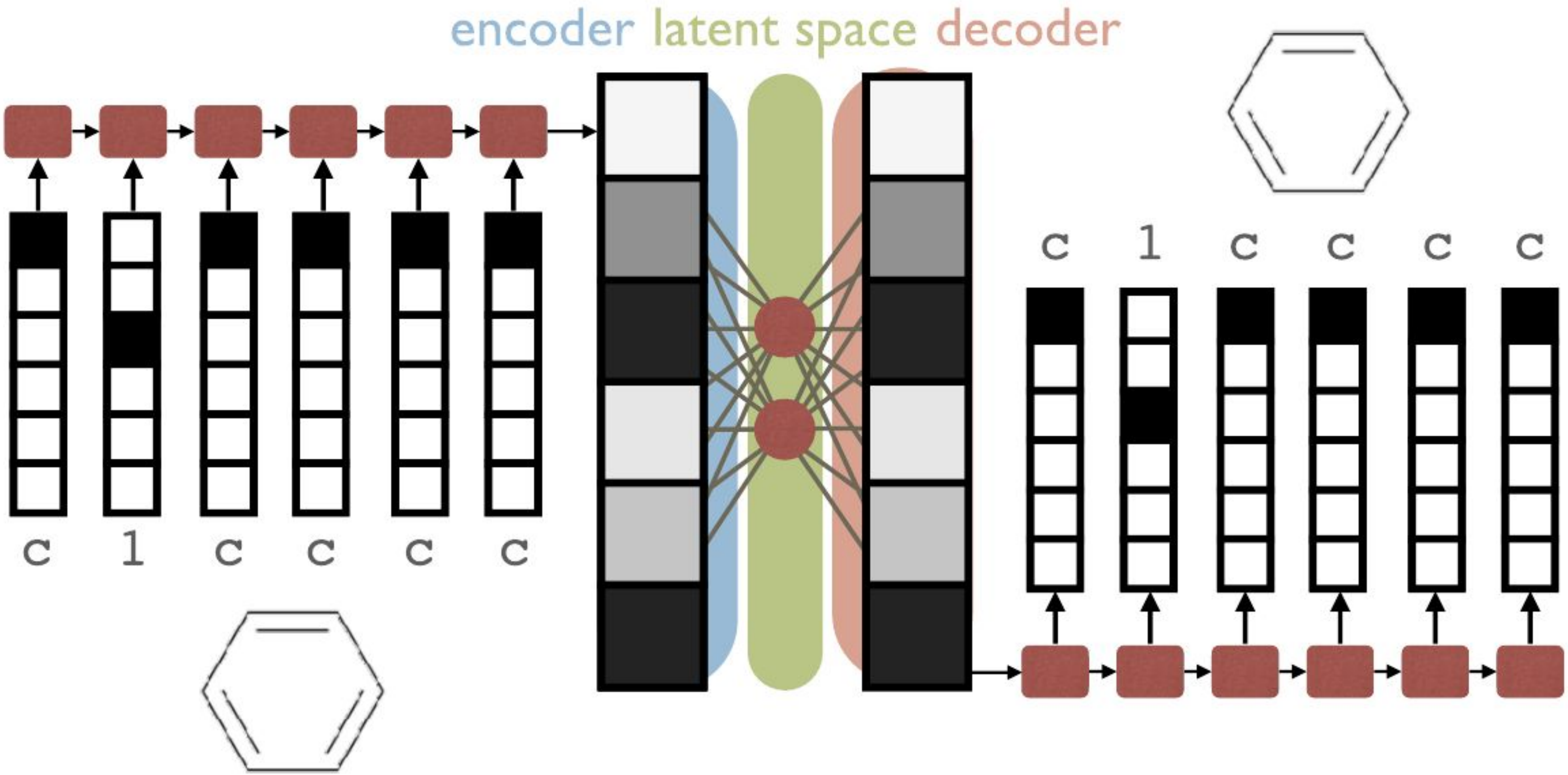
SMALL CHANGES RUIN EVERYTHING



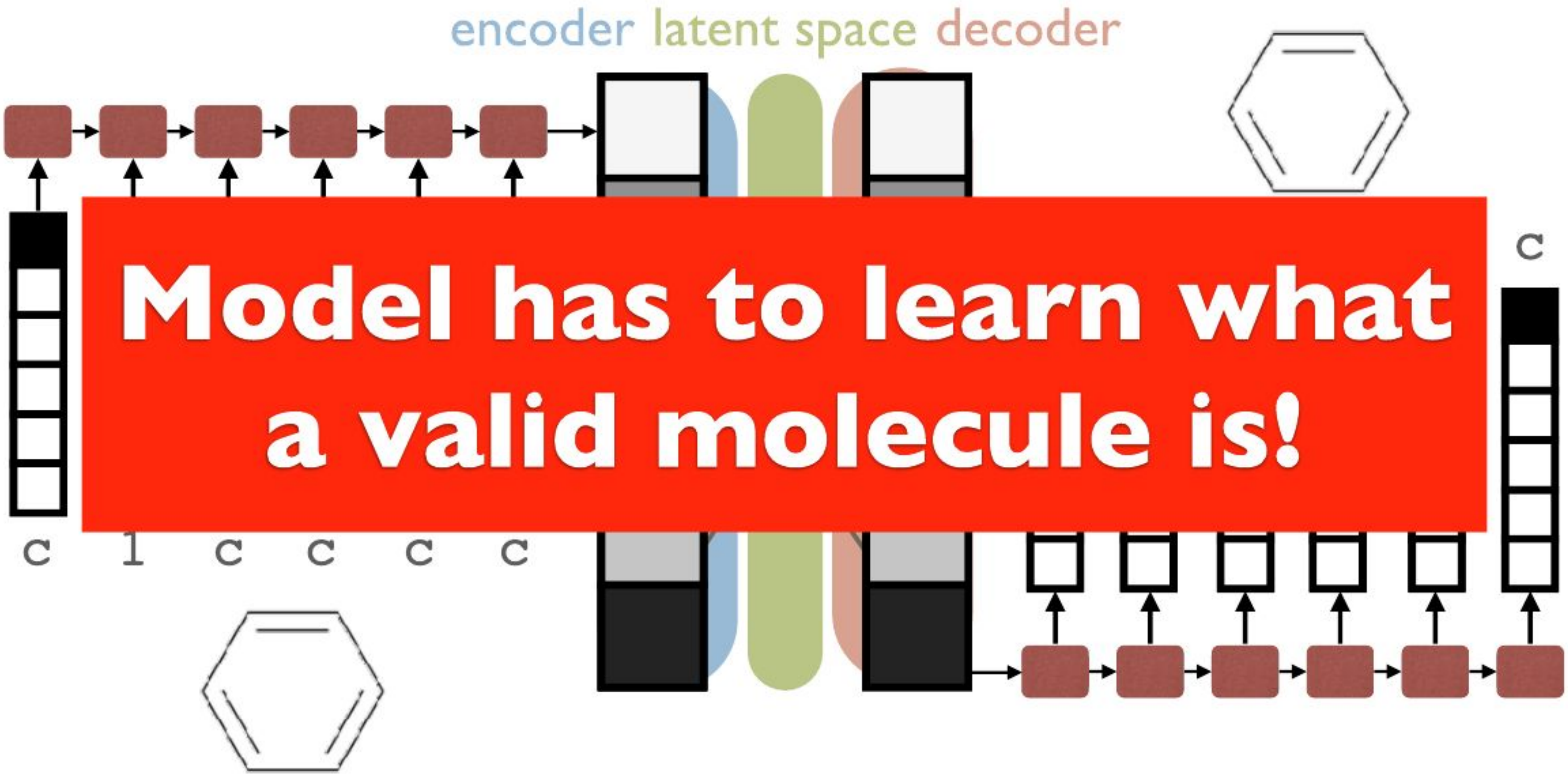
SMALL CHANGES RUIN EVERYTHING



DISCRETE GENERATIVE MODELS: CAN WE DO BETTER?



DISCRETE GENERATIVE MODELS: CAN WE DO BETTER?



Model has to learn what a valid molecule is!

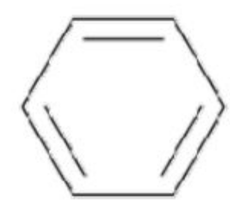
A NEW REPRESENTATION

[Kusner, Paige, Hernández-Lobato, 2017]

Context-Free Grammar

[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'



c1ccccc1

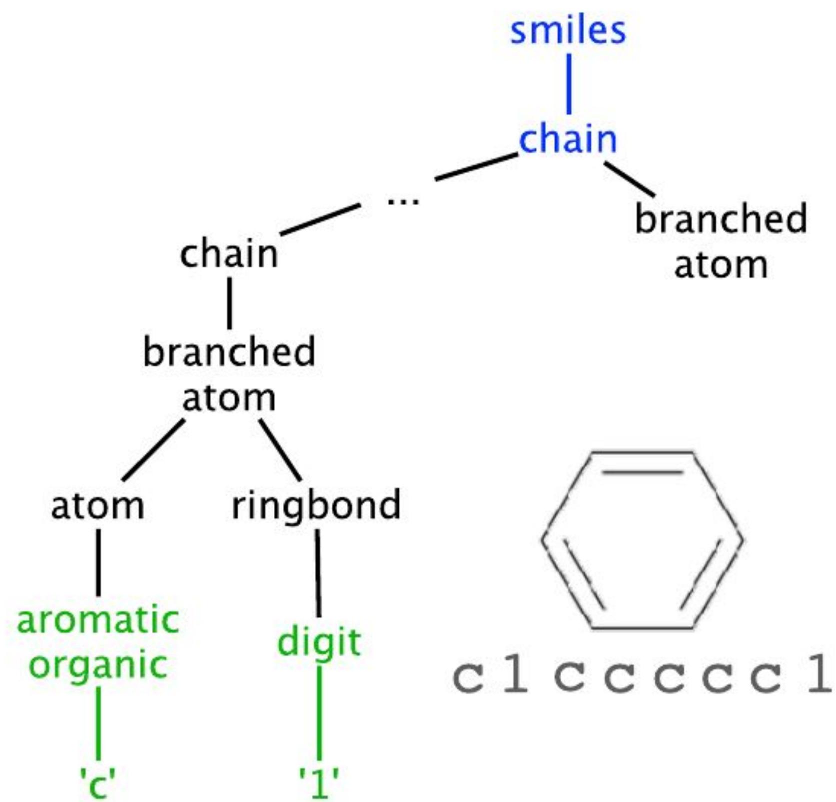
A NEW REPRESENTATION

[Kusner, Paige, Hernández-Lobato, 2017]

Context-Free Grammar

[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'



c 1 c c c c c 1

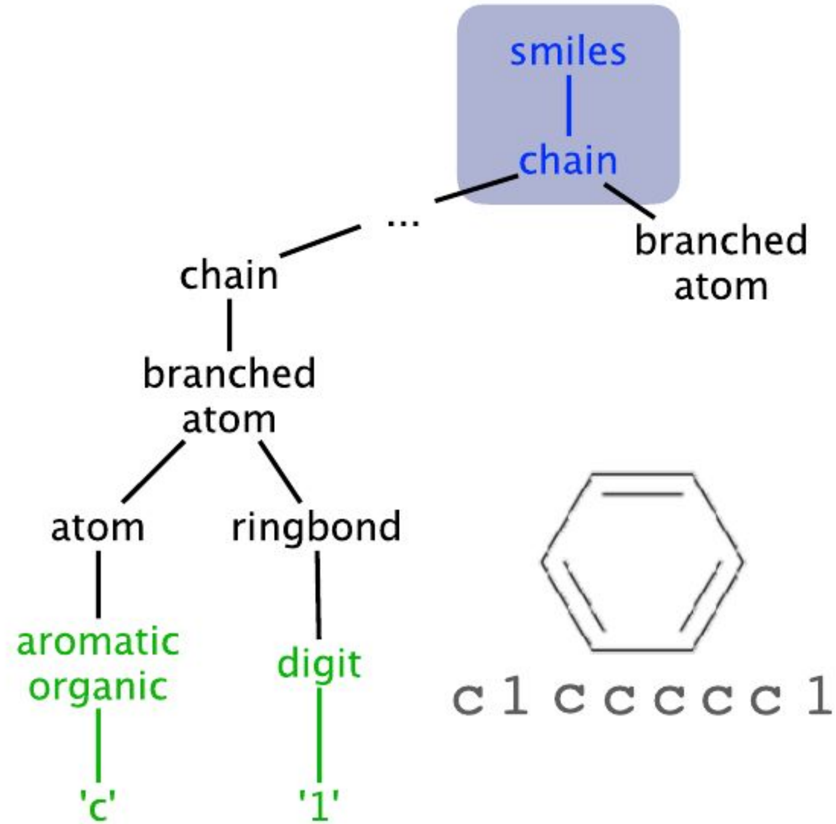
A NEW SEQUENCE!

[Kusner, Paige, Hernández-Lobato, 2017]

Context-Free Grammar

[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'



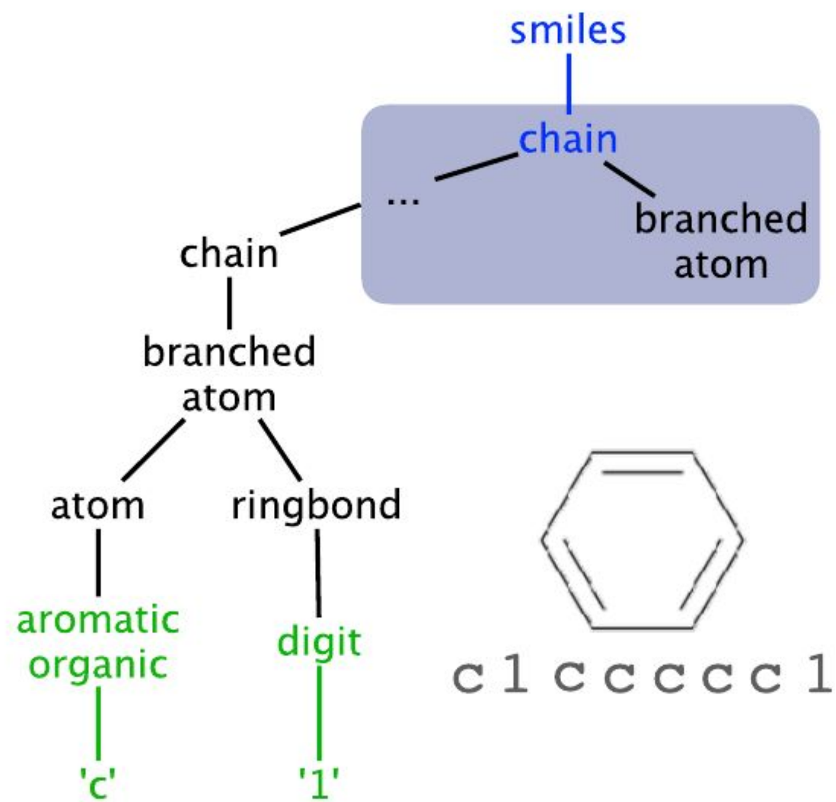
A NEW SEQUENCE!

[Kusner, Paige, Hernández-Lobato, 2017]

Context-Free Grammar

[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'



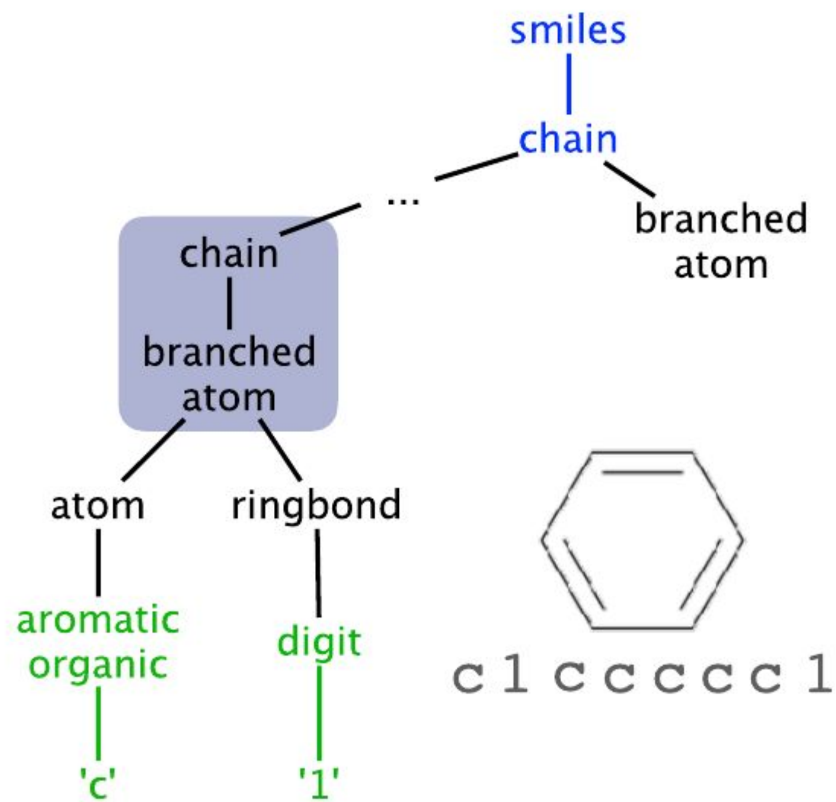
A NEW SEQUENCE!

[Kusner, Paige, Hernández-Lobato, 2017]

Context-Free Grammar

[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'



c 1 c c c c c 1

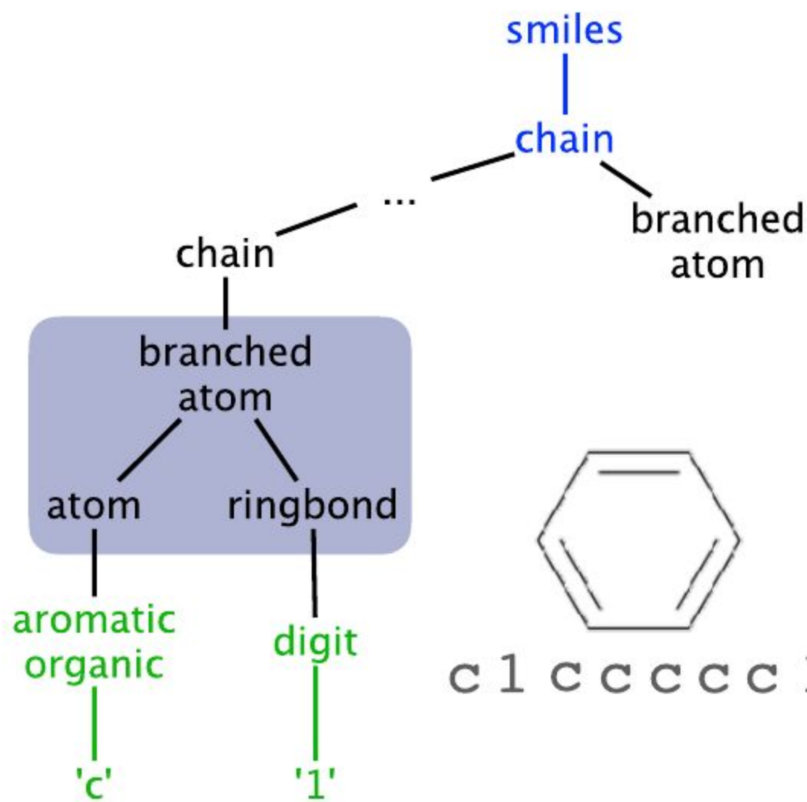
A NEW SEQUENCE!

[Kusner, Paige, Hernández-Lobato, 2017]

Context-Free Grammar

[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'



c 1 c c c c c 1

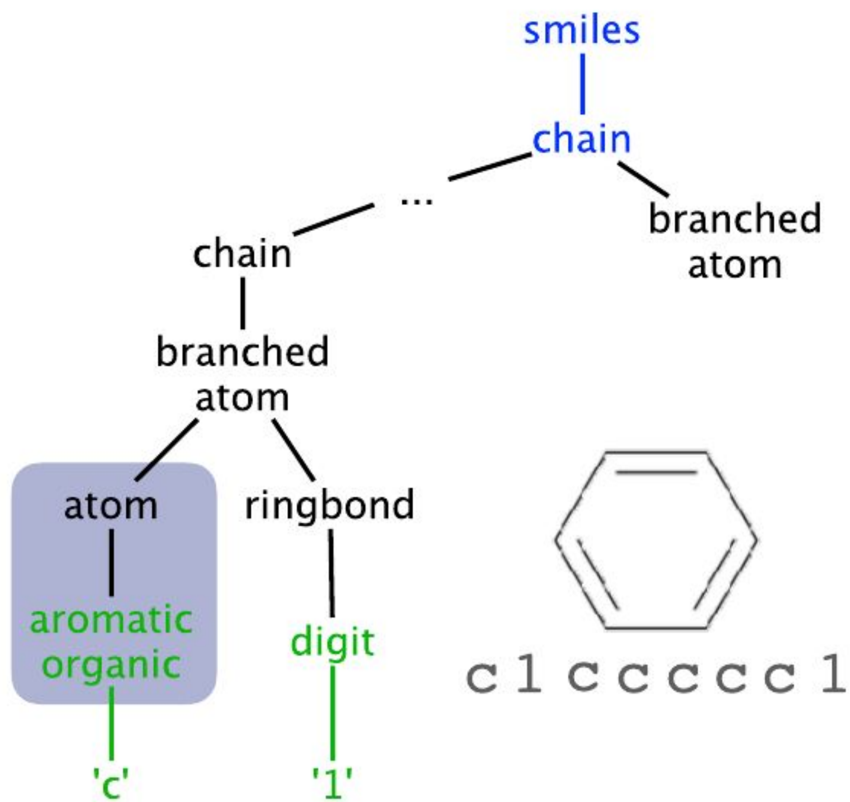
A NEW SEQUENCE!

[Kusner, Paige, Hernández-Lobato, 2017]

Context-Free Grammar

[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'



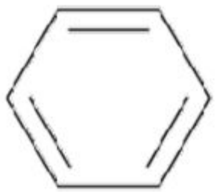
A NEW SEQUENCE!

[Kusner, Paige, Hernández-Lobato, 2017]

Context-Free Grammar

[James et al., 2015]

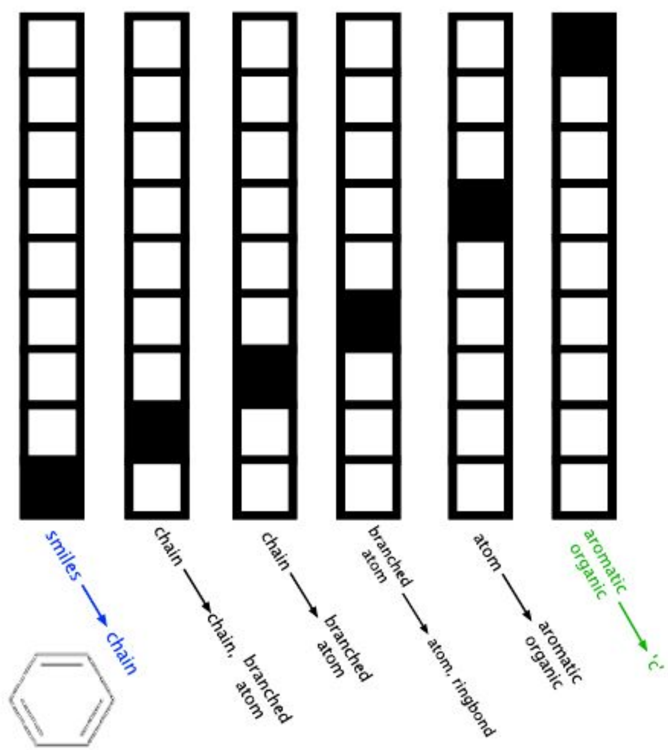
smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'



smiles → chain
chain → chain, branched atom
...
chain → branched atom
branched atom → atom, ringbond
atom → aromatic organic
aromatic organic → 'c'
ringbond → digit
digit → '1'
...

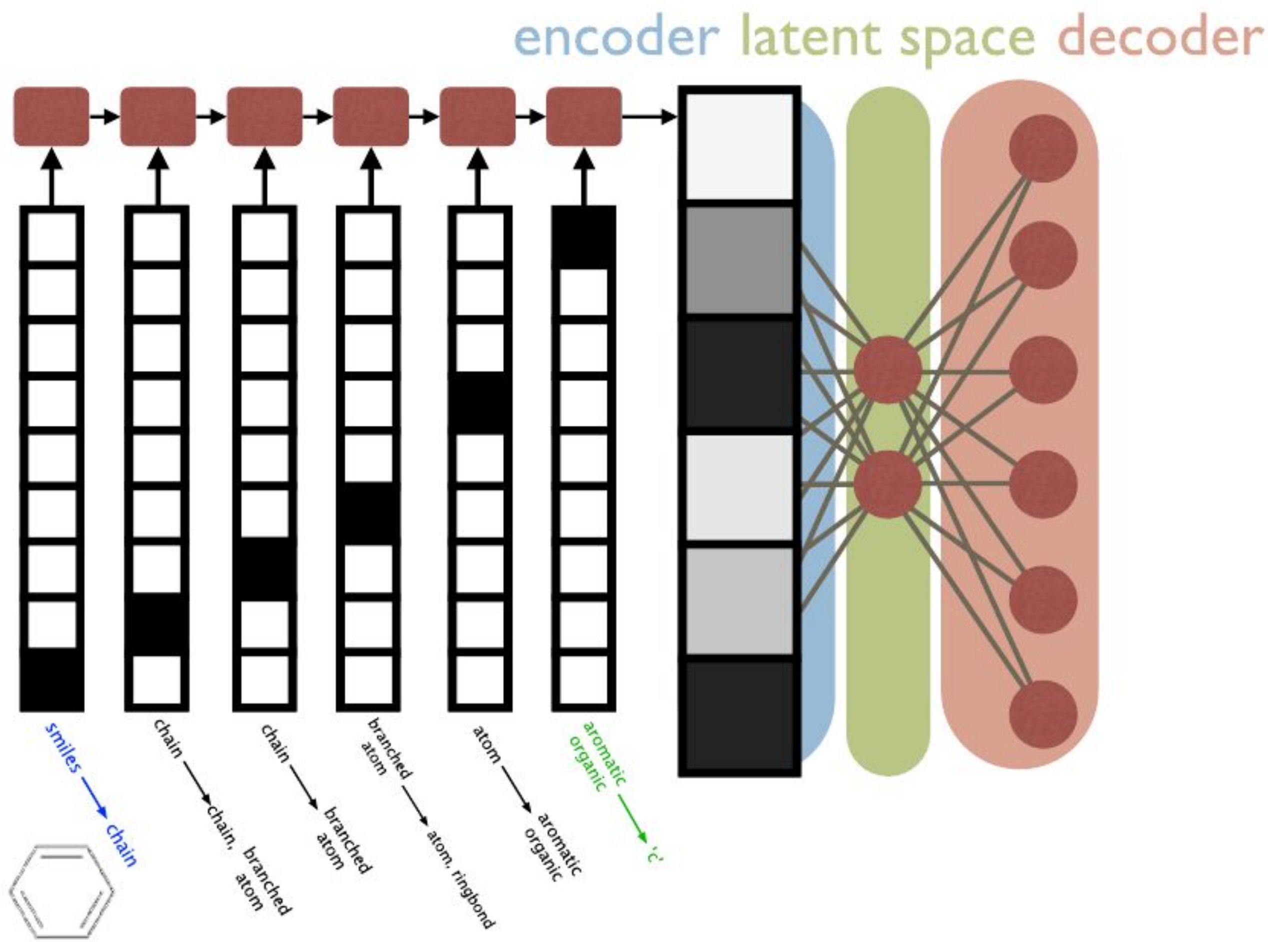
ENCODER

[Kusner, Paige, Hernández-Lobato, 2017]



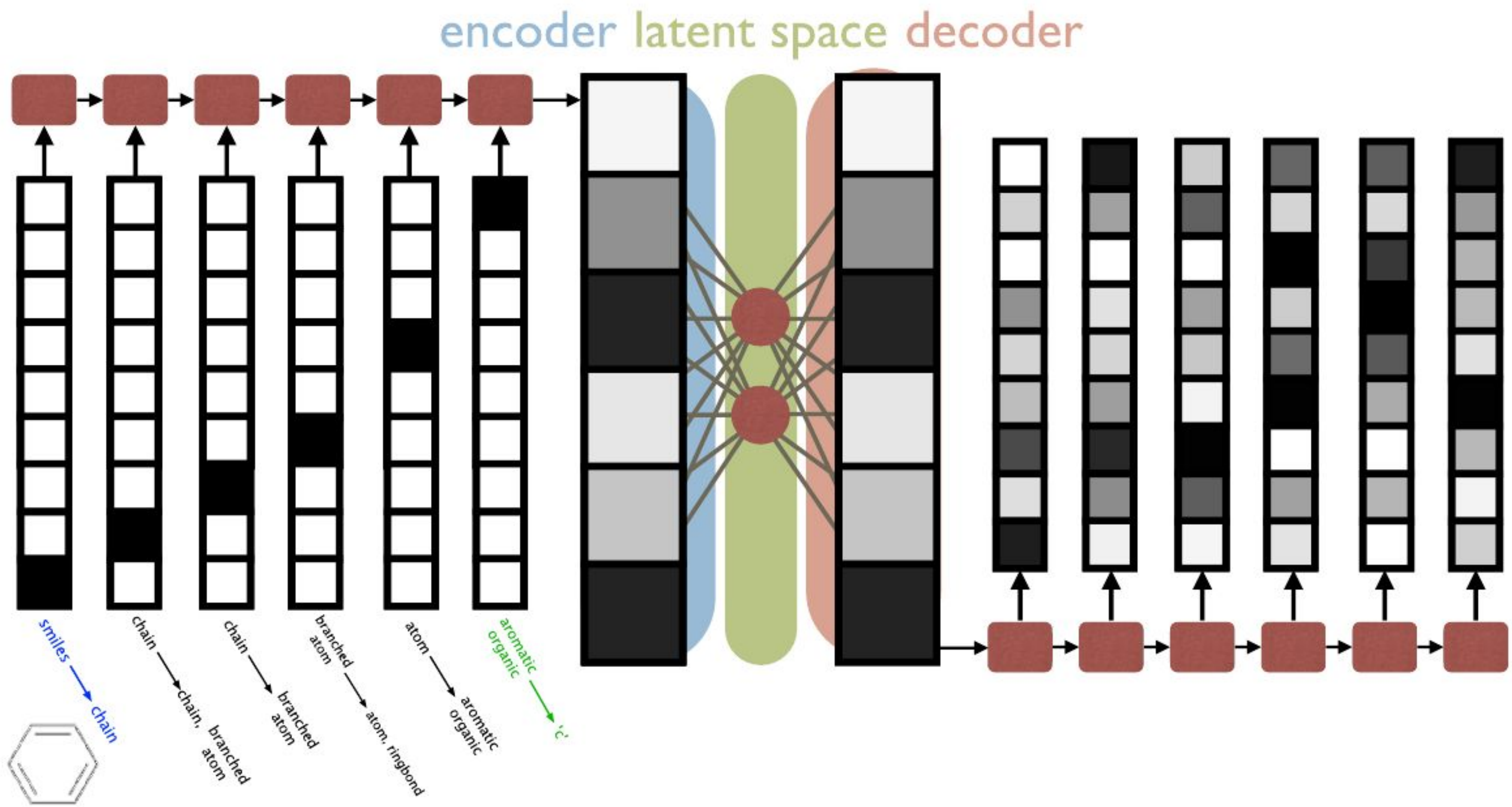
ENCODER

[Kusner, Paige, Hernández-Lobato, 2017]



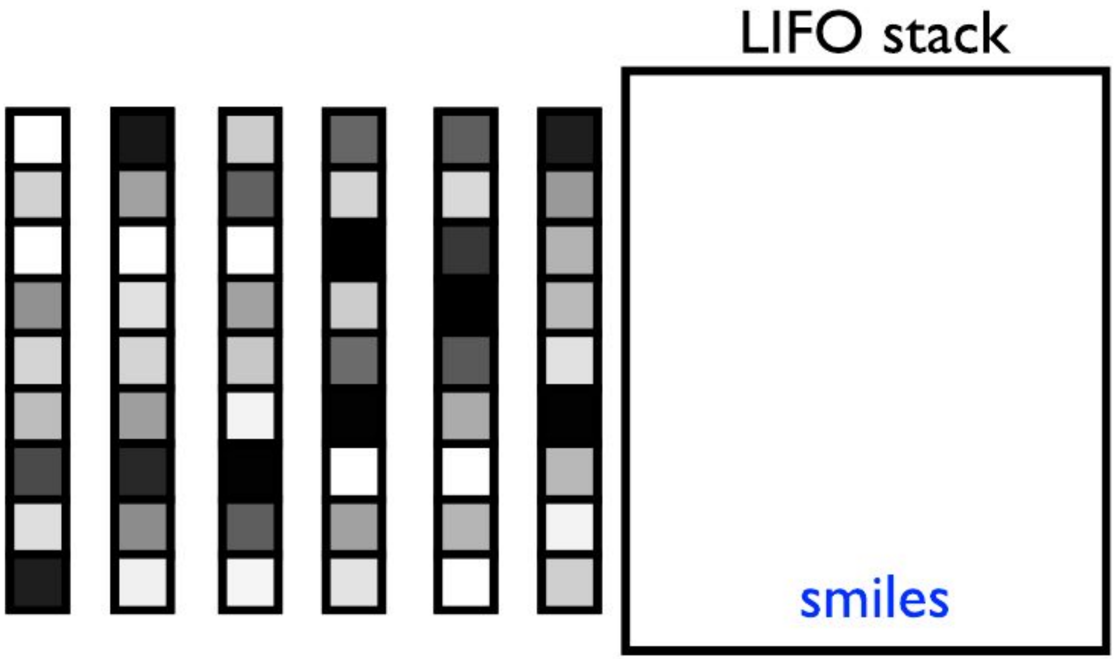
DECODER

[Kusner, Paige, Hernández-Lobato, 2017]



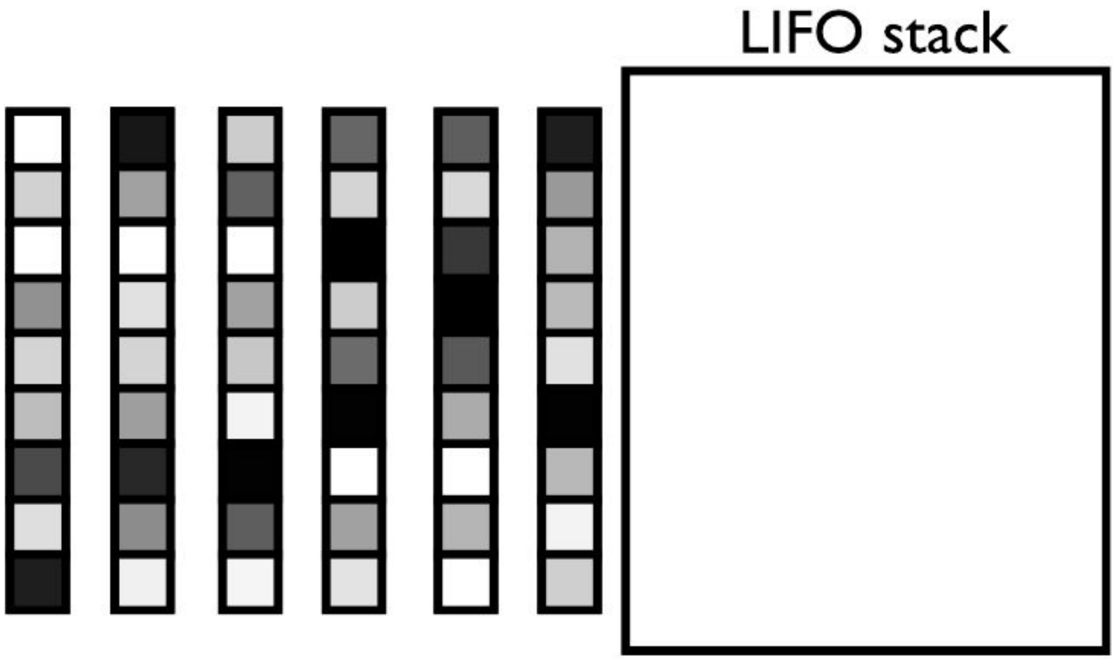
DECODER

[Kusner, Paige, Hernández-Lobato, 2017]



DECODER

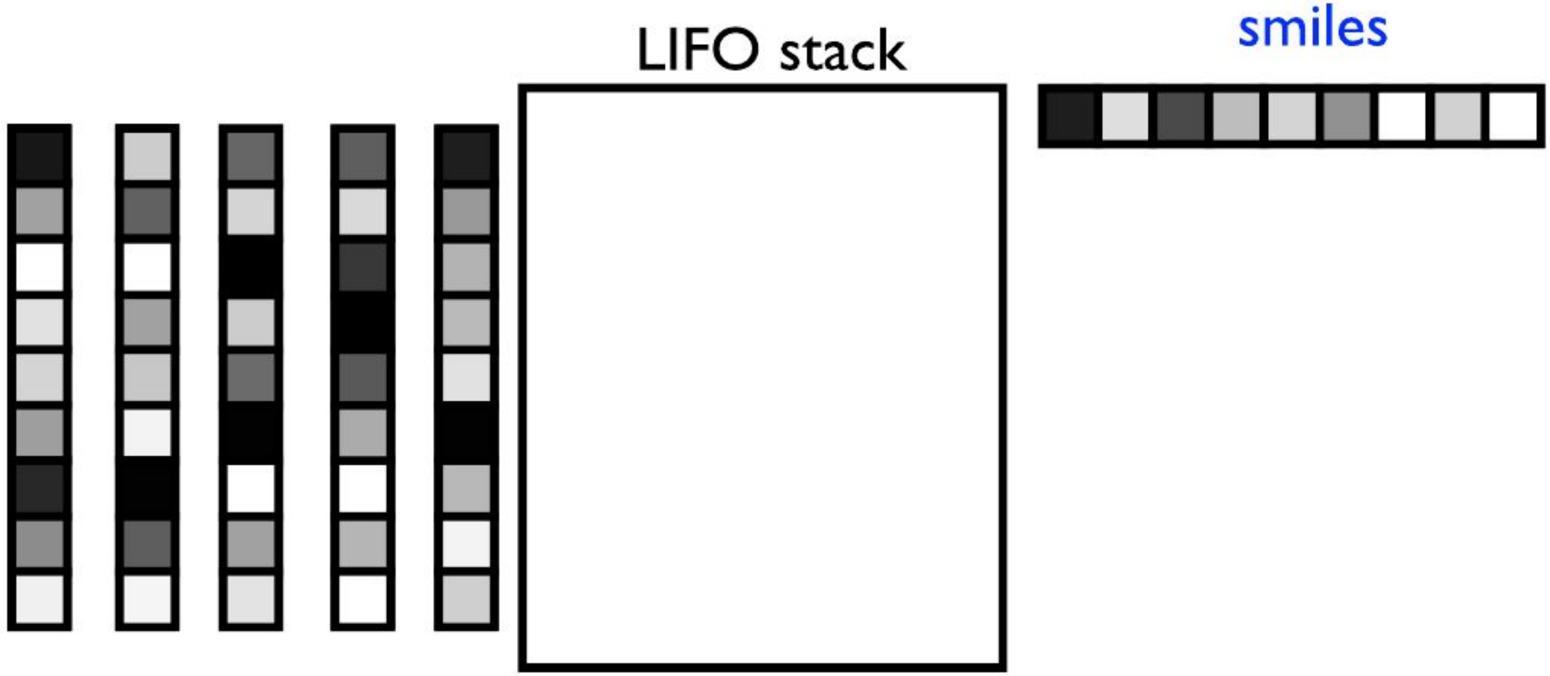
[Kusner, Paige, Hernández-Lobato, 2017]



smiles

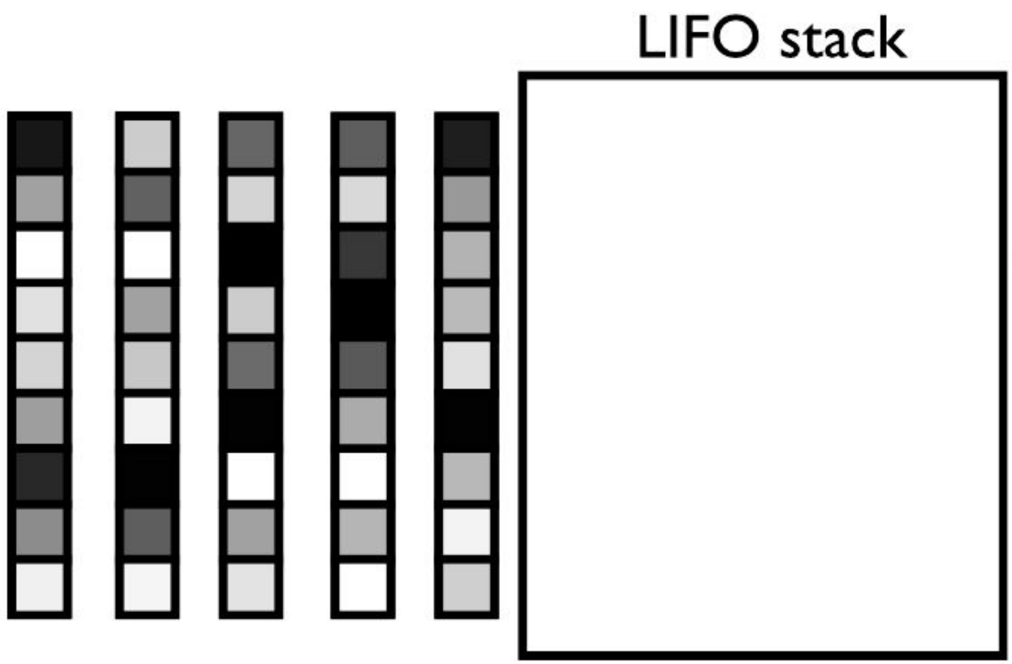
DECODER

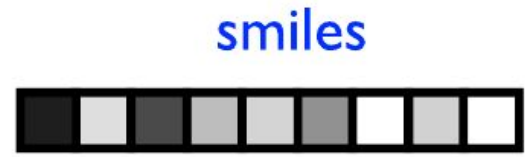
[Kusner, Paige, Hernández-Lobato, 2017]



DECODER

[Kusner, Paige, Hernández-Lobato, 2017]





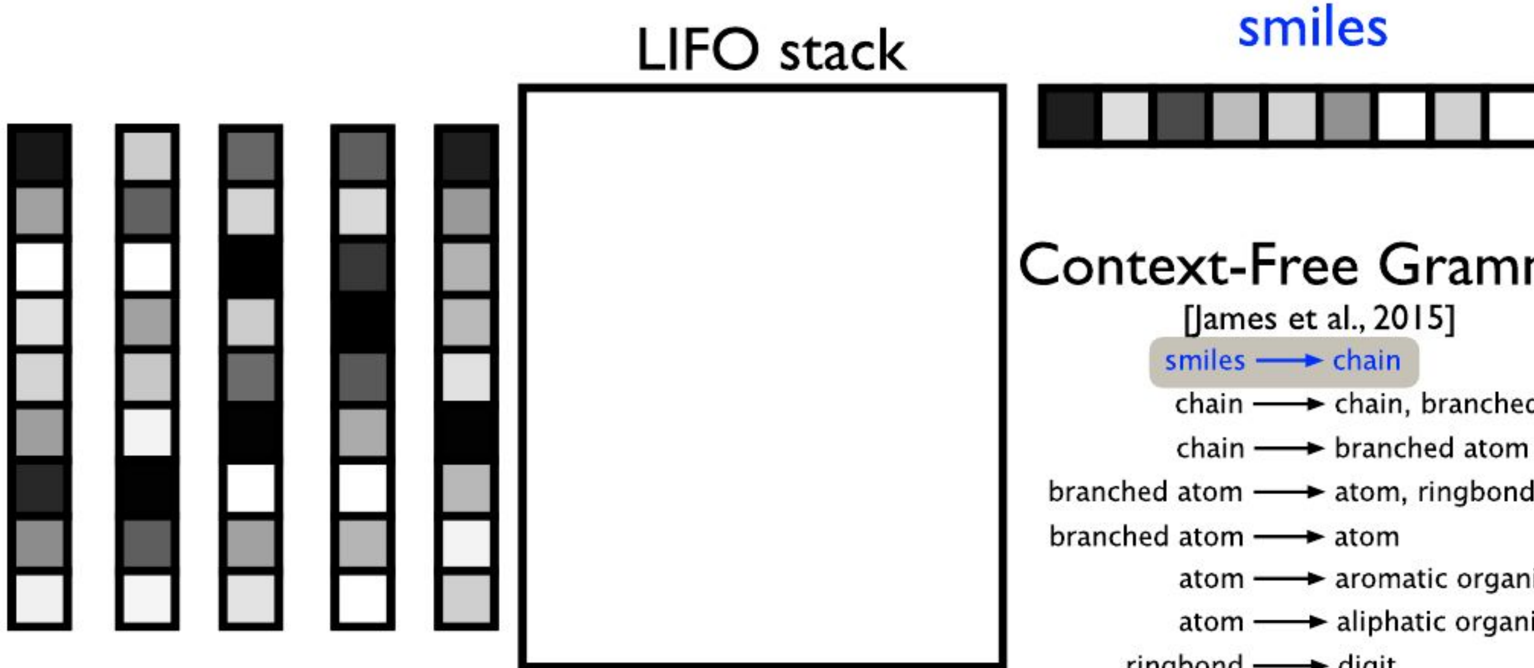
Context-Free Grammar

[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'

DECODER

[Kusner, Paige, Hernández-Lobato, 2017]



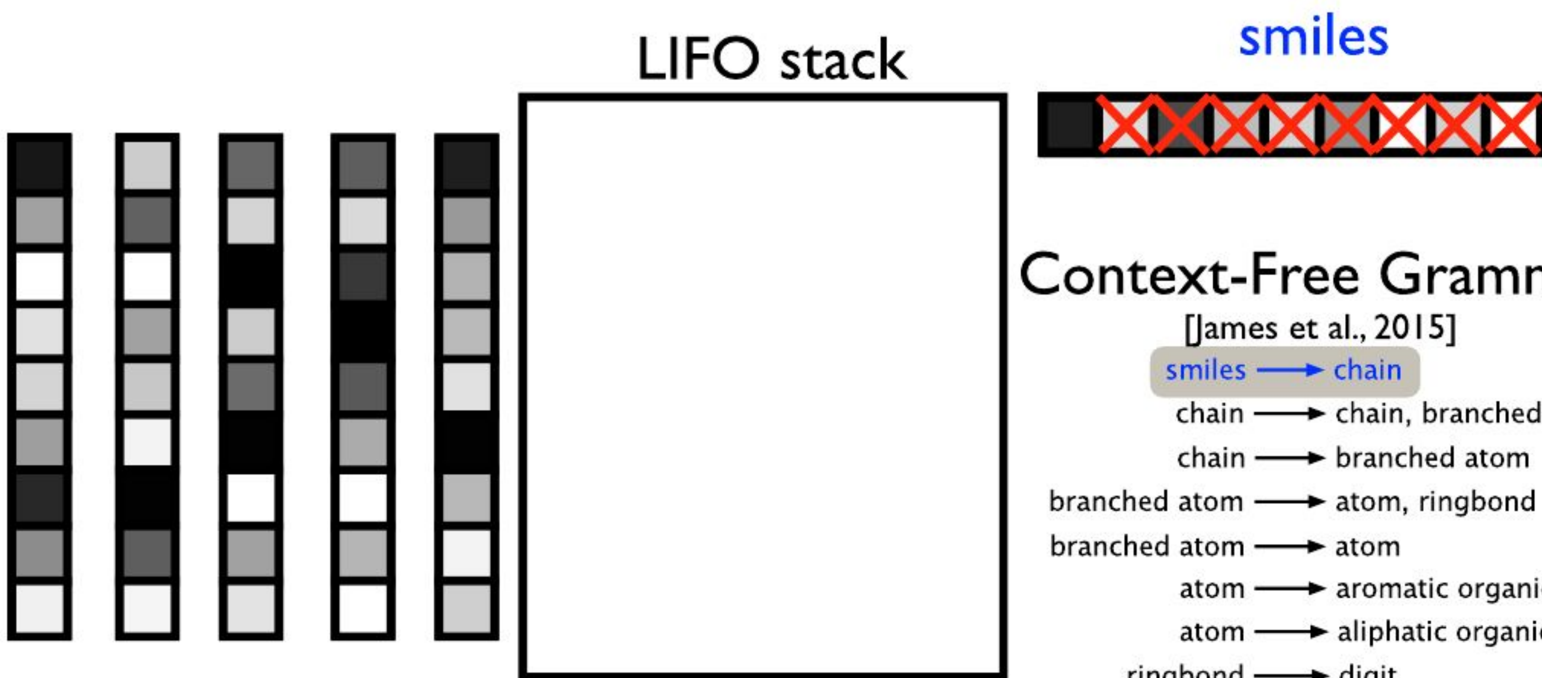
Context-Free Grammar

[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'

DECODER

[Kusner, Paige, Hernández-Lobato, 2017]



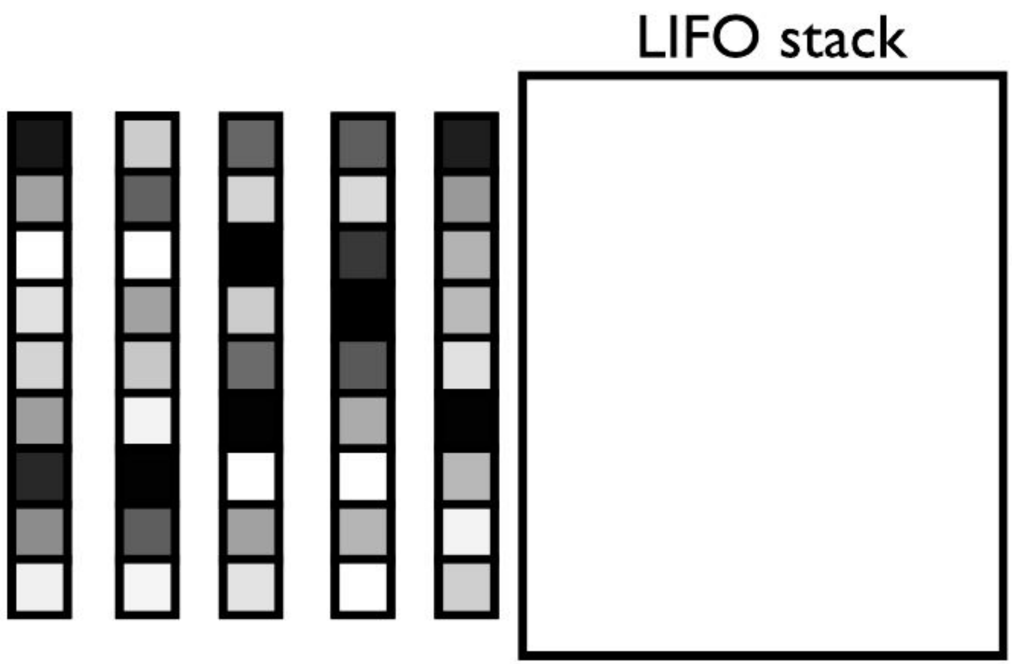
Context-Free Grammar

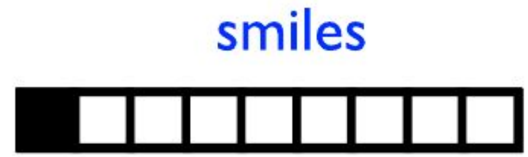
[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'

DECODER

[Kusner, Paige, Hernández-Lobato, 2017]





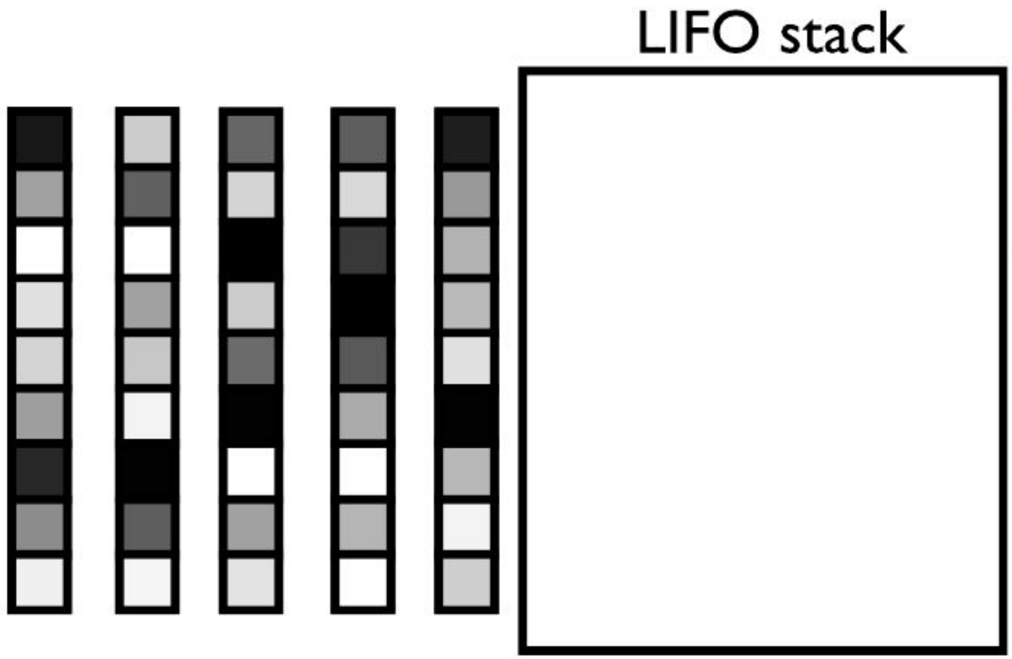
Context-Free Grammar

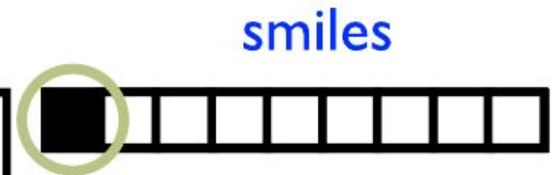
[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'

DECODER

[Kusner, Paige, Hernández-Lobato, 2017]





smiles → chain

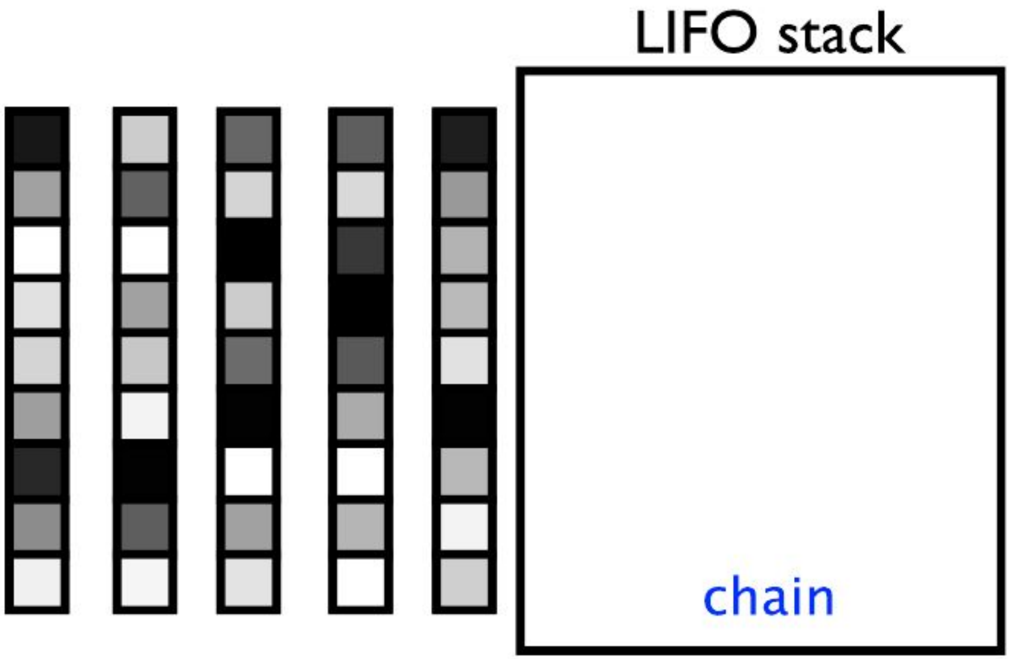
Context-Free Grammar

[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'

DECODER

[Kusner, Paige, Hernández-Lobato, 2017]



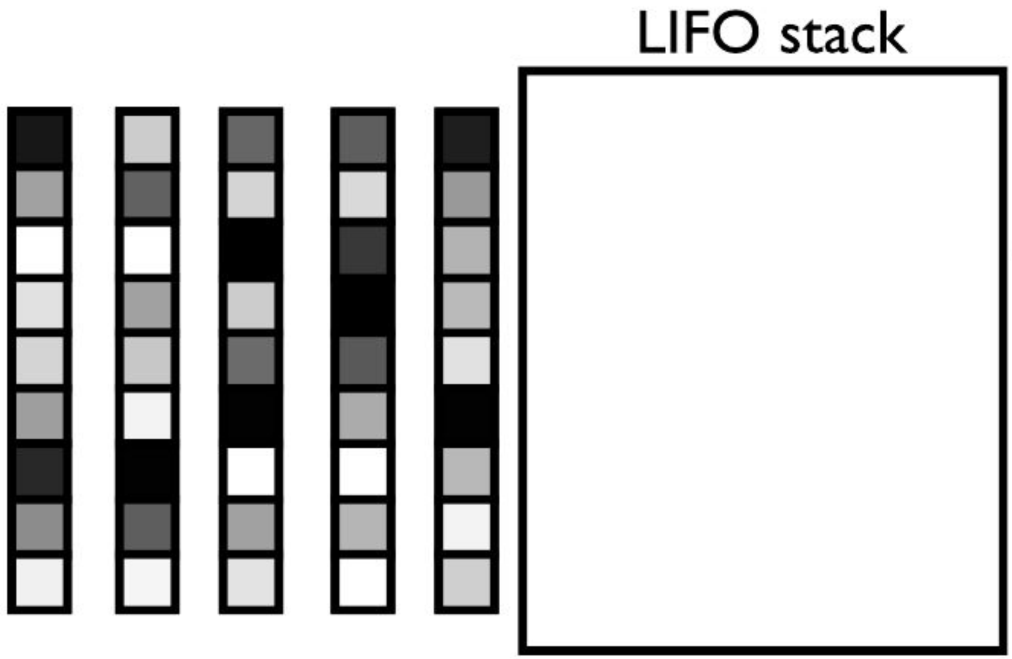
Context-Free Grammar

[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'

DECODER

[Kusner, Paige, Hernández-Lobato, 2017]



chain

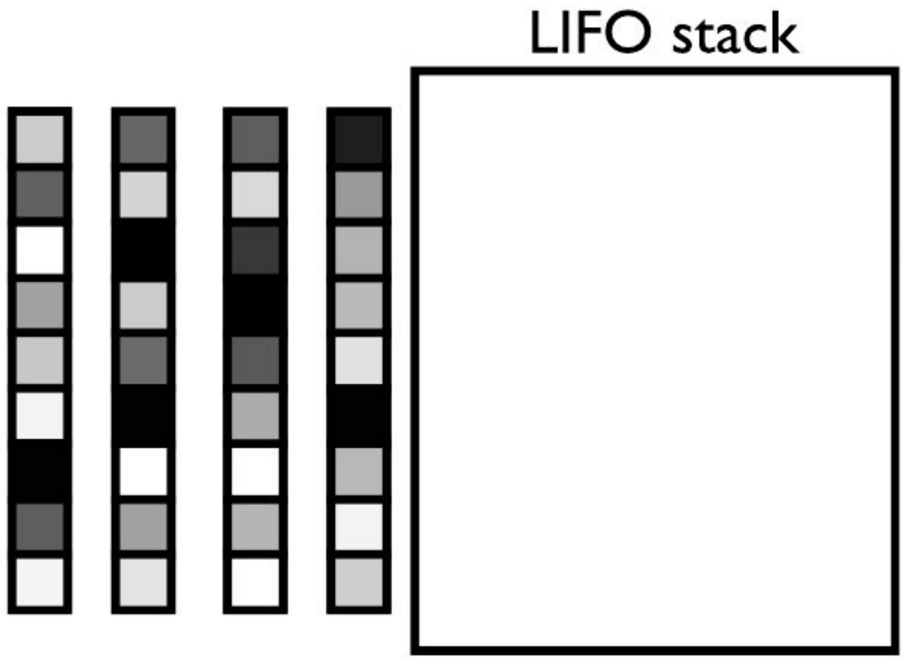
Context-Free Grammar

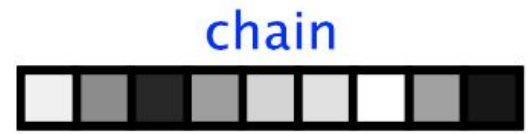
[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'

DECODER

[Kusner, Paige, Hernández-Lobato, 2017]





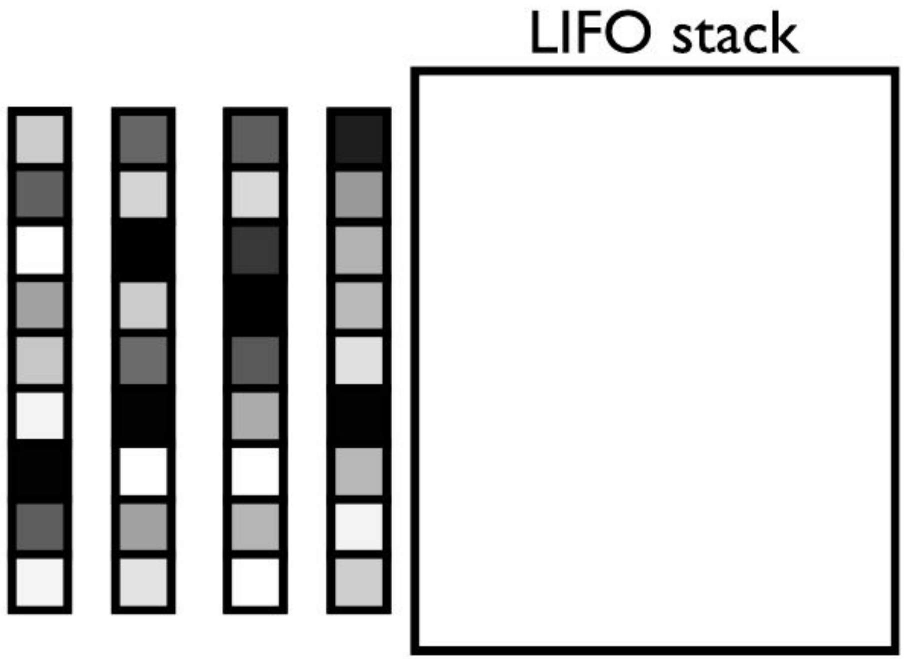
Context-Free Grammar

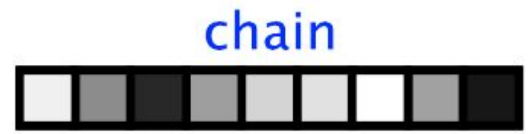
[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'

DECODER

[Kusner, Paige, Hernández-Lobato, 2017]





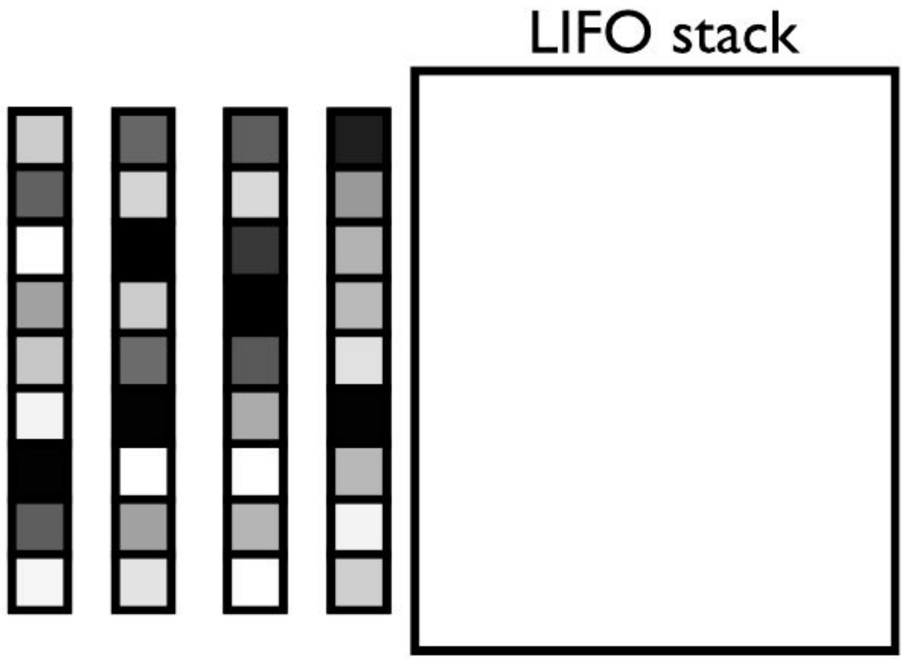
Context-Free Grammar

[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'

DECODER

[Kusner, Paige, Hernández-Lobato, 2017]



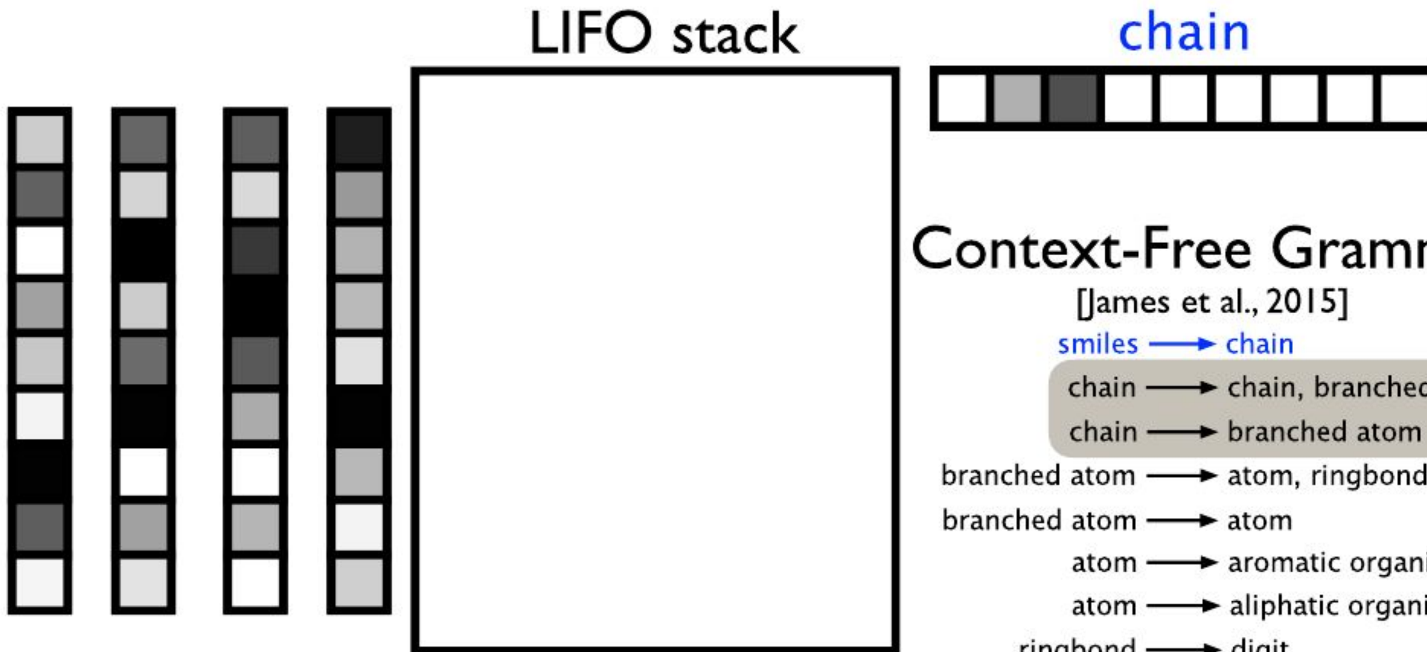
Context-Free Grammar

[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'

DECODER

[Kusner, Paige, Hernández-Lobato, 2017]



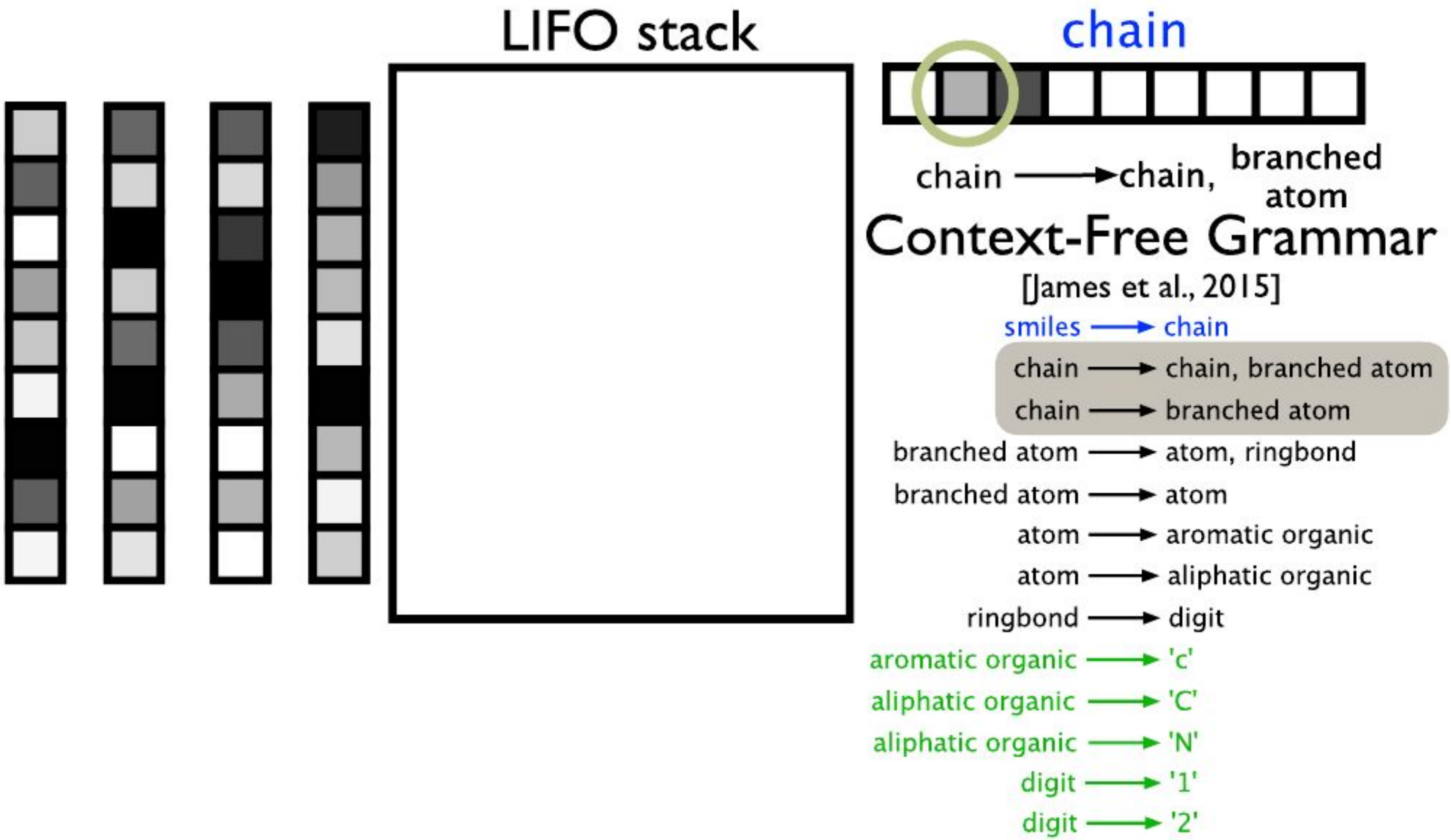
Context-Free Grammar

[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'

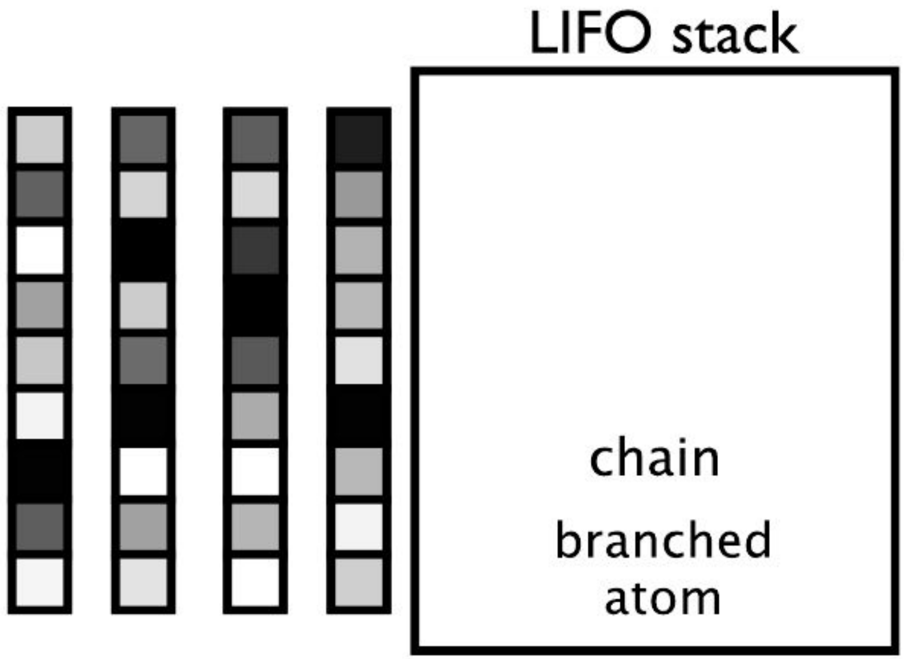
DECODER

[Kusner, Paige, Hernández-Lobato, 2017]



DECODER

[Kusner, Paige, Hernández-Lobato, 2017]



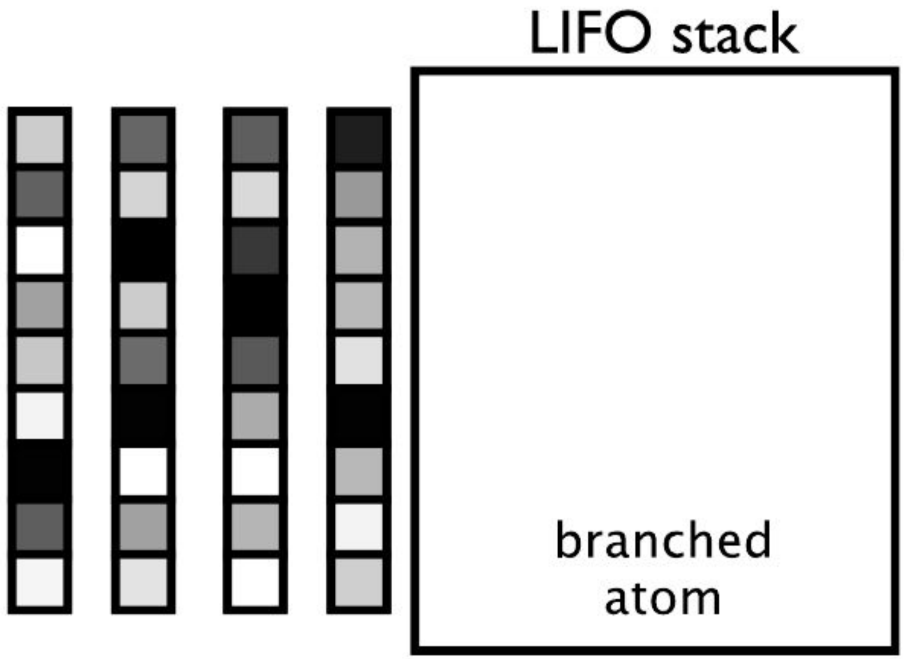
Context-Free Grammar

[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'

DECODER

[Kusner, Paige, Hernández-Lobato, 2017]



chain

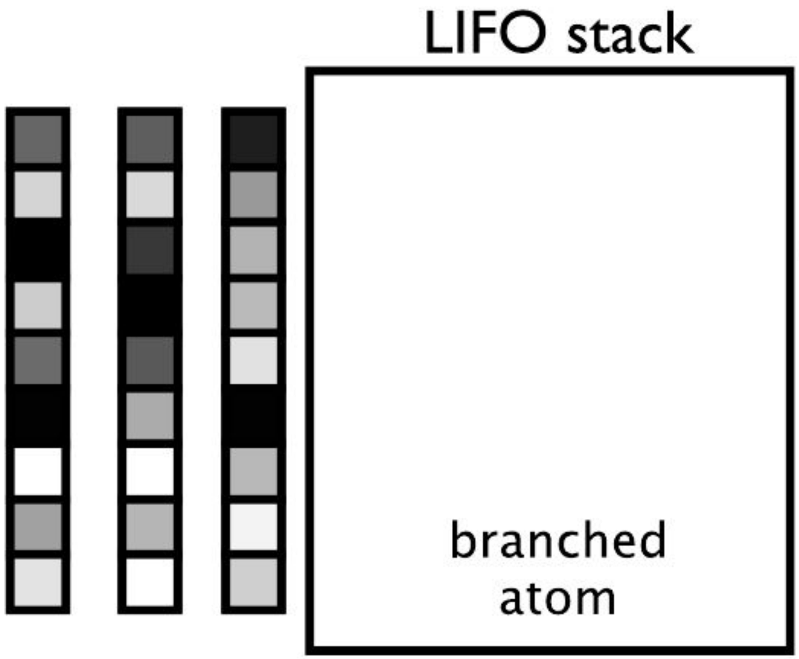
Context-Free Grammar

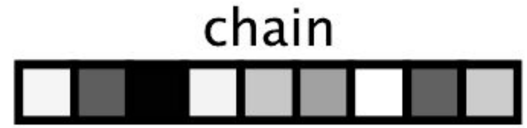
[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'

DECODER

[Kusner, Paige, Hernández-Lobato, 2017]





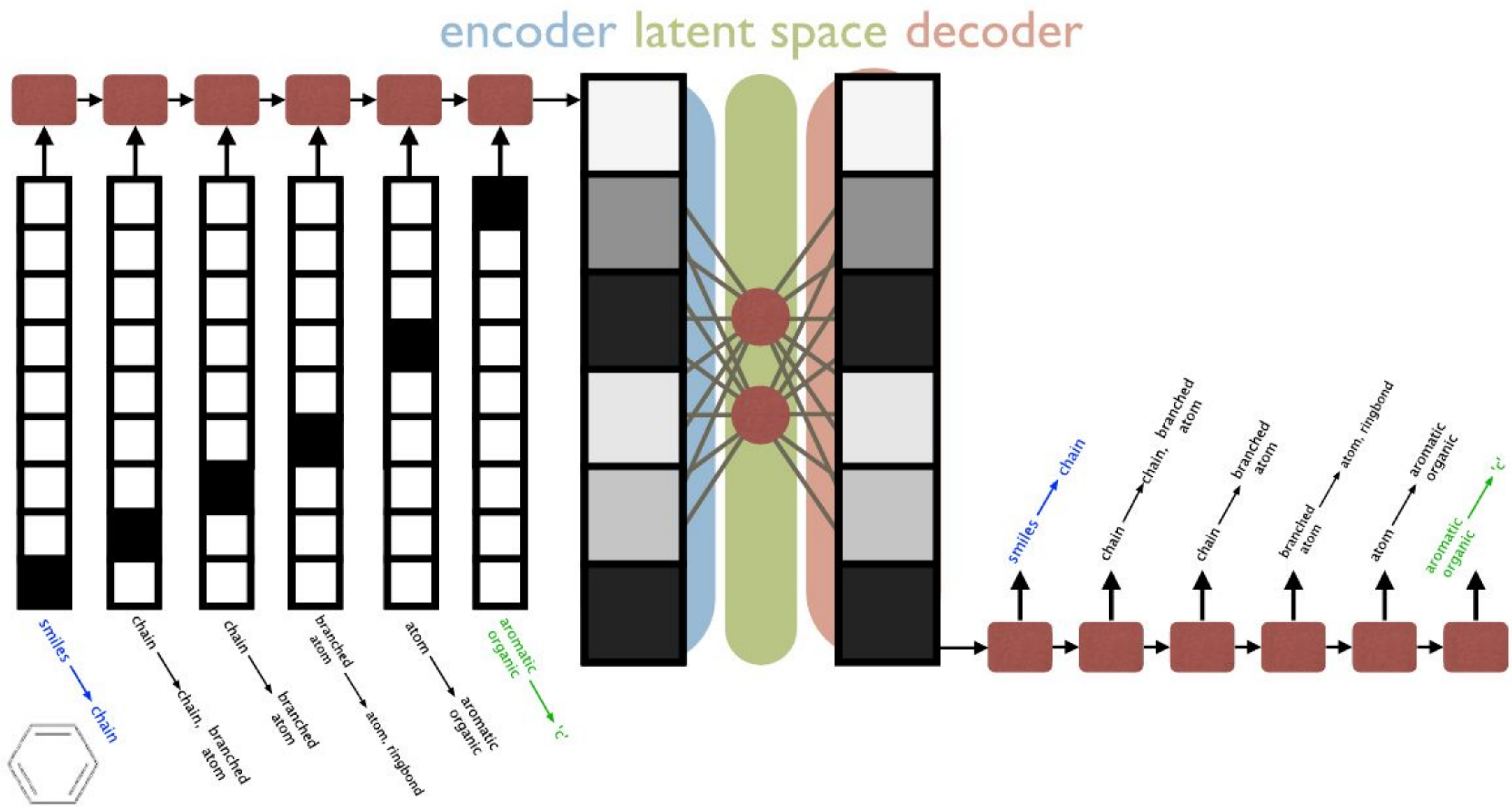
Context-Free Grammar

[James et al., 2015]

smiles → chain
chain → chain, branched atom
chain → branched atom
branched atom → atom, ringbond
branched atom → atom
atom → aromatic organic
atom → aliphatic organic
ringbond → digit
aromatic organic → 'c'
aliphatic organic → 'C'
aliphatic organic → 'N'
digit → '1'
digit → '2'

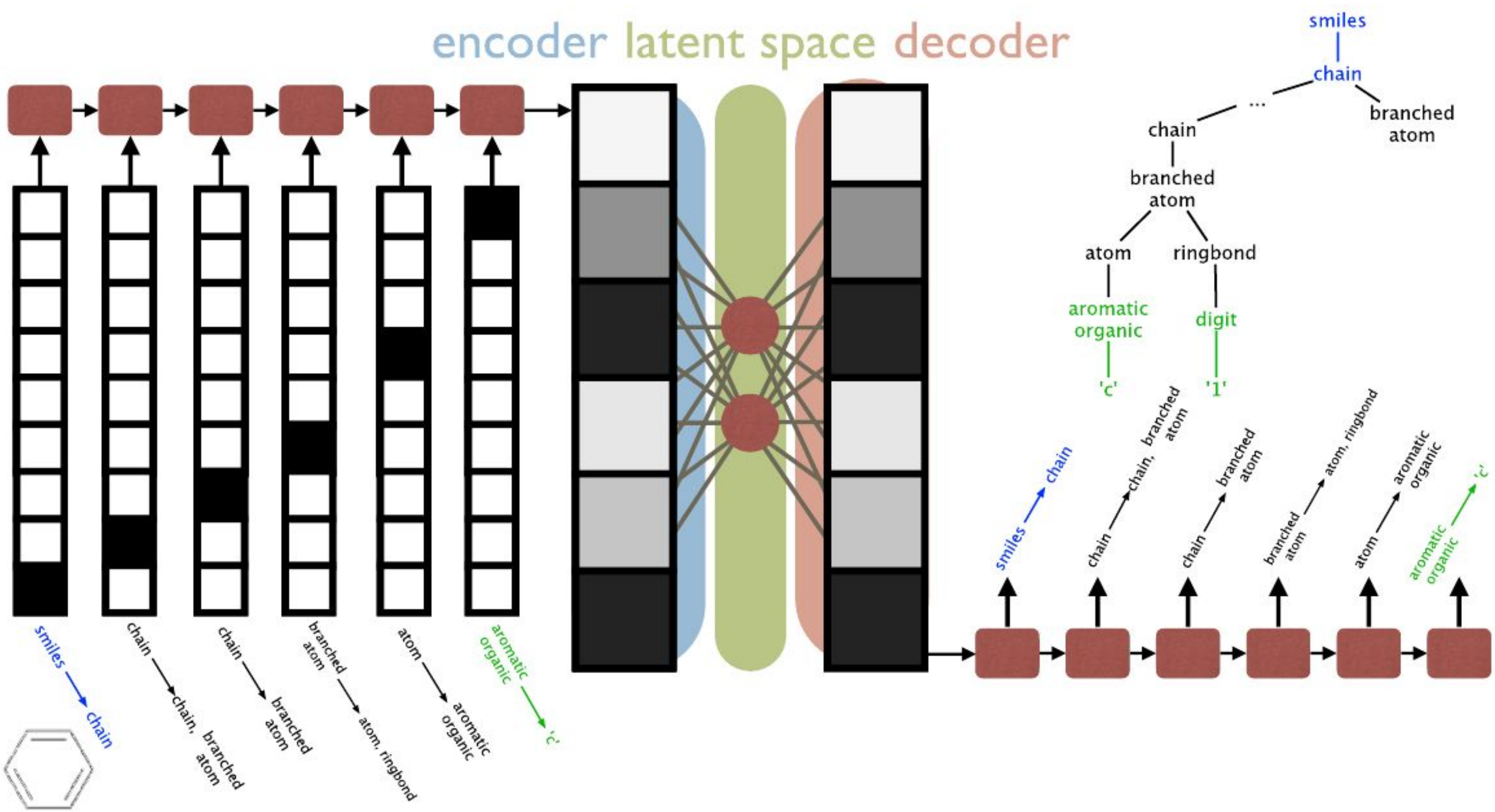
DECODER

[Kusner, Paige, Hernández-Lobato, 2017]



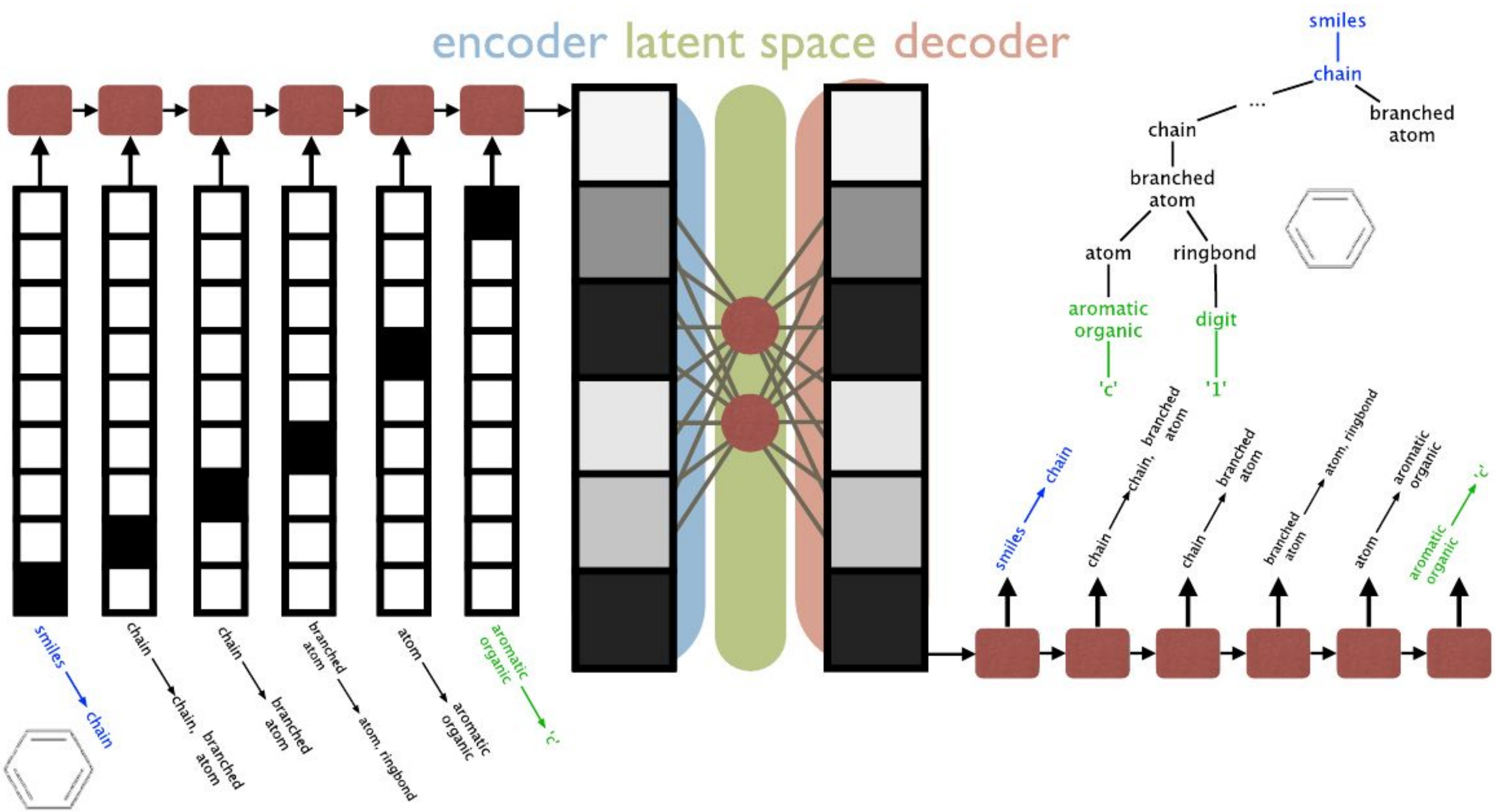
DECODER

[Kusner, Paige, Hernández-Lobato, 2017]



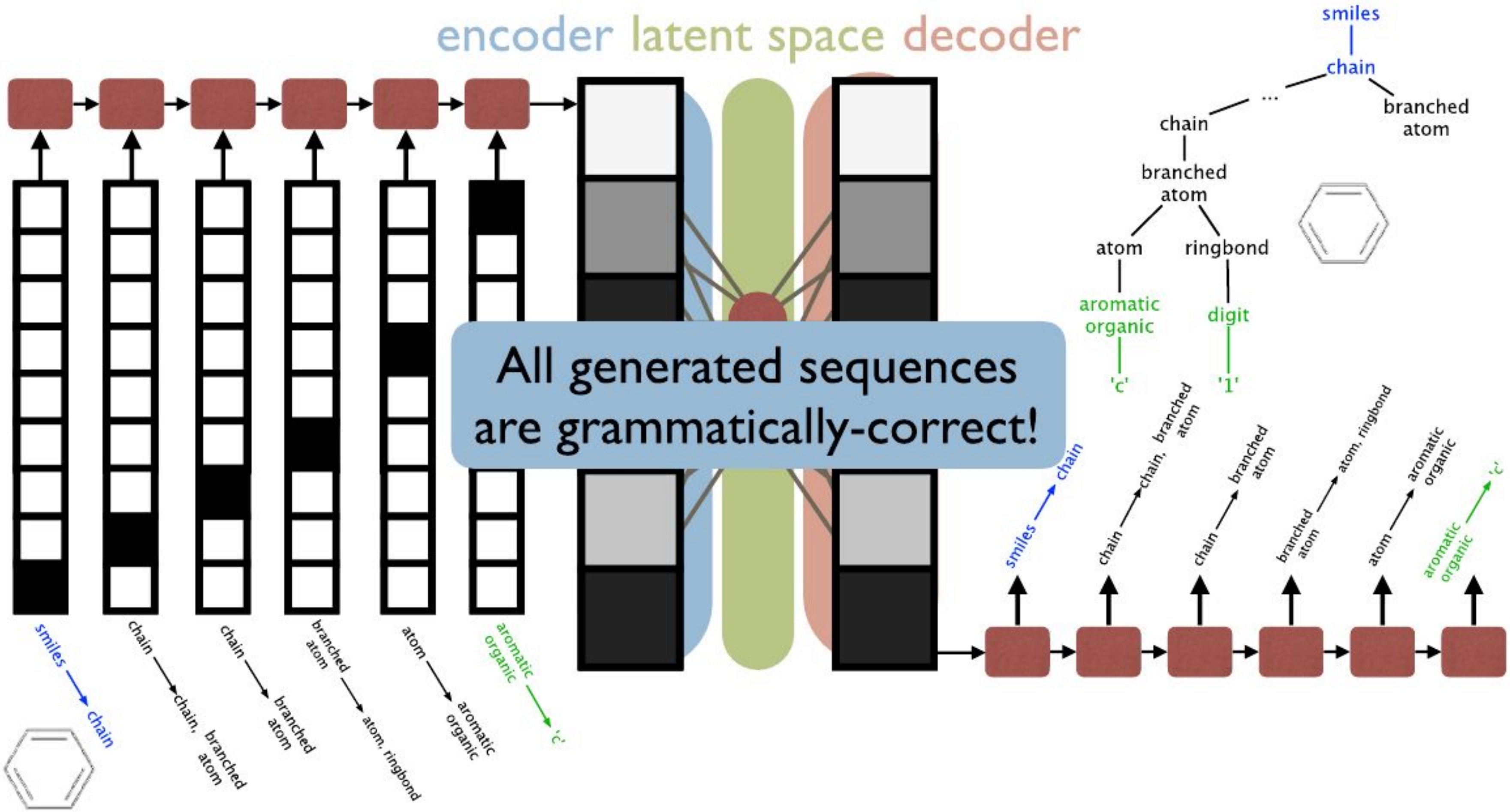
DECODER

[Kusner, Paige, Hernández-Lobato, 2017]



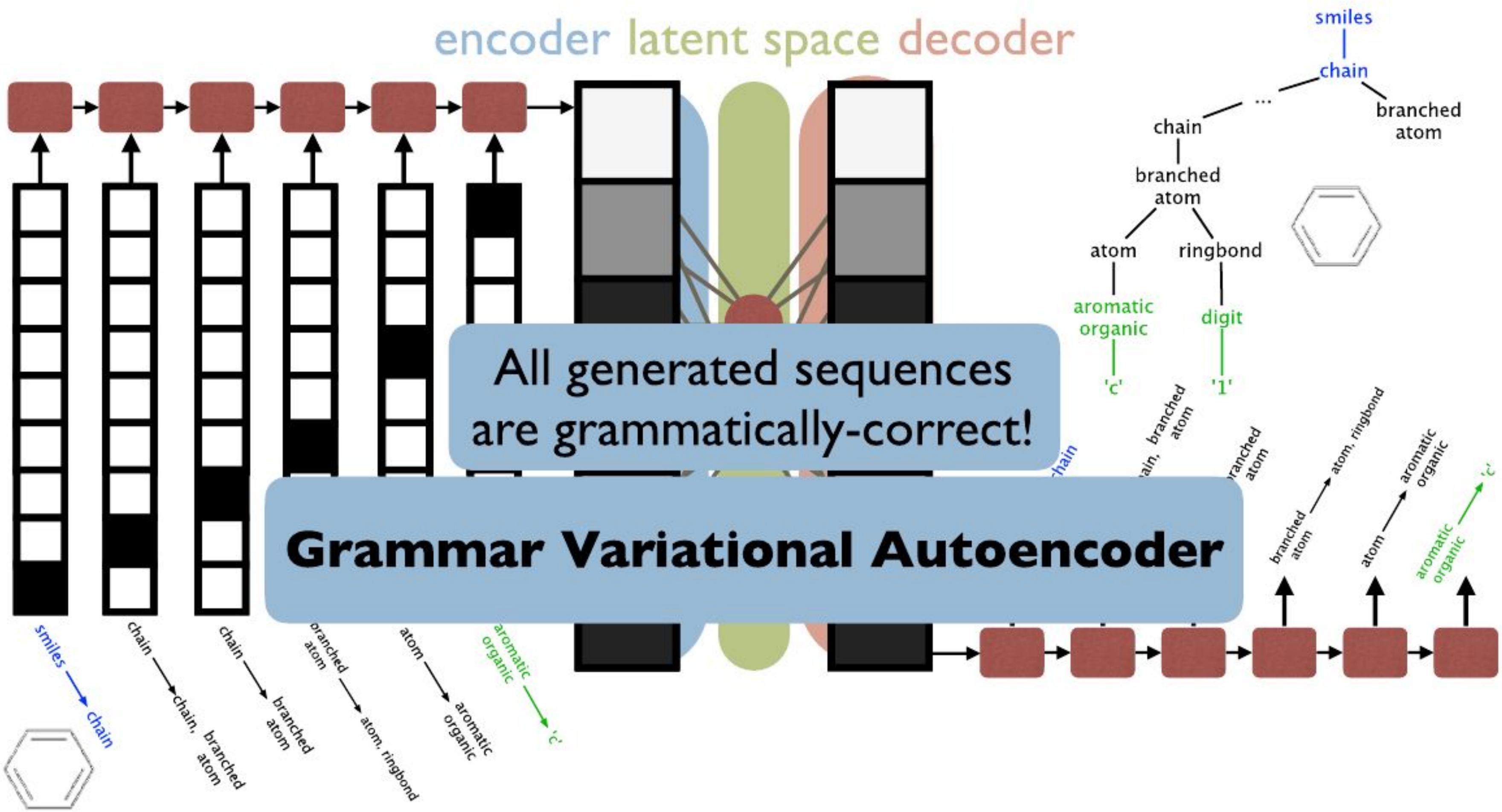
DECODER

[Kusner, Paige, Hernández-Lobato, 2017]



DECODER

[Kusner, Paige, Hernández-Lobato, 2017]



All generated sequences are grammatically-correct!

Grammar Variational Autoencoder

HOW WELL DOES THIS WORK?

Reconstruction accuracy and sample validity results.

Method	% Reconstruct	% Prior Valid
GVAE	**53.7**	**7.2**
CVAE	44.6	0.70

Results finding best molecule

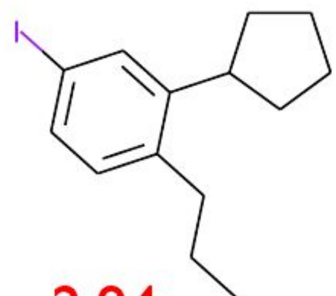
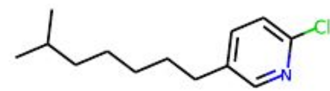
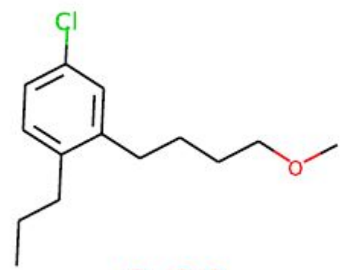
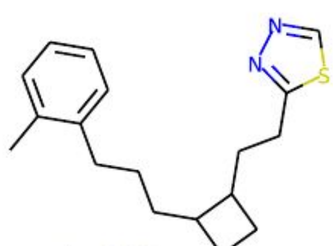
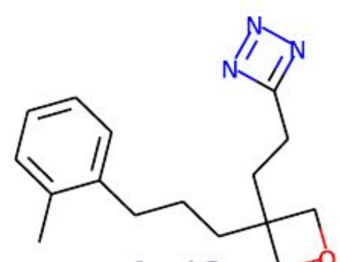
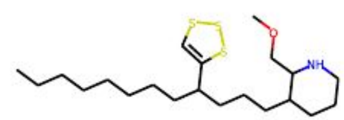
Method	Frac. valid	Avg. score
GVAE	**0.31±0.07**	**-9.57 ±1.77**
CVAE	0.17±0.05	-54.66±2.66

Test Log-likelihood (LL) and RMSE for predictions from latent space

Objective	Method	Molecules
LL	GVAE	**-1.739 ±0.004**
	CVAE	-1.812±0.004
RMSE	GVAE	**1.404 ±0.006**
	CVAE	1.504±0.006

DRUG SEARCH

Trained on **ZINC**
250K molecules
maximum 120 chars

	1st	2nd	3rd
GVAE [Kusner, Paige, Hernández-Lobato, 2017]			
drug-likeness:	2.94	2.89	2.80
CVAE [Gómez-Bombarelli et al., 2016]			
drug-likeness:	1.98	1.42	1.19

Take home messages

Generative models of molecues...

1. **Data-driven approach** to molecule generation, no expertise needed.
2. Sampled molecules are **realistic**, unlike those generated with rules.
3. Create a **continuous latent space** which is useful for optimization.

Grammar variational autoencoder...

1. Produces a **larger fraction of valid molecules** when decoding.
2. Produces **better predictions of molecule properties** from latent space.
3. **Molecule optimization** results are **improved**.

A lot of work going on this area...

Adding semantic constraints into GVAE:

- Dai, Tian, Dai, Skiena, and Song, 2018.

Generative models of graphs:

- Li, Vinyals, Dyer, Pascanu and Battaglia, 2018.
- Liu, Allamanis, Brockschmidt and Gaunt, 2018.
- Jin, Barzilay and Jaakkola, 2018.

Semi-supervised generative models:

- Kang and Cho, 2018.

Using GANs insead of VAEs:

- Guimaraes, Sanchez-Lengeling, Outeiral, Farias and Aspuru-Guzik, 2017.
- De Cao and Kipf, 2018.

and many more!

Collaborators


R. P. Adams


J. Aguilera-Ipaguirre


A. Aspuru-Guzik


D. Duvenaud


R. Gomez-Bonbardelli


T. Hirzel


M. Kusner

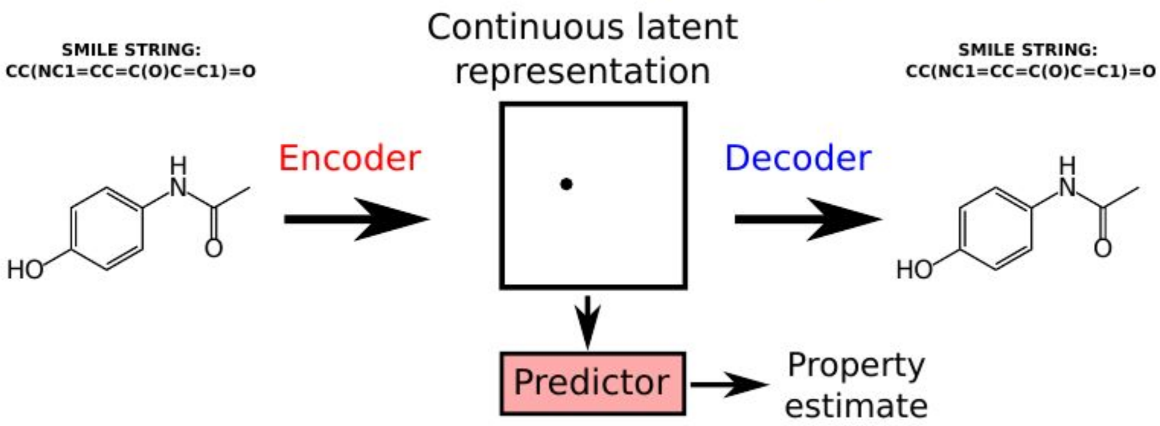

B. Paige


B. Sánchez-Lengeling

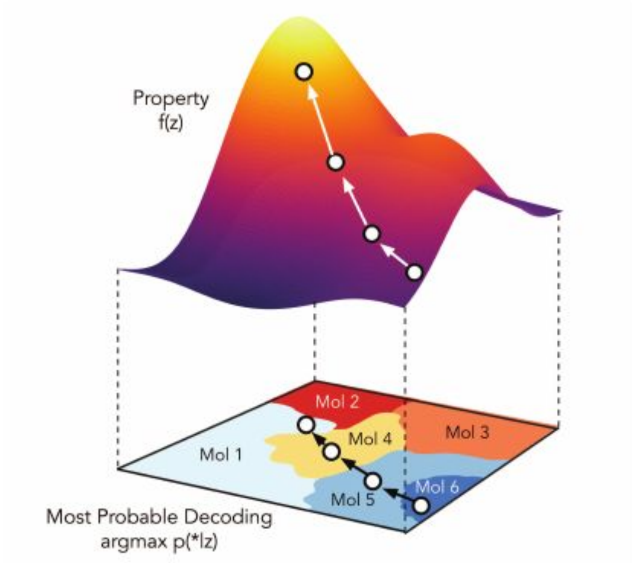

D. Sheberla


J. Wey

Latent space optimization (LSO)



Data Efficient optimization possible in the latent space of a **deep generative model**.



We avoid directly working with structured high-dimensional input spaces.

Off-the-shelf data efficient optimization methods can be used in latent space.

The generated points will exihibit similar patterns to those found in training set.

However, several **limitations** still exist.

Gómez-Bombarelli*, Wei*, Duvenaud*, Hernández-Lobato*, Sánchez-Lengeling, Sheberla, Aguilera-Iparraguirre, Hirzel, Adams and Aspuru-Guzik, 2018. (⋆ equal contributors).

Weighted Retraining

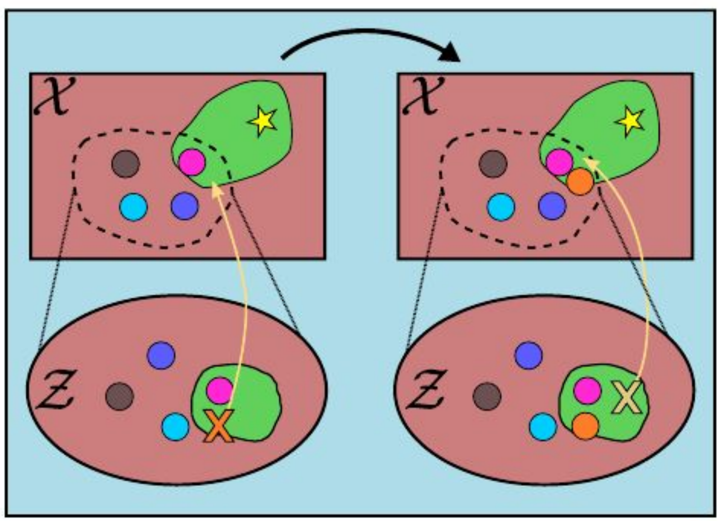
Austin Tripp⋆, Erik Daxberger⋆ and José Miguel Hernández-Lobato,
Sample-Efficient Optimization in the Latent Space of Deep Generative Models via Weighted Retraining,
In NeurIPS 2020. (⋆ equal contributors).

Limitations of current LSO methods

❶ The generative model **is not retrained** after collecting new data.

LSO cannot find an optimum lying far from the training data.

❷ Even if the model is retrained, **the new data has a diminishing effect**.



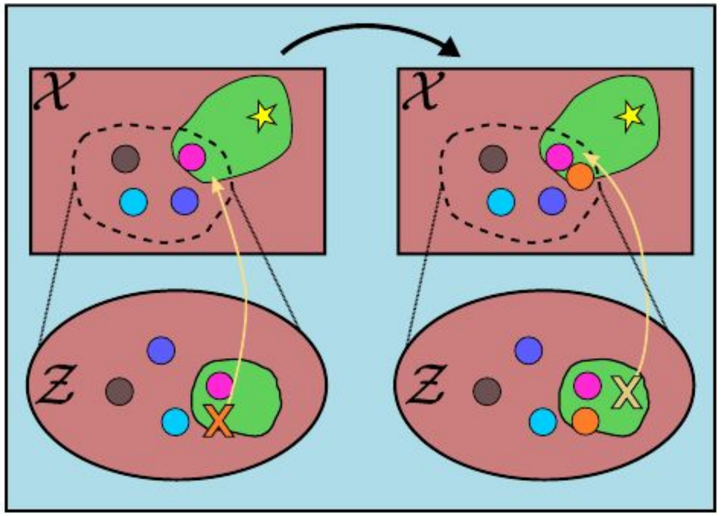
Standard Latent Space Optimization

Limitations of current LSO methods

❶ The generative model **is not retrained** after collecting new data.

LSO cannot find an optimum lying far from the training data.

❷ Even if the model is retrained, **the new data has a diminishing effect**.



Standard Latent Space Optimization

We address

❶ by **retraining** the generative model

and

❷ by **weighting** the data.

Weighted retraining

- We **retrain** the generative model after collecting new data (e.g. 50 points).
- We fit the model by optimizing the **weighted** objective $\sum_{n=1}^{N} w_n \mathcal{L}(\mathrm{x}_n)$,

where

$$w_n \propto \frac{1}{kN + \mathrm{rank}(\mathrm{x}_n) - 1} \, ,$$

k controls the weighting:

$k = \infty$ is uniform weighting, $k = 0$ places all weight on single point with highest value.

Weighted retraining

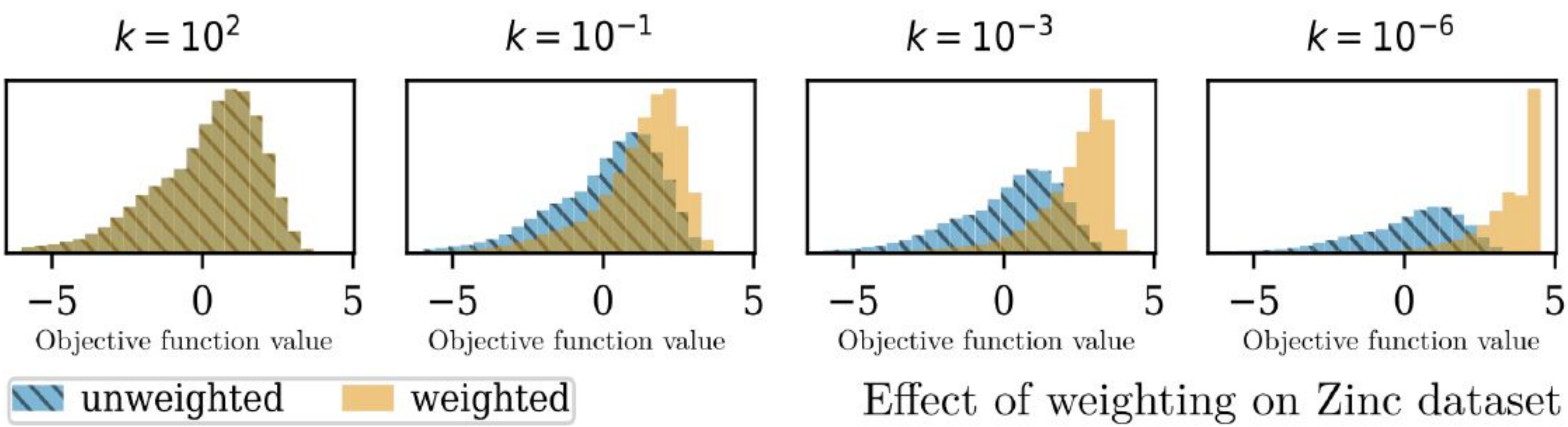
- We **retrain** the generative model after collecting new data (e.g. 50 points).
- We fit the model by optimizing the **weighted** objective $\sum_{n=1}^{N} w_n \mathcal{L}(\mathrm{x}_n)$,

where

$$w_n \propto \frac{1}{kN + \mathrm{rank}(\mathrm{x}_n) - 1},$$

k controls the weighting:

$k = \infty$ is uniform weighting, $k = 0$ places all weight on single point with highest value.



Effect of weighting on Zinc dataset

Weighted retraining

- We **retrain** the generative model after collecting new data (e.g. 50 points).
- We fit the model by optimizing the **weighted** objective $\sum_{n=1}^{N} w_n \mathcal{L}(x_n)$,

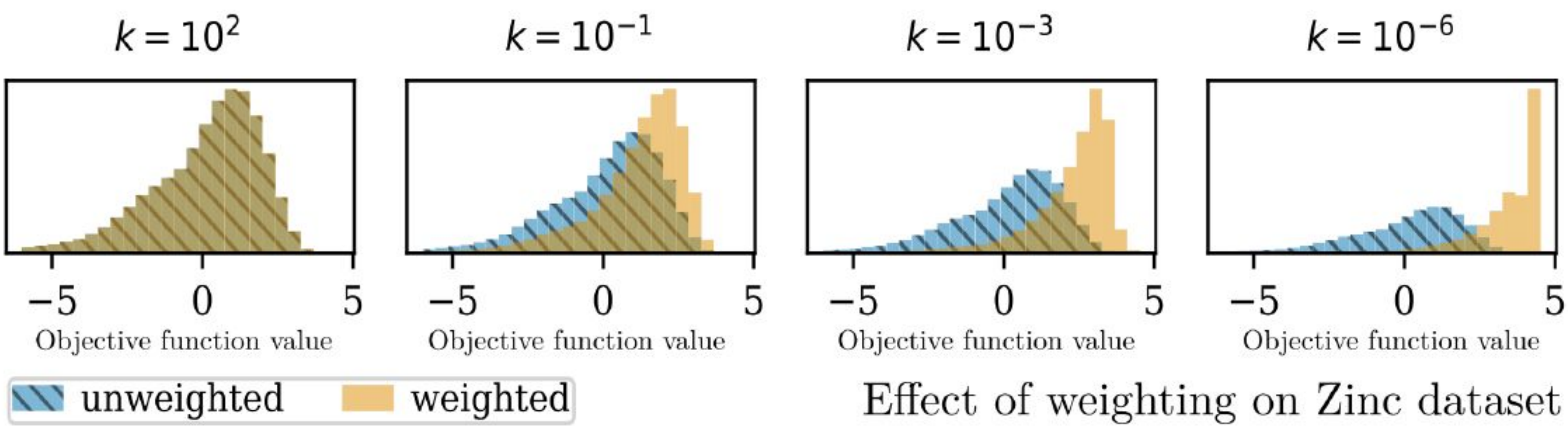
where

$$w_n \propto \frac{1}{kN + \mathrm{rank}(x_n) - 1},$$

k controls the weighting:

$k = \infty$ is uniform weighting, $k = 0$ places all weight on single point with highest value.

When $kN \gg 1$, the weight distribution per data fraction **depends on** k but not N.



Effect of weighting on Zinc dataset

Weighted retraining

- We **retrain** the generative model after collecting new data (e.g. 50 points).
- We fit the model by optimizing the **weighted** objective $\sum_{n=1}^{N} w_n \mathcal{L}(x_n)$,

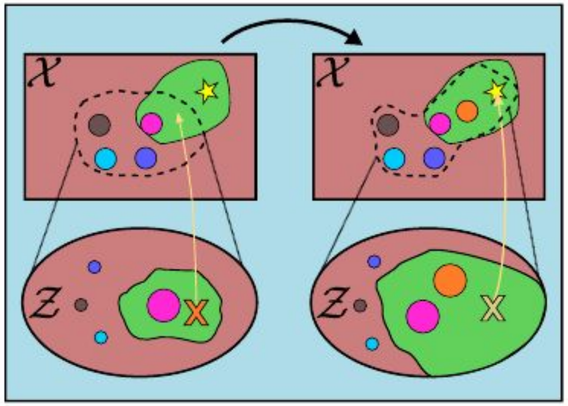
where

$$w_n \propto \frac{1}{kN + \text{rank}(x_n) - 1},$$

k controls the weighting:

$k = \infty$ is uniform weighting, $k = 0$ places all weight on single point with highest value.

When $kN \gg 1$, the weight distribution per data fraction **depends on** k but not N.



with Weighted Retraining

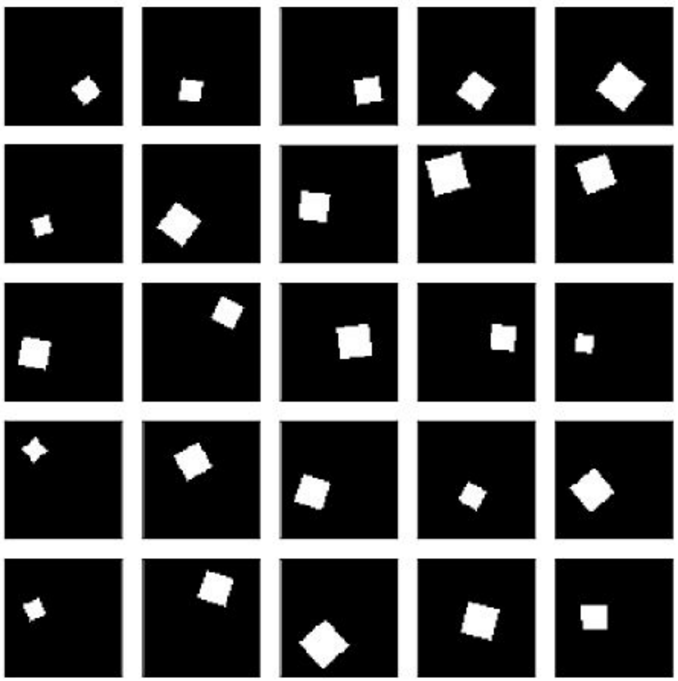
With our approach, a larger fraction of latent space maps into regions of input space with high objective values.

2D shape area maximization

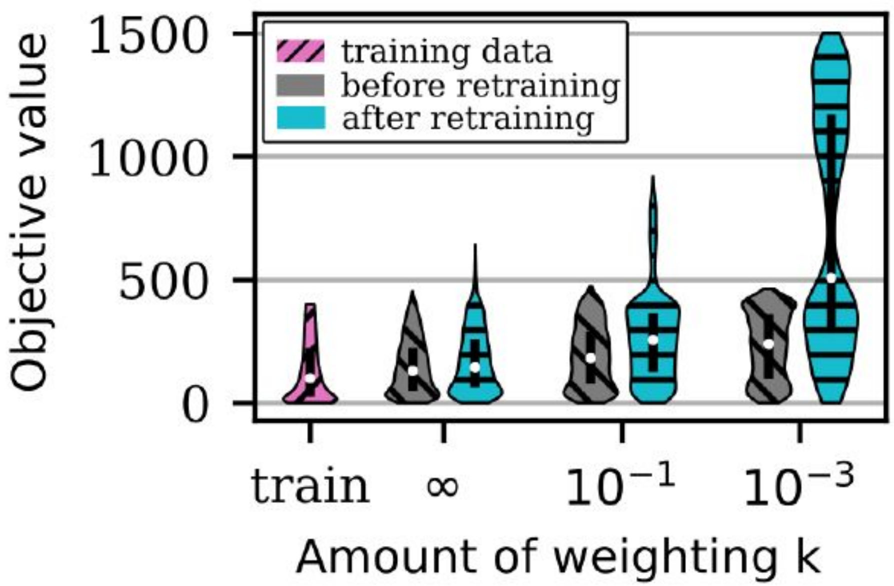
We train a conv-VAE on the dSprites dataset with 245,760 **square shapes**.

Goal: generate shape with **max. area**. Exhaustive search in 5D latent space.

We retrain after each 50 iterations.



Objective values for data sampled from the conv-VAE after 500 function evaluations.

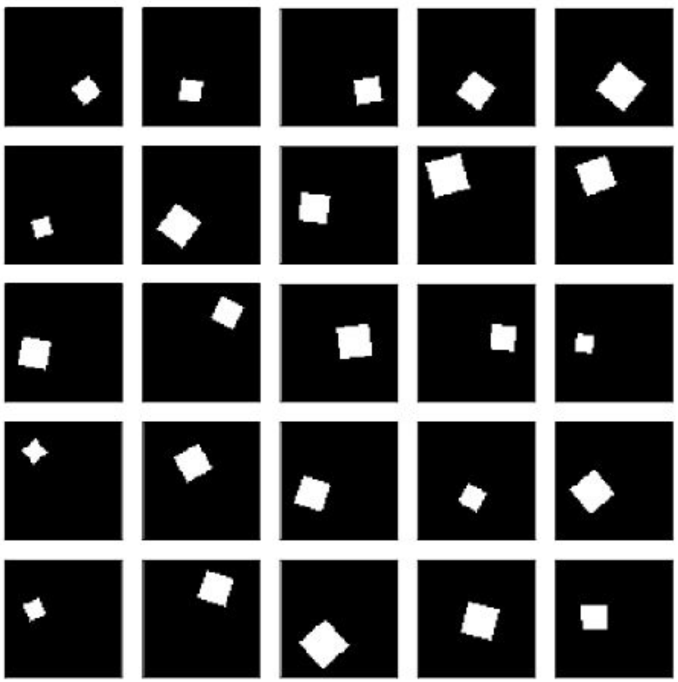


2D shape area maximization

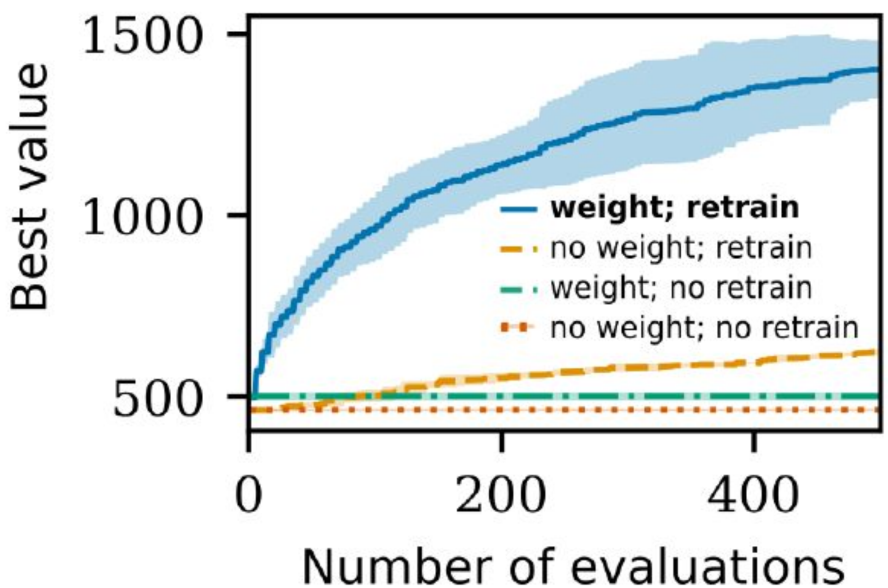
We train a conv-VAE on the dSprites dataset with 245,760 **square shapes**.

Goal: generate shape with **max. area**. Exhaustive search in 5D latent space.

We retrain after each 50 iterations.



Best objective value of the shapes generated during the optimization process, $k = 10^{-3}$.

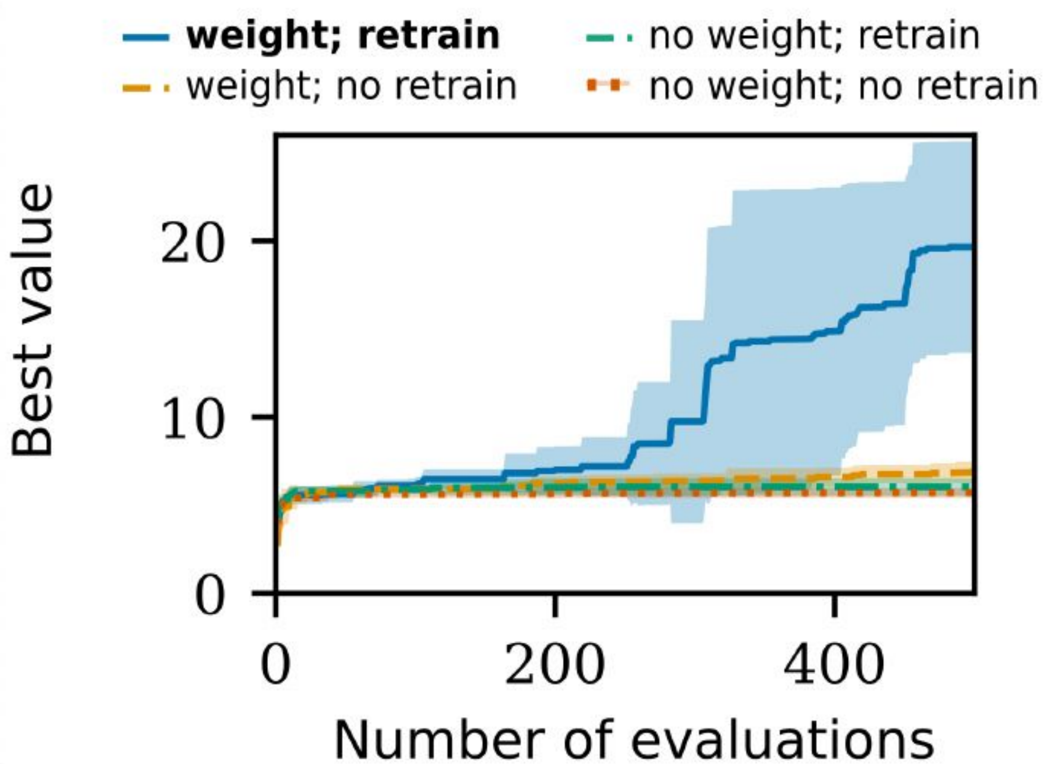


Chemical design task

Optimize the water-octanol partition coefficient (logP).

Initial data: 250,000 molecules from ZINC database.

We use $k = 10^{-3}$ and a **junction tree** VAE with **sparse GPs**.



Take home messages

Weighted retraining is an approach for latent space optimization that

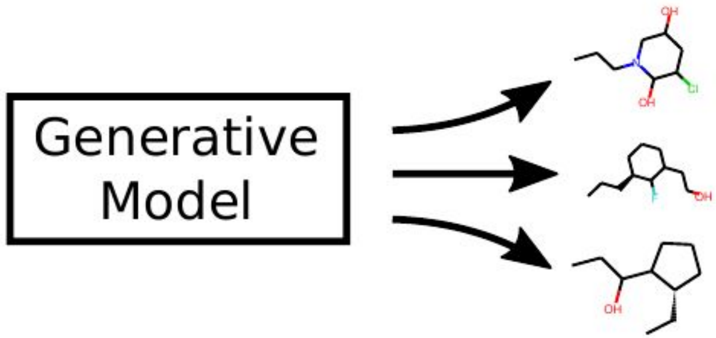
1. results in **latent spaces** that are highly **useful for optimization**.
2. is conceptually **simple and easy** to implement.
3. significantly improves **data efficiency and final performance**.

Deep Generative Models of Molecules via Chemical Reactions

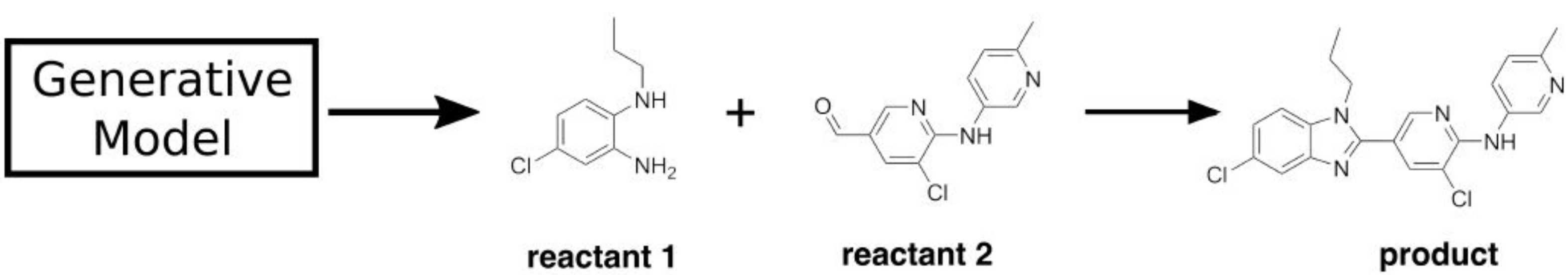
John Bradshaw, Brooks Paige, Matt J. Kusner, Marwin Segler and José Miguel Hernández-Lobato,
Barking up the right tree: an approach to search over molecule synthesis DAGs,
In NeurIPS 2020.

Limitations of generative models of molecules

Most models will not tell you how to **synthesize** the generated molecules:



We want models that generate molecules via **chemical reactions**, like humans!



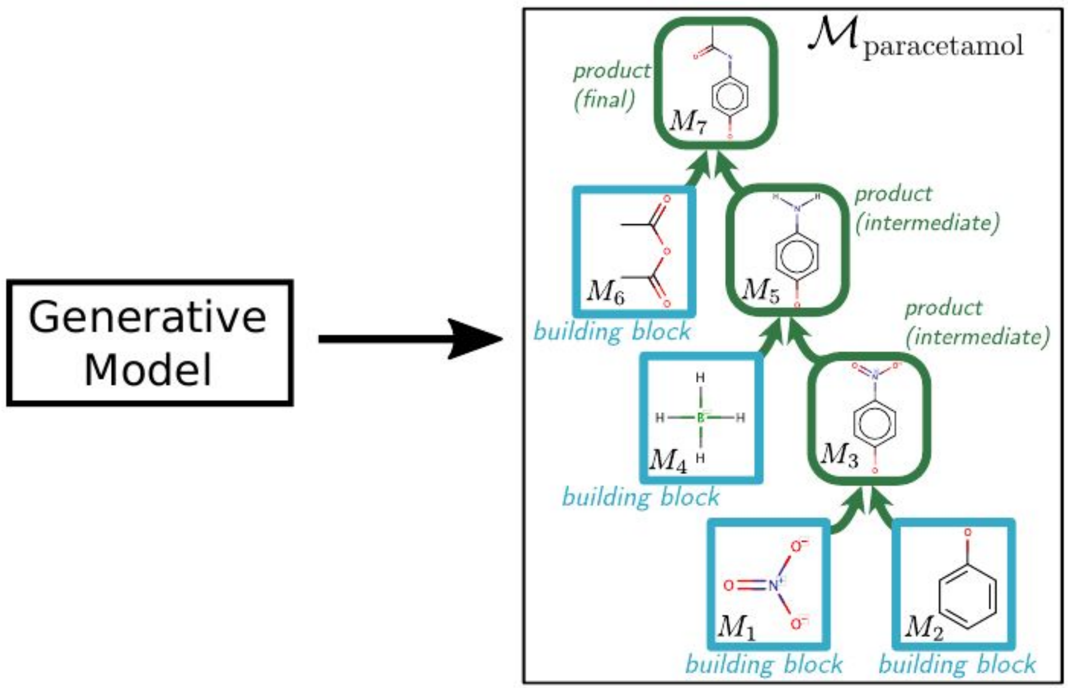
Advantages:

- Generated molecules more likely to be synthesizable in practice.
- Synthesis information is available right away.
- Generated molecules exhibit more realistic properties (e.g. improved stability).

Proposed approach

A multi-step reaction can be encoded as a **directed acyclic graph (DAG)**.

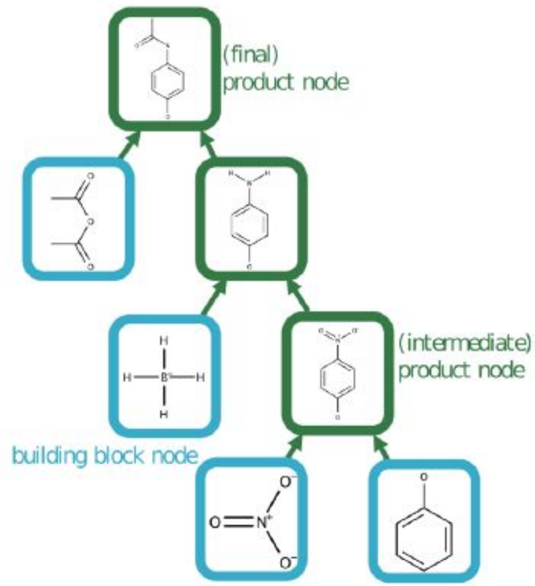
We generate the whole molecule synthesis DAG:



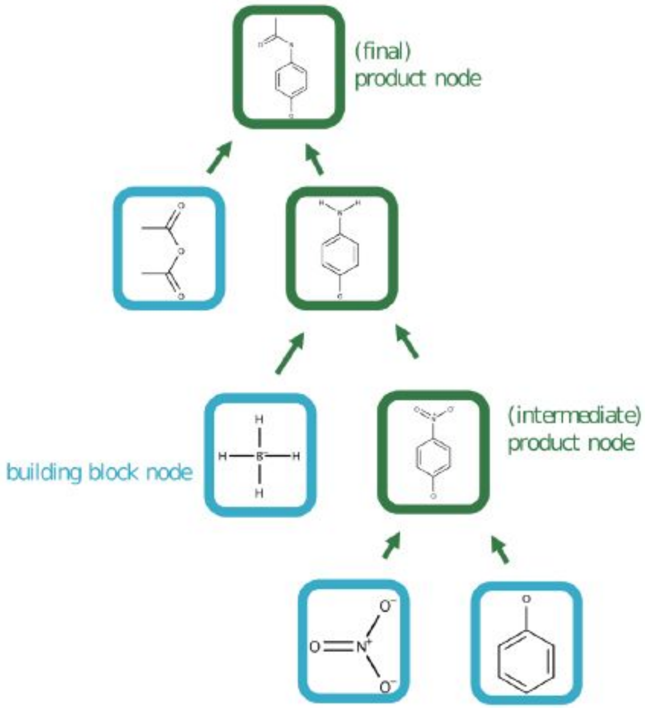
We call our model **DoG** since it generates DAGs Of (molecular) Graphs (DoGs).

We use as training data 72,008 synthesis DAGs obtained from the **USPTO dataset**.

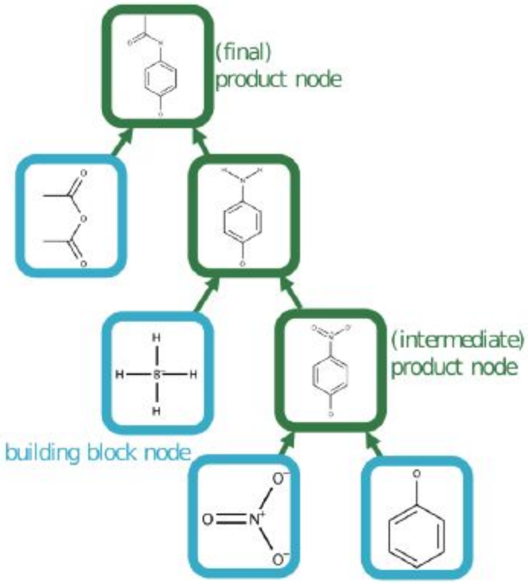
Serializing molecule synthesis DAGs



Serializing molecule synthesis DAGs



Serializing molecule synthesis DAGs

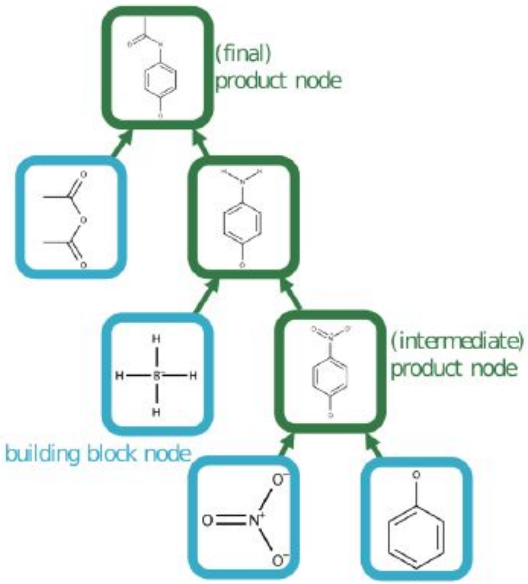


Target DAG

DAG so far

Actions

Serializing molecule synthesis DAGs



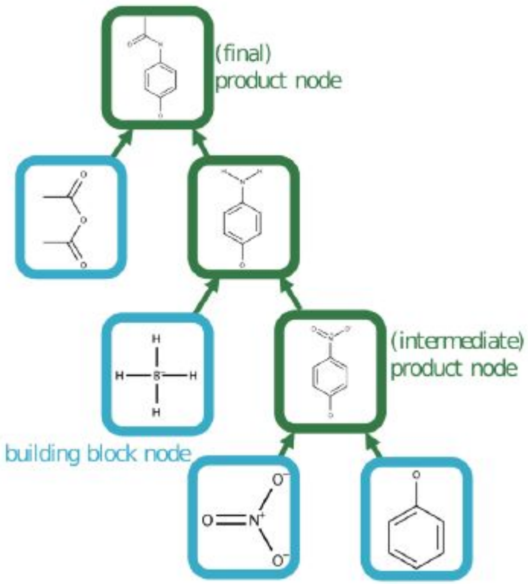
Target DAG

DAG so far

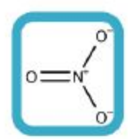
'B'

Actions

Serializing molecule synthesis DAGs



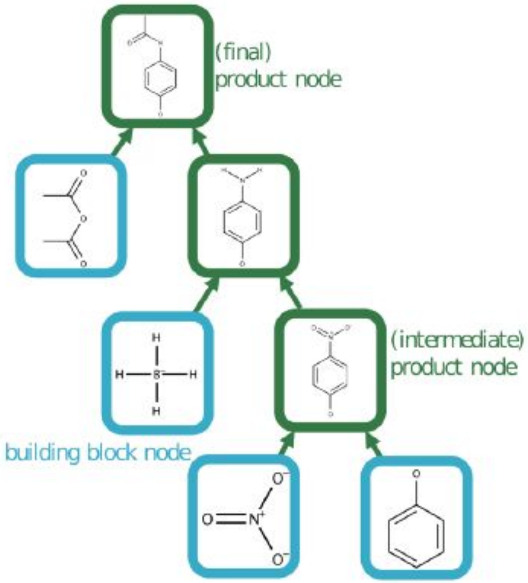
Target DAG

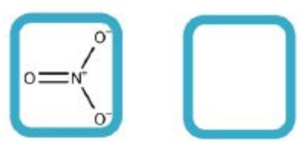


DAG so far

Actions

Serializing molecule synthesis DAGs



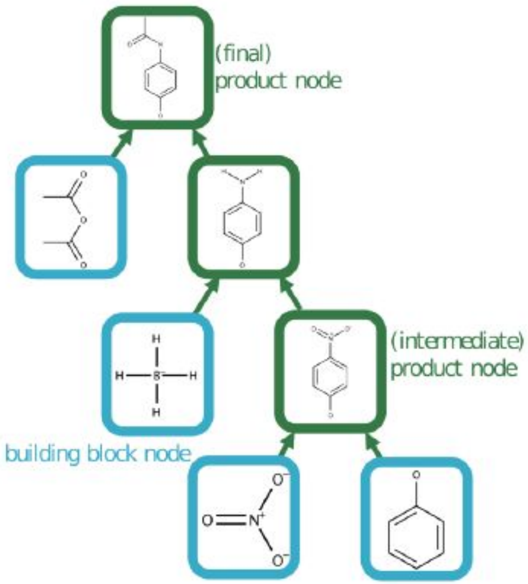


Target DAG

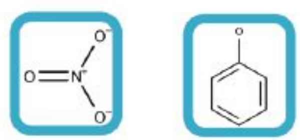
DAG so far

Actions

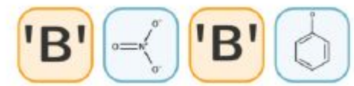
Serializing molecule synthesis DAGs



Target DAG

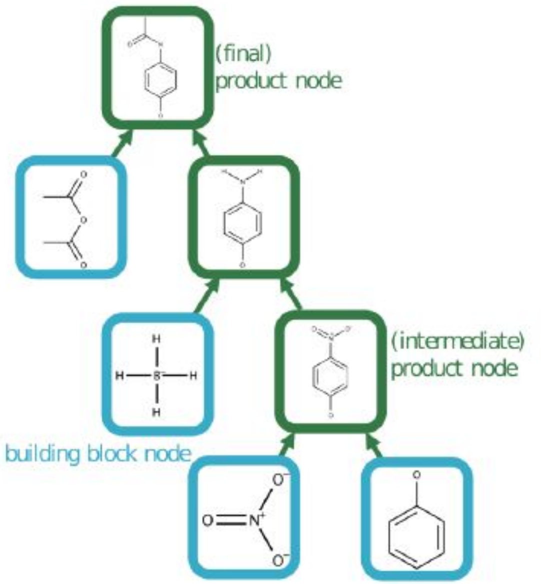


DAG so far

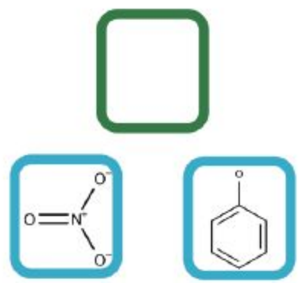


Actions

Serializing molecule synthesis DAGs



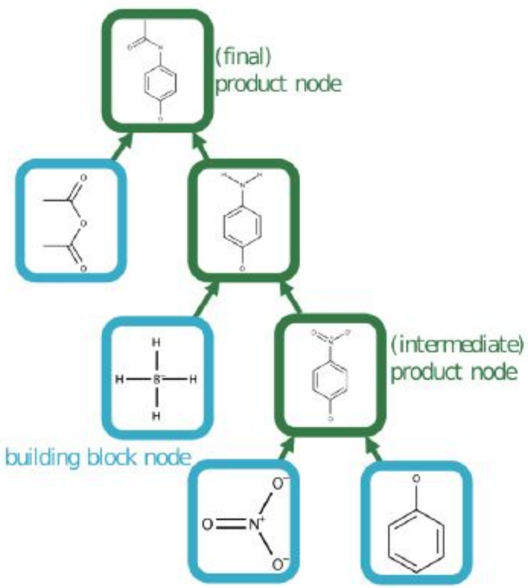
Target DAG



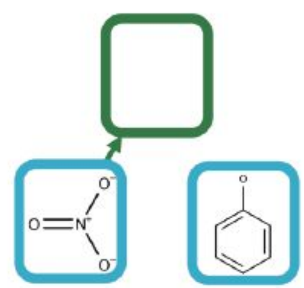
DAG so far

Actions

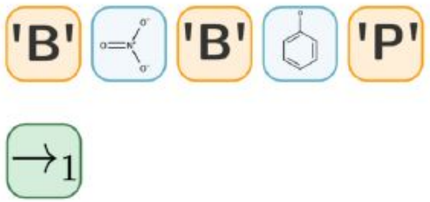
Serializing molecule synthesis DAGs



Target DAG

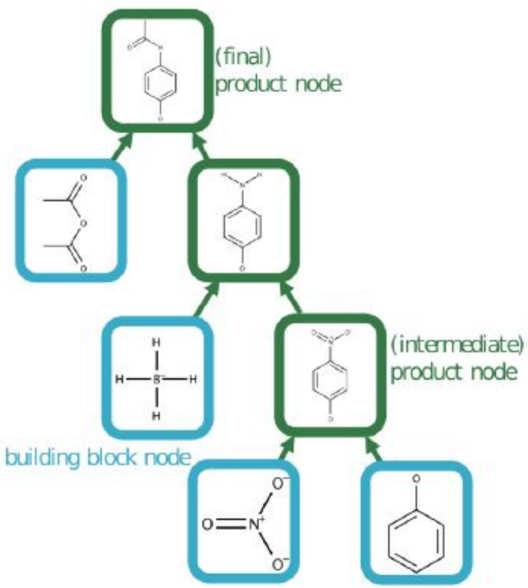


DAG so far

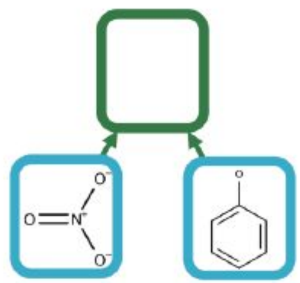


Actions

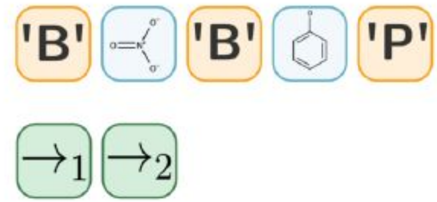
Serializing molecule synthesis DAGs



Target DAG

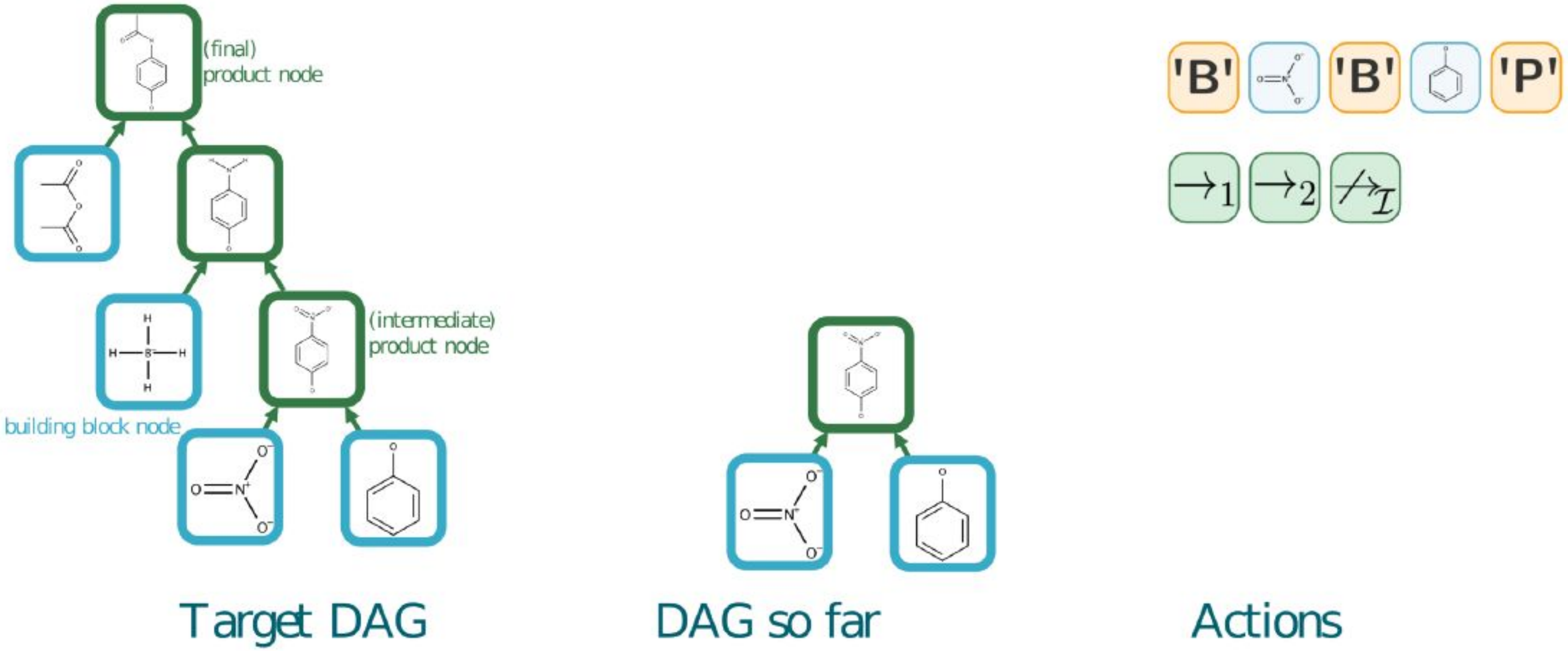


DAG so far

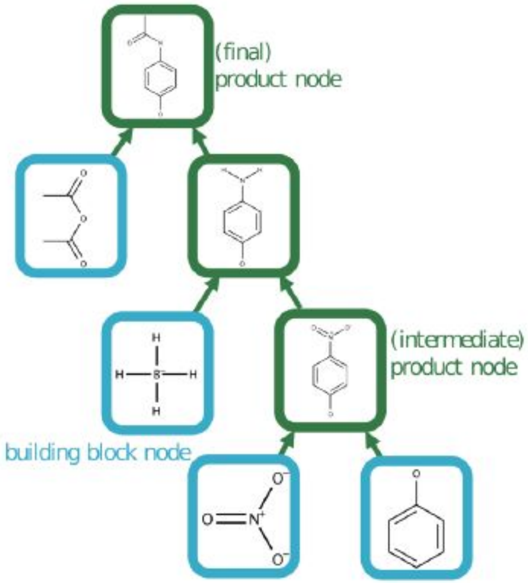


Actions

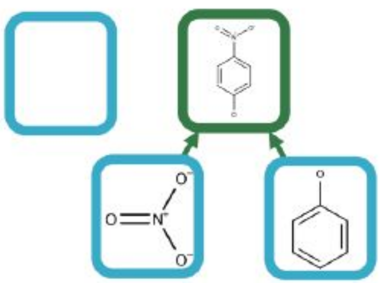
Serializing molecule synthesis DAGs



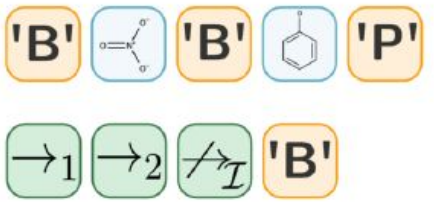
Serializing molecule synthesis DAGs



Target DAG

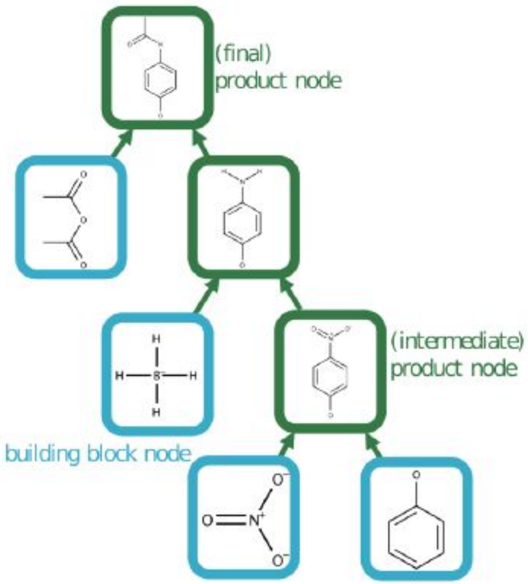


DAG so far

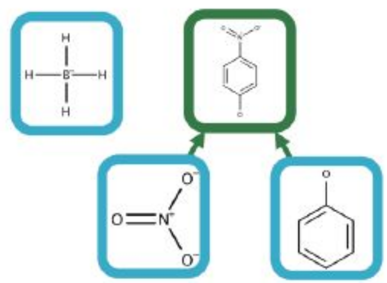


Actions

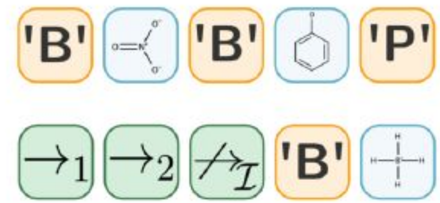
Serializing molecule synthesis DAGs



Target DAG

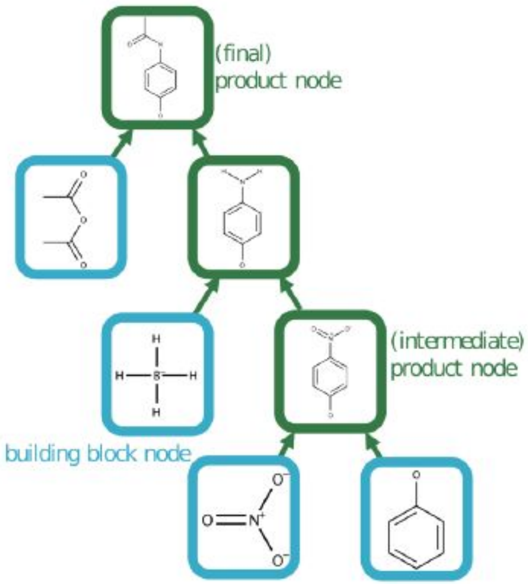


DAG so far

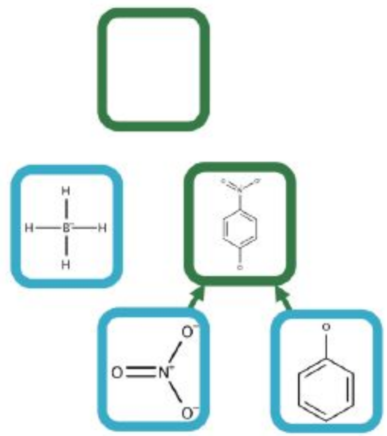


Actions

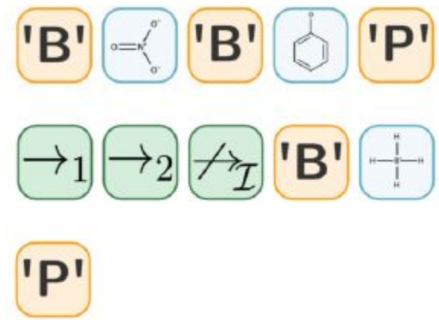
Serializing molecule synthesis DAGs



Target DAG

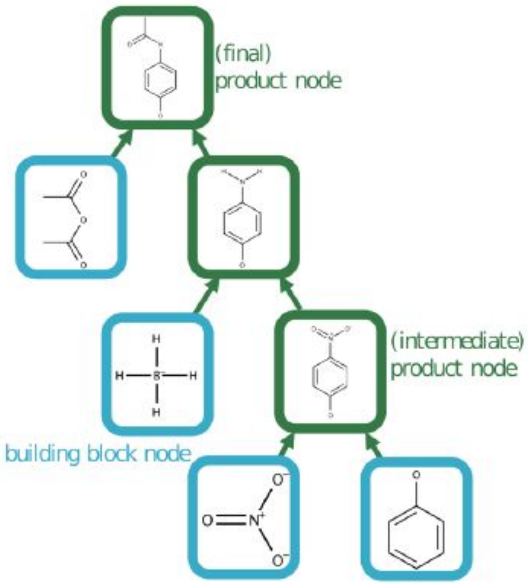


DAG so far

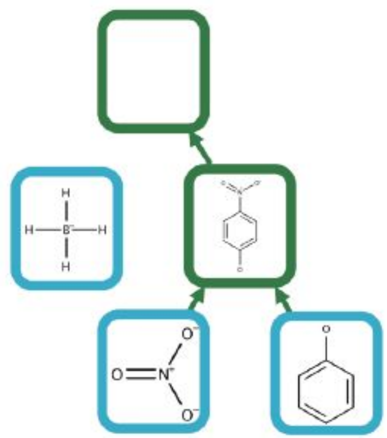


Actions

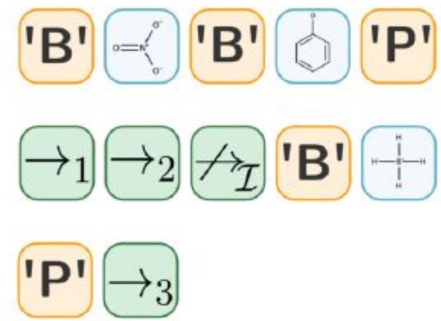
Serializing molecule synthesis DAGs



Target DAG

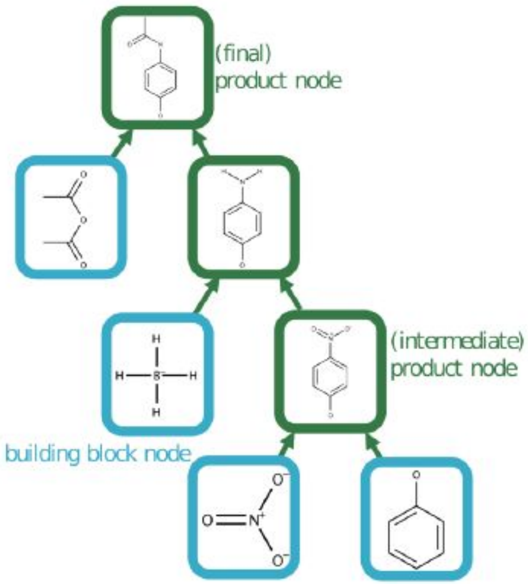


DAG so far

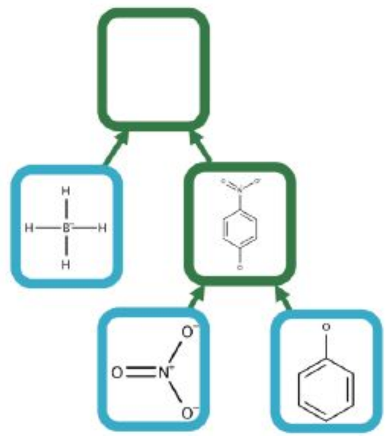


Actions

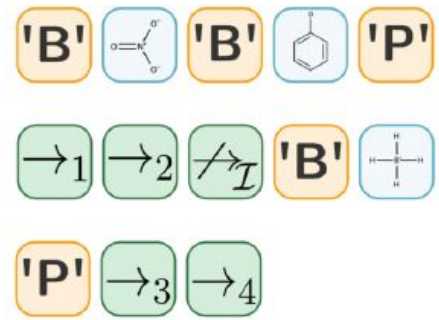
Serializing molecule synthesis DAGs



Target DAG

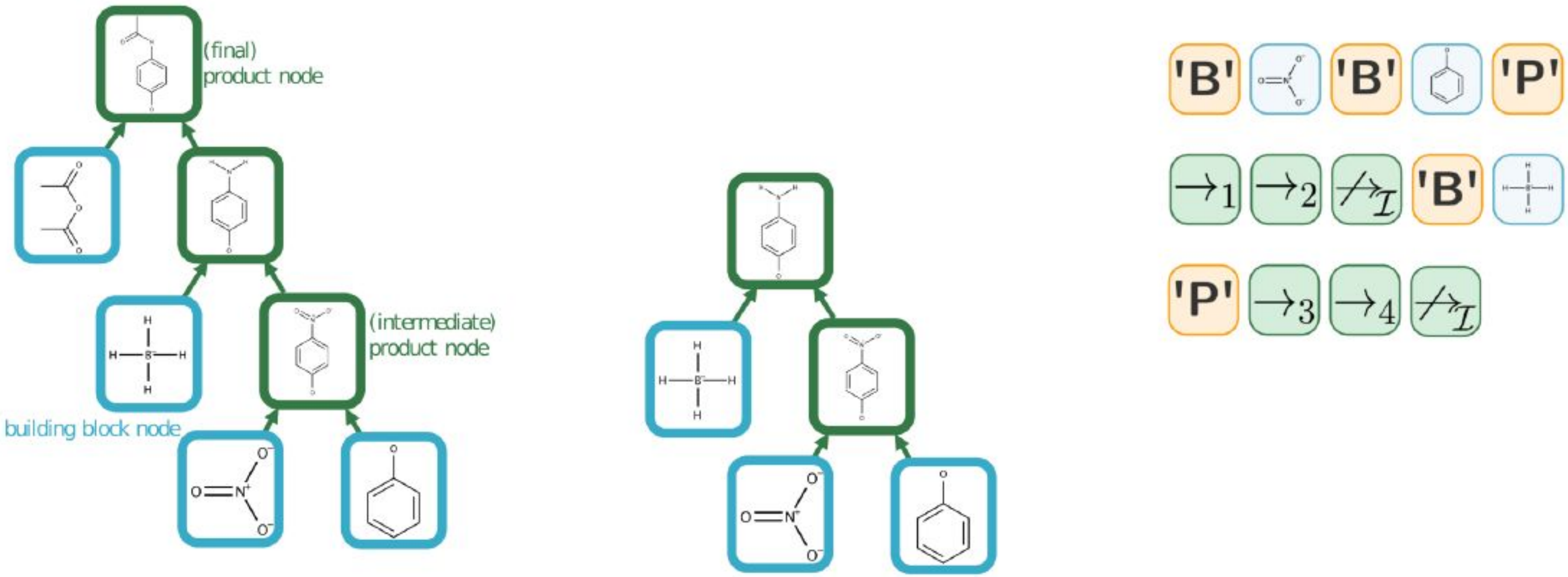


DAG so far



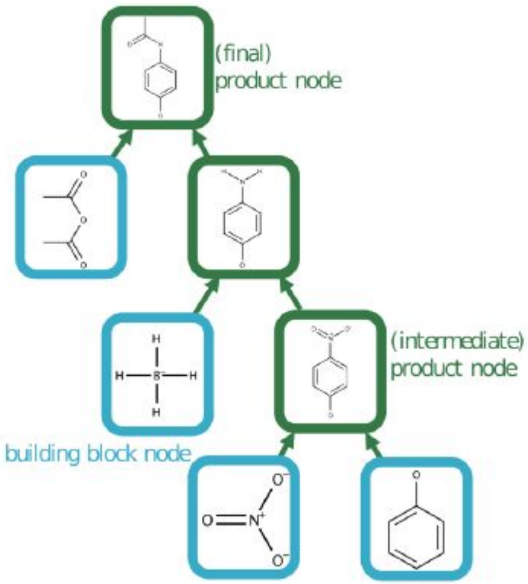
Actions

Serializing molecule synthesis DAGs

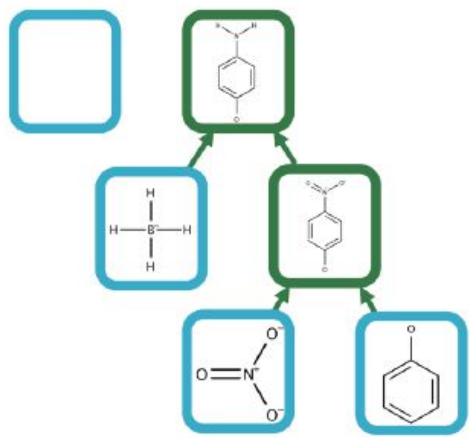


Target DAG | DAG so far | Actions

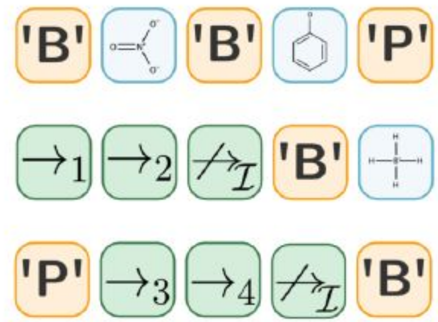
Serializing molecule synthesis DAGs



Target DAG

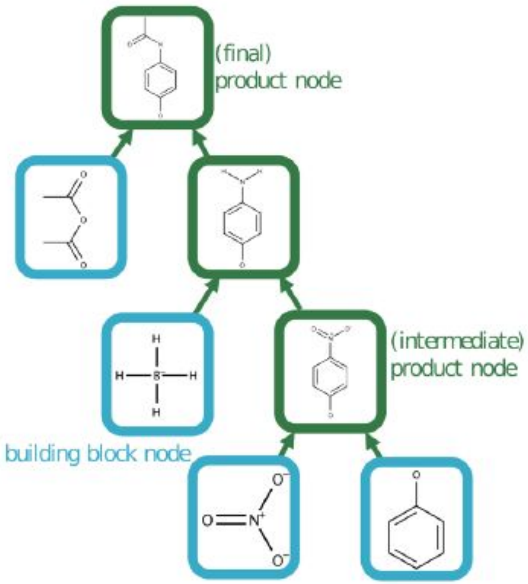


DAG so far

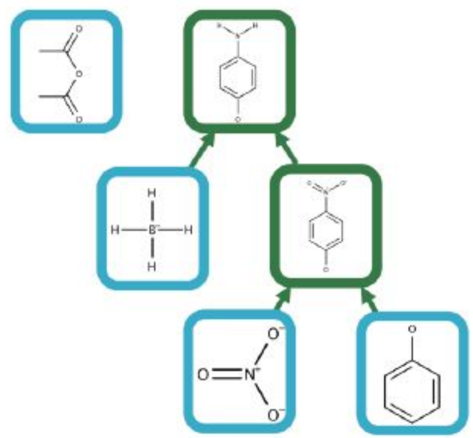


Actions

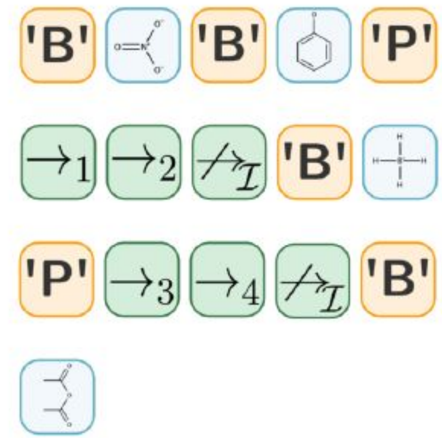
Serializing molecule synthesis DAGs



Target DAG

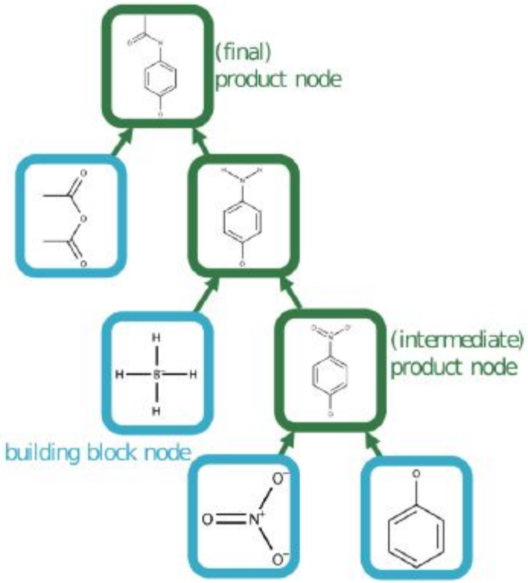


DAG so far

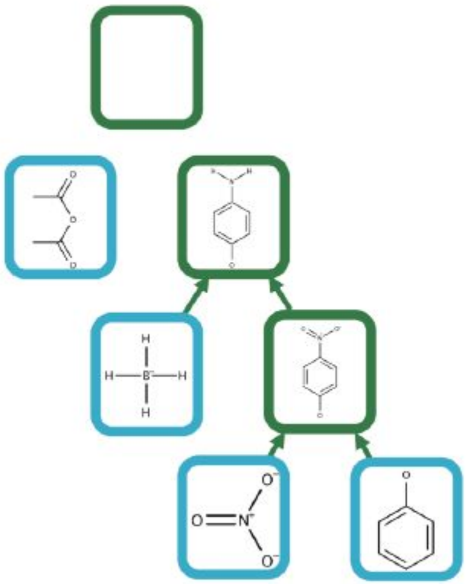


Actions

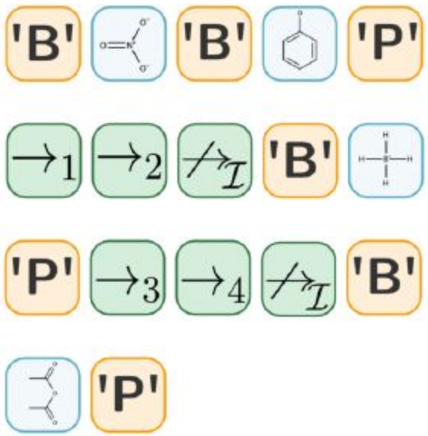
Serializing molecule synthesis DAGs



Target DAG

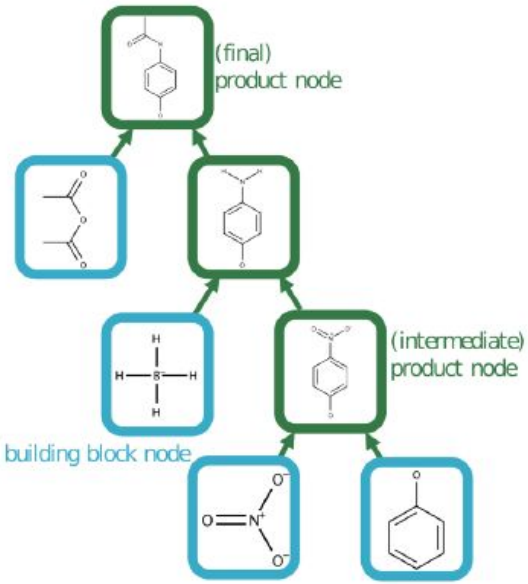


DAG so far

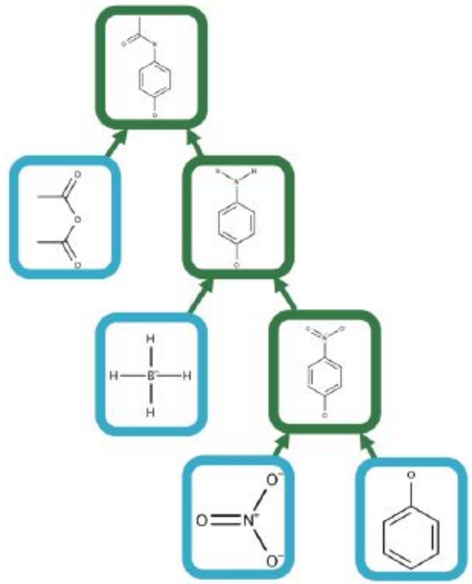


Actions

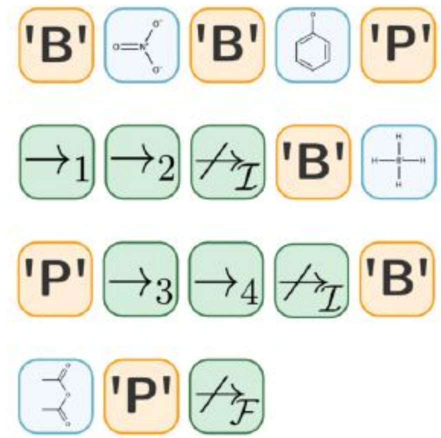
Serializing molecule synthesis DAGs



Target DAG



DAG so far



Actions

Modeling the sequential generation of DAGs

An **autoregressive model** describes the sequential generation of the syntehsis DAG:

$$p_\theta(\mathrm{DAG}) = \prod_{t=1}^{T} p_\theta(\mathrm{action}_t | \mathrm{action}_{t-1}, \mathrm{action}_{t-2}, \ldots) \,.$$

Modeling the sequential generation of DAGs

An **autoregressive model** describes the sequential generation of the syntehsis DAG:

$$p_\theta(\mathrm{DAG}) = \prod_{t=1}^{T} p_\theta(\mathrm{action}_t | \mathrm{action}_{t-1}, \mathrm{action}_{t-2}, \ldots) \, .$$

Recurrent neural networks (RNNs) are used to condition on past actions.

Molecule embeddings are generated using **graph neural networks**.

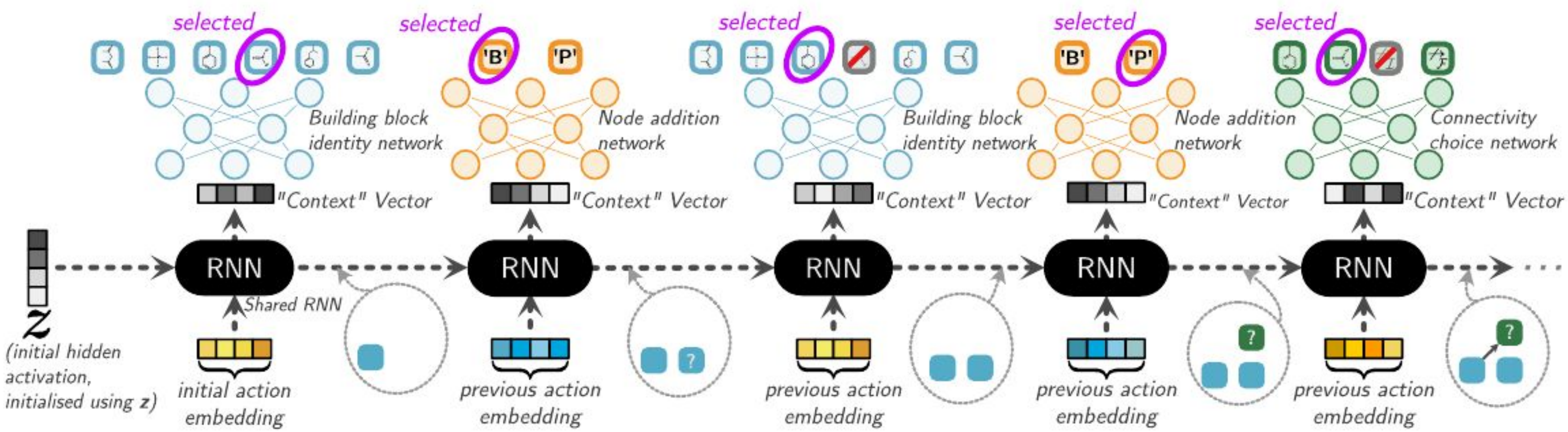
Modeling the sequential generation of DAGs

An **autoregressive model** describes the sequential generation of the syntehsis DAG:

$$p_\theta(\mathrm{DAG}) = \prod_{t=1}^{T} p_\theta(\mathrm{action}_t | \mathrm{action}_{t-1}, \mathrm{action}_{t-2}, \ldots)\,.$$

Recurrent neural networks (RNNs) are used to condition on past actions.

Molecule embeddings are generated using **graph neural networks**.



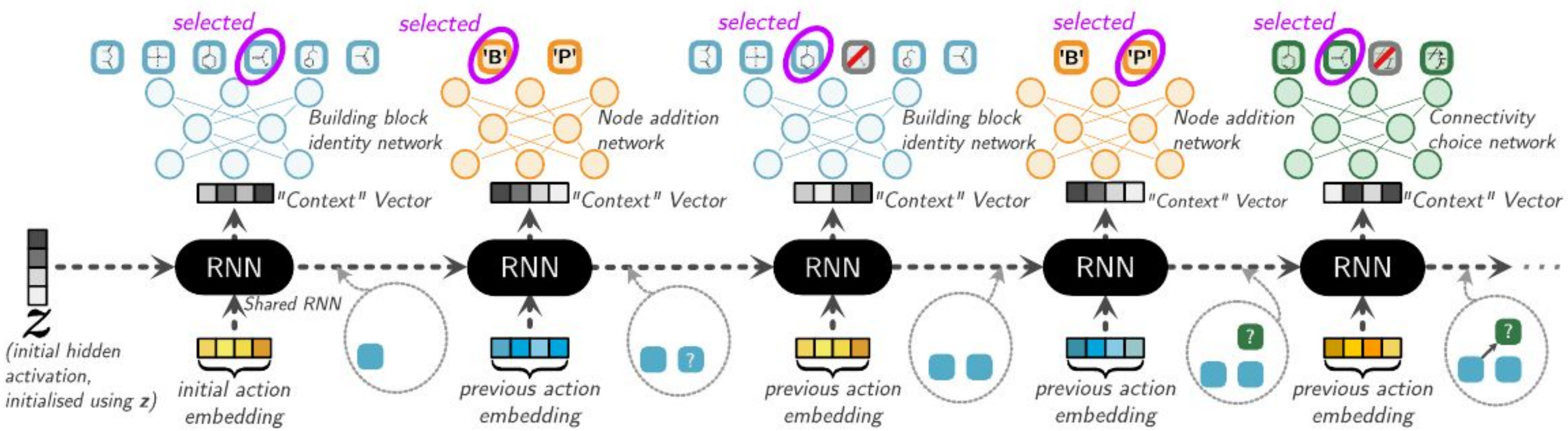
Modeling the sequential generation of DAGs

An **autoregressive model** describes the sequential generation of the syntehsis DAG:

$$p_\theta(\mathrm{DAG}) = \prod_{t=1}^{T} p_\theta(\mathrm{action}_t | \mathrm{action}_{t-1}, \mathrm{action}_{t-2}, \ldots)\,.$$

Recurrent neural networks (RNNs) are used to condition on past actions.

Molecule embeddings are generated using **graph neural networks**.



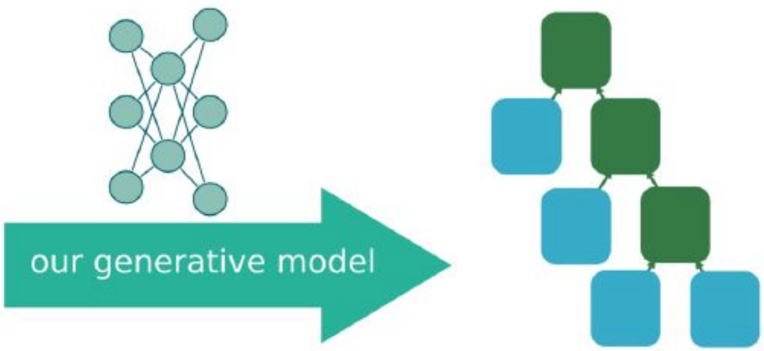
Content of product nodes generated by calling the **molecular transformer** [1].

[1] Schwaller et al. ACS Central Science, 2019.

Two deep generative approaches

1. **DoG-Gen**, an autoregressive model

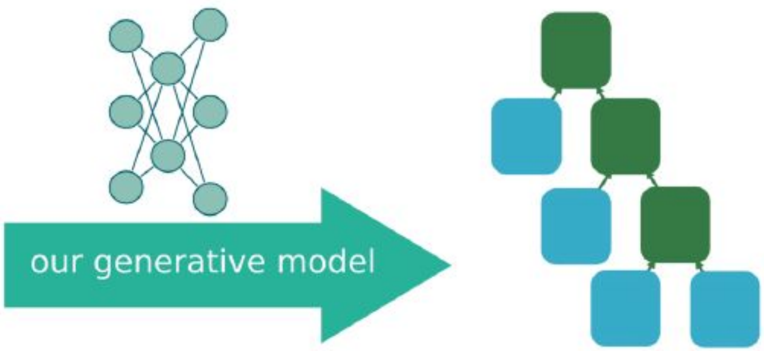
   Can still be used for optimization
   (e.g. cross entropy method) [1,2]

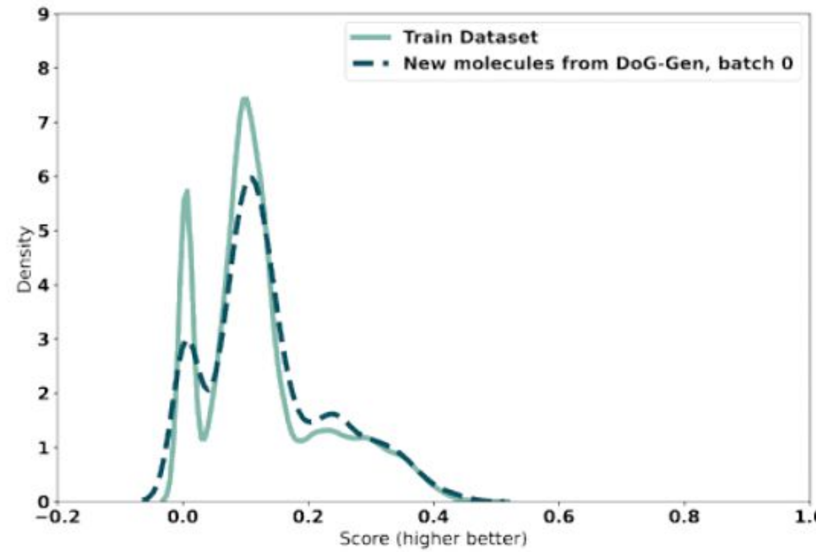


Two deep generative approaches

1. **DoG-Gen**, an autoregressive model

Can still be used for optimization (e.g. cross entropy method) [1,2]

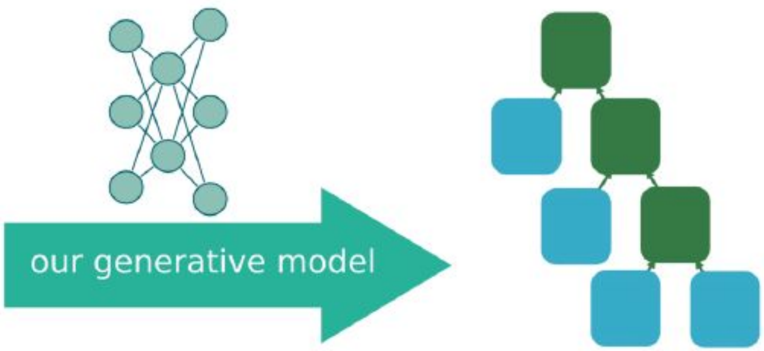


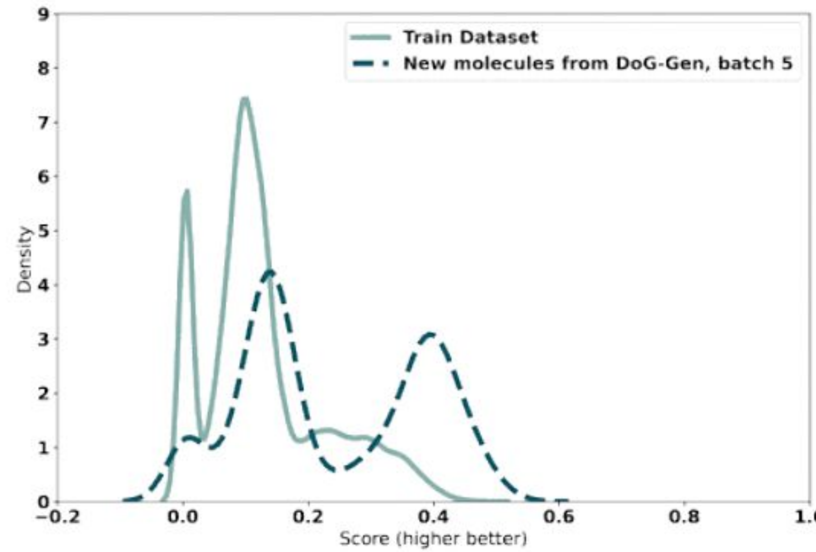


Two deep generative approaches

1. **DoG-Gen**, an autoregressive model

Can still be used for optimization
(e.g. cross entropy method) [1,2]

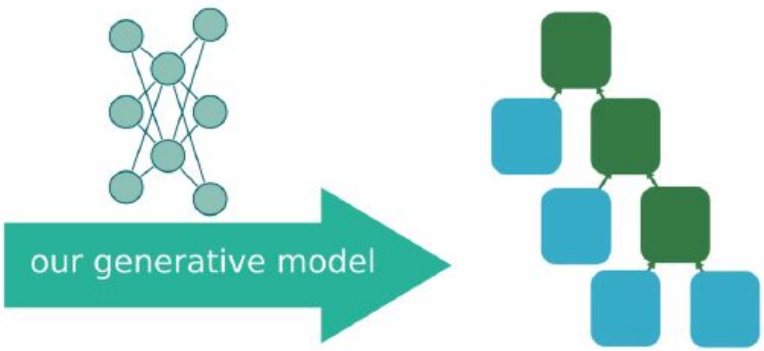


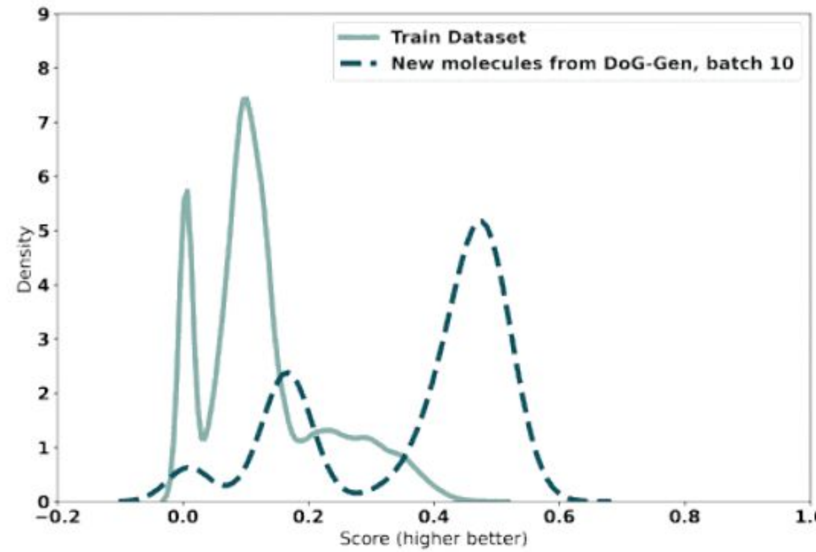


Two deep generative approaches

1. **DoG-Gen**, an autoregressive model

Can still be used for optimization (e.g. cross entropy method) [1,2]

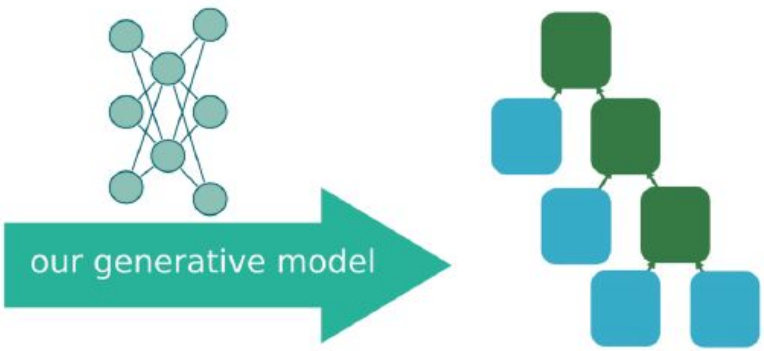


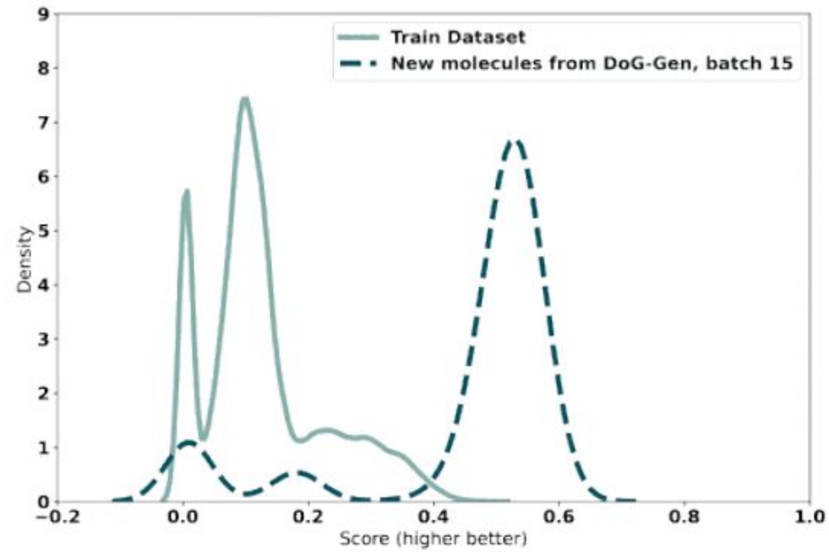


Two deep generative approaches

1. **DoG-Gen**, an autoregressive model

Can still be used for optimization (e.g. cross entropy method) [1,2]

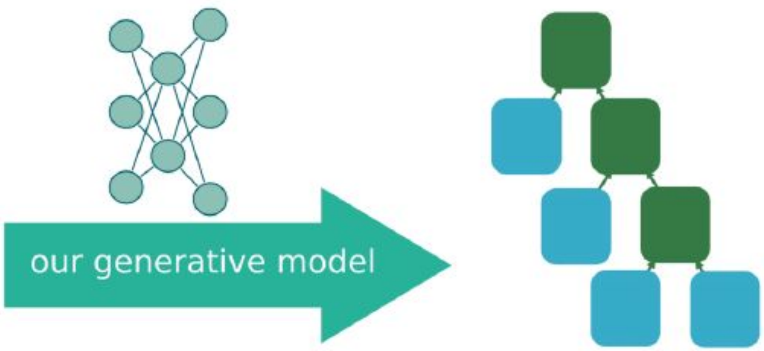


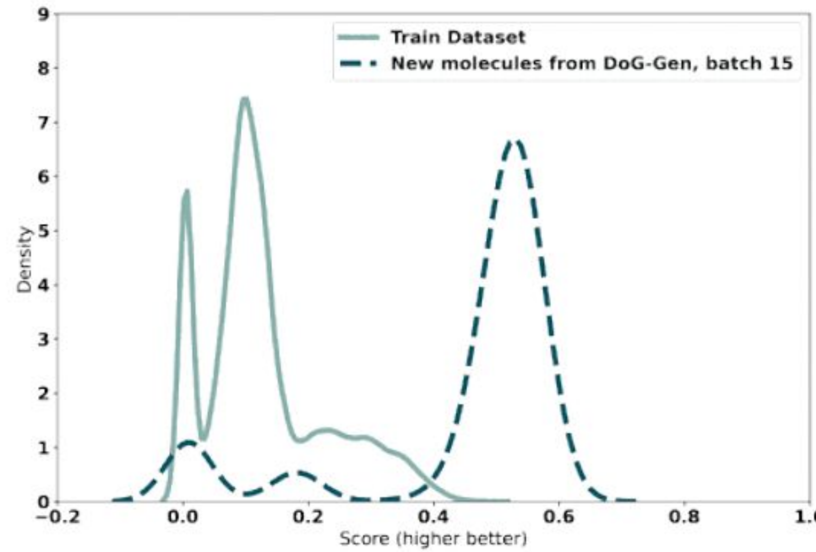


Two deep generative approaches

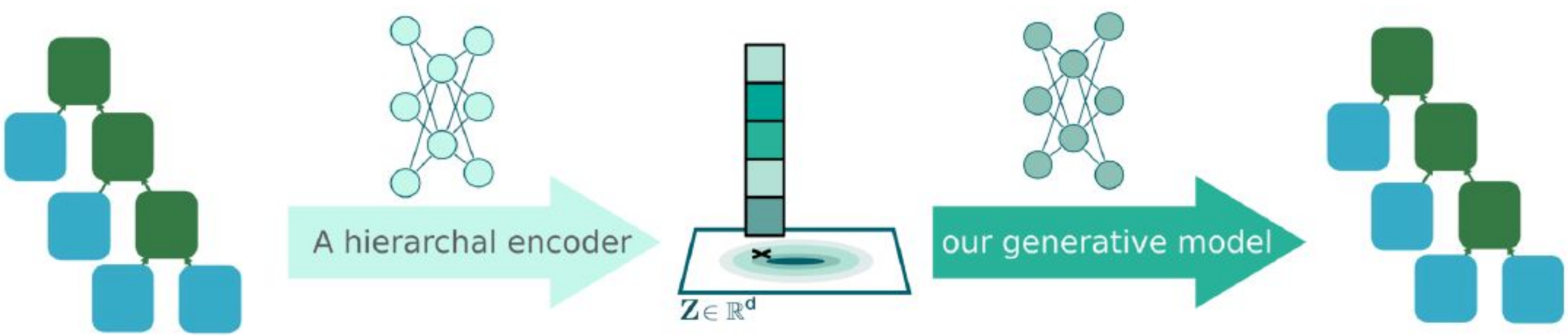
1. **DoG-Gen**, an autoregressive model

Can still be used for optimization
(e.g. cross entropy method) [1,2]



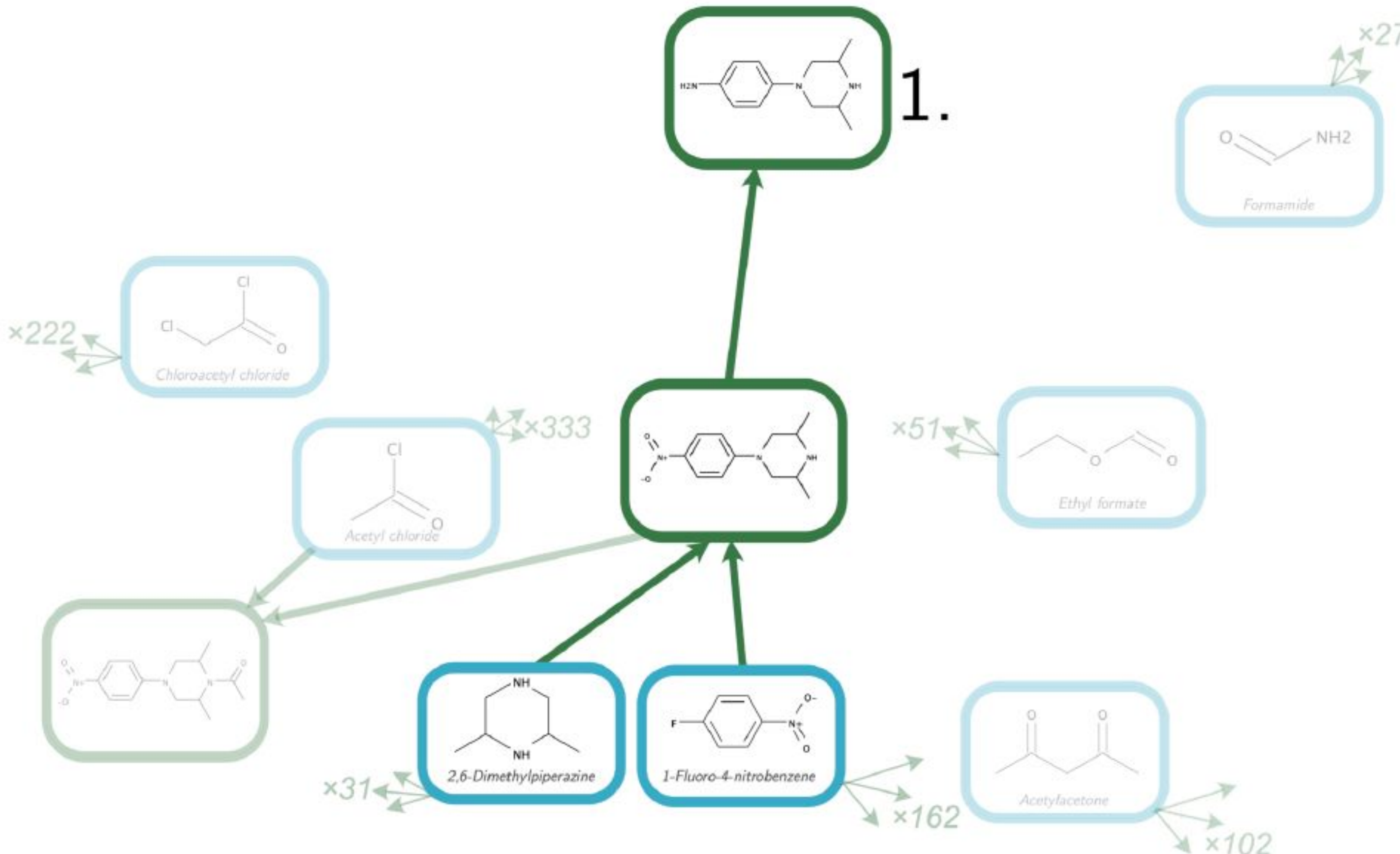


2. **DoG-AE**, an autoencoder model

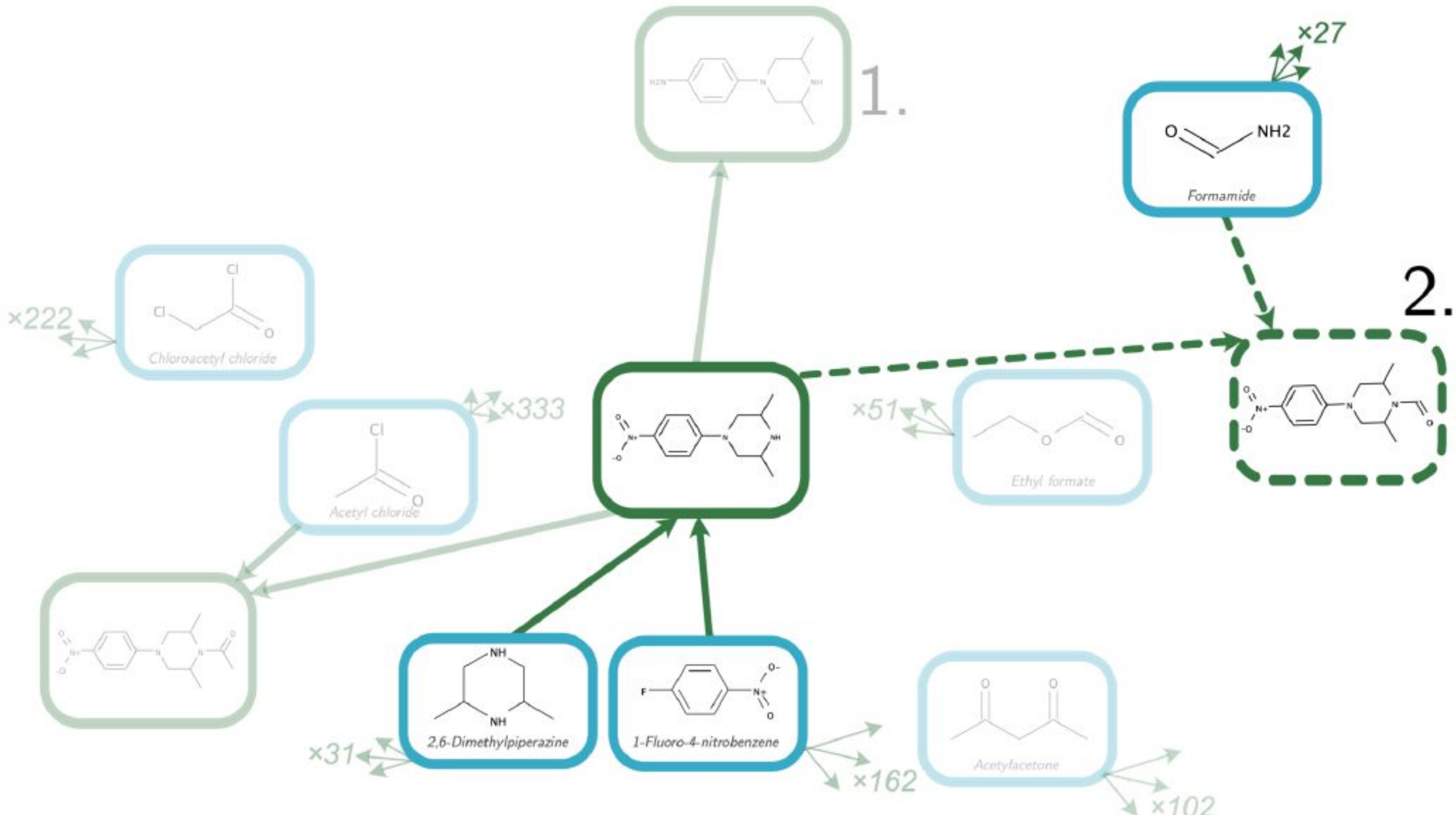


[1] (Segler et al., 2017) doi.org/10.1021/acscentsci.7b00512 [2] (de Boer et al., 2005) doi.org/10.1007/s10479-005-5724-z

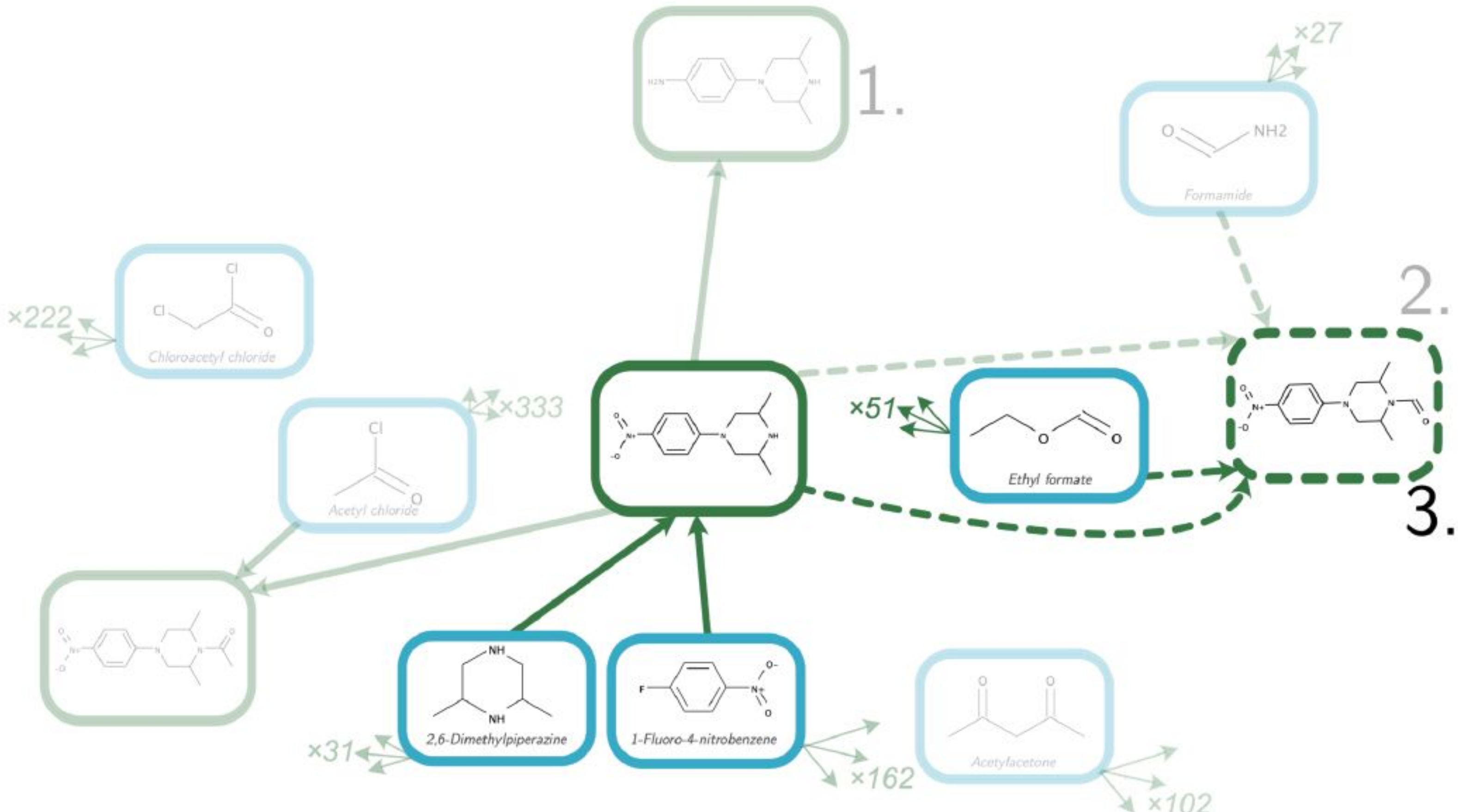
Latent space interpolation



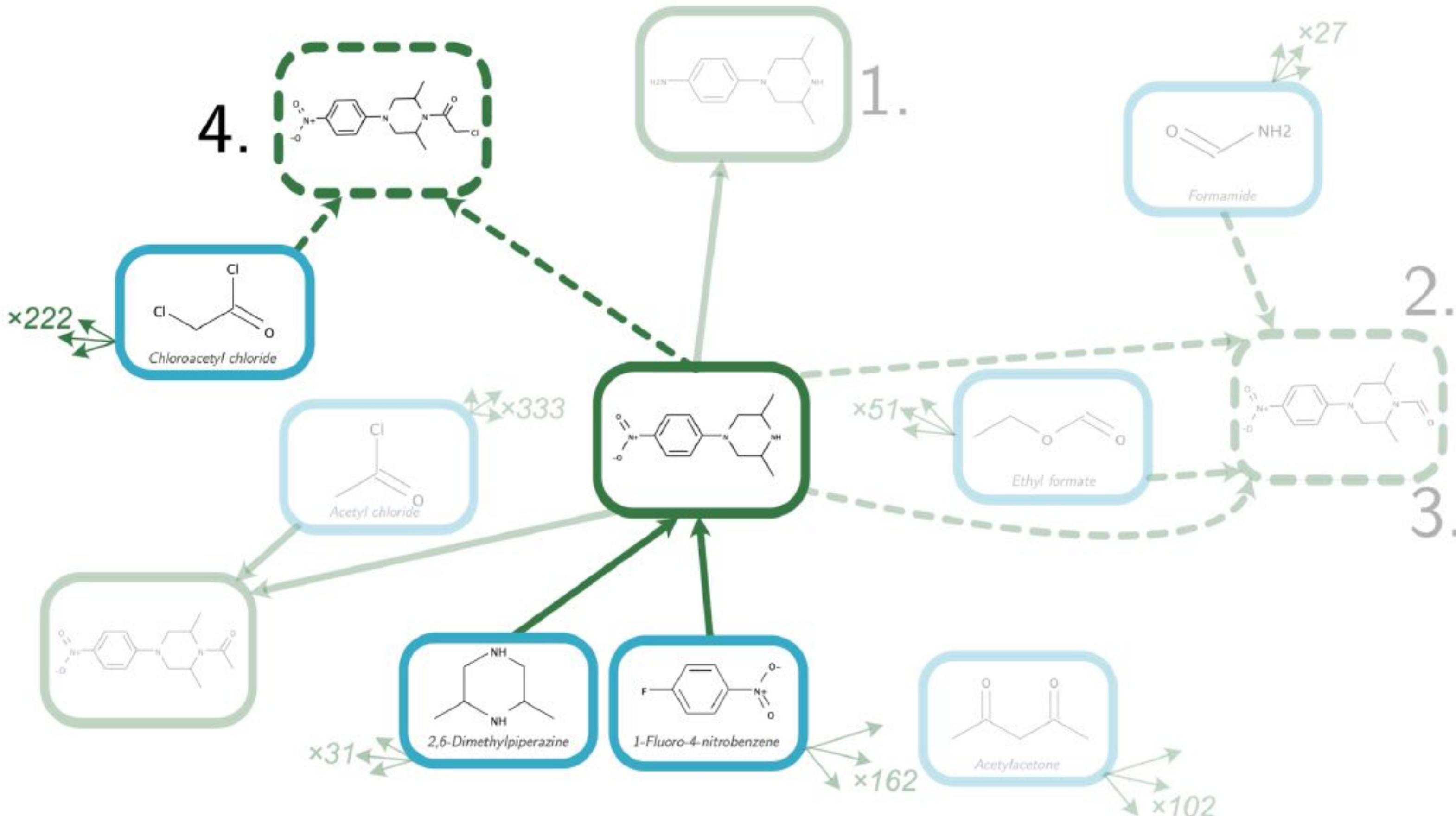
Latent space interpolation



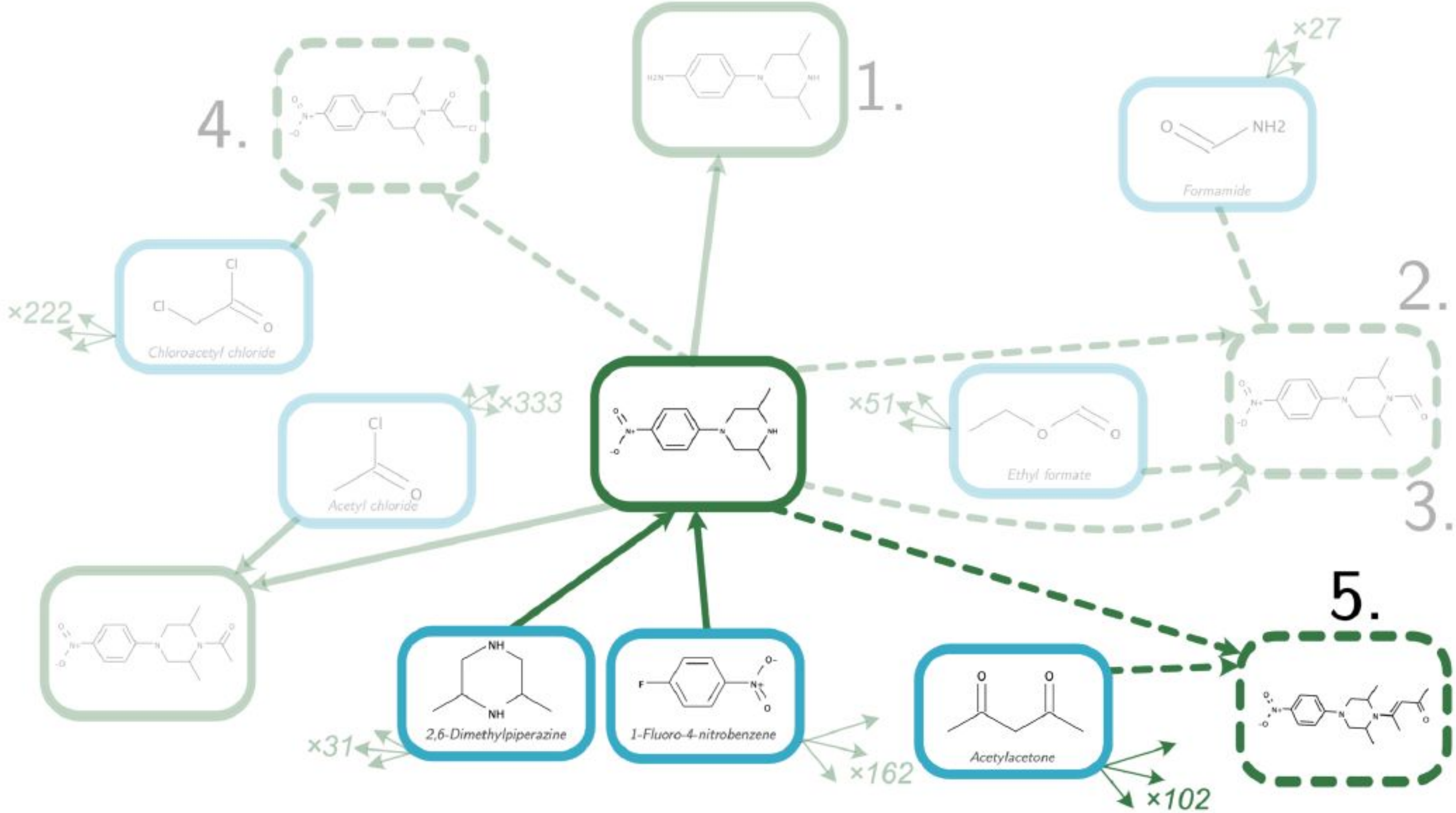
Latent space interpolation



Latent space interpolation

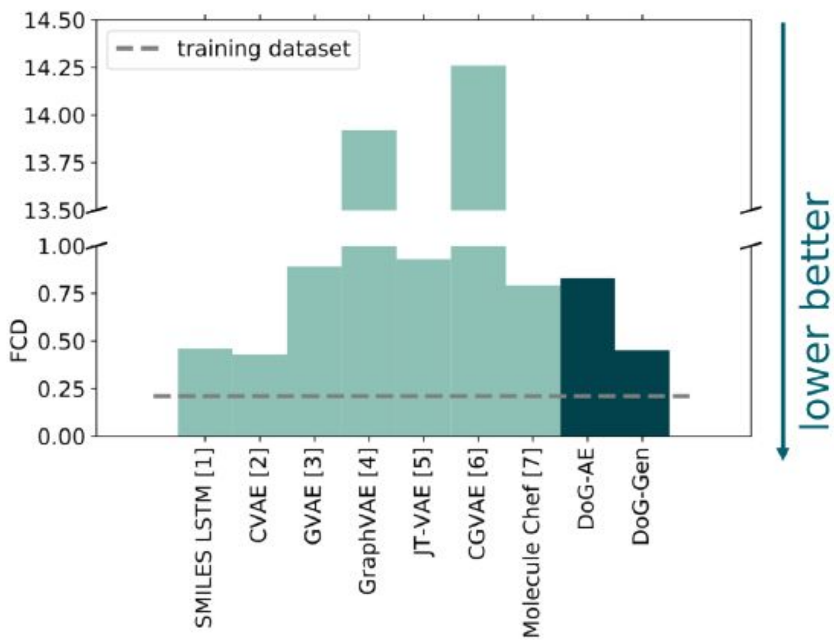


Latent space interpolation

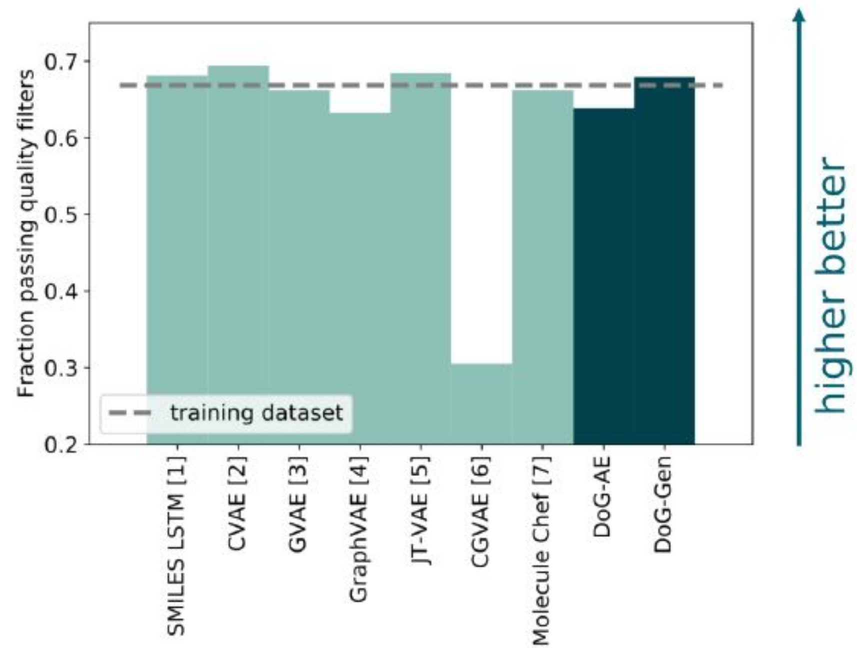


How good are our generated molecules?

With our model, **constrained to generate synthesizable molecules**, we match the quality and space coverage of many previous, **more flexible**, generative models:



Fréchet ChemNet Distance [9]

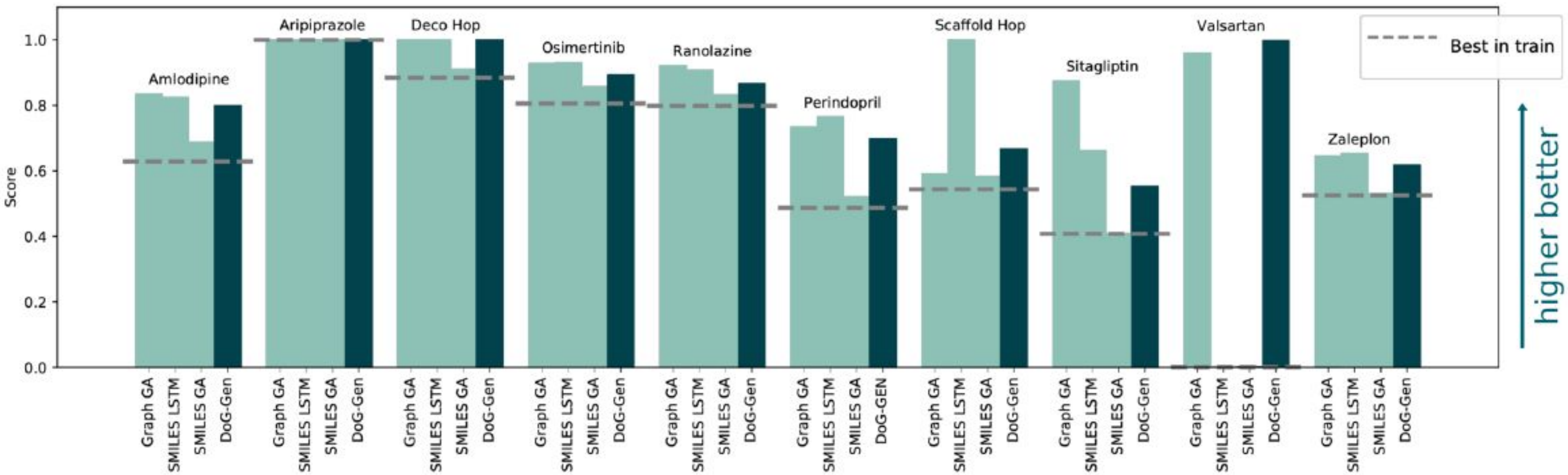


Quality filters (assess stability) [8]

[1] (Segler et al., 2017) doi.org/10.1021/acscentsci.7b00512; [2] (Gómez-Bombarelli et al., 2018) doi.org/10.1021/acscentsci.7b00572; [3] (Kusner et al., 2017) proceedings.mlr.press/v70/kusner17a; [4] (Simonovsky and Komodakis, 2018) arxiv.org/abs/1802.03480; [5] (Jin et al., 2018) proceedings.mlr.press/v80/jin18a.html; [6] (Liu et al., 2018) arxiv.org/abs/1805.09076; [7] (Bradshaw et al., 2019b) arxiv.org/abs/1906.05221; [8] (Brown et al, 2019) doi.org/10.1021/acs.jcim.8b00839; [9] (Preuer et al, 2018) doi.org/10.1021/acs.jcim.8b00234

How good is our molecule optimization?

We obtain competitive results when compared to strong baselines that **do not consider synthesizability**:

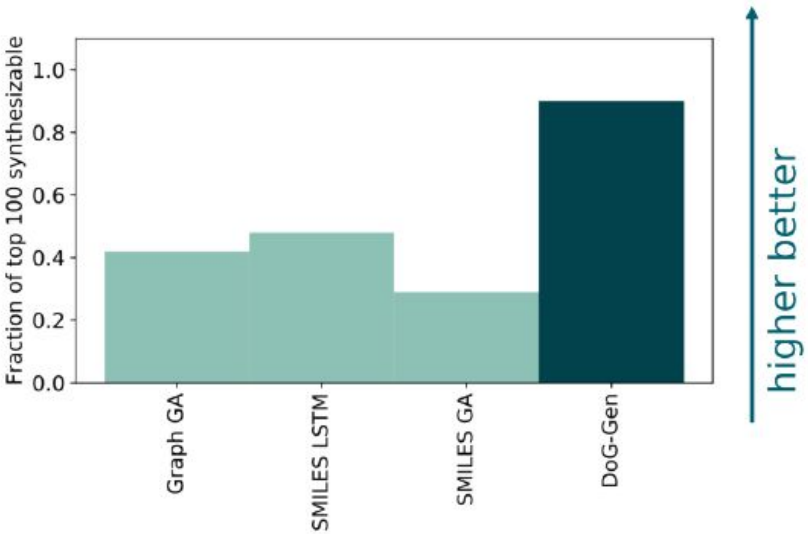


10 GuacaMol tasks [1]

[1] (Brown et al, 2019) doi.org/10.1021/acs.jcim.8b00839

How good are our optimized molecules?

We suggest molecules that are **synthesizable** and that **pass the quality checks**:



Plot showing fraction synthesizable as judged by running through retrosynthetic pipeline [1,2]

Plot showing fraction of optimized molecules passing quality filters (assess stability) [3]

[1] (Segler et al, 2018) nature.com/articles/nature25978 ; [2] (Gao and Coley, 2020) doi.org/10.1021/acs.jcim.0c00174 ; [3] (Brown et al, 2019) doi.org/10.1021/acs.jcim.8b00839

Take home messages

DoG is a deep generative model of molecules that

1. generates **synthesizable** molecules up-front.
2. works by generating **molecular synthesis DAGs** in a serial way.
3. obtains **competitive results** in molecule generation and optimization.
4. produces optimal molecules with higher degrees of **stability and synthesizability**.

Collaborators

Austin Tripp

Erik Daxberger

John Bradshaw

Matt Kusner

Brooks Paige

Marwin Segler

Thanks!

Questions?