Compound representation, ADMET profiles and automatic optimization

Deep learning for computational chemistry

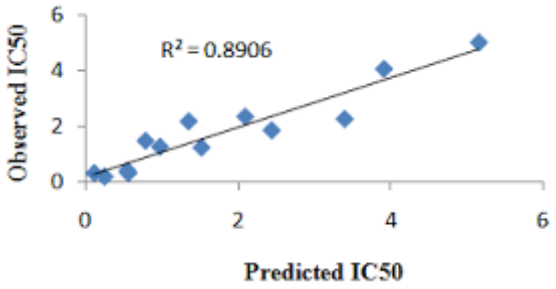
Floriane Montanari

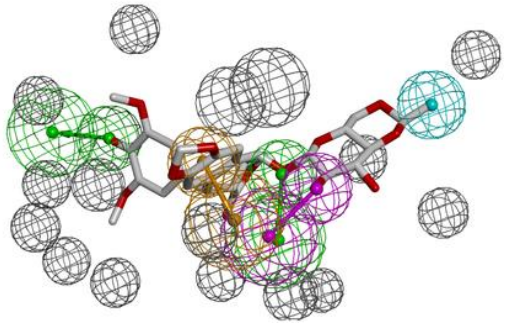
ICGEB-TRAIN 15.05.2019 Bled, Slovenia

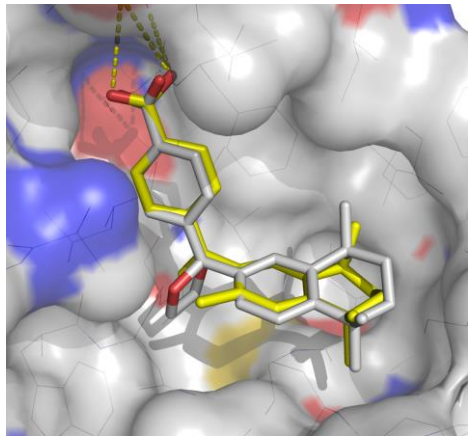
Introduction

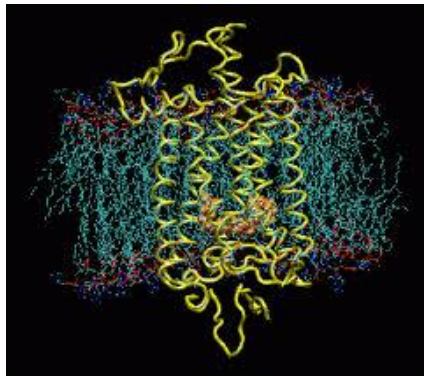
Computational chemistry

HTS screen **Hit to lead** **Lead optimization**





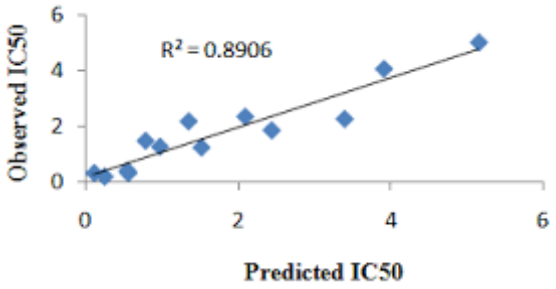


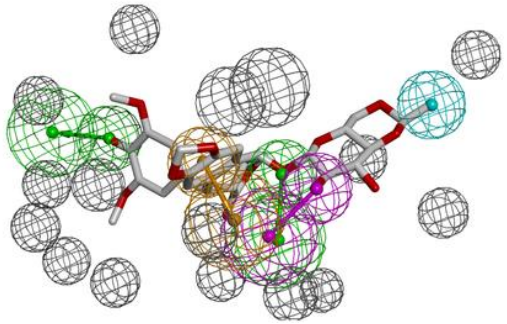


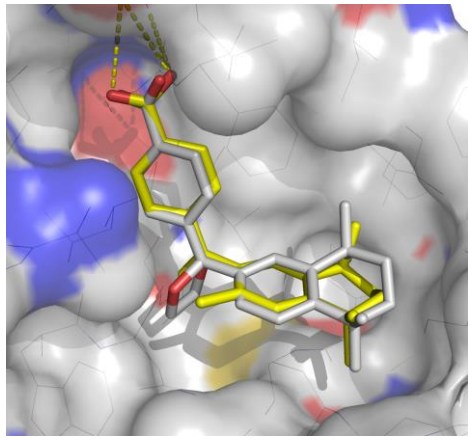
Computational chemistry

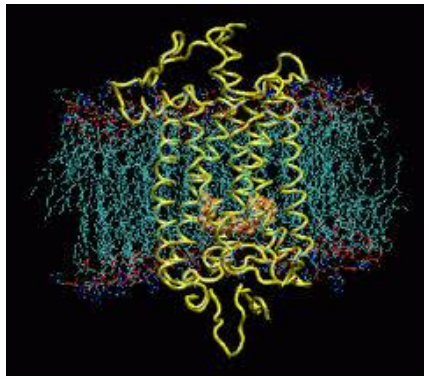
Library design

HTS screen **Hit to lead** **Lead optimization**







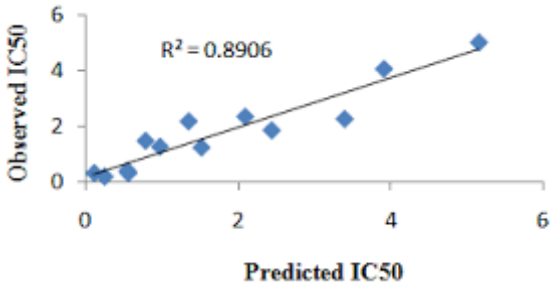


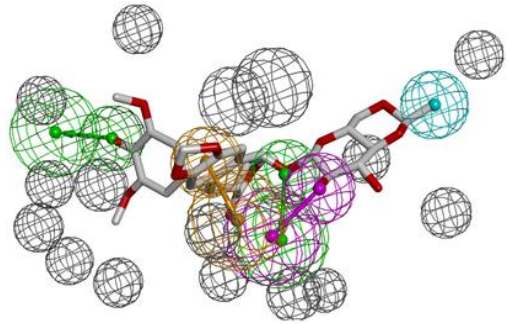
Computational chemistry

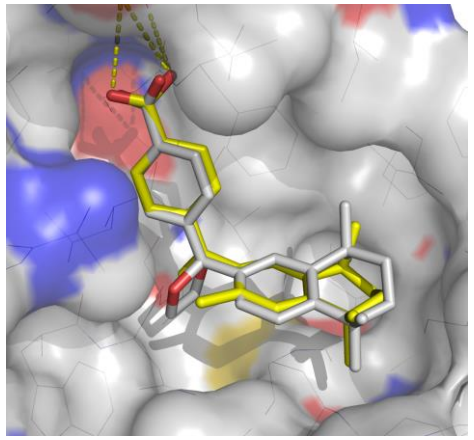
Library design

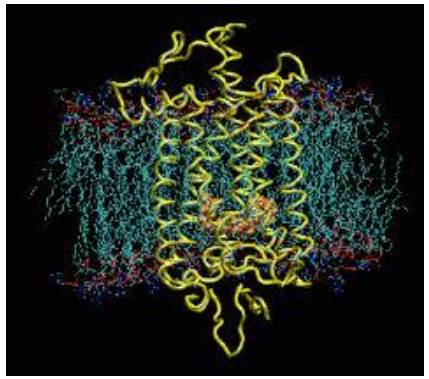
Hit selection,
Hit expansion

HTS screen → **Hit to lead** → **Lead optimization**









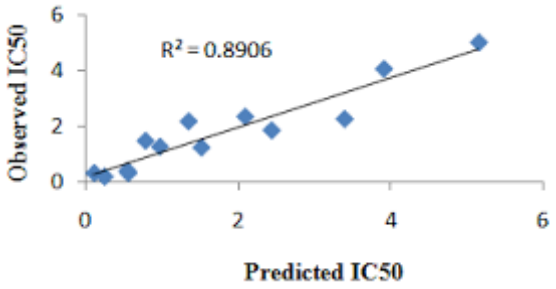
Computational chemistry

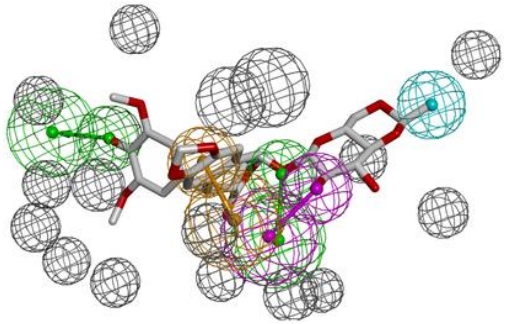
Library design

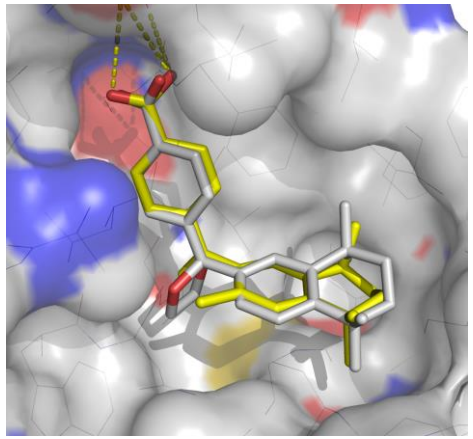
Hit selection,
Hit expansion

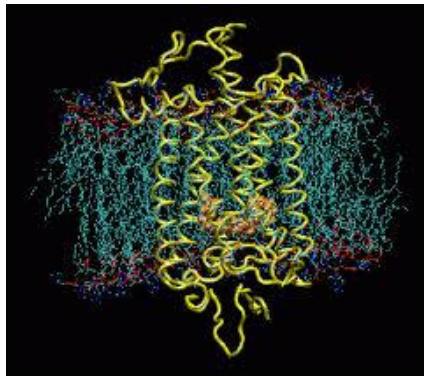
Understanding SAR
and binding

HTS screen → **Hit to lead** → **Lead optimization**









Computational chemistry

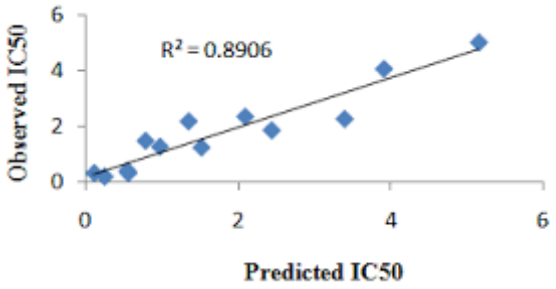
Library design

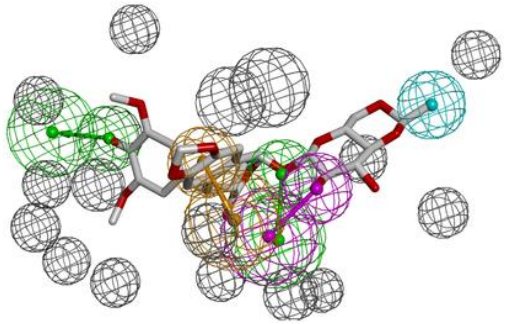
Hit selection,
Hit expansion

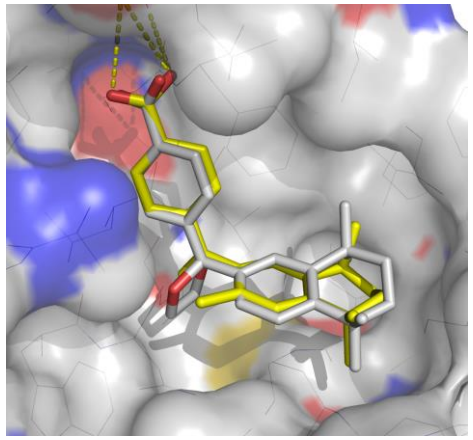
Understanding SAR
and binding

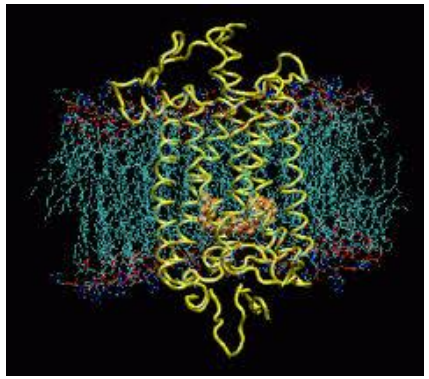
ADMET profile
prediction

HTS screen | **Hit to lead** | **Lead optimization**









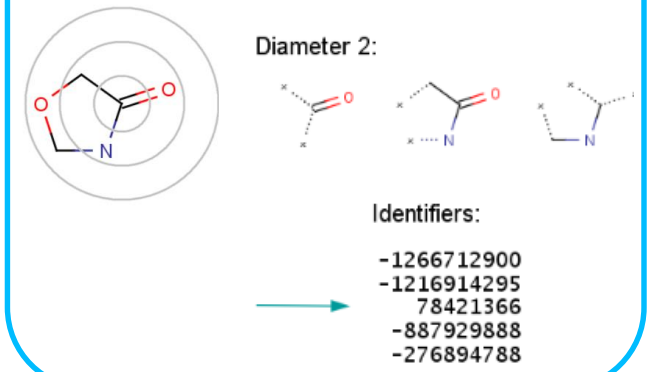
Describing the chemical matter

Molecular descriptors

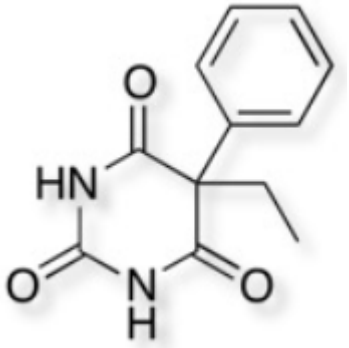
- Number of oxygen atoms
- Molecular weight
- Greasiness
- Number of aromatic bonds

…

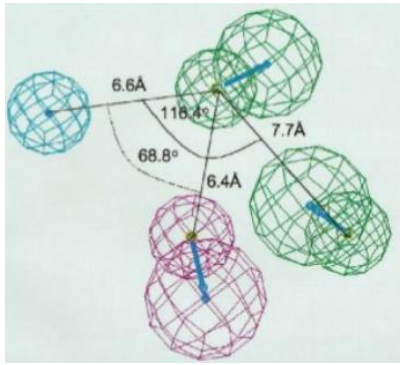
Circular fingerprints



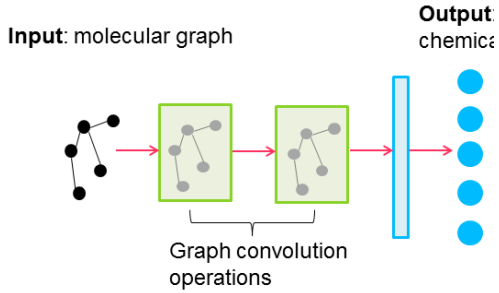
Pictures



3D descriptors



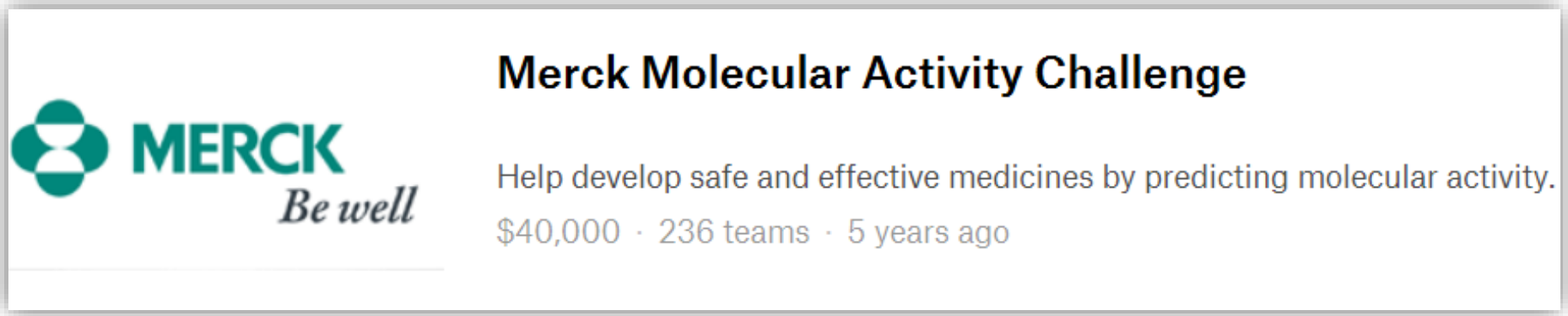
Neural fingerprints (graph convolutions)



Line notation

- Phenobarbital
- CCC1(C(=O)NC(=O)NC1=O)C1=CC=CC=C1
- InChI=1S/C12H12N2O3/c1-2-12(8-6-4-3-5-7-8)9(15)13-11(17)14-10(12)16/h3-7H,2H2,1H3,(H2,13,14,15,16,17)
- 5-ethyl-5-phenyl-1,3-diazinane-2,4,6-trione

History of deep learning for computational chemistry



2012

Winning team: gggg

- George Dahl, Toronto, Canada
- Ruslan Salakhutdinov, Toronto, Canada
- Navdeep Jaitly, Toronto, Canada
- Chris Jordan-Squire, Seattle, Washington
- Geoffrey Hinton, Toronto, Canada

No deep learning involved →

#	△pub	Team Name	Score
1	—	**gggg**	0.49409
2	—	**DataRobot**	0.48811
3	▲2	.	0.48209

Their solution:
Mixture of single task neural networks, multitask neural networks, Gaussian processes and boosted trees

History of deep learning for computational chemistry

JOURNAL OF CHEMICAL INFORMATION AND MODELING
2013
Article
pubs.acs.org/jcim

Deep Architectures and Deep Learning in Chemoinformatics: The Prediction of Aqueous Solubility for Drug-Like Molecules

Alessandro Lusci,*[,†] Gianluca Pollastri,[†] and Pierre Baldi*[,‡]

[†]School of Computer Science and Informatics, University College Dublin, Belfield, Dublin 4, Ireland
[‡]Department of Computer Science, University of California, Irvine, Irvine, California 92697, United States

Multi-task Neural Networks for QSAR Predictions

George E. Dahl gdahl@cs.toronto.edu
Navdeep Jaitly ndjaitly@cs.toronto.edu
Ruslan Salakhutdinov rsalakhu@cs.toronto.edu
Department of Computer Science, University of Toronto,
6 King's College Rd, Toronto, Ontario M5S 3G4, Canada

2014

Deep Neural Nets as a Method for Quantitative Structure–Activity Relationships

Junshui Ma*†, Robert P. Sheridan‡, Andy Liaw†, George E. Dahl§, and Vladimir Svetnik†
†Biometrics Research Department and ‡Structural Chemistry Department, Merck Research Laboratories, Rahw
Jersey 07065, United States
§ Computer Science Department, University of Toronto, Toronto, Ontario ON M5S, Canada

J. Chem. Inf. Model., **2015**, *55* (2), pp 263–274
DOI: 10.1021/ci500747n
Publication Date (Web): January 30, 2015

2015

Toxicity Prediction using Deep Learning

Thomas Unterthiner, Andreas Mayr, Günter Klambauer, Sepp Hochreiter
Published 2015 in ArXiv

2015

Deep Learning as an Opportunity in Virtual ening

RISC Software GmbH &
Institute of Bioinformatics
Johannes Kepler University Linz, Austria
unterthiner@bioinf.jku.at

Andreas Mayr *
RISC Software GmbH &
Institute of Bioinformatics
Johannes Kepler University Linz, Austria
mayr@bioinf.jku.at

Günter Klambauer
Institute of Bioinformatics
s Kepler University Linz, Austria
mbauer@bioinf.jku.at

Marvin Steijaert
OpenAnalytics, Belgium
marvin.steijaert@openanalytics.eu

Jörg K. Wegner
Johnson & Johnson
naceutical Research & Development
jwegner@its.jnj.com

Hugo Ceulemans
Johnson & Johnson
Pharmaceutical Research & Development
hceulema@its.jnj.com

Sepp Hochreiter
Institute of Bioinformatics
Johannes Kepler University Linz, Austria
hochreit@bioinf.jku.at

2014

2016

Modeling Industrial ADMET Data with Multitask Networks

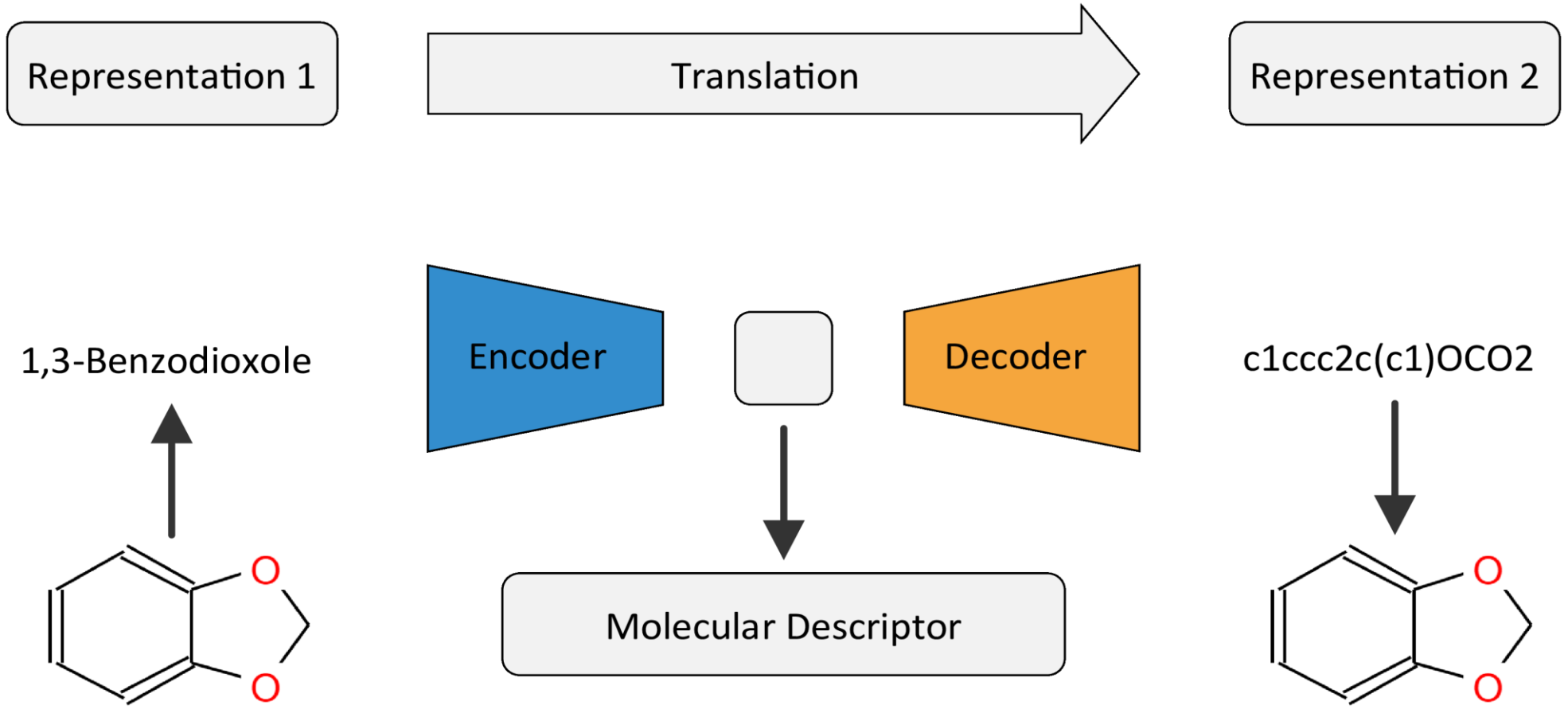
Steven Kearnes
Stanford University
kearnes@stanford.edu

Brian Goldman
Vertex Pharmaceuticals Inc.
brian_goldman@vrtx.com

Vijay Pande
Stanford University
pande@stanford.edu

Continuous, data-driven molecular descriptors

General idea



Winter et al., *Chemical Sciences*, 2019

Data and representations

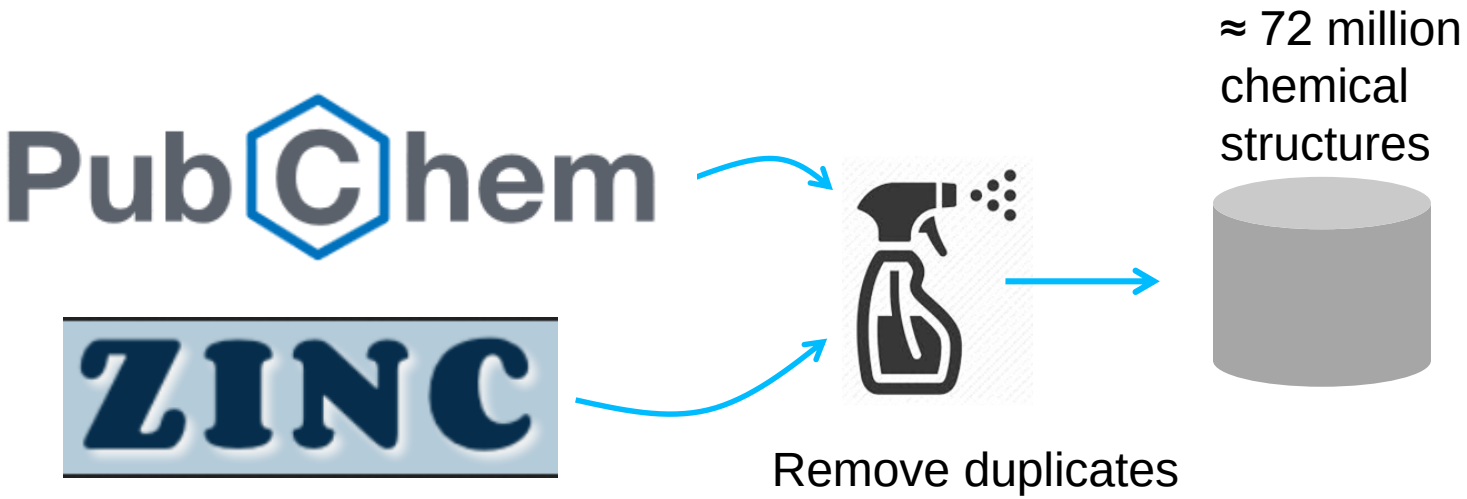
Winter et al., *Chemical Sciences*, 2019

Data and representations

Remove duplicates
Remove inorganic molecules
12 < MW < 600
>3 heavy atoms
-7 < logP < 5
Remove stereochemistry
Keep largest fragments

Winter et al., *Chemical Sciences*, 2019

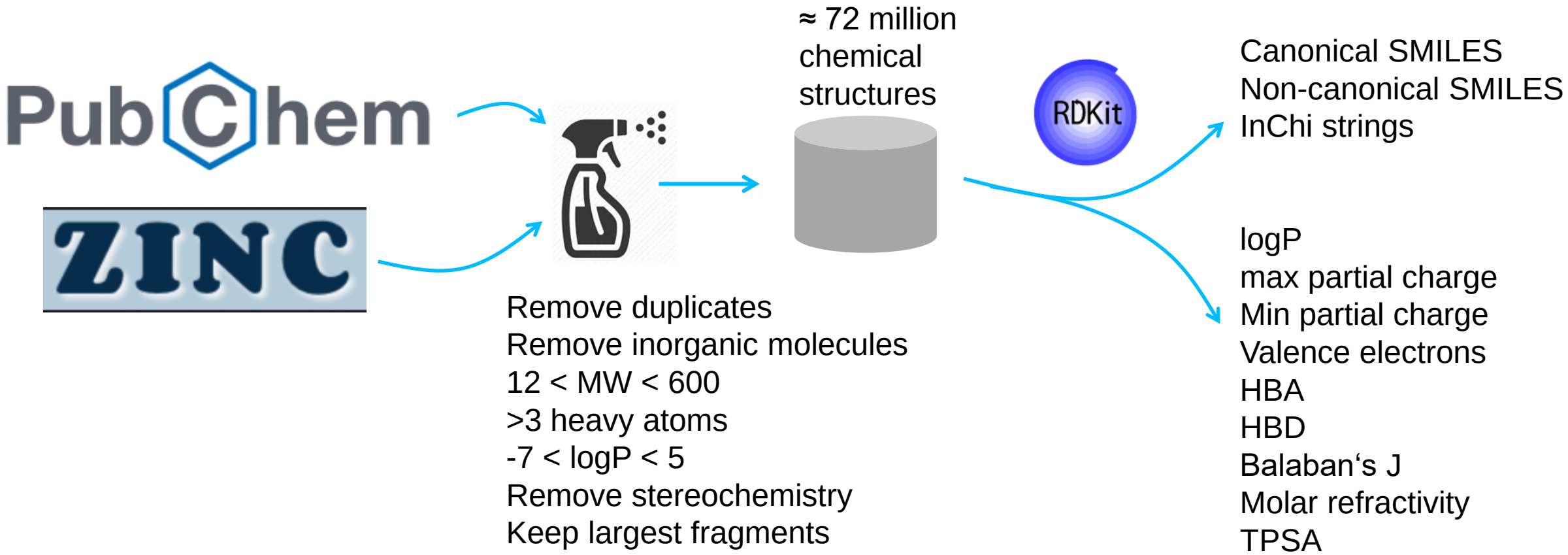
Data and representations



Remove duplicates
Remove inorganic molecules
12 < MW < 600
>3 heavy atoms
-7 < logP < 5
Remove stereochemistry
Keep largest fragments

Winter et al., *Chemical Sciences*, 2019

Data and representations

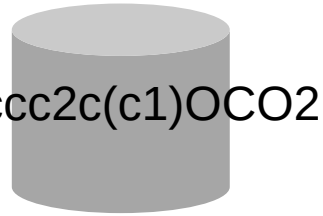


Winter et al., *Chemical Sciences*, 2019

Training the translation model

Input tokenization

Database represented
by SMILES

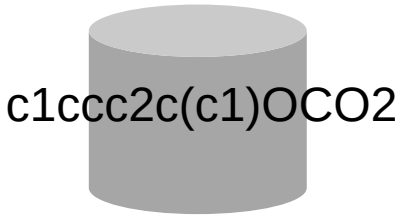


Winter et al., *Chemical Sciences*, 2019

Training the translation model

Input tokenization

Database represented by SMILES



Lookup table

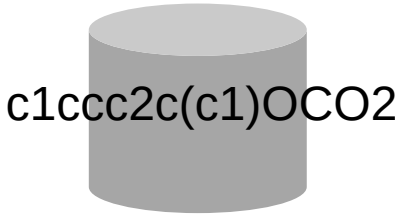
Token	Index
c	0
1	1
2	2
(	3
)	4
O	5
C	6
…	
Br	37

Winter et al., *Chemical Sciences*, 2019

Training the translation model

Input tokenization

Database represented by SMILES



Lookup table

Token	Index
c	0
1	1
2	2
(	3
)	4
O	5
C	6
…	
Br	37

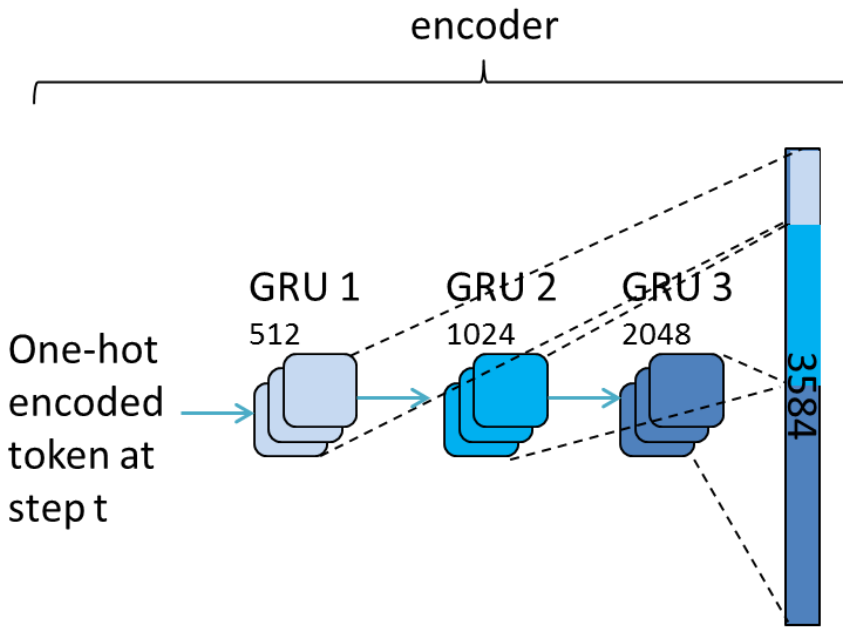
One-hot encoding of tokens

„c“: [1, 0, 0, 0, …, 0]
„(„: [0, 0. 0 ,1, …, 0]
…

Winter et al., *Chemical Sciences*, 2019

Training the translation model

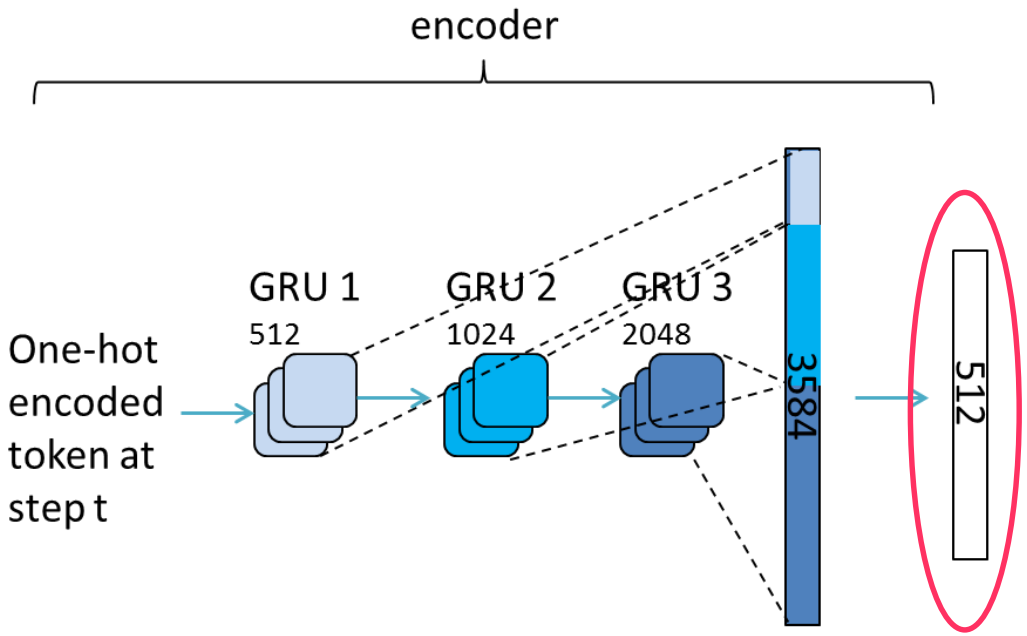
Model architecture



Winter et al., *Chemical Sciences*, 2019

Training the translation model

Model architecture

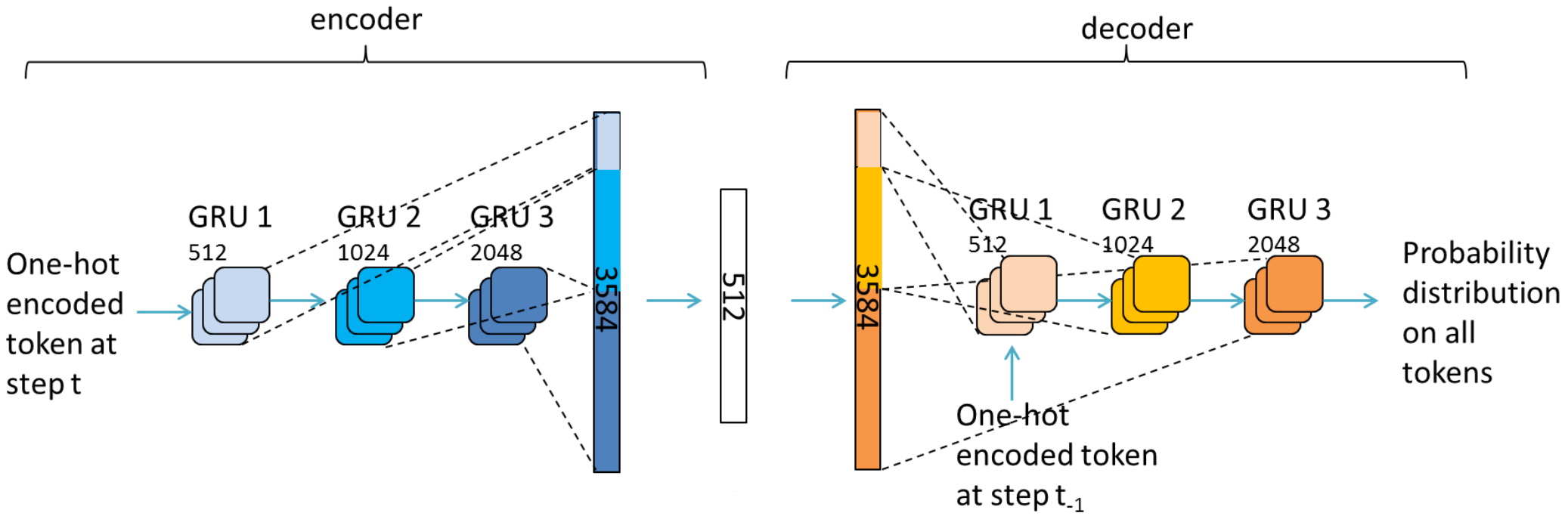


bottleneck: embedding for the input SMILES string

Winter et al., *Chemical Sciences*, 2019

Training the translation model

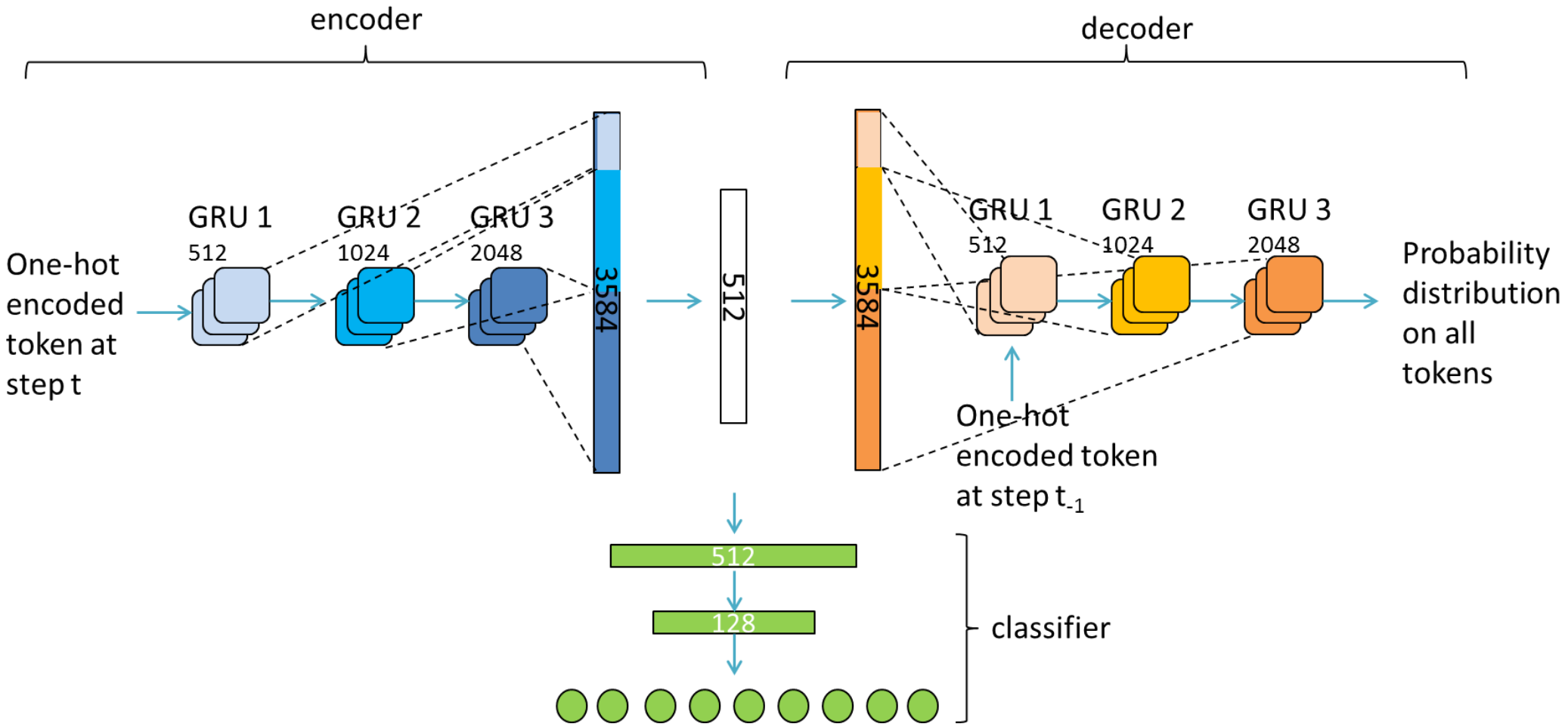
Model architecture



Winter et al., *Chemical Sciences*, 2019

Training the translation model

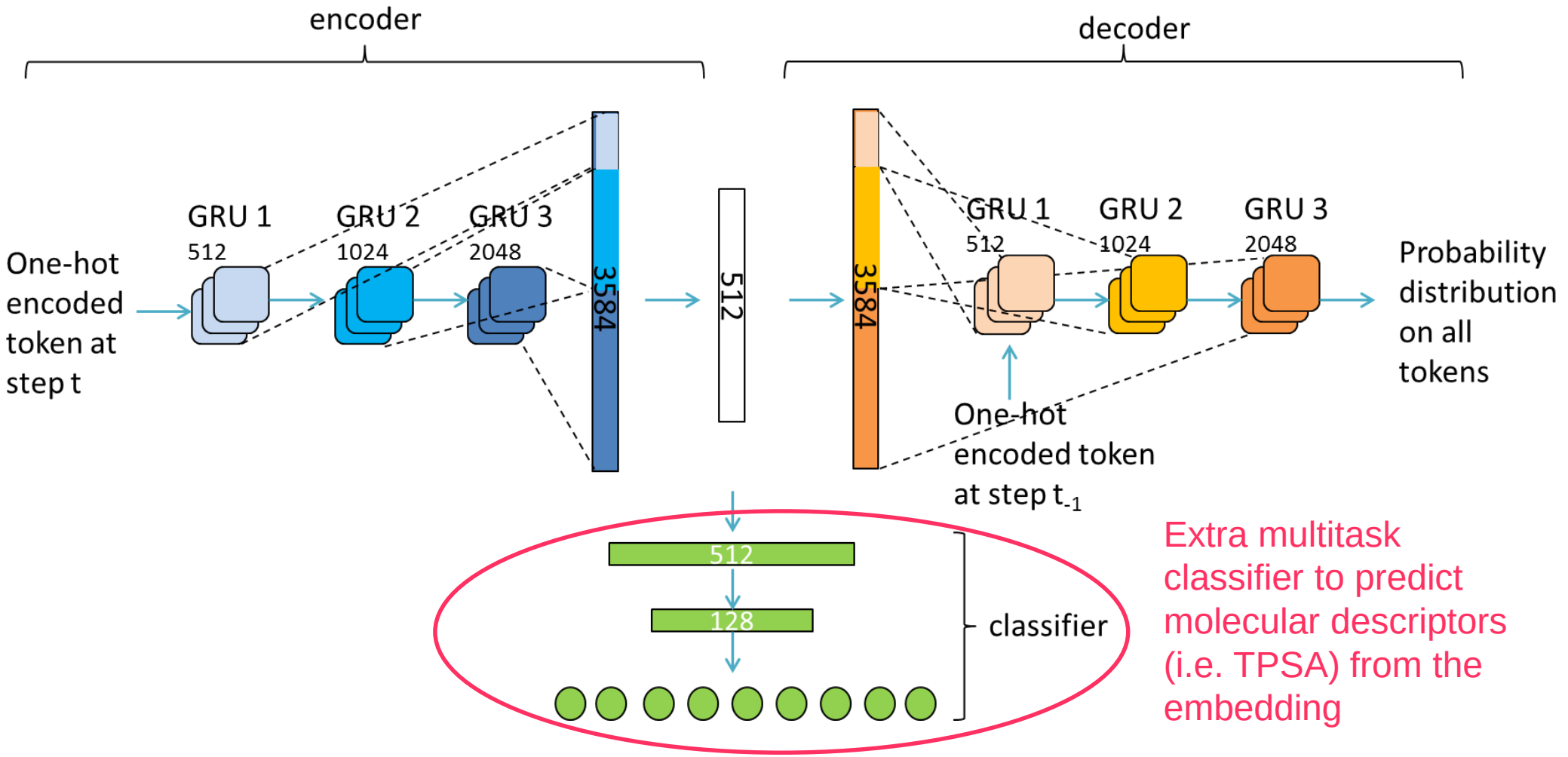
Model architecture



Winter et al., *Chemical Sciences*, 2019

Training the translation model

Model architecture

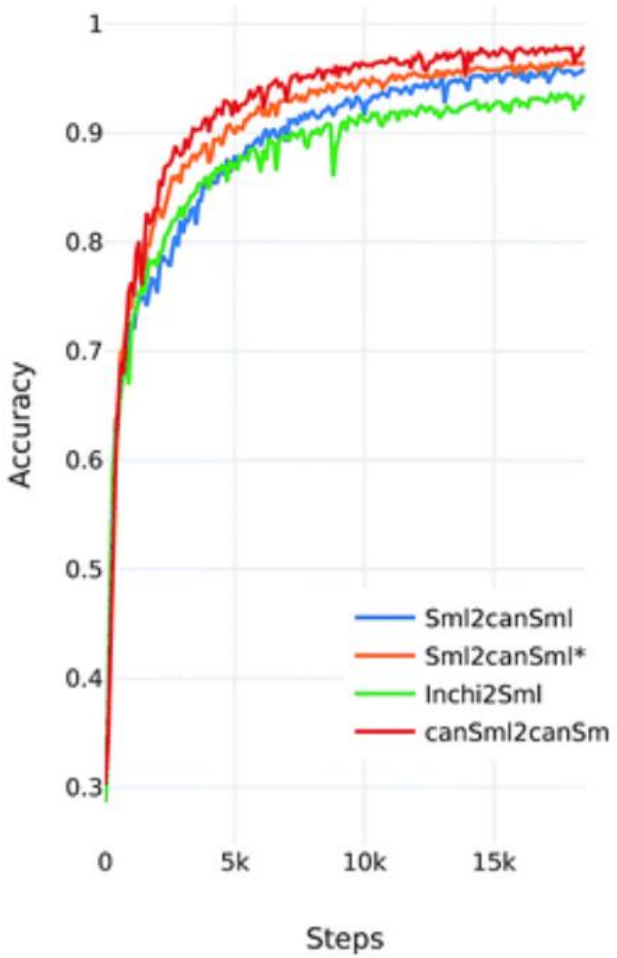


Extra multitask classifier to predict molecular descriptors (i.e. TPSA) from the embedding

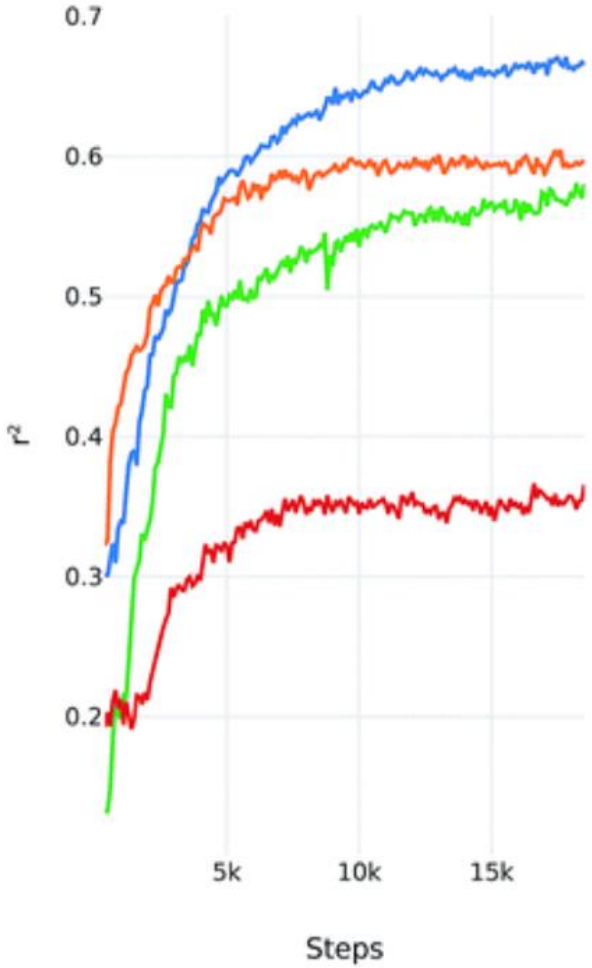
Winter et al., *Chemical Sciences*, 2019

Training the translation model

Translation accuracy



Performance of the embedding
as descriptor for lipophilicity prediction

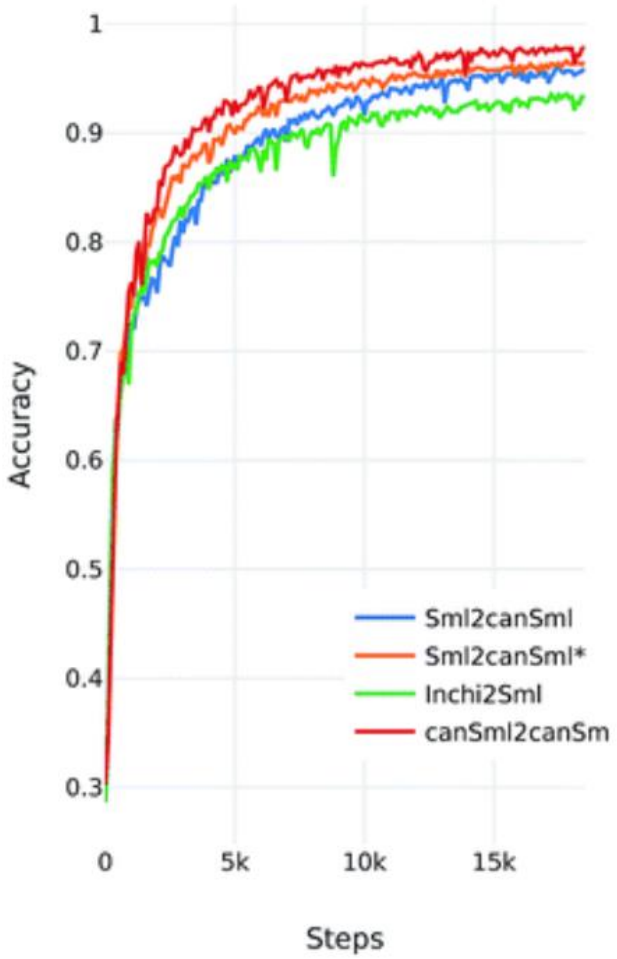


- canonical SMILES to canonical SMILES
- SMILES to canonical SMILES, **no helper tasks**
- SMILES to canonical SMILES
- InChi to canonical SMILES

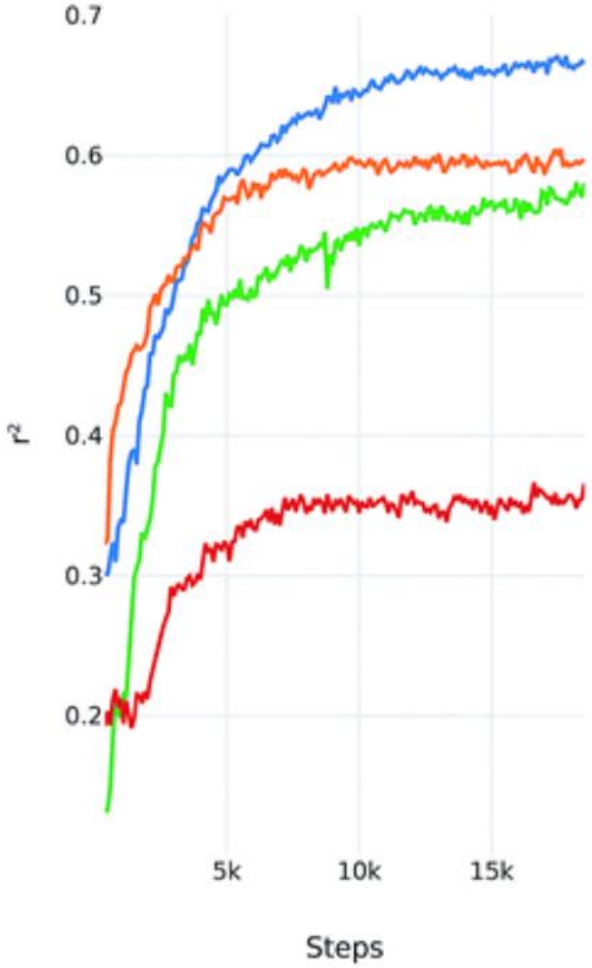
Winter et al., *Chemical Sciences*, 2019

Training the translation model

Translation accuracy



Performance of the embedding
as descriptor for lipophilicity prediction



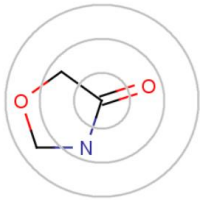
- canonical SMILES to canonical SMILES
- SMILES to canonical SMILES, **no helper tasks**
- SMILES to canonical SMILES
- InChi to canonical SMILES

Best model for reconstruction: canSMI to canSMI. It is also the poorest in terms of quality of the embedding for downstream tasks. The „helper tasks“ (i.e. additional loss for the embedding to predict simple molecular properties) boosts the performance on downstream tasks.

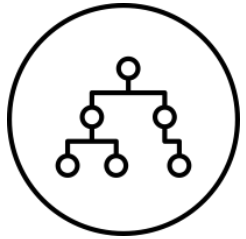
Winter et al., *Chemical Sciences*, 2019

Performance of the autoencoder embedding as molecular descriptor

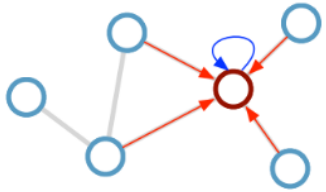
Dataset	Acronym	Task	Split	Number of compounds
Ames mutagenicity	ames	Classification	Validation	6130
HERG inhibition	herg	Classification	Test	3440
Blood–brain barrier penetration	bbbp	Classification	Test	1879
β-Secretase 1 inhibition	bace	Classification	Test	1483
Toxicity in honeybees	beet	Classification	Test	188
Epidermal growth factor inhibition	egfr	Regression	Test	4451
Plasmodium falciparum inhibition	plasmo	Regression	Test	3999
Lipophilicity	lipo	Regression	Validation	3817
Aqueous solubility	esol	Regression	Test	1056
Melting point	melt	Regression	Test	184



- CDDD
- Fingerprints of diameter 2, 4, 6 and fold size 512, 1024 or 2048



- Random Forest
- SVM
- Gradient boosting

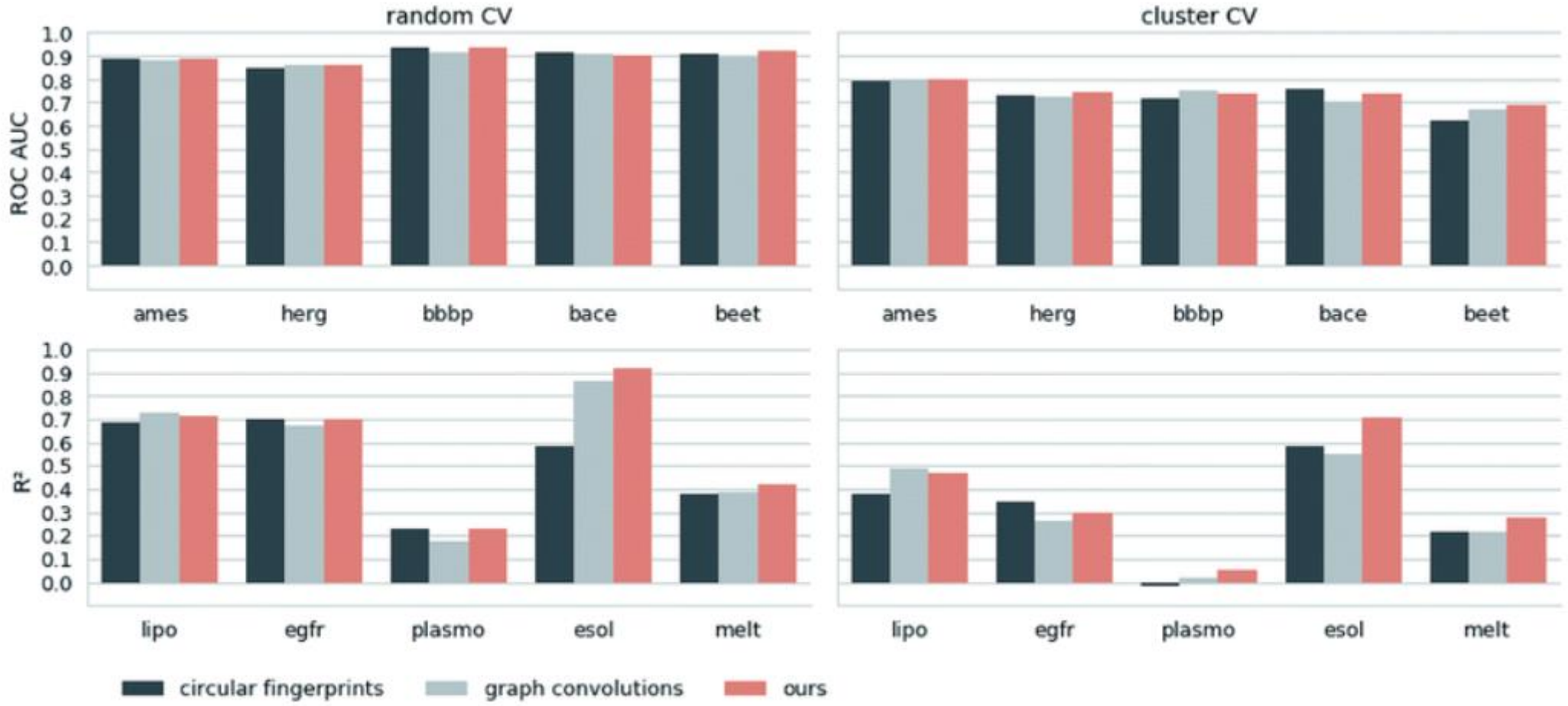


Graph convolutional network

Winter et al., *Chemical Sciences*, 2019

Performance of the autoencoder embedding as molecular descriptor

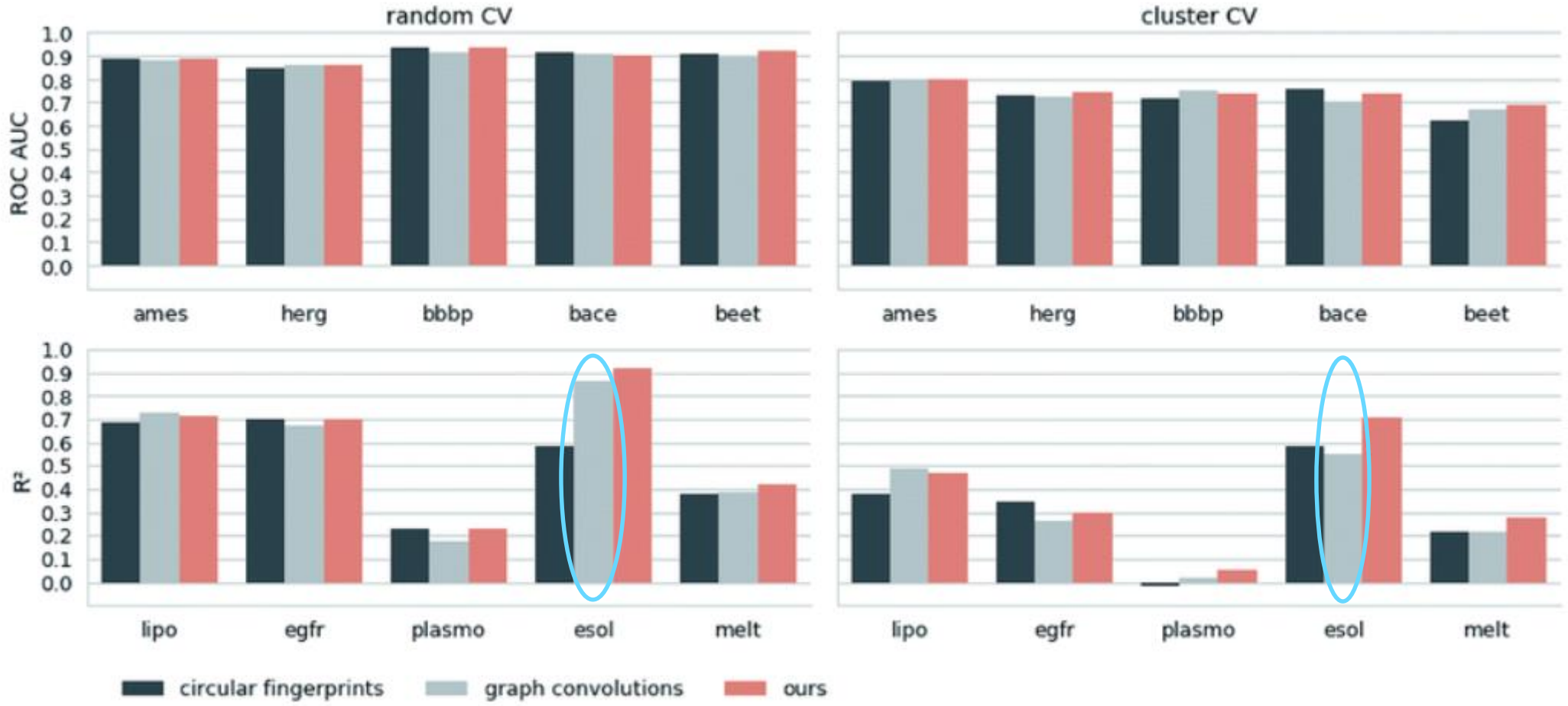
QSAR models



Winter et al., *Chemical Sciences*, 2019

Performance of the autoencoder embedding as molecular descriptor

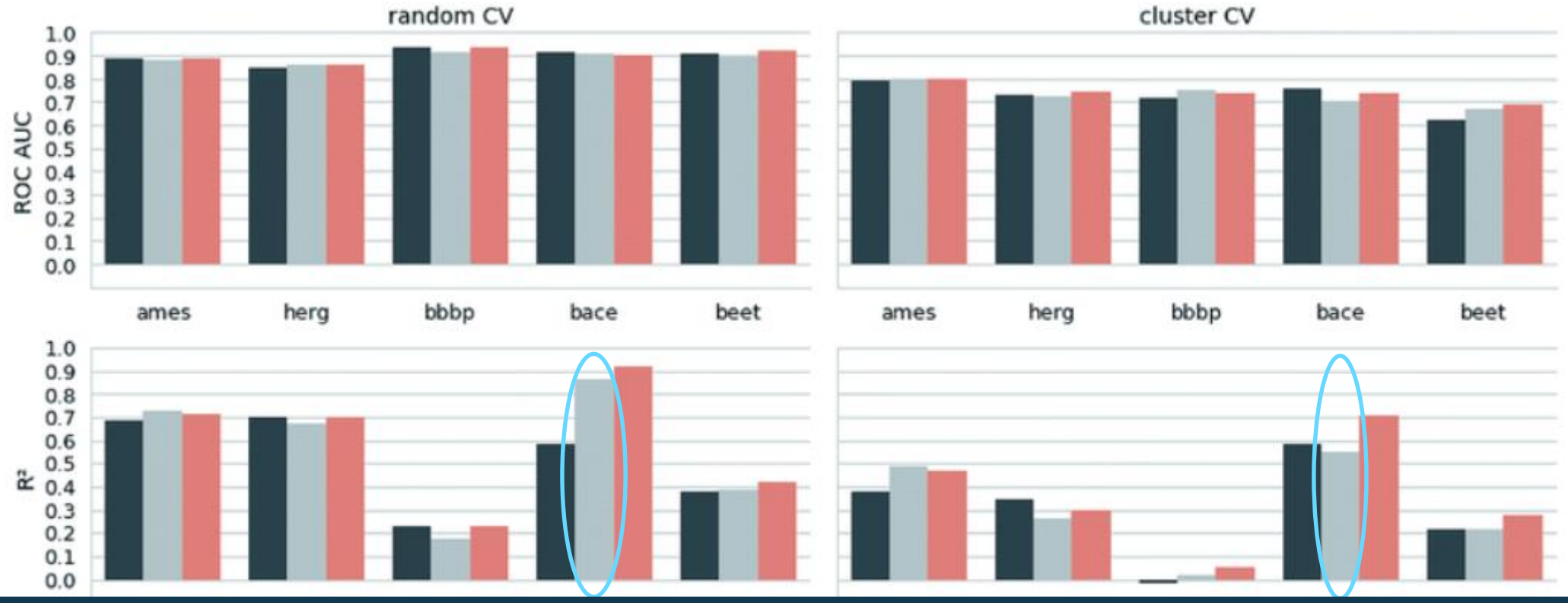
QSAR models



Winter et al., *Chemical Sciences*, 2019

Performance of the autoencoder embedding as molecular descriptor

QSAR models

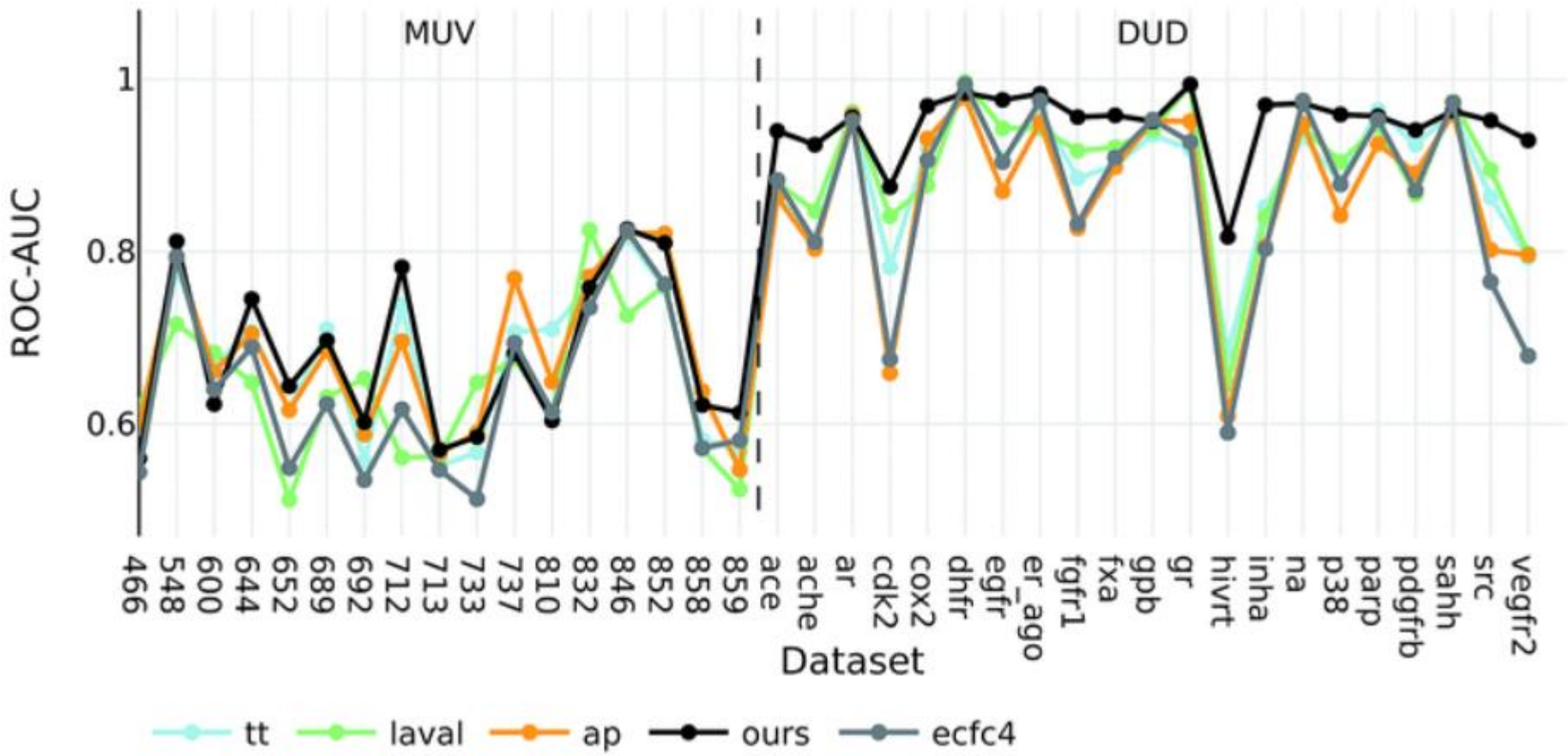


Competitive performance of CDDD with SVM against other optimized methods. Performance stable under cluster split evaluation. Graph convolutional approaches tend to suffer of overfitting.

Winter et al., *Chemical Sciences*, 2019

Performance of the autoencoder embedding as molecular descriptor

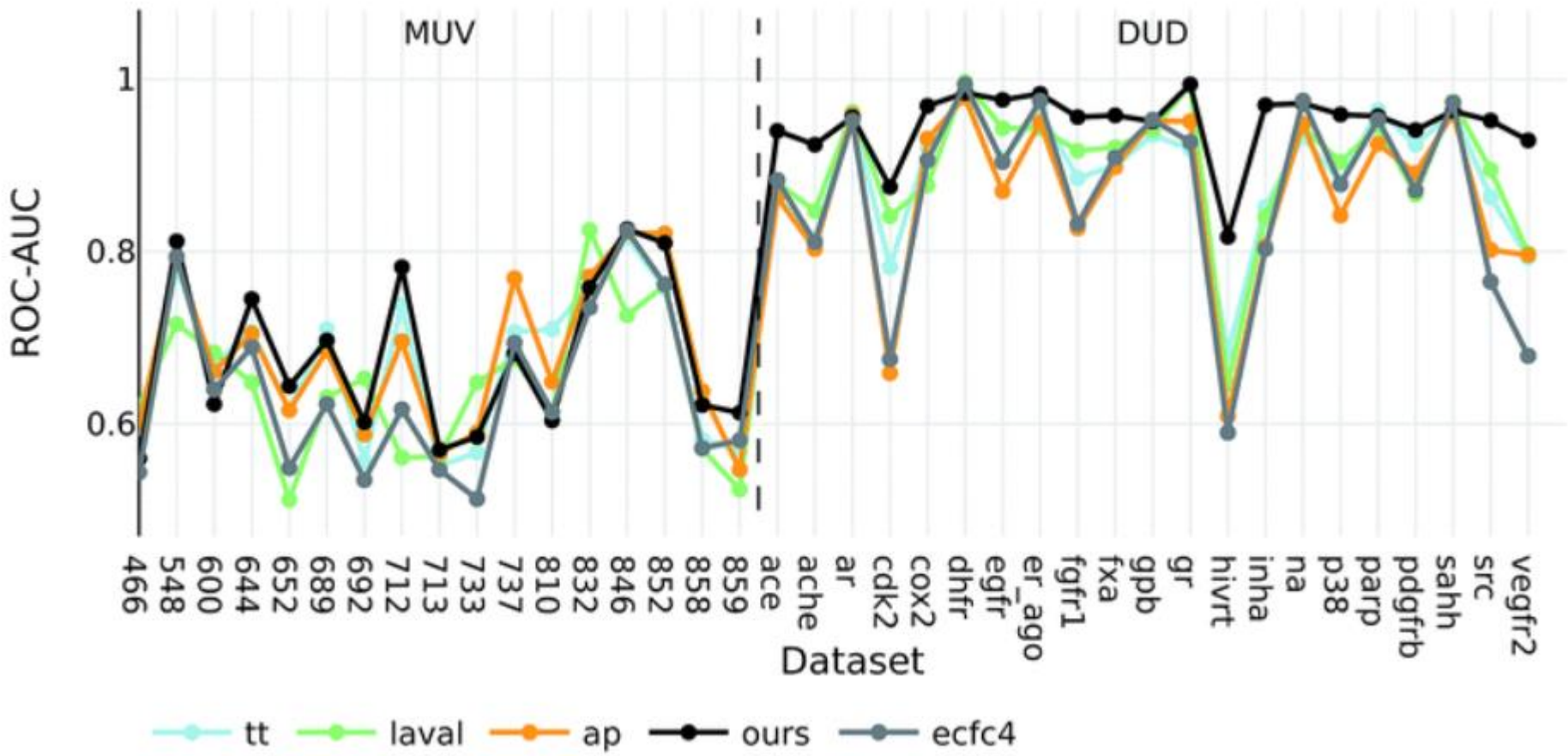
Virtual screening



Winter et al., *Chemical Sciences*, 2019

Performance of the autoencoder embedding as molecular descriptor

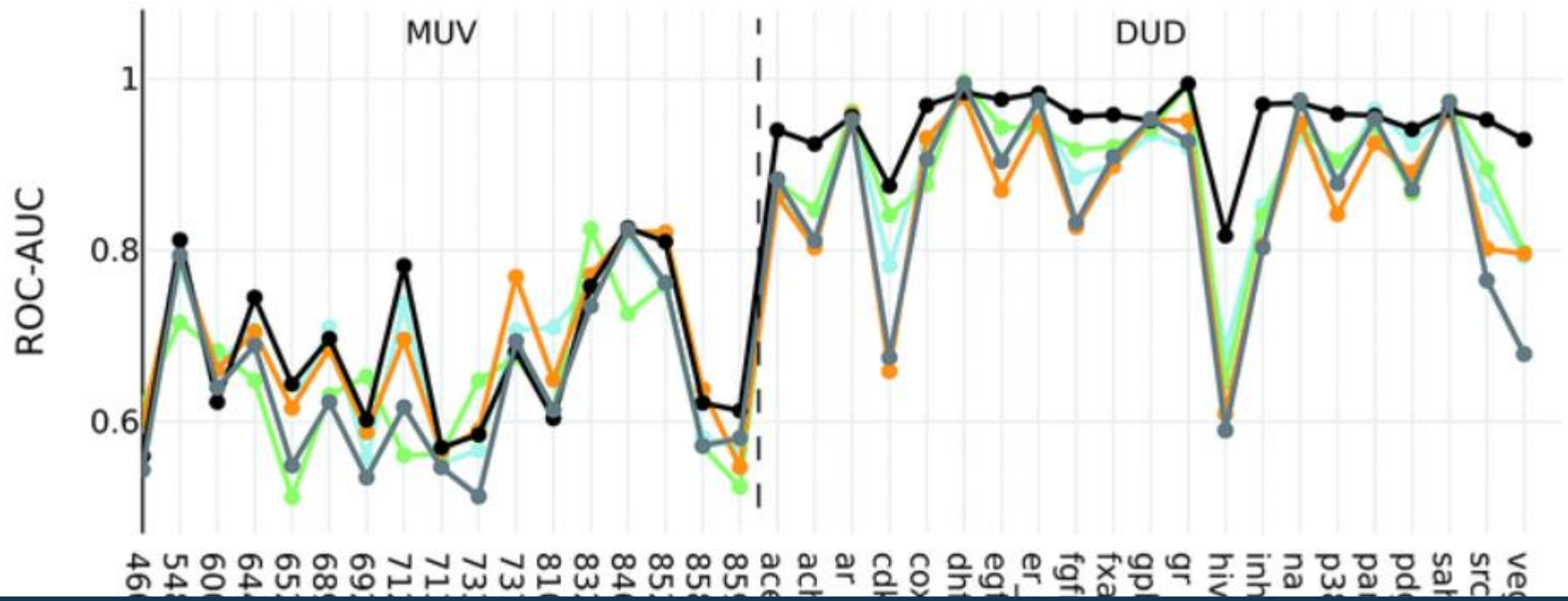
Virtual screening



Winter et al., *Chemical Sciences*, 2019

Performance of the autoencoder embedding as molecular descriptor

Virtual screening



Overall datasets, ranking using CDDD significantly outperforms the second best descriptors ($p < 0.05$). Similarities as measured by the distance between CDDD embeddings seem to correlate well with biological activity.

Winter et al., *Chemical Sciences*, 2019

Wrap-up

Inspired by language translation models, we pre-train an autoencoder on 72 million chemical structures.

The bottleneck of the model can be used to describe compounds.

The descriptors work very well in combination with SVM for building QSAR models.

The descriptors outperform the benchmarked descriptors on the MUV and DUD virtual screening datasets.

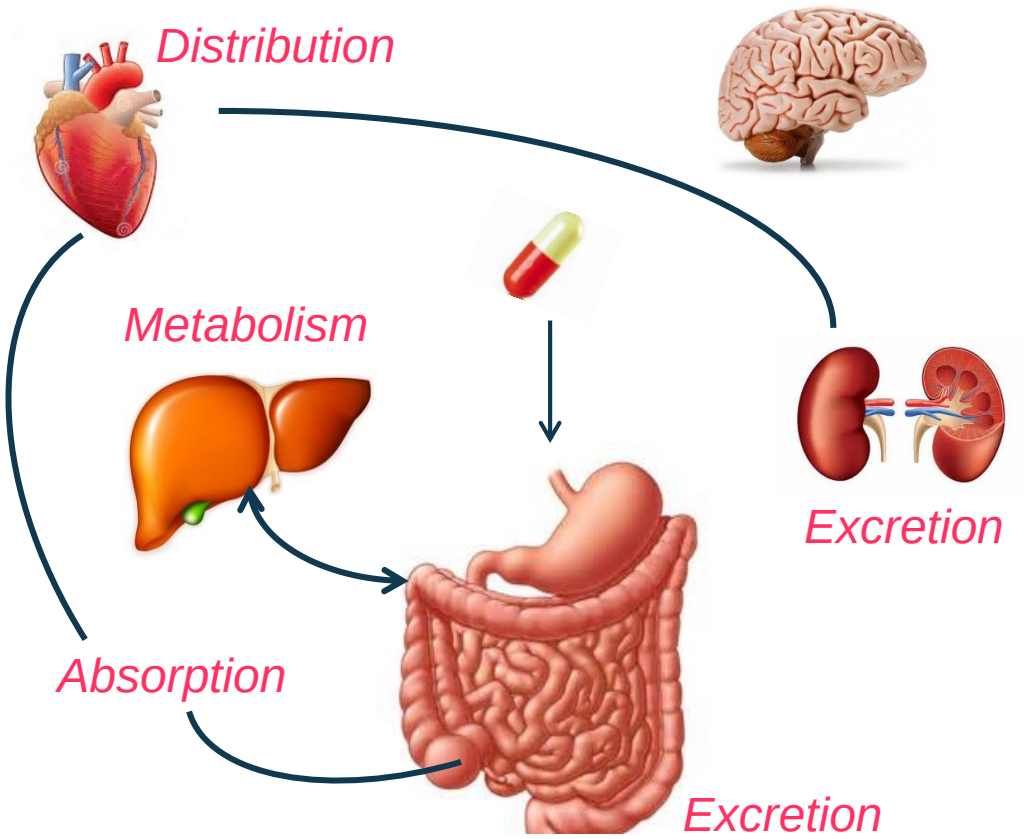
One crucial point is that the embedding is a continuous space, and that one can use the decoder to reverse the embedding to a molecule (not possible with other types of descriptors). More on that in the last part!

Trained model is available on github! https://github.com/jrwnter/cddd

Winter et al., *Chemical Sciences*, 2019

Multitask learning for ADMET prediction

Absorption – Distribution – Metabolism – Excretion – Toxicity



Where will it go?

How much will get there?

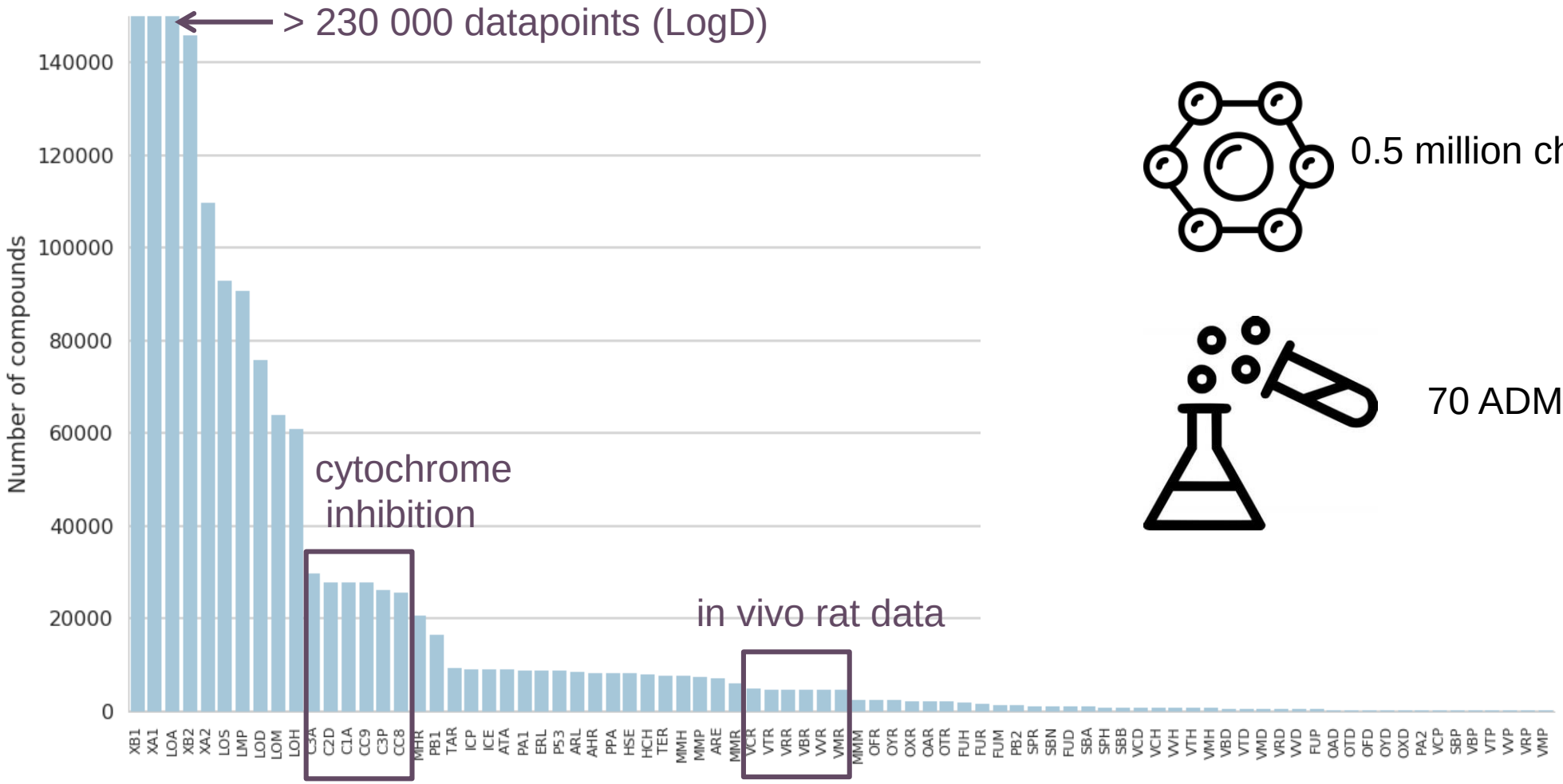
How long will it stay?

Will it be transformed?

How will it be removed?

Will it reach unwanted sites?

The data

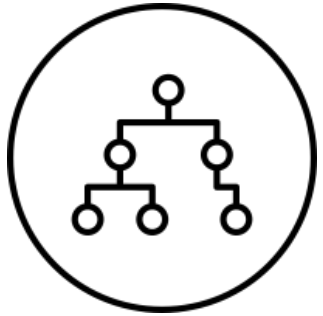


0.5 million chemical structures

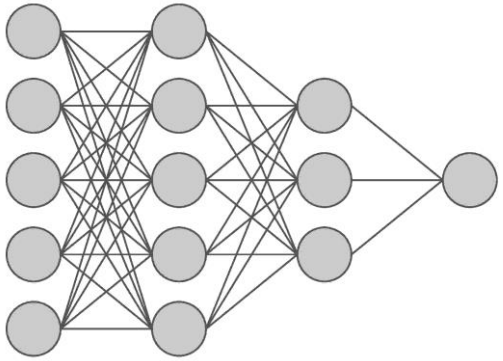
70 ADMET endpoints

Different methods compared

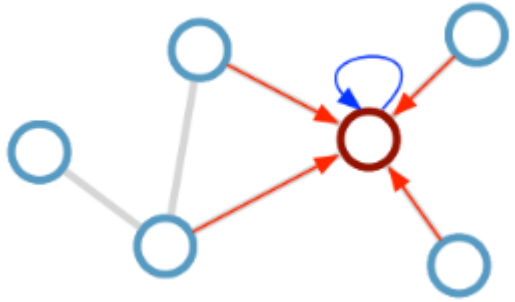
Baseline



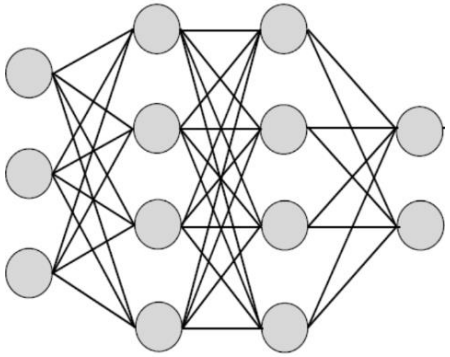
Single task neural networks



Graph convolutional neural networks



Multitask neural networks



Multitask learning

Not the same as multiclass or multilabel learning!!

Multiclass

New instance to predict can belong to only one class among several options. Classical examples: MNIST, Iris, Imagenet, …

Multitask learning

Not the same as multiclass or multilabel learning!!

Multiclass

New instance to predict can belong to only one class among several options. Classical examples: MNIST, Iris, Imagenet, …

Multilabel

- Action
- Drama
- Sci-fi

New instance to predict can belong to more than one category.

Multitask learning

Not the same as multiclass or multilabel learning!!

Multiclass

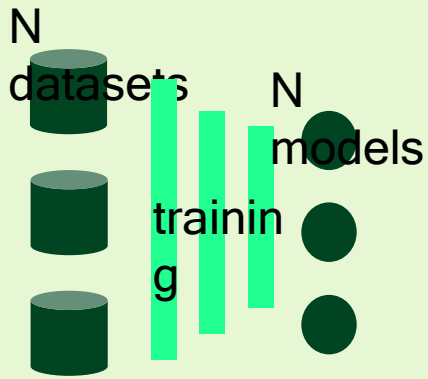
New instance to predict can belong to only one class among several options. Classical examples: MNIST, Iris, Imagenet, …

Multilabel

- Action
- Drama
- Sci-fi

New instance to predict can belong to more than one category.

Multitask



Multiple related learning tasks are learned simultaneously using a shared representation. Ex: text translation in multiple languages.

Multitask learning

Not the same as multiclass or multilabel learning!!

Multiclass

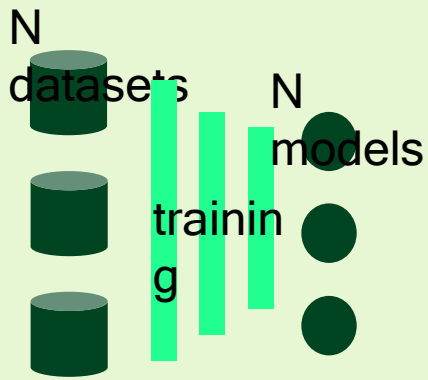
New instance to predict can belong to only one class among several options. Classical examples: MNIST, Iris, Imagenet, …

Multilabel

- Action
- Drama
- Sci-fi

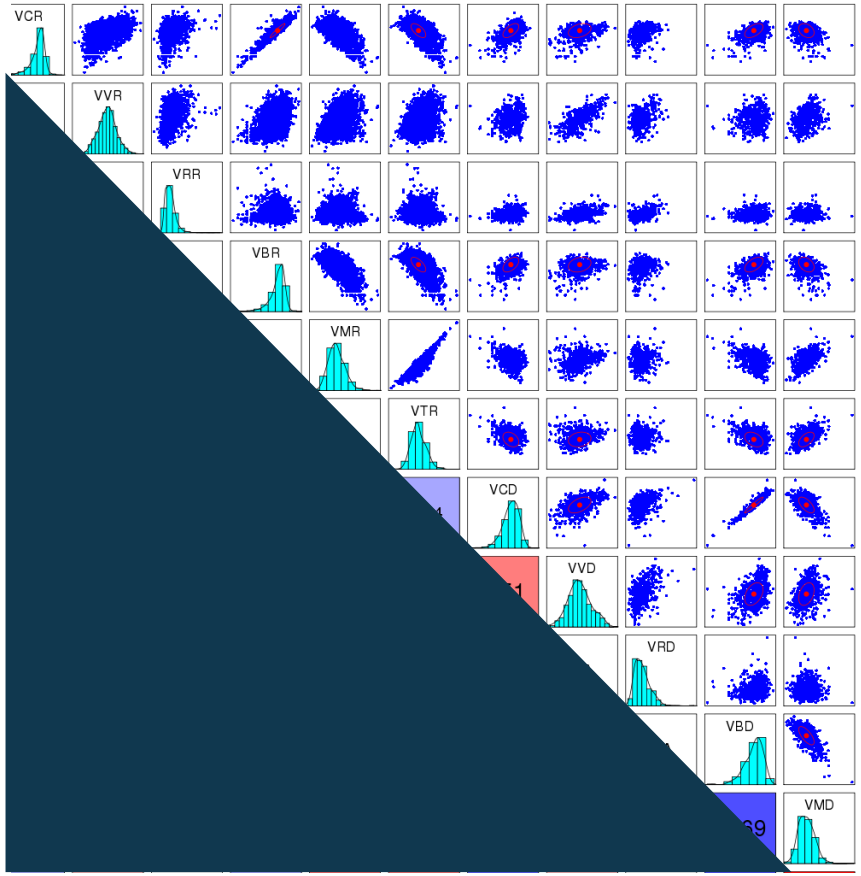
New instance to predict can belong to more than one category.

Multitask



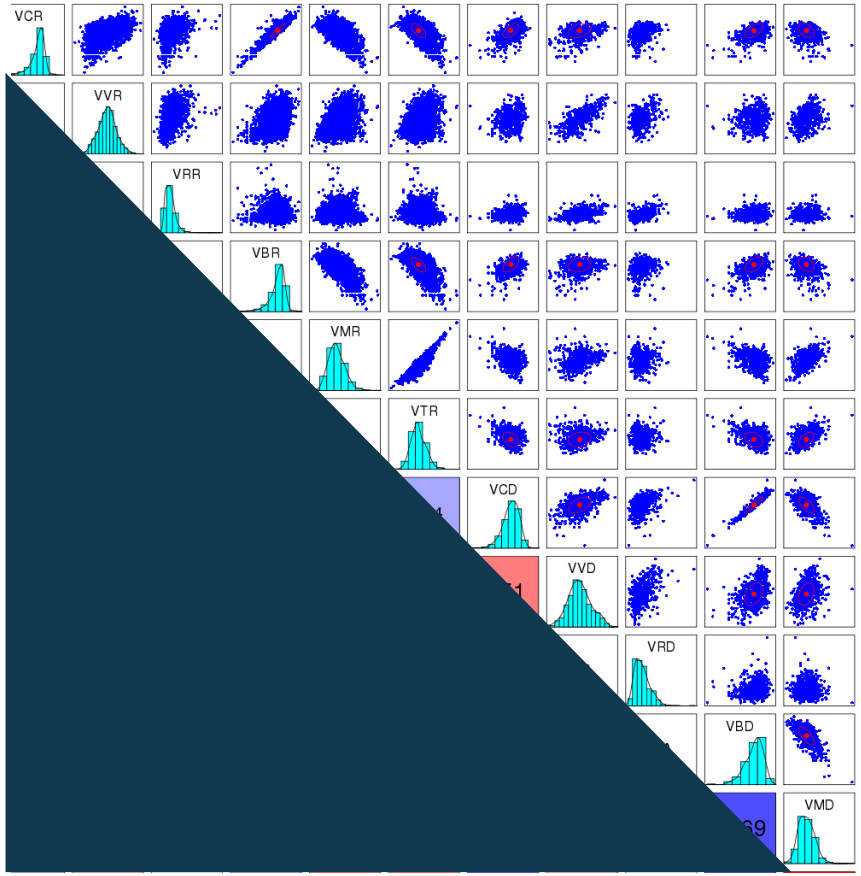
Multiple related learning tasks are learned simultaneously using a shared representation. Ex: text translation in multiple languages.

Motivation for a multitask approach and expected benefits



Some endpoints are (weakly) correlated
Some endpoints are complementary in a biological sense
Some endpoints are obtained from the same biological experiment

Motivation for a multitask approach and expected benefits

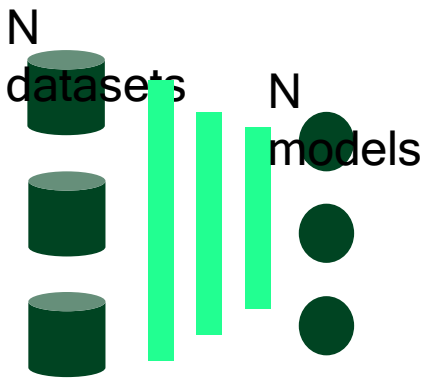


Some endpoints are (weakly) correlated
Some endpoints are complementary in a biological sense
Some endpoints are obtained from the same biological experiment

Benefits

- **Larger training set**: endpoints with less compounds benefit from the chemical space of endpoints with more compounds.
- **Exploits correlations between endpoints**
- **Regularization** method

Multitask learning in practice

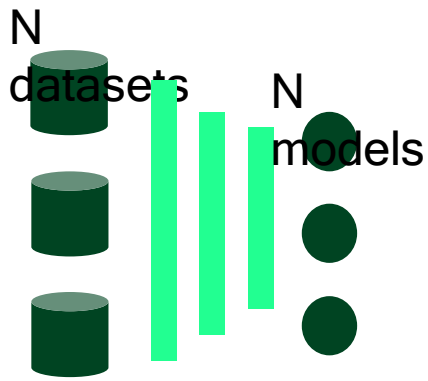


- Choose N datasets that can be learned together:
LogD (X_1, y_1)
Inhibition of CYP3A4 (X_2, y_2)
Caco2 permeability (X_3, y_3)

- Combine them into a multitask training set (X_4, Y)
X_4 contains *U* unique compounds from X_1, X_2 and X_3
Y is of shape (U, N) with missing values when a given *u* doesn‘t have a measurement

There is no requirement of overlap between the different datasets. Some overlap helps!

Multitask learning in practice

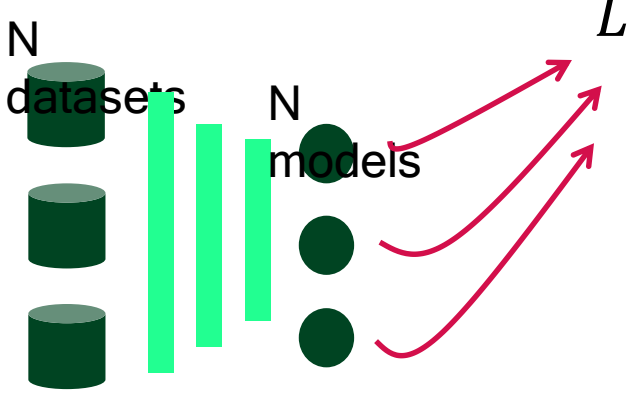


Loss calculation:

$$L_t = \frac{1}{|B|}\sum_{i \in B}(yi - \hat{y}_i)^2$$

Consequence: the N tasks must be giving outputs in the same range! Necessary to scale all y_t (z-scaling works well in practice)

Multitask learning in practice



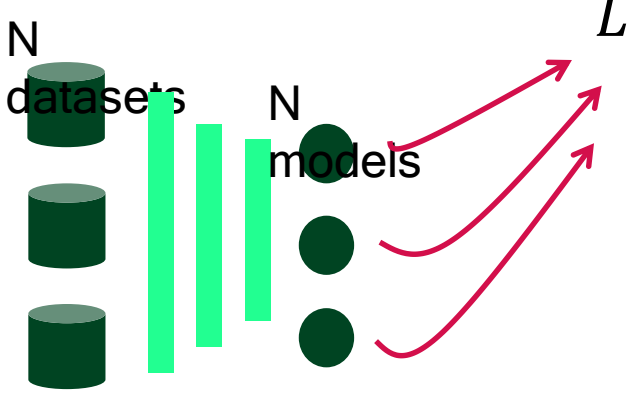
Loss calculation:

$$L_t = \frac{1}{|B|}\sum_{i \in B}(yi - \hat{y}_i)^2$$

Within a minibatch *B*
Independent for each task *t*
Ignore missing values

Consequence: the N tasks must be giving outputs in the same range! Necessary to scale all y_t (z-scaling works well in practice)

Multitask learning in practice



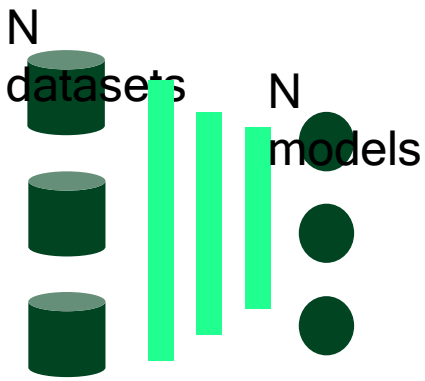
Loss calculation:

$$L_t = \frac{1}{|B|}\sum_{i \in B}(yi - \hat{y}_i)^2 \quad \Rightarrow \quad L_{global} = \frac{\sum_t w_t L_t}{\sum w_t}$$

Within a minibatch *B*
Independent for each task *t*
Ignore missing values

Consequence: the N tasks must be giving outputs in the same range! Necessary to scale all y_t (z-scaling works well in practice)

Multitask learning in practice



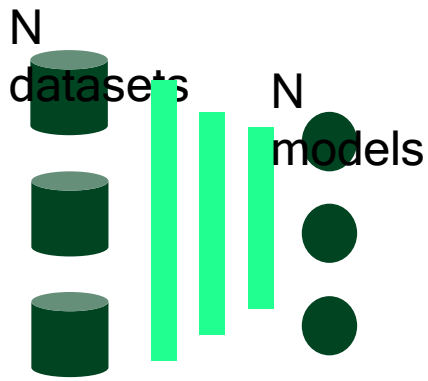
Loss calculation:

$$L_{global} = \frac{\sum_t w_t L_t}{\sum w_t}$$

Considerations on w_t

Typical choice would be $1/N$

Multitask learning in practice

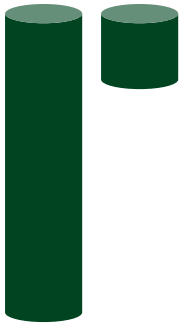


Loss calculation:

$$L_{global} = \frac{\sum_t w_t L_t}{\sum w_t}$$

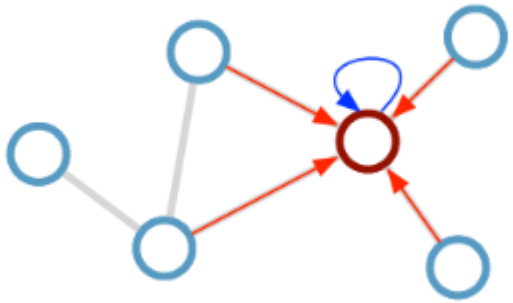
Considerations on w_t

Typical choice would be $1/N$



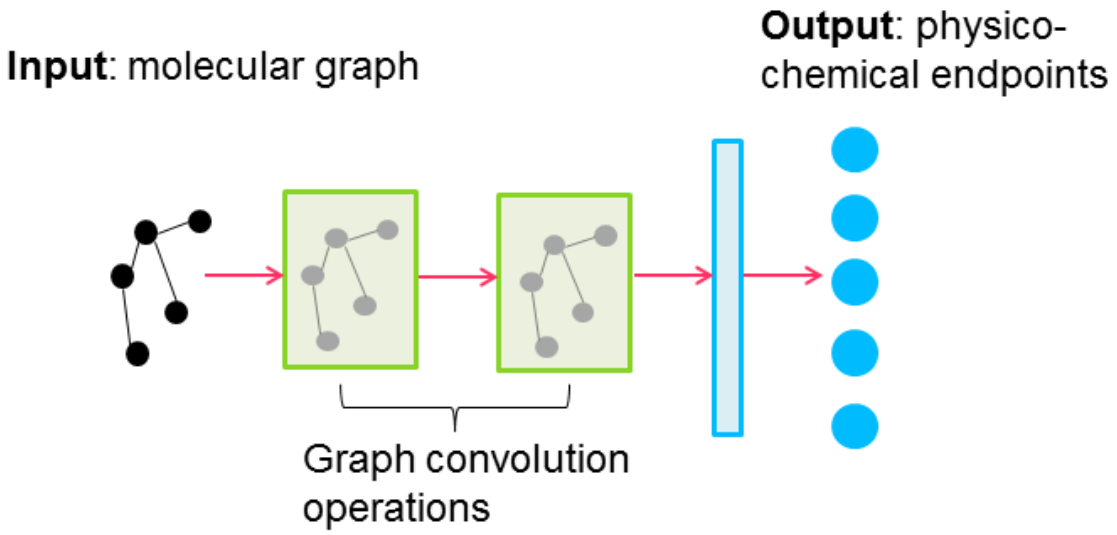
In case of highly varying task sizes, it might be useful to increase the weight on the smaller task so that it participates more in the global loss.

Graph convolutional networks for chemical data

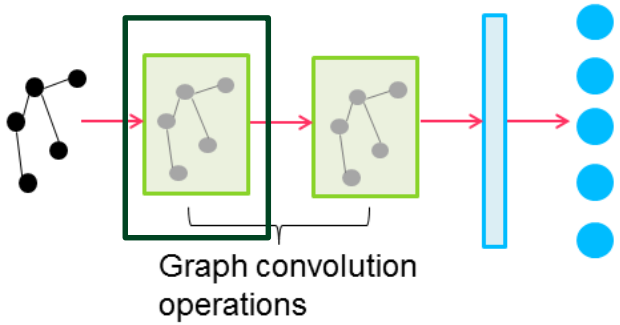


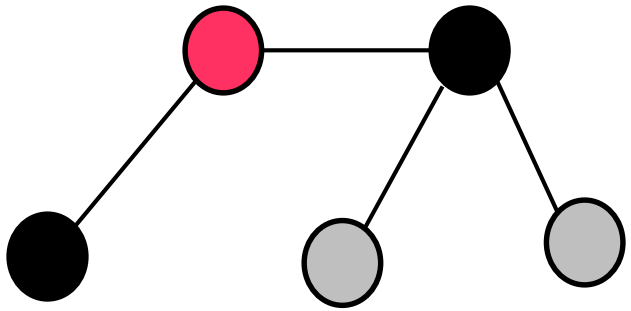
Concept: represent the molecules as graphs (nodes = atoms, edges = bonds)

Learn node (atom) representations that will help with the task at hand (end-to-end learning)



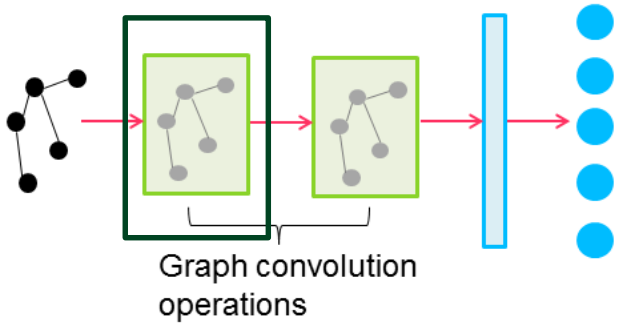
Graph convolution operations

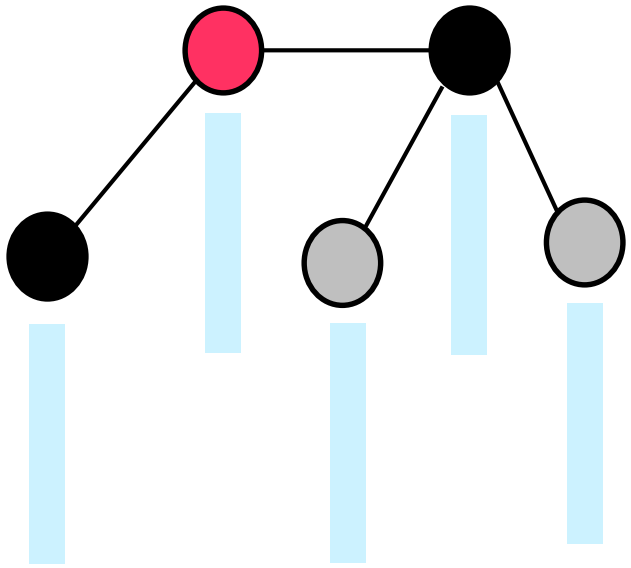




1. Feature propagation along the graph and affine transformation

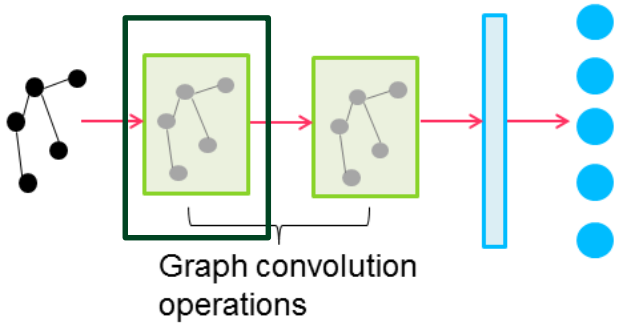
Graph convolution operations

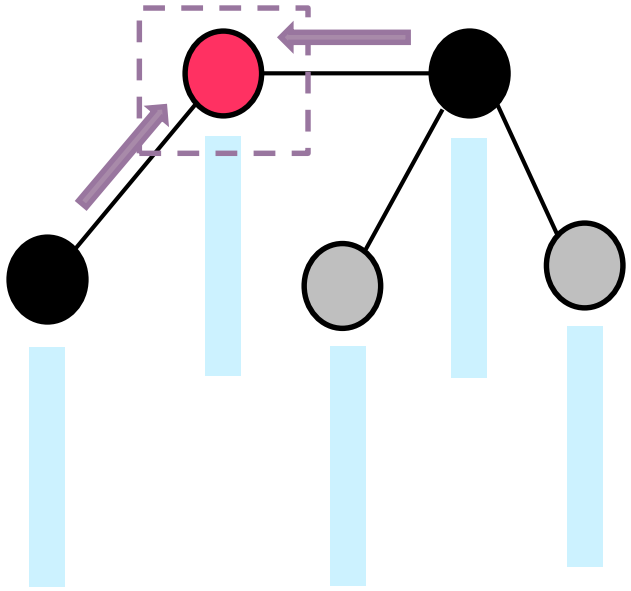




1. Feature propagation along the graph and affine transformation

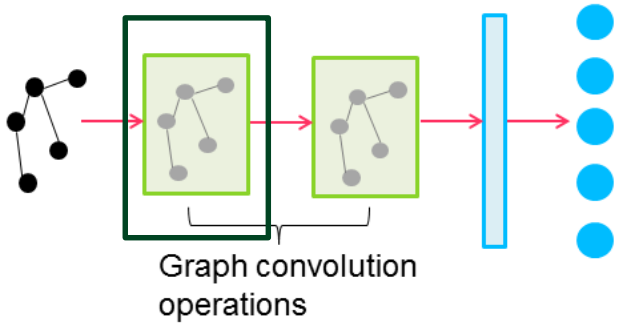
Graph convolution operations

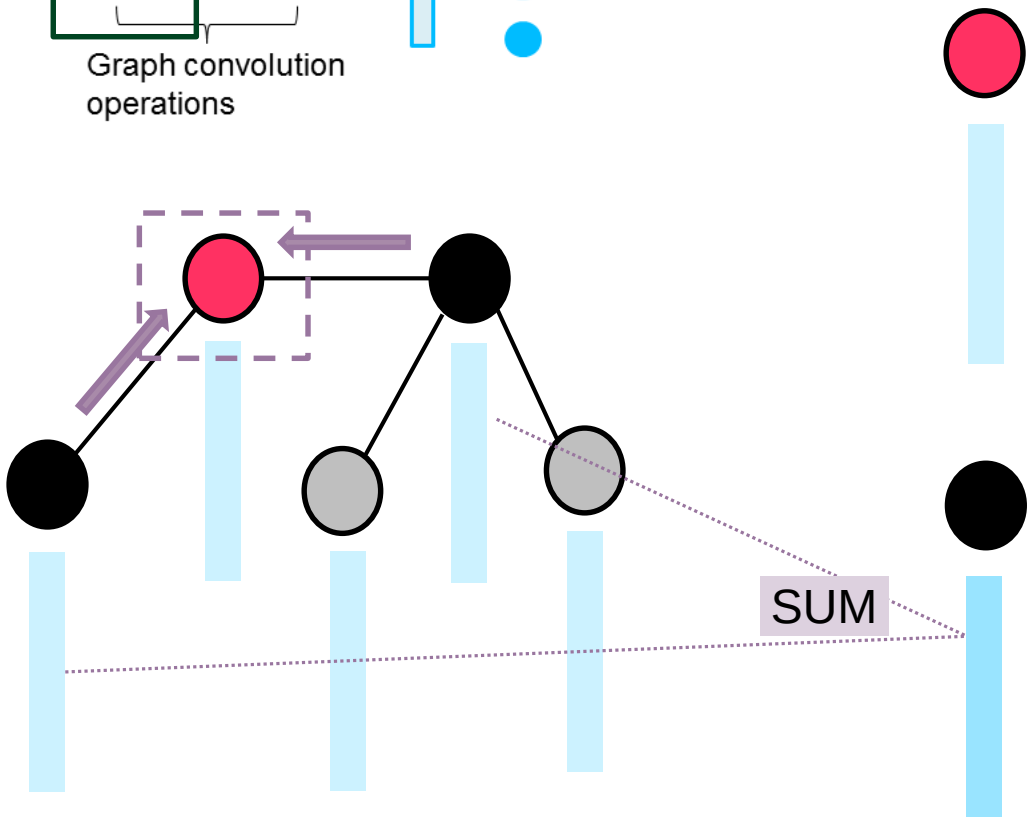




1. Feature propagation along the graph and affine transformation

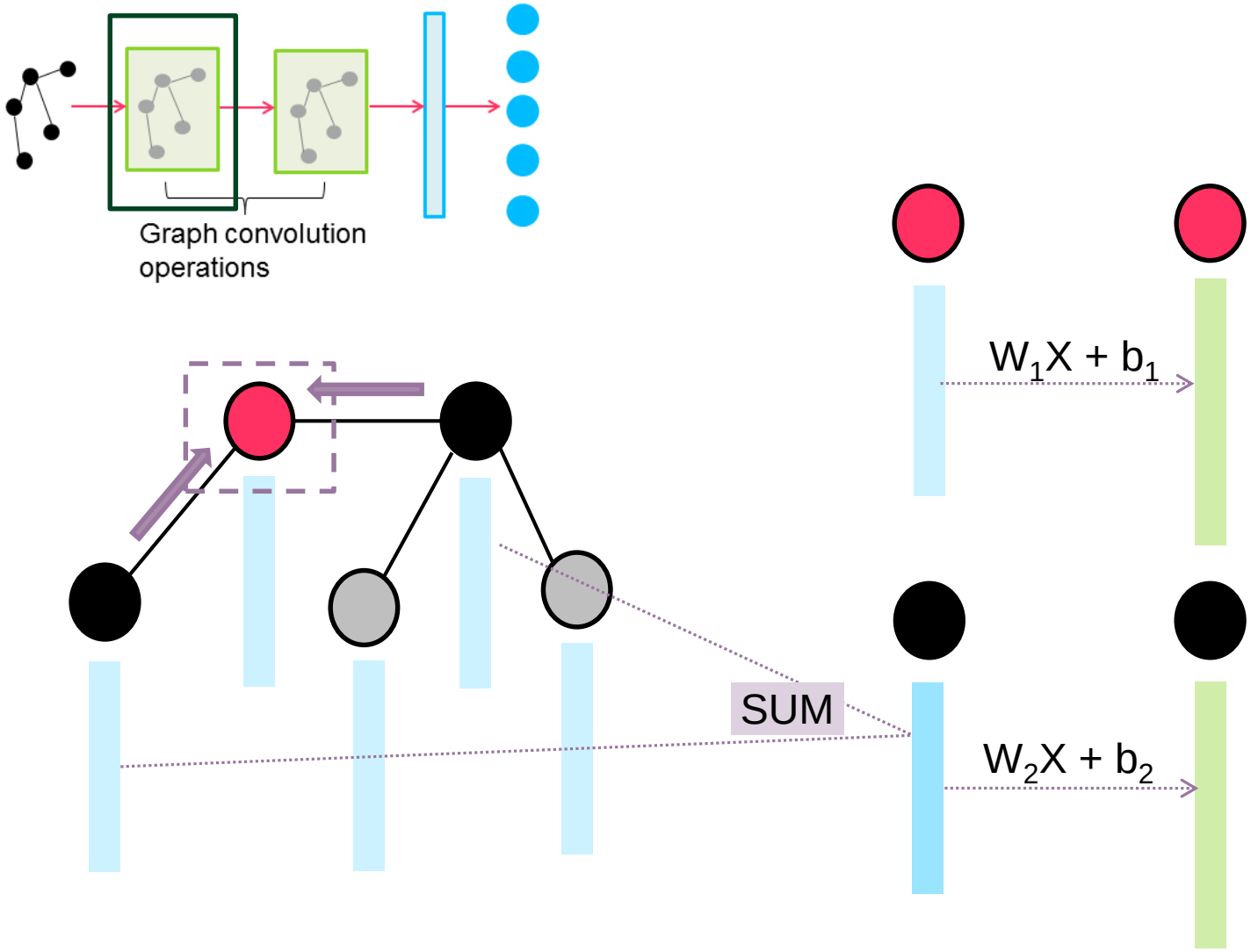
Graph convolution operations





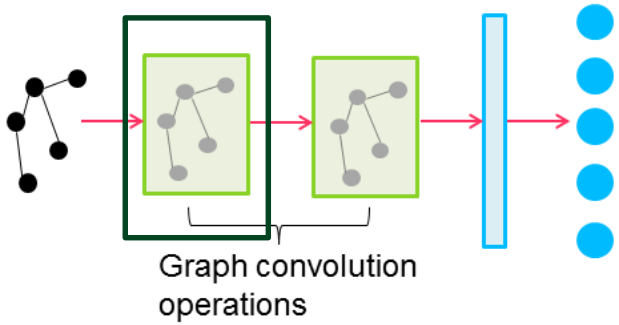
1. Feature propagation along the graph and affine transformation

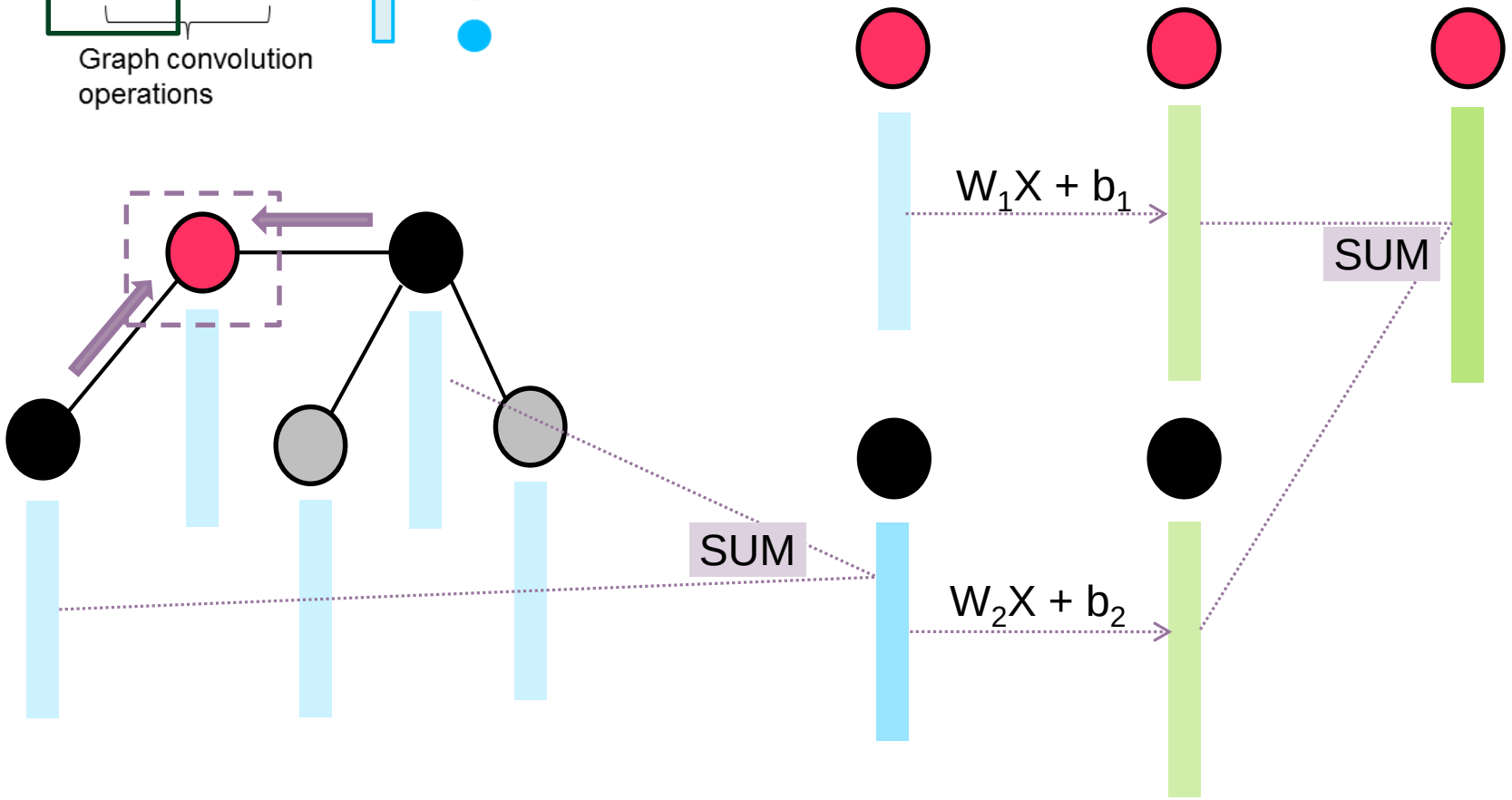
Graph convolution operations



1. Feature propagation along the graph and affine transformation

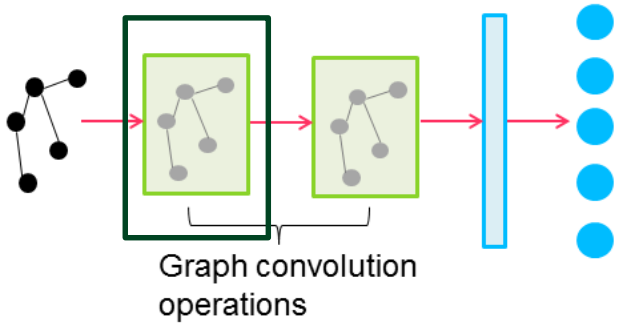
Graph convolution operations

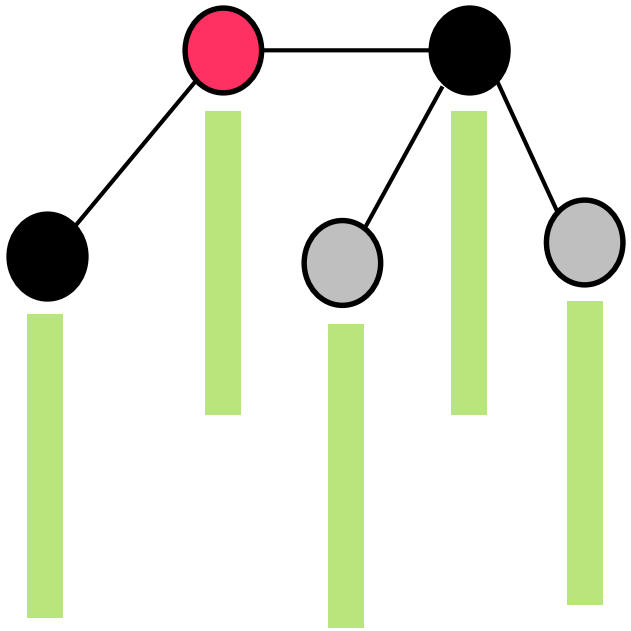




1. Feature propagation along the graph and affine transformation

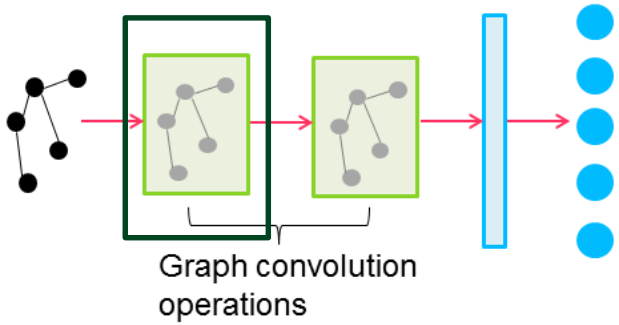
Graph convolution operations



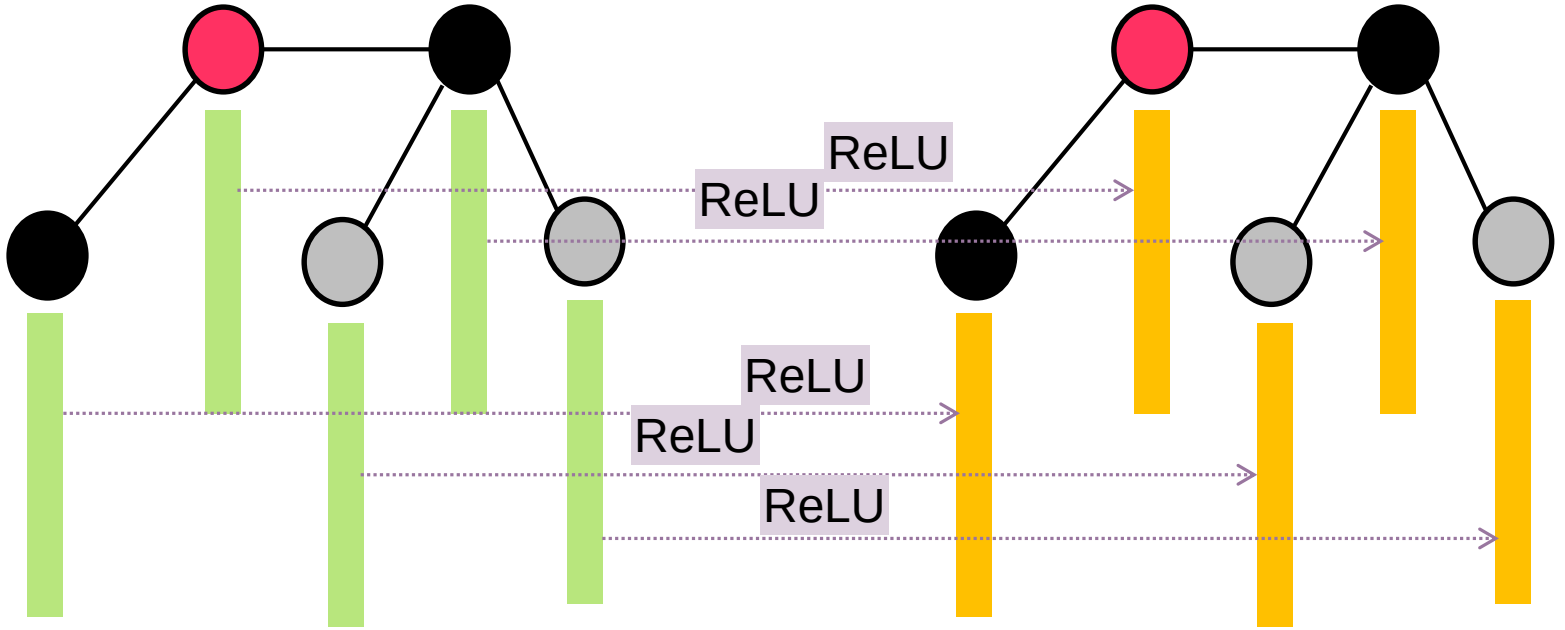


2. Non linearity
(3. Batch norm
4. Dropout)

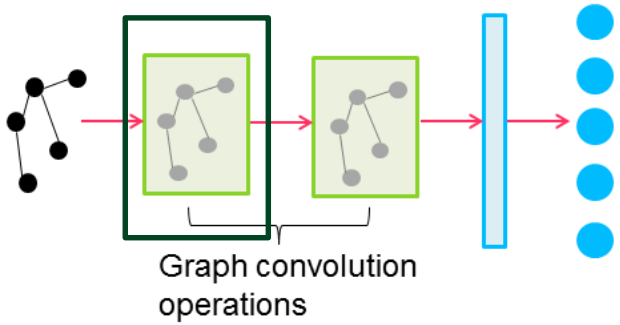
Graph convolution operations



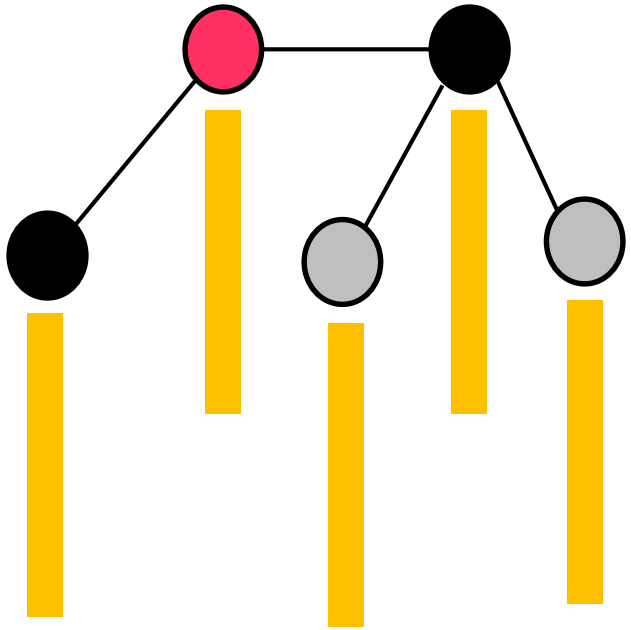
2. Non linearity
(3. Batch norm
4. Dropout)



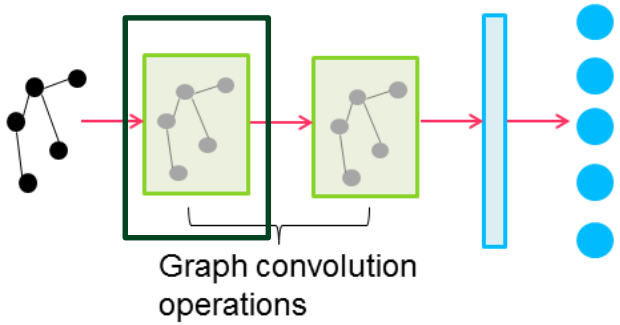
Graph convolution operations

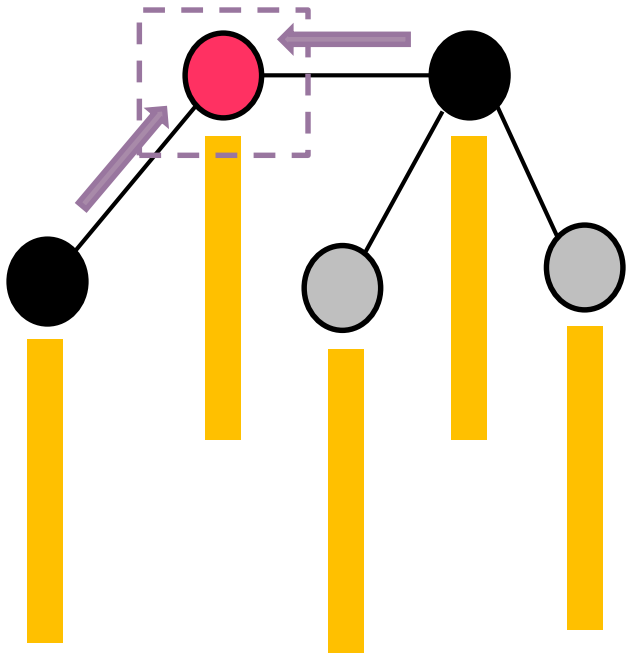


5. Graph pooling



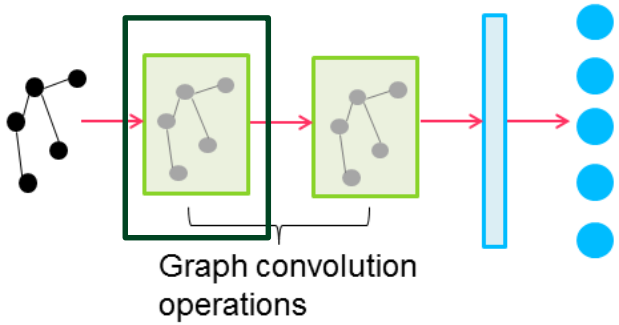
Graph convolution operations



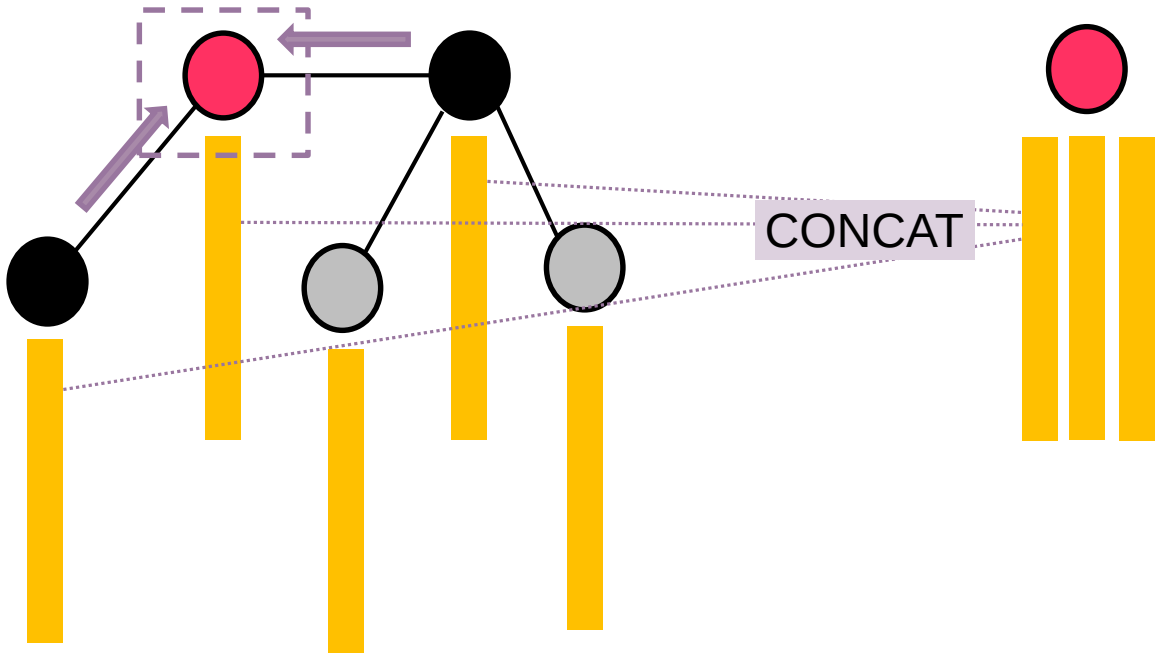


5. Graph pooling

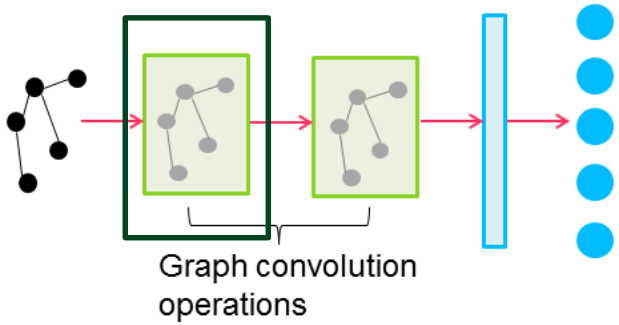
Graph convolution operations



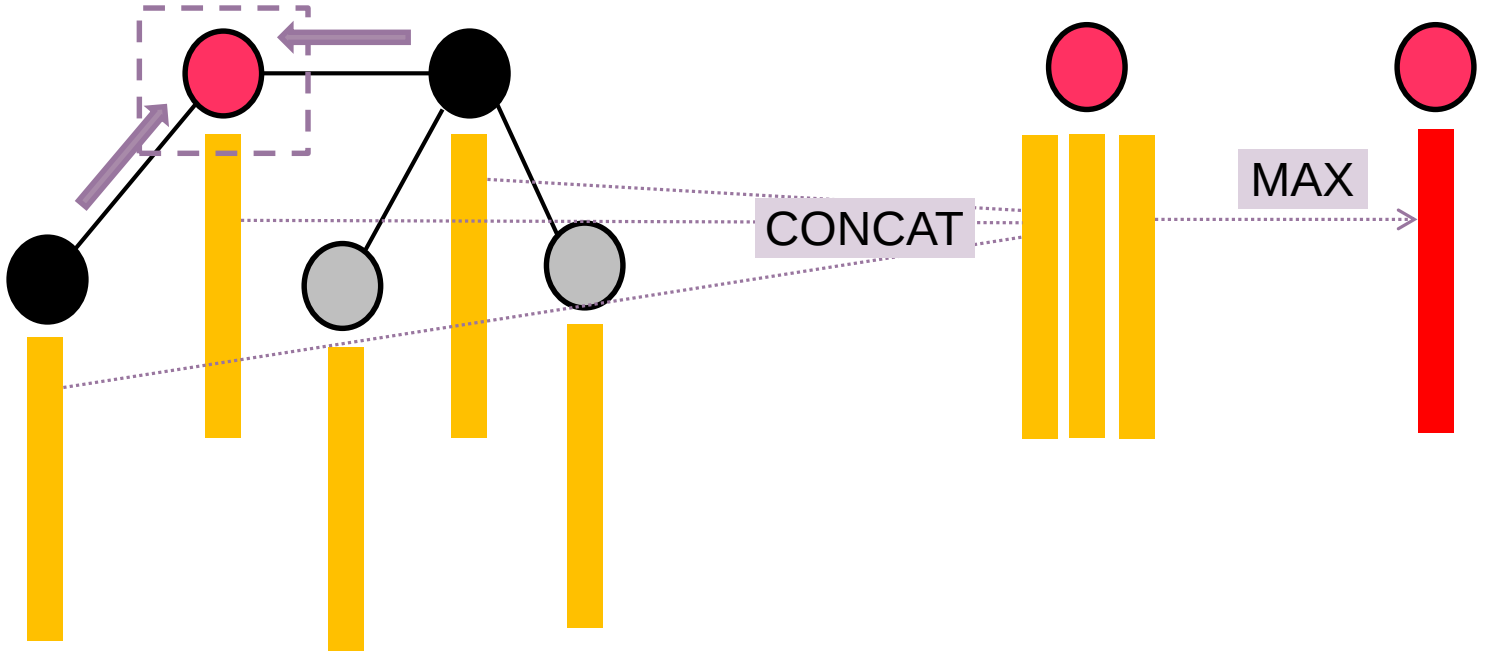
5. Graph pooling



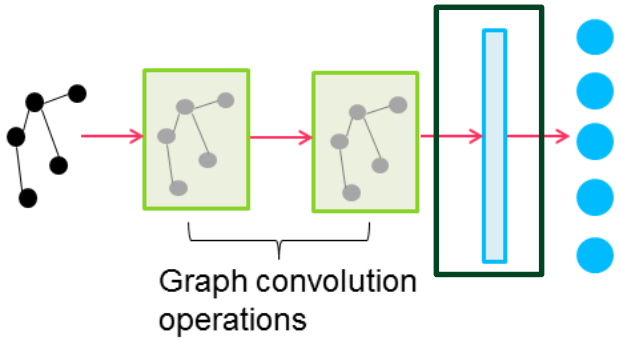
Graph convolution operations



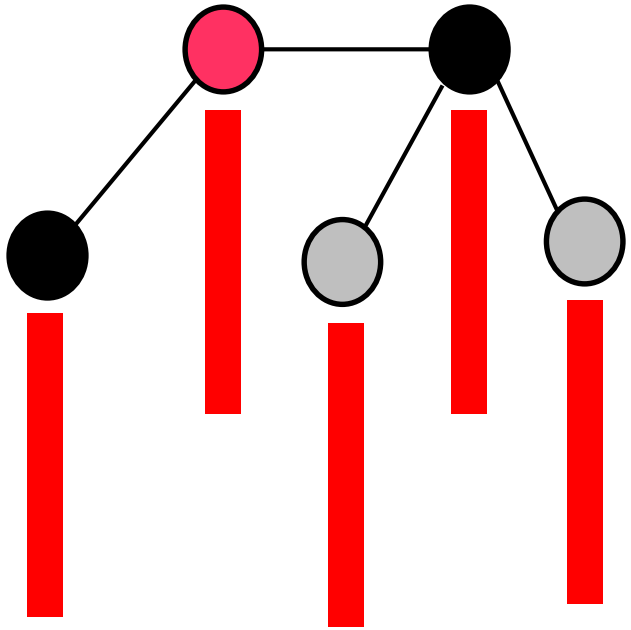
5. Graph pooling



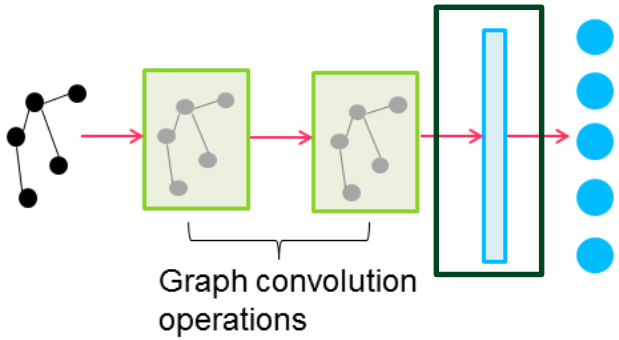
Graph convolution operations



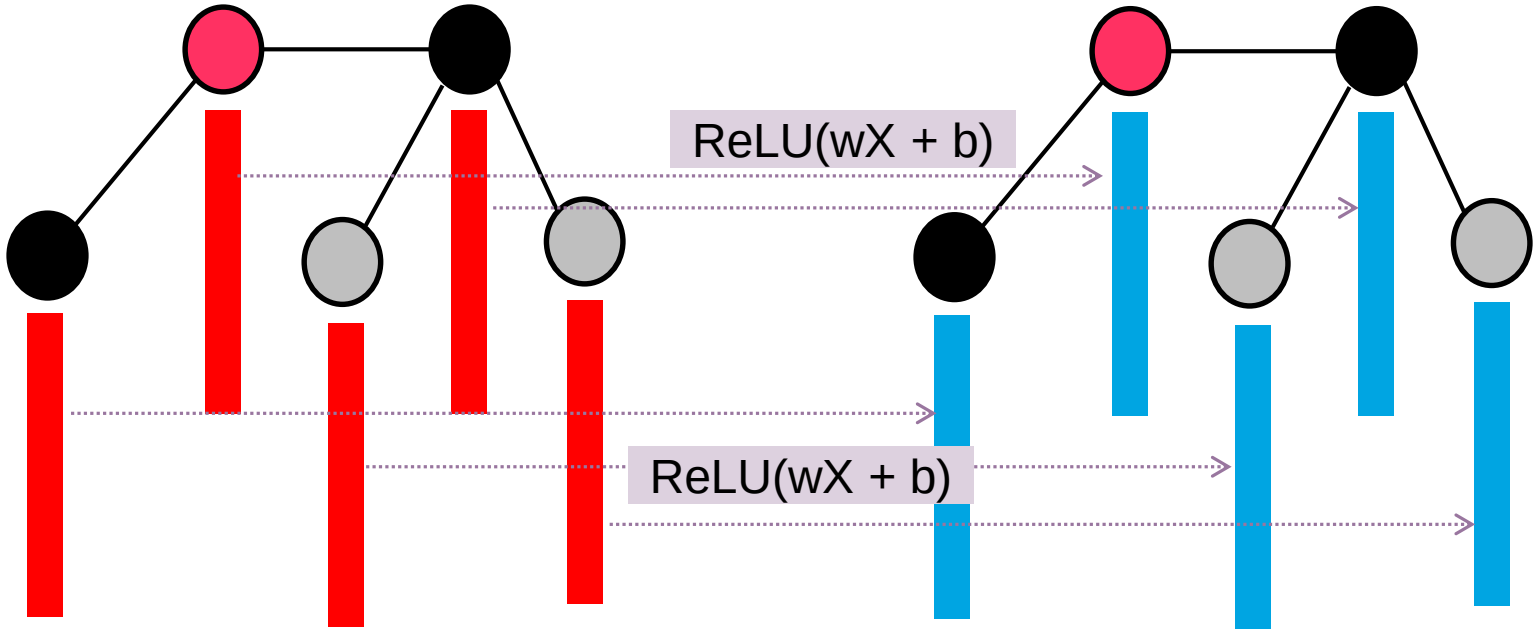
5. Atom-level dense layer
(6. Batch Norm
7. Dropout)



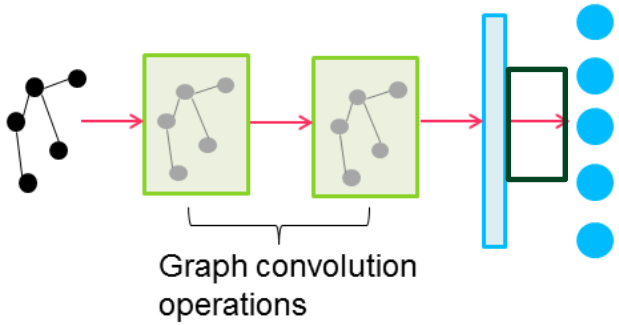
Graph convolution operations



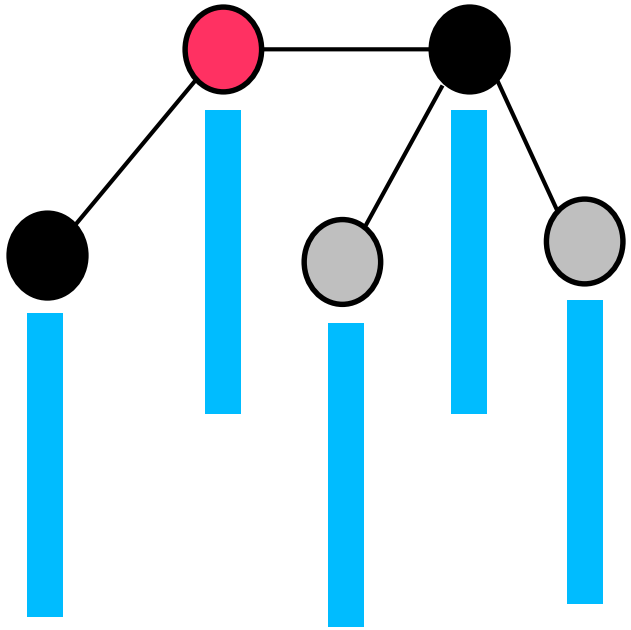
5. Atom-level dense layer
(6. Batch Norm
7. Dropout)



Going back to a molecule-level representation

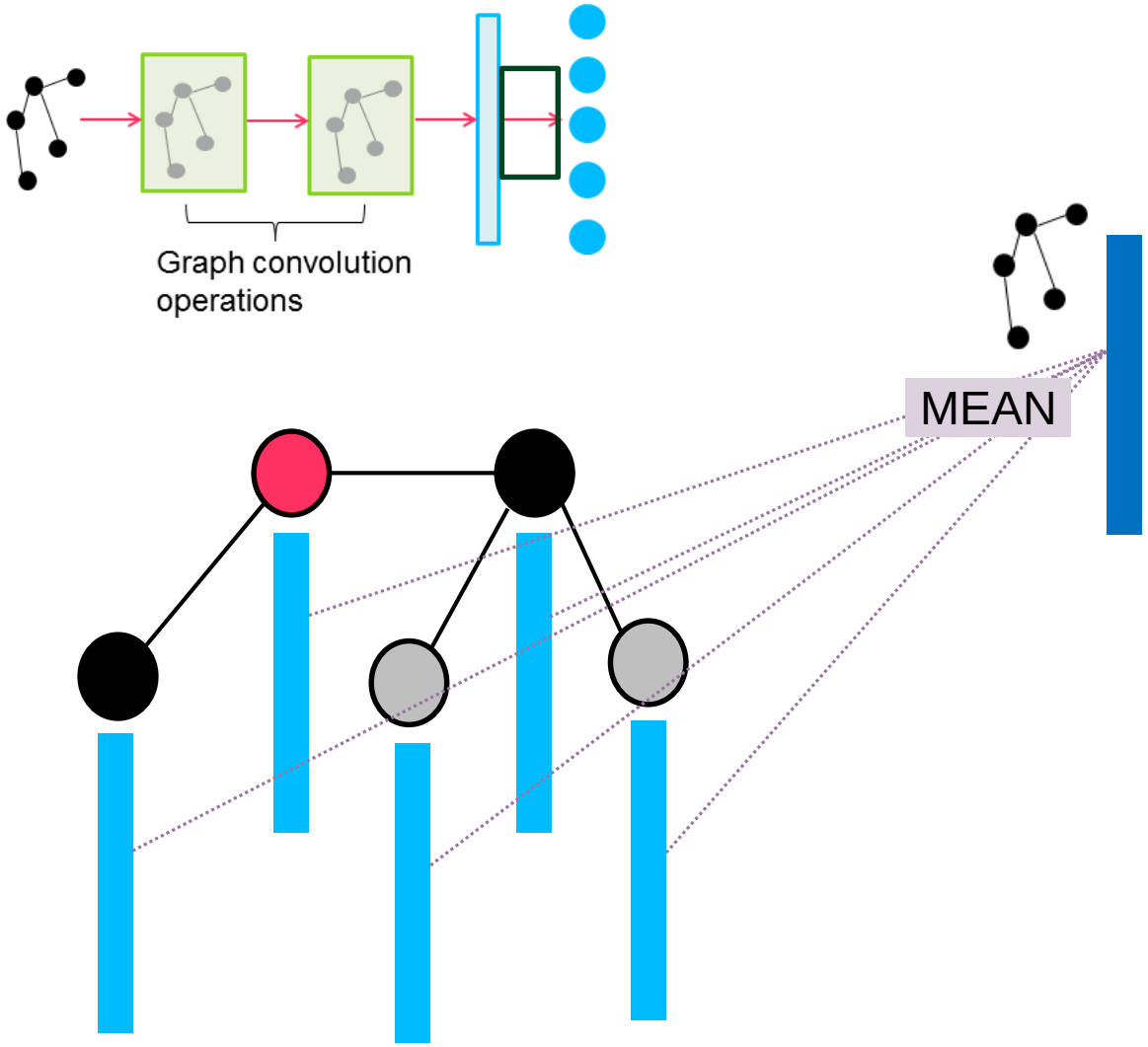


8. GraphGather operation



DeepChem 2.1
https://github.com/deepchem

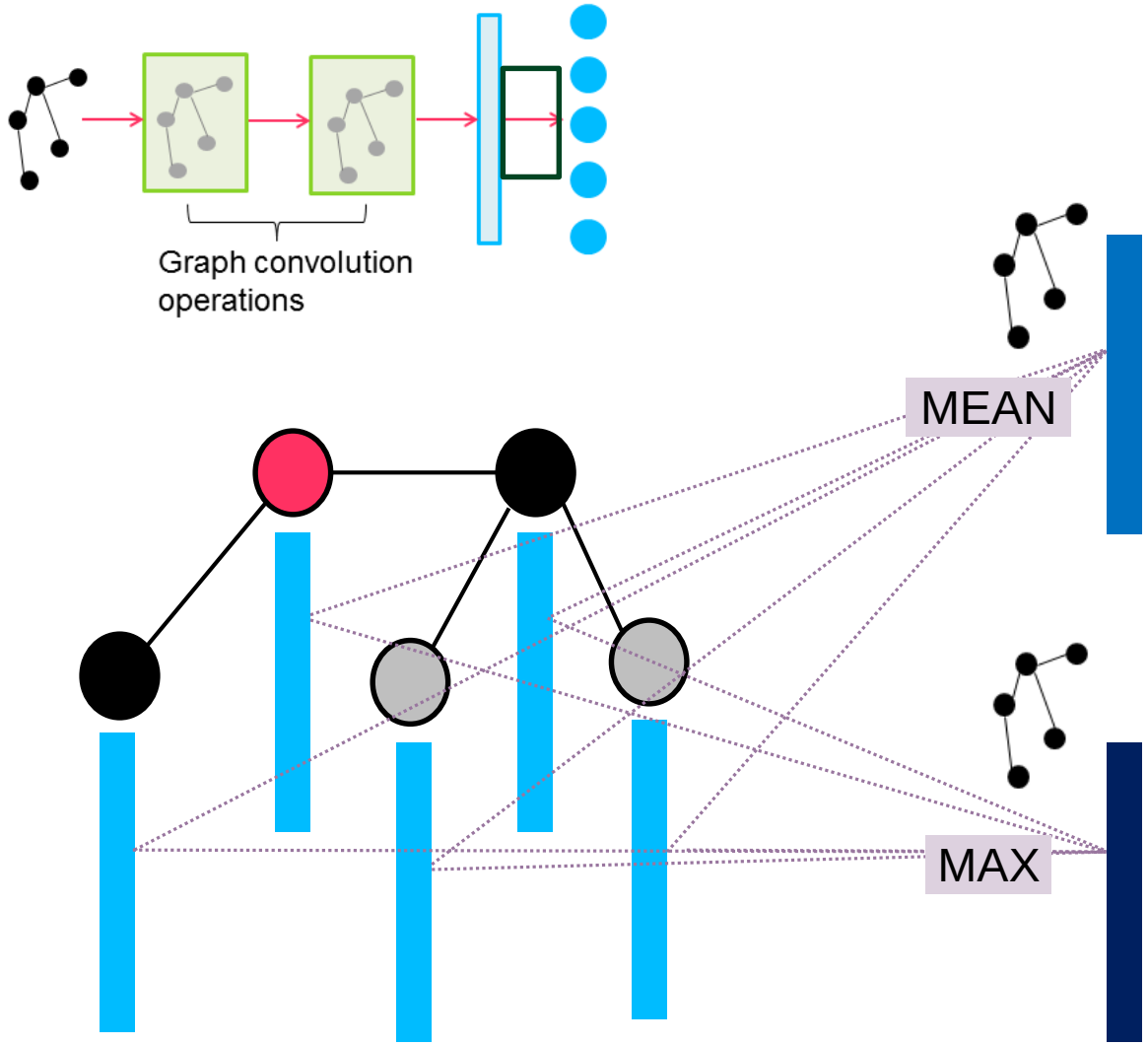
Going back to a molecule-level representation



8. GraphGather operation

DeepChem 2.1
https://github.com/deepchem

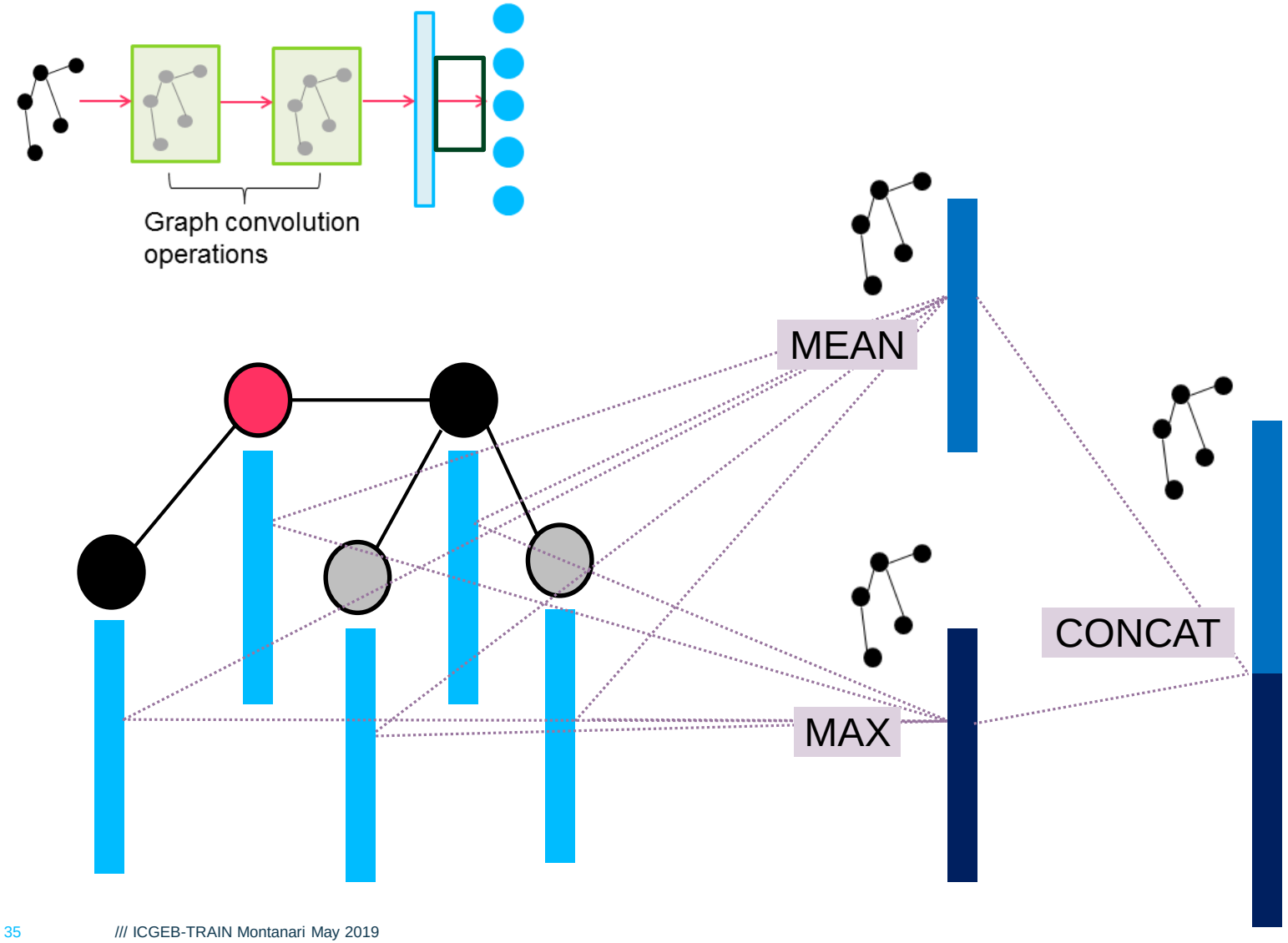
Going back to a molecule-level representation



8. GraphGather operation

DeepChem 2.1
https://github.com/deepchem

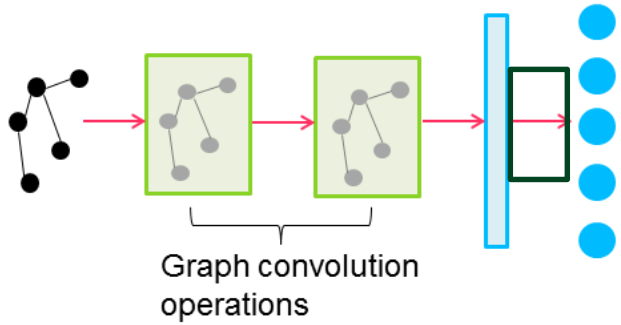
Going back to a molecule-level representation



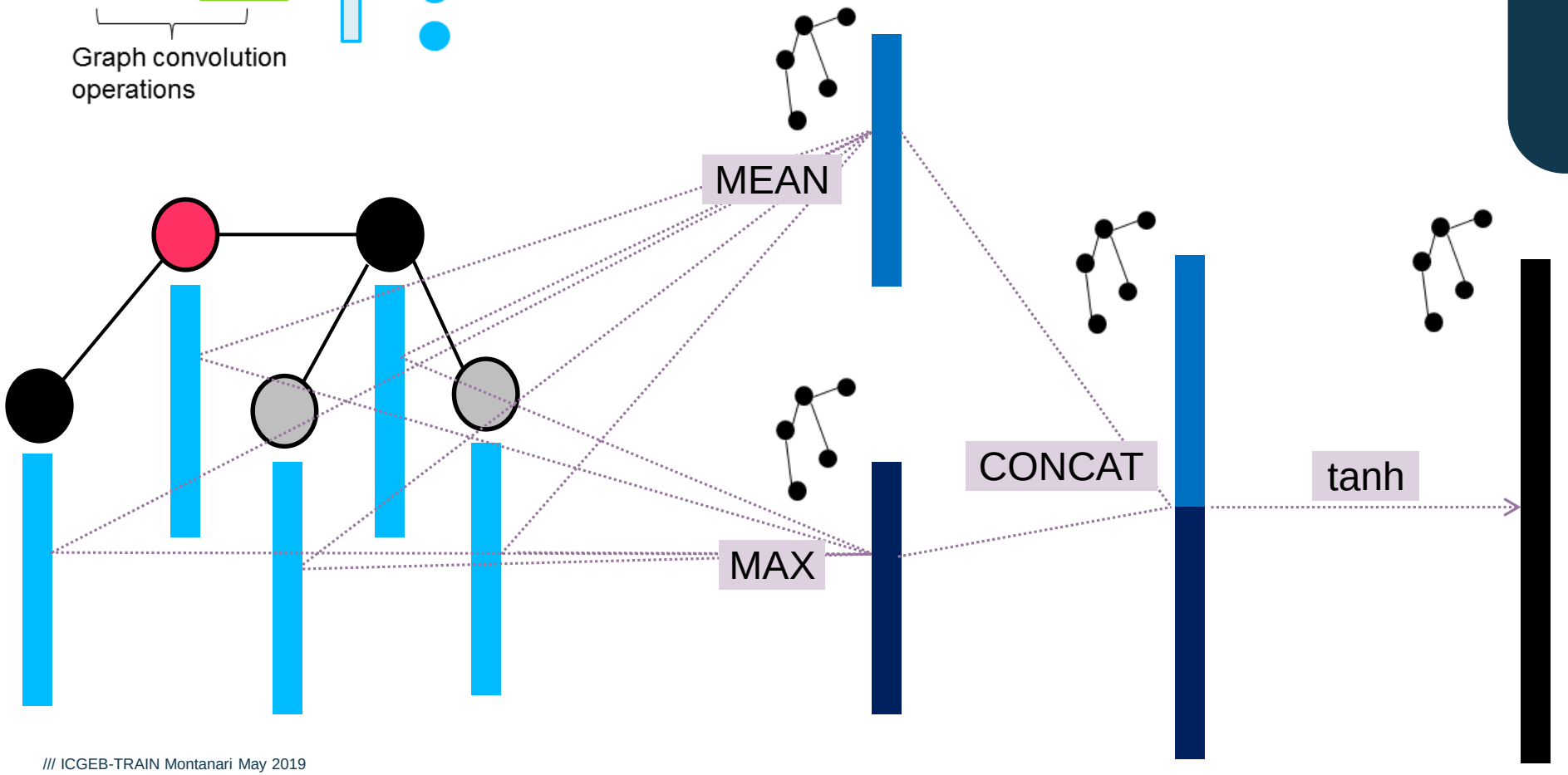
8. GraphGather operation

DeepChem 2.1
https://github.com/deepchem

Going back to a molecule-level representation



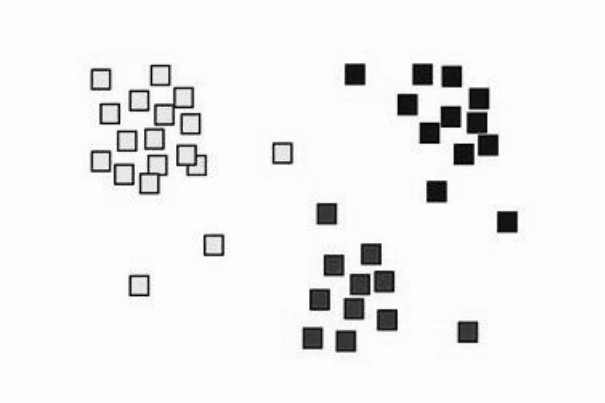
8. GraphGather operation



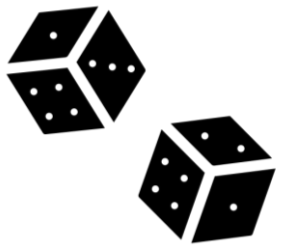
DeepChem 2.1
https://github.com/deepchem

Model evaluation

Cross-validation



Cluster CV



Random CV

Separate test set

Strict or taskwise time split

Metrics

R^2 (in CV or test set)
Spearman‘s rho (ranking capability)

Absorption – Distribution: physico-chemical properties

Solubility

- *Nephelometry*:	88 000 measured cpds
- PBS pH 6.5 from DMSO	38 800 measured cpds
- PBS pH 6.5 from Powder	2 300 measured cpds
- *PBS pH 6.5 from DMSO* *not fully dissolved*	7 300 measured cpds
- *PBS pH 6.5 unknown starting point*	50 000 measured cpds

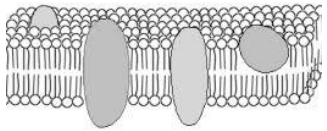
LogD

- pH 7.5	88 000 measured cpds
- pH 2.3	236 000 measured cpds

Melting point

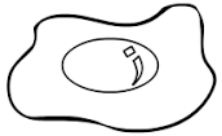
92 000 compounds

Membrane affinity



66 800 compounds

Serum albumin binding

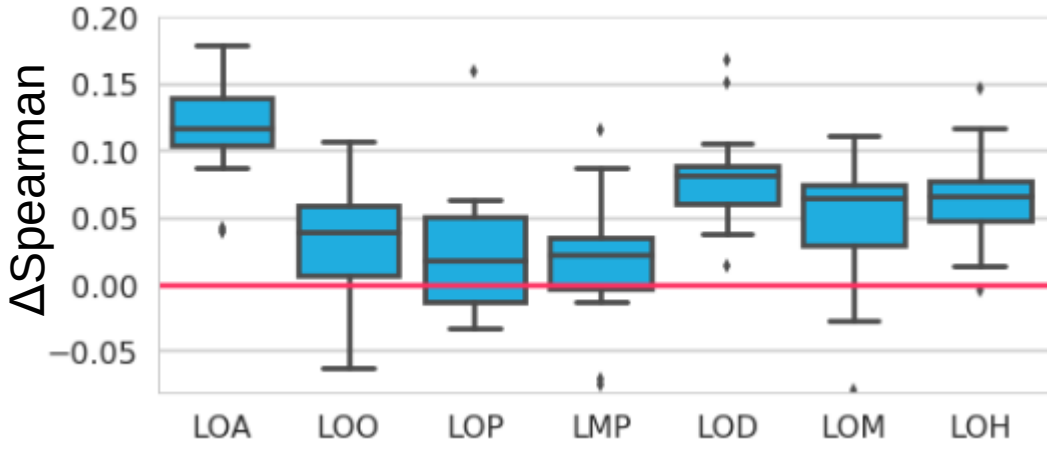


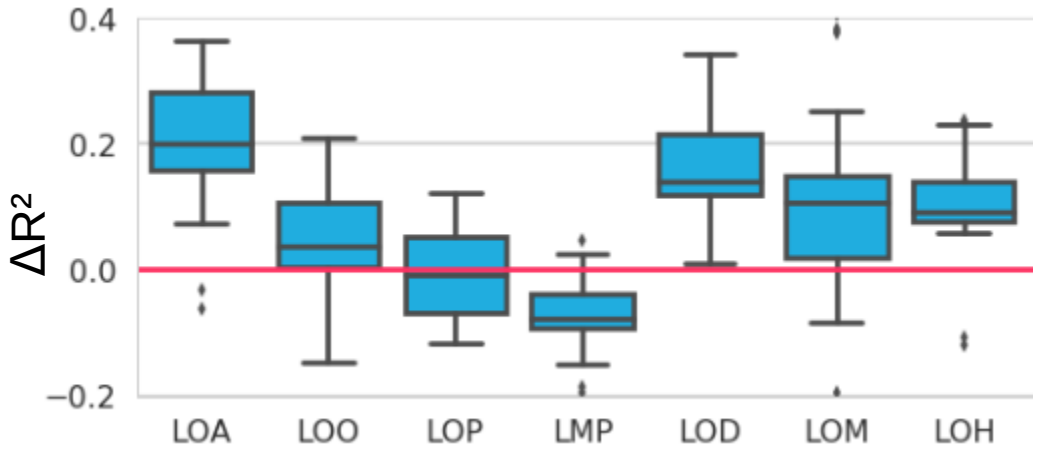
64 000 compounds

Absorption – Distribution: physico-chemical properties

Random Forest versus single task neural networks

Average over 4 leave-cluster-out CV experiments, networks hyperparamters were only optimized on task *LMP, Spearman*



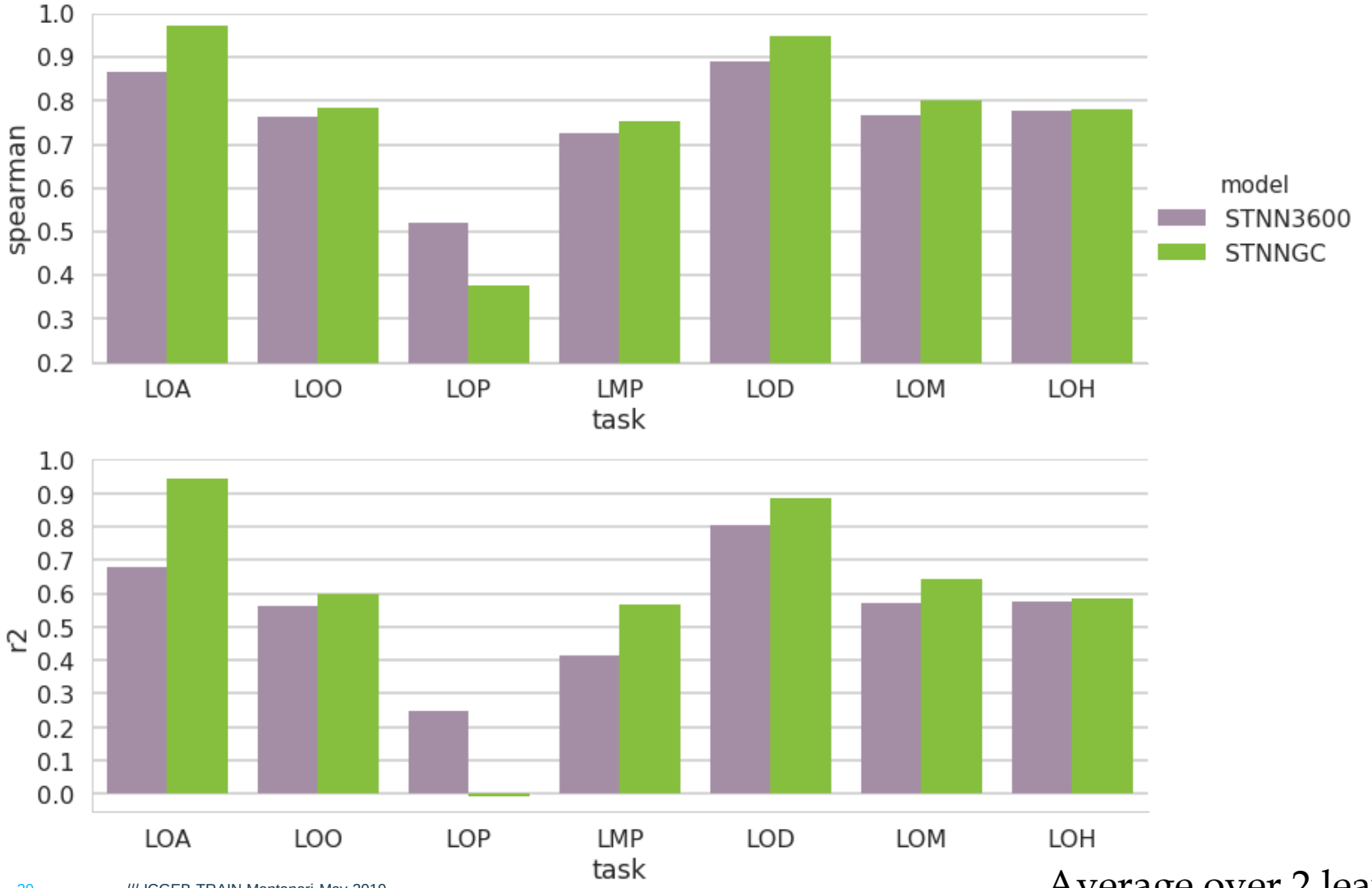


Neural networks (fully connected, same features as RF) overperform RF for the physchem properties

LogD acid / Solubility DMSO / Solubility powder / melting point / LogD / Membrane affinity / HSA binding

Absorption – Distribution: physico-chemical properties

Fully connected networks versus graph convolutional networks



Graph convolution brings better performance on average, especially true for the larger tasks. LOP is very small (≈2000 cpds) so probably graph conv is overfitting.

Average over 2 leave-cluster-out CV experiments

Absorption – Distribution: physico-chemical properties

Best model: multitask graph convolutional network

Average over 2 leave-cluster-out CV experiments

	R^2	Spearman	RMSE
LogD pH 7.5	0.88	0.94	0.34
LogD pH 2.3	0.91	0.96	0.36
Membrane affinity	0.71	0.84	0.51
hSA binding	0.63	0.82	0.50
Melting point	0.53	0.74	39
Solubility DMSO	0.58	0.77	0.83
Solubility Powder	0.55	0.75	0.79

Absorption – Distribution: physico-chemical properties

Best model: multitask graph convolutional network

Average over 2 leave-cluster-out CV experiments

	R^2	Spearman	RMSE
LogD pH 7.5	0.88	0.94	0.34
LogD pH 2.3	0.91	0.96	0.36
Membrane affinity	0.71	0.84	0.51
hSA binding	0.63	0.82	0.50
Melting point	0.53	0.74	39
Solubility DMSO	0.58	0.77	0.83
Solubility Powder	0.55	0.75	0.79

Excellent performance for all modeled endpoints and significant improvement over models previously in production.

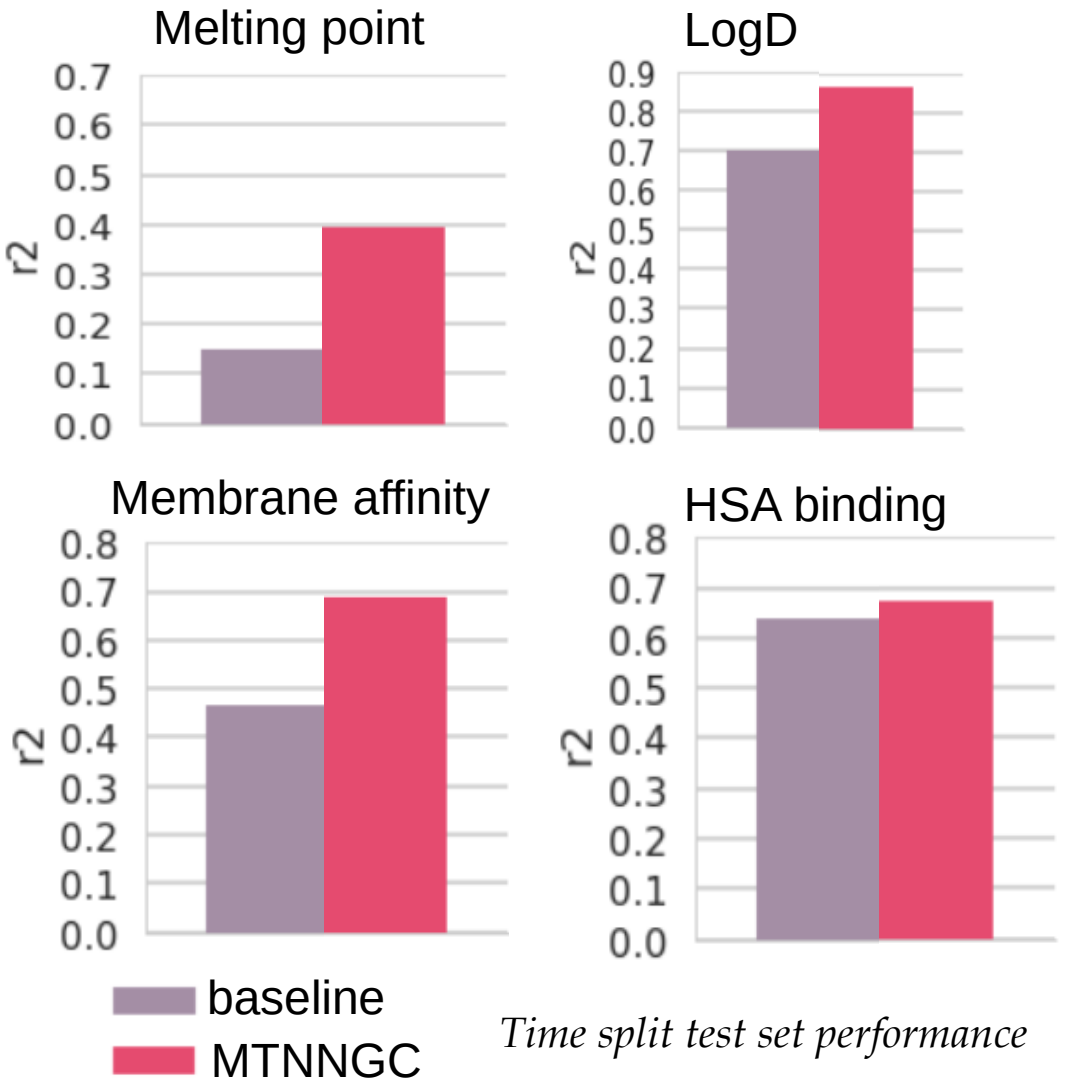
Absorption – Distribution: physico-chemical properties

Best model: multitask graph convolutional network

Average over 2 leave-cluster-out CV experiments

	R^2	Spearman	RMSE
LogD pH 7.5	0.88	0.94	0.34
LogD pH 2.3	0.91	0.96	0.36
Membrane affinity	0.71	0.84	0.51
hSA binding	0.63	0.82	0.50
Melting point	0.53	0.74	39
Solubility DMSO	0.58	0.77	0.83
Solubility Powder	0.55	0.75	0.79

Excellent performance for all modeled endpoints and significant improvement over models previously in production.



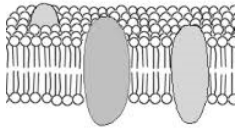
Time split test set performance

Absorption – Distribution: physico-chemical properties

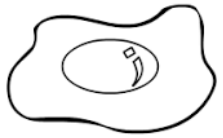
Impact of helper tasks and effect of multitask learning on small vs large tasks

LogD acid

LogD neutral



Memb. affinity



HSA binding

Melting point

Solubility

In multitask

With helper task

Helper tasks (other solubility assays) help slightly with the solubility endpoints. They do not really influence other endpoints.
Multitask learning penalizes the largest task (LogD acid) but in general benefits the smaller tasks.

Wrap-up

With the current amount of data for physico-chemical properties, Deep Learning boosts performance with respect to classical ML models.

Graph convolutional networks are very poweful for those assays once the training set size is large enough.

Multitask learning improves the performance on all but the largest task, and adding more related tasks also can help.

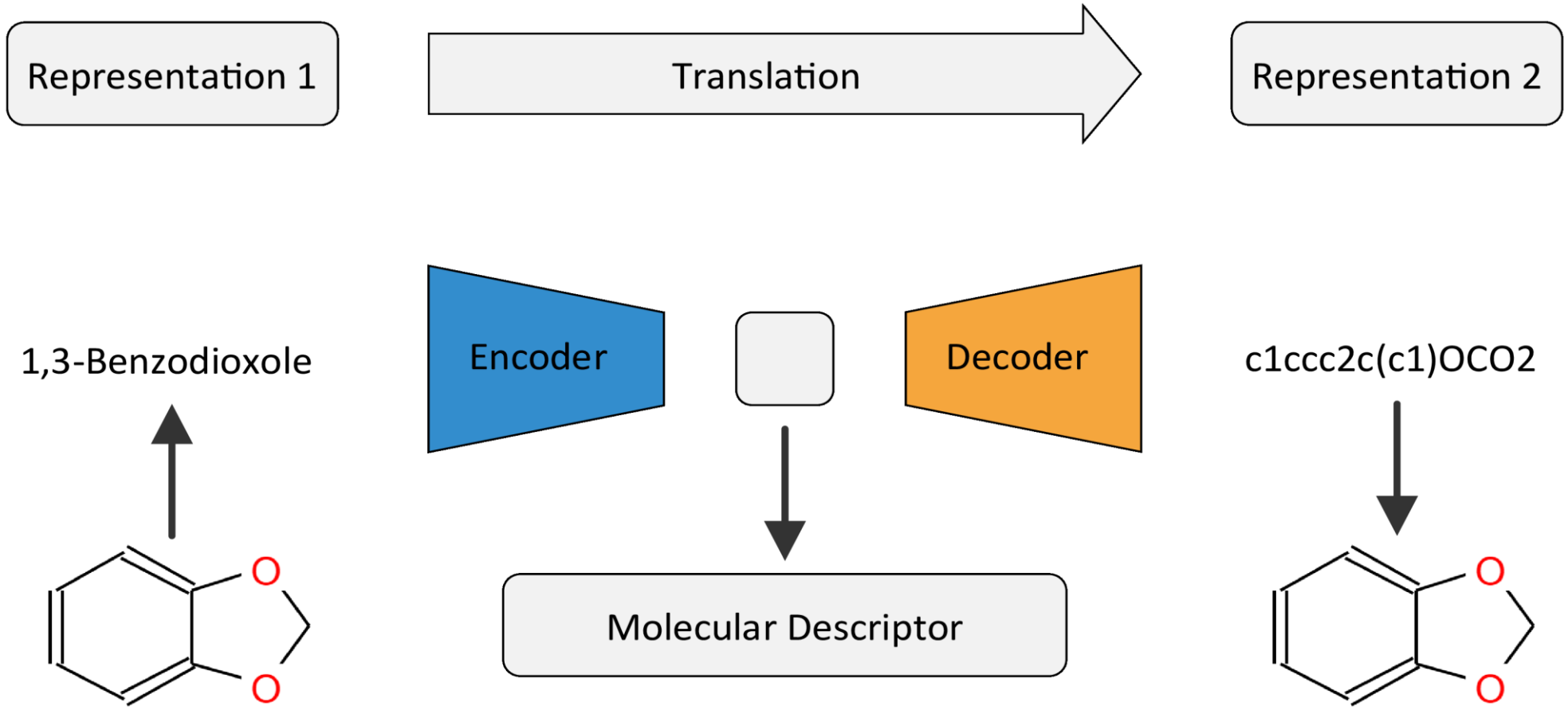
When doing multitask learning, one has to take care of a few things: that the outputs are in the same range and maybe that some smaller task losses must be over-weighted.

Current model is in production and used by the medicinal chemists at Bayer.

Other endpoints have been modeled, graph convolutional networks are not always the best! and it is hard to know how to group tasks, but overall one can get better performance with DL compared to Random Forest.

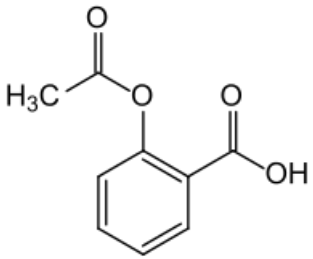
Molecule Swarm Optimization (MSO)

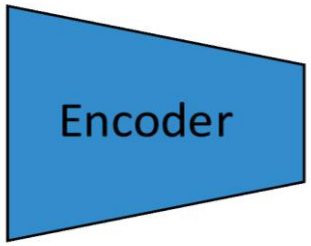
Introduction: navigating the CDDD chemical space



Winter et al., *ChemRxiv*, 2019

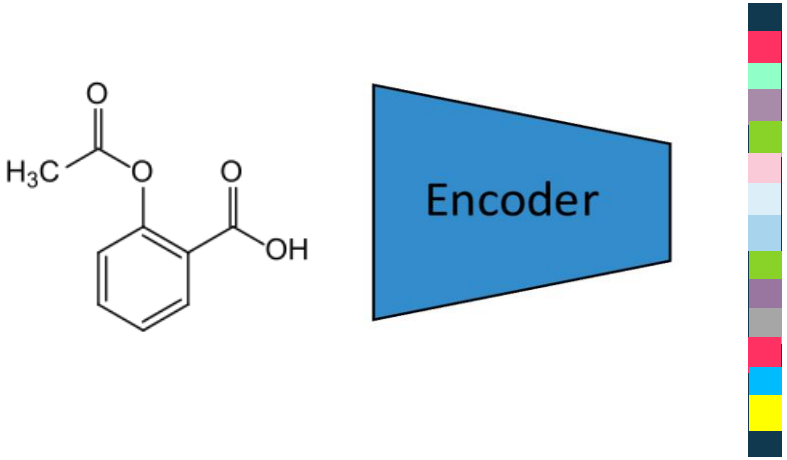
Introduction: navigating the CDDD chemical space





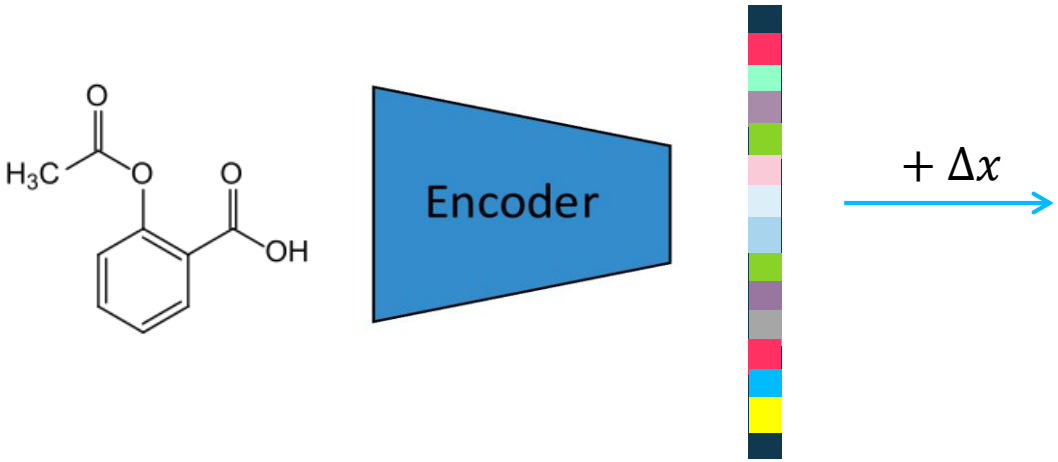
Winter et al., *ChemRxiv*, 2019

Introduction: navigating the CDDD chemical space



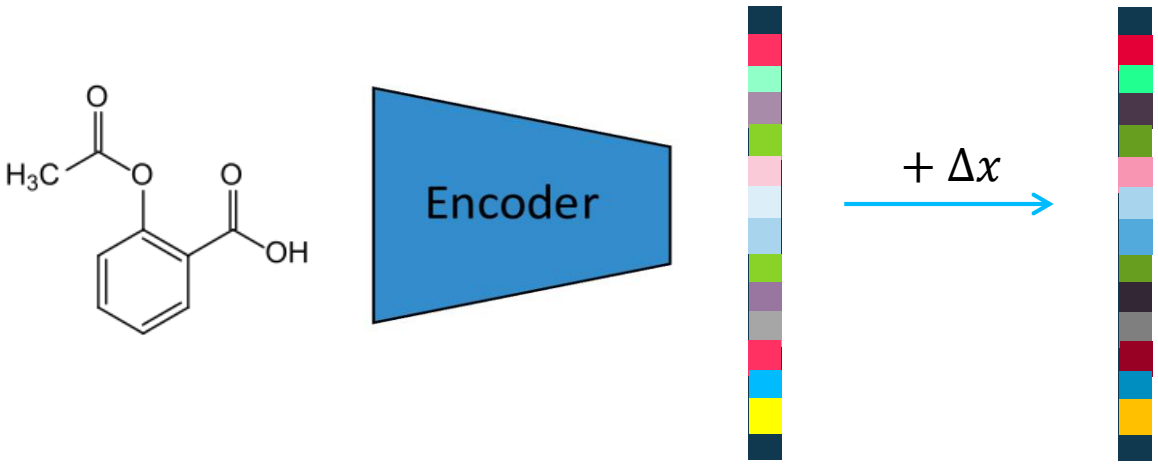
Winter et al., *ChemRxiv*, 2019

Introduction: navigating the CDDD chemical space



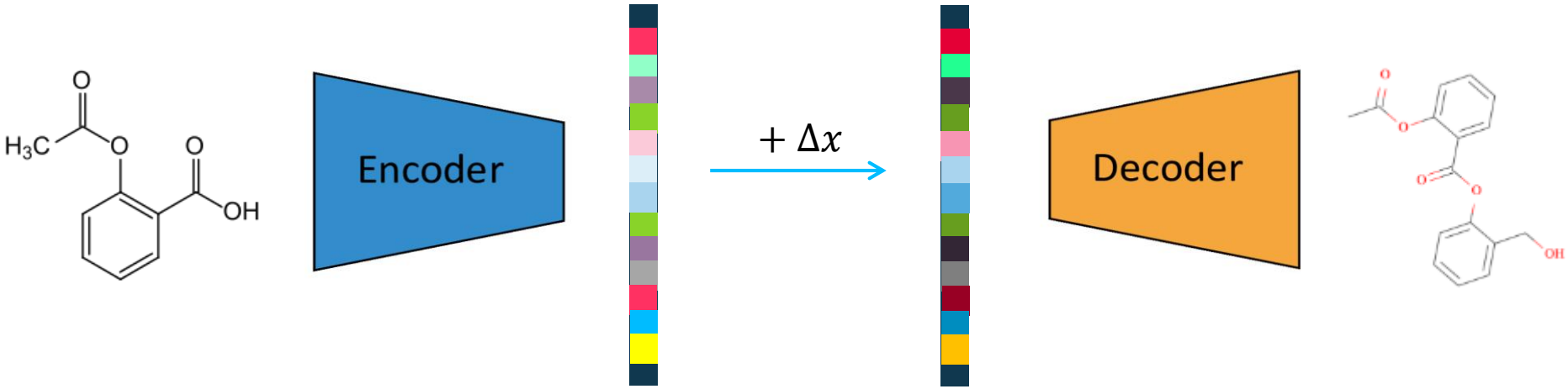
Winter et al., *ChemRxiv*, 2019

Introduction: navigating the CDDD chemical space



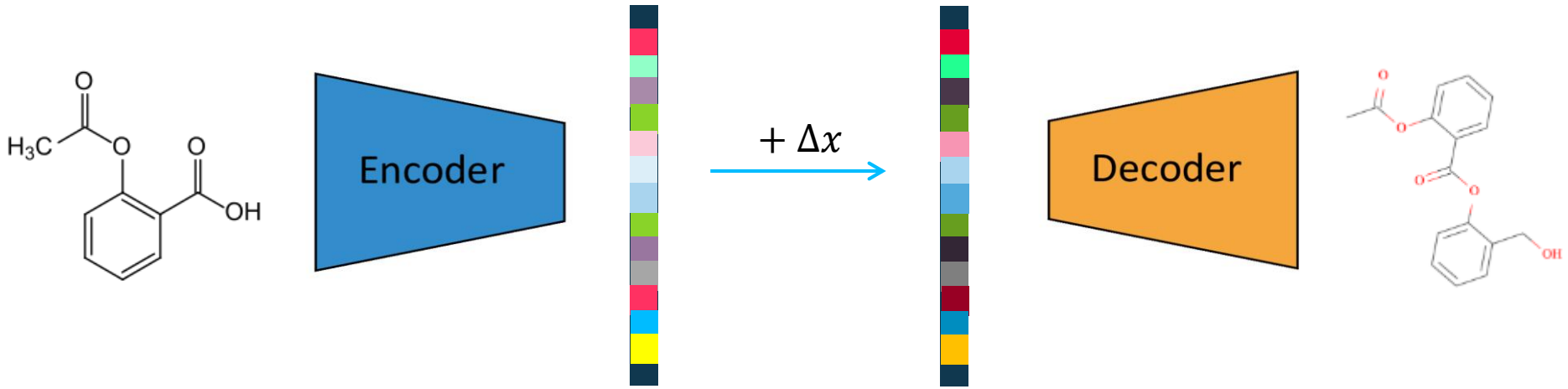
Winter et al., *ChemRxiv*, 2019

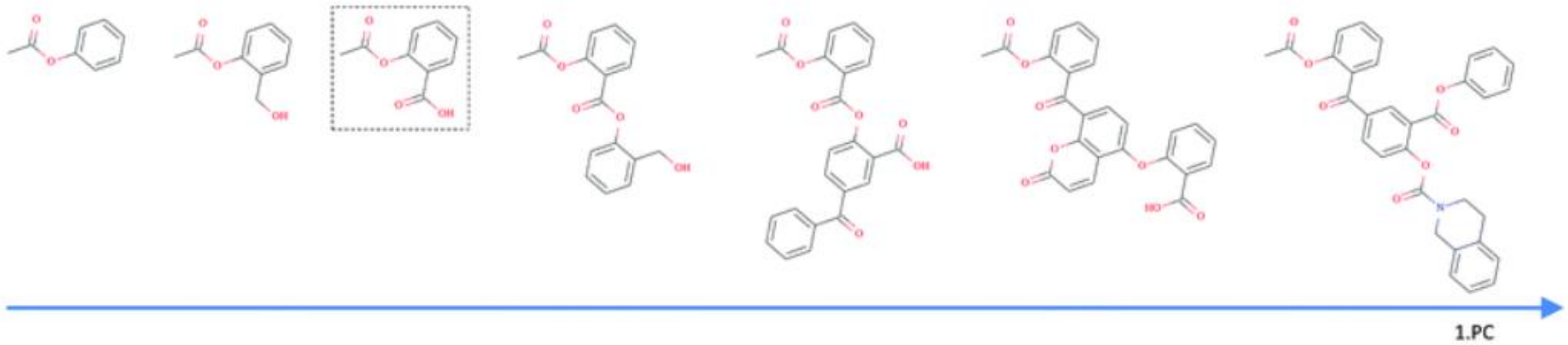
Introduction: navigating the CDDD chemical space



Winter et al., *ChemRxiv*, 2019

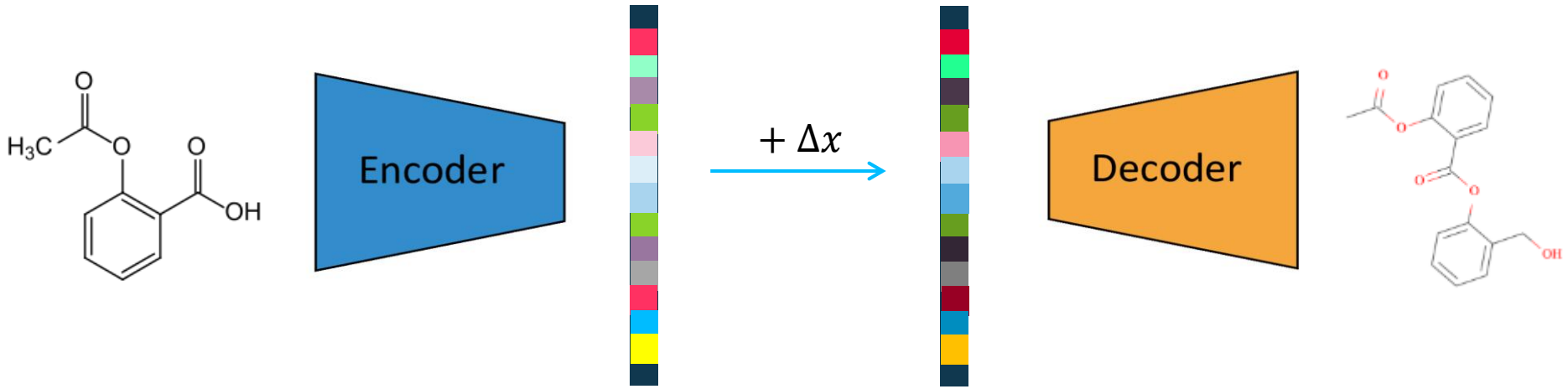
Introduction: navigating the CDDD chemical space



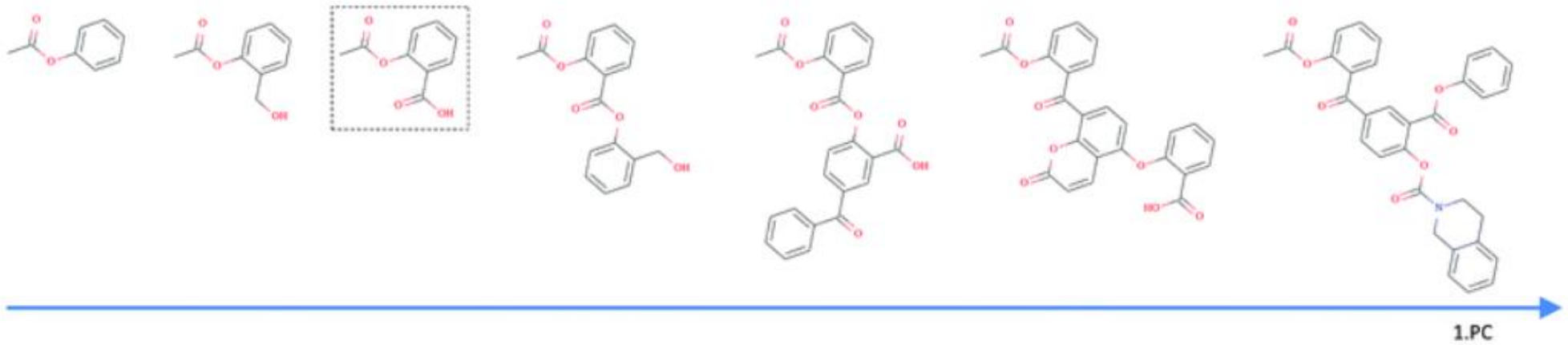


Winter et al., *ChemRxiv*, 2019

Introduction: navigating the CDDD chemical space



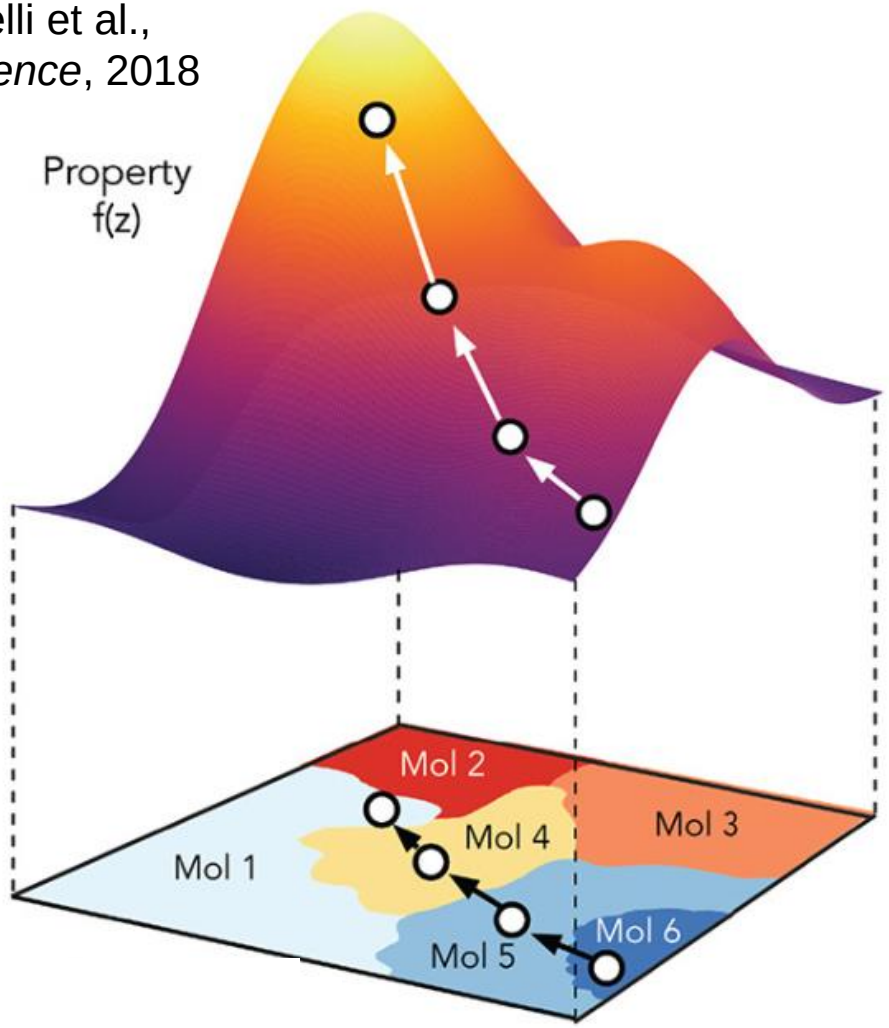
Moving in the direction of the first P.C. of the training set embedding results in smooth chemical transitions towards larger molecules.



Winter et al., *ChemRxiv*, 2019

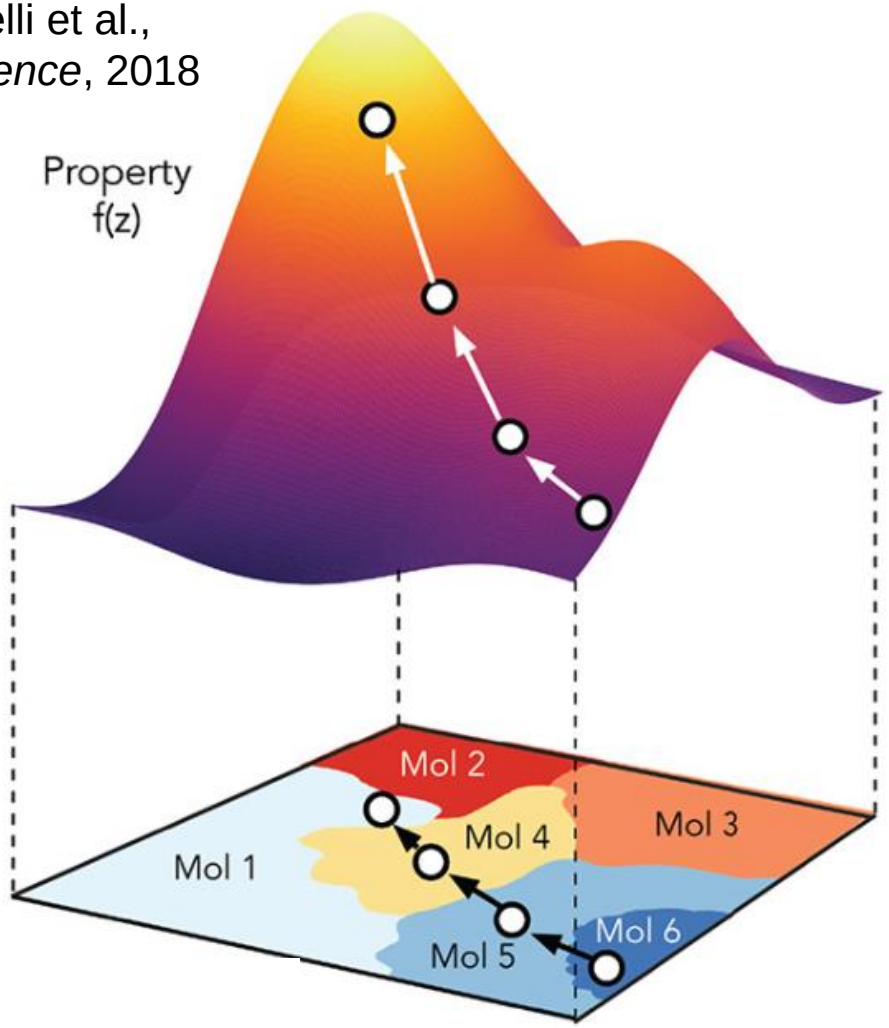
How to steer the navigation towards useful chemistry?

Gomez-Bombarelli et al.,
ACS Central Science, 2018



How to steer the navigation towards useful chemistry?

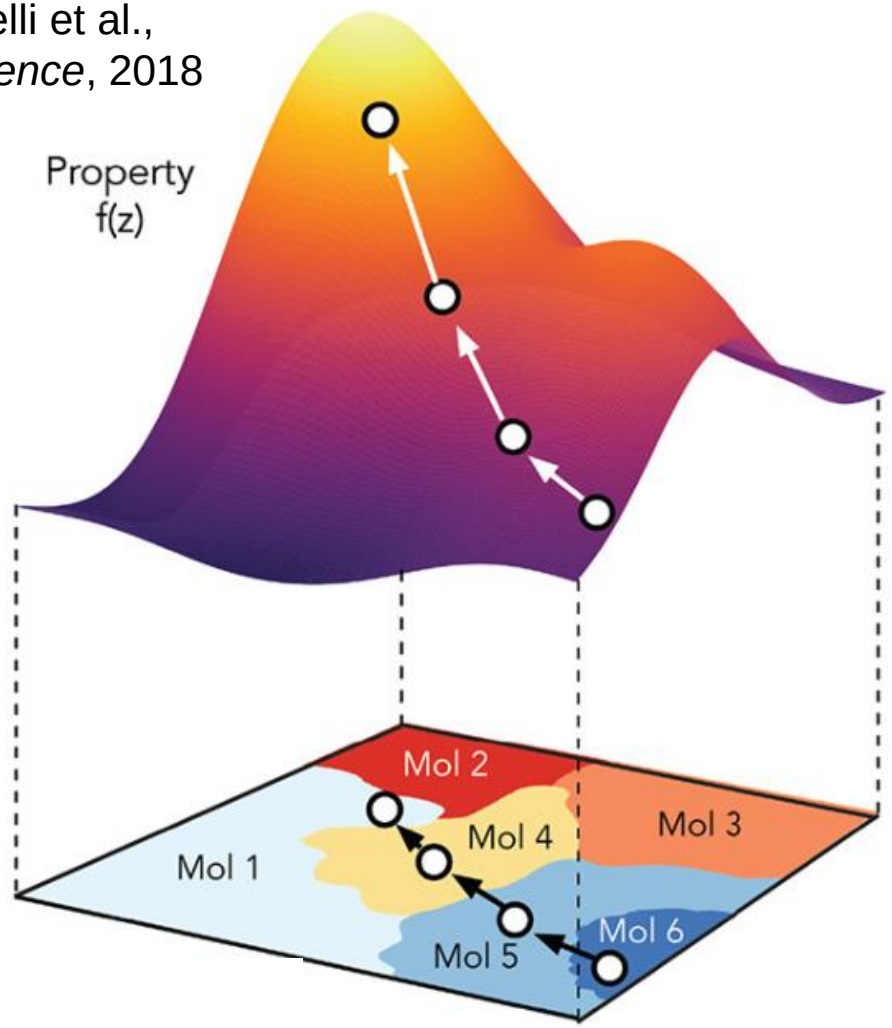
Gomez-Bombarelli et al.,
ACS Central Science, 2018



Enumerate large amounts of virtual compounds. Prioritize them using a predictive model for the property of interest.

How to steer the navigation towards useful chemistry?

Gomez-Bombarelli et al.,
ACS Central Science, 2018

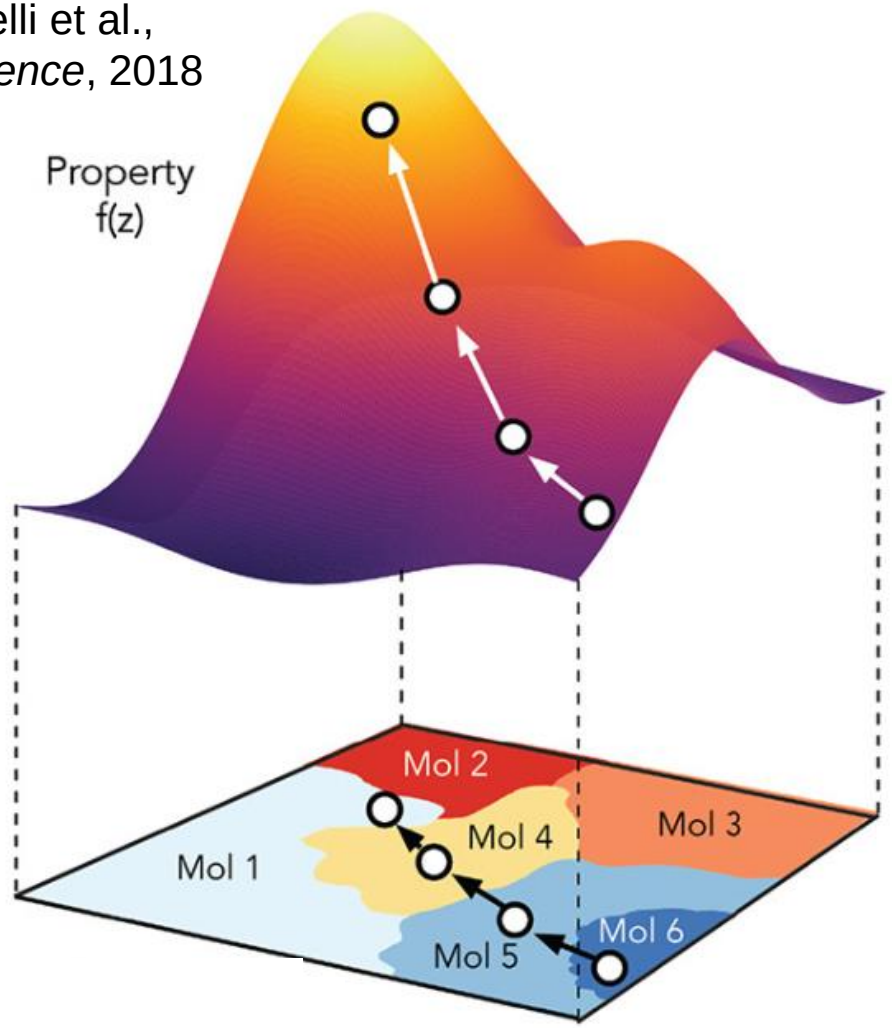


Enumerate large amounts of virtual compounds. Prioritize them using a predictive model for the property of interest.

Fine tune a pre-trained generative model to distort the generation towards desired properties.

How to steer the navigation towards useful chemistry?

Gomez-Bombarelli et al.,
ACS Central Science, 2018



Enumerate large amounts of virtual compounds. Prioritize them using a predictive model for the property of interest.

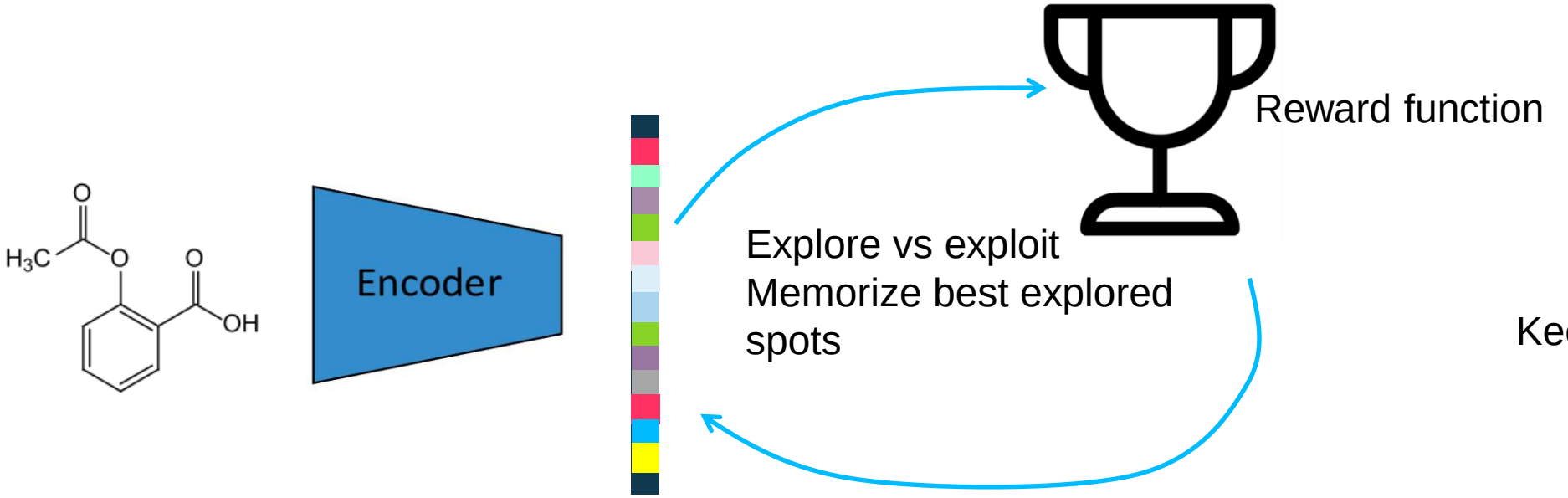
Fine tune a pre-trained generative model to distort the generation towards desired properties.

Use Reinforcement Learning to force the generative model to take decisions that will maximize its reward.

How to steer the navigation towards useful chemistry?

Our approach

- ✓ Do not retrain the autoencoder
- ✓ Do not depend on the particular set of reward functions
- ✓ Do not depend on enumeration of virtual compounds



Winter et al., *ChemRxiv*, 2019

Particle swarm optimization

0. Starting point: x (position in the 512-dimensional space)
Draw N random velocities to move N particles starting from x

Winter et al., *ChemRxiv*, 2019

Particle swarm optimization

0. Starting point: x (position in the 512-dimensional space)
Draw N random velocities to move N particles starting from x

1. Update the positions of the particles

$$x_i^{k+1} = x_i^k + v_i^{k+1}$$

Winter et al., *ChemRxiv*, 2019

Particle swarm optimization

0. Starting point: x (position in the 512-dimensional space)
Draw N random velocities to move N particles starting from x

1. Update the positions of the particles

$$x_i^{k+1} = x_i^k + v_i^{k+1}$$

2. Evaluate every particle with the reward function f

Winter et al., *ChemRxiv*, 2019

Particle swarm optimization

0. Starting point: x (position in the 512-dimensional space)
Draw N random velocities to move N particles starting from x

1. Update the positions of the particles

$$x_i^{k+1} = x_i^k + v_i^{k+1}$$

2. Evaluate every particle with the reward function f

3. Update the velocities using information from the previous steps and from the other N-1 particles

current position for particle i

$$v_i^{k+1} = \underbrace{w v_i^k}_{\text{inertia}} + c_1 r_1 \left(x_i^{\text{best}} - x_i^k \right) + c_2 r_2 \left(x^{\text{best}} - x_i^k \right)$$

Winter et al., *ChemRxiv*, 2019

Particle swarm optimization

0. Starting point: x (position in the 512-dimensional space)
Draw N random velocities to move N particles starting from x

1. Update the positions of the particles

$$x_i^{k+1} = x_i^k + v_i^{k+1}$$

2. Evaluate every particle with the reward function f

3. Update the velocities using information from the previous steps and from the other N-1 particles

$x_i^{\text{best}} = \text{argmax} f(x_i^k)$ best position explored by particle i

$x^{\text{best}} = \text{argmax} f(x_i^{\text{best}})$ overall best position ever explored

current position for particle i

$$v_i^{k+1} = \underbrace{w v_i^k}_{\text{inertia}} + c_1 r_1 \left(x_i^{\text{best}} - x_i^k \right) + c_2 r_2 \left(x^{\text{best}} - x_i^k \right)$$

Winter et al., *ChemRxiv*, 2019

Particle swarm optimization

0. Starting point: x (position in the 512-dimensional space)
Draw N random velocities to move N particles starting from x

1. Update the positions of the particles

$$x_i^{k+1} = x_i^k + v_i^{k+1}$$

2. Evaluate every particle with the reward function f

3. Update the velocities using information from the previous steps and from the other N-1 particles

$x_i^{\text{best}} = \text{argmax} f(x_i^k)$ best position explored by particle i

$x^{\text{best}} = \text{argmax} f(x_i^{\text{best}})$ overall best position ever explored

current position for particle i

$$v_i^{k+1} = \underbrace{w v_i^k}_{\text{inertia}} + c_1 r_1 \left(x_i^{\text{best}} - x_i^k\right) + c_2 r_2 \left(x^{\text{best}} - x_i^k\right)$$

Winter et al., *ChemRxiv*, 2019

Single parameter optimization

	ORGAN	JT-VAE	GCPN	MSO
reference	13	18	16	ours
penalized logP	3.63	5.30	7.98	**26.1**
QED	0.896	0.925	**0.948**	**0.948**
EGFR [pIC_{50}]	-	-	-	10.3
BACE1 [pIC_{50}]	-	-	-	11.5
run time	~1 d	~1 d	~8 h	~10 m

Winter et al., *ChemRxiv*, 2019

Single parameter optimization

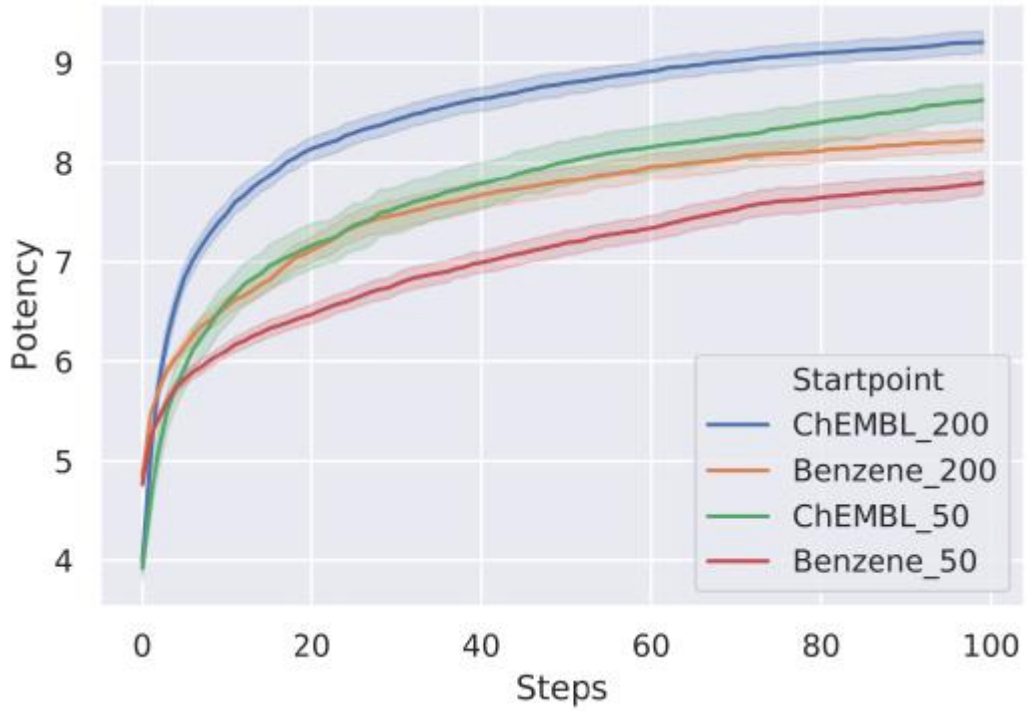
	ORGAN	JT-VAE	GCPN	MSO
reference	13	18	16	ours
penalized logP	3.63	5.30	7.98	**26.1**
QED	0.896	0.925	**0.948**	**0.948**
EGFR [pIC_{50}]	-	-	-	10.3
BACE1 [pIC_{50}]	-	-	-	11.5
run time	~1 d	~1 d	~8 h	~10 m

Winter et al., *ChemRxiv*, 2019

Single parameter optimization

	ORGAN	JT-VAE	GCPN	MSO
reference	13	18	16	ours
penalized logP	3.63	5.30	7.98	**26.1**
QED	0.896	0.925	**0.948**	**0.948**
EGFR [pIC_{50}]	-	-	-	10.3
BACE1 [pIC_{50}]	-	-	-	11.5
run time	~1 d	~1 d	~8 h	~10 m

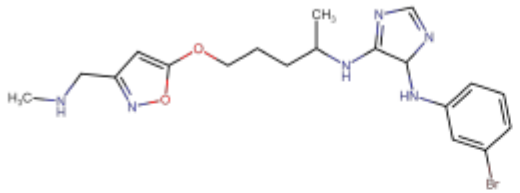
EGFR activity optimization

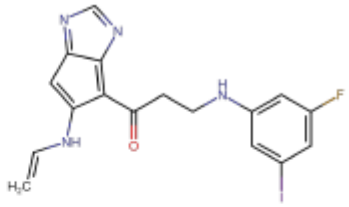


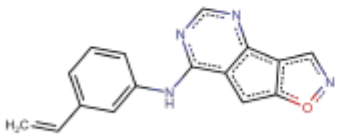
Winter et al., *ChemRxiv*, 2019

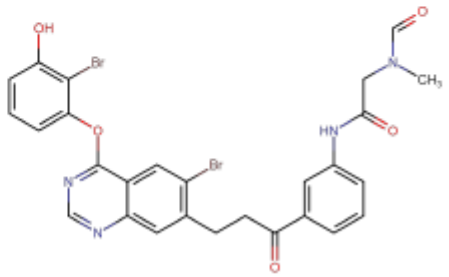
Single parameter optimization

	ORGAN	JT-VAE	GCPN	MSO
reference	13	18	16	ours
penalized logP	3.63	5.30	7.98	**26.1**
QED	0.896	0.925	**0.948**	**0.948**
EGFR [pIC_{50}]	-	-	-	10.3
BACE1 [pIC_{50}]	-	-	-	11.5
run time	~1 d	~1 d	~8 h	~10 m

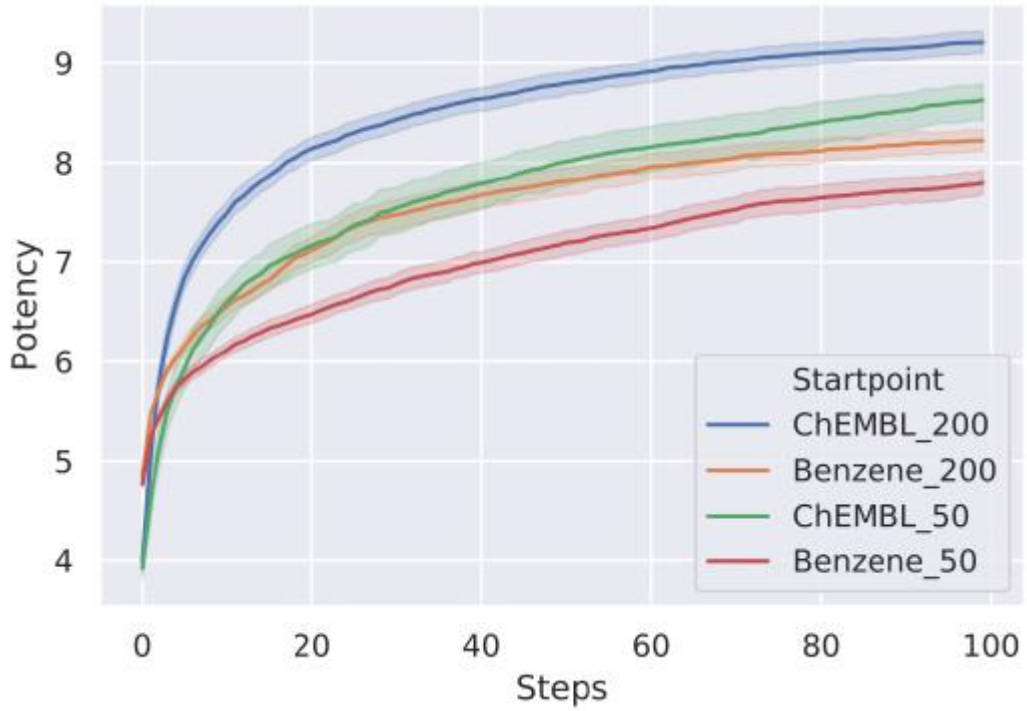








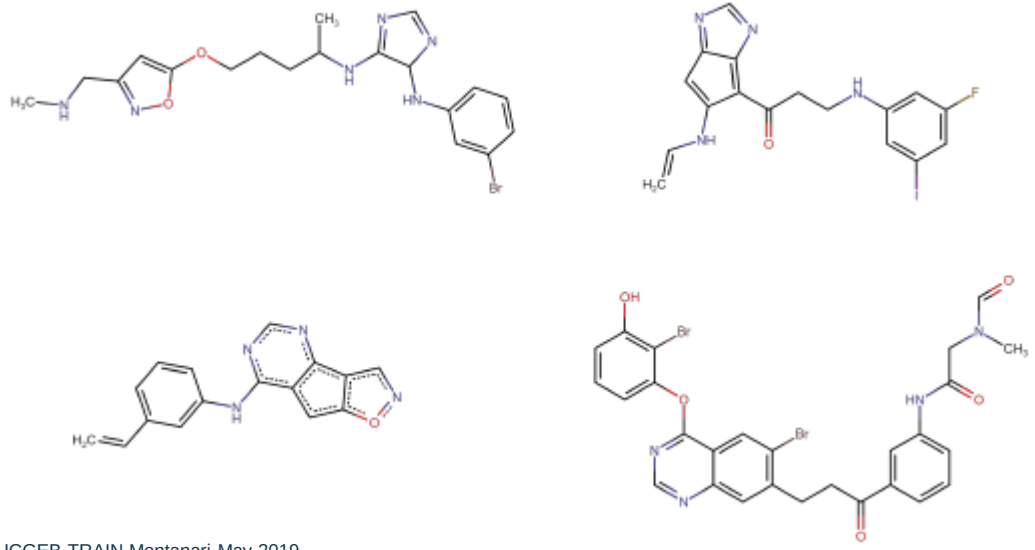
EGFR activity optimization



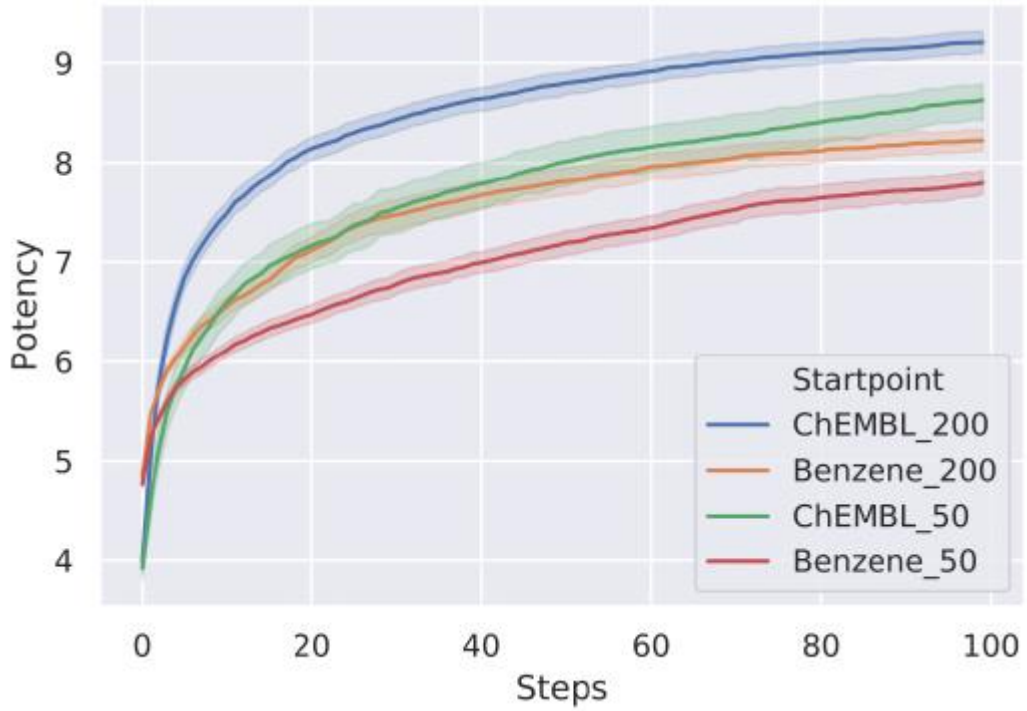
Winter et al., *ChemRxiv*, 2019

Single parameter optimization

	ORGAN	JT-VAE	GCPN	MSO
reference	13	18	16	ours
penalized logP	3.63	5.30	7.98	**26.1**
QED	0.896	0.925	**0.948**	**0.948**
EGFR [pIC_{50}]	-	-	-	10.3
BACE1 [pIC_{50}]	-	-	-	11.5
run time	~1 d	~1 d	~8 h	~10 m



EGFR activity optimization



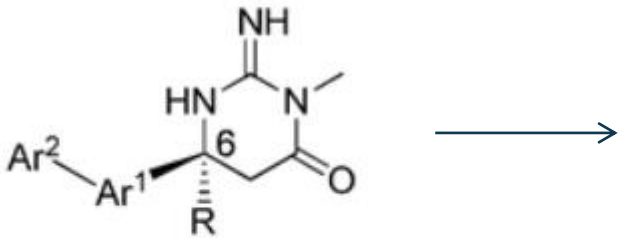
MSO method is able to optimize random starting points towards single objectives like predicted activity towards EGFR. The embedding does not need to be retrained. At this stage no control over the explored chemical space.

Winter et al., *ChemRxiv*, 2019

Restraining the chemical space

Discovery of an Orally Available, Brain Penetrant BACE1 Inhibitor That Affords Robust CNS Aβ Reduction

Andrew W. Stamford*†, Jack D. Scott†, Sarah W. Li†, Suresh Babu‡, Dawit Tadesse‡, Rachael Hunter‡, Yusheng Wu†, Jeffrey Misiaszek†, Jared N. Cumming†, Eric J. Gilbert†, Chunli Huang†, Brian A. McKittrick†, Liwu Hong†, Tao Guo‡, Zhaoning Zhu†, Corey Strickland#, Peter Orth#, Johannes H. Voigt#, Matthew E. Kennedy§, Xia Chen§, Reshma Kuvelkar§, Robert Hodgson§, Lynn A. Hyde§, Kathleen Cox±, Leonard Favreau±, Eric M. Parker§, and William J. Greenlee†



3

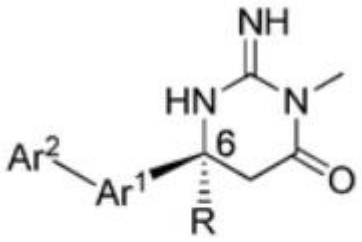
iminopyrimidinone scaffold

Stamford et al., *ACS MedChem Letters*, 2012

Restraining the chemical space

Discovery of an Orally Available, Brain Penetrant BACE1 Inhibitor That Affords Robust CNS Aβ Reduction

Andrew W. Stamford*†, Jack D. Scott†, Sarah W. Li†, Suresh Babu‡, Dawit Tadesse‡, Rachael Hunter‡, Yusheng Wu†, Jeffrey Misiaszek†, Jared N. Cumming†, Eric J. Gilbert†, Chunli Huang†, Brian A. McKittrick†, Liwu Hong†, Tao Guo‡, Zhaoning Zhu†, Corey Strickland#, Peter Orth#, Johannes H. Voigt#, Matthew E. Kennedy§, Xia Chen§, Reshma Kuvelkar§, Robert Hodgson§, Lynn A. Hyde§, Kathleen Cox±, Leonard Favreau±, Eric M. Parker§, and William J. Greenlee†

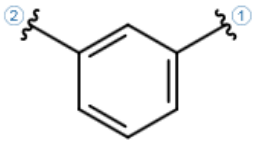


3

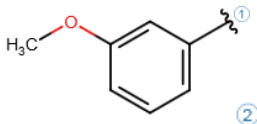
iminopyrimidinone scaffold

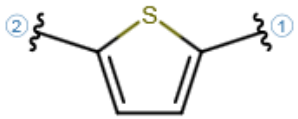
Enumerate possible R, Ar^1 and Ar^2 groups

Me



Phe



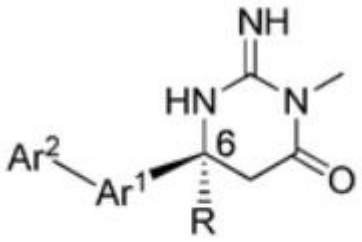


Stamford et al., *ACS MedChem Letters*, 2012

Restraining the chemical space

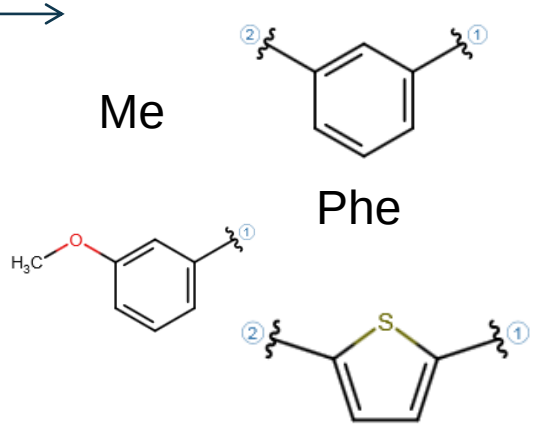
Discovery of an Orally Available, Brain Penetrant BACE1 Inhibitor That Affords Robust CNS Aβ Reduction

Andrew W. Stamford*†, Jack D. Scott†, Sarah W. Li†, Suresh Babu‡, Dawit Tadesse‡, Rachael Hunter‡, Yusheng Wu†, Jeffrey Misiaszek†, Jared N. Cumming†, Eric J. Gilbert†, Chunli Huang†, Brian A. McKittrick†, Liwu Hong†, Tao Guo‡, Zhaoning Zhu†, Corey Strickland#, Peter Orth#, Johannes H. Voigt#, Matthew E. Kennedy§, Xia Chen§, Reshma Kuvelkar§, Robert Hodgson§, Lynn A. Hyde§, Kathleen Cox±, Leonard Favreau±, Eric M. Parker§, and William J. Greenlee†



3
iminopyrimidinone scaffold

Enumerate possible R, Ar^1 and Ar^2 groups



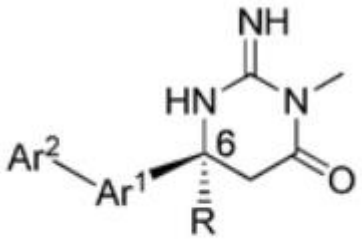
14 compounds made and tested

Stamford et al., *ACS MedChem Letters*, 2012

Restraining the chemical space

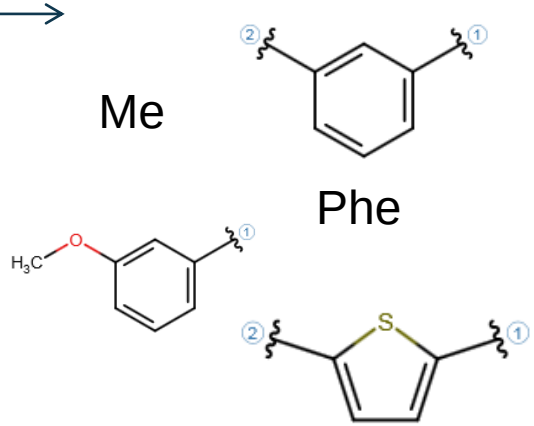
Discovery of an Orally Available, Brain Penetrant BACE1 Inhibitor That Affords Robust CNS Aβ Reduction

Andrew W. Stamford*†, Jack D. Scott†, Sarah W. Li†, Suresh Babu‡, Dawit Tadesse‡, Rachael Hunter‡, Yusheng Wu†, Jeffrey Misiaszek†, Jared N. Cumming†, Eric J. Gilbert†, Chunli Huang†, Brian A. McKittrick†, Liwu Hong†, Tao Guo‡, Zhaoning Zhu†, Corey Strickland#, Peter Orth#, Johannes H. Voigt#, Matthew E. Kennedy§, Xia Chen§, Reshma Kuvelkar§, Robert Hodgson§, Lynn A. Hyde§, Kathleen Cox±, Leonard Favreau±, Eric M. Parker§, and William J. Greenlee†

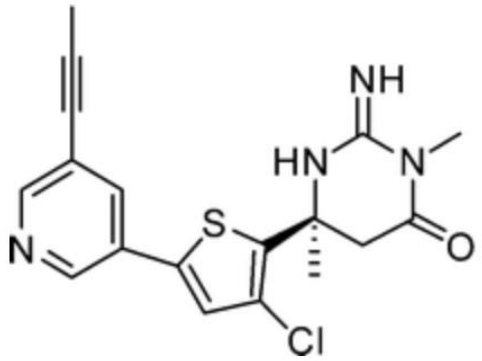


3
iminopyrimidinone scaffold

Enumerate possible R, Ar^1 and Ar^2 groups



14 compounds made and tested

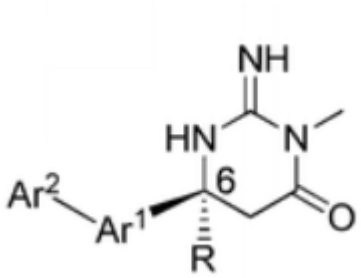


IC_{50}: 11 nM

Stamford et al., *ACS MedChem Letters*, 2012

Restraining the chemical space

Building the reward function:



This scaffold has to be present

SVM model to predict BACE1 activity

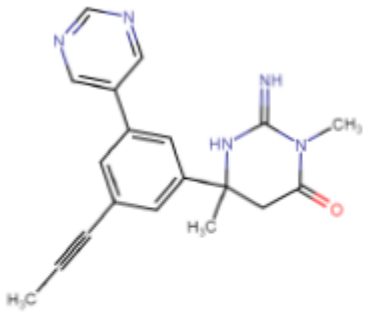
Training set: from ChEMBL, removing all compounds containing that scaffold

Chemistry health check-up:
- No more than 26 heavy atoms
- No toxic moiety
- No rare substructure

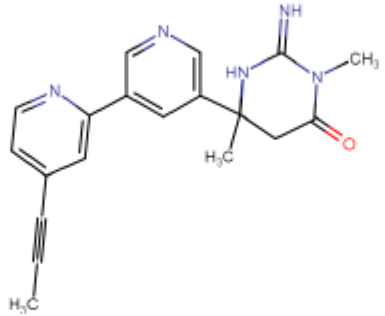
(i.e. not occurring in ChEMBL)

Winter et al., *ChemRxiv*, 2019

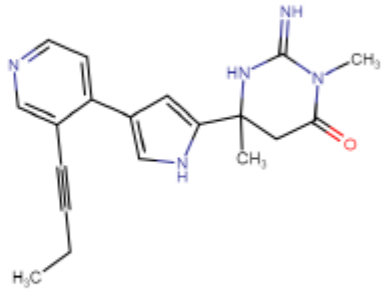
Restraining the chemical space



(d) 37 nM



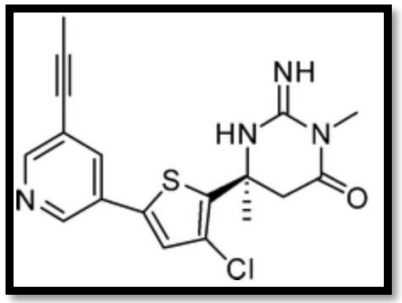
(e) 52 nM

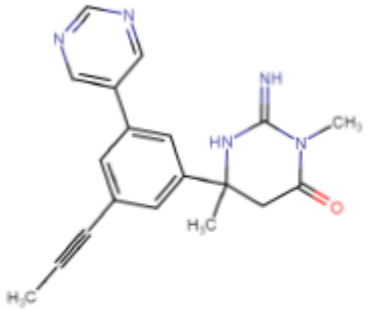


(f) 53 nM

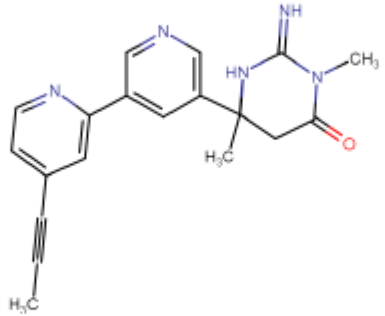
Winter et al., *ChemRxiv*, 2019

Restraining the chemical space

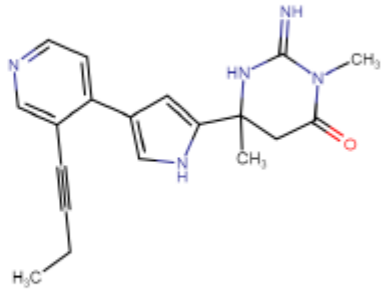




(d) 37 nM



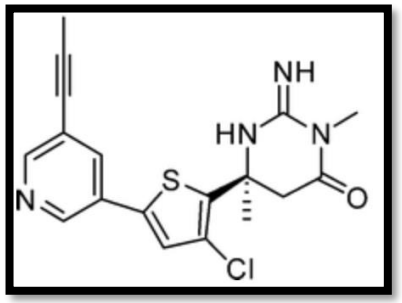
(e) 52 nM

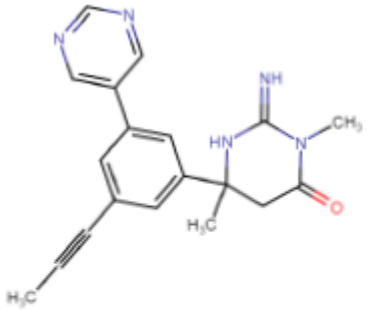


(f) 53 nM

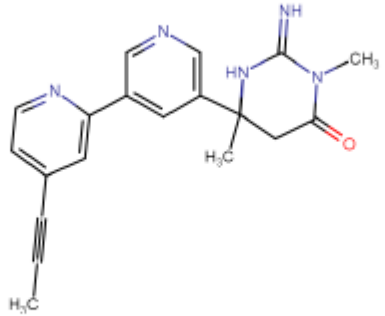
Winter et al., *ChemRxiv*, 2019

Restraining the chemical space

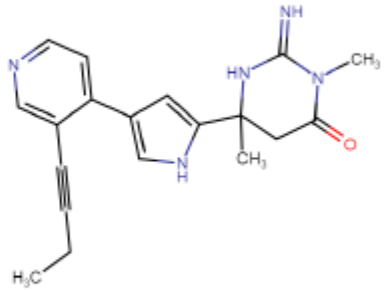




(d) 37 nM



(e) 52 nM



(f) 53 nM

MSO-optimized compounds are in the chemical vicinity of the Stamford et al. reported best compound. The exact compound was not found among the best particles because the BACE1 QSAR model gives it a worse prediction than the final candidates (170 nM). MSO is able to produce compounds in a given restricted chemical space.

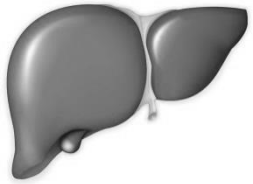
Winter et al., *ChemRxiv*, 2019

Multi-parameter optimization

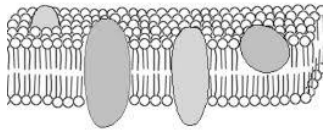
SVM model to predict BACE1 activity

SVM model to predict EGFR activity

Solubility



Metabolic stability



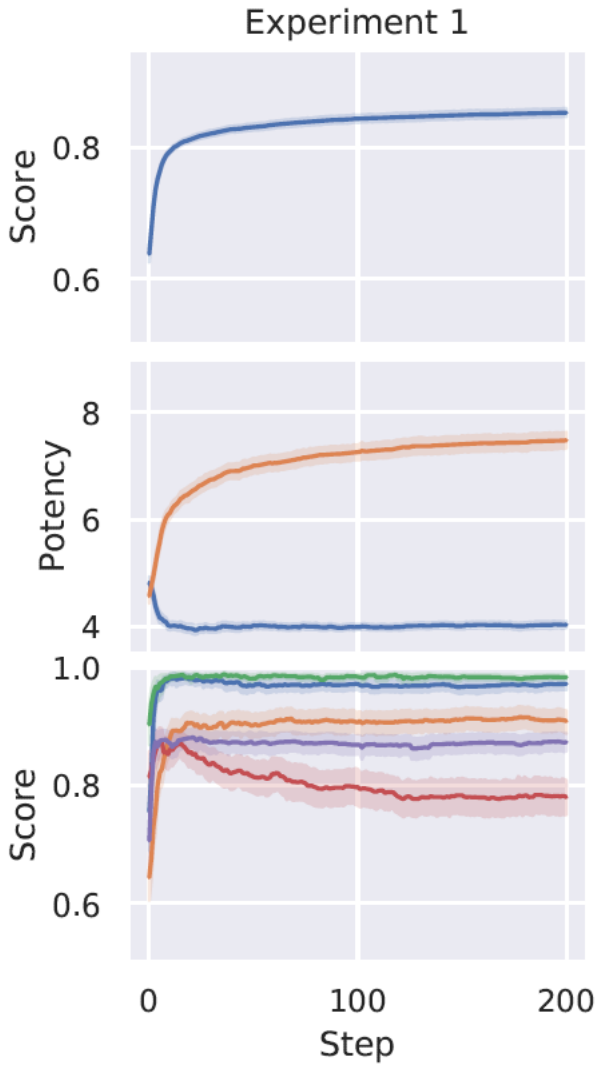
Cell permeability

Drug likeness
Synthetic accessibility
No toxic substructures
No rare substructures
200 < MW < 600

10 individual objectives to fulfill, scaled between 0 and 1. Different weights can be applied to the different objectives. Final reward function is a weighted average of the individual objectives.

Winter et al., *ChemRxiv*, 2019

Multi-parameter optimization



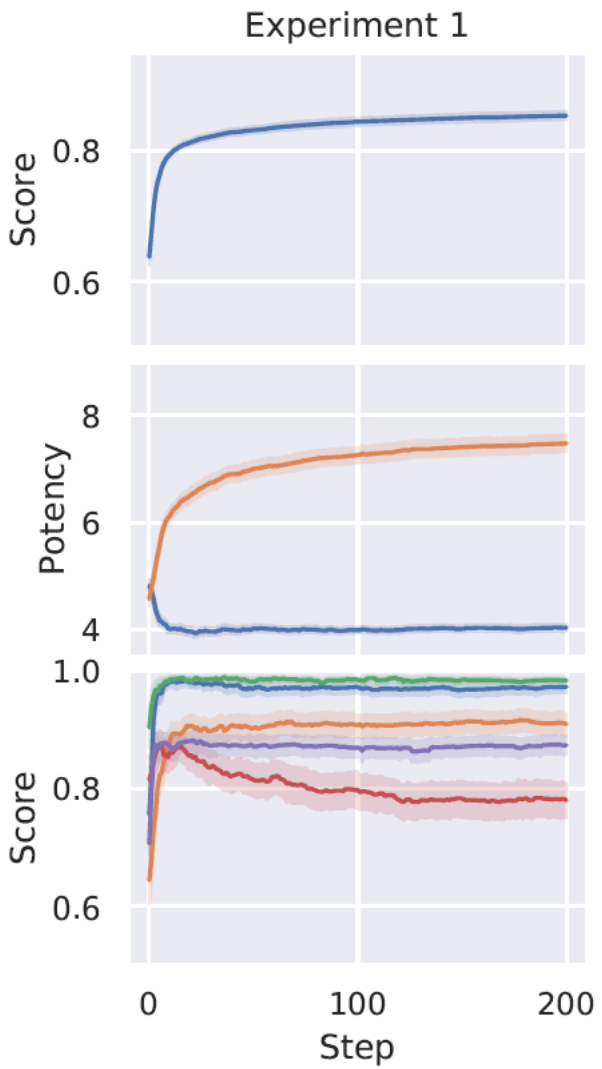
Overall score of the best particle

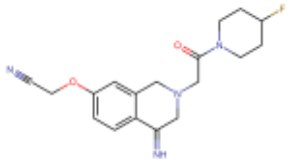
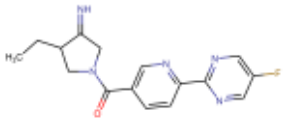
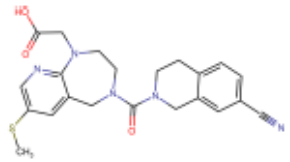
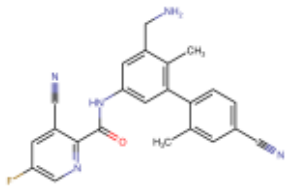
Orange: potency on BACE1
Blue: potency on EGFR

Green: permeability score
Blue: solubility score
Orange: metabolic stability score
Purple: drug likeness
Red: synthetic accessibility

Winter et al., *ChemRxiv*, 2019

Multi-parameter optimization



Compound	EGFR	BACE	stab	sol	perm	QED	SA
	3.7	8.0	86	390	72	0.90	3.0
	4.4	8.5	86	500	130	0.94	3.4
	4.1	8.3	78	570	90	0.73	3.0
	4.3	8.3	86	25	69	0.69	2.8

Winter et al., *ChemRxiv*, 2019

Wrap-up

It is possible to use the pre-trained autoencoder for compound optimization by combining reward functions with a Particle Swarm Optimization heuristic.

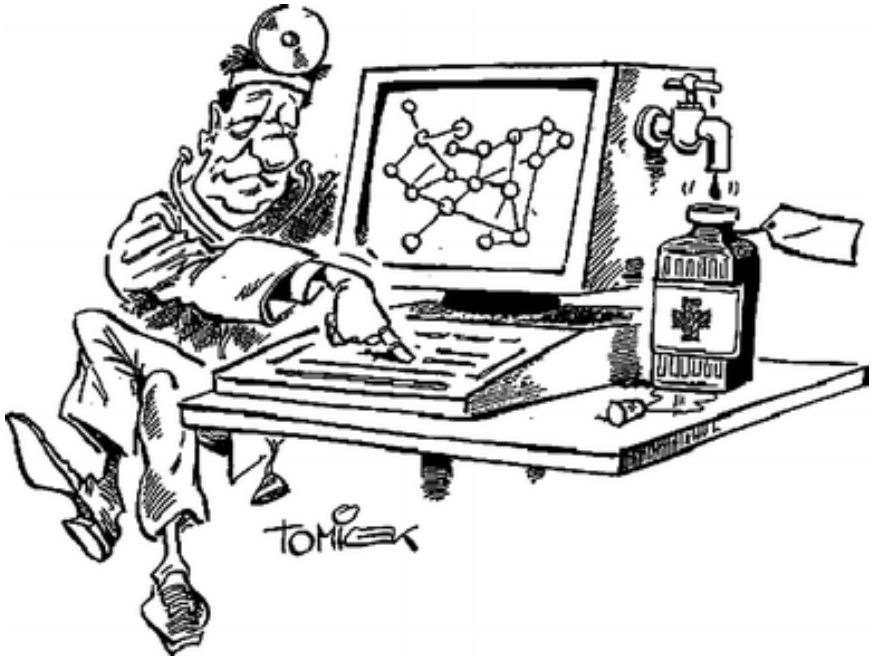
It is possible to find solutions for complex optimization problems with multiple, possibly contradicting objectives.

The method is fast and flexible: one does not need retraining of the autoencoder when the reward function changes.

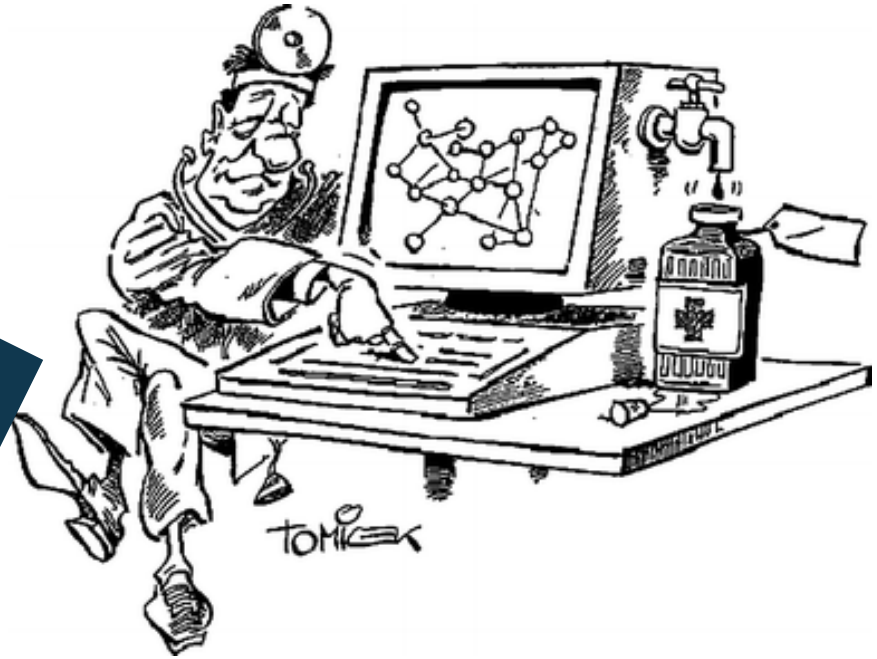
The tool relies on useful reward functions: one need to build strong QSAR models to steer compounds into a meaningful direction.

Article is submitted to Chemical Sciences but already available on ChemRxiv!
https://chemrxiv.org/articles/Efficient_Multi-Objective_Molecular_Optimization_in_a_Continuous_Latent_Space/7971101

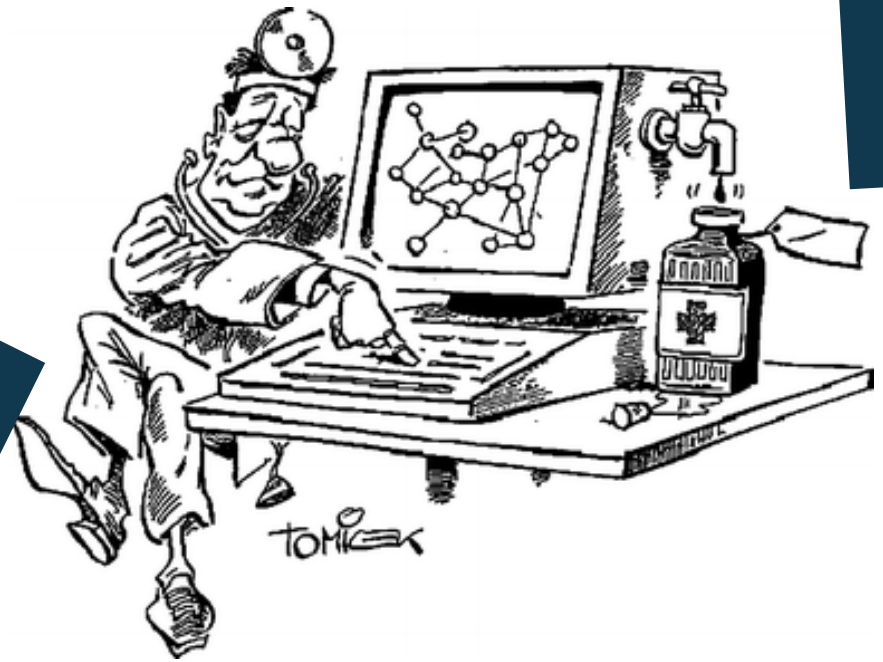
Conclusions


TOMICEK

Continuous and reversible representation of the chemical space



Better ADMET modeling using new Deep Learning approaches



Continuous and reversible representation of the chemical space

Better ADMET modeling using new Deep Learning approaches

Continuous and reversible representation of the chemical space

New tool to propose optimized solutions in drug discovery projects

Thank you!

Robin Winter

Djork-Arné Clevert
Lara Kuhnke
Antonius ter Laak

We are recruiting!!

Stellenbezeichnung: Research Scientist - Machine learning/Deep learning (m/f/d)

Research Scientist - Machine learning/Deep learning (m/f/d)

Bayer is looking for a highly creative and motivated Research Scientist with strong expertise in deep learning and computational life science to join the machine learning research team in R&D in Berlin, Germany. The position will be advertised within the EU-funded Innovative Medicines Initiative (IMI) project MELLODDY. With MELLODDY, ten leading European biopharmaceutical companies have come together to exchange their research data in a privacy preserving matter to improve the predictive performance of their machine learning models by federated learning. The position will involve research in direct collaboration with scientists from toxicology, medicinal chemistry, high-throughput image analysis, computer scientists, as well as leading European research groups in both academia and industry.

The successful applicant will be part of a cross-organizational team, applying deep learning within the R&D organization on our existing big data sets. She / he will contribute to the implementation of a deep learning platform providing impact on drug discovery and will be responsible for developing, testing and continually improving deep learning methods to predict properties of novel molecules (e.g. drug toxicity, assay bioactivity).

Where? Berlin
How long? 3 years
What? IMI project MELLODDY
When? July 2019

Drop us an email with your CV!
floriane.montanari@bayer.com
djork-arne.clevert@bayer.com

https://www.career.bayer.com/en/career/job-search/?fulltext=&accessLevel=&functional_area=&country=*&location=sap_lo_1013353&division=sap_di_PH